



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2022년10월19일

(11) 등록번호 10-2457147

(24) 등록일자 2022년10월17일

(51) 국제특허분류(Int. Cl.)  
C12Q 1/68 (2018.01) G01N 33/487 (2006.01)

(52) CPC특허분류  
C12Q 1/6869 (2018.05)  
G01N 33/48721 (2013.01)

(21) 출원번호 10-2017-7013136

(22) 출원일자(국제) 2015년10월19일

심사청구일자 2020년10월15일

(85) 번역문제출일자 2017년05월16일

(65) 공개번호 10-2017-0072915

(43) 공개일자 2017년06월27일

(86) 국제출원번호 PCT/GB2015/053097

(87) 국제공개번호 WO 2016/059436

국제공개일자 2016년04월21일

(30) 우선권주장

1418459.2 2014년10월17일 영국(GB)

(뒷면에 계속)

(56) 선행기술조사문헌

KR1020140050067 A

(73) 특허권자

옥스포드 나노포어 테크놀로지즈 피엘씨

영국, 옥스퍼드셔 오엑스4 4디큐, 옥스퍼드, 옥스퍼드 사이언스 파크, 에드먼드 홀리 로드, 고슬링 빌딩

(72) 발명자

제럴드, 다니엘 라이언

영국 옥스포드 오엑스4 4취에이 옥스포드 사이언스 파크 로버트 로빈슨 애비뉴 4 에드먼드 카트 라이트 하우스 옥스포드 나노포어 테크놀로지즈 리미티드

해론, 앤드류 존

영국 옥스포드 오엑스4 4취에이 옥스포드 사이언스 파크 로버트 로빈슨 애비뉴 4 에드먼드 카트 라이트 하우스 옥스포드 나노포어 테크놀로지즈 리미티드

(뒷면에 계속)

(74) 대리인

양영준, 김영

전체 청구항 수 : 총 13 항

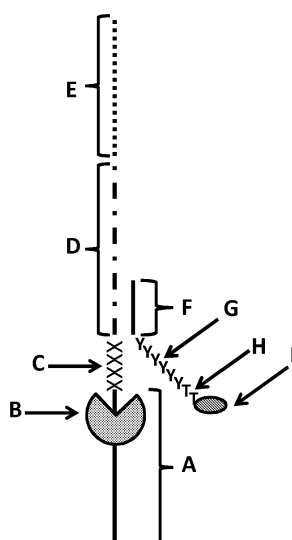
심사관 : 이준혁

(54) 발명의 명칭 나노 세공 RNA 특징 규명을 위한 방법

## (57) 요약

본 발명은 표적 RNA 폴리뉴클레오티드가 막횡단 세공에 대하여 이동함에 따라 하나 이상의 측정치를 취함으로써 표적 RNA 폴리뉴클레오티드의 특징을 규명하는 새로운 방법에 관한 것이다. 상기 이동은 DNA 헬리카제 효소에 의해 제어된다. 본 발명은 또한, RNA 폴리뉴클레오티드를, 이와 결합하는 DNA 헬리카제를 증가시키도록 변형시킨 변형된 RNA 구축물에 관한 것이다.

대표도 - 도3



(52) CPC특허분류

C12Q 2521/513 (2013.01)

C12Q 2525/121 (2013.01)

C12Q 2525/197 (2013.01)

C12Q 2565/631 (2013.01)

(72) 발명자

**자야징헤, 라크말**

영국 옥스포드 오엑스4 4쥐에이 옥스포드 사이언스  
파크 로베르트 로빈슨 애비뉴 4 에드먼드 카트라이  
트 하우스 옥스포드 나노포어 테크놀로지즈 리미티  
드

**터너, 다니엘 존**

영국 옥스포드 오엑스4 4쥐에이 옥스포드 사이언스  
파크 로베르트 로빈슨 애비뉴 4 에드먼드 카트라이  
트 하우스 옥스포드 나노포어 테크놀로지즈 리미티  
드

**화이트, 제임스**

영국 옥스포드 오엑스4 4쥐에이 옥스포드 사이언스  
파크 로베르트 로빈슨 애비뉴 4 에드먼드 카트라이  
트 하우스 옥스포드 나노포어 테크놀로지즈 리미티  
드

(30) 우선권주장

1508270.4 2015년05월14일 영국(GB)

1517634.0 2015년10월06일 영국(GB)

PCT/GB2015/052916 2015년10월06일 영국(GB)

## 명세서

### 청구범위

#### 청구항 1

- a) (i) DNA 헬리카제에 대한 결합 부위를 포함하는 DNA 리더 서열을 포함하도록 변형시킨 RNA 폴리뉴클레오티드 및 (ii) DNA 헬리카제 효소를 제공하는 단계;
- b) a)에서 제공된 RNA 폴리뉴클레오티드 및 DNA 헬리카제 효소를 막횡단 세공과 접촉시켜, DNA 헬리카제가 막횡단 세공을 통한 RNA 폴리뉴클레오티드의 이동을 제어하도록 하는 단계; 및
- c) RNA 폴리뉴클레오티드가 막횡단 세공에 대하여 이동함에 따라, RNA 폴리뉴클레오티드의 하나 이상의 특징을 나타내는 하나 이상의 측정치를 취하고, 여기서 하나 이상의 특징이 (i) RNA 폴리뉴클레오티드의 길이, (ii) RNA 폴리뉴클레오티드의 실체(identity), (iii) RNA 폴리뉴클레오티드의 서열, (iv) RNA 폴리뉴클레오티드의 2차 구조 및 (v) RNA 폴리뉴클레오티드가 변형되는 지의 여부로부터 선택되는 것이고, 이에 따라 표적 RNA 폴리뉴클레오티드의 특징을 규명하는 단계

를 포함하는, 표적 RNA 폴리뉴클레오티드의 특징을 규명하는 방법.

#### 청구항 2

제1항에 있어서, 리더 서열이 세공 내로 삽입되는 것인 방법.

#### 청구항 3

제1항에 있어서, (i) DNA 리더 서열이, 각각의 RNA 폴리뉴클레오티드와 DNA 리더 서열 상의 적어도 하나의 반응성 기 간에 형성된 공유 결합에 의해 RNA 폴리뉴클레오티드에 부착되는 것이거나; (ii) DNA 리더 서열이 화학적 또는 효소적 라이게이션에 의해 RNA 폴리뉴클레오티드에 라이게이션되는 것이거나; 또는 (iii) DNA 리더 서열이 RNA 폴리뉴클레오티드와 혼성화되는 것인 방법.

#### 청구항 4

제1항에 있어서, RNA 폴리뉴클레오티드의 하나 이상의 특징이 전기적 및/또는 광학적 측정에 의해 측정되는 것인 방법.

#### 청구항 5

제1항에 있어서, 단계 c)가, RNA 폴리뉴클레오티드가 막횡단 세공에 대하여 이동함에 따라 막횡단 세공을 관통하는 전류를 측정하고, 여기서 전류가 RNA 폴리뉴클레오티드의 하나 이상의 특징을 나타내며, 이로써 RNA 폴리뉴클레오티드의 특징을 규명하는 것을 포함하는 것인 방법.

#### 청구항 6

제1항에 있어서, RNA 폴리뉴클레오티드가 메틸화에 의한, 산화에 의한, 손상에 의한, 하나 이상의 단백질, 염기 유사체를 이용하거나 또는 하나 이상의 표지, 태그 또는 스페이서를 이용한 변형을 포함하는 것인 방법.

#### 청구항 7

제1항에 있어서, RNA 폴리뉴클레오티드가 하나 이상의 앵커를 사용하여 막에 커플링되는 것인 방법.

#### 청구항 8

제1항에 있어서, DNA 헬리카제가 폴리뉴클레오티드 결합성 도메인 내의 개구부의 크기를 감소시키는 변형을 포함하며, 이를 통하여 적어도 하나의 입체 형태적 상태에서는 RNA 폴리뉴클레오티드가 헬리카제로부터 결합 해제될 수 있는 것인 방법.

#### 청구항 9

제1항에 있어서, 하나 이상의 헬리카제가

(a) He1308 헬리카제, RecD 헬리카제, XPD 헬리카제 또는 Dda 헬리카제;

(b) 서열식별번호: 15 (Trwc Cba), 서열식별번호: 8 (He1308 Mbu) 또는 서열식별번호: 14 (Dda)를 포함하거나; 또는 (1) E94C 및 A360C 또는 (2) E94C, A360C, C109A 및 C136A를 포함하는 서열식별번호: 14 (Dda)의 변이체를 포함하는 헬리카제; 또는

(c) (a) 및/또는 (b)에서의 헬리카제 중 임의의 것의 조합물인 방법.

#### 청구항 10

제9항에 있어서, 하나 이상의 헬리카제가 (1) E94C 및 A360C 또는 (2) E94C, A360C, C109A 및 C136A를 포함하고, M1의 결실 및 G1의 부가를 추가로 포함하는 서열식별번호: 14 (Dda)의 변이체를 포함하는 것인 방법.

#### 청구항 11

제1항에 있어서, 헬리카제로부터 유래되고; 폴리뉴클레오티드와는 결합하지만 헬리카제로서 기능하지 않도록 변형되는 하나 이상의 분자 브레이크를 추가로 포함하는 방법.

#### 청구항 12

제1항에 있어서, 막횡단 세공이 단백질 세공이거나, 또는 막횡단 세공이 고체 상태 세공인 방법.

#### 청구항 13

제12항에 있어서, 막횡단 단백질 세공이 단백질 세공이고, 헤모리신, 류코시딘, 미코박테리움 스메그마티스 (*Mycobacterium smegmatis*) 포린 A (MspA), MspB, MspC, MspD, CsgG, 리세닌, 외막 포린 F (OmpF), 외막 포린 G (OmpG), 외막 포스포리파제 A, 네이세리아 (*Neisseria*) 자동 수송체 지질단백질 (NalP) 및 WZA로 구성된 그룹으로부터 선택된 어느 하나로부터 유래되는 것인 방법.

#### 청구항 14

삭제

#### 청구항 15

삭제

#### 청구항 16

삭제

#### 청구항 17

삭제

#### 청구항 18

삭제

#### 청구항 19

삭제

#### 청구항 20

삭제

#### 청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

**발명의 설명**

**기술 분야**

[0001] 본 발명은 표적 RNA 폴리뉴클레오티드가 막횡단 세공에 대하여 이동함에 따라 하나 이상의 측정치를 취함으로써 표적 RNA 폴리뉴클레오티드의 특징을 규명하는 새로운 방법에 관한 것이다. 표적 RNA 폴리뉴클레오티드가 막횡단 세공에 대하여 이동하는 것은 DNA 헬리카제 효소에 의해 제어되고, 표적 RNA 폴리뉴클레오티드는 이와 결합하는 DNA 헬리카제를 증가시키도록 변형된다. 본 발명은 또한, 표적 RNA 폴리뉴클레오티드를, 이와 결합하는 DNA 헬리카제를 증가시키도록 변형시킨 변형된 RNA 구축물에 관한 것이다.

**배경 기술**

[0002] 현재, 광범위한 응용 분야에 걸쳐 신속하고 저렴한 폴리뉴클레오티드 (예를 들어, DNA 또는 RNA) 서열 분석 및 확인 기술이 필요하다. 기존 기술은 주로, 다량의 폴리뉴클레오티드를 생성하기 위해 증폭 기술에 의존하고 시그널 검출을 위한 다량의 전문 형광 화학물질을 필요로 하기 때문에 느리고 비용이 많이 든다.

[0003] 막횡단 세공 (나노 세공)은 중합체 및 각종의 작은 분자에 대한 직접 전기 바이오센서(biosensor)로서 큰 잠재력을 가지고 있다. 특히, 최근에는 잠재적인 DNA 서열 분석 기술로서 나노 세공에 초점이 맞추어져 왔다.

[0004] 전위가 나노 세공에 걸쳐 인가될 때, 분석물, 예컨대 뉴클레오티드가 특정 기간 동안 배럴에 일시적으로 존재할 때 전류 흐름 상의 변화가 있다. 뉴클레오티드의 나노 세공 검출은 공지된 시그니처(signature)의 전류 변화와 지속 기간을 제공한다. 가닥 서열 분석 방법에서는, 단일 폴리뉴클레오티드 가닥이 상기 세공을 관통하고 뉴클레오티드의 실체가 도출된다. 가닥 서열 분석은 세공을 통한 폴리 뉴클레오티드의 이동을 제어하기 위한 폴리 뉴클레오티드 결합성 단백질의 사용을 포함할 수 있다.

[0005] 메신저 RNA는 유기체의 역동적인 상태에 대한 시각을 제공하며 건강 검진에서의 사용, 예를 들어 특정 암에서의 전이 진행 및 심장 질환을 포함한 직접적인 RNA 서열 분석의 혜택과 적용이 광범위하다. 직접적인 RNA 서열 분석은 또한 작물의 질환 저항성을 조사하여, 스트레스, 예를 들어 가뭄, 자외선 및 염도에 대한 반응을 결정하는데 있어서 적용되어 왔고, 배아 발생 동안의 세포 분화 및 결정에 적용되어 왔다.

[0006] RNA의 직접적인 서열 분석, 특히 500개 이상의 뉴클레오티드의 서열 분석에서 발생하는 문제는 막횡단 세공을 통한 RNA의 전위를 제어할 수 있는 적합한 분자 모터를 찾는 것이다. 지금까지, RNA와 맞물려서 일관된 이동을

제공하는 분자 모터는 밝혀지지 않았다. RNA 중합체의 일관된 이동과, 중합체의 긴 절편을 관측할 수 있는 능력이 폴리뉴클레오타이드의 특징을 규명하거나 또는 서열 분석하는 데 바람직하다.

[0007] 국제 특허 출원 번호 PCT/GB2014/053121 (WO 2015/056028)에는 상보적 폴리뉴클레오타이드를 형성한 다음, 막횡단 세공을 이용하여 이러한 상보적 폴리뉴클레오타이드의 특징을 규명하는 것을 포함하는, 표적 리보핵산 (RNA)의 특징을 규명하는 방법이 기재되어 있다. 이러한 간접적인 RNA 특징 규명은 오류가 나기 쉽고, 예를 들어 RNA의 메틸화 상태에 관한 중요한 정보의 상실을 초래할 수 있다. 다른 중요한 변형이 또한, RNA를 cDNA로 전환할 때 감춰질 수 있다.

### 발명의 내용

[0008] 본 발명자들은 놀랍게도, DNA 헬리카제 효소의 제어하에 표적 RNA 폴리뉴클레오타이드가 막횡단 세공에 대하여 이동함에 따라 하나 이상의 측정치를 취함으로써 표적 RNA 폴리뉴클레오타이드의 특징을 규명하는 것이 가능하다는 것을 입증하였다. 따라서, 한 실시양태에서, 세공을 통하여 서열 분석될 RNA의 능력 또는 효율을 증가시키기 위한 방법이 제공된다. 또 다른 실시양태에서, 변형되지 않은 형태의 RNA 보다 더 높은 효율로 세공을 통하여 서열 분석될 수 있는 변형된 RNA를 생성시키는 방법이 또한 제공된다. 따라서 본 발명은

[0009] a) (i) RNA 폴리뉴클레오타이드 (여기서, 이러한 RNA 폴리뉴클레오타이드는 비-RNA 폴리뉴클레오타이드를 포함하도록 변형된다) 및 (ii) DNA 헬리카제 효소를 제공하는 단계;

[0010] b) a)에서 제공된 RNA 폴리뉴클레오타이드 및 DNA 헬리카제 효소를 막횡단 세공과 접촉시켜, 이러한 DNA 헬리카제가 막횡단 세공을 통한 RNA 폴리뉴클레오타이드의 이동을 제어하도록 하는 단계; 및

[0011] c) 표적 RNA 폴리뉴클레오타이드가 막횡단 세공에 대하여 이동함에 따라 하나 이상의 측정치를 취함으로써 (여기서, 이러한 측정치는 RNA 폴리뉴클레오타이드의 하나 이상의 특징을 나타낸다), 표적 RNA 폴리뉴클레오타이드의 특징을 규명하는 단계

[0012] 를 포함하는, 표적 RNA 폴리뉴클레오타이드의 특징을 규명하는 방법을 제공한다.

[0013] 비-RNA 폴리뉴클레오타이드 (예컨대 폴리뉴클레오타이드 영역 또는 서열 또는 구조물)를 포함하도록 RNA 폴리뉴클레오타이드를 변형시키면, 이와 결합하는 DNA 헬리카제가 증가된다. 본원에 정의된 바와 같은 "비-RNA 폴리뉴클레오타이드"는 폴리뉴클레오타이드의 적어도 하나의 뉴클레오타이드가 리보뉴클레오타이드가 아닌, 즉 RNA로부터의 것이 아닌 폴리뉴클레오타이드이다. 따라서 비-RNA 폴리뉴클레오타이드는 적어도 하나의 리보뉴클레오타이드 (또는 RNA 뉴클레오타이드)를 포함할 수 있지만, 비-RNA 뉴클레오타이드 또는 서열, 즉 RNA가 아닌 뉴클레오타이드 또는 뉴클레오타이드의 서열을 부가적으로 포함해야만 한다. 본 발명의 바람직한 실시양태에서 비-RNA 폴리뉴클레오타이드 (이는 적어도 하나의 리보뉴클레오타이드 또는 RNA 뉴클레오타이드를 포함할 수 있거나 또는 포함하지 않을 수 있다)는 DNA 또는 DNA 유사체, 바람직하게 DNA 헬리카제 결합 부위 또는 DNA 어댑터를 포함한다. 바람직하게 비-RNA 폴리뉴클레오타이드는 세공 내로 우선적으로 삽입되는 리더 서열을 포함한다. 따라서 비-RNA 폴리뉴클레오타이드를 먼저 관측한 다음, 표적 RNA 서열의 특징을 규명해야 한다.

[0014] 본 발명은 (i) RNA 폴리뉴클레오타이드 및 (ii) 비-RNA 폴리뉴클레오타이드를 포함하는 "혼성체" 폴리뉴클레오타이드인 구조물을 제공한다. 바람직하게 비-RNA 폴리뉴클레오타이드는 DNA 폴리뉴클레오타이드를 포함하는데, 여기서 DNA 폴리뉴클레오타이드는 DNA 헬리카제 결합 부위를 포함하거나 또는 DNA 헬리카제 결합 부위만을 포함한다. 바람직하게 비-RNA 폴리뉴클레오타이드는 나노 세공 내로 우선적으로 삽입되는 리더 서열을 추가로 포함한다. 보다 바람직하게 비-RNA 폴리뉴클레오타이드는 폴리뉴클레오타이드 가닥 상의 바코딩 섹션(barcoding section)을 추가로 포함한다. 이러한 바코딩 섹션은 바람직하게 리더 서열과 DNA 헬리카제 결합 부위 사이에 위치한다. 바코딩 섹션은 분석물의 명백한 확인을 가능하게 하는데, 즉 여러 샘플 중 어느 것이 서열 분석되고 있는지를 사용자에게 알려준다.

[0015] 비-RNA 폴리뉴클레오타이드를 포함하도록 표적 RNA 폴리뉴클레오타이드를 변형시키는 것은 본원에 기재된 부착 방법 중 하나 이상을 포함한 임의의 적합한 부착 방법을 이용하여 비-RNA 폴리뉴클레오타이드 (가능하게는 RNA 서열 또는 적어도 하나의 리보뉴클레오타이드를 포함함)를 표적 RNA 폴리뉴클레오타이드에 부착시키는 것을 포함할 수 있다. 본원에 기재된 바와 같이, 비-RNA 폴리뉴클레오타이드를 표적 RNA에 부착시키는 것은 표적 RNA를 비-RNA 폴리뉴클레오타이드에 부착시키는 것과 동의어다. 비-RNA 폴리뉴클레오타이드가 리보뉴클레오타이드 또는 RNA 서열을 포함하는 경우, 비-RNA 폴리뉴클레오타이드는 이러한 비-RNA 폴리뉴클레오타이드 내에 포함되는 리보뉴클레오타이드 또는 RNA를 통하여 표적 RNA 폴리뉴클레오타이드에 부착될 수 있다.

- [0016] 비-RNA 폴리뉴클레오티드는, 각각의 표적 RNA 폴리뉴클레오티드와 비-RNA 폴리뉴클레오티드 상의 적어도 하나의 반응성 기 간에 형성된 공유 결합에 의해 표적 RNA 폴리뉴클레오티드에 부착될 수 있다. 비-RNA 폴리뉴클레오티드는 RNA 폴리뉴클레오티드에 화학적으로 또는 효소적으로 라이게이션될 수 있다. 비-RNA 폴리뉴클레오티드는 부가적으로 또는 또 다른 한편으로, 혼성화에 의해 및/또는 하나 이상의 토포이소머라제를 사용함으로써 RNA 폴리뉴클레오티드에 부착될 수 있다. 바람직하게 상기 방법에 의해 결정될 하나 이상의 특징은 (i) RNA 폴리뉴클레오티드의 길이, (ii) RNA 폴리뉴클레오티드의 실체, (iii) RNA 폴리뉴클레오티드의 서열, (iv) RNA 폴리뉴클레오티드의 2차 구조 및 (v) RNA 폴리뉴클레오티드가 변형되는 지의 여부로부터 선택된다. RNA 폴리뉴클레오티드의 하나 이상의 특징은 전기적 및/또는 광학적 측정에 의해 측정될 수 있다. 바람직하게 단계 c)는 RNA 폴리뉴클레오티드가 막형단 세공에 대하여 이동함에 따라 막형단 세공을 관통하는 전류를 측정함으로써 (여기서, 이러한 전류는 RNA 폴리뉴클레오티드의 하나 이상의 특징을 나타낸다), RNA 폴리뉴클레오티드의 특징을 규명하는 것을 포함한다.
- [0017] 그의 특징을 규명하고자 하는 표적 RNA 폴리뉴클레오티드는 메틸화에 의해, 산화에 의해, 손상에 의해, 하나 이상의 단백질을 이용하거나 또는 하나 이상의 표지, 태그 또는 스페이서를 이용하여 부가적으로 또는 추가로 변형시킬 수 있다. 표적 RNA는 염기 유사체를 함유할 수 있다. RNA 폴리뉴클레오티드는 하나 이상의 앵커를 이용하여 막과 커플링될 수 있다.
- [0018] 바람직하게 DNA 헬리카제는 폴리뉴클레오티드 결합성 도메인 내의 개구부의 크기를 감소시키는 변형을 포함하는데, 이를 통하여 적어도 하나의 입체 형태적 상태에서는 RNA 폴리뉴클레오티드가 헬리카제로부터 결합 해제될 수 있다. 본 발명의 한 실시양태에서, RNA 폴리뉴클레오티드의 이동은 일련의 하나 이상의 DNA 헬리카제에 의해 제어된다. 하나 이상의 헬리카제는 a) Hel308 헬리카제, RecD 헬리카제, XPD 헬리카제 또는 Dda 헬리카제; (b) (a)에서의 헬리카제 중 임의의 것으로부터 유래된 헬리카제; 또는 (c) (a) 및/또는 (b)에서의 헬리카제 중 임의의 것의 조합물이다. 본 방법은 헬리카제로부터 유래되고; 폴리뉴클레오티드와는 결합하지만 헬리카제로서 기능하지 않도록 변형되는 하나 이상의 분자 브레이크(brake)의 사용을 추가로 포함할 수 있다.
- [0019] 막형단 세공은 단백질 세공 또는 고체 상태 세공일 수 있다. 바람직하게 막형단 단백질 세공은 단백질 세공이고, 이는 헤모리신, 류코시딘, 미코박테리움 스메그마티스 (*Mycobacterium smegmatis*) 포린 A (MspA), MspB, MspC, MspD, 리세닌, CsgG, 외막 포린 F (OmpF), 외막 포린 G (OmpG), 외막 포스포리파제 A, 네이세리아 (*Neisseria*) 자동 수송체 지질단백질 (NalP) 및 WZA로부터 유래된다.
- [0020] 본 발명은 또한,
- [0021] a) (i) RNA 폴리뉴클레오티드 (여기서, 이러한 RNA 폴리뉴클레오티드는 비-RNA 폴리뉴클레오티드를 포함하도록 변형된다) 및 (ii) DNA 헬리카제 효소를 제공하는 단계; 및
- [0022] b) a)에서 제공된 RNA 폴리뉴클레오티드 및 DNA 헬리카제 효소를 막형단 세공과 접촉시켜, 이러한 DNA 헬리카제가 막형단 세공에 대한 RNA 폴리뉴클레오티드의 이동을 제어하도록 하는 단계
- [0023] 를 포함하는, 표적 RNA 폴리뉴클레오티드의 이동이 DNA 헬리카제 효소에 의해 제어될 때, 표적 RNA 폴리뉴클레오티드를 막형단 세공에 대하여 이동시키는 방법을 제공한다.
- [0024] 비-RNA 폴리뉴클레오티드를 포함하도록 RNA 폴리뉴클레오티드를 변형시키면, 이와 결합하는 DNA 헬리카제가 증가된다. 본 발명의 한 실시양태에서 상기 방법은 접촉 단계 전에, DNA 헬리카제를 상기 변형된 RNA 폴리뉴클레오티드와 미리 결합시키는 것을 포함한다.
- [0025] 본 발명의 방법은 막형단 세공에 대한 RNA 폴리뉴클레오티드의 보다 일관된 이동을 제공한다. 본 발명은 또한 (i) RNA 폴리뉴클레오티드와 (ii) 비-RNA 폴리뉴클레오티드를 포함하는 "혼성체" 폴리뉴클레오티드를 제공한다. 바람직하게 비-RNA 폴리뉴클레오티드는 DNA 폴리뉴클레오티드를 포함하는데, 여기서 DNA 폴리뉴클레오티드는 DNA 헬리카제 결합 부위를 포함하거나 또는 DNA 헬리카제 결합 부위만을 포함한다. 바람직하게 비-RNA 폴리뉴클레오티드는 나노 세공 내로 우선적으로 삽입되는 리더 서열을 추가로 포함한다. 보다 바람직하게 비-RNA 폴리뉴클레오티드는 폴리뉴클레오티드 가닥 상의 바코딩 섹션을 추가로 포함한다. 이러한 바코딩 섹션은 바람직하게 리더 서열과 DNA 헬리카제 결합 부위 사이에 위치한다. 바코딩 섹션은 분석물의 명백한 확인을 가능하게 하는데, 즉 여러 샘플 중 어느 것이 서열 분석되고 있는지를 사용자에게 알려준다.
- [0026] 본 발명은 또한, 표적 RNA 폴리뉴클레오티드와 DNA 헬리카제의 조합물을 제공하는데, 여기서 RNA 폴리뉴클레오티드의 일부가 DNA 헬리카제와 상호 작용하거나 또는 결합하도록 변형되었다. 바람직하게 RNA는 비-RNA 폴리뉴



클레오티드를 포함하도록 변형되고, 가장 바람직하게 비-RNA 폴리뉴클레오티드는 DNA 또는 DNA 유사체이다.

[0027] 본 발명은 또한, 표적 RNA 폴리뉴클레오티드의 특징을 규명하기 위한 키트를 제공한다. 바람직하게 이러한 키트는 그의 특징을 규명하고자 하는 표적 RNA에 부착되도록 개조시킨 비-RNA 폴리뉴클레오티드를 포함한다. 본 발명은 또한 샘플 중의 표적 RNA 폴리뉴클레오티드의 특징을 규명하기 위한 장치를 제공한다.

[0028] 본원에 기재된 변형된 RNA 구축물은 DNA 헬리카제 효소에 대한 결합 부위를 제공한다. DNA 헬리카제는 본질적으로 "속임수"를 당하여 RNA 폴리뉴클레오티드를 판독하는데: 일단 DNA 헬리카제가 비-RNA 폴리뉴클레오티드와 결합하게 되면, 이는 RNA 폴리뉴클레오티드를 따라 전이될 수 있다. 헬리카제는 용액 중의 RNA를 따라 전이될 수 있다. 헬리카제가 RNA 폴리뉴클레오티드를 따라 이동하는 것을 촉진시키기 위해서는 나노 세공의 존재를 필요로 할 수 있다. 헬리카제가 RNA를 따라 이동하는 것은 나노 세공에 의해 촉진될 수 있고, 인가된 장을 나노 세공에 걸쳐 인가하는 것을 통하여 촉진될 수 있다.

[0029] DNA 헬리카제의 제어하에 나노 세공을 통하여 전체 RNA 구축물을 전위시킬 수 있는 능력은 공지된 방법에 비해 개선된 정확도 및 속도로 RNA 폴리뉴클레오티드의 특징, 예컨대 그의 서열을 추정할 수 있게 해준다: 본 발명의 방법은 선행 기술분야의 방법과 비교해서 PCR 바이어스(bias)가 없다. RNA에서의 변형이 검출될 수 있고, 비-코딩 이소형 및 스플라이스-변이체가 정확하게 확인될 수 있다. RNA 염기 유사체가 나노 세공에서 직접적으로 검출될 수 있다. 본원에 기재된 구축물은 500개 이상의 뉴클레오티드, 예를 들어 1,000개, 5,000개, 10,000개, 20,000개, 50,000개, 100,000개 또는 그 초과 뉴클레오티드의 표적 RNA의 전위를 달성하는 데 특히 유효하다.

[0030] 본 발명의 방법에서 DNA 헬리카제는 본질적으로 "속임수"를 당하여, 비-RNA 리더 서열의 존재를 통해서 표적 RNA 서열을 판독한다. 일단 DNA 헬리카제의 이동이 비-RNA 폴리뉴클레오티드 (이는 DNA 또는 DNA 유사체를 포함할 수 있다)에 의해 개시되면, 이는 RNA를 따라 지속적으로 이동할 수 있다. 본 발명의 방법은 또한, RNA 서열과 DNA 서열이 동일한 경우일지라도, 평균 진폭 및 범위의 함수로서 RNA와 DNA를 서로 차별하기 위한 수단을 제공한다.

### 도면의 간단한 설명

[0031] 도 1은 진핵 RNA의 가닥 (파선으로서 제시됨)을 DNA의 가닥 (실선으로서 제시됨)에 부착시키는 방법을 나타내는 만화를 도시한 것이다. 진핵 RNA는 5' 말단에 7-메틸구아노신 캡을 갖는다 (별 모양으로서 제시되고 A로 표시됨). 반응 단계 1은 7-메틸구아노신 캡을 제자리에 두고 진핵 RNA의 5' 말단에 반응성 기 (B로 표시되고 정사각형으로서 제시됨)를 부가한다. 반응성 기는 하이퍼메틸라제 효소, 예를 들어 변형된 S아데노실 메티아민을 수반한 트리메틸 구아노신 신타제를 사용하여 부가된다. 단계 2는 부착된 C로 표시되고 원형으로서 제시된 반응성 기를 또한 갖는 DNA 가닥을 다른 반응성 기 (B)와 반응시켜 공유 결합을 형성시키는 화학적 반응 단계를 나타낸다. 단계 3은 담배 산 포스파타제를 사용하여 7-메틸구아노신 캡을 제거하는데, 이로써 D로 표시된 RNA 가닥이 생성된다. 이어서, RNA (7-메틸구아노신 캡이 제거되고, D로 표시됨)를 수많은 상이한 방식으로 처리하여 DNA의 가닥에 부착된 RNA의 가닥을 생성시키는 것이 가능하다 (단계 4, 6, 7, 8 또는 11). 단계 4에서는 반응성 기 (B로 표시되고 사각형으로서 제시됨)를 RNA 가닥 D의 5' 말단에 부가한다. 단계 5는 부착된 C로 표시되고 원형으로서 제시된 반응성 기를 또한 갖는 DNA 가닥을 다른 반응성 기 (B)와 반응시켜 공유 결합을 형성시키는 화학적 반응 단계를 나타낸다. 단계 6 및 7은 DNA 가닥 (E로 표시됨)을, 예를 들어 T4 RNA 폴리머라제 1, T4 RNA 폴리머라제 2, 열안정성 5' App DNA/RNA 리가제 등을 사용하여 RNA 상으로 직접 라이게이션할 수 있다는 것을 나타낸다. 단계 7에서는 효소 (G로 표시됨)를 DNA와 미리 결합시키는 반면, 단계 6에서는 어떠한 효소도 미리 결합시키지 않는다. 단계 8 및 11은 리더를 수반한 DNA 프라이머 (F로 표시됨)를 RNA 가닥 D와 혼성화시키는 것을 나타낸다. 단계 11에서는 효소 (G로 표시됨)를 DNA 프라이머와 미리 결합시키는 반면, 단계 8에서는 어떠한 효소도 미리 결합시키지 않는다. 단계 9 및 12는 RNA 가닥 D의 역전사를 나타내는데, 이로써 3개의 C의 3' 오버행이 초래된다. 단계 10 및 13은 DNA 헤어핀 (H로 표시됨)을 이중 가닥 DNA/RNA (I로 표시됨)에 라이게이션하는 것을 나타낸다.

도 2는 원핵 RNA의 가닥 (파선으로서 제시됨)을 DNA의 가닥 (실선으로서 제시됨)에 부착시키는 방법을 나타내는 만화를 도시한 것이다. 반응 단계 1은, 예를 들어 폴리A 폴리머라제 효소를 사용하여 폴리(dA) 테일(tail) (A로 표시됨)을 RNA 가닥의 3' 말단 상으로 부가하는데, 이로써 가닥 B가 생성된다. 단계 2 및 5는 리더를 수반한 DNA 프라이머 (C로 표시됨)를 폴리(dA) 영역 (A로 표시됨)과 혼성화시키는 것을 나타낸다. 단계 5에서는 효소 (D로 표시됨)를 DNA 프라이머와 미리 결합시키는 반면, 단계 2에서는 어떠한 효소도 미리 결합시키지 않는다. 단계 3 및 6은 RNA 가닥 B의 역전사를 나타내는데, 이로써 1개 내지 3개의 C의 3' 오버행이 초래된다. 단계 4 및 7은 DNA 헤어핀 (E로 표시됨)을 이중 가닥 DNA/RNA (F로 표시됨)에 라이게이션하는 것을 나타낸다.



반응 단계 8, 10 및 11은 폴리(dA) 영역을 RNA에 부가함으로써 수행되기보다는 오히려 원핵 RNA 상에서 직접적으로 일어난다. 단계 8에서는 반응성 기 (G로 표지되고 사각형으로서 제시됨)를 RNA 가닥의 5' 말단에 부가한다. 단계 9는 부착된 반응성 기 (H로 표지되고; 반응성 기는 원형으로서 제시됨)를 또한 갖는 DNA 가닥을 다른 반응성 기 (G)와 반응시켜 공유 결합을 형성시키는 화학적 반응 단계를 나타낸다. 단계 10 및 11은 DNA 가닥 (I로 표지됨)을, 예를 들어 T4 RNA 폴리머라제 1, T4 RNA 폴리머라제 2, 열안정성 5' App DNA/RNA 리가제 등을 사용하여 RNA 상으로 직접 라이게이션할 수 있다는 것을 나타낸다. 단계 11에서는 효소 (D로 표지됨)를 DNA와 미리 결합시키는 반면, 단계 10에서는 어떠한 효소도 미리 결합시키지 않는다.

도 3은 실시예 1에서 DNA 헬리카제 [T4 Dda - E94C/A360C (돌연변이 E94C/A360C에 이어서 ( $\Delta$ M1)G1을 수반한 서열식별번호 (SEQ ID NO): 14)]를 사용하여 MspA 나노 세공을 통하여 전위시킨 DNA/RNA 가닥을 나타내는 만화를 도시한 것이다. 영역 A는 DNA 헬리카제 (B로 표지됨)와 결합하는 40개의 뉴클레오타이드 폴리(dT) 리더 (서열식별번호: 16)에 상응한다. 영역 A는 4개의 iSpC3 스페이서 (X로서 제시되고 C로 표지됨)에 부착된다. 영역 D는 합성 RNA 영역 (서열식별번호: 17)에 상응한다. 영역 E는 가변 길이 폴리(U)에 상응한다. 영역 F는 영역 D와 혼성화되는 DNA (서열식별번호: 18)이다. 이러한 DNA (서열식별번호: 18)에 6개의 iSp18 스페이서 (Y로서 제시되고 G로 표지됨), 2개의 티민 (T로서 제시되고 H로 표지됨) 및 3' 콜레스테롤 TEG (I로 표지됨)가 부착된다.

도 4는 RNA 폴리(U) 폴리머라제 연장 단계 (실시예 1 단계 1.1) 후 5% PAGE TBE 바이오래드(BioRad) 기준 겔 (140 V에서 40분 동안 실행됨)을 도시한 것이다. 레인 1은 100 bp 트리다이 래더(TriDye ladder)를 나타낸다. 레인 2는 중합 전의 합성 DNA/RNA 1 (서열식별번호: 17의 5' 말단에 반대 말단에서 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 16)을 나타낸다. 레인 3은 폴리(U) 중합 혼합물을 나타낸다. 레인 4는 폴리(U) 연장된 DNA/RNA 2를 함유하는 정제된 샘플 1을 나타낸다. 화살표 X는 비-연장된 DNA/RNA 1에 상응하고, 화살표 Y는 폴리(U) 연장된 DNA/RNA 2에 상응한다.

도 5는 DNA 헬리카제 [T4 Dda - E94C/A360C (돌연변이 E94C/A360C에 이어서 ( $\Delta$ M1)G1을 수반한 서열식별번호: 14)]가 DNA/RNA 2 (도 3에 도시된 만화)의 이동을 제어한 경우의 헬리카제 제어된 DNA 이동의 추적 예 [y-축 = 전류 (pA), x-축 = 시간 (s)]를 도시한 것이다. 영역 1은 폴리(dT) 리더 (서열식별번호: 16)에 상응하고, 영역 2는 iSpC3 스페이서 (이러한 스페이서는 더 많은 양의 전류가 DNA 또는 RNA 영역보다는 나노 세공을 통하여 흐를 수 있게 하였다)에 상응하며, 영역 3은 RNA 서열 (서열식별번호: 17)에 상응하고, 영역 4는 실시예 1 단계 1.1에서 부가된 가변 길이 폴리(U) RNA 영역에 상응한다.

도 6은 토포이소머라제를 사용하여 dsDNA를 RNA의 가닥 상으로 부착시킬 수 있는 방법을 도시한 것이다. 본 만화 도면에서는, 그의 5' 말단에 유리 히드록실을 갖는 RNA의 가닥에 dsDNA가 부착된다 (진핵 RNA의 5' 말단에 존재하는 7-메틸구아노신 캡 및 5' 포스페이트는, 예를 들어 포스파타제, 예컨대 남극(antarctic) 포스파타제 또는 알칼리성 포스파타제를 사용하여 제거시킬 필요가 있을 것이다). 백시니아 토포이소머라제는 제시된 서열에서 dsDNA 상으로 결합한다 (단계 1). 적색 화살표는 두 번째 티민 반대쪽의 dsDNA의 하부 가닥에 닉(nick)이 있다는 것을 강조 표시한다. 일단 토포이소머라제가 결합되면, 이는 두 번째 티민 다음의 DNA의 상부 가닥을 커팅하고, dsDNA와 결합된 채로 있다 (단계 2). 이어서, dsDNA와 결합되는 토포이소머라제를, 유리 5' 히드록실을 갖는 RNA와 함께 인큐베이션하였다 (단계 3). 이때 토포이소머라제는 상기 RNA를 dsDNA와 연결시켜 준다.

도 7은 토포이소머라제를 사용하여 dsDNA를 RNA의 가닥 상으로 부착시킬 수 있는 또 다른 방법을 도시한 것이다. 본 만화 도면에서는, RNA와 혼성화하고 RNA 부착에 도움을 주는 DNA의 단일 가닥 영역을 이용하여 dsDNA를 RNA의 가닥에 부착시킨다. 백시니아 토포이소머라제가 제시된 서열에서 dsDNA 상으로 결합된다 (단계 1). DNA의 하부 가닥은 그 내부에 닉을 갖지 않는다. 일단 토포이소머라제가 결합되면, 이는 두 번째 티민 다음의, DNA의 상부 가닥만을 커팅하고, dsDNA/ssDNA와 결합된 채로 있다 (단계 2). 이어서, dsDNA와 결합되는 토포이소머라제를, 유리 5' 히드록실을 갖는 RNA와 함께 인큐베이션하였다 (단계 3). 이때 토포이소머라제는 상기 RNA를 dsDNA와 연결시켜 준다. ssDNA 영역은 상보적 RNA 서열을 DNA에 대한 부착 지점으로 유인하는 데 도움을 준다.

도 8은 실시예 2로부터의 샘플을 포함하는, 140 V에서 60분 동안 실행된 10% PAGE TBE-우레아 변성 기준 겔을 도시한 것이다. 레인 1은 100 bp 트리다이 래더를 나타낸다. 레인 2는 T4 DNA 리가제의 부재하에 실시예 2에 기재된 라이게이션 반응을 진행시킨 후의 DNA 올리고 1 (1X 농도, 서열식별번호: 21), RNA 올리고 1 (1X 농도, 서열식별번호: 19) 및 DNA 스플린트 (0.5X 농도, 서열식별번호: 20)를 나타낸다. 레인 3은 T4 DNA 리가제의 부재하에 실시예 2에 기재된 라이게이션 반응을 진행시킨 후의 DNA 올리고 1 (1X 농도, 서열식별번호: 21), RNA

올리고 1 (1X 농도, 서열식별번호: 19) 및 DNA 스플린트 (4X 농도, 서열식별번호: 20)를 나타낸다. 레인 4는 대조군으로서 함께 혼합된, 반응되지 않은 DNA 올리고 1 (서열식별번호: 21) 및 DNA 스플린트 (서열식별번호: 20)를 나타낸다. 레인 5는 T4 DNA 리가제의 존재하에 실시예 2에 기재된 라이게이션 반응을 진행시킨 후의 DNA 올리고 1 (1X 농도, 서열식별번호: 21), RNA 올리고 1 (1X 농도, 서열식별번호: 19) 및 DNA 스플린트 (0.5X 농도, 서열식별번호: 20)를 나타낸다. 레인 6은 T4 DNA 리가제의 존재하에 실시예 2에 기재된 라이게이션 반응을 진행시킨 후의 DNA 올리고 1 (1X 농도, 서열식별번호: 21), RNA 올리고 1 (1X 농도, 서열식별번호: 19) 및 DNA 스플린트 (1X 농도, 서열식별번호: 20)를 나타낸다. 레인 7은 T4 DNA 리가제의 존재하에 실시예 2에 기재된 라이게이션 반응을 진행시킨 후의 DNA 올리고 1 (1X 농도, 서열식별번호: 21), RNA 올리고 1 (1X 농도, 서열식별번호: 19) 및 DNA 스플린트 (2X 농도, 서열식별번호: 20)를 나타낸다. 레인 8은 T4 DNA 리가제의 존재하에 실시예 2에 기재된 라이게이션 반응을 진행시킨 후의 DNA 올리고 1 (1X 농도, 서열식별번호: 21), RNA 올리고 1 (1X 농도, 서열식별번호: 19) 및 DNA 스플린트 (4X 농도, 서열식별번호: 20)를 나타낸다. 레인 9는 T4 DNA 리가제의 존재하에 실시예 2에 기재된 라이게이션 반응을 진행시켰고, 이러한 라이게이션 단계 후에 추가의 DNA 스플린트 (4.5X 농도, 서열식별번호: 20)와 함께 인큐베이션시킨 후의 DNA 올리고 1 (1X 농도, 서열식별번호: 21), RNA 올리고 1 (1X 농도, 서열식별번호: 19) 및 DNA 스플린트 (0.5X 농도, 서열식별번호: 20)를 나타낸다. 레인 10은 T4 DNA 리가제의 존재하에 실시예 2에 기재된 라이게이션 반응을 진행시켰고, 추가로 열 처리시키며 ExoI에 노출시킨 후의 DNA 올리고 1 (1X 농도, 서열식별번호: 21), RNA 올리고 1 (1X 농도, 서열식별번호: 19) 및 DNA 스플린트 (1X 농도, 서열식별번호: 20)를 나타낸다. 레인 11은 대조군으로서의 DNA 올리고 1 (서열식별번호: 21)을 나타낸다. 레인 12는 대조군으로서의 RNA 올리고 1을 나타낸다. 레인 13은 대조군으로서 함께 혼합된, 반응되지 않은 RNA 올리고 1 (서열식별번호: 19) 및 DNA 스플린트 (서열식별번호: 20)를 나타낸다. 1로 표시된 밴드는 RNA 올리고 1 (서열식별번호: 19)에 상응한다. 2로 표시된 밴드는 스플린트 (서열식별번호: 20)와 혼성화된 RNA 올리고 1 (서열식별번호: 19)에 상응한다. 3으로 표시된 밴드는 스플린트 (서열식별번호: 20)에 상응한다. 4로 표시된 밴드는 DNA 올리고 1 (서열식별번호: 21)에 상응한다. 5로 표시된 밴드는 스플린트 (서열식별번호: 20)와 혼성화된 DNA 올리고 1 (서열식별번호: 21)에 상응한다. A로 표시된 영역은 스플린트의 존재하에 라이게이션된 기질에 상응한다 (스플린트와 혼성화된, RNA 올리고 1에 라이게이션된 DNA 올리고 1). B로 표시된 영역은 스플린트의 부재하에 라이게이션된 기질에 상응한다 (RNA 올리고 1에 라이게이션된 DNA 올리고 1).

도 9는 실시예 3으로부터의 샘플을 포함하는, (A) SYBR 염색 전 및 (B) SYBR 염색 후 60분 동안 140 mV에서 실행된 5% PAGE TBE 바이오래드 기준 겔을 도시한 것이다. 레인 1 및 7은 트리다이 1 kB 래더를 나타낸다. 레인 2 및 8은 실시예 3A로부터 생성된 생성물 [DNA 스플린트 X1 (서열식별번호: 24)의 존재하에 5'-헥시닐-G를 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오티드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임)을 수반한 mRNA)와 반응시킨 DNA X1 (3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 23의 5' 말단에 그의 반대 말단에서 부착된 4개의 iSp18 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 22)]을 나타낸다. 레인 3 및 9는 실시예 3B로부터 생성된 생성물 [DNA 스플린트 X1 (서열식별번호: 24)의 존재하에 5'-헥시닐-G를 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오티드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임)을 수반한 mRNA)와 반응시킨 DNA X2 (그의 5' 말단에 부착된 CY3을 갖고 그의 3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 25)]을 나타낸다. 레인 4 및 10은 5'-헥시닐-G를 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오티드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임)을 수반한 mRNA)와 혼합시킨 DNA X2 (그의 5' 말단에 부착된 CY3을 갖고 그의 3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 25)를 나타낸다. 레인 5 및 11은 DNA X1 (3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 23의 5' 말단에 그의 반대 말단에서 부착된 4개의 iSp18 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 22)을 나타낸다. 레인 6 및 12는 5'-헥시닐-G를 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오티드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임)을 수반한 mRNA)를 나타낸다. 1로 표시된 밴드는, 부착된 DNA를 수반하는 경우 및 수반하지 않은 경우의 RNA X1에 상응한다. 밴드 2는 반응되지 않은 DNA X1 (3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 23의 5' 말단에 그의 반대 말단에서 부착된 4개의 iSp18 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 22)에 상응한다. 밴드 3은 반응되지 않은 DNA X2 (그의 5' 말단에 부착된 CY3을 갖고 그의 3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 25)에 상응한다. CY3 표지를 함유한 DNA만이 비-SYBR 염색된 겔 상에서 가시적이었다.

도 10은 DNA 헬리카제 [T4 Dda - E94C/A360C (돌연변이 E94C/A360C에 이어서 ( $\Delta$ M1)G1)을 수반한 서열식별번호: 14]]를 사용하여 MspA 나노 세공을 통하여 전위시킨, 실시예 3A에서 생성된 DNA/RNA 가닥을 나타내는 만화를 도

시한 것이다. 영역 A는 DNA 헬리카제 (B로 표지됨)와 결합하는 DNA 리더 (서열식별번호: 22)에 상응한다. 영역 A는 4개의 iSp18 스페이서 (X로서 제시되고 C로 표지됨)에 부착된다. 영역 D는 제2 DNA 서열 (서열식별번호: 23)에 상응한다. 영역 E는 5'-헥시닐-G 영역을 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오타이드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA)에 상응한다. 영역 D와 E는 클릭(click) 화학에 의해 부착된다. 반응된 아지드 및 헥시닐 기는 박스 (F로 표지됨)로써 나타낸다. 영역 G는 영역 D와 혼성화되는 DNA (서열식별번호: 18)이다. DNA (서열식별번호: 18)에 6개의 iSp18 스페이서 (Y로서 제시되고 H로 표지됨), 2개의 티민 (T로서 제시되고 I로 표지됨) 및 3' 콜레스테롤 TEG (J로 표지됨)가 부착된다.

도 11은 DNA 헬리카제 [T4 Dda - E94C/A360C (돌연변이 E94C/A360C에 이어서 ( $\Delta$ M1)G1을 수반한 서열식별번호: 14)]가 실시예 3A에 의해 생성된 DNA/RNA 생성물 (도 3에 도시된 만화)의 이동을 제어한 경우의 헬리카제 제어된 DNA 이동의 추적 예 [y-축 = 전류 (pA), x-축 = 시간 (s)]를 도시한 것이다. 영역 1은 RNA 상으로 라이게이션되지 않은 DNA 리더에 상응하고 (DNA X1), 영역 2는 4개의 iSp18 스페이서 (이러한 스페이서는 더 많은 양의 전류가 DNA 또는 RNA 영역보다는 나노 세공을 통하여 흐를 수 있게 한다)에 부착된 서열식별번호: 22에 상응하며, 영역 3은 DNA 서열 (서열식별번호: 23)에 상응하고, 영역 4는 반딧불이 루시페라제 mRNA 영역 (RNA X1; 5'의 최외각 뉴클레오타이드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA)에 상응한다. \*로 표지된 전류 상의 피크는 DNA와 RNA가 연결될 때 만들어지는 클릭 연쇄의 전위에 상응한다.

도 12는 진핵 RNA의 가닥 (파선으로서 제시됨)을 DNA의 가닥 (실선으로서 제시됨)에 부착시키는 방법을 나타내는 만화를 도시한 것이다. 진핵 RNA는 나머지 RNA 가닥과 반대 방향으로 실행되는 7-메틸구아노신 캡을 갖는다 (즉, 5' 말단은 역전된 염기를 포함한다; 별 모양으로서 제시되고 A로 표지됨). 진핵 RNA는 역전된 DNA 염기 (n개의 무작위 서열로서 제시되고 C로 표지됨)의 영역을 포함하는 DNA의 가닥 (B로 표지됨)에 라이게이션된다. 역전된 염기의 영역에서는, 이러한 염기가 또한, 반대 방향으로 실행된다. 이는 RNA 및 DNA 상의 역전된 염기의 두 영역이 함께 라이게이션될 수 있다는 것을 예시한다.

도 13은 실시예 4에서 생성된 RNA/cDNA 구축물을 나타내는 만화를 도시한 것이다. 영역 B는 서열식별번호: 30에 상응한다. 영역 C는 서열식별번호: 30 내의 마지막 3개의 아데닌에 상응한다. 영역 D는 3T 헤어핀과 10T 헤어핀 둘 다 내의 서열식별번호: 29에 상응한다. 영역 E는 4개의 iSpC3 스페이서 (X로서 제시됨)에 상응한다. 영역 F는 3T 헤어핀 내의 서열식별번호 27 및 10T 헤어핀 내의 서열식별번호: 28에 상응한다. 영역 G는 서열식별번호: 27 또는 28 내의 마지막 3개의 티민에 상응한다. 영역 H는 RNA 가닥 (서열식별번호: 30)의 역전사 동안 생성된 cDNA에 상응한다.

도 14는 헤어핀 라이게이션 및 역전사 전 및 후에 실시예 4로부터의 각종 샘플을 나타내는 10% PAGE TBE-우레아 변성 겔 (140 V에서 60분 동안 실행됨)을 도시한 것이다. 레인 1은 트리다이 래더를 나타낸다. 레인 2는 라이게이션 또는 역전사 전에 실시예 4에서 사용된 RNA 가닥 (서열식별번호: 30)을 나타낸다. 레인 3은 라이게이션 또는 역전사 전의 10T 헤어핀 대조군 (서열식별번호: 29는 그의 5' 말단에서 포스페이트 기에 부착되고; 반대 말단에서 서열식별번호: 28의 5' 말단에 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된다)을 나타낸다. 레인 4는 프라이머를 이용한 역전사 후의 RNA 가닥 (서열식별번호: 30)을 나타낸다. 레인 5는 3T 헤어핀 (서열식별번호: 29는 그의 5' 말단에서 포스페이트 기에 부착되고; 반대 말단에서 서열식별번호: 27의 5' 말단에 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된다)에 라이게이션된 후의 RNA 가닥 (서열식별번호: 30)을 나타낸다. 레인 6은 3T 헤어핀 (서열식별번호: 29는 그의 5' 말단에서 포스페이트 기에 부착되고; 반대 말단에서 서열식별번호: 27의 5' 말단에 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된다)에 라이게이션되고 연속해서 역전사된 후의 RNA 가닥 (서열식별번호: 30)을 나타낸다. 레인 7은 10T 헤어핀 (서열식별번호: 29는 그의 5' 말단에서 포스페이트 기에 부착되고; 반대 말단에서 서열식별번호: 28의 5' 말단에 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된다)에 라이게이션된 후의 RNA 가닥 (서열식별번호: 30)을 나타낸다. 레인 8은 10T 헤어핀 (서열식별번호: 29는 그의 5' 말단에서 포스페이트 기에 부착되고; 반대 말단에서 서열식별번호: 28의 5' 말단에 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된다)에 라이게이션되고 연속해서 역전사된 후의 RNA 가닥 (서열식별번호: 30)을 나타낸다. 레인 9는 RNA 가닥과 10T 헤어핀을 리가제의 부재하에 함께 인큐베이션한 대조군 실험을 나타낸다. 화살표 A는 실시예 4에서 사용된 RNA 가닥 (서열식별번호: 30)에 상응한다.

도 15는 DNA 헬리카제 [T4 Dda - E94C/C109A/C136A/A360C (돌연변이 E94C/C109A/C136A/A360C에 이어서 ( $\Delta$ M1)G1을 수반한 서열식별번호: 14)]를 사용하여 MspA 나노 세공을 통하여 전위시킨, 실시예 5에서 생성된

DNA/RNA/cDNA 가닥을 나타내는 만화를 도시한 것이다. 영역 A는 DNA 헬리카제 (B로 표지됨)와 결합하는 DNA 리더 (서열식별번호: 22)에 상응한다. 영역 A가 4개의 iSp18 스페이서 (X로서 제시되고 C로 표지됨)에 부착된다. 영역 D는 제2의 DNA 서열 (서열식별번호: 23)에 상응한다. 영역 E는 5'-헥시닐-G 영역을 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오타이드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA)에 상응한다. 영역 D와 E는 클릭 화학에 의해 부착된다. 반응된 아지드 및 헥시닐 기는 박스 (F로 표지됨)로써 나타난다. 영역 G는 영역 D와 혼성화되는 DNA (서열식별번호: 18)이다. DNA (서열식별번호: 18)에 6개의 iSp18 스페이서 (Y로서 제시되고 H로 표지됨), 2개의 티민 (T로서 제시되고 I로 표지됨) 및 3' 콜레스테롤 TEG (J로 표지됨)가 부착된다. 영역 K는 FLuc mRNA (RNA X1; 5' 최외각 뉴클레오타이드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA) 내의 마지막 3개의 아데닌에 상응한다. 영역 L은 10T 헤어핀 내의 서열식별번호: 29에 상응한다. 영역 M은 4개의 iSpC3 스페이서 (X로서 제시됨)에 상응한다. 영역 N은 10T 헤어핀 내의 서열식별번호: 28에 상응한다. 영역 O는 서열식별번호: 28 내의 마지막 3개의 티민에 상응한다. 영역 P는 RNA 가닥 (RNA X1; 5'의 최외각 뉴클레오타이드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA)의 역전사 동안 생성된 cDNA에 상응한다.

도 16은 DNA 헬리카제 [T4 Dda - E94C/C109A/C136A/A360C (돌연변이 E94C/C109A/C136A/A360C에 이어서 ( $\Delta$ M1)G1을 수반한 서열식별번호: 14)]가 실시예 5에서 생성된 DNA/RNA/cDNA 생성물 (도 15에 도시된 만화)의 이동을 제어한 경우의 헬리카제 제어된 DNA 이동의 추적 예 [y-축 = 전류 (pA), x-축 = 시간 (s)]를 도시한 것이다. 영역 1은 DNA 리더 (DNA X1)에 상응하고, 영역 2는 반딧불이 루시페라제 mRNA 영역 (RNA X1; 5'의 최외각 뉴클레오타이드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA)에 상응하며, 영역 3은 4개의 iSpC3 스페이서에 상응하고, 영역 4는 10T 헤어핀의 폴리T 영역에 상응하며, 영역 5는 mRNA의 역전사에 의해 생성된 cDNA에 상응한다.

도 17은 탈캡핑(decapping) 및 라이게이션 전 및 후에 실시예 5로부터의 각종 샘플을 나타내는 5% PAGE TBE-우레아 변성 겔 (140 V에서 60분 동안 실행됨)을 도시한 것이다. 레인 1은 트리다이 래더를 나타낸다. 레인 2는 캡핑된 RNA 가닥 (5'에서 5'로의 트리포스페이트 연쇄에 의해 상기 가닥의 5' 말단과 연결된 7-메틸구아노신 캡을 갖는 서열식별번호: 30)을 나타낸다. 레인 3은 비-RNA 폴리뉴클레오타이드 (RNA 서열 CAAGGG에 반대 말단에서 부착된 4개의 5-니트로인돌에 3' 말단에서 부착된 서열식별번호: 32의 5' 말단에 반대 말단에서 부착된 4개의 iSp18 스페이서에 3' 말단에서 부착된 서열식별번호: 31의 5' 말단에 부착된 30개의 SpC3 스페이서)를 나타낸다. 레인 4는 비-RNA 폴리뉴클레오타이드 (RNA 서열 CAAGGG에 반대 말단에서 부착된 4개의 5-니트로인돌에 3' 말단에서 부착된 서열식별번호: 32의 5' 말단에 반대 말단에서 부착된 4개의 iSp18 스페이서에 3' 말단에서 부착된 서열식별번호: 31의 5' 말단에 부착된 30개의 SpC3 스페이서)에 라이게이션된 후의 RNA 가닥 (서열식별번호: 30)을 나타낸다. 화살표 A는 RNA 가닥 (서열식별번호: 30)에 상응한다. 화살표 B는 비-RNA 폴리뉴클레오타이드에 상응한다. 화살표 C는 RNA 가닥을 비-RNA 폴리뉴클레오타이드에 라이게이션시킨 경우의 라이게이션된 생성물에 상응한다.

도 18은 변형된 염기를 수반한 경우 및 수반하지 않은 경우의 RNA로부터의 컨센서스 전류 수준을 예시한 것이다. 본 도면은 변형된 염기가 몇 가지 연속되는 전류 수준에 영향을 미친다는 것을 보여준다.

도 19는 나노 세공을 통한 RNA-DNA 2D 가닥 (RNA-센스 및 DNA-안티센스)의 DNA 헬리카제 제어된 전위를 도시한 것이다. 도 19는 RNA 대 DNA에 대하여 관찰된 상이한 평균 진폭 및 범위를 명확하게 보여준다.

도 20은 DNA 헬리카제 [He1308Mbu-E284C/S615C (돌연변이 E284C/S615C를 수반한 서열식별번호: 8)]가 실시예 8에 의해 생성된 DNA/RNA 생성물의 이동을 제어한 경우의 헬리카제 제어된 DNA 이동의 추적 예 [y-축 = 전류 (pA), x-축 = 시간 (s)]를 도시한 것이다.

도 21은 DNA 헬리카제 [T4 Dda - E94C/A360C (0.36  $\mu$ l, 3.8  $\mu$ M), 돌연변이 E94C/A360C에 이어서 ( $\Delta$ M1)G1을 수반한 서열식별번호: 14]가 실시예 3에 의해 생성된 DNA/RNA 생성물의 이동을 제어한 경우에 리세린 돌연변이체를 통한 헬리카제 제어된 DNA 이동의 추적 예 [y-축 = 전류 (pA), x-축 = 시간 (s)]를 도시한 것이다.

도 22는 DNA 헬리카제 [He1308Mbu-E284C/S615C (돌연변이 E284C/S615C를 수반한 서열식별번호: 8)]가 실시예 5에 의해 생성된 DNA/RNA 생성물의 이동을 제어한 경우에 CsgG 돌연변이체 세공 (CsgG-Eco-(Y51T/F56Q)-StrepII(C))9 (StepII(C)가 서열식별번호: 45이고 C-말단에서 부착되는 돌연변이 Y51T/F56Q를 수반한 서열식별번호: 44)을 통한 헬리카제 제어된 DNA 이동의 추적 예 [y-축 = 전류 (pA), x-축 = 시간 (s)]를 도시한 것이다.

## 서열 목록에 관한 설명

서열식별번호: 1은 MS-B1 돌연변이체 MspA 단량체를 코딩하는 코돈 최적화된 폴리뉴클레오타이드 서열을 제시한다. 이러한 돌연변이체에는 시그널 서열이 결여되고 다음 돌연변이를 포함한다: D90N, D91N, D93N, D118R, D134R 및 E139K.

서열식별번호: 2는 MspA 단량체의 MS-B1 돌연변이체의 성숙한 형태의 아미노산 서열을 제시한다. 이러한 돌연변이체에는 시그널 서열이 결여되고 다음 돌연변이를 포함한다: D90N, D91N, D93N, D118R, D134R 및 E139K.

서열식별번호: 3은  $\alpha$ -헤모리신-E111N/K147N ( $\alpha$ -HL-NN; 문헌 [Stoddart et al., PNAS, 2009; 106(19):7702-7707])의 하나의 단량체를 코딩하는 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 4는  $\alpha$ -HL-NN의 하나의 단량체의 아미노산 서열을 제시한다.

서열식별번호: 5 내지 7은 MspB, C 및 D의 아미노산 서열을 제시한다.

서열식별번호: 8은 He1308 Mbu의 아미노산 서열을 제시한다.

서열식별번호: 9는 He1308 Csy의 아미노산 서열을 제시한다.

서열식별번호: 10은 He1308 Tga의 아미노산 서열을 제시한다.

서열식별번호: 11은 He1308 Mhu의 아미노산 서열을 제시한다.

서열식별번호: 12는 TraI Eco의 아미노산 서열을 제시한다.

서열식별번호: 13은 XPD Mbu의 아미노산 서열을 제시한다.

서열식별번호: 14는 Dda 1993의 아미노산 서열을 제시한다.

서열식별번호: 15는 Trwc Cba의 아미노산 서열을 제시한다.

서열식별번호: 16은 실시예 1에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 17은 실시예 1에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 18은 실시예 1 및 3에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 19는 실시예 2에서 사용된 폴리뉴클레오타이드 서열을 제시한다. 이러한 서열은 5' 포스페이트 기를 갖는다.

서열식별번호: 20은 실시예 2에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 21은 실시예 2에서 사용된 폴리뉴클레오타이드 서열을 제시한다. 이러한 서열은 5' CY3 기를 갖는다.

서열식별번호: 22는 실시예 3에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 23은 실시예 3에서 사용된 폴리뉴클레오타이드 서열을 제시한다. 이러한 서열은 3' 아지드N 기를 갖는다.

서열식별번호: 24는 실시예 3에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 25는 실시예 3에서 사용된 폴리뉴클레오타이드 서열을 제시한다. 이러한 서열은 3' 아지드N 기 및 5' CY3 기를 갖는다.

서열식별번호: 26은 실시예 3 및 5에서 사용된 폴리뉴클레오타이드 서열의 오픈 리딩 프레임을 제시한다. 이러한 서열은 서열 내의 첫 번째 G에 부착된 5'-헥시닐 기를 갖는다.

서열식별번호: 27은 실시예 4 및 5에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 28은 실시예 4에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 29는 실시예 4 및 5에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 30은 실시예 4 및 6에서 사용된 폴리뉴클레오타이드 서열을 제시한다.

서열식별번호: 31은 실시예 6에서 사용된 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 32는 실시예 6에서 사용된 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 33은 단독중합체 관독을 예시하기 위해 사용된 서열을 제시한다.

서열식별번호: 34는 단독중합체 관독을 예시하기 위해 사용된 서열을 제시한다.

서열식별번호: 35는 실시예 8에서 사용된 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 36은 실시예 8에서 사용된 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 37은 실시예 8에서 사용된 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 38은 실시예 8에서 사용된 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 39는 실시예 8에서 사용된 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 40은 실시예 8에서 사용된 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 41은 리세닌 단량체를 코딩하는 폴리뉴클레오티드 서열을 제시한다.

서열식별번호: 42는 리세닌 단량체의 아미노산 서열을 제시한다.

서열식별번호: 43은 에스케리키아 콜라이 (*Escherichia coli*) Str. K-12 substr. MC4100으로부터의 야생형 CsgG 단량체를 코딩하는 코돈 최적화된 폴리뉴클레오티드 서열을 제시한다. 이러한 단량체에는 시그널 서열이 결여된다.

서열식별번호: 44는 에스케리키아 콜라이 Str. K-12 substr. MC4100으로부터의 야생형 CsgG 단량체의 성숙한 형태의 아미노산 서열을 제시한다. 이러한 단량체에는 시그널 서열이 결여된다. 이러한 CsgG에 대해 사용된 약어는 CsgG-Eco이다.

서열식별번호: 45는 StepII(C)의 아미노산 서열을 제시한다.

서열식별번호: 46 내지 64는 본 명세서에 기재된 폴리뉴클레오티드 서열이다.

서열식별번호: 65 및 66은 본 실시예에서 사용된 폴리뉴클레오티드 서열이다.

### 발명을 실시하기 위한 구체적인 내용

[0032] 개시된 생성물 및 방법의 상이한 적용이 관련 기술분야의 특이적 요구에 맞춰질 수 있음을 이해해야 한다. 본원에서 사용된 용어는 단지 본 발명의 특별한 실시양태를 설명하기 위한 것이며, 본 발명을 제한하는 것으로 의도되지 않는다는 것을 또한 이해해야 한다.

[0033] 또한 본 명세서 및 첨부된 청구범위에서 사용된 바와 같이, 단수 형태는 그 내용이 명확하게 달리 지시하지 않는 한 복수 대상을 포함한다. 따라서, 예를 들어 "폴리뉴클레오티드"에 대한 언급은 2개 이상의 폴리뉴클레오티드를 포함하고, "폴리뉴클레오티드 결합성 단백질"에 대한 언급은 2개 이상의 상기 단백질을 포함하며, "헬리카제"에 대한 언급은 2개 이상의 헬리카제를 포함하고, "단량체"에 대한 언급은 2개 이상의 단량체를 지칭하며, "세공"에 대한 언급은 2개 이상의 세공을 포함한다.

[0034] 본원에 인용된 모든 공보, 특허 및 특허 출원은 상기 또는 후속 여부에 관계없이 그 전문이 본원에 참조로 포함된다.

[0035] 표적 RNA 폴리뉴클레오티드의 특징을 규명함

[0036] 본 발명의 방법은 표적 RNA 폴리뉴클레오티드의 특징을 규명하는 것을 포함한다. 이러한 RNA 폴리뉴클레오티드가 막횡단 세공에 전달되고, 이러한 세공을 사용하여 RNA 폴리뉴클레오티드의 특징을 규명한다. 본 발명은 DNA 헬리카제 효소의 제어하에 표적 RNA 폴리뉴클레오티드가 막횡단 세공에 대하여 이동함에 따라 하나 이상의 측정치를 취함으로써 표적 리보핵산 (RNA) 폴리뉴클레오티드의 특징을 규명하는 방법을 제공한다.

[0037] 막횡단 세공은 표적 폴리뉴클레오티드의 단일 분자를 검출할 수 있기 때문에, 표적 RNA 폴리뉴클레오티드를 증폭시킬 필요가 없다. 본 방법은 전형적으로, 폴리머라제 연쇄 반응 (PCR) 또는 역전사 PCR (RT-PCR)을 포함하지 않는다. 이는 표적 RNA 폴리뉴클레오티드의 특징을 규명하는 데 필요한 작업 흐름의 양을 상당히 감소시켜



준다. 이는 또한 PCR에 의해 도입된 임의의 바이어스 및 인공물을 피하게 해준다.

- [0038] 본 발명의 방법은 RNA 폴리뉴클레오타이드의 하나 이상의 특징을 결정하거나 또는 측정하는 것에 관한 것일 수 있다. 이러한 방법은 RNA 폴리뉴클레오타이드의 1개, 2개, 3개, 4개 또는 5개 또는 그 초과 특징을 결정하거나 또는 측정하는 것을 포함할 수 있다. 하나 이상의 특징은 바람직하게, (i) RNA 폴리뉴클레오타이드의 길이 또는 크기, (ii) RNA 폴리뉴클레오타이드의 실체, (iii) RNA 폴리뉴클레오타이드의 서열, (iv) RNA 폴리뉴클레오타이드의 2차 구조 및 (v) RNA 폴리뉴클레오타이드가 변형되는 지의 여부로부터 선택된다. (i) 내지 (v)의 임의의 조합, 예컨대 {i}, {ii}, {iii}, {iv}, {v}, {i,ii}, {i,iii}, {i,iv}, {i,v}, {ii,iii}, {ii,iv}, {ii,v}, {iii,iv}, {iii,v}, {iv,v}, {i,ii,iii}, {i,ii,iv}, {i,ii,v}, {i,iii,iv}, {i,iii,v}, {i,iv,v}, {ii,iii,iv}, {ii,iii,v}, {ii,iv,v}, {iii,iv,v}, {i,ii,iii,iv}, {i,ii,iii,v}, {i,ii,iv,v}, {i,iii,iv,v}, {ii,iii,iv,v} 또는 {i,ii,iii,iv,v}이 본 발명에 따라서 측정될 수 있다. 상기 열거된 조합 중 임의의 것을 포함한 (i) 내지 (v)의 상이한 조합을 측정할 수 있다. 본 발명의 방법은 바람직하게, RNA 폴리뉴클레오타이드의 서열을 추정하거나 또는 RNA 폴리뉴클레오타이드를 서열 분석하는 것을 포함한다.
- [0039] (i)의 경우, RNA 폴리뉴클레오타이드의 길이는, 예를 들어 RNA 폴리뉴클레오타이드와 세공 간의 상호 작용의 수 또는 RNA 폴리뉴클레오타이드와 세공 간의 상호 작용의 지속 기간을 결정함으로써 측정될 수 있다.
- [0040] (ii)의 경우, RNA 폴리뉴클레오타이드의 실체는 수많은 방식으로 측정될 수 있다. RNA 폴리뉴클레오타이드의 실체는 RNA 폴리뉴클레오타이드의 서열의 측정과 연계하거나 또는 RNA 폴리뉴클레오타이드의 서열의 측정과 연계하지 않고서 측정될 수 있다. 전자가 간단한데; RNA 폴리뉴클레오타이드를 서열 분석함으로써 실체를 확인한다. 후자는 여러 방식으로 수행될 수 있다. 예를 들어, RNA 폴리뉴클레오타이드 내의 특별한 모티프의 존재를 측정할 수 있다 (RNA 폴리뉴클레오타이드의 나머지 서열은 측정하지 않는다). 또 다른 한편으로, 상기 방법에서 특별한 전기적 및/또는 광학 시그널의 측정은 RNA 폴리뉴클레오타이드를 특별한 공급원으로부터 유래한 것으로서 확인할 수 있다.
- [0041] (iii)의 경우, RNA 폴리뉴클레오타이드의 서열은 이전에 기재된 바와 같이 결정될 수 있다. 적합한 서열 분석 방법, 특히 전기적 측정을 이용하는 방법이 문헌 [Stoddart D et al., Proc Natl Acad Sci, 12;106(19):7702-7, Lieberman KR et al., J Am Chem Soc. 2010;132(50):17961-72] 및 국제 출원 WO 2000/28312에 기재되어 있다.
- [0042] (iv)의 경우, 2차 구조는 각종 방식으로 측정될 수 있다. 예를 들어, 상기 방법이 전기적 측정을 포함하는 경우, 2차 구조는 세공을 통하여 흐르는 전류 상의 변화 또는 체류 시간 상의 변화를 이용하여 측정될 수 있다. 이로써, 단일 가닥 RNA 폴리뉴클레오타이드의 영역과 이중 가닥 RNA 폴리뉴클레오타이드의 영역이 구별될 수 있다.
- [0043] (v)의 경우, 임의의 변형의 존재 또는 부재를 측정할 수 있다. 그 방법은 바람직하게, 폴리뉴클레오타이드가 메틸화에 의해, 산화에 의해, 손상에 의해, 하나 이상의 단백질을 이용하거나 또는 하나 이상의 표지, 태그 또는 스페이서를 이용하여 변형되는 지의 여부를 결정하는 것을 포함한다. 특이적 변형은 다음에 기재되는 방법을 이용하여 측정될 수 있는 세공과의 특이적 상호 작용을 초래할 것이다. 예를 들어, 메틸시토신은 각 리보뉴클레오타이드와의 상호 작용 동안 세공을 통하여 흐르는 전류를 근거로 하여 시토신과 구별될 수 있다. 본 발명의 방법은 심지어 단일 샘플 중에서의 RNA와 DNA 간을 구별하기 위해 사용될 수 있다: RNA 및 DNA는 RNA 서열과 DNA 서열이 동일한 경우일지라도 평균 진폭 및 범위의 함수로서 서로 차별될 수 있다.
- [0044] 본 방법은 세공이 막에 존재하는 막/세공 시스템을 조사하는 데 적합한 임의의 장치를 사용하여 수행될 수 있다. 이러한 방법은 막형단 세공 감지에 적합한 임의의 장치를 이용하여 수행될 수 있다. 예를 들어, 상기 장치는 수성 용액을 포함하는 챔버; 및 이러한 챔버를 2개의 섹션으로 분리시켜 주는 장벽을 포함한다. 장벽은 전형적으로, 세공을 함유하는 막이 형성되는 구멍을 갖는다. 또 다른 한편으로, 장벽은 세공이 존재하는 막을 형성한다.
- [0045] 상기 방법은 국제 출원 번호 PCT/GB08/000562 (WO 2008/102120)에 기재된 장치를 이용하여 수행될 수 있다.
- [0046] 본 방법은 RNA 폴리뉴클레오타이드가 세공에 대하여 이동함에 따라 세공을 관통하는 전류를 측정하는 것을 포함할 수 있다. 따라서, 상기 장치는 또한, 전위를 인가할 수 있고 막 및 세공에 걸쳐 전기적 시그널을 측정할 수 있는 전기 회로를 포함할 수 있다. 상기 방법은 패치 클램프 또는 전압 클램프를 이용하여 수행될 수 있다. 상기 방법은 바람직하게, 전압 클램프의 사용을 포함한다.
- [0047] 본 발명의 방법은 RNA 폴리뉴클레오타이드가 세공에 대하여 이동함에 따라 세공을 관통하는 전류를 측정하는 것을 포함할 수 있다. 상기 폴리뉴클레오타이드가 세공에 대하여 이동함에 따라 세공을 관통하는 전류를 이용하여, 표적 RNA 폴리뉴클레오타이드의 서열을 결정한다. 이것이 가닥 서열 분석이다. 막형단 단백질 세공을 통하여 이온

전류를 측정하는 데 적합한 조건은 관련 기술분야에 공지되어 있고 본 실시예에 개시된다. 상기 방법은 전형적으로, 막 및 세공에 걸쳐 인가된 전압을 이용하여 수행된다. 사용된 전압은 전형적으로 +5 V 내지 -5 V, 예컨대 +4 V 내지 -4 V, +3 V 내지 -3 V 또는 +2 V 내지 -2 V이다. 사용된 전압은 전형적으로 -600 mV 내지 +600 mV 또는 -400 mV 내지 +400 mV이다. 사용된 전압은 바람직하게, -400 mV, -300 mV, -200 mV, -150 mV, -100 mV, -50 mV, -20 mV 및 0 mV로부터 선택된 하한치 및 +10 mV, +20 mV, +50 mV, +100 mV, +150 mV, +200 mV, +300 mV 및 +400 mV로부터 독립적으로 선택된 상한치를 갖는 범위 내에 있다. 사용된 전압은 보다 바람직하게, 100 mV 내지 240 mV의 범위 및 가장 바람직하게 120 mV 내지 220 mV의 범위 내에 있다. 증가된 인가된 전위를 이용함으로써 세공에 의해 상이한 리보뉴클레오타이드들 간의 판별을 증가시킬 수 있다.

[0048] 본 방법은 전형적으로, 임의의 전하 담체, 예컨대 금속 염, 예를 들어 알칼리 금속 염, 할라이드 염, 예를 들어 클로라이드 염, 예컨대 알칼리 금속 클로라이드 염의 존재하에 수행된다. 전하 담체는 이온성 액체 또는 유기 염, 예를 들어 테트라메틸 암모늄 클로라이드, 트리메틸페닐 암모늄 클로라이드, 페닐트리메틸 암모늄 클로라이드, 또는 1-에틸-3-메틸 이미다졸리움 클로라이드를 포함할 수 있다. 상기 논의된 예시적인 장치에서, 염은 챔버 내의 수성 용액에 존재한다. 염화칼륨 (KCl), 염화나트륨 (NaCl), 염화세슘 (CsCl) 또는 페로시아화칼륨과 페리시아화칼륨의 혼합물이 전형적으로 사용된다. KCl, NaCl 및 페로시아화칼륨과 페리시아화칼륨의 혼합물이 바람직하다. 전하 담체는 막에 걸쳐 비대칭적일 수 있다. 예를 들어, 전하 담체의 유형 및/또는 농도는 막의 각 측면에서 상이할 수 있다.

[0049] 염 농도는 포화 상태일 수 있다. 염 농도는 3 M 이하일 수 있고, 전형적으로 0.1 내지 2.5 M, 0.3 내지 1.9 M, 0.5 내지 1.8 M, 0.7 내지 1.7 M, 0.9 내지 1.6 M 또는 1 M 내지 1.4 M이다. 염 농도는 바람직하게 150 mM 내지 1 M이다. 본 방법은 바람직하게, 적어도 0.3 M, 예컨대 적어도 0.4 M, 적어도 0.5 M, 적어도 0.6 M, 적어도 0.8 M, 적어도 1.0 M, 적어도 1.5 M, 적어도 2.0 M, 적어도 2.5 M 또는 적어도 3.0 M의 염 농도를 이용하여 수행된다. 높은 염 농도는 높은 시그널 대 소음 비를 제공하고, 정상적인 전류 변동의 배경에 대항하여 확인될 리보뉴클레오타이드의 존재를 나타내는 전류를 허용한다.

[0050] 본 방법은 전형적으로, 완충제의 존재하에 수행된다. 상기 논의된 예시적인 장치에서, 완충제는 챔버 내의 수성 용액에 존재한다. 임의의 완충제가 본 발명의 방법에 사용될 수 있다. 전형적으로, 완충제는 인산염 완충제이다. 다른 적합한 완충제는 HEPES 및 트리스-HCl 완충제이다. 상기 방법은 전형적으로, 4.0 내지 12.0, 4.5 내지 10.0, 5.0 내지 9.0, 5.5 내지 8.8, 6.0 내지 8.7 또는 7.0 내지 8.8 또는 7.5 내지 8.5의 pH에서 수행된다. 사용된 pH는 바람직하게 약 7.5이다.

[0051] 본 방법은 0°C 내지 100°C, 15°C 내지 95°C, 16°C 내지 90°C, 17°C 내지 85°C, 18°C 내지 80°C, 19°C 내지 70°C, 또는 20°C 내지 60°C에서 수행될 수 있다. 상기 방법은 전형적으로 실온에서 수행된다. 상기 방법은 임의로, 효소 기능을 지지해 주는 온도, 예컨대 약 37°C에서 수행된다.

[0052] 본 방법은 유리 뉴클레오타이드 또는 유리 뉴클레오타이드 유사체의 존재하에 및/또는 헬리카제 또는 구축물의 작용을 촉진시켜 주는 효소 보조인자의 존재하에 수행될 수 있다. 이러한 방법은 또한, 유리 뉴클레오타이드 또는 유리 뉴클레오타이드 유사체의 부재하에 및 효소 보조인자의 부재하에 수행될 수 있다. 유리 뉴클레오타이드는 상기 논의된 개개의 뉴클레오타이드 중 임의의 것의 하나 이상일 수 있다. 유리 뉴클레오타이드는 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), 사이클릭 아데노신 모노포스페이트 (cAMP), 사이클릭 구아노신 모노포스페이트 (cGMP), 데옥시아데노신 모노포스페이트 (dAMP), 데옥시아데노신 디포스페이트 (dADP), 데옥시아데노신 트리포스페이트 (dATP), 데옥시구아노신 모노포스페이트 (dGMP), 데옥시구아노신 디포스페이트 (dGDP), 데옥시구아노신 트리포스페이트 (dGTP), 데옥시티미딘 모노포스페이트 (dTMP), 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 포함하지만, 이에 제한되지 않는다. 유리 뉴클레오타이드는 바람직하게 AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP 또는 dCMP로부터 선택된다. 유리 뉴클레오타이드는 바람직하게 아데노신 트리포스페이트 (ATP)이다. 효소 보조인자는 헬리카제 또는 구축물이 기능할 수 있도록 해주는 인자이다. 효소 보조인자는 바람직하게 2가 금속 양이온이다. 2가 금속 양이온은 바람직하게  $Mg^{2+}$ ,  $Mn^{2+}$ ,  $Ca^{2+}$  또는  $Co^{2+}$ 이다. 효소 보조인자

는 가장 바람직하게  $Mg^{2+}$ 이다.

- [0053] 표적 RNA
- [0054] RNA는 2개 이상의 리보뉴클레오타이드를 포함하는 거대 분자이다. 표적 RNA 폴리뉴클레오타이드는 진핵 또는 원핵 RNA일 수 있다. 표적 RNA 폴리뉴클레오타이드는 임의의 리보뉴클레오타이드의 임의의 조합을 포함할 수 있다. 리보뉴클레오타이드는 자연 발생적 또는 인공적일 수 있다. 표적 RNA 폴리뉴클레오타이드 내의 하나 이상의 리보뉴클레오타이드는 산화되거나 또는 메틸화될 수 있다. 표적 RNA 내의 하나 이상의 리보뉴클레오타이드가 손상될 수 있다. 예를 들어, 표적 RNA는 피리미딘 이량체, 예컨대 우라실 이량체를 포함할 수 있다. 이러한 이량체는 전형적으로, 자외선에 의한 손상과 연관되고, 피부 흑색종의 주요 원인이다. 표적 RNA 폴리뉴클레오타이드 내의 하나 이상의 리보뉴클레오타이드는, 예를 들어 표지 또는 태그로 변형시킬 수 있다. 적합한 표지는 다음에 기재된다. 표적 RNA는 하나 이상의 스페이서를 포함할 수 있다.
- [0055] 리보뉴클레오타이드는 전형적으로, 핵염기, 리보스 당 및 적어도 하나의 포스페이트 기를 함유한다. 핵염기는 전형적으로 헤테로사이클릭이다. 핵염기는 퓨린 및 피리미딘 및 보다 구체적으로 아데닌, 구아닌, 티민, 우라실 및 시토신을 포함하지만, 이에 제한되지 않는다. 뉴클레오타이드는 전형적으로 모노포스페이트, 디포스페이트 또는 트리포스페이트를 함유한다. 포스페이트는 뉴클레오타이드의 5' 또는 3' 측면 상에 부착될 수 있다.
- [0056] 리보뉴클레오타이드는 아데노신 모노포스페이트 (AMP), 구아노신 모노포스페이트 (GMP), 티미딘 모노포스페이트 (TMP), 우리딘 모노포스페이트 (UMP), 시티딘 모노포스페이트 (CMP), 5-메틸시티딘 모노포스페이트, 5-메틸시티딘 디포스페이트, 5-메틸시티딘 트리포스페이트, 5-히드록시메틸시티딘 모노포스페이트, 5-히드록시메틸시티딘 디포스페이트 및 5-히드록시메틸시티딘 트리포스페이트를 포함하지만, 이에 제한되지 않는다. 뉴클레오타이드는 바람직하게 AMP, TMP, GMP, CMP 및 UMP로부터 선택된다.
- [0057] 리보뉴클레오타이드 탈염기성일 수 있다 (즉, 핵염기가 결여된다). 리보뉴클레오타이드는 또한, 핵염기 및 당이 결여될 수 있다 (즉, 이는 C3 스페이서이다).
- [0058] 표적 RNA 폴리뉴클레오타이드 내의 리보뉴클레오타이드는 임의의 방식으로 서로 부착될 수 있다. 리보뉴클레오타이드는 전형적으로, 핵산에서와 같이 그의 당 및 포스페이트 기에 의해 부착된다. 리보뉴클레오타이드는 피리미딘 이량체에서와 같이 그의 핵염기를 통해 연결될 수 있다.
- [0059] RNA는 극도로 다양한 분자이다. 표적 RNA 폴리뉴클레오타이드는 임의의 자연 발생적 또는 합성 리보뉴클레오타이드 분자, 예를 들어, RNA, 메신저 RNA (mRNA), 리보솜 RNA (rRNA), 이질적 핵 RNA (hnRNA), 전이 RNA (tRNA), 전이-메신저 RNA (tmRNA), 마이크로 RNA (miRNA), 작은 핵 RNA (snRNA), 작은 핵소체 RNA (snoRNA), 시그널 인식 입자 (SRP RNA), SmY RNA, 작은 카잘 몸체(Cajal body)-특이적 RNA (scaRNA), 가이드 RNA (gRNA), 스플라이스된 리더 RNA (SL RNA), 안티센스 RNA (asRNA), 긴 비코딩 RNA (lncRNA), 피위(Piwi)-상호 작용성 RNA (piRNA), 저분자 간섭 RNA (siRNA), 트랜스-작용성 siRNA (tasiRNA), 반복 서열 관련 siRNA (rasiRNA), Y RNA, 바이러스 RNA 또는 염색체 RNA일 수 있는데, 이들 모두는 경우에 따라, 단일 가닥, 이중 가닥 또는 삼중 가닥일 수 있다.
- [0060] 표적 RNA 폴리뉴클레오타이드는 바람직하게 메신저 RNA (mRNA)이다. 표적 mRNA는 대체 스플라이스 변이체일 수 있다. 변경된 양 (또는 수준)의 mRNA 및/또는 대체 mRNA 스플라이스 변이체는 질환 또는 병태와 연관될 수 있다.
- [0061] 또 다른 한편으로 표적 RNA 폴리뉴클레오타이드는 마이크로RNA (또는 miRNA)이다. 저 농도로는 검출하기가 어려운 RNA의 한 군은 마이크로-리보핵산 (마이크로-RNA 또는 miRNA)이다. miRNA는 전사 후에 단백질 생성을 조절할 수 있는, 고도로 안정적인 RNA 올리고머이다. 이는 2가지 기전 중 한 가지에 의해 작용한다. 식물에서는, miRNA가 주로 메신저 RNA의 절단을 지시함으로써 작용하는 것으로 밝혀진 반면, 동물에서는 miRNA에 의한 유전자 조절이 전형적으로, 번역을 방해하는 메신저 RNA의 3' UTR에 대한 miRNA의 혼성화를 포함한다 (Lee et al., Cell 75, 843-54 (1993); Wightman et al., Cell 75, 855-62 (1993); and Esquela-Kerscher et al., Cancer 6, 259-69 (2006)). miRNA는 종종, 그의 표적과 불완전한 상보성으로 결합한다. 이들은 200개 이상 정도의 많은 유전자 표적과 각각 결합하고 모든 인간 유전자의 1/3을 초과하여 조절하는 것으로 예상되었다 (Lewis et al., Cell 120, 15-20 (2005)).
- [0062] 본 발명에서 사용하기 적합한 miRNA는 관련 기술분야에 널리 공지되어 있다. 예를 들어, 적합한 miRNA는 공개적으로 입수 가능한 데이터베이스 상에 저장된다 (Jiang Q., Wang Y., Hao Y., Juan L., Teng M., Zhang X., Li M., Wang G., Liu Y., (2009) miR2Disease: a manually curated database for microRNA deregulation in

human disease. Nucleic Acids Res.). 특정의 마이크로RNA의 발현 수준은 종양에서 변하는 것으로 공지되므로, 상이한 종양 유형에 특징적인 패턴의 마이크로RNA 발현이 제공된다 (Rosenfeld, N. et al., Nature Biotechnology 26, 462-9 (2008)). 또한, miRNA 프로파일은 메신저 RNA 프로파일보다 더 큰 정확도로 종양 발생 단계를 밝혀낼 수 있는 것으로 나타났다 (Lu et al., Nature 435, 834-8 (2005) and Barshack et al., The International Journal of Biochemistry & Cell Biology 42, 1355-62 (2010)). 이러한 발견은 miRNA의 높은 안정성 및 혈청 및 혈장 중에서의 순환성 miRNA를 검출할 수 있는 능력과 함께 (문헌 [Wang et al., Biochemical and Biophysical Research Communications 394, 184-8 (2010); Gilad et al., PloS One 3, e3148 (2008); and Keller et al., Nature Methods 8, 841-3 (2011)]), 암 바이오마커로서의 마이크로RNA의 잠재적 이용에 상당한 관심을 갖게 해주었다. 치료가 효과적이기 위해서는, 암을 정확하게 분류하고 상이하게 치료할 필요가 있지만, 분류 수단으로서의 종양 형태학 평가의 효능은 많은 상이한 유형의 암이 형태학적 특징을 공유한다는 사실에 의해 저하된다. miRNA는 잠재적으로 보다 신뢰성 있고 덜 침습적인 해결책을 제공한다.

- [0063] 질환 또는 병태를 진단 또는 예측하기 위해 mRNA 및 miRNA를 사용하는 것이 다음에 보다 상세히 논의된다.
- [0064] 임의의 수의 RNA를 조사할 수 있다. 예를 들어, 본 발명의 방법은 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, 100 개 또는 그 초과인 RNA 분자의 존재, 부재 또는 하나 이상의 특징을 결정하는 것에 관한 것일 수 있다.
- [0065] 폴리뉴클레오티드는 자연 발생적 또는 인공적일 수 있다. 예를 들어, 본 방법은 2개 이상의 제작된 올리고뉴클레오티드의 서열을 검증하기 위해 사용될 수 있다. 이러한 방법은 전형적으로 시험관 내에서 수행된다.
- [0066] 표적 RNA 폴리뉴클레오티드는 임의의 길이일 수 있다. 예를 들어, RNA 폴리뉴클레오티드는 그 길이가 적어도 10개, 적어도 50개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 400개 또는 적어도 500개의 리보뉴클레오티드일 수 있다. 표적 RNA는 그 길이가 1,000개 이상의 리보뉴클레오티드, 5,000개 이상의 리보뉴클레오티드 또는 100,000개 이상의 리보뉴클레오티드일 수 있다. 이러한 방법을 이용하여 표적 RNA의 전체 또는 단지 일부의 특징을 규명할 수 있다. 서열 분석되는 RNA의 일부는 바람직하게 모든 표적 분자를 포함하지만, 예를 들어 전체 분자보다 작을 수 있는데, 예를 들어, 4개 염기 내지 1 kb, 예를 들어, 4 내지 100개 염기일 수 있다.
- [0067] 표적 RNA 폴리뉴클레오티드는 전형적으로 임의의 적합한 샘플에 존재하거나 또는 이러한 샘플로부터 유래된다. 본 발명은 전형적으로, 표적 RNA 폴리뉴클레오티드를 함유하는 것으로 공지되거나 또는 이를 함유하는 것으로 의심되는 샘플 상에서 수행된다. 또 다른 한편으로, 본 발명은 특정 샘플 중에서의 그의 존재가 공지되어 있거나 또는 의심되는 하나 이상의 표적 RNA의 실체를 확증하기 위해 상기 샘플 상에서 수행될 수 있다.
- [0068] 샘플은 생물학적 샘플일 수 있다. 본 발명은 임의의 유기체 또는 미생물로부터 획득되거나 또는 추출된 샘플 상에서 시험관내 수행될 수 있다. 이러한 유기체 또는 미생물은 전형적으로, 고세균, 원핵 생물 또는 진핵 생물이며 전형적으로, 다음 5개 계 중 하나에 속한다: 식물계, 동물계, 진균, 모네라계 및 원생 생물계. 표적 RNA 폴리뉴클레오티드는 진핵 세포로부터 유래될 수 있거나 또는 진핵 세포의 전사 기구를 이용하여 바이러스로부터 유래될 수 있다. 본 발명은 임의의 바이러스로부터 획득되거나 또는 추출된 샘플 상에서 시험관내 수행될 수 있다.
- [0069] 샘플은 바람직하게 유체 샘플이다. 샘플은 전형적으로, 환자의 체액을 포함한다. 샘플은 소변, 림프액, 타액, 점액 또는 양수일 수 있지만 혈액, 혈장 또는 혈청이 바람직하다. 전형적으로, 샘플은 그 기원이 인간이지만, 또 다른 한편으로 또 다른 포유류 동물, 예컨대 상업적으로 사육된 동물, 예컨대 말, 소, 양 또는 돼지로부터 유래될 수 있거나 또는 또 다른 한편으로, 애완동물, 예컨대 고양이 또는 개로부터 유래될 수 있다. 또 다른 한편으로, 식물 기원의 샘플은 전형적으로, 상업 작물, 예컨대 곡물, 콩과 식물, 과일 또는 야채, 예를 들어 밀, 보리, 귀리, 카놀라, 옥수수, 콩, 쌀, 바나나, 사과, 토마토, 감자, 포도, 담배, 콩류, 렌틸, 사탕 수수, 코코아 또는 면화로부터 획득된다.
- [0070] 샘플은 비-생물학적 샘플일 수 있다. 비-생물학적 샘플은 바람직하게 유체 샘플이다. 비-생물학적 샘플의 예는 수술용 유체, 물, 예컨대 식수, 해수 또는 강물, 및 실험실 시험용 시약을 포함한다.
- [0071] 샘플은 전형적으로, 검정되기 전에, 예를 들어 원심 분리에 의해 또는 적혈구와 같은 원치 않는 분자 또는 세포를 걸러내는 막을 통과함으로써 처리된다. 샘플은 채취한 즉시 측정할 수 있다. 샘플은 또한 전형적으로, 검정 이전에, 바람직하게는 -70°C 아래에서 저장될 수 있다. 표적 RNA 폴리뉴클레오티드는 전형적으로, 본 발명의 방법에 사용하기 전에, 샘플로부터 추출된다. RNA 추출용 키트는, 예를 들어 뉴 잉글랜드 바이오랩스(New England Biolabs®) 및 인비트로젠(Invitrogen®)으로부터 시판되고 있다.



[0072] 표적 RNA의 변형

[0073] RNA 폴리뉴클레오타이드에 대한 변형은 DNA 헬리카제 결합을 촉진시켜 주고/주거나 변형된 RNA 폴리뉴클레오타이드에 대한 증가된 DNA 헬리카제 결합 효과를 유발시키거나 또는 이러한 효과를 갖는 임의의 변형일 수 있다. 본원에 사용된 바와 같은 용어 "결합"은 친화도, 또는 DNA 헬리카제와 기질 폴리뉴클레오타이드가 임의의 소정의 시간에 결합될 개연성을 지칭한다. 생화학적으로, 이러한 친화도 상의 증가는 "작동 속도" 또는 결합 속도의 증가 또는 "이탈 속도" 또는 결합 해제 속도의 감소에 의해 유발되거나 또는 "작동 속도"의 증가와 "이탈 속도"의 감소 둘 다에 의해 유발될 수 있었다.

[0074] RNA 폴리뉴클레오타이드의 변형은 변형된 RNA에 대한 DNA 헬리카제의 친화도를 적어도 10%, 바람직하게 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80% 또는 적어도 90% 만큼 증가시킬 수 있다. 가장 바람직하게 RNA 폴리뉴클레오타이드의 변형은 DNA 헬리카제의 친화도를 95% 이상 증가시킨다. 변형된 RNA에 대한 DNA 헬리카제 결합이 촉진된다는 것은 DNA 헬리카제가 비-변형되거나 또는 변형되지 않은 RNA 폴리뉴클레오타이드와 비교해서 변형된 RNA 폴리뉴클레오타이드와 더 용이하게 결합하는 상황으로서 정의된다. 변형된 RNA 폴리뉴클레오타이드에 대한 DNA 헬리카제 결합이 증가된다는 것은 DNA 헬리카제 결합의 양 또는 수준이, 비-변형되거나 또는 변형되지 않은 RNA 폴리뉴클레오타이드, 즉 본 발명의 변형 방법에 따라서 변형되지 않은 RNA에 대하여 관찰되는 DNA 헬리카제 결합의 양 또는 수준보다 더 크거나 또는 더 많은 것으로서 정의된다. 표적 RNA 폴리뉴클레오타이드에 대한 DNA 헬리카제의 결합 수준은 관련 기술분야의 통상의 기술자에게 공지되고 일상적인 일상적인 방법을 이용하여 용이하게 시험될 수 있다.

[0075] 표적 RNA 폴리뉴클레오타이드는 비-RNA 폴리뉴클레오타이드, 예컨대 폴리뉴클레오타이드 영역 또는 서열 또는 구조물을 포함하도록 변형시킨다. 비-RNA 폴리뉴클레오타이드의 적어도 하나의 뉴클레오타이드는 RNA가 아니다. 따라서 비-RNA 폴리뉴클레오타이드는 리보뉴클레오타이드 또는 RNA 뉴클레오타이드를 포함할 수 있지만, 또한 비-RNA 뉴클레오타이드 또는 서열, 즉 RNA가 아닌 뉴클레오타이드 또는 서열을 포함해야만 한다. 표적 RNA 폴리뉴클레오타이드는 표적 RNA를 비-RNA 폴리뉴클레오타이드 (이는 리보뉴클레오타이드 또는 RNA 뉴클레오타이드를 포함할 수 있거나 또는 포함하지 않을 수 있다)에 부가 또는 부착시킴으로써 변형시켜 본 발명의 구조물을 형성한다. 비-RNA 폴리뉴클레오타이드를 RNA 폴리뉴클레오타이드에 부가 또는 부착하는 것은, 비-RNA 폴리뉴클레오타이드를 부착시키지 않은 변형되지 않은 형태의 RNA 폴리뉴클레오타이드와 DNA 헬리카제 간에 일어나는 상호 작용과 비교해서, DNA 헬리카제와 변형된 RNA 구조물간의 상호 작용이 증가된다는 것을 의미한다. 부가적으로 또는 또 다른 한편으로, 비-RNA 폴리뉴클레오타이드를 RNA 폴리뉴클레오타이드에 부가 또는 부착하는 것은, 비-RNA 폴리뉴클레오타이드를 부착시키지 않은 변형되지 않은 형태의 RNA 폴리뉴클레오타이드에 대한 DNA 헬리카제의 특이성과 비교해서, 변형된 RNA 구조물에 대한 DNA 헬리카제의 특이성이 증가된다는 것을 의미한다. 부가적으로 또는 또 다른 한편으로, 비-RNA 폴리뉴클레오타이드를 RNA 폴리뉴클레오타이드에 부가 또는 부착하는 것은, 비-RNA 폴리뉴클레오타이드를 부착시키지 않은 변형되지 않은 형태의 RNA 폴리뉴클레오타이드와 DNA 헬리카제 간에 일어나는 결합과 비교해서, 변형된 RNA 구조물에 대한 DNA 헬리카제 결합이 촉진되고/되거나 변형된 RNA 구조물에 대한 DNA 헬리카제 결합이 증가된다는 것을 의미한다. 부가적으로 또는 또 다른 한편으로, 비-RNA 폴리뉴클레오타이드를 RNA 폴리뉴클레오타이드에 부가 또는 부착하는 것은, 비-RNA 폴리뉴클레오타이드를 부착시키지 않은 변형되지 않은 형태의 RNA 폴리뉴클레오타이드와 DNA 헬리카제 간에 일어나는 결합과 비교해서, DNA 헬리카제가 변형된 RNA 구조물과 더 효율적이거나 또는 더 강력하게 결합하고 변형된 구조물로부터 이탈될 가능성이 더 적다는 것을 의미한다. 바람직하게 RNA 폴리뉴클레오타이드의 변형은 DNA 헬리카제가 RNA 폴리뉴클레오타이드로부터 결합 해제되는 것을 적어도 10%, 바람직하게 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80% 또는 적어도 90% 만큼 감소시켜 준다. 가장 바람직하게 RNA 폴리뉴클레오타이드의 변형은 DNA 헬리카제가 RNA 폴리뉴클레오타이드로부터 결합 해제되는 것을 95% 이상 감소시켜 준다.

[0076] 비-RNA 폴리뉴클레오타이드

[0077] 표적 RNA 폴리뉴클레오타이드 서열은 비-RNA 폴리뉴클레오타이드를 포함하도록 변형시킨다. 표적 RNA를 비-RNA 폴리뉴클레오타이드에 부착시킨다. 표적 RNA는 바람직하게, 비-RNA 폴리뉴클레오타이드에 공유적으로 부착시킨다. 비-RNA 폴리뉴클레오타이드는 리보뉴클레오타이드가 아닌, 즉 RNA로부터 유래되지 않은 적어도 하나의 뉴클레오타이드를 포함해야만 한다. 비-RNA 폴리뉴클레오타이드는 부가적으로 리보뉴클레오타이드 또는 RNA를 포함할 수 있지만, 이는 또한 적어도 하나의 비-RNA 뉴클레오타이드, 즉 RNA가 아닌 뉴클레오타이드를 포함해야만 한다. 전형적으로 RNA를 포함하는 비-RNA 폴리뉴클레오타이드는 20개 미만의 RNA 뉴클레오타이드, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 또는 19개의 RNA 뉴클레오타이드를 포함한다. 따라서 비-RNA 폴리뉴클레오타이드는, 예를 들어 RNA와 또 다른 폴리뉴클레오타이드, 예컨대 DNA 또는 DNA 유사체를 포함하는 "혼성체" 폴

리뉴클레오티드일 수 있다. 비-RNA는 또한 DNA 스페이서 등을 포함할 수 있다. 통상의 기술자는 본 발명의 변형된 RNA 구축물을 제조하는 데 적합한 것으로 기재된 부착 방법 중 임의의 것이, 비-RNA 폴리뉴클레오티드를 제조하는 데 동등하게 적합하다는 것을 알고 있을 것인데, 여기서는 2가지 이상의 유형의 핵산 서열을 조합할 수 있다.

- [0078] 바람직하게 비-RNA 폴리뉴클레오티드는 리더 서열을 포함한다. 리더 서열은 세공 내로 우선적으로 삽입된다.
- [0079] 바람직하게 표적 RNA 폴리뉴클레오티드는 비-RNA 리더 서열을 RNA에 부착시킴으로써 변형시킨다. 리더 서열은 본 발명의 특징 규명 방법을 촉진시켜 준다. 리더 서열은 세공 내로 우선적으로 삽입됨으로써 세공을 통한 폴리뉴클레오티드의 이동을 촉진시키도록 설계된다. 리더 서열은 또한, 표적 RNA 폴리뉴클레오티드를 다음에 논의되는 바와 같은 하나 이상의 앵커에 연결시키기 위해 사용될 수 있다. 리더 서열은 표적 RNA 폴리뉴클레오티드에 연결될 수 있다.
- [0080] 리더 서열 전형적으로 중합체 영역을 포함한다. 중합체 영역은 바람직하게 음전하를 띤다. 중합체는 바람직하게 폴리뉴클레오티드, 예컨대 DNA, 변형된 폴리뉴클레오티드 (예컨대 탈염기성 DNA), PNA, LNA, 폴레에틸렌 글리콜 (PEG) 또는 폴리펩티드이다. 리더 서열은 바람직하게 단일 가닥 폴리뉴클레오티드를 포함한다.
- [0081] 단일 가닥 리더 서열은 단일 가닥의 DNA, iSpC3, 폴리 dT 섹션 또는 폴리 dC 섹션을 포함할 수 있다. 리더 서열은 바람직하게 하나 이상의 스페이서를 포함한다. 리더 서열은 바람직하게, 다음에 정의된 바와 같은 Y 어댑터의 일부이다.
- [0082] 리더 서열은 임의의 길이일 수 있지만, 전형적으로 그 길이가 10개 내지 200개의 뉴클레오티드이다. 본 발명의 한 실시양태에서 비-RNA 리더 서열은 iSpC3을 포함한다. 가장 바람직하게 비-RNA 리더 서열은 iSpC3이거나 또는 그 길이가 대략 40개의 뉴클레오티드인 C와 A의 반복 서열 [예를 들어, 4x(9C 및 1A)]이다. 리더의 길이는 전형적으로, 본 방법에 사용된 막형단 세공에 좌우된다.
- [0083] 바람직하게 비-RNA 폴리뉴클레오티드는 DNA 헬리카제와 결합할 수 있는 영역 (DNA 헬리카제 결합 부위) 또는 DNA 어댑터를 포함한다. 표적 RNA 폴리뉴클레오티드는 막형단 세공을 통한 표적 RNA의 이동을 제어하는 DNA 헬리카제에 대한 DNA 결합 부위를 포함하도록 변형될 수 있다. 본원에 사용된 바와 같은 용어 "DNA 헬리카제 결합 부위"는, 하나 이상의 DNA 헬리카제가 이와 결합할 수 있기에 충분한 크기/길이의 DNA 또는 DNA 유사체 서열을 포함한다. 상기 결합 부위의 길이는, 이와 결합되어야 하는 헬리카제의 수에 좌우된다. DNA 헬리카제와 결합할 수 있는 영역은 바람직하게 폴리뉴클레오티드, 예컨대 DNA, 변형된 폴리뉴클레오티드 (예컨대 탈염기성 DNA), PNA, LNA, 폴레에틸렌 글리콜 (PEG)이다. 바람직하게 DNA 헬리카제 결합 부위는 단일 가닥의 혼성화되지 않은 영역이다. 이러한 영역은 리더 서열에 상응할 수 있다. 또 다른 한편으로, 상기 영역은 리더 서열과 별개일 수 있다. DNA 헬리카제는 다음에 보다 상세히 논의되는 바와 같이 세공을 통한 RNA 폴리뉴클레오티드의 이동을 제어하는 데 도움을 줄 수 있다.
- [0084] 바람직하게 비-RNA 폴리뉴클레오티드에는, 하나 이상의 헬리카제가 이동될 곳과 반대의 말단에서, DNA 결합 부위에 인접하거나 또는 아주 근접하여 위치될 수 있는 차단 부위 또는 차단성 분자가 추가로 제공된다. 이러한 차단성 분자는 헬리카제가 뒤로 이동하는 것을 방지시켜 주고, 그것이 구축물로부터 미끄러지는 것을 방지시켜 준다.
- [0085] 표적 RNA 서열에 대한 부착을 위한 비-RNA 폴리뉴클레오티드 구축물은 바람직하게 다음을 포함한다: 부분 (i) 바람직하게 세공에 의한 표적 폴리뉴클레오티드의 포획을 제공하는 5개 이상의 하전된 단위의 중합체; 및/또는 부분 (ii) 대략 20개의 뉴클레오티드 길이의 차단성-가닥 혼성화 부위; 및/또는 부분 (iii) 1개 이상의 비-RNA 뉴클레오티드, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 또는 50개의 비-RNA 뉴클레오티드의 DNA-헬리카제 결합 부위; 및/또는 부분 (iv) 1개 이상의 단위, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과 단위,의 자연 화학, 예컨대 본원에 참조로 포함되는 WO2014/135838에 기재된 바와 같은 Sp18; 및/또는 부분 (v) 대략 30개의 뉴클레오티드 길이의 테더 혼성화 부위; 및/또는 부분 (vi) RNA 폴리뉴클레오티드에 대한 비-RNA 폴리뉴클레오티드의 라이게이션을 촉진시켜 주는 서열. 따라서 표적 RNA 서열에 부착시키기 위한 비-RNA 폴리뉴클레오티드의 총 길이는 대략 50개 내지 200개의 뉴클레오티드를 포함할 수 있다.
- [0086] 예를 들어, 한 실시양태에서, 비-RNA 폴리뉴클레오티드는 다음 중 적어도 하나를 포함할 수 있다: (i) 5개 이상의 하전된 단위의 중합체; (ii) 대략 20개의 뉴클레오티드 길이의 차단성-가닥 혼성화 부위; (iii) 1개 이상의 비-RNA 뉴클레오티드, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,



30, 40, 또는 50개의 비-RNA 뉴클레오티드의 DNA-헬리카제 결합 부위; (iv) 1개 이상의 단위, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과 단위,의 자연 화학, 예컨대 본원에 참조로 포함되는 W02014/135838에 기재된 바와 같은 Sp18; (v) 대략 30개의 뉴클레오티드 길이의 테더 혼성화 부위; 및/또는 (vi) 이전의 섹션에 기재된 바와 같이, RNA 폴리뉴클레오티드에 대한 비-RNA 폴리뉴클레오티드의 라이게이션을 촉진시켜 주는 서열.

[0087] 각각의 (i) 내지 (vi)가 다음에 보다 상세히 논의된다:

[0088]           부분 (i)

[0089] 비-RNA 폴리뉴클레오티드의 부분 (i)은 바람직하게, 순 음전하를 갖는 중합체이다. 이러한 중합체는 리더 서열에 대하여 상기 논의된 것들 중 임의의 것일 수 있다. 바람직하게 상기 중합체에는 핵염기가 결여되거나 또는 뉴클레오시드가 결여된다. 상기 중합체는 다음에 논의된 스페이서 중 임의의 것일 수 있다. 이들 기준을 충족시키는 서열의 대표적인 예는 다음과 같다:

[0090] TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT (서열식별번호: 16)

[0091] CCCCCCCCCACCCCCCCCCACCCCCCCCCACCCCCCCCCA (서열식별번호: 46)

[0092] 666

[0093] 777

[0094] 888

[0095] 상기에서, 6 = 1,2-데데옥시 뉴클레오티드 모노포스페이트

[0096] 7 = n-프로필렌 포스페이트 (스페이서 3 기)

[0097] 8 = PEG3 포스페이트 (스페이서 9 기).

[0098]           부분 (ii)

[0099] 비-RNA 폴리뉴클레오티드의 부분 (ii)는 차단성 가닥의 혼성화를 허용해 주고 DNA 또는 RNA 또는 유사체, 예컨대 PNA, GNA, TNA, BNA, LNA 또는 모르폴리노를 포함할 수 있는 임의의 폴리뉴클레오티드 서열이다. 이들 기준을 충족시키는 서열의 대표적인 예는 다음과 같다:

[0100] ACTCGCAGATCATTACGATC (서열식별번호: 47)

[0101] rArCrUrCrGrCrArGrArUrCrArUrUrArCrGARUrC (서열식별번호: 48)

[0102] 서열식별번호: 47의 서열을 수반한 PNA

[0103] 서열식별번호: 47의 서열을 수반한 BNA

[0104] CGATTGACTAAGCTATACGC (서열식별번호: 49)

[0105] rCrGrArUrUrGrArCrUrArArGrCrUrArUrArCrGrC (서열식별번호: 50)

[0106] 서열식별번호: 49의 서열을 수반한 PNA

[0107] 서열식별번호: 49의 서열을 수반한 BNA.

[0108]           부분 (iii)

[0109] 비-RNA 폴리뉴클레오티드의 부분 (iii)은 사용되는 특이적 DNA 헬리카제의 결합을 허용하기에 충분한 길이여야 하고, 유사체, 예컨대 PNA, GNA, TNA, BNA, LNA 또는 모르폴리노를 포함한 비-RNA 뉴클레오티드로 구성되어야 한다. 본 발명의 바람직한 실시양태에서, 부분 (iii)은 DNA이다. 이들 기준을 충족시키는 서열의 대표적인 예는 다음과 같다:

[0110] TTTTTTTTTT (서열식별번호: 51)

[0111] TTTTTTTTTTTTTTTTTT (서열식별번호: 52)

[0112] CCCCCCCCCA (서열식별번호: 31)

- [0113] CCCCCCCCCCCCCCCCCCA (서열식별번호: 53)
- [0114] XXXXXXXXX (여기서, X는 A, T, G 또는 C로부터 독립적으로 선택된다).
- [0115] **부분 (iv)**
- [0116] 비-RNA 폴리뉴클레오타이드의 부분 (iv)은 세공의 힘 없이도 DNA 헬리카제의 ATP-매개된 전위를 방지하거나 또는 느리게 해야 한다. 이들 기준을 충족시키는 서열의 대표적인 예는 다음과 같다:
- [0117] rArArArArArArArArA (서열식별번호: 54)
- [0118] rUrCrCrArUrArCrGrArA (서열식별번호: 55)
- [0119] 9999
- [0120] 상기에서 9 = PEG6 포스페이트 (스페이서 9 (iSp9) 기).
- [0121] **부분 (v)**
- [0122] 비-RNA 폴리뉴클레오타이드의 부분 (v)은 안정한 혼성체를 형성하기에 충분히 높은 TM을 갖는 테더링 올리고를 혼성화시킬 수 있어야 한다. 이들 기준을 충족시키는 서열의 대표적인 예는 다음과 같다:
- [0123] AACTACTAGGATCATCGATGTATCTGCTCA (서열식별번호: 56)
- [0124] AGCTTAACATACGATACTCTTAGCTAACCA (서열식별번호: 57)
- [0125] rArArCrUrArCrUrArGrGrArUrCrArUrCrGrArUrGrUrArUrCrUrGrCrUrCrA (서열식별번호: 58)
- [0126] rArGrCrUrUrArArCrArUrArCrGrArUrArCrUrUrArGrCrUrArArCrCrA (서열식별번호: 59)
- [0127] 서열식별번호: 56의 서열을 수반한 PNA
- [0128] 서열식별번호: 57의 서열을 수반한 PNA.
- [0129] **부분 (vi)**
- [0130] 비-RNA 폴리뉴클레오타이드의 부분 (vi)은 RNA 폴리뉴클레오타이드에 대한 비-RNA 폴리뉴클레오타이드의 라이게이션을 촉진시켜야 한다. 이들 기준을 충족시키는 서열의 대표적인 예는 다음과 같다:
- [0131] ACTCTGAACC (서열식별번호: 60)
- [0132] ACTCTrGrArArCrC (SEQ ID NO: 61)
- [0133] GCACAATGAT (서열식별번호: 62)
- [0134] GCACArArTrGrArT (서열식별번호: 63).
- [0135] (i) 내지 (vi)의 임의의 조합이 본 발명에 따라서 만들어질 수 있다: 바람직하게 비-RNA 폴리뉴클레오타이드는 {i}, {ii}, {iv}, {v}, {i,ii}, {i,iv}, {i,v}, {ii,iv}, {ii,v}, {iv,v}, {i,ii,iv}, {i,ii,v}, {i,iv,v}, {ii,iv,v}, {i,ii,iv,v}와 조합하여 (iii) 및 (vi)를 포함한다.
- [0136] 부분 (i) 내지 (vi) 각각에 대해 상기 제시된 상이한 대표적인 예들 중 임의의 것을 상호 교환적으로 사용하여, 표적 RNA 서열에 부착하기 위한 비-RNA 폴리뉴클레오타이드를 형성할 수 있다.
- [0137] 비-RNA 폴리뉴클레오타이드의 7개의 상이한 대표적인 예는 다음과 같다:
- [0138] 구조물 1: (40x SpC3)(ACTCGAGATCATTACGATC)(10x dT)(4x Sp18)  
(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)
- [0139] 즉, (40x SpC3)(서열식별번호: 47)(서열식별번호: 51)(4x Sp18)(서열식별번호: 56)(서열식별번호: 60).
- [0140] 구조물 2: (40x SpC3)(ACTCGAGATCATTACGATC)(20x dT)(4x Sp18)  
(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)
- [0141] 즉, (40x SpC3)(서열식별번호: 47)(서열식별번호: 52)(4x Sp18)(서열식별번호: 56)(서열식별번호: 60).
- [0142] 구조물 3: (40x rU)(ACTCGAGATCATTACGATC)(10x dT)(4x Sp18)(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)

- [0143] 즉, (40x rU)(서열식별번호: 47)(서열식별번호: 51)(4x Sp18)(서열식별번호: 56)(서열식별번호: 60).
- [0144] 구축물 4: (40x SpC3)(ACTCGCAGATCATTACGATC)(10x dT)(4x Sp18)(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTrGrArArCrC)
- [0145] 즉, (40x SpC3)(서열식별번호: 47)(서열식별번호: 51)(4x Sp18)(서열식별번호: 56)(서열식별번호: 61).
- [0146] 구축물 5: (40x SpC3)(ACTCGCAGATCATTACGATC)(10x dT)(rArArArCrUrArCrGrCrU)(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)
- [0147] 즉, (40x SpC3)(서열식별번호: 47)(서열식별번호: 51)(서열식별번호: 64)(서열식별번호: 56)(서열식별번호: 60).
- [0148] 구축물 6: (40x SpC3)(ACTCGCAGATCATTACGATC)(10x dT)(4x Sp18)(AGCTTAACATACGATACTCTTAGCTAACCA)(ACTCTGAACC)
- [0149] 즉, (40x SpC3)(서열식별번호: 47)(서열식별번호: 51)(4x Sp18)(서열식별번호: 57)(서열식별번호: 60).
- [0150] 구축물 7: (40x SpC3)(서열식별번호: 47의 서열을 수반한 PNA)(10x dT)(4x Sp18)(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)
- [0151] 즉, (40x SpC3)(서열식별번호: 47의 서열을 수반한 PNA)(서열식별번호: 51)(4x Sp18)(서열식별번호: 56)(서열식별번호: 60).
- [0152] 부착
- [0153] 표적 RNA 폴리뉴클레오타이드를 비-RNA 폴리뉴클레오타이드에 부착시키고/시키거나 비-RNA 폴리뉴클레오타이드를 표적 RNA 폴리뉴클레오타이드에 부착시켜 변형된 RNA 폴리뉴클레오타이드를 형성시킨다. 비-RNA 폴리뉴클레오타이드가 리보뉴클레오타이드 또는 RNA 서열을 포함하는 경우, 비-RNA 폴리뉴클레오타이드는 이러한 비-RNA 폴리뉴클레오타이드 내에 포함되는 리보뉴클레오타이드 또는 RNA 서열을 통하여 표적 RNA 폴리뉴클레오타이드에 부착될 수 있다.
- [0154] 본 방법은 단계 (a)에서 변형된 RNA 폴리뉴클레오타이드 및 DNA 헬리카제 효소를 제공하는 것을 포함한다. 상기 방법은 단계 (a) 전에, RNA 폴리뉴클레오타이드를 비-RNA 폴리뉴클레오타이드에 부착시키고/시키거나 비-RNA 폴리뉴클레오타이드를 RNA 폴리뉴클레오타이드에 부착시켜 변형된 RNA 폴리뉴클레오타이드를 형성시키는 것을 추가로 포함할 수 있다.
- [0155] 표적 RNA 폴리뉴클레오타이드와 비-RNA 폴리뉴클레오타이드 (이는 DNA 또는 DNA 유사체일 수 있다)는 임의의 방법 또는 관련 기술분야에 공지된 방법을 이용하여 임의의 방식으로 서로 부착시킬 수 있다. 바람직하게 표적 RNA 폴리뉴클레오타이드와 비-RNA 폴리뉴클레오타이드는 다음에 기재된 각종 방법 중 하나 이상을 이용하여 부착시킨다. 표적 RNA 폴리뉴클레오타이드는, 예를 들어 공유 결합에 의해 비-RNA 폴리뉴클레오타이드에 화학적으로 부착시킬 수 있다. 표적 RNA 폴리뉴클레오타이드는 화학적 또는 효소적 라이게이션에 의해 비-RNA 폴리뉴클레오타이드에 부착시킬 수 있다. 표적 RNA 폴리뉴클레오타이드는 혼성화 및/또는 합성 방법에 의해 비-RNA 폴리뉴클레오타이드에 부착시킬 수 있다. RNA 폴리뉴클레오타이드는 토포이소머라제를 사용하여 비-RNA 폴리뉴클레오타이드에 부착시킬 수 있다. RNA 폴리뉴클레오타이드는 2개 이상, 예컨대 2개 또는 3개의 지점에서 비-RNA 폴리뉴클레오타이드에 부착시킬 수 있다. 부착 방법은 1개, 2개, 3개, 4개, 5개 또는 그 초과와 상이한 부착 방법을 포함할 수 있다. 다음에 기재된 부착 방법의 임의의 조합이 본 발명에 따라서 사용될 수 있다.
- [0156] RNA 폴리뉴클레오타이드와 비-RNA 폴리뉴클레오타이드 (이는 리보뉴클레오타이드 또는 RNA를 포함할 수 있거나 또는 포함하지 않을 수 있다)는 별도로 생성한 다음, 함께 부착시킬 수 있다. 이들 두 성분은 임의의 입체 배치로 부착시킬 수 있다. 예를 들어, 이들은 그들의 단말 (즉, 5' 또는 3')을 통하여 부착시킬 수 있다. 적합한 입체 배치는 비-RNA 폴리뉴클레오타이드의 3' 말단에 부착되는 RNA 폴리뉴클레오타이드의 5' 말단 및 그 반대의 경우를 포함하지만, 이에 제한되지 않는다. 또 다른 한편으로, 상기 두 성분은 그들의 서열 내의 뉴클레오타이드를 통하여 부착시킬 수 있다.
- [0157] 폴리뉴클레오타이드는 그의 자연 발생적 뉴클레오타이드를 통하여 부착시킬 수 있다. 자연 발생적 뉴클레오타이드는 부착을 촉진시키도록 변형될 수 있다. 예를 들어, 자연 발생적 핵산은, 예를 들어 mRNA 캡에 대한 트리메틸 구아노신 신타제에 의해 변형될 수 있다. 다른 적합한 변형은 관련 기술분야에 공지되어 있다. 변형은 치환에 의해 도입될 수 있다. RNA 폴리뉴클레오타이드는 링커 분자를 통하여 비-RNA 폴리뉴클레오타이드에 부착시킬 수 있다. RNA 폴리뉴클레오타이드는 하나 이상, 예컨대 2개 또는 3개의 링커를 이용하여 비-RNA 폴리뉴클레오타이드에

부착시킬 수 있다. 링커는 필요한 거리에 걸쳐 연장되는 임의의 분자를 포함할 수 있다. 링커는 그 길이가 하나의 탄소 (포스젠-유형 링커)에서부터 많은 옹스트롬까지 다양할 수 있다. 링커로서 사용하기에 적합한 선형 분자의 예는 폴리에틸렌 글리콜 (PEG), 폴리펩티드, 폴리사카라이드, 데옥시리보핵산 (DNA), 펩티드 핵산 (PNA), 트레오스 핵산 (TNA), 글리세롤 핵산 (GNA), 포화 및 불포화 탄화수소, 폴리아미드를 포함하지만, 이에 제한되지 않는다. 이들 링커는 불활성 또는 반응성일 수 있는데, 특히 이들은 규정된 위치에서 화학적으로 절단 가능할 수 있거나, 또는 형광단 또는 리간드를 이용하여 그 자체를 변형시킬 수 있다. 링커는 바람직하게, 디티오프레이틀 (DTT)에 대해 저항성이다. 바람직한 가요성 펩티드 링커는 2 내지 20개, 예컨대 4, 6, 8, 10 또는 16개의 세린 및/또는 글리신 아미노산의 연장물이다. 보다 바람직한 가요성 링커는 (SG)<sub>1</sub>, (SG)<sub>2</sub>, (SG)<sub>3</sub>, (SG)<sub>4</sub>, (SG)<sub>5</sub>, (SG)<sub>8</sub>, (SG)<sub>10</sub>, (SG)<sub>15</sub> 또는 (SG)<sub>20</sub> (여기서, S는 세린이고 G는 글리신이다)을 포함한다. 바람직한 강직 링커는 2 내지 30개, 예컨대 4, 6, 8, 16 또는 24개의 프롤린 아미노산의 연장물이다. 보다 바람직한 강직 링커는 (P)<sub>12</sub> (여기서, P는 프롤린이다)를 포함한다.

[0158] RNA 폴리뉴클레오타이드는 하나 이상의 화학적 가교 결합제 또는 하나 이상의 펩티드 링커를 이용하여 비-RNA 폴리뉴클레오타이드에 부착될 수 있다. 적합한 화학적 가교 결합제는 관련 기술분야에 널리 공지되어 있다. 적합한 화학적 가교 결합제는 다음 관능기를 포함한 것을 포함하지만, 그에 제한되지 않는다: 말레이미드, 활성 에스테르, 숙신이미드, 아지드, 알킨 (예컨대, 디벤조시클로옥텐올 (DIBO 또는 DBCO), 디플루오로 시클로알킨 및 선형 알킨), 포스핀 [예컨대, 흔적없는 및 흔적이 없지 않은 슈타우딩거(Staudinger) 라이게이션에 사용된 것], 할로아세틸 (예컨대, 아이오도아세트아미드), 포스젠 유형 시약, 술폰일 클로라이드 시약, 이소티오시아네이트, 아실 할라이드, 히드라진, 디술피드, 비닐 술폰, 아지리딘 및 광반응성 시약 (예컨대, 아릴 아지드, 디아지리딘).

[0159] RNA 폴리뉴클레오타이드와 비-RNA 폴리뉴클레오타이드 간의 반응은 자발적일 수 있거나, 또는 아지드와 선형 알킨을 연결시키기 위한 외부 시약, 예컨대 Cu(I)이 필요할 수 있다.

[0160] 바람직한 가교 결합제는 2,5-디옥소피롤리딘-1-일 3-(피리딘-2-일디술팜닐)프로파노에이트, 2,5-디옥소피롤리딘-1-일 4-(피리딘-2-일디술팜닐)부타노에이트, 2,5-디옥소피롤리딘-1-일 8-(피리딘-2-일디술팜닐)옥타노에이트, 디-말레이미드 PEG 1k, 디-말레이미드 PEG 3.4k, 디-말레이미드 PEG 5k, 디-말레이미드 PEG 10k, 비스(말레이미도)에탄 (BMOE), 비스-말레이미도핵산 (BMH), 1,4-비스-말레이미도부탄 (BMB), 1,4 비스-말레이미드-2,3-디히드록시부탄 (BMDDB), BM[PEO]2 (1,8-비스-말레이미도디에틸렌글리콜), BM[PEO]3 (1,11-비스-말레이미도트리에틸렌 글리콜), 트리스[2-말레이미도에틸]아민 (TMEA), DTME 디티오비스말레이미도에탄, 비스-말레이미드 PEG3, 비스-말레이미드 PEG11, DBCO-말레이미드, DBCO-PEG4-말레이미드, DBCO-PEG4-NH<sub>2</sub>, DBCO-PEG4-NHS, DBCO-NHS, DBCO-PEG-DBCO 2.8kDa, DBCO-PEG-DBCO 4.0kDa, DBCO-15개 원자-DBCO, DBCO-26개 원자-DBCO, DBCO-35개 원자-DBCO, DBCO-PEG4-S-S-PEG3-비오틴, DBCO-S-S-PEG3-비오틴 및 DBCO-S-S-PEG11-비오틴을 포함한다. 가장 바람직한 가교 결합제는 숙신이미드 3-(2-피리딜디티오)프로피오네이트 (SPDP) 및 말레이미드-PEG(2kDa)-말레이미드 (알파,오메가-비스-말레이미도 폴리(에틸렌 글리콜))이다.

[0161] 링커는 표지될 수 있다. 적합한 표지는 형광성 분자 [예컨대, Cy3 또는 알렉사플루오르(AlexaFluor®) 555], 방사성 동위원소, 예를 들어 <sup>125</sup>I, <sup>35</sup>S, 효소, 항체, 항원, 폴리뉴클레오타이드 및 리간드, 예컨대 비오틴을 포함하지만, 이에 제한되지 않는다. 이러한 표지는 링커의 양을 정량화시킬 수 있다. 표지는 또한, 절단 가능한 정제 태그, 예컨대 비오틴, 또는 확인 방법에서 그 모습을 나타내는 특이적 서열일 수 있다.

[0162] 상당히 과량의 RNA 폴리뉴클레오타이드 및/또는 비-RNA 폴리뉴클레오타이드에서 링커의 농도를 유지함으로써 RNA 폴리뉴클레오타이드 또는 비-RNA 폴리뉴클레오타이드 자체의 가교 결합을 방지시킬 수 있다. 또 다른 한편으로, 2개의 링커가 사용되는 "자물쇠와 열쇠" 배열을 사용할 수 있다. 각 링커의 단 하나의 말단이 함께 반응하여 더 긴 링커를 형성할 수 있고, 링커의 다른 말단은 각각 구축물 (즉, RNA 폴리뉴클레오타이드 또는 비-RNA 폴리뉴클레오타이드)의 상이한 부분과 반응한다.

[0163] 비-RNA 폴리뉴클레오타이드에 대한 RNA 폴리뉴클레오타이드의 부착은 영구적이거나 또는 안정적일 수 있다 (즉, RNA 폴리뉴클레오타이드는 본 발명의 방법에서 비-RNA 폴리뉴클레오타이드로부터 분리되지 않는다). 바람직한 영구적 또는 안정적 부착은 공유적 부착이다.

[0164] 또 다른 한편으로, 부착은 일시적인데, 즉 RNA 폴리뉴클레오타이드가 비-RNA 폴리뉴클레오타이드로부터 분리될 수 있다. 관련 기술분야의 통상의 기술자는, 본원에 기재된 방법 중 임의의 것이 비-RNA 폴리뉴클레오타이드를 표적 RNA에 부착하는 데 적합하고, 또한 상기 언급된 바와 같이 2가지 이상의 유형의 핵산, 예를 들어 DNA 및 RNA의

혼성체일 수 있는 비-RNA 폴리뉴클레오티드 자체를 구축하는 데 동등하게 적합하다는 것을 이해할 것이다.

[0165] *클릭 화학*

[0166] 표적 RNA 폴리뉴클레오티드는 비-RNA 폴리뉴클레오티드에 공유적으로 부착시킬 수 있다. 비-RNA 폴리뉴클레오티드는 미리 결합된 DNA 헬리카제 효소를 포함할 수 있거나 또는 포함하지 않을 수 있다. 바람직한 실시양태에서, RNA 폴리뉴클레오티드와 비-RNA 폴리뉴클레오티드, 예를 들어, DNA 리더 서열 간의 공유 결합은 구리 무함유 클릭 화학 또는 구리 촉매된 클릭 화학을 이용하여 이루어질 수 있다. 클릭 화학은 다양한 빌딩 블록 간의 공유 결합을 창출하기 위한 그의 바람직한 특성 및 범위 때문에 이러한 적용 분야에 사용되어 왔다. 예를 들어, 이는 신속하고, 깨끗하며 독성이 없으므로, 해를 끼치지 않는 부산물만을 생성시킨다. 클릭 화학은 소규모 및 대규모 적용 분야 모두에서 믿을 수 있게 작동하는 강력하고 선택적인 모듈러 빌딩 블록의 확장 세트를 설명하기 위해 콜브 등(Kolb et al.)에 의해 2001년도에 처음 도입된 용어이다 (Kolb HC, Finn, MG, Sharpless KB, Click chemistry: diverse chemical function from a few good reactions, Angew. Chem. Int. Ed. 40 (2001) 2004-2021). 그들은 클릭 화학에 대한 일련의 엄격한 기준을 다음과 같이 정의하였다: "그 반응은 모듈화되고, 범위가 넓어야 하며, 매우 높은 수율을 제공해야만 하고, 마이크로마토그래피 방법에 의해 제거될 수 있는 해를 끼치지 않는 부산물만을 생성해야 하며 입체 특이적이어야 한다 (그러나 반드시 예난티오선택적일 필요는 없다). 요구되는 공정 특징은 간단한 반응 조건 (이상적으로는 공정이 산소와 물에 민감하지 않아야 한다), 용이하게 입수 가능한 출발 물질 및 시약, 용매를 사용하지 않거나 또는 양성 (예컨대 물)이거나 쉽게 제거되는 용매의 사용, 및 간단한 생성물 분리를 포함한다. 필요한 경우 정제는 마이크로마토그래피 방법, 예컨대 결정화 또는 증류에 의해서 수행해야만 하며, 생성물은 생리적 조건하에 안정적이어야만 한다".

[0167] 클릭 화학의 적합한 예는 다음을 포함하지만, 그에 제한되지 않는다:

[0168] (a) 아지드가 압박하에, 예를 들어 시클로옥탄 환에서 알킨과 반응하는 1,3 양극성 부가환화 반응의 구리-무함유 변이체;

[0169] (b) 하나의 링커 상의 산소 친핵체와 다른 링커 상의 에폭시드 또는 아지리딘 반응성 모이어티(moiety)의 반응; 및

[0170] (c) 알킨 모이어티를 아릴 포스핀으로 대체시켜, 아지드와의 특이적 반응을 초래하여 아마이드 결합을 수득할 수 있는 슈타우딩거 라이게이션.

[0171] 바람직하게 클릭 화학 반응은 알킨과 아지드 간의 Cu (I) 촉매된 1,3 양극성 부가 환화 반응이다. 바람직한 실시양태에서, 제1 기는 아지드 기이고 제2 기는 알킨 기이다. 아지드 및 알킨 기를 바람직한 위치에 혼입시킨 핵산 염기는 이미 합성되었다 (예를 들어, 문헌 [Kocalka P, El-Sagheer AH, Brown T, Rapid and efficient DNA strand cross-linking by click chemistry, ChemBiochem. 2008. 9(8):1280-5]). 알킨 기는 베리 어소시에이트 (Berry Associates; 미국 미시간주)로부터 시판되고 있고, 아지드 기는 ATD바이오 또는 IDT 바이오에 의해 합성된다.

[0172] 링커의 핵산 영역 내의 뉴클레오티드가, 공유 결합을 형성할 수 있는 기를 포함하도록 변형되는 경우, 이와 같이 변형된 뉴클레오티드는 바람직하게, 링크를 달성하기 위해 하나의 뉴클레오티드에 의해 서로 오프셋된다. 이는 톰 브라운(Tom Brown)의 출간된 저작물에 따른다 (Kocalka et al. (2008) ChemBiochem 9 8 1280-1285).

[0173] RNA 전사체가 형성될 때 클릭-반응성 염기를 표적 RNA 폴리뉴클레오티드에 부가할 수 있다. 또 다른 한편으로 하이퍼메틸라제 효소 (캡핑된 mRNA의 경우)에 의해 클릭 기를 표적 RNA 폴리뉴클레오티드에 부가할 수 있다.

[0174] 바람직하게 반응성 기는 아지드 및 핵시닐 기, 예컨대 3아지드N 및 5'-핵시닐-G이다. 바람직하게 아지드 기는, 바람직하게 DNA인 비-RNA 폴리뉴클레오티드 (이는 리보뉴클레오티드 또는 RNA 서열을 포함할 수 있거나 포함하지 않을 수 있다)에 부착시키고, 핵시닐 기는 표적 RNA 폴리뉴클레오티드에 부착시킨다.

[0175] 실시예 3 및 5는 비-RNA 폴리뉴클레오티드 (예를 들어 DNA를 포함함)를 표적 RNA 폴리뉴클레오티드에 연결시키기 위하여 클릭 반응을 사용하는 것을 예시한다. 실시예 5에는 2차원 (2D) RNA-cDNA 구축물을 수득하기 위하여 브릿징 모이어티를 사용하는 것이 추가로 기재되어 있다. 도 15에 예시되는 실시예 5의 구축물을, 브릿징 모이어티를 포함하지 않는 실시예 3의 구축물 (도 10에 예시된 1D 구축물)과 비교할 수 있다.

[0176] *라이게이션*

[0177] 표적 RNA 폴리뉴클레오티드를 비-RNA 폴리뉴클레오티드 (이러한 RNA 폴리뉴클레오티드는 리보뉴클레오티드 또는



RNA 서열을 포함할 수 있거나 포함하지 않을 수 있다)에 라이게이션할 수 있다. 라이게이션은 가장 통상적으로 효소의 작용을 통하여 또는 화학적 수단에 의해 두 핵산 단편을 연결하는 것이다. 하나의 RNA 또는 DNA 말단의 3'-히드록실과 다른 RNA 또는 DNA 말단의 5'-포스포릴 간에 포스포디에스테르 결합을 형성함으로써 RNA 단편의 말단과 DNA 단편의 말단을 함께 연결시킨다. 보조인자가 일반적으로 이러한 반응에 관여하고, 이는 통상적으로 ATP 또는  $NAD^+$ 이다. RNA 또는 비-RNA 폴리뉴클레오타이드, 예컨대 DNA, PNA, 글리세롤 핵산 (GNA), 트레오스 핵산 (TNA) 또는 고정 핵산 (LNA)의 스플린트가 라이게이션 반응에 사용되어, 이러한 스플린트와의 혼성화에 의해 RNA 폴리뉴클레오타이드와 비-RNA 폴리뉴클레오타이드를 서로 인접하게 유지시킴으로써 라이게이션을 촉진시킬 수 있다. RNA 폴리뉴클레오타이드에 부착될 비-RNA 폴리뉴클레오타이드는 미리 결합된 DNA 헬리카제 효소를 포함할 수 있거나 또는 포함하지 않을 수 있다.

[0178] 비-RNA 폴리뉴클레오타이드는 RNA 폴리뉴클레오타이드의 어느 한 말단, 즉 5' 또는 3' 말단에 라이게이션될 수 있다. 비-RNA 폴리뉴클레오타이드는 표적 RNA 폴리뉴클레오타이드의 양 말단에 라이게이션될 수 있다. 바람직하게 비-RNA 폴리뉴클레오타이드는 표적 RNA 폴리뉴클레오타이드의 5' 말단에 라이게이션된다. 비-RNA 폴리뉴클레오타이드는 관련 기술분야에 공지된 임의의 방법을 이용하여 RNA 폴리뉴클레오타이드에 라이게이션될 수 있다. 하나 이상의 비-RNA 폴리뉴클레오타이드는 리가제, 예컨대 T4 DNA 리가제, 이. 콜라이 (*E. coli*) DNA 리가제, Taq DNA 리가제, Tma DNA 리가제, 9° N DNA 리가제, T4 폴리머라제 I, T4 폴리머라제 2, 열안정성 5' App DNA/RNA 리가제, 스플린트R, circ 리가제, T4 RNA 리가제 1 또는 T4 RNA 리가제 2를 사용하여 라이게이션될 수 있다. 하나 이상의 비-RNA 폴리뉴클레오타이드는 ATP의 부재하에 또는 ATP 대신 감마-S-ATP (ATP  $\gamma$  S)를 사용하여 RNA 폴리뉴클레오타이드에 라이게이션될 수 있다 (또는 그 반대의 경우도 마찬가지이다).

[0179] 본 방법은 바람직하게, 상기 방법 조건으로부터 리가제를 제거하는 것을 추가로 포함한다.

[0180] 실시예 2에는 T4 DNA 리가제를 사용하여 RNA 폴리뉴클레오타이드 DNA에 라이게이션하는 것이 예시되어 있다.

[0181] **합성 방법**

[0182] 올리고뉴클레오타이드 또는 프라이머를 표적 RNA 폴리뉴클레오타이드의 임의의 영역과 혼성화시키는 데 사용할 수 있는데, 이는 DNA 합성을 위한 출발 지점으로서 작용한다. 올리고뉴클레오타이드 또는 프라이머는 미리 결합된 DNA 헬리카제 효소를 포함할 수 있거나 또는 포함하지 않을 수 있다.

[0183] 진핵 RNA는 전형적으로 폴리A 테일, 즉 연속되는 아데노신 모노포스페이트의 연장물을 포함한다. 폴리A 테일은 전형적으로 RNA의 3' 말단에 있다. 폴리 A 폴리머라제 또는 말단 트랜스퍼라제 효소를 활용하여, 필요한 경우 폴리(dA) 테일을 원핵 RNA 가닥의 3' 말단 상으로 부가할 수 있다. 프라이머를 표적 RNA의 폴리A 테일과 혼성화시키고, 이를 합성을 위한 출발 지점으로서 사용할 수 있다. 프라이머는 바람직하게 폴리T 영역, 즉 티민에 기반한 뉴클레오타이드만을 함유하는 영역, 또는 폴리U 영역, 즉 우라실에 기반한 뉴클레오타이드만을 함유하는 영역을 포함한다. 폴리U 영역은 UMP 또는 dUMP를 함유할 수 있다. 폴리U 영역은 임의의 길이, 예컨대 적어도 10개, 적어도 15개, 적어도 20개, 적어도 25개 또는 그 초과일 수 있다.

[0184] 본 발명의 한 실시양태에서 비-RNA 폴리뉴클레오타이드는 폴리(dA) 영역에서 RNA 가닥과 혼성화되는 리더 서열을 수반한 DNA 프라이머를 포함한다. 하나 이상의 DNA 헬리카제 효소는 RNA 가닥과 혼성화되어야 하는 DNA 프라이머와 미리 결합될 수 있다. 또 다른 한편으로 RNA와 혼성화되어야 하는 DNA 프라이머는 미리 결합된 DNA 헬리카제 효소를 함유하지 않거나 또는 포함하지 않는다. DNA 프라이머로부터 RNA 가닥의 역전사가 발생할 수 있고, 이어서 브릿징 모이어티, 예컨대 DNA 헤어핀이 이중 가닥 DNA/RNA에 라이게이션될 수 있다. 이러한 브릿징 모이어티, 예를 들어 헤어핀 루프 어댑터를 본 발명의 변형된 구축물 또는 임의의 이중 가닥 표적 RNA 폴리뉴클레오타이드 (RNA/RNA 또는 RNA/DNA)에 부가할 수 있다. 브릿징 모이어티는 단일 가닥으로의 연결에 의해 주형과 보체 가닥 둘 다의 연속되는 서열 분석을 허용한다. 바람직하게 리더 서열을 함유하는 어댑터 (예를 들어, Y 어댑터)는 RNA의 한 말단에 부착시키고, 브릿징 모이어티 어댑터는 다른 말단에 부착시킨다. 리더 서열은 나노 세공 내로 우선적으로 삽입되고, 두 가닥을 연결하는 (이는 RNA/RNA 또는 RNA/DNA일 수 있다) 브릿징 모이어티는, 폴리뉴클레오타이드가 분해되고 양 가닥 (브릿징 모이어티를 통하여 연결됨)이 세공을 통해 이동함에 따라 양 가닥의 특징을 규명할 수 있게 해준다. 이는 이중 가닥 폴리뉴클레오타이드로부터 취득된 정보의 양을 2배로 해주기 때문에 유리하다. 더욱이, 두 가닥 내의 서열은 상보적이기 때문에, 이러한 두 가닥으로부터의 정보는 그 정보에 따라 조합될 수 있다. 이러한 기전은 더 높은 신뢰도의 관찰을 제공하는 교정 능력을 제공한다.

[0185] 또 다른 한편으로 브릿징 모이어티는 보체의 합성 이전에 표적 RNA 폴리뉴클레오타이드에 부가할 수 있고, 다음에



보다 상세히 논의되는 바와 같이, 보체 합성을 위한 프라이머로서 사용될 수 있다.

- [0186] 본 발명의 한 실시양태에서, 이중 가닥 RNA 폴리뉴클레오티드 또는 RNA/DNA 이중체의 가닥들 (예를 들어, RNA 및 cDNA)은 브릿징 모이어티를 이용하여 연결시킨다. 이때 본 발명에 따르는 표적 RNA 폴리뉴클레오티드의 특징을 규명하는 방법은 바람직하게, (i) 표적 RNA 폴리뉴클레오티드를 포함하는 연결된 구축물 (여기서, 상기 RNA 폴리뉴클레오티드는 비-RNA 폴리뉴클레오티드를 포함하도록 변형시킨다) 및 (ii) DNA 헬리카제 효소를 막형단 세공과 접촉시켜, 표적 RNA가 세공을 통하여 이동하도록 하는 것을 포함한다. 상기 방법은 바람직하게, 표적 RNA가 세공에 대하여 이동함에 따라 하나 이상의 측정치를 취함으로써 (여기서, 이러한 측정치는 상보적 폴리뉴클레오티드 (RNA 또는 cDNA) 및 표적 RNA의 하나 이상의 특징을 나타낸다), 표적 이중 가닥 폴리뉴클레오티드의 특징을 규명하는 것을 포함한다.
- [0187] 이러한 방식으로 표적 dsRNA 또는 RNA/DNA 이중체의 양 가닥을 연결하고 심문하는 것은 그 특징 규명의 효율과 정확도를 증가시켜 준다.
- [0188] 브릿징 모이어티는 표적 dsRNA 폴리뉴클레오티드 또는 RNA/DNA 이중체의 두 가닥을 연결시킬 수 있다. 브릿징 모이어티는 전형적으로, 표적 dsRNA 폴리뉴클레오티드 또는 RNA/DNA 이중체의 두 가닥을 공유적으로 연결시킨다. 브릿징 모이어티는 표적 dsRNA 폴리뉴클레오티드 또는 RNA/DNA 이중체의 두 가닥을 연결시킬 수 있는 어떠한 것일 수도 있는데, 단 브릿징 모이어티는 막형단 세공을 통한 RNA 폴리뉴클레오티드의 이동을 방해하지 않아야 한다.
- [0189] 브릿징 모이어티는 관련 기술분야에 공지된 임의의 적합한 수단에 의해 표적 폴리뉴클레오티드에 연결될 수 있다. 브릿징 모이어티는 별도로 합성될 수 있고, RNA 표적 폴리뉴클레오티드에 화학적으로 부착되거나 또는 효소적으로 라이게이션될 수 있다. 또 다른 한편으로, 브릿징 모이어티는 표적 폴리뉴클레오티드의 프로세싱에서 생성될 수 있다.
- [0190] 브릿징 모이어티는 표적 폴리뉴클레오티드의 한 말단에서 또는 그 근처에서 표적 폴리뉴클레오티드에 연결된다. 브릿징 모이어티는 바람직하게, 표적 폴리뉴클레오티드의 말단의 10개 뉴클레오티드 이내에 표적 폴리뉴클레오티드에 연결된다.
- [0191] 적합한 브릿징 모이어티는 중합체성 링커, 화학적 링커, 폴리뉴클레오티드 또는 폴리펩티드를 포함하지만, 이에 제한되지 않는다. 바람직하게, 브릿징 모이어티는 DNA, RNA, 변형된 DNA (예컨대 탈염기성 DNA), RNA, PNA, LNA 또는 PEG를 포함한다. 브릿징 모이어티는 보다 바람직하게 DNA 또는 RNA이다.
- [0192] 브릿징 모이어티는 가장 바람직하게 헤어핀 루프 또는 헤어핀 루프 어댑터이다. 적합한 헤어핀 어댑터는 관련 기술분야에 공지된 방법을 이용하여 설계할 수 있다. 헤어핀 루프는 임의의 길이일 수 있다. 헤어핀 루프는 그 길이가 전형적으로, 110개 이하의 뉴클레오티드, 예컨대 100개 이하의 뉴클레오티드, 90개 이하의 뉴클레오티드, 80개 이하의 뉴클레오티드, 70개 이하의 뉴클레오티드, 60개 이하의 뉴클레오티드, 50개 이하의 뉴클레오티드, 40개 이하의 뉴클레오티드, 30개 이하의 뉴클레오티드, 20개 이하의 뉴클레오티드 또는 10개 이하의 뉴클레오티드이다. 헤어핀 루프는 그 길이가 바람직하게, 약 1 내지 110개, 2 내지 100개, 5 내지 80개 또는 6 내지 50개의 뉴클레오티드이다. 헤어핀 루프가 어댑터의 차별 선택 가능성에 관련한 경우에는, 더 긴 길이의 헤어핀 루프, 예컨대 50 내지 110개의 뉴클레오티드가 바람직하다. 유사하게, 헤어핀 루프가 선택 가능한 결합에 관여하지 않는 경우에는, 더 짧은 길이의 헤어핀 루프, 예컨대 1 내지 5개의 뉴클레오티드가 바람직하다.
- [0193] 헤어핀 어댑터는 표적 폴리뉴클레오티드의 어느 한 말단, 즉 5' 또는 3' 말단에 라이게이션될 수 있다. 헤어핀 어댑터는 관련 기술분야에 공지된 임의의 방법을 이용하여 라이게이션될 수 있다. 헤어핀 어댑터는 리가제, 예컨대 T4 DNA 리가제, 이. 콜라이 DNA 리가제, Taq DNA 리가제, Tma DNA 리가제 및 9° N DNA 리가제를 사용하여 라이게이션될 수 있다.
- [0194] 상보적 폴리뉴클레오티드 (RNA 또는 cDNA)와 표적 RNA는, 연결된 구축물을 본 발명에 따라서 세공과 접촉시킨 후 또는 전에 분리될 수 있다. 이들은 세공을 통한 폴리뉴클레오티드 이동이 폴리뉴클레오티드 결합성 단백질, 예컨대 헬리카제, 또는 분자 브레이크에 의해 제어됨에 따라 분리될 수 있다.
- [0195] 상보적 폴리뉴클레오티드 (RNA 또는 cDNA)와 표적 RNA는 관련 기술분야에 공지된 임의의 방법을 이용하여 분리될 수 있다. 예를 들어, 이들은 폴리뉴클레오티드 결합성 단백질에 의해 분리될 수 있거나 또는 탈혼성화를 선호하는 조건 (탈혼성화를 선호하는 조건의 예는 고온, 높은 pH, 및 수소 결합 또는 염기쌍 형성을 붕괴시킬 수 있는 작용제, 예컨대 포름아미드 및 우레아의 부가를 포함하지만, 이에 제한되지 않는다)을 이용하여 분리될 수

있다.

- [0196] 헤어핀 어댑터는 바람직하게 선택 가능한 결합성 모이어티를 포함한다. 이로써 상기 연결된 구축물이 정제 또는 분리될 수 있다. 선택 가능한 결합성 모이어티는 그의 결합 특성을 기준으로 하여 선택될 수 있는 모이어티이다. 따라서, 선택 가능한 결합성 모이어티는 바람직하게, 표면과 특이적으로 결합하는 모이어티이다. 선택 가능한 결합성 모이어티가 본 발명에서 사용된 임의의 다른 모이어티보다 훨씬 더 큰 정도로 표면과 결합하는 경우에, 상기 선택 가능한 결합성 모이어티는 그 표면과 특이적으로 결합한다. 바람직한 실시양태에서, 상기 모이어티는 본 발명에서 사용된 다른 모이어티와 결합하지 않는 표면과 결합한다.
- [0197] 적합한 선택적 결합성 모이어티는 관련 기술분야에 공지되어 있다. 바람직한 선택적 결합성 모이어티는 비오틴, 폴리뉴클레오타이드 서열, 항체, 항체 단편, 예컨대 Fab 및 ScSv, 항원, 폴리뉴클레오타이드 결합성 단백질, 폴리 히스티딘 테일 및 GST 태그를 포함하지만, 이에 제한되지 않는다. 가장 바람직한 선택적 결합성 모이어티는 비오틴 및 선택 가능한 폴리뉴클레오타이드 서열이다. 비오틴은 아비딘으로 코팅된 표면과 특이적으로 결합한다. 선택 가능한 폴리뉴클레오타이드 서열은 상동 서열로 코팅된 표면과 특이적으로 결합한다 (즉, 혼성화한다). 또 다른 한편으로, 선택 가능한 폴리뉴클레오타이드 서열은 폴리뉴클레오타이드 결합성 단백질로 코팅된 표면과 특이적으로 결합한다.
- [0198] 헤어핀 어댑터 및/또는 선택 가능한 결합성 모이어티는 커팅되거나, 닉킹되거나, 절단되거나 또는 가수분해될 수 있는 영역을 포함할 수 있다. 이러한 영역은 상보적 폴리뉴클레오타이드 및/또는 표적 RNA가, 정제 또는 분리 후에 이와 결합되는 표면으로부터 제거될 수 있도록 설계될 수 있다. 적합한 영역이 관련 기술분야에 공지되어 있다. 적합한 영역은 RNA 영역, 데스티오비오틴 및 스트렙타비딘을 포함하는 영역, 디설피드 결합 및 광절단 가능한 영역을 포함하지만, 이에 제한되지 않는다.
- [0199] 상기 연결된 구축물은 바람직하게, 브릿징 모이어티, 예컨대 헤어핀 루프 또는 헤어핀 루프 어댑터로부터의 반대 말단에 리더 서열을 포함한다.
- [0200] 본 발명의 한 실시양태에서, 브릿징 모이어티, 예컨대 헤어핀-형성성 올리고뉴클레오타이드가 표적 RNA 가닥에 부착된다. 브릿징 모이어티의 부착 이전에, 표적 RNA 가닥이 비-RNA 폴리뉴클레오타이드를 포함하도록 변형시켰을 수 있다. 예를 들어 표적 RNA 가닥은 화학적 부착을 이용하여, 예를 들어 공유 결합, 클릭 화학, 화학적 또는 효소적 라이게이션에 의해, 혼성화 및/또는 합성 방법에 의해 비-RNA 폴리뉴클레오타이드에 부착시켰을 수 있다. RNA 폴리뉴클레오타이드는 토포이소머라제를 사용하여 비-RNA 폴리뉴클레오타이드에 부착시켰을 수 있다. 또 다른 한편으로 표적 RNA 가닥이 비-RNA 폴리뉴클레오타이드를 포함하도록 변형시키기 전에, 브릿징 모이어티를 이러한 표적 RNA 가닥에 부착시킬 수 있다. 유사하게, 브릿징 모이어티, 예컨대 헤어핀-형성성 올리고뉴클레오타이드는 본원에 기재된 부착 방법 중 임의의 것에 의해 표적 RNA 가닥에 부착시킬 수 있다. 바람직하게 브릿징 모이어티는 라이게이션에 의해 표적 RNA 가닥에 부착된다. 상기 언급된 임의의 적합한 리가제, 예를 들어, T4 DNA 리가제, 이. 콜라이 DNA 리가제, Taq DNA 리가제, Tma DNA 리가제, 9° N DNA 리가제, T4 폴리머라제 I, T4 폴리머라제 2, 열안정성 5' App DNA/RNA 리가제, 스플린트R, circ 리가제, T4 RNA 리가제 1 또는 T4 RNA 리가제 2를, 브릿징 모이어티를 표적 RNA 가닥에 라이게이션하기 위해 사용할 수 있다.
- [0201] 바람직하게 브릿징 모이어티는 표적 RNA 가닥의 3' 또는 5' 말단에 부착되고, 가장 바람직하게 브릿징 모이어티는 표적 RNA 가닥의 3' 말단에 부착된다. 브릿징 모이어티, 예컨대 헤어핀-형성성 올리고뉴클레오타이드로부터 역전사되면, 헤어핀에 의해 연결되는 RNA-cDNA 구축물이 형성된다. 이러한 실시양태에서 브릿징 모이어티는 역전사를 위한 프라이머로서 작용한다: 브릿징 모이어티, 예컨대 헤어핀 형성성 올리고뉴클레오타이드는 그 자체가 역전사를 위한 프라이머로서 사용되어 RNA-cDNA 구축물을 생성시키고 2D 판독을 가능하게 한다. 바람직하게 역전사는 브릿징 모이어티의 3' 말단에서 개시된다. 브릿징 모이어티 상의 폴리-T 오버행이 RNA의 폴리 A-테일과 혼성화된다. RNA의 폴리 A-테일은 전형적으로 그 길이가 50개 내지 300개의 뉴클레오타이드이다. 바람직하게 브릿징 모이어티의 폴리-T 오버행은 100개 미만의 뉴클레오타이드, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 92, 93, 94, 95, 96, 97, 98, 또는 99개의 뉴클레오타이드를 포함한다. 바람직하게 폴리 T 오버행은 브릿징 모이어티의 3' 말단 상에 있다. 본 발명의 이러한 실시양태에서, 브릿징 모이어티가 표적 RNA (이러한 표적 RNA는 비-RNA 폴리뉴클레오타이드, 바람직하게 DNA 폴리뉴클레오타이드를 포함하도록 이미 변형시켰거나 또는 아직 변형되지 않았다)에 부착되고 브릿징 모이어티가 역전사를 위한 프라이머로서 작용하는 경우,

DNA/RNA/cDNA 구축물이 창출된다 (실시에 5 및 도 15 참조). 이러한 구축물은 2D 판독을 허용하는데, 여기서 표적 RNA와 그의 상보적 DNA 서열 (역전사에 의해 형성됨) 둘 다는 브릿징 모이어티의 존재에 기인하여 막횡단 세공에 의해 서열 분석될 수 있다. 이러한 방법은 단일의 이중 가닥 표적 폴리뉴클레오티드 구축물로부터 수득된 정보의 양을 2배로 해주기 때문에 유리하다. 더욱이, 상보적 (cDNA) 가닥 내의 서열은 주형 RNA 가닥의 서열과 반드시 직교하기 때문에, 두 가닥으로부터의 정보는 그 정보에 따라 조합될 수 있다. 따라서, 이러한 기전은 더 높은 신뢰도의 관찰을 제공하는 직교 교정 능력을 제공한다.

[0202] 더욱이, 본 발명의 방법의 다른 주요 이점은 다음과 같다:

[0203] 1) 누락된 뉴클레오티드의 포함 범위: 상기 방법은 임의의 누락된 뉴클레오티드 또는 뉴클레오티드 군의 문제 (예를 들어, RNA 가닥이 세공을 빠져나가는 것과 같은 이동 상의 문제에 기인함)를 실질적으로 최소화하는데, 이는 한 가닥에서 누락될 수도 있는 임의의 상태가, 그의 보체 영역으로부터 수득된 직교 정보에 의해 커버될 가능성이 있기 때문이다.

[0204] 2) 문제가 있는 서열 모티프의 포함 범위: 어떠한 어려운 서열 모티프라도 상보적 가닥 내의 직교 및 반대 정보에 의해 커버되는데, 상이한 서열을 갖는 것은 동일한 서열 의존성 문제를 야기시키지 않을 것이다. 예를 들어, 이는 특히, 전류 상의 작은 변화만을 생성시키거나 또는 유사한 전류 수준을 갖는 서열 모티프와 관련이 있는데, 즉 나노 세공을 통해 이동할 때 동일한 전류 블록을 생성시키므로, 전류 상의 변화 단계가 없는 경우에는 관찰되지 않는 연속되는 염기 모티프와 관련이 있다. 하나의 서열 모티프로부터의 임의의 유사한 전류 수준은 보체 가닥 내의 그의 직교 서열로부터 수득된 완전히 상이한 전류 수준에 의해 커버될 것이다.

[0205] 상기 논의된 이점 이외에도, 이중 가닥 폴리뉴클레오티드의 양 가닥을 판독하는 개념을 활용하여 추가의 혜택을 제공할 수 있는 다수의 특별한 사례가 있다:

[0206] 1. 후성 유전 정보

[0207] 자연 RNA 가닥 내의 손상된 염기 또는 후성 유전 정보 (예컨대 5-메틸시토신 또는 5-히드록시메틸시토신 뉴클레오티드)를 확인할 수 있는 것이 광범위한 적용 분야에서 바람직하다. 본 발명의 방법을 이용하여, 화학적 처리 또는 증폭 없이 (이들 둘 다는 오류를 도입할 수 있다) 상기 정보를 수득한다. 나노 세공 서열 분석 동안, 뉴클레오티드가 나노 세공을 관통함에 따라 전류 수준 상의 변화를 측정한다. 전류 수준은 단일 뉴클레오티드에 의해서라기 보다는 오히려 여러 개의 염기에 의해 영향을 받는다. 변형된 뉴클레오티드가 나노 세공을 관통하는 경우, 이는 여러 연속적인 수준에 영향을 미친다: 이는 본 발명자들이 그것을 검출할 때 가질 수 있는 자신감을 증가시켜 준다 (도 18 참조).

[0208] 나노 세공 서열 분석은 또한, 단일 분자 서열 분석 기술이므로, 증폭시킬 필요없이 수행될 수 있다. 나노 세공은 표준 4개의 RNA 뉴클레오티드에 대한 변형을 검출할 수 있는 것으로 밝혀졌다. 폴리뉴클레오티드의 양 가닥을 판독하는 것이, 변형된 염기가 또 다른 염기와 유사한 방식으로 행동하는 (유사한 전류 시그널을 생성시킨다) 상황에서 RNA 변형을 검출하는 데 유용할 수 있다. 예를 들어 메틸시토신 (mC)이 우라실과 유사한 방식으로 행동하는 경우에는, mC를 U에 매핑하는 것과 연관된 오류가 있다. 주형 가닥에서는, 그러한 염기가 mC 또는 U로 지칭될 개연성이 있다. 그러나, 보체 가닥에서는, 상응하는 염기가 G로서 보여질 개연성은 더 높을 수 있다. 따라서, 보체 가닥을 "교정"함으로써, 주형 가닥 내의 염기가 U가 아닌 mC일 가능성이 높다.

[0209] 주형 및 보체 가닥을 판독하는 것은 증폭 또는 복제의 필요성 없이도 수행될 수 있다. 그러나, 증폭 또는 복제가 샘플 제조의 일부로서 부가되어 후성 유전 정보의 검출을 도와줄 수 있다.

[0210] 표적 RNA 폴리뉴클레오티드를 포함하는 연결된 가닥은, 서열 분석을 수행하기 전의 임의의 단계에서 필요한 만큼 수차례 분리 및 복제시킬 수 있다. 예를 들어, 상기 언급된 바와 같은 제1 RNA/cDNA 폴리뉴클레오티드 구축물의 연결된 두 가닥을 분리시킨 후, 이로써 생성되는 단일 가닥 폴리뉴클레오티드에 대한 상보적 가닥을 생성시켜 또 다른 이중 가닥 폴리뉴클레오티드를 형성할 수 있다. 이어서, 이러한 이중 가닥 폴리뉴클레오티드의 두 가닥을, 브릿징 모이어티를 이용하여 연결시켜 제2 구축물을 형성할 수 있다. 이는 본원에서 "DUO" 방법으로서 지칭될 수 있다. 이어서, 이러한 구축물을 본 발명에 사용할 수 있다. 이러한 실시양태에서는, 상기 생성된 구축물 내의 이중 가닥 폴리뉴클레오티드의 한 가닥이 브릿징 모이어티에 의해 연결된 원래의 표적 이중 가닥 폴리뉴클레오티드 (제1 구축물 내에 있음)의 양 가닥을 함유한다. 원래의 표적 이중 가닥 폴리뉴클레오티드의 서열 또는 보체 가닥을 추정하거나 또는 결정할 수 있다. 이러한 복제 과정은 필요한 만큼 수차례 반복될 수 있고, 표적 폴리뉴클레오티드가 실제로 수차례 판독되므로 부가의 교정을 제공한다.

[0211] 다음 정보가 나노 세공을 통하여 다음 순서로 판독되는 경우에 뉴클레오티드 가닥이 구축될 수 있다: 주형 RNA

(원래), 보체 cDNA (원래), -브릿징 모이어티-, 주형 RNA (보체), 보체 cDNA (보체).

[0212] 이러한 계획에서는, 메틸화 염기에 관한 정보가 4배로 수득될 것이다. 후성 유전 염기가 원래의 주형 가닥 내에 있는 경우 (본 경우에는, mC이다), 다음 정보가 높은 확률로 수득될 것이다: 주형 (원래)-mC, 보체 (원래)-G, 주형 (보체)-C, 및 보체 (보체)-G. 원래의 주형 판독과 복제된 주형 판독이 상이한 결과를 제공할 것인 반면, 양 상보적 판독은 동일한 염기 호출을 산출한다는 것은 명백하다. 이러한 정보를 이용하여 원래 주형 가닥 내의 후성 유전 염기의 위치를 표시할 수 있다.

## [0213] 2. 단독중합체 판독

[0214] 단독중합체 판독은 단일 분자 나노 세공 서열 분석에 문제가 될 수 있다. 단독중합체 영역이 세공의 판독 섹션보다 더 긴 경우, 단독중합체 섹션의 길이를 결정하기가 어려워질 것이다.

[0215] 이러한 단독중합체 판독의 문제점을 극복하기 위해, 상이한/변형된 염기를 원래의 dTTP, dGTP, dATP, dCTP와 조합하여 부가하여 cDNA 가닥을 합성할 수 있다. 이는 자연 염기 유사체, 예컨대 이노신 (I)일 수 있었다. 그 염기는 정확한 자연 염기와 비교해서 무작위로 혼입될 기회가 있을 것이며, 삽입률은 트리포스페이트 중의 농도를 변화시킴으로써 제어될 수 있다.

[0216] 대체 염기의 부가를 통하여, 대체 염기가 단독중합체 영역의 역보체 내로 삽입될 개연성이 있을 것이다. 이에 따른 결과는 단독중합체 실행의 길이가 나노 세공의 판독 섹션에 의해 판독될 수 있는 지점으로 감소될 것이라는 사실이다. 예를 들어, AAAAAAAAAA (서열식별번호: 33)의 단독중합체 군은 대체 염기의 무작위 삽입을 가질 것이고 TTTITTTIITTTI (서열식별번호: 34) (여기서, I는 이노신이다)를 제공할 수 있다.

[0217] 개개의 뉴클레오타이드 또는 뉴클레오타이드 군이 추정되거나 또는 결정될 수 있도록 단독중합체 연장물이 감소된다. 주형 가닥은 자연 RNA 가닥인 반면, 상보적/cDNA 가닥은 자연 염기와 염기 유사체의 혼합물을 함유할 것이다. 주형 판독 및 상보적 판독으로부터의 데이터의 조합을 이용하여, 원래의 RNA 섹션 내에서의 단독중합체 실행의 길이를 추정할 수 있다.

## [0218] 토포이소머라제

[0219] 토포이소머라제는 단일 가닥 또는 이중 가닥 DNA 중 하나와 결합하고 DNA의 포스페이트 주쇄를 커팅한다. 이러한 중간 단계는 DNA가 풀릴 수 있게 해주고, 이 과정이 끝날 무렵, DNA 주쇄는 다시 재밀봉된다. 적합한 토포이소머라제 결합 전략이 도 6 및 7에 예시된다.

## [0220] 다른 방법

[0221] RNA 표적 폴리뉴클레오타이드를 비-RNA 폴리뉴클레오타이드에 부착시키는 대체 방법은 7 메틸 구아노신 캡을 반대 방향으로 부가함으로써 변형되는 진행 RNA의 5' 말단을 활용하는 것을 포함한다. 따라서 진행 RNA의 5' 말단은 역전된 염기를 함유하거나 또는 포함하는데, 즉 RNA 가닥의 5' 말단에서 개개의 염기는 5' 말단과 반대로 3' 말단이 없다. 표적 RNA 폴리뉴클레오타이드는 역전된 염기를 수반한 영역을 또한 갖는 비-RNA 폴리뉴클레오타이드(이는 리보뉴클레오타이드 또는 RNA 서열을 포함할 수 있거나 또는 포함하지 않을 수 있다)에 라이게이션되어, 도 12 (여기서, 비-RNA 폴리뉴클레오타이드는 DNA이다)에 의해 예시된 바와 같이 반대 방향으로 실행되도록 할 수 있다. 또 다른 한편으로 비-RNA 폴리뉴클레오타이드는 한 말단에 역전된 염기를 수반한 RNA의 섹션을 가질 수 있다. 이러한 역전된 RNA 염기의 섹션은 라이게이션에 의해 진행 RNA의 5' 말단에서 역전된 염기에 부착될 수 있다.

[0222] 비-RNA 폴리뉴클레오타이드는 미리 결합된 DNA 헬리카제 효소를 포함할 수 있거나 또는 포함하지 않을 수 있다.

## [0223] 진행 RNA

[0224] 진행 생물에서는, 핵에서 생성된 1차 RNA 전사체를, 번역 기구를 프로그래밍하기 위해 사용되는 세포질로 수송하기 전에 여러 방식으로 프로세싱한다. 먼저, 트리포스페이트 결합에 의해 전사체의 5' 말단에 연결된 7-메틸 구아노신 잔기로 이루어진 캡을 전사 동안 부가한다 (캡핑). 캡핑은 5'-5' 트리포스페이트 연쇄를 포함한다. 캡은 번역 기구에 의해 인식되고, 성장하고 있는 RNA 쇄가 뉴클레아제에 의해 분해되지 못하게 한다. 이어서, 아데노신 잔기의 연장물을 3' 말단에 부가한다 (폴리아데닐화). 이들 폴리A 테일은 150개 내지 200개 잔기 길이이다. 이들 변형 후, RNA 스플라이싱이 개재 서열 (즉, 인트론)을 제거한다.

[0225] 본 발명의 한 실시양태에서 5' 캡을 제자리에 두고 적어도 하나의 화학적으로 반응성인 기를 표적 RNA 폴리뉴클레오타이드의 5' 말단에 부가한다. 화학적 부착의 임의의 방법을 본 발명의 방법에서 사용할 수 있는데, 단 a) 이러한 화학적 반응은 RNA 또는 DNA 구조를 손상시키지 않아야 하고, b) 생성되는 반응성 연쇄는 DNA 헬리카제



효소가 그를 따라 이동할 수 없거나 또는 그를 통과할 수 없을 정도로 벌키하지 않아야 한다. 적어도 하나의 반응성 기를, 하이퍼메틸라제 효소를 사용하여 표적 RNA 폴리뉴클레오타이드에 부가할 수 있다. 본 발명의 한 실시양태에서 RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기가 클릭 반응성 기이긴 하지만, 이것이 필수적인 것은 아니다. RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기는 또 다른 한편으로 임의의 적합한 반응성 기, 예컨대 티올일 수 있다. 적어도 하나의 반응성 기는 또한, 비-RNA 폴리뉴클레오타이드 (이는 리보뉴클레오타이드 또는 RNA 서열을 포함할 수 있거나 또는 포함하지 않을 수 있다)의 말단에 부착된다. 바람직하게 적어도 하나의 반응성 기는 비-RNA 폴리뉴클레오타이드의 3' 말단에 부착된다. 바람직하게 비-RNA 폴리뉴클레오타이드는 DNA 가닥이다. 비-RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기는 하이퍼메틸라제 효소를 사용하여 부가할 수 있다. 본 발명의 한 실시양태에서 비-RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기가 클릭 반응성 기이긴 하지만, 이것이 필수적인 것은 아니다. 이어서, 각각의 RNA 폴리뉴클레오타이드 및 비-RNA 폴리뉴클레오타이드 상의 하나 이상의 반응성 기는 공유 결합을 형성하기에 적합한 조건하에 접촉시킨다. ATP 감마 S 및 효소를 사용하여 티오포스페이트를 DNA에 부가할 수 있고, 이어서 이를, 말레이미드가 부착된 RNA에 부착시킬 수 있다.

[0226] 본 발명의 대체 실시양태에서 7-메틸구아노신 캡은, 바람직하게 담배 산 피로포스파타제를 사용하여 제거하여 탈캡핑된 RNA 가닥을 형성시킨다. 이어서, 이와 같이 탈캡핑된 표적 RNA 폴리뉴클레오타이드를 수많은 상이한 방식으로 처리하여, 비-RNA 폴리뉴클레오타이드에 부착된 RNA 폴리뉴클레오타이드의 가닥을 생성시킬 수 있다. 한 실시양태에서, 적어도 하나의 반응성 기를 상기 탈캡핑된 RNA 폴리뉴클레오타이드의 5' 말단에 부가할 수 있다. 적어도 하나의 반응성 기는 하이퍼메틸라제 효소를 사용하여 RNA 폴리뉴클레오타이드에 부가할 수 있다. 본 발명의 한 실시양태에서 RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기가 클릭 반응성 기이긴 하지만, 이것이 필수적인 것은 아니다. RNA 폴리뉴클레오타이드에 부가된 반응성 기는 또 다른 한편으로 임의의 적합한 반응성 기, 예컨대 티올일 수 있다. 적어도 하나의 반응성 기는 또한, 비-RNA 폴리뉴클레오타이드의 말단에 부착된다. 바람직하게 적어도 하나의 반응성 기는 비-RNA 폴리뉴클레오타이드의 3' 말단에 부착된다. 바람직하게 비-RNA 폴리뉴클레오타이드는 DNA 가닥이다. 비-RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기는 하이퍼메틸라제 효소를 사용하여 부가할 수 있다. 본 발명의 한 실시양태에서 비-RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기가 클릭 반응성 기이긴 하지만, 이것이 필수적인 것은 아니다. 이어서, 각각의 RNA 폴리뉴클레오타이드 및 비-RNA 폴리뉴클레오타이드 상의 하나 이상의 반응성 기는 공유 결합을 형성하기에 적합한 조건하에 접촉시킨다.

[0227] 또 다른 한편으로, 비-RNA 폴리뉴클레오타이드의 가닥은 리가제, 예컨대 T4 DNA 리가제, 이. 콜라이 DNA 리가제, Taq DNA 리가제, Tma DNA 리가제, 9° N DNA 리가제, T4 폴리머라제 1, T4 폴리머라제 2, 열안정성 5' App DNA/RNA 리가제, 스플린트R, circ 리가제, T4 RNA 리가제 1, T4 RNA 리가제 2를 사용하여 표적 RNA 폴리뉴클레오타이드 상으로 직접 라이게이션할 수 있다. 바람직하게 비-RNA 폴리뉴클레오타이드는 DNA 가닥이다. 본 발명의 한 실시양태에서, 하나 이상의 효소는, RNA에 라이게이션되어야 하는 비-RNA 폴리뉴클레오타이드와 미리 결합될 수 있다. 효소를 비-RNA 폴리뉴클레오타이드 상에 미리 부하하면 샘플 제조 공정의 속도가 높아지는데, 이는 더 적은 수의 튜브가 사용된다는 것을 의미한다. 또 다른 한편으로 RNA 폴리뉴클레오타이드에 라이게이션되어야 하는 비-RNA 폴리뉴클레오타이드는 미리 결합된 효소를 함유하지 않거나 포함하지 않는다.

[0228] 본 발명의 대체 실시양태에서, 비-RNA 폴리뉴클레오타이드는 RNA 폴리뉴클레오타이드와 혼성화되는 리더 서열을 수반한 DNA 프라이머이다. 하나 이상의 효소는 RNA 폴리뉴클레오타이드와 혼성화되어야 하는 DNA 프라이머와 미리 결합될 수 있다. 또 다른 한편으로 RNA 폴리뉴클레오타이드와 혼성화되어야 하는 DNA 프라이머는 미리 결합된 효소를 함유하지 않거나 포함하지 않는다. DNA 프라이머로부터 RNA 폴리뉴클레오타이드를 역전사시키면, 사용된 역전사효소에 따라서 1개 내지 3개의 C의 3' 오버행이 생성된다. 이어서, DNA 헤어핀을 이중 가닥 DNA/RNA에 라이게이션할 수 있다. 혼성화를 허용하는 조건은 관련 기술분야에 널리 공지되어 있다 (예를 들어, 문헌 [Sambrook et al., 2001, Molecular Cloning: a laboratory manual, 3rd edition, Cold Spring Harbour Laboratory Press; and Current Protocols in Molecular Biology, Chapter 2, Ausubel et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995)]). 혼성화는 저 수준의 엄격성 조건하에, 예를 들어 37°C 하에 30 내지 35% 포름아미드, 1 M NaCl 및 1% SDS (소듐 도데실 술페이트)의 완충 용액의 존재하에, 이어서 50°C 하에 1 X (0.1650 M Na+) 내지 2 X (0.33 M Na+) SSC (표준 소듐 시트레이트)에서 세척함으로써 수행될 수 있다. 혼성화는 중간 수준의 엄격성 조건하에, 예를 들어 37°C 하에 40 내지 45% 포름아미드, 1 M NaCl, 및 1% SDS의 완충 용액의 존재하에, 이어서 55°C 하에 0.5 X (0.0825 M Na+) 내지 1 X (0.1650 M Na+) SSC에서 세척함으로써 수행될 수 있다. 혼성화는 고 수준의 엄격성 조건하에, 예를 들어 37°C 하에 50% 포름아미드, 1 M NaCl 및 1% SDS의 완충 용액의 존재하에, 이어서 60°C 하에 0.1 X (0.0165 M Na+) SSC에서 세척함으로써 수

행될 수 있다. 특히, 상기 조건은 바람직하게, 10 mM 트리스-HCl, 50 mM NaCl, pH 7 중의 10  $\mu$ M 올리고머 및 98°C로의 가열 후, 분당 2°C로 18°C까지 냉각시킨다.

[0229] 진핵 RNA는 전형적으로 폴리A 테일, 즉 연속되는 아데노신 모노포스페이트의 연장물을 포함한다. 폴리A 테일은 전형적으로 RNA의 3' 말단에 있다. 이러한 실시양태에서, 특정 프라이머가 표적 RNA의 폴리A 테일과 혼성화될 수 있다. 이러한 프라이머는 바람직하게 폴리U 영역, 즉 우라실에 근거한 뉴클레오타이드만을 함유하는 영역을 포함한다. 폴리U 영역은 UMP 또는 dUMP를 함유할 수 있다. 폴리U 영역은 임의의 길이, 예컨대 적어도 10개, 적어도 15개, 적어도 20개, 적어도 25개 또는 그 초과일 수 있다. 또 다른 한편으로 프라이머는 폴리T 영역을 포함한다. 폴리T 영역은 임의의 길이, 예컨대 적어도 10개, 적어도 15개, 적어도 20개, 적어도 25개 또는 그 초과일 수 있다. 진핵 RNA의 가닥을 DNA의 가닥에 부착시키는 방법이 도 1에 예시되어 있다.

[0230] 본 발명의 대체 실시양태에서, 비-RNA 폴리뉴클레오타이드는 토포이소머라제-기반 전략을 이용하여 표적 RNA 폴리뉴클레오타이드에 부착시킬 수 있다 (Cheng and Shuman, 2000, Nucleic Acids Research, Vol. 28, No. 9 1893-1898; 본원에 참조로 포함된다). DNA 토포이소머라제는 이중체 DNA와 결합하고, 특이적 표적 부위에서 한 가닥의 포스포디에스테르 주쇄를 절개한다. 이중체 DNA의 다른 가닥은 상응하는 위치에서 서열 내에 닉 또는 절단물(break)을 포함한다. 일단 토포이소머라제가 결합되고 DNA의 상부 가닥을 커팅하면, 이는 dsDNA와 결합된 채로 있다. dsDNA와 결합되는 토포이소머라제는 DNA 가닥을 5' OH 말단화 RNA 가닥으로 전이시켜, 이를 유리 5' 히드록실을 갖는 RNA와 함께 인큐베이션한 경우에는 탠덤(tandem) DNA-RNA 공중합체를 형성시킬 수 있다. 표적 RNA는 유리 5' 히드록실을 가져야만 한다: 진핵 RNA는 탈캡핑시켜, 유리 5' 히드록실을 수반한 RNA를 생성시켜야만 하는 반면, 마이크로RNA는 유리 5' 히드록실 기를 갖는다 (도 6 참조).

[0231] 대체 토포이소머라제-기반 전략에서는, RNA와 혼성화하고 RNA 부착을 도와주는 DNA의 단일 가닥 영역을 이용하여, dsDNA를 RNA의 가닥에 부착시킨다 (Sekiguchi et al., 1997, The Journal of Biological Chemistry, Vol 272, No. 25, 15721-15728; 본원에 참조로 포함된다). 토포이소머라제는 특이적 표적 서열에서 dsDNA 상으로 결합된다. DNA의 하부 가닥은 그 내부에 닉을 갖지 않는다. 일단 토포이소머라제가 결합되면, 이는 단지 DNA의 상부 가닥을 커팅하고 dsDNA/ssDNA 혼성체와 결합된 채로 있다. 이어서, dsDNA와 결합되는 토포이소머라제를, 유리 5' 히드록실을 갖는 RNA와 함께 인큐베이션하면 RNA가 dsDNA와 연결된다. ssDNA 영역은 상보적 RNA 서열을 DNA에 대한 부착 지점으로 유인하는 데 도움을 준다 (도 7 참조).

[0232] RNA 폴리뉴클레오타이드를 비-RNA 폴리뉴클레오타이드에 부착시키는 대체 방법은 반대 방향으로 실행되는 7-메틸구아노신 캡의 부가에 의해 변형되는 진핵 RNA의 5' 말단을 활용하는 것을 포함한다. 따라서 진핵 RNA의 5' 말단은 역전시킨 구아닌 염기를 함유하거나 또는 포함하는데, 즉 이것은 RNA 가닥의 5' 말단에 있지만 개개의 염기는 5' 말단과 반대로 3' 말단이 없다. 표적 RNA는 역전된 염기를 수반한 영역을 또한 갖는 DNA (또는 RNA) 서열에 라이게이션되어, 반대 방향으로 실행되도록 할 수 있다 (도 12 참조).

[0233] *원핵 RNA*

[0234] 폴리 A 폴리머라제 효소를 활용하여 폴리(dA) 테일을 RNA 폴리뉴클레오타이드의 3' 말단 상으로 부가할 수 있다. 본 발명의 한 실시양태에서, 비-RNA 폴리뉴클레오타이드는 RNA 폴리뉴클레오타이드와 혼성화되는 리더 서열을 수반한 DNA 프라이머이다. 리더 서열을 수반한 DNA 프라이머는 폴리(dA) 영역과 혼성화될 수 있다. 하나 이상의 효소는 RNA 폴리뉴클레오타이드와 혼성화되어야 하는 DNA 프라이머와 미리 결합될 수 있다. 또 다른 한편으로 RNA 폴리뉴클레오타이드와 혼성화되어야 하는 DNA 프라이머는 미리 결합된 효소를 함유하지 않거나 포함하지 않는다. DNA 프라이머로부터 RNA 폴리뉴클레오타이드를 역전사시키면, 사용된 역전사효소에 따라서 1개 내지 3개의 C의 3' 오버행이 생성된다 [슈퍼스크립트(SuperScript) II (MMLV)의 경우에는 3개 내지 4개] (문헌 [p1192 Biotechniques Vol 29 No 6 (2000)] 참조). 이어서, DNA 헤어핀을 이중 가닥 DNA/RNA에 라이게이션할 수 있다.

[0235] 본 발명의 대체 실시양태에서, 폴리 (dA) 영역은 RNA 폴리뉴클레오타이드에 부가하지 않고, 원핵 RNA 폴리뉴클레오타이드는 반응 단계에 직접 사용된다. 반응성 기는 RNA 폴리뉴클레오타이드의 5' 또는 3' 말단에 부가할 수 있다. 바람직하게 적어도 하나의 반응성 기는 RNA 폴리뉴클레오타이드의 5' 말단에 부가한다. RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기가 클릭 반응성 기일 수 있지만, 이것이 필수적인 것은 아니다. RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기는 또 다른 한편으로 임의의 적합한 반응성 기, 예컨대 티올일 수 있다. 적어도 하나의 반응성 기는 또한, 비-RNA 폴리뉴클레오타이드의 말단에 부착된다. 바람직하게 적어도 하나의 반응성 기는 비-RNA 폴리뉴클레오타이드의 3' 말단에 부착된다. 바람직하게 비-RNA 폴리뉴클레오타이드는 DNA 가닥을 포함한다. 비-RNA 폴리뉴클레오타이드에 부가된 적어도 하나의 반응성 기는 하이퍼메틸라제 효소



를 사용하여 부가할 수 있다. 본 발명의 한 실시양태에서 비-RNA 폴리뉴클레오티드에 부가된 적어도 하나의 반응성 기가 클릭 반응성 기이긴 하지만, 이것이 필수적인 것은 아니다. 이어서, 각각의 RNA 폴리뉴클레오티드 및 비-RNA 폴리뉴클레오티드 상의 하나 이상의 반응성 기는 공유 결합을 형성하기에 적합한 조건하에 접촉시킨다.

[0236] 본 발명의 대체 실시양태에서, 비-RNA 폴리뉴클레오티드의 가닥은 리가제, 예컨대 T4 DNA 리가제, 이. 콜라이 DNA 리가제, Taq DNA 리가제, Tma DNA 리가제, 9° N DNA 리가제, T4 폴리머라제 I, T4 폴리머라제 2, 열안정성 5' App DNA/RNA 리가제, 스플린트R, circ 리가제, T4 RNA 리가제 1, T4 RNA 리가제 2를 사용하여 표적 RNA 폴리뉴클레오티드 상으로 직접 라이게이션할 수 있다. 바람직하게 비-RNA 폴리뉴클레오티드는 DNA 가닥을 포함한다. 본 발명의 한 실시양태에서, 하나 이상의 효소는, RNA 폴리뉴클레오티드에 라이게이션되어야 하는 비-RNA 폴리뉴클레오티드와 미리 결합될 수 있다. 또 다른 한편으로 RNA 폴리뉴클레오티드에 라이게이션되어야 하는 비-RNA 폴리뉴클레오티드는 미리 결합된 효소를 함유하지 않거나 포함하지 않는다. 원핵 RNA 폴리뉴클레오티드를 비-RNA (예를 들어, DNA) 폴리뉴클레오티드에 부착시키는 방법은 도 2에 예시되어 있다.

[0237] 본 발명의 대체 실시양태에서, 비-RNA 폴리뉴클레오티드는 토포이소머라제-기반 전략을 이용하여 표적 RNA 폴리뉴클레오티드에 부착시킬 수 있다 (Cheng and Shuman, 2000, Nucleic Acids Research, Vol. 28, No. 9 1893-1898; 본원에 참조로 포함된다). DNA 토포이소머라제는 이중체 DNA와 결합하고, 특이적 표적 부위에서 한 가닥의 포스포디에스테르 주쇄를 절개한다. 이중체 DNA의 다른 가닥은 상응하는 위치에서 서열 내에 Nick 또는 절단물을 포함한다. 일단 토포이소머라제가 결합되고 DNA의 상부 가닥을 커팅하면, 이는 dsDNA와 결합된 채로 있다. dsDNA와 결합되는 토포이소머라제는 DNA 가닥을 5' OH 말단화 RNA 가닥으로 전이시켜, 이를 유리 5' 히드록실을 갖는 RNA와 함께 인큐베이션한 경우에는 탠덤 DNA-RNA 공중합체를 형성시킬 수 있다. 표적 RNA는 유리 5' 히드록실을 가져야만 한다. 대체 토포이소머라제-기반 전략에서는, RNA와 혼성화하고 RNA 부착을 도와주는 DNA의 단일 가닥 영역을 이용하여, dsDNA를 RNA의 가닥에 부착시킨다 (Sekiguchi et al., 1997, The Journal of Biological Chemistry, Vol 272, No. 25, 15721-15728; 본원에 참조로 포함된다). 토포이소머라제는 특이적 표적 서열에서 dsDNA 상으로 결합된다. DNA의 하부 가닥은 그 내부에 Nick를 갖지 않는다. 일단 토포이소머라제가 결합되면, 이는 단지 DNA의 상부 가닥을 커팅하고 dsDNA/ssDNA 혼성체와 결합된 채로 있다. 이어서, dsDNA와 결합되는 토포이소머라제를, 유리 5' 히드록실을 갖는 RNA와 함께 인큐베이션하면 RNA가 dsDNA와 연결된다. ssDNA 영역은 상보적 RNA 서열을 DNA에 대한 부착 지점으로 유인하는 데 도움을 준다 (도 7 참조).

[0238] 증폭 없음

[0239] 표적 RNA 폴리뉴클레오티드는 전형적으로, 본 발명의 방법에서 증폭되지 않는다. 본 방법은 전형적으로, 다수 카피의 표적 RNA를 만드는 것을 포함하지 않는다.

[0240] 본 방법은 바람직하게 폴리머라제 연쇄 반응 (PCR) 또는 역전사 PCR (RT-PCR)을 포함하지 않는다.

[0241] 상기 논의된 바와 같이, 본 발명의 한 실시양태에서 RNA 폴리뉴클레오티드는 DNA 프라이머 및 역전사효소를 사용하여 생성시킨 DNA 보체를 포함한다. 브릿징 모이어티를 이용하여 RNA 가닥과 DNA 가닥을 연결시키면, 2D 판독이 가능해 진다.

[0242] 본 발명의 대체 실시양태에서, 헤어핀 형성성 올리고뉴클레오티드는 그 자체가 역전사를 위한 프라이머로서 사용되어 RNA-cDNA 구축물을 생성시키고 2차원 판독 (즉, RNA 및 cDNA 가닥)을 가능하게 한다.

[0243] DNA 헬리카제(들) 및 분자 브레이크(들)

[0244] DNA 헬리카제는 세공을 통한 RNA 폴리뉴클레오티드의 이동을 제어하기 위해 사용된다. DNA 헬리카제 효소는, 이것이 표적 RNA 폴리뉴클레오티드와 결합할 수 있고 세공을 통한 표적 RNA 폴리뉴클레오티드의 이동을 제어할 수 있는 한은 효소 활성을 나타내지 않아도 된다. 예를 들어, 상기 효소는 그의 효소 활성을 제거하도록 변형될 수 있거나 또는 효소로서 작용하지 못하게 하는 조건하에 사용될 수 있다.

[0245] 헬리카제는 Hel308 헬리카제, RecD 헬리카제, 예컨대 TraI 헬리카제 또는 TrwC 헬리카제, XPD 헬리카제 또는 Dda 헬리카제일 수 있거나 또는 이들로부터 유래될 수 있다. DNA 헬리카제는 Hel308 Mbu (서열식별번호: 8), Hel308 Csy (서열식별번호: 9), Hel308 Tga (서열식별번호: 10), Hel308 Mhu (서열식별번호: 11), TraI Eco (서열식별번호: 12), XPD Mbu (서열식별번호: 13), Dda 1993 (서열식별번호:14) 또는 그의 변이체로부터 유래될 수 있다.

[0246] 헬리카제는 국제 출원 번호 PCT/GB2012/052579 (WO 2013/057495로서 공개됨); PCT/GB2012/053274 (WO

2013/098562로서 공개됨); PCT/GB2012/053273 (WO2013098561로서 공개됨); PCT/GB2013/051925 (WO 2014/013260으로서 공개됨); PCT/GB2013/051924 (WO 2014/013259로서 공개됨); PCT/GB2013/051928 (WO 2014/013262로서 공개됨) 및 PCT/GB2014/052736에 개시된 DNA 헬리카제, 변형된 DNA 헬리카제 또는 DNA 헬리카제 구축물 중 임의의 것일 수 있다.

[0247] 헬리카제는 세공에 대하여 두 가지 모드로 작동할 수 있다. 첫째, 인가된 전압으로부터 비롯되는 장을 이용하여 세공을 통해 RNA 폴리뉴클레오타이드의 이동을 제어시키도록 DNA 헬리카제를 사용하여 상기 방법을 수행하는 것이 바람직하다. 이러한 모드에서는, RNA 폴리뉴클레오타이드의 5' 말단을 먼저, 세공 내에 포획시키고; 상기 효소가 세공 내로의 RNA 폴리뉴클레오타이드의 이동을 제어시켜, RNA 폴리뉴클레오타이드가 막의 트랜스 측으로 최종적으로 전위될 때까지 상기 장을 이용하여 RNA 폴리뉴클레오타이드가 세공을 관통하도록 한다. 또 다른 한편으로, 상기 방법은 인가된 전압으로부터 비롯되는 장에 대항하여 DNA 헬리카제 효소가 세공을 통한 RNA 폴리뉴클레오타이드의 이동을 제어시키도록 수행하는 것이 바람직하다. 이러한 모드에서는, RNA 폴리뉴클레오타이드의 3' 말단을 먼저, 세공 내에 포획시키고; 상기 효소가 세공을 통한 RNA 폴리뉴클레오타이드의 이동을 제어시켜, RNA 폴리뉴클레오타이드가 최종적으로 막의 시스 측으로 다시 방출될 때까지 인가된 장에 대항하여 세공으로부터 인출되도록 한다.

[0248] DNA 헬리카제는 바람직하게 서열식별번호: 15에 제시된 서열 (Trwc Cba) 또는 그의 변이체, 서열식별번호: 8에 제시된 서열 (HeI308 Mbu) 또는 그의 변이체 또는 서열식별번호: 14에 제시된 서열 (Dda) 또는 그의 변이체를 포함한다. 변이체는 막횡단 세공에 관하여 다음에 논의된 방식 중 임의의 것에서 천연 서열과 상이할 수 있다. 서열식별번호: 14의 바람직한 변이체는 (a) E94C 및 A360C 또는 (b) E94C, A360C, C109A 및 C136A를 포함한 다음, 임의로 ( $\Delta$ M1)G1 (즉, M1을 결실시킨 다음, G1을 부가한다)을 포함한다.

[0249] 임의의 수의 헬리카제를 본 발명에 따라서 사용될 수 있다. 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과 헬리카제를 사용할 수 있다. 일부 실시양태에서, 상이한 수의 헬리카제를 사용할 수 있다. 본 발명의 한 실시양태에서 DNA 헬리카제는 비-RNA 폴리뉴클레오타이드와 미리 결합된다.

[0250] 본 발명의 방법은 바람직하게, 폴리뉴클레오타이드를 2개 이상의 헬리카제와 접촉시키는 것을 포함한다. 2개 이상의 헬리카제는 전형적으로 동일한 헬리카제이다. 2개 이상의 헬리카제는 상이한 헬리카제일 수 있다.

[0251] 2개 이상의 헬리카제는 상기 언급된 헬리카제의 임의의 조합일 수 있다. 2개 이상의 헬리카제는 2개 이상의 Dda 헬리카제일 수 있다. 2개 이상의 헬리카제는 하나 이상의 Dda 헬리카제 및 하나 이상의 TrwC 헬리카제일 수 있다. 2개 이상의 헬리카제는 동일한 헬리카제의 상이한 변이체일 수 있다.

[0252] 2개 이상의 헬리카제는 바람직하게 서로 부착된다. 2개 이상의 헬리카제는 보다 바람직하게 서로 공유적으로 부착된다. 헬리카제는 임의의 순서로 임의의 방법을 이용하여 부착될 수 있다. 본 발명에 사용하기 바람직한 헬리카제 구축물은 국제 출원 번호 PCT/GB2013/051925 (WO 2014/013260으로서 공개됨); PCT/GB2013/051924 (WO 2014/013259로서 공개됨); PCT/GB2013/051928 (WO 2014/013262로서 공개됨) 및 PCT/GB2014/052736에 기재되어 있다.

[0253] 서열식별번호: 8, 9, 10, 11, 12, 13, 14 또는 15의 변이체는 서열식별번호: 8, 9, 10, 11, 12, 13, 14 또는 15의 서열에서 벗어나고 폴리뉴클레오타이드 결합 능력을 보유하고 있는 아미노산 서열을 갖는 효소이다. 이는 관련 기술분야에 공개된 임의의 방법을 이용하여 측정될 수 있다. 예를 들어, 변이체는 폴리뉴클레오타이드와 접촉시킬 수 있고, 폴리뉴클레오타이드와 결합하여 이를 따라 이동할 수 있는 그의 능력을 측정할 수 있다. 변이체는 폴리뉴클레오타이드의 결합을 촉진시키고/시키거나 높은 염 농도 및/또는 실온에서 그의 활성을 촉진시키는 변형을 포함할 수 있다. 변이체는 이것이 폴리뉴클레오타이드와 결합하지만 (즉, 폴리뉴클레오타이드 결합 능력을 보유한다), 헬리카제로서 기능하지 않도록 (즉, 이동을 촉진시키는 데 필요한 성분, 예를 들어 ATP 및  $Mg^{2+}$  모두가 제공될 때 폴리뉴클레오타이드를 따라 이동하지 않도록) 변형될 수 있다. 이러한 변형은 관련 기술분야에 공지되어 있다. 예를 들어, 헬리카제 내에서의  $Mg^{2+}$  결합성 도메인의 변형은 전형적으로, 헬리카제로서 기능하지 않는 변이체를 생성시킨다. 이러한 유형의 변이체가 분자 브레이크 (하기 참조)로서 작용될 수 있다.

[0254] 서열식별번호: 8, 9, 10, 11, 12, 13, 14 또는 15의 아미노산 서열의 전체 길이에 걸쳐, 변이체는 바람직하게, 아미노산 유사성 또는 동일성에 근거하여 상기 서열과 적어도 50% 상동일 것이다. 보다 바람직하게, 변이체 폴리펩티드는 전체 서열에 걸쳐 서열식별번호: 8, 9, 10, 11, 12, 13, 14 또는 15의 아미노산 서열과의 아미노산 유사성 또는 동일성에 근거하여 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 및 보다 바람직하게 적어도 95%, 97% 또는 99% 상동일 수 있다. 200개 이상, 예를 들어

230, 250, 270, 280, 300, 400, 500, 600, 700, 800, 900 또는 1,000개 또는 그 초과와 연속되는 아미노산의 연장물에 걸쳐 적어도 80%, 예를 들어 적어도 85%, 90% 또는 95%의 아미노산 유사성 또는 동일성이 존재할 수 있다 ("하드 상동성"). 상동성은 상기 언급된 바와 같이 결정된다. 변이체는 상기 서열식별번호: 2 및 4와 관련하여 상기 논의된 방식 중 임의의 것에서 야생형 서열과 상이할 수 있다. 효소는 세공에 공유적으로 부착될 수 있다. 임의의 방법을 이용하여 효소를 세공에 공유적으로 부착시킬 수 있다.

[0255] 바람직한 실시양태에서, 본 방법은

[0256] (a) RNA 폴리뉴클레오타이드에 하나 이상의 DNA 헬리카제 및 하나 이상의 분자 브레이크를 제공하는 단계 (여기서, RNA 폴리뉴클레오타이드는 비-RNA 폴리뉴클레오타이드를 포함하고 이에 대한 DNA 헬리카제 결합을 증가시키도록 변형된다);

[0257] (b) RNA 폴리뉴클레오타이드를 막횡단 세공과 접촉시키고, 이러한 세공에 걸쳐 전위를 인가하여, 하나 이상의 DNA 헬리카제와 하나 이상의 분자 브레이크를 합치게 하고 둘 다가 세공을 통한 RNA 폴리뉴클레오타이드의 이동을 제어하도록 하는 단계; 및

[0258] (c) RNA 폴리뉴클레오타이드가 세공에 대하여 이동함에 따라 하나 이상의 측정치를 취함으로써 (여기서, 이러한 측정치는 RNA 폴리뉴클레오타이드의 하나 이상의 특징을 나타낸다), RNA 폴리뉴클레오타이드의 특징을 규명하는 단계

[0259] 를 포함한다.

[0260] 이러한 유형의 방법은 UK 출원 번호 1406151.9에 상세히 논의되어 있다. 바람직한 분자 브레이크는 TrwC Cba-Q594A (돌연변이 Q594A를 수반한 서열식별번호: 15)이다. 이러한 변이체는 헬리카제로서 기능하지 않는다 (즉, 이동을 촉진시키는 데 필요한 성분, 예를 들어 ATP 및  $Mg^{2+}$  모두가 제공될 때 폴리뉴클레오타이드와 결합하지만 그를 따라 이동하지 않는다).

[0261] 하나 이상의 헬리카제는 상기 논의된 것 중 임의의 것일 수 있다. 하나 이상의 분자 브레이크는 RNA 폴리뉴클레오타이드와 결합하고 세공을 통한 RNA 폴리뉴클레오타이드의 이동을 느리게 하는 임의의 화합물 또는 분자일 수 있다. 하나 이상의 분자 브레이크는 바람직하게, RNA 폴리뉴클레오타이드와 결합하는 하나 이상의 화합물을 포함한다. 하나 이상의 화합물은 바람직하게 하나 이상의 매크로사이클(macrocyclic)이다. 적합한 매크로사이클은 시클로텍스트린, 칼릭사렌, 사이클릭 펩티드, 크라운 에테르, 쿠쿠르비투릴, 필라라렌, 그의 유도체 또는 그의 조합물을 포함하지만, 이에 제한되지 않는다. 시클로텍스트린 또는 그의 유도체는 문헌 [Eliseev, A. V., and Schneider, H-J. (1994) *J. Am. Chem. Soc.* 116, 6081-6088]에 개시된 것 중 임의의 것일 수 있다. 상기 작용제는 보다 바람직하게 헵타키스-6-아미노-β-시클로텍스트린 ( $am_7-\beta CD$ ), 6-모노데옥시-6-모노아미노-β-시클로텍스트린 ( $am_1-\beta CD$ ) 또는 헵타키스-(6-데옥시-6-구아니디노)-시클로텍스트린 ( $gu_7-\beta CD$ )이다.

[0262] 하나 이상의 분자 브레이크는 바람직하게, 하나 이상의 단일 가닥 결합성 단백질 (SSB)이 아니다. 하나 이상의 분자 브레이크는 보다 바람직하게, 순 음전하를 갖지 않는 카르복시-말단 (C-말단) 영역을 포함하는 단일 가닥 결합성 단백질 (SSB)이 아니거나 또는 C-말단 영역의 순 음전하를 감소시키는 그의 C-말단 영역 내에 하나 이상의 변형을 포함하는 변형된 SSB가 아니다. 하나 이상의 분자 브레이크는 가장 바람직하게, 국제 출원 번호 PCT/GB2013/051924 (WO 2014/013259로서 공개됨)에 개시된 SSB 중 임의의 것이 아니다.

[0263] 하나 이상의 분자 브레이크는 바람직하게 하나 이상의 폴리뉴클레오타이드 결합성 단백질이다. 폴리뉴클레오타이드 결합성 단백질은 RNA 폴리뉴클레오타이드와 결합할 수 있고, RNA 폴리뉴클레오타이드가 세공을 통하여 이동하는 것을 제어할 수 있는 임의의 단백질일 수 있다. 특정 단백질이 폴리뉴클레오타이드와 결합하는 지의 여부를 결정하는 것은 관련 기술분야에서 간단한 일이다. 이러한 단백질은 전형적으로, 폴리뉴클레오타이드와 상호 작용하고 폴리뉴클레오타이드의 적어도 하나의 특성을 변형시킨다. 상기 단백질은 폴리뉴클레오타이드를 절단하여 개개의 뉴클레오타이드 또는 더 짧은 쇄의 뉴클레오타이드, 예컨대 디- 또는 트리뉴클레오타이드를 형성시킴으로써, 이러한 폴리뉴클레오타이드를 변형시킬 수 있다. 모이어터는 폴리뉴클레오타이드를 특이적 위치로 배향시키거나 또는 이동하게 함으로써, 즉 그의 이동을 제어함으로써 폴리뉴클레오타이드를 변형시킬 수 있다.

[0264] 폴리뉴클레오타이드 결합성 단백질은 바람직하게, 폴리뉴클레오타이드 핸들링 효소로부터 유래된다. 하나 이상의 분자 브레이크는 상기 논의된 폴리뉴클레오타이드 핸들링 효소 중 임의의 것으로부터 유래될 수 있다. 분자 브레이크로서 작용하는 Phi29 폴리머라제 (서열식별번호: 8)의 변형된 버전이 미국 특허 번호 5,576,204에

개시된다. 하나 이상의 분자 브레이크는 바람직하게 헬리카제로부터 유래된다.

- [0265] 헬리카제로부터 유래된 임의의 수의 분자 브레이크를 사용할 수 있다. 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과인 헬리카제를 분자 브레이크로서 사용할 수 있다. 2개 이상의 헬리카제가 분자 브레이크로서 사용되는 경우, 이러한 2개 이상의 헬리카제는 전형적으로 동일한 헬리카제이다. 2개 이상의 헬리카제는 상이한 헬리카제일 수 있다.
- [0266] 2개 이상의 헬리카제는 상기 언급된 헬리카제의 임의의 조합물일 수 있다. 2개 이상의 헬리카제는 2개 이상의 Dda 헬리카제일 수 있다. 2개 이상의 헬리카제는 하나 이상의 Dda 헬리카제와 하나 이상의 TrwC 헬리카제일 수 있다. 2개 이상의 헬리카제는 동일한 헬리카제의 상이한 변이체일 수 있다.
- [0267] 2개 이상의 헬리카제는 바람직하게 서로 부착된다. 2개 이상의 헬리카제는 보다 바람직하게 서로 공유적으로 부착된다. 헬리카제는 임의의 순서로 임의의 방법을 이용하여 부착될 수 있다. 헬리카제로부터 유래된 하나 이상의 분자 브레이크는 폴리뉴클레오티드 결합성 도메인 내의 개구부의 크기를 감소시키도록 변형시키는데, 이를 통하여 적어도 하나의 입체 형태적 상태에서는 폴리뉴클레오티드가 헬리카제로부터 결합 해제될 수 있다. 이는 WO 2014/013260에 개시된다.
- [0268] 본 발명에 사용하기 바람직한 헬리카제 구축물이 국제 출원 번호 PCT/GB2013/051925 (WO 2014/013260으로서 공개됨); PCT/GB2013/051924 (WO 2014/013259로서 공개됨) 및 PCT/GB2013/051928 (WO 2014/013262로서 공개됨); 및 UK 출원 번호 1318464.3 (2013년 10월 18일에 출원됨)에 기재되어 있다.
- [0269] 하나 이상의 헬리카제가 활성 모드로 사용되는 경우 (즉, 하나 이상의 헬리카제에, 이동을 촉진시키는 데 필요한 성분, 예를 들어 ATP 및  $Mg^{2+}$  모두가 제공되는 경우), 하나 이상의 분자 브레이크는 바람직하게 (a) 불활성 모드로 사용되거나 (즉, 이동을 촉진시키는 데 필요한 성분의 부재하에 사용되거나 또는 활성적으로 이동할 수 없다), (b) 하나 이상의 분자 브레이크가 하나 이상의 헬리카제와 반대 방향으로 이동하는 활성 모드로 사용되거나 또는 (c) 하나 이상의 분자 브레이크가 하나 이상의 헬리카제와 동일한 방향으로 이동하고 하나 이상의 헬리카제보다 더 느리게 이동하는 활성 모드로 사용된다.
- [0270] 하나 이상의 헬리카제가 불활성 모드로 사용되는 경우 (즉, 하나 이상의 헬리카제에, 이동을 촉진시키는 데 필요한 성분, 예를 들어 ATP 및  $Mg^{2+}$  모두가 제공되지 않거나 또는 활성적으로 이동할 수 없는 경우), 하나 이상의 분자 브레이크는 바람직하게 (a) 불활성 모드로 사용되거나 (즉, 이동을 촉진시키는 데 필요한 성분의 부재하에 사용되거나 또는 활성적으로 이동할 수 없다) 또는 (b) 하나 이상의 분자 브레이크가 세공을 통한 폴리뉴클레오티드와 동일한 방향으로 폴리뉴클레오티드를 따라 이동하는 활성 모드로 사용된다.
- [0271] 하나 이상의 헬리카제와 하나 이상의 분자 브레이크는 임의의 위치에서 RNA에 부착시켜, 이들을 합치게 하고 둘다가 세공을 통한 RNA의 이동을 제어하도록 할 수 있다. 하나 이상의 헬리카제와 하나 이상의 분자 브레이크는 적어도 하나의 뉴클레오티드 정도 떨어져 있고, 예컨대 적어도 5개, 적어도 10개, 적어도 50개, 적어도 100개, 적어도 500개, 적어도 1,000개, 적어도 5,000개, 적어도 10,000개, 적어도 50,000개 또는 그 초과인 뉴클레오티드 정도 떨어져 있다. 본 방법이 한 말단에 Y 어댑터가 제공되고 다른 말단에 헤어핀 루프 어댑터가 제공된 이중 가닥 RNA 폴리뉴클레오티드의 특징을 규명하는 것에 관한 것이면, 하나 이상의 헬리카제는 바람직하게 Y 어댑터에 부착되고, 하나 이상의 분자 브레이크는 바람직하게 헤어핀 루프 어댑터에 부착된다. 이러한 실시양태에서, 하나 이상의 분자 브레이크는 바람직하게, 헬리카제가 RNA 폴리뉴클레오티드와는 결합하지만, 헬리카제로서 기능하지 않도록 변형되는 하나 이상의 헬리카제이다. Y 어댑터에 부착된 하나 이상의 DNA 헬리카제는 바람직하게, 다음에 보다 상세히 논의되는 바와 같은 스페이스에서 지연된다. 헤어핀 루프 어댑터에 부착된 하나 이상의 분자 브레이크는 바람직하게, 스페이스에서 지연되지 않는다. 하나 이상의 DNA 헬리카제가 헤어핀 루프에 도달할 때 하나 이상의 DNA 헬리카제와 하나 이상의 분자 브레이크를 합치는 것이 바람직하다. Y 어댑터를 폴리뉴클레오티드에 부착시키기 전 또는 Y 어댑터를 폴리뉴클레오티드에 부착시킨 후에 하나 이상의 DNA 헬리카제를 Y 어댑터에 부착시킬 수 있다. 헤어핀 루프 어댑터를 폴리뉴클레오티드에 부착시키기 전 또는 헤어핀 루프 어댑터를 폴리뉴클레오티드에 부착시킨 후에 하나 이상의 분자 브레이크를 헤어핀 루프 어댑터에 부착시킬 수 있다.
- [0272] 하나 이상의 헬리카제와 하나 이상의 분자 브레이크는 바람직하게 서로 부착시키지 않는다. 하나 이상의 헬리카제와 하나 이상의 분자 브레이크는 보다 바람직하게 서로 공유적으로 부착시키지 않는다. 하나 이상의 헬리카제와 하나 이상의 분자 브레이크는 바람직하게, 국제 출원 번호 PCT/GB2013/051925 (WO 2014/013260으로서 공개됨); PCT/GB2013/051924 (WO 2014/013259로서 공개됨) 및 PCT/GB2013/051928 (WO 2014/013262로서 공개됨);



및 UK 출원 번호 1318464.3 (2013년 10월 18일에 출원됨)에 기재된 바와 같이 부착시키지 않는다.

[0273] 스페이서(들)

[0274] 하나 이상의 스페이서가 본 발명의 구축물에 포함될 수 있다. RNA 폴리뉴클레오타이드의 일부가 세공 내로 유입되고 인가된 전위로부터 발생한 장을 따라 세공을 통해 이동하는 경우, RNA 폴리뉴클레오타이드가 세공을 통해 이동함에 따라, 하나 이상의 헬리카제가 세공에 의해 스페이서를 지나 이동된다. 이는 RNA 폴리뉴클레오타이드 (하나 이상의 스페이서를 포함함)가 세공을 통해 이동되고 하나 이상의 헬리카제가 세공의 상단에 남아있기 때문이다. 하나 이상의 DNA 헬리카제는 국제 출원 번호 PCT/GB2014/050175 (WO 2014/135838로서 공개됨)에 논의된 바와 같이 하나 이상의 스페이서에서 지연될 수 있다. 상기 국제 출원에 개시된 하나 이상의 헬리카제 및 하나 이상의 스페이서의 임의의 입체 배치를 본 발명에 사용할 수 있다.

[0275] 하나 이상의 스페이서는 표적 RNA 폴리뉴클레오타이드의 일부일 수 있고, 예를 들어 상기 폴리뉴클레오타이드 서열을 방해한다. 하나 이상의 스페이서는 바람직하게, 표적 RNA와 혼성화된 하나 이상의 차단성 분자, 예컨대 과속 방지턱의 일부가 아니다. 하나 이상의 스페이서는 비-RNA 폴리뉴클레오타이드(예를 들어, DNA 폴리뉴클레오타이드)의 일부일 수 있고, 예를 들어 상기 폴리뉴클레오타이드 서열을 방해한다. 하나 이상의 스페이서는 RNA 폴리뉴클레오타이드의 일부일 수 있다. 하나 이상의 스페이서는 표적 RNA 폴리뉴클레오타이드 및/또는 비-RNA 폴리뉴클레오타이드에 부착될 수 있다. 하나 이상의 스페이서는 RNA 폴리뉴클레오타이드 또는 비-RNA 폴리뉴클레오타이드의 말단에 위치 설정될 수 있고/있거나 하나 이상의 스페이서는 RNA 폴리뉴클레오타이드 또는 비-RNA 폴리뉴클레오타이드 내에 위치 설정될 수 있다.

[0276] 임의의 수의 스페이서, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과인 스페이서가 표적 RNA 폴리뉴클레오타이드 또는 비-RNA 폴리뉴클레오타이드에 존재할 수 있다. 바람직하게 2개, 4개 또는 6개의 스페이서가 본 발명의 구축물에 존재한다. 이러한 구축물의 상이한 영역 내에 하나 이상의 스페이서가 존재할 수 있는데, 예컨대 리더 서열 내에 하나의 스페이서가 존재하고 헤어핀 루프 내에 하나의 스페이서가 존재할 수 있다.

[0277] 하나 이상의 스페이서는 각각 하나 이상의 헬리카제가 활성 모드에서도 극복할 수 없는 에너지 장벽을 제공한다. 하나 이상의 스페이서는 헬리카제의 견인을 저하시킴으로써 (예를 들어, 표적 RNA 폴리뉴클레오타이드 또는 비-RNA 폴리뉴클레오타이드 내의 뉴클레오타이드로부터 염기를 제거함으로써) 또는 하나 이상의 헬리카제의 이동을 물리적으로 차단시킴으로써 (예를 들어, 벌기 화학적 기를 이용함), 하나 이상의 헬리카제를 지연시킬 수 있다.

[0278] 하나 이상의 스페이서는 하나 이상의 헬리카제를 지연시키는 임의의 분자 또는 분자의 조합물을 포함할 수 있다. 하나 이상의 스페이서는 하나 이상의 헬리카제가 표적 RNA 폴리뉴클레오타이드를 따라 이동하지 못하게 하는 임의의 분자 또는 분자의 조합물을 포함할 수 있다. 하나 이상의 헬리카제가 막힘단 세공 및 인가된 전위의 부재하에 하나 이상의 스페이서에서 지연되는 지의 여부를 결정하는 것은 간단한 일이다. 예를 들어, 헬리카제가 스페이서를 지나 이동할 수 있는 능력은 PAGE에 의해 측정될 수 있다.

[0279] 하나 이상의 스페이서는 전형적으로 선형 분자, 예컨대 중합체를 포함한다. 하나 이상의 스페이서는 전형적으로, 표적 RNA 폴리뉴클레오타이드 또는 비-RNA 폴리뉴클레오타이드와 상이한 구조를 갖는다. 예를 들어, 하나 이상의 스페이서는 전형적으로 RNA가 아니다. 특히, 하나 이상의 스페이서는 바람직하게 펩티드 핵산 (PNA), 글리세롤 핵산 (GNA), 트레오스 핵산 (TNA), 고정 핵산 (LNA), 또는 뉴클레오타이드 측쇄를 수반한 합성 중합체를 포함한다. 하나 이상의 스페이서는 폴리뉴클레오타이드로부터 반대 방향으로 하나 이상의 뉴클레오타이드를 포함할 수 있다. 예를 들어, 폴리뉴클레오타이드가 5'에서 3' 방향으로 있는 경우에는, 하나 이상의 스페이서가 하나 이상의 뉴클레오타이드를 3'에서 5' 방향으로 포함할 수 있다. 이러한 뉴클레오타이드는 상기 논의된 것들 중 임의의 것일 수 있다.

[0280] 하나 이상의 스페이서는 바람직하게 하나 이상의 니트로인돌, 예컨대 하나 이상의 5-니트로인돌, 하나 이상의 이노신, 하나 이상의 아크리딘, 하나 이상의 2-아미노퓨린, 하나 이상의 2-6-디아미노퓨린, 하나 이상의 5-브로모-데옥시우리딘, 하나 이상의 역위 티미딘 (역위 dT), 하나 이상의 역위 디데옥시-티미딘 (ddT), 하나 이상의 디데옥시-시티딘 (ddC), 하나 이상의 5-메틸시티딘, 하나 이상의 5-히드록시메틸시티딘, 하나 이상의 2'-O-메틸 RNA 염기, 하나 이상의 이소-데옥시시티딘 (이소-dC), 하나 이상의 이소-데옥시구아노신 (이소-dG), 하나 이상의 iSpC3 기 (즉, 당 및 염기가 결합된 뉴클레오타이드), 하나 이상의 광절단 가능한 (PC) 기, 하나 이상의 핵산 디올 기, 하나 이상의 스페이서 9 (iSp9) 기, 하나 이상의 스페이서 18 (iSp18) 기, 중합체 또는 하나 이상의 티올 연결물을 포함한다. 하나 이상의 스페이서는 이들 기의 임의의 조합물을 포함할 수 있다. 이들 기 중 다



수가 IDT® [인티그레이티드 DNA 테크놀로지(Integrated DNA Technologies®)]로부터 시판되고 있다.

- [0281] 하나 이상의 스페이서는 임의의 수의 이들 기를 함유할 수 있다. 예를 들어, 2-아미노퓨린, 2-6-디아미노퓨린, 5-브로모-데옥시우리딘, 역위 dT, ddT, ddC, 5-메틸시티딘, 5-히드록시메틸시티딘, 2'-O-메틸 RNA 염기, 이소-dC, 이소-dG, iSpC3 기, PC 기, 핵산디올 기 및 티올 연결물의 경우에는, 하나 이상의 스페이서가 바람직하게 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12개 또는 그 초과를 포함한다. 하나 이상의 스페이서는 바람직하게 2, 3, 4, 5, 6, 7, 8개 또는 그 초과 iSp9 기를 함유한다. 하나 이상의 스페이서는 바람직하게 2, 3, 4, 5 또는 6 개 또는 그 초과 iSp18 기를 포함한다. 가장 바람직한 스페이서는 4개의 iSpC3 기이다.
- [0282] 중합체는 바람직하게 폴리펩티드 또는 폴레테틸렌 글리콜 (PEG)이다. 폴리펩티드는 바람직하게 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12개 또는 그 초과 아미노산을 포함한다. PEG는 바람직하게 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12개 또는 그 초과 단량체 단위를 포함한다.
- [0283] 하나 이상의 스페이서는 바람직하게 하나 이상의 탈염기성 뉴클레오티드 (즉, 핵염기가 결합된 뉴클레오티드), 예컨대 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12개 또는 그 초과 탈염기성 뉴클레오티드를 포함한다. 핵염기는 탈염기성 뉴클레오티드 내의 -H (idSp) 또는 -OH에 의해 대체될 수 있다. 탈염기성 스페이서는 하나 이상의 인접한 뉴클레오티드로부터 핵염기를 제거함으로써 표적 폴리뉴클레오티드 내로 삽입할 수 있다. 예를 들어, 폴리뉴클레오티드는 3-메틸아데닌, 7-메틸구아닌, 1,N6-에테노아데닌 이노신 또는 히포크산틴을 포함하도록 변형시킬 수 있고, 핵염기는 인간 알킬아데닌 DNA 글리코실라제 (hAAG)를 사용하여 이들 뉴클레오티드로부터 제거할 수 있다. 또 다른 한편으로, 폴리뉴클레오티드는 우라실을 포함하도록 변형시킬 수 있고, 핵염기는 우라실-DNA 글리코실라제 (UDG)를 사용하여 제거한다. 한 실시양태에서, 하나 이상의 스페이서는 임의의 탈염기성 뉴클레오티드를 포함하지 않는다.
- [0284] 하나 이상의 DNA 헬리카제는 각 선형 분자 스페이서에 의해 (즉, 스페이서 앞에서) 또는 이러한 스페이서 상에서 지연될 수 있다. 선형 분자 스페이서가 사용되는 경우, 구축물에는 바람직하게, 하나 이상의 헬리카제가 그를 지나 이동되어야 하는 각 스페이서의 말단에 인접한 폴리뉴클레오티드의 이중 가닥 영역이 제공된다. 이중 가닥 영역은 전형적으로, 상기 인접한 스페이서 상의 하나 이상의 헬리카제를 지연시키는 데 도움을 준다. 본 방법이 약 100 mM 이하의 염 농도에서 수행되는 경우에는, 이중 가닥 영역(들)의 존재가 특히 바람직하다. 각 이중 가닥 영역은 전형적으로, 그 길이가 적어도 10개, 예컨대 적어도 12개의 뉴클레오티드이다. 본 발명에 사용된 표적 폴리뉴클레오티드가 단일 가닥인 경우, 더 짧은 폴리뉴클레오티드를 스페이서에 인접한 영역과 혼성화시킴으로써 이중 가닥 영역이 형성될 수 있다. 이러한 더 짧은 폴리뉴클레오티드는 전형적으로, 표적 폴리뉴클레오티드와 동일한 뉴클레오티드로부터 형성되지만, 상이한 뉴클레오티드로부터 형성될 수 있다. 예를 들어, 더 짧은 폴리뉴클레오티드는 LNA로부터 형성될 수 있다.
- [0285] 선형 분자 스페이서가 사용되는 경우, 구축물에는 바람직하게, 하나 이상의 헬리카제가 그를 지나 이동되어야 하는 말단과 반대의 각 스페이서의 말단에 차단성 분자가 제공된다. 이는 하나 이상의 헬리카제가 각 스페이서 상에 여전히 지연된 채로 있다는 것을 보장하는 데 도움을 줄 수 있다. 이는 또한, 하나 이상의 헬리카제가 용액 중에 확산되는 경우에, 이러한 하나 이상의 헬리카제가 구축물 상에 보유되는 것을 도와줄 수 있다. 차단성 분자는 하나 이상의 헬리카제를 물리적으로 지연시켜 주는, 다음에 논의된 화학적 기 중 임의의 것일 수 있다. 차단성 분자는 폴리뉴클레오티드의 이중 가닥 영역일 수 있다.
- [0286] 하나 이상의 스페이서는 바람직하게, 하나 이상의 헬리카제를 물리적으로 지연시켜 주는 하나 이상의 화학적 기를 포함한다. 하나 이상의 화학적 기는 바람직하게 하나 이상의 펜던트 화학적 기이다. 하나 이상의 화학적 기는 표적 폴리뉴클레오티드 내의 하나 이상의 핵염기에 부착될 수 있다. 하나 이상의 화학적 기는 표적 폴리뉴클레오티드 주쇄에 부착될 수 있다. 임의의 수, 예컨대 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12개 또는 그 초과 이들 화학적 기가 존재할 수 있다. 적합한 기는 형광단, 스트렙타비딘 및/또는 비오틴, 콜레스테롤, 메틸렌 블루, 디니트로페놀 (DNP), 디곡시게닌 및/또는 항-디곡시게닌 및 디벤질시클로옥틴 기를 포함하지만, 이에 제한되지 않는다.
- [0287] 표적 폴리뉴클레오티드 내의 상이한 스페이서는 상이한 지연성 분자를 포함할 수 있다. 예를 들어, 하나의 스페이서는 상기 논의된 선형 분자 중 하나를 포함할 수 있고, 또 다른 스페이서는 하나 이상의 헬리카제를 물리적으로 지연시켜 주는 하나 이상의 화학적 기를 포함할 수 있다. 스페이서는 상기 언급된 선형 분자, 및 하나 이상의 헬리카제를 물리적으로 지연시켜 주는 하나 이상의 화학적 기, 예컨대 하나 이상의 탈염기물 및 형광단 중 임의의 것을 포함할 수 있다.

[0288] 적합한 스페이서는 표적 폴리뉴클레오티드의 유형 및 본 발명의 방법이 수행되는 조건에 따라서 설계될 수 있다. 대부분의 헬리카제는 DNA와 결합하고 이를 따라 이동하므로, DNA가 아닌 어떠한 것을 사용해서도 지연시킬 수 있다. 적합한 분자는 상기 논의되어 있다.

[0289] 본 발명의 방법은 바람직하게, 유리 뉴클레오티드의 존재하에 및/또는 헬리카제 보조인자의 존재하에 수행된다. 이는 다음에 보다 상세히 논의된다. 막횡단 세공 및 인가된 전위의 부재하에서는, 하나 이상의 스페이서가 바람직하게, 유리 뉴클레오티드의 존재하에 및/또는 헬리카제 보조인자의 존재하에 하나 이상의 헬리카제를 지연시킬 수 있다.

[0290] 본 발명의 방법이 다음에 논의되는 바와 같은 유리 뉴클레오티드 및 헬리카제 보조인자의 존재하에 수행되는 경우 (이로써 하나 이상의 헬리카제는 활성 모드로 있게 된다), 하나 이상의 더 긴 스페이서가 전형적으로 사용되어, 하나 이상의 헬리카제를 막횡단 세공과 접촉시키고 전위를 인가하기 전에, 하나 이상의 헬리카제가 표적 폴리뉴클레오티드 상에서 지연되도록 보장한다. 하나 이상의 더 짧은 스페이서는 유리 뉴클레오티드 및 헬리카제 보조인자의 부재하에 사용될 수 있다 (이로써 하나 이상의 헬리카제는 불활성 모드로 있게 된다).

[0291] 염 농도는 또한, 하나 이상의 헬리카제를 지연시킬 수 있는 하나 이상의 스페이서의 능력에 영향을 미친다. 막횡단 세공 및 인가된 전위의 부재하에서는, 하나 이상의 스페이서가 바람직하게, 약 100 mM 이하의 염 농도에서 하나 이상의 헬리카제를 지연시킬 수 있다. 본 발명의 방법에 사용된 염 농도가 더 높을 수록, 전형적으로 사용되는 하나 이상의 스페이서가 더 짧아지고, 그 반대의 경우도 마찬가지이다.

[0292] 특징의 바람직한 조합이 다음 표 1에 제시된다.

[0293] <표 1>

폴리뉴클레오티드	스페이서 조성*	스페이서 길이 (즉, *의 수)	염 []	유리 뉴클레오티드?	헬리카제 보조인자?
DNA	iSpC3	4	1 M	있음	있음
DNA	iSp18	4	100-1000 mM	있음	있음
DNA	iSp18	6	<100-1000 mM	있음	있음
DNA	iSp18	2	1 M	있음	있음
DNA	iSpC3	12	<100-1000 mM	있음	있음
DNA	iSpC3	20	<100-1000 mM	있음	있음
DNA	iSp9	6	100-1000 mM	있음	있음
DNA	idSp	4	1 M	있음	있음

[0294]

[0295] 본 방법은 2개 이상의 헬리카제가 스페이서를 지나 이동하는 것에 관한 것일 수 있다. 이러한 경우, 스페이서의 길이는 전형적으로 후행 헬리카제가, 세공 및 인가된 전위의 부재하에 선도 헬리카제가 스페이서를 지나치지 못하도록 증가시킨다. 상기 방법이 2개 이상의 헬리카제가 하나 이상의 스페이서를 지나 이동하는 것에 관한 것이라면, 상기 논의된 스페이서 길이는 적어도 1.5배, 예컨대 2배, 2.5배 또는 3배 증가될 수 있다. 예를 들어, 상기 방법이 2개 이상의 헬리카제가 하나 이상의 스페이서를 지나 이동하는 것에 관한 것이라면, 상기 표 1의 세 번째 칼럼 내의 스페이서 길이는 1.5배, 2배, 2.5배 또는 3배 증가될 수 있다.

[0296] 막횡단 세공

[0297] 막횡단 세공은 막을 어느 정도 가로지르는 구조이다. 이는 인가된 전위에 의해 구동되는 수화된 이온이 막을

가로 질러 또는 막 내에서 흐르도록 허용한다. 막횡단 세공은 전형적으로, 전체 막을 가로지르므로, 수화된 이온이 막의 한 측면에서 막의 다른 측면으로 흐를 수 있다. 그러나, 막횡단 세공은 막을 가로지를 필요가 없다. 한쪽 말단이 닫혀있을 수 있다. 예를 들어, 상기 세공은 그를 따라 또는 그 내부로 수화된 이온이 흐를 수 있는 막 내의 웰, 갭, 채널, 트렌치 또는 슬릿(slit)일 수 있다.

[0298] 임의의 막횡단 세공이 본 발명에 사용될 수 있다. 세공은 생물학적 또는 인공적일 수 있다. 적합한 세공은 단백질 세공, 폴리뉴클레오티드 세공 및 고체 상태 세공을 포함하지만, 이에 제한되지 않는다. 세공은 DNA 오리гами(origami) 세공일 수 있다 (Langecker et al., Science, 2012; 338: 932-936).

[0299] 막횡단 세공은 바람직하게 막횡단 단백질 세공이다. 막횡단 단백질 세공은 수화된 이온, 예컨대 분석물이 막의 한 측면에서 막의 다른 측면으로 흐르게 허용하는 폴리펩티드 또는 폴리펩티드의 컬렉션이다. 본 발명에서, 막횡단 단백질 세공은 인가된 전위에 의해 구동되는 수화된 이온이 막의 한 측면에서 다른 측면으로 흐를 수 있게 허용하는 세공을 형성할 수 있다. 막횡단 단백질 세공은 바람직하게, 분석물, 예컨대 뉴클레오티드가 막, 예컨대 지질 이중 층의 한 측면에서 다른 측면으로 흐르게 허용한다. 막횡단 단백질 세공은 폴리뉴클레오티드 또는 핵산, 예컨대 DNA 또는 RNA가 세공을 통하여 이동할 수 있도록 한다.

[0300] 막횡단 단백질 세공은 단량체 또는 올리고머일 수 있다. 이러한 세공은 바람직하게, 여러 개의 반복되는 서브유닛, 예컨대 6, 7, 8 또는 9개의 서브유닛으로 구성된다. 세공은 바람직하게 6량체, 7량체, 8량체 또는 9량체 세공이다.

[0301] 막횡단 단백질 세공은 전형적으로, 그를 통하여 이온이 흐를 수 있는 배럴 또는 채널을 포함한다. 이러한 세공의 서브유닛은 전형적으로, 중심 축을 둘러싸고 있고 바닥이 막횡단  $\beta$  배럴 또는 채널 또는 막횡단  $\alpha$ -나선 번들 또는 채널에 기인한다.

[0302] 막횡단 단백질 세공의 배럴 또는 채널은 전형적으로, 분석물, 예컨대 뉴클레오티드, 폴리뉴클레오티드 또는 핵산과의 상호 작용을 촉진시켜 주는 아미노산을 포함한다. 이들 아미노산은 바람직하게, 상기 배럴 또는 채널의 협착 근처에 위치한다. 막횡단 단백질 세공은 전형적으로, 하나 이상의 양전하를 띤 아미노산, 예컨대 아르기닌, 리신 또는 히스티딘, 또는 방향족 아미노산, 예컨대 티로신 또는 트립토판을 포함한다. 이들 아미노산은 전형적으로, 세공과 뉴클레오티드, 폴리뉴클레오티드 또는 핵산 간의 상호 작용을 촉진시켜 준다.

[0303] 본 발명에 따라서 사용하기 위한 막횡단 단백질 세공은  $\beta$ -배럴 세공 또는  $\alpha$ -나선 번들 세공으로부터 유래될 수 있다.  $\beta$ -배럴 세공은  $\beta$ -가닥으로부터 형성되는 배럴 또는 채널을 포함한다. 적합한  $\beta$ -배럴 세공은  $\beta$ -독소, 예컨대  $\alpha$ -헤모리신, 탄저균 독소 및 류코시딘, 및 박테리아의 외막 단백질/포린, 예컨대 미코박테리움 스메그마티스 포린 (Msp), 예를 들어 MspA, MspB, MspC 또는 MspD, CsgG, 외막 포린 F (OmpF), 외막 포린 G (OmpG), 외막 포스포리파제 A 및 네이세리아 자동 수송체 지질단백질 (NalP)을 포함하지만, 이에 제한되지 않는다.  $\alpha$ -나선 번들 세공은  $\alpha$ -나선으로부터 형성되는 배럴 또는 채널을 포함한다. 적합한  $\alpha$ -나선 번들 세공은 내막 단백질 및  $\alpha$  외막 단백질, 예컨대 WZA 및 ClyA 독소를 포함하지만, 이에 제한되지 않는다. 막횡단 세공은 Msp 또는  $\alpha$ -헤모리신 ( $\alpha$ -HL)으로부터 유래될 수 있다.

[0304] 막횡단 단백질 세공은 바람직하게 Msp, 바람직하게 MspA로부터 유래된다. 이러한 세공은 올리고머성일 것이고, 전형적으로 Msp로부터 유래된 7, 8, 9 또는 10개의 단량체를 포함한다. 상기 세공은 동일한 단량체를 포함하는 Msp로부터 유래된 동종-올리고머성 세공일 수 있다. 또 다른 한편으로, 상기 세공은 다른 것과 상이한 적어도 하나의 단량체를 포함하는 Msp로부터 유래된 이종-올리고머성 세공일 수 있다. 바람직하게 상기 세공은 MspA 또는 그의 상동체 또는 파라로그(paralog)로부터 유래된다.

[0305] Msp로부터 유래된 단량체는 전형적으로, 서열식별번호: 2에 제시된 서열 또는 그의 변이체를 포함한다. 서열식별번호: 2는 MspA 단량체의 MS-(B1)8 돌연변이체이다. 이는 다음 돌연변이를 포함한다: D90N, D91N, D93N, D118R, D134R 및 E139K. 서열식별번호: 2의 변이체는 서열식별번호: 2의 서열에서 벗어나고 세공을 형성할 수 있는 그의 능력을 보유하고 있는 아미노산 서열을 갖는 폴리펩티드이다. 세공을 형성할 수 있는 변이체의 능력은 관련 기술분야에 공지된 임의의 방법을 이용하여 검정할 수 있다. 예를 들어, 상기 변이체는 다른 적절한 서브유닛과 함께 친양쪽성 층 내로 삽입할 수 있고, 세공을 형성하기 위해 올리고머화할 수 있는 그의 능력을 결정할 수 있다. 서브유닛을 막, 예컨대 친양쪽성 층 내로 삽입하기 위한 방법이 관련 기술분야에 공지되어 있다. 예를 들어, 서브유닛은 지질 이중 층을 함유하는 용액 중에 정제된 형태로 현탁시켜, 이것이 지질 이중 층으로 확산되고 지질 이중 층과의 결합과 기능적 상태로의 어셈블리에 의해 삽입되도록 한다. 또 다른 한편으로, 서브유닛은 문헌 [M.A. Holden, H. Bayley, J. Am. Chem. Soc. 2005, 127, 6502-6503] 및 국제 출원

번호 PCT/GB2006/001057 (WO 2006/100484로서 공개됨)에 기재된 "선택 및 배치" 방법을 이용하여 막 내로 직접 삽입할 수 있다.

- [0306] 서열식별번호: 2의 아미노산 서열의 전체 길이에 걸쳐, 변이체는 바람직하게, 아미노산 유사성 또는 동일성에 근거하여 상기 서열과 적어도 50% 상동일 것이다. 보다 바람직하게, 변이체는 전체 서열에 걸쳐 서열식별번호: 2의 아미노산 서열과의 아미노산 유사성 또는 동일성에 근거하여 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 및 보다 바람직하게 적어도 95%, 97% 또는 99% 상동일 수 있다. 100개 이상, 예를 들어 125, 150, 175 또는 200개 또는 그 초과에 연속되는 아미노산의 연장물에 걸쳐 적어도 80%, 예를 들어 적어도 85%, 90% 또는 95%의 아미노산 유사성 또는 동일성이 존재할 수 있다 ("하드 상동성").
- [0307] 관련 기술분야의 표준 방법을 이용하여 상동성을 결정할 수 있다. 예를 들어 UWGCG 패키지는 상동성을 계산하기 위해 사용될 수 있는, 예를 들어 그의 디폴트 환경에 근거하여 사용될 수 있는 BESTFIT 프로그램을 제공한다 (Devereux et al. (1984) *Nucleic Acids Research* 12, p387-395). 예를 들어, 문헌 [Altschul S. F. (1993) *J Mol Evol* 36:290-300; Altschul, S.F et al. (1990) *J Mol Biol* 215:403-10]에 기재된 바와 같이, PILEUP 및 BLAST 알고리즘을 사용하여 상동성을 계산하거나 또는 서열을 정렬할 수 있다 [예컨대, (전형적으로 그들의 디폴트 환경에 근거하여) 등가의 잔기 또는 상응하는 서열을 확인한다]. BLAST 분석을 수행하기 위한 소프트웨어는 미국 국립 생명공학 정보 센터 (<http://www.ncbi.nlm.nih.gov/>)를 통하여 공개적으로 입수 가능하다.
- [0308] 유사성은 쌍을 이룬 동일성을 이용하거나 또는 스코어링 매트릭스, 예컨대 BLOSUM62를 적용하고 이를 등가의 동일성으로 전환시킴으로써 측정할 수 있다. 진화된 변화 라기보다는 기능적 변화를 나타내기 때문에, 의도적으로 돌연변이된 위치는 상동성을 결정할 때 가려질 것이다. 유사성은, 예를 들어 단백질 서열의 포괄적 데이터 베이스에 대한 PSIBLAST를 사용하는 위치 특이적 스코어링 매트릭스의 적용에 의해 보다 민감하게 결정될 수 있다. 진화론적 시간 척도 (예를 들어, 전하)에 걸쳐 치환 빈도가 아닌 아미노산의 화학 물리적 특성을 반영하는 상이한 스코어링 매트릭스를 사용할 수 있었다.
- [0309] 서열식별번호: 2는 MspA 단량체의 MS-(B1)8 돌연변이체이다. 변이체는 MspA와 비교해서 MspB, C 또는 D 단량체 내에 임의의 돌연변이를 포함할 수 있다. 성숙한 형태의 MspB, C 및 D가 서열식별번호: 5 내지 7에 제시된다. 특히, 변이체는 MspB에 존재하는 다음 치환을 포함할 수 있다: A138P. 변이체는 MspC에 존재하는 다음 치환 중 하나 이상을 포함할 수 있다: A96G, N102E 및 A138P. 변이체는 MspD에 존재하는 다음 돌연변이 중 하나 이상을 포함할 수 있다: G1의 결실, L2V, E5Q, L8V, D13G, W21A, D22E, K47T, I49H, I68V, D91G, A96Q, N102D, S103T, V104I, S136K 및 G141A. 변이체는 Msp B, C 및 D로부터의 돌연변이 및 치환 중 하나 이상의 조합을 포함할 수 있다. 변이체는 바람직하게 돌연변이 L88N을 포함한다. 서열식별번호: 2의 변이체는 MS-(B1)8의 모든 돌연변이 이외에 돌연변이 L88N을 가지는데, 이는 MS-(B2)8로 지칭된다. 본 발명에 사용된 세공은 바람직하게 MS-(B2)8이다. 추가의 바람직한 변이체는 돌연변이 G75S/G77S/L88N/Q126R을 포함한다. 서열식별번호: 2의 변이체는 MS-(B1)8의 모든 돌연변이 이외에 돌연변이 G75S/G77S/L88N/Q126R을 갖는데, 이는 MS-(B2C)8로 지칭된다. 본 발명에 사용된 세공은 바람직하게 MS-(B2)8 또는 MS-(B2C)8이다.
- [0310] Msp로부터 유래된 단량체는 관련 기술분야에 공지된 표준 방법을 이용하여 생성될 수 있다. Msp로부터 유래된 단량체는 합성적으로 또는 재조합 수단에 의해 만들어질 수 있다. 예를 들어, 세공은 시험관내 번역 및 전사 (IVTT)에 의해 합성될 수 있다. 세공을 생성하는 데 적합한 방법은 국제 출원 번호 PCT/GB09/001690 (WO 2010/004273으로서 공개됨), PCT/GB09/001679 (WO 2010/004265로서 공개됨) 또는 PCT/GB10/000133 (WO 2010/086603으로서 공개됨)에 논의되어 있다. 세공을 막 내로 삽입하는 방법이 또한 상기 문헌에 논의되어 있다.
- [0311] 막횡단 단백질 세공은 또한 바람직하게  $\alpha$ -헤모리신 ( $\alpha$ -HL)으로부터 유래된다. 야생형  $\alpha$ -HL 세공은 7개의 동일한 단량체 또는 서브유닛으로 형성된다 (즉, 이는 7량체이다).  $\alpha$ -헤모리신-NN의 하나의 단량체 또는 서브유닛의 서열이 서열식별번호: 4에 제시된다. 막횡단 단백질 세공은 바람직하게, 각각 서열식별번호: 4에 제시된 서열 또는 그의 변이체를 포함하는 7개의 단량체를 포함한다. 서열식별번호: 4의 아미노산 1, 7 내지 21, 31 내지 34, 45 내지 51, 63 내지 66, 72, 92 내지 97, 104 내지 111, 124 내지 136, 149 내지 153, 160 내지 164, 173 내지 206, 210 내지 213, 217, 218, 223 내지 228, 236 내지 242, 262 내지 265, 272 내지 274, 287 내지 290 및 294가 루프 영역을 형성한다. 서열식별번호: 4의 잔기 113 및 147은  $\alpha$ -HL의 배럴 또는 채널의 협착의 일부를 형성한다.
- [0312] 이러한 실시양태에서, 각각 서열식별번호: 4에 제시된 서열 또는 그의 변이체를 포함하는 7개의 단백질 또는 단



량체를 포함하는 세공이 바람직하게 본 발명의 방법에 사용된다. 이러한 7개의 단백질은 동일하거나 (동종-7량체) 또는 상이할 수 있다 (이종-7량체).

[0313] 변이체는 헬리카제 또는 구축물에 대한 공유 부착 또는 헬리카제 또는 구축물과의 상호 작용을 촉진시켜 주는 변형을 포함할 수 있다. 변이체는 바람직하게, 헬리카제 또는 구축물에 대한 부착을 촉진시켜 주는 하나 이상의 반응성 시스테인 잔기를 포함한다. 예를 들어, 변이체는 서열식별번호: 4의 위치 8, 9, 17, 18, 19, 44, 45, 50, 51, 237, 239 및 287 중 하나 이상에서 및/또는 아미노 또는 카르복시 말단 상에 시스테인을 포함할 수 있다. 바람직한 변이체는 서열식별번호: 4의 위치 8, 9, 17, 237, 239 및 287에서의 잔기를 시스테인으로 치환시킨 것을 포함한다 (A8C, T9C, N17C, K237C, S239C 또는 E287C). 변이체는 바람직하게, 국제 출원 번호 PCT/GB09/001690 (WO 2010/004273으로서 공개됨), PCT/GB09/001679 (WO 2010/004265로서 공개됨) 또는 PCT/GB10/000133 (WO 2010/086603으로서 공개됨)에 기재된 변이체 중 어느 하나이다.

[0314] 변이체는 또한, 뉴클레오티드와의 임의의 상호 작용을 촉진시켜 주는 변형을 포함할 수 있다.

[0315] 변이체는 유기체, 예를 들어 스태필로코쿠스 (*Staphylococcus*) 박테리움에 의해 자연적으로 발현되는 자연 발생적 변이체일 수 있다. 또 다른 한편으로, 변이체는 시험관 내에서 발현되거나 또는 박테리움, 예컨대 에스케리키아 콜라이에 의해 재조합적으로 발현될 수 있다. 변이체는 또한, 재조합 기술에 의해 생성된 비-자연 발생적 변이체를 포함한다. 서열식별번호: 4의 아미노산 서열의 전체 길이에 걸쳐, 변이체는 바람직하게, 아미노산 동일성에 근거하여 상기 서열과 적어도 50% 상동일 것이다. 보다 바람직하게, 변이체 폴리펩티드는 전체 서열에 걸쳐 서열식별번호: 4의 아미노산 서열과의 아미노산 동일성에 근거하여 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 및 보다 바람직하게 적어도 95%, 97% 또는 99% 상동일 수 있다. 200개 이상, 예를 들어 230, 250, 270 또는 280개 또는 그 초과에 연속되는 아미노산의 연장물에 걸쳐 적어도 80%, 예를 들어 적어도 85%, 90% 또는 95%의 아미노산 동일성이 존재할 수 있다 ("하드 상동성"). 상동성은 상기 논의된 바와 같이 결정될 수 있다.

[0316] 아미노산 치환은 상기 논의된 것 이외에도, 서열식별번호: 4의 아미노산 서열에 대해 이루어질 수 있는데, 예를 들어 1, 2, 3, 4, 5, 10, 20 또는 30개 이하의 치환이 이루어질 수 있다. 보존적 치환이 이루어질 수 있다.

[0317] 서열식별번호: 4의 아미노산 서열의 하나 이상의 아미노산 잔기는 부가적으로, 상기 언급된 폴리펩티드로부터 결실될 수 있다. 1, 2, 3, 4, 5, 10, 20 또는 30개 이하의 잔기 또는 그 초과에 잔기가 결실될 수 있다.

[0318] 변이체는 서열식별번호: 4의 단편일 수 있다. 이러한 단편은 세공-형성 활성을 보유한다. 단편은 그 길이가 적어도 50, 100, 200 또는 250개의 아미노산일 수 있다. 단편은 바람직하게, 서열식별번호: 4의 세공-형성성 도메인을 포함한다. 단편은 전형적으로 서열식별번호: 4의 잔기 119, 121, 135, 113 및 139를 포함한다.

[0319] 하나 이상의 아미노산은 또 다른 한편으로 또는 부가적으로, 상기 언급된 폴리펩티드에 부가될 수 있다. 연장은 서열식별번호: 4의 아미노산 서열 또는 그의 변이체 또는 단편의 아미노 말단 또는 카르복시 말단에 제공될 수 있다. 연장은 다소 짧을 수 있는데, 예를 들어 그 길이가 1개 내지 10개의 아미노산일 수 있다. 또 다른 한편으로, 연장은 더 길 수 있는데, 예를 들어 50개 또는 100개 이하의 아미노산일 수 있다. 담체 단백질을 세공 또는 변이체와 융합시킬 수 있다.

[0320] 상기 논의된 바와 같이, 서열식별번호: 4의 변이체는 서열식별번호: 4의 서열에서 벗어나고 세공을 형성할 수 있는 그의 능력을 보유하고 있는 아미노산 서열을 갖는 서브유닛이다. 변이체는 전형적으로, 세공 형성에 대해 책임이 있는 서열식별번호: 4의 영역을 함유한다.  $\beta$ -배럴을 함유하는  $\alpha$ -HL의 세공 형성 능력은 각 서브유닛 내의  $\beta$ -가닥에 의해 제공된다. 서열식별번호: 4의 변이체는 전형적으로,  $\beta$ -가닥을 형성하는 서열식별번호: 4 내의 영역을 포함한다.  $\beta$ -가닥을 형성하는 서열식별번호: 4의 아미노산은 상기 논의되어 있다. 하나 이상의 변형은  $\beta$ -가닥을 형성하는 서열식별번호: 4의 영역에 대해 이루어질 수 있는데, 단 이로써 생성되는 변이체는 세공을 형성할 수 있는 그의 능력을 보유해야 한다. 서열식별번호: 4의  $\beta$ -가닥 영역에 대해 이루어질 수 있는 특이적 변형은 상기 논의되어 있다.

[0321] 서열식별번호: 4의 변이체는 바람직하게, 그의  $\alpha$ -나선 및/또는 루프 영역 내에서의 하나 이상의 변형, 예컨대 치환, 부가 또는 결실을 포함한다.  $\alpha$ -나선 및 루프를 형성하는 아미노산은 상기 논의되어 있다.

[0322] 변이체는 상기 논의된 바와 같이 그의 확인 또는 정제를 도와주기 위해 변형될 수 있다.

[0323]  $\alpha$ -HL로부터 유래된 세공은 Msp로부터 유래된 세공과 관련하여 상기 논의된 바와 같이 만들어질 수 있다.



- [0324] 막
- [0325] 임의의 막이 본 발명에 따라서 사용될 수 있다. 적합한 막은 관련 기술분야에 널리 공지되어 있다. 막은 바람직하게 친양쪽성 층이다. 친양쪽성 층은 적어도 하나의 친수성 부분과 적어도 하나의 친지성 또는 소수성 부분 둘 다를 갖는 친양쪽성 분자, 예컨대 인지질로부터 형성된 층이다. 친양쪽성 분자는 합성 또는 자연 발생적일 수 있다. 비-자연 발생적 친양쪽성체, 및 단층을 형성하는 친양쪽성체는 관련 기술분야에 공지되어 있고, 예를 들어 블록 공중합체를 포함한다 (Gonzalez-Perez et al., Langmuir, 2009, 25, 10447-10450). 블록 공중합체는 2개 이상의 단량체 서브유닛을 함께 중합시켜 단일 중합체 쇄를 창출시키는 중합체성 물질이다. 블록 공중합체는 전형적으로, 각 단량체 서브유닛에 의해 기인되는 특성을 갖는다. 그러나, 블록 공중합체는 개개의 서브유닛으로부터 형성된 중합체가 보유하고 있지 않은 독특한 특성을 가질 수 있다. 블록 공중합체는 단량체 서브유닛 중 하나가 소수성 (즉, 친지성)인 반면, 다른 서브유닛(들)이 수성 매질 중에서 친수성이 되도록 조작될 수 있다. 이러한 경우, 블록 공중합체는 친양쪽성 특성을 보유할 수 있고, 생물학적 막을 모방하는 구조를 형성할 수 있다. 블록 공중합체는 디블록 (2개의 단량체 서브유닛으로 이루어짐)일 수 있지만, 2개 초과인 단량체 서브유닛으로부터 구축되어, 친양쪽성체로서 행동하는 더 복잡한 배열을 형성할 수 있다. 상기 공중합체는 트리블록, 테트라블록 또는 펜타블록 공중합체일 수 있다.
- [0326] 친양쪽성 층은 단층 또는 이중 층일 수 있다. 친양쪽성 층은 전형적으로 평면 지질 이중 층 또는 지지된 이중 층이다.
- [0327] 친양쪽성 층은 전형적으로 지질 이중 층이다. 지질 이중 층은 세포 막의 모델이고, 일정 범위의 실험 연구를 위한 탁월한 플랫폼으로서 제공된다. 예를 들어, 지질 이중 층은 단일-채널 기록에 의한 막 단백질의 시험관내 조사를 위해 사용될 수 있다. 또 다른 한편으로, 지질 이중 층은 일정 범위의 물질의 존재를 검출하기 위한 바이오센서로서 사용될 수 있다. 지질 이중 층은 임의의 지질 이중 층일 수 있다. 적합한 지질 이중 층은 평면 지질 이중 층, 지지된 이중 층 또는 리포솜을 포함하지만, 이에 제한되지 않는다. 지질 이중 층은 바람직하게 평면 지질 이중 층이다. 적합한 지질 이중 층은 국제 출원 번호 PCT/GB08/000563 (WO 2008/102121로서 공개됨), 국제 출원 번호 PCT/GB08/004127 (WO 2009/077734로서 공개됨) 및 국제 출원 번호 PCT/GB2006/001057 (WO 2006/100484로서 공개됨)에 개시된다.
- [0328] 지질 이중 층을 형성하는 방법은 관련 기술분야에 공지되어 있다. 적합한 방법이 본 실시예에 개시된다. 지질 이중 층은 통상적으로 문헌 [Montal and Mueller, Proc. Natl. Acad. Sci. USA., 1972; 69: 3561-3566]의 방법에 의해 형성되는데, 여기서는 지질 단층이 그 계면에 수직인 구멍의 어느 한쪽을 지나서 수성 용액/공기 계면에 운반된다.
- [0329] 상기 문헌[Montal & Mueller]의 방법이 대중적인데, 이는 단백질 세공 삽입에 적합한 양질의 지질 이중 층을 형성하는 데 있어 비용면에서 효과적이고 비교적 간단한 방법이기 때문이다. 이중 층 형성의 다른 통상적 방법은 틱-디핑, 페인팅 이중 층 및 리포솜 이중 층의 패치-클램핑을 포함한다.
- [0330] 바람직한 실시양태에서, 지질 이중 층은 국제 출원 번호 PCT/GB08/004127 (WO 2009/077734로서 공개됨)에 기재된 바와 같이 형성된다.
- [0331] 또 다른 바람직한 실시양태에서, 막은 고체 상태 층이다. 고체 상태 층은 생물학적 기원의 것이 아니다. 다시 말하면, 고체 상태 층은 생물학적 환경, 예컨대 유기체 또는 세포, 또는 생물학적으로 이용 가능한 구조의 합성적으로 제작된 버전으로부터 유래되지 않거나 또는 단리되지 않는다. 고체 상태 층은 마이크로 전자 재료, 절연 재료, 예컨대  $\text{Si}_3\text{N}_4$ ,  $\text{Al}_2\text{O}_3$ , 및  $\text{SiO}_2$ , 유기 및 무기 중합체, 예컨대 폴리아미드, 플라스틱, 예컨대 테플론 (Teflon®) 또는 탄성중합체, 예컨대 2-성분 부가 경화 실리콘 고무 및 유리를 포함하지만, 이에 제한되지 않는 유기 재료와 무기 재료 둘 다로부터 형성될 수 있다. 고체 상태 층은 단일 원자 층, 예컨대 그래핀, 또는 단지 몇 개의 원자 두께의 층으로부터 형성될 수 있다. 적합한 그래핀 층은 국제 출원 번호 PCT/US2008/010637 (WO 2009/035647로서 공개됨)에 개시된다.
- [0332] 본 방법은 전형적으로, (i) 세공을 포함하는 인공적 친양쪽성 층, (ii) 세공을 포함하는 단리된 자연 발생적 지질 이중 층, 또는 (iii) 그 내부에 세공이 삽입된 세포를 이용하여 수행된다. 상기 방법은 전형적으로, 인공적 친양쪽성 층, 예컨대 인공적 지질 이중 층을 이용하여 수행된다. 이러한 층은 세공 이외에도, 다른 막횡단 및/또는 막내 단백질 뿐만 아니라 다른 분자를 포함할 수 있다. 적합한 장치 및 조건이 다음에 논의된다. 본 발명의 방법은 전형적으로 시험관 내에서 수행된다.

- [0333] 커플링
- [0334] 표적 RNA 폴리뉴클레오타이드는 바람직하게, 막횡단 세공을 포함하는 막과 커플링시킨다. 이는 임의의 공지된 방법을 이용하여 수행될 수 있다. 본 방법은 표적 RNA 폴리뉴클레오타이드를, 막횡단 세공을 포함하는 막과 커플링시키는 것을 포함할 수 있다. RNA 폴리뉴클레오타이드는 바람직하게, 하나 이상의 앵커를 사용하여 막과 커플링시킨다. RNA 폴리뉴클레오타이드는 임의의 공지된 방법을 이용하여 막과 커플링시킬 수 있다.
- [0335] 각 앵커는 RNA 폴리뉴클레오타이드와 커플링 (또는 결합)되는 기 및 막과 커플링 (또는 결합)되는 기를 포함한다. 각 앵커는 RNA 폴리뉴클레오타이드 및/또는 막과 공유적으로 커플링 (또는 결합)될 수 있다. Y 어댑터 및/또는 헤어핀 루프 어댑터가 사용되는 경우, RNA는 바람직하게 어댑터(들)를 사용하여 막과 커플링된다.
- [0336] RNA 폴리뉴클레오타이드는 임의의 수의 앵커, 예컨대 2, 3, 4개 또는 그 초과 앵커를 이용하여 막과 커플링시킬 수 있다. 예를 들어, RNA 폴리뉴클레오타이드는, 각각 RNA 폴리뉴클레오타이드와 막 둘 다와 별도로 커플링 (또는 결합)되는 2개의 앵커를 이용하여 막과 커플링시킬 수 있다.
- [0337] 하나 이상의 앵커는 상기 논의된 하나 이상의 DNA 헬리카제 및/또는 하나 이상의 분자 브레이크를 포함할 수 있다.
- [0338] 막이 친양쪽성 층, 예컨대 지질 이중 층 (상기에 상세히 논의된 바와 같음)인 경우, RNA는 바람직하게, 막에 존재하는 폴리펩티드를 통하여 또는 막에 존재하는 소수성 앵커를 통하여 막과 커플링시킨다. 소수성 앵커는 바람직하게 지질, 지방산, 스테롤, 탄소 나노튜브 또는 아미노산이다.
- [0339] RNA 폴리뉴클레오타이드는 막과 직접적으로 커플링시킬 수 있다. RNA 폴리뉴클레오타이드는 국제 출원 번호 PCT/GB2012/051191 (WO 2012/164270으로서 공개됨)에 개시된 방법 중 임의의 것을 이용하여 막과 커플링시킬 수 있다. RNA 폴리뉴클레오타이드는 바람직하게 링커를 통하여 막과 커플링된다. 바람직한 링커는 중합체, 예컨대 폴리뉴클레오타이드, 폴레에틸렌 글리콜 (PEG) 및 폴리펩티드를 포함하지만, 이에 제한되지 않는다. RNA가 막과 직접적으로 커플링되면, 특정 규명 실행이 막과 세공 및/또는 폴리뉴클레오타이드 결합성 단백질 간의 거리로 인해 RNA의 말단까지 지속될 수 없기 때문에 일부 데이터가 상실될 것이다. 링커를 사용하면, RNA에 대한 처리를 완료할 수 있다. 링커를 사용하면, 이러한 링커는 임의의 위치에서 RNA에 부착될 수 있다. 링커는 전형적으로, 테일 중합체에서 RNA에 부착된다.
- [0340] 커플링은 안정적 또는 일시적일 수 있다. 특정의 적용의 경우에는, 커플링의 일시적 성질이 바람직하다. 안정적인 커플링 분자가 RNA의 5' 또는 3' 말단 중 하나에 직접적으로 부착된 경우, 특정 규명 실행이 막과 세공 및/또는 폴리뉴클레오타이드 결합성 단백질 간의 거리로 인해 상보적 폴리뉴클레오타이드의 말단까지 지속될 수 없기 때문에 일부 데이터가 상실될 것이다. 커플링이 일시적인 경우, 커플링된 말단이 무작위로 막에 없게 될 때, RNA 폴리뉴클레오타이드에 대한 처리를 완료할 수 있다. 막과 안정적 또는 일시적 링크를 형성하는 화학적 기는 다음에 보다 상세히 논의된다. RNA 폴리뉴클레오타이드는 콜레스테롤 또는 지방 아실 쇄를 이용하여 친양쪽성 층, 예컨대 지질 이중 층과 일시적으로 커플링될 수 있다. 6개 내지 30개 탄소 원자의 길이를 갖는 임의의 지방 아실 쇄, 예컨대 헥사데칸산을 사용할 수 있다.
- [0341] 폴리뉴클레오타이드를 합성 지질 이중 층과 커플링시키는 것은 각종 상이한 테더링 전략을 이용하여 기존에 수행되었다. 이들이 다음 표 2에 요약되어 있다.

[0342] <표 2>

부착 기	커플링의 유형	참고 문헌
티올	안정적	Yoshina-Ishii, C. and S. G. Boxer (2003). "Arrays of mobile tethered vesicles on supported lipid bilayers." <u>J Am Chem Soc</u> 125(13): 3696-7.
비오틴	안정적	Nikolov, V., R. Lipowsky, et al. (2007). "Behavior of giant vesicles with anchored DNA molecules." <u>Biophys J</u> 92(12): 4356-68
콜레스테롤	일시적	Pfeiffer, I. and F. Hook (2004). "Bivalent cholesterol-based coupling of oligonucleotides to lipid membrane assemblies." <u>J Am Chem Soc</u> 126(33): 10224-5
지질	안정적	van Lengerich, B., R. J. Rawle, et al. "Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions." <u>Langmuir</u> 26(11): 8666-72

[0343]

[0344]

합성 폴리뉴클레오티드는 합성 반응에서 변형된 포스포르아미다이트를 이용하여 기능화할 수 있는데, 이는 적합한 앵커링 기, 예컨대 티올, 콜레스테롤, 지질 및 비오틴 기를 추가하는 데 용이하게 화합성이다. 이러한 상이한 부착 화학은 폴리뉴클레오티드에 부착하기 위한 일련의 옵션을 제공한다. 각각의 상이한 변형 기는 약간 상이한 방식으로 폴리뉴클레오티드와 커플링되고, 커플링이 항상 영구적인 것은 아니므로 막에 대한 폴리뉴클레오티드의 체류 시간이 상이하다. 일시적 커플링의 이점이 상기 논의되어 있다.

[0345]

RNA 폴리뉴클레오티드의 커플링은 또한, 수많은 다른 수단에 의해 달성될 수 있는데, 단 반응성 기를 RNA 폴리뉴클레오티드에 부착할 수 있다.

[0346]

또 다른 한편으로, 반응성 기는 막과 이미 커플링된 것에 상보적인 RNA 폴리뉴클레오티드 내의 짧은 영역인 것으로 간주될 수 있으므로, 부착은 혼성화를 통해 달성될 수 있다. 이러한 영역은 RNA 폴리뉴클레오티드의 일부일 수 있거나 또는 이에 라이게이션될 수 있다. T4 RNA 리가제 I을 사용하여 ssDNA의 짧은 조각을 라이게이션하는 것이 보고되었다 (Troutt, A. B., M. G. McHeyzer-Williams, et al. (1992). "Ligation-anchored PCR: a simple amplification technique with single-sided specificity." Proc Natl Acad Sci U S A 89(20): 9823-5).

[0347]

가장 바람직하게, RNA는, 그에 부착된 RNA 폴리뉴클레오티드 또는 비-RNA 폴리뉴클레오티드와 혼성화되는 콜레스테롤-태그부착된 폴리뉴클레오티드를 이용하여 막과 커플링시킨다.

[0348]

질환 또는 병태의 진단 또는 예측

[0349]

mRNA는 바람직하게, 특정 질환 또는 병태를 진단 또는 예측하기 위해 본 발명에 사용된다. 일부 질환 또는 병태는 변경된 양 (또는 수준)의 mRNA와 연관이 있다. 이러한 mRNA는 정상 또는 야생형 mRNA일 수 있는데, 즉 대체 스플라이싱되지 않을 수 있다. mRNA의 양 (또는 수준)은 질환 또는 병태가 없는 환자에서의 양 (또는 수준)과 비교해서 상기 질환 또는 병태에서 증가되거나 또는 감소될 수 있다. 이러한 질환 또는 병태는 본 발명의 방법을 이용하여 환자로부터 획득된 샘플 중의 mRNA의 양을 결정함으로써 진단 또는 예측할 수 있다.

[0350]

많은 유전 질환 또는 병태는 대체 mRNA 스플라이싱, 예컨대 mRNA 스플라이싱 결함을 유발시키는 돌연변이에 의해 야기된다. 수많은 질환 또는 병태는 명백한 돌연변이에 기인하지 않는 대체 mRNA 스플라이싱과 연관이 있다. 대체 스플라이싱의 존재 또는 부재는 본 발명의 방법을 이용하여 환자로부터의 샘플 중의 대체 스플라이싱된 mRNA의 존재 또는 부재를 결정함으로써 확인될 수 있다. 일부 경우에, 대체 mRNA 스플라이싱은 세포의 정상적 기능일 수 있다. 이러한 경우에, 정상적인 양 (즉, 질환 또는 병태가 없는 환자 내에서의 양)과 비교해서 증가되거나 또는 감소된 양 (또는 수준)의 대체 스플라이싱된 mRNA를 사용하여 질환 또는 병태를 진단 또는 예측할 수 있다.

[0351]

본 발명은 특정 환자에게서 메신저 RNA (mRNA)의 변경된 양 및/또는 대체 스플라이싱과 연관된 질환 또는 병태를 진단 또는 예측하는 방법을 제공한다. 본 발명은 환자에게 메신저 RNA (mRNA)의 변경된 양 및/또는 대체 스플라이싱과 연관된 질환 또는 병태가 있는지의 여부 또는 이러한 질환 또는 병태가 발생할 위험이 있는지의

여부를 결정하는 방법을 제공한다. 각 경우에, 이러한 방법은 본 발명의 방법을 이용하여 환자로부터의 샘플 중의 mRNA의 양 및/또는 실체를 결정하는 것을 포함한다. 질환 또는 병태는 다음에 논의된 것 중 임의의 것일 수 있다. 질환 또는 병태는 바람직하게, 낭포성 섬유증, 가족성 자율신경이상증, 전두측두엽 치매, 근위축성 측삭 경화증, 휴킨슨-길포드(Hutchinson-Gilford) 조로증 증후군, 중간 쇠 아실-CoA 데히드로게나제 (MCAD) 결핍증, 근긴장성 디스트로피, 프래더-윌리(Prader-Willi) 증후군, 척수 근디스트로피, 타우병증(tauopathy), 고콜레스테롤혈증 또는 암이다. 이들 질환, 그의 원인 및 가능한 치료법이 문헌 [Tazi et al., Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, Volume 1792, Issue 1, January 2009, Pages 14-26]에 논의되어 있다.

[0352] 환자로부터의 샘플 중의 mRNA의 변경된 (즉, 증가되거나 또는 감소된) 양 (또는 수준)의 존재는 전형적으로 질환 또는 병태를 진단 또는 예측해 주는데, 즉 환자에게 이러한 질환 또는 병태가 있거나 또는 질환 또는 병태가 발생할 위험이 있다는 것을 표시한다. 환자로부터의 샘플 중의 mRNA의 변경된 (즉, 증가되거나 또는 감소된) 양 (또는 수준)의 부재는 전형적으로, 환자에게 이러한 질환 또는 병태가 없거나 또는 질환 또는 병태가 발생할 위험이 없다는 것을 표시한다. mRNA의 양은 상기 논의된 바와 같이 결정될 수 있다.

[0353] 환자로부터의 샘플 중의 대체 스플라이싱된 mRNA의 존재는 전형적으로 질환 또는 병태를 진단 또는 예측해 주는데, 즉 환자에게 이러한 질환 또는 병태가 있거나 또는 질환 또는 병태가 발생할 위험이 있다는 것을 표시한다. 환자로부터의 샘플 중의 대체 스플라이싱된 mRNA의 부재는 전형적으로, 환자에게 이러한 질환 또는 병태가 없거나 또는 질환 또는 병태가 발생할 위험이 없다는 것을 표시한다. 대체 스플라이싱된 mRNA의 존재 또는 부재는 상기 논의된 바와 같이 샘플 중의 RNA를 확인함으로써 결정될 수 있다.

[0354] 환자로부터의 샘플 중의 대체 스플라이싱된 mRNA의 증가되거나 또는 감소된 양 (또는 수준)은 전형적으로 질환 또는 병태를 진단 또는 예측해 주는데, 즉 환자에게 이러한 질환 또는 병태가 있거나 또는 질환 또는 병태가 발생할 위험이 있다는 것을 표시한다. (질환 또는 병태가 없는 환자에서의 양 또는 수준과 비교해서) 환자로부터의 샘플 중의 대체 스플라이싱된 mRNA의 양에 있어 변화가 없다는 것은 전형적으로, 환자에게 이러한 질환 또는 병태가 없거나 또는 질환 또는 병태가 발생할 위험이 없다는 것을 표시한다. 대체 스플라이싱된 mRNA의 양은 상기 논의된 바와 같이 결정될 수 있다.

[0355] miRNA는 바람직하게, 특정 질환 또는 병태를 진단 또는 예측하기 위해 본 발명에 사용된다. 본 발명은 miRNA와 연관된 질환 또는 병태를 진단 또는 예측하는 방법을 제공한다. 본 발명은 환자에게 miRNA와 연관된 질환 또는 병태가 있는지의 여부 또는 이러한 질환 또는 병태가 발생할 위험이 있는지의 여부를 결정하는 방법을 제공한다. 이러한 방법은 본 발명의 방법을 이용하여 환자로부터의 샘플 중의 miRNA의 존재 또는 부재를 결정하는 것을 포함한다. 이러한 질환 또는 병태는 다음에 논의된 것 중 임의의 것일 수 있다.

[0356] 환자로부터의 샘플 중의 miRNA의 존재는 전형적으로, 환자에게 이러한 질환 또는 병태가 있거나 또는 질환 또는 병태가 발생할 위험이 있다는 것을 표시한다. 환자로부터의 샘플 중의 miRNA의 부재는 전형적으로, 환자에게 이러한 질환 또는 병태가 없거나 또는 질환 또는 병태가 발생할 위험이 없다는 것을 표시한다. miRNA의 존재 또는 부재는 상기 논의된 바와 같이 샘플 중의 miRNA를 확인함으로써 결정될 수 있다.

[0357] 상기 질환 또는 병태는 바람직하게 암, 관상 동맥성 심장 질환, 심혈관계 질환 또는 패혈증이다. 이러한 질환 또는 병태는 보다 바람직하게, 복부 대동맥류, 급성 림프모구성 백혈병 (ALL), 급성 골수성 백혈병 (AML), 급성 심근 경색, 급성 전골수구성 백혈병 (APL), 선종, 부신피질 암종, 알콜성 간 질환, 알츠하이머병(Alzheimer's disease), 악성 갑상선 암종 (ATC), 불안 장애, 천식, 성상세포종, 아토피성 피부염, 자폐 스펙트럼 장애 (ASD), B-세포 만성 림프구성 백혈병, B-세포 림프종, 벡커(Becker) 근디스트로피 (BMD), 방광암, 뇌 신생물, 유방암, 버킷(Burkitt) 림프종, 심장 비대증, 심근병증, 심혈관계 질환, 소뇌 신경변성증, 자궁 경부암, 담관암종, 진주종(cholesteatoma), 용모막 암종, 만성 림프구성 백혈병, 만성 골수성 백혈병, 만성 췌장염, 결장암종, 결장직장암, 울혈성 심장 질환, 관상 동맥 질환, 코텐(cowden) 증후군, 피부근염 (DM), 당뇨병성 신증, 설사 우세한 과민성 장 증후군, 미만성 큰 B-세포 림프종, 확장된 심근병증, 다운 증후군 (DS), 뒤시엔느(duchenne) 근디스트로피 (DMD), 자궁내막암, 자궁 내막 모양의 자궁 내막 선암종, 자궁내막증, 상피성 난소암, 식도암, 식도 편평 세포 암종, 본태성 혈소판 증가증 (ET), 안검 상완 근디스트로피 (FSHD), 여포성 림프종 (FL), 여포성 갑상선 암종 (FTC), 전두 측두엽 치매, 위암 (위암), 교모세포종, 다발성 교모세포종 (GBM), 신경교종, 사구체 질환, 사구체경화증, 파오종, HBV-관련 간경변, HCV 감염, 두경부암, 두경부 편평 세포 암종 (HNSCC), 청력 상실, 심장 질환, 심부전증, B형 간염, C형 간염, 간세포 암종 (HCC), 폐색성 담관암종, 호지킨(Hodgkin) 림프종, 동종접합성 겸상 적혈구 질환 (HbSS), 헌팅톤병(Huntington's disease) (HD), 고혈압, 하인



두암, 봉입체 근염 (IBM), 인슐린종, 간내 담관암종 (ICC), 신장암, 신장 질환, 후두 암종, 늦은 불면증 (수면 병), 폐의 평활근종, 백혈병, 사지-거들 근디스트로피 유형 2A (LGMD2A), 지방종, 폐 선암종, 폐암, 림프 세포 증식성 질환, 악성 림프종, 악성 흑색종, 악성 중피종 (MM), 외투 세포 림프종 (MCL), 수모세포종, 흑색종, 수 막종, 대사 질환, 미요시형 근병증 (MM), 다발성 골수종 (MM), 다발성 경화증, MYC-재배열된 림프종, 골수형성 이상 증후군, 골수증식성 장애, 심근 경색, 심근 손상, 근종, 비인두 암종 (NPC), 네발린 근병증 (NM), 신염, 신경모세포종 (NB), 호중구 증가증, 니만-피크(Niemann-Pick) 유형 C (NPC) 질환, 비알콜성 지방 간 질환 (NAFLD), 비소세포 폐암 (NSCLC), 비만, 구강암 골육종 난소 암 (OC), 췌장암, 췌장 도관 선암종 (PDAC), 췌장 신생물, 통증 질환, 유두상 갑상선 암종 (PTC), 파킨슨병(Parkinson's disease), PFV-1 감염, 인두 질환, 뇌하 수체 선종, 다낭성 신장 질환, 다낭성 간 질환, 진성 적혈구증가증 (PV), 다발성 근염 (PM), 원발성 담즙성 간 경변 (PBC), 원발성 골수섬유증, 프리온(prion) 질환, 전립선암, 건선성 관절염, 건선, 폐 고혈압, 재발성 난소 암, 신세포 암종, 신장 투명 세포 암종, 색소성 망막염 (RP), 망막모세포종, 횡문근육종, 류마티스성 심장 질환 및 심방 세동, 류마티스성 관절염, 육종, 정신분열증, 폐혈증, 장액성 난소암, 세자리(Sezary) 증후군, 피부 질환, 소세포 폐암, 척수손실 실조, 편평 상피암, T-세포 백혈병, 기형암종, 고환 생식 세포 종양, 지중해 빈혈, 갑상선암, 혀 편평 세포 암종, 투렛(tourette) 증후군, 유형 2 당뇨병, 궤양성 결장염 (UC), 자궁 평활근종 (ULM), 포도막 흑색종, 혈관 질환, 수포성 구내염 또는 발덴스트롬(Waldenstrom) 마크로글로불린혈증 (WM)이다.

[0358] 환자는 상기 언급된 포유류 중 임의의 것일 수 있다. 환자는 바람직하게 인간이다. 환자는 개개인이다.

[0359] 샘플은 상기 언급된 것 중 임의의 것일 수 있다. 샘플은 전형적으로 임의의 조직 또는 체액으로부터 유래된다. 샘플은 전형적으로, 환자의 체액 및/또는 세포를 포함하고, 예를 들어 면봉, 예컨대 구강 면봉을 이용하여 수득 할 수 있다. 샘플은 혈액, 소변, 타액, 피부, 땀 세포 또는 모근 샘플일 수 있거나 또는 이로부터 유래될 수 있다. 표적 RNA는 전형적으로, 이를 본 발명의 방법에 사용하기 전에 샘플로부터 추출된다.

[0360] 본 방법은 환자에게서 질환 또는 병태를 진단하는 것에 관한 것일 수 있는데, 즉 환자에게 질환 또는 병태가 있는지의 여부를 결정하는 것일 수 있다. 환자는 증상성일 수 있다.

[0361] 본 방법은 환자에게서 질환 또는 병태를 예측하는 것에 관한 것일 수 있는데, 즉 환자에게 질환 또는 병태가 발생할 것으로 예상되는지의 여부를 결정하는 것일 수 있다. 환자는 무증상성일 수 있다. 환자는 질환 또는 병태에 대한 유전적 소인을 가질 수 있다. 환자의 가족 구성원 중 1명 이상에게서 질환 또는 병태가 있을 수 있다.

[0362] RNA 폴리뉴클레오타이드의 이동을 개선시키는 방법

[0363] 본 발명은 또한,

[0364] a) (i) RNA 폴리뉴클레오타이드 (여기서, RNA는 비-RNA 폴리뉴클레오타이드를 포함하도록 변형시킨다) 및 (ii) DNA 헬리카제 효소를 제공하는 단계; 및

[0365] b) a)에서 제공된 RNA 폴리뉴클레오타이드 및 DNA 헬리카제 효소를 막형단 세공과 접촉시켜, 이러한 DNA 헬리카제가 막형단 세공에 대한 RNA 폴리뉴클레오타이드의 이동을 제어하도록 하는 단계

[0366] 를 포함하는, 표적 RNA 폴리뉴클레오타이드의 이동이 DNA 헬리카제 효소에 의해 제어될 때, 표적 RNA 폴리뉴클레오타이드를 막형단 세공에 대하여 이동시키는 방법을 제공한다.

[0367] RNA 폴리뉴클레오타이드를 변형시키면, 그에 대한 DNA 헬리카제 결합이 증가된다. 변형된 RNA 폴리뉴클레오타이드에 대한 증가된 DNA 헬리카제 결합은 DNA 헬리카제 결합의 양 또는 수준이, 비-변형된 또는 변형되지 않은 RNA 폴리뉴클레오타이드, 즉 본 발명의 변형 방법에 따라서 변형시키지 않은 RNA에 대해 관찰되는 DNA 헬리카제 결합의 양 또는 수준보다 더 크거나 또는 더 많은 것으로서 정의된다. 표적 RNA 폴리뉴클레오타이드에 대한 DNA 헬리카제의 결합 수준은 관련 기술분야의 통상의 기술자에게 공지되고 일상적인 일상적인 방법을 이용하여 용이하게 시험될 수 있다.

[0368] 바람직하게 DNA 헬리카제 효소는 비-RNA 폴리뉴클레오타이드와 미리 결합된다. 상기 언급된 실시양태 중 임의의 것이 또한, 상기 방법에 적용된다. 예를 들어, 한 실시양태에서, 비-RNA 폴리뉴클레오타이드는 다음 중 적어도 하나를 포함할 수 있다: (i) 5개 이상의 하전된 단위의 중합체; (ii) 대략 20개의 뉴클레오타이드 길이의 차단성-가닥 혼성화 부위; (iii) 1개 이상의 비-RNA 뉴클레오타이드, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 또는 50개의 뉴클레오타이드의 DNA-헬리카제 결합 부위; (iv) 1개 이상의 단위, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과 단위,의 지연 화학, 예컨대 본원에



참조로 포함되는 WO2014/135838에 기재된 바와 같은 Sp18; (v) 대략 30개의 뉴클레오타이드 길이의 테더 혼성화 부위; 및/또는 (vi) 이전의 섹션에 기재된 바와 같이, RNA 폴리뉴클레오타이드에 대한 비-RNA 폴리뉴클레오타이드의 라이게이션을 촉진시켜 주는 서열.

[0369] 본 실시예는 나노 세공을 통한 DNA/RNA 가닥의 이동을 제어하기 위해 DNA 헬리카제를 사용하는 것을 추가로 예시한다. 따라서, 한 실시양태에서, 본원에는 세공을 통하여 서열 분석될 RNA의 능력 또는 효율을 증가시키는 방법이 제공된다.

[0370] 본 발명의 구축물의 생성 방법

[0371] 구축물의 생성 방법은 표적 RNA 폴리뉴클레오타이드를 비-RNA 폴리뉴클레오타이드에 부착시키는 것을 포함한다. 비-RNA 폴리뉴클레오타이드의 적어도 하나의 뉴클레오타이드는 리보뉴클레오타이드가 아닌데, 즉 RNA로부터의 것이 아니다. 따라서 비-RNA 폴리뉴클레오타이드는 적어도 하나의 리보뉴클레오타이드 (또는 RNA 뉴클레오타이드)를 포함할 수 있지만, 또한 부가적으로 비-RNA 뉴클레오타이드 또는 서열, 즉 RNA가 아닌 뉴클레오타이드 또는 뉴클레오타이드의 서열을 포함해야만 한다. 비-RNA 폴리뉴클레오타이드는 상기 언급된 바와 같은 실시양태 중 임의의 것을 포함할 수 있다. 바람직하게 비-RNA 폴리뉴클레오타이드는 DNA 헬리카제 결합 부위 또는 DNA 어댑터를 포함한다. 보다 바람직하게 비-RNA 폴리뉴클레오타이드는 리더 서열을 포함한다.

[0372] 부착 부위 및 방법은 상기 논의된 바와 같이 선택된다. 바람직하게, 부착 방법은 화학적 부착, 공유 부착, 효소적 부착, 혼성화, 합성 방법, 또는 토포이소머라제를 사용하는 것으로부터 선택될 수 있다. RNA 폴리뉴클레오타이드는 2개 이상, 예컨대 2개 또는 3개의 지점에서 비-RNA 폴리뉴클레오타이드에 부착될 수 있다. 부착 방법은 1개, 2개, 3개, 4개, 5개 또는 그 초과와 상이한 부착 방법을 포함할 수 있다. 상기 언급된 부착 방법의 임의의 조합을 본 발명에 따라서 사용할 수 있다.

[0373] 본 방법은 추가로, 구축물이 DNA 헬리카제의 제어하에 나노 세공을 통하여 제어 이동할 수 있는지의 여부를 결정하는 것을 포함할 수 있다. 이를 시험하기 위한 검정은 관련 기술분야의 통상의 기술자에게 공지되어 있다. 나노 세공을 통한 RNA 폴리뉴클레오타이드의 이동이 제어될 수 있는 경우, 본 발명의 구축물이 생성되었다. RNA 폴리뉴클레오타이드의 이동이 제어될 수 없는 경우에는, 본 발명의 구축물이 생성되지 않았다.

[0374] 본 발명의 한 실시양태에서 표적 RNA 폴리뉴클레오타이드를 비-RNA 폴리뉴클레오타이드에 부착시키는 것을 포함하는 구축물의 생성 방법은 접촉 단계 (b) 전에 수행된다. 또 다른 실시양태에서, 본원에는 변형되지 않은 형태의 RNA보다 더 큰 효율로 세공을 통해 서열 분석될 수 있는 변형된 RNA를 생성시키는 방법이 제공된다.

[0375] 본 발명의 변형 방법

[0376] 본 발명은 특정 규명을 위하여, 예컨대 서열 분석을 위하여 표적 RNA 폴리뉴클레오타이드를 변형시키는 방법을 제공한다. 이와 같이 변형된 RNA 폴리뉴클레오타이드는 본 발명에 따라서 그 특징을 규명하거나 또는 서열 분석한다. 이는 상기에 보다 상세히 논의되어 있다.

[0377] 본 방법은 기재된 방법 중 하나 이상을 이용하여 하나 이상의 변형된 RNA 폴리뉴클레오타이드를 형성하는 것을 포함한다. 하나 이상의 변형된 폴리뉴클레오타이드는, 특히 가닥 서열 분석을 이용하여 변형되지 않은 폴리뉴클레오타이드보다 더 용이하게 그 특징이 규명된다.

[0378] 본 발명의 생성물/구축물

[0379] 본 발명은 또한, 본 발명의 변형 방법을 이용하여 변형된 RNA 폴리뉴클레오타이드를 제공한다. 표적 RNA 폴리뉴클레오타이드는 이러한 RNA 폴리뉴클레오타이드를 비-RNA 폴리뉴클레오타이드에 부착시킴으로써 변형시켜 본 발명의 구축물을 형성한다. 표적 RNA의 변형으로 인해, 비-RNA 폴리뉴클레오타이드가 부착되지 않은 변형되지 않은 형태의 RNA 폴리뉴클레오타이드와 DNA 헬리카제 간에 일어나는 상호 작용과 비교해서, DNA 헬리카제와 변형된 RNA 구축물 간의 상호 작용이 반드시 증가된다. 부가적으로 또는 또 다른 한편으로, 비-RNA 폴리뉴클레오타이드를 RNA 폴리뉴클레오타이드에 부가 또는 부착하는 것은, 비-RNA 폴리뉴클레오타이드를 부착시키지 않은 변형되지 않은 형태의 RNA 폴리뉴클레오타이드에 대한 DNA 헬리카제의 특이성과 비교해서, 변형된 RNA 구축물에 대한 DNA 헬리카제의 특이성이 증가된다는 것을 의미한다. 부가적으로 또는 또 다른 한편으로, 비-RNA 폴리뉴클레오타이드를 RNA 폴리뉴클레오타이드에 부가 또는 부착하는 것은, 비-RNA 폴리뉴클레오타이드를 부착시키지 않은 변형되지 않은 형태의 RNA 폴리뉴클레오타이드와 DNA 헬리카제 간에 일어나는 결합과 비교해서, 변형된 RNA 구축물에 대한 DNA 헬리카제 결합이 촉진되고/되거나 변형된 RNA 구축물에 대한 DNA 헬리카제 결합이 증가된다는 것을 의미한다. 부가적으로 또는 또 다른 한편으로, 비-RNA 폴리뉴클레오타이드를 RNA 폴리뉴클레오타이드에 부가 또는 부착하는 것은, 비-

RNA 폴리뉴클레오타이드를 부착시키지 않은 변형되지 않은 형태의 RNA 폴리뉴클레오타이드와 DNA 헬리카제 간에 일어나는 결합과 비교해서, DNA 헬리카제가 변형된 RNA 구축물과 더 효율적이거나 또는 더 강력하게 결합하고 변형된 구축물로부터 이탈될 가능성이 더 적다는 것을 의미한다.

- [0380] 비-RNA 폴리뉴클레오타이드는 RNA가 아닌 임의의 폴리뉴클레오타이드일 수 있다. 비-RNA 폴리뉴클레오타이드는 적어도 하나의 리보뉴클레오타이드를 포함할 수 있지만, 비-RNA 뉴클레오타이드 또는 서열, 즉 RNA가 아닌 뉴클레오타이드 또는 뉴클레오타이드의 서열을 또한 부가적으로 포함해야만 한다. 부착 부위 및 방법은 상기 논의된 바와 같이 선택된다. 비-RNA 폴리뉴클레오타이드는 미리 결합된 DNA 헬리카제를 포함할 수 있거나 또는 포함하지 않을 수 있다. 바람직하게 비-RNA 폴리뉴클레오타이드는 DNA 헬리카제와 결합할 수 있는 영역 (DNA 헬리카제 결합 부위) 또는 DNA 어댑터를 포함한다. 보다 바람직하게 DNA 헬리카제 결합 부위 또는 DNA 어댑터는 나노 세공 내로 우선적으로 삽입되는 리더 서열을 포함한다. 리더 서열은 또한, 상기 논의된 바와 같이 표적 RNA를 하나 이상의 앵커에 연결시키기 위해 사용될 수 있다. 리더 서열은 표적 RNA 폴리뉴클레오타이드에 연결될 수 있다.
- [0381] 본 발명의 구축물은 바람직하게, DNA 헬리카제를 사용하여 MspA 나노 세공을 통하여 전위될 수 있는 DNA/RNA 혼성체 가닥이다.
- [0382] 이러한 구축물은 폴리뉴클레오타이드 가닥 상의 바코딩 섹션을 추가로 포함할 수 있다. 폴리뉴클레오타이드 바코드는 관련 기술분야에 널리 공지되어 있다 (Kozarewa, I. et al., (2011), *Methods Mol. Biol.* 733, p279-298). 바코드는 특이적이고 공지된 방식으로 세공을 통하여 흐르는 전류에 영향을 미치는 폴리뉴클레오타이드의 특이적 서열이다. 바코딩 섹션은 분석물의 명백한 확인을 가능하게 해준다. 바람직하게 바코딩 섹션은 리더 서열과 DNA 헬리카제 결합 부위 사이에 위치한다.
- [0383] 앵커, 예를 들어 DNA 앵커는 상기 언급된 바와 같이, RNA 폴리뉴클레오타이드 또는 비-RNA 폴리뉴클레오타이드와 혼성화될 수 있다. DNA 앵커는 스페이서 및 콜레스테롤을 추가로 포함할 수 있다.
- [0384] RNA 폴리뉴클레오타이드는 폴리 (U) 폴리머라제를 사용하여 추가로 연장될 수 있다. 이로써 완전한 길이의 RNA를 관독할 수 있다.
- [0385] 변형된 RNA 폴리뉴클레오타이드는 사용된 본 발명의 변형 방법(들)에 따라서 각종 형태일 수 있다. 가능한 형태는 다음 중 하나 이상을 포함하지만, 그에 제한되지 않는다:
- [0386] - 예를 들어, 클릭 화학을 이용하여 비-RNA 폴리뉴클레오타이드에 화학적으로 부착된 RNA 폴리뉴클레오타이드;
- [0387] - 비-RNA 폴리뉴클레오타이드에 라이게이션된 RNA 폴리뉴클레오타이드;
- [0388] - 브릿징 모이어티를 이용하거나 이용하지 않으면서, 비-RNA 리더 서열과 혼성화된 RNA 폴리뉴클레오타이드;
- [0389] - 브릿징 모이어티를 이용하거나 이용하지 않으면서, cDNA 서열과 혼성화된 RNA 폴리뉴클레오타이드;
- [0390] - 토포이소머라제를 사용하여 비-RNA 폴리뉴클레오타이드에 부착된 RNA 폴리뉴클레오타이드;
- [0391] - 나머지 비-RNA 폴리뉴클레오타이드와 반대 방향으로 실행되도록 역전된 염기를 수반한 영역을 갖는 비-RNA 폴리뉴클레오타이드에 라이게이션된 RNA 폴리뉴클레오타이드. 바람직하게 비-RNA 폴리뉴클레오타이드는 DNA 폴리뉴클레오타이드이다.
- [0392] 키트
- [0393] 본 발명은 또한, 표적 RNA 폴리뉴클레오타이드의 특징을 규명하기 위한 키트를 제공한다. 이러한 키트는 특징 규명을 위하여 임의의 표적 RNA 폴리뉴클레오타이드에 부착되도록 개조시킨 비-RNA 폴리뉴클레오타이드를 포함한다.
- [0394] 바람직하게, 특징 규명을 위하여 임의의 표적 RNA 폴리뉴클레오타이드에 부착되도록 개조시킨 비-RNA 폴리뉴클레오타이드는 부착된 반응성 기, 예를 들어 클릭 반응성 기를 갖는다. 부착된 반응성 기를 수반한 비-RNA 폴리뉴클레오타이드는 궁극적으로, 선택된 임의의 표적 RNA 폴리뉴클레오타이드와 반응하여 공유 결합을 형성하도록 최종 사용자에게 의해 사용될 수 있다. 바람직하게 표적 RNA 폴리뉴클레오타이드는 또한, 비-RNA 폴리뉴클레오타이드에 부착된 반응성 기와 반응하여 공유 결합을 형성하는, 부착된 반응성 기, 예를 들어 클릭 반응성 기를 갖는다.
- [0395] 또 다른 한편으로 특징 규명을 위하여 임의의 표적 RNA 폴리뉴클레오타이드에 부착되도록 개조시킨 비-RNA 폴리뉴클레오타이드에 리가제가 제공되거나 또는 이러한 비-RNA 폴리뉴클레오타이드는 선택된 표적 RNA 폴리뉴클레오타이드의 임의의 영역과 혼성화되고 cDNA 합성을 위한 출발 지점으로서 작용되기 위해 최종 사용자에게 의해 사용될 수 있는 올리고뉴클레오타이드 또는 프라이머이다.

- [0396] 또 다른 한편으로 특정 규명을 위하여 임의의 표적 RNA 폴리뉴클레오타이드에 부착되도록 개조시킨 비-RNA 폴리뉴클레오타이드는 특이적 DNA 폴리뉴클레오타이드와 결합된 토포이소머라제를 포함한다. 토포이소머라제 결합된 DNA는 궁극적으로, 선택된 임의의 표적 RNA 폴리뉴클레오타이드에 비-RNA (DNA) 폴리뉴클레오타이드를 부착시키기 위해 최종 사용자에게 의해 사용될 수 있다. 최종 사용자는 유리 5' 히드록실을 갖는 RNA와 함께, 상기 토포이소머라제 결합된 DNA를 인큐베이션할 수 있다. 이때 토포이소머라제는 RNA를 DNA에 연결시켜 준다.
- [0397] 또 다른 한편으로 특정 규명을 위하여 임의의 표적 RNA 폴리뉴클레오타이드에 부착되도록 개조시킨 비-RNA 폴리뉴클레오타이드는 역전된 염기의 영역 (예를 들어, 역전된 염기의 DNA 영역)을 포함한다. 이러한 역전된 영역은 반대 방향으로 실행되는 7-메틸구아노신 캡의 부가에 의해 변형되는 진행 RNA의 5' 말단에 부착될 수 있다.
- [0398] 본 발명의 방법과 관련하여 상기 논의된 실시양태 중 임의의 것이 키트에도 동등하게 적용된다. 키트는 비-RNA 폴리뉴클레오타이드와 미리 결합될 수 있는 DNA 헬리카제 결합성 단백질을 추가로 포함할 수 있다. 키트는 세공 및 막의 성분, 예컨대 친양쪽성 층, 예컨대 지질 이중 층을 형성하는 데 필요한 인지질을 추가로 포함할 수 있다.
- [0399] 본 발명의 키트는 상기 언급된 실시양태 중 임의의 것을 수행할 수 있게 해주는 하나 이상의 다른 시약 또는 기구를 부가적으로 포함할 수 있다. 이러한 시약 또는 기구는 다음 중 하나 이상을 포함한다: 적합한 완충제(들) (수성 용액), 대상체로부터 샘플을 수득하기 위한 수단 (예컨대, 용기, 또는 바늘을 포함하는 기구), 상기 정의된 바와 같은 막, 또는 전압 또는 패치 클램프 장치. 시약은 키트 내에 건조 상태로 존재할 수 있으므로, 유체 샘플이 시약을 재현탁시키도록 한다. 키트는 또한 임의로, 이러한 키트가 본 발명의 방법에 사용될 수 있도록 해주는 설명서, 또는 본 발명의 방법이 사용될 수 있는 환자에 관한 세부 사항을 포함할 수 있다. 키트는 전형적으로, 뉴클레오타이드를 포함한다. 키트는 바람직하게 dAMP, dTMP, dGMP 및 dCMP를 포함한다. 키트는 바람직하게, 폴리뉴클레오타이드를 증폭 및/또는 발현시키기 위한 수단을 포함하지 않는다.
- [0400] 다음 실시예가 본 발명을 예시한다.
- [0401] **실시예 1**
- [0402] 본 실시예는 1) RNA/DNA 가닥의 RNA 영역을 연장시키고, 2) 앵커를 어닐링시키며, 3) 효소를 결합시킨 다음, 4) 이로써 생성된 가닥을 전기 생리학 실험에서 시험하는 샘플 제조 과정을 보여준다. 본 실시예는 DNA 헬리카제 [T4 Dda - E94C/A360C (돌연변이 E94C/A360C에 이어서 ( $\Delta M1$ )G1을 수반한 서열식별번호: 14)]를 사용하여, MspA 나노 세공을 통한 합성 DNA/RNA 가닥 (RNA 가닥에 부착된 DNA 리더; 도 3에 도시됨)의 이동을 제어하는 것이 가능하였다는 것을 예시하였다.
- [0403] **물질 및 방법**
- [0404] **1.1 폴리(U) 폴리머라제를 사용하여 DNA/RNA 가닥의 3' 말단을 연장시킨다**
- [0405] 다음 표 3에 열거된 시약을 혼합하고 37°C 하에 10분 동안 인큐베이션하였다. 이어서, 이 혼합물을 샘플 1  $\mu$ L 당 1.8  $\mu$ L SPRI 비드의 비율로 아젠코트 앰퓨어(Agencourt Ampure) SPRI 비드를 이용하여 정제하였다. 이러한 샘플은 샘플 1 (DNA/RNA 2)로서 공지되었다. 도 4는 본 도면에서 Y로 표지된 넓은 밴드가 가변적으로 연장된 DNA/RNA 1에 상응함에 따라, 폴리머라제 연장 반응이 성공적이었다는 것을 도시한 것이다.

[0406] <표 3>

시약	용적	스톡의 농도	최종 농도
합성 DNA/RNA 1 (서열식별번호: 17의 5' 말단에 반대 말단에서 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 16)	0.4 $\mu$ l	100 $\mu$ M	1 $\mu$ M
rUTP	0.4 $\mu$ l	100 mM	1 mM
NE 완충제	4 $\mu$ l	10 x	1 x
폴리 U Pol (NEB)	2 $\mu$ l	2 U/ $\mu$ l	4 U
NF H <sub>2</sub> O	33.2 $\mu$ l		
총	40 $\mu$ l		

[0407]

[0408] **1.2 앵커를 어닐링한다**

[0409] 다음 표 4에 열거된 시약을 혼합하고 65℃ 하에 인큐베이션한 다음, 초당 0.1℃의 속도로 4℃로 냉각시켰다. 이러한 샘플은 샘플 2로서 공지되었다.

[0410] <표 4>

시약	용적	스톡의 농도	최종 농도
샘플 1	9 $\mu$ l	~1 $\mu$ M	942 $\mu$ M
앵커 (그의 3' 말단에서 6 개의 iSp18 스페이서, 2 개의 티민 및 3' 콜레스테롤 TEG 에 부착된 서열식별번호: 18)	0.36 $\mu$ l	100 $\mu$ M	3.77 $\mu$ M
10 mM 트리스 pH 7.5 50 mM NaCl	0.19 $\mu$ l	50 x	1 x
총	9.55 $\mu$ l		

[0411]

[0412] **1.3 DNA 헬리카제를 결합시킨다**

[0413] 샘플 2 (0.28  $\mu$ L)를 실온하에 1시간 동안 완충제 (10 mM 트리스 pH 7.5, 50 mM NaCl) 중의 T4 Dda - E94C/A360C (0.36  $\mu$ L, 3.8  $\mu$ M; 돌연변이 E94C/A360C에 있어서 ( $\Delta$ M1)G1을 수반한 서열식별번호: 14)와 함께 인큐베이션하였다. 이러한 샘플은 샘플 3으로서 공지되었다.

[0414] **1.4 전기 생리학**

[0415] 샘플 3을 완충제 (1221  $\mu$ L의 600 mM KCl, 50 mM HEPES pH 8.0, 463 mM 글리세롤) 내로 희석시켰다. MgCl<sub>2</sub> (13  $\mu$ L, 1 M) 및 ATP (65  $\mu$ L, 100 mM)를 샘플 3 완충제 혼합물에 부가하여 총 1300  $\mu$ L의 용적을 제공하였다.

[0416] 완충제 [25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, 약 pH 8.0] 중의 블록 공중합체 내에 삽입된 단일 MspA 나노 세공으로부터 전기적 측정치를 획득하였다. 상기 블록 공중합체 내에 삽입된 단일 세공을 달성한 후, 완충제 [2 mL, 25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, pH 8.0]를 시스템 내로 유동시켜 임의의 과도한 MspA 나노 세공을 제거하였다.

[0417] 과도한 KCl 완충제 (600 mM KCl, 50 mM HEPES pH 약 8, 463 mM 글리세롤)를 시스템 내로 유동시키고, 이러한 KCl 완충제를 아가로스 브릿지에 의해 전극 완충제 (25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, pH 8.0)로부터 분리시켰다.

[0418] 이 실험은 -120 mV에서 실행하였고, 헬리카제-제어된 DNA 이동을 모니터링하였다.

[0419] **결과**

[0420] 샘플 3 (DNA/RNA 2를 나타낸 만화가 도 3에 도시된다)을 나노 세공 시스템에 부가한 경우에, 헬리카제 제어된 DNA 이동이 관찰되었다. 샘플 3에 대한 헬리카제 제어된 DNA 이동의 한 예가 도 5에 도시된다. 합성 가닥의 각종 영역이, 나노 세공을 통하여 전위된 가닥으로서 확인되었다 [영역 1 = 폴리(dT) 리더 (서열식별번호: 16), 영역 2 = iSpC3 스페이서, 영역 3 = RNA 서열 (서열식별번호: 17) 및 영역 4 = 가변 길이 폴리(U) RNA]. 본 실시예는 DNA 헬리카제를 사용하여, MspA 나노 세공을 통한 DNA/RNA 가닥 (도 3에 도시된 만화)의 이동을 제어하는 것이 가능하였다는 것을 보여준다. 폴리 (U) 폴리머라제 연장 단계는 완전한 길이의 RNA가 관측되었다는 것을 보장해 주었다.

[0421] **실시예 2**

[0422] 본 실시예는 T4 DNA 리가제를 사용하여 DNA 가닥 (서열식별번호: 21)을 RNA 가닥 (서열식별번호: 19)에 라이게이션하는 것을 보여준다.

[0423] **물질 및 방법**

[0424] **2.1 T4 DNA 리가제를 사용하여 DNA 가닥을 RNA 가닥에 라이게이션한다**

[0425] 다음 표 5에 열거된 시약을 혼합하고, 이를 열순환기 위에 놓아두었다. 열순환기는 다음 표 6에서의 프로그램

으로 설정하였다. 이어서, 140 V에서 60분 동안 실행된 10% PAGE TBE-우레아 변성 바이오래드 기준 겔을 이용하여 샘플을 분석하였다.

<표 5>

시약	용적	스톡 농도	최종 농도
5' 포스페이트 RNA (서열식별번호: 19)	0.3 ul	50 uM	1 uM
T4 DNA 리가제 완충제	1.5	10 x	1 x
T4 DNA 리가제	1 ul	10 U/ul	10 U
DNA 스플린트 (서열식별번호: 20)	0.3 ul	50 uM	1 uM
Cy3 DNA (서열식별번호: 21)	0.6 ul	50 uM	2 uM
ATP	0.3 ul	50 mM	1 mM
NF H <sub>2</sub> O	11 ul		
총	15 ul		

<표 6>

순환 수	단계	온도 (°C)	시간
1	라이게이션한다	25	1:00:00
1	변성시킨다	65	0:10:00

## 결과

TBE-우레아 변성 겔을 이용하여, RNA (서열식별번호: 19)에 대한 CY3 DNA (서열식별번호: 21)의 라이게이션을 분석하였다. 도 8, 레인 2 내지 3은 T4 DNA 리가제의 부재하에 증가 농도의 DNA 스플린트 (서열식별번호: 20)를 이용하는 라이게이션 단계의 대조군 반응을 도시하였다. 이들 대조군 반응의 경우에, 영역 A 및 B 내에 밴드가 전혀 관찰되지 않았는데, 이는 이들 조건하에서는 라이게이션 반응이 발생하지 않았다는 것을 표시한다. 레인 5 내지 8은 T4 DNA 리가제의 존재하에 증가 농도의 DNA 스플린트 (서열식별번호: 20)를 이용하는 라이게이션 단계를 도시하였다. 레인 5 내지 8 모두에 대해서는, 영역 A와 영역 B 둘 다에서 밴드가 가시적이었는데, 이는 혼성화된 DNA 스플린트를 수반한 라이게이션된 기질 (A) 및 혼성화된 DNA 스플린트를 수반하지 않은 라이게이션된 기질 (B)에 상응하였다. 스플린트의 농도가 증가함에 따라, 밴드 A의 세기도 증가하였다. 레인 9 및 10에 제시된 2개의 추가의 대조군 반응을 수행하였다. 레인 9는 라이게이션 단계 후 부가된 여분의 DNA 스플린트 (4.5x)의 부가를 이용하여 추가로 처리시킨, 레인 5에 제시된 바와 동일한 샘플에 상응하였다. 이는 레인 5와 비교해서, 상부 밴드 (혼성화된 스플린트를 수반한 라이게이션된 생성물에 상응함)의 세기에 있어서의 증가가 예상되고 하부 밴드 B (혼성화된 스플린트를 수반하지 않은 라이게이션된 생성물에 상응함)의 세기에 있어서의 감소가 예상된다는 것을 보여주었다. 레인 10은 37°C에서 30분 동안 ExoI로 추가로 처리시킨, 레인 6에 제시된 바와 동일한 샘플에 상응하였다. ExoI 처리 동안 가열하면, 상기 라이게이션된 가닥과 DNA 스플린트의 분리가 초래되었다. DNA 스플린트는 ExoI에 의해 우선적으로 소화되었는데, 이는 3' DNA 말단을 갖고 있었기 때문이다. 이로써, DNA 스플린트와 혼성화된 라이게이션된 생성물에 상응하는 영역 A 내의 밴드가 사라졌는데, 이는 DNA 스플린트가 ExoI에 의해 소화되었기 때문이다. 영역 B 내의 밴드는 ExoI을 이용한 소화 후에도 여전히 가시적이었다. 이는 DNA가 RNA에 라이게이션되지 않은 경우, 이와 같이 라이게이션되지 않은 DNA가 ExoI에 의해 소화되었을 것이고 영역 B 내에서는 어떠한 밴드도 가시적이지 않을 것이기 있기 때문에, 라이게이션 단계는 성공적이었다는 것을 의미한다. 따라서, 본 실시예는 T4 DNA 리가제를 사용하여 DNA를 RNA에 라이게이션시키는 것이 가능하였다는 것을 보여주었다.

## 실시예 3

본 실시예는 클릭 화학을 이용하여 DNA 가닥을 RNA 가닥에 화학적으로 부착시키는 샘플 제조 과정을 보여준다. 이는 2개의 상이한 샘플 상에서 수행되었는데, 그 중 하나는 겔 전기영동을 이용하여 화학적 부착 단계를 확증하기 위하여, DNA에 부착된 형광성 기를 갖고 있었다. 이어서, 형광성 기가 부착되지 않은 DNA/RNA 가닥을 전기 생리학 실험에서 시험하였다. 본 실시예는 DNA 헬리카제 [T4 Dda - E94C/A360C (돌연변이 E94C/A360C에 이어서 (ΔM1)G1을 수반한 서열식별번호: 14)]를 사용하여 MspA 또는 리세닌 나노 세공을 통한, 구리-매개된 클릭-화학에 의해 비-RNA 폴리뉴클레오타이드에 부착되는 RNA 가닥 (구축물을 나타내는 만화가 도 10에 도시된다)의



이동을 제어하는 것이 가능하였다는 것을 예시하였다.

#### [0434] 물질 및 방법

#### [0435] 3A.1 RNA X1에 대한 DNA X1의 클릭 반응

[0436] RNA X1, DNA X1 및 스플린트 X1 (다음 표 7에 열거됨)을 완충제 (트리스-NaCl (500 mM-2.5 M) pH 8)에서 혼합하였다. DNA X1, RNA X1 및 스플린트 X1을 PCR 기계에서 어닐링시켰다 (55℃로 가열하고 0.1℃/s 하에 4℃로 냉각시키는 프로토콜). 이어서, CuSO<sub>4</sub>, 트리스(3-히드록시프로필트리아졸릴메틸아민) 및 아스코르브산나트륨 [시그마(Sigma) A4034]를 DNA X1/RNA X1/스플린트 X1 혼합물에 부가한 다음, 샘플을 열순환기 위에 놓아두었다. 열순환기는 다음 표 8에서의 프로그램으로 설정하였다. 이어서, 상기 샘플을 샘플 1 μL당 1.8 μL SPRI 비드의 비율로 아젠코트 앰플어 SPRI 비드를 이용하여 정제하였다. 이러한 샘플은 샘플 3A (DNA/RNA 3A)로서 공지되었다. 이어서, 이러한 샘플을 다음 3.3에 기재된 바와 같이, 5% PAGE TBE 바이오래드 기준 겔 상에서 전기생리학으로 분석하였다.

[0437] <표 7>

시약	용적	스톡 농도	최종 농도
트리링크 바이오테크 (미국 캘리포니아주)로부터의 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오타이드로서 5'-헥시닐-G 를 갖고 3' 폴리 A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임을 수반한 mRNA)	12.9 ul	1.1 uM	0.94 uM
아지드를 수반한 DNA (DNA X1; 3' 말단에 부착된 3 아지드 N 을 갖는 서열식별번호: 23의 5' 말단에 그의 반대 말단에서 부착된 4 개의 iSp18 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 22)	0.3 ul	50 uM	1 uM
스플린트 X1 (서열식별번호: 24)	0.6 ul	50 uM	2 uM
CuSO <sub>4</sub>	0.3 ul	50 mM	1 mM
트리스(3-히드록시프로필트리아졸릴메틸아민)	0.3 ul	100 mM	2 mM
아스코르브산나트륨 (시그마 A4034)	0.3 ul	200 mM	4 mM
트리스-NaCl (500mM-2.5M) pH 8	0.3 ul	500 mM	10 mM
총	15 ul		

[0439] <표 8>

순환 수	단계	온도 (°C)	시간
1	클릭하다	23	0:30:00
1	클릭하다	37	0:30:00

#### [0441] 3B.1 RNA X1에 대한 DNA X2의 클릭 반응

[0442] RNA X1, DNA X2 및 스플린트 X1 (다음 표 9에 열거됨)을 완충제 (트리스 또는 MOPS (500 mM-2.5 M) pH 6.8-7)에서 혼합하였다. DNA X2, RNA X1 및 스플린트 X1을 PCR 기계에서 어닐링시켰다 (55℃로 가열하고 0.1℃/s 하에 4℃로 냉각시키는 프로토콜). 이어서, CuSO<sub>4</sub>, 트리스(3-히드록시프로필트리아졸릴메틸아민) 및 아스코르브산나트륨 (시그마 A4034)을 DNA X2/RNA X1/스플린트 X1 혼합물에 부가한 다음, 샘플을 열순환기 위에 놓아두었다. 열순환기는 다음 표 10에서의 프로그램으로 설정하였다. 이어서, 상기 샘플을 샘플 1 μL당 1.8 μL SPRI 비드의 비율로 아젠코트 앰플어 SPRI 비드를 이용하여 정제하였다. 이러한 샘플은 샘플 3B (DNA/RNA 3B)로서 공지되었다. 이어서, 이러한 샘플을 5% PAGE TBE 바이오래드 기준 겔 상에서 분석하였다.

[0443] <표 9>

시약	용적	스톡 농도	최종 농도
5'-헥시닐-G를 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오타이드로서 5'-헥시닐-G를 갖고 3' 폴리 A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임 수반한 mRNA)	12.9 ul	1.1 uM	0.94 uM
아지드를 수반한 Cy3 DNA (DNA X2; 그의 5' 말단에 부착된 Cy3을 갖고 3' 말단에 부착된 3 아지드 N을 갖는 서열식별번호: 25)	0.3 ul	50 uM	1 uM
스플린트 X1 (서열식별번호: 24)	0.6 ul	50 uM	2 uM
CuSO <sub>4</sub>	0.3 ul	50 mM	1 mM
Cu 리간드	0.3 ul	100 mM	2 mM
아스코르브산나트륨 (시그마 A4034)	0.3 ul	200 mM	4 mM
트리스-NaCl (500mM-2.5M) pH 8	0.3 ul	500 mM	10 mM
총	15 ul		

[0444]

[0445] <표 10>

순환 수	단계	온도 (°C)	시간
1	클릭하다	23	0:30:00
1	클릭하다	37	0:30:00

[0446]

### [0447] 3.3 전기 생리학

[0448] 테더를 실시예 1에 따라서 샘플 3A와 어닐링시켰다 (섹션 1.2 참조). 샘플 3A (0.28  $\mu$ l) 및 T4 Dda - E94C/A360C (0.36  $\mu$ l, 3.8  $\mu$ M; 돌연변이 E94C/A360C에 있어서 ( $\Delta$ M1)G1을 수반한 서열식별번호: 14)를 완충제 (1221  $\mu$ L의 500 mM KCl, 25 mM 인산칼륨 pH 8.0) 내로 희석시켰다. MgCl<sub>2</sub> (13  $\mu$ L, 1 M) 및 ATP (65  $\mu$ L, 100 mM)를 샘플 3A 완충제 혼합물에 부가하여 총 1300  $\mu$ L의 용적을 제공하였다.

[0449] 완충제 [25 mM 인산칼륨 완충제, 150 mM 페로시아화칼륨, 150 mM 페리시아화칼륨, 약 pH 8.0] 중의 블록 공중합체 내에 삽입된 단일 MspA 또는 레세닌 나노 세공으로부터 전기적 측정치를 획득하였다. 상기 블록 공중합체 내에 삽입된 단일 세공을 달성한 후, 완충제 [2 mL, 25 mM 인산칼륨 완충제, 150 mM 페로시아화칼륨, 150 mM 페리시아화칼륨, pH 8.0]를 시스템 내로 유동시켜 임의의 과도한 MspA 또는 레세닌 나노 세공을 제거하였다.

[0450] 샘플을 부가하기 전에, 과도한 완충제 (500 mM KCl, 25 mM 인산칼륨 pH 8.0)를 시스템 내로 유동시켰다. 최종적으로, 샘플 3A와 결합된 T4 Dda - E94C/A360C를 나노 세공 시스템에 부가하고, 실험을 -120 mV에서 실행하였고, 헬리카제-제어된 DNA 이동을 모니터링하였다.

### [0451] 결과

[0452] TBE-우레아 변성 겔을 사용하여, RNA X1에 대한 DNA [DNA X1 (실시예 3A; 3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 23의 5' 말단에 그의 반대 말단에서 부착된 4개의 iSp18 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 22) 또는 DNA X2 (실시예 3B; 그의 5' 말단에 부착된 CY3을 갖고 그의 3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 25)]의 클릭 반응을 분석하였다. 실시예 3B에 사용되었던 DNA는, 부착된 Cy3 기로 인해 SYBR 염색 전에 가시적이었다. 도 9, 레인 3 및 9는 실시예 3B의 클릭 반응 후에 생성된 샘플을 제시하였다. 레인 3의 상부에서의 형광성 밴드 (도 9에서 백색 박스로써 강조 표시됨)는 Cy3 표지를 수반한 DNA가 mRNA에 부착되었다는 것을 나타내었다 (DNA-단독에 대한 형광성 밴드는 화살표 3에 있고, mRNA와 결합된 DNA는 화살표 1에 있다). 레인 2 및 8은 실시예 3A에서 클릭 반응 후에 생성된 샘플을 나타내었다 (DNA 상에 형광성 기가 없다). 레인 11 내의 밴드는 반응되지 않은 DNA에 상응하였고 (화살표 2), mRNA와 결합된 DNA는 화살표 1이었다. 따라서, 이러한 겔은 클릭 반응이 DNA [형광성 표지를 수반하거나 (X1) 또는 형광성 표지를 수반하지 않음

(X2)]을 mRNA (RNA X1)에 연결시키는 데 성공적이었다는 것을 보여주었다.

[0453] 샘플 3A (DNA/RNA 3A를 나타낸 만화가 도 10에 도시된다)를 나노 세공 시스템에 부가한 경우에, 헬리카제-제어된 RNA 이동이 관찰되었다. MspA를 통한 헬리카제 제어된 RNA 이동의 한 예가 도 11에 도시되고 (영역 2, 3 및 4), 리세닌 돌연변이체를 통한 이동이 도 21에 도시된다. 도 11에서의 가닥의 각종 영역이 나노 세공을 통하여 전위된 가닥으로서 확인되었다. 취합해 보면, 영역 2, 3 및 4는 클릭 반응에 의해 표적 RNA와 반응된 DNA 리더를 나타낸다 (영역 2 = 4개의 iSp18 스페이서와 연결된 서열식별번호: 22; 영역 3 = 서열식별번호: 23; 영역 4 = 서열식별번호: 26의 오픈 리딩 프레임에 수반한 반딧불이 루시페라제 mRNA; 및 \*로 표시된 화살표는 클릭 반응에 의해 형성된 연쇄를 강조 표시한다). 영역 1은 RNA와 반응되지 않은 DNA 리더의 별도의 나노 세공 전위를 나타낸다 (서열식별번호: 23과 연결된 4개의 iSp18 스페이서와 연결된 서열식별번호: 22). 유사한 특색이 도 21에서 리세닌 주적에 대하여 확인될 수 있었다. 본 실시예는 DNA 헬리카제를 사용하여, MspA 또는 리세닌 나노 세공을 통하여 비-RNA 폴리뉴클레오티드 (서열식별번호: 23과 연결된 4개의 iSp18 스페이서와 연결된 서열식별번호: 22)를 긴 mRNA 가닥 (도 10에 도시된 만화)에 연결시킴으로써 나노 세공을 통한 긴 mRNA 가닥의 이동을 제어하는 것이 가능하였다는 것을 보여주었다.

#### [0454] 실시예 4

[0455] 본 실시예는 T4 DNA 리가제를 사용하여 헤어핀-형성성 올리고 (3T 헤어핀 = 서열식별번호: 29는 그의 5' 말단에서 포스페이트 기에 부착되고; 서열식별번호: 27의 5' 말단에 반대 말단에서 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착되거나, 또는 10T 헤어핀 = 서열식별번호: 29는 그의 5' 말단에서 포스페이트 기에 부착되고; 서열식별번호: 28의 5' 말단에 반대 말단에서 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된다)를 RNA 가닥 (서열식별번호: 30)의 3' 말단에 라이게이션한 다음, 상기 헤어핀-형성성 올리고의 3' 말단으로부터 후속 역전사시키는 것을 보여준다. 이러한 라이게이션 및 역전사는 RNA/cDNA 구축물을 구축하는 방법을 명확하게 보여주었다 (도 13 참조).

[0456] 헤어핀-형성성 올리고 (3T 헤어핀 또는 10T 헤어핀)의 3' 말단에서의 폴리 T 오버행은 mRNA의 폴리 A-테일과 혼성화되고, DNA를 RNA에 효율적으로 라이게이션하기 위한 스플린트로서 작용한다. 상기 헤어핀은 후속 역전사를 위한 프라이머로서 작용할 수 있다.

#### [0457] 물질 및 방법

[0458] 다음 표 11에 열거된 시약을 혼합하고, 이를 열순환기 위에 놓아두었다. 열순환기는 다음 표 12에서의 프로그램으로 설정하였다. 이어서, 상기 혼합물을 샘플 1  $\mu$ L당 1.8  $\mu$ L SPRI 비드의 비율로 아젠코트 앰프어 SPRI 비드를 이용하여 정제하였다. 정제 후, 라이프 테크놀로지(Life Technologies) 슈퍼 스크립트 II를 이용하여 역전사를 수행하였다: 표 13에서의 시약을 제조업자의 프로토콜에 따라서 혼합하고, 표 14에서의 프로그램으로 설정된 열순환기 위에 놓아두었다. 이어서, 140 V에서 60분 동안 실행된 10% PAGE TBE-우레아 변성 바이오래드 기준 겔을 이용하여 샘플을 분석하였다.

[0459] <표 11>

시약	용적	스톡 농도	최종 농도
RNA 가닥 (서열식별번호: 30)	0.72 ul	556ng/ul	0.2 uM
폴리 T 헤어핀 (상기 언급된 3T 헤어핀 또는 10T 헤어핀)	0.4 ul	50uM	1 uM
T4 DNA 리가제 완충제	4 ul	5 x	1 x
T4 DNA 리가제	1 ul	2000U	2000 U
NF H <sub>2</sub> O	13.88 ul		
총	20 ul		

[0460]

[0461] <표 12>

순환 수	단계	온도 (°C)	시간
1	라이게이션하다	16	2:00:00

[0462]

&lt;표 13&gt;

시약	용적	스톡 농도	최종 농도
폴리 T 헤어핀 라이게이션 후의 RNA	7 ul	35.5ng/ul	248.5 ng/반응
dNTPs	1 ul	각각 10 uM	0.5 uM
NF H <sub>2</sub> O	5 ul		
제 1-가닥 완충제	4 ul	5x	1x
0.1M DTT	2 ul	0.1M	0.01M
슈퍼 스크립트 II	1ul	200U	200U
총	20 ul		

&lt;표 14&gt;

순환 수	단계	온도 (°C)	시간
1	역전사	42	0:50:00
2	변성	70	0:15:00

**결과**

TBE-우레아 변성 겔 (도 14 참조)을 이용하여 3T 및 10T 헤어핀 라이게이션 및 역전사를 분석하였다. 레인 5에서는, 3T-헤어핀의 부분적 라이게이션이, RNA 가닥 밴드 (서열식별번호: 30; RNA 가닥이 도 14에서 밴드 A로서 표지된다)의 상향 시프트의 형태로 가시적이었다. 역전사 (레인 6) 후, 밴드는 RNA/cDNA의 이중 가닥 구축물 (도 13에 도시됨)이 단일 가닥 RNA 가닥 (도 14에서 밴드 A로서 제시된 서열식별번호: 30)보다 더 신속하게 이주하였다는 사실 때문에 RNA 가닥 (서열식별번호: 30; 도 14에서의 밴드 A)의 수준 아래로 시프트되었다. 레인 7에서는, 10T 헤어핀의 라이게이션이 거의 100% 효율로 발생하였고, 단일 가닥 RNA 가닥 (도 14에서 밴드 A로서 제시된 서열식별번호: 30)의 수준 위의 단일 밴드로서 가시적이었다. 레인 8은 10T 헤어핀에 의해 프라이밍된, 역전사 후의 샘플을 나타내는데, 여기서 라이게이션되고 역전사된 생성물은 단일 가닥 RNA 밴드 (도 14에서 밴드 A로서 제시된 서열식별번호: 30) 아래로 하향 시프트되었다. 레인 9에서의 혼성화 대조군 샘플 (T4 DNA 리가제를 수반하지 않은 반응 혼합물)은 단일 가닥 RNA (도 14에서 밴드 A로서 제시된 서열식별번호: 30)의 수준에서 가시적이었다고, 라이게이션된 생성물에 대해서만 시프트가 관찰된 것으로 나타났다.

**실시예 5**

본 실시예는 RNA/cDNA 구축물을 전기 생리학 실험에서 분석하는 방법을 나타낸다. RNA/cDNA 구축물을 수득하기 위하여, 10T 헤어핀 (서열식별번호: 29는 그의 5' 말단에서 포스페이트 기에 부착되고; 서열식별번호: 28의 5' 말단에 반대 말단에서 부착되는 4개의 iSpC3 스페이서에 그의 3' 말단에서 부착된다)를 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오티드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA)의 3' 말단에 라이게이션하였다. 이러한 라이게이션 이전에, 클릭 화학을 이용하여 실시예 3에 기재된 바와 같이, 아지드를 수반한 DNA (DNA X1; 3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 23의 5' 말단에 그의 반대 말단에서 부착된 4개의 iSp18 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 22)를 5'-헥시닐-G를 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오티드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA)의 5' 말단에 부착시켰다. 헤어핀 라이게이션 후, 역전사를 수행하였다.

**물질 및 방법****클릭 반응**

실시예 3B 클릭 반응에 기재된 바와 같이, 5'-헥시닐-G를 수반한 반딧불이 루시페라제 mRNA (RNA X1; 5'의 최외각 뉴클레오티드로서 5'-헥시닐-G를 갖고 3' 폴리A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임에 수반한 mRNA)의 RNA를, 아지드를 수반한 DNA (DNA X1; 3' 말단에 부착된 3아지드N을 갖는 서열식별번호: 23의 5' 말단에 그의 반대 말단에서 부착된 4개의 iSp18 스페이서에 그의 3' 말단에서 부착된 서열식별번호: 22)에 라이게이션하였다. 이러한 샘플은 샘플 3B로서 공지되었다.

**헤어핀 라이게이션 및 RT**

[0475] 이어서, 실시예 4에 기재된 바와 같이, 샘플 3B (DNA/RNA 3B)를 10T 헤어핀에 라이게이션하고 역전사시켰다. 시약의 용적과 양이 다음 표 15 및 17에 제시되고, 열순환 조건은 표 16 및 18에 제시된다. 라이게이션/역전사 후에 생성된 샘플은 샘플 5B로서 공지되었다. 샘플 구축물을 나타내는 만화가 도 15에 도시된다.

[0476] <표 15>

시약	용적	스톡 농도	최종 농도
샘플 3B	10 ul	22ng/ul	0.017 uM
폴리 T 헤어핀 (상기 언급된 3T 헤어핀 또는 10T 헤어핀)	0.4 ul	50uM	1 uM
T4 DNA 리가제 완충제	4 ul	5 x	1 x
T4 DNA 리가제	1 ul	2000U/ul	2000 U
NF H <sub>2</sub> O	4.6 ul		
총	20 ul		

[0477]

[0478] <표 16>

순환 수	단계	온도 (°C)	시간
1	라이게이션하다	16	2:00:00

[0479]

[0480] <표 17>

시약	용적	스톡 농도	최종 농도
역전사 후 샘플 3B	12 ul	20ng/ul	240ng/반응
dNTPs	1 ul	각각 10uM	0.5 uM
제 1-가닥 완충제	4 ul	5x	1x
0.1M DTT	2 ul	0.1M	0.01M
슈퍼 스크립트 II	1ul	200U/ul	200U
총	20 ul		

[0481]

[0482] <표 18>

순환 수	단계	온도 (°C)	시간
1	역전사	42	0:50:00
2	변성	70	0:15:00

[0483]

[0484] 전기 생리학

[0485] 테더를 실시예 1에 기재된 바와 같이 샘플 5B와 어닐링시켰다 (섹션 1.2 참조). 샘플 5B (4 μl) 및 T4 Dda - E94C/C109A/C136A/A360C (0.36 μl, 3.8 μM; 돌연변이 E94C/C109A/C136A/A360C에 이어서 (ΔM1)G1을 수반한 서열식별번호: 14)를 완충제 (1221 μL의 500 mM KCl, 25 mM 인산칼륨 pH 8.0) 내로 희석시켰다. MgCl<sub>2</sub> (13 μL, 1 M) 및 ATP (65 μL, 100 mM)를 샘플 5B (DNA/RNA/cDNA 5B) 완충제 혼합물에 부가하여 총 1300 μL의 용적을 제공하였다.

[0486] 완충제 [25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, 약 pH 8.0] 중의 블록 공중합체 내에 삽입된 단일 MspA 또는 CsgG-Eco-(Y51T/F56Q)-StrepII(C))9 [StepII(C)가 서열식별번호: 45이고 C-말단에 부착되는, 돌연변이 Y51T/F56Q를 수반한 서열식별번호: 44] 나노 세공으로부터 전기적 측정치를 획득하였다. 상기 블록 공중합체 내에 삽입된 단일 세공을 달성한 후, 완충제 [2 mL, 25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, pH 8.0]를 시스템 내로 유동시켜 임의의 과도한 MspA 또는 CsgG 나노 세공을 제거하였다.

[0487] 샘플을 부가하기 전에, 과도한 완충제 (500 mM KCl, 25 mM 인산칼륨 pH 8.0)를 시스템 내로 유동시켰다. 최종적으로, 샘플 5B (DNA/RNA/cDNA 5B)와 결합된 T4 Dda - E94C/C109A/C136A/A360C를 나노 세공 시스템에 부가하



고, 실험을 -140 mV에서 실행하였고, 헬리카제-제어된 DNA 이동을 모니터링하였다.

## 결과

샘플 5B를 나노 세공 시스템에 부가한 경우에, 헬리카제-제어된 DNA/RNA/cDNA 이동이 관찰되었다. MspA의 경우에는 헬리카제 제어된 DNA/RNA/cDNA 이동 현상의 한 예가 도 16에 도시되거나 (영역 1 내지 5) 또는 CsgG의 경우에는 도 22에 도시된다. 도 16에서의 가닥의 각종 영역이 나노 세공을 통하여 전위된 가닥으로서 확인되었다; 영역 1은 클릭 반응에 의해 표적 RNA와 반응된 비-RNA 폴리뉴클레오티드를 나타내고; 영역 2 = 반딧불이 루시페라제 mRNA; 영역 3 = mRNA에 라이게이션된 10T 헤어핀에 존재하는 iSpC3 스페이서; 영역 4 = mRNA에 라이게이션된 10T 헤어핀의 폴리T 영역; 영역 5 = mRNA의 역전사에 의해 생성된 cDNA. 유사한 특색이 도 22에서 CsgG 추적에 대하여 확인될 수 있다. 본 실시예는 DNA 헬리카제를 사용하여, MspA 또는 CsgG 나노 세공을 통하여 가닥 DNA/RNA/cDNA (DNA 가닥을 mRNA 가닥에 연결시키고, 헤어핀을 그러한 mRNA 가닥에 라이게이션하며, 헤어핀을 사용하여 mRNA를 역전사시킴으로써 형성됨) (도 15에 도시된 만화)의 이동을 제어하는 것이 가능하였다는 것을 보여주었다.

## 실시예 6

본 실시예는 비-RNA 폴리뉴클레오티드를, 라이게이션 전에 캡을 제거함으로써 5' 메틸구아노신 캡 (일부 천연 세포성 mRNA와 상동)을 갖는 RNA 가닥에 부착시킬 수 있는 방법을 보여준다. 이러한 경우에, 캡핑된 RNA 가닥 (5'에서 5'로의 트리포스페이트 연쇄에 의해 상기 가닥의 5' 말단과 연결된 7-메틸구아노신 캡을 갖는 서열식별번호: 30)을 사용하였는데, 이어서 RNA 5' 피로포스포히드롤라제 (RppH)를 사용하여 탈캡핑한 후, 연속해서 T4 RNA 리가제 1을 사용하여 비-RNA 폴리뉴클레오티드 (RNA 서열 CAAGGG에 반대 말단에서 부착된 4개의 5-니트로인돌에 3' 말단에서 부착된 서열식별번호: 32의 5' 말단에 반대 말단에서 부착된 4개의 iSp18 스페이서에 3' 말단에서 부착된 서열식별번호: 31의 5' 말단에 부착된 30개의 SpC3 스페이서)에 라이게이션하였다.

## 물질 및 방법

비-RNA 폴리뉴클레오티드를 캡핑된 RNA 가닥 (5'에서 5'로의 트리포스페이트 연쇄에 의해 상기 가닥의 5' 말단과 연결된 7-메틸구아노신 캡을 갖는 서열식별번호: 30)의 5' 말단에 라이게이션하기 위하여, mRNA를 먼저, 탈캡핑 효소로서 RppH를 사용하여 탈캡핑시켰다. 다음 표 19에 열거된 시약을 혼합한 다음, 반응 혼합물을, 표 20에서의 프로그램으로 설정된 열순환기 위에 놓아두었다. 이어서, 이로써 생성된 반응 혼합물을 샘플 1  $\mu$ L당 1.8  $\mu$ L SPRI 비드의 비율로 아젠코트 앰플어 SPRI 비드를 이용하여 정제하였다. 연속해서, 표 21에 열거된 시약을 혼합하고, 이러한 혼합물을 표 22에서의 프로그램으로 설정된 열순환기 위에 놓아둠으로써, 비-RNA 폴리뉴클레오티드 (RNA 서열 CAAGGG에 반대 말단에서 부착된 4개의 5-니트로인돌에 3' 말단에서 부착된 서열식별번호: 32의 5' 말단에 반대 말단에서 부착된 4개의 iSp18 스페이서에 3' 말단에서 부착된 서열식별번호: 31의 5' 말단에 부착된 30개의 SpC3 스페이서)를 RNA에 라이게이션하였다. 이어서, 반응 혼합물을 140 V에서 60분 동안 실행된 5% PAGE TBE-우레아 변성 바이오래드 기준 겔을 이용하여 분석하였다.

<표 19>

시약	용적	스톡 농도	최종 농도
캡핑된 RNA 가닥 (5'에서 5'로의 트리포스페이트 연쇄에 의해 상기 가닥의 5' 말단과 연결된 7-메틸구아노신 캡을 갖는 서열식별번호: 30)	5.39 $\mu$ l	556ng/ $\mu$ l	2996.8ng/반응
NEB2 완충제	2 $\mu$ l	10x	1x
RNA 5' 피로포스포히드롤라제	2 $\mu$ l	5U/ $\mu$ l	10U
NF H <sub>2</sub> O	10.61 $\mu$ l		
총	20 $\mu$ l		

<표 20>

순환 수	단계	온도 (°C)	시간
1	탈캡핑	37	1:00:00

[0498] <표 21>

시약	용적	스톡 농도	최종 농도
탈캡핑된 RNA 가닥 (서열식별번호: 30)	1.5 ul	166ng/ul	250ng/반응
T4 RNA 리가제 1 반응 완충제	2 ul	10x	1x
비-RNA 폴리뉴클레오티드 (RNA 서열 CAAGGG 에 반대 말단에서 부착된 4 개의 5-니트로인들에 3' 말단에서 부착된 서열식별번호: 32 의 5' 말단에 반대 말단에서 부착된 4 개의 iSp18 스페이서에 3' 말단에서 부착된 서열식별번호: 31 의 5' 말단에 부착된 30 개의 SpC3 스페이서)	2.4 ul	50uM	8.33uM
ATP	0.4 ul	50mM	1mM
NF H <sub>2</sub> O	0.8 ul		
T4 RNA 리가제 1	2.9ul	10U/ul	29U
PEG 8k	10 ul	50%	25%
총	20 ul		

[0499]

[0500] <표 22>

순환 수	단계	온도 (°C)	시간
1	라이게이션	16	4:00:00

[0501]

[0502] **결과**

[0503] TBE-우레아 변성 겔 (도 17 참조)을 이용하여, RNA 가닥 (서열식별번호: 30)의 탈캡핑, 및 비-RNA 폴리뉴클레오티드 (RNA 서열 CAAGGG에 반대 말단에서 부착된 4개의 5-니트로인들에 3' 말단에서 부착된 서열식별번호: 32의 5' 말단에 반대 말단에서 부착된 4개의 iSp18 스페이서에 3' 말단에서 부착된 서열식별번호: 31의 5' 말단에 부착된 30개의 SpC3 스페이서)와의 후속 라이게이션의 성공을 분석하였다. 도 17의 레인 4는 RNA 가닥 (서열식별번호: 30; 레인 2에 제시되고 A로 표지된 대조군)의 수준 위에서 관찰된 부가 밴드 (C로 표지됨)를 나타내는데, 이는 약 40%의 효율로 일어난 비-RNA 폴리뉴클레오티드 (RNA 서열 CAAGGG에 반대 말단에서 부착된 4개의 5-니트로인들에 3' 말단에서 부착된 서열식별번호: 32의 5' 말단에 반대 말단에서 부착된 4개의 iSp18 스페이서에 3' 말단에서 부착된 서열식별번호: 31의 5' 말단에 부착된 30개의 SpC3 스페이서)의 성공적인 라이게이션을 표시한다.

[0504] **실시예 7**

[0505] 본 실시예는 나노 세공을 통한 천연 사카로마이세스 세레비지아에 (*Saccharomyces cerevisiae*) mRNA 가닥의 DNA 헬리카제 제어된 이동을 허용하기 위하여 사카로마이세스 세레비지아로부터 추출된 세포성 mRNA로부터 2D (RNA-센스 및 DNA-안티센스) 라이브러리를 제조하는 방법을 보여준다.

[0506] **물질 및 방법**

[0507] 헤어핀을 mRNA의 3'에 라이게이션하고, mRNA의 5' 말단을 탈캡핑하며, 역전사함으로써 DNA 보체를 창출시키며, 비-RNA 폴리뉴클레오티드를 mRNA의 5' 말단에 라이게이션함으로써, 사카로마이세스 세레비지아로부터 추출된 세포성 mRNA로부터 2D (RNA-센스 및 DNA-안티센스) 라이브러리를 제조하였다.

[0508] **3' 헤어핀을 라이게이션하다**

[0509] 사카로마이세스 세레비지아로부터의 폴리A+ mRNA는 클론테크(Clontech)로부터 구입하였다. 다음 표 23에 제시된 시약을 혼합하고 이러한 혼합물을 다음 표 24에 제시된 프로그램으로 열순환기에 놓아둠으로써, mRNA의 폴리A 테일과 혼성화되는 헤어핀을 mRNA에 라이게이션하였다. 이어서, 혼합물을 샘플 1  $\mu$ L당 1.8  $\mu$ L SPRI 비드의 비율로 아젠코트 앰플어 SPRI 비드를 이용하여 정제하고, 16 ul의 NF H<sub>2</sub>O에서 용출시켰다.

[0510] <표 23>

시약	용적	스톡 농도	최종 농도
폴리 A+ mRNA	0.5 ul	1 ug/ul	500 ng
폴리 T 헤어핀 (서열식별번호: 29 는 그의 5' 말단에서 포스페이트 기에 부착되고; 반대 말단에서 서열식별번호: 28 의 5' 말단에 부착되는 4 개의 iSpC3 스페이서에 그의 3' 말단에서 부착된다)	0.4 ul	25uM	1 uM
T4 DNA 리가제 완충제	4 ul	5 x	1 x
T4 DNA 리가제	1 ul	2000U/ul	2000 U
NF H <sub>2</sub> O	14.1 ul		
총	20 ul		

[0511]

[0512] <표 24>

순환 수	단계	온도 (°C)	시간
1	라이게이션하다	16	0:30:00

[0513]

[0514] 탈캡핑하다

[0515] 비-RNA 폴리뉴클레오타이드를, 5'에서 5'로의 트리포스페이트 연쇄를 통하여 그의 5' 말단과 연결된 7-메틸구아노신 캡을 갖는 RNA 가닥 (서열식별번호: 30)의 5' 말단에 라이게이션하기 전에, RNA-5'-피로포스포히드롤라제 (RppH)에 의해 상기 캡을 효소적으로 제거하였다. 다음 표 25에서의 시약을 혼합하고, 이러한 혼합물을, 다음 표 26에서의 프로그램으로 설정된 열순환기에 놓아둠으로써 탈캡핑을 달성하였다. 이어서, 이로써 생성된 반응 혼합물을 샘플 1 μL당 1.8 μL SPRI 비드의 비율로 아젠코트 앰플어 SPRI 비드를 이용하여 정제하고, 12 ul의 NF H<sub>2</sub>O에서 용출시켰다.

[0516] <표 25>

시약	용적	스톡 농도	최종 농도
효모 폴리 A+ mRNA (클론테크)	16 ul	28ng ug/ul	500 ng/반응
NEB2 완충제	2 ul	10x	1x
RNA 5' 피로포스포히드롤라제	2 ul	5U/ul	10U
총	20 ul		

[0517]

[0518] <표 26>

순환 수	단계	온도 (°C)	시간
1	탈캡핑	37	1:00:00

[0519]

[0520] RT

[0521] 탈캡핑 후, 라이프 테크놀로지 슈퍼 스크립트 II를 이용하여 역전사를 수행하였다: 다음 표 27에서의 시약을 제조업자의 프로토콜에 따라서 혼합하고, 표 28에서의 프로그램으로 설정된 열순환기 위에 놓아두었다.

[0522] <표 27>

시약	용적	스톡 농도	최종 농도
역전사 후 3B (DNA/RNA 3B)	12 ul	~20ng/ul	240ng/반응
dNTPs	1 ul	각각 10 uM	0.5 uM
제 1-가닥 완충제	4 ul	5x	1x
0.1M DTT	2 ul	0.1M	0.01M
슈퍼 스크립트 II	1ul	200U/ul	200U
총	20 ul		

[0523]

[0524] <표 28>

순환 수	단계	온도 (°C)	시간
1	역전사	42	0:50:00
2	변성	70	0:15:00

[0525]

[0526] 이어서, 상기 혼합물을 샘플 1  $\mu\text{L}$ 당 1.8  $\mu\text{L}$  SPRI 비드의 비율로 아젠코트 앰플어 SPRI 비드를 이용하여 정제하고, 10 ul의 NF H<sub>2</sub>O에서 용출시켰다.

[0527] 비-RNA 폴리뉴클레오타이드 라이게이션

[0528] 표 29에 열거된 시약을 혼합하고, 이러한 혼합물을 표 30에서의 프로그램으로 설정된 열순환기에 놓아둠으로써, 비-RNA 폴리뉴클레오타이드 (RNA 서열 CAAGGG에 반대 말단에서 부착된 4개의 5-니트로인돌에 3' 말단에서 부착된 서열식별번호: 32의 5' 말단에 반대 말단에서 부착된 4개의 iSp18 스페이서에 3' 말단에서 부착된 서열식별번호: 31의 5' 말단에 부착된 30개의 SpC3 스페이서)를 역전사된 mRNA에 라이게이션하였다. 비-RNA 폴리뉴클레오타이드 라이게이션 후에 생성된 샘플은 샘플 7A로서 공지되었다.

[0529] 5' 리더는 다음에 의해 라이게이션되었다:

[0530] <표 29>

시약	용적	스톡 농도	최종 농도
역전사된 mRNA	10.0 ul		
비-RNA 폴리뉴클레오타이드 (상기 언급됨)	1.2 ul	50 uM	2 ul
T4 RNA 리가제 완충제	3.0 ul	10x	1x
ATP	0.6 ul	50 mM	1 mM
8k PEG	12.0 ul	50%	
T4 RNA 리가제 1	3.2 ul	10 U/ul	
총	30.0 ul		

[0531]

[0532] <표 30>

순환 수	단계	온도 (°C)	시간
1	라이게이션하다	25	0:30:00

[0533]

[0534] 헤어핀 정제

[0535] 서열 특이적 테더 및 라이프테크(Lifetech) 마이원 C1 스트랩타비딘 비드를 통하여 리가제 및 반응되지 않은 성분으로부터 샘플 EX를 정제하였다.

[0536] 1.25 ul의 100 uM 테더를 30 ul의 샘플 7A 및 8.75 ul의 NF H<sub>2</sub>O와 혼합하고, 이 혼합물을 실온 하에 15분 동안 인큐베이션함으로써, 샘플 7A를 테더 (/5테스티오비오틴/TT/iSp18//iSp18//iSp18//iSp18//iSp18//iSp18/(서열

식별번호: 65)/iSp18//iSp18//iSp18//iSp18//iSp18//iSp18/TT/3 CholTEG/)와 혼성화시켰다.

[0537] 자기 랙을 이용하여, 20  $\mu$ l의 마이윈 C1 스트랩타비딘 비드를 제조업자에 의해 명시된 바와 같이 200  $\mu$ l의 1x 결합 및 세척 완충제 (B&W 완충제)로 세척한 다음, 40  $\mu$ l의 2x B&W 완충제에 재현탁시켰다.

[0538] 라이프테크에 의해 명시된 2x B&W 완충제:

[0539] 10 mM 트리스-HCl pH 7.5

[0540] 1 mM EDTA

[0541] 2 M NaCl.

[0542] 40  $\mu$ l의 테더링된 샘플 7A를 40  $\mu$ l의 스트랩타비딘 비드에 부가하고, 15분 동안 물리 상에서 인큐베이션하였다. 이어서, 용액을 자기 랙 상에 놓아두고, 제조업자의 지시에 따라서 1x B&W 완충제로 2회 세척하였다.

[0543] H<sub>2</sub>O 중의 15  $\mu$ l의 133  $\mu$ M 비오틴을 스트랩타비딘 비드에 부가하고 10분 동안 37°C로 가열함으로써, 샘플을 상기 비드로부터 용출시켰다. 튜브를 자기 랙 위에 신속하게 놓아두고 상등액을 상기 비드로부터 제거하였다. 이러한 생성물은 정제된 샘플이고 샘플 7B로서 공지되었다.

[0544] 전기 생리학

[0545] 3  $\mu$ l의 [1  $\mu$ M 테더링 올리고 (서열식별번호: 66/iSp18//iSp18//iSp18//iSp18//iSp18//iSp18/TT/3CholTEG/), 750 mM KCl, 125 mM 인산칼륨 완충제 pH 7, 및 5 mM EDTA]를 혼합하고, 이 혼합물을 실온 하에 20분 동안 인큐베이션함으로써, 샘플 7B (9  $\mu$ l)를 테더링 올리고와 혼성화시켰다.

[0546] 이어서, 이와 같이 테더링된 샘플 7B를 2  $\mu$ l의 17.4  $\mu$ M T4 Dda [E94C/F98W/C109A/C136A/K194L/A360C(돌연변이 E94C/F98W/C109A/C136A/K194L/A360C에 이어서 ( $\Delta$ M1)G1을 수반한 서열식별번호: 8)]과 함께 15분 동안 인큐베이션하였다. 이어서, 2.1  $\mu$ l의 800  $\mu$ M TMAD를 상기 인큐베이션된 혼합물에 부가하고, 실온 하에 10분 동안 유지시켰다. 그 다음, 이러한 샘플을 완충제 (282  $\mu$ L의 500 mM KCl, 25 mM 인산칼륨 pH 8.0) 내로 희석시키고, 2  $\mu$ l의 (70  $\mu$ M ATP 및 75  $\mu$ M MgCl<sub>2</sub>)와 혼합하였다.

[0547] 완충제 [25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, 약 pH 8.0] 중의 블록 공중합체 내에 삽입된 단일 MspA 나노 세공으로부터 전기적 측정치를 획득하였다. 상기 블록 공중합체 내에 삽입된 단일 세공을 달성한 후, 완충제 [2 mL, 25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, pH 8.0]를 시스템 내로 유동시켜 임의의 과도한 MspA 나노 세공을 제거하였다.

[0548] 샘플을 부가하기 전에, 과도한 완충제 (500 mM KCl, 25 mM 인산칼륨 pH 8.0)를 시스템 내로 유동시켰다. 최종적으로, 샘플 3A와 결합된 T4 Dda - E94C/A360C를 나노 세공 시스템에 부가하고, 실험을 -120 mV에서 실행하였고, 헬리카제-제어된 DNA 이동을 모니터링하였다.

[0549] 결과

[0550] 도 19는 샘플 7B를 부가한 경우 단일 나노 세공으로부터의 이온 전류 기록을 도시한 것이다. 전기 생리학 실험은 다양한 길이의 2D 가닥 (상이한 천연 세포성 mRNA 전사체 길이에 상응함)을 수반한, 우수한 처리량을 나타내었다. 도 19는 또한, RNA가 DNA와 비교해서 상이한 평균 진폭 및 범위를 갖고 있다는 것을 명확하게 보여준다. 따라서, RNA와 DNA는, RNA 서열과 DNA 서열이 동일한 경우일지라도 평균 진폭 및 범위의 함수로서 서로 차별될 수 있다.

[0551] 실시예 8

[0552] 본 실시예는 DNA-함유 리더를 메신저 RNA (mRNA)에 부착시켜, DNA 헬리카제인 He1308Mbu-E284C/S615C (돌연변이 E284C/S615C를 수반한 서열식별번호: 8)의 부하를 촉진시키고 나노 세공을 통한 RNA의 후속 헬리카제 제어된 이동을 촉진시키는 방법을 보여준다. 1.9 kb mRNA를 트리링크 바이오테크(Trilink Biotech)로부터 구입하였다. 상기 DNA-함유 리더를 mRNA의 3' 말단에 라이게이션하였다. 이어서, He1308을 리더 내의 DNA 결합 부위 상으로 부가하고, 그 기질을 나노 세공에 의해 분석하였다.

[0553] 물질 및 방법

[0554] DNA-함유 리더를 여러 개의 올리고 [서열식별번호: 35 (5' 포스페이트를 갖는다)는 서열식별번호: 37, 38 및 39 (3' 콜레스테롤 TEG를 갖는다)와 어닐링시킨 30개의 iSpC3 스페이서에 그의 5' 말단에서 부착되는



서열식별번호: 36의 5' 말단에 반대 말단에서 부착되는 4개의 iSp18 스페이서에 그의 3' 말단에서 부착되었다]와 미리 어닐링시키고, 이와 같이 미리 어닐링시킨 리더를, 표 31에 열거된 시약을 혼합하고 이 혼합물을 표 32에서의 프로그램으로 설정된 열순환기 위에 놓아둠으로써 mRNA에 라이게이션하였다.

<표 31>

시약	용적	스톡 농도	최종 농도
FLuc mRNA (RNA X1; 5'의 최외각 뉴클레오타이드로서 5'-헥시닐-G를 갖고 3' 폴리 A 테일을 갖는 서열식별번호: 26의 오픈 리딩 프레임을 수반한 mRNA)	2.5 ul	1 ug/ul	2500 ng/반응
미리 어닐링된 리더 1:1:1:1	3.5 ul	각각 2 uM	1x
NEB Quick 리가제 완충제 (라이프테크)	4.0 ul	5 x	1x
NEB Quick 리가제	2.0 ul	2,000 U/ul	4,000 U
스플린트 (서열식별번호: 40)	0.2 ul	100 uM	1 uM
물	8 ul		
총	20.2 ul		

<표 32>

순환 수	단계	온도 (°C)	시간
1	라이게이션하다	16	0:30:00

이어서, 상기 혼합물을 샘플 1  $\mu$ L당 1.8  $\mu$ L SPRI 비드의 비율로 아젠코트 앰플어 SPRI 비드를 이용하여 정제하였다.

#### 역전사

표 33에 열거된 시약을 혼합하고 이 혼합물을 표 34에서의 프로그램으로 설정된 열순환기 위에 놓아둠으로써, 슈퍼스크립트 II 키트를 이용하여 상기 샘플을 역전사시켰다.

<표 33>

시약	용적	스톡 농도	최종 농도
라이게이션으로부터의 샘플	2.2 ul		500 ng
dNTPs	1 ul	각각 10 mM	
제 1 가닥 완충제 (라이프테크)	4 ul	5 x	1 x
DTT	2 ul	0.1 M	
SSII 역전사효소	1 ul		
NF H <sub>2</sub> O	9.8 ul		
총	20 ul		

[0564] <표 34>

순환 수	단계	온도 (°C)	시간
1	역전사	42	0:50:00
2	변성	70	0:15:00

[0565]

[0566]

이어서, 상기 혼합물을 샘플 1  $\mu\text{L}$ 당 1.8  $\mu\text{L}$  SPRI 비드의 비율로 아젠코트 앰프어 SPRI 비드를 이용하여 정제하였다. He1308Mbu-E284C/S615C (돌연변이 E284C/S615C를 수반한 서열식별번호: 8)를, 7 kda 제바(Zeba) 칼럼을 이용하여 50 mM HEPES pH 8, 100 mM KAc로 완충제 교환시켰다. 역전사된 RNA 샘플을 등 용적의 100 mM HEPES pH 8, 200 mM KAc와 혼합한 다음, 1:100의 물 비로 He1308Mbu-E284C/S615C와 혼합하고, 100 mM 인산칼륨 pH 8, 100 mM NaCl, 5 mM EDTA, 0.1% 트윈(TWEEN)으로 완충제 교환시켰다. DMF에 용해된 20 mM BMOE를 100 mM 인산칼륨 pH 8, 100 mM NaCl, 5 mM EDTA, 0.1% 트윈에 부가하여 5 mM의 농도가 되도록 하였다. 이러한 5 mM BMOE 용액을 RNA와 He1308의 혼합물에 부가하여 40  $\mu\text{M}$  BMOE의 최종 농도가 되도록 하였다. 이 용액을 실온 하에 2시간 동안 인큐베이션하였다. 그 다음, 상기 용액을 샘플 1  $\mu\text{L}$ 당 1.8  $\mu\text{L}$  SPRI 비드의 비율로 아젠코트 앰프어 SPRI 비드에 결합시켰다. SPRI 비드를 EtOH 혼합물로 세척하는 것 대신, 20% PEG, 2.5 M NaCl, 50 mM 트리스의 변형된 세척 완충제를 이용한 단일 세척을 이용하였다. 샘플을 30  $\mu\text{L}$ 의 500 mM KCl, 25 mM 인산칼륨 pH 8에서 용출시켰다. 이것은 샘플 6으로서 공지되었다.

[0567]

#### 전기 생리학

[0568]

샘플 6 (4  $\mu\text{L}$ )을 완충제 (500 mM KCl, 25 mM 인산칼륨 pH 8, 2 mM ATP, 2 mM  $\text{MgCl}_2$ , 295  $\mu\text{L}$ )와 혼합하였다.

[0569]

완충제 [25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, 약 pH 8.0] 중의 블록 공중합체 내에 삽입된 단일 MspA 나노 세공으로부터 전기적 측정치를 획득하였다. 상기 블록 공중합체 내에 삽입된 단일 세공을 달성한 후, 완충제 [2 mL, 25 mM 인산칼륨 완충제, 150 mM 페로시아나화칼륨, 150 mM 페리시아나화칼륨, pH 8.0]를 시스템 내로 유동시켜 임의의 과도한 MspA 나노 세공을 제거하였다.

[0570]

샘플 6을 부가하기 전에, 과도한 완충제 (500 mM KCl, 25 mM 인산칼륨 pH 8.0)를 시스템 내로 유동시켰다. 최종적으로, 샘플 6을 나노 세공 시스템에 부가하고, 실험을 -140 mV에서 실행하였고, 헬리카제-제어된 DNA 이동을 모니터링하였다.

[0571]

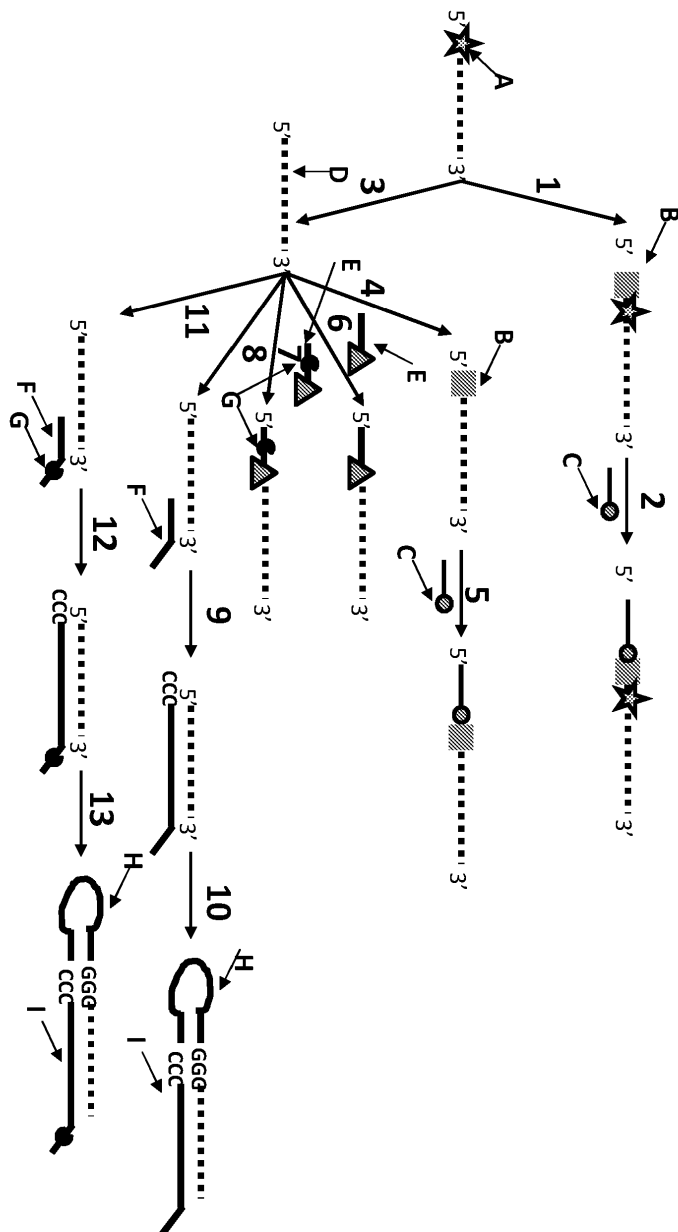
#### 결과

[0572]

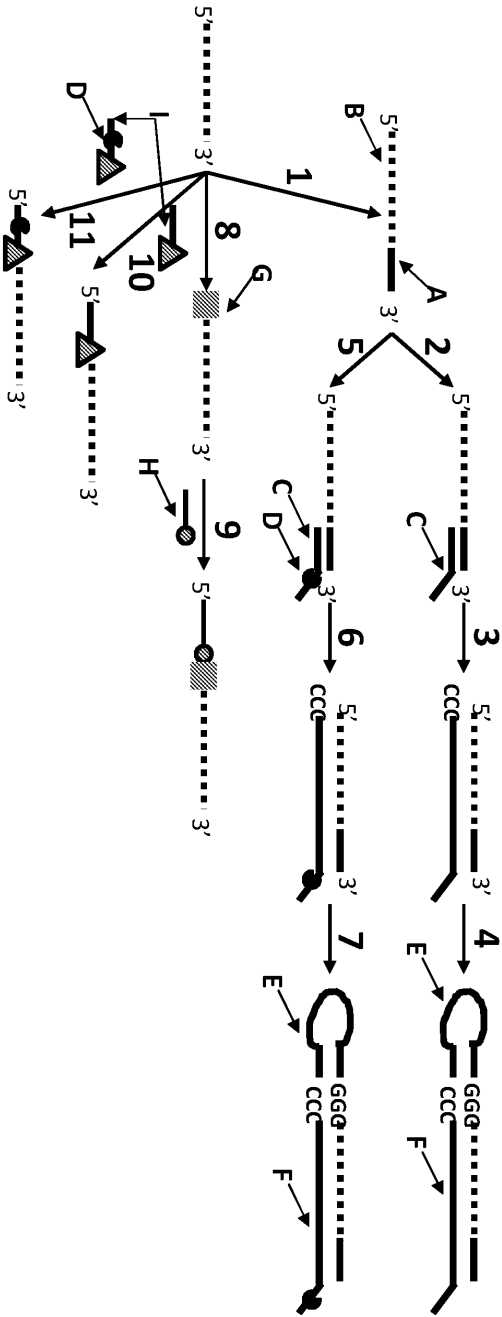
본 실시예는 DNA-함유 리더를 메신저 RNA (mRNA)에 부착시켜, DNA 헬리카제인 He1308Mbu-E284C/S615C (돌연변이 E284C/S615C를 수반한 서열식별번호: 8)의 부하를 촉진시키고, RNA의 후속 헬리카제 제어된 이동을 관찰하는 방법을 보여준다. 헬리카제 제어된 RNA 이동의 한 예가 도 20에 도시된다.

도면

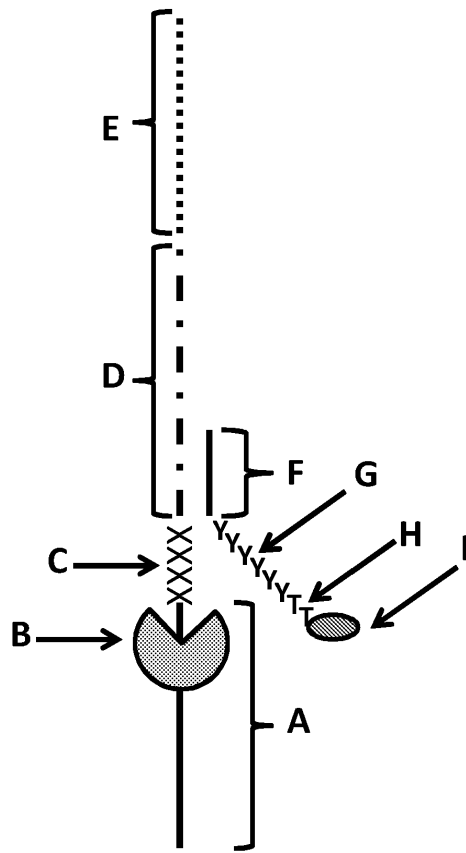
도면1



도면2

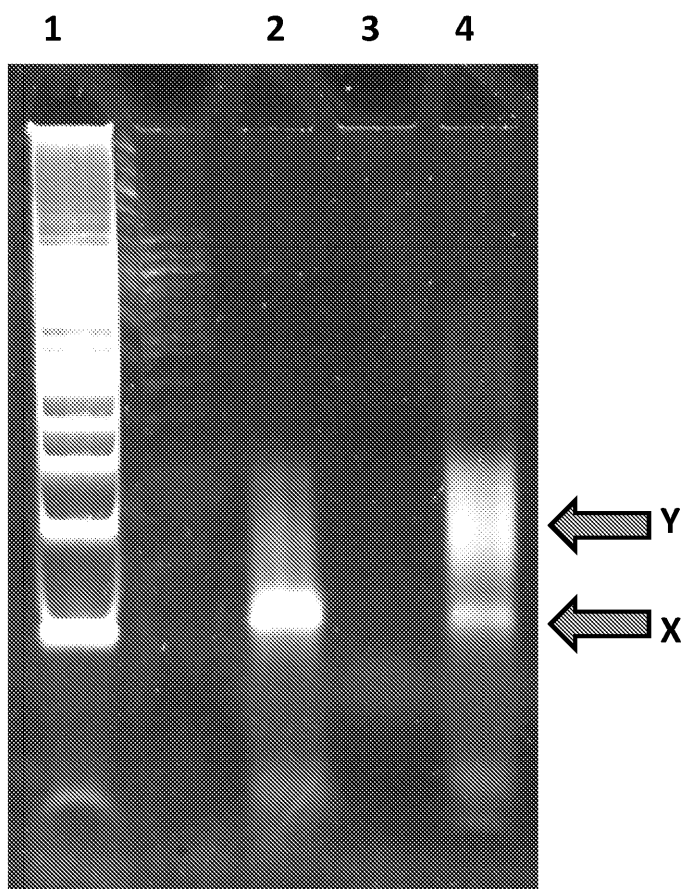


도면3

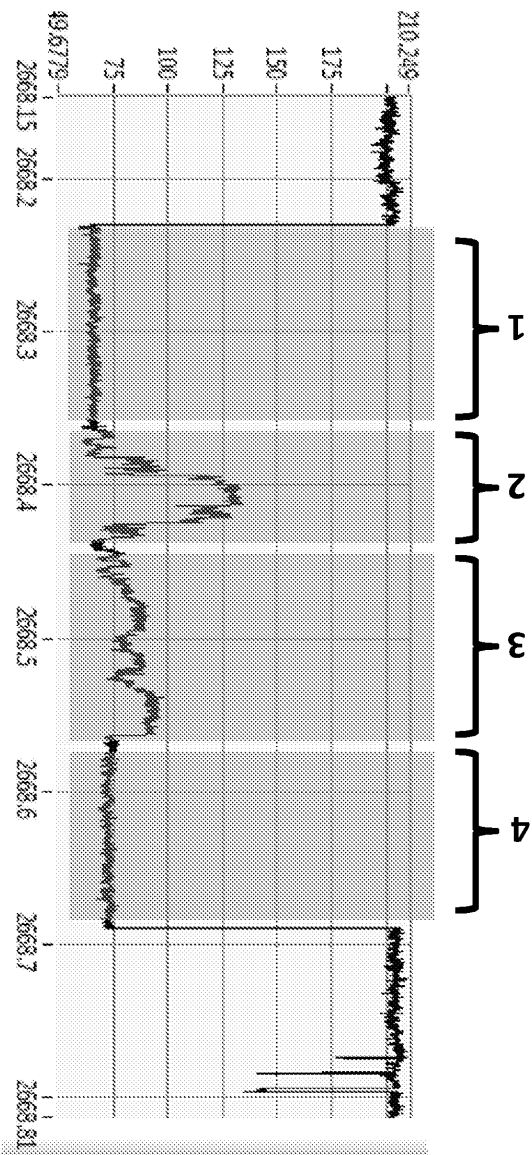




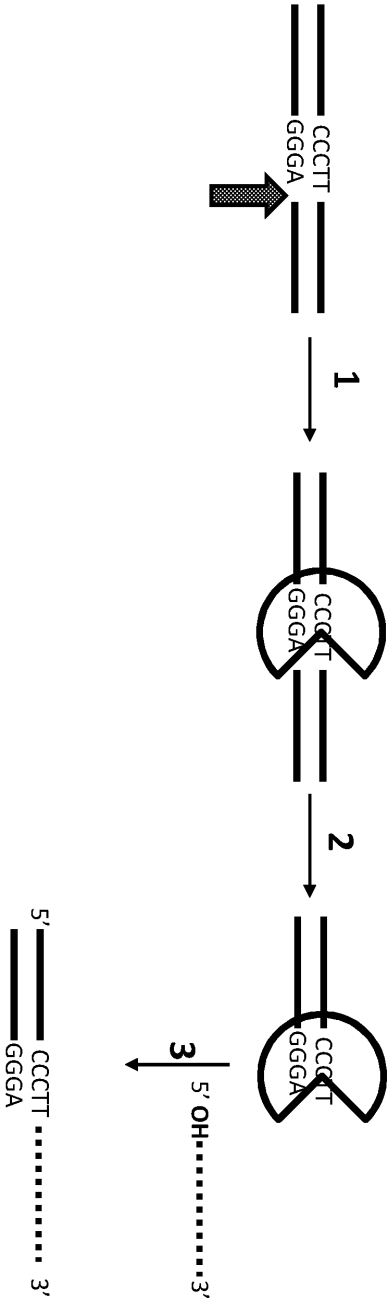
도면4



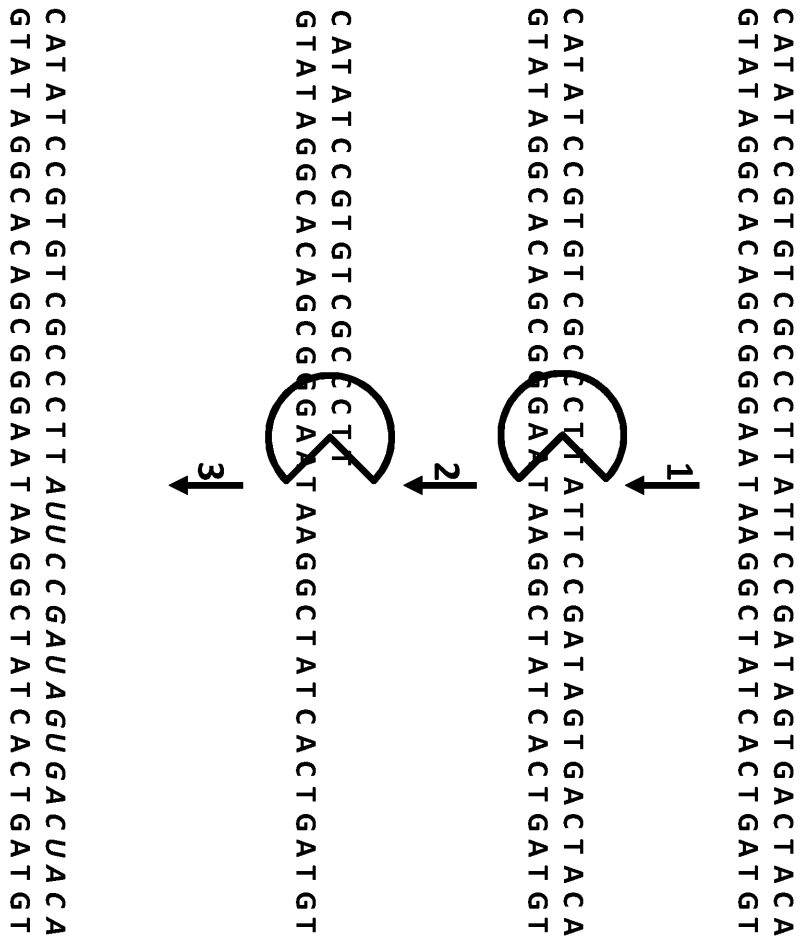
도면5



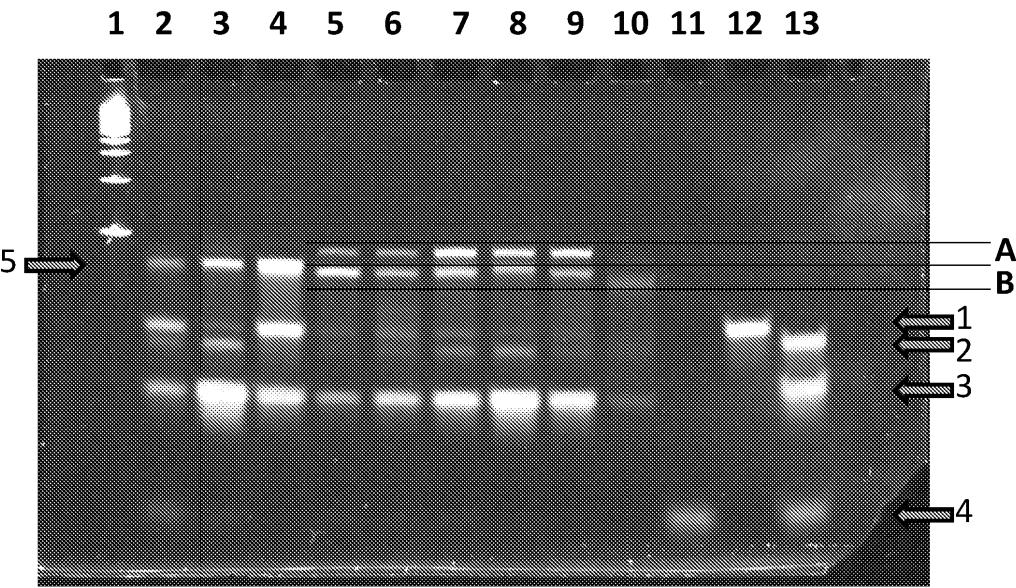
도면6



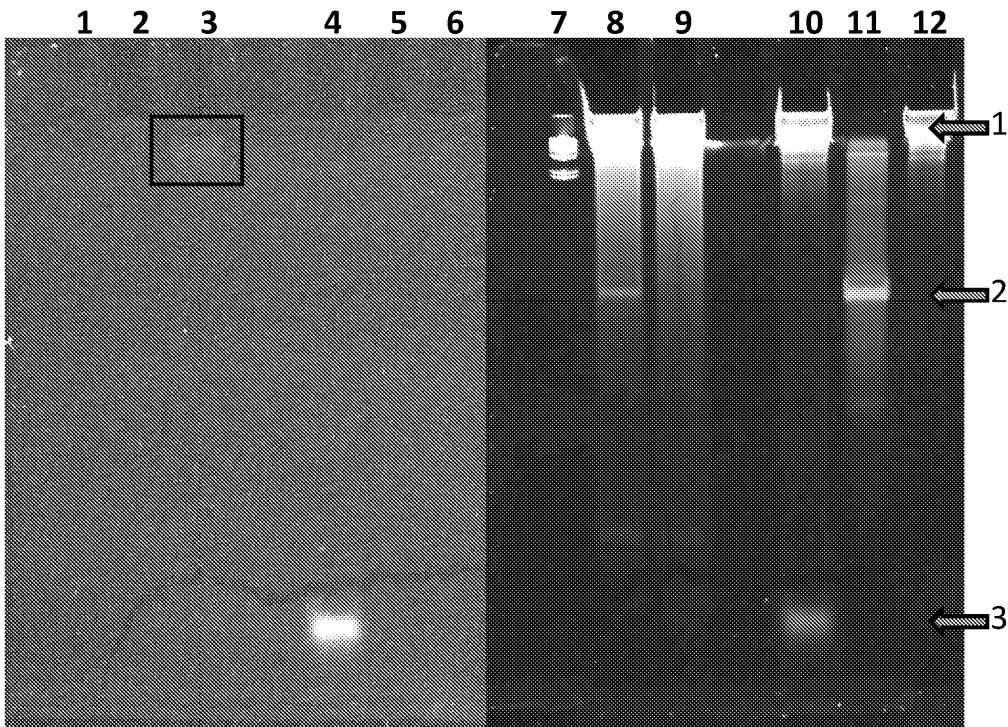
도면7



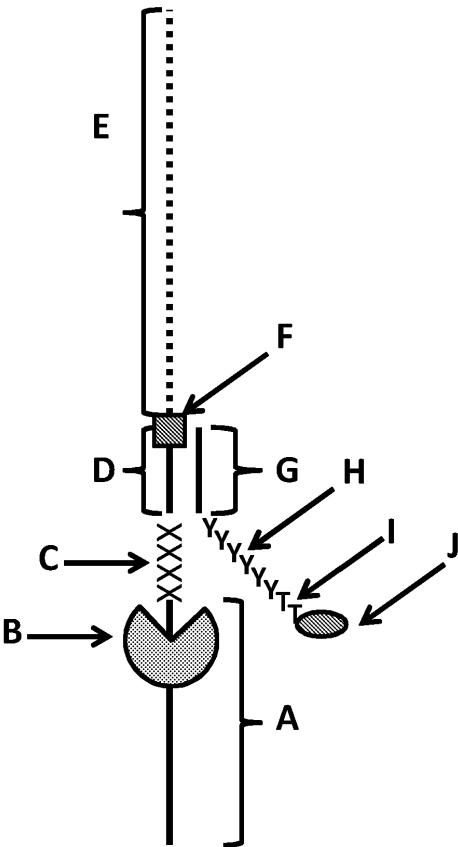
도면8



도면9

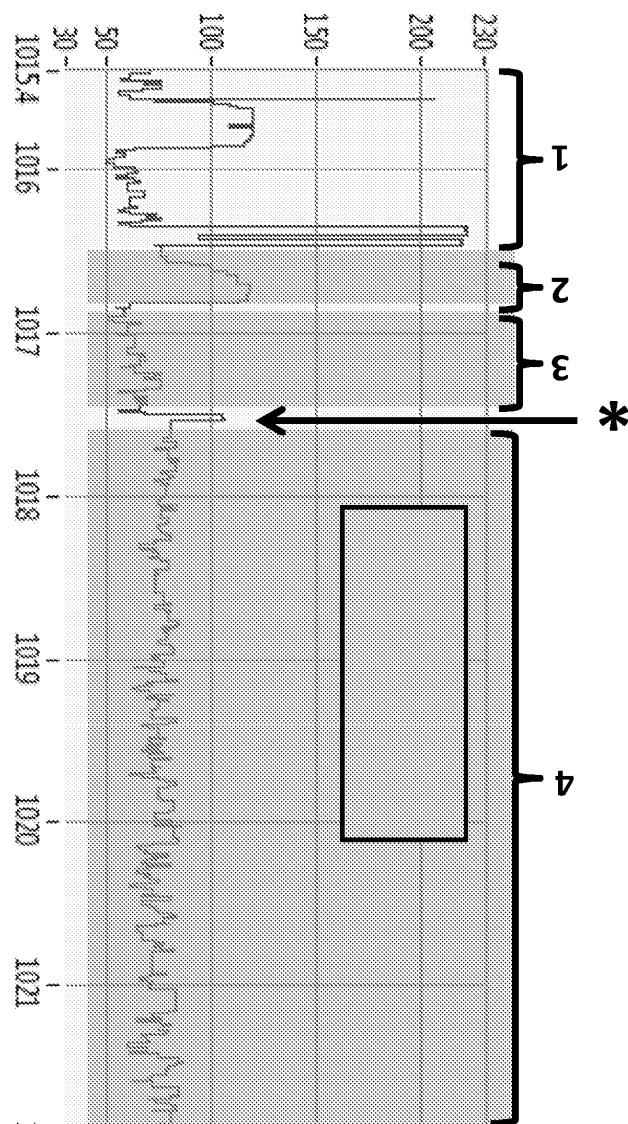


도면10

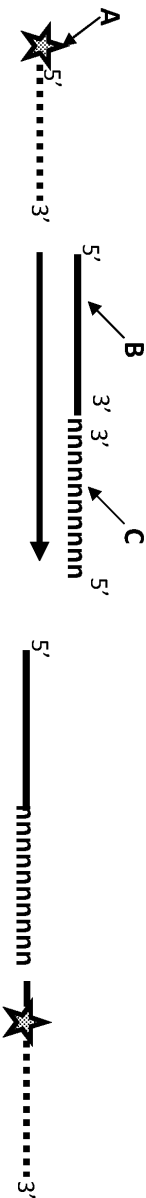




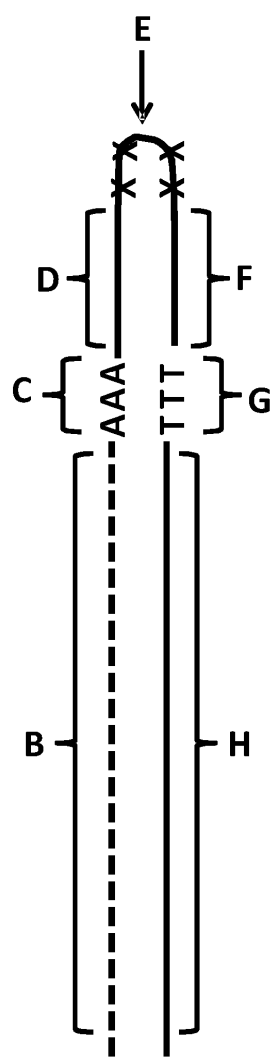
도면11



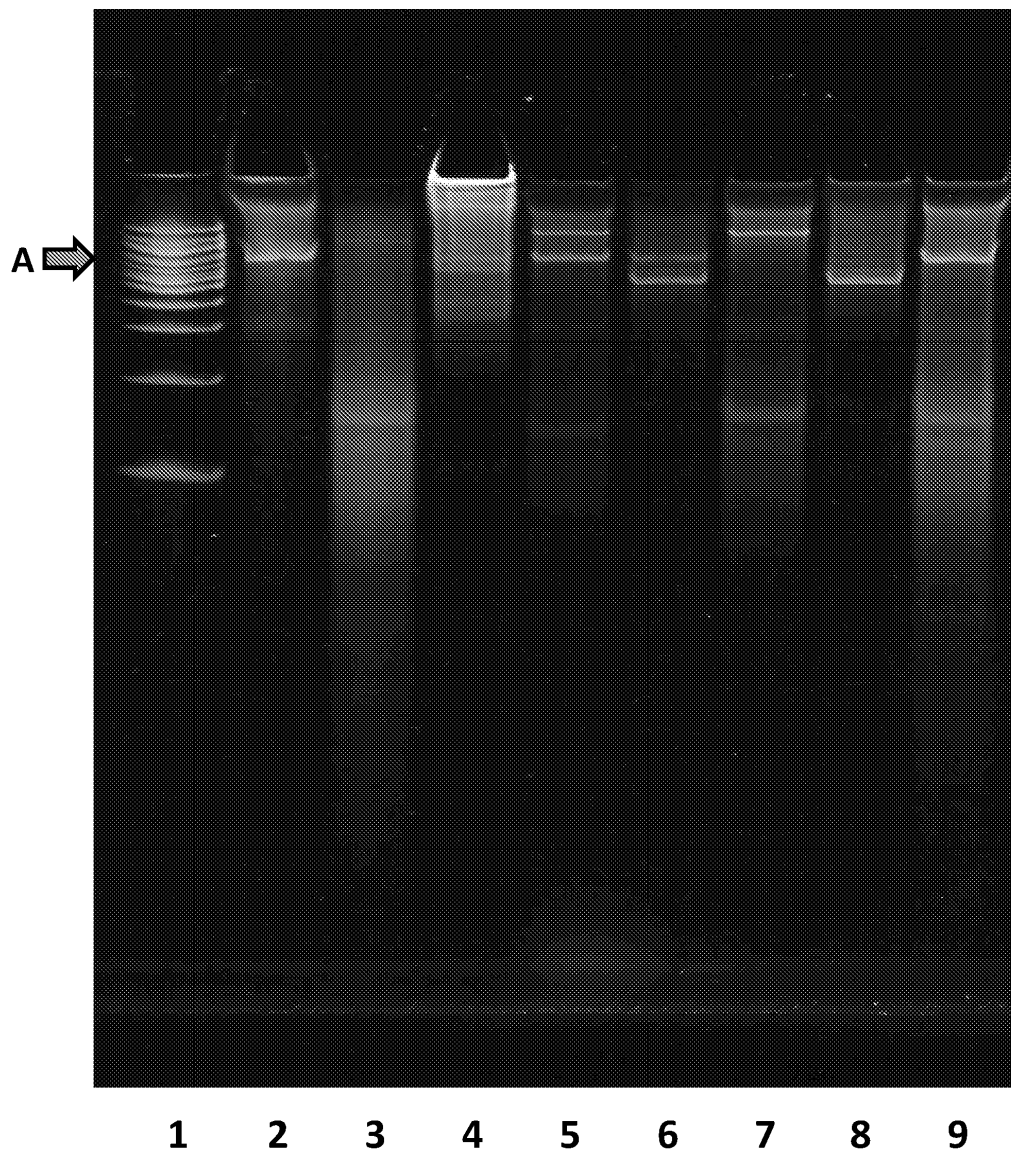
도면12



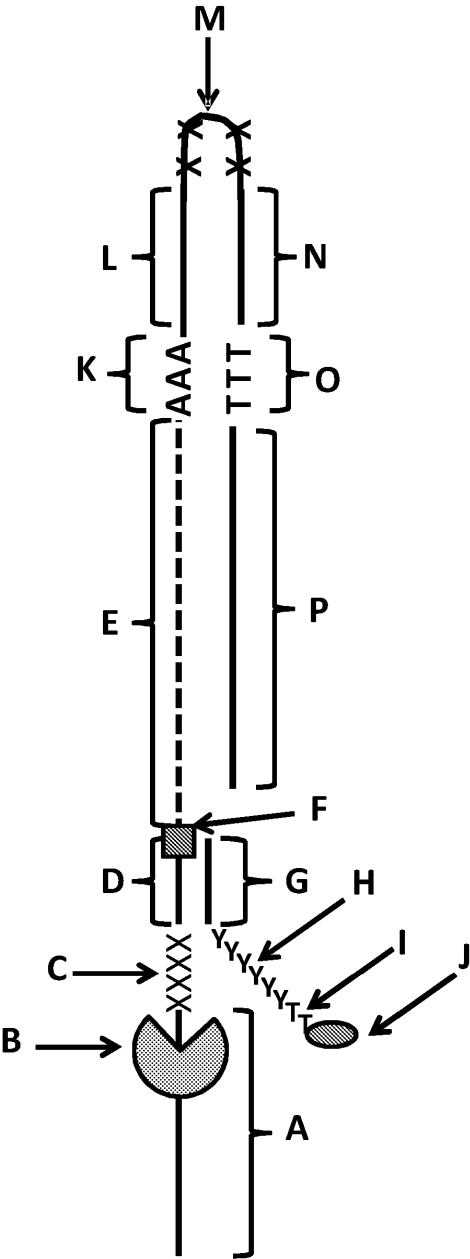
도면13



도면14

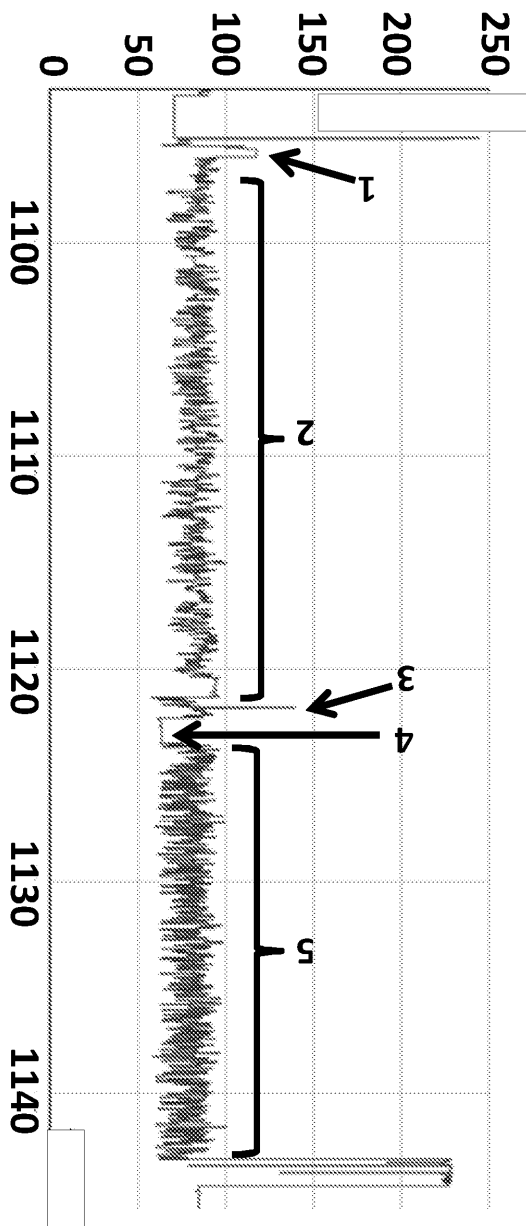


도면15

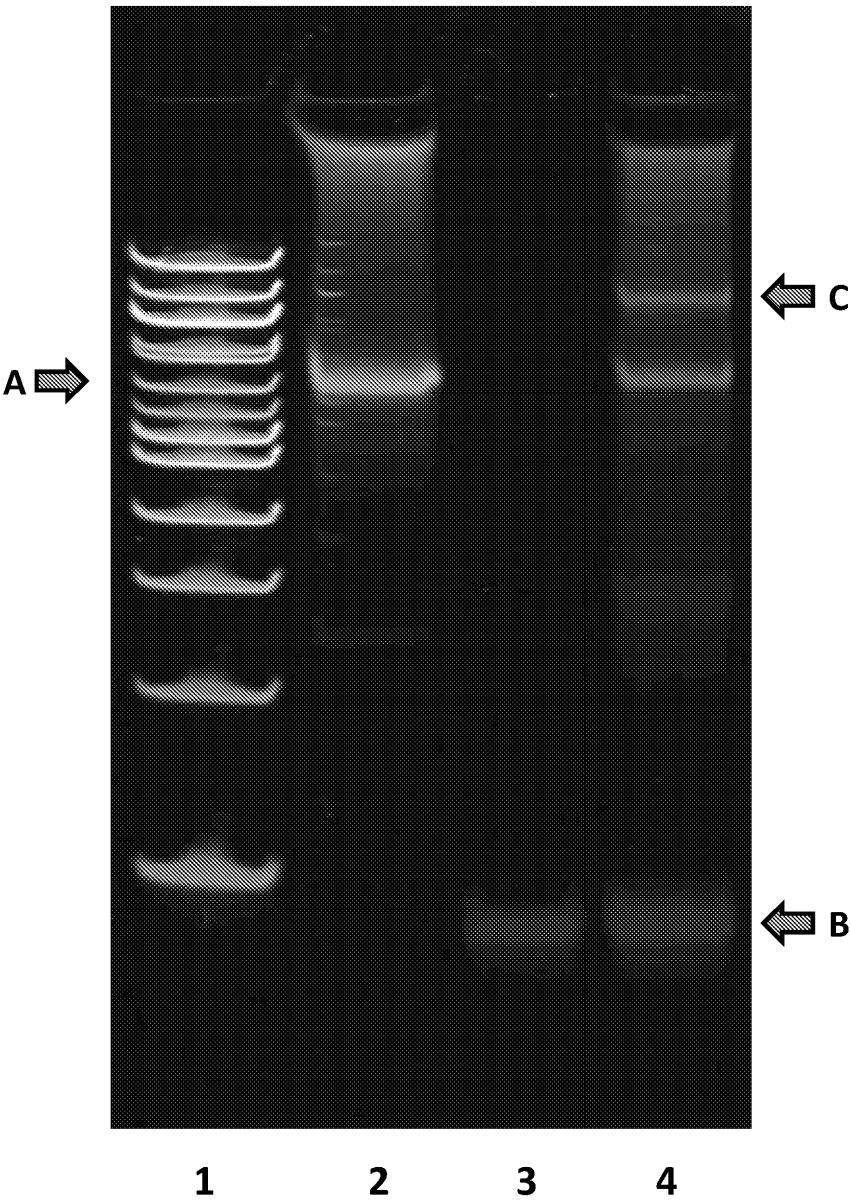




도면 16

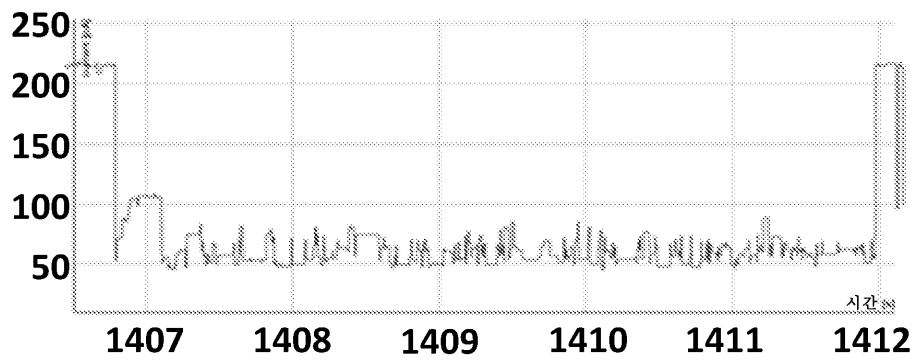


도면17

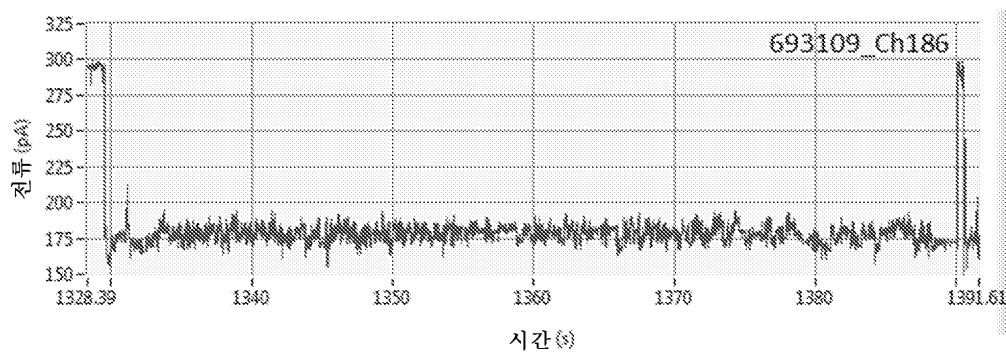




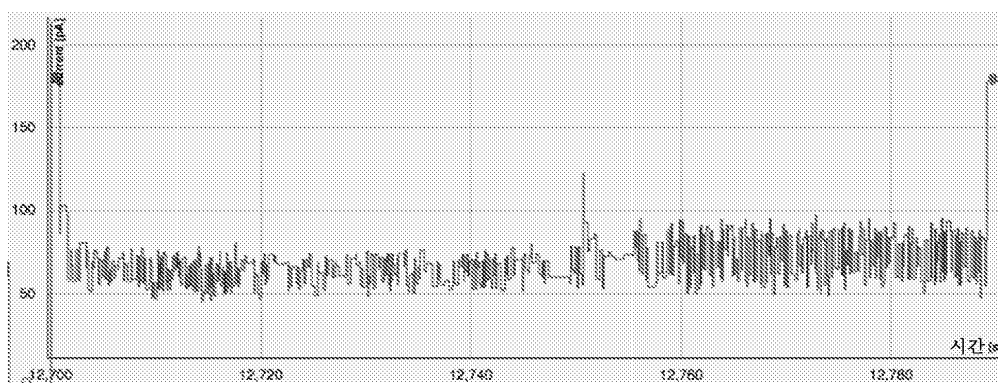
도면20



도면21



도면22



## 서열 목록

SEQUENCE LISTING

<110> Oxford Nanopore Technologies Limited

<120> Method

<130> N404137W0

<150> GB1418459.2  
 <151> 2014-10-17  
 <150> GB1508270.4  
 <151> 2015-05-14  
 <150> GB1517634.0  
 <151> 2015-10-06  
 <160> 66  
 <170> PatentIn version 3.5  
 <210> 1  
 <211> 558  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Mycobacterium smegmatis porin A mutant  
 (D90N/D91N/D93N/D118R/D134R/E193K)  
 <400> 1  
 atgggtctgg ataatgaact gagcctgggtg gacggtcaag atcgtaccct gacggtgcaa 60  
  
 caatgggata cttttctgaa tggcgttttt ccgctggatc gtaatcgct gacccgtgaa 120  
 tggtttcatt ccggtcgcgc aaaatataatc gtcgcaggcc cgggtgctga cgaattcgaa 180  
 ggcacgctgg aactgggtta tcagattggc tttccgtggt cactgggcgt tggatatcaac 240  
 ttctcgtaca ccacgccgaa tattctgata aacaatggta acattaccgc accgccgttt 300  
 ggctgaaca gcgtgattac gccgaacctg tttccgggtg ttagcatctc tgcccgtctg 360  
 ggcaatggtc cgggcattca agaagtggca accttttagtg tgcgcgtttc cggcgctaaa 420  
 ggcggtgtcg cgggtgtctaa cgcccacggt accgttacgg gcgcggccgg cgggtgtcctg 480  
  
 ctgcgtccgt tcgcgcgcct gattgcctct accggcgaca gcgttacgac ctatggcgaa 540  
 ccgtggaata tgaactaa 558  
 <210> 2  
 <211> 184  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Mycobacterium smegmatis porin A mutant  
 (D90N/D91N/D93N/D118R/D134R/E139K)  
 <400> 2  
 Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu



1                      5                      10                      15  
 Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp  
  
                     20                      25                      30  
 Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr  
                     35                      40                      45  
 Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu  
                     50                      55                      60  
 Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe  
 65                      70                      75                      80  
 Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asn Asn Gly Asn Ile Thr Ala  
  
                     85                      90                      95  
 Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly  
                     100                      105                      110  
 Val Ser Ile Ser Ala Arg Leu Gly Asn Gly Pro Gly Ile Gln Glu Val  
                     115                      120                      125  
 Ala Thr Phe Ser Val Arg Val Ser Gly Ala Lys Gly Gly Val Ala Val  
                     130                      135                      140  
 Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu  
  
 145                      150                      155                      160  
 Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr  
                     165                      170                      175  
 Tyr Gly Glu Pro Trp Asn Met Asn  
                     180  
 <210> 3  
 <211> 885  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> alpha-hemolysin mutant (E111N/K147N)  
 <400> 3  
 atggcagatt ctgatattaa tattaacc ggtactacag atattggaag caatactaca 60  
 gtaaaaacag gtgatttagt cacttatgat aaagaaaatg gcatgcacaa aaaagtattt 120

tatagtttta tcgatgataa aaatcacaaat aaaaaactgc tagttattag aacaaaaggt 180  
accattgctg gtcaatatag agtttatagc gaagaagggtg ctaacaaaag tggtttagcc 240  
tggccttcag cctttaaggt acagttgcaa ctacctgata atgaagtagc tcaaatatct 300  
gattactatc caagaaattc gattgataca aaaaaactata tgagtacttt aacttatgga 360  
ttcaacggta atgttactgg tgatgataca ggaaaaattg gcggccttat tggtgcaaat 420  
gtttcgattg gtcatacaact gaactatgtt caacctgatt tcaaaacaat tttagagagc 480  
ccaactgata aaaaagtagg ctggaaagtg atatttaaca atatggtgaa tcaaaattgg 540

ggaccatacg atcgagattc ttggaacccg gtatatggca atcaactttt catgaaaact 600  
agaaatgggt ctatgaaagc agcagataac ttccttgatc ctaacaaagc aagttctcta 660  
ttatcttcag ggttttcacc agacttcgct acagttatta ctatggatag aaaagcatcc 720  
aaacaacaaa caaatataga tgtaatatag gaacgagttc gtgatgatta ccaattgcat 780  
tggacttcaa caaattggaa aggtaccaat actaaagata aatggacaga tcgtttctca 840  
gaaagatata aaatcgattg ggaaaaagaa gaaatgacaa attaa 885

<210> 4

<211> 293

<212> PRT

<213> Artificial Sequence

<220><223> alpha-hemolysin mutant (E111N/K147N)

<400> 4

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 Asn 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 Asn 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> 5

<211> 184

<212> PRT

<213> Mycobacterium smegmatis

<400> 5

Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu

1                    5                    10                    15  
 Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp  
                   20                    25                    30  
 Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr  
                   35                    40                    45  
 Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu  
                   50                    55                    60  
 Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe

65                    70                    75                    80  
 Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala  
                   85                    90                    95  
 Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly  
                   100                    105                    110  
 Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val  
                   115                    120                    125  
 Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val

130                    135                    140  
 Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu  
 145                    150                    155                    160  
 Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr  
                   165                    170                    175  
 Tyr Gly Glu Pro Trp Asn Met Asn  
                   180

<210> 6

<211> 184

<212> PRT

<213> *Mycobacterium smegmatis*

<400> 6

Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu

1                    5                    10                    15  
 Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp  
                   20                    25                    30

Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr  
 35 40 45  
 Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu  
 50 55 60  
 Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe  
 65 70 75 80  
 Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Gly  
 85 90 95  
 Pro Pro Phe Gly Leu Glu Ser Val Ile Thr Pro Asn Leu Phe Pro Gly  
 100 105 110  
 Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val  
 115 120 125  
 Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val  
 130 135 140  
 Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu  
 145 150 155 160  
 Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr  
 165 170 175  
 Tyr Gly Glu Pro Trp Asn Met Asn  
 180  
 <210> 7  
 <211> 183  
 <212> PRT  
 <213> *Mycobacterium smegmatis*  
 <400> 7  
 Val Asp Asn Gln Leu Ser Val Val Asp Gly Gln Gly Arg Thr Leu Thr  
 1 5 10 15  
 Val Gln Gln Ala Glu Thr Phe Leu Asn Gly Val Phe Pro Leu Asp Arg  
 20 25 30  
 Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Thr Tyr His  
 35 40 45  
 Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu Gly



50 55 60  
Tyr Gln Val Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe Ser

65 70 75 80  
Tyr Thr Thr Pro Asn Ile Leu Ile Asp Gly Gly Asp Ile Thr Gln Pro

85 90 95  
Pro Phe Gly Leu Asp Thr Ile Ile Thr Pro Asn Leu Phe Pro Gly Val

100 105 110  
Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala

115 120 125  
Thr Phe Ser Val Asp Val Lys Gly Ala Lys Gly Ala Val Ala Val Ser

130 135 140  
Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg

145 150 155 160  
Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr

165 170 175  
Gly Glu Pro Trp Asn Met Asn

180

<210> 8

<211> 760

<212> PRT

<213> Methanococcoides burtonii

<400> 8

Met Met Ile Arg Glu Leu Asp Ile Pro Arg Asp Ile Ile Gly Phe Tyr

1 5 10 15  
Glu Asp Ser Gly Ile Lys Glu Leu Tyr Pro Pro Gln Ala Glu Ala Ile

20 25 30  
Glu Met Gly Leu Leu Glu Lys Lys Asn Leu Leu Ala Ala Ile Pro Thr

35 40 45  
Ala Ser Gly Lys Thr Leu Leu Ala Glu Leu Ala Met Ile Lys Ala Ile

50 55 60  
Arg Glu Gly Gly Lys Ala Leu Tyr Ile Val Pro Leu Arg Ala Leu Ala

65	70	75	80
Ser	Glu	Lys	Phe
Glu	Arg	Phe	Lys
Glu	Leu	Ala	Pro
Phe	Gly	Ile	Lys
85	90	95	
Val	Gly	Ile	Ser
Thr	Gly	Asp	Leu
Asp	Ser	Arg	Ala
Asp	Trp	Leu	Gly
100	105	110	
Val	Asn	Asp	Ile
Ile	Val	Ala	Thr
Ser	Glu	Lys	Thr
Asp	Ser	Leu	Leu
115	120	125	
Arg	Asn	Gly	Thr
Ser	Trp	Met	Asp
Glu	Ile	Thr	Thr
Val	Val	Val	Asp
130	135	140	
Glu	Ile	His	Leu
Leu	Asp	Ser	Lys
Asn	Arg	Gly	Pro
Thr	Leu	Glu	Val
145	150	155	160
Thr	Ile	Thr	Lys
Leu	Met	Arg	Leu
Asn	Pro	Asp	Val
Gln	Val	Val	Ala
165	170	175	
Leu	Ser	Ala	Thr
Val	Gly	Asn	Ala
Arg	Glu	Met	Ala
Asp	Trp	Leu	Gly
180	185	190	
Ala	Ala	Leu	Val
Leu	Ser	Glu	Trp
Arg	Pro	Thr	Asp
Leu	His	Glu	Gly
195	200	205	
Val	Leu	Phe	Gly
Asp	Ala	Ile	Asn
Phe	Pro	Gly	Ser
Gln	Lys	Lys	Ile
210	215	220	
Asp	Arg	Leu	Glu
Lys	Asp	Ala	Val
Asn	Leu	Val	Leu
Asp	Thr	Ile	
225	230	235	240
Lys	Ala	Glu	Gly
Gln	Cys	Leu	Val
Phe	Glu	Ser	Ser
Arg	Arg	Asn	Cys
245	250	255	
Ala	Gly	Phe	Ala
Lys	Thr	Ala	Ser
Ser	Lys	Val	Ala
Lys	Ile	Leu	Asp
260	265	270	
Asn	Asp	Ile	Met
Ile	Lys	Leu	Ala
Gly	Ile	Ala	Glu
Glu	Val	Glu	Ser
275	280	285	
Thr	Gly	Glu	Thr
Asp	Thr	Ala	Ile
Val	Leu	Ala	Asn
Cys	Ile	Arg	Lys
290	295	300	
Gly	Val	Ala	Phe
His	His	Ala	Gly
Leu	Asn	Ser	Asn
His	Arg	Lys	Leu
305	310	315	320

Val Glu Asn Gly Phe Arg Gln Asn Leu Ile Lys Val Ile Ser Ser Thr

325 330 335

Pro Thr Leu Ala Ala Gly Leu Asn Leu Pro Ala Arg Arg Val Ile Ile

340 345 350

Arg Ser Tyr Arg Arg Phe Asp Ser Asn Phe Gly Met Gln Pro Ile Pro

355 360 365

Val Leu Glu Tyr Lys Gln Met Ala Gly Arg Ala Gly Arg Pro His Leu

370 375 380

Asp Pro Tyr Gly Glu Ser Val Leu Leu Ala Lys Thr Tyr Asp Glu Phe

385 390 395 400

Ala Gln Leu Met Glu Asn Tyr Val Glu Ala Asp Ala Glu Asp Ile Trp

405 410 415

Ser Lys Leu Gly Thr Glu Asn Ala Leu Arg Thr His Val Leu Ser Thr

420 425 430

Ile Val Asn Gly Phe Ala Ser Thr Arg Gln Glu Leu Phe Asp Phe Phe

435 440 445

Gly Ala Thr Phe Phe Ala Tyr Gln Gln Asp Lys Trp Met Leu Glu Glu

450 455 460

Val Ile Asn Asp Cys Leu Glu Phe Leu Ile Asp Lys Ala Met Val Ser

465 470 475 480

Glu Thr Glu Asp Ile Glu Asp Ala Ser Lys Leu Phe Leu Arg Gly Thr

485 490 495

Arg Leu Gly Ser Leu Val Ser Met Leu Tyr Ile Asp Pro Leu Ser Gly

500 505 510

Ser Lys Ile Val Asp Gly Phe Lys Asp Ile Gly Lys Ser Thr Gly Gly

515 520 525

Asn Met Gly Ser Leu Glu Asp Asp Lys Gly Asp Asp Ile Thr Val Thr

530 535 540

Asp Met Thr Leu Leu His Leu Val Cys Ser Thr Pro Asp Met Arg Gln

545 550 555 560

Leu Tyr Leu Arg Asn Thr Asp Tyr Thr Ile Val Asn Glu Tyr Ile Val

565                      570                      575  
 Ala His Ser Asp Glu Phe His Glu Ile Pro Asp Lys Leu Lys Glu Thr  
  
 580                      585                      590  
 Asp Tyr Glu Trp Phe Met Gly Glu Val Lys Thr Ala Met Leu Leu Glu  
 595                      600                      605  
 Glu Trp Val Thr Glu Val Ser Ala Glu Asp Ile Thr Arg His Phe Asn  
 610                      615                      620  
 Val Gly Glu Gly Asp Ile His Ala Leu Ala Asp Thr Ser Glu Trp Leu  
 625                      630                      635                      640  
 Met His Ala Ala Ala Lys Leu Ala Glu Leu Leu Gly Val Glu Tyr Ser  
  
 645                      650                      655  
 Ser His Ala Tyr Ser Leu Glu Lys Arg Ile Arg Tyr Gly Ser Gly Leu  
 660                      665                      670  
 Asp Leu Met Glu Leu Val Gly Ile Arg Gly Val Gly Arg Val Arg Ala  
 675                      680                      685  
 Arg Lys Leu Tyr Asn Ala Gly Phe Val Ser Val Ala Lys Leu Lys Gly  
 690                      695                      700  
 Ala Asp Ile Ser Val Leu Ser Lys Leu Val Gly Pro Lys Val Ala Tyr  
  
 705                      710                      715                      720  
 Asn Ile Leu Ser Gly Ile Gly Val Arg Val Asn Asp Lys His Phe Asn  
 725                      730                      735  
 Ser Ala Pro Ile Ser Ser Asn Thr Leu Asp Thr Leu Leu Asp Lys Asn  
 740                      745                      750  
 Gln Lys Thr Phe Asn Asp Phe Gln  
 755                      760  
 <210> 9  
 <211> 707  
 <212> PRT  
 <213> Cenarchaeum symbiosum  
 <400> 9  
 Met Arg Ile Ser Glu Leu Asp Ile Pro Arg Pro Ala Ile Glu Phe Leu

1	5	10	15
Glu Gly Glu Gly Tyr Lys Lys Leu Tyr Pro Pro Gln Ala Ala Ala Ala			
	20	25	30
Lys Ala Gly Leu Thr Asp Gly Lys Ser Val Leu Val Ser Ala Pro Thr			
	35	40	45
Ala Ser Gly Lys Thr Leu Ile Ala Ala Ile Ala Met Ile Ser His Leu			
	50	55	60
Ser Arg Asn Arg Gly Lys Ala Val Tyr Leu Ser Pro Leu Arg Ala Leu			
65	70	75	80
Ala Ala Glu Lys Phe Ala Glu Phe Gly Lys Ile Gly Gly Ile Pro Leu			
	85	90	95
Gly Arg Pro Val Arg Val Gly Val Ser Thr Gly Asp Phe Glu Lys Ala			
	100	105	110
Gly Arg Ser Leu Gly Asn Asn Asp Ile Leu Val Leu Thr Asn Glu Arg			
	115	120	125
Met Asp Ser Leu Ile Arg Arg Arg Pro Asp Trp Met Asp Glu Val Gly			
130	135	140	
Leu Val Ile Ala Asp Glu Ile His Leu Ile Gly Asp Arg Ser Arg Gly			
145	150	155	160
Pro Thr Leu Glu Met Val Leu Thr Lys Leu Arg Gly Leu Arg Ser Ser			
	165	170	175
Pro Gln Val Val Ala Leu Ser Ala Thr Ile Ser Asn Ala Asp Glu Ile			
	180	185	190
Ala Gly Trp Leu Asp Cys Thr Leu Val His Ser Thr Trp Arg Pro Val			
195	200	205	
Pro Leu Ser Glu Gly Val Tyr Gln Asp Gly Glu Val Ala Met Gly Asp			
	210	215	220
Gly Ser Arg His Glu Val Ala Ala Thr Gly Gly Gly Pro Ala Val Asp			
225	230	235	240
Leu Ala Ala Glu Ser Val Ala Glu Gly Gly Gln Ser Leu Ile Phe Ala			
	245	250	255



Asp Thr Arg Ala Arg Ser Ala Ser Leu Ala Ala Lys Ala Ser Ala Val

260 265 270

Ile Pro Glu Ala Lys Gly Ala Asp Ala Ala Lys Leu Ala Ala Ala Ala

275 280 285

Lys Lys Ile Ile Ser Ser Gly Gly Glu Thr Lys Leu Ala Lys Thr Leu

290 295 300

Ala Glu Leu Val Glu Lys Gly Ala Ala Phe His His Ala Gly Leu Asn

305 310 315 320

Gln Asp Cys Arg Ser Val Val Glu Glu Glu Phe Arg Ser Gly Arg Ile

325 330 335

Arg Leu Leu Ala Ser Thr Pro Thr Leu Ala Ala Gly Val Asn Leu Pro

340 345 350

Ala Arg Arg Val Val Ile Ser Ser Val Met Arg Tyr Asn Ser Ser Ser

355 360 365

Gly Met Ser Glu Pro Ile Ser Ile Leu Glu Tyr Lys Gln Leu Cys Gly

370 375 380

Arg Ala Gly Arg Pro Gln Tyr Asp Lys Ser Gly Glu Ala Ile Val Val

385 390 395 400

Gly Gly Val Asn Ala Asp Glu Ile Phe Asp Arg Tyr Ile Gly Gly Glu

405 410 415

Pro Glu Pro Ile Arg Ser Ala Met Val Asp Asp Arg Ala Leu Arg Ile

420 425 430

His Val Leu Ser Leu Val Thr Thr Ser Pro Gly Ile Lys Glu Asp Asp

435 440 445

Val Thr Glu Phe Phe Leu Gly Thr Leu Gly Gly Gln Gln Ser Gly Glu

450 455 460

Ser Thr Val Lys Phe Ser Val Ala Val Ala Leu Arg Phe Leu Gln Glu

465 470 475 480

Glu Gly Met Leu Gly Arg Arg Gly Gly Arg Leu Ala Ala Thr Lys Met

485 490 495

Gly Arg Leu Val Ser Arg Leu Tyr Met Asp Pro Met Thr Ala Val Thr

500 505 510  
Leu Arg Asp Ala Val Gly Glu Ala Ser Pro Gly Arg Met His Thr Leu

515 520 525  
Gly Phe Leu His Leu Val Ser Glu Cys Ser Glu Phe Met Pro Arg Phe

530 535 540  
Ala Leu Arg Gln Lys Asp His Glu Val Ala Glu Met Met Leu Glu Ala

545 550 555 560  
Gly Arg Gly Glu Leu Leu Arg Pro Val Tyr Ser Tyr Glu Cys Gly Arg

565 570 575  
Gly Leu Leu Ala Leu His Arg Trp Ile Gly Glu Ser Pro Glu Ala Lys

580 585 590  
Leu Ala Glu Asp Leu Lys Phe Glu Ser Gly Asp Val His Arg Met Val

595 600 605  
Glu Ser Ser Gly Trp Leu Leu Arg Cys Ile Trp Glu Ile Ser Lys His

610 615 620  
Gln Glu Arg Pro Asp Leu Leu Gly Glu Leu Asp Val Leu Arg Ser Arg

625 630 635 640  
Val Ala Tyr Gly Ile Lys Ala Glu Leu Val Pro Leu Val Ser Ile Lys

645 650 655  
Gly Ile Gly Arg Val Arg Ser Arg Arg Leu Phe Arg Gly Gly Ile Lys

660 665 670  
Gly Pro Gly Asp Leu Ala Ala Val Pro Val Glu Arg Leu Ser Arg Val

675 680 685  
Glu Gly Ile Gly Ala Thr Leu Ala Asn Asn Ile Lys Ser Gln Leu Arg

690 695 700  
Lys Gly Gly

705

<210> 10

<211> 720

<212> PRT

<213> Thermococcus gammatolerans

<400> 10

Met Lys Val Asp Glu Leu Pro Val Asp Glu Arg Leu Lys Ala Val Leu

1 5 10 15

Lys Glu Arg Gly Ile Glu Glu Leu Tyr Pro Pro Gln Ala Glu Ala Leu

20 25 30

Lys Ser Gly Ala Leu Glu Gly Arg Asn Leu Val Leu Ala Ile Pro Thr

35 40 45

Ala Ser Gly Lys Thr Leu Val Ser Glu Ile Val Met Val Asn Lys Leu

50 55 60

Ile Gln Glu Gly Gly Lys Ala Val Tyr Leu Val Pro Leu Lys Ala Leu

65 70 75 80

Ala Glu Glu Lys Tyr Arg Glu Phe Lys Glu Trp Glu Lys Leu Gly Leu

85 90 95

Lys Val Ala Ala Thr Thr Gly Asp Tyr Asp Ser Thr Asp Asp Trp Leu

100 105 110

Gly Arg Tyr Asp Ile Ile Val Ala Thr Ala Glu Lys Phe Asp Ser Leu

115 120 125

Leu Arg His Gly Ala Arg Trp Ile Asn Asp Val Lys Leu Val Val Ala

130 135 140

Asp Glu Val His Leu Ile Gly Ser Tyr Asp Arg Gly Ala Thr Leu Glu

145 150 155 160

Met Ile Leu Thr His Met Leu Gly Arg Ala Gln Ile Leu Ala Leu Ser

165 170 175

Ala Thr Val Gly Asn Ala Glu Glu Leu Ala Glu Trp Leu Asp Ala Ser

180 185 190

Leu Val Val Ser Asp Trp Arg Pro Val Gln Leu Arg Arg Gly Val Phe

195 200 205

His Leu Gly Thr Leu Ile Trp Glu Asp Gly Lys Val Glu Ser Tyr Pro

210 215 220

Glu Asn Trp Tyr Ser Leu Val Val Asp Ala Val Lys Arg Gly Lys Gly

225 230 235 240

Ala Leu Val Phe Val Asn Thr Arg Arg Ser Ala Glu Lys Glu Ala Leu

245 250 255

Ala Leu Ser Lys Leu Val Ser Ser His Leu Thr Lys Pro Glu Lys Arg

260 265 270

Ala Leu Glu Ser Leu Ala Ser Gln Leu Glu Asp Asn Pro Thr Ser Glu

275 280 285

Lys Leu Lys Arg Ala Leu Arg Gly Gly Val Ala Phe His His Ala Gly

290 295 300

Leu Ser Arg Val Glu Arg Thr Leu Ile Glu Asp Ala Phe Arg Glu Gly

305 310 315 320

Leu Ile Lys Val Ile Thr Ala Thr Pro Thr Leu Ser Ala Gly Val Asn

325 330 335

Leu Pro Ser Phe Arg Val Ile Ile Arg Asp Thr Lys Arg Tyr Ala Gly

340 345 350

Phe Gly Trp Thr Asp Ile Pro Val Leu Glu Ile Gln Gln Met Met Gly

355 360 365

Arg Ala Gly Arg Pro Arg Tyr Asp Lys Tyr Gly Glu Ala Ile Ile Val

370 375 380

Ala Arg Thr Asp Glu Pro Gly Lys Leu Met Glu Arg Tyr Ile Arg Gly

385 390 395 400

Lys Pro Glu Lys Leu Phe Ser Met Leu Ala Asn Glu Gln Ala Phe Arg

405 410 415

Ser Gln Val Leu Ala Leu Ile Thr Asn Phe Gly Ile Arg Ser Phe Pro

420 425 430

Glu Leu Val Arg Phe Leu Glu Arg Thr Phe Tyr Ala His Gln Arg Lys

435 440 445

Asp Leu Ser Ser Leu Glu Tyr Lys Ala Lys Glu Val Val Tyr Phe Leu

450 455 460

Ile Glu Asn Glu Phe Ile Asp Leu Asp Leu Glu Asp Arg Phe Ile Pro

465 470 475 480

Leu Pro Phe Gly Lys Arg Thr Ser Gln Leu Tyr Ile Asp Pro Leu Thr

485 490 495  
Ala Lys Lys Phe Lys Asp Ala Phe Pro Ala Ile Glu Arg Asn Pro Asn

500 505 510  
Pro Phe Gly Ile Phe Gln Leu Ile Ala Ser Thr Pro Asp Met Ala Thr

515 520 525  
Leu Thr Ala Arg Arg Arg Glu Met Glu Asp Tyr Leu Asp Leu Ala Tyr

530 535 540  
Glu Leu Glu Asp Lys Leu Tyr Ala Ser Ile Pro Tyr Tyr Glu Asp Ser

545 550 555 560  
Arg Phe Gln Gly Phe Leu Gly Gln Val Lys Thr Ala Lys Val Leu Leu

565 570 575  
Asp Trp Ile Asn Glu Val Pro Glu Ala Arg Ile Tyr Glu Thr Tyr Ser

580 585 590  
Ile Asp Pro Gly Asp Leu Tyr Arg Leu Leu Glu Leu Ala Asp Trp Leu

595 600 605  
Met Tyr Ser Leu Ile Glu Leu Tyr Lys Leu Phe Glu Pro Lys Glu Glu

610 615 620  
Ile Leu Asn Tyr Leu Arg Asp Leu His Leu Arg Leu Arg His Gly Val

625 630 635 640  
Arg Glu Glu Leu Leu Glu Leu Val Arg Leu Pro Asn Ile Gly Arg Lys

645 650 655  
Arg Ala Arg Ala Leu Tyr Asn Ala Gly Phe Arg Ser Val Glu Ala Ile

660 665 670  
Ala Asn Ala Lys Pro Ala Glu Leu Leu Ala Val Glu Gly Ile Gly Ala

675 680 685  
Lys Ile Leu Asp Gly Ile Tyr Arg His Leu Gly Ile Glu Lys Arg Val

690 695 700  
Thr Glu Glu Lys Pro Lys Arg Lys Gly Thr Leu Glu Asp Phe Leu Arg

705 710 715 720

<210> 11

<211> 799

<212> PRT

<213> Methanospirillum hungatei

<400> 11

Met Glu Ile Ala Ser Leu Pro Leu Pro Asp Ser Phe Ile Arg Ala Cys

1 5 10 15

His Ala Lys Gly Ile Arg Ser Leu Tyr Pro Pro Gln Ala Glu Cys Ile

20 25 30

Glu Lys Gly Leu Leu Glu Gly Lys Asn Leu Leu Ile Ser Ile Pro Thr

35 40 45

Ala Ser Gly Lys Thr Leu Leu Ala Glu Met Ala Met Trp Ser Arg Ile

50 55 60

Ala Ala Gly Gly Lys Cys Leu Tyr Ile Val Pro Leu Arg Ala Leu Ala

65 70 75 80

Ser Glu Lys Tyr Asp Glu Phe Ser Lys Lys Gly Val Ile Arg Val Gly

85 90 95

Ile Ala Thr Gly Asp Leu Asp Arg Thr Asp Ala Tyr Leu Gly Glu Asn

100 105 110

Asp Ile Ile Val Ala Thr Ser Glu Lys Thr Asp Ser Leu Leu Arg Asn

115 120 125

Arg Thr Pro Trp Leu Ser Gln Ile Thr Cys Ile Val Leu Asp Glu Val

130 135 140

His Leu Ile Gly Ser Glu Asn Arg Gly Ala Thr Leu Glu Met Val Ile

145 150 155 160

Thr Lys Leu Arg Tyr Thr Asn Pro Val Met Gln Ile Ile Gly Leu Ser

165 170 175

Ala Thr Ile Gly Asn Pro Ala Gln Leu Ala Glu Trp Leu Asp Ala Thr

180 185 190

Leu Ile Thr Ser Thr Trp Arg Pro Val Asp Leu Arg Gln Gly Val Tyr

195 200 205

Tyr Asn Gly Lys Ile Arg Phe Ser Asp Ser Glu Arg Pro Ile Gln Gly

210 215 220



Lys Thr Lys His Asp Asp Leu Asn Leu Cys Leu Asp Thr Ile Glu Glu  
 225 230 235 240  
 Gly Gly Gln Cys Leu Val Phe Val Ser Ser Arg Arg Asn Ala Glu Gly  
 245 250 255  
 Phe Ala Lys Lys Ala Ala Gly Ala Leu Lys Ala Gly Ser Pro Asp Ser  
 260 265 270  
 Lys Ala Leu Ala Gln Glu Leu Arg Arg Leu Arg Asp Arg Asp Glu Gly  
 275 280 285  
  
 Asn Val Leu Ala Asp Cys Val Glu Arg Gly Ala Ala Phe His His Ala  
 290 295 300  
 Gly Leu Ile Arg Gln Glu Arg Thr Ile Ile Glu Glu Gly Phe Arg Asn  
 305 310 315 320  
 Gly Tyr Ile Glu Val Ile Ala Ala Thr Pro Thr Leu Ala Ala Gly Leu  
 325 330 335  
 Asn Leu Pro Ala Arg Arg Val Ile Ile Arg Asp Tyr Asn Arg Phe Ala  
 340 345 350  
  
 Ser Gly Leu Gly Met Val Pro Ile Pro Val Gly Glu Tyr His Gln Met  
 355 360 365  
 Ala Gly Arg Ala Gly Arg Pro His Leu Asp Pro Tyr Gly Glu Ala Val  
 370 375 380  
 Leu Leu Ala Lys Asp Ala Pro Ser Val Glu Arg Leu Phe Glu Thr Phe  
 385 390 395 400  
 Ile Asp Ala Glu Ala Glu Arg Val Asp Ser Gln Cys Val Asp Asp Ala  
 405 410 415  
  
 Ser Leu Cys Ala His Ile Leu Ser Leu Ile Ala Thr Gly Phe Ala His  
 420 425 430  
 Asp Gln Glu Ala Leu Ser Ser Phe Met Glu Arg Thr Phe Tyr Phe Phe  
 435 440 445  
 Gln His Pro Lys Thr Arg Ser Leu Pro Arg Leu Val Ala Asp Ala Ile  
 450 455 460  
 Arg Phe Leu Thr Thr Ala Gly Met Val Glu Glu Arg Glu Asn Thr Leu

465					470								475								480				
Ser	Ala	Thr	Arg	Leu	Gly	Ser	Leu	Val	Ser	Arg	Leu	Tyr	Leu	Asn	Pro										
				485								490								495					
Cys	Thr	Ala	Arg	Leu	Ile	Leu	Asp	Ser	Leu	Lys	Ser	Cys	Lys	Thr	Pro										
				500								505								510					
Thr	Leu	Ile	Gly	Leu	Leu	His	Val	Ile	Cys	Val	Ser	Pro	Asp	Met	Gln										
				515								520								525					
Arg	Leu	Tyr	Leu	Lys	Ala	Ala	Asp	Thr	Gln	Leu	Leu	Arg	Thr	Phe	Leu										
				530								535								540					
Phe	Lys	His	Lys	Asp	Asp	Leu	Ile	Leu	Pro	Leu	Pro	Phe	Glu	Gln	Glu										
				545								550								555					560
Glu	Glu	Glu	Leu	Trp	Leu	Ser	Gly	Leu	Lys	Thr	Ala	Leu	Val	Leu	Thr										
				565								570								575					
Asp	Trp	Ala	Asp	Glu	Phe	Ser	Glu	Gly	Met	Ile	Glu	Glu	Arg	Tyr	Gly										
				580								585								590					
Ile	Gly	Ala	Gly	Asp	Leu	Tyr	Asn	Ile	Val	Asp	Ser	Gly	Lys	Trp	Leu										
				595								600								605					
Leu	His	Gly	Thr	Glu	Arg	Leu	Val	Ser	Val	Glu	Met	Pro	Glu	Met	Ser										
				610								615								620					
Gln	Val	Val	Lys	Thr	Leu	Ser	Val	Arg	Val	His	His	Gly	Val	Lys	Ser										
				625								630								635					640
Glu	Leu	Leu	Pro	Leu	Val	Ala	Leu	Arg	Asn	Ile	Gly	Arg	Val	Arg	Ala										
				645								650								655					
Arg	Thr	Leu	Tyr	Asn	Ala	Gly	Tyr	Pro	Asp	Pro	Glu	Ala	Val	Ala	Arg										
				660								665								670					
Ala	Gly	Leu	Ser	Thr	Ile	Ala	Arg	Ile	Ile	Gly	Glu	Gly	Ile	Ala	Arg										
				675								680								685					
Gln	Val	Ile	Asp	Glu	Ile	Thr	Gly	Val	Lys	Arg	Ser	Gly	Ile	His	Ser										
				690								695								700					
Ser	Asp	Asp	Asp	Tyr	Gln	Gln	Lys	Thr	Pro	Glu	Leu	Leu	Thr	Asp	Ile										
				705								710								715					720

Pro Gly Ile Gly Lys Lys Met Ala Glu Lys Leu Gln Asn Ala Gly Ile  
725 730 735

Ile Thr Val Ser Asp Leu Leu Thr Ala Asp Glu Val Leu Leu Ser Asp  
740 745 750

Val Leu Gly Ala Ala Arg Ala Arg Lys Val Leu Ala Phe Leu Ser Asn  
755 760 765

Ser Glu Lys Glu Asn Ser Ser Ser Asp Lys Thr Glu Glu Ile Pro Asp  
770 775 780

Thr Gln Lys Ile Arg Gly Gln Ser Ser Trp Glu Asp Phe Gly Cys  
785 790 795

<210> 12

<211> 1756

<212> PRT

<213> Escherichia coli

<400> 12

Met Met Ser Ile Ala Gln Val Arg Ser Ala Gly Ser Ala Gly Asn Tyr  
1 5 10 15

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

Trp Ala Gly Lys Gly Ala Glu Gln Leu Gly Leu Gln Gly Ser Val Asp  
35 40 45

Lys Asp Val Phe Thr Arg Leu Leu Glu Gly Arg Leu Pro Asp Gly Ala

50 55 60

Asp Leu Ser Arg Met Gln Asp Gly Ser Asn Lys His Arg Pro Gly Tyr  
65 70 75 80

Asp Leu Thr Phe Ser Ala Pro Lys Ser Val Ser Met Met Ala Met Leu  
85 90 95

Gly Gly Asp Lys Arg Leu Ile Asp Ala His Asn Gln Ala Val Asp Phe  
100 105 110

Ala Val Arg Gln Val Glu Ala Leu Ala Ser Thr Arg Val Met Thr Asp

115 120 125

Gly Gln Ser Glu Thr Val Leu Thr Gly Asn Leu Val Met Ala Leu Phe  
 130 135 140  
 Asn His Asp Thr Ser Arg Asp Gln Glu Pro Gln Leu His Thr His Ala  
 145 150 155 160  
 Val Val Ala Asn Val Thr Gln His Asn Gly Glu Trp Lys Thr Leu Ser  
 165 170 175  
 Ser Asp Lys Val Gly Lys Thr Gly Phe Ile Glu Asn Val Tyr Ala Asn  
 180 185 190  
 Gln Ile Ala Phe Gly Arg Leu Tyr Arg Glu Lys Leu Lys Glu Gln Val  
 195 200 205  
 Glu Ala Leu Gly Tyr Glu Thr Glu Val Val Gly Lys His Gly Met Trp  
 210 215 220  
 Glu Met Pro Gly Val Pro Val Glu Ala Phe Ser Gly Arg Ser Gln Ala  
 225 230 235 240  
 Ile Arg Glu Ala Val Gly Glu Asp Ala Ser Leu Lys Ser Arg Asp Val  
 245 250 255  
 Ala Ala Leu Asp Thr Arg Lys Ser Lys Gln His Val Asp Pro Glu Ile  
 260 265 270  
 Arg Met Ala Glu Trp Met Gln Thr Leu Lys Glu Thr Gly Phe Asp Ile  
 275 280 285  
 Arg Ala Tyr Arg Asp Ala Ala Asp Gln Arg Thr Glu Ile Arg Thr Gln  
 290 295 300  
 Ala Pro Gly Pro Ala Ser Gln Asp Gly Pro Asp Val Gln Gln Ala Val  
 305 310 315 320  
 Thr Gln Ala Ile Ala Gly Leu Ser Glu Arg Lys Val Gln Phe Thr Tyr  
 325 330 335  
 Thr Asp Val Leu Ala Arg Thr Val Gly Ile Leu Pro Pro Glu Asn Gly  
 340 345 350  
 Val Ile Glu Arg Ala Arg Ala Gly Ile Asp Glu Ala Ile Ser Arg Glu  
 355 360 365  
 Gln Leu Ile Pro Leu Asp Arg Glu Lys Gly Leu Phe Thr Ser Gly Ile

370                      375                      380  
 His Val Leu Asp Glu Leu Ser Val Arg Ala Leu Ser Arg Asp Ile Met  
 385                      390                      395                      400  
 Lys Gln Asn Arg Val Thr Val His Pro Glu Lys Ser Val Pro Arg Thr  
                          405                      410                      415  
 Ala Gly Tyr Ser Asp Ala Val Ser Val Leu Ala Gln Asp Arg Pro Ser  
                          420                      425                      430  
 Leu Ala Ile Val Ser Gly Gln Gly Gly Ala Ala Gly Gln Arg Glu Arg  
  
                          435                      440                      445  
 Val Ala Glu Leu Val Met Met Ala Arg Glu Gln Gly Arg Glu Val Gln  
                          450                      455                      460  
 Ile Ile Ala Ala Asp Arg Arg Ser Gln Met Asn Leu Lys Gln Asp Glu  
 465                      470                      475                      480  
 Arg Leu Ser Gly Glu Leu Ile Thr Gly Arg Arg Gln Leu Leu Glu Gly  
                          485                      490                      495  
 Met Ala Phe Thr Pro Gly Ser Thr Val Ile Val Asp Gln Gly Glu Lys  
  
                          500                      505                      510  
 Leu Ser Leu Lys Glu Thr Leu Thr Leu Leu Asp Gly Ala Ala Arg His  
                          515                      520                      525  
 Asn Val Gln Val Leu Ile Thr Asp Ser Gly Gln Arg Thr Gly Thr Gly  
                          530                      535                      540  
 Ser Ala Leu Met Ala Met Lys Asp Ala Gly Val Asn Thr Tyr Arg Trp  
 545                      550                      555                      560  
 Gln Gly Gly Glu Gln Arg Pro Ala Thr Ile Ile Ser Glu Pro Asp Arg  
  
                          565                      570                      575  
 Asn Val Arg Tyr Ala Arg Leu Ala Gly Asp Phe Ala Ala Ser Val Lys  
                          580                      585                      590  
 Ala Gly Glu Glu Ser Val Ala Gln Val Ser Gly Val Arg Glu Gln Ala  
                          595                      600                      605  
 Ile Leu Thr Gln Ala Ile Arg Ser Glu Leu Lys Thr Gln Gly Val Leu  
                          610                      615                      620

Gly His Pro Glu Val Thr Met Thr Ala Leu Ser Pro Val Trp Leu Asp

625 630 635 640

Ser Arg Ser Arg Tyr Leu Arg Asp Met Tyr Arg Pro Gly Met Val Met

645 650 655

Glu Gln Trp Asn Pro Glu Thr Arg Ser His Asp Arg Tyr Val Ile Asp

660 665 670

Arg Val Thr Ala Gln Ser His Ser Leu Thr Leu Arg Asp Ala Gln Gly

675 680 685

Glu Thr Gln Val Val Arg Ile Ser Ser Leu Asp Ser Ser Trp Ser Leu

690 695 700

Phe Arg Pro Glu Lys Met Pro Val Ala Asp Gly Glu Arg Leu Arg Val

705 710 715 720

Thr Gly Lys Ile Pro Gly Leu Arg Val Ser Gly Gly Asp Arg Leu Gln

725 730 735

Val Ala Ser Val Ser Glu Asp Ala Met Thr Val Val Val Pro Gly Arg

740 745 750

Ala Glu Pro Ala Ser Leu Pro Val Ser Asp Ser Pro Phe Thr Ala Leu

755 760 765

Lys Leu Glu Asn Gly Trp Val Glu Thr Pro Gly His Ser Val Ser Asp

770 775 780

Ser Ala Thr Val Phe Ala Ser Val Thr Gln Met Ala Met Asp Asn Ala

785 790 795 800

Thr Leu Asn Gly Leu Ala Arg Ser Gly Arg Asp Val Arg Leu Tyr Ser

805 810 815

Ser Leu Asp Glu Thr Arg Thr Ala Glu Lys Leu Ala Arg His Pro Ser

820 825 830

Phe Thr Val Val Ser Glu Gln Ile Lys Ala Arg Ala Gly Glu Thr Leu

835 840 845

Leu Glu Thr Ala Ile Ser Leu Gln Lys Ala Gly Leu His Thr Pro Ala

850 855 860

Gln Gln Ala Ile His Leu Ala Leu Pro Val Leu Glu Ser Lys Asn Leu



865                      870                      875                      880  
 Ala Phe Ser Met Val Asp Leu Leu Thr Glu Ala Lys Ser Phe Ala Ala  
  
                          885                      890                      895  
 Glu Gly Thr Gly Phe Thr Glu Leu Gly Gly Glu Ile Asn Ala Gln Ile  
                          900                      905                      910  
 Lys Arg Gly Asp Leu Leu Tyr Val Asp Val Ala Lys Gly Tyr Gly Thr  
                          915                      920                      925  
 Gly Leu Leu Val Ser Arg Ala Ser Tyr Glu Ala Glu Lys Ser Ile Leu  
                          930                      935                      940  
 Arg His Ile Leu Glu Gly Lys Glu Ala Val Thr Pro Leu Met Glu Arg  
  
 945                      950                      955                      960  
 Val Pro Gly Glu Leu Met Glu Thr Leu Thr Ser Gly Gln Arg Ala Ala  
                          965                      970                      975  
 Thr Arg Met Ile Leu Glu Thr Ser Asp Arg Phe Thr Val Val Gln Gly  
                          980                      985                      990  
 Tyr Ala Gly Val Gly Lys Thr Thr Gln Phe Arg Ala Val Met Ser Ala  
                          995                      1000                      1005  
 Val Asn Met Leu Pro Ala Ser Glu Arg Pro Arg Val Val Gly Leu  
  
 1010                      1015                      1020  
 Gly Pro Thr His Arg Ala Val Gly Glu Met Arg Ser Ala Gly Val  
                          1025                      1030                      1035  
 Asp Ala Gln Thr Leu Ala Ser Phe Leu His Asp Thr Gln Leu Gln  
                          1040                      1045                      1050  
 Gln Arg Ser Gly Glu Thr Pro Asp Phe Ser Asn Thr Leu Phe Leu  
                          1055                      1060                      1065  
 Leu Asp Glu Ser Ser Met Val Gly Asn Thr Glu Met Ala Arg Ala  
  
 1070                      1075                      1080  
 Tyr Ala Leu Ile Ala Ala Gly Gly Gly Arg Ala Val Ala Ser Gly  
                          1085                      1090                      1095  
 Asp Thr Asp Gln Leu Gln Ala Ile Ala Pro Gly Gln Ser Phe Arg  
                          1100                      1105                      1110

Leu Gln	Gln Thr Arg Ser Ala	Ala Asp Val Val Ile	Met Lys Glu
1115	1120	1125	
Ile Val	Arg Gln Thr Pro Glu	Leu Arg Glu Ala Val	Tyr Ser Leu
1130	1135	1140	
Ile Asn	Arg Asp Val Glu Arg	Ala Leu Ser Gly Leu	Glu Ser Val
1145	1150	1155	
Lys Pro	Ser Gln Val Pro Arg	Leu Glu Gly Ala Trp	Ala Pro Glu
1160	1165	1170	
His Ser	Val Thr Glu Phe Ser	His Ser Gln Glu Ala	Lys Leu Ala
1175	1180	1185	
Glu Ala	Gln Gln Lys Ala Met	Leu Lys Gly Glu Ala	Phe Pro Asp
1190	1195	1200	
Ile Pro	Met Thr Leu Tyr Glu	Ala Ile Val Arg Asp	Tyr Thr Gly
1205	1210	1215	
Arg Thr	Pro Glu Ala Arg Glu	Gln Thr Leu Ile Val	Thr His Leu
1220	1225	1230	
Asn Glu	Asp Arg Arg Val Leu	Asn Ser Met Ile His	Asp Ala Arg
1235	1240	1245	
Glu Lys	Ala Gly Glu Leu Gly	Lys Glu Gln Val Met	Val Pro Val
1250	1255	1260	
Leu Asn	Thr Ala Asn Ile Arg	Asp Gly Glu Leu Arg	Arg Leu Ser
1265	1270	1275	
Thr Trp	Glu Lys Asn Pro Asp	Ala Leu Ala Leu Val	Asp Asn Val
1280	1285	1290	
Tyr His	Arg Ile Ala Gly Ile	Ser Lys Asp Asp Gly	Leu Ile Thr
1295	1300	1305	
Leu Gln	Asp Ala Glu Gly Asn	Thr Arg Leu Ile Ser	Pro Arg Glu
1310	1315	1320	
Ala Val	Ala Glu Gly Val Thr	Leu Tyr Thr Pro Asp	Lys Ile Arg
1325	1330	1335	
Val Gly	Thr Gly Asp Arg Met	Arg Phe Thr Lys Ser	Asp Arg Glu

1340	1345	1350
Arg Gly Tyr Val Ala Asn Ser	Val Trp Thr Val Thr	Ala Val Ser
1355	1360	1365
Gly Asp Ser Val Thr Leu Ser	Asp Gly Gln Gln Thr	Arg Val Ile
1370	1375	1380
Arg Pro Gly Gln Glu Arg Ala	Glu Gln His Ile Asp	Leu Ala Tyr
1385	1390	1395
Ala Ile Thr Ala His Gly Ala	Gln Gly Ala Ser Glu	Thr Phe Ala
1400	1405	1410
Ile Ala Leu Glu Gly Thr Glu	Gly Asn Arg Lys Leu	Met Ala Gly
1415	1420	1425
Phe Glu Ser Ala Tyr Val Ala	Leu Ser Arg Met Lys	Gln His Val
1430	1435	1440
Gln Val Tyr Thr Asp Asn Arg	Gln Gly Trp Thr Asp	Ala Ile Asn
1445	1450	1455
Asn Ala Val Gln Lys Gly Thr	Ala His Asp Val Leu	Glu Pro Lys
1460	1465	1470
Pro Asp Arg Glu Val Met Asn	Ala Gln Arg Leu Phe	Ser Thr Ala
1475	1480	1485
Arg Glu Leu Arg Asp Val Ala	Ala Gly Arg Ala Val	Leu Arg Gln
1490	1495	1500
Ala Gly Leu Ala Gly Gly Asp	Ser Pro Ala Arg Phe	Ile Ala Pro
1505	1510	1515
Gly Arg Lys Tyr Pro Gln Pro	Tyr Val Ala Leu Pro	Ala Phe Asp
1520	1525	1530
Arg Asn Gly Lys Ser Ala Gly	Ile Trp Leu Asn Pro	Leu Thr Thr
1535	1540	1545
Asp Asp Gly Asn Gly Leu Arg	Gly Phe Ser Gly Glu	Gly Arg Val
1550	1555	1560
Lys Gly Ser Gly Asp Ala Gln	Phe Val Ala Leu Gln	Gly Ser Arg
1565	1570	1575

Asn Gly Glu Ser Leu Leu Ala Asp Asn Met Gln Asp Gly Val Arg  
 1580 1585 1590  
 Ile Ala Arg Asp Asn Pro Asp Ser Gly Val Val Val Arg Ile Ala  
 1595 1600 1605  
 Gly Glu Gly Arg Pro Trp Asn Pro Gly Ala Ile Thr Gly Gly Arg  
 1610 1615 1620  
 Val Trp Gly Asp Ile Pro Asp Asn Ser Val Gln Pro Gly Ala Gly  
 1625 1630 1635  
 Asn Gly Glu Pro Val Thr Ala Glu Val Leu Ala Gln Arg Gln Ala  
 1640 1645 1650  
 Glu Glu Ala Ile Arg Arg Glu Thr Glu Arg Arg Ala Asp Glu Ile  
 1655 1660 1665  
 Val Arg Lys Met Ala Glu Asn Lys Pro Asp Leu Pro Asp Gly Lys  
 1670 1675 1680  
 Thr Glu Leu Ala Val Arg Asp Ile Ala Gly Gln Glu Arg Asp Arg  
 1685 1690 1695  
 Ser Ala Ile Ser Glu Arg Glu Thr Ala Leu Pro Glu Ser Val Leu  
 1700 1705 1710  
 Arg Glu Ser Gln Arg Glu Arg Glu Ala Val Arg Glu Val Ala Arg  
 1715 1720 1725  
 Glu Asn Leu Leu Gln Glu Arg Leu Gln Gln Met Glu Arg Asp Met  
 1730 1735 1740  
 Val Arg Asp Leu Gln Lys Glu Lys Thr Leu Gly Gly Asp  
 1745 1750 1755  
 <210> 13  
 <211> 726  
 <212> PRT  
 <213> Methanococcoides burtonii  
 <400> 13  
 Met Ser Asp Lys Pro Ala Phe Met Lys Tyr Phe Thr Gln Ser Ser Cys  
 1 5 10 15  
 Tyr Pro Asn Gln Gln Glu Ala Met Asp Arg Ile His Ser Ala Leu Met

20	25	30
Gln Gln Gln Leu Val Leu Phe Glu Gly Ala Cys Gly Thr Gly Lys Thr		
35	40	45
Leu Ser Ala Leu Val Pro Ala Leu His Val Gly Lys Met Leu Gly Lys		
50	55	60
Thr Val Ile Ile Ala Thr Asn Val His Gln Gln Met Val Gln Phe Ile		
65	70	75
Asn Glu Ala Arg Asp Ile Lys Lys Val Gln Asp Val Lys Val Ala Val		
85	90	95
Ile Lys Gly Lys Thr Ala Met Cys Pro Gln Glu Ala Asp Tyr Glu Glu		
100	105	110
Cys Ser Val Lys Arg Glu Asn Thr Phe Glu Leu Met Glu Thr Glu Arg		
115	120	125
Glu Ile Tyr Leu Lys Arg Gln Glu Leu Asn Ser Ala Arg Asp Ser Tyr		
130	135	140
Lys Lys Ser His Asp Pro Ala Phe Val Thr Leu Arg Asp Glu Leu Ser		
145	150	155
Lys Glu Ile Asp Ala Val Glu Glu Lys Ala Arg Gly Leu Arg Asp Arg		
165	170	175
Ala Cys Asn Asp Leu Tyr Glu Val Leu Arg Ser Asp Ser Glu Lys Phe		
180	185	190
Arg Glu Trp Leu Tyr Lys Glu Val Arg Ser Pro Glu Glu Ile Asn Asp		
195	200	205
His Ala Ile Lys Asp Gly Met Cys Gly Tyr Glu Leu Val Lys Arg Glu		
210	215	220
Leu Lys His Ala Asp Leu Leu Ile Cys Asn Tyr His His Val Leu Asn		
225	230	235
Pro Asp Ile Phe Ser Thr Val Leu Gly Trp Ile Glu Lys Glu Pro Gln		
245	250	255
Glu Thr Ile Val Ile Phe Asp Glu Ala His Asn Leu Glu Ser Ala Ala		
260	265	270

Arg Ser His Ser Ser Leu Ser Leu Thr Glu His Ser Ile Glu Lys Ala  
275 280 285

Ile Thr Glu Leu Glu Ala Asn Leu Asp Leu Leu Ala Asp Asp Asn Ile  
290 295 300

His Asn Leu Phe Asn Ile Phe Leu Glu Val Ile Ser Asp Thr Tyr Asn  
305 310 315 320

Ser Arg Phe Lys Phe Gly Glu Arg Glu Arg Val Arg Lys Asn Trp Tyr  
325 330 335

Asp Ile Arg Ile Ser Asp Pro Tyr Glu Arg Asn Asp Ile Val Arg Gly  
340 345 350

Lys Phe Leu Arg Gln Ala Lys Gly Asp Phe Gly Glu Lys Asp Asp Ile  
355 360 365

Gln Ile Leu Leu Ser Glu Ala Ser Glu Leu Gly Ala Lys Leu Asp Glu  
370 375 380

Thr Tyr Arg Asp Gln Tyr Lys Lys Gly Leu Ser Ser Val Met Lys Arg  
385 390 395 400

Ser His Ile Arg Tyr Val Ala Asp Phe Met Ser Ala Tyr Ile Glu Leu  
405 410 415

Ser His Asn Leu Asn Tyr Tyr Pro Ile Leu Asn Val Arg Arg Asp Met  
420 425 430

Asn Asp Glu Ile Tyr Gly Arg Val Glu Leu Phe Thr Cys Ile Pro Lys  
435 440 445

Asn Val Thr Glu Pro Leu Phe Asn Ser Leu Phe Ser Val Ile Leu Met  
450 455 460

Ser Ala Thr Leu His Pro Phe Glu Met Val Lys Lys Thr Leu Gly Ile  
465 470 475 480

Thr Arg Asp Thr Cys Glu Met Ser Tyr Gly Thr Ser Phe Pro Glu Glu  
485 490 495

Lys Arg Leu Ser Ile Ala Val Ser Ile Pro Pro Leu Phe Ala Lys Asn  
500 505 510

Arg Asp Asp Arg His Val Thr Glu Leu Leu Glu Gln Val Leu Leu Asp



515                      520                      525  
 Ser Ile Glu Asn Ser Lys Gly Asn Val Ile Leu Phe Phe Gln Ser Ala  
 530                      535                      540

Phe Glu Ala Lys Arg Tyr Tyr Ser Lys Ile Glu Pro Leu Val Asn Val  
 545                      550                      555                      560  
 Pro Val Phe Leu Asp Glu Val Gly Ile Ser Ser Gln Asp Val Arg Glu  
 565                      570                      575  
 Glu Phe Phe Ser Ile Gly Glu Glu Asn Gly Lys Ala Val Leu Leu Ser  
 580                      585                      590  
 Tyr Leu Trp Gly Thr Leu Ser Glu Gly Ile Asp Tyr Arg Asp Gly Arg  
 595                      600                      605

Gly Arg Thr Val Ile Ile Ile Gly Val Gly Tyr Pro Ala Leu Asn Asp  
 610                      615                      620  
 Arg Met Asn Ala Val Glu Ser Ala Tyr Asp His Val Phe Gly Tyr Gly  
 625                      630                      635                      640  
 Ala Gly Trp Glu Phe Ala Ile Gln Val Pro Thr Ile Arg Lys Ile Arg  
 645                      650                      655  
 Gln Ala Met Gly Arg Val Val Arg Ser Pro Thr Asp Tyr Gly Ala Arg  
 660                      665                      670

Ile Leu Leu Asp Gly Arg Phe Leu Thr Asp Ser Lys Lys Arg Phe Gly  
 675                      680                      685  
 Lys Phe Ser Val Phe Glu Val Phe Pro Pro Ala Glu Arg Ser Glu Phe  
 690                      695                      700  
 Val Asp Val Asp Pro Glu Lys Val Lys Tyr Ser Leu Met Asn Phe Phe  
 705                      710                      715                      720  
 Met Asp Asn Asp Glu Gln  
 725

<210> 14

<211> 439

<212> PRT

<213> Enterobacteria phage T4

<400> 14

Met Thr Phe Asp Asp Leu Thr Glu Gly Gln Lys Asn Ala Phe Asn Ile  
1 5 10 15

Val Met Lys Ala Ile Lys Glu Lys Lys His His Val Thr Ile Asn Gly  
20 25 30

Pro Ala Gly Thr Gly Lys Thr Thr Leu Thr Lys Phe Ile Ile Glu Ala  
35 40 45

Leu Ile Ser Thr Gly Glu Thr Gly Ile Ile Leu Ala Ala Pro Thr His  
50 55 60

Ala Ala Lys Lys Ile Leu Ser Lys Leu Ser Gly Lys Glu Ala Ser Thr  
65 70 75 80

Ile His Ser Ile Leu Lys Ile Asn Pro Val Thr Tyr Glu Glu Asn Val  
85 90 95

Leu Phe Glu Gln Lys Glu Val Pro Asp Leu Ala Lys Cys Arg Val Leu  
100 105 110

Ile Cys Asp Glu Val Ser Met Tyr Asp Arg Lys Leu Phe Lys Ile Leu  
115 120 125

Leu Ser Thr Ile Pro Pro Trp Cys Thr Ile Ile Gly Ile Gly Asp Asn  
130 135 140

Lys Gln Ile Arg Pro Val Asp Pro Gly Glu Asn Thr Ala Tyr Ile Ser  
145 150 155 160

Pro Phe Phe Thr His Lys Asp Phe Tyr Gln Cys Glu Leu Thr Glu Val  
165 170 175

Lys Arg Ser Asn Ala Pro Ile Ile Asp Val Ala Thr Asp Val Arg Asn  
180 185 190

Gly Lys Trp Ile Tyr Asp Lys Val Val Asp Gly His Gly Val Arg Gly  
195 200 205

Phe Thr Gly Asp Thr Ala Leu Arg Asp Phe Met Val Asn Tyr Phe Ser  
210 215 220

Ile Val Lys Ser Leu Asp Asp Leu Phe Glu Asn Arg Val Met Ala Phe  
225 230 235 240

Thr Asn Lys Ser Val Asp Lys Leu Asn Ser Ile Ile Arg Lys Lys Ile

245 250 255

Phe Glu Thr Asp Lys Asp Phe Ile Val Gly Glu Ile Ile Val Met Gln

260 265 270

Glu Pro Leu Phe Lys Thr Tyr Lys Ile Asp Gly Lys Pro Val Ser Glu

275 280 285

Ile Ile Phe Asn Asn Gly Gln Leu Val Arg Ile Ile Glu Ala Glu Tyr

290 295 300

Thr Ser Thr Phe Val Lys Ala Arg Gly Val Pro Gly Glu Tyr Leu Ile

305 310 315 320

Arg	His	Trp	Asp	Leu	Thr	Val	Glu	Thr	Tyr	Gly	Asp	Asp	Glu	Tyr	Tyr	
				325					330					335		
Arg	Glu	Lys	Ile	Lys	Ile	Ile	Ser	Ser	Asp	Glu	Glu	Leu	Tyr	Lys	Phe	
				340					345					350		
Asn	Leu	Phe	Leu	Gly	Lys	Thr	Ala	Glu	Thr	Tyr	Lys	Asn	Trp	Asn	Lys	
				355				360						365		
Gly	Gly	Lys	Ala	Pro	Trp	Ser	Asp	Phe	Trp	Asp	Ala	Lys	Ser	Gln	Phe	
				370			375				380					

Ser	Lys	Val	Lys	Ala	Leu	Pro	Ala	Ser	Thr	Phe	His	Lys	Ala	Gln	Gly
385					390					395					400
Met	Ser	Val	Asp	Arg	Ala	Phe	Ile	Tyr	Thr	Pro	Cys	Ile	His	Tyr	Ala
				405					410					415	
Asp	Val	Glu	Leu	Ala	Gln	Gln	Leu	Leu	Tyr	Val	Gly	Val	Thr	Arg	Gly
			420					425					430		
Arg	Tyr	Asp	Val	Phe	Tyr	Val									

<210> 15

<211> 970

<212> PRT

<213> Clostridium botulinum

<400> 15

Met Leu Ser Val Ala Asn Val Arg Ser Pro Ser Ala Ala Ala Ser Tyr

1	5	10	15												
Phe	Ala	Ser	Asp	Asn	Tyr	Tyr	Ala	Ser	Ala	Asp	Ala	Asp	Arg	Ser	Gly
	20		25		30										
Gln	Trp	Ile	Gly	Asp	Gly	Ala	Lys	Arg	Leu	Gly	Leu	Glu	Gly	Lys	Val
	35		40		45										
Glu	Ala	Arg	Ala	Phe	Asp	Ala	Leu	Leu	Arg	Gly	Glu	Leu	Pro	Asp	Gly
	50		55		60										
Ser	Ser	Val	Gly	Asn	Pro	Gly	Gln	Ala	His	Arg	Pro	Gly	Thr	Asp	Leu
65			70				75							80	
Thr	Phe	Ser	Val	Pro	Lys	Ser	Trp	Ser	Leu	Leu	Ala	Leu	Val	Gly	Lys
	85		90				95								
Asp	Glu	Arg	Ile	Ile	Ala	Ala	Tyr	Arg	Glu	Ala	Val	Val	Glu	Ala	Leu
	100		105				110								
His	Trp	Ala	Glu	Lys	Asn	Ala	Ala	Glu	Thr	Arg	Val	Val	Glu	Lys	Gly
	115		120				125								
Met	Val	Val	Thr	Gln	Ala	Thr	Gly	Asn	Leu	Ala	Ile	Gly	Leu	Phe	Gln
	130		135				140								
His	Asp	Thr	Asn	Arg	Asn	Gln	Glu	Pro	Asn	Leu	His	Phe	His	Ala	Val
145			150				155							160	
Ile	Ala	Asn	Val	Thr	Gln	Gly	Lys	Asp	Gly	Lys	Trp	Arg	Thr	Leu	Lys
	165		170				175								
Asn	Asp	Arg	Leu	Trp	Gln	Leu	Asn	Thr	Thr	Leu	Asn	Ser	Ile	Ala	Met
	180		185				190								
Ala	Arg	Phe	Arg	Val	Ala	Val	Glu	Lys	Leu	Gly	Tyr	Glu	Pro	Gly	Pro
	195		200				205								
Val	Leu	Lys	His	Gly	Asn	Phe	Glu	Ala	Arg	Gly	Ile	Ser	Arg	Glu	Gln
	210		215				220								
Val	Met	Ala	Phe	Ser	Thr	Arg	Arg	Lys	Glu	Val	Leu	Glu	Ala	Arg	Arg
225			230				235							240	
Gly	Pro	Gly	Leu	Asp	Ala	Gly	Arg	Ile	Ala	Ala	Leu	Asp	Thr	Arg	Ala
	245		250				255								

Ser Lys Glu Gly Ile Glu Asp Arg Ala Thr Leu Ser Lys Gln Trp Ser  
260 265 270

Glu Ala Ala Gln Ser Ile Gly Leu Asp Leu Lys Pro Leu Val Asp Arg  
275 280 285

Ala Arg Thr Lys Ala Leu Gly Gln Gly Met Glu Ala Thr Arg Ile Gly  
290 295 300

Ser Leu Val Glu Arg Gly Arg Ala Trp Leu Ser Arg Phe Ala Ala His  
305 310 315 320

Val Arg Gly Asp Pro Ala Asp Pro Leu Val Pro Pro Ser Val Leu Lys  
325 330 335

Gln Asp Arg Gln Thr Ile Ala Ala Ala Gln Ala Val Ala Ser Ala Val  
340 345 350

Arg His Leu Ser Gln Arg Glu Ala Ala Phe Glu Arg Thr Ala Leu Tyr  
355 360 365

Lys Ala Ala Leu Asp Phe Gly Leu Pro Thr Thr Ile Ala Asp Val Glu  
370 375 380

Lys Arg Thr Arg Ala Leu Val Arg Ser Gly Asp Leu Ile Ala Gly Lys  
385 390 395 400

Gly Glu His Lys Gly Trp Leu Ala Ser Arg Asp Ala Val Val Thr Glu  
405 410 415

Gln Arg Ile Leu Ser Glu Val Ala Ala Gly Lys Gly Asp Ser Ser Pro  
420 425 430

Ala Ile Thr Pro Gln Lys Ala Ala Ala Ser Val Gln Ala Ala Ala Leu  
435 440 445

Thr Gly Gln Gly Phe Arg Leu Asn Glu Gly Gln Leu Ala Ala Ala Arg  
450 455 460

Leu Ile Leu Ile Ser Lys Asp Arg Thr Ile Ala Val Gln Gly Ile Ala  
465 470 475 480

Gly Ala Gly Lys Ser Ser Val Leu Lys Pro Val Ala Glu Val Leu Arg  
485 490 495

Asp Glu Gly His Pro Val Ile Gly Leu Ala Ile Gln Asn Thr Leu Val

500	505	510	
Gln Met Leu Glu Arg Asp Thr Gly Ile Gly Ser Gln Thr Leu Ala Arg			
515	520	525	
Phe Leu Gly Gly Trp Asn Lys Leu Leu Asp Asp Pro Gly Asn Val Ala			
530	535	540	
Leu Arg Ala Glu Ala Gln Ala Ser Leu Lys Asp His Val Leu Val Leu			
545	550	555	560
Asp Glu Ala Ser Met Val Ser Asn Glu Asp Lys Glu Lys Leu Val Arg			
565	570	575	
Leu Ala Asn Leu Ala Gly Val His Arg Leu Val Leu Ile Gly Asp Arg			
580	585	590	
Lys Gln Leu Gly Ala Val Asp Ala Gly Lys Pro Phe Ala Leu Leu Gln			
595	600	605	
Arg Ala Gly Ile Ala Arg Ala Glu Met Ala Thr Asn Leu Arg Ala Arg			
610	615	620	
Asp Pro Val Val Arg Glu Ala Gln Ala Ala Ala Gln Ala Gly Asp Val			
625	630	635	640
Arg Lys Ala Leu Arg His Leu Lys Ser His Thr Val Glu Ala Arg Gly			
645	650	655	
Asp Gly Ala Gln Val Ala Ala Glu Thr Trp Leu Ala Leu Asp Lys Glu			
660	665	670	
Thr Arg Ala Arg Thr Ser Ile Tyr Ala Ser Gly Arg Ala Ile Arg Ser			
675	680	685	
Ala Val Asn Ala Ala Val Gln Gln Gly Leu Leu Ala Ser Arg Glu Ile			
690	695	700	
Gly Pro Ala Lys Met Lys Leu Glu Val Leu Asp Arg Val Asn Thr Thr			
705	710	715	720
Arg Glu Glu Leu Arg His Leu Pro Ala Tyr Arg Ala Gly Arg Val Leu			
725	730	735	
Glu Val Ser Arg Lys Gln Gln Ala Leu Gly Leu Phe Ile Gly Glu Tyr			
740	745	750	



Arg Val Ile Gly Gln Asp Arg Lys Gly Lys Leu Val Glu Val Glu Asp  
755 760 765

Lys Arg Gly Lys Arg Phe Arg Phe Asp Pro Ala Arg Ile Arg Ala Gly  
770 775 780

Lys Gly Asp Asp Asn Leu Thr Leu Leu Glu Pro Arg Lys Leu Glu Ile  
785 790 795 800

His Glu Gly Asp Arg Ile Arg Trp Thr Arg Asn Asp His Arg Arg Gly  
805 810 815

Leu Phe Asn Ala Asp Gln Ala Arg Val Val Glu Ile Ala Asn Gly Lys  
820 825 830

Val Thr Phe Glu Thr Ser Lys Gly Asp Leu Val Glu Leu Lys Lys Asp  
835 840 845

Asp Pro Met Leu Lys Arg Ile Asp Leu Ala Tyr Ala Leu Asn Val His  
850 855 860

Met Ala Gln Gly Leu Thr Ser Asp Arg Gly Ile Ala Val Met Asp Ser  
865 870 875 880

Arg Glu Arg Asn Leu Ser Asn Gln Lys Thr Phe Leu Val Thr Val Thr  
885 890 895

Arg Leu Arg Asp His Leu Thr Leu Val Val Asp Ser Ala Asp Lys Leu  
900 905 910

Gly Ala Ala Val Ala Arg Asn Lys Gly Glu Lys Ala Ser Ala Ile Glu  
915 920 925

Val Thr Gly Ser Val Lys Pro Thr Ala Thr Lys Gly Ser Gly Val Asp  
930 935 940

Gln Pro Lys Ser Val Glu Ala Asn Lys Ala Glu Lys Glu Leu Thr Arg  
945 950 955 960

Ser Lys Ser Lys Thr Leu Asp Phe Gly Ile  
965 970

<210> 16

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Sequence used in Example 1.

<400> 16

tttttttttt tttttttttt tttttttttt tttttttttt 40

<210> 17

<211> 40

<212> RNA

<213> Artificial Sequence

<220><223> Sequence used in Example 1.

<400> 17

gguuguuuucu guuggugcug auauugcauu uaagagcauu 40

<210> 18

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Sequence used in Examples 1 and 3.

<400> 18

gcaatatcag caccaacaga aacaacct 28

<210> 19

<211> 55

<212> RNA

<213> Artificial Sequence

<220><223> Sequence used in Example 2.

<400> 19

gggaaauaag agagaccaca ccaccacac caccacacc acccacacca cacca 55

<210> 20

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Sequence used in Example 2.

<400> 20

tctctcttat ttcccagggtt aaacacccaa gcagacgcc 39

<210> 21

<211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence used in Example 2.  
 <400> 21  
 ggcgtctgct tgggtgttta acct 24  
 <210> 22  
 <211> 40  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence used in Example 3.  
 <400> 22  
 cccccccca cccccccca cccccccca cccccccca 40  
 <210> 23  
 <211> 51  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence used in Example 3.  
 <400> 23  
 ggttggttct gttggtgctg atattgcggc gtctgcttgg gtgtttaacc t 51  
 <210> 24  
 <211> 32  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence used in Example 3.  
 <400> 24  
 tctctcttat cccaacacc caagcagacg cc 32  
 <210> 25  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence used in Example 3.  
 <400> 25  
 ggcgtctgct tgggtgttta acct 24

<210> 26

<211> 1653

<212> RNA

<213>

> Artificial Sequence

<220><223> Sequence used in Examples 3 and 5.

<400> 26

```

auggaagacg ccaaaaacau aaagaaaggc ccggcgccau ucuaucgcu ggaagaugga      60
accgcuggag agcaacugca uaaggcuaua aagagauacg ccugguucc uggaacaauu      120
gcuuuuacag augcacauau cgagguggac aucacuuacg cugaguacuu cgaaaugucc      180
guucgguggg cagaagcuau gaaacgauau gggcugaaua caaaucacag aaucgucgua      240
ugcagugaaa acucucuua auucuuuaua ccgguguugg gcgcguuuuu uaucggaguu      300
gcaguugcgc ccgcgaacga cauuuauaua gaacgugaau ugcucaacag uaugggcauu      360

ucgcagccua ccgugguguu cguuucaaaa aaggguugc aaaaauuuu gaacgugcaa      420
aaaaagcucc caaucaucca aaaaauuuu aucauggauu cuaaaacgga uuaccaggga      480
uuucagucga uguacacguu cgucacauau caucuaccuc ccgguuuuaa ugaauacgau      540
uuugucccag aguccuucga uagggacaag acaauugcac ugaucaugaa cuccucugga      600
ucuacugguc ugccuaaagg ugucgcucug ccucauagaa cugccugcgu gagauucugc      660
caugccagag auccuauuuu uggcaaucaa aucauuccgg auacugcgau uuaaaguguu      720
guuccauucc aucacgguuu uggaauuuu acuaacacug gauauuugau auguggauuu      780

cgagucgucu uaauguauag auuugaagaa gaggcuguuuc ugaggagccu ucaggauuac      840
aagauucaa ggcgcugcu ggugccaacc cuauucuccu ucuucgcaa aagcacucug      900
auugacaaau acgauuuuac uauuuuacac gaaauugcuu cugguggcgc uccccucucu      960
aaggaagucg gggaagcggg ugccaagagg uucaucugc cagguaucag gcaaggauau      1020
gggcucacug agacuacauc agcuauucug auuacacccg agggggauga uaaaccgggc      1080
gcggucggua aaguuguucc auuuuuugaa gcgaagguug uggaucugga uaccgggaaa      1140
acgcugggcg uaaaucaaag aggcgaacug ugugugagag guccuauau uauuccggu      1200

uauuuuacg uauuuuacg uauuuuacg uauuuuacg uauuuuacg uauuuuacg      1260
ggagacauag cuuacuggga cgaagacgaa cacuucuua ucguugaccg ccugaagucu      1320
cugauuuagu acaaaggcua ucagguggcu cccgcugaau uggaauccau cuugcuccaa      1380
caccucaaca ucuucgacgc agguugcgca ggucuucccg acgaugacgc cggugaacuu      1440
ccgcgcgccg uuguuuuuu ggagcacgga aagacgauga cggaaaaaga gaucugggau      1500

```

uacgucgcca gucaaguaac aaccgcgaaa aaguugcgcg gaggaguugu guuuguggac 1560

gaaguaccga aaggucuuac cggaaaacuc gacgcaagaa aaucagaga gauccucaua 1620

aaggccaaga agggcggaaa gaucgccgug uaa 1653

<210> 27

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Sequence used in Examples 4 and 5.

<400> 27

cccccccca cccccccca cccccccca agaacataa acagaacgtt ttttttt 58

<210> 28

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Sequence used in Example 4.

<400> 28

cccccccca cccccccca cccccccca agaacataa acagaacgtt t 51

<210> 29

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Sequence used in Examples 4 and 5.

<400> 29

cgttctgttt atgtttcttg gacactgatt gacacggttt agtagaac 48

<210> 30

<211> 288

<212> RNA

<213> Artificial Sequence

<220><223> Sequence used in Examples 4 and 6.

<400> 30

gggaaauaag agagaaaaga agaguaagaa gaaauauaag agccaccaug guuuauauug 60

cggccgcuaa auuaagcugc cuucugcggg gcuugccuuc uggccaugcc cuucucucu 120

cccuugcacc uguaccucu ggucuuugaa uaaagccuga guaggaagaa aaaaaaaaa	180
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	240
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa	288
<210> 31	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence used in Example 6.	
<400> 31	
cccccccca	10
<210> 32	
<211> 27	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence used in Example 6.	
<400> 32	
ggttgtttct gttggtgctg atattgc	27
<210> 33	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence used to illustrate homopolymer read.	
<400> 33	
aaaaaaaaaa aa	12
<210> 34	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence used to illustrate homopolymer read.	
<220><221> n	
<222> (4)..(4)	
<223> Inosine	
<220><221> misc_feature	

<222> (4)..(4)

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

<220><221> n

<222> (7)..(8)

<223> Inosine

<220><221> misc\_feature

<222> (7)..(8)

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

<220><221> n

<222> (12)..(12)

<223> Inosine

<220><221> misc\_feature

<222> (12)..(12)

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

<400> 34

tttntntntt tn

12

<210> 35

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Sequence used in Example 7.

<400> 35

gcaatatcag caccaacaga aarcrcararc rc

32

<210> 36

<211> 48

<212>

> DNA

<213> Artificial Sequence

<220><223> Sequence used in Example 7.

<400> 36

ttctgaatta ggaaccattt tttttttttt atgatgcaag atacgcac

48

<210> 37

<211> 36

<212> DNA



<213> Artificial Sequence	
<220><223> Sequence used in Example 7.	
<400> 37	
gaggcgagcg gtcaatttgt gcgtatcttg catcat	36
<210> 38	
<211> 18	
<212> DNA	
<213> Artificial Sequence	
<220><223> 39	
<400> 38	
atggttctcta attcagaa	18
<210> 39	
<211> 15	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence used in Example 7.	
<400> 39	
ttgaccgctc gcctc	15
<210> 40	
<211> 37	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence used in Example 7.	
<400> 40	
ggttggtttct gttggtgctg atattgcttt tttttt	37
<210> 41	
<211> 897	
<212> DNA	
<213> Eisenia fetida	
<400> 41	
atgagtgcga aggctgctga aggttatgaa caaatcgaag ttgatgtggt tgctgtgtgg	60
aaggaagggtt atgtgtatga aaatcgtggt agtacctccg tggatcaaaa aattaccatc	120
acgaaaggca tgaagaacgt taatagcgaa acccgtagcg tcaccgcgac gcattctatt	180

ggcagtacca tciccacggg tgacgccttt gaaatcggct ccgtggaagt ttcataattcg 240  
 catagccacg aagaatcaca agtttcgatg accgaaacgg aagcttacga atcaaaagtg 300  
 attgaacaca ccattacgat cccgccgacc tcgaagtcca cgcgctggca gctgaacgca 360  
 gatgtcggcg gtgctgacat tgaatataatg tacctgatcg atgaagttac cccgattggc 420  
 ggtacgcaga gtattccgca agtgatcacc tcccgtgcaa aaattatcgt tggtcgccag 480

attatcctgg gcaagaccga aattcgtatc aaacatgctg aacgcaagga atatatgacc 540  
 gtggttagcc gtaaactctg gccggcggcc acgctgggtc acagtaaact gtttaagttc 600  
 gtgctgtacg aagattgggg cggttttcgc atcaaaaccc tgaatacgaat gtattctggt 660  
 tatgaatacg cgtatagctc tgaccagggc ggtatctact tcgatcaagg caccgacaac 720  
 ccgaaacagc gtggggccat taataagagc ctgccgctgc gccatggtga tgtcgtgacc 780  
 tttatgaaca aatacttcac gcgttctggt ctgtgctatg atgacggccc ggcgaccaat 840  
 gtgtattgtc tggataaacg cgaagacaag tggattctgg aagttgtcgg ctaatga 897

<210> 42

<211> 297

<212> PRT

<213> Eisenia fetida

<400> 42

Met Ser Ala Lys Ala Ala Glu Gly Tyr Glu Gln Ile Glu Val Asp Val

1 5 10 15

Val Ala Val Trp Lys Glu Gly Tyr Val Tyr Glu Asn Arg Gly Ser Thr

20 25 30

Ser Val Asp Gln Lys Ile Thr Ile Thr Lys Gly Met Lys Asn Val Asn

35 40 45

Ser Glu Thr Arg Thr Val Thr Ala Thr His Ser Ile Gly Ser Thr Ile

50 55 60

Ser Thr Gly Asp Ala Phe Glu Ile Gly Ser Val Glu Val Ser Tyr Ser

65 70 75 80

His Ser His Glu Glu Ser Gln Val Ser Met Thr Glu Thr Glu Val Tyr

85 90 95

Glu Ser Lys Val Ile Glu His Thr Ile Thr Ile Pro Pro Thr Ser Lys

100 105 110

Phe Thr Arg Trp Gln Leu Asn Ala Asp Val Gly Gly Ala Asp Ile Glu

115 120 125

Tyr Met Tyr Leu Ile Asp Glu Val Thr Pro Ile Gly Gly Thr Gln Ser

130 135 140

Ile Pro Gln Val Ile Thr Ser Arg Ala Lys Ile Ile Val Gly Arg Gln

145 150 155 160

Ile Ile Leu Gly Lys Thr Glu Ile Arg Ile Lys His Ala Glu Arg Lys

165 170 175

Glu Tyr Met Thr Val Val Ser Arg Lys Ser Trp Pro Ala Ala Thr Leu

180 185 190

Gly His Ser Lys Leu Phe Lys Phe Val Leu Tyr Glu Asp Trp Gly Gly

195 200 205

Phe Arg Ile Lys Thr Leu Asn Thr Met Tyr Ser Gly Tyr Glu Tyr Ala

210 215 220

Tyr Ser Ser Asp Gln Gly Gly Ile Tyr Phe Asp Gln Gly Thr Asp Asn

225 230 235 240

Pro Lys Gln Arg Trp Ala Ile Asn Lys Ser Leu Pro Leu Arg His Gly

245 250 255

Asp Val Val Thr Phe Met Asn Lys Tyr Phe Thr Arg Ser Gly Leu Cys

260 265 270

Tyr Asp Asp Gly Pro Ala Thr Asn Val Tyr Cys Leu Asp Lys Arg Glu

275 280 285

Asp Lys Trp Ile Leu Glu Val Val Gly

290 295

<210> 43

<211> 786

<212> DNA

<213> Escherichia coli

<400> 43

tgtctgaccg caccgccgaa agaagcggca cgtccgaccc tgatgccgcg tgcacagtct 60

tataaagatc tgacccatct gccggtccg acgggcaaaa tttttgtag cgtctataac 120

atccaggacg aaaccggta atttaaaccg tacccggcga gtaattttctc cacggccgtt 180  
ccgcagagtg caaccgctat gctgggtcacg gactgaaag attcccggtg gttcattccg 240  
ctggaacgcc agggcctgca aaacctgctg aatgaacgta aaattatccg cgcagctcag 300  
gaaaacggta ccgtggccat taacaatcgt attccgctgc aaagcctgac cgccgcaaac 360  
atcatgggtt aaggctctat catcggttac gaatcaaacg tcaaatcggg cgggtgtgggc 420  
gcacgttatt ttggcattgg tgctgatacc cagtaccaac tggaccagat cgcagttaac 480

ctgcgcgtgg ttaatgtcag caccggcgaa attctgagct ctgtgaatac cagcaaaacg 540  
atcctgtctt acgaagtga ggctgggtgtt ttctgtttca ttgattatca acgctgctg 600  
gaaggcgaag tcggttacac ctcaaacgaa ccggtgatgc tgtgtctgat gtcggcgatt 660  
gaaacgggtg ttattttctt gatcaatgat ggcatcgacc gtggtctgtg ggatctgcag 720  
aacaagccg aacgtcaaaa tgacattctg gtgaaatacc gccacatgag tgttcgccg 780  
gaatcc 786

<210> 44

<211> 262

<212> PRT

<213> Escherichia coli

<400> 44

Cys Leu Thr Ala Pro Pro Lys Glu Ala Ala Arg Pro Thr Leu Met Pro

1 5 10 15

Arg Ala Gln Ser Tyr Lys Asp Leu Thr His Leu Pro Ala Pro Thr Gly

20 25 30

Lys Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe

35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala

50 55 60

Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Ile Pro

65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile

85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Ile Asn Asn Arg Ile Pro

100 105 110

Leu Gln Ser Leu Thr Ala Ala Asn Ile Met Val Glu Gly Ser Ile Ile

115					120					125					
Gly	Tyr	Glu	Ser	Asn	Val	Lys	Ser	Gly	Gly	Val	Gly	Ala	Arg	Tyr	Phe
130					135					140					
Gly	Ile	Gly	Ala	Asp	Thr	Gln	Tyr	Gln	Leu	Asp	Gln	Ile	Ala	Val	Asn
145					150					155					160
Leu	Arg	Val	Val	Asn	Val	Ser	Thr	Gly	Glu	Ile	Leu	Ser	Ser	Val	Asn
165					170					175					
Thr	Ser	Lys	Thr	Ile	Leu	Ser	Tyr	Glu	Val	Gln	Ala	Gly	Val	Phe	Arg
180					185					190					
Phe	Ile	Asp	Tyr	Gln	Arg	Leu	Leu	Glu	Gly	Glu	Val	Gly	Tyr	Thr	Ser
195					200					205					
Asn	Glu	Pro	Val	Met	Leu	Cys	Leu	Met	Ser	Ala	Ile	Glu	Thr	Gly	Val
210					215					220					
Ile	Phe	Leu	Ile	Asn	Asp	Gly	Ile	Asp	Arg	Gly	Leu	Trp	Asp	Leu	Gln
225					230					235					240
Asn	Lys	Ala	Glu	Arg	Gln	Asn	Asp	Ile	Leu	Val	Lys	Tyr	Arg	His	Met
245					250					255					

<400> 46  
 cccccccca cccccccca cccccccca cccccccca 40  
 <210> 47  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 47  
 actcgcagat cattacgac 20  
 <210> 48  
 <211> 39  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 48  
 rarcrcrg rcrargraru rcrarurura rcrgarurc 39  
 <210> 49  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 49  
 cgattgacta agctatacgc 20  
 <210> 50  
 <211> 40  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 50  
 rcrcraruru rgrarcrura rargrcrura rurarcrgc 40  
 <210> 51  
 <211> 10  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 51  
 tttttttttt 10  
 <210> 52  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 52  
 tttttttttt tttttttttt 20  
 <210> 53  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 53  
 ccccccccca ccccccccca 20  
 <210> 54  
 <211> 20  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 54  
 rarararara rarararara 20  
 <210> 55  
 <211> 20  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 55  
 rurcrararu rarcrgrara 20  
 <210> 56



<211> 30  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 56  
 aactactagg atcatcgatg tatctgctca 30  
 <210> 57  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
  
 <400> 57  
 agcttaacat acgatactct tagctaacca 30  
 <210> 58  
 <211> 60  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 58  
 rararcrua rcrurargrg rarurcraru rcrgrarurg rurarurcru rgrcrurcra 60  
 <210> 59  
 <211> 60  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 59  
 rargrcruu rararcruu rarcgraru rarcruu rurargrcru rararcra 60  
 <210> 60  
  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Sequence described in the description.  
 <400> 60

actctgaacc	10
<210> 61	
<211> 15	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence described in the description.	
<400> 61	
actctrgrar arcrc	15
<210> 62	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence described in the description.	
<400> 62	
gcacaatgat	10
<210> 63	
<211> 15	
<212> DNA	
<213> Artificial Sequence	
<220><223> Sequence described in the description.	
<400> 63	
gcacarartr grart	15
<210> 64	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Sequence described in the description.	
<400> 64	
rarararcru rarcgrcru	20
<210> 65	
<211> 32	
<212> DNA	
<213> Artificial Sequence	

<220><223> Sequence used in Example 7.

<400> 65

ttgttctact aaaccgtgtc aatcagtggtc tt 32

<210> 66

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Sequence used in Example 7.

<400> 66

gcaatattcag caccaacaga aacaacctt 29