



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0005082
(43) 공개일자 2025년01월09일

(51) 국제특허분류(Int. Cl.)
C12N 9/14 (2006.01) C07K 14/245 (2006.01)
C12Q 1/6869 (2018.01)
(52) CPC특허분류
C12N 9/14 (2013.01)
C07K 14/245 (2013.01)
(21) 출원번호 10-2024-7032256
(22) 출원일자(국제) 2023년04월14일
심사청구일자 없음
(85) 번역문제출일자 2024년09월26일
(86) 국제출원번호 PCT/EP2023/059821
(87) 국제공개번호 WO 2023/198911
국제공개일자 2023년10월19일
(30) 우선권주장
2205617.0 2022년04월14일 영국(GB)

(71) 출원인
옥스포드 나노포어 테크놀로지즈 피엘씨
영국, 옥스퍼드셔 오엑스4 4디큐, 옥스퍼드, 옥스퍼드 사이언스 파크, 에드먼드 홀리 로드, 고슬링 빌딩
(72) 발명자
보웬, 레베카 빅토리아
영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나노포어 테크놀로지즈 피엘씨
브루스, 마크 존
영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나노포어 테크놀로지즈 피엘씨
(뒷면에 계속)
(74) 대리인
양영준, 김영

전체 청구항 수 : 총 63 항

(54) 발명의 명칭 신규 변형된 단백질 가공 및 효소

(57) 요약

본 발명은 폴리뉴클레오티드와 같은 분석물의 이동을 제어하는 데 사용될 수 있는 변형된 Dda 헬리카제에 관한 것이다. 변형된 Dda 헬리카제는 분석물 검출 및 특성화에 사용된다. 본 발명은 또한 신규 단백질 가공 및 분석물 검출 및 특성화에서 이들의 용도에 관한 것이다. 본 발명은 특히 CsgG-유사 가공 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체에 의해 형성된 단리된 가공 복합체에 관한 것이며, 이에 의해 나노기공에 추가적인 채널 수축 또는 판독 헤드를 혼입한다.

(52) CPC특허분류

C12Q 1/6869 (2018.05)
C07K 2319/02 (2013.01)
C07K 2319/21 (2013.01)
C07K 2319/22 (2013.01)
C07K 2319/50 (2013.01)
C12Q 2521/513 (2013.01)
C12Q 2565/631 (2013.01)

(72) 발명자

월리스, 엘리자베스 제인

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

무디, 폴 리차드

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

부르사, 프란시스

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

페이지, 데이비드 크리스토퍼

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

모사예비, 마지드

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

헤런, 앤드류 존

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

리에라, 알베르토

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

유드, 크리스토퍼 피터

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

포스터, 리차드 찰스

영국 옥스포드 오엑스4 4디큐 옥스포드 사이언스
 파크 에드먼드 홀리 로드 고슬링 빌딩 옥스포드 나
 노포어 테크놀로지즈 피엘씨

명세서

청구범위

청구항 1

변형된 DNA 의존적 ATPase(Dda) 헬리카제로서, 상기 헬리카제가 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 위치 중 하나 이상에서 변형 또는 치환을 포함하는 것인, 변형된 DNA 의존적 ATPase(Dda) 헬리카제.

청구항 2

제1항에 있어서, 상기 (a) 위치 55에 상응하는 아미노산이 D, E, K, N 또는 S로 치환되고/되거나, (b) 위치 114에 상응하는 아미노산이 A, V, I, L, M, F, Y, W, G, P, S, T, N 또는 Q로 치환되고/되거나, (c) 위치 156에 상응하는 아미노산이 A, E, F, G, I, L, M, P, S, V, Y, D, K 또는 N으로 치환되고/되거나, (d) 위치 177에 상응하는 아미노산이 D, E, F, G, H, I, L, M, N, Q, R, S, T, V, W 또는 Y로 치환되고/되거나, (e), 위치 210에 상응하는 아미노산이 D, E, K, S, N, R, H 또는 Y로 치환되고/되거나 (f), 위치 221에서 아미노산이 D, K, E, Q, R, A, H, L, T 또는 Y로 치환되고/되거나, (g) 위치 350에 상응하는 아미노산이 A, D, E, G, K, L, N, Q, R, T, V, H 또는 M으로 치환되거나 D, E, A, V, I, L, M, F, W, R, H, K, L, S, T, N 또는 Q로 치환되고/되거나 (h) 위치 358에 상응하는 아미노산이 D, E, A, V, I, L, M, F, Y, W, R, H, L, S, T, N 또는 Q로 치환되는 것인, 변형된 DNA 의존적 ATPase(Dda) 헬리카제.

청구항 3

제1항 또는 제2항에 있어서, 상기 헬리카제가 Dda 1993의 아미노산 위치 40에 상응하는 위치에서 변형 또는 치환을 추가로 포함하는 것인, 변형된 Dda 헬리카제.

청구항 4

제3항에 있어서, 상기 위치 40에 상응하는 아미노산이 A, V, I, L, M, F, Y 또는 W로 치환되는 것인, 변형된 DNA 의존적 ATPase(Dda) 헬리카제.

청구항 5

제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 및 추가적인 폴리뉴클레오티드 결합 모이어티를 포함하는 작제물로서, 상기 헬리카제가 폴리뉴클레오티드 결합 모이어티에 부착되고 작제물이 분석물의 이동을 제어하는 능력을 갖는 것인, 작제물.

청구항 6

제5항에 있어서, 상기 작제물이 제1항 내지 제4항 중 어느 한 항에 따른 2개 이상의 헬리카제를 포함하는 것인, 작제물.

청구항 7

제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물을 암호화하는 서열을 포함하는 폴리뉴클레오티드.

청구항 8

프로모터에 작동가능하게 연결된 제7항에 따른 폴리뉴클레오티드를 포함하는 벡터.

청구항 9

제8항에 따른 벡터를 포함하는 숙주 세포.

청구항 10

제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물을 제조하는

방법으로서, 제7항에 따른 폴리뉴클레오티드를 발현시키거나, 제8항에 따른 벡터로 세포를 형질감염시키거나 제9항에 따른 숙주 세포를 배양하는 단계를 포함하는, 방법.

청구항 11

분석물의 이동을 제어하는 방법으로서, 분석물을 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물과 접촉시켜 분석물의 이동을 제어하는 단계를 포함하는, 방법.

청구항 12

제11항에 있어서, 상기 방법이 막관통 기공을 통해 분석물의 이동을 제어하기 위한 것인, 방법.

청구항 13

표적 분석물을 특성화하는 방법으로서,

(a) 표적 분석물을 막관통 기공 및 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물과 접촉시켜 헬리카제 또는 작제물이 기공을 통해 표적 분석물의 이동을 제어하도록 하는 단계; 및

(b) 폴리뉴클레오티드가 기공과 관련하여 이동할 때 하나 이상의 측정을 수행하여 측정이 표적 분석물의 하나 이상의 특성을 나타내는 것이고 이에 의해 표적 분석물을 특성화하는 단계

를 포함하는, 방법.

청구항 14

제13항에 있어서, 상기 하나 이상의 특성이 (i) 표적 분석물의 길이, (ii) 표적 분석물의 동일성, (iii) 표적 분석물의 서열, (iv) 표적 분석물의 2차 구조 및 (v) 표적 분석물이 변형되는지 여부로부터 선택되는 것인, 방법.

청구항 15

제14항에 있어서, 상기 표적 분석물이 메틸화, 산화, 손상, 하나 이상의 단백질 또는 하나 이상의 라벨, 태그 또는 스페이서에 의해 변형되는 것인, 방법.

청구항 16

제13항 내지 제15항 중 어느 한 항에 있어서, 상기 표적 분석물의 하나 이상의 특성이 전기적 측정 및/또는 광학적 측정에 의해 측정되는 것인, 방법.

청구항 17

제16항에 있어서, 상기 전기적 측정이 전류 측정, 임피던스 측정, 터널링 측정 또는 전계 효과 트랜지스터(FET) 측정인, 방법.

청구항 18

제17항에 있어서, 상기 방법이

(a) 표적 분석물을 막관통 기공 및 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물과 접촉시켜 헬리카제 또는 작제물이 기공을 통해 표적 분석물의 이동을 제어하도록 하는 단계; 및

(b) 폴리뉴클레오티드가 기공에 대해 이동할 때 기공을 통해 통과하는 전류를 측정하여 전류가 표적 분석물의 하나 이상의 특성을 나타내는 것이고 이에 의해 표적 분석물을 특성화하는 단계

를 포함하는 것인, 방법.

청구항 19

제13항 내지 제18항 중 어느 한 항에 있어서, 상기 방법이 기공을 가로질러 전압을 인가하여 기공과 헬리카제 또는 작제물 사이에 복합체를 형성하는 단계를 추가로 포함하는 것인, 방법.

청구항 20

제13항 내지 제20항 중 어느 한 항에 있어서, 상기 폴리뉴클레오티드의 적어도 일부가 이중 가닥인, 방법.

청구항 21

제13항 내지 제20항 중 어느 한 항에 있어서, 상기 기공이 막관통 단백질 기공 또는 고체 상태 기공인, 방법.

청구항 22

제21항에 있어서, 상기 막관통 단백질 기공이 용혈소, 백혈구독, 바이코박테리움 스메그마티스 포린 A(MspA), MspB, MspC, MspD, 리세닌, 외부 막 포린 F(OmpF), 외부 막 포린 G(OmpG), 외부 막 포스포리파제 A, 네이세리아 자동수송체 지단백질(NalP) 및 WZA, CsgG, CsgG/CsgF, 리세닌, ClyA, Sp1 또는 용혈성 단백질 프라가세아톡신 C(FraC)로부터 유래되는 것인, 방법.

청구항 23

제22항에 있어서, 상기 막관통 단백질이 서열번호: 3에 제시된 바와 같은 6 내지 10개의 동일한 서브유닛 또는 (b) 서브유닛 중 하나 이상이 전체 서열에 걸쳐 서열번호: 3에 대해 적어도 40% 동일성을 갖고 기공 활성을 갖는 이의 변이체로 형성되는 것인, 방법.

청구항 24

표적 분석물을 특성화하기 위한 센서를 형성하는 방법으로서, (a) 기공과 (b) 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물 사이에 복합체를 형성하여 표적 분석물의 특성화를 위한 센서를 형성하는 단계를 포함하는, 방법.

청구항 25

제24항에 있어서, 상기 복합체가 (a) 표적 분석물의 존재 하에 기공과 헬리카제 또는 작제물을 접촉시키고 (a) 기공을 가로질러 전위를 인가함으로써 형성되는 것인, 방법.

청구항 26

제25항에 있어서, 상기 전위가 전압 전위 또는 화학 전위인, 방법.

청구항 27

제26항에 있어서, 상기 복합체가 기공을 헬리카제 또는 작제물에 공유적으로 부착함으로써 형성되는 것인, 방법.

청구항 28

표적 분석물을 특성화하기 위한 센서로서, (a) 기공과 (b) 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물 사이에 복합체를 포함하는, 센서.

청구항 29

기공을 통해 표적 분석물의 이동을 제어하기 위한 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물의 용도.

청구항 30

표적 분석물을 특성화하기 위한 키트로서,

(a) 기공 및 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물; 또는

(b) 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제 또는 제5항 또는 제6항에 따른 작제물 및 하나 이상의 로딩 모이어티

를 포함하는, 키트.

청구항 31

샘플 내 표적 분석물을 특성화하기 위한 장치로서, (a) 복수의 기공 및 (b) 제1항 내지 제4항 중 어느 한 항에 따른 복수의 헬리카제 또는 제5항 또는 제6항에 따른 복수의 작제물을 포함하는, 장치.

청구항 32

제1항 내지 제4항 중 어느 한 항에 따른 헬리카제를 생산하는 방법으로서,

(a) 헬리카제를 제공하는 단계; 및

(b) 헬리카제를 변형시켜 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제를 생산하는 단계를 포함하는, 방법.

청구항 33

제32항에 있어서, 상기 방법이 (c) 생성된 헬리카제가 폴리뉴클레오티드의 이동을 제어할 수 있는지 여부를 결정하는 단계를 추가로 포함하는 것인, 방법.

청구항 34

제5항 또는 제6항에 따른 작제물을 생산하는 방법으로서, 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제를 추가적인 폴리뉴클레오티드 결합 모이어티에 부착시켜 작제물을 생산하는 단계를 포함하는, 방법.

청구항 35

제34항에 있어서, 상기 방법이 생성된 작제물이 폴리뉴클레오티드의 이동을 제어할 수 있는지 여부를 결정하는 단계를 추가로 포함하는 것인, 방법.

청구항 36

폴리뉴클레오티드에 부착된 일련의 2개 이상의 헬리카제로서, 상기 2개 이상의 헬리카제 중 적어도 하나가 제1항 내지 제4항 중 어느 한 항에 따른 헬리카제인, 일련의 2개 이상의 헬리카제.

청구항 37

이동이 DNA 의존적 ATPase(Dda) 헬리카제에 의해 제어될 때 막관통 기공에 대한 표적 분석물의 이동을 개선하는 방법으로서, 상기 DNA 의존적 ATPase(Dda) 헬리카제가 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 위치 및/또는 Dda 1993의 아미노산 위치 40에 상응하는 위치 중 하나 이상에서 치환을 포함하도록 변형되어 막관통 기공에 대한 표적 분석물의 이동을 개선하는 것인, 방법.

청구항 38

단리된 CsgG 기공 또는 이의 상동체 또는 돌연변이체, 또는 CsgG 기공, 또는 이의 상동체 또는 돌연변이체, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는 단리된 기공 복합체로서, 상기 CsgG 기공이 서열번호: 117의 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 적어도 하나의 단량체를 포함하는 것인, 단리된 기공 복합체.

청구항 39

제38항에 있어서, 상기 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체가 CsgG 기공, 또는 이의 상동체 또는 돌연변이체의 루멘에 삽입되는 것인, 단리된 기공 복합체.

청구항 40

제38항 또는 제39항에 있어서, 상기 기공 복합체가 CsgG 채널 수축 및 CsgF 채널 수축을 포함하는 2개 이상의 채널 수축을 갖는 것인, 단리된 기공 복합체.

청구항 41

제40항에 있어서, 상기 CsgF 채널 수축이 0.5 nm 내지 2.0 nm 범위의 직경을 갖는 것인, 단리된 기공 복합체.

청구항 42

제38항 내지 제41항 중 어느 한 항에 따른 기공 또는 기공 복합체, 및 막의 구성요소를 포함하는 단리된 막관통 기공 복합체.

청구항 43

제38항 내지 제41항 중 어느 한 항에 따른 기공 또는 기공 복합체를 포함하는 막.

청구항 44

제43항에 따른 복수의 막을 포함하는 어레이.

청구항 45

(a) 제43항에 따른 막 또는 제44항에 따른 어레이, (b) 막(들)을 가로질러 전위를 인가하기 위한 수단 및 (c) 막(들)을 가로질러 전기적 또는 광학적 신호를 검출하기 위한 수단을 포함하는 시스템.

청구항 46

제42항에 따른 막관통 기공 복합체를 생산하는 방법으로서, 적합한 숙주 세포에서 CsgG 기공, 또는 이의 상동체 또는 돌연변이체, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 공동 발현하여 생체 내 막관통 기공 복합체 형성을 허용하는 단계를 포함하는, 방법.

청구항 47

제38항 내지 제41항 중 어느 한 항에 따른 단리된 기공 복합체를 생산하는 방법으로서, CsgG 단량체, 또는 이의 상동체 또는 돌연변이체를 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체와 접촉시켜 단리된 기공 복합체의 시험관 내 재구성을 허용하는 단계를 포함하는, 방법.

청구항 48

표적 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 방법으로서,

(a) 표적 분석물을 제38항 내지 제41항 중 어느 한 항에 따른 단리된 기공 또는 단리된 기공 복합체 또는 제42항에 따른 막관통 기공 복합체와 접촉시켜 표적 분석물이 기공 복합체에 대해 또는 내로 이동하도록 하는 단계; 및

(b) 분석물이 기공 복합체를 통해 이동할 때 하나 이상의 측정을 수행하여 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 단계

를 포함하는, 방법.

청구항 49

제48항에 있어서, 상기 기공 복합체에 대한 표적 분석물의 이동이 폴리뉴클레오티드 결합 단백질, 제1항 내지 제4항 중 어느 한 항에 따른 Dda 헬리카제 또는 제5항 또는 제6항에 따른 작제물에 의해 제어되는 것인, 방법.

청구항 50

제48항 또는 제49항에 있어서, 상기 분석물이 폴리뉴클레오티드인, 방법.

청구항 51

제48항 또는 제49항에 있어서, 상기 분석물이 (폴리)펩티드인, 방법.

청구항 52

제48항 또는 제49항에 있어서, 상기 분석물이 이당류인, 방법.

청구항 53

제50항에 있어서, (i) 폴리뉴클레오티드의 길이, (ii) 폴리뉴클레오티드의 동일성, (iii) 폴리뉴클레오티드의 서열, (iv) 폴리뉴클레오티드의 2차 구조 및 (v) 폴리뉴클레오티드가 변형되는지 여부로부터 선택된 하나 이상의 특성을 결정하는 단계를 포함하는, 방법.

청구항 54

제38항 내지 제41항 중 어느 한 항에 따른 단리된 기공 또는 단리된 기공 복합체 또는 제42항에 따른 막관통 기공 복합체를 사용하여 폴리뉴클레오티드 또는 (폴리)펩티드를 특성화하는 방법.

청구항 55

제46항 내지 제54항 중 어느 한 항에 있어서, 상기 CsgG 기공, 또는 이의 상동체 또는 돌연변이체가 6 내지 10 개의 단량체를 포함하는 것인, 방법.

청구항 56

표적 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하기 위한 제38항 내지 제41항 중 어느 한 항에 따른 단리된 기공 또는 단리된 기공 복합체 또는 제42항에 따른 막관통 기공 복합체의 용도.

청구항 57

표적 분석물이 기공을 통해 통과하는 속도를 변경하는 방법으로서, 표적 분석물을 제38항 내지 제41항 중 어느 한 항에 따른 단리된 기공 또는 단리된 기공 복합체 또는 제42항에 따른 막관통 기공 복합체와 접촉시켜 표적 분석물이 기공 복합체에 대해 또는 내로 이동하도록 하는 단계를 포함하는, 방법.

청구항 58

제57항에 있어서, 상기 기공 복합체에 대한 표적 분석물의 이동이 제1항 내지 제4항 중 어느 한 항에 따른 Dda 헬리카제 또는 제5항 또는 제6항에 따른 작제물에 의해 제어되는 것인, 방법.

청구항 59

(a) 제38항 내지 제41항 중 어느 한 항에 따른 단리된 기공 또는 단리된 기공 복합체 및 (b) 막의 구성요소 및 (c) 폴리뉴클레오티드 결합 단백질 중 하나 또는 둘 다를 포함하는 표적 분석물을 특성화하기 위한 키트.

청구항 60

표적 분석물을 특성화하는 방법으로서,

(a) 표적 분석물을 제38항 내지 제41항 중 어느 한 항에 따른 단리된 기공 또는 단리된 기공 복합체 및 제1항 내지 제4항 중 어느 한 항에 따른 DNA 의존적 ATPase(Dda) 헬리카제 또는 제5항 또는 제6항에 따른 헬리카제 작제물과 접촉시켜 헬리카제 또는 작제물이 기공 또는 기공 복합체를 통해 표적 분석물의 이동을 제어하도록 하는 단계; 및

(b) 폴리뉴클레오티드가 기공 또는 기공 복합체에 대해 이동할 때 하나 이상의 측정을 수행하여 상기 측정이 표적 분석물의 하나 이상의 특성을 나타내고 이에 의해 표적 분석물을 특성화하는 단계

를 포함하는, 방법.

청구항 61

(a) 제1항 내지 제4항 중 어느 한 항에 따른 DNA 의존적 ATPase(Dda) 헬리카제 또는 제5항 또는 제6항에 따른 헬리카제 작제물 및 (b) 제38항 내지 제41항 중 어느 한 항에 따른 단리된 CsgG 기공 또는 단리된 기공 복합체를 포함하는 표적 분석물을 특성화하기 위한 키트.

청구항 62

시험관 내 막에 삽입된 제38항 내지 제41항 중 어느 한 항에 따른 막관통 단백질 기공 또는 기공 복합체를 포함

하는 장치.

청구항 63

(i) 제38항 내지 제41항 중 어느 한 항에 따른 단리된 기공 또는 단리된 기공 복합체를 수득하는 단계 및 (ii) 단리된 기공 또는 단리된 기공 복합체를 시험관 내 막과 접촉시켜 기공이 시험관 내 막에 삽입되도록 하는 단계를 포함하는 방법에 의해 생산된 장치.

발명의 설명

기술 분야

[0001] 본 발명은 폴리뉴클레오티드와 같은 분석물의 이동을 제어하는 데 사용될 수 있는 변형된 Dda 헬리카제에 관한 것이다. 변형된 Dda 헬리카제는 분석물 검출 및 특성화에 사용된다. 본 발명은 또한 신규 단백질 기공 또는 기공 복합체 및 분석물 검출 및 특성화에서 이들의 용도에 관한 것이다.

배경 기술

[0002] 나노기공 감지를 사용한 분석물, 특히 중합체 특성화의 2가지 필수 구성요소는 (1) 기공을 통한 중합체 이동 제어 및 (2) 중합체가 기공을 통해 이동할 때 구성 빌딩 블록의 구별이다. 나노기공 감지 동안, 기공의 가장 좁은 부분은 전형적으로 나노기공에 대한 분석물 이동의 함수로서 측정 신호의 변화에 대한 나노기공의 가장 구별되는 부분에 상응한다. CsgG는 에스케리키아 콜라이(*Escherichia coli*)에서 게이팅되지 않은 비선택적 단백질 채널로서 식별되었으며(Goyal 등, 2014) 분석물을 검출하고 특성화하기 위한 나노기공으로 사용되었다. 이 맥락에서 기공의 특성을 개선하는 야생형 CsgG에 대한 돌연변이가 또한 개시되었다(WO2016/034591, WO2017/149316, WO2017/149317, WO2017/149318, WO2018/211241, 및 WO2019/002893, 이들의 전문이 본원에 참조로 포함됨). 또한 전문이 본원에 참조로 포함된 WO2015/055981, WO2015/166276 및 WO2016/055777은 본원에 기재된 CsgG 기공과 같은 막관통 단백질 기공에 대한 분석물의 이동을 제어하는 데 사용될 수 있는 폴리뉴클레오티드 결합 단백질, 구체적으로 Dda 헬리카제를 기재한다.

발명의 내용

[0003] 본 발명자들은 놀랍게도 기공을 통해 분석물의 이동을 제어하는 능력이 개선된 특이적 Dda 돌연변이체를 식별하였다. 기공을 사용하여 폴리뉴클레오티드를 서열분석할 때, 시스템은 기공을 통해 통과하는 염기/뉴클레오티드의 수 및 동일성을 공동으로 추정한다. 이동 속도의 가변성에 대한 더 나은 제어는 통계적 잡음의 원인 중 하나를 감소시키고 추정 작업을 단순화할 수 있다. 기공에서 폴리뉴클레오티드의 연속적인 짧은 체류의 실행은 기본적인 뉴클레오티드/염기 호출 실패를 촉발하여 삭제 오류를 초래할 수 있다. 비정상적으로 긴 체류는 삽입 오류로 이어질 수 있다. 각 뉴클레오티드/염기가 기공에서 충분한 시간 간격을 보내도록 보장하면 잡음 신호 수준으로부터 뉴클레오티드/염기 동일성의 통계적 불확실성을 해결하는 데 도움이 된다. 예를 들어 모터 효소와의 상호작용을 통해 뉴클레오티드/염기 동일성에 대한 체류 시간의 중속성으로부터 추가 정보를 추출할 수 있다. 체류 시간의 전반적인 가변성을 줄이면 이 채널을 통해 보다 정확한 정보를 추출하는 데 도움이 될 수 있다. 신호 수준이 이동에 관한 제한된 정보를 제공하는 영역(예를 들어, 긴 동중중합체 영역) 중에서 다중-뉴클레오티드/염기 체류 시간을 사용하여 기공을 가로지르는 염기의 수를 추론할 수 있다. 체류 시간의 가변성을 줄이면 이러한 추론을 보다 정확하게 만들 수 있다.

[0004] 일부 구현예에서 본 발명의 돌연변이체는 막관통 기공을 통해 분석물의 이동을 제어하는 방법 및 막관통 기공을 사용하여 분석물을 특성화하는 방법에서 사용될 때 개선된 정확도를 나타낸다. 분석물 특성화(특히 폴리뉴클레오티드)의 맥락에서, 정확도는 원시 판독 단순 정확도; 즉 막관통 기공을 통한 단일 분자의 단일 통과를 의미하는 것으로 해석된다. 정확도는 서열분석 디바이스(device)의 플랫폼 개선을 추적하는 데 유용한 척도이다. 정확도는 또한 공통 정확도 또는 예를 들어 폴리뉴클레오티드 분석물에서 돌연변이와 같은 특이적인 것을 감지하는 정확도를 지칭할 수 있다. 추가적으로 또는 대안적으로, 정확도는 특정 신뢰 수준을 초과하는 염기의 백분율을 의미하는 것으로 해석되며, 여기서 신뢰 수준은 미리 보정되었다. 일부 구현예에서 본 발명의 돌연변이체는 속도의 변화가 최소이거나 전혀 없이 개선된 정확도를 나타낸다. 일부 구현예에서 정확도는 10% 미만 오류, 5% 미만 오류, 4% 미만 오류, 3% 미만 오류, 2% 미만 오류, 1% 미만 오류, 0.1% 미만 오류를 제공하도록 개선된다. 본 발명자들에 의해 식별된 돌연변이체는 전형적으로 돌연변이 조합, 즉 막관통 기공과 상호작용하는 돌연변이체의 일부에서 하나 이상의 변형을 포함한다. 정확도는 또한 중합체가 효소 제어 하에 기공을 전위하는 속도에

의해 영향을 받을 수 있고 속도는 효소에 제공된 ATP의 농도를 변경함으로써 변경될 수 있다. 본 발명자들은 놀랍게도 효소가 정확도 감소를 초래할 수 있는 동일한 조건 하에 동일한 서열분석 실행 내에서 연속적인 중합체 전위 동안 속도의 변화를 나타낼 수 있다는 것을 인식하였다.

- [0005] 정확도는 나노기공 모양 및 조성, 효소뿐만 아니라 효소와 나노기공 사이의 상호작용과 같은 다수의 요인에 의해 영향을 받을 수 있다. 또한 중합체가 효소 제어 하에 기공을 전위하는 속도에 의해 영향을 받으며 전위 속도는 효소에 제공된 ATP의 농도를 변경함으로써 증가되거나 저하될 수 있다. 본 발명자들은 놀랍게도 동일한 서열분석 조건 하에 동일한 서열분석 실행 내에서 연속적인 중합체 전위 동안 속도 변화가 발생하여, 서열분석 정확도의 감소를 초래할 수 있다는 것을 인식하였다. 다수의 중합체에 대한 서열분석 속도의 변화를 측정하여 정규화된 속도 분포를 획득할 수 있고 본 발명자들은 놀랍게도 일부 변형된 효소가 더 낮은 정규화된 속도 분포를 초래하므로 서열분석 정확도를 증가시킬 수 있다는 것을 인식하였다.
- [0006] 본 발명은 다음을 제공한다:
- [0007] - 변형된 DNA 의존적 ATPase(Dda) 헬리카제, 여기서 헬리카제는 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 위치 중 하나 이상에서 변형 또는 치환을 포함함;
- [0008] - 본 발명의 헬리카제 및 추가적인 폴리뉴클레오티드 결합 모이어티를 포함하는 작제물, 여기서 헬리카제는 폴리뉴클레오티드 결합 모이어티에 부착되고 작제물은 분석물의 이동을 제어하는 능력을 가짐;
- [0009] - 본 발명의 헬리카제 또는 본 발명의 작제물을 암호화하는 서열을 포함하는 폴리뉴클레오티드;
- [0010] - 프로모터에 작동가능하게 연결된 본 발명의 폴리뉴클레오티드를 포함하는 벡터;
- [0011] - 본 발명의 벡터를 포함하는 숙주 세포;
- [0012] - 본 발명의 폴리뉴클레오티드를 발현시키거나, 세포를 본 발명의 벡터로 형질감염시키거나 본 발명의 숙주 세포를 배양하는 단계를 포함하는, 본 발명의 헬리카제 또는 본 발명의 작제물을 제조하는 방법;
- [0013] - 분석물을 본 발명의 헬리카제 또는 본 발명의 작제물과 접촉시켜 분석물의 이동을 제어하는 단계를 포함하는, 분석물의 이동을 제어하는 방법;
- [0014] - 다음을 포함하는, 표적 분석물을 특성화하는 방법:
- [0015] (a) 표적 분석물을 막관통 기공 및 본 발명의 헬리카제 또는 본 발명의 작제물과 접촉시켜 헬리카제 또는 작제물이 기공을 통해 표적 분석물의 이동을 제어하도록 하는 단계; 및
- [0016] (b) 폴리뉴클레오티드가 기공에 대해 이동할 때 하나 이상의 측정을 수행하여 측정이 표적 분석물의 하나 이상의 특성화를 나타내는 것이고 이에 의해 표적 분석물을 특성화하는 단계;
- [0017] - (a) 기공과 (b) 본 발명의 헬리카제 또는 본 발명의 작제물 사이에 복합체를 형성하여 표적 분석물을 특성화하기 위한 센서를 형성하는 단계를 포함하는, 표적 분석물을 특성화하기 위한 센서를 형성하는 방법;
- [0018] - (a) 기공과 (b) 본 발명의 헬리카제 또는 본 발명의 작제물 사이에 복합체를 포함하는, 표적 분석물을 특성화하기 위한 센서;
- [0019] - 기공을 통해 표적 분석물의 이동을 제어하기 위한 본 발명의 헬리카제 또는 본 발명의 작제물의 용도;
- [0020] - 다음을 포함하는, 표적 분석물을 특성화하기 위한 키트:
- [0021] (a) 기공 및 본 발명의 헬리카제 또는 본 발명의 작제물; 또는
- [0022] (b) 본 발명의 헬리카제 또는 본 발명의 작제물 및 하나 이상의 로딩 모이어티;
- [0023] - (a) 복수의 기공 및 (b) 본 발명의 복수의 헬리카제 또는 본 발명의 복수의 작제물을 포함하는, 샘플에서 표적 분석물을 특성화하기 위한 장치;
- [0024] - 다음을 포함하는, 본 발명의 헬리카제를 생산하는 방법:
- [0025] (a) 헬리카제를 제공하는 단계; 및
- [0026] (b) 헬리카제를 변형시켜 본 발명의 헬리카제를 생산하는 단계;
- [0027] - 본 발명의 헬리카제를 추가적인 폴리뉴클레오티드 결합 모이어티에 부착하여 작제물을 생산하는 단계를 포함

하는, 본 발명의 작제물을 생산하는 방법;

- [0028] - 폴리뉴클레오티드에 부착된 일련의 2개 이상의 헬리카제, 여기서 2개 이상의 헬리카제 중 적어도 하나는 본 발명의 헬리카제임; 및
- [0029] - 이동이 DNA 의존적 ATPase(Dda) 헬리카제에 의해 제어될 때 막관통 기공에 대한 표적 분석물의 이동을 개선하는 방법으로서, 여기서 DNA 의존적 ATPase(Dda) 헬리카제는 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하고/하거나 Dda 1993의 아미노산 위치 40에 상응하는 위치 중 하나에서 치환을 포함하도록 변형되어 막관통 기공에 대한 표적 분석물의 이동을 개선함.
- [0030] 본 발명자들은 또한 놀랍게도 분석물이 이를 통과하는/상대적인 속도를 개선하거나 변경하는 신규 막관통 기공 돌연변이를 식별하였으며, 바람직하게는 여기서 분석물의 이동은 폴리뉴클레오티드 결합 단백질의 제어 하에 있다. 본 발명의 하나의 구현예에서 막관통 기공 돌연변이는 분석물이 이를 통과하는/상대적인 속도를 증가시킨다. 또 다른 구현예에서 막관통 기공 돌연변이는 분석물이 이를 통해 통과하는/상대적인 속도를 감소시킨다. 분석물이 기공을 통해 통과하는/상대적인 속도는 분석물이 본 발명의 돌연변이를 포함하지 않는 기공에 대해 이동하는 속도에 비해 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 100%, 150%, 200% 또는 300% 또는 그 이상 증가할 수 있다. 분석물이 기공을 통해 통과하는/상대적인 속도는 분석물이 본 발명의 돌연변이를 포함하지 않는 기공에 대해 이동하는 속도에 비해 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90 % 감소될 수 있다. 본 발명자들을 놀랍게도 기공에 대한 변형으로 야기된 속도 증가 또는 감소와 같은 이러한 속도 변경이 정확도 관독에 최소한 또는 전혀 영향을 미치지 않는다는 것을 발견하였다. 이는 분석물을 기공 및 폴리뉴클레오티드 결합 단백질을 예컨대 본 발명의 헬리카제와 접촉시켜, 폴리뉴클레오티드 결합 단백질이 기공을 통해/상대적인 표적 분석물의 이동을 제어하도록 하는 분석물을 특성화하는 방법에서 특히 유리하다. 하나의 구현예에서, 돌연변이체 기공은 돌연변이를 포함하지 않는 다른 막관통 기공과는 상이한 방식으로 폴리뉴클레오티드 결합 단백질과 상호작용한다. 기공 돌연변이체는 DNA를 기공을 통해 전위함으로써 속도 분포를 변경시켜 속도 분포가 더 밀집되어 돌연변이를 포함하지 않는 다른 막관통 기공과 비교할 때 서열분석 오류의 감소를 초래하도록 할 수 있다. 본 발명의 바람직한 구현예에서 본 발명의 변형된 DNA 의존적 ATPase(Dda) 헬리카제는 본 발명의 막관통 기공을 통한 폴리뉴클레오티드와 같은 분석물의 이동을 제어하는 데 사용된다.
- [0031] 본 발명은 단리된 CsgG 기공 또는 이의 상동체 또는 돌연변이체, 또는 CsgG 기공, 또는 이의 상동체 또는 돌연변이체, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는 단리된 기공 복합체를 제공하며, 여기서 CsgG 기공은 서열번호: 117의 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 적어도 하나의 단량체를 포함한다;
- [0032] 하나의 측면에서, CsgF 펩티드는 CsgG-결합 영역, 및 기공에 수축을 형성하는 영역을 포함한다. 하나의 측면에서 CsgF 펩티드는 CsgF의 C-말단 헤드 도메인이 결여된 절두된 CsgF 펩티드이다. 또 다른 측면에서 CsgF 펩티드는 CsgF의 C-말단 헤드 및 넥 도메인의 일부가 결여된 절두된 CsgF 펩티드이다. 또 다른 측면에서 CsgF CsgF의 C-말단 헤드 및 넥 도메인이 결여된 절두된 CsgF 펩티드이다. CsgG/CsgF 기공은 또한 본원에서 기공 복합체 및 단리된 기공 복합체로 언급된다. 단리된 기공 복합체는 CsgG 기공, 또는 이의 상동체 또는 돌연변이체, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체, 특히 절두된 CsgF 단편, 또는 이의 상동체 또는 돌연변이체를 포함한다. 하나의 구현예에서, 상기 변형된 CsgF 펩티드, 또는 상동체 또는 돌연변이체는 CsgG 기공, 또는 이의 상동체 또는 돌연변이체의 루멘(lumen)에 위치한다. 또 다른 구현예에서, 상기 단리된 기공 복합체는 2개 이상의 채널 수축을 가지며, 하나는 수축 루프에 의해 형성된 CsgG 기공에 의해 위치하거나 제공되고, 또 다른 추가적인 채널 수축 또는 관독 헤드가 변형된 CsgF 펩티드 또는 이의 상동체 또는 돌연변이체에 의해 도입된다. 하나의 구현예에서, 상기 CsgG-기공 또는 CsgG-유사 기공은 야생형 기공이 아니라, 돌연변이체 CsgG 기공이며, 특정 구현예에서 돌연변이는 예를 들어, 상기 채널 수축 루프에 존재한다. 다른 구현예에서 돌연변이는 기공이 폴리뉴클레오티드 결합 단백질과 상호작용하는 영역에서 기공의 상단에 대안적으로 또는 추가적으로 존재한다. 돌연변이는 폴리뉴클레오티드 결합 단백질이 기공과 상호작용하는 방식 및/또는 기공이 폴리뉴클레오티드 결합 단백질과 상호작용하는 방식에 영향을 미칠 수 있다. 또 다른 구현예에서, 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는 단리된 기공 복합체는 직경이 0.5 nm 내지 2.0 nm 범위의 CsgF 채널 수축을 갖는다. 하나의 구현예에서, 기공 복합체는 다음을 포함한다: (i) 첫번째 개구, 베타 배럴을 포함하는 중간 섹션, 두번째 개구, 및 첫번째 개구에서 중간 섹션을 통해 두번째 개구까지 확장하는 루멘을 포함하는 CsgG 기공, 여기서 중간 섹션의 루멘 표면은 CsgG 수축을 정의함; 및 (ii) 복수의 변형된 CsgF 펩티드로, 각각은 CsgF 수축 영역 및 CsgF 결합 영역(또한 본원에서 CsgG-결합 도메인 또는 CsgF 영역으로 언급됨)을 가지며, 여기서 변형된

CsgF 펩티드는 CsgG 기공의 베타 배럴 내에서 CsgF 수축을 형성하고 여기서 CsgG 수축 및 CsgF 수축은 CsgG 기공의 베타 배럴 내에서 동축으로 간격을 두고 있음. CsgG 기공의 루멘 표면은 CsgG 수축을 정의하는 CsgG 단량체의 하나 이상의 루프 영역을 포함할 수 있다. CsgF 수축 영역 및 CsgF 결합 영역은 전형적으로 CsgF 성숙 펩티드의 N-말단 부분에 상응한다. 하나의 구현예에서, 기공 복합체는 CsgA, CsgB 및 CsgE를 제외한다.

[0033] 하나의 구현예는 CsgG 기공 및 변형된 CsgF 펩티드를 포함하는 기공에 관한 것이며, 여기서 변형된 CsgF 펩티드는 CsgG에 결합되고 기공에서 수축을 형성하고 여기서 기공은 기공 및 폴리뉴클레오티드 결합 효소의 상호작용을 변경하도록 돌연변이되고/되거나 상기 기공은 분석물이 기공을 통해 통과하는 속도를 개선하도록 돌연변이된다. 본 발명의 하나의 구현예에서, 분석물이 기공을 통해 통과하는 속도는 증가된다. 본 발명의 또 다른 구현예에서, 분석물이 기공을 통해 통과하는 속도는 감소된다.

[0034] 또 다른 구현예는 단리된 기공 복합체에 관한 것이며 여기서 변형된 CsgF 펩티드 및 CsgG 기공 또는 상기 기공의 단량체, 또는 이의 상동체 또는 돌연변이체는 공유적으로 결합된다. 보다 더 특히, 상기 결합은 서열번호: 117 또는 서열번호: 3의 132, 133, 136, 138, 140, 142, 144, 145, 147, 149, 151, 153, 155, 183, 185, 187, 189, 191, 201, 203, 205, 207 또는 209에 상응하는 위치에서 CsgG 단량체, 또는 이의 상동체에서 시스테인 잔기를 통해 또는 비천연 반응성 또는 광반응성 아미노산을 통해 이루어진다.

[0035] 본 발명은 또한 단리된 막관통 기공 또는 기공 복합체, 또는 본 발명의 단리된 기공 또는 기공 복합체를 포함하는 막성 조성물, 및 막의 구성요소를 제공한다. 특히, 상기 막관통 기공 또는 기공 복합체 또는 막성 조성물은 본 발명의 단리된 기공 또는 기공 복합체, 및 막 또는 절연 층의 구성요소로 이루어진다.

[0036] 본 발명은 또한 다음을 제공한다:

[0037] - 본 발명의 기공 또는 기공 복합체를 포함하는 막;

[0038] - 본 발명의 복수의 막을 포함하는 어레이;

[0039] - (a) 본 발명의 막 또는 본 발명의 어레이, (b) 막(들)을 가로질러 전위를 인가하기 위한 수단 및 (c) 막(들)을 가로질러 전기적 또는 광학적 신호를 검출하기 위한 수단을 포함하는 시스템.

[0040] 본 발명은 또한 적합한 숙주 세포에서 CsgG 기공, 또는 이의 상동체 또는 돌연변이체, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 공동 발현시켜, 생체 내 막관통 기공 복합체 형성을 허용하는 단계를 포함하는, 본 발명의 막관통 기공 복합체를 생산하는 방법을 제공한다.

[0041] 본 발명은 또한 CsgG 단량체, 또는 이의 상동체 또는 돌연변이체를 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체와 접촉시켜, 단리된 기공 복합체의 시험관 내 재구성을 허용하는 단계를 포함하는, 본 발명의 단리된 기공 복합체를 생산하는 방법을 제공한다. 변형된 CsgF 펩티드는 기공의 형성 전 또는 후에 절단되는 아미노산 서열의 적합한 위치에 효소 절단 부위를 포함하는 펩티드일 수 있다.

[0042] 구체적 구현예에서, 상기 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체는 서열번호: 12 또는 서열번호: 14, 또는 이의 상동체 또는 돌연변이체를 포함한다. 특정 구현예에서, 상기 방법의 변형된 CsgF 펩티드는 서열번호: 15 또는 서열번호: 16, 또는 이의 상동체 또는 돌연변이체를 포함한다.

[0043] 본 발명은 또한 다음 단계를 포함하는, 표적 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 방법을 제공한다:

[0044] (i) 표적 분석물을 본 발명의 단리된 기공 또는 단리된 기공 복합체 또는 본 발명의 막관통 기공 복합체와 접촉시켜 표적 분석물이 기공 채널로 이동하도록 하는 단계; 및

[0045] (ii) 분석물이 기공 채널을 통해 이동할 때 하나 이상의 측정을 수행하여 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 단계. 하나의 구현예에서, 상기 분석물은 폴리뉴클레오티드이다. 특히, 폴리뉴클레오티드를 분석물로 사용하는 상기 방법은 대안적으로 (i) 분석물 또는 폴리뉴클레오티드의 길이, (ii) 분석물 또는 폴리뉴클레오티드의 동일성, (iii) 분석물 또는 폴리뉴클레오티드의 서열, (iv) 분석물 또는 폴리뉴클레오티드의 2차 구조 및 (v) 분석물 또는 폴리뉴클레오티드가 변형되는지 여부로부터 선택된 하나 이상의 특성을 결정하는 단계를 포함한다.

[0046] 또 다른 구현예에서, 분석물은 단백질, (폴리)펩티드 또는 펩티드이다. 추가 구현예에서, 상기 분석물은 중합체, 올리고당류, 이당류, 또는 예를 들면 약학적으로 활성인 화합물, 독성 화합물 및 및 오염물질과 같으나 이에 제한되지 않는 작은 유기 또는 무기 화합물이다.

- [0047] 본 발명은 또한 본 발명의 단리된 기공 또는 단리된 기공 복합체 또는 본 발명의 막관통 기공 복합체를 사용하여 폴리뉴클레오티드 또는 (폴리)펩티드를 특성화하는 방법을 제공한다. 특히, 상기 CsgG 기공, 또는 이의 상동체 또는 돌연변이체는 CsgG 기공 채널을 형성하는 6 내지 10개의 CsgG 단량체를 포함한다.
- [0048] 본 발명은 또한 표적 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하기 위해 본 발명의 단리된 기공 또는 단리된 기공 복합체 또는 본 발명의 막관통 기공 복합체의 용도를 제공한다. 더욱이, 본 발명은 또한 (a) 상기 단리된 기공 또는 기공 복합체 및 (b) 막의 구성요소를 포함하는 표적 분석물을 특성화하기 위한 키트에 관한 것이다.
- [0049] 본 발명은 또한 다음을 제공한다:
- [0050] - 표적 분석물을 본 발명의 단리된 기공 또는 단리된 기공 복합체 또는 본 발명의 막관통 기공 복합체와 접촉시켜, 표적 분석물이 기공 복합체에 대해 또는 내로 이동하도록 하는 단계를 포함하는 표적 분석물이 기공을 통해 통과하는 속도를 변경하는 방법;
- [0051] - (a) 본 발명의 단리된 기공 또는 단리된 기공 복합체 및 (b) 막의 구성요소 및 (c) 폴리뉴클레오티드 결합 단백질 중 하나 또는 둘 다를 포함하는 표적 분석물을 특성화하기 위한 키트;
- [0052] - 다음을 포함하는, 표적 분석물을 특성화하는 방법:
- [0053] (a) 표적 분석물을 본 발명의 단리된 기공 또는 단리된 기공 복합체 및 본 발명의 DNA 의존적 ATPase(Dda) 헬리카제 또는 본 발명의 헬리카제 작제물과 접촉시켜 헬리카제 또는 작제물이 기공 또는 기공 복합체를 통해 표적 분석물의 이동을 제어하도록 하는 단계; 및
- [0054] (b) 폴리뉴클레오티드가 기공 또는 기공 복합체에 대해 이동할 때 하나 이상의 측정을 수행하여 측정이 표적 분석물의 하나 이상의 특성을 나타내는 것이고 이에 의해 표적 분석물을 특성화하는 단계;
- [0055] - (a) 본 발명의 DNA 의존적 ATPase(Dda) 헬리카제 또는 본 발명의 헬리카제 작제물 (b) 본 발명의 단리된 CsgG 기공 또는 단리된 기공 복합체를 포함하는 표적 분석물을 특성화하기 위한 키트;
- [0056] - 시험관 내 막에 삽입된 본 발명의 기공 또는 기공 복합체를 포함하는 장치;
- [0057] - (i) 본 발명의 단리된 기공 또는 단리된 기공 복합체를 수득하는 단계 및 (ii) 단리된 기공 또는 단리된 기공 복합체를 시험관 내 막과 접촉시켜 기공이 시험관 내 막에 삽입되도록 하는 단계를 포함하는 방법에 의해 생산된 장치.

발명을 실시하기 위한 구체적인 내용

- [0058] 본원에 인용된 모든 간행물, 특허 및 특허 출원은 상기된 하기든 이들의 전문이 본원에 참조로 포함된다. 본 명세서에 언급된 모든 간행물, 특허, 및 특허 출원은 각 개별 간행물, 특허, 또는 특허 출원이 참조로 혼입되도록 구체적으로 및 개별적으로 지시되는 것과 동일한 정도로 본원에 참조로 포함된다. 참조로 포함된 간행물 및 특허 또는 특허 출원이 명세서에 함유된 개시내용과 모순되는 경우, 명세서는 임의의 이러한 모순되는 자료를 대체하고/하거나 우선하도록 의도된다.
- [0059] 본 발명은 특정 구현예에 대해 특정 도면을 참조하여 기재될 것이지만 본 발명은 이에 제한되지 않고 청구범위에 의해서만 제한된다. 청구범위에서 임의의 참조 기호는 범위를 제한하는 것으로 해석되어서는 안 된다. 물론, 반드시 모든 측면 또는 이점이 본 발명의 임의의 특정 구현예에 따라 달성될 수 있지 않다는 것이 이해되어야 한다. 따라서, 예를 들어 당업자는 본 발명이 본원에 교시되거나 제안될 수 있는 바와 같이 다른 측면 또는 이점을 반드시 달성하지 않고 본원에 교시된 바와 같은 하나의 이점 또는 이점 그룹을 달성하거나 최적화하는 방식으로 구현되거나 수행될 수 있다는 것을 인식할 것이다.
- [0060] 본 발명은 이의 특징 및 이점과 함께 조직화 및 작동 방법 둘 다에 대해 첨부 도면과 함께 읽을 때 다음 상세한 설명을 참조하여 가장 잘 이해될 수 있다. 본 발명의 측면 및 이점은 이하에 기재된 구현예(들)로부터 명백하고 이를 참조하여 설명될 것이다. 본 명세서 전반에 걸쳐 "하나의 구현예" 또는 "구현예"에 대한 언급은 구현예와 관련하여 기재된 특정 특징, 구조 또는 특성이 본 발명의 적어도 하나의 구현예에 포함된다는 것을 의미한다. 따라서, 본 명세서 전반에 걸쳐 다양한 곳에서 "하나의 구현예에서" 또는 "구현예에서"라는 어구의 출현은 반드시 모두 동일한 구현예를 지칭하는 것은 아니지만, 그럴 수 있다. 유사하게, 본 발명의 예시적인 구현예의 설명에서, 본 발명의 다양한 특징은 때때로 개시내용을 간소화하고 다양한 독창적인 측면 중 하나 이상의 이해를 돕

기 위한 목적으로 단일 구현예, 또는 이의 설명에서 함께 그룹화된다는 것이 이해되어야 한다. 그러나 본 개시 내용의 이러한 방법은 청구된 발명이 각 청구항에 명시적으로 언급된 것보다 더 많은 특징을 필요로 한다는 의도를 반영하는 것으로 해석되어서는 안 된다. 오히려, 다음 청구항을 반영하듯이, 독창적인 측면은 단일 전술한 개시된 구현예의 모든 특징보다 적게 있다.

[0061] 본 명세서 및 첨부된 청구범위에 사용된 것 외에도, 단수 형태는 내용상 달리 분명히 명시되지 않는 한 복수 지시대상을 포함한다. 따라서, 예를 들어, "폴리뉴클레오타이드"에 대한 언급은 2개 이상의 폴리뉴클레오타이드를 포함하고, "폴리뉴클레오타이드 결합 단백질"에 대한 언급은 2개 이상의 이러한 단백질을 포함하고, "헬리카제"에 대한 언급은 2개 이상의 헬리카제를 포함하고, "단량체"에 대한 언급은 2개 이상의 단량체를 지칭하고, "기공"에 대한 언급은 2개 이상의 기공을 포함하는 것 등이다.

[0062] 본원의 모든 논의에서, 아미노산에 대한 표준 1글자 코드가 사용된다. 이는 다음과 같다: 알라닌(A), 아르기닌(R), 아스파라긴(N), 아스파르트산(D), 시스테인(C), 글루탐산(E), 글루타민(Q), 글리신(G), 히스티딘(H), 이소류신(I), 류신(L), 리신(K), 메티오닌(M), 페닐알라닌(F), 프롤린(P), 세린(S), 트레오닌(T), 트립토판(W), 티로신(Y) 및 발린(V). 표준 치환 표기법이 또한 사용되며, 즉 Q42R은 위치 42의 Q가 R로 대체된다는 것을 의미한다.

[0063] 본원의 문단에서 특정 위치에 있는 상이한 아미노산이 / 기호로 구분되는 경우, / 기호는 "또는"을 의미한다. 예를 들면, Q87R/K는 Q87R 또는 Q87K를 의미한다. 본원의 문단에서 상이한 위치가 / 기호로 구분되는 경우, / 기호는 "및"을 의미하므로 Y51/N55는 Y51 및 N55이다.

[0064] 정의

[0065] 다음 용어 또는 정의는 오로지 본 발명의 이해를 돕기 위해 제공된다. 달리 구체적으로 본원에서 정의되지 않는 한, 본원에 사용되는 모든 용어는 본 발명의 기술 분야의 숙련자에게 동일한 의미를 갖는다. 기술분야의 정의 및 용어에 대해 실무자들은 특히 Sambrook 등, Molecular Cloning: A Laboratory Manual, 4th ed., Cold Spring Harbor Press, Plainsview, New York(2012); 및 Ausubel 등, Current Protocols in Molecular Biology(Supplement 114), John Wiley & Sons, New York(2016)에 주의를 기울인다. 본원에 제공된 정의는 당업자에 의해 이해되는 것보다 적은 범위를 갖는 것으로 해석되어서는 안 된다.

[0066] 본원에 사용된 바와 같은 "약"은 양, 시간적 지속기간 등과 같은 측정가능한 값을 언급할 때 명시된 값에서 $\pm 20\%$ 또는 $\pm 10\%$, 보다 바람직하게는 $\pm 5\%$, 보다 더 바람직하게는 $\pm 1\%$, 훨씬 더 바람직하게는 $\pm 0.1\%$ 의 변동을 포함하는 것을 의미하며, 이러한 변동은 개시된 방법을 수행하는 데 적절하다.

[0067] 본원에 사용된 바와 같은 "뉴클레오타이드 서열", "DNA 서열" 또는 "핵산 분자(들)"는 임의의 길이의 뉴클레오타이드의 중합체 형태, 리보뉴클레오타이드 또는 데옥시리보뉴클레오타이드를 지칭한다. 이 용어는 분자의 1차 구조만을 지칭한다. 따라서, 이 용어는 이중 가닥 및 단일 가닥 DNA, 및 RNA를 포함한다. 본원에 사용된 바와 같은 용어 "핵산"은 뉴클레오타이드의 단일 또는 이중 가닥 공유 결합된 서열이며 여기서 각 뉴클레오타이드 상의 3' 및 5' 단부는 포스포디에스테르 결합으로 연결된다. 폴리뉴클레오타이드는 데옥시리보뉴클레오타이드 염기 또는 리보뉴클레오타이드 염기로 구성될 수 있다. 핵산은 시험관 내에서 합성적으로 제조되거나 천연 공급원으로부터 단리될 수 있다. 핵산은 변형된 DNA 또는 RNA, 예를 들어 메틸화된 DNA 또는 RNA, 또는 번역 후 변형, 예를 들어 7-메틸гу아노신으로 5'-캡핑, 절단 및 폴리아데닐화와 같은 3'-처리, 및 스플라이싱을 겪은 RNA를 추가로 포함할 수 있다. 핵산은 또한 합성 핵산(XNA), 예컨대 핵시톨 핵산(HNA), 사이클로핵산(CeNA), 트레오스 핵산(TNA), 글리세롤 핵산(GNA), 잠금 핵산(LNA) 및 펩티드 핵산(PNA)을 포함할 수 있다. 본원에서 "폴리뉴클레오타이드"로도 언급되는 핵산의 크기는 전형적으로 이중 가닥 폴리뉴클레오타이드의 경우 염기 쌍(bp) 수로 표현되거나, 단일 가닥 폴리뉴클레오타이드의 경우 뉴클레오타이드(nt) 수로 표현된다. 1000개의 bp 또는 nt는 킬로베이스(kb)와 같다. 약 40개 미만의 뉴클레오타이드 길이의 폴리뉴클레오타이드는 전형적으로 "올리고뉴클레오타이드"라고 하고 중합효소 연쇄 반응(PCR)을 통해서와 같이 DNA 조작에 사용하기 위한 프라이머를 포함할 수 있다.

[0068] 여기에 사용된 바와 같은 "유전자"는 유전자의 프로모터 영역 뿐만 아니라 코딩 서열을 둘 다 포함한다. 이는 프로모터 서열에 작동가능하게 연결된 게놈 서열(가능한 인트론 포함) 뿐만 아니라 스플라이스된 메신저에서 유래된 cDNA를 둘 다 지칭한다.

[0069] "코딩 서열"은 적절한 조절 서열의 제어 하에 배치될 때 mRNA로 전사되고/되거나 폴리펩티드로 번역되는 뉴클레오타이드 서열이다. 코딩 서열의 경계는 5'-말단에 있는 번역 시작 코돈 및 3'-말단에 있는 번역 정지 코돈에 의해 결정된다. 코딩 서열은 mRNA, cDNA, 재조합 뉴클레오타이드 서열 또는 게놈 DNA를 포함할 수 있으나 이에 제

한되지 않는 반면, 인트론 또한 특정 상황 하에 존재할 수 있다.

[0070] 본 개시내용의 맥락에서 용어 "아미노산"은 광범위한 의미로 사용되며 각 아미노산에 특이적인 측쇄(예를 들어, R 기)와 함께 아민(NH₂) 및 카르복실(COOH) 작용 기를 함유하는 유기 화합물을 포함하는 것을 의미한다. 일부 구현예에서, 아미노산은 자연 발생 L α-아미노산 또는 잔기를 지칭한다. 자연 발생 아미노산에 대해 통상적으로 사용되는 1 및 3 글자 약어가 본원에 사용된다: A=Ala; C=Cys; D=Asp; E=Glu; F=Phe; G=Gly; H=His; I=Ile; K=Lys; L=Leu; M=Met; N=Asn; P=Pro; Q=Gln; R=Arg; S=Ser; T=Thr; V=Val; W=Trp; 및 Y=Tyr(Lehninger, A. L., (1975) Biochemistry, 2d ed., pp. 71-92, Worth Publishers, New York). 일반적인 용어 "아미노산"은 D-아미노산, 레트로-인버소(retro-inverso) 아미노산 뿐만 아니라 아미노산 유사체와 같은 화학적으로 변형된 아미노산, 노르류신과 같은 일반적으로 단백질에 혼입되지 않는 자연 발생 아미노산, 및 β-아미노산과 같은 아미노산의 특성인 것으로 당업계에 알려진 특성을 갖는 화학적으로 합성된 화합물을 추가로 포함한다. 예를 들어, 천연 Phe 또는 Pro와 같이 펩티드 화합물의 동일한 형태적 제한을 허용하는 페닐알라닌 또는 프롤린의 유사체 또는 모방체는 아미노산의 정의 내에 포함된다. 이러한 유사체 및 모방체는 본원에서 각각의 아미노산의 "기능적 등가물"로 언급된다. 아미노산의 다른 예는 Roberts and Vellaccio, The Peptides: Analysis, Synthesis, Biology, Gross and Meiehofer, eds., Vol. 5 p. 341, Academic Press, Inc., N.Y. 1983에 의해 나열되어 있으며, 이는 본원에 참조로 포함된다.

[0071] 용어 "단백질", "폴리펩티드", 및 "펩티드"는 아미노산 잔기의 중합체 및 이의 변이체 및 합성 유사체를 지칭하기 위해 본원에서 추가로 상호교환가능하게 사용된다. 따라서, 이러한 용어는 하나 이상의 아미노산 잔기가 상응하는 자연 발생 아미노산의 화학적 유사체와 같은 합성 비자연 발생 아미노산인 아미노산 중합체, 뿐만 아니라 자연 발생 아미노산 중합체에 적용된다. 폴리펩티드는 또한 글리코실화, 단백질분해 절단, 지질화, 신호 펩티드 절단, 프로펩티드 절단, 인산화 등을 포함할 수 있으나 이에 제한되지 않는 성숙 또는 번역 후 변형 과정을 거칠 수 있다. "재조합 폴리펩티드"란 재조합 기술을 사용하여, 예를 들어, 재조합 또는 합성 폴리뉴클레오타이드의 발현을 통해 제조된 폴리펩티드를 의미한다. 키메라 폴리펩티드 또는 이의 생물학적으로 활성인 부분이 재조합적으로 생산되는 경우, 또한 바람직하게는 배양 배지가 실질적으로 없으며, 예를 들어, 배양 배지는 단백질 제제 부피의 약 20 % 미만, 보다 바람직하게는 약 10 % 미만, 가장 바람직하게는 약 5 % 미만을 나타낸다. "단리된"이란 천연 상태에서 정상적으로 수반되는 구성요소가 실질적으로 또는 본질적으로 없는 재료를 의미한다. 예를 들어, 본원에 사용된 바와 같은 "단리된 폴리펩티드"는 자연 발생 상태에서 옆에 있는 분자에서 정지된 폴리펩티드, 예를 들어, 상기 폴리펩티드에 인접한 생산 숙주에 존재하는 분자에서 제거된 단백질 복합체 또는 CsgF 펩티드를 지칭한다. 단리된 CsgF 펩티드(임의적으로 절두된 CsgF 펩티드)는 아미노산 화학적 합성에 의해 생성될 수 있거나 재조합 생산에 의해 생성될 수 있다. 단리된 복합체는 복합체의 구성요소, 예를 들어 CsgG 기공 및 CsgF 펩티드(들)의 정제 후에 시험관 내 재구성에 의해 생성될 수 있거나, 재조합 공동 발현에 의해 생성될 수 있다.

[0072] "오솔로그(Orthologue)" 및 "패럴로그(paralogue)"는 유전자의 조상 관계를 설명하는 데 사용되는 진화적 개념을 포괄한다. 패럴로그는 조상 유전자의 복제를 통해 기원된 동일한 종 내의 유전자이며; 오솔로그는 종 분화를 통해 기원된 상이한 유기체의 유전자이고, 또한 공통 조상 유전자로부터 유래된다.

[0073] 단백질의 "상동체", "상동체들"은 해당 비변형 또는 야생형 단백질에 비해 아미노산 치환, 결실 및/또는 삽입을 갖고 유래된 비변형된 단백질과 유사한 생물학적 및 기능적 활성을 갖는 펩티드, 올리고펩티드, 폴리펩티드, 단백질 및 효소를 포괄한다. 본원에 사용된 바와 같은 용어 "아미노산 동일성"은 서열이 비교 창에 걸쳐 아미노산 별 기준으로 동일한 정도를 지칭한다. 따라서, "서열 동일성 백분율"은 비교 창에 걸쳐 2개의 최적으로 정렬된 서열을 비교하고, 동일한 아미노산 잔기(예를 들어, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys 및 Met)가 두 서열에서 발생하여 일치된 위치 수를 산출하는 위치 수를 결정하고, 일치된 위치 수를 비교 창의 총 위치 수(즉, 창 크기)로 나누고, 결과에 100을 곱하여 서열 동일성 백분율을 산출함으로써 계산된다.

[0074] 용어 "CsgG 기공"은 다중 CsgG 단량체를 포함하는 기공을 정의한다. 각 CsgG 단량체는 콜라이(*E. coli*)의 야생형 단량체(서열번호: 3), 이. 콜라이 CsgG의 야생형 상동체, 예컨대 예를 들어, 서열번호: 68 내지 88에 제시된 아미노산 서열 중 임의의 하나를 갖는 단량체. 또는 이의 임의의 변이체(예를 들어 서열번호: 3, 117 및 68 내지 88 중 임의의 하나의 변이체)일 수 있다. 변이체 CsgG 단량체는 또한 변형된 CsgG 단량체 또는 돌연변이체 CsgG 단량체로 언급될 수 있다. 변이체에서 변형, 또는 돌연변이는 본원에 개시된 변형 중 임의의 하나 이상, 또는 상기 변형의 조합을 포함하나 이에 제한되지 않는다.

- [0075] 본 발명의 모든 측면 및 구현예에 대해, CsgG 상동체는 서열번호: 117 또는 서열번호: 3에 제시된 바와 같은 야생형 이. 콜라이 CsgG와 적어도 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95% 또는 99% 완전한 서열 동일성을 갖는 폴리펩티드로 언급된다. CsgG 상동체는 또한 CsgG-유사 단백질의 특성인 PFAM 도메인 PF03783을 함유하는 폴리펩티드로 언급된다. 현재 알려진 CsgG 상동체 및 CsgG 아키텍처의 목록은 <http://pfam.xfam.org/family/PF03783>에서 찾을 수 있다. 마찬가지로, CsgG 상동 폴리뉴클레오티드는 서열번호: 1에 제시된 바와 같은 야생형 이. 콜라이 CsgG와 적어도 40%, 50%, 60%, 70%, 80%, 90%, 95% 또는 99% 완전한 서열 동일성을 갖는 폴리뉴클레오티드를 포함할 수 있다. 서열번호:3에 제시된 CsgG의 상동체의 예는 서열번호: 68 내지 88에 제시된 서열을 갖는다.
- [0076] 용어 "변형된 CsgF 펩티드" 또는 "CsgF 펩티드"는 C-말단 단부에서 절두된 CsgF 펩티드(예를 들어 N-말단 단편임)를 정의하고/하거나 절단 부위를 포함하도록 변형된다. CsgF 펩티드는 야생형 이. 콜라이 CsgF(서열번호: 5 또는 서열번호: 6)의 단편, 또는 이. 콜라이 CsgF의 야생형 상동체의 단편, 예컨대 예를 들어, 서열번호: 17 내지 36에 제시된 아미노산 서열 중 임의의 하나를 포함하는 펩티드. 또는 이의 임의의 변이체(예를 들어 절단 부위를 포함하도록 변형된 것)일 수 있다.
- [0077] 본 발명의 모든 측면 및 구현예에 대해, CsgF 상동체는 서열번호: 6에 제시된 바와 같은 야생형 이. 콜라이 CsgF와 적어도 40%, 50%, 60%, 70%, 80%, 90%, 95% 또는 99% 완전한 서열 동일성을 갖는 폴리펩티드로 언급된다. 일부 구현예에서, CsgF 상동체는 또한 CsgF-유사 단백질의 특성인 PFAM 도메인 PF10614를 함유하는 폴리펩티드로 언급된다. 현재 알려진 CsgF 상동체 및 CsgF 아키텍처의 목록은 <http://pfam.xfam.org/family/PF10614>에서 찾을 수 있다. 마찬가지로, CsgF 상동 폴리뉴클레오티드는 서열번호: 4에 제시된 바와 같은 야생형 이. 콜라이 CsgF와 적어도 40%, 50%, 60%, 70%, 80%, 90%, 95% 또는 99% 완전한 서열 동일성을 갖는 폴리뉴클레오티드를 포함할 수 있다. 서열번호:6에 제시된 CsgF 상동체의 절두된 영역의 예는 서열번호:17 내지 36에 제시된 서열을 갖는다.
- [0078] 용어 "CsgF 성숙 펩티드의 N-말단 부분"은 CsgF 성숙 펩티드(신호 서열 없음)의 N-말단에서 시작하여 처음 60개, 50개, 또는 40개 아미노산 잔기에 상응하는 아미노산 서열을 갖는 펩티드를 지칭한다. CsgF 성숙 펩티드는 야생형 또는 돌연변이체(예를 들어, 하나 이상의 돌연변이 있음)일 수 있다.
- [0079] 서열 동일성은 또한 전장 폴리뉴클레오티드 또는 폴리펩티드의 단편 또는 일부에 대한 것일 수 있다. 따라서, 서열은 전장 참조 서열과 50 % 전체 서열 동일성만 가질 수 있지만, 특정 영역, 도메인 또는 서브유닛의 서열은 참조 서열과 80 %, 90 %, 또는 최대한 99 % 서열 동일성을 공유할 수 있다. 각각 CsgG 상동체의 경우 서열번호: 1 또는 CsgF 상동체의 경우 서열번호:4의 핵산 서열에 대한 상동성은 단순히 서열 동일성에 제한되지 않는다. 많은 핵산 서열은 길보기에 낮은 서열 동일성을 가짐에도 불구하고 서로 생물학적으로 유의한 상동성을 입증할 수 있다. 상동 핵산 서열은 낮은 엄격성 조건 하에 서로 혼성화할 것으로 간주된다(M.R. Green, J. Sambrook, 2012, Molecular Cloning: A Laboratory Manual, Fourth Edition, Books 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
- [0080] 용어 "야생형"은 자연 발생 공급원으로부터 단리된 유전자 또는 유전자 산물을 지칭한다. 야생형 유전자는 집단에서 가장 빈번하게 관찰되는 것이며 따라서 임의로 유전자의 "정상" 또는 "야생형" 형태로 설계된다. 대조적으로, 용어 "변형된", "돌연변이체" 또는 "변이체"는 야생형 유전자 또는 유전자 산물과 비교할 때 서열의 변형(예를 들어, 치환, 절두, 또는 삽입), 번역 후 변형 및/또는 기능적 특성(예를 들어, 변형된 특성)을 나타내는 유전자 또는 유전자 산물을 지칭한다. 자연 발생 돌연변이체는 단리될 수 있다는 점에 유의하며; 이들은 야생형 유전자 또는 유전자 산물과 비교할 때 변경된 특성을 갖는다는 사실에 의해 식별된다. 자연 발생 아미노산을 도입하거나 치환하는 방법은 당업계에 잘 알려져 있다. 예를 들면, 메티오닌(M)은 돌연변이체 단량체를 암호화하는 폴리뉴클레오티드의 관련 위치에서 메티오닌에 대한 코돈(ATG)을 아르기닌에 대한 코돈(CGT)으로 대체함으로써 아르기닌(R)으로 치환될 수 있다. 비자연 발생 아미노산을 도입하거나 치환하는 방법이 또한 당업계에 잘 알려져 있다. 예를 들면, 비자연 발생 아미노산은 돌연변이체 단량체를 발현하는 데 사용되는 IVTT 시스템에 합성 아미노아실-tRNA를 포함함으로써 도입될 수 있다. 대안적으로, 이들은 특이적 아미노산의 합성(즉 비자연 발생) 유사체의 존재 하에 특이적 아미노산에 대한 영양요구성이 있는 이. 콜라이에서 돌연변이체 단량체를 발현함으로써 도입될 수 있다. 이들은 또한 돌연변이체 단량체가 부분적 펩티드 합성을 사용하여 생산되는 경우 네이키드(naked) 결찰에 의해 생산될 수 있다. 보존적 치환은 아미노산을 유사한 화학적 구조, 유사한 화학적 특성 또는 유사한 측쇄 부피의 다른 아미노산으로 대체한다. 도입된 아미노산은 이들을 대체하는 아미노산에 대해 유사한 극성, 친수성, 소수성, 염기성, 산성, 중성 또는 전하를 가질 수 있다. 대안적으로, 보존적 치환은 기존의 방향족 또는 지방족 아미노산 대신 방향족 또는 지방족인 또 다른 아미노산을 도입할 수 있다. 보존적 아미노산

변화는 당업계에 잘 알려져 있고 아래 표 1에 정의된 바와 같은 20개 주요 아미노산의 특성에 따라 선택될 수 있다. 아미노산이 유사한 극성을 갖는 경우, 이들은 또한 표 2에서 아미노산 측쇄에 대한 소수성 척도(hydrophathy scale)를 참조하여 결정될 수 있다.

[0081] 표 1 - 아미노산의 화학적 특성

Ala	지방족, 소수성, 중성	Met	소수성, 중성
Cys	극성, 소수성, 중성	Asn	극성, 친수성, 중성
Asp	극성, 친수성, 전하(-)	Pro	소수성, 중성
Glu	극성, 친수성, 전하(-)	Gln	극성, 친수성, 중성
Phe	방향족, 소수성, 중성	Arg	극성, 친수성, 전하(+)
Gly	지방족, 중성	Ser	극성, 친수성, 중성
His	방향족, 극성, 친수성, 전하(+)	Thr	극성, 친수성, 중성
Ile	지방족, 소수성, 중성	Val	지방족, 소수성, 중성
Lys	극성, 친수성, 전하(+)	Trp	방향족, 소수성, 중성
Leu	지방족, 소수성, 중성	Tyr	방향족, 극성, 소수성

[0082]

[0083] 표 2 - 소수성 척도

측쇄	소수성
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Trp	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5
Gln	-3.5
Asp	-3.5
Asn	-3.5
Lys	-3.9
Arg	-4.5

[0084]

[0085] 돌연변이체 또는 변형된 단백질, 단량체 또는 펩티드는 또한 임의의 방식으로 임의의 부위에서 화학적으로 변형될 수 있다. 돌연변이체 또는 변형된 단량체 또는 펩티드는 바람직하게는 분자를 하나 이상의 시스테인에 부착시키거나(시스테인 결합), 분자를 하나 이상의 리신에 부착시키거나, 분자를 하나 이상의 비천연 아미노산에 부착시키거나, 에피토프의 효소를 변형시키거나 말단을 변형시켜 화학적으로 변형된다. 이러한 변형을 수행하는 데 적합한 방법은 당업계에 잘 알려져 있다. 변형된 단백질, 단량체 또는 펩티드의 돌연변이체는 임의의 분자에 부착하여 화학적으로 변형될 수 있다. 예를 들면, 변형된 단백질, 단량체 또는 펩티드의 돌연변이체는 염료 또는 형광단을 부착하여 화학적으로 변형될 수 있다. 일부 구현예에서, 돌연변이체 또는 변형된 단량체 또는 펩티드는 단량체 또는 펩티드를 포함하는 기공과 표적 뉴클레오타이드 또는 표적 폴리뉴클레오타이드 서열 사이의 상호작용을 용이하게 하는 분자 어댑터로 화학적으로 변형된다. 분자 어댑터는 바람직하게는 환식 분자, 사이클로텍스트린, 혼성화할 수 있는 중, DNA 결합체 또는 인터칼레이터(interchelator), 펩티드 또는 펩티드 유사체, 합성 중합체, 방향족 평면 분자, 작은 양으로 하전된 분자 또는 수소 결합할 수 있는 소분자이다.

- [0086] 어댑터의 존재는 기공 및 뉴클레오티드 또는 폴리뉴클레오티드 서열의 호스트-게스트 화학을 개선시키고 이에 의해 돌연변이체 단량체로부터 형성된 기공의 서열분석 능력을 개선시킨다. 호스트-게스트 화학의 원리는 당업계에 잘 알려져 있다. 어댑터는 뉴클레오티드 또는 폴리뉴클레오티드 서열과의 상호작용을 개선하는 기공의 물리적 또는 화학적 특성에 영향을 미친다. 어댑터는 기공의 배열 또는 채널의 전하를 변경하거나 뉴클레오티드 또는 폴리뉴클레오티드 서열과 특이적으로 상호작용하거나 이에 결합하여 기공과의 상호작용을 용이하게 할 수 있다. 따라서 개시내용에 제공된 바와 같은 변형된 CsgF 펩티드는 효소 또는 단백질에 결합되어 상기 단백질 또는 효소를 기공에 더 근접하게 제공할 수 있으며, 이는 변형된 CsgF 펩티드를 포함하는 기공 복합체의 특정 적용을 용이하게 한다.
- [0087] 이 맥락에서, 단백질은 또한 융합 단백질일 수 있으며, 특히 예를 들어, 재조합 DNA 기술로 제조된 유전적 융합을 언급한다. 단백질은 또한 본원에서 사용된 바와 같이 접합되거나, 또는 "접합"될 수 있으며, 특히 안정한 공유 결합을 초래하는 화학적 및/또는 효소적 접합을 지칭한다.
- [0088] 단백질은 여러 폴리펩티드 또는 단백질 단량체가 서로 결합하거나 상호작용할 때 단백질 복합체를 형성할 수 있다. "결합"은 직접적이든 간접적이든 임의의 상호작용을 의미한다. 직접 상호작용은 예를 들면 공유 결합 또는 커플링을 통한 결합 파트너 사이의 접촉을 의미한다. 간접 상호작용은 상호작용 파트너가 2개 초과 화합물 복합체에서 상호작용하는 임의의 상호작용을 의미한다. 상호작용은 하나 이상의 가교 분자의 도움으로 완전히 간접적일 수 있거나, 파트너 사이에 여전히 직접 접촉이 있는 경우 부분적으로 간접적일 수 있으며, 이는 하나 이상의 화합물의 추가적인 상호작용에 의해 안정화된다. 본 개시내용에서 언급된 바와 같은 "복합체"는 상이한 기능을 가질 수 있는 2개 이상의 회합된 단백질 그룹으로 정의된다. 단백질 복합체의 상이한 폴리펩티드(polyptide) 사이의 회합은 소수성 또는 이온성 힘과 같은 비공유 상호작용을 통해 이루어질 수 있거나, 또한 이황화물 가교, 또는 펩티드 결합과 같은 공유 결합 또는 커플링일 수 있다. 공유 "결합" 또는 "커플링"은 본원에서 상호교환가능하게 사용되고, 또한 각각 시스테인 사이 또는 (광)반응성 아미노산 사이의 생물접합을 언급하는 "시스테인 커플링" 또는 "반응성 또는 광반응성 아미노산 커플링"을 수반할 수 있으며, 이는 안정한 복합체를 형성하는 화학적 공유 결합이다. 광반응성 아미노산의 예는 아지도호모알라닌, 호모프로파르길글리신, 호모알렐글리신, p-아세틸-Phe, p-아지도-Phe, p-프로파르길옥시-Phe 및 p-벤조일-Phe를 포함한다(Wang 등 2012, in Protein Engineering, DOI: 10.5772/28719; Chin 등 2002, Proc. Nat. Acad. Sci. USA 99(17); 11020-24).
- [0089] "생물학적 기공"은 막의 한쪽에서 다른 쪽으로 분자 및 이온의 전위를 허용하는 채널 또는 구멍을 정의하는 막관통 단백질 구조이다. 기공을 통한 이온 종의 전위는 기공의 어느 한쪽에 인가되는 전기적 전위차에 의해 구동될 수 있다. "나노기공"은 분자 또는 이온이 통과하는 채널의 최소 직경이 대략 나노미터(10-9 나노미터)인 생물학적 기공이다. 일부 구현예에서, 생물학적 기공은 막관통 단백질 기공일 수 있다. 생물학적 기공의 막관통 단백질 구조는 자연에서 단량체 또는 올리고머일 수 있다. 전형적으로, 기공은 중심 축 주위에 배열된 복수의 폴리펩티드 서브유닛을 포함하며 이에 의해 나노기공이 상주하는 막에 실질적으로 수직으로 확장되는 단백질이 늘어선 채널을 형성한다. 폴리펩티드 서브유닛의 수는 제한되지 않는다. 전형적으로, 서브유닛의 수는 5 내지 최대 30개이며, 적합하게는 서브유닛의 수는 6 내지 10개이다. 대안적으로, 서브유닛의 수는 펩티드글리신 또는 관련 대형 막 기공의 경우와 같이 정의되지 않는다. 단백질이 늘어선 채널을 형성하는 나노기공 내의 단백질 서브유닛의 부분은 전형적으로 하나 이상의 막관통 β -배럴, 및/또는 α -나선 섹션을 포함할 수 있는 2차 구조적 모티프를 포함한다.
- [0090] 본원에서 상호교환가능하게 사용된 바와 같은 용어 "기공", "기공 복합체", 또는 "복합 기공"은 올리고머 기공을 지칭하며, 여기서 예를 들면 적어도 CsgG 단량체(예를 들어, 하나 이상의 CsgG 단량체 예컨대 2개 이상의 CsgG 단량체, 3개 이상의 CsgG 단량체 포함) 또는 CsgG 기공(CsgG 단량체로 구성됨), 및 CsgF 펩티드(예를 들어, 변형된 또는 절두된 CsgF 펩티드)는 복합체로 회합되어 함께 기공 또는 나노기공을 형성한다. 본 개시내용의 기공 복합체는 생물학적 기공의 특징을 가지고 있으며, 즉 전형적인 막관통 단백질 구조를 갖는다. 기공 복합체가 막 구성요소, 막, 세포, 또는 절연 층을 갖는 환경에서 제공될 때, 기공 복합체는 막 또는 절연 층에 삽입되고, "막관통 기공 복합체"를 형성할 것이다.
- [0091] 본 개시내용의 기공, 기공 복합체, 막관통 기공 또는 막관통 기공 복합체는 분석물 특성화에 적합하다. 일부 구현예에서, 본원에 기재된 기공, 기공 복합체, 막관통 기공 또는 막관통 기공 복합체는 예를 들어, 높은 감도로 상이한 뉴클레오티드 사이를 구별할 수 있기 때문에 폴리뉴클레오티드 서열을 서열분석하는 데 사용될 수 있다. 기공 또는 기공 복합체는 단리되거나, 실질적으로 단리되거나, 정제되거나 실질적으로 정제될 수 있다. 기공 또는 기공 복합체는 지질 또는 다른 기공, 또는 정상적으로 천연 상태에서 회합되는 다른 단백질 예를 들어, CsgE, CsgA CsgB와 같은 임의의 다른 구성요소가 완전히 없는 경우, 또는 막성 구획에서 충분히 풍부화된 경우

"단리"되거나 정제된다. 기공 또는 기공 복합체는 의도된 용도를 방해하지 않을 담체 또는 희석제와 혼합되는 경우 실질적으로 단리된다. 예를 들면, 기공 또는 기공 복합체는 트리블록 공중합체, 지질 또는 다른 기공과 같은 다른 구성요소의 10% 미만, 5% 미만, 2% 미만 또는 1% 미만을 포함하는 형태로 존재하는 경우 실질적으로 단리되거나 실질적으로 정제된다. 대안적으로, 본 개시내용의 기공 복합체는 막에 존재할 때 막관통 기공 또는 막관통 기공 복합체일 수 있다. 본 개시내용은 동일한 돌연변이체 단량체를 포함하는 CsgG에서 유래된 호모올리고머 기공을 포함하는 단리된 기공 및 단리된 기공 복합체를 제공하며, 이는 또한 이의 상동체로서 CsgG 단량체의 돌연변이체 형태를 함유할 수 있다. 대안적으로, 이종올리고머 CsgG 기공을 포함하는 단리된 기공 또는 단리된 기공 복합체가 제공되며, 이는 돌연변이체 및 야생형 CsgG 단량체, 또는 상이한 형태의 CsgG 변이체, 돌연변이체 또는 상동체로 이루어진 CsgG 기공일 수 있다. 단리된 기공 복합체는 전형적으로 적어도 7개, 적어도 8개, 적어도 9개 또는 적어도 10개의 CsgG 단량체, 및 1개 이상의 (변형된) CsgF 펩티드, 예컨대 2, 3, 4, 5, 6, 7, 8, 9, 10개의 CsgF 펩티드를 포함한다. 기공 복합체는 임의의 비율의 CsG 단량체:CsgF 펩티드를 포함할 수 있다. 하나의 구현예에서, CsG 단량체:CsgF 펩티드의 비율은 1:1이다.

[0092] 본원에서 상호교환가능하게 사용된 바와 같은 "수축", "오리피스", "수축 영역", "채널 수축", 또는 "수축 부위"는 기공 또는 기공 복합체의 루멘 표면에 의해 정의된 구멍을 지칭하며, 이는 이온 및 표적 분자(예를 들어, 폴리뉴클레오티드 또는 개별 뉴클레오티드에 제한되지 않음)가 기공 또는 기공 복합체 채널을 통해 통과하는 것을 허용하지만 다른 비표적 분자는 그렇지 않도록 작용한다. 일부 구현예에서, 수축(들)은 기공 또는 기공 복합체 내의 가장 좁은 구멍(들)이다. 이 구현예에서, 수축(들)은 분자가 기공을 통해 통과하는 것을 제한하는 역할을 할 수 있다. 수축의 크기는 전형적으로 핵산 서열분석 적용을 위한 나노기공의 적합성을 결정하는 데 있어서 핵심 요인이다. 수축이 너무 작으면, 서열분석될 분자가 통과할 수 없을 것이다. 그러나, 채널을 통한 이온 흐름에 최대 효과를 달성하기 위해, 수축은 너무 크지 않아야 한다. 예를 들어, 수축은 표적 분석물의 용매 접근가능한 횡단 직경보다 더 넓지 않아야 한다. 이상적으로, 임의의 수축은 분석물이 통과하는 횡단 직경과 가능한 한 가까운 직경이어야 한다. 핵산 및 핵산 염기의 서열분석을 위해, 적합한 수축 직경은 나노미터 범위(10⁻⁹ 미터 범위)이다. 적합하게는, 직경은 0.5 내지 2.0 nm 범위여야 하며, 전형적으로, 직경은 0.7 내지 1.2 nm 범위이다. 야생형 이. 콜라이 CsgG에서 수축은 대략 9 Å(0.9 nm)의 직경을 갖는다. CsgG-유사 기공 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는 기공 복합체에서 형성된 CsgF 수축은 0.5 내지 2 nm 범위 또는 0.7 내지 1.2 nm 범위의 직경을 가지며 따라서 핵산 서열분석에 적합하다.

[0093] 2개 이상의 수축이 존재하고 일정한 간격으로 있는 경우 각 수축은 핵산 가닥 내의 별개의 뉴클레오티드를 동시에 상호작용하거나 "판독"할 수 있다. 이 상황에서, 채널을 통한 이온 흐름의 감소는 뉴클레오티드를 함유하는 모든 수축의 흐름에서 조합된 제한의 결과일 것이다. 따라서, 일부 경우에 이중 수축은 복합 전류 신호로 이어질 수 있다. 특정 상황에서, 2개의 이러한 판독 헤드가 존재하는 경우 하나의 수축, 또는 "판독 헤드"에 대한 전류 판독은 개별적으로 결정될 수 있지 않을 수 있다. 야생형 이. 콜라이 CsgG(서열번호:3)의 수축은 인접한 단백질 단량체의 위치 51에 티로신 잔기(Tyr 51), 및 또한 각각 위치 56 및 55에 페닐알라닌 및 아스파라긴 잔기(Phe 56 및 Asn 55)의 병치에 의해 형성된 2개의 환형 고리로 구성된다. CsgG의 야생형 기공 구조는 대부분의 경우 재조합 유전 기술을 통해 재조작되어 CsgG 수축(본원에서 "CsgG 채널 수축"으로 언급됨)을 구성하는 2개의 환형 고리 중 하나를 넓히거나, 변경하거나, 제거하여, 단일 잘 정의된 판독 헤드를 남겨둔다. CsgG 올리고머 기공의 수축 모티프는 서열번호: 3에 도시된 야생형 단량체 이. 콜라이 CsgG 폴리펩티드의 위치 38 내지 63에서 아미노산 잔기에 위치한다. 이 영역을 고려하면, 아미노산 잔기 위치 50 내지 53, 54 내지 56 및 58 내지 59 중 임의의 것에 있는 돌연변이, 뿐만 아니라 야생형 CsgG 구조의 채널 내에서 Tyr51, Asn55, 및 Phe56의 측쇄 위치 지정의 핵심은 판독 헤드의 특성을 변형시키거나 변경하는 데 유리한 것으로 제시되었다. CsgG-기공 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는 기공 복합체와 관련된 본 개시내용은 놀랍게도 또 다른 수축(본원에서 "CsgF 채널 수축"으로 언급됨)을 CsgG-함유 기공 복합체에 첨가하여, 변형된 CsgF 펩티드와의 복합체 형성을 통해 기공에서 적합한 추가적인 두번째 판독 헤드를 형성한다. 상기 추가적인 CsgF 채널 수축 또는 판독 헤드는 CsgG 기공, 또는 돌연변이된 GcsG 기공의 수축 루프에 인접하게 위치한다. 상기 추가적인 CsgF 채널 수축 또는 판독 헤드는 CsgG 기공, 또는 돌연변이된 GcsG 기공의 수축 루프에서 대략 10nm 미만, 예컨대 5nm 미만, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9 nm에 위치한다. 본 개시내용의 기공 복합체 또는 막관통 기공 복합체는 2개의 판독 헤드를 갖는 기공 복합체를 포함하며, 이는 채널 수축이 다른 수축 채널 판독 헤드의 정확도를 방해하지 않고 적합한 별개의 판독 헤드를 제공하는 방식으로 위치한다는 것을 의미한다. 따라서 상기 기공 복합체는 CsgG 돌연변이체 기공 W02016/034591, W02017/149316, W02017/149317, W02019/002893, W02017/149318, W0 2018/211241, W02019/002893(모두 그 전문이 본원에 참조로 포함됨) 이들 각각은 기공의 특성을 개선하는 야생형 CsgG 기공에 대한 돌연변이를 나열함) 뿐만 아니라 변형된 CsgF 펩티드, 또는 이의 상동

체 또는 돌연변이체와 함께 야생형 CsgG 기공, 또는 이의 상동체를 포함할 수 있으며, 여기서 상기 CsgF 펩티드는 판독 헤드를 형성하는 또 다른 수축 채널을 갖는다.

[0094] 기공 및 기공 복합체

[0095] 본 발명은 단리된 CsgG 기공 또는 이의 상동체 또는 돌연변이체, 또는 CsgG 기공, 또는 이의 상동체 또는 돌연변이체, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는 단리된 기공 복합체를 제공하며, 여기서 CsgG 기공은 서열번호: 117의 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 적어도 하나의 돌연변이체 CsgG 단량체를 포함한다. CsgG 기공은 서열번호: 3 또는 117의 기공 또는 이의 상동체 또는 돌연변이체일 수 있다. 적어도 하나의 돌연변이체 단량체는 바람직하게는 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 서열번호: 117의 변이체를 포함한다.

[0096] 본 발명은 단리된 CsgG 기공 또는 이의 상동체 또는 돌연변이체를 제공하며, 여기서 CsgG 기공은 서열번호: 117의 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 적어도 하나의 돌연변이체 CsgG 단량체를 포함한다. CsgG 기공은 서열번호: 3 또는 117의 기공 또는 이의 상동체 또는 돌연변이체일 수 있다. 적어도 하나의 돌연변이체 단량체는 바람직하게는 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 서열번호: 117의 변이체를 포함한다.

[0097] 본 발명은 CsgG 기공, 또는 이의 상동체 또는 돌연변이체, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는 단리된 기공 복합체를 제공하며, 여기서 CsgG 기공은 서열번호: 117의 위치 W97, Q100, E101, N102 및 T104 중 하나 이상에서 변형을 포함하는 적어도 하나의 돌연변이체 CsgG 단량체를 포함한다. CsgG 기공은 서열번호: 3 또는 117의 기공 또는 이의 상동체 또는 돌연변이체일 수 있다. 적어도 하나의 돌연변이체 단량체는 바람직하게는 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 서열번호: 117의 변이체를 포함한다.

[0098] 적어도 하나의 돌연변이체 단량체 또는 변이체는 서열번호: 117의 위치 (a) W97, (b) Q100, (c) E101, (d) N102 및 (e) T104 중 하나 이상에 변형의 임의의 수 및 조합을 포함할 수 있다. 예를 들면, 적어도 하나의 돌연변이체 단량체는 (a); (b); (c); (d); (e); (a) 및 (b); (a) 및 (c); (a) 및 (d); (a) 및 (e); (b) 및 (c); (b) 및 (d); (b) 및 (e); (c) 및 (d); (c) 및 (e); (d) 및 (e); (a), (b) 및 (c); (a), (b) 및 (d); (a), (b) 및 (e); (a), (c) 및 (d); (a), (c) 및 (e); (a), (d) 및 (e); (b), (c) 및 (d); (b), (c) 및 (e); (b), (d) 및 (e); (c), (d) 및 (e); (a), (b), (c) 및 (d); (a), (b), (c) 및 (e); (a), (b), (d) 및 (e); (a), (c), (d) 및 (e); (b), (c), (d) 및 (e); 및 (a), (b), (c), (d) 및 (e)에 변형을 포함할 수 있다. 적어도 하나의 돌연변이체 단량체 또는 변이체는 바람직하게는 서열번호: 117의 하나 이상의 위치 (a) W97, (b) Q100, (c) E101 및 (d) N102에 변형을 포함하며, 위에 제시된 (a) 내지 (d)의 모든 조합을 포함한다.

[0099] 서열번호: 117의 위치 W97, Q100, E101, N102 및 T104 중 하나 이상에서의 변형은 아래에 보다 상세하게 논의된 변형 중 임의의 것일 수 있다. 변형은 E101의 결실과 같은 결실일 수 있다. E101의 결실은 기공을 통해 이동 속도를 증가시킨다(실시예 1). 변형은 바람직하게는 치환이다.

[0100] 위치 97의 W는 바람직하게는 R, H, K, A, V, I, L, M, F, Y, S, T, Q, D, E, N, C, P 또는 G로 치환된다. 위치 97의 W는 보다 바람직하게는 D, G, N, R 또는 S로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시킨다(실시예 1). 위치 97의 W는 보다 바람직하게는 D 또는 R로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시키고 정규화된 속도 분포를 증가시킨다(실시예 1). 위치 97의 W는 보다 바람직하게는 G, N 또는 S로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시키고 정규화된 속도 분포를 감소시킨다(실시예 1).

[0101] 위치 100의 Q는 바람직하게는 R, H, K, W, A, V, I, L, M, F, Y, T, N 또는 S로 치환된다. 위치 100의 Q는 보다 바람직하게는 A, K 또는 S로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시킨다(실시예 1). 위치 100의 Q는 보다 바람직하게는 K(Q100K)로 치환된다. 이 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시키고 정규화된 속도 분포를 증가시킨다(실시예 1). 위치 100의 Q는 보다 바람직하게는 A 또는 S(Q100A 또는 Q100S)로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시키고 정규화된 속도 분포를 감소시킨다(실시예 1). 위치 100의 Q는 가장 바람직하게는 A(Q100A)로 치환된다. 이 치환은 또한 본 발명의 변형된 Dda 헬리카제와 함께 사용될 때 실시예 4에 제시된 효과를 갖는다.

[0102] 위치 101의 E는 바람직하게는 A, V, I, L, M, F, Y 또는 W로 치환된다. 위치 101의 E는 보다 바람직하게는 A(E101A)로 치환된다. 이 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 감소시키고 정규화된 속도 분

포를 감소시킨다(실시예 1).

[0103] 위치 101의 E는 바람직하게는 S, T, N, Q, C, G 또는 P로 치환된다. 위치 101의 E는 보다 바람직하게는 G 또는 S로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시킨다(실시예 1). 위치 101의 E는 보다 바람직하게는 G(E101G)로 치환된다. 이 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시키고 정규화된 속도 분포를 감소시킨다(실시예 1). 위치 101의 E는 보다 바람직하게는 S(E101S)로 치환된다. 이 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시키고 정규화된 속도 분포를 증가시킨다(실시예 1).

[0104] 위치 102의 N은 바람직하게는 D, E, R, H, K, S, T, Q, V, I, L, M, F, Y, W 또는 A로 치환된다. 위치 102의 N은 보다 바람직하게는 A, D, R, S 또는 W로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시킨다(실시예 1). 위치 102의 N은 바람직하게는 A, R 또는 S로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시키고 정규화된 속도 분포를 감소시킨다(실시예 1). 위치 102의 N은 바람직하게는 D 또는 W(N102D 또는 N102W)로 치환된다. 이러한 치환은 분석물이 기공을 통해 통과하는/상대적인 속도를 증가시키고 정규화된 속도 분포를 증가시킨다(실시예 1). 위치 102의 N은 가장 바람직하게는 A 또는 S(N102A 또는 N102S)로 치환된다. 이러한 치환은 또한 본 발명의 변형된 Dda 헬리카제와 함께 사용될 때 실시예 4에 제시된 효과를 갖는다.

[0105] 위치 104의 T는 바람직하게는 R, H 또는 K로 치환된다.

[0106] 아래에 보다 상세하게 설명된 바와 같이, 기공은 바람직하게는 6 내지 10개의 단량체를 포함한다. 이들 중 임의의 수, 예컨대 6, 7, 8, 9 또는 10은 서열번호: 117의 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 돌연변이체 단량체일 수 있거나 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 서열번호: 117의 변이체를 포함할 수 있다. 6 내지 10개의 단량체는 모두 서열번호: 117의 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함할 수 있거나 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 서열번호: 117의 변이체를 포함할 수 있다.

[0107] 돌연변이체 CsgG 단량체는 서열이 야생형 CsgG 단량체의 서열과 다르고 기공을 형성하는 능력을 유지하는 단량체이다. 돌연변이체 단량체는 또한 본원에서 변이체로 언급될 수 있다. 기공을 형성하는 돌연변이체 단량체의 능력을 확인하기 위한 방법은 당업계에 잘 알려져 있고 아래에 보다 상세하게 논의되어 있다. 적어도 하나의 돌연변이체 단량체 또는 변이체는 아래에 제시된 서열번호: 117 또는 서열번호: 3에 대한 상동성/서열 동일성의 % 중 임의의 것을 가질 수 있다. 적어도 하나의 돌연변이체 단량체는 본 발명의 Dda 헬리카제에 관하여 기재된 변형 및 치환의 유형을 포함하여, 아래에 기재된 추가적인 변형, 돌연변이 또는 치환 중 임의의 것을 함유할 수 있다. 적어도 하나의 돌연변이체 단량체는 W02016/034591, W02017/149316, W02017/149317 및 W02017/149318, W02018/211241, 및 W02019/002893(모두 그 전문이 본원에 참조로 포함됨)에 기재된 추가적인 변형, 돌연변이 또는 치환 중 임의의 것을 함유할 수 있다.

[0108] 본 발명은 임의적으로 세포외적으로 위치한 CsgF 펩티드와 복합체화되어 놀랍게도 기공 복합체에 추가적인 채널 수축 또는 판독 헤드를 도입하는 CsgG 기공에 관한 것이다. 더욱이, 본 개시내용은 기공 복합체 내에서 CsgF 펩티드에 의해 만들어진 수축에 대한 위치 정보를 제공하며, 펩티드는 CsgG 기공의 루멘에 삽입되고, 수축 부위는 CsgF 단백질의 N-말단 부분에 있다. 더욱이, 본 개시내용의 변형된 또는 절두된 CsgF 펩티드는 기공 복합체 형성에 충분한 것으로 나타났고, 생물감지 적용을 위한 수단 및 방법을 제공한다. 본 개시내용은 임의적으로 변형된 또는 절두된 CsgF 펩티드 및 이의 돌연변이체 또는 상동체와 조합된, 야생형 및 돌연변이체 CsgG 기공(예를 들어, W02016/034591, W02017/149316, W02017/149317, W02017/149318 및 국제 특허 출원 번호 PCT/GB2018/051191에 개시된 바와 같음), 또는 이의 상동체 또는 돌연변이체를 포함하며, 모두 함께 폴리뉴클레오티드와 같은 분석물과 상호작용하는 CsgG 기공 또는 CsgG-유사 기공 복합체의 능력을 개선시킨다. (변형된 또는 절두된) CsgF 펩티드와의 복합체 형성에 의해 CsgG-유사 나노기공 채널에 도입된 추가적인 수축은 통과하는 분석물과의 접촉 표면을 확장하고 분석물 검출 및 특성화를 위한 두번째 판독 헤드로서 역할을 할 수 있다. CsgF의 신규 돌연변이체 또는 변형된 형태와 조합된 돌연변이체 CsgG 단량체를 포함하는 기공은 폴리뉴클레오티드와 같은 분석물의 특성화를 개선시켜, 폴리뉴클레오티드가 기공을 통해 이동할 때 관찰된 전류 사이에 보다 구별되는 직접적인 관계를 제공할 수 있다. 특히, 정의된 거리에서 간격을 둔 2개의 적재된 판독 헤드를 가짐으로써, CsgG:CsgF 기공 복합체는 적어도 하나의 동중중합체 스트레치, 예를 들어, 달리 단일 CsgG 판독 헤드의 상호작용 길이를 초과하는 동일한 뉴클레오티드의 여러 연속적인 사본을 함유하는 폴리뉴클레오티드의 특성화를 용이하게 할 수 있다. 추가적으로, 정의된 거리에서 2개의 적재된 수축을 가짐으로써, CsgG:CsgF 복합 기공을

통해 통과하는 유기 또는 무기 약물 및 오염물질을 포함하는 소분자 분석물은 2개의 독립적인 판독 헤드를 연속적으로 통과할 것이다. 어느 한 판독 헤드의 화학적 속성은 독립적으로 변형될 수 있으며, 각각은 분석물과 고유한 상호작용 특성을 제공하므로, 분석물 검출 동안 추가적인 구별력을 제공한다.

[0109] 본 발명은 CsgG 기공, 또는 이의 상동체 또는 돌연변이체, 또는 CsgG-유사 기공, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는 단리된 기공 복합체에 관한 것이다. 실제로, 본 개시내용은 변형된 CsgF 펩티드를 포함하며 이의 절두된 돌연변이체 및/또는 변이체일 수 있는 변형된 CsgG 생물학적 기공에 관한 것이다. 하나의 구현예에서, 상기 변형된 CsgF 펩티드 또는 이의 상동체 또는 돌연변이체 사이의 상호작용 영역은 CsgG 기공, 또는 이의 상동체 또는 돌연변이체의 루멘에 위치한다. 또 다른 구현예에서, 기공 복합체는 CsgG 기공의 적어도 하나의 수축에 의해 제공되는 2개 이상의 수축 부위 또는 판독 헤드를 가지며, 적어도 하나는 CsgF 펩티드에 의해 도입되어, CsgG 기공과 복합체를 형성한다. 서열번호:5의 아미노산 잔기 39-64, 또는 보다 특히 서열번호:5의 아미노산 잔기 49-64의 범위에 위치를 포함하는 N-말단 CsgF 위치는 검출가능한 양의 안정적인 CsgG:CsgF 복합체를 허용하는 것으로 나타났다. 하나의 구현예에서, 변형된 CsgF 펩티드(예를 들어, 본원에 기재된 것)에 의해 생산된 CsgF 수축은 기공 복합체의 CsgG 기공에서 첫번째 수축과 인접하거나 직접 대면한다. CsgG 또는 CsgG-유사 단백질 기공의 경우, 수축 부위는 베타 가닥의 루프 영역에 의해 형성되는 것으로 결정되었다.

[0110] 하나의 구현예에서, 변형된 CsgF 펩티드는 상기 변형이 특히 수축 영역을 함유하고 CsgG 단량체, 또는 이의 상동체 또는 돌연변이체에 결합하도록 제한하여 정의된 N-말단 CsgF 펩티드 단편을 포함하는 절두된 CsgF 단백질 또는 단편을 지칭하는 펩티드이다. 상기 변형된 CsgF 펩티드는 추가적으로 돌연변이 또는 상동 서열을 포함할 수 있으며, 이는 기공 복합체의 특정 특성을 용이하게 할 수 있다. 특정 구현예에서, 변형된 CsgF 펩티드는 야생형 전단백질(서열번호:5) 또는 성숙 단백질(서열번호:6) 서열, 또는 이의 상동체와 비교하여 CsgF 단백질 절두를 포함한다. 이러한 변형된 펩티드는 CsgG 및 변형된 또는 절두된 CsgF 펩티드에 의해 형성된 CsgG-유사 기공 내에 추가적인 수축 부위 또는 판독 헤드를 도입하는 기공 복합체 구성요소로서 기능하도록 의도된다. 절두된 변형된 펩티드의 예는 아래에 기재되어 있다.

[0111] 변형된 CsgF 펩티드의 상동체의 예는 W02019/002893(그 전문이 본원에 참조로 포함됨)에 개시되어 있고 상이한 박테리아 균주에서 상동 또는 유사한 수축 영역을 포함하는 CsgF-유사 단백질 또는 CsgF 펩티드를 보여주며, 이는 유사한 기공 복합체의 사용에 유용할 수 있다. 다양한 CsgF 상동체로부터 유래된 CsgF 펩티드의 구조적 특성 및 CsgG-결합 요소는 보존되므로, CsgF 펩티드가 상이한 야생형 또는 돌연변이체 CsgG 기공과 조합하여 사용될 수 있게 한다. 이는 CsgG 기공과 비동종 CsgF의 복합체를 포함하며, 이는 CsgG 기공 및 CsgF가 유래되는 모 CsgF 상동체가 동일한 오펜, 박테리아 종 또는 균주에서 기원할 필요가 없다는 것을 의미한다.

[0112] 대안적인 구현예에서, 기공 복합체 내의 CsgG 기공은 야생형 기공이 아니지만, 기공 특성을 증가시키기 위한 돌연변이 또는 변형을 또한 포함한다. CsgG 기공, 또는 이의 상동체, 및 변형된 CsgF 펩티드, 또는 이의 상동체에 의해 형성된 본 개시내용의 단리된 기공 복합체는 CsgG 기공의 야생형 형태에 의해 형성될 수 있거나 특정 아미노산 잔기의 지시된 돌연변이생성에 의해서와 같이 CsgG 기공에서 추가로 변형되어, 기공 복합체 내에서 사용하기 위한 CsgG 기공의 원하는 특성을 추가로 향상시킬 수 있다. 예를 들어, 본 발명의 구현예에서 돌연변이는 채널 내에서 수축의 수, 크기, 모양, 배치 또는 방향을 변경하는 것으로 고려된다. 변형된 돌연변이체 CsgG 기공을 포함하는 기공 복합체는 폴리펩티드 서열에서 특이적 표적화된 아미노산 잔기의 삽입, 치환 및/또는 결실을 초래하는 알려진 유전자 조작 기술에 의해 제조될 수 있다. 올리고머 CsgG 기공의 경우, 돌연변이는 각 단량체 폴리펩티드 서브유닛, 또는 단량체 중 임의의 하나, 또는 모든 단량체에서 제조될 수 있다. 적합하게는, 본 발명의 하나의 구현예에서 기재된 돌연변이는 올리고머 단백질 구조 내의 모든 단량체 폴리펩티드로 이루어진다. 돌연변이체 CsgG 단량체는 서열이 야생형 CsgG 단량체의 서열과 다르고 기공을 형성하는 능력을 유지하는 단량체이다. 기공을 형성하는 돌연변이체 단량체의 능력을 확인하는 방법은 당업계에 잘 알려져 있다. 본 개시내용은 변형된 또는 절두된 CsgF 펩티드 및 이들의 돌연변이체 또는 상동체와 조합된 야생형 및 돌연변이체 CsgG 기공(예를 들어, W02016/034591, W02017/149316, W02017/149317, W02017/149318 및 국제 특허 출원 번호 PCT/GB2018/051191에 개시된 바와 같음), 또는 이의 상동체를 포함하며, 모두 함께 폴리뉴클레오티드와 같은 분석물과 상호작용하는 CsgG-유사 기공 복합체의 능력을 개선시킨다. 돌연변이체 CsgG 기공은 하나 이상의 돌연변이체 단량체를 포함할 수 있다. CsgG 기공은 동일한 단량체를 포함하는 동종중합체, 또는 2개 이상의 상이한 단량체를 포함하는 이종중합체일 수 있다. 단량체는 아래에 기재된 돌연변이 중 하나 이상을 임의의 조합으로 가질 수 있다.

[0113] 변형된 CsgF 펩티드가 특정 구현예에서 야생형 CsgF 단백질의 N-말단 단편 또는 절두만을 포함하므로 변형된

CsgF 펩티드를 포함하는 나노기공 복합체는 서열번호:6에 묘사된 야생형 CsgF 단백질과 비교하여 상이하다. 그러나 변형된 CsgF 펩티드는 추가적으로 또는 대안적으로 아미노산 치환으로서 돌연변이가 CsgG 기공 및 변형된 CsgF 펩티드를 포함하는 복합체에 의해 형성된 기공에서 더 나은 두번째 수축 부위를 허용한다는 의미에서 돌연변이된 CsgF 펩티드일 수 있다. 이와 같이 돌연변이체 단량체는 상기 복합체가 뉴클레오티드 서열분석에 사용될 때 개선된 폴리뉴클레오티드 판독 특성을 가질 수 있으며, 즉 2개의 판독 헤드를 포함하는 복합체의 개선된 특징 외에도 개선된 폴리뉴클레오티드 포획 및 뉴클레오티드 구별을 나타낼 수 있다. 특히, 돌연변이체 펩티드로부터 구성된 기공은 야생형보다 뉴클레오티드 및 폴리뉴클레오티드를 보다 용이하게 포획한다. 게다가, 돌연변이체 펩티드로부터 구성된 기공은 상이한 뉴클레오티드를 구별하는 것을 더 쉽게 만드는 증가된 전류 범위, 및 신호 대 잡음비를 증가시키는 감소된 상태 변화를 나타낼 수 있다. 게다가, 폴리뉴클레오티드가 돌연변이체로부터 구성된 기공을 통해 이동할 때 전류에 기여하는 뉴클레오티드의 수는 감소될 수 있다. 이는 폴리뉴클레오티드가 기공을 통해 이동할 때 관찰된 전류와 폴리뉴클레오티드 서열 사이의 직접적인 관계를 식별하는 것을 더 용이하게 만든다. 게다가, 돌연변이체 펩티드로부터 구성된 기공은 증가된 처리량을 나타낼 수 있으며, 예를 들어, 폴리뉴클레오티드와 같은 분석물과 상호작용할 가능성이 더 높다. 이는 기공을 사용하여 분석물을 특성화하는 것을 더 용이하게 만든다. 돌연변이체 펩티드로부터 구성된 기공은 막에 더 용이하게 삽입될 수 있거나, 기공 복합체의 인근에 추가적인 단백질을 유지하는 더 쉬운 방식을 제공할 수 있다.

[0114] 대안적인 구현예에서, 본 발명의 기공 복합체에 제공된 CsgF 수축 부위는 0.5 nm 내지 2.0 nm 범위의 직경을 가지므로, 위에 기재된 바와 같이 핵산 서열분석에 적합한 기공 복합체를 제공한다.

[0115] 기공은 CsgF 펩티드를 CsgG 기공에 공유 부착함으로써 안정화될 수 있다. 공유 결합은 예를 들어 이황화물 결합, 또는 클릭 화학일 수 있다. CsgF 펩티드 및 CsgG 기공은 예를 들어, 각각 서열번호: 6 및 서열번호: 3 또는 서열번호: 117의 다음 위치 쌍 중 하나 이상에 상응하는 위치에서 잔기를 통해 공유적으로 연결될 수 있다: 1 및 153, 4 및 133, 5 및 136, 8 및 187, 8 및 203, 9 및 203, 11 및 142, 11 및 201, 12 및 149, 12 및 203, 26 및 191, 및 29 및 144.

[0116] 기공에서, CsgF 펩티드와 CsgG 기공 사이의 상호작용은 예를 들어, 각각 서열번호: 6 및 서열번호: 3 또는 서열번호: 117의 다음 위치 쌍 중 하나 이상에 상응하는 위치에서 소수성 상호작용 또는 정전기적 상호작용에 의해 안정화될 수 있다: 1 및 153, 4 및 133, 5 및 136, 8 및 187, 8 및 203, 9 및 203, 11 및 142, 11 및 201, 12 및 149, 12 및 203, 26 및 191, 및 29 및 144.

[0117] 위에 나열된 위치 중 하나 이상에서 CsgF 및/또는 CsgG의 잔기는 기공에서 CsgG와 CsgF 사이의 상호작용을 향상시키기 위해 변형될 수 있다.

[0118] 하나의 구현예에서, 본 발명의 기공은 단리되거나, 실질적으로 단리되거나, 정제되거나 실질적으로 정제될 수 있다. 본 발명의 기공은 지질 또는 다른 기공과 같은 임의의 다른 구성요소가 완전히 없는 경우 단리되거나 정제된다. 기공은 의도된 용도를 방해하지 않을 담체 또는 희석제와 혼합된 경우 실질적으로 단리된다. 예를 들면, 기공은 트리블록 공중합체, 지질 또는 다른 기공과 같은 다른 구성요소의 10% 미만, 5% 미만, 2% 미만 또는 1% 미만을 포함하는 형태로 존재하는 경우 실질적으로 단리되거나 실질적으로 정제된다. 대안적으로, 본 발명의 기공은 막에 존재할 수 있다. 적합한 막은 아래에 논의되어 있다.

[0119] 본 발명의 기공은 개별 또는 단일 기공으로 존재할 수 있다. 대안적으로, 본 발명의 기공은 2개 이상의 기공의 상동 또는 이종 집단으로 존재할 수 있다.

[0120] CsgF 펩티드

[0121] 본 발명의 단리된 기공 복합체는 변형된 CsgF 단량체(펩티드), 또는 절두된 CsgF 단백질, 또는 CsgF 상동체 또는 돌연변이체의 변형된 또는 절두된 펩티드를 포함한다. 이러한 신규 변형된 CsgF 펩티드는 두번째 또는 추가적인 판독 헤드를 통합하기 위해 기공 복합체에서 사용될 수 있다. 상기 변형 또는 절두는 바람직하게는 야생형 CsgF, 또는 돌연변이체 또는 상동체 CsgF 단백질의 단편, 보다 바람직하게는 N-말단 단편을 생성한다. 서열번호 5의 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체는 W02016/034591, W02017/149316, W02017/149317, W02017/149318, W02018/211241, 및 W02019/002893(모두 그 전문이 본원에 참조로 포함됨)에 개시된 것들 중 임의의 것일 수 있다.

[0122] 본 발명의 일부를 형성하는 CsgF 펩티드는 C-말단 헤드가 결여되거나; CsgF의 C-말단 헤드 및 넥 도메인의 일부가 결여되거나(예를 들어, 절두된 CsgF 펩티드는 CsgF의 넥 도메인의 일부만 포함할 수 있음); CsgF의 C-말단 헤드 및 넥 도메인이 결여된 절두된 CsgF 펩티드이다. CsgF 펩티드는 CsgF 넥 도메인의 일부가 결여될 수 있

며, 예를 들어 CsgF 펩티드는 예컨대 예를 들어, 넥 도메인의 N-말단 단부에 있는 아미노산 잔기 36부터(서열번호:6 참조)(예를 들어 서열번호: 6의 잔기 36-40, 36-41, 36-42, 36-43, 36-45, 36-46 잔기 36-50 또는 36-60 까지) 넥 도메인의 일부를 포함할 수 있다. CsgF 펩티드는 바람직하게는 CsgG-결합 영역 및 기공에서 수축을 형성하는 영역을 포함한다. CsgG-결합 영역은 전형적으로 CsgF 단백질(서열번호: 6 또는 또 다른 종의 상동체)의 잔기 1 내지 8 및/또는 29 내지 32를 포함하고 하나 이상의 변형을 포함할 수 있다. 기공에 수축을 형성하는 영역은 전형적으로 CsgF 단백질(서열번호: 6 또는 또 다른 종의 상동체)의 잔기 9 내지 28을 포함하고 하나 이상의 변형을 포함할 수 있다. 잔기 9 내지 17은 보존된 모티프 N9PXFGGXXX17을 포함하고 회전 영역을 형성한다. 잔기 9 내지 28은 알파-나선을 형성한다. X17(서열번호: 6의 N17)은 기공에서 CsgF 수축의 가장 좁은 부분에 상응하는 수축 영역의 정점을 형성한다. CsgF 수축 영역은 또한 주로 서열번호: 6의 잔기 9, 11, 12, 18, 21 및 22에서 CsgG 베타-배럴과의 접촉을 안정화시킨다.

[0123] CsgF 펩티드는 전형적으로 28 내지 50개 아미노산, 예컨대 29 내지 49개, 30 내지 45개 또는 32 내지 40개 아미노산 길이를 갖는다. 바람직하게는 CsgF 펩티드는 29 내지 35개 아미노산, 또는 29 내지 45개 아미노산을 포함한다. CsgF 펩티드는 서열번호: 6의 잔기 1 내지 35에 상응하는 FCP의 전부 또는 일부를 포함한다. CsgF 펩티드가 FCP보다 더 짧은 경우, 절두는 바람직하게는 C-말단 단부에서 이루어진다.

[0124] 서열번호:6 또는 이의 상동체 또는 돌연변이체의 CsgF 단편은 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 또는 55개 아미노산 길이를 가질 수 있다.

[0125] CsgF 펩티드는 잔기 1부터 잔기 25 내지 60, 예컨대 27 내지 50, 예를 들어, 서열번호: 6의 28 내지 45 중 임의의 하나 까지의 아미노산 서열, 또는 서열번호: 6의 상동체, 또는 이의 어느 하나의 변이체로부터 상응하는 잔기를 포함할 수 있다. 보다 구체적으로, CsgF 펩티드는 서열번호: 39(서열번호: 6의 잔기 1 내지 29), 또는 이의 상동체 또는 변이체를 포함할 수 있다.

[0126] 이러한 CsgF 펩티드의 예는 서열번호: 15(서열번호: 6의 잔기 1 내지 34), 서열번호: 54(서열번호: 6의 잔기 1 내지 30), 서열번호: 40(서열번호: 6의 잔기 1 내지 45), 또는 서열번호: 55(서열번호: 6의 잔기 1 내지 35) 및 이의 임의의 상동체 또는 변이체를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다. CsgF 펩티드의 다른 예는 서열번호: 7, 서열번호: 8, 서열번호: 9, 서열번호: 10, 서열번호: 11, 서열번호: 12, 서열번호: 13, 서열번호: 14, 서열번호: 16을 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다.

[0127] CsgF 펩티드에서, 예를 들어, 서열번호: 15, 서열번호: 39, 서열번호: 40, 서열번호: 54, 또는 서열번호: 55의 하나 이상의 잔기가 변형될 수 있다.

[0128] 예를 들어, CsgF 펩티드는 서열번호: 6의 다음 위치 중 하나 이상에 상응하는 위치에 변형을 포함할 수 있다: G1, T4, F5, R8, N9, N11, F12, A26 및 Q29.

[0129] CsgF 펩티드는 예를 들어 서열번호: 6의 다음 위치 중 하나 이상에 상응하는 위치에 시스테인, 소수성 아미노산, 하전된 아미노산, 비천연 반응성 아미노산, 또는 광반응성 아미노산을 도입하도록 변형될 수 있다: G1, T4, F5, R8, N9, N11, F12, A26 및 Q29.

[0130] 예를 들어, CsgF 펩티드는 서열번호: 6의 다음 위치 중 하나 이상에 상응하는 위치에 변형을 포함할 수 있다: N15, N17, A20, N24 및 A28. CsgF 펩티드는 CsgG-CsgF 복합체를 안정화하기 위해 D34에 상응하는 위치에 변형을 포함할 수 있다. 특정 구현예에서, CsgF 펩티드는 다음 치환 중 하나 이상을 포함한다: N15S/A/T/Q/G/L/V/I/F/Y/W/R/K/D/C, N17S/A/T/Q/G/L/V/I/F/Y/W/R/K/D/C, A20S/T/Q/N/G/L/V/I/F/Y/W/R/K/D/C, N24S/T/Q/A/G/L/V/I/F/Y/W/R/K/D/C, A28S/T/Q/N/G/L/V/I/F/Y/W/R/K/D/C 및 D34F/Y/W/R/K/N/Q/C. CsgF 펩티드는 예를 들어, 다음 치환 중 하나 이상을 포함할 수 있다: G1C, T4C, N17S, 및 D34Y 또는 D34N.

[0131] CsgF 펩티드는 효소를 사용하여 전장 CsgF와 같은 더 긴 단백질의 절단에 의해 생산될 수 있다. 특정 부위에서 절단은 전장 CsgF와 같은 더 긴 단백질을 변형시켜, 적절한 위치에 효소 절단 부위를 포함하도록 지시될 수 있다. 이러한 효소 절단 부위를 포함하도록 변형된 CsgF 아미노산 서열의 예는 서열번호: 56 내지 67에 제시되어 있다. 절단 후 첨가된 효소 절단 부위의 전부 또는 일부는 CsgG와 회합하여 기공을 형성하는 CsgF 펩티드에 존재할 수 있다. 따라서 CsgF 펩티드는 C-말단 단부에 효소 절단 부위의 전부 또는 일부를 추가로 포함할 수 있다.

[0132] 적합한 CsgF 펩티드의 일부 예는 아래 표 3에 제시되어 있다:

[0133] 표 3: CsgF 펩티드

기술적 설명	단백질 서열	서열번호:
CsgF-(1-45)	GTMTFQFRNPNGGNPNNGAFLNSAQAQN SYKDPSYNDDFGIET	40
CsgF-(1-29)	GTMTFQFRNPNGGNPNNGAFLNSAQAQ	39
CsgF-(1-35)	GTMTFQFRNPNGGNPNNGAFLNSAQAQN SYKDP	55
CsgF-G1C-(1-45)	CTMTFQFRNPNGGNPNNGAFLNSAQAQN SYKDPSYND DFGIET	89
CsgF-G1C-(1-30)	CTMTFQFRNPNGGNPNNGAFLNSAQAQ	90
CsgF-G1C-(1-35)	CTMTFQFRNPNGGNPNNGAFLNSAQAQN SYKDP	91
CsgF-N17S-(1-35)	GTMTFQFRNPNGGNPSNGAFLNSAQAQN SYKDP	92
CsgF-G1C/D34Y-(1-35)	CTMTFQFRNPNGGNPNNGAFLNSAQAQN SYKYP	93
CsgF-N17S-(1-45)	GTMTFQFRNPNGGNPSNGAFLNSAQAQN SYKDPSYND DFGIET	94
CsgF-G1C/N17S/D34Y-(1-35)	CTMTFQFRNPNGGNPSNGAFLNSAQAQN SYKYP	95
CsgF-G1C-(1-30)-H10	CTMTFQFRNPNGGNPNNGAFLNSAQAQH HHHHHHHHH	96
CsgF-G1C-(1-35)-H10	CTMTFQFRNPNGGNPNNGAFLNSAQAQN SYKDPPHHHHHHHHH	97
CsgF-N17S/D34Y-(1-35)	GTMTFQFRNPNGGNPSNGAFLNSAQAQN SYKYP	98
CsgF-N17S/D34N-(1-35)	GTMTFQFRNPNGGNPSNGAFLNSAQAQN SYKNP	99
CsgF-T4C/N17S-(1-35)	GTMCFQFRNPNGGNPSNGAFLNSAQAQN SYKDP	100
CsgF-T4C/N17S/D34Y-(1-35)	GTMCFQFRNPNGGNPSNGAFLNSAQAQN SYKYP	101
CsgF-T4C/N17S/D34N-(1-35)	GTMCFQFRNPNGGNPSNGAFLNSAQAQN SYKNP	102

[0134]

CsgF-G1C/T4C/N17S-(1-35)	CTMCFQFRNPNGGNPSNGAFLNSAQAQN SYKDP	103
CsgF-G1C/T4C/N17S/D34N-(1-35)	CTMCFQFRNPNGGNPSNGAFLNSAQAQN SYKNP	104
CsgF-G1C/T4C/N17S/D34Y-(1-35)	CTMCFQFRNPNGGNPSNGAFLNSAQAQN SYKYP	105
CsgF-N17V-(1-35)	GTMTFQFRNPNGGNPVNGAFLNSAQAQN SYKDP	106
CsgF-N17A-(1-35)	GTMTFQFRNPNGGNPANGAFLNSAQAQN SYKDP	107
CsgF-N17F-(1-35)	GTMTFQFRNPNGGNPFNGAFLNSAQAQN SYKDP	108
CsgF-N17S-(1-30)	GTMTFQFRNPNGGNPSNGAFLNSAQAQN	109
CsgF-N17V-(1-30)	GTMTFQFRNPNGGNPVNGAFLNSAQAQN	110
CsgF-N17A-(1-30)	GTMTFQFRNPNGGNPANGAFLNSAQAQN	111
CsgF-N17F-(1-30)	GTMTFQFRNPNGGNPFNGAFLNSAQAQN	112
CsgF-N17V/D34Y-(1-35)	GTMTFQFRNPNGGNPVNGAFLNSAQAQN SYKYP	113
CsgF-N17A/D34Y-(1-35)	GTMTFQFRNPNGGNPANGAFLNSAQAQN SYKYP	114
CsgF-N17F/D34Y-(1-35)	GTMTFQFRNPNGGNPFNGAFLNSAQAQN SYKYP	115
CsgF-N17S/A20Q-(1-35)	GTMTFQFRNPNGGNPSNGQFLNSAQAQN SYKDP	116

[0135]

[0136] 특정 구현예에서, 상기 CsgF 단편은 아미노산 서열 서열번호:39, 또는 이의 돌연변이체 또는 상동체를

포함한다. 특히, 서열번호:39는 성숙 CsgF 펩티드(서열번호:6)의 처음 29개 아미노산을 포함한다. 또 다른 구현예에서, 본 발명의 변형된 CsgF 펩티드는 서열번호:40을 포함하는 절두된 펩티드이다. 특히, 서열번호:40은 성숙 CsgF 펩티드(서열번호:6)의 처음 45개 아미노산을 포함한다. 특히, CsgG에 대한 CsgF 수축 부위 및 결합 부위는 N-말단 CsgF 펩티드 영역 내에 위치하며, 서열번호:5의 아미노산 39 내지 64(서열번호:39 및 서열번호:40에 존재), 또는 특히 서열번호:5의 아미노산 49 내지 64(서열번호:40에 존재하지만, 서열번호:39에는 존재하지 않으며, 서열번호:39에 의해 암호화된 후자의 단편은 CsgG와의 더 약한 상호작용을 보임(실시에 참조))가 복합체에 더 높은 안정성을 부여한다는 점에서 추가로 특성화된다. 따라서, 본 개시내용은 단백질을 상기 펩티드 또는 상기 N-말단 단편 또는 수축 부위 영역을 포함하는 펩티드로 절두시켜 생체 내에서 CsgG 기공, 또는 이의 상동체 또는 돌연변이체와의 복합체 형성을 허용함으로써 CsgF 단백질의 변형을 제공한다. 서열번호:37 또는 서열번호:38을 포함하는 변형된 CsgF 펩티드와 관련하여 하나의 구현예에서 추가 제한이 제공된다. 마지막으로, 특히 수축 영역 내에 정렬된 CsgF 상동 펩티드(FCP 펩티드)의 식별은 또한 상기 단리된 복합체의 일부를 형성할 수 있는 변형된 CsgF 펩티드 상동체를 제공한다.

[0137] 추가의 구현예는 서열번호:15를 포함하는 변형된 또는 절두된 CsgF 펩티드에 관한 것이며, 여기서 상기 서열번호:15는 CsgG 또는 이의 상동체, 및 변형된 CsgF 펩티드를 포함하는 복합 기공의 시험관 내 재구성에 충분한 CsgG 결합 및/또는 수축 부위의 영역으로부터 여러 잔기를 포함하는 CsgF 단백질의 영역을 함유하여, CsgF 채널 수축을 포함하는 단리된 기공 복합체를 생성한다. 또 다른 구현예는 CsgF 단백질의 N-말단 단편을 함유하는 서열번호:16을 포함하는 상기 변형된 CsgF 펩티드, 및 (합성) 펩티드의 용해도 및 안정성을 증가시킬 뿐만 아니라, 상기 복합 기공의 시험관 내 재구성을 허용하는 2개의 추가적인 아미노산(KD)을 기재한다. 상기 변형된 CsgF 펩티드가 서열번호:15, 서열번호:16 또는 이의 상동체 또는 돌연변이체를 포함하는 추가의 구현예가 제공되며, 여기서 상기 변형된 CsgF 펩티드는 추가로 돌연변이되지만, 상기 서열번호:15 또는 16에 상응하는 변형된 CsgF 펩티드 영역 내에서 각각 서열번호:15, 또는 서열번호:16에 대해 최소 35 % 아미노산 동일성 예를 들어, 40%, 50%, 60%, 70%, 80% 85%, 90% 아미노산 동일성을 여전히 유지한다. 상기 변형된 CsgF 펩티드가 서열번호:15, 서열번호:16 또는 이의 상동체 또는 돌연변이체를 포함하는 추가의 구현예가 제공되며, 여기서 상기 변형된 CsgF 펩티드는 추가로 돌연변이되지만, 상기 서열번호:15 또는 16에 상응하는 변형된 CsgF 펩티드 영역 내에서 각각 서열번호:15, 또는 서열번호:16에 대해 최소 40%, 45%, 50%, 60%, 70%, 80% 85% 또는 90% 아미노산 동일성을 여전히 유지한다. 이러한 돌연변이된 영역은 위에 논의된 바와 같이 CsgF 수축 부위의 특성을 변경하고/하거나 개선하도록 의도되므로, 예를 들면 보다 정확한 표적 분석이 수득될 수 있다. 또 다른 구현예는 서열번호:39, 서열번호:40, 서열번호:54 또는 서열번호:55를 포함하는 영역에서 하나 이상의 위치가 변형되고, 상기 돌연변이(들)가 서열번호:39, 서열번호:40, 서열번호:54 또는 서열번호:55를 포함하는 영역에 상응하는 펩티드 단편에서 서열번호:39, 서열번호:40, 서열번호:54 또는 서열번호:55에 대해 최소 35 % 아미노산 동일성, 또는 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95% 아미노산 동일성을 유지하는 변형된 CsgF 펩티드를 개시한다.

[0138] 추가적인 구현예는 단리된 기공 복합체에 관한 것이며, 여기서 상기 CsgG 기공은 적어도 하나의 단량체를 통해, 변형된 CsgF 펩티드는 공유 결합을 통해 커플링된다. 상기 공유 연결 또는 결합은 하나의 경우에 시스테인 연결을 통해 가능하며, 여기서 시스테인의 술폰드릴 측기는 또 다른 아미노산 잔기 또는 모이어티와 공유적으로 연결된다. 두번째 가능성에서, 공유 결합은 비천연 (광)반응성 아미노산 사이의 상호작용을 통해 수득된다. (광)반응성 아미노산은 단백질 복합체의 가교에 사용될 수 있는 천연 아미노산의 인공 유사체를 언급하며, 생체 내 또는 시험관 내에서 단백질 및 펩티드에 혼입될 수 있다. 일반적으로 사용되는 광반응성 아미노산 유사체는 류신 및 메티오닌, 및 파라-벤조일-페닐-알라닌, 뿐만 아니라 아지도호모알라닌, 호모프로파르길글리신, 호모알릴글리신, p-아세틸-Phe, p-아지도-Phe, p-프로파르길옥시-Phe 및 p-벤조일-Phe에 대한 광보호성 디아지린 유사체이다(Wang 등 2012; Chin 등 2002). 자외선에 노출 시, 이들은 활성화되고 광반응성 아미노산 유사체의 몇 옹스트롬 내에 있는 상호작용 단백질에 공유적으로 결합한다. 그러나, 상기 공유 결합이 발생할 수 있는 CsgG 단량체의 위치는 변형된 CsgF 펩티드에 대한 노출에 따라 달라진다. 여러 아미노산이 공유 결합을 제공하는 위치, 즉 서열번호: 3 또는 서열번호: 117, 또는 이의 상동체의 위치 132, 133, 136, 138, 140, 142, 144, 145, 147, 149, 151, 153, 155, 183, 185, 187, 189, 191, 201, 203, 205, 207 또는 209 of 서열번호: 3 또는 서열번호: 117에 있다.

[0139] 본 발명의 또 다른 측면은 상기 변형된 CsgF 펩티드를 포함하는 작제물에 관한 것이며, 여기서 상기 펩티드는 공유적으로 부착된다. "작제물"은 변형된 CsgF 및/또는 CsgG, 또는 이의 상동체로부터 유래된 2개 이상의 공유적으로 부착된 단량체를 포함한다. 다시 말해서, 작제물은 하나 초과와 단량체를 함유할 수 있다. 또 다른 측면에서, 본 발명은 또한 본 발명의 적어도 하나의 작제물을 포함하는 기공 복합체를 제공한다. 기공 복합체는 기공을 형성하기에 충분한 작제물 및, 필요한 경우, 단량체를 함유한다. 예를 들면, 팔량체 기공은 (a) 각각 2개

의 단량체를 포함하는 4개의 작제물, (b) 각각 4개의 단량체를 포함하는 2개의 작제물, (c) 2개의 단량체 및 작제물의 일부를 형성하지 않는 6개의 단량체를 포함하는 1개의 작제물, 또는 (d) 1개의 작제물에서 1 또는 2개의 CsgF 단량체, 및 6 내지 7개의 CsgG 단량체를 갖는 1개의 작제물 또는 심지어 (e) CsgG 단량체를 단독으로 포함하는 또 다른 작제물 외에 CsgF 및 CsgG 단량체를 갖는 작제물을 포함할 수 있다. 예를 들면 구량체 기공에 대한 동일하고 추가적인 가능성이 제공된다. 작제물 및 단량체의 다른 조합이 당업자에 의해 구상될 수 있다. 본 발명의 하나 이상의 작제물은 폴리뉴클레오티드를 서열분석하는 것과 같은 특성화를 위한 기공 복합체를 형성하는 데 사용될 수 있다. 작제물은 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개 또는 적어도 10개의 단량체를 포함할 수 있다. 작제물은 바람직하게는 2개의 단량체를 포함한다. 2개 이상의 단량체는 동일하거나 상이할 수 있거나, CsgF, CsgG, CsgG/CsgF 융합 단량체 또는 이의 상동체일 수 있거나, 이의 임의의 조합일 수 있다.

[0140] 또 다른 구현예는 본 발명의 상기 기공 또는 기공 복합체를 암호화하는 폴리뉴클레오티드 또는 핵산 분자, 또는 이의 상동체 또는 돌연변이체, 또는 위에 기재된 바와 같은 작제물을 암호화하는 폴리뉴클레오티드에 관한 것이다.

[0141] 특정 구현예는 단리된 기공 복합체 또는 본 발명의 단리된 기공 복합체, 및 막의 구성요소를 포함하는 단리된 막관통 기공 복합체에 관한 것이다. 상기 단리된 막관통 기공 복합체는 핵산 서열분석과 같은 분자 감지에 사용하기 위해 직접 적용가능하다. 대안적으로, 본 발명의 단리된 기공 복합체에 따라 본원에 기재된 바와 같은 변형된 CsgG/CsgF 생물학적 기공, 및 막, 막 구성요소, 또는 절연 층을 포함하는 막성 조성물이 제공된다. 하나의 구현예는 본 발명에 따른 단리된 기공 복합체, 및 막의 구성요소로 이루어진 단리된 막관통 기공 복합체에 관한 것이다.

[0142] CsgG:CsgF 복합체가 매우 안정하지만, CsgF가 절두된 경우, CsgG:CsgF 복합체의 안정성은 전장 CsgF를 포함하는 복합체와 비교하여 감소한다. 따라서, 예를 들어 본원에서 식별된 위치에 시스테인 잔기를 도입한 후, CsgG와 CsgF 사이에 이황화물 결합이 이루어져 복합체를 더욱 안정하게 만들 수 있다. 기공 복합체는 이전에 언급된 방법 중 임의의 것으로 이루어질 수 있고 이황화물 결합 형성은 산화제(예: 구리-오르토포넌트올린)를 사용하여 유도될 수 있다. 다른 상호작용(예: 소수성 상호작용, 전하-전하 상호작용/정전기적 상호작용)이 또한 시스테인 상호작용 대신 해당 위치에 사용될 수 있다.

[0143] 또 다른 구현예에서, 비천연 아미노산이 또한 해당 위치에 혼입될 수 있다. 이 구현예에서, 만들어진 공유 결합은 클릭 화학을 통해 이루어진다. 예를 들어, 아지드 또는 알킨을 갖거나 디벤조사이클로옥틴(DBCO) 기 및/또는 비사이클로[6.1.0]노닌(BCN) 기를 갖는 비천연 아미노산이 이러한 위치 중 하나 이상에 도입될 수 있다.

[0144] 이러한 안정화 돌연변이는 CsgG 및/또는 CsgF에 대한 임의의 다른 변형, 예를 들어 본원에 개시된 변형과 조합될 수 있다.

[0145] CsgG 기공은 CsgF 펩티드에 대한 부착을 용이하게 하도록 변형된 적어도 하나, 예컨대 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 CsgG 단량체를 포함할 수 있다. 예를 들어 시스테인 잔기는 CsgG에 대한 공유 부착을 용이하게 하기 위해 서열번호: 3 또는 서열번호: 117의 위치 132, 133, 136, 138, 140, 142, 144, 145, 147, 149, 151, 153, 155, 183, 185, 187, 189, 191, 201, 203, 205, 207 및 209에 상응하는 위치 중 하나 이상에 도입될 수 있다. 시스테인 잔기를 통한 공유 부착에 대한 대안 또는 추가로서, 기공은 소수성 상호작용 또는 정전기적 상호작용에 의해 안정화될 수 있다. 이러한 상호작용을 용이하게 하기 위해, 서열번호: 3 또는 서열번호: 117의 위치 132, 133, 136, 138, 140, 142, 144, 145, 147, 149, 151, 153, 155, 183, 185, 187, 189, 191, 201, 203, 205, 207 및 209 중 하나 이상에 상응하는 위치에 비천연 반응성 또는 광반응성 아미노산.

[0146] CsgF 펩티드는 CsgG 기공에 대한 부착을 용이하게 하도록 변형될 수 있다. 예를 들어 시스테인 잔기는 CsgG에 대한 공유 부착을 용이하게 하기 위해 서열번호: 6의 위치 1, 4, 5, 8, 9, 11, 12, 26 또는 29에 상응하는 위치 중 하나 이상에 도입될 수 있다. 시스테인 잔기를 통한 공유 부착에 대한 대안 또는 추가로서, 기공은 소수성 상호작용 또는 정전기적 상호작용에 의해 안정화될 수 있다. 이러한 상호작용을 용이하게 하기 위해, 서열번호: 6의 위치 1, 4, 5, 8, 9, 11, 12, 26 또는 29 중 하나 이상에 상응하는 위치에 비천연 반응성 또는 광반응성 아미노산.

[0147] 바람직한 예시적인 CsgF 펩티드는 서열번호: 6에 대한 다음 돌연변이를 포함하는 것을 포함한다: N15X1/N17X2/A20X3/N24X4/A28X5/D34X6, 여기서 X1은 N/S/A/T/Q/G/L/V/I/F/Y/W/R/K/D/C이고, X2는 N/S/A/T/Q/G/L/V/I/F/Y/W/R/K/D/C이고, X3은 A/S/T/Q/N/G/L/V/I/F/Y/W/R/K/D/C이고, X4는

N/S/T/Q/A/G/L/V/I/F/Y/W/R/K/D/C이고, X5는 A/S/T/Q/N/G/L/V/I/F/Y/W/R/K/D/C이고 X5는 D/F/Y/W/R/K/N/Q/C이다. 위치 N15, N17, A20, N24 및 A28에서 돌연변이는 수축 돌연변이이고 위치 34에서 돌연변이는 pF CsgF와 CsgG 기공 바닥의 상호작용에 영향을 미쳐 상호작용을 안정화한다.

[0148] CsgG 기공

[0149] CsgG 기공은 본 발명의 동일한 돌연변이체 단량체를 포함하는 호모올리고머 기공일 수 있다. CsgG 기공은 예를 들어 본원에 개시된 바와 같은 적어도 하나의 돌연변이체 단량체를 포함하는 CsgG로부터 유래된 이종올리고머 기공일 수 있다.

[0150] CsgG 기공은 임의의 수의 돌연변이체 단량체를 함유할 수 있다. 기공은 전형적으로 적어도 7개, 적어도 8개, 적어도 9개 또는 적어도 10개의 동일한 돌연변이체 단량체, 예컨대 7, 8, 9 또는 10개의 돌연변이체 단량체를 포함한다. CsgG 기공은 바람직하게는 8 또는 9개의 동일한 돌연변이체 단량체를 포함한다.

[0151] 바람직한 구현예에서, 이종올리고머 CsgG 기공의 모든 단량체(예컨대 10, 9, 8 또는 7개의 단량체)는 본원에 개시된 바와 같은 돌연변이체 단량체이며, 여기서 이들 중 적어도 하나는 다른 것과 상이하다. 이들은 모두 서로 상이할 수 있다.

[0152] CsgG 기공의 돌연변이체 단량체는 바람직하게는 모두 대략 동일한 길이이거나 동일한 길이이다. 기공 내 본 발명의 돌연변이 단량체의 배열은 바람직하게는 대략 동일한 길이이거나 동일한 길이이다. 길이는 아미노산의 수 및/또는 길이의 단위로 측정될 수 있다.

[0153] 돌연변이체 단량체는 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서 변형을 포함하는 서열번호: 3 또는 서열번호: 117의 변이체일 수 있다. 서열번호: 3 또는 서열번호: 117의 아미노산 서열의 전체 길이에 걸쳐, 변이체는 바람직하게는 아미노산 동일성을 기준으로 해당 서열에 대해 적어도 40% 상동성일 것이다. 보다 바람직하게는, 변이체는 전체 서열에 걸쳐 서열번호: 3 또는 서열번호: 117의 아미노산 서열에 대한 아미노산 동일성을 기준으로 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 보다 바람직하게는 적어도 95%, 97% 또는 99% 상동성일 수 있다. 서열번호: 3 또는 서열번호: 117의 아미노산 서열의 전체 길이에 걸쳐, 변이체는 바람직하게는 해당 서열과 적어도 40% 동일할 것이다. 보다 바람직하게는, 변이체는 전체 서열에 걸쳐 서열번호: 3 또는 서열번호: 117과 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 및 보다 바람직하게는 적어도 95%, 97% 또는 99% 동일할 수 있다. 100개 이상, 예를 들어 125, 150, 175 또는 200개 이상의 연속 아미노산의 스트레치에 걸쳐 적어도 80%, 예를 들어 적어도 85%, 90% 또는 95%, 아미노산 동일성이 있을 수 있다("정질 상동성").

[0154] CsgG 단량체는 고도로 보존되어 있다(W02017/149317의 도면 45 내지 47에서 용이하게 이해될 수 있음). 더욱이, 서열번호: 3 또는 서열번호: 117에 관한 돌연변이에 대한 지식으로부터 서열번호: 3 또는 서열번호: 117 이외의 CsgG 단량체의 돌연변이에 대한 동등한 위치를 결정하는 것이 가능하다.

[0155] 따라서 서열번호: 3 또는 서열번호: 117에 제시된 바와 같은 서열의 변이체 및 청구범위와 명세서의 다른 곳에 제시된 바와 같은 이의 특이적 아미노산 돌연변이를 포함하는 돌연변이체 CsgG 단량체에 대한 언급은 또한 서열번호: 68 내지 88에 제시된 바와 같은 서열의 변이체 및 이의 상응하는 아미노산 돌연변이를 포함하는 돌연변이체 CsgG 단량체를 포괄한다. 마찬가지로 서열번호: 3 또는 서열번호: 117에 제시된 바와 같은 서열의 변이체 및 청구범위와 명세서의 다른 곳에 제시된 바와 같은 이의 특이적 아미노산 변이체를 포함하는 돌연변이체 CsgG 단량체에 관한 기공의 사용을 수반하는 작제물, 기공 또는 방법에 대한 언급은 또한 상기 개시된 서열번호에 따른 서열의 변이체 및 이의 상응하는 아미노산 돌연변이를 포함하는 돌연변이체 CsgG 단량체에 관한 작제물, 기공 또는 방법을 포괄한다. 만약 본 발명은 고도로 보존된 영역을 보여주는 명세서에서 명시적으로 식별되지 않은 다른 변이체 CsgG 단량체로 확장된다는 것이 추가로 이해될 것이다.

[0156] 당업계의 표준 방법은 상동성을 결정하는 데 사용될 수 있다. 예를 들어 UWGCG 패키지는 예를 들어 기본 설정으로 사용되는 상동성을 계산하는 데 사용될 수 있는 BESTFIT 프로그램을 제공한다(Devereux 등 (1984) Nucleic Acids Research 12, p387-395). PILEUP 및 BLAST 알고리즘은 예를 들어 Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S.F 등 (1990) J Mol Biol 215:403-10에 기재된 바와 같이 상동성을 계산하거나 또는 서열을 정렬하는 데 사용될 수 있다(예컨대 (전형적으로 그들의 기본 설정으로) 동등한 잔기 또는 상응하는 서열 식별). BLAST 분석을 수행하기 위한 소프트웨어는 미국 국립생물공학정보센터(National Center for Biotechnology Information)를 통해 공개적으로 이용가능하다(<http://www.ncbi.nlm.nih.gov/>).

- [0157] 서열번호: 3은 에스케리키아 콜라이 Str. K-12 substr. MC4100의 야생형 CsgG 단량체이다. 서열번호: 3 또는 서열번호: 117의 변이체는 또 다른 CsgG 상동체에 존재하는 임의의 치환을 포함할 수 있다. 바람직한 CsgG 상동체는 서열번호: 68 내지 88에 제시되어 있다. 변이체는 서열번호: 3 또는 서열번호: 117과 비교하여 서열번호: 68 내지 88에 존재하는 치환 중 하나 이상의 조합을 포함할 수 있다. 예를 들어, 돌연변이는 서열번호: 3 또는 서열번호: 117과 서열번호: 68 내지 88 중 임의의 하나 사이에 상이한 서열번호: 3 또는 서열번호: 117의 위치 중 임의의 하나 이상에서 이루어질 수 있다. 이러한 돌연변이는 서열번호: 3 또는 서열번호: 117의 아미노산을 서열번호: 68 내지 88 중 임의의 하나의 상응하는 위치의 아미노산으로 치환하는 것일 수 있다. 대안적으로, 이러한 위치 중 임의의 하나에 있는 돌연변이는 임의의 아미노산으로의 치환일 수 있거나, 1 내지 10개 아미노산, 예컨대 2 내지 8개 또는 3 내지 6개 아미노산의 결실 또는 삽입과 같은 결실 또는 삽입 돌연변이일 수 있다. 본원에 개시된 돌연변이 외에도, 서열번호: 3 또는 서열번호: 117과 모든 서열번호: 66 내지 88 사이에 보존된 아미노산은 바람직하게는 본 발명의 변이체에 존재한다. 그러나, 보존적 돌연변이는 서열번호: 3 또는 서열번호: 117과 모든 서열번호: 66 내지 88 사이에 보존된 이러한 위치 중 임의의 하나 이상에서 이루어질 수 있다.
- [0158] 본 발명은 서열번호: 3 또는 서열번호: 117의 특이적 위치에 상응하는 CsgG 단량체 구조의 위치에서 서열번호: 3 또는 서열번호: 117의 특이적 위치로 치환될 때 본원에 기재된 아미노산 중 임의의 하나 이상을 포함하는 기공 형성 CsgG 돌연변이체 단량체를 제공한다. 상응하는 위치는 당업계의 표준 기술에 의해 결정될 수 있다. 예를 들어, 상기 언급된 PILEUP 및 BLAST 알고리즘을 사용하여 CsgG 단량체의 서열을 서열번호: 3 또는 서열번호: 117과 정렬하고 따라서 상응하는 잔기를 식별할 수 있다.
- [0159] 기공 형성 돌연변이체 단량체는 전형적으로 야생형 CsgG 단량체와 동일한 3D 구조, 예컨대 서열번호: 3 또는 서열번호: 117의 서열을 갖는 CsgG 단량체와 동일한 3D 구조를 형성하는 능력을 유지한다. CsgG의 3D 구조는 당업계에 알려져 있고 예를 들어, Goyal 등 (2014) Nature 516(7530):250-3에 개시되어 있다. 임의의 수의 돌연변이가 본원에 기재된 돌연변이 외에도 야생형 CsgG 서열에서 이루어질 수 있으며 단 CsgG 돌연변이체 단량체는 본 발명의 돌연변이에 의해 부여된 개선된 특성을 유지한다.
- [0160] 전형적으로 CsgG 단량체는 3개의 알파 나선 및 5개의 베타 시트를 포함하는 구조를 형성하는 능력을 유지할 것이다. 돌연변이는 막관통 기공을 형성하는 CsgG 단량체의 능력에 영향을 미치지 않으면서, 적어도 첫번째 알파 나선(서열번호:3의 S63에서 시작)에 N-말단인 CsgG 영역, 두번째 알파 나선(서열번호: 3 또는 서열번호: 117의 G85에서 A99까지), 두번째 알파 나선과 첫번째 베타 시트 사이의 루프(서열번호: 3 또는 서열번호: 117의 Q100에서 N120까지), 네번째 및 다섯번째 베타 시트(서열번호: 3 또는 서열번호: 117의 각각 S173 내지 R192 및 R198 내지 T107) 및 네번째와 다섯번째 베타 시트 사이의 루프(서열번호: 3 또는 서열번호: 117의 F193 내지 Q197)에서 이루어질 수 있으며, 막관통 기공은 폴리펩티드를 전위시킬 수 있다. 따라서, 폴리뉴클레오티드를 전위시킬 수 있는 기공을 형성하는 단량체의 능력에 영향을 미치지 않으면서 임의의 CsgG 단량체의 임의의 이러한 영역에서 추가 돌연변이가 이루어질 수 있는 것으로 생각된다. 또한 폴리뉴클레오티드를 전위시킬 수 있는 기공을 형성하는 단량체의 능력에 영향을 미치지 않으면서 임의의 알파 나선(서열번호: 3 또는 서열번호: 117의 S63 내지 R76, G85 내지 A99 또는 V211 내지 L236) 또는 임의의 베타 시트(서열번호: 3 또는 서열번호: 117의 I121 내지 N133, K135 내지 R142, I146 내지 R162, S173 내지 R192 또는 R198 내지 T107)와 같은 다른 영역에서 돌연변이가 이루어질 수 있는 것으로 예상된다. 또한 폴리뉴클레오티드를 전위시킬 수 있는 기공을 형성하는 단량체의 능력에 영향을 미치지 않으면서 알파 나선 및 베타 시트를 연결하는 임의의 루프 영역 및/또는 CsgG 단량체의 N-말단 및/또는 C-말단 영역에서 하나 이상의 아미노산의 결실이 이루어질 수 있는 것으로 예상된다.
- [0161] 아미노산 치환은 위에 논의된 것 외에도, 예를 들어 최대 1, 2, 3, 4, 5, 10, 20 또는 30개의 치환이 서열번호: 3 또는 서열번호: 117의 아미노산 서열에 대해 이루어질 수 있다. 보존적 치환은 아미노산을 유사한 화학적 구조, 유사한 화학적 특성 또는 유사한 측쇄 부피의 다른 아미노산으로 대체한다. 도입된 아미노산은 이들이 대체하는 아미노산에 대해 유사한 극성, 친수성, 소수성, 염기성, 산성, 중성 또는 전하를 가질 수 있다. 대안적으로, 보존적 치환은 기존의 방향족 또는 지방족 아미노산 대신 방향족 또는 지방족인 또 다른 아미노산을 도입할 수 있다. 보존적 아미노산 변화는 당업계에 잘 알려져 있고 위의 표 1에 정의된 바와 같은 20개의 주요 아미노산의 특성에 따라 선택될 수 있다. 아미노산이 유사한 극성을 갖는 경우, 이는 또한 표 2의 아미노산 측쇄에 대한 소수성 척도를 참조하여 결정될 수 있다.
- [0162] 서열번호: 3 또는 서열번호: 117의 아미노산 서열의 하나 이상의 아미노산 잔기는 추가적으로 위에 기재된 폴리펩티드로부터 결실될 수 있다. 최대 1, 2, 3, 4, 5, 10, 20 또는 30개 이상의 잔기가 결실될 수 있다.
- [0163] 변이체는 서열번호: 3 또는 서열번호: 117의 단편을 포함할 수 있다. 이러한 단편은 기공 형성 활성을

유지한다. 단편은 적어도 50개, 적어도 100개, 적어도 150개, 적어도 200개 또는 적어도 250개 아미노산 길이일 수 있다. 이러한 단편은 기공을 생산하는 데 사용될 수 있다. 단편은 바람직하게는 서열번호: 3 또는 서열번호: 117의 막 스페닝 도메인, 즉 K135-Q153 및 S183-S208을 포함한다.

[0164] 하나 이상의 아미노산은 대안적으로 또는 추가적으로 위에 기재된 폴리펩티드에 첨가될 수 있다. 연장은 서열번호: 3 또는 서열번호: 117의 아미노산 서열 또는 이의 폴리펩티드 변이체 또는 단편의 아미노 말단 또는 카르복시 말단에 제공될 수 있다. 연장은 예를 들어 1 내지 10개 아미노산 길이로 상당히 짧을 수 있다. 대안적으로, 연장은 예를 들어 최대 50 또는 100개 아미노산으로 더 길 수 있다. 담체 단백질은 본 발명에 따른 아미노산 서열에 융합될 수 있다. 다른 융합 단백질은 아래에 보다 상세하게 논의되어 있다.

[0165] 본원에 기재된 바와 같은 CsgG 기공은 야생형 CsgG 기공, 또는 이의 상동체 또는 돌연변이체/변이체를 포함한다. 변이체는 서열번호: 3 또는 서열번호: 117의 아미노산 서열과 다른 아미노산 서열을 갖고 기공을 형성하는 능력을 유지하는 폴리펩티드이다. 변이체는 전형적으로 기공 형성을 담당하는 서열번호: 3 또는 서열번호: 117의 영역을 함유한다. β -배럴을 함유하는 CsgG의 기공 형성 능력은 각 서브유닛의 β -시트에 의해 제공된다. 서열번호: 3 또는 서열번호: 117의 변이체는 전형적으로 β -시트를 형성하는 서열번호: 3 또는 서열번호: 117의 영역, 즉 K134-Q154 및 S183-S208을 포함한다. 하나 이상의 변형은 생성된 변이체가 기공을 형성하는 능력을 유지하는 한 β -시트를 형성하는 서열번호: 3 또는 서열번호: 117의 영역에 이루어질 수 있다. 서열번호: 3 또는 서열번호: 117의 변이체는 바람직하게는 α -나선 및/또는 루프 영역 내에 치환, 부가 또는 결실과 같은 하나 이상의 변형을 포함한다.

[0166] 돌연변이체 CsgG 단량체는 돌연변이체 CsgG 단량체일 수 있으며, 이는 서열이 야생형 CsgG 단량체의 서열과 다른 단량체이고 기공을 형성하는 능력을 유지한다. 돌연변이체 단량체는 또한 본원에서 변이체로 언급될 수 있다. 돌연변이체 단량체가 기공을 형성하는 능력을 확인하기 위한 방법은 당업계에 잘 알려져 있고 아래에 보다 상세하게 논의되어 있다.

[0167] 본 발명의 CsgG 기공 또는 기공 복합체에서 적어도 하나의 단량체, 또는 임의의 또는 모든 6 내지 10개의 단량체는 W02016/034591, W02017/149316, W02017/149317, W02017/149318, W02018/211241, 및 W02019/002893(모두 그 전문이 본원에 참조로 포함됨)에 개시된 특정 변형 또는 치환 중 임의의 것을 가질 수 있다.

[0168] 본 발명의 기공 또는 기공 복합체에서 서열번호: 117의 적어도 하나의 단량체/변이체에서의 바람직한 추가적인 변형은 다음 중 하나 이상, 예컨대 2개 이상, 3개 이상, 4개 이상, 5개 이상, 6개 이상, 7개 이상, 8개 이상 또는 전부를 포함하나 이에 제한되지 않는다:

[0169] (a) 위치 Y51에서 치환, 예컨대 Y51A, Y51I, Y51L 또는 Y51T;

[0170] (b) 위치 N55에서 치환, 예컨대 N55V, N55Q, N55S 또는 N55A;

[0171] (c) 위치 F56에서 치환, 예컨대 F56Q, F56A 또는 F56N;

[0172] (d) 위치 L90에서 치환, 예컨대 L90N, L90R 또는 L90K;

[0173] (e) 위치 N91에서 치환, 예컨대 N91N, N91R 또는 N91K;

[0174] (f) 위치 K94에서 치환, 예컨대 K94Q, K94R, K94F, K94Y, K94W, K94L, K94S 또는 K94N;

[0175] (g) 위치 R192에서 치환, 예컨대 R192D, R192Q, R192F, R192S 또는 R192T;

[0176] (h) 위치 Q153에서 치환, 예컨대 Q153C; 및

[0177] (i) 위치 C215에서 치환, 예컨대 C215T.

[0178] 서열번호: 117의 적어도 하나의 단량체/변이체는 하나 이상의 위치의 결실, 예컨대 V105-I107의 결실, F193-L199의 결실 또는 F195-L199의 결실을 추가로 포함할 수 있다.

[0179] 기공 또는 기공 복합체에서 6, 7, 8, 9 또는 10개와 같은 임의의 수의 단량체는 서열번호: 117의 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서의 변형 외에도 이러한 추가적인 변형 중 하나 이상을 추가로 포함하는 서열번호: 117의 돌연변이체 단량체/변이체일 수 있다. 기공 또는 기공 복합체에서 모든 6 내지 10개의 단량체는 서열번호: 117의 위치 W97, Q100, E101, N102, 및 T104 중 하나 이상에서의 변형 외에도 이러한 추가적인 치환 중 하나 이상을 추가로 포함하는 서열번호: 117의 돌연변이체 단량체/변이체일 수 있다.

[0180] 이중 기공 복합체

[0181] 본 발명의 기공 또는 기공 복합체는 첫번째 기공 또는 복합체 및 두번째 기공 또는 복합체를 포함하는 이중 기공 복합체일 수 있다. 이중 기공 복합체는 기공-기공, 기공-기공 복합체 또는 기공 복합체-기공 복합체를 포함할 수 있다. 기공 또는 복합체 중 임의의 것은 본 발명의 기공 또는 복합체일 수 있다. 하나의 구현예에서 첫번째 기공 복합체 및 두번째 기공 복합체는 둘 다 본 발명의 CsgG/CsgF 기공 복합체이다. 또 다른 구현예에서 첫번째 기공 및 두번째 기공은 둘 다 본 발명의 CsgG 기공이다. 첫번째 및 두번째 기공 또는 기공 복합체는 동일하거나 상이할 수 있다. 본원에 개시된 돌연변이 중 임의의 것 외에도, 이중 기공에서, 적어도 하나의 CsgG 단량체는 W02016/034591, W02017/149316, W02017/149317 및, W02017/149318, W02018/211241, 및 W02019/002893 (모두 그 전문이 본원에 참조로 포함됨)에 기재된 추가적인 돌연변이 중 하나 이상을 포함할 수 있다. 적어도 하나의 CsgG 단량체는 바람직하게는 위에 개시된 추가적인 치환 중 임의의 것을 포함한다.

[0182] 변형된 단백질을 제조하는 방법

[0183] 비자연 발생 아미노산을 도입하거나 치환하는 방법이 또한 당업계에 잘 알려져 있다. 예를 들면, 비자연 발생 아미노산은 돌연변이체 단량체를 발현하는 데 사용되는 IVTT 시스템에 합성 아미노아실-tRNA를 포함함으로써 도입될 수 있다. 대안적으로, 이들은 이러한 특이적 아미노산의 합성(즉 비자연 발생) 유사체의 존재 하에 특이적 아미노산에 대한 영양요구성이 있는 이. 콜라이에서 돌연변이체 단량체를 발현함으로써 도입될 수 있다. 이들은 또한 돌연변이체 단량체가 부분적인 펩티드 합성을 사용하여 생산되는 경우 네이키드 결찰에 의해 생산될 수 있다.

[0184] CsgG로부터 유래된 단량체는 식별 또는 정제를 돕기 위해, 예를 들어 스트렙타비딘 태그를 첨가하거나 신호 서열을 첨가하여 단량체가 이러한 서열을 자연적으로 함유하지 않는 세포에서 분비를 촉진하도록 변형될 수 있다. 다른 적합한 태그는 아래에 보다 상세하게 논의되어 있다. 단량체는 노출 라벨로 라벨링될 수 있다. 노출 라벨은 단량체 검출을 허용하는 임의의 적합한 라벨일 수 있다. 적합한 라벨은 아래에 기재되어 있다.

[0185] CsgG로부터 유래된 단량체는 또한 D-아미노산을 사용하여 생산될 수 있다. 예를 들면, CsgG로부터 유래된 단량체는 L-아미노산 및 D-아미노산의 혼합물을 포함할 수 있다. 이는 이러한 단백질 또는 펩티드를 생산하기 위한 기술 분야에서 통상적이다.

[0186] CsgG로부터 유래된 단량체는 뉴클레오티드 구별을 용이하게 하기 위한 하나 이상의 특이적 변형을 함유한다. CsgG로부터 유래된 단량체는 또한 기공 형성을 방해하지 않는 한 다른 비특이적 변형을 함유할 수 있다. 다수의 비특이적 측쇄 변형이 당업계에 알려져 있고 CsgG로부터 유래된 단량체의 측쇄에 대해 이루어질 수 있다. 이러한 변형은 예를 들어, 알데하이드와의 반응에 의한 아미노산의 환원성 알킬화 이어서 NaBH₄로의 환원, 메틸아세트이미데이트로의 아미딘화 또는 아세트산 무수물로의 아실화를 포함한다.

[0187] CsgG로부터 유래된 단량체는 당업계에 알려진 표준 방법을 사용하여 생산될 수 있다. CsgG로부터 유래된 단량체는 합성적으로 또는 재조합 수단에 의해 만들어질 수 있다. 예를 들어, 단량체는 시험관 내 번역 및 전사(IVT)에 의해 합성될 수 있다. 기공 및 단량체를 생산하는 적합한 방법은 국제 출원 WO 2010/004273, WO 2010/004265, 또는 WO 2010/086603(이들의 전문이 본원에 참조로 포함됨)에 논의되어 있다. 기공을 막에 삽입하는 방법은 알려져 있다.

[0188] 기공 내 2개 이상의 CsgG 단량체는 서로 공유적으로 부착될 수 있다. 예를 들어, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개 또는 적어도 10개의 단량체가 공유적으로 부착될 수 있다. 공유적으로 부착된 단량체는 동일하거나 상이할 수 있다.

[0189] 단량체는 임의적으로 링커를 통해 유전적으로 융합되거나, 예를 들면 화학적 가교제를 통해 화학적으로 융합될 수 있다. 단량체를 공유적으로 부착하는 방법은 W02017/149316, W02017/149317, 및 W02017/149318(이들의 전문이 본원에 참조로 포함됨)에 개시되어 있다.

[0190] 일부 구현예에서, 돌연변이체 단량체는 화학적으로 변형된다. 돌연변이체 단량체는 임의의 방식으로 임의의 부위에 화학적으로 변형될 수 있다. 돌연변이체 단량체는 바람직하게는 하나 이상의 시스테인에 분자의 부착(시스테인 연결), 하나 이상의 리신에 분자의 부착, 하나 이상의 비천연 아미노산에 분자의 부착, 에피토프의 효소 변형 또는 말단의 변형에 의해 화학적으로 변형된다. 이러한 변형을 수행하는 적합한 방법은 당업계에 잘 알려져 있다. 돌연변이체 단량체는 임의의 분자의 부착에 의해 화학적으로 변형될 수 있다. 예를 들면, 돌연변이체 단량체는 염료 또는 형광단의 부착에 의해 화학적으로 변형될 수 있다.

- [0191] 일부 구현예에서, 돌연변이체 단량체는 단량체를 포함하는 기공과 표적 뉴클레오타이드 또는 표적 폴리뉴클레오타이드 서열 사이의 상호작용을 용이하게 하는 분자 어댑터로 화학적으로 변형된다. 어댑터의 존재는 기공 및 뉴클레오타이드 또는 폴리뉴클레오타이드 서열의 호스트-게스트 화학을 개선하고 이에 의해 돌연변이체 단량체로부터 형성된 기공의 서열분석 능력을 개선한다. 호스트-게스트 화학의 원리는 당업계에 잘 알려져 있다. 어댑터는 뉴클레오타이드 또는 폴리뉴클레오타이드 서열과의 상호작용을 개선하는 기공의 물리적 또는 화학적 특성에 영향을 미친다. 어댑터는 기공의 배럴 또는 채널의 전하를 변경하거나 뉴클레오타이드 또는 폴리뉴클레오타이드 서열과 특이적으로 상호작용하거나 이에 결합하여 기공과의 상호작용을 용이하게 할 수 있다.
- [0192] 분자 어댑터는 바람직하게는 환식 분자, 사이클로텍스트린, 혼성화할 수 있는 중, DNA 결합제 또는 인터칼레이터, 펩티드 또는 펩티드 유사체, 합성 중합체, 방향족 평면 분자, 작은 양으로-하전된 분자 또는 수소 결합할 수 있는 소분자이다.
- [0193] 어댑터는 환식일 수 있다. 환식 어댑터는 바람직하게는 기공과 동일한 대칭을 갖는다. 어댑터는 바람직하게는 CsgG가 전형적으로 중심 축 주위에 8 또는 9개 서브유닛을 가지므로 8-배 또는 9-배 대칭을 갖는다. 이는 아래에 보다 상세하게 논의되어 있다.
- [0194] 어댑터는 전형적으로 호스트-게스트 화학을 통해 뉴클레오타이드 또는 폴리뉴클레오타이드 서열과 상호작용한다. 어댑터는 전형적으로 뉴클레오타이드 또는 폴리뉴클레오타이드 서열과 상호작용할 수 있다. 어댑터는 뉴클레오타이드 또는 폴리뉴클레오타이드 서열과 상호작용할 수 있는 하나 이상의 화학 기를 포함한다. 하나 이상의 화학 기는 바람직하게는 소수성 상호작용, 수소 결합, 반데르발스 힘, π -양이온 상호작용 및/또는 정전기력과 같은 비공유 상호작용에 의해 뉴클레오타이드 또는 폴리뉴클레오타이드 서열과 상호작용한다. 뉴클레오타이드 또는 폴리뉴클레오타이드 서열과 상호작용할 수 있는 하나 이상의 화학 기는 바람직하게는 양으로 하전된다. 뉴클레오타이드 또는 폴리뉴클레오타이드 서열과 상호작용할 수 있는 하나 이상의 화학 기는 보다 바람직하게는 아미노 기를 포함한다. 아미노 기는 1차, 2차 또는 3차 탄소 원자에 부착될 수 있다. 어댑터는 보다 더 바람직하게는 6, 7 또는 8개 아미노 기의 고리와 같은 아미노 기의 고리를 포함한다. 어댑터는 가장 바람직하게는 8개 아미노 기의 고리를 포함한다. 양성자화된 아미노 기의 고리는 뉴클레오타이드 또는 폴리뉴클레오타이드 서열에서 음으로 하전된 포스페이트 기와 상호작용할 수 있다.
- [0195] 기공 내에서 어댑터의 올바른 위치지정은 어댑터와 및 돌연변이체 단량체를 포함하는 기공 사이의 호스트-게스트 화학에 의해 용이하게 될 수 있다. 어댑터는 바람직하게는 기공 내 하나 이상의 아미노산과 상호작용할 수 있는 하나 이상의 화학 기를 포함한다. 어댑터는 보다 바람직하게는 소수성 상호작용, 수소 결합, 반데르발스 힘, π -양이온 상호작용 및/또는 정전기력과 같은 비공유 상호작용을 통해 기공 내 하나 이상의 아미노산과 상호작용할 수 있는 하나 이상의 화학 기를 포함한다. 기공 내 하나 이상의 아미노산과 상호작용할 수 있는 화학 기는 전형적으로 하이드록실 또는 아민이다. 하이드록실 기는 1차, 2차 또는 3차 탄소 원자에 부착될 수 있다. 하이드록실 기는 기공 내 비하전된 아미노산과 수소 결합을 형성할 수 있다. 기공과 뉴클레오타이드 또는 폴리뉴클레오타이드 서열 사이의 상호작용을 용이하게 하는 임의의 어댑터가 사용될 수 있다.
- [0196] 적합한 어댑터는 사이클로텍스트린, 환식 펩티드 및 쿠쿠르비투릴을 포함하나 이에 제한되지 않는다. 어댑터는 바람직하게는 사이클로텍스트린 또는 이의 유도체이다. 사이클로텍스트린 또는 이의 유도체는 Eliseev, A. V., and Schneider, H-J.(1994) J. Am. Chem. Soc. 116, 6081-6088에 개시된 것 중 임의의 것일 수 있다. 어댑터는 보다 바람직하게는 헵타키스-6-아미노- β -사이클로텍스트린(am7- β CD), 6-모노데옥시-6-모노아미노- β -사이클로텍스트린(am1- β CD) 또는 헵타키스-(6-데옥시-6-구아니디노)-사이클로텍스트린(gu7- β CD)이다. gu7- β CD의 구아니디노 기는 am7- β CD의 1차 아민보다 훨씬 더 높은 pKa를 가지므로 더 많이 양으로 하전된다. 이 gu7- β CD 어댑터는 기공에서 뉴클레오타이드의 체류 시간을 증가시키고, 측정된 잔류 전류의 정확도를 증가시킬 뿐만 아니라, 고온 또는 낮은 데이터 획득 속도로 염기 검출물을 증가시키는 데 사용될 수 있다.
- [0197] 숙신이미딜 3-(2-피리딜디티오)프로피오네이트(SPDP) 가교제가 아래에 보다 상세하게 논의된 바와 같이 사용되는 경우, 어댑터는 바람직하게는 헵타키스(6-데옥시-6-아미노)-6-N-모노(2-피리딜)디티오프로파노일- β -사이클로텍스트린(am6amPDP1- β CD)이다.
- [0198] 보다 적합한 어댑터는 9개의 당 단위를 포함하는(따라서 9-배 대칭을 갖는) γ -사이클로텍스트린을 포함한다. γ -사이클로텍스트린은 링커 분자를 함유할 수 있거나 위에서 논의된 β -사이클로텍스트린 예에서 사용되는 모든 또는 더 많은 변형된 당 단위를 포함하도록 변형될 수 있다.
- [0199] 분자 어댑터는 돌연변이체 단량체에 공유적으로 부착될 수 있다. 어댑터는 당업계에 알려진 임의의 방법을 사용

하여 기공에 공유적으로 부착될 수 있다. 어댑터는 전형적으로 화학적 연결을 통해 부착된다. 분자 어댑터가 시스템 연결을 통해 부착되는 경우, 하나 이상의 시스템은 바람직하게는 예를 들면 배럴에서 치환에 의해 돌연변이체에 도입되었다. 돌연변이체 단량체는 돌연변이체 단량체의 하나 이상의 시스템에 분자 어댑터를 부착 시킴으로써 화학적으로 변형될 수 있다. 하나 이상의 시스템은 자연 발생할 수 있으며, 즉 서열번호: 3 또는 서열번호: 117의 위치 1 및/또는 215에 있을 수 있다. 대안적으로, 돌연변이체 단량체는 다른 위치에 도입된 하나 이상의 시스템에 분자를 부착시킴으로써 화학적으로 변형될 수 있다. 위치 215의 시스템은 예를 들면 치환에 의해 제거되어 분자 어댑터가 위치 1의 시스템 또는 또 다른 위치에 도입된 시스템 이외의 해당 위치에 부착되지 않는다는 것을 보장할 수 있다.

[0200] 시스템 잔기의 반응성은 인접한 잔기의 변형에 의해 향상될 수 있다. 예를 들면, 측면 아르기닌, 히스티딘 또는 리신 잔기의 염기성 기는 시스템 티올 기의 pKa를 더 반응성 S- 기의 pKa로 바꿀 것이다. 시스템 잔기의 반응성은 dTNB와 같은 티올 보호 기에 의해 보호될 수 있다. 이들은 링커가 부착되기 전에 돌연변이체 단량체의 하나 이상의 시스템 잔기와 반응할 수 있다.

[0201] 분자는 돌연변이체 단량체에 직접 부착될 수 있다. 분자는 바람직하게는 화학적 가교제 또는 펩티드 링커와 같은 링커를 사용하여 돌연변이체 단량체에 부착된다.

[0202] 적합한 화학적 가교제는 당업계에서 잘 알려져 있다. 바람직한 가교제는 2,5-디옥소피롤리딘-1-일 3-(피리딘-2-일 디술폰닐)프로파노에이트, 2,5-디옥소피롤리딘-1-일 4-(피리딘-2-일 디술폰닐)부타노에이트 및 2,5-디옥소피롤리딘-1-일 8-(피리딘-2-일 디술폰닐)옥타노에이트를 포함한다. 가장 바람직한 가교제는 숙신이미딜 3-(2-피리딜 디티오)프로피오네이트(SPDP)이다. 전형적으로, 분자는 분자/가교제 복합체가 돌연변이체 단량체에 공유적으로 부착되기 전에 이작용성 가교제에 공유적으로 부착되지만 이작용성 가교제/단량체 복합체가 분자에 부착되기 전에 이작용성 가교제를 단량체에 공유적으로 부착하는 것이 또한 가능하다.

[0203] 링커는 바람직하게는 디티오프레이톨(DTT)에 저항성이 있다. 적합한 링커는 요오도아세트아미드-기반 및 말레이미드-기반 링커를 포함하나 이에 제한되지 않는다.

[0204] 다른 구현예에서, 단량체는 폴리뉴클레오타이드 결합 단백질에 부착될 수 있다. 이는 본 발명의 서열분석 방법에서 사용될 수 있는 모듈식 서열분석 시스템을 형성한다. 폴리뉴클레오타이드 결합 단백질은 아래에 논의되어 있다.

[0205] 폴리뉴클레오타이드 결합 단백질은 바람직하게는 돌연변이체 단량체에 공유적으로 부착된다. 단백질은 당업계에 알려진 임의의 방법을 사용하여 단량체에 공유적으로 부착될 수 있다. 단량체 및 단백질은 화학적으로 융합되거나 유전적으로 융합될 수 있다. 단량체 및 단백질은 전체 작제물이 단일 폴리뉴클레오타이드 서열로부터 발현되는 경우 유전적으로 융합된다. 폴리뉴클레오타이드 결합 단백질에 단량체의 유전적 융합은 WO 2010/004265(이의 전문이 본원에 참조로 포함됨)에 논의되어 있다.

[0206] 폴리뉴클레오타이드 결합 단백질이 시스템 연결을 통해 부착되는 경우, 하나 이상의 시스템은 바람직하게는 치환에 의해 돌연변이체에 도입되었다. 하나 이상의 시스템은 바람직하게는 상동체 사이에서 낮은 보존성을 갖는 루프 영역에 도입되며 이는 돌연변이 또는 삽입이 허용될 수 있다는 것을 나타낸다. 따라서 이들은 폴리뉴클레오타이드 결합 단백질을 부착하는 데 적합하다. 이러한 구현예에서, 위치 251에서 자연 발생 시스템은 제거될 수 있다. 시스템 잔기의 반응성은 위에 기재된 바와 같이 변형에 의해 향상될 수 있다.

[0207] 폴리뉴클레오타이드 결합 단백질은 돌연변이체 단량체에 직접 부착되거나 하나 이상의 링커를 통해 부착될 수 있다. 분자는 WO 2010/086602(이의 전문이 본원에 참조로 포함됨)에 기재된 혼성화 링커를 사용하여 돌연변이체 단량체에 부착될 수 있다. 대안적으로, 펩티드 링커가 사용될 수 있다. 펩티드 링커는 아미노산 서열이다. 펩티드 링커의 길이, 유연성 및 친수성은 전형적으로 단량체 및 분자의 기능을 방해하지 않도록 설계된다. 바람직한 유연한 펩티드 링커는 2 내지 20개, 예컨대 4, 6, 8, 10 또는 16개의 세린 및/또는 글리신 아미노산의 스트레치이다. 보다 바람직한 유연한 링커는 (SG)1, (SG)2, (SG)3, (SG)4, (SG)5 및 (SG)8을 포함하며 여기서 S는 세린이고 G는 글리신이다. 바람직한 견고한 링커는 2 내지 30개, 예컨대 4, 6, 8, 16 또는 24개의 프롤린 아미노산의 스트레치이다. 보다 바람직한 견고한 링커는 (P)12를 포함하며 여기서 P는 프롤린이다.

[0208] 화학적 변형

[0209] 돌연변이체 CsgG 단량체 또는 CsgF 펩티드는 분자 어댑터 및 폴리뉴클레오타이드 결합 단백질로 화학적으로 변형될 수 있다.

- [0210] 분자(단량체 또는 펩티드가 화학적으로 변형된 것 함유)는 단량체 또는 펩티드에 직접 부착되거나 WO 2010/004273, WO 2010/004265 또는 WO 2010/086603(이들의 전문이 본원에 참조로 포함됨)에 개시된 바와 같이 링커를 통해 부착될 수 있다.
- [0211] CsgG 단량체 및/또는 CsgF 펩티드와 같은 본원에 기재된 임의의 단백질은 예를 들어 히스티딘 잔기(his 태그), 아스파르트산 잔기(asp 태그), 스트렙타비딘 태그, 플래그(flag) 태그, SUMO 태그, GST 태그 또는 MBP 태그의 첨가, 또는 폴리펩티드가 이러한 서열을 자연적으로 함유하지 않는 세포로부터 분비를 촉진하기 위한 신호 서열의 첨가에 의해 식별 또는 정제를 돕도록 변형될 수 있다. 유전자 태그를 도입하는 것에 대한 대안은 단백질 상의 천연 또는 조작된 위치에 태그를 화학적으로 반응시키는 것이다. 이의 예는 겔-이동 시약을 단백질 외부에서 조작된 시스템에 반응시키는 것일 것이다. 이는 용혈소 헥테로-올리고머를 분리하는 방법으로 입증되었다(Chem Biol. 1997 Jul;4(7):497-505).
- [0212] CsgG 단량체 및/또는 CsgF 펩티드와 같은 본원에 기재된 임의의 단백질은 노출 라벨로 라벨링될 수 있다. 노출 라벨은 단백질 검출을 허용하는 임의의 적합한 라벨일 수 있다. 적합한 라벨은 형광 분자, 방사성 동위원소, 예를 들어 ¹²⁵I, ³⁵S, 효소, 항체, 항원, 폴리뉴클레오타이드 및 비오틴과 같은 리간드를 포함하나 이에 제한되지 않는다.
- [0213] CsgG 단량체 및/또는 CsgF 펩티드와 같은 본원에 기재된 임의의 단백질은 합성적으로 또는 재조합 수단에 의해 제조될 수 있다. 예를 들어, 단백질은 시험관 내 번역 및 전사(IVTT)에 의해 합성될 수 있다. 단백질의 아미노산 서열은 비자연 발생 아미노산을 포함하거나 단백질의 안정성을 증가시키도록 변형될 수 있다. 단백질이 합성 수단에 의해 생산될 때, 이러한 아미노산은 생산 동안 도입될 수 있다. 단백질은 또한 합성 또는 재조합 생산 후에 변경될 수 있다.
- [0214] 단백질은 또한 D-아미노산을 사용하여 생산될 수 있다. 예를 들면, 단백질은 L-아미노산 및 D-아미노산의 혼합물을 포함할 수 있다. 이는 이러한 단백질 또는 펩티드를 생산하기 위한 기술 분야에서 통상적이다.
- [0215] 단백질은 또한 단백질의 기능을 방해하지 않는 한 다른 비특이적 변형을 함유할 수 있다. 다수의 비특이적 측쇄 변형은 당업계에 알려져 있고 단백질(들)의 측쇄에 대해 이루어질 수 있다. 이러한 변형은 예를 들어, 알데하이드와의 반응에 의한 아미노산의 환원성 아킬화 이어서 NaBH₄로의 환원, 메틸아세티미데이트로의 아미딘화 또는 아세트산 무수물로의 아실화를 포함한다.
- [0216] CsgG 단량체 및/또는 CsgF 펩티드와 같은 본원에 기재된 임의의 단백질은 당업계에 알려진 표준 방법을 사용하여 생산될 수 있다. 단백질을 암호화하는 폴리뉴클레오타이드 서열은 당업계의 표준 방법을 사용하여 유래 및 복제될 수 있다. 단백질을 암호화하는 폴리뉴클레오타이드 서열은 당업계의 표준 기술을 사용하여 박테리아 숙주 세포에서 발현될 수 있다. 단백질은 재조합 발현 벡터로부터 폴리펩티드의 제자리(in situ) 발현에 의해 세포에서 생산될 수 있다. 발현 벡터는 임의적으로 폴리펩티드의 발현을 제어하기 위해 유도성 프로모터를 운반한다. 이러한 방법은 Sambrook, J. and Russell, D.(2001). Molecular Cloning: A Laboratory Manual, 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY에 기재되어 있다.
- [0217] 단백질은 단백질 생산 유기체로부터 임의의 단백질 액체 크로마토그래피 시스템에 의한 정제 후 또는 재조합 발현 후 대규모로 생산될 수 있다. 전형적인 단백질 액체 크로마토그래피 시스템은 FPLC, AKTA 시스템, Bio-Cad 시스템, Bio-Rad BioLogic 시스템 및 Gilson HPLC 시스템을 포함한다.
- [0218] 기공을 생산하는 방법
- [0219] 본 발명은 생체 내 및 시험관 내에서 2개 이상의 수축 부위를 보유하는 CsgG: 변형된 CsgF 기공 복합체를 생산하는 방법을 제공한다. 하나의 구현예는 공동 발현을 통해 CsgG 기공, 또는 이의 상동체 또는 돌연변이체 형태, 및 변형된 CsgF 펩티드, 또는 이의 상동체 또는 돌연변이체를 포함하는, 막관통 기공 복합체를 생산하는 방법을 제공한다. 적합한 숙주 세포에서 CsgG 단량체(서열번호: 2에 제공된 전단백질, 또는 이의 상동체 또는 돌연변이체로 표현됨)를 발현하고, 변형된 또는 절단된 CsgF 단량체를 발현시켜, 생체 내 복합 기공 형성을 허용하는 단계를 포함하는 상기 방법. 상기 복합체는 CsgG 기공과의 복합체를 이루는 변형된 CsgF 펩티드를 포함하여 기공에 추가적인 관독 헤드를 제공한다. 변형된 CsgF 펩티드를 사용하여 상기 방법에 의해 생산된 생성된 기공 복합체는 핵산 서열분석과 같은 표적 분석물의 특성화에서 기공 복합체를 사용하기에 충분한 구조를 제공하는 데, 이는 분석물, 특히 폴리뉴클레오타이드 가닥의 통과를 허용하고, 상기 적용을 위한 적절한 설정으로 사용될 때 상기 폴리뉴클레오타이드 서열의 개선된 관독을 위한 2개 이상의 관독 헤드를 포함하기 때문이다. 본 발명의 기공 및 복합체를 제조하는 방법 및 이들을 태그하는 방식은 WO2016/034591, WO2017/149316, WO2017/149317 및,

W02017/149318, W02018/211241, 및 W02019/002893(모두 그 전문이 본원에 참조로 포함됨)에 개시되어 있다.

[0220] 분석물을 특성화하는 방법

[0221] 본 발명은 표적 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 방법을 제공한다. 방법은 표적 분석물을 본 발명의 기공과 같은 단리된 기공 복합체, 또는 막관통 기공과 접촉시켜, 표적 분석물이 기공 채널과 관련하여, 예컨대 기공 채널 내에 또는 기공 채널을 통해 이동하고 분석물이 기공과 관련하여 이동할 때 하나 이상의 측정을 수행하여 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 단계를 수반한다. 표적 분석물은 또한 주형 분석물 또는 관심 분석물이라고 할 수 있다. 단리된 기공 복합체는 전형적으로 적어도 7개, 적어도 8개, 적어도 9개 또는 적어도 10개의 단량체, 예컨대 7, 8, 9 또는 10개의 CsgG 단량체를 포함한다. 단리된 기공 복합체는 바람직하게는 8 또는 9개의 동일한 CsgG 단량체를 포함한다. 하나 이상, 예컨대 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 CsgG 단량체는 바람직하게는 화학적으로 변형되거나, CsgF 펩티드는 화학적으로 변형된다. 단리된 기공 복합체 단량체, 예컨대 CsgG 단량체, 또는 이의 상동체 또는 돌연변이체, 및 변형된 CsgF 단량체, 또는 이의 상동체 또는 돌연변이체는 임의의 유기체로부터 유래될 수 있다. 분석물은 CsgG 수축을 통해 통과한 후, CsgF 수축을 통해 통과할 수 있다. 대안적인 구현예에서 분석물은 막에서 CsgG/CsgF 복합체의 방향에 따라, CsgF 수축을 통해 통과한 후, CsgG 수축을 통해 통과할 수 있다.

[0222] 방법은 표적 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하기 위한 것이다. 방법은 적어도 하나의 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하기 위한 것일 수 있다. 방법은 2개 이상의 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 것과 관련될 수 있다. 방법은 임의의 수의 분석물, 예컨대 2, 5, 10, 15, 20, 30, 40, 50, 100개 이상의 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 것을 포함할 수 있다. 1, 2, 3, 4, 5, 10개 이상의 특성과 같이 하나 이상의 분석물의 임의의 수의 특성이 결정될 수 있다.

[0223] 기공 복합체의 채널, 또는 채널의 어느 한 개구 부근에서 분자의 결합은 기공을 통해 개방 채널 이온 흐름에 영향을 미칠 것이며, 이는 기공 채널의 "분자 감지"에 필수적이다. 핵산 서열분석 적용과 유사한 방식으로, 개방 채널 이온 흐름의 변동은 전류의 변화에 의해 적합한 측정 기술을 사용하여 측정될 수 있다(예를 들어, W0 2000/28312 및 D. Stoddart 등, Proc. Natl. Acad. Sci., 2010, 106, 7702-7 또는 W0 2009/077734; 모두 이들의 전문이 본원에 참조로 포함됨). 전류 감소에 의해 측정된 바와 같이 이온 흐름의 감소 정도는 기공 내, 또는 부근의 장애물 크기와 관련된다. 따라서 기공 내 또는 근처에서 "분석물"로도 언급되는 관심 분자의 결합은 검출가능하고 측정가능한 이벤트를 제공하여, "생물학적 센서"의 기초를 형성한다. 나노기공 감지에 적합한 분자는 핵산; 단백질; 펩티드; 이당류 및 소분자(여기서 저분자량(예를 들어, < 900Da 또는 < 500Da) 유기 또는 무기 화합물을 지칭함) 예컨대 의약품, 독소, 사이토카인, 및 오염물질을 포함한다. 생물학적 분자의 존재를 검출하는 것은 개인화된 약물 개발, 의학, 진단, 생명 과학 연구, 환경적 모니터링, 및 보안 및/또는 방위 산업에 적용된다.

[0224] 또 다른 측면에서, 야생형 또는 변형된 이. 콜라이 CsgG 나노기공, 또는 이의 상동체 또는 돌연변이체, 및 복합체 내의 기공에 채널 수축을 제공하는 변형된 CsgF 펩티드를 함유하는 단리된 기공 복합체, 또는 막관통 기공 복합체는 분자 또는 생물학적 센서로서 역할을 할 수 있다. 일부 구현예에서, CsgG 나노기공은 박테리아 단백질(예를 들어, 이. 콜라이, 살모넬라 타이피(Salmonella typhi))로부터 유래되거나 단리될 수 있다. 일부 구현예에서, CsgG 나노기공은 재조합적으로 생산될 수 있다. 분석물 검출을 위한 절차는 Howorka 등 Nature Biotechnology(2012) Jun 7; 30(6):506-7에 기재되어 있다. 검출될 분석물 분자는 채널의 어느 한 면, 또는 채널 자체의 루멘 내에 결합될 수 있다. 결합 위치는 감지될 분자의 크기에 의해 결정될 수 있다.

[0225] 표적 분석물은 바람직하게는 금속 이온, 무기 염, 중합체, 아미노산, 펩티드, 폴리펩티드, 단백질, 뉴클레오타이드, 올리고뉴클레오타이드, 폴리뉴클레오타이드, 이당류, 염료, 표백제, 의약품, 진단제, 기본 전환 약물, 폭발물, 독성 화합물, 또는 환경적 오염물질이다. 방법은 2개 이상의 단백질, 2개 이상의 뉴클레오타이드 또는 2개 이상의 의약품과 같이 동일한 유형의 2개 이상의 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 것에 관한 것일 수 있다. 대안적으로, 방법은 하나 이상의 단백질, 하나 이상의 뉴클레오타이드 및 하나 이상의 의약품과 같이 상이한 유형의 2개 이상의 분석물의 존재, 부재 또는 하나 이상의 특성을 결정하는 것에 관한 것일 수 있다.

[0226] 표적 분석물은 세포로부터 분비될 수 있다. 대안적으로, 표적 분석물은 세포 내부에 존재하는 분석물일 수 있으므로 분석물은 방법을 수행하기 전에 세포로부터 추출되어야 한다.

[0227] 야생형 기공은 센서로서 작용할 수 있지만, 종종 감지될 분자의 결합 강도, 결합 위치, 또는 결합 특이성을 증가시키기 위한 재조합 또는 화학적 방법을 통해 변형된다. 전형적인 변형은 감지될 분자의 구조에 상보적인 특

이적 결합 모이어티의 첨가를 포함한다. 분석물 분자가 핵산을 포함하는 경우, 이 결합 모이어티는 사이클로텍스트린 또는 올리고뉴클레오타이드를 포함할 수 있거나; 소분자의 경우 이는 알려진 상보적 결합 영역, 예를 들어 단일쇄 가변 단편(scFv) 영역 또는 T-세포 수용체(TCR)의 항원 인식 도메인을 포함하여, 항체 또는 비항체 분자의 항원 결합 부분일 수 있거나; 단백질의 경우, 표적 단백질의 알려진 리간드일 수 있다. 이 방식으로 야생형 또는 변형된 이. 콜라이 CsgG 나노기공, 또는 이의 상동체는 수용체, 고휘도 중앙 또는 혈액학적 암 세포(예를 들어 림프종(lymphoma) 또는 백혈병(leukaemia))의 마커, 바이러스 항원, 박테리아 항원, 원생동물 항원, 알레르겐, 알레르기 관련 분자, 알부민(예를 들어 인간, 설치류, 또는 소), 형광 분자(플루오레세인 포함), 혈액형 항원, 소분자, 약물, 효소, 효소 또는 효소 기질의 촉매 부위, 및 효소 기질의 전이 상태 유사체를 포함하여, 세포 표면 항원을 포함할 수 있는 적합한 항원(에피토프 포함)의 샘플에서 존재를 검출하기 위한 분자 센서로서 작용할 수 있게 제공될 수 있다. 위에 기재된 바와 같이, 변형은 알려진 유전자 조작 및 재조합 DNA 기술을 사용하여 달성될 수 있다. 임의의 적응에 대한 위치지정은 감지될 분자의 속성, 예를 들어, 크기, 3차원 구조, 및 생화학적 속성에 따라 달라질 것이다. 적응된 구조의 선택은 컴퓨터 구조 설계를 사용할 수 있다. 단백질-단백질 상호작용 또는 단백질-소분자 상호작용의 결정 및 최적화는 표면 플라즈몬 공명(BIAcore, Inc., 뉴저지주 피스카타웨이 소재; 또한 www.biacore.com 참조)을 사용하여 분자 상호작용을 검출하는 BIAcore®와 같은 기술을 사용하여 조사될 수 있다.

[0228] 하나의 구현예에서, 분석물은 아미노산, 펩티드, 폴리펩티드 또는 단백질이다. 아미노산, 펩티드, 폴리펩티드 또는 단백질은 자연 발생 또는 비자연 발생일 수 있다. 폴리펩티드 또는 단백질은 그들 내에 합성 또는 변형된 아미노산을 포함할 수 있다. 아미노산에 대한 여러 상이한 유형의 변형은 당업계에 알려져 있다. 적합한 아미노산 및 이의 변형은 위에 나와 있다. 표적 분석물은 당업계에서 이용가능한 임의의 방법에 의해 변형될 수 있는 것으로 이해되어야 한다.

[0229] 또 다른 구현예에서, 분석물은 핵산과 같은 폴리뉴클레오타이드이며, 이는 2개 이상의 뉴클레오타이드를 포함하는 거대분자로 정의된다. 핵산은 나노기공 서열분석에 특히 적합하다. DNA 및 RNA에서 자연 발생 핵산 염기는 물리적 크기에 의해 구별될 수 있다. 핵산 분자, 또는 개별 염기가 나노기공의 채널을 통해 통과할 때, 염기 사이의 크기 차이는 채널을 통한 이온 흐름에서 직접 관련된 감소를 야기한다. 이온 흐름의 변동은 기록될 수 있다. 이온 흐름 변동을 기록하기 위한 적합한 전기적 측정 기술은 위에 논의되어 있다. 적합한 보정을 통해, 이온 흐름의 특징적인 감소를 사용하여 실시간으로 채널을 가로지르는 특정 뉴클레오타이드 및 관련 염기를 식별할 수 있다. 전형적인 나노기공 핵산 서열분석에서, 관심 핵 서열의 개별 뉴클레오타이드가 뉴클레오타이드에 의한 채널의 부분적 차단으로 인해 나노기공의 채널을 통해 순차적으로 통과할 때 개방 채널 이온 흐름이 감소된다. 이온 흐름의 이러한 감소는 위에 기재된 적합한 기록 기술을 사용하여 측정된다. 이온 흐름의 감소는 채널을 통해 알려진 뉴클레오타이드에 대한 측정된 이온 흐름의 감소를 보정하여 뉴클레오타이드가 채널을 통해 통과하는 것을 결정하는 수단을 초래하며, 따라서, 순차적으로 수행될 때, 나노기공을 통해 통과하는 핵산의 뉴클레오타이드 서열을 결정하는 방식일 수 있다. 개별 뉴클레오타이드의 정확한 결정을 위해, 전형적으로 채널을 통한 이온 흐름의 감소가 수축(또는 "판독 헤드")을 통해 통과하는 개별 뉴클레오타이드의 크기와 직접 상관관계가 있을 필요가 있다. 예를 들어 연관된 중합효소의 작용을 통해 기공을 통해 '스레드'된 온전한 핵산 중합체에 대해 서열분석이 수행될 수 있다는 것을 알 것이다. 대안적으로, 서열은 기공에 근접한 표적 핵산으로부터 순차적으로 제거된 뉴클레오타이드 트리포스페이트 염기의 통로에 의해 결정될 수 있다(예를 들어 그 전문이 본원에 참조로 포함된 WO 2014/187924 참조).

[0230] 폴리뉴클레오타이드 또는 핵산은 임의의 뉴클레오타이드의 임의의 조합을 포함할 수 있다. 뉴클레오타이드는 자연 발생 또는 인공적일 수 있다. 폴리뉴클레오타이드에서 하나 이상의 뉴클레오타이드는 산화되거나 메틸화될 수 있다. 폴리뉴클레오타이드에서 하나 이상의 뉴클레오타이드는 손상될 수 있다. 예를 들면, 폴리뉴클레오타이드는 피리미딘 이량체를 포함할 수 있다. 이러한 이량체는 전형적으로 자외선에 의한 손상과 연관되고 피부 흑색종(melanoma)의 주요 원인이다. 폴리뉴클레오타이드에서 하나 이상의 뉴클레오타이드는 예를 들면 라벨 또는 태그로 변형될 수 있으며, 이 경우 적합한 예는 당업자에게 알려져 있다. 폴리뉴클레오타이드는 하나 이상의 스페이서를 포함할 수 있다. 뉴클레오타이드는 전형적으로 핵염기, 당 및 적어도 하나의 포스페이트 기를 함유한다. 핵염기 및 당은 뉴클레오타이드를 형성한다. 핵염기는 전형적으로 헤테로환식이다. 핵염기는 퓨린 및 피리미딘, 보다 구체적으로 아데닌(A), 구아닌(G), 티민(T), 우라실(U) 및 시토신(C)을 포함하나 이에 제한되지 않는다. 당은 전형적으로 펜토스 당이다. 뉴클레오타이드 당은 리보스 및 데옥시리보스를 포함하나 이에 제한되지 않는다. 당은 바람직하게는 데옥시리보스이다. 폴리뉴클레오타이드는 바람직하게는 다음 뉴클레오타이드를 포함한다: 데옥시아데노신(dA), 데옥시우리딘(dU) 및/또는 티미딘(dT), 데옥시구아노신(dG) 및 데옥시시티딘(dC). 뉴클레오타이드는 전형적으로 리보뉴클레오타이드 또는 데옥시리보뉴클레오타이드이다. 뉴클레오타이드는 전형적으로 모노포스페이트, 디포스페이트 또는

는 트리포스페이트를 함유한다. 뉴클레오티드는 3개 초과와 포스페이트, 예컨대 4 또는 5개의 포스페이트를 포함할 수 있다. 포스페이트는 뉴클레오티드의 5' 또는 3' 쪽에 부착될 수 있다. 폴리뉴클레오티드에서 뉴클레오티드는 임의의 방식으로 서로 부착될 수 있다. 뉴클레오티드는 전형적으로 핵산에서와 같이 당 및 포스페이트기에 의해 부착된다. 뉴클레오티드는 피리미딘 이량체에서와 같이 핵염기를 통해 연결될 수 있다. 폴리뉴클레오티드는 단일 가닥 또는 이중 가닥일 수 있다. 폴리뉴클레오티드의 적어도 일부는 바람직하게는 이중 가닥이다. 폴리뉴클레오티드는 가장 바람직하게는 리보핵 핵산(RNA) 또는 데옥시리보핵산(DNA)이다. 특히, 폴리뉴클레오티드를 분석물로서 사용하는 상기 방법은 대안적으로 (i) 폴리뉴클레오티드 길이, (ii) 폴리뉴클레오티드 동일성, (iii) 폴리뉴클레오티드 서열, (iv) 폴리뉴클레오티드의 2차 구조 및 (v) 폴리뉴클레오티드가 변형되는 여부로부터 선택된 하나 이상의 특성을 결정하는 것을 포함한다.

[0231] 폴리뉴클레오티드는 임의의 길이 (i)일 수 있다. 예를 들어, 폴리뉴클레오티드는 적어도 10개, 적어도 50개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 400개 또는 적어도 500개 뉴클레오티드 또는 뉴클레오티드 쌍 길이일 수 있다. 폴리뉴클레오티드는 1000개 이상의 뉴클레오티드 또는 뉴클레오티드 쌍, 5000개 이상의 뉴클레오티드 또는 뉴클레오티드 쌍 길이 또는 100000개 이상의 뉴클레오티드 또는 뉴클레오티드 쌍 길이일 수 있다. 임의의 수의 폴리뉴클레오티드가 조사될 수 있다. 예를 들면, 방법은 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, 100개 이상의 폴리뉴클레오티드를 특성화하는 것에 관한 것일 수 있다. 2개 이상의 폴리뉴클레오티드가 특성화되면, 이들은 상이한 폴리뉴클레오티드이거나 동일한 폴리뉴클레오티드의 2가지 경우일 수 있다. 폴리뉴클레오티드는 자연 발생 또는 인공적일 수 있다. 예를 들면, 방법은 제조된 올리고뉴클레오티드의 서열을 확인하는 데 사용될 수 있다. 방법은 전형적으로 시험관 내에서 수행된다.

[0232] 뉴클레오티드는 임의의 동일성(ii)를 가질 수 있고, 아데노신 모노포스페이트(AMP), 구아노신 모노포스페이트(GMP), 티미딘 모노포스페이트(TMP), 우리딘 모노포스페이트(UMP), 5-메틸시티딘 모노포스페이트, 5-하이드록시메틸시티딘 모노포스페이트, 시티딘 모노포스페이트(CMP), 사이클릭 아데노신 모노포스페이트(cAMP), 사이클릭 구아노신 모노포스페이트(cGMP), 데옥시아데노신 모노포스페이트(dAMP), 데옥시구아노신 모노포스페이트(dGMP), 데옥시티미딘 모노포스페이트(dTMP), 데옥시우리딘 모노포스페이트(dUMP), 데옥시시티딘 모노포스페이트(dCMP) 및 데옥시메틸시티딘 모노포스페이트를 포함하나 이에 제한되지 않는다. 뉴클레오티드는 바람직하게는 AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP, dCMP 및 dUMP로부터 선택된다. 뉴클레오티드는 무염기성(즉 핵염기 결여)일 수 있다. 뉴클레오티드는 또한 핵염기 및 당이 결여될 수 있다(즉 C3 스페이서임). 뉴클레오티드의 서열(iii)은 가닥의 5'에서 3' 방향으로 폴리뉴클레오티드 군주 전반에 걸쳐 서로 부착된 다음 뉴클레오티드의 연속적인 동일성에 의해 결정된다.

[0233] CsgG 기공 및 CsgF 펩티드를 포함하는 기공은 동중중합체를 분석하는 데 특히 유용하다. 예를 들어, 기공은 동일한 2개 이상, 예컨대 적어도 3, 4, 5, 6, 7, 8, 9 또는 10개의 연속적인 뉴클레오티드를 포함하는 폴리뉴클레오티드의 서열을 결정하는 데 사용된다. 예를 들어, 기공은 polyA, polyT, polyG 및/또는 polyC 영역을 포함하는 폴리뉴클레오티드를 서열분석하는 데 사용될 수 있다.

[0234] CsgG 기공 수축은 서열번호: 3 또는 서열번호: 117의 51, 55 및 56 위치에 있는 잔기로 이루어진다. CsgG의 판독 헤드 및 이의 수축 돌연변이체는 일반적으로 날카롭다. DNA가 수축을 통해 통과할 때, 임의의 주어진 시간에 기공의 판독 헤드와 DNA의 대략 5개 염기의 상호작용은 전류 신호가 우세하다. 이러한 더 날카로운 판독 헤드는 DNA의 혼합된 서열 영역을 판독하는 데 매우 우수하지만(A, T, G 및 C가 혼합된 경우), DNA 내의 동중중합체 영역(예 polyT, polyG, polyA, polyC)이 있는 경우 신호가 평평하게 되고 정보가 부족하다. 5개의 염기가 CsgG 및 이의 수축 돌연변이체의 신호가 우세하기 때문에, 추가적인 체류 시간 정보를 사용하지 않고 5개보다 더 긴 동중중합체를 구별하기가 어렵다. 그러나, DNA가 두번째 판독 헤드를 통해 통과하는 경우, 더 많은 DNA 염기가 조합된 판독 헤드와 상호작용하여, 구별할 수 있는 동중중합체의 길이를 증가시킬 것이다.

[0235] 변형된 Dda 헬리카제

[0236] 본 발명은 변형된 Dda 헬리카제를 제공한다. 하나 이상의 특이적 변형은 아래에 보다 상세하게 논의되어 있다. 본 발명에 따른 변형은 아래에 논의된 바와 같은 하나 이상의 치환을 포함한다.

[0237] Dda 1993의 위치 55, 114, 156, 177, 210, 221, 350 및 358 중 하나 이상

[0238] 본 발명은 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 위치 중 하나 이상이 변형되거나 치환된 변형된 DNA 의존적 ATPase(Dda) 헬리카제를 제공한다. Dda 헬리카제 및 Dda 1993의 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 위치는 아래에 보다 상세하게 논의되어 있다. 위치 55,

[illegible]

(g) 및 (h); (a), (c), (d), (e), (f) 및 (g); (a), (c), (d), (e), (f) 및 (h); (a), (c), (d), (e), (g) 및 (h); (a), (c), (d), (f), (g) 및 (h); (a), (c), (e), (f), (g) 및 (h); (a), (d), (e), (f), (g) 및 (h); (b), (c), (d), (e), (f) 및 (g); (b), (c), (d), (e), (f) 및 (h); (b), (c), (d), (e), (g) 및 (h); (b), (c), (d), (f), (g) 및 (h); (b), (c), (e), (f), (g) 및 (h); (b), (d), (e), (f), (g) 및 (h); (c), (d), (e), (f), (g) 및 (h); (a), (b), (c), (d), (e), (f) 및 (g); (a), (b), (c), (d), (e), (f) 및 (h); (a), (b), (c), (d), (e), (g) 및 (h); (a), (b), (c), (d), (f), (g) 및 (h); (a), (b), (c), (e), (f), (g) 및 (h); (a), (b), (d), (e), (f), (g) 및 (h); (a), (c), (d), (e), (f), (g) 및 (h); (b), (c), (d), (e), (f), (g) 및 (h); 또는 (a), (b), (c), (d), (e), (f), (g) 및 (h).

[0239] 본 발명은 Dda 1993의 아미노산 위치 114, 177, 350 및 358에 상응하는 위치 중 하나 이상이 변형되거나 치환된 변형된 DNA 의존적 ATPase(Dda) 헬리카제를 제공한다. Dda 헬리카제 및 Dda 1993의 위치 114, 177, 350 및 358에 상응하는 위치는 아래에 보다 상세하게 논의되어 있다. 위치 114 및 177은 Dda 1993의 1A 도메인에 있다. 위치 Y350 및 K358은 Dda 1993의 타워 도메인에 있다. 본 발명의 변형된 Dda 헬리카제는 다음을 포함하여, Dda 1993의 아미노산 위치 (a) 114, (b) 177, (c) 350 및 (d) 358에 상응하는 위치의 임의의 수 및 조합에서 변형 또는 치환을 포함할 수 있다: (a); (b); (c); (d); (a) 및 (b); (a) 및 (c); (a) 및 (d); (b) 및 (c); (b) 및 (d); (c) 및 (d); (a), (b) 및 (c); (a), (b) 및 (d); (a), (c) 및 (d); (b), (c) 및 (d); 또는 (a), (b), (c) 및 (d).

[0240] Dda 1993의 아미노산 위치 55에 상응하는 위치는 바람직하게는 D, E, K, N 또는 S로 치환된다. Dda 1993의 아미노산 위치 114에 상응하는 위치는 바람직하게는 K(T55K)로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시키고 정확도를 증가시킨다(실시예 5). 이러한 치환은 또한 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정규화된 속도 분포를 감소시킨다(실시예 5).

[0241] Dda 1993의 아미노산 위치 114에 상응하는 위치는 바람직하게는 A, V, I, L, M, F, Y, W, G, P, S, T, N 또는 Q로 치환된다. Dda 1993의 아미노산 위치 114에 상응하는 위치는 바람직하게는 A, G, I, L, M, P, S, T 또는 V로 치환된다. Dda 1993의 아미노산 위치 114에 상응하는 위치는 바람직하게는 G, L, S 또는 T로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 감소시킨다(실시예 2). Dda 1993의 아미노산 위치 114에 상응하는 위치는 바람직하게는 A, I, M, P, 또는 V로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시킨다(실시예 2). Dda 1993의 아미노산 위치 114에 상응하는 위치는 바람직하게는 G(C11G)로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 감소시키고 정확도를 증가시킨다(실시예 2). Dda 1993의 아미노산 위치 114에 상응하는 위치는 바람직하게는 I 또는 P로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시키고 정확도를 감소시킨다(실시예 2). Dda 1993의 아미노산 위치 114에 상응하는 위치는 바람직하게는 G, I 또는 P로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정규화된 속도 분포를 감소시킨다(실시예 2). Dda 1993의 아미노산 위치 114에 상응하는 위치는 가장 바람직하게는 I(C114I)로 치환된다.

[0242] Dda 1993의 아미노산 위치 156에 상응하는 위치는 바람직하게는 A, E, F, G, I, L, M, P, S, V, Y, D, K 또는 N으로 치환된다. Dda 1993의 아미노산 위치 156에 상응하는 위치는 바람직하게는 F(T156F)로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시키고 정확도를 증가시킨다(실시예 5). 이 치환은 또한 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정규화된 속도 분포를 감소시킨다(실시예 5).

[0243] Dda 1993의 아미노산 위치 177에 상응하는 위치는 바람직하게는 D, E, F, G, H, I, L, M, N, Q, R, S, T, V, W 또는 Y로 치환된다. Dda 1993의 아미노산 위치 177에 상응하는 위치는 바람직하게는 F, G, S, V, W 또는 Y로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 감소시킨다(실시예 2). Dda 1993의 아미노산 위치 177에 상응하는 위치는 바람직하게는 D, E, G, H, I, L, M, N, Q, R, 또는 T로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시킨다(실시예 2). Dda 1993의 아미노산 위치 177에 상응하는 위치는 바람직하게는 F, H, I, L, M, N 또는 W로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정확도 및 정규화된 속도 분포를 감소시킨다(실시예 2). 이들은 속도에 상이한 효과를 미친다(실시예 2). Dda 1993의 아미노산 위치 177에 상응하는 위치는 바람직하게는 N(K177N)으로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정확도를 감소시키고 정규화된 속도 분포를 증가시킨다(실시예 2). Dda 1993의 아미노산 위치 177에 상응하는 위치는 가장 바람직하게는 M(K177M)으로 치환된다.

[0244] Dda 1993의 아미노산 위치 210에 상응하는 위치는 바람직하게는 D, E, K, S, N, R, H 또는 Y로 치환된다. Dda

1993의 아미노산 위치 210에 상응하는 위치는 바람직하게는 R(T210R), H(T210H) 또는 K(T210K)로 치환된다. Dda 1993의 아미노산 위치 210에 상응하는 위치는 바람직하게는 K(T210K)로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시키고 정확도를 증가시킨다(실시예 5). 이 치환은 또한 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정규화된 속도 분포를 감소시킨다(실시예 5).

[0245] Dda 1993의 아미노산 위치 221에 상응하는 위치는 바람직하게는 D, K, E, Q, R, A, H, L, T 또는 Y로 치환된다. Dda 1993의 아미노산 위치 221에 상응하는 위치는 바람직하게는 D(N221D) 또는 E(N221E)로 치환된다. Dda 1993의 아미노산 위치 221에 상응하는 위치는 바람직하게는 E(N221E)로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시키고 정확도를 증가시킨다(실시예 5). 이 치환은 또한 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정규화된 속도 분포를 감소시킨다(실시예 5).

[0246] Dda 1993의 아미노산 위치 350에 상응하는 위치는 바람직하게는 D, E, A, V, I, L, M, F, W, R, H, K, L, S, T, N 또는 Q로 치환된다. Dda 1993의 아미노산 위치 350에 상응하는 위치는 바람직하게는 I, F, W 또는 S로 치환된다. Dda 1993의 아미노산 위치 350에 상응하는 위치는 바람직하게는 I 또는 S(Y350I 또는 Y350S)로 치환된다. Dda 1993의 아미노산 위치 350에 상응하는 위치는 바람직하게는 I(Y350I)로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시키고 정확도 및 정규화된 속도 분포를 감소시킨다(실시예 3). Dda 1993의 아미노산 위치 350에 상응하는 위치는 바람직하게는 I 또는 S(Y350I 또는 Y350S)로 치환된다. 이러한 치환은 본 발명의 기공 복합체와 함께 사용될 때 실시예 4에 제시된 효과를 갖는다.

[0247] Dda 1993의 아미노산 위치 350에 상응하는 위치는 바람직하게는 A, D, E, G, K, L, N, Q, R, T, V, H 또는 M으로 치환된다. Dda 1993의 아미노산 위치 350에 상응하는 위치는 바람직하게는 D(Y350D) 또는 E(Y350E)로 치환된다. Dda 1993의 아미노산 위치 350에 상응하는 위치는 바람직하게는 E(Y350E)로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시키고 정확도를 증가시킨다(실시예 5). 이 치환은 또한 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정규화된 속도 분포를 감소시킨다(실시예 5).

[0248] Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 D, E, A, V, I, L, M, F, Y, W, R, H, L, S, T, N 또는 Q로 치환된다.

[0249] Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 E, I, L 또는 M으로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 감소시킨다(실시예 2). Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 I 또는 M으로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 감소시키고 정확도를 증가시킨다(실시예 2). Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 M(K358M)으로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 감소시키고 정확도 및 정규화된 속도 분포를 증가시킨다(실시예 2). Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 I(K358I)로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도 및 정규화된 속도 분포를 감소시키고 정확도를 증가시킨다(실시예 2). 실시예 2는 CsgF 펩티드가 없는 CsgG 기공을 사용한다.

[0250] Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 A, E, F, I, M 또는 S로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정확도를 증가시킨다(실시예 3). Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 A, E, F, I 또는 M으로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 감소시킨다(실시예 3). Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 S(K358S)로 치환된다. 이 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 증가시킨다(실시예 3). Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 A, E, I, M 또는 S로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정규화된 속도 분포를 감소시킨다(실시예 3). Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 (K358F)로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 정규화된 속도 분포를 증가시킨다(실시예 3). 실시예 3은 CsgF 펩티드를 함유하는 CsgG 기공 복합체를 사용한다.

[0251] Dda 1993의 아미노산 위치 358에 상응하는 위치는 바람직하게는 I, L 또는 Q로 치환된다. 이러한 치환은 폴리뉴클레오티드 분석물을 특성화하는 데 사용될 때 속도를 감소시키고 정확도 및 정규화된 속도 분포를 증가시킨다(실시예 4). 실시예 4는 본 발명의 기공 복합체를 사용한다.

[0252] Dda 1993의 아미노산 위치 358에 상응하는 위치는 가장 바람직하게는 I(K358I)로 치환된다.

[0253] 본 발명의 변형된 헬리카제는 아래에 논의된 변형, 돌연변이 또는 치환 중 임의의 것을 추가로 포함할 수 있다.

[0254] 본 발명에 따라 변형된 Dda 헬리카제는 서열번호: 118 내지 133 중 임의의 것일 수 있다. 서열번호: 118은 Dda 1993이다. 변형된 헬리카제는 바람직하게는 서열번호: 118 내지 133 중 임의의 것의 변이체를 포함한다. 변이체는 아래 제시된 서열번호: 118 내지 113 중 임의의 것에 대한 임의의 %의 서열 상동성/동일성을 가질 수 있다.

[0255] 아래 표 4는 본 발명에 따라 변형될 수 있는 바람직한 Dda 헬리카제를 요약한다.

Dda 상동체 (서열번호:)		서식지	Uniprot	길이	1993과의 서열 동일성 / %	D/E 대 K/R 아미노산 수	# C
Rma- DSM (서열번호: 119)	로도씨머스 마리누스(Rhodothermus marinus)	온화한 호염균, 중온성 호염균 > 65 °C	D0MKQ2	678	21	-84/+85	2
Csp (서열번호: 120)	시아노테세 종(Cyanothece sp.)(균주 ATCC 51142)	해양 세균	B1X365	496	24	-76/+76	5
Sru (서열번호: 121)	살리나박터 루버(Salinibacter ruber)	극도로 호염균, 35-45 °C	Q2S429	421	26	-78/+54	3
Sgo (서열번호: 122)	술푸리모나스 고틀란드리카(Sulfurimonas gotlandica) GD1	서식지: 열수 분출구, 연안 퇴적물	B6BJ43	500	27	-72/+64	2
Vph12B8 (서열번호: 123)	비브리오 파지 헨리에테(Vibrio phage henriette) 12B8	바닷물, 장염에서 발견되는 속주	M4MBC3	450	27	-62/+47	6
Vph (서열번호: 124)	비브리오 파지 phi-pp2	바닷물, 장염에서 발견되는 속주	I6XGX8	421	39	-55/+45	5
Aph65 (서열번호: 125)	아에로모나스 파지(Aeromonas phage) 65	담수/기수, 장염에서 발견되는 속주	E5DRP6	434	40	-57/+48	4
AphCC2 (서열번호: 126)	아에로모나스 파지 CC2	담수/기수, 장염에서 발견되는 속주	I6XH64	420	41	-53/+44	4
Cph (서열번호: 127)	크로노박터 파지(Cronobacter phage)	장내세균의 속주 구성원	K4FBD0	443	42	-59/+57	4

[0256]

127)	phage) vB CsaM GAP161						
Kph (서열번호: 128)	클렙시엘라 파지(Klebsiella phage) KP15	장내세균의 숙주 구성원	D5JF67	442	44	-59/+58	5
SphIME13 (서열번호: 129)	스테노트로포모나스 파지(Stenotrophomonas phage) IME13	토양에서 발견되는 숙주	J7HXT5	438	51	-58/+59	7
AphAc42 (서열번호: 130)	아시네토박터 파지(Acinetobacter phage) Ac42	토양에서 발견되는 숙주	E5EYE6	442	59	-53/+49	9
SphSP18 (서열번호: 131)	시겔라 파지(Shigella phage) SP18	장내세균의 숙주 구성원	E3SFA5	442	59	-55/+55	9
Yph (서열번호: 132)	예르시니아 파지(Yersinia phage) phiR1-RT	장내세균의 숙주 구성원	I7J3V8	439	64	-52/+52	7
SphS16 (서열번호: 133)	살모넬라 파지(Salmonella phage) S16	장내세균의 숙주 구성원	M1EA88	441	72	-56/+55	5
1993 (서열번호: 118)	엔테로박테리아 파지(Enterobacteria phage) T4	장내세균의 숙주 구성원	P32270	439	100	-57/+58	5

[0257]

[0258]

표 5는 서열번호: 118의 위치 40, 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 서열번호: 119 내지 133의 아미노산을 보여준다.

서열번호:									
118	T40	T55	C114	T156	K177	T210	N221	Y350	K358
119	L40	I54	V124 또는 D125	P187	Y211	A255	Q266	L378	L386
120	F98	Y112	I180 또는 D171	I223	Q238	I274	Q285	E407	E415
121	Y43	C57	V119 또는 D120	P160	H173	A207	S218	R334 또는 R343	A340
122	F56	F80	V147 또는 D148	E191	V205	T245	D256	A367	S375
123	T46	K59	I125 또는 D126	E165	V194	I234	S245	E354 또는 E256	A362
124	F33	K47	V107 또는 E108	P147	V166	L200	T211	E332	V340
125	F38	L51	V110 또는 E111	Q150	M171	Y205	S216	L339	E347
126	F25	L38	I97 또는 E98	P137	M158	Y192	S203	A326 또는 A337	E334 또는 E345
127	T42	S57	V119 또는 E120	R160	M181	F213	Q224	D351 또는 D326	Y359 또는 Y334
128	T42	S57	V119 또는 E120	I160	M181 또는 M182	M213	R224	E350	Y358
129	T40	Q55	C114	T155	K176	-	K218	Y349	I357

[0259]

			또는 D115					또는 Y347	또는 I355
130	T46	R61	C120 또는 D121	E162	M183	T216	K227	N355 또는 N353	K363 또는 K361
131	T42	M57	C116 또는 D117	Q158	M179	T212	N223	P352 또는 P355	K360 또는 K363
132	T41	T56	C115 또는 D116	E157	K178	N211	R222	T352 또는 T350	K360 또는 K358
133	T41	T56	C115 또는 D116	E157	M178	T211	Q222	N351	K359

[0260]

[0261] 본 발명의 헬리카제는 바람직하게는 다음과 같이 (a)-(h) 중 하나 이상을 포함하는 서열번호: 118의 변이체를 포함한다:

[0262] (a)

[0263] T55D, T55E, T55K, T55N 또는 T55S, 또는

[0264] T55K.

[0265] (b)

[0266] C114A, C114V, C114I, C114L, C114M, C114F, C114Y, C114W, C114G, C114P, C114S, C114T, C114N 또는 C114Q,

- [0267] C114A, C114G, C114I, C114L, C114M, C114P, C114S, C114T 또는 C114V,
- [0268] C114G, C114L, C114S 또는 C114T,
- [0269] C114A, C114I, C114M, C114P, 또는 C114V
- [0270] C11G,
- [0271] C114I 또는 C114P
- [0272] C114G, C114I 또는 C114P, 또는
- [0273] C114I;
- [0274] (c)
- [0275] T156A, T156E, T156F, T156G, T156I, T156L, T156M, T156P, T156S, T156V, T156Y, T156D, T156K 또는 T156N, 또는
- [0276] T156F;
- [0277] (d)
- [0278] K177D, K177E, K177F, K177G, K177H, K177I, K177L, K177M, K177N, K177Q, K177R, K177S, K177T, K177V, K177W 또는 K177Y,
- [0279] K177F, K177G, K177S, K177V, K177W 또는 K177Y,
- [0280] K177D, K177E, K177G, K177H, K177I, K177L, K177M, K177N, K177Q, K177R, 또는 K177T,
- [0281] K177F, K177H, K177I, K177L, K177M, K177N 또는 K177W,
- [0282] K177N, 또는
- [0283] K177M;
- [0284] (e)
- [0285] T210D, T210E, T210K, T210S, T210N, T210R, T210H 또는 T210Y,
- [0286] T210R, T210H 또는 T210Y, 또는
- [0287] T210K;
- [0288] (f)
- [0289] N221D, N221K, N221E, N221Q, N221R, N221A, N221H, N221L, N221T 또는 N221Y,
- [0290] N221D 또는 N221E, 또는
- [0291] N221E;
- [0292] (g)
- [0293] Y350D, Y350E, Y350A, Y350V, Y350I, Y350L, Y350M, Y350F, Y350W, Y350R, Y350H, Y350K, Y350L, Y350S, Y350T, Y350N 또는 Y350Q,
- [0294] Y350I 또는 Y350S,
- [0295] Y350I,
- [0296] Y350S,
- [0297] Y350I, Y350F, Y350W 또는 Y350S,
- [0298] Y350A, Y350D, Y350E, Y350G, Y350K, Y350L, Y350N, Y350Q, Y350R, Y350T, Y350V, Y350H 또는 Y350M,
- [0299] Y350D 또는 Y350E, 또는

- [0300] Y350E; 및
- [0301] (h)
- [0302] K358D, K358E, K358A, K358V, K358I, K358L, K358M, K358F, K358Y, K358W, K358R, K358H, K358L, K358S, K358T, K358N 또는 K358Q,
- [0303] K358E, K358I, K358L 또는 K358M,
- [0304] K358I 또는 K358M,
- [0305] K358M,
- [0306] K358I,
- [0307] K358A, K358E, K358F, K358I, K358M 또는 K358S,
- [0308] K358A, K358E, K358F, K358I 또는 K358M,
- [0309] K358S,
- [0310] K358A, K358E, K358I, K358M 또는 K358S,
- [0311] K358F,
- [0312] K358I, K358L 또는 K358Q, 또는
- [0313] K358I.
- [0314] 변이체는 위에 제시된 바와 같은 (a)-(h)의 임의의 조합 및 순열을 포함할 수 있다.
- [0315] 본 발명의 헬리카제는 바람직하게는 다음과 같이 (a), (b), (c) 및 (d) 중 하나 이상을 포함하는 서열번호: 118의 변이체를 포함한다:
- [0316] (a)
- [0317] C114A, C114V, C114I, C114L, C114M, C114F, C114Y, C114W, C114G, C114P, C114S, C114T, C114N 또는 C114Q,
- [0318] C114A, C114G, C114I, C114L, C114M, C114P, C114S, C114T 또는 C114V,
- [0319] C114G, C114L, C114S 또는 C114T,
- [0320] C114A, C114I, C114M, C114P, 또는 C114V
- [0321] C11G,
- [0322] C114I 또는 C114P
- [0323] C114G, C114I 또는 C114P, 또는
- [0324] C114I;
- [0325] (b)
- [0326] K177D, K177E, K177F, K177G, K177H, K177I, K177L, K177M, K177N, K177Q, K177R, K177S, K177T, K177V, K177W 또는 K177Y,
- [0327] K177F, K177G, K177S, K177V, K177W 또는 K177Y,
- [0328] K177D, K177E, K177G, K177H, K177I, K177L, K177M, K177N, K177Q, K177R, 또는 K177T,
- [0329] K177F, K177H, K177I, K177L, K177M, K177N 또는 K177W,
- [0330] K177N, 또는
- [0331] K177M;
- [0332] (c)
- [0333] Y350D, Y350E, Y350A, Y350V, Y350I, Y350L, Y350M, Y350F, Y350W, Y350R, Y350H, Y350K, Y350L, Y350S,

Y350T, Y350N 또는 Y350Q,

- [0334] Y350I 또는 Y350S,
- [0335] Y350I, 또는
- [0336] Y350S; 및
- [0337] (d)
- [0338] K358D, K358E, K358A, K358V, K358I, K358L, K358M, K358F, K358Y, K358W, K358R, K358H, K358L, K358S, K358T, K358N 또는 K358Q,
- [0339] K358E, K358I, K358L 또는 K358M,
- [0340] K358I 또는 K358M,
- [0341] K358M,
- [0342] K358I,
- [0343] K358A, K358E, K358F, K358I, K358M 또는 K358S,
- [0344] K358A, K358E, K358F, K358I 또는 K358M,
- [0345] K358S,
- [0346] K358A, K358E, K358I, K358M 또는 K358S,
- [0347] K358F,
- [0348] K358I, K358L 또는 K358Q, 또는
- [0349] K358I.
- [0350] 변이체는 (a); (b); (c); (d); (a) 및 (b); (a) 및 (c); (a) 및 (d); (b) 및 (c); (b) 및 (d); (c) 및 (d); (a), (b) 및 (c); (a), (b) 및 (d); (a), (c) 및 (d); (b), (c) 및 (d); 또는 (a), (b), (c) 및 (d)를 포함할 수 있다.
- [0351] 서열번호: 118의 바람직한 변이체는 C114I; K177M; Y350I; K358I; C114I 및 K177M; C114I 및 Y350I; C114I 및 K358I; K177M 및 Y350I; K177M 및 K358I; Y350I 및 K358I; C114I, K177M 및 Y350I; C114I, K177M 및 K358I; C114I, Y350I 및 K358I; K177M, Y350I 및 K358I; 또는 C114I, K177M, Y350I 및 K358I를 포함한다.
- [0352] 헬리카제는 바람직하게는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함하며 여기서 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 위치 중 하나 이상은 위에 정의된 바와 같이 변형되거나 치환된다(특이적 치환 포함). Dda 1993의 위치 55, 114, 156, 177, 210, 221, 350 및 358 중 하나 이상의 다양한 조합 및 순열은 (a)-(h)를 참조하여 위에 정의되어 있다.
- [0353] 헬리카제는 바람직하게는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함하며 여기서 Dda 1993의 아미노산 위치 114, 177, 350 및 358에 상응하는 위치 중 하나 이상은 위에 정의된 바와 같이 변형되거나 치환된다(특이적 치환 포함). Dda 1993의 위치 114, 177, 350 및 358 중 하나 이상의 다양한 조합 및 순열은 (a)-(d)를 참조하여 위에 정의되어 있다.
- [0354] 헬리카제는 바람직하게는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함하며 여기서 Dda 1993의 아미노산 위치 114, 177 및 358에 상응하는 위치 중 하나 이상은 위에 정의된 바와 같이 변형되거나 치환된다(특이적 치환 포함).
- [0355] *Dda 1993의 위치 40*
- [0356] 본 발명의 임의의 변형된 헬리카제는 Dda 1993의 아미노산 위치 40에 상응하는 위치에서 변형 또는 치환을 추가로 포함할 수 있다. 위치 40 또는 상응하는 위치는 A, V, I, L, M, F, Y 또는 W로 치환될 수 있다. Dda 1993의 위치 T40에 상응하는 위치는 위의 표 5에 제시되어 있다. 본 발명의 헬리카제는 바람직하게는 위에 제시된 변형/치환 외에도, T40A, T40V, T40I, T40L, T40M, T40F, T40Y 또는 T40W와 같은 T40에서의 치환을 추가로 포함하

는 서열번호: 118의 변이체를 포함한다. 치환은 바람직하게는 T40Y이다.

- [0357] 본 발명은 Dda 1993의 아미노산 위치 40에 상응하는 위치가 변형되거나 치환된 변형된 DNA 의존적 ATPase(Dda) 헬리카제를 제공한다. 위치 T40은 Dda 1993의 타워 도메인에 있다. 본 발명의 헬리카제는 바람직하게는 T40A, T40V, T40I, T40L, T40M, T40F, T40Y 또는 T40W와 같은 T40에서의 치환을 포함하는 서열번호: 118의 변이체를 포함한다. 치환은 바람직하게는 T40Y이다. 본 발명의 변형된 Dda 헬리카제는 위에 제시된 (a)-(h)의 임의의 조합 및 순열을 포함하여, 아미노산 위치 (a) 55, (b) 114, (c) 156, (d) 177, (e) 210, (f) 221, (g) 350 및 (h) 358에 상응하는 위치 중 하나 이상에서 변형 또는 치환을 추가로 포함할 수 있다. 본 발명의 변형된 Dda 헬리카제는 다음을 포함하여, Dda 1993의 아미노산 위치 (a) 114, (b) 177, (c) 350 및 (d) 358에 상응하는 위치 중 하나 이상에서 변형 또는 치환을 추가로 포함할 수 있다: (a); (b); (c); (d); (a) 및 (b); (a) 및 (c); (a) 및 (d); (b) 및 (c); (b) 및 (d); (c) 및 (d); (a), (b) 및 (c); (a), (b) 및 (d); (a), (c) 및 (d); (b), (c) 및 (d); 또는 (a), (b), (c) 및 (d). 헬리카제는 바람직하게는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함하며 여기서 Dda 1993의 아미노산 위치 40에 상응하는 위치는 위에 정의된 바와 같이 변형되거나 치환된다(특이적 치환 포함).
- [0358] 본 발명의 변형된 헬리카제는 아래에 논의된 변형, 치환, 변형의 조합 또는 치환의 조합 중 임의의 것을 추가로 포함할 수 있다.
- [0359] *본 발명의 다른 헬리카제*
- [0360] 본 발명은 또한 별개로 아래에 논의된 변형, 치환, 변형의 조합 또는 치환의 조합 중 임의의 것을 갖는 변형된 DNA 의존적 ATPase(Dda) 헬리카제를 제공한다. 다시 말해서, 본 발명의 이러한 헬리카제는 Dda 1993의 아미노산 위치 40, 55, 114, 156, 177, 210, 221, 350 및 358 중 임의의 것 또는 Dda 1993의 아미노산 위치 40, 114, 350, 177 및 K358 중 임의의 것에 상응하는 위치에 치환을 반드시 가질 필요가 없다. 본 발명의 이러한 변형된 헬리카제는 바람직하게는 서열번호: 118 내지 133 중 임의의 것의 변이체이다. 변이체는 아래에 제시된 서열번호: 118 내지 133 중 임의의 것에 대한 임의의 %의 서열 상동성/동일성을 가질 수 있다.
- [0361] 분석물 특성화에서 Dda 헬리카제의 사용은 W02015/055981, W02015/166276 및 W02016/055777(모두 참조로 포함됨)에 기재되어 있다.
- [0362] 본 발명의 변형된 헬리카제는 막관통 기공에 관해 예컨대 이를 통해 표적 분석물의 보다 일관된 이동을 제공하여 정확도 개선으로 이어진다. 헬리카제는 바람직하게는 폴리뉴클레오티드와 같은 표적 분석물이 기공에 관해 예컨대 이를 통해 이동할 때 하나의 k-mer에서 또 다른 k-mer로 또는 k-mer에서 k-mer로 보다 일관된 이동을 제공한다. 헬리카제는 표적 폴리뉴클레오티드와 같은 표적 분석물이 막관통 기공과 관련하여, 예컨대 이를 통해 보다 순조롭게 이동하도록 한다. 헬리카제는 바람직하게는 막관통 기공에 관해 예컨대 이를 통해 표적 폴리뉴클레오티드와 같은 표적 분석물의 보다 규칙적이거나 덜 불규칙적인 이동을 제공한다.
- [0363] 변형(들)은 전형적으로 변형이 없는 헬리카제와 비교하여 정확도를 적어도 0.1%, 적어도 0.5%, 적어도 1%, 적어도 5%, 적어도 10%, 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80% 또는 적어도 90% 증가시킨다.
- [0364] 폴리뉴클레오티드의 이동을 제어하는 헬리카제의 능력은 실시예에 기재된 바와 같이 결정될 수 있다.
- [0365] 변형된 헬리카제는 폴리뉴클레오티드의 이동을 제어하는 능력을 갖는다. 폴리뉴클레오티드의 이동을 제어하는 헬리카제의 능력은 당업계에 알려진 임의의 방법을 사용하여 검정될 수 있다. 예를 들면, 헬리카제는 폴리뉴클레오티드와 접촉할 수 있고 폴리뉴클레오티드의 위치는 표준 방법을 사용하여 결정할 수 있다. 폴리뉴클레오티드의 이동을 제어하는 변형된 헬리카제의 능력은 전형적으로 아래에 기재된 것, 특히, 실시예에 기재된 바와 같은 것들과 같은 나노기공 시스템으로 검정된다.
- [0366] 본 발명의 변형된 헬리카제는 단리되거나, 실질적으로 단리되거나, 정제되거나 실질적으로 정제될 수 있다. 헬리카제는 지질, 폴리뉴클레오티드, 기공 단량체 또는 다른 단백질과 같은 임의의 다른 구성요소가 완전히 없는 경우 단리되거나 정제된다. 헬리카제는 의도된 용도를 방해하지 않을 담체 또는 희석제와 혼합되는 경우 실질적으로 단리된다. 예를 들면, 헬리카제는 지질, 폴리뉴클레오티드, 기공 단량체 또는 다른 단백질과 같은 다른 구성요소의 10% 미만, 5% 미만, 2% 미만 또는 1% 미만을 포함하는 형태로 존재하는 경우 실질적으로 단리되거나 실질적으로 정제된다.
- [0367] *Dda 헬리카제*

- [0368] 임의의 Dda 헬리카제는 본 발명에 따라 변형될 수 있다. 바람직한 Dda 헬리카제는 아래에 논의되어 있고 W02015/055981, W02015/166276 및 W02016/055777(모두 참조로 포함됨)에 기재되어 있다.
- [0369] Dda 헬리카제는 전형적으로 다음 5개의 도메인을 포함한다: 1A(RecA-유사 모터) 도메인, 2A(RecA-유사 모터) 도메인, 타워 도메인, 핀 도메인 및 후크 도메인(Xiaoping He 등, 2012, Structure; 20: 1189-1200). 도메인은 단백질 모델링, 결정질 상태 단백질의 x-선 회절 측정(Rupp B(2009). Biomolecular Crystallography: Principles, Practice and Application to Structural Biology. New York: Garland Science.), 용액 중 단백질의 핵 자기 공명(NMR) 분광법(Mark Rance; Cavanagh, John; Wayne J. Fairbrother; Arthur W. Hunt III; Skelton, NNicholas J.(2007). Protein NMR spectroscopy: principles and practice(2nd ed.). Boston: Academic Press.) 또는 동결 수화된 상태 단백질의 극저온 전자 현미경(van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A(2000). "Single-particle electron cryo-microscopy: towards atomic resolution.". Q Rev Biophys. 33: 307-69)을 사용하여 식별될 수 있다. 위에 언급된 방법에 의해 결정된 단백질의 구조적 정보는 단백질 은행(PDB) 데이터베이스에서 공개적으로 이용가능하다.
- [0370] Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또는 치환, Dda 1993의 아미노산 위치 114, 177, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또는 치환, Dda 1993의 아미노산 위치 114, 177 및 350에 상응하는 하나 이상의 위치에서 변형 또는 치환 및/또는 Dda 1993의 위치 40에 상응하는 위치에서 변형 또는 치환 외에도, 본 발명의 변형된 헬리카제는 바람직하게는 다음 추가적인 변형, 치환, 변형의 조합 또는 치환의 조합 중 임의의 것을 포함한다. 위에 설명된 바와 같이, 본 발명은 또한 별도로 아래에 제시된 변형, 치환, 변형의 조합 또는 치환의 조합 중 임의의 것을 갖는 변형된 헬리카제를 제공한다(즉, Dda 1993의 위치 40, 55, 114, 156, 177, 210, 221, 350 및 358 중 임의의 것 또는 Dda 1993의 위치 40, 114, 177, 350 및 358 중 임의의 것에서 치환을 반드시 갖지 않음).
- [0371] *본 발명의 변형*
- [0372] 본 발명의 헬리카제는 막관통 기공과 상호작용하는 적어도 하나의 아미노산이 치환될 것일 수 있다. 1개 이상, 2개 이상, 3개 이상, 4개 이상, 5개 이상 또는 6개 이상의 아미노산과 같은 임의의 수의 아미노산이 치환될 수 있다. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 아미노산이 치환될 수 있다. 막관통 기공과 상호작용하는 아미노산은 위에 논의된 바와 같은 단백질 모델링을 사용하여 식별될 수 있다.
- [0373] *염기 및/또는 당 상호작용*
- [0374] 본 발명의 헬리카제는 바람직하게는 단일 가닥 DNA(ssDNA)에서 하나 이상의 뉴클레오티드의 당 및/또는 염기와 상호작용하는 적어도 하나의 아미노산이 더 큰 측쇄(R 기)를 포함하는 아미노산으로 치환된 것이다. 1개 이상, 2개 이상, 3개 이상, 4개 이상, 5개 이상 또는 6개 이상의 아미노산과 같은 임의의 수의 아미노산이 치환될 수 있다. 각 아미노산은 염기, 당 또는 염기 및 당과 상호작용할 수 있다. 단일 가닥 DNA에서 하나 이상의 뉴클레오티드의 당 및/또는 염기와 상호작용하는 아미노산은 위에 논의된 바와 같은 단백질 모델링을 사용하여 식별될 수 있다.
- [0375] 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 변이체를 포함하며 여기서 ssDNA에서 하나 이상의 뉴클레오티드의 당 및/또는 염기와 상호작용하는 적어도 하나의 아미노산은 H82, N88, P89, F98, D121, V150, P152, F240, F276, S287, H396 및 Y415 중 적어도 하나이다. 이러한 숫자는 서열번호: 118의 관련 위치에 상응하고 서열번호: 118과 비교하여 하나 이상의 아미노산이 삽입되거나 결실된 변이체의 경우 변형되어야 할 수 있다. 당 염기는 위에 논의된 바와 같은 변이체에서 상응하는 위치를 결정할 수 있다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 변이체를 포함하며 여기서 ssDNA에서 하나 이상의 뉴클레오티드의 당 및/또는 염기와 상호작용하는 적어도 하나의 아미노산은 F98 및 H82, N88, P89, D121, V150, P152, F240, F276, S287, H396 및 Y415 중 하나 이상, 예컨대 F98/H82, F98/N88, F98/P89, F98/D121, F98/V150, F98/P152, F98/F240, F98/F276, F98/S287 또는 F98/H396이다.
- [0376] 본 발명의 헬리카제는 바람직하게는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체이며 여기서 ssDNA에서 하나 이상의 뉴클레오티드의 당 및/또는 염기와 상호작용하는 적어도 하나의 아미노산은 서열번호: 118의 H82, N88, P89, F98, D121, V150, P152, F240, F276, S287, H396 및 Y415에 상응하는 아미노산 중 적어도 하나이다. 본 발명의 헬리카제는 바람직하게는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함하며 여기서

ssDNA에서 하나 이상의 뉴클레오타이드의 당 및/또는 염기와 상호작용하는 적어도 하나의 아미노산은 서열번호: 118의 F98에 상응하는 아미노산 및 서열번호: 118의 H82, N88, P89, D121, V150, P152, F240, F276, S287, H396 및 Y415에 상응하는 아미노산, 예컨대 F98/H82, F98/N88, F98/P89, F98/D121, F98/V150, F98/ P152, F98/F240, F98/F276, F98/S287 또는 F98/H396에 상응하는 아미노산 중 하나 이상이다.

[0377] 표 6은 서열번호: 118의 H82, N88, P89, F98, D121, V150, P152, F240, F276, S287, H396 및 Y415에 상응하는 서열번호: 119 내지 133의 아미노산을 보여준다.

서열번호:												
118	H82	N88	P89	F98	D121	V150	P152	F240	F276	S287	H396	Y415
119	H81	F87	D88	S105	S131	V181	Q183	R274	H313	G314	H428	H447
120	A144	Q150	P151	G158	N187	V217	K218	W307	F344	K355	H455	H473
121	H87	R93	L94	G100	G126	V154	N155	Y240	Y277	G280	H377	R395
122	H112	K118	P119	E128	G154	I185	N187	Y273	F309	K310	H414	H433
123	H93	V99	M100	D106	N132	I159	Q160	Y267	M302	G303	H400	K419
124	H74	H80	P81	F90	D114	V143	H145	Y230	M266	P273	H378	Y397
125	H78	H84	P85	F94	S117	E147	A149	Y235	M271	I279	H387	Y406
126	H65	H71	P72	F81	S104	V133	H135	F222	M258	I266	H373	Y392
127	H84	S90	P91	F100	D126	V155	A157	Y243	V279	T290	H399	A418
128	H84	S90	P91	F100	D126	V155	T157	Y243	V279	V290	H398	A417
129	Q82	N88	P89	F98	T121	V150	E152	Y237	M274	K285	H393	Q412
130	H88	N94	P95	F104	D127	V156	P158	F246	I282	S293	H399	K418
131	H84	N90	P91	F100	D123	V152	P154	Y242	M278	S289	H399	M418
132	H83	N89	P90	F99	D122	V151	P153	Y241	M277	H288	H396	M415
133	H83	N89	P90	F99	D122	V151	P153	F241	M277	H288	H397	M416

[0378]

[0379] ssDNA에서 하나 이상의 뉴클레오타이드의 당 및/또는 염기와 상호작용하는 적어도 하나의 아미노산은 바람직하게는 ssDNA에서 뉴클레오타이드 사이에 끼워넣는 적어도 하나의 아미노산이다. ssDNA에서 뉴클레오타이드 사이에 끼워넣는 아미노산은 위에 논의된 바와 같이 모델링될 수 있다. ssDNA에서 뉴클레오타이드 사이에 끼워넣는 적어도 하나의 아미노산은 바람직하게는 서열번호: 118의 P89, F98 및 V150 중 적어도 하나, 예컨대 P89, F98, V150, P89/F98, P89/V150, F98/V150 또는 P89/F98/V150이다.

[0380] 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 ssDNA에서 뉴클레오타이드 사이에 끼워넣는 적어도 하나의 아미노산은 바람직하게는 서열번호: 118의 P89, F98 및 V150에 상응하는 아미노산 중 적어도 하나, 예컨대 P89, F98, V150, P89/F98, P89/V150, F98/V150 또는 P89/F98/V150이다. 상응하는 아미노산은 위의 표 6에 제시되어 있다.

[0381] 더 큰 R 기

[0382] 더 큰 측쇄(R 기)는 바람직하게는 (a) 증가된 수의 탄소 원자를 함유하고/하거나, (b) 증가된 길이를 갖고/갖거나, (c) 증가된 분자 부피를 갖고/갖거나 (d) 증가된 반데르발스 부피를 갖는다. 더 큰 측쇄(R 기) 바람직하게는 (a); (b); (c); (d); (a) 및 (b); (a) 및 (c); (a) 및 (d); (b) 및 (c); (b) 및 (d); (c) 및 (d); (a), (b) 및 (c); (a), (b) 및 (d); (a), (c) 및 (d); (b), (c) 및 (d); 또는 (a), (b), (c) 및 (d). (a) 내지 (d) 각각은 당염계의 표준 방법을 사용하여 측정될 수 있다.

[0383] 더 큰 측쇄(R 기)는 바람직하게는 ssDNA에서 적어도 하나의 아미노산과 하나 이상의 뉴클레오타이드 사이의 (i) 정전기적 상호작용 (ii) 수소 결합 및/또는 (iii) 양이온-파이(양이온- π) 상호작용을 증가시키며, 예컨대 (i); (ii); (iii); (i) 및 (ii); (i) 및 (iii); (ii) 및 (iii); 및 (i), (ii) 및 (iii)을 증가시킨다. 당염자

는 R 기가 이러한 상호작용 중 임의의 것을 증가시키는지 결정할 수 있다. 예를 들면 (i)에서, 아르기닌(R), 히스티딘(H) 및 리신(K)과 같은 양으로 하전된 아미노산은 정전기적 상호작용을 증가시키는 R 기를 갖는다. 예를 들면 (ii)에서, 아스파라긴(N), 세린(S), 글루타민(Q), 트레오닌(T) 및 히스티딘(H)과 같은 아미노산은 수소 결합을 증가시키는 R 기를 갖는다. 예를 들면 (iii)에서, 페닐알라닌(F), 트립토판(W), 티로신(Y) 또는 히스티딘(H)과 같은 방향족 아미노산은 양이온-파이(양이온- π) 상호작용을 증가시키는 R 기를 갖는다. 아래 특이적 치환은 이러한 변화를 반영하기 위해 (i) 내지 (iii)으로 라벨링된다. 다른 가능한 치환은 (iv)로 라벨링된다. 이러한 (iv) 치환은 전형적으로 측쇄(R 기)의 길이를 증가시킨다.

- [0384] 더 큰 측쇄(R)를 포함하는 아미노산은 비천연 아미노산일 수 있다. 비천연 아미노산은 아래 논의된 것들 중 임의의 것일 수 있다.
- [0385] 더 큰 측쇄(R 기)를 포함하는 아미노산은 바람직하게는 알라닌(A), 시스테인(C), 글리신(G), 셀레노시스테인(U), 메티오닌(M), 아스파르트산(D) 또는 글루탐산(E)이 아니다.
- [0386] 히스티딘(H)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q) 또는 아스파라긴(N) 또는 (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W)으로 치환된다. 히스티딘(H)은 보다 바람직하게는 (a) N, Q 또는 W 또는 (b) Y, F, Q 또는 K로 치환된다.
- [0387] 아스파라긴(N)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q) 또는 히스티딘(H) 또는 (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W)으로 치환된다. 아스파라긴(N)은 보다 바람직하게는 R, H, W 또는 Y로 치환된다.
- [0388] 프롤린(P)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N), 트레오닌(T) 또는 히스티딘(H), (iii) 티로신(Y), 페닐알라닌(F) 또는 트립토판(W) 또는 (iv) 류신(L), 발린(V) 또는 이소류신(I)으로 치환된다. 프롤린(P)은 보다 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N), 트레오닌(T) 또는 히스티딘(H), (iii) 페닐알라닌(F) 또는 트립토판(W) 또는 (iv) 류신(L), 발린(V) 또는 이소류신(I)으로 치환된다. 프롤린(P)은 보다 바람직하게는 (a) F, (b) L, V, I, T 또는 F 또는 (c) W, F, Y, H, I, L 또는 V로 치환된다.
- [0389] 발린(V)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N) 또는 히스티딘(H), (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W) 또는 (iv) 이소류신(I) 또는 류신(L)으로 치환된다. 발린(V)은 보다 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N) 또는 히스티딘(H), (iii) 티로신(Y) 또는 트립토판(W) 또는 (iv) 이소류신(I) 또는 류신(L)으로 치환된다. 발린(V)은 보다 바람직하게는 I 또는 H 또는 I, L, N, W 또는 H로 치환된다.
- [0390] 페닐알라닌(F)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 히스티딘(H) 또는 (iii) 티로신(Y) 또는 트립토판(W)으로 치환된다. 페닐알라닌(F)은 보다 바람직하게는 (a) W, (b) W, Y 또는 H, (c) W, R 또는 K 또는 (d) K, H, W 또는 R로 치환된다.
- [0391] 글루타민(Q)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K) 또는 (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W)으로 치환된다.
- [0392] 알라닌(A)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N) 또는 히스티딘(H), (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W) 또는 (iv) 이소류신(I) 또는 류신(L)으로 치환된다.
- [0393] 세린(S)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N) 또는 히스티딘(H), (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W) 또는 (iv) 이소류신(I) 또는 류신(L)으로 치환된다. 세린(S)은 바람직하게는 K, R, W 또는 F로 치환된다.
- [0394] 리신(K)은 바람직하게는 (i) 아르기닌(R) 또는 (iii) 티로신(Y) 또는 트립토판(W)으로 치환된다.
- [0395] 아르기닌(R)은 바람직하게는 (iii) 티로신(Y) 또는 트립토판(W)으로 치환된다.
- [0396] 메티오닌(M)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N) 또는 히스티딘(H) 또는 (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W)으로 치환된다.
- [0397] 류신(L)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q) 또는 아스파라긴(N) 또는 (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W)으로 치환된다.

- [0398] 아스파르트산(D)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N) 또는 히스티딘(H) 또는 (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W)으로 치환된다. 아스파르트산(D)은 보다 바람직하게는 H, Y 또는 K로 치환된다.
- [0399] 글루탐산(E)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N) 또는 히스티딘(H) 또는 (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W)으로 치환된다.
- [0400] 이소류신(I)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 글루타민(Q), 아스파라긴(N) 또는 히스티딘(H), (iii) 페닐알라닌(F), 티로신(Y) 또는 트립토판(W) 또는 (iv) 류신(L)으로 치환된다.
- [0401] 티로신(Y)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K) 또는 (iii) 트립토판(W)으로 치환된다. 티로신(Y)은 보다 바람직하게는 W 또는 R로 치환된다.
- [0402] 헬리카제는 보다 바람직하게는 서열번호: 118의 변이체를 포함하고 (a) P89F, (b) F98W, (c) V150I, (d) V150H, (e) P89F 및 F98W, (f) P89F 및 V150I, (g) P89F 및 V150H, (h) F98W 및 V150I, (i) F98W 및 V150H (j) P89F, F98W 및 V150I 또는 (k) P89F, F98W 및 V150H를 포함한다.
- [0403] 헬리카제는 보다 바람직하게는 다음을 포함하는 서열번호: 118의 변이체를 포함한다: H82N; H82Q; H82W; N88R; N88H; N88W; N88Y; P89L; P89V; P89I ; P89E; P89T; P89F; D121H; D121Y; D121K; V150I; V150L; V150N; V150W; V150H; P152W; P152F; P152Y; P152H; P152I; P152L; P152V; F240W; F240Y; F240H; F276W; F276R; F276K; F276H; S287K; S287R; S287W; S287F; H396Y; H396F; H396Q; H396K; Y415W; Y415R; F98W/H82N; F98W/H82Q; F98W/H82W; F98W/N88R; F98W/N88H; F98W/N88W; F98W/N88Y; F98W/P89L; F98W/P89V; F98W/P89I; F98W/P89T; F98W/P89F; F98W/D121H; F98W/D121Y; F98W/D121K; F98W/V150I; F98W/V150L; F98W/V150N; F98W/V150W; F98W/V150H; F98W/P152W; F98W/P152F; F98W/P152Y; F98W/P152H; F98W/P152I; F98W/P152L; F98W/P152V; F98W/F240W; F98W/F240Y; F98W/F240H; F98W/F276W; F98W/F276R; F98W/F276K; F98W/F276H; F98W/S287K; F98W/S287R; F98W/S287W; F98W/S287F; F98W/H396Y; F98W/H396F; F98W/H396Q; F98W/Y415W; 또는 F98W/Y415R.
- [0404] *포스페이트 상호작용*
- [0405] 본 발명의 헬리카제는 바람직하게는 ssDNA에서 하나 이상의 뉴클레오티드의 하나 이상의 포스페이트 기와 상호작용하는 적어도 하나의 아미노산이 치환된 것이다. 1개 이상, 2개 이상, 3개 이상, 4개 이상, 5개 이상 또는 6개 이상의 아미노산과 같은 임의의 수의 아미노산이 치환될 수 있다. ssDNA에서 뉴클레오티드는 각각 3개의 포스페이트 기를 포함한다. 치환되는 각 아미노산은 한번에 1, 2 또는 3개의 포스페이트 기와 같이 한번에 임의의 수의 포스페이트 기와 상호작용할 수 있다. 하나 이상의 포스페이트 기와 상호작용하는 아미노산은 위에 논의된 바와 같은 단백질 모델링을 사용하여 식별될 수 있다.
- [0406] 치환은 바람직하게는 ssDNA에서 적어도 하나의 아미노산과 하나 이상의 포스페이트 기 사이의 (i) 정전기적 상호작용, (ii) 수소 결합 및/또는 (iii) 양이온-파이(양이온- π) 상호작용을 증가시킨다. (i), (ii) 및 (iii)을 증가시키는 바람직한 치환은 라벨링 (i), (ii) 및 (iii)을 사용하여 아래에 논의되어 있다.
- [0407] 치환은 바람직하게는 위치의 순 양전하를 증가시킨다. 임의의 위치에서 순 전하는 당업계에 알려진 방법을 사용하여 측정될 수 있다. 예를 들면, 등전점을 사용하여 아미노산의 순 전하를 정의할 수 있다. 순 전하는 전형적으로 약 7.5에서 측정된다. 치환은 바람직하게는 음으로 하전된 아미노산을 양으로 하전, 비하전, 비극성 또는 방향족 아미노산으로 치환하는 것이다. 음으로 하전된 아미노산은 순 음전하를 갖는 아미노산이다. 음으로 하전된 아미노산은 아스파르트산(D) 및 글루탐산(E)을 포함하나 이에 제한되지 않는다. 양으로 하전된 아미노산은 순 양전하를 갖는 아미노산이다. 양으로 하전된 아미노산은 자연 발생 또는 비자연 발생일 수 있다. 양으로 하전된 아미노산은 합성 또는 변형될 수 있다. 예를 들면, 순 양전하를 갖는 변형된 아미노산은 본 발명에 사용하기 위해 특이적으로 설계될 수 있다. 아미노산에 대한 다수의 상이한 유형의 변형은 당업계에 잘 알려져 있다. 바람직한 자연 발생 양으로 하전된 아미노산은 히스티딘(H), 리신(K) 및 아르기닌(R)을 포함하나 이에 제한되지 않는다.
- [0408] 비하전된 아미노산, 비극성 아미노산 또는 방향족 아미노산은 자연 발생 또는 비자연 발생일 수 있다. 이는 합성 또는 변형될 수 있다. 비하전된 아미노산은 순 전하가 없다. 적합한 비하전된 아미노산은 시스테인(C), 세린(S), 트레오닌(T), 메티오닌(M), 아스파라긴(N) 및 글루타민(Q)을 포함하나 이에 제한되지 않는다. 비극성 아미노산은 비극성 측쇄를 갖는다. 적합한 비극성 아미노산은 글리신(G), 알라닌(A), 프롤린(P), 이소류신(I), 류신

(L) 및 발린(V)을 포함하나 이에 제한되지 않는다. 방향족 아미노산은 방향족 측쇄를 갖는다. 적합한 방향족 아미노산은 히스티딘(H), 페닐알라닌(F), 트립토판(W) 및 티로신(Y)을 포함하나 이에 제한되지 않는다.

[0409] 헬리카제는 바람직하게는 서열번호: 118의 변이체를 포함하며 여기서 ssDNA에서 하나 이상의 뉴클레오타이드의 하나 이상의 포스페이트와 상호작용하는 적어도 하나의 아미노산은 H64, T80, S83, N242, K243, N293, T394 및 K397 중 적어도 하나이다. 이들 숫자는 서열번호: 89의 관련 위치에 상응하고 서열번호: 118과 비교하여 하나 이상의 아미노산이 삽입되거나 결실된 변이체의 경우 변경될 필요가 있을 수 있다. 당업자는 위에 논의된 바와 같은 변이체에서 상응하는 위치를 결정할 수 있다.

[0410] 헬리카제는 바람직하게는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함하고 여기서 ssDNA에서 하나 이상의 뉴클레오타이드의 하나 이상의 포스페이트와 상호작용하는 적어도 하나의 아미노산은 서열번호: 118의 H64, T80, S83, N242, K243, N293, T394 및 K397에 상응하는 아미노산 중 적어도 하나이다.

[0411] 표 7은 서열번호: 118의 H64, T80, S83, N242, K243, N293, T 394 및 K397에 상응하는 서열번호: 119 내지 133의 아미노산을 보여준다.

서열번호:								
118	H64	T80	S83	N242	K243	N293	T394	K397
119	G63	T79	S82	N276	A277	N317	T426	K429
120	N121	T142	K145	N309	K310	N361	T453	K456
121	H66	T85	S88	N242	K243	N286	T375	R378
122	H89	T110	S113	N275	K276	V315	T412	K415
123	H68	T91	S94	N269	A270	N308	T398	K401
124	H56	T72	S75	N232	D233	N279	T376	K379
125	H60	T76	S79	N237	K238	N285	T385	K388
126	H47	T63	S66	N224	K225	N272	T371	K374
127	H66	T82	S85	N245	D246	N296	T397	K400
128	H66	T82	S85	N245	A246	N296	T396	K399
129	H64	T80	S83	N239	N240	N291	T391	K394
130	H70	T86	S89	N248	K249	N299	T397	K400
131	H66	T82	S85	N244	K245	N295	T397	K400
132	H65	T81	S84	N243	K244	N294	T394	K397
133	H65	T81	K84	N243	K244	N294	T395	K398

[0412]

[0413] 히스티딘(H)은 바람직하게는 (i) 아르기닌(R) 또는 리신(K), (ii) 아스파라긴(N), 세린(S), 글루타민(Q) 또는 트레오닌(T), (iii) 페닐알라닌(F), 트립토판(W) 또는 티로신(Y)으로 치환된다. 히스티딘(H)은 바람직하게는 (a) N, Q, K 또는 F 또는 (b) N, Q 또는 W로 치환된다.

[0414] 트레오닌(T)은 바람직하게는 (i) 아르기닌(R), 히스티딘(H) 또는 리신(K), (ii) 아스파라긴(N), 세린(S), 글루타민(Q) 또는 히스티딘(H) 또는 (iii) 페닐알라닌(F), 트립토판(W), 티로신(Y) 또는 히스티딘(H)으로 치환된다. 트레오닌(T)은 보다 바람직하게는 (a) K, Q 또는 N 또는 (b) K, H 또는 N으로 치환된다.

[0415] 세린(s)은 바람직하게는 (i) 아르기닌(R), 히스티딘(H) 또는 리신(K), (ii) 아스파라긴(N), 글루타민(Q), 트레오닌(T) 또는 히스티딘(H) 또는 (iii) 페닐알라닌(F), 트립토판(W), 티로신(Y) 또는 히스티딘(H)으로 치환된다. 세린(S)은 보다 바람직하게는 H, N, K, T, R 또는 Q로 치환된다.

[0416] 아스파라긴(N)은 바람직하게는 (i) 아르기닌(R), 히스티딘(H) 또는 리신(K), (ii) 세린(S), 글루타민(Q), 트레오닌(T) 또는 히스티딘(H) 또는 (iii) 페닐알라닌(F), 트립토판(W), 티로신(Y) 또는 히스티딘(H)으로 치환된다.

아스파라긴(N)은 보다 바람직하게는 (a) H 또는 Q 또는 (b) Q, K 또는 H로 치환된다.

- [0417] 리신(K)은 바람직하게는 (i) 아르기닌(R) 또는 히스티딘(H), (ii) 아스파라긴(N), 세린(S), 글루타민(Q), 트레오닌(T) 또는 히스티딘(H) 또는 (iii) 페닐알라닌(F), 트립토판(W), 티로신(Y) 또는 히스티딘(H)으로 치환된다. 리신(K)은 보다 바람직하게는 (a) Q 또는 H 또는 (b) R, H 또는 Y로 치환된다.
- [0418] 헬리카제는 보다 바람직하게는 서열번호: 118의 변이체를 포함하고 (a) H64N, H64Q, H64K 또는 H64F, (b) T80K, T80Q 또는 T80N, (c) S83H, S83N, S83K, S83T, S83R, 또는 S83Q(d) N242H 또는 N242Q, (e) K243Q 또는 K243H, (f) N293Q, N293K 또는 N293H, (g) T394K, T394H 또는 T394N 또는 (h) K397R, K397H 또는 K397Y 중 하나 이상, 예컨대 전부를 포함한다.
- [0419] 조합
- [0420] 헬리카제는 바람직하게는 다음에서의 치환을 포함하는 서열번호: 118의 변이체이다:
- [0421] - F98/H64, 예컨대 F98W/H64N, F98W/H64Q, F98W/H64K 또는 F98W/H64F;
- [0422] - F98/T80, 예컨대 F98W/T80K, F98W/T80Q, F98W/T80N;
- [0423] - F98/H82, 예컨대 F98W/H82N, F98W/H82Q 또는 F98W/H82W;
- [0424] - F98/S83, 예컨대 F98W/S83H, F98W/S83N, F98W/S83K, F98W/S83T, F98W/S83R 또는 F98W/S83Q;
- [0425] - F98/N242, 예컨대 F98W/N242H, F98W/N242Q, F98W/K243Q 또는 F98W/K243H;
- [0426] - F98/N293, 예컨대 F98W/N293Q, F98W/N293K, F98W/N293H, F98W/T394K, F98W/T394H, F98W/T394N, F98W/H396Y, F98W/H396F, F98W/H396Q 또는 F98W/H396K; 또는
- [0427] - F98/K397, 예컨대 F98W/K397R, F98W/K397H 또는 F98W/K397Y.
- [0428] 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 바람직한 조합은 위에 나열된 서열번호: 118의 조합에 상응하는 아미노산의 조합을 포함한다.
- [0429] 기공 상호작용
- [0430] 본 발명의 헬리카제는 추가로 막관통 기공과 상호작용하는 헬리카제의 부분이 하나 이상의 변형, 바람직하게는 하나 이상의 치환을 포함하는 것이다. 막관통 기공과 상호작용하는 헬리카제의 부분은 전형적으로 예를 들면 아래에 보다 상세하게 논의된 바와 같이 헬리카제가 기공을 통해 폴리뉴클레오티드의 이동을 제어하는 데 사용되는 경우 막관통 기공과 상호작용하는 헬리카제의 부분이다. 부분은 전형적으로 예를 들면 아래에 보다 상세하게 논의된 바와 같이 헬리카제가 기공을 통해 폴리뉴클레오티드의 이동을 제어하는 데 사용되는 경우 기공과 상호작용하거나 접촉하는 아미노산을 포함한다. 부분은 전형적으로 헬리카제가 인가된 전위 하에 기공을 통해 이동하는 폴리뉴클레오티드와 같은 분석물에 결합되거나 부착되는 경우 기공과 상호작용하거나 접촉하는 아미노산을 포함한다.
- [0431] 서열번호: 118에서, 막관통 기공과 상호작용하는 부분은 전형적으로 위치 1, 2, 3, 4, 5, 6, 51, 176, 177, 178, 179, 180, 181, 185, 189, 191, 193, 194, 195, 197, 198, 199, 200, 201, 202, 203, 204, 207, 208, 209, 210, 211, 212, 213, 216, 219, 220, 221, 223, 224, 226, 227, 228, 229, 247, 254, 255, 256, 257, 258, 259, 260, 261, 298, 300, 304, 308, 318, 319, 321, 337, 347, 350, 351, 405, 415, 422, 434, 437, 438에 아미노산을 포함한다. 이들 숫자는 서열번호: 118의 관련 위치와 상응하고 서열번호: 118과 비교하여 하나 이상의 아미노산이 삽입되거나 결실된 변이체의 경우 변형될 필요가 있을 수 있다. 당업자는 위에 논의된 바와 같은 변이체에서 상응하는 위치를 결정할 수 있다. 막관통 기공과 상호작용하는 부분은 바람직하게는 다음에서의 아미노산을 포함한다:
- [0432] (a) 서열번호: 118의 위치 1, 2, 4, 51, 177, 178, 179, 180, 185, 193, 195, 197, 198, 199, 200, 202, 203, 204, 207, 208, 209, 210, 211, 212, 216, 221, 223, 224, 226, 227, 228, 229, 254, 255, 256, 257, 258, 260, 304, 318, 321, 347, 350, 351, 405, 415, 422, 434, 437 및 438; 또는
- [0433] (b) 서열번호: 118의 위치 1, 2, 178, 179, 180, 185, 195, 197, 198, 199, 200, 202, 203, 207, 209, 210, 212, 216, 221, 223, 226, 227, 255, 258, 260, 304, 350 및 438.
- [0434] 막관통 기공과 상호작용하는 부분은 바람직하게는 서열번호: 118의 위치 K194, W195, K198, K199 및 E258에 하

나 이상, 예컨대 2, 3, 4, 또는 5개의 아미노산을 포함한다. 서열번호: 118의 변이체는 바람직하게는 (a), K194, (b) W195, (c) D198, (d) K199 및 (d) E258 중 하나 이상에서 변형을 포함한다. 서열번호: 118의 변이체는 바람직하게는 (a) K194, 예컨대 K194L, (b) W195, 예컨대 W195A, (c) D198, 예컨대 D198V, (d) K199, 예컨대 K199L 및 (e) E258, 예컨대 E258L 중 하나 이상에 치환을 포함한다. 변이체는 {a}; {b}; {c}; {d}; {e}; {a,b}; {a,c}; {a,d}; {a,e}; {b,c}; {b,d}; {b,e}; {c,d}; {c,e}; {d,e}; {a,b,c}; {a,b,d}; {a,b,e}; {a,c,d}; {a,c,e}; {a,d,e}; {b,c,d}; {b,c,e}; {b,d,e}; {c,d,e}; {a,b,c,d}; {a,b,c,e}; {a,b,d,e}; {a,c,d,e}; {b,c,d,e}; 또는 {a,b,c,d,e}를 포함할 수 있다. 이 문단에 제시된 변형 또는 치환은 변형된 폴리뉴클레오티드 결합 단백질이 MspA로부터 유래된 기공, 특히 아래에 논의된 변형된 기공 중 임의의 것과 상호작용할 때 바람직하다.

[0435] 막관통 기공과 상호작용하는 폴리뉴클레오티드 결합 단백질의 부분은 바람직하게는 서열번호: 118의 위치 194 또는 199에 아미노산을 포함한다. 변이체는 바람직하게는 K194A, K194V, K194F, K194D, K194S, K194W 또는 K194L 및/또는 K199A, K199V, K199F, K199D, K199S, K199W 또는 K199L을 포함한다. 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133에서, 막관통 기공과 상호작용하는 부분은 전형적으로 위에 나열된 서열번호: 118에 상응하는 위치에 아미노산을 포함한다. 서열번호: 118의 이러한 위치에 상응하는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 및 133의 아미노산은 아래 표 8의 정렬을 사용하여 식별될 수 있다.

서열번호: 118	K194	W195	D198	K199	E258
서열번호: 119	L230	E231	H234	Y235	R293
서열번호: 120	W259	N260	T263	Y264	E326
서열번호: 121	A192	D193	F196	G197	A259
서열번호: 122	I224	K225	D228	F229	Q292
서열번호: 123	Q213	D214	Y217	A218	A286
서열번호: 124	Q185	W186	T189	N190	N248
서열번호: 125	G190	W191	P194	N195	K253
서열번호: 126	G177	W178	Q181	N182	K240
서열번호: 127	K200	M201	P204	M205	K261
서열번호: 128	K200	P201	P204	L205	K261
서열번호: 129	K193	W194	E197	K198	A256
서열번호: 130	N200	W201	E204	N205	N264
서열번호: 131	G196	W197	D200	C201	E260
서열번호: 132	G195	W196	E199	N200	E259
서열번호: 133	S195	W196	E199	K200	Q259

[0436]
[0437] 바람직한 조합

[0438] 본 발명의 헬리카제는 바람직하게는 F98R, F98K, F98Q, F98N, F98H, F98Y, F98F 또는 F98W와 같은 F98에서의 치환, 및 K194A, K194V, K194F, K194D, K194S, K194W 또는 K194L과 같은 K194에서의 치환, 및/또는 K199, 예컨대 K199A, K199V, K199F, K199D, K199S, K199W 또는 K199L과 같은 K199에서의 치환을 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 F98에 상응하는 위치에서 치환 및 서열번호: 118의 K194 및/또는 K199에 상응하는 위치(들)에서의 치환을 포함하는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함한다. 이러한 상응하는 위치는 서열번호: 118의 F98, K194 및 K199에 대해 상기 나열된 아미노산 중 임의의 것으로 대체될 수 있다.

[0439] 헬리카제는 바람직하게는 다음에서의 치환을 포함하는 서열번호: 118의 변이체이다:

- [0440] - F98/K194/H64, 예컨대 F98W/K194L/H64N, F98W/K194L/H64Q, F98W/K194L/H64K 또는 F98W/K194L/H64F;
- [0441] - F98/K194/T80, 예컨대 F98W/K194L/T80K, F98W/K194L/T80Q 또는 F98W/K194L/T80N;
- [0442] - F98/K194/H82, 예컨대 F98W/K194L/H82N, F98W/K194L/H82Q 또는 F98W/K194L/H82W
- [0443] - F98/S83/K194, 예컨대 F98W/S83H/K194L, F98W/S83T/K194L, F98W/S83R/K194L, F98W/S83Q/K194L, F98W/S83N/K194L, F98W/S83K/K194L, F98W/N88R/K194L, F98W/N88H/K194L, F98W/N88W/K194L 또는 F98W/N88Y/K194L;
- [0444] - F98/S83/K194/F276, 예컨대 F98W/S83H/K194L/F276K;
- [0445] - F98/P89/K194, 예컨대 F98W/P89L/K194L, F98W/P89V/K194L, F98W/P89I/K194L 또는 F98W/P89T/K194L;
- [0446] - F98/D121/K194, 예컨대 F98W/D121H/K194L, F98W/D121Y/K194L 또는 F98W/D121K/K194L;
- [0447] - F98/V150/K194, 예컨대 F98W/V150I/K194L, F98W/V150L/K194L, F98W/V150N/K194L, F98W/V150W/K194L 또는 F98W/V150H/K194L;
- [0448] - F98/P152/K194, 예컨대 F98W/P152W/K194L, F98W/P152F/K194L, F98W/P152Y/K194L, F98W/P152H/K194L, F98W/P152I/K194L, F98W/P152L/K194L 또는 F98W/P152V/K194L;
- [0449] - F98/F240/K194, 예컨대 F98W/F240W/K194L, F98W/F240Y/K194L 또는 F98W/F240H/K194L;
- [0450] - F98/N242/K194, 예컨대 F98W/N242H/K194L 또는 F98W/N242Q/K194L;
- [0451] - F98/K194/F276, 예컨대 F98W/K194L/F276K, F98W/K194L/F276H, F98W/K194L/F276W 또는 F98W/K194L/F276R;
- [0452] - F98/K194/S287, 예컨대 F98W/K194L/S287K, F98W/K194L/S287R, F98W/K194L/S287W 또는 F98W/K194L/S287F;
- [0453] - F98/N293/K194, 예컨대 F98W/N293Q/K194L, F98W/N293K/K194L 또는 F98W/N293H/K194L;
- [0454] - F98/T394/K194, 예컨대 F98W/T394K/K194L, F98W/T394H/K194L 또는 F98W/T394N/K194L;
- [0455] - F98/H396/K194, 예컨대 F98W/H396Y/K194L, F98W/H396F/K194L, F98W/H396Q/K194L 또는 F98W/H396K/K194L;
- [0456] - F98/K397/K194, 예컨대 F98W/K397R/K194L, F98W/K397H/K194L 또는 F98W/K397Y/K194L; 또는
- [0457] - F98/Y415/K194, 예컨대 F98W/Y415W/K194L 또는 F98W/Y415R/K194L.
- [0458] 임의의 상기 조합에서, K194는 W195, D198, K199 및 E258 중 임의의 것으로 대체될 수 있다.
- [0459] 변형된 헬리카제는 바람직하게는 Dda 1993의 아미노산 위치 98 및/또는 194에 상응하는 위치(들)에서 변형 또는 치환을 포함한다. 이는 바람직하게는 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 위치 중 하나 이상에서 변형 또는 치환, Dda 1993의 아미노산 위치 114, 177, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또는 치환 및/또는 Dda 1993의 위치 40에 상응하는 위치에서 변형 또는 치환에 더해진다. 위치 98 또는 상응하는 위치는 R, H, K, S, T, N, Q, A, V, I, L, M, Y 또는 W로 치환될 수 있다. 위치 98 또는 상응하는 위치는 바람직하게는 R, K, Q, N, H, Y 또는 W로 치환된다. 위치 194 또는 상응하는 위치는 A, V, I, L, M, F, Y, W, D, E, S, T, N 또는 Q로 치환될 수 있다. 위치 194 또는 상응하는 위치는 바람직하게는 A, V, F, D, S, W 또는 L로 치환된다. 본 발명의 헬리카제는 바람직하게는 F98R, F98K, F98Q, F98N, F98H, F98Y 또는 F98W와 같은 F98에서의 치환, 및/또는 K194A, K194V, K194F, K194D, K194S, K194W 또는 K194L과 같은 K194에서의 치환을 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 F98W 및 K194L을 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 F98에 상응하는 위치에서 치환 및/또는 서열번호: 118의 K194에 상응하는 위치에서 치환을 포함하는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함한다.
- [0460] 임의의 상기 조합에서, K194는 W195, D198, K199 및 E258 중 임의의 것으로 대체될 수 있다.
- [0461] 타워 도메인 및/또는 핀 도메인 및/또는 1A 도메인에서의 변형
- [0462] 변형된 헬리카제는 바람직하게는 Dda 1993의 아미노산 위치 360에 상응하는 위치에서 변형 또는 치환을 포함한다. 이는 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또는 치환, Dda 1993의 아미노산 위치 114, 177, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또

는 치환 및/또는 Dda 1993의 위치 40에 상응하는 위치에서 변형 또는 치환에 더해질 수 있다. A360은 Y350 및 K358과 같이 Dda 1993의 타워 도메인에 있다. 위치 360 또는 상응하는 위치는 C, G, P, A, V, I, L, M, F, Y 또는 W로 치환될 수 있다. 위치 360 또는 상응하는 위치는 바람직하게는 C 또는 Y로 치환된다. 본 발명의 헬리카제는 바람직하게는 A360C 또는 A360Y와 같은 A360에서의 치환을 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 K358I 및 A360C를 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 A360에 상응하는 위치에서 치환을 포함하는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함한다.

[0463] 변형된 헬리카제는 바람직하게는 위치(들) 94, 98, 109, 94 및 98, 94 및 109, 98 및 109 및 94, 98 및 109과 같은 Dda 1993의 아미노산 위치 94, 98 및 109에 상응하는 위치 중 하나 이상에서 변형 또는 치환을 포함한다. 이는 Dda 1993의 아미노산(amino acid) 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또는 치환, Dda 1993의 아미노산 위치 114, 177, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또는 치환 및/또는 Dda 1993의 위치 40에 상응하는 위치에서 변형 또는 치환에 더해질 수 있다. 이들 위치는 모두 핀 도메인에 있다. 위치 94 또는 상응하는 위치는 C, G, P, A, V, I, L, M, F, Y 또는 W로 치환될 수 있다. 위치 94 또는 상응하는 위치는 바람직하게는 C 또는 Y로 치환된다. 위치 98 또는 상응하는 위치는 R, H, K, S, T, N, Q, A, V, I, L, MY 또는 W로 치환될 수 있다. 위치 98 또는 상응하는 위치는 바람직하게는 R, K, Q, N, H, Y 또는 W로 치환된다. 위치 109 또는 상응하는 위치는 A, V, I, L, M, F, Y 또는 W로 치환될 수 있다. 위치 109 또는 상응하는 위치는 바람직하게는 A 또는 V로 치환된다. 본 발명의 헬리카제는 바람직하게는 (위에 제시된 모든 조합을 포함하여) E94, F98 및 C109 중 하나 이상에서 치환을 포함하는 서열번호: 118의 변이체를 포함한다. 바람직한 변이체는 다음에서의 치환을 포함한다:

[0464] - E94 및 F98, 예컨대 E94C 또는 E94Y 및 F98R, F98K, F98Q, F98N, F98H, F98Y, F98F 또는 F98W;

[0465] - E94 및 C109, 예컨대 E94C 또는 E94Y 및 C109A 또는 C109V;

[0466] - F98 및 C109, 예컨대 F98R, F98K, F98Q, F98N, F98H, F98Y, F98F 또는 F98W 및 C109A 또는 C109V; 또는

[0467] - E94, F98 및 C109, 예컨대 E94C 또는 E94Y 및 F98R, F98K, F98Q, F98N, F98H, F98Y, F98F 또는 F98W 및 C109A 또는 C109V.

[0468] 보다 바람직한 변이체는 E94C 및 F98W; E94C 및 C109A; F98W 및 C109A; 또는 E94C, F98W 및 C109A를 포함한다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 E94, F98 및 C109 중 하나 이상에 상응하는 위치(들)에서 치환을 포함하는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함한다.

[0469] 표 9는 E94, C109, C136 및 A360에 대한 정보를 포함한다(위 및 아래에 개시된 변형된 헬리카제 참조).

서열번호: 118	E94	C109	C136	A360
서열번호: 119	E95	V124	T167	I388
서열번호: 120	D156	I180	S204	A417
서열번호: 121	E101	V119	V140	E342
서열번호: 122	E128	V147	V172	A377
서열번호: 123	E107	I125	S145	A364
서열번호: 124	D86	V107	C129	A342
서열번호: 125	D90	V110	C132	A349
서열번호: 126	D77	I89	C119	A347
서열번호: 127	E96	V119	A141	A336
서열번호: 128	E96	V119	A141	A360
서열번호: 129	E94	C114	C136	A357
서열번호: 130	D100	C120	C142	A363
서열번호: 131	D96	C116	C138	A365
서열번호: 132	E95	C115	C137	A360
서열번호: 133	E95	C115	C137	A361

[0470]

[0471] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기(즉 하나 이상의 시스테인 잔기) 및/또는 적어

도 하나의 비천연 아미노산(즉 하나 이상의 비천연 아미노산)이 (i) 타워 도메인 및/또는 (ii) 핀 도메인 및/또는 (iii) 1A(RecA-유사 모터) 도메인에 도입된 것이며, 여기서 헬리카제는 폴리뉴클레오티드의 이동을 제어하는 능력을 갖는다. 이러한 유형의 변형은 WO 2015/055981(이의 전문이 본원에 참조로 포함됨)에 개시되어 있다. 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산은 타워 도메인, 핀 도메인, 1A 도메인, 타워 도메인 및 핀 도메인, 타워 도메인 및 1A 도메인 또는 타워 도메인, 핀 도메인 및 1A 도메인에 도입될 수 있다.

[0472] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인 및 (ii) 핀 도메인 및/또는 1A(RecA-유사 모터) 도메인 각각, 즉 타워 도메인 및 핀 도메인, 타워 도메인 및 1A 도메인 또는 타워 도메인, 핀 도메인 및 1A 도메인에 도입된 것이다.

[0473] 임의의 수의 시스테인 잔기 및/또는 비천연 아미노산은 각 도메인에 도입될 수 있다. 예를 들면, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 시스테인 잔기가 도입될 수 있고/있거나 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 비천연 아미노산이 도입될 수 있다. 하나 이상의 시스테인 잔기만이 도입될 수 있다. 하나 이상의 비천연 아미노산만이 도입될 수 있다. 하나 이상의 시스테인 잔기 및 하나 이상의 비천연 아미노산의 조합이 도입될 수 있다.

[0474] 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 치환에 의해 도입된다. 이를 수행하는 방법은 당업계에 알려져 있다.

[0475] 이러한 변형은 헬리카제가 폴리뉴클레오티드에 결합하는 것을 막지 않는다. 이러한 변형은 폴리뉴클레오티드가 헬리카제로부터 결합해제되거나 풀리는 능력을 감소시킨다. 다시 말해서, 하나 이상의 변형은 폴리뉴클레오티드가닥으로부터 해리되는 것을 방지함으로써 헬리카제의 진행성을 증가시킨다. 효소의 열적 안정성은 전형적으로 또한 하나 이상의 변형에 의해 또한 증가되어 가닥 서열분석(Strand Sequencing)에 유익한 개선된 구조적 안정성을 제공한다.

[0476] 비천연 아미노산은 헬리카제에서 자연적으로 발견되지 않는 아미노이다. 비천연 아미노산은 바람직하게는 히스티딘, 알라닌, 이소류신, 아르기닌, 류신, 아스파라긴, 리신, 아스파르트산, 메티오닌, 시스테인, 페닐알라닌, 글루탐산, 트레오닌, 글루타민, 트립토판, 글리신, 발린, 프롤린, 세린 또는 티로신이 아니다. 비천연 아미노산은 보다 바람직하게는 이전 문장의 20개 아미노산 중 임의의 것 또는 셀레노시스테인이 아니다.

[0477] 본 발명에 사용하기에 바람직한 비천연 아미노산은 다음을 포함하나 이에 제한되지 않는다: 4-아지도-L-페닐알라닌(Faz), 4-아세틸-L-페닐알라닌, 3-아세틸-L-페닐알라닌, 4-아세토아세틸-L-페닐알라닌, 0-알릴-L-티로신, 3-(페닐셀라닐)-L-알라닌, 0-2-프로판-1-일-L-티로신, 4-(디하이드록시보틸)-L-페닐알라닌, 4-[(에틸술팜)카르보닐]-L-페닐알라닌, (2S)-2-아미노-3-4-[(프로판-2-일술팜)카르보닐]페닐; 프로판산, (2S)-2-아미노-3-4-[(2-아미노-3-술팜)프로판노일]아미노]페닐; 프로판산, 0-메틸-L-티로신, 4-아미노-L-페닐알라닌, 4-시아노-L-페닐알라닌, 3-시아노-L-페닐알라닌, 4-플루오로-L-페닐알라닌, 4-요오도-L-페닐알라닌, 4-브로모-L-페닐알라닌, 0-(트리플루오로메틸)티로신, 4-니트로-L-페닐알라닌, 3-하이드록시-L-티로신, 3-아미노-L-티로신, 3-요오도-L-티로신, 4-이소프로필-L-페닐알라닌, 3-(2-나프틸)-L-알라닌, 4-페닐-L-페닐알라닌, (2S)-2-아미노-3-(나프탈렌-2-일아미노)프로판산, 6-(메틸술팜)노르류신, 6-옥소-L-리신, D-티로신, (2R)-2-하이드록시-3-(4-하이드록시페닐)프로판산, (2R)-2-암모니오옥타노에이트-3-(2,2'-비피리딘-5-일)-D-알라닌, 2-아미노-3-(8-하이드록시-3-퀴놀릴)프로판산, 4-벤조일-L-페닐알라닌, S-(2-니트로벤질)시스테인, (2R)-2-아미노-3-[(2-니트로벤질)술팜]프로판산, (2S)-2-아미노-3-[(2-니트로벤질)옥시]프로판산, 0-(4,5-디메톡시-2-니트로벤질)-L-세린, (2S)-2-아미노-6-[(2-니트로벤질)옥시]카르보닐; 아미노]헥산산, 0-(2-니트로벤질)-L-티로신, 2-니트로페닐알라닌, 4-[(E)-페닐디아제닐]-L-페닐알라닌, 4-[3-(트리플루오로메틸)-3H-디아지렌-3-일]-D-페닐알라닌, 2-아미노-3-[[5-(디메틸아미노)-1-나프틸]술팜]아미노]프로판산, (2S)-2-아미노-4-(7-하이드록시-2-옥소-2H-크로멘-4-일)부탄산, (2S)-3-[(6-아세틸나프탈렌-2-일)아미노]-2-아미노프로판산, 4-(카르복시메틸)페닐알라닌, 3-니트로-L-티로신, 0-술포-L-티로신, (2R)-6-아세트아미도-2-암모니오헥사노에이트, 1-메틸히스티딘, 2-아미노노난산, 2-아미노데칸산, L-호모시스테인, 5-술팜노르발린, 6-술팜-L-노르류신, 5-(메틸술팜)노르발린, N⁶-[(2R,3R)-3-메틸-3,4-디하이드로-2H-피롤-2-일]카르보닐; -L-리신, N⁶-[(벤질옥시)카르보닐]리신, (2S)-2-아미노-6-[(사이클로펜틸카르보닐)아미노]헥산산, N⁶-[(사이클로펜틸옥시)카르보닐]-L-리신, (2S)-2-아미노-6-[(2R)-테트라하이드로푸란-2-일카르보닐]아미노; 헥산산, (2S)-2-아미노-8-[(2R,3S)-3-에틸테트라하이드로푸란-2-일]-8-옥소옥탄산, N⁶-(tert-부톡시카르보닐)-L-리신, (2S)-2-하이드록시-6-[(2-메틸-2-프로판)옥시]카르

보닐;아미노)헥산산, N^6 -[(알릴옥시)카르보닐]리신, (2S)-2-아미노-6-([(2-아지도벤질)옥시]카르보닐;아미노)헥산산, N^6 -L-프롤릴-L-리신, (2S)-2-아미노-6-[(프로프-2-인-1-일옥시)카르보닐]아미노;헥산산 및 N^6 -[(2-아지도에톡시)카르보닐]-L-리신. 가장 바람직한 비천연 아미노산은 4-아지도-L-페닐알라닌(Faz)이다.

[0478] 아래 표 10(이는 2개 부분으로 구분됨)은 각 Dda 상동체의 각 도메인을 구성하는 잔기를 식별한다(서열번호: 118 내지 133).

상동체	서열번호	1A	2A
Dda-Rma-DSM	119	M1-I84 + R113-Y211	R212-E294 + G422-S678
Dda-Csp	120	M1-L147 + S166-V240	R241-N327 + A449-G496
Dda-Sru	121	M1-L90 + E108-H173	R174-D260 + A371-V421
Dda-Sgo	122	M1-L115 + N136-V205	R206-K293 + I408-L500
Dda-Vph12B8	123	M1-L96 + F114-V194	R195-D287 + V394-Q450
Dda-Vph	124	M1-L77 + V96-V166	R167-T249 + L372-N421
Dda-Aph65	125	M1-M81 + L99-M171	R172-T254 + L381-K434
Dda-AphCC2	126	M1-M68 + M86-M158	R159-T241 + L367-K420
Dda-Cph	127	M1-L87 + A108-M181	R182-T262 + L393-V443
Dda-Kph	128	M1-L87 + A108-M181	R182-T262 + L392-V442
Dda-SphIME13	129	M1-L85 + T103-K176	R177-N257 + L387-V438
Dda-AphAc42	130	M1-L91 + V109-M183	R184-T265 + L393-I442
Dda-SphSP18	131	M1-L87 + M105-M179	R180-T261 + L393-V442
Dda-Yph	132	M1-L86 + V104-K178	R179-T260 + L390-I439
Dda-SphS16	133	M1-L86 + V104-M178	R179-T260 + L391-V441
Dda-1993	118	M1-L85 + V103-K177	R178-T259 + L390-V439

[0479]

상동체	서열번호	타워	핀	후크
Dda-Rma-DSM	119	G295-N309 + F316-Y421	Y85-L112	A310-L315
Dda-Csp	120	V328-P342 + N360-Y448	K148-N165	V343-L359
Dda-Sru	121	A261-T275 + T285-Y370	G91-E107	W276-L284
Dda-Sgo	122	G294-I307 + T314-Y407	G116-T135	R308-Y313
Dda-Vph12B8	123	V288-E301 + N307-N393	G97-P113	M302-W306
Dda-Vph	124	S250-P264 + E278-S371	K78-E95	V265-I277
Dda-Aph65	125	K255-P269 + T284-S380	K82-K98	V270-F283
Dda-AphCC2	126	D242-P256 + T271-S366	K69-K85	V257-F270
Dda-Cph	127	T263-P277 + N295-P392	K88-K107	L278-Y294
Dda-Kph	128	D263-P277 + N295-A391	K88-K107	L278-Y294
Dda-SphIME13	129	A258-P272 + N290-P386	K86-G102	L273-F289
Dda-AphAc42	130	L266-P280 + N298-A392	K92-D108	L281-F297
Dda-SphSP18	131	D262-P276 + N294-A392	K88-E104	H277-F293
Dda-Yph	132	D261-P275 + N293-A389	K87-E103	L276-F292
Dda-SphS16	133	E261-P275 + T293-A390	K87-E103	L276-F292
Dda-1993	118	D260-P274 + N292-A389	K86-E102	L275-F291

[0480]

[0481] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 D260-P274 및 N292-A389) 및/또는 (ii) 핀 도메인(잔기 K86-E102) 및/또는 (iii) 1A 도메인(잔기 M1-L85 및 V103-K177)에 도입된 서열번호: 118의 변이체를 포함한다. 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N292-A389에 도입된다.

[0482] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 G295-N309 및 F316-Y421) 및/또는 (ii) 핀 도메인(잔기 Y85-L112) 및/또는 (iii) 1A 도메인(잔기 M1-I84 및 R113-Y211)에 도입된 서열번호: 119의 변이체를 포함한다. 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 F316-Y421에 도입된다.

[0483] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 V328-P342 및 N360-Y448) 및/또는 (ii) 핀 도메인(잔기 K148-N165) 및/또는 (iii) 1A 도메인(잔기 M1-L147 및 S166-V240)에 도입된 서열번호: 120의 변이체를 포함한다. 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N360-Y448에 도입된다.

[0484] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 A261-T275 및 T285-Y370) 및/또는 (ii) 핀 도메인(잔기 G91-E107) 및/또는 (iii) 1A 도메인(잔기 M1-L90 및 E108-H173)에 도입된 서열번호: 121의 변이체를 포함한다. 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 T285-Y370에 도입된다.

- [0485] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 G294-I307 및 T314-Y407) 및/또는 (ii) 핀 도메인(잔기 G116-T135) 및/또는 (iii) 1A 도메인(잔기 M1-L115 및 N136-V205)에 도입된 서열번호: 122의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 T314-Y407에 도입된다.
- [0486] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 V288-E301 및 N307-N393) 및/또는 (ii) 핀 도메인(잔기 G97-P113) 및/또는 (iii) 1A 도메인(잔기 M1-L96 및 F114-V194)에 도입된 서열번호: 123의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N307-N393에 도입된다.
- [0487] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 S250-P264 및 E278-S371) 및/또는 (ii) 핀 도메인(잔기 K78-E95) 및/또는 (iii) 1A 도메인(잔기 M1-L77 및 V96-V166)에 도입된 서열번호: 124의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 E278-S371에 도입된다.
- [0488] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 K255-P269 및 T284-S380) 및/또는 (ii) 핀 도메인(잔기 K82-K98) 및/또는 (iii) 1A 도메인(잔기 M1-M81 및 L99-M171)에 도입된 서열번호: 125의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 T284-S380에 도입된다.
- [0489] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 D242-P256 및 T271-S366) 및/또는 (ii) 핀 도메인(잔기 K69-K85) 및/또는 (iii) 1A 도메인(잔기 M1-M68 및 M86-M158)에 도입된 서열번호: 126의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 T271-S366에 도입된다.
- [0490] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 T263-P277 및 N295-P392) 및/또는 (ii) 핀 도메인(잔기 K88-K107) 및/또는 (iii) 1A 도메인(잔기 M1-L87 및 A108-M181)에 도입된 서열번호: 127의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N295-P392에 도입된다.
- [0491] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 D263-P277 및 N295-A391) 및/또는 (ii) 핀 도메인(잔기 K88-K107) 및/또는 (iii) 1A 도메인(잔기 M1-L87 및 A108-M181)에 도입된 서열번호: 128의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N295-A391에 도입된다.
- [0492] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 A258-P272 및 N290-P386) 및/또는 (ii) 핀 도메인(잔기 K86-G102) 및/또는 (iii) 1A 도메인(잔기 M1-L85 및 T103-K176)에 도입된 서열번호: 129의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N290-P386에 도입된다.
- [0493] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 L266-P280 및 N298-A392) 및/또는 (ii) 핀 도메인(잔기 K92-D108) 및/또는 (iii) 1A 도메인(잔기 M1-L91 및 V109-M183)에 도입된 서열번호: 130의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N298-A392에 도입된다.
- [0494] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 D262-P276 및 N294-A392) 및/또는 (ii) 핀 도메인(잔기 K88-E104) 및/또는 (iii) 1A 도메인(잔기 M1-L87 및 M105-M179)에 도입된 서열번호: 131의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N294-A392에 도입된다.
- [0495] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 D261-P275 및 N293-A389) 및/또는 (ii) 핀 도메인(잔기 K87-E103) 및/또는 (iii) 1A 도메인(잔기 M1-L86 및 V104-K178)에 도입된 서열번호: 132의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 N293-A389에 도입된다.
- [0496] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인(잔기 E261-P275 및 T293-A390) 및/또는 (ii) 핀 도메인(잔기 K87-E103) 및/또는 (iii) 1A 도

메인(잔기 M1-L86 및 V104-M178)에 도입된 서열번호: 133의 변이체를 포함한다. 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 타워 도메인의 잔기 T293-A390에 도입된다.

[0497] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인 및 (ii) 핀 도메인 및/또는 1A 도메인 각각에 도입된 서열번호: 118 내지 133 중 임의의 하나의 변이체를 포함한다. 본 발명의 헬리카제는 보다 바람직하게는 적어도 하나의 시스템인 잔기 및/또는 적어도 하나의 비천연 아미노산이 (i) 타워 도메인, (ii) 핀 도메인 및 (iii) 1A 도메인 각각에 도입된 서열번호: 118 내지 133 중 임의의 하나의 변이체를 포함한다. 시스템인 잔기 및 비천연 아미노산의 임의의 수 및 조합은 위에 논의된 바와 같이 도입될 수 있다.

[0498] 본 발명의 헬리카제는 바람직하게는 다음을 포함하는 서열번호: 118의 변이체를 포함한다: (i) E94C 및/또는 A360C; (ii) E93C 및/또는 K358C; (iii) E93C 및/또는 A360C; (iv) E93C 및/또는 E361C; (v) E93C 및/또는 K364C; (vi) E94C 및/또는 L354C; (vii) E94C 및/또는 K358C; (viii) E93C 및/또는 L354C; (ix) E94C 및/또는 E361C; (x) E94C 및/또는 K364C; (xi) L97C 및/또는 L354C; (xii) L97C 및/또는 K358C; (xiii) L97C 및/또는 A360C; (xiv) L97C 및/또는 E361C; (xv) L97C 및/또는 K364C; (xvi) K123C 및/또는 L354C; (xvii) K123C 및/또는 K358C; (xviii) K123C 및/또는 A360C; (xix) K123C 및/또는 E361C; (xx) K123C 및/또는 K364C; (xxi) N155C 및/또는 L354C; (xxii) N155C 및/또는 K358C; (xxiii) N155C 및/또는 A360C; (xxiv) N155C 및/또는 E361C; (xxv) N155C 및/또는 K364C; (xxvi) (i) 내지 (xxv) 중 임의의 것 및 G357C; (xxvii) (i) 내지 (xxv) 중 임의의 것 및 Q100C; (xxviii) (i) 내지 (xxv) 중 임의의 것 및 I127C; (xxix) (i) 내지 (xxv) 중 임의의 것 및 Q100C 및 I127C; (xxx) E94C 및/또는 F377C; (xxxi) N95C; (xxxii) T91C; (xxxiii) Y92L, E94Y, Y350N, A360C 및 Y363N; (xxxiv) E94Y 및 A360C; (xxxv) A360C; (xxxvi) Y92L, E94C, Y350N, A360Y 및 Y363N; (xxxvii) Y92L, E94C 및 A360Y; (xxxviii) E94C 및/또는 A360C 및 F276A; (xxxix) E94C 및/또는 L356C; (xl) E93C 및/또는 E356C; (xli) E93C 및/또는 G357C; (xlii) E93C 및/또는 A360C; (xliii) N95C 및/또는 W378C; (xliv) T91C 및/또는 S382C; (xlv) T91C 및/또는 W378C; (xlvi) E93C 및/또는 N353C; (xlvii) E93C 및/또는 S382C; (xlviii) E93C 및/또는 K381C; (xlix) E93C 및/또는 D379C; (l) E93C 및/또는 S375C; (li) E93C 및/또는 W378C; (lii) E93C 및/또는 W374C; (liii) E94C 및/또는 N353C; (liv) E94C 및/또는 S382C; (lv) E94C 및/또는 K381C; (lvi) E94C 및/또는 D379C; (lvii) E94C 및/또는 S375C; (lviii) E94C 및/또는 W378C; (lix) E94C 및/또는 W374C; (lx) E94C 및 A360Y; (lxi) E94C, G357C 및 A360C 또는 (lxii) T2C, E94C 및 A360C. (i) 내지 (lxii) 중 임의의 하나에서, 및/또는 바람직하게는 및 이다.

[0499] 본 발명의 헬리카제는 바람직하게는 (i) 내지 (lxii) 중 임의의 것에 정의된 바와 같은 서열번호: 118의 위치에 상응하는 위치에서 시스템인 잔기를 포함하는 서열번호: 119 내지 133 중 임의의 하나의 변이체를 포함한다. 서열번호: 118의 위치에 상응하는 서열번호: 119 내지 133 중 임의의 하나의 위치는 아래 서열번호: 118 내지 133의 정렬을 사용하여 식별될 수 있다. 본 발명의 헬리카제는 바람직하게는 (a) D99C 및/또는 L341C, (b) Q98C 및/또는 L341C 또는 (d) Q98C 및/또는 A340C를 포함하는 서열번호: 92의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 D90C 및/또는 A349C를 포함하는 서열번호: 96의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 D96C 및/또는 A362C를 포함하는 서열번호: 102의 변이체를 포함한다.

[0500] 본 발명의 헬리카제는 바람직하게는 Faz가 시스템인 대신 특이적 위치 중 하나 이상에 도입된 (i) 내지 (lxii) 중 임의의 하나에 정의된 바와 같은 서열번호: 118 내지 133 중 임의의 하나의 변이체를 포함한다. Faz는 시스템인 대신 각 특이적 위치에 도입될 수 있다. 본 발명의 헬리카제는 바람직하게는 (i) E94Faz 및/또는 A360C; (ii) E94C 및/또는 A360Faz; (iii) E94Faz 및/또는 A360Faz; (iv) Y92L, E94Y, Y350N, A360Faz 및 Y363N; (v) A360Faz; (vi) E94Y 및 A360Faz; (vii) Y92L, E94Faz, Y350N, A360Y 및 Y363N; (viii) Y92L, E94Faz 및 A360Y; (ix) E94Faz 및 A360Y; 및 (x) E94C, G357Faz 및 A360C를 포함하는 서열번호: 118의 변이체를 포함한다.

[0501] 본 발명의 헬리카제는 바람직하게는 핀 도메인으로부터 하나 이상의 단일 아미노산 결실을 추가로 포함한다. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상과 같은 임의의 수의 단일 아미노산 결실이 이루어질 수 있다. 헬리카제는 보다 바람직하게는 E93의 결실, E95의 결실 또는 E93 및 E95의 결실을 포함하는 서열번호: 118의 변이체를 포함한다. 헬리카제는 보다 바람직하게는 (a) E94C, N95의 결실 및 A360C; (b) E93의 결실, E94의 결실, N95의 결실 및 A360C; (c) E93의 결실, E94C, N95의 결실 및 A360C 또는 (d) E93C, N95의 결실 및 A360C을 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 E93에 상응하는 위치의 결실, 서열번호: 118의 E95에 상응하는 위치의 결실 또는 서열번호: 118의 E93 및 E95에 상응하는 위치의 결실

을 포함하는 서열번호: 119 내지 133 중 임의의 하나의 변이체를 포함한다.

- [0502] 본 발명의 헬리카제는 바람직하게는 후크 도메인에서 하나 이상의 단일 아미노산 결실을 추가로 포함한다. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상과 같은 임의의 수의 단일 아미노산 결실이 이루어질 수 있다. 헬리카제는 보다 바람직하게는 위치 T278 내지 S287의 임의의 수의 결실을 포함하는 서열번호: 118의 변이체를 포함한다. 헬리카제는 보다 바람직하게는 (a) E94C, Y279 내지 K284의 결실 및 A360C, (b) E94C, T278, Y279, V286 및 S287의 결실 및 A360C, (c) E94C, I281 및 K284의 결실 및 단일 G로의 대체 및 A360C, (d) E94C, K280 및 P2845의 결실 및 단일 G로의 대체 및 A360C, 또는 (e) Y279 내지 K284의 결실, E94C, F276A 및 A230C를 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 278 내지 287에 상응하는 위치의 임의의 수의 결실을 포함하는 서열번호: 119 내지 133 중 임의의 하나의 변이체를 포함한다.
- [0503] 본 발명의 헬리카제는 바람직하게는 핀 도메인에서 하나 이상의 단일 아미노산 결실 및 후크 도메인에서 하나 이상의 단일 아미노산 결실을 추가로 포함한다.
- [0504] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인 및/또는 2A(RecA-유사) 도메인에 추가로 도입된 것이다. 타워, 핀 및 1A 도메인에 대해 위에 논의된 바와 같이 시스테인 잔기 및 비천연 아미노산의 임의의 수 및 조합이 도입될 수 있다.
- [0505] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 L275-F291) 및/또는 2A(RecA-유사) 도메인(잔기 R178-T259 및 L390-V439)에 추가로 도입된 서열번호: 118의 변이체를 포함한다.
- [0506] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 A310-L315) 및/또는 2A(RecA-유사) 도메인(잔기 R212-E294 및 G422-S678)에 추가로 도입된 서열번호: 119의 변이체를 포함한다.
- [0507] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 V343-L359) 및/또는 2A(RecA-유사) 도메인(잔기 R241-N327 및 A449-G496)에 추가로 도입된 서열번호: 120의 변이체를 포함한다.
- [0508] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 W276-L284) 및/또는 2A(RecA-유사) 도메인(잔기 R174-D260 및 A371-V421)에 추가로 도입된 서열번호: 121의 변이체를 포함한다.
- [0509] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 R308-Y313) 및/또는 2A(RecA-유사) 도메인(잔기 R206-K293 및 I408-L500)에 추가로 도입된 서열번호: 122의 변이체를 포함한다.
- [0510] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 M302-W306) 및/또는 2A(RecA-유사) 도메인(잔기 R195-D287 및 V394-Q450)에 추가로 도입된 서열번호: 123의 변이체를 포함한다.
- [0511] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 V265-I277) 및/또는 2A(RecA-유사) 도메인(잔기 R167-T249 및 L372-N421)에 추가로 도입된 서열번호: 124의 변이체를 포함한다.
- [0512] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 V270-F283) 및/또는 2A(RecA-유사) 도메인(잔기 R172-T254 및 L381-K434)에 추가로 도입된 서열번호: 125의 변이체를 포함한다.
- [0513] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 V257-F270) 및/또는 2A(RecA-유사) 도메인(잔기 R159-T241 및 L367-K420)에 추가로 도입된 서열번호: 126의 변이체를 포함한다.
- [0514] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 L278-Y294) 및/또는 2A(RecA-유사) 도메인(잔기 R182-T262 및 L393-V443)에 추가로 도입된 서열번호: 127의 변이체를 포함한다.
- [0515] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후

크 도메인(잔기 L278-Y294) 및/또는 2A(RecA-유사) 도메인(잔기 R182-T262 및 L392-V442)에 추가로 도입된 서열번호: 128의 변이체를 포함한다.

[0516] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 L273-F289) 및/또는 2A(RecA-유사) 도메인(잔기 R177-N257 및 L387-V438)에 추가로 도입된 서열번호: 129의 변이체를 포함한다.

[0517] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 L281-F297) 및/또는 2A(RecA-유사) 도메인(잔기 R184-T265 및 L393-I442)에 추가로 도입된 서열번호: 130의 변이체를 포함한다.

[0518] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 H277-F293) 및/또는 2A(RecA-유사) 도메인(잔기 R180-T261 및 L393-V442)에 추가로 도입된 서열번호: 131의 변이체를 포함한다.

[0519] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 L276-F292) 및/또는 2A(RecA-유사) 도메인(잔기 R179-T260 및 L390-I439)에 추가로 도입된 서열번호: 132의 변이체를 포함한다.

[0520] 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인(잔기 L276-F292) 및/또는 2A(RecA-유사) 도메인(잔기 R179-T260 및 L391-V441)에 추가로 도입된 서열번호: 133의 변이체를 포함한다.

[0521] 본 발명의 헬리카제는 바람직하게는 (i) I181C; (ii) Y279C; (iii) I281C; 및 (iv) E288C 중 하나 이상을 포함하는 서열번호: 118의 변이체를 포함한다. 헬리카제는 (i) 내지 (iv)의 임의의 조합, 예컨대 (i); (ii); (iii); (iv); (i) 및 (ii); (i) 및 (iii); (i) 및 (iv); (ii) 및 (iii); (ii) 및 (iv); (iii) 및 (iv); 또는 (i), (ii), (iii) 및 (iv)를 포함할 수 있다. 헬리카제는 보다 바람직하게는 (a) E94C, I281C 및 A360C 또는 (b) E94C, I281C, G357C 및 A360C를 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 (i) 내지 (iv), (a) 및 (b)에 정의된 바와 같은 서열번호: 118의 위치에 상응하는 위치(들) 중 하나 이상에서 시스테인 잔기를 포함하는 서열번호: 119 내지 133 중 임의의 하나의 변이체를 포함한다. 헬리카제는 Faz가 시스테인 대신 특이적 위치 중 하나 이상(또는 각 특이적 위치)에 도입된 이러한 변이체 중 임의의 것을 포함할 수 있다.

[0522] 본 발명의 헬리카제는 표면 음전하를 감소시키도록 추가로 변형된다. 표면 잔기는 위에 개시된 Dda 도메인과 동일한 방식으로 식별될 수 있다. 표면 음전하는 전형적으로 아스파르트산(D) 및 글루탐산(E)과 같은 표면 음으로-하전된 아미노산이다.

[0523] 헬리카제는 바람직하게는 하나 이상의 음으로 하전된 아미노산을 하나 이상의 양으로 하전된 아미노산, 비하전된 아미노산, 비극성 아미노산 및/또는 방향족 아미노산으로 치환하거나 하나 이상의 양으로 하전된 아미노산을 바람직하게는 하나 이상의 음으로 하전된 아미노산에 인접하게 도입함으로써 하나 이상의 표면 음전하를 중화하도록 변형된다. 적합한 양으로 하전된 아미노산은 히스티딘(H), 리신(K) 및 아르기닌(R)을 포함하나 이에 제한되지 않는다. 비하전된 아미노산은 순 전하가 없다. 적합한 비하전된 아미노산은 시스테인(C), 세린(S), 트레오닌(T), 메티오닌(M), 아스파라긴(N) 및 글루타민(Q)을 포함하나 이에 제한되지 않는다. 비극성 아미노산은 비극성 측쇄를 갖는다. 적합한 비극성 아미노산은 글리신(G), 알라닌(A), 프롤린(P), 이소류신(I), 류신(L) 및 발린(V)을 포함하나 이에 제한되지 않는다. 방향족 아미노산은 방향족 측쇄를 갖는다. 적합한 방향족 아미노산은 히스티딘(H), 페닐알라닌(F), 트립토판(W) 및 티로신(Y)을 포함하나 이에 제한되지 않는다.

[0524] 바람직한 치환은 E를 R로 치환, E를 K로 치환, E를 N으로 치환, D를 K로 치환 및 D를 R로 치환을 포함하나 이에 제한되지 않는다.

[0525] 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 변이체를 포함하고 하나 이상의 음으로 하전된 아미노산은 D5, E8, E23, E47, D167, E172, D202, D212 및 E273 중 하나 이상이다. 이러한 아미노산 중 1, 2, 3, 4, 5, 6, 7 또는 8개와 같은 임의의 수의 아미노산이 중화될 수 있다. 임의의 조합이 중화될 수 있다. 본 발명의 헬리카제는 바람직하게는 서열번호: 119 내지 133 중 임의의 하나의 변이체를 포함하고 하나 이상의 음으로 하전된 아미노산은 서열번호: 118의 D5, E8, E23, E47, D167, E172, D202, D212 및 E273 중 하나 이상에 상응한다. 서열번호: 118의 D5, E8, E23, E47, D167, E172, D202, D212 및 E273에 상응하는 서열번호: 119 내지 133의 아미노산은 W02015/055981의 정렬을 사용하여 결정될 수 있다. 본 발명의 헬리카제는 바람직하게는 (a) E94C, E273G

및 A360C 또는 (b) E94C, E273G, N292G 및 A360C를 포함하는 서열번호: 118의 변이체를 포함한다.

[0526] 본 발명의 헬리카제는 바람직하게는 하나 이상의 천연 시스테인 잔기를 제거하여 추가로 변형된다. 임의의 수의 천연 시스테인 잔기가 제거될 수 있다. 하나 이상의 시스테인 잔기는 바람직하게는 치환에 의해 제거된다. 하나 이상의 시스테인 잔기는 바람직하게는 알라닌(A), 세린(S) 또는 발린(V)으로 치환된다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 변이체를 포함하고 하나 이상의 천연 시스테인 잔기는 C109, C114, C136, C171 및 C412 중 하나 이상이다. 이러한 시스테인 잔기의 임의의 수 및 조합이 제거될 수 있다. 예를 들면, 서열번호: 118의 변이체는 C109; C114; C136; C171; C412; C109 및 C114; C109 및 C136; C109 및 C171; C109 및 C412; C114 및 C136; C114 및 C171; C114 및 C412; C136 및 C171; C136 및 C412; C171 및 C412; C109, C114 및 C136; C109, C114 및 C171; C109, C114 및 C412; C109, C136 및 C171; C109, C136 및 C412; C109, C171 및 C412; C114, C136 및 C171; C114, C136 및 C412; C114, C171 및 C412; C136, C171 및 C412; C109, C114, C136 및 C171; C109, C114, C136 및 C412; C109, C114, C171 및 C412; C109, C136, C171 및 C412; C114, C136, C171 및 C412; 또는 C109, C114, C136, C171 및 C412를 포함할 수 있다.

[0527] 변형된 헬리카제는 바람직하게는 Dda 1993의 아미노산 위치(들) 109 및/또는 136에 상응하는 위치(들)에서 변형 또는 치환을 포함한다. 이는 1 또는 2개의 시스테인 잔기를 제거한다. 이는 Dda 1993의 아미노산 위치 55, 114, 156, 177, 210, 221, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또는 치환, Dda 1993의 아미노산 위치 114, 177, 350 및 358에 상응하는 하나 이상의 위치에서 변형 또는 치환 및/또는 Dda 1993의 위치 40에 상응하는 위치에서 변형 또는 치환에 더해질 수 있다. 위치 109 또는 상응하는 위치는 A, V, I, L, M, F, Y 또는 W로 치환될 수 있다. 위치 109 또는 상응하는 위치는 바람직하게는 A 또는 V로 치환된다. 위치 136 또는 상응하는 위치는 A, V, I, L, M, F, Y 또는 W로 치환될 수 있다. 위치 136 또는 상응하는 위치는 바람직하게는 A 또는 V로 치환된다. 본 발명의 헬리카제는 바람직하게는 C109A, C109V, C109I, C109L, C109M, C109F, C109Y 또는 C109W와 같은 C109 및/또는 C136, 예컨대 C136A, C136V, C136I, C136L, C136M, C136F, C136Y 또는 C136W와 같은 C136에서 치환을 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 C109A 및/또는 C136A를 포함하는 서열번호: 118의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 서열번호: 118의 C109 및/또는 C136에 상응하는 위치(들)에서 치환을 포함하는 서열번호: 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132 또는 133의 변이체를 포함한다. 본 발명의 헬리카제는 바람직하게는 적어도 하나의 시스테인 잔기(즉 하나 이상의 시스테인 잔기) 및/또는 적어도 하나의 비천연 아미노산(즉 하나 이상의 비천연 아미노산)이 타워 도메인에만 도입된 것이다. 적합한 변형은 위에 논의되어 있다.

[0528] 본 발명의 헬리카제는 바람직하게는 다음 돌연변이를 포함하는 서열번호: 118의 변이체를 포함한다: E93C 및 K364C; E94C 및 K364C; E94C 및 A360C; L97C 및 E361C; L97C 및 E361C 및 C412A; K123C 및 E361C; K123C, E361C 및 C412A; N155C 및 K358C; N155C, K358C 및 C412A; N155C 및 L354C; N155C, L354C 및 C412A; 델타E93, E94C, 델타N95 및 A360C; E94C, 델타N95 및 A360C; E94C, Q100C, I127C 및 A360C; L354C; G357C; E94C, G357C 및 A360C; E94C, Y279C 및 A360C; E94C, I281C 및 A360C; E94C, Y279Faz 및 A360C; Y279C 및 G357C; I281C 및 G357C; E94C, Y279C, G357C 및 A360C; E94C, I281C, G357C 및 A360C; E8R, E47K, E94C, D202K 및 A360C; D5K, E23N, E94C, D167K, E172R, D212R 및 A360C; D5K, E8R, E23N, E47K, E94C, D167K, E172R, D202K, D212R 및 A360C; E94C, C114A, C171A, A360C 및 C412D; E94C, C114A, C171A, A360C 및 C412S; E94C, C109A, C136A 및 A360C; E94C, C109A, C114A, C136A, C171A, A360C 및 C412S; E94C, C109V, C114V, C171A, A360C 및 C412S; C109A, C114A, C136A, G153C, C171A, E361C 및 C412A; C109A, C114A, C136A, G153C, C171A, E361C 및 C412D; C109A, C114A, C136A, G153C, C171A, E361C 및 C412S; C109A, C114A, C136A, G153C, C171A, K358C 및 C412A; C109A, C114A, C136A, G153C, C171A, K358C 및 C412D; C109A, C114A, C136A, G153C, C171A, K358C 및 C412S; C109A, C114A, C136A, N155C, C171A, K358C 및 C412A; C109A, C114A, C136A, N155C, C171A, K358C 및 C412D; C109A, C114A, C136A, N155C, C171A, K358C 및 C412S; C109A, C114A, C136A, N155C, C171A, L354C 및 C412A; C109A, C114A, C136A, N155C, C171A, L354C 및 C412D; C109A, C114A, C136A, N155C, C171A, L354C 및 C412S; C109A, C114A, K123C, C136A, C171A, E361C 및 C412A; C109A, C114A, K123C, C136A, C171A, E361C 및 C412D; C109A, C114A, K123C, C136A, C171A, K358C 및 C412A; C109A, C114A, K123C, C136A, C171A, K358C 및 C412D; C109A, C114A, K123C, C136A, C171A, K358C 및 C412S; C109A, C114A, K123C, C136A, C171A, K358C 및 C412S; C109A, C114A, C136A, G153C, C171A, E361C 및 C412A; E94C, C109A, C114A, C136A, C171A, A360C 및 C412D; E94C, C109A, C114V, C136A, C171A, A360C 및 C412D; E94C, C109V, C114A, C136A, C171A, A360C 및 C412D; L97C, C109A, C114A, C136A, C171A, E361C 및 C412A; L97C, C109A, C114A, C136A, C171A, E361C 및 C412D; 또는 L97C, C109A, C114A, C136A, C171A, E361C 및 C412S.

- [0529] 후크 도메인 및/또는 2A 도메인에서의 변형
- [0530] 하나의 구현예에서, 본 발명의 헬리카제는 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산이 후크 도메인 및/또는 2A(RecA-유사 모터) 도메인에 도입된 것이며, 여기서 헬리카제는 폴리뉴클레오티드의 이동을 제어하는 능력을 갖는다. 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 후크 도메인 및 2A(RecA-유사 모터) 도메인에 도입된다.
- [0531] 임의의 수의 시스테인 잔기 및/또는 비천연 아미노산이 각 도메인에 도입될 수 있다. 예를 들면, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 시스테인 잔기가 도입될 수 있고/있거나 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 비천연 아미노산이 도입될 수 있다. 하나 이상의 시스테인 잔기만이 도입될 수 있다. 하나 이상의 비천연 아미노산만이 도입될 수 있다. 하나 이상의 시스테인 잔기 및 하나 이상의 비천연 아미노산의 조합이 도입될 수 있다.
- [0532] 적어도 하나의 시스테인 잔기 및/또는 적어도 하나의 비천연 아미노산은 바람직하게는 치환에 의해 도입된다. 이를 수행하는 방법은 당업계에 알려져 있다. 후크 도메인 및/또는 2A(RecA-유사 모터) 도메인의 적합한 변형은 위에 논의되어 있다.
- [0533] 본 발명의 헬리카제는 바람직하게는 (a) Y279C, I181C, E288C, Y279C 및 I181C, (b) Y279C 및 E288C, (c) I181C 및 E288C 또는 (d) Y279C, I181C 및 E288C를 포함하는 서열번호: 118의 변이체이다. 본 발명의 헬리카제는 바람직하게는 (a) 내지 (d)에 정의된 바와 같이 서열번호: 118의 위치에 상응하는 위치(들) 중 하나 이상에서 돌연변이를 포함하는 서열번호: 199 내지 133 중 임의의 하나의 변이체를 포함한다.
- [0534] 표면 변형
- [0535] 하나의 구현예에서, 헬리카제는 표면 음전하를 감소시키도록 변형되며, 여기서 헬리카제는 폴리뉴클레오티드의 이동을 제어하는 능력을 갖는다. 적합한 변형은 위에 논의되어 있다. 임의의 수의 표면 음전하가 중화될 수 있다.
- [0536] 본 발명의 헬리카제는 바람직하게는 다음 돌연변이를 포함하는 서열번호: 118의 변이체를 포함한다: E273G; E8R, E47K 및 D202K; D5K, E23N, D167K, E172R 및 D212R; 또는 D5K, E8R, E23N, E47K, D167K, E172R, D202K 및 D212R.
- [0537] 다른 변형된 헬리카제
- [0538] 하나의 구현예에서, 본 발명의 헬리카제는 A360K; Y92L 및/또는 A360Y; Y92L, Y350N 및 Y363N; Y92L 및/또는 Y363N; 또는 Y92L을 포함하는 서열번호: 118의 변이체를 포함한다.
- [0539] 다른 변형
- [0540] 위에 개시된 특이적 돌연변이 외에도, 서열번호: 118의 변이체는 다음 돌연변이 중 하나 이상을 포함할 수 있다: K38A; T91F; T91N; T91Q; T91W; V96E; V96F; V96L; V96Q; V96R; V96W; V96Y; P274G; V286F; V286W; V286Y; F291G; N292F; N292G; N292P; N292Y; G294Y; G294F; K364A; 및 W378A.
- [0541] 위에 논의된 특이적 돌연변이 외에도, 서열번호: 118의 변이체는 다음을 포함할 수 있다: K38A, E94C 및 A360C; H64K; E94C 및 A360C; H64N; E94C 및 A360C; H64Q; E94C 및 A360C; H64S; E94C 및 A360C; H64W, E94C 및 A360C; T80K, E94C 및 A360C; T80K, S83K, E94C, N242K, N293K 및 A360C; T80K, S83K, E94C, N242K, N293K, A360C 및 T394K; T80K, S83K, E94C, N293K 및 A360C; T80K, S83K, E94C, A360C 및 T394K; T80K, S83K, E94C, A360C 및 T394N; T80K, E94C, N242K 및 A360C; T80K, E94C, N242K, N293K 및 A360C; T80K, E94C, N293K 및 A360C; T80N, E94C 및 A360C; H82A, E94C 및 A360C; H82A, P89A, E94C, F98A 및 A360C; H82F, E94C 및 A360C; H82Q, E94C, A360C; H82R, E94C 및 A360C; H82W, E94C 및 A360C; H82W, P89W, E94C, F98W 및 A360C; H82Y, E94C 및 A360C; S83K, E94C 및 A360C; S83K, T80K, E94C, A360C 및 T394K; S83N, E94C 및 A360C; S83T, E94C 및 A360C; N88H, E94C 및 A360C; N88Q, E94C 및 A360C; P89A, E94C 및 A360C; P89A, F98W, E94C 및 A360C; P89A, E94C, F98Y 및 A360C; P89A, E94C, F98A 및 A360C; P89F, E94C 및 A360C; P89S, E94C 및 A360C; P89T, E94C 및 A360C; P89W, E94C, F98W 및 A360C; P89Y, E94C 및 A360C; T91F, E94C 및 A360C; T91N, E94C 및 A360C; T91Q, E94C 및 A360C; T91W, E94C 및 A360C; E94C, V96E 및 A360C; E94C, V96F 및 A360C; E94C, V96L 및 A360C; E94C, V96Q 및 A360C; E94C, V96R 및 A360C; E94C, V96W 및 A360C; E94C, V96Y 및 A360C; E94C, F98A 및 A360C; E94C, F98L 및 A360C; E94C, F98V 및 A360C; E94C, F98Y 및 A360C; E94C, F98W 및 A360C; E94C, V150A 및 A360C; E94C, V150F 및 A360C; E94C, V150I 및 A360C; E94C, V150K 및 A360C; E94C, V150L 및

A360C; E94C, V150S 및 A360C; E94C, V150T 및 A360C; E94C, V150W 및 A360C; E94C, V150Y 및 A360C; E94C, F240Y 및 A360C; E94C, F240W 및 A360C; E94C, N242K 및 A360C; E94C, N242K, N293K 및 A360C; E94C, P274G 및 A360C; E94C, L275G 및 A360C; E94C, F276A 및 A360C; E94C, F276I 및 A360C; E94C, F276M 및 A360C; E94C, F276V 및 A360C; E94C, F276W 및 A360C; E94C, F276Y 및 A360C; E94C, V286F 및 A360C; E94C, V286W 및 A360C; E94C, V286Y 및 A360C; E94C, S287F 및 A360C; E94C, S287W 및 A360C; E94C, S287Y 및 A360C; E94C, F291G 및 A360C; E94C, N292F 및 A360C; E94C, N292G 및 A360C; E94C, N292P 및 A360C; E94C, N292Y 및 A360C; E94C, N293F 및 A360C; E94C, N293K 및 A360C; E94C, N293Q 및 A360C; E94C, N293Y 및 A360C; E94C, G294F 및 A360C; E94C, G294Y 및 A360C; E94C, A36C 및 K364A; E94C, A360C, W378A; E94C, A360C 및 T394K; E94C, A360C 및 H396Q; E94C, A360C 및 H396S; E94C, A360C 및 H396W; E94C, A360C 및 Y415F; E94C, A360C 및 Y415K; E94C, A360C 및 Y415M; 또는 E94C, A360C 및 Y415W.

- [0542] 본 발명의 헬리카제는 바람직하게는 (a) E94C/A360C/W378A 또는 (b) E94C/A360C/C109A/C136A/W378A 또는 (d) E94C/A360C/C109A/C136A/W378A를 포함한 다음 ($\Delta M1$)G1G2(즉 M1의 결실 이어서 G1 및 G2 첨가)를 포함하는 서열번호: 118의 변이체를 포함한다.
- [0543] 서열번호: 118 내지 133 중 임의의 하나의 바람직한 변이체는 (본 발명의 변형 외에도) 하나의 글리신 잔기(G)로 대체된 N-말단 메티오닌(M)을 갖는다. 예에서 이는 ($\Delta M1$)G1로 나타낸다. 또한 M1G라고 부를 수 있다. 위에 논의된 변이체 중 임의의 것은 M1G를 추가로 포함할 수 있다.
- [0544] 본 발명의 가장 바람직한 헬리카제는 다음을 포함하는 서열번호: 118의 변이체를 포함한다: (a) E94C/F98W/A360C/C109A/C136A/K194L, (b) M1G/E94C/F98W/A360C/C109A/C136A/K194L; (c) E94C/F98W/A360C/C109A/C136A/K199L; 또는 (d) M1G/E94C/F98W/A360C/C109A/C136A/K199L.
- [0545] 다른 바람직한 헬리카제
- [0546] 본 발명의 다른 바람직한 헬리카제는 다음에서의 치환을 포함하는 서열번호: 118의 변이체를 포함한다:
- [0547] T40/E94/F98, 예컨대 T40Y/E94C/F98W;
- [0548] T40/E94/F98/C114, 예컨대 T40Y/E94C/F98W/C114I;
- [0549] T40/E94/F98/K177, 예컨대 T40Y/E94C/F98W/K177M;
- [0550] T40/E94/F98/Y350, 예컨대 T40Y/E94C/F98W/Y350, T40Y/E94C/F98W/Y350I 또는 T40Y/E94C/F98W/Y350E;
- [0551] T40/E94/F98/K358, 예컨대 T40Y/E94C/F98W/K358I;
- [0552] E94/F98/C114, 예컨대 E94C/F98W/C114I;
- [0553] E94/F98/K177, 예컨대 E94C/F98W/K177M;
- [0554] E94/F98/Y350, 예컨대 E94C/F98W/Y350I 또는 E94C/F98W/Y350E;
- [0555] E94/F98/K358, 예컨대 E94C/F98W/K358I;
- [0556] T40/E94/F98/C114/K177, 예컨대 T40Y/E94C/F98W/C114I/K177M;
- [0557] T40/E94/F98/C114/Y350, 예컨대 T40Y/E94C/F98W/C114I/Y350I 또는 T40Y/E94C/F98W/C114I/Y350E;
- [0558] T40/E94/F98/C114/K358, 예컨대 T40Y/E94C/F98W/C114I/K358I;
- [0559] T40/E94/F98/K177/Y350, 예컨대 T40Y/E94C/F98W/K177M/Y350I 또는 T40Y/E94C/F98W/K177M/Y350E;
- [0560] T40/E94/F98/K177/K358, 예컨대 T40Y/E94C/F98W/K177M/K358I;
- [0561] T40/E94/F98/Y350/K358, 예컨대 T40Y/E94C/F98W/Y350I/K358I 또는 T40Y/E94C/F98W/Y350E/K358I;
- [0562] E94/F98/C114/K177, 예컨대 E94C/F98W/C114I/K177M;
- [0563] E94/F98/C114/Y350, 예컨대 E94C/F98W/C114I/Y350I 또는 E94C/F98W/C114I/Y350E;
- [0564] E94/F98/C114/K358, 예컨대 E94C/F98W/C114I/K358I;
- [0565] E94/F98/K177/Y350, 예컨대 E94C/F98W/K177M/Y350I 또는 E94C/F98W/K177M/Y350E;

- [0566] E94/F98/K177/K358, 예컨대 E94C/F98W/K177M/K358I;
- [0567] E94/F98/Y350/K358, 예컨대 E94C/F98W/Y350I/K358I 또는 E94C/F98W/Y350E/K358I;
- [0568] T40/E94/F98/C114/K177/Y350, 예컨대 T40Y/E94C/F98W/C114I/K177M/Y350I 또는 T40Y/E94C/F98W/C114I/K177M/Y350E;
- [0569] T40/E94/F98/C114/K177/K358, 예컨대 T40Y/E94C/F98W/C114I/K177M/K358I;
- [0570] T40/E94/F98/C114/Y350/K358, 예컨대 T40Y/E94C/F98W/C114I/Y350I/K358I 또는 T40Y/E94C/F98W/C114I/Y350E/K358I;
- [0571] T40/E94/F98/K177/Y350/K358, 예컨대 T40Y/E94C/F98W/K177M/Y350I/K358I 또는 T40Y/E94C/F98W/K177M/Y350E/K358I;
- [0572] E94/F98/C114/K177/Y350, 예컨대 E94C/F98W/C114I/K177M/Y350I 또는 E94C/F98W/C114I/K177M/Y350E;
- [0573] E94/F98/C114/K177/K358, 예컨대 E94C/F98W/C114I/K177M/K358I;
- [0574] E94/F98/C114/Y350/K358, 예컨대 E94C/F98W/C114I/Y350I/K358I 또는 E94C/F98W/C114I/Y350E/K358I;
- [0575] E94/F98/K177/Y350/K358, 예컨대 E94C/F98W/K177M/Y350I/K358I 또는 E94C/F98W/K177M/Y350E/K358I;
- [0576] T40/E94/F98/C114/K177/Y350/K358, 예컨대 T40Y/E94C/F98W/C114I/K177M/Y350I/K358I 또는 T40Y/E94C/F98W/C114I/K177M/Y350E/K358I;
- [0577] E94/F98/C114/K177/Y350/K358, 예컨대 E94C/F98W/C114I/K177M/Y350I/K358I 또는 E94C/F98W/C114I/K177M/Y350E/K358I;
- [0578] T40/F98/K194, 예컨대 T40Y/E94C/F98W/K194L;
- [0579] T40/F98/C114/K194, 예컨대 T40Y/E94C/F98W/C114I K194L;
- [0580] T40/F98/K177/K194, 예컨대 T40Y/E94C/F98W/K177M K194L;
- [0581] T40/F98/K194/Y350, 예컨대 T40Y/E94C/F98W/K194L/Y350I 또는 T40Y/E94C/F98W/K194L/Y350E;
- [0582] T40/F98/K194/K358, 예컨대 T40Y/E94C/F98W/K194L/K358I;
- [0583] F98/C114/K194, 예컨대 E94C/F98W/C114I/K194L;
- [0584] F98/K177/K194, 예컨대 E94C/F98W/K177M/K194L;
- [0585] F98/K194/Y350, 예컨대 E94C/F98W/K194L/Y350I 또는 E94C/F98W/K194L/Y350E;
- [0586] F98/K194/K358, 예컨대 E94C/F98W/K194L/K358I;
- [0587] T40/F98/C114/K177/K194, 예컨대 T40Y/E94C/F98W/C114I/K177M/K194L;
- [0588] T40/F98/C114/K194/Y350, 예컨대 T40Y/E94C/F98W/C114I/K194L/Y350I 또는 T40Y/E94C/F98W/C114I/K194L/Y350E;
- [0589] T40/F98/C114/K194/K358, 예컨대 T40Y/E94C/F98W/C114I/K194L/K358I;
- [0590] T40/F98/K177/K194/Y350, 예컨대 T40Y/E94C/F98W/K177M/K194L/Y350I 또는 T40Y/E94C/F98W/K177M/K194L/Y350E;
- [0591] T40/F98/K177/K194/K358, 예컨대 T40Y/E94C/F98W/K177M/K194L/K358I;
- [0592] T40/F98/K194/Y350/K358, 예컨대 T40Y/E94C/F98W/K194L/Y350I/K358I 또는 T40Y/E94C/F98W/K194L/Y350E/K358I;
- [0593] F98/C114/K177/K194, 예컨대 E94C/F98W/C114I/K177M/K194L;
- [0594] F98/C114/K194/Y350, 예컨대 E94C/F98W/C114I/K194L/Y350I 또는 E94C/F98W/C114I/K194L/Y350E;
- [0595] F98/C114/K194/K358, 예컨대 E94C/F98W/C114I/K194L/K358I;
- [0596] F98/K177/K194/Y350, 예컨대 E94C/F98W/K177M/K194L/Y350I 또는 E94C/F98W/K177M/K194L/Y350E;
- [0597] F98/K177/K194/K358, 예컨대 E94C/F98W/K177M/K194L/K358I;

[0598]	F98/K194/Y350/K358, 예컨대 E94C/F98W/K194L/Y350I/K358I 또는 E94C/F98W/K194L/Y350E/K358I;		
[0599]	T40/F98/C114/K177/K194/Y350, 예컨대 T40Y/E94C/F98W/C114I/K177M/K194L/Y350I 또는 T40Y/E94C/F98W/C114I/K177M/K194L/Y350E;		
[0600]	T40/F98/C114/K177/K194/K358, 예컨대 T40Y/E94C/F98W/C114I/K177M/K194L/K358I;		
[0601]	T40/F98/C114/K194/Y350/K358, 예컨대 T40Y/E94C/F98W/C114I/K194L/Y350I/K358I 또는 T40Y/E94C/F98W/C114I/K194L/Y350E/K358I;		
[0602]	T40/F98/K177/K194/Y350/K358, 예컨대 T40Y/E94C/F98W/K177M/K194L/Y350I/K358I 또는 T40Y/E94C/F98W/K177M/K194L/Y350E/K358I;		
[0603]	F98/C114/K177/K194/Y350, 예컨대 E94C/F98W/C114I/K177M/K194L/Y350I 또는 E94C/F98W/C114I/K177M/K194L/Y350E;		
[0604]	F98/C114/K177/K194/K358, 예컨대 E94C/F98W/C114I/K177M/K194L/K358I;		
[0605]	F98/C114/K194/Y350/K358, 예컨대 E94C/F98W/C114I/K194L/Y350I/K358I 또는 E94C/F98W/C114I/K194L/Y350E/K358I;		
[0606]	F98/K177/K194/Y350/K358, 예컨대 E94C/F98W/K177M/K194L/Y350I/K358I 또는 E94C/F98W/K177M/K194L/Y350E/K358I;		
[0607]	T40/F98/C114/K177/K194/Y350/K358, 예컨대 T40Y/E94C/F98W/C114I/K177M/K194L/Y350I/K358I 또는 T40Y/E94C/F98W/C114I/K177M/K194L/Y350E/K358I;		
[0608]	F98/C114/K177/K194/Y350/K358, 예컨대 E94C/F98W/C114I/K177M/K194L/Y350I/K358I 또는 E94C/F98W/C114I/K177M/K194L/Y350E/K358I;		
[0609]	T40/A360, 예컨대 T40Y/A360C;		
[0610]	T40/C114/A360, 예컨대 T40Y/C114I/A360C;		
[0611]	T40/K177/A360, 예컨대 T40Y/K177M/A360C;		
[0612]	T40/Y350/A360, 예컨대 T40Y/Y350I/A360C 또는 T40Y/Y350E/A360C;		
[0613]	T40/K358/A360, 예컨대 T40Y/K358I/A360C;		
[0614]	C114/A360, 예컨대 C114I/A360C;		
[0615]	K177/A360, 예컨대 K177M/A360C;		
[0616]	Y350/A360, 예컨대 Y350I/A360C 또는 Y350E/A360C;		
[0617]	K358/A360, 예컨대 K358I/A360C;		
[0618]	T40/C114/K177/A360, 예컨대 T40Y/C114I/K177M/A360C;		
[0619]	T40/C114/Y350/A360, 예컨대 T40Y/C114I/Y350I/A360C 또는 T40Y/C114I/Y350E/A360C;		
[0620]	T40/C114/K358/A360, 예컨대 T40Y/C114I/K358I/A360C;		
[0621]	T40/K177/Y350/A360, 예컨대 T40Y/K177M/Y350I/A360C 또는 T40Y/K177M/Y350E/A360C;		
[0622]	T40/K177/K358/A360, 예컨대 T40Y/K177M/K358I/A360C;		
[0623]	T40/Y350/K358/A360, 예컨대 T40Y/Y350I/K358I/A360C 또는 T40Y/Y350E/K358I/A360C;		
[0624]	C114/K177/A360, 예컨대 C114I/K177M/A360C;		
[0625]	C114/Y350/A360, 예컨대 C114I/Y350I/A360C 또는 C114I/Y350E/A360C;		
[0626]	C114/K358/A360, 예컨대 C114I/K358I/A360C;		
[0627]	K177/Y350/A360, 예컨대 K177M/Y350I/A360C 또는 K177M/Y350E/A360C;		

- [0628] K177/K358/A360, 예컨대 K177M/K358I/A360C;
- [0629] Y350/K358/A360, 예컨대 Y350I/K358I/A360C 또는 Y350E/K358I/A360C;
- [0630] T40/C114/K177/Y350/A360, 예컨대 T40Y/C114I/K177M/Y350I/A360C 또는 T40Y/C114I/K177M/Y350E/A360C;
- [0631] T40/C114/K177/K358/A360, 예컨대 T40Y/C114I/K177M/K358I/A360C;
- [0632] T40/C114/Y350/K358/A360, 예컨대 T40Y/C114I/Y350I/K358I/A360C 또는 T40Y/C114I/Y350E/K358I/A360C;
- [0633] T40/K177/Y350/K358/A360, 예컨대 T40Y/K177M/Y350I/K358I/A360C 또는 T40Y/K177M/Y350E/K358I/A360C;
- [0634] C114/K177/Y350/A360, 예컨대 C114I/K177M/Y350I/A360C 또는 C114I/K177M/Y350E/A360C;
- [0635] C114/K177/K358/A360, 예컨대 C114I/K177M/K358I/A360C;
- [0636] C114/Y350/K358/A360, 예컨대 C114I/Y350I/K358I/A360C 또는 C114I/Y350E/K358I/A360C;
- [0637] K177/Y350/K358/A360, 예컨대 K177M/Y350I/K358I/A360C 또는 K177M/Y350E/K358I/A360C;
- [0638] T40/C114/K177/Y350/K358/A360, 예컨대 T40Y/C114I/K177M/Y350I/K358I/A360C 또는 K177M/Y350E/K358I/A360C;
- [0639] C114/K177/Y350/K358/A360, 예컨대 C114I/K177M/Y350I/K358I/A360C 또는 C114I/K177M/Y350E/K358I/A360C;
- [0640] T40/E94/F98/C109, 예컨대 T40Y/E94C/F98W/C109A;
- [0641] T40/E94/F98/C109/C114, 예컨대 T40Y/E94C/F98W/C109A /C114I;
- [0642] T40/E94/F98/C109/K177, 예컨대 T40Y/E94C/F98W/C109A /K177M;
- [0643] T40/E94/F98/C109/Y350, 예컨대 T40Y/E94C/F98W/C109A /Y350I 또는 T40Y/E94C/F98W/C109A /Y350E;
- [0644] T40/E94/F98/C109/K358, 예컨대 T40Y/E94C/F98W/C109A /K358I;
- [0645] E94/F98/C109/C114, 예컨대 E94C/F98W/C109A /C114I;
- [0646] E94/F98/C109/K177, 예컨대 E94C/F98W/C109A /K177M;
- [0647] E94/F98/C109/Y350, 예컨대 E94C/F98W/C109A /Y350I 또는 E94C/F98W/C109A /Y350E;
- [0648] E94/F98/C109/K358, 예컨대 E94C/F98W/C109A /K358I;
- [0649] T40/E94/F98/C109/C114/K177, 예컨대 T40Y/E94C/F98W/C109A/C114I/K177M;
- [0650] T40/E94/F98/C109/C114/Y350, 예컨대 T40Y/E94C/F98W/C109A /C114I/Y350I 또는 T40Y/E94C/F98W/C109A /C114I/Y350E;
- [0651] T40/E94/F98/C109/C114/K358, 예컨대 T40Y/E94C/F98W/C109A /C114I/K358I;
- [0652] T40/E94/F98/C109/K177/Y350, 예컨대 T40Y/E94C/F98W/C109A /K177M/Y350I 또는 T40Y/E94C/F98W/C109A /K177M/Y350E;
- [0653] T40/E94/F98/C109/K177/K358, 예컨대 T40Y/E94C/F98W/C109A /K177M/K358I;
- [0654] T40/E94/F98/C109/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A /Y350I/K358I 또는 T40Y/E94C/F98W/C109A /Y350E/K358I;
- [0655] E94/F98/C109/C114/K177, 예컨대 E94C/F98W/C109A/C114I/K177M;
- [0656] E94/F98/C109/C114/Y350, 예컨대 E94C/F98W/C109A /C114I/Y350I 또는 E94C/F98W/C109A /C114I/Y350E;
- [0657] E94/F98/C109/C114/K358, 예컨대 E94C/F98W/C109A /C114I/K358I;
- [0658] E94/F98/C109/K177/Y350, 예컨대 E94C/F98W/C109A /K177M/Y350I 또는 E94C/F98W/C109A /K177M/Y350E;
- [0659] E94/F98/C109/K177/K358, 예컨대 E94C/F98W/C109A /K177M/K358I;
- [0660] E94/F98/C109/Y350/K358, 예컨대 E94C/F98W/C109A /Y350I/K358I 또는 E94C/F98W/C109A /Y350E/K358I;
- [0661] T40/E94/F98/C109/C114/K177/Y350, 예컨대 T40Y/E94C/F98W/C109A /C114I/K177M/Y350I 또는

T40Y/E94C/F98W/C109A /C114I/K177M/Y350E;

- [0662] T40/E94/F98/C109/C114/K177/K358, 예컨대 T40Y/E94C/F98W/C109A /C114I/K177M/K358I;
- [0663] T40/E94/F98/C109/C114/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A /C114I/Y350I/K358I 또는 T40Y/E94C/F98W/C109A /C114I/Y350E/K358I;
- [0664] T40/E94/F98/C109/K177/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A /K177M/Y350I/K358I 또는 T40Y/E94C/F98W/C109A /K177M/Y350E/K358I;
- [0665] E94/F98/C109/C114/K177/Y350, 예컨대 E94C/F98W/C109A /C114I/K177M/Y350I 또는 E94C/F98W/C109A /C114I/K177M/Y350E;
- [0666] E94/F98/C109/C114/K177/K358, 예컨대 E94C/F98W/C109A /C114I/K177M/K358I;
- [0667] E94/F98/C109/C114/Y350/K358, 예컨대 E94C/F98W/C109A /C114I/Y350I/K358I 또는 E94C/F98W/C109A /C114I/Y350E/K358I;
- [0668] E94/F98/C109/K177/Y350/K358, 예컨대 E94C/F98W/C109A /K177M/Y350I/K358I 또는 E94C/F98W/C109A /K177M/Y350E/K358I;
- [0669] T40/E94/F98/C109/C114/K177/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A/C114I/K177M/Y350I/K358I 또는 T40Y/E94C/F98W/C109A/C114I/K177M/Y350E/K358I;
- [0670] E94/F98/C109/C114/K177/Y350/K358, 예컨대 E94C/F98W/C109A/C114I/K177M/Y350I/K358I 또는 E94C/F98W/C109A/C114I/K177M/Y350E/K358I;
- [0671] T40/C109/C136, 예컨대 T40Y/C109A/C136A;
- [0672] T40/C109/C114/C136, 예컨대 T40Y/C109A/C114I/C136A;
- [0673] T40/C109/C136/K177, 예컨대 T40Y/C109A/C136A/K177M;
- [0674] T40/C109/C136/Y350, 예컨대 T40Y/C109A/C136A/Y350I 또는 T40Y/C109A/C136A/Y350E;
- [0675] T40/C109/C136/K358, 예컨대 T40Y/C109A/C136A/K358I;
- [0676] C109/C114/C136, 예컨대 C109A/C136A/C114I;
- [0677] C109/C136/K177, 예컨대 C109A/C136A/K177M;
- [0678] C109/C136/Y350, 예컨대 C109A/C136A/Y350I 또는 C109A/C136A/Y350E;
- [0679] C109/C136/K358, 예컨대 C109A/C136A/K358I;
- [0680] T40/C109/C114/C136/K177, 예컨대 T40Y/C109A/C114I/C136A/K177M;
- [0681] T40/C109/C114/C136/Y350, 예컨대 T40Y/C109A/C114I/C136A/Y350 또는 T40Y/C109A/C114I/C136A/Y350I I ;
- [0682] T40/C109/C114/C136/K358, 예컨대 T40Y/C109A/C114I/C136A/K358I;
- [0683] T40/C109/C136/K177/Y350, 예컨대 T40Y/C109A/C136A/K177M/Y350E 또는 T40Y/C109A/C136A/K177M/Y350I;
- [0684] T40/C109/C136/K177/K358, 예컨대 T40Y/C109A/C136A/K177M/K358I;
- [0685] T40/C109/C136/Y350/K358, 예컨대 T40Y/C109A/C136A/Y350I/K358I 또는 T40Y/C109A/C136A/Y350E/K358I;
- [0686] C109/C114/C136/K177, 예컨대 C109A/C114I/C136A/K177M;
- [0687] C109/C114/C136/Y350, 예컨대 C109A/C114I/C136A/Y350I 또는 C109A/C114I/C136A/Y350E;
- [0688] C109/C114/C136/K358, 예컨대 C109A/C114I/C136A/K358I;
- [0689] C109/C136/K177/Y350, 예컨대 C109A/C136A/K177M/Y350I 또는 C109A/C136A/K177M/Y350E;
- [0690] C109/C136/K177/K358, 예컨대 C109A/C136A/K177M/K358I;
- [0691] C109/C136/Y350/K358, 예컨대 C109A/C136A/Y350I/K358I 또는 C109A/C136A/Y350E/K358I;

[0692]	T40/C109/C114/C136/K177/Y350, T40Y/C109A/C114I/C136A/K177M/Y350E;	예컨대	T40Y/C109A/C114I/C136A/K177M/Y350I	또는
[0693]	T40/C109/C114/C136/K177/K358,	예컨대	T40Y/C109A/C114I/C136A/ K177M/K358I;	
[0694]	T40/C109/C114/C136/Y350/K358, T40Y/C109A/C114I/C136AY350E/K358I;	예컨대	T40Y/C109A/C114I/C136AY350I/K358I	또는
[0695]	T40/C109/C136/K177/Y350/K358, T40Y/C109A/C136A/K177M/Y350E/K358I;	예컨대	T40Y/C109A/C136A/K177M/Y350I/K358I	또는
[0696]	C109/C114/C136/K177/Y350,	예컨대	C109A/C114I/C136A/K177M/Y350I 또는 C109A/C114I/C136A/K177M/Y350E;	
[0697]	C109/C114/C136/K177/K358,	예컨대	C109A/C114I/C136A/ K177M/K358I;	
[0698]	C109/C114/C136/Y350/K358,	예컨대	C109A/C114I/C136AY350I/K358I 또는 C109A/C114I/C136AY350E/K358I;	
[0699]	C109/C136/K177/Y350/K358,	예컨대	C109A/C136A/K177M/Y350I/K358I 또는 C109A/C136A/K177M/Y350E/K358I;	
[0700]	T40/C109/C114/C136/K177/Y350/K358, T40Y/C109A/C114I/C136A/K177M/Y350E/K358I;	예컨대	T40Y/C109A/C114I/C136A/K177M/Y350I/K358I	또는
[0701]	C109/C114/C136/K177/Y350/K358, C109A/C114I/C136A/K177M/Y350E/K358I;	예컨대	C109A/C114I/C136A/K177M/Y350I/K358I	
[0702]	T40/E94/F98/C109/K194,	예컨대	T40Y/E94C/F98W/C109A/K194L;	
[0703]	T40/E94/F98/C109/C114/K194,	예컨대	T40Y/E94C/F98W/C109A/C114I/K194L;	
[0704]	T40/E94/F98/C109/K177/K194,	예컨대	T40Y/E94C/F98W/C109A/K177M/K194L;	
[0705]	T40/E94/F98/C109/K194/Y350, T40Y/E94C/F98W/C109A/K194L/Y350E;	예컨대	T40Y/E94C/F98W/C109A/K194L/Y350I	또는
[0706]	T40/E94/F98/C109/K194/K358,	예컨대	T40Y/E94C/F98W/C109A/K194L/K358I;	
[0707]	E94/F98/C109/C114/K194,	예컨대	E94C/F98W/C109A /C114I/K194L;	
[0708]	E94/F98/C109/K177/K194,	예컨대	E94C/F98W/C109A /K177M/K194L;	
[0709]	E94/F98/C109/K194/Y350,	예컨대	E94C/F98W/C109A/K194L/Y350I 또는 E94C/F98W/C109A/K194L/Y350E;	
[0710]	E94/F98/C109/K194/K358,	예컨대	E94C/F98W/C109A/K194L/K358I;	
[0711]	T40/E94/F98/C109/C114/K177/K194,	예컨대	T40Y/E94C/F98W/C109A/C114I/K194L/K177M;	
[0712]	T40/E94/F98/C109/C114/K194/Y350, T40Y/E94C/F98W/C109A/C114I/K194L /Y350E;	예컨대	T40Y/E94C/F98W/C109A/C114I/K194L /Y350I	또는
[0713]	T40/E94/F98/C109/C114/K194/K358,	예컨대	T40Y/E94C/F98W/C109A/C114I/K194L /K358I;	
[0714]	T40/E94/F98/C109/K177/K194/Y350, T40Y/E94C/F98W/C109A/K177M/K194L /Y350E;	예컨대	T40Y/E94C/F98W/C109A/K177M/K194L /Y350I	또는
[0715]	T40/E94/F98/C109/K177/K194/K358,	예컨대	T40Y/E94C/F98W/C109A/K177M/K194L /K358I;	
[0716]	T40/E94/F98/C109/K194/Y350/K358, T40Y/E94C/F98W/C109A/K194L /Y350E/K358I ;	예컨대	T40Y/E94C/F98W/C109A/K194L /Y350I/K358Ior	
[0717]	E94/F98/C109/C114/K177/K194,	예컨대	E94C/F98W/C109A/C114I/K194L/K177M;	
[0718]	E94/F98/C109/C114/K194/Y350, E94C/F98W/C109A/C114I/K194L /Y350E;	예컨대	E94C/F98W/C109A/C114I/K194L /Y350I	또는
[0719]	E94/F98/C109/C114/K194/K358,	예컨대	E94C/F98W/C109A/C114I/K194L /K358I;	
[0720]	E94/F98/C109/K177/K194/Y350,	예컨대	E94C/F98W/C109A/K177M/K194L /Y350I o E94C/F98W/C109A/K177M/K194L	

/Y350E;

- [0721] E94/F98/C109/K177/K194/K358, 예컨대 E94C/F98W/C109A/K177M/K194L /K358I;
- [0722] E94/F98/C109/K194/Y350/K358, 예컨대 E94C/F98W/C109A/K194L /Y350I/K358I 또는 E94C/F98W/C109A/K194L /Y350E/K358I;
- [0723] T40/E94/F98/C109/C114/K177/K194/Y350, 예컨대 T40Y/E94C/F98W/C109A /C114I/K177M/K194L/Y350I 또는 T40Y/E94C/F98W/C109A /C114I/K177M/K194L/Y350E;
- [0724] T40/E94/F98/C109/C114/K177/K194/K358, 예컨대 T40Y/E94C/F98W/C109A /C114I/K177M/K194L/K358I;
- [0725] T40/E94/F98/C109/C114/K194/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A /C114I/K194L/Y350I/K358I 또는 T40Y/E94C/F98W/C109A /C114I/K194L/Y350E/K358I;
- [0726] T40/E94/F98/C109/K177/K194/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A /K177M/K194L/Y350I/K358I 또는 T40Y/E94C/F98W/C109A /K177M/K194L/Y350E/K358I;
- [0727] E94/F98/C109/C114/K177/K194/Y350, 예컨대 E94C/F98W/C109A /C114I/K177M/K194L/Y350I 또는 E94C/F98W/C109A /C114I/K177M/K194L/Y350E;
- [0728] E94/F98/C109/C114/K177/K194/K358, 예컨대 E94C/F98W/C109A /C114I/K177M/K194L/K358I;
- [0729] E94/F98/C109/C114/K194/Y350/K358, 예컨대 E94C/F98W/C109A /C114I/K194L/Y350I/K358I 또는 E94C/F98W/C109A /C114I/K194L/Y350E/K358I;
- [0730] E94/F98/C109/K177/K194/Y350/K358, 예컨대 E94C/F98W/C109A /K177M/K194L/Y350I/K358I 또는 E94C/F98W/C109A /K177M/K194L/Y350E/K358I;
- [0731] T40/E94/F98/C109/C114/K177/K194/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A/C114I/K177M/K194L/Y350I/K358I 또는 T40Y/E94C/F98W/C109A/C114I/K177M/K194L/Y350E/K358I;
- [0732] E94/F98/C109/C114/K177/K194/Y350/K358, 예컨대 E94C/F98W/C109A/C114I/K177M/K194L/Y350I/K358I 또는 E94C/F98W/C109A/C114I/K177M/K194L/Y350E/K358I;
- [0733] T40/E94/F98/C109/C136, 예컨대 T40Y/E94C/F98W/C109A/C136A;
- [0734] T40/E94/F98/C109/C114/C136, 예컨대 T40Y/E94C/F98W/C109A /C114I/C136A;
- [0735] T40/E94/F98/C109/C136/K177, 예컨대 T40Y/E94C/F98W/C109A/C136A/K177M;
- [0736] T40/E94/F98/C109/C136/Y350, 예컨대 T40Y/E94C/F98W/C109A/C136A/Y350I 또는 T40Y/E94C/F98W/C109A/C136A/Y350E;
- [0737] T40/E94/F98/C109/C136/K358, 예컨대 T40Y/E94C/F98W/C109A/C136A/K358I;
- [0738] E94/F98/C109/C114/C136, 예컨대 E94C/F98W/C109A /C114I/C136A;
- [0739] E94/F98/C109/C136/K177, 예컨대 E94C/F98W/C109A/C136A/K177M;
- [0740] E94/F98/C109/C136/Y350, 예컨대 E94C/F98W/C109A/C136A/Y350I 또는 T40Y/E94C/F98W/C109A/C136A/Y350E;
- [0741] E94/F98/C109/C136/K358, 예컨대 E94C/F98W/C109A/C136A/K358I;
- [0742] T40/E94/F98/C109/C114/C136/K177, 예컨대 T40Y/E94C/F98W/C109A/C114I/C136A/K177M;
- [0743] T40/E94/F98/C109/C114/C136/Y350, 예컨대 T40Y/E94C/F98W/C109A/C114I/C136A/Y350I 또는 T40Y/E94C/F98W/C109A/C114I/C136A/Y350E;
- [0744] T40/E94/F98/C109/C114/C136/K358, 예컨대 T40Y/E94C/F98W/C109A/C114I/C136A/K358I;
- [0745] T40/E94/F98/C109/C136/K177/Y350, 예컨대 T40Y/E94C/F98W/C109A/C136AK177M/Y350I 또는 T40Y/E94C/F98W/C109A/C136AK177M/Y350E;
- [0746] T40/E94/F98/C109/C136/K177/K358, 예컨대 T40Y/E94C/F98W/C109A/C136AK177M/K358I;

[0747]	T40/E94/F98/C109/C136/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A/C136AY350I/K358I 또는 T40Y/E94C/F98W/C109A/C136AY350E/K358I;	
[0748]	E94/F98/C109/C114/C136/K177, 예컨대 E94C/F98W/C109A/C114I/C136A/K177M;	
[0749]	E94/F98/C109/C114/C136/Y350, 예컨대 E94C/F98W/C109A/C114I/C136A/Y350I 또는 E94C/F98W/C109A/C114I/C136A/Y350E;	
[0750]	E94/F98/C109/C114/C136/K358, 예컨대 E94C/F98W/C109A/C114I/C136A/K358I;	
[0751]	E94/F98/C109/C136/K177/Y350, 예컨대 E94C/F98W/C109A/C136AK177M/Y350I 또는 E94C/F98W/C109A/C136AK177M/Y350E;	
[0752]	E94/F98/C109/C136/K177/K358, 예컨대 E94C/F98W/C109A/C136AK177M/K358I;	
[0753]	E94/F98/C109/C136/Y350/K358, 예컨대 E94C/F98W/C109A/C136AY350I/K358I 또는 E94C/F98W/C109A/C136AY350E/K358I;	
[0754]	T40/E94/F98/C109/C114/C136/K177/Y350, 예컨대 T40Y/E94C/F98W/C109A/C114I/C136A/K177M/Y350I 또는 T40Y/E94C/F98W/C109A/C114I/C136A/K177M/Y350E;	
[0755]	T40/E94/F98/C109/C114/C136/K177/K358, 예컨대 T40Y/E94C/F98W/C109A/C114I/C136A/K177M/K358I;	
[0756]	T40/E94/F98/C109/C114/C136/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A/C114I/C136A/Y350I/K358I 또는 T40Y/E94C/F98W/C109A/C114I/C136A/Y350E/K358I;	
[0757]	T40/E94/F98/C109/C136/K177/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A/C136A/K177M/Y350I/K358I 또는 T40Y/E94C/F98W/C109A/C136A/K177M/Y350E/K358I;	
[0758]	E94/F98/C109/C114/C136/K177/Y350, 예컨대 E94C/F98W/C109A/C114I/C136A/K177M/Y350I 또는 E94C/F98W/C109A/C114I/C136A/K177M/Y350E;	
[0759]	E94/F98/C109/C114/C136/K177/K358, 예컨대 E94C/F98W/C109A/C114I/C136A/K177M/K358I;	
[0760]	E94/F98/C109/C114/C136/Y350/K358, 예컨대 E94C/F98W/C109A/C114I/C136A/Y350I/K358I 또는 E94C/F98W/C109A/C114I/C136A/Y350E/K358I;	
[0761]	E94/F98/C109/C136/K177/Y350/K358, 예컨대 E94C/F98W/C109A/C136A/K177M/Y350I/K358I 또는 E94C/F98W/C109A/C136A/K177M/Y350E/K358I;	
[0762]	T40/E94/F98/C109/C114/C136/K177/Y350/K358, 예컨대 T40Y/E94C/F98W/C109A/C114I/C136A/K177M/Y350I/K358I 또는 T40Y/E94C/F98W/C109A/C114I/C136A/K177M/Y350E/K358I;	
[0763]	E94/F98/C109/C114/C136/K177/Y350/K358, 예컨대 E94C/F98W/C109A/C114I/C136A/K177M/Y350I/K358I 또는 E94C/F98W/C109A/C114I/C136A/K177M/Y350E/K358I;	
[0764]	T40/E94/F98/C109/A360, 예컨대 T40Y/E94C/F98W/C109A/A360C;	
[0765]	T40/E94/F98/C109/C114/A360, 예컨대 T40Y/E94C/F98W/C109A/C114I/A360C;	
[0766]	T40/E94/F98/C109/K177/A360, 예컨대 T40Y/E94C/F98W/C109A/K177M/A360C;	
[0767]	T40/E94/F98/C109/Y350/A360, 예컨대 T40Y/E94C/F98W/C109A/Y350I/A360C 또는 T40Y/E94C/F98W/C109A/Y350E/A360C;	
[0768]	T40/E94/F98/C109/K358/A360, 예컨대 T40Y/E94C/F98W/C109A/K358I/A360C;	
[0769]	E94/F98/C109/C114/A360, 예컨대 E94C/F98W/C109A /C114I/A360C;	
[0770]	E94/F98/C109/K177/A360, 예컨대 E94C/F98W/C109A /K177M/A360C;	
[0771]	E94/F98/C109/Y350/A360, 예컨대 E94C/F98W/C109A /Y350I/A360C 또는 E94C/F98W/C109A /Y350E/A360C;	
[0772]	E94/F98/C109/K358/A360, 예컨대 E94C/F98W/C109A /K358I/A360C;	

- [0773] T40/E94/F98/C109/C114/K177/A360, 예컨대 T40Y/E94C/F98W/C109A/C114I/K177M/A360C;
- [0774] T40/E94/F98/C109/C114/Y350/A360, 예컨대 T40Y/E94C/F98W/C109A /C114I/Y350I/A360C 또는 T40Y/E94C/F98W/C109A /C114I/Y350E/A360C;
- [0775] T40/E94/F98/C109/C114/K358/A360, 예컨대 T40Y/E94C/F98W/C109A /C114I/K358I/A360C;
- [0776] T40/E94/F98/C109/K177/Y350/A360, 예컨대 T40Y/E94C/F98W/C109A /K177M/Y350I/A360C 또는 T40Y/E94C/F98W/C109A /K177M/Y350E/A360C;
- [0777] /A360C
- [0778] T40/E94/F98/C109/Y350/K358/A360, 예컨대 T40Y/E94C/F98W/C109A /Y350I/K358I/A360C 또는 T40Y/E94C/F98W/C109A /Y350E/K358I/A360C;
- [0779] E94/F98/C109/C114/K177/A360, 예컨대 E94C/F98W/C109A/C114I/K177M/A360C;
- [0780] E94/F98/C109/C114/Y350/A360, 예컨대 E94C/F98W/C109A /C114I/Y350I/A360C 또는 E94C/F98W/C109A /C114I/Y350E/A360C;
- [0781] E94/F98/C109/C114/K358/A360, 예컨대 E94C/F98W/C109A /C114I/K358I/A360C;
- [0782] E94/F98/C109/K177/Y350/A360, 예컨대 E94C/F98W/C109A /K177M/Y350I/A360C 또는 E94C/F98W/C109A /K177M/Y350E/A360C;
- [0783] E94/F98/C109/Y350/K358/A360, 예컨대 E94C/F98W/C109A /Y350I/K358I/A360C 또는 E94C/F98W/C109A /Y350E/K358I/A360C;
- [0784] T40/E94/F98/C109/C114/K177/Y350/A360, 예컨대 T40Y/E94C/F98W/C109A /C114I/K177M/Y350I/A360C 또는 T40Y/E94C/F98W/C109A /C114I/K177M/Y350E/A360C;
- [0785] T40/E94/F98/C109/C114/K177/K358/A360, 예컨대 T40Y/E94C/F98W/C109A /C114I/K177M/K358I/A360C;
- [0786] T40/E94/F98/C109/C114/Y350/K358/A360, 예컨대 T40Y/E94C/F98W/C109A /C114I/Y350I/K358I/A360C 또는 T40Y/E94C/F98W/C109A /C114I/Y350E/K358I/A360C;
- [0787] T40/E94/F98/C109/K177/Y350/K358/A360, 예컨대 T40Y/E94C/F98W/C109A /K177M/Y350I/K358I/A360C 또는 T40Y/E94C/F98W/C109A /K177M/Y350E/K358I/A360C;
- [0788] E94/F98/C109/C114/K177/Y350/A360, 예컨대 E94C/F98W/C109A /C114I/K177M/Y350I/A360C 또는 E94C/F98W/C109A /C114I/K177M/Y350E/A360C;
- [0789] E94/F98/C109/C114/K177/K358/A360, 예컨대 E94C/F98W/C109A /C114I/K177M/K358I/A360C;
- [0790] E94/F98/C109/C114/Y350/K358/A360, 예컨대 E94C/F98W/C109A /C114I/Y350I/K358I/A360C 또는 E94C/F98W/C109A /C114I/Y350E/K358I/A360C;
- [0791] E94/F98/C109/K177/Y350/K358/A360, 예컨대 E94C/F98W/C109A /K177M/Y350I/K358I/A360C 또는 E94C/F98W/C109A /K177M/Y350E/K358I/A360C;
- [0792] T40/E94/F98/C109/C114/K177/Y350/K358/A360, 예컨대 T40Y/E94C/F98W/C109A/C114I/K177M/Y350I/K358I/A360C 또는 T40Y/E94C/F98W/C109A/C114I/K177M/Y350E/K358I/A360C; 또는
- [0793] E94/F98/C109/C114/K177/Y350/K358/A360, 예컨대 E94C/F98W/C109A/C114I/K177M/Y350I/K358I/A360C 또는 E94C/F98W/C109A/C114I/K177M/Y350E/K358I/A360C.
- [0794] 서열번호: 118의 이러한 변이체 중 임의의 것은 다음을 포함하여, 위치 (a) 55, (b) 156, (c) 210 및 (d) 221의 임의의 수 및 조합에서 변형 또는 치환을 추가로 포함할 수 있다: (a); (b); (c); (d); (a) 및 (b); (a) 및 (c); (a) 및 (d); (b) 및 (c); (b) 및 (d); (c) 및 (d); (a), (b) 및 (c); (a), (b) 및 (d); (a), (c) 및 (d); (b), (c) 및 (d); 또는 (a), (b), (c) 및 (d). 서열번호: 118의 이러한 변이체 중 임의의 것은 다음을 포함하여, (a) T55K, (b) T156F, (c) T210K 및 (d) N221E의 임의의 수 및 조합을 추가로 포함할 수 있다: (a); (b); (c); (d); (a) 및 (b); (a) 및 (c); (a) 및 (d); (b) 및 (c); (b) 및 (d); (c) 및 (d); (a), (b) 및 (c); (a), (b) 및 (d); (a), (c) 및 (d); (b), (c) 및 (d); 또는 (a), (b), (c) 및 (d).

- [0795] 추가적인 변형/치환
- [0796] 본 발명은 또한 Dda 1993의 다음 아미노산 위치에 상응하는 위치 중 하나 이상이 변형되거나 치환된 변형된 DNA 의존적 ATPase(Dda) 헬리카제를 제공한다: 86, 90, 92, 97, 101, 102, 273, 293, 300, 301, 303, 305, 308, 310, 312 317, 323, 328, 332, 334, 335, 336, 337, 339, 351, 354, 359, 361, 364, 366, 368, 371, 374, 376, 377, 379 및 388. 본 발명은 또한 Dda 1993의 다음 아미노산 위치에 상응하는 위치 중 하나 이상이 변형되거나 치환된 변형된 DNA 의존적 ATPase(Dda) 헬리카제를 제공한다: 351, 354 및 361. 351, 354, 361, 351 및 354, 351 및 361, 354 및 361 또는 351, 354 및 361을 포함하여, 이러한 변형/치환의 임의의 수 및 조합이 이루어질 수 있다. 이러한 위치는 별개로 또는 위의 발명의 변형 또는 치환 중 임의의 것과 조합하여 변형되거나 치환될 수 있다.
- [0797] 본 발명의 헬리카제는 바람직하게는 다음 치환 중 하나 이상을 포함하는 서열번호: 118의 변이체를 포함한다: K86A, V90T, Y92A, Y92D, Y92F, Y92G, Y92H, Y92N, Y92Q, Y92S, Y92T, Y92V, Y92W, L97H, K101A, E102G, E273A, N293H, I200A, I300F, E301I, E303I, Y305L, F308I, F308L, K310A, K310I, K310L, R312I, R312L, R312M, E317I, E317Y, W323H, E328I, D332A, D332L, E334A, E334I, E334Y, Y335A, Y335I, Y335L, Y336L, R337I, R337L, R337M, K339A, K339I, K339L, K351I, K351Q, L354Q, L354A, T359L, E361T, E361I, E361Q, K364A, K364R, W366L, K368A, K368I, K368L, K371I, K371L, K371M, W374A, W374L, W374L, D376S, F377A, D379A, D379L 또는 K388R. 본 발명의 헬리카제는 바람직하게는 다음 치환 중 하나 이상을 포함하는 서열번호: 118의 변이체를 포함한다: (a) K351I 또는 K351Q, (b) L354A 또는 L354Q 및 (c) E361I 또는 E361Q. 변이체는 (a), (b), (c), (a) 및 (b), (a) 및 (c), (b) 및 (c) 또는 (a), (b) 및 (c)를 포함할 수 있다. 이러한 치환은 별개로 또는 위의 발명의 변형 또는 치환 중 임의의 것과 조합하여 이루어질 수 있다.
- [0798] 변이체
- [0799] 헬리카제의 변이체는 야생형 헬리카제의 아미노산 서열과 다른 아미노산 서열을 갖고 폴리뉴클레오티드 결합 활성을 갖는 효소이다. 특히, 서열번호: 118 내지 133 중 임의의 하나의 변이체는 서열번호: 118 내지 133 중 임의의 하나의 아미노산 서열과 다른 아미노산 서열을 갖고 폴리뉴클레오티드 결합 활성을 갖는 효소이다. 폴리뉴클레오티드 결합 활성은 당업계에 알려진 방법을 사용하여 결정될 수 있다. 적합한 방법은 형광 이방성, 트립토판 형광 및 전기영동 이동성 변화 검정(EMSA)을 포함하나 이에 제한되지 않는다. 예를 들면, 단일 가닥 폴리뉴클레오티드에 결합하는 변이체의 능력은 실시예에 기재된 바와 같이 결정될 수 있다.
- [0800] 변이체는 헬리카제 활성을 갖는다. 이는 다양한 방식으로 측정될 수 있다. 예를 들면, 폴리뉴클레오티드를 따라 전위하는 변이체의 능력은 전기생리학, 형광 검정 또는 ATP 가수분해를 사용하여 측정될 수 있다.
- [0801] 변이체는 헬리카제를 암호화하는 폴리뉴클레오티드의 취급을 용이하게 하고/하거나 높은 염 농도 및/또는 실온에서 활성을 용이하게 하는 변형을 포함할 수 있다.
- [0802] 서열번호: 118 내지 133 중 임의의 하나의 아미노산 서열의 전체 길이에 걸쳐, 변이체는 바람직하게는 아미노산 유사성 또는 동일성을 기준으로 해당 서열과 적어도 20% 상동성일 것이다. 보다 바람직하게는, 변이체 폴리펩티드는 전체 서열에 걸쳐 서열번호: 118 내지 133 중 임의의 하나의 아미노산 서열에 대한 아미노산 동일성을 기준으로 적어도 30%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 및 보다 바람직하게는 적어도 95%, 97% 또는 99% 상동성일 수 있다. 보다 바람직하게는, 변이체 폴리펩티드는 전체 서열에 걸쳐 서열번호: 118 내지 133 중 임의의 하나의 아미노산 서열과 적어도 30%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 및 보다 바람직하게는 적어도 95%, 97% 또는 99% 동일할 수 있다. 100개 이상, 예를 들어 150, 200, 300, 400 또는 500개 이상의 연속 아미노산의 스트레치에 걸쳐 적어도 70%, 예를 들어 적어도 80%, 적어도 85%, 적어도 90% 또는 적어도 95%, 아미노산 동일성이 있을 수 있다 ("경질 상동성"). 상동성은 아래에 기재된 바와 같이 결정된다. 특히, 위에 논의된 특이적 변형 외에도, 서열번호: 118 내지 133 중 임의의 하나의 변이체는 아래에 논의된 바와 같은 하나 이상의 치환, 하나 이상의 결실 및/또는 하나 이상의 첨가를 포함할 수 있다.
- [0803] 서열번호: 118 내지 133 중 임의의 하나의 바람직한 변이체는 아미노-(N-) 말단 및/또는 카르복시(C-) 말단에서 Faz와 같은 비천연 아미노산을 갖는다. 서열번호: 118 내지 133 중 임의의 하나의 바람직한 변이체는 아미노-(N-) 말단 및/또는 카르복시(C-) 말단에 시스테인 잔기를 갖는다. 서열번호: 118 내지 133 중 임의의 하나의 바람직한 변이체는 아미노-(N-) 말단에 시스테인 잔기 및 카르복시(C-) 말단에 Faz와 같은 비천연 아미노산을 갖

거나 그 반대의 경우도 마찬가지이다.

- [0804] 서열번호: 118의 바람직한 변이체는 다음 변형 E54G, D151E, I196N 및 G357A 중 하나 이상, 예컨대 전부를 함유한다.
- [0805] 연결 없음
- [0806] 하나의 바람직한 구현예에서, 본 발명의 변형된 헬리카제에서 도입된 시스테인 및/또는 비천연 아미노산은 어떤 것도 서로 연결되지 않는다.
- [0807] 도입된 시스테인 및/또는 비천연 아미노산 중 2개 이상의 연결
- [0808] 또 다른 바람직한 구현예에서, 본 발명의 변형된 헬리카제에서 도입된 시스테인 및/또는 비천연 아미노산 중 2개 이상이 서로 연결된다. 이는 전형적으로 폴리뉴클레오티드로부터 결합해제하는 본 발명의 헬리카제의 능력을 감소시킨다.
- [0809] 2개 이상의 도입된 시스테인 및/또는 비천연 아미노산의 임의의 수 및 조합은 서로 연결될 수 있다. 예를 들면, 3, 4, 5, 6, 7, 8개 이상의 시스테인 및/또는 비천연 아미노산이 서로 연결될 수 있다. 하나 이상의 시스테인은 하나 이상의 시스테인에 연결될 수 있다. 하나 이상의 시스테인은 Faz와 같은 하나 이상의 비천연 아미노산에 연결될 수 있다. Faz와 같은 하나 이상의 비천연 아미노산은 Faz와 같은 하나 이상의 비천연 아미노산에 연결될 수 있다.
- [0810] 2개 이상의 시스테인 및/또는 비천연 아미노산은 어떠한 방식으로든 연결될 수 있다. 연결은 일시적일 수 있으며, 예를 들어 비공유일 수 있다. 일시적인 연결조차도 헬리카제로부터 폴리뉴클레오티드의 결합해제를 감소시킬 것이다.
- [0811] 2개 이상의 시스테인 및/또는 비천연 아미노산은 바람직하게는 친화성 분자에 연결된다. 적합한 친화성 분자는 당업계에 알려져 있다. 친화성 분자는 바람직하게는 (a) 상보적 폴리뉴클레오티드(WO 2010/086602, 전문이 본원에 참조로 포함됨), (b) 항체 또는 이의 단편 및 상보적 에피토프(Biochemistry 6thEd, W.H. Freeman and co (2007) pp953-954), (c) 펩티드 지퍼(O'Shea 등, Science 254(5031): 539-544), (d) β -시트 층대에 의해 상호 작용할 수 있는 것(Remaut and Waksman Trends Biochem. Sci.(2006) 31 436-444), (e) 수소 결합, 파이-적재 또는 염 가교를 형성할 수 있는 것, (f) 로탁산(Xiang Ma and He Tian Chem. Soc. Rev., 2010,39, 70-80), (g) 압타머 및 상보적 단백질(James, W. in Encyclopedia of Analytical Chemistry, R.A. Meyers(Ed.) pp. 4848-4871 John Wiley & Sons Ltd, Chichester, 2000) 또는 (h) 반킬레이터(Hammerstein 등 J Biol Chem. 2011 April 22; 286(16): 14324-14334)이다. (e)의 경우, 수소 결합은 전자음성 원자에 결합된 양성자와 또 다른 전자음성 원자 사이에서 발생한다. 파이-적재는 고리의 평면이 평행한 경우 함께 적재할 수 있는 2개의 방향족 고리를 필요로 한다. 염 가교는 여러 원자에 걸쳐 그들의 전자를 탈국소화할 수 있는 기 사이, 예를 들어 아스파르트레이트와 아르기닌 사이에 있다.
- [0812] 2개 이상의 부분은 6-his 태그 또는 Ni-NTA에 의해 일시적으로 연결될 수 있다.
- [0813] 2개 이상의 시스테인 및/또는 비천연 아미노산은 바람직하게는 영구적으로 연결된다. 본 발명의 맥락에서, 헬리카제가 사용되는 동안 연결이 끊어지지 않거나 -S-S- 결합을 열기 위해 환원을 사용하는 것과 같이 사용자의 부분에 대한 개입 없이 끊어질 수 없는 경우 연결은 영구적이다.
- [0814] 2개 이상의 시스테인 및/또는 비천연 아미노산은 바람직하게는 공유적으로 부착된다. 2개 이상의 시스테인 및/또는 비천연 아미노산은 당업계에 알려진 임의의 방법을 사용하여 공유적으로 부착될 수 있다.
- [0815] 2개 이상의 시스테인 및/또는 비천연 아미노산은 시스테인, 트레오닌, 세린, 아스파르트레이트, 아스파라긴, 글루타메이트 및 글루타민과 같은 자연 발생 아미노산을 통해 공유적으로 부착될 수 있다. 자연 발생 아미노산은 부착을 용이하게 하도록 변형될 수 있다. 예를 들면, 자연 발생 아미노산은 아실화, 인산화, 글리코실화 또는 파르네실화에 의해 변형될 수 있다. 다른 적합한 변형이 당업계에 알려져 있다. 자연 발생 아미노산에 대한 변형은 번역 후 변형일 수 있다. 2개 이상의 시스테인 및/또는 비천연 아미노산은 서열에 도입된 아미노산을 통해 부착될 수 있다. 이러한 아미노산은 바람직하게는 치환에 의해 도입된다. 도입된 아미노산은 부착을 용이하게 하는 시스테인 또는 비천연 아미노산일 수 있다. 적합한 비천연 아미노산은 4-아지도-L-페닐알라닌(Faz), Liu C. C. and Schultz P. G., Annu. Rev. Biochem., 2010, 79, 413-444의 도면 1에 포함된 1-71로 번호지정된 아미노산 중 임의의 하나 또는 아래에 나열된 아미노산 중 임의의 하나를 포함하나 이에 제한되지 않는다. 도입된

아미노산은 위에 논의된 바와 같이 변형될 수 있다.

- [0816] 바람직한 구현예에서, 2개 이상의 시스테인 및/또는 비천연 아미노산은 링커를 사용하여 연결된다. 링커 분자는 아래에 보다 상세하게 논의되어 있다. 하나의 적합한 연결 방법은 시스테인 연결이다. 이는 아래에 보다 상세하게 논의되어 있다. 2개 이상의 시스테인 및/또는 비천연 아미노산은 바람직하게는 하나 이상, 예컨대 2 또는 3개의 링커를 사용하여 연결된다. 하나 이상의 링커는 위에 논의된 바와 같이 개구의 크기를 줄이거나 단도록 설계될 수 있다. 하나 이상의 링커가 위에 논의된 바와 같이 개구를 닫는 데 사용되는 경우, 하나 이상의 링커의 적어도 일부는 바람직하게는 헬리카제에 의해 결합될 때 폴리뉴클레오티드와 평행하지 않도록 배향된다. 보다 바람직하게는, 모든 링커는 이러한 방식으로 배향된다. 하나 이상의 링커가 위에 논의된 바와 같이 개구를 닫는 데 사용되는 경우, 하나 이상의 링커의 적어도 일부는 바람직하게는 헬리카제에 의해 결합될 때 폴리뉴클레오티드와 평행하지 않는 배향으로 개구를 가로지른다. 보다 바람직하게는, 모든 링커는 이러한 방식으로 개구를 가로지른다. 이러한 구현예에서, 하나 이상의 링커의 적어도 일부는 폴리뉴클레오티드에 수직일 수 있다. 이러한 배향은 폴리뉴클레오티드가 개구를 통해 헬리카제로부터 결합해제할 수 없도록 개구를 효과적으로 닫는다.
- [0817] 각 링커는 2개 이상의 기능적 단부, 예컨대 2, 3 또는 4개의 기능적 단부를 가질 수 있다. 링커에서 단부의 적합한 구성은 당업계에 잘 알려져 있다.
- [0818] 하나 이상의 링커의 하나 이상의 단부는 바람직하게는 헬리카제에 공유적으로 부착된다. 하나의 단부가 공유적으로 부착된 경우, 하나 이상의 링커는 위에 논의된 바와 같은 2개 이상의 시스테인 및/또는 비천연 아미노산을 일시적으로 연결할 수 있다. 두 말단 또는 모든 말단이 공유적으로 부착된 경우, 하나 이상의 링커는 2개 이상의 시스테인 및/또는 비천연 아미노산을 영구적으로 연결한다.
- [0819] 하나 이상의 링커는 바람직하게는 아미노산 서열 및/또는 화학적 가교제이다.
- [0820] 펩티드 링커와 같은 적합한 아미노산 링커는 당업계에 알려져 있다. 아미노산 또는 펩티드 링커의 길이, 유연성 및 친수성은 전형적으로 개구의 크기를 줄이지만, 헬리카제의 기능을 방해하지 않도록 설계된다. 바람직한 유연한 펩티드 링커는 2 내지 20개, 예컨대 4, 6, 8, 10 또는 16개의 세린 및/또는 글리신 아미노산의 스트레치이다. 보다 바람직한 유연한 링커는 (SG)₁, (SG)₂, (SG)₃, (SG)₄, (SG)₅, (SG)₈, (SG)₁₀, (SG)₁₅ 또는 (SG)₂₀을 포함하며 여기서 S는 세린이고 G는 글리신이다. 바람직한 견고한 링커는 2 내지 30개, 예컨대 4, 6, 8, 16 또는 24개의 프롤린 아미노산의 스트레치이다. 보다 바람직한 견고한 링커는 (P)₁₂를 포함하며 여기서 P는 프롤린이다. 링커의 아미노산 서열은 바람직하게는 폴리뉴클레오티드 결합 모이어터를 포함한다. 이러한 모이어터 및 이들의 사용과 연관된 이점은 아래에 논의되어 있다.
- [0821] 적합한 화학적 가교제는 당업계에 잘 알려져 있다. 적합한 화학적 가교제는 다음 작용 기를 포함하는 것을 포함하나 이에 제한되지 않는다: 말레이미드, 활성 에스테르, 숙신이미드, 아지드, 알킨(예컨대 디벤조사이클로옥티놀(DIBO) 또는 DBCO), 디플루오로 사이클로알킨 및 선형 알킨), 포스핀(예컨대 흔적이 있고 흔적이 없는 Staudinger 결합에 사용되는 것들), 할로아세틸(예컨대 요오도아세트아미드), 포스젠 유형 시약, 술폰일 클로라이드 시약, 이소티오시아네이트, 아실 할라이드, 하이드라진, 디술파이드, 비닐 술폰, 아지리딘 및 광반응성 시약(예컨대 아릴 아지드, 디아지리딘).
- [0822] 아미노산과 작용 기 사이의 반응은 시스테인/말레이미드와 같이 자발적일 수 있거나, 아지드 및 선형 알킨을 연결하기 위한 Cu(I)와 같은 외부 시약을 필요로 할 수 있다.
- [0823] 링커는 필요한 거리를 가로질러 늘어나는 임의의 분자를 포함할 수 있다. 링커는 하나의 탄소(포스젠 유형 링커)에서 많은 옹스트롱까지 길이가 달라질 수 있다. 선형 분자의 예는 폴리에틸렌글리콜(PEG), 폴리펩티드, 이당류, 데옥시리보핵산(DNA), 펩티드 핵산(PNA), 트레오스 핵산(TNA), 글리세롤 핵산(GNA), 포화 및 불포화 탄화수소, 폴리아미드를 포함하나 이에 제한되지 않는다. 이러한 링커는 불활성 또는 반응성일 수 있으며, 특히 이들은 정의된 위치에서 화학적으로 절단가능할 수 있거나, 그 자체가 형광단 또는 리간드로 변형될 수 있다. 링커는 바람직하게는 디티오프레이톨(DTT)에 저항성이 있다.
- [0824] 바람직한 가교제는 2,5-디옥소피롤리딘-1-일 3-(피리딘-2-일디술파닐)프로파노에이트, 2,5-디옥소피롤리딘-1-일 4-(피리딘-2-일디술파닐)부타노에이트 및 2,5-디옥소피롤리딘-1-일 8-(피리딘-2-일디술파닐)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-디하이드록시부탄(BMDB), BM[PEO]2(1,8-비스-말레이미도디에틸렌글리콜), BM[PEO]3(1,11-비스-말레이미도트

리에틸렌 글리콜), 트리스[2-말레이미도에틸]아민(TMEA), DTME 디티오비스말레이미도에탄, 비스-말레이미드 PEG3, 비스-말레이미드 PEG11, DBCO-말레이미드, DBCO-PEG4-말레이미드, DBCO-PEG4-NH₂, DBCO-PEG4-NHS, DBCO-PEG-DBCO 2.8kDa, DBCO-PEG-DBCO 4.0kDa, DBCO-15 원자-DBCO, DBCO-26 원자-DBCO, DBCO-35 원자-DBCO, DBCO-PEG4-S-S-PEG3-비오틴, DBCO-S-S-PEG3-비오틴, DBCO-S-S-PEG11-비오틴, (숙신이미딜 3-(2-피리딜디티오)프로피오네이트(SPDP) 및 말레이미드-PEG(2kDa)-말레이미드(알파,오메가-비스-말레이미도 폴리(에틸렌 글리콜))를 포함한다. 가장 바람직한 가교제는 말레이미드-프로필-SRDFWRS-(1,2-디아미노에탄)-프로필-말레이미드이다.

[0825] 하나 이상의 링커는 절단가능할 수 있다. 이는 아래에 보다 상세하게 논의되어 있다.

[0826] 2개 이상의 시스테인 및/또는 비천연 아미노산은 서로에 특이적인 2개의 상이한 링커를 사용하여 연결될 수 있다. 링커 중 하나는 하나의 부분에 부착되고 다른 것은 또 다른 부분에 부착된다. 링커는 본 발명의 변형된 헬리카제를 형성하도록 반응해야 한다. 2개 이상의 시스테인 및/또는 비천연 아미노산은 WO 2010/086602(이의 전문이 본원에 참조로 포함됨)에 기재된 혼성화 링커를 사용하여 연결될 수 있다. 특히, 2개 이상의 시스테인 및/또는 비천연 아미노산은 각각 혼성화가능한 영역 및 공유 결합을 형성할 수 있는 기를 포함하는 2개 이상의 링커를 사용하여 연결될 수 있다. 링커의 혼성화가능한 영역은 2개 이상의 시스테인 및/또는 비천연 아미노산을 혼성화하고 연결한다. 그런 다음 연결된 시스테인 및/또는 비천연 아미노산은 기 사이에 공유 결합 형성을 통해 결합된다. WO 2010/086602(이의 전문이 본원에 참조로 포함됨)에 개시된 특이적 링커 중 임의의 것이 본 발명에 따라 사용될 수 있다.

[0827] 2개 이상의 시스테인 및/또는 비천연 아미노산은 변형된 다음 2개의 변형에 특이적인 화학적 가교제를 사용하여 부착될 수 있다. 위에 논의된 가교제 중 임의의 것이 사용될 수 있다.

[0828] 링커는 라벨링될 수 있다. 적합한 라벨은 형광 분자(예컨대 Cy3 또는 AlexaFluor®555), 방사성 동위원소, 예를 들어 ¹²⁵I, ³⁵S, 효소, 항체, 항원, 폴리뉴클레오타이드 및 비오틴과 같은 리간드를 포함하나 이에 제한되지 않는다. 이러한 라벨은 링커의 양을 정량화할 수 있게 한다. 라벨은 또한 비오틴과 같은 절단가능한 정제 태그 이거나, 단백질 자체에 존재하지 않지만, 트립신 소화에 의해 방출되는 펩티드와 같은 식별 방법으로 나타나는 특이적 서열일 수 있다.

[0829] 2개 이상의 시스테인을 연결하는 바람직한 방법은 시스테인 연결을 통한 것이다. 이는 이작용성 화학적 가교제 또는 말단에 제시된 시스테인 잔기가 있는 아미노산 링커에 의해 매개될 수 있다.

[0830] 임의의 이작용성 링커의 길이, 반응성, 특이성, 강성 및 용해성은 개구의 크기가 충분히 감소되고 헬리카제의 기능이 유지되도록 설계될 수 있다. 적합한 링커는 비스말레이미드 가교제, 예컨대 1,4-비스(말레이미도)부탄(BMB) 또는 비스(말레이미도)헥산을 포함한다. 이작용성 링커의 하나의 단점은 특이적 부위에서 부착이 바람직한 경우 헬리카제의 요건이 더 이상 표면 접근가능한 시스테인 잔기를 함유하지 않아야 한다는 점인데, 이작용성 링커와 표면 접근가능한 시스테인 잔기의 결합이 제거하기 어려울 수 있고 기질 결합 또는 활성에 영향을 미칠 수 있기 때문이다. 헬리카제가 여러 접근가능한 시스테인 잔기를 함유하는 경우, 변형이 헬리카제의 접합 또는 활성에 영향을 미치지 않도록 하면서 이를 제거하기 위해 헬리카제의 변형이 필요할 수 있다. 이는 WO 2010/086603(이의 전문이 본원에 참조로 포함됨)에 논의되어 있다. 시스테인 잔기의 반응성은 예를 들어 펩티드 링커에서 인접한 잔기의 변형에 의해 향상될 수 있다. 예를 들면, 측면 아르기닌, 히스티딘 또는 리신 잔기의 염기성 기는 시스테인 티올 기의 pKa를 더 반응성 S⁻ 기의 pKa로 변경할 것이다. 시스테인 잔기의 반응성은 5,5'-디티오비스-(2-니트로벤조산)(dTNB)과 같은 티올 보호 기에 의해 보호될 수 있다. 이들은 링커가 부착되기 전에 헬리카제의 하나 이상의 시스테인 잔기와 반응할 수 있다. 표면 접근가능한 시스테인의 선택적 탈보호는 비드에 고정화된 환원 시약(예를 들어 고정화된 트리스(2-카르복시에틸) 포스핀, TCEP)을 사용하여 가능할 수 있다. 시스테인 연결은 아래에 보다 상세하게 논의되어 있다.

[0831] Faz 연결을 통한 또 다른 바람직한 부착 방법. 이는 이작용성 화학적 링커 또는 말단에 제시된 Faz 잔기가 있는 폴리펩티드 링커에 의해 매개될 수 있다.

[0832] 다른 변형

[0833] 본 발명의 헬리카제는 또한 (i) 타워 도메인 및 (ii) 핀 도메인 및/또는 1A 도메인 사이의 인력을 증가시키도록 변형될 수 있다. 본 발명에 따라 임의의 알려진 화학적 변형이 이루어질 수 있다. 이러한 유형의 변형은 WO 2015/055981(그 전문이 본원에 참조로 포함됨)에 개시되어 있다.

- [0834] 특히, 본 발명은 적어도 하나의 하전된 아미노산이 (i) 타워 도메인 및/또는 (ii) 핀 도메인 및/또는 (iii) 1A(RecA-유사 모터) 도메인에 도입된 본 발명의 헬리카제를 제공하며, 여기서 헬리카제는 폴리뉴클레오티드의 이동을 제어하는 능력을 갖는다. 헬리카제가 폴리뉴클레오티드의 이동을 제어하는 능력은 위에 논의된 바와 같이 측정될 수 있다. 본 발명은 바람직하게는 적어도 하나의 하전된 아미노산이 (i) 타워 도메인 및 (ii) 핀 도메인 및/또는 1A 도메인에 도입된 본 발명의 헬리카제를 제공한다.
- [0835] 적어도 하나의 하전된 아미노산은 음으로 하전되거나 양으로 하전될 수 있다. 적어도 하나의 하전된 아미노산은 바람직하게는 헬리카제에서 상호작용하는 임의의 아미노산(들)과 반대로 하전된다. 예를 들면, 적어도 하나의 양으로 하전된 아미노산은 핀 도메인의 음으로 하전된 아미노산과 상호작용하는 위치에서 타워 도메인에 도입될 수 있다. 적어도 하나의 하전된 아미노산은 전형적으로 야생형(즉 비변형된) 헬리카제에서 하전되지 않은 위치에 도입된다. 적어도 하나의 하전된 아미노산은 헬리카제에서 적어도 하나의 반대로 하전된 아미노산을 대체하는 데 사용될 수 있다. 예를 들면, 양으로 하전된 아미노산은 음으로 하전된 아미노산을 대체하는 데 사용될 수 있다.
- [0836] 적합한 하전된 아미노산은 위에 논의되어 있다. 적어도 하나의 하전된 아미노산은 아르기닌(R), 히스티딘(H), 리신(K), 아스파르트산(D) 또는 글루탐산(D)과 같이 천연일 수 있다. 대안적으로, 적어도 하나의 하전된 아미노산은 인공적 또는 비천연일 수 있다. 임의의 수의 하전된 아미노산이 각 도메인에 도입될 수 있다. 예를 들면, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 하전된 아미노산이 각 도메인에 도입될 수 있다.
- [0837] 헬리카제는 바람직하게는 다음 위치 중 하나 이상에서 양으로 하전된 아미노산을 포함하는 서열번호: 118의 변이체를 포함한다: (i) 93; (ii) 354; (iii) 360; (iv) 361; (v) 94; (vi) 97; (vii) 155; (viii) 357; (ix) 100; 및 (x) 127. 헬리카제는 바람직하게는 다음 위치 중 하나 이상에서 음으로 하전된 아미노산을 포함하는 서열번호: 118의 변이체를 포함한다: (i) 354; (ii) 358; (iii) 360; (iv) 364; (v) 97; (vi) 123; (vii) 155; (viii); 357; (ix) 100; 및 (x) 127. 헬리카제는 바람직하게는 (i) 내지 (x) 중 임의의 것에 정의된 바와 같이 서열번호: 118의 위치에 상응하는 위치에서 양으로 하전된 아미노산 또는 음으로 하전된 아미노산을 포함하는 서열번호: 119 내지 133 중 임의의 하나의 변이체를 포함한다. 서열번호: 118의 위치에 상응하는 서열번호: 119 내지 133 중 임의의 하나의 위치는 아래의 서열번호: 118 내지 133의 정렬을 사용하여 식별될 수 있다.
- [0838] 헬리카제는 바람직하게는 다음 위치에서 반대로 하전된 아미노산을 포함하도록 적어도 하나의 하전된 아미노산을 도입하여 변형된 서열번호: 118의 변이체를 포함한다: (i) 93 및 354; (ii) 93 및 358; (iii) 93 및 360; (iv) 93 및 361; (v) 93 및 364; (vi) 94 및 354; (vii) 94 및 358; (viii) 94 및 360; (ix) 94 및 361; (x) 94 및 364; (xi) 97 및 354; (xii) 97 및 358; (xiii) 97 및 360; (xiv) 97 및 361; (xv) 97 및 364; (xvi) 123 및 354; (xvii) 123 및 358; (xviii) 123 및 360; (xix) 123 및 361; (xx) 123 및 364; (xxi) 155 및 354; (xxii) 155 및 358; (xxiii) 155 및 360; (xxiv) 155 및 361; (xxv) 155 및 364. 본 발명의 헬리카제는 바람직하게는 (i) 내지 (xxv) 중 임의의 것에 정의된 바와 같이 서열번호: 118의 위치에 상응하는 위치에서 반대로 하전된 아미노산을 포함하는 서열번호: 119 내지 133 중 임의의 하나의 변이체를 포함한다.
- [0839] 본 발명은 또한 (i) 적어도 하나의 하전된 아미노산이 타워 도메인에 도입되고 (ii) 적어도 하나의 반대로 하전된 아미노산이 핀 도메인 및/또는 1A(RecA-유사 모터) 도메인에 도입된 헬리카제를 제공하며, 여기서 헬리카제는 폴리뉴클레오티드의 이동을 제어하는 능력을 갖는다. 적어도 하나의 하전된 아미노산은 음으로 하전될 수 있고 적어도 하나의 반대로 하전된 아미노산은 양으로 하전될 수 있거나 그 반대의 경우도 마찬가지이다. 적합한 하전된 아미노산은 위에 논의되어 있다. 임의의 수의 하전된 아미노산 및 임의의 수의 반대로 하전된 아미노산이 도입될 수 있다. 예를 들면, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 하전된 아미노산이 도입될 수 있고/있거나 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 반대로 하전된 아미노산이 도입될 수 있다.
- [0840] 하전된 아미노산은 전형적으로 야생형 헬리카제에서 하전되지 않은 위치에 도입된다. 하전된 아미노산 중 하나 또는 둘 다는 헬리카제에서 하전된 아미노산을 대체하는 데 사용될 수 있다. 예를 들면, 양으로 하전된 아미노산은 음으로 하전된 아미노산을 대체하는 데 사용될 수 있다. 하전된 아미노산은 위에 논의된 (i) 타워 도메인 및 (ii) 핀 도메인 및/또는 1A 도메인의 위치 중 임의의 것에 도입될 수 있다. 반대로 하전된 아미노산은 전형적으로 생성된 헬리카제에서 상호작용하도록 도입된다. 헬리카제는 바람직하게는 반대로 하전된 아미노산이 다음 위치에서 도입된 서열번호: 118의 변이체를 포함한다: (i) 97 및 354; (ii) 97 및 360; (iii) 155 및 354; 또는 (iv) 155 및 360. 본 발명의 헬리카제는 바람직하게는 (i) 내지 (iv) 중 임의의 것에 정의된 바와 같이 서열번호: 118의 위치에 상응하는 위치에서 반대로 하전된 아미노산을 포함하는 서열번호: 119 내지 133 중 임의의 하나의 변이체를 포함한다.

- [0841] 작제물
- [0842] 본 발명은 또한 본 발명의 변형된 헬리카제 및 추가적인 폴리뉴클레오티드 결합 모이어티를 포함하는 작제물을 제공하며, 여기서 헬리카제는 폴리뉴클레오티드 결합 모이어티에 부착되고 작제물은 폴리뉴클레오티드의 이동을 제어하는 능력을 갖는다. 작제물은 인공적 또는 비천연이다.
- [0843] 본 발명의 작제물은 가닥 서열분석 동안 폴리뉴클레오티드의 이동을 제어하는 데 유용한 도구이다. 본 발명의 작제물은 본 발명의 변형된 헬리카제보다 서열분석되는 폴리뉴클레오티드로부터 풀릴 가능성이 훨씬 더 적다. 작제물은 나노기공을 통해 폴리뉴클레오티드의 전위를 제어하기 때문에 폴리뉴클레오티드의 더 큰 판독 길이를 제공할 수 있다.
- [0844] 특이적 폴리뉴클레오티드 서열에 결합하는 표적화된 작제물이 또한 설계될 수 있다. 아래에 보다 상세하게 논의된 바와 같이, 폴리뉴클레오티드 결합 모이어티는 특이적 폴리뉴클레오티드 서열에 결합하여 작제물의 헬리카제 부분을 특이적 서열에 표적할 수 있다.
- [0845] 작제물은 폴리뉴클레오티드의 이동을 제어하는 능력을 갖는다. 이는 위에 논의된 바와 같이 결정될 수 있다.
- [0846] 본 발명의 작제물은 단리되거나, 실질적으로 단리되거나, 정제되거나 실질적으로 정제될 수 있다. 작제물은 지질, 폴리뉴클레오티드 또는 기공 단량체와 같은 임의의 다른 구성요소가 완전히 없는 경우 단리되거나 정제된다. 작제물은 의도된 용도를 방해하지 않을 담체 또는 희석제와 혼합된 경우 실질적으로 단리된다. 예를 들면, 작제물은 지질, 폴리뉴클레오티드 또는 기공 단량체와 같은 다른 구성요소의 10% 미만, 5% 미만, 2% 미만 또는 1% 미만을 포함하는 형태로 존재하는 경우 실질적으로 단리되거나 실질적으로 정제된다.
- [0847] 헬리카제는 위에 논의된 본 발명의 헬리카제 중 임의의 것일 수 있다.
- [0848] 헬리카제는 바람직하게는 추가적인 폴리뉴클레오티드 결합 모이어티에 공유적으로 부착된다. 헬리카제는 하나 초과, 예컨대 2 또는 3개의 지점에서 모이어티에 부착될 수 있다.
- [0849] 헬리카제는 당업계에 알려진 임의의 방법을 사용하여 모이어티에 공유적으로 부착될 수 있다. 적합한 방법은 2개 이상의 부분을 연결하는 것과 관련하여 위에 논의되어 있다.
- [0850] 헬리카제 및 모이어티는 별도로 생산된 다음 함께 부착될 수 있다. 2가지 구성요소는 임의의 구성으로 부착될 수 있다. 예를 들면, 이들은 말단(즉 아미노 또는 카르복시 말단) 아미노산을 통해 부착될 수 있다. 적합한 구성은 모이어티의 아미노 말단이 헬리카제의 카르복시 말단에 부착되는 것 및 그 반대의 경우도 포함하나 이에 제한되지 않는다. 대안적으로, 2가지 구성요소는 서열 내의 아미노산을 통해 부착될 수 있다. 예를 들면, 모이어티는 헬리카제의 루프 영역에서 하나 이상의 아미노산에 부착될 수 있다. 바람직한 구현예에서, 모이어티의 말단 아미노산은 헬리카제의 루프 영역에서 하나 이상의 아미노산에 부착된다.
- [0851] 바람직한 구현예에서, 헬리카제는 예를 들면 위에 논의된 바와 같은 하나 이상의 링커 분자를 통해 모이어티에 화학적으로 부착된다. 또 다른 바람직한 구현예에서, 헬리카제는 모이어티에 유전적으로 융합된다. 헬리카제는 전체 작제물이 단일 폴리뉴클레오티드 서열로부터 발현되는 경우 모이어티에 유전적으로 융합된다. 헬리카제 및 모이어티의 코딩 서열은 임의의 방식으로 조합되어 작제물을 암호화하는 단일 폴리뉴클레오티드 서열을 형성할 수 있다. 핵산 결합 단백질에 기공의 유전적 융합은 WO 2010/004265(이의 전문이 본원에 참조로 포함됨)에 논의되어 있다.
- [0852] 헬리카제 및 모이어티는 임의의 구성으로 유전적으로 융합될 수 있다. 헬리카제 및 모이어티는 말단 아미노산을 통해 융합될 수 있다. 예를 들면, 모이어티의 아미노 말단은 헬리카제의 카르복시 말단에 융합될 수 있고 그 반대의 경우도 마찬가지이다. 모이어티의 아미노산 서열은 바람직하게는 헬리카제의 아미노산 서열에 프레임 내에서 첨가된다. 다시 말해서, 모이어티는 바람직하게는 헬리카제의 서열 내에 삽입된다. 이러한 구현예에서, 헬리카제 및 모이어티는 전형적으로 2개 지점에서, 즉 모이어티의 아미노 및 카르복시 말단 아미노산을 통해 부착된다. 모이어티가 헬리카제의 서열 내에 삽입되는 경우, 모이어티의 아미노 및 카르복시 말단 아미노산이 근접해 있고 각각 헬리카제 또는 이의 변이체의 서열에서 인접한 아미노산에 부착되는 것이 바람직하다. 바람직한 구현예에서, 모이어티는 헬리카제의 루프 영역에 삽입된다.
- [0853] 헬리카제는 모이어티에 직접 부착될 수 있다. 헬리카제는 바람직하게는 위에 논의된 바와 같이 하나 이상, 예컨대 2 또는 3개의 링커를 사용하여 모이어티에 부착된다. 하나 이상의 링커는 모이어티의 이동성을 제약하도록 설계될 수 있다. 헬리카제 및/또는 모이어티는 위에 논의된 바와 같이 하나 이상의 링커의 부착을 용이하게 하

도록 변형될 수 있다.

- [0854] 절단가능한 링커는 비부착된 구성요소로부터 작제물의 분리를 돕는 것으로 사용될 수 있고 합성 반응을 추가로 제어하는 데 사용될 수 있다. 예를 들어, 이중-이작용성 링커는 헬리카제와 반응할 수 있지만, 모이어티와는 반응하지 않을 수 있다. 링커의 자유 단부를 사용하여 헬리카제 단백질을 표면에 결합할 수 있는 경우, 첫번째 반응의 반응하지 않은 헬리카제를 혼합물에서 제거할 수 있다. 후속적으로, 링커를 절단하여 모이어티와 반응하는 기에 노출시킬 수 있다. 게다가, 연결 반응의 이러한 순서를 따름으로써, 조건을 먼저 헬리카제에 대한 반응에 최적화한 다음, 링커의 절단 후에 모이어티에 대한 반응에 최적화할 수 있다. 두번째 반응은 또한 링커가 이미 부착된 영역에 한정되기 때문에 모이어티와의 올바른 반응 부위를 향해 훨씬 더 지시될 수 있다.
- [0855] 헬리카제는 헬리카제/가교제 복합체가 모이어티에 공유적으로 부착되기 전에 이작용성 가교제에 공유적으로 부착될 수 있다. 대안적으로, 모이어티는 이작용성 가교제/모이어티 복합체가 헬리카제에 부착되기 전에 이작용성 가교제에 공유적으로 부착될 수 있다. 헬리카제 및 모이어티는 동시에 화학적 가교제에 공유적으로 부착될 수 있다.
- [0856] 헬리카제를 모이어티에 부착하는 바람직한 방법은 위에 기재된 바와 같은 시스템 연결 및 Faz 연결이다. 바람직한 구현예에서, 반응성 시스템은 모이어티에 유전적으로 부착된 캡티드 링커에 존재한다. 이는 모이어티로부터 다른 접근가능한 시스템 잔기를 제거하기 위해 추가적인 변형이 반드시 필요하지 않음을 의미한다.
- [0857] 헬리카제 또는 모이어티를 그들 자체에 교차 연결하는 것은 엄청난 과잉 헬리카제 및/또는 모이어티에서 링커의 농도를 유지함으로써 방지될 수 있다. 대안적으로, 2개의 링커가 사용되는 "자물쇠 및 열쇠" 배열이 사용될 수 있다. 각 링커의 한쪽 단부만 함께 반응하여 더 긴 링커를 형성할 수 있고 링커의 다른 단부는 각각 작제물의 상이한 부분(즉 헬리카제 또는 모이어티)과 반응할 수 있다. 이는 아래에 보다 상세하게 논의되어 있다.
- [0858] 부착 부위는 작제물 폴리뉴클레오티드와 접촉할 때, 헬리카제 및 모이어티 둘 다 폴리뉴클레오티드에 결합하고 이동을 제어할 수 있도록 선택된다.
- [0859] 본 발명의 변형된 헬리카제 및 작제물을 수반하는 구현예
- [0860] 본 발명은 폴리뉴클레오티드, 벡터 및 숙주 세포를 제공한다. 이들은 위에 논의되어 있다.
- [0861] 본 발명은 특히 분석물의 이동을 제어하고 표적 분석물을 특성화하기 위해 본 발명의 Dda 헬리카제를 사용하는 다양한 방법을 제공한다. 이러한 방법은 상기 및 W02016/034591, W02017/149316, W02017/149317, W02017/149318, W02018/211241, W02019/002893, W02015/055981, W02015/166276 및 W02016/055777(모두 본원에 참조로 포함됨)에 기재되어 있다.
- [0862] 본 발명의 변형된 Dda 헬리카제 및 작제물은 표적 분석물을 특성화하기 위한 센서를 형성하는 데 사용될 수 있다. 이러한 센서는 표적 분석물의 존재 하에 기공 및 헬리카제 또는 작제물을 접촉시키고 기공을 가로질러 전위를 인가함으로써 형성될 수 있다. 헬리카제 또는 작제물은 예를 들면 위에 기재된 바와 같이 기공에 공유적으로 부착될 수 있다.
- [0863] 이러한 방법 및 센서에서, 본 발명의 변형된 Dda 헬리카제 및 작제물은 본 발명의 기공과 조합/사용될 수 있다. 대안적으로, 이들은 고체 상태 기공과 함께 사용될 수 있거나 막관통 단백질 기공은 용혈소, 백혈구독, 바이크 박테리움 스메그마티스(Mycobacterium smegmatis) 포린 A(MspA), MspB, MspC, MspD, 리세닌, 외부 막 포린 F(OmpF), 외부 막 포린 G(OmpG), 외부 막 포스포리파제 A, 네이세리아(Neisseria) 자동수송체 지단백질(NaIP) 및 WZA, CsgG, CsgG/CsgF, 리세닌, ClyA, Sp1 또는 용혈성 단백질 프라가세아톡신 C(FraC)로부터 유래된다. 기공 및 헬리카제의 바람직한 조합은 아래에 보다 상세하게 논의되어 있다.
- [0864] 키트
- [0865] 본 발명은 또한 (a) 본 발명의 기공 및 헬리카제 또는 작제물 또는 (b) 본 발명의 헬리카제 또는 작제물 및 하나 이상의 로딩 모이어티를 포함하는 표적 분석물을 특성화하기 위한 키트를 제공한다. 기공은 본 발명의 기공을 포함하여 위에 논의된 것들 중 임의의 것일 수 있다. 기공 및 헬리카제의 바람직한 조합은 아래에 보다 상세하게 논의되어 있다.
- [0866] 본 발명은 또한 (a) 본 발명의 헬리카제 또는 작제물 및 (b) 본 발명의 단리된 CsgG 기공 또는 단리된 기공 복합체를 포함하는 표적 분석물, 바람직하게는 표적 폴리뉴클레오티드를 특성화하기 위한 키트를 제공한다. 기공 및 헬리카제의 바람직한 조합은 아래에 보다 상세하게 논의되어 있다.

- [0867] 키트는 바람직하게는 막의 구성요소를 추가로 포함한다. 키트는 양친매성 층 또는 트리블록 공중합체 막과 같은 임의의 유형의 막의 구성요소를 포함할 수 있다. 키트는 표적 분석물을 막에 결합하기 위한 하나 이상의 앵커(anchor), 예컨대 콜레스테롤을 추가로 포함할 수 있다. 키트는 폴리뉴클레오티드의 특성화를 용이하게 하기 위해 표적 폴리뉴클레오티드에 부착될 수 있는 하나 이상의 폴리뉴클레오티드 어댑터를 추가로 포함할 수 있다. 하나의 구현예에서, 콜레스테롤과 같은 앵커는 폴리뉴클레오티드 어댑터에 부착된다.
- [0868] 본 발명은 (a) 본 발명의 단리된 기공 또는 단리된 기공 복합체 및 (b) 막의 구성요소 및 (c) 폴리뉴클레오티드 결합 단백질 중 하나 또는 둘 다를 포함하는 표적 분석물을 특성화하기 위한 키트는 제공한다.
- [0869] 바람직한 폴리뉴클레오티드 결합 단백질은 중합효소, 엑소뉴클레아제, 헬리카제 및 토포이소머라제, 예컨대 자이라제이다. 적합한 효소는 이. 콜라이의 엑소뉴클레아제 I, 이. 콜라이의 엑소뉴클레아제 III 효소, 티. 썬모필러스(*T. thermophilus*)의 RecJ 및 박테리오파지 람다 엑소뉴클레아제, TatD 엑소뉴클레아제 및 이의 변이체를 포함하나 이에 제한되지 않는다. 티. 썬모필러스의 RecJ 서열 또는 이의 변이체를 포함하는 3개의 서브유닛은 상호작용하여 삼량체 엑소뉴클레아제를 형성한다. 중합효소는 PyroPhage® 3173 DNA 중합효소(Lucigen® Corporation에서 상업적으로 입수가능), SD 중합효소(Bioron®에서 상업적으로 입수가능) 또는 이의 변이체일 수 있다. 효소는 Phi29 DNA 중합효소(서열번호: 7) 또는 이의 변이체일 수 있다. 토포이소머라제는 바람직하게는 모이어티 분류(EC) 그룹 5.99.1.2 및 5.99.1.3 중 임의의 것의 구성원이다.
- [0870] 효소는 가장 바람직하게는 Hel308 Mbu, Hel308 Csy, Hel308 Tga, Hel308 Mhu, TraI Eco, XPD Mbu 또는 이의 변이체와 같은 헬리카제로부터 유래된다. 임의의 헬리카제가 본 발명에서 사용될 수 있다. 헬리카제는 Hel308 헬리카제, RecD 헬리카제, 예컨대 TraI 헬리카제 또는 TrwC 헬리카제, XPD 헬리카제 또는 Dda 헬리카제일 수 있거나 이로부터 유래될 수 있다. 헬리카제는 WO 2013/057495; WO 2013/098562; WO2013098561; WO 2014/013260; WO 2014/013259; WO 2014/013262 및 WO 2015/055981에 개시된 헬리카제, 변형된 헬리카제 또는 헬리카제 작제물 중 임의의 것일 수 있다. 이들 모두는 그 전문이 본원에 참조로 포함된다.
- [0871] 키트는 위에 언급된 임의의 구현예가 수행될 수 있도록 하는 하나 이상의 다른 시약 또는 기구를 추가로 포함할 수 있다. 이러한 시약 또는 기구는 다음 중 하나 이상을 포함한다: 적합한 완충액(들)(수용액), 대상체로부터 샘플을 획득하기 위한 수단(예컨대 바늘을 포함하는 용기 또는 기구), 폴리뉴클레오티드를 증폭 및/또는 발현하기 위한 수단 또는 전압 또는 패치 클램프 장치. 시약은 유체 샘플이 시약을 재현탁하도록 건조 상태로 키트에 존재할 수 있다. 키트는 또한 임의적으로 키트가 본 발명의 방법에서 사용될 수 있게 하는 지침 또는 방법이 사용될 수 있는 유기체에 관한 세부사항을 포함할 수 있다. 마지막으로, 키트는 또한 분석물 특성화에 유용한 추가적인 구성요소를 포함할 수 있다.
- [0872] 장치
- [0873] 본 발명은 또한 (a) 복수의 기공 및 (b) 본 발명의 복수의 헬리카제 또는 복수의 작제물을 포함하는, 샘플에서 표적 분석물을 특성화하기 위한 장치를 제공한다. 복수의 기공은 본 발명의 기공을 포함하여 위에 논의된 것들 중 임의의 것일 수 있다. 기공 및 헬리카제의 바람직한 조합은 아래에 보다 상세하게 논의되어 있다.
- [0874] 본 발명은 또한 시험관 내 막에 삽입된 본 발명의 막관통 단백질 기공 또는 기공 복합체를 포함하는 장치를 제공한다.
- [0875] 본 발명은 또한 (i) 본 발명의 단리된 기공 또는 단리된 기공 복합체를 획득하는 단계 및 (ii) 단리된 기공 또는 단리된 기공 복합체를 시험관 내 막과 접촉시켜 기공이 시험관 내 막에 삽입되도록 하는 단계를 포함하는 방법에 의해 생산된 장치를 제공한다.
- [0876] 특히 본 발명의 기공, 기공 복합체, 헬리카제 및 작제물과 관련하여 위에 논의된 임의의 특이적 구현예는 본 발명의 장치에 동일하게 적용가능하다.
- [0877] 본 발명의 막
- [0878] 본 발명은 또한 본 발명의 기공 또는 기공 복합체를 포함하는 막을 제공한다. 기공은 바람직하게는 막에 존재하여, 함께 막관통 기공을 형성한다. 막은 양친매성 층 또는 트리블록 공중합체 막과 같은 임의의 유형의 막의 구성요소를 포함할 수 있다. 막은 기공에 부착된 폴리뉴클레오티드 결합 단백질 또는 폴리펩티드 취급 효소를 추가로 포함할 수 있다. 막은 폴리뉴클레오티드 또는 폴리펩티드를 막에 결합하기 위한 하나 이상의 앵커를 추가로 포함할 수 있다. 기공은 위에 논의된 것들 중 임의의 것일 수 있다.

- [0879] 어레이
- [0880] 본 발명은 또한 본 발명의 복수의 막을 포함하는 어레이를 제공한다. 본 발명의 막에 대해 위에 논의된 임의의 구현에는 본 발명의 어레이에 동일하게 적용된다. 어레이는 아래에 기재된 방법 중 임의의 것을 수행하도록 설정될 수 있다.
- [0881] 바람직한 구현예에서, 어레이에서 각 막은 하나의 기공을 포함한다. 어레이가 형성된 방식으로 인해, 예를 들어, 어레이는 기공을 포함하지 않는 하나 이상의 막, 및/또는 2개 이상의 기공을 포함하는 하나 이상의 막을 포함할 수 있다. 어레이는 약 2 내지 약 1000개, 예컨대 약 10 내지 약 800개, 약 20 내지 약 600개 또는 약 30 내지 약 500개의 막을 포함할 수 있다.
- [0882] 시스템
- [0883] 본 발명은 (a) 본 발명의 막 또는 본 발명의 어레이, (b) 막(들)을 가로질러 전위를 인가하기 위한 수단 및 (c) 막(들)을 가로질러 전기적 또는 광학적 신호를 검출하기 위한 수단을 포함하는 시스템을 제공한다.
- [0884] 기공 및 막은 위 및 아래에 기재된 바와 같은 것일 수 있다.
- [0885] 하나의 구현예에서, 시스템은 제1 챔버 및 제2 챔버를 추가로 포함하며, 여기서 제1 및 제2 챔버는 막(들)에 의해 분리된다. 표적 분석물을 특성화하는 데 사용되는 경우, 시스템은 표적 분석물을 추가로 포함할 수 있으며, 여기서 표적 분석물은 연속 채널 내에 일시적으로 위치하고 여기서 표적 분석물의 한쪽 단부는 제1 챔버에 위치하고 표적 분석물의 한쪽 단부는 제2 챔버에 위치한다. 표적 분석물은 바람직하게는 표적 폴리펩티드 또는 표적 폴리뉴클레오티드이다.
- [0886] 하나의 구현예에서, 시스템은 기공(들)과 접촉하는 전기 전도성 용액, 막(들)을 가로질러 전위를 제공하는 전극, 및 기공(들)을 통해 전류를 측정하기 위한 측정 시스템을 추가로 포함한다. 하나의 구현예에서, 막 및 기공을 가로질러 인가된 전압은 +5 V 내지 -5 V, 예컨대 -600 mV 내지 +600mV 또는 -400 mV 내지 +400 mV이다. 사용된 전압은 바람직하게는 100 mV 내지 240 mV 범위, 보다 바람직하게는 120 mV 내지 220 mV 범위이다. 증가된 인가된 전위를 사용함으로써 기공에 의해 상이한 아미노산 또는 뉴클레오티드 사이의 구별을 증가시키는 것이 가능하다. 임의의 적합한 전기 전도성 용액이 사용될 수 있다. 예를 들어, 용액은 전하 운반자, 예컨대 금속 염, 예를 들어 알칼리 금속 염, 할라이드 염, 예를 들어 클로라이드 염, 예컨대 알칼리 금속 클로라이드 염을 포함할 수 있다. 전하 운반자는 이온성 액체 또는 유기 염, 예를 들어 테트라메틸 암모늄 클로라이드, 트리메틸 페닐 암모늄 클로라이드, 페닐트리메틸 암모늄 클로라이드, 또는 1-에틸-3-메틸 이미다졸륨 클로라이드를 포함할 수 있다. 예시적인 시스템에서, 염은 챔버 내에 수용액으로 존재한다. 칼륨 클로라이드(KCl), 나트륨 클로라이드(NaCl), 세슘 클로라이드(CsCl) 또는 칼륨 페로시아나이드 및 칼륨 페리시아나이드의 혼합물이 전형적으로 사용된다. KCl, NaCl 및 칼륨 페로시아나이드 및 칼륨 페리시아나이드의 혼합물이 바람직하다. 전하 운반자는 막을 가로질러 비대칭일 수 있다. 예를 들면, 전하 운반자의 유형 및/또는 농도는 예를 들어, 각 챔버에서 막의 각 측면에서 상이할 수 있다.
- [0887] 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 바람직하게는 150 mM 내지 1 M이다. 방법은 바람직하게는 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 사용하여 수행된다. 높은 염 농도는 높은 신호 대 잡음비를 제공하고 아미노산 또는 뉴클레오티드의 존재를 나타내는 전류가 정상적인 전류 변동의 배경에 대해 식별되게 한다.
- [0888] 완충액은 전기 전도성 용액에 존재할 수 있다. 전형적으로, 완충액은 포스페이트 완충액이다. 다른 적합한 완충액은 HEPES 및 Tris-HCl 완충액이다. 전기 전도성 용액의 pH는 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 8.5일 수 있다. 사용된 pH는 바람직하게는 약 7.5이다.
- [0889] 시스템은 장치에 포함될 수 있다. 장치는 분석물 분석을 위한 임의의 통상적인 장치, 예컨대 어레이 또는 칩일 수 있다. 장치는 바람직하게는 개시된 방법을 수행하도록 설정된다. 예를 들어, 장치는 수용액을 포함하는 챔버 및 챔버를 2개 섹션으로 분리하는 장벽을 포함할 수 있다. 장벽은 전형적으로 기공(들)을 함유하는 막(들)이 형성되는 구멍을 갖는다. 대안적으로, 장벽은 기공이 존재하는 막을 형성한다.
- [0890] 장치는 또한 막 및 기공을 가로질러 전위를 인가하고 전기 신호를 측정할 수 있는 전기 회로를 포함할 수 있다.

- [0891] 장치는 WO 2008/102120, WO 2009/077734, WO 2010/122293, WO 2011/067559, 또는 WO 00/28312(모두 그 전문이 본원에 참조로 포함됨)에 개시된 것들 중 임의의 것일 수 있다.
- [0892] 막
- [0893] 임의의 적합한 막이 시스템에서 사용될 수 있다. 막은 바람직하게는 양친매성 층이다. 양친매성 층은 친수성 및 친유성 특성을 둘 다 갖는 인지질과 같은 양친매성 분자로부터 형성된 층이다. 양친매성 분자는 합성적이거나 자연 발생할 수 있다. 비자연 발생 양친매성 물질 및 단층을 형성하는 양친매성 물질은 당업계에 알려져 있으며 예를 들어, 블록 공중합체를 포함한다(Gonzalez-Perez 등, Langmuir, 2009, 25, 10447-10450). 블록 공중합체는 2개 이상의 단량체 서브유닛이 함께 중합되어 단일 중합체 쇄를 생성하는 중합체 물질이다. 블록 공중합체는 전형적으로 각 단량체 서브유닛에 의해 기여된 특성을 갖는다. 그러나, 블록 공중합체는 개별 서브유닛으로부터 형성된 중합체를 보유하지 않는 고유한 특성을 가질 수 있다. 블록 공중합체는 단량체 서브유닛 중 하나가 소수성(즉, 친유성)이고, 다른 서브유닛(들이 수성 매체 중에서 친수성이도록 조작될 수 있다. 이 경우, 블록 공중합체는 양친매성 특성을 보유할 수 있고 생물학적 막을 모방하는 구조를 형성할 수 있다. 블록 공중합체 디블록(2개의 단량체 서브유닛으로 이루어짐)일 수 있지만 또한 2개 초과인 단량체 서브유닛으로부터 구성되어 암피필(amphipile)로서 거동하는 보다 복잡한 배열을 형성할 수 있다. 공중합체는 트리블록, 테트라블록 또는 펜타블록 공중합체일 수 있다. 막은 바람직하게는 트리블록 공중합체 막이다.
- [0894] 막은 가장 바람직하게는 국제 출원 번호 W02014/064443 또는 W02014/064444에 개시된 막 중 하나이다.
- [0895] 양친매성 분자는 폴리뉴클레오티드의 결합을 용이하게 하기 위해 화학적으로 변형되거나 작용화될 수 있다. 양친매성 층은 단층 또는 이중층일 수 있다. 양친매성 층은 전형적으로 평면이다. 양친매성 층은 곡선화될 수 있다. 양친매성 층은 지지될 수 있다.
- [0896] 양친매성 막은 전형적으로 자연적으로 이동성이며, 본질적으로 대략 10^{-8} cm s⁻¹의 지질 확산 속도로 2차원 유체로서 작용한다. 이는 기공 및 결합된 폴리뉴클레오티드가 전형적으로 양친매성 막 내에서 이동할 수 있음을 의미한다.
- [0897] 막은 지질 이중층일 수 있다. 지질 이중층은 세포 막의 모델이며 다양한 실험 연구를 위한 훌륭한 플랫폼으로서 역할을 한다. 예를 들어, 지질 이중층은 단일-채널 기록에 의해 막 단백질의 시험관 내 조사에 사용될 수 있다. 대안적으로, 지질 이중층은 다양한 물질의 존재를 검출하기 위한 바이오센서로서 사용될 수 있다. 지질 이중층은 임의의 지질 이중층일 수 있다. 적합한 지질 이중층은 평면 지질 이중층, 지지된 이중층, 또는 리포솜을 포함하나 이에 제한되지 않는다. 지질 이중층은 바람직하게는 평면 지질 이중층이다. 적합한 지질 이중층은 WO 2008/102121, WO 2009/077734, 및 WO 2006/100484(모두 그 전문이 본원에 참조로 포함됨)에 개시되어 있다.
- [0898] 또 다른 바람직한 구현예에서, 막은 고체 상태 층을 포함한다. 고체 상태 층은 마이크로전자 재료, Si₃N₄, Al₂O₃, 및 SiO와 같은 절연 재료, 폴리아미드와 같은 유기 및 무기 중합체, Teflon®과 같은 플라스틱 또는 2-구성요소 첨가-경화 실리콘 고무와 같은 엘라스토머, 및 유리를 포함하나 이에 제한되지 않는 유기 및 무기 재료들로부터 형성될 수 있다. 고체 상태 층은 그래핀으로부터 형성될 수 있다. 적합한 그래핀 층은 WO 2009/035647(이의 전문이 본원에 참조로 포함됨)에 개시되어 있다. 막이 고체 상태 층을 포함하는 경우, 기공은 전형적으로 고체 상태 층 내에 함유된 양친매성 막 또는 층에, 예를 들면 고체 상태 층 내에서 구멍, 벽, 갭, 채널, 트렌치 또는 슬릿 내에 존재한다. 당업자는 적합한 고체 상태/양친매성 하이브리드 시스템을 제조할 수 있다. 적합한 시스템은 WO 2009/020682 및 WO 2012/005857(둘 다 그 전문이 본원에 참조로 포함됨)에 개시되어 있다. 위에 논의된 양친매성 막 또는 층 중 임의의 것이 사용될 수 있다.
- [0899] 방법은 전형적으로 (i) 기공을 포함하는 인공 양친매성 층, (ii) 기공을 포함하는 단리된 자연 발생 지질 이중층, 또는 (iii) 그 안에 삽입된 기공을 갖는 세포를 사용하여 수행된다. 방법은 전형적으로 인공 트리블록 공중합체 층과 같은 인공 양친매성 층을 사용하여 수행된다. 층은 다른 막관통 및/또는 막내 단백질 뿐만 아니라 기공 외의 다른 분자를 포함할 수 있다. 적합한 장치 및 조건은 아래에 논의되어 있다. 본 발명의 방법은 전형적으로 시험관 내에서 수행된다.
- [0900] 기공 및 헬리카제의 바람직한 조합
- [0901] 위의 여러 구현예에서, 본 발명의 변형된 헬리카제는 기공 또는 막관통 기공과 함께 사용될 수 있다. 본 발명의 여러 구현예에서, 본 발명의 단리된 기공 또는 본 발명의 단리된 기공 복합체는 본 발명의 변형된 헬리카제와

조합하여 사용될 수 있다. 다음 조합이 본 발명에서 바람직하다.

- [0902] 본 발명의 변형된 헬리카제는 W02016/034591, W02017/149316, W02017/149317 및, W02017/149318, W02018/211241, 및 W02019/002893(모두 그 전문이 본원에 참조로 포함됨)에 기재된 임의의 기공과 조합하여 사용될 수 있다. 본 발명의 변형된 헬리카제는 바람직하게는 실시예 2에서와 같이 W02016/034591(본원에 참조로 포함됨)에 기재된 바와 같은 CsgG 기공 또는 W02019/002893(참조로 본원에 포함됨)에 기재된 바와 같은 CsgG:CsgF 기공 복합체와 함께 사용된다.
- [0903] 본 발명의 단리된 기공 또는 본 발명의 단리된 기공 복합체는 W02015/055981, W02015/166276 및 W02016/055777(모두 그 전문이 본원에 참조로 포함됨)에 개시된 임의의 헬리카제와 조합하여 사용될 수 있다.
- [0904] 본 발명의 헬리카제는 바람직하게는 본 발명의 CsgG 기공 또는 CsgG:CsgF 기공 복합체와 함께 사용되고 실시예 4에서와 같이 그 반대의 경우도 마찬가지이다. 본 발명의 CsgG 기공 또는 CsgG:CsgF 기공 복합체는 바람직하게는 Q100A 및/또는 N102A 또는 N102S를 포함하고 헬리카제는 바람직하게는 실시예 4에서의 임의의 치환, 바람직하게는 Y350I 및/또는 K358I를 포함한다.
- [0905] 특정 구현예, 특이적 구성 뿐만 아니라 재료 및/또는 분자가 본 발명에 따라 조작된 세포 및 방법에 대해 본원에서 논의되었지만, 본 발명의 범위 및 취지를 벗어나지 않고 형태 및 세부사항의 다양한 변화 또는 변형이 이루어질 수 있다는 것이 이해되어야 한다. 다음 예는 특정 구현예를 더 잘 예시하기 위해 제공되며, 이들은 출원을 제한하는 것으로 간주되어서는 안 된다. 출원은 청구범위에 의해서만 제한된다.
- [0906] 서열
- [0907] 서열번호:1은 신호 서열을 포함한 균주 K12에서 야생형 이. 콜라이 CsgG의 폴리뉴클레오티드 서열(유전자 ID: 945619)을 보여준다.
- [0908] 서열번호:2는 신호 서열을 포함한 야생형 이. 콜라이 CsgG의 아미노산 서열(Uniprot 수탁 번호 P0AEA2)을 보여준다.
- [0909] 서열번호:3은 성숙 단백질로서 야생형 이. 콜라이 CsgG의 아미노산 서열(Uniprot 수탁 번호 P0AEA2)을 보여준다.
- [0910] 서열번호:4는 신호 서열을 포함한 균주 K12에서 야생형 이. 콜라이 CsgF의 폴리뉴클레오티드 서열(유전자 ID: 945622)을 보여준다.
- [0911] 서열번호:5는 신호 서열을 포함한 야생형 이. 콜라이 CsgF의 아미노산 서열(Uniprot 수탁 번호 P0AE98)을 보여준다.
- [0912] 서열번호:6은 성숙 단백질로서 야생형 이. 콜라이 CsgF의 아미노산 서열(Uniprot 수탁 번호 P0AE98)을 보여준다.
- [0913] 서열번호:7은 아미노산 1 내지 27 및 C-말단 6 His 태그를 암호화하는 야생형 이. 콜라이 CsgF의 단편의 폴리뉴클레오티드 서열을 보여준다.
- [0914] 서열번호:8은 아미노산 1 내지 27 및 C-말단 6 His 태그를 포함하는 야생형 이. 콜라이 CsgF의 단편의 아미노산 서열을 보여준다.
- [0915] 서열번호:9는 아미노산 1 내지 38 및 C-말단 6 His 태그를 암호화하는 야생형 이. 콜라이 CsgF의 단편의 폴리뉴클레오티드 서열을 보여준다.
- [0916] 서열번호:10은 아미노산 1 내지 38 및 C-말단 6 His 태그를 포함하는 야생형 이. 콜라이 CsgF의 단편의 아미노산 서열을 보여준다.
- [0917] 서열번호:11은 아미노산 1 내지 48 및 C-말단 6 His 태그를 암호화하는 야생형 이. 콜라이 CsgF의 단편의 폴리뉴클레오티드 서열을 보여준다.
- [0918] 서열번호:12는 아미노산 1 내지 48 및 C-말단 6 His 태그를 포함하는 야생형 이. 콜라이 CsgF의 단편의 아미노산 서열을 보여준다.
- [0919] 서열번호:13은 아미노산 1 내지 64 및 C-말단 6 His 태그를 암호화하는 야생형 이. 콜라이 CsgF의 단편의 폴리뉴클레오티드 서열을 보여준다.

- [0920] 서열번호:14는 아미노산 1 내지 64 및 C-말단 6 His 태그를 포함하는 야생형 이. 콜라이 CsgF의 단편의 아미노산 서열을 보여준다.
- [0921] 서열번호:15는 이. 콜라이 CsgF의 잔기 20 내지 53에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0922] 서열번호:16은 C-말단에 KD를 포함하는 이. 콜라이 CsgF의 잔기 20 내지 42에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0923] 서열번호:17은 CsgF 상동체 Q88H88의 잔기 23 내지 55에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0924] 서열번호:18은 CsgF 상동체 A0A143HJA0의 잔기 25 내지 57에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0925] 서열번호:19는 CsgF 상동체 Q5E245의 잔기 21 내지 53에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0926] 서열번호:20은 CsgF 상동체 Q084E5의 잔기 19 내지 51에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0927] 서열번호:21은 CsgF 상동체 F0LZU2의 잔기 15 내지 47에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0928] 서열번호:22는 CsgF 상동체 A0A136HQR0의 잔기 26 내지 58에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0929] 서열번호:23은 CsgF 상동체 A0A0W1SRL3의 잔기 21 내지 53에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0930] 서열번호:24는 CsgF 상동체 B0UH01의 잔기 26 내지 59에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0931] 서열번호:25는 CsgF 상동체 Q6NAU5의 잔기 22 내지 53에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0932] 서열번호:26은 CsgF 상동체 G8PUY5의 잔기 7 내지 38에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0933] 서열번호:27은 CsgF 상동체 A0A0S2ETP7의 잔기 25 내지 57에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0934] 서열번호:28은 CsgF 상동체 E3I1Z1의 잔기 19 내지 51에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0935] 서열번호:29는 CsgF 상동체 F3Z094의 잔기 24 내지 55에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0936] 서열번호:30은 CsgF 상동체 A0A176T7M2의 잔기 21 내지 53에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0937] 서열번호:31은 CsgF 상동체 D2QPP8의 잔기 14 내지 45에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0938] 서열번호:32는 CsgF 상동체 N2IYT1의 잔기 28 내지 58에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0939] 서열번호:33은 CsgF 상동체 W7QHV5의 잔기 26 내지 58에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0940] 서열번호:34는 CsgF 상동체 D4ZLW2의 잔기 23 내지 55에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0941] 서열번호:35는 CsgF 상동체 D2QT92의 잔기 21 내지 53에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0942] 서열번호:36은 CsgF 상동체 A0A167UJA2의 잔기 20 내지 51에 상응하는 펩티드의 아미노산 서열을 보여준다
- [0943] 서열번호:37은 아미노산 20 내지 27을 포함하는 야생형 이. 콜라이 CsgF의 단편의 아미노산 서열을 보여준다.
- [0944] 서열번호:38은 아미노산 20 내지 38을 포함하는 야생형 이. 콜라이 CsgF의 단편의 아미노산 서열을 보여준다.
- [0945] 서열번호:39:는 아미노산 20 내지 48을 포함하는 야생형 이. 콜라이 CsgF의 단편의 아미노산 서열을 보여준다.
- [0946] 서열번호:40은 아미노산 20 내지 64를 포함하는 야생형 이. 콜라이 CsgF의 단편의 아미노산 서열을 보여준다.
- [0947] 서열번호:41은 프라이머 CsgF_d27_단부의 뉴클레오티드 서열을 보여준다
- [0948] 서열번호:42는 프라이머 CsgF_d38_단부의 뉴클레오티드 서열을 보여준다
- [0949] 서열번호:43은 프라이머 CsgF_d48_단부의 뉴클레오티드 서열을 보여준다
- [0950] 서열번호:44는 프라이머 CsgF_d64_단부의 뉴클레오티드 서열을 보여준다
- [0951] 서열번호:45는 프라이머 pNa62_CsgF_his태그_Fw의 뉴클레오티드 서열을 보여준다
- [0952] 서열번호:46은 프라이머 CsgF-His_pET22b_FW의 뉴클레오티드 서열을 보여준다
- [0953] 서열번호:47은 프라이머 CsgF-His_pET22b_Rev의 뉴클레오티드 서열을 보여준다

- [0954] 서열번호:48은 프라이머 csgEFG_pDONR221_FW의 뉴클레오티드 서열을 보여준다
- [0955] 서열번호:49는 프라이머 csgEFG_pDONR221_Rev의 뉴클레오티드 서열을 보여준다
- [0956] 서열번호:50은 프라이머 Mut_csgF_His_FW의 뉴클레오티드 서열을 보여준다
- [0957] 서열번호:51은 프라이머 Mut_csgF_His_Rev의 뉴클레오티드 서열을 보여준다
- [0958] 서열번호:52는 프라이머 DelCsgE_Rev의 뉴클레오티드 서열을 보여준다
- [0959] 서열번호:53은 프라이머 DelCsgE FW의 뉴클레오티드 서열을 보여준다
- [0960] 서열번호: 54는 성숙 이. 콜라이 CsgF의 잔기 1 내지 30의 아미노산 서열을 보여준다
- [0961] 서열번호: 55는 성숙 이. 콜라이 CsgF의 잔기 1 내지 35의 아미노산 서열을 보여준다
- [0962] 서열번호: 56은 신호 서열을 갖는 돌연변이된 (T4C/N17S) CsgF 서열, 및 성숙 단백질 서열의 잔기 35와 36 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS)의 아미노산 서열을 보여준다.
- [0963] 서열번호: 57은 신호 서열을 갖는 돌연변이된 (N17S-Del) CsgF 서열, 및 성숙 단백질 서열의 잔기 35와 36 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS)의 아미노산 서열을 보여준다.
- [0964] 서열번호: 58은 신호 서열을 갖는 돌연변이된 (G1C/N17S) CsgF 서열, 및 성숙 단백질 서열의 잔기 35와 36 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS)의 아미노산 서열을 보여준다.
- [0965] 서열번호: 59는 신호 서열을 갖는 돌연변이된 (G1C) CsgF 서열, 및 성숙 단백질 서열의 잔기 35와 36 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS)의 아미노산 서열을 보여준다.
- [0966] 서열번호: 60은 신호 서열을 갖는 CsgF 서열, 성숙 단백질 서열의 잔기 45와 46 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS), 및 C-말단에 있는 His₁₀ 태그의 아미노산 서열을 보여준다.
- [0967] 서열번호: 61은 신호 서열을 갖는 CsgF 서열, 성숙 단백질 서열의 잔기 35와 36 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS), 및 C-말단에 있는 His₁₀ 태그의 아미노산 서열을 보여준다.
- [0968] 서열번호: 62는 신호 서열을 갖는 CsgF 서열, 성숙 단백질 서열의 잔기 30과 31 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS), 및 C-말단에 있는 His₁₀ 태그의 아미노산 서열을 보여준다.
- [0969] 서열번호: 63은 신호 서열을 갖는 CsgF 서열, 성숙 단백질 서열의 잔기 45와 51 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS), 및 C-말단에 있는 His₁₀ 태그의 아미노산 서열을 보여준다.
- [0970] 서열번호: 64는 신호 서열을 갖는 CsgF 서열, 성숙 단백질 서열의 잔기 30과 37 사이에 삽입된 TEV 프로테아제 절단 부위(ENLYFQS), 및 C-말단에 있는 His₁₀ 태그의 아미노산 서열을 보여준다.
- [0971] 서열번호: 65는 신호 서열을 갖는 CsgF 서열, 성숙 단백질 서열의 잔기 34와 36 사이에 삽입된 HCV C3 프로테아제 절단 부위(LEVLFGQP), 및 C-말단에 있는 His₁₀ 태그의 아미노산 서열을 보여준다.
- [0972] 서열번호: 66은 신호 서열을 갖는 CsgF 서열, 성숙 단백질 서열의 잔기 42와 43 사이에 삽입된 HCV C3 프로테아제 절단 부위(LEVLFGQP), 및 C-말단에 있는 His₁₀ 태그의 아미노산 서열을 보여준다.
- [0973] 서열번호: 67은 신호 서열을 갖는 CsgF 서열, 성숙 단백질 서열의 잔기 38과 47 사이에 삽입된 HCV C3 프로테아제 절단 부위(LEVLFGQP), 및 C-말단에 있는 His₁₀ 태그의 아미노산 서열을 보여준다.
- [0974] 서열번호: 68은 가설 단백질 CKO_02032 [시트로박터 코세리(Citrobacter koseri) ATCC BAA-895]의 YP_001453594.1: 1-248의 아미노산 서열을 보여주며, 이는 서열번호: 3과 99% 동일하다.
- [0975] 서열번호: 69는 컬리(curli) 생산 조립/수송 구성요소 CsgG, 부분적 [살모넬라 엔테리카(Salmonella enterica)]의 WP_001787128.1: 16-238의 아미노산 서열을 보여주며, 이는 서열번호: 3에 대해 98%이다.
- [0976] 서열번호: 70은 컬리 생산 조립/수송 단백질 CsgG [시트로박터 아말로나티쿠스(Citrobacter amalonaticus)]의 KEY44978.1: 16-277의 아미노산 서열을 보여주며, 이는 서열번호: 3과 98% 동일하다.
- [0977] 서열번호: 71은 컬리 생산 조립/수송 구성요소 [시트로박터 로덴튬(Citrobacter rodentium) ICC168]의

YP_003364699.1: 16-277의 아미노산 서열을 보여주며, 이는 서열번호: 3과 97% 동일하다.

- [0978] 서열번호: 72는 칼리 생산 조립/수송 구성요소 CsgG [*엔테로박터 아스부리아*(*Enterobacter asburiae*) LF7a]의 YP_004828099.1: 16-277의 아미노산 서열을 보여주며, 이는 서열번호: 3과 94% 동일하다.
- [0979] 서열번호: 73은 수송체 [요케넬라 레겐스부르케이(*Yokenella regensburgei*)]의 WP_006819418.1: 19-280의 아미노산 서열을 보여주며, 이는 서열번호: 3과 91% 동일하다.
- [0980] 서열번호: 74는 칼리 생산 조립/수송 단백질 CsgG [*크로노박터 풀베리스*(*Cronobacter pulveris*)]의 WP_024556654.1: 16-277의 아미노산 서열을 보여주며, 이는 서열번호: 3과 89% 동일하다.
- [0981] 서열번호: 75는 칼리 생산 조립/수송 단백질 CsgG [*라호넬라 아쿠아틸리스*(*Rahnella aquatilis*) HX2]의 YP_005400916.1: 16-277의 아미노산 서열을 보여주며, 이는 서열번호: 3과 84% 동일하다.
- [0982] 서열번호: 76은 CsgG 계열 칼리 생산 조립/수송 구성요소 [*클루이베라 아스코르바타*(*Kluyvera ascorbata*) ATCC 33433]의 KFC99297.1: 20-278의 아미노산 서열을 보여주며, 이는 서열번호: 3과 82% 동일하다.
- [0983] 서열번호: 77은 CsgG 계열 칼리 생산 조립/수송 구성요소 [*하프니아 알베이*(*Hafnia alvei*) ATCC 13337]의 KFC86716.1: 16-274의 아미노산 서열을 보여주며, 이는 서열번호: 3과 81% 동일하다.
- [0984] 서열번호: 78은 칼리 중합체의 형성에 수반되는 비특성화된 단백질 [도 57의 엔테로박테리아세에 박테리움(*Enterobacteriaceae bacterium*) 균주 FGI 57]의 YP_007340845.1: 16-270의 아미노산 서열을 보여주며, 이는 서열번호: 3과 76% 동일하다.
- [0985] 서열번호: 79는 칼리 생산 조립/수송 단백질 CsgG [*플레이오모나스 시겔로이데스*(*Plesiomonas shigelloides*)]의 WP_010861740.1: 17-274의 아미노산 서열을 보여주며, 이는 서열번호: 3과 70% 동일하다.
- [0986] 서열번호: 80은 칼리 생산 조립/수송 외부 막 지단백질 구성요소 CsgG [*비브리오 피스케리*(*Vibrio fischeri*) ES114]의 YP_205788.1 : 23-270의 아미노산 서열을 보여주며, 이는 서열번호: 3과 60% 동일하다.
- [0987] 서열번호: 81은 칼리 생산 조립 단백질 CsgG [*알리이비브리오 로케이*(*Aliivibrio logei*)]의 WP_017023479.1: 23-270의 아미노산 서열을 보여주며, 이는 서열번호: 3과 59% 동일하다.
- [0988] 서열번호: 82는 칼리 생산 조립/수송 구성요소 CsgG [*포토박테리움 종*(*Photobacterium sp.*) AK15]의 WP_007470398.1: 22-275의 아미노산 서열을 보여주며, 이는 서열번호: 3과 57% 동일하다.
- [0989] 서열번호: 83은 칼리 생산 조립 단백질 CsgG [*아에로모나스 베로니이*(*Aeromonas veronii*)]의 WP_021231638.1: 17-277의 아미노산 서열을 보여주며, 이는 서열번호: 3과 56% 동일하다.
- [0990] 서열번호: 84는 칼리 생산 조립/수송 단백질 CsgG [*쉬와넬라 종*(*Shewanella sp.*) ECSMB14101]의 WP_033538267.1: 27-265의 아미노산 서열을 보여주며, 이는 서열번호: 3과 56% 동일하다.
- [0991] 서열번호: 85는 칼리 생산 조립 단백질 CsgG [*슈도모나스 푸티다*(*Pseudomonas putida*)]의 WP_003247972.1: 30-262의 아미노산 서열을 보여주며, 이는 서열번호: 3과 54% 동일하다.
- [0992] 서열번호: 86은 칼리 생산 조립/수송 구성요소 CsgG [*쉬와넬라 비올라세아*(*Shewanella violacea*) DSS12]의 YP_003557438.1: 1-234의 아미노산 서열을 보여주며, 이는 서열번호: 3과 53% 동일하다.
- [0993] 서열번호: 87은 칼리 생산 조립/수송 단백질 CsgG [*마리노박테리움 자나쉬이*(*Marinobacterium jannaschii*)]의 WP_027859066.1: 36-280의 아미노산 서열을 보여주며, 이는 서열번호: 3과 53% 동일하다.
- [0994] 서열번호: 88은 칼리 생산 조립/수송 구성요소 CsgG [*크리세오박테리움 오라니멘세*(*Chryseobacterium oranimentense*) G311]의 CEJ70222.1: 29-262의 아미노산 서열을 보여주며, 이는 서열번호: 3과 50% 동일하다.
- [0995] 서열번호: 89 내지 104는 표 4의 서열을 보여준다.

- [0996] 서열번호: 1(>P0AEA2; 이. 콜라이 K12의 WT CsgG에 대한 코딩 서열)
- ATGCAGCGCTTATTTCTTTGGTTGCCGTCATGTTACTGAGCGGATGCTTAACCGCCCCGCCTAAAG
AAGCCGCCAGACCGACATTAATGCCTCGTGCTCAGAGCTACAAAGATTTGACCCATCTGCCAGCGCC
GACGGGTAATACTTTGTTTCGGTATACAACATTCAGGACGAAACCGGGCAATTTAAACCTACCCG
GCAAGTAACTTCTCCACTGCTGTTCCGCAAAGCGCCACGGCAATGCTGGTCACGGCACTGAAAGATT
CTCGCTGGTTTATACCGCTGGAGCGCCAGGGCTTACAAAACCTGCTTAACGAGCGCAAGATTATTCG
TGCGGCACAAGAAAACGGCACGGTTGCCATTAATAACCGAATCCCGCTGCAATCTTTAACGGCGGCA
AATATCATGGTTGAAGGTTTCGATTATCGGTTATGAAAGCAACGTCAAATCTGGCGGGGTTGGGGCAA
GATATTTTGGCATCGGTGCCGACACGCAATACCAGCTCGATCAGATTGCCGTGAACCTGCGCGTCGT
CAATGTGAGTACCGGCGAGATCCTTTCTCGGTGAACACAGTAAGACGATACTTTCCTATGAAGTT
CAGGCCGGGGTTTTCCGCTTTATTGACTACCAGCGCTTGCTTGAAGGGGAAGTGGGTTACACCTCGA
ACGAACCTGTTATGCTGTGCCTGATGTCGGCTATCGAAACAGGGGTCATTTTCTGATTAATGATGG
TATCGACCGTGGTCTGTGGATTGTGAAAATAAAGCAGAACGGCAGAATGACATTCTGGTGAAATAC
CGCCATATGTCGGTTCACCGGAATCCTGA
- [0997]
- [0998] 서열번호: 2(>P0AEA2(1:277); 이. 콜라이 K12의 WT prepro CsgG)
- MQRLFLLVAVMLLSGCLTAPPKEAARPTLMPRAQSYKDLTHLPAPTGKIFVSVYNIQDETGFQKPYPASNF
STAVPQSATAMLVTKDSRWFIPLERQGLQNLNLERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSI
IGYESNVKSGGVGARYFGIGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQ
RLLEGEVGYTSNEPVMLCLMSAIETGVIFLINDGIDRGLWDLQNKARQNDILVKYRHMSVPPES
- [0999]
- [1000] 서열번호: 3(>P0AEA2(16:277); 이. 콜라이 K12의 성숙 CsgG)
- CLTAPPKEAARPTLMPRAQSYKDLTHLPAPTGKIFVSVYNIQDETGFQKPYPASNFSTAVPQSATAMLV
ALKDSRWFIPLERQGLQNLNLERKIIRAAQENGTVAINNRIPLQSLTAANIMVEGSIIGYESNVKSGGVG
- [1001]
- ARYFGIGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEVGYTSNEP
VMLCLMSAIETGVIFLINDGIDRGLWDLQNKARQNDILVKYRHMSVPPES
- [1002]
- [1003] 서열번호: 4(>P0AE98; 이. 콜라이 K12의 WT CsgF에 대한 코딩 서열)
- ATGCGTGTCAAACATGCAGTAGTTCTACTCATGCTTATTTCCGCCATTAAGTTGGGCTGGAACCATGAC
TTTCCAGTTCCTGAATCCAACTTTGGTGGTAACCCAAATAATGGCGCTTTTTTATAAATAGCGCTC
AGGCCCAAACCTTTATAAAGATCCGAGCTATAACGATGACTTTGGTATTGAAACACCCTCAGCGTTA
GATAACTTTACTAGGCCATCCAGTCACAAATTTAGGTGGGCTACTGTCGAATATTAATACCGGTAA
ACCGGGCCGCATGGTGACCAACGATTATATTGTCGATATTGCCAACCGCGATGGTCAATTGCAGTTG
AACGTGACAGATCGTAAACCGGACAAACCTCGACCATCCAGTTTCGGGTTTACAAAATAACTCAA
CCGATTTT
- [1004]
- [1005] 서열번호: 5(>P0AE98(1:138); 이. 콜라이 K12의 WT pre CsgF)
- MRVKHAVVLLMLISPLSWAGTMTFQFRNPNGGNPNNGAFLNLSAQNSYKDPSYNDDFGIETPSAL
DNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTD
F
- [1006]
- [1007] 서열번호: 6(>P0AE98(20:138); 이. 콜라이 K12의 WT 성숙 CsgF)
- GTMTFQFRNPNGGNPNNGAFLNLSAQNSYKDPSYNDDFGIETPSALDNFTQAIQSQILGGLLSNIN
TGKPGRMVTNDYIVDIANRDGQLQLNVTDRKTGQTSTIQVSGLQNNSTDF
- [1008]
- [1009] 서열번호: 7(>P0AE98; CsgF 1:27_6His에 대한 코딩 서열)
- ATGCGTGTCAAACATGCAGTAGTTCTACTCATGCTTATTTCCGCCATTAAGTTGGGCTGGAACCATGAC
TTTCCAGTTCCTGCATCACCATCACCATCACTAAGCCC
- [1010]

- [1011] 서열번호:8(>P0AE98(1:28); CsgF 20:27_6His의 전단백질)
- [1012] MRVKHAVVLLMLISPLSWA GTMTFQFR HHHHHH
- [1013] 서열번호:9(>P0AE98; CsgF 1:38_6His에 대한 코딩 서열)
- [1014] ATGCGTGTCAAACATGCAGTAGTTCTACTCATGCTTATTTGCGCCATTAAGTTGGGCTGGAACCATGAC
TTTCCAGTTCCGTAATCCAAACTTTGGTGGTAACCCAAATAATGGCCATCACCATCACCATCACTAAG
CCC
- [1015] 서열번호:10(>P0AE98(1:39); CsgF 20:38_6His의 전단백질)
- [1016] MRVKHAVVLLMLISPLSWAGTMTFQFRNPFGGNPNNG HHHHHH
- [1017] 서열번호:11(>P0AE98; CsgF 1:48_6His에 대한 코딩 서열)
- [1018] ATGCGTGTCAAACATGCAGTAGTTCTACTCATGCTTATTTGCGCCATTAAGTTGGGCTGGAACCATGAC
TTTCCAGTTCCGTAATCCAAACTTTGGTGGTAACCCAAATAATGGCGCTTTTTATTAAATAGCGCTC
AGGCCCAACATCACCATCACCATCACTAAGCCC
- [1019] 서열번호:12(>P0AE98(1:49); CsgF 20:48_6His의 전단백질)
- [1020] MRVKHAVVLLMLISPLSWAGTMTFQFRNPFGGNPNNGAFLNLSAQQA HHHHHH
- [1021] 서열번호:13(>P0AE98; CsgF 1:64_6His에 대한 코딩 서열)
- [1022] ATGCGTGTCAAACATGCAGTAGTTCTACTCATGCTTATTTGCGCCATTAAGTTGGGCTGGAACCATGAC
TTTCCAGTTCCGTAATCCAAACTTTGGTGGTAACCCAAATAATGGCGCTTTTTATTAAATAGCGCTC
AGGCCCAAACTCTTATAAAGATCCGAGCTATAACGATGACTTTGGTATTGAAACA
CATCACCATCACCATCACTAAGCCC
- [1023] 서열번호:14(>P0AE98(1:65); CsgF 20:64_6His의 전단백질)
- [1024] MRVKHAVVLLMLISPLSWAGTMTFQFRNPFGGNPNNGAFLNLSAQQAQNSYKDPYNDDFGIETHHHH
HH
- [1025] 서열번호:15(>P0AE98(20:53); CsgF 20:53의 성숙 펩티드)
- [1026] GTMTFQFRNPFGGNPNNGAFLNLSAQQAQNSYKD
- [1027] 서열번호:16(>P0AE98(20:42); CsgF 20:42+KD의 성숙 펩티드)
- [1028] GTMTFQFRNPFGGNPNNGAFLNKD
- [1029] 서열번호:17(>Q88H88_PSEPK(23:55))
- [1030] TELVYTPVNPAFGGNPLNGTWLLNNAQAQNDY
- [1031] 서열번호:18(>A0A143HJAO_9GAMM(25:57))
- [1032] TELIYEPVNPFGGNPLNGSYLLNNAQAQDRH
- [1033] 서열번호:19(>Q5E245_VIBF1(21:53))
- [1034] SELVYTPVNPFGGNPLNTSHLFGGANAINDY

- [1035] 서열번호:20(>Q084E5_SHEFN(19:51))
TQLVYTPVNPFAFGGSYLNGSYLLANASAEQNEH
- [1036]
- [1037] 서열번호:21(>F0LZU2_VIBFN(15:47))
SSLVYEPVNPFTFGGNPLNTTHLFSRAEAINDY
- [1038]
- [1039] 서열번호:22(>A0A136HQR0_9ALTE(26:58))
TELVYEPINPSFGGNPLNGSFLLSKANSQNAH
- [1040]
- [1041] 서열번호:23(>A0A0W1SRL3_9GAMM(21:53))
TEIVYQPINPSFGGNPMNGSFLLQKAQSQNAH
- [1042]
- [1043] 서열번호:24(>B0UH01_METS4(26:59))
SSLVYQPVNPFAFGGPQLNGSWLQAEANAQNIPO
- [1044]
- [1045] 서열번호:25(>Q6NAU5_RHOPA(22:53))
GSLVYTPTNPAFGGSPLNGSWQMQQATAGNH
- [1046]
- [1047] 서열번호:26(>G8PUY5_PSEUV(7:38))
QQLIYQPTNPSFGGYAANTTHLFATANAQKTA
- [1048]
- [1049] 서열번호:27(>A0A0S2ETP7_9RHIZ(25:57))
GDLVYTPVNPSPFGGSPLNSAHLISIAGAQKNA
- [1050]
- [1051] 서열번호:28(>E3I1Z1_RHOVT(19:51))
AELGYTPVNPSPFGGSPLNGSTLLSEASAQKPN
- [1052]
- [1053] 서열번호:29(>F3Z094_DESAF(24:55))
TELVFSFTNPSFGGDPMIGNFLLNKADSQKR
- [1054]
- [1055] 서열번호:30(>A0A176T7M2_9FLA0(21:53))
QQLVYKSINPFFGGGDSFAYQQLLASANAQND
- [1056]
- [1057] 서열번호:31(>D2QPP8_SPILD(14:45))
QALVYHPNNPAFGGNTFNYQWMLSSAQADQR
- [1058]
- [1059] 서열번호:32(>N2IYT1_9PSED(26:58))
TELVYTPKNPAFGGSPLNGSYLLGNAQAQNDY
- [1060]
- [1061] 서열번호:33(>W7QHV5_9GAMM(26:58))
GQLIYQPINPSFGGDPLLGNHLLNKAQAQDTK
- [1062]

- [1063] 서열번호:34(>D4ZLW2_SHEVD(23:55))
TQLIYTPVNPFGGSYLNGSYLLANASVQNDH
- [1064]
- [1065] 서열번호:35(>D2QT92_SPILD(21:53))
QAFVYHPNPNFGGNTFNYSWMLSSAQADRT
- [1066]
- [1067] 서열번호:36(>A0A167UJA2_9FLAO(20:51))
QGLIYKPKNPAFGGDTFNYQWLASSAESQNK
- [1068]
- [1069] 서열번호:37(>P0AE98(20:28); CsgF 20:27의 성숙 펩티드)
GTMTFQFR
- [1070]
- [1071] 서열번호:38(>P0AE98(20:39); CsgF 20:38의 성숙 펩티드)
GTMTFQFRNPFGGNPNNG
- [1072]
- [1073] 서열번호:39(>P0AE98(20:49); CsgF 20:48의 성숙 펩티드)
GTMTFQFRNPFGGNPNNGAFLLSAQAAQ
- [1074]
- [1075] 서열번호:40(>P0AE98(20:65); CsgF 20:64의 성숙 펩티드)
GTMTFQFRNPFGGNPNNGAFLLSAQAAQNSYKDPSYNDDFGIET
- [1076]
- [1077] 서열번호:41(CsgF_d27_단부)
ACGGAACTGGAAAGTCATGGTTCC
- [1078]
- [1079] 서열번호:42(CsgF_d38_단부)
GCCATTATTTGGGTTACCACCAAAGTTTGG
- [1080]
- [1081] 서열번호:43(CsgF_d48_단부)
TTGGGCCTGAGCGCTATTTAATAAAAAAGC
- [1082]
- [1083] 서열번호:44(CsgF_d64_단부)
TGTTTCAATACCAAAGTCATCGTTATAGCTCGG
- [1084]
- [1085] 서열번호:45(pNa62_CsgF_his태그_Fw)
CATCACCATCACCATCACTAAGCCC
- [1086]
- [1087] 서열번호:46(CsgF-His_pET22b_FW)
cccccatatgGGAACCATGACTTTCCAGTTCC
- [1088]
- [1089] 서열번호:47: (CsgF-His_pET22b_Rev)
ccccGAATTCCTAatggtgatggtgatggtGTAAAAATCGGTTGAGTTATTTTG
- [1090]

- [1091] 서열번호:48: (csgEFG_pDONR221_FW)
GGGGACAAGTTTGTACAAAAAGCAGGCTACCTCAGGCGATAAAGCCATGAAACGTTA
- [1092]
- [1093] 서열번호:49: (csgEFG_pDONR221_Rev)
GGGGACCACTTTGTACAAGAAAGCTGGGTGTTTAAACTCATTTTTCGAACTGCGGGTGGCTCCAAGC
GCTGG
- [1094]
- [1095] 서열번호:50: (Mut_csgF_His_FW)
CAAAATAACTCAACCGATTTTcatcaccatcaccatcacTAAGCCCCAGCTTCATAAGG
- [1096]
- [1097] 서열번호:51: (Mut_csgF_His_Rev)
CCTTATGAAGCTGGGCTTAgtgatggtgatggtgatgAAAATCGGTTGAGTTATTTTG
- [1098]
- [1099] 서열번호:52: (DelCsgE_Rev)
AGCCTGCTTTTTGTACAAAC
- [1100]
- [1101] 서열번호:53: (DelCsgE FW)
ATAAAAAATTGTTCTGGAGGCTGC
- [1102]
- [1103] 서열번호:54(>POAE98(20:50); CsgF 1:30의 성숙 펩티드)
GTMTFQFRNPFGGNPNNGAFLNLSAQAN
- [1104]
- [1105] 서열번호:55(>POAE98(20:54); CsgF 1:35의 성숙 펩티드)
GTMTFQFRNPFGGNPNNGAFLNLSAQANSYKDP
- [1106]
- [1107] 단백질로 만들어진 프로테아제 절단 부위가 있는 CsgF 서열의 예. 신호 펩티드는 **굵은 글씨**로 표시되고 TEV 프로테아제 절단 부위는 **굵은 글씨 및 밑줄**로 표시되고 HCV C3 프로테아제 절단 부위는 밑줄로 표시된다. StrepII는 C 말단에 있는 Strep 태그를 나타내고, H10은 C 말단에 있는 10x히스티딘 태그를 나타내고 **는 정지 코돈을 나타낸다.
- [1108] 서열번호:56 Pro-CsgF-Eco-(WT-T4C/N17S/P35-TEV-S36)-StrepII
MRVKHAVVLLMLISPLSWAGTMCFQFRNPFGGNPSNGAFLNLSAQANSYKDP**ENLYFOSSYND**
DFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTDRKTGQTSTIQV
SGLQNNSTDFSAWSHPQFEK**
- [1109]
- [1110] 서열번호:57 Pro-CsgF-Eco-(WT-N17S-Del(P35-[TEV]-S36)-StrepII
MRVKHAVVLLMLISPLSWAGTMTFQFRNPFGGNPSNGAFLNLSAQANSYKDP**ENLYFOSSYND**
DFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTDRKTGQTSTIQV
SGLQNNSTDFSAWSHPQFEK**
- [1111]
- [1112] 서열번호:58 Pro-CsgF-Eco-(WT-G1C/N17S/P35-[TEV]-S36)-StrepII
MRVKHAVVLLMLISPLSWACTMTFQFRNPFGGNPSNGAFLNLSAQANSYKDP**ENLYFOSSYND**
DFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTDRKTGQTSTIQV
SGLQNNSTDFSAWSHPQFEK**
- [1113]

- [1114] 서열번호:59 Pro-CsgF-Eco-(WT-G1C/P35-[TEV]-S36)-StrepII
- MRVKHAVVLLMLISPLSWACTMTFQFRNP**FGGNPNNGAFLNSAQANSYKDP**ENLYFQSSYND**
DFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~
SGLQNNSTDFSAWSHPQFEK**
- [1115]
- [1116] 서열번호:60 Pro-CsgF-Eco-(WT-T45-TEV-P46)-H10
- MRVKHAVVLLMLISPLSWAGTMTFQFRNP**FGGNPNNGAFLNSAQANSYKDPSYNDDFGIET**E**
NLYFQSPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~
VSGLQNNSTDFHHHHHHHHHH**
- [1117]
- [1118] 서열번호:61 Pro-CsgF-Eco-(WT-P35-TEV-S36)-H10
- MRVKHAVVLLMLISPLSWAGTMTFQFRNP**FGGNPNNGAFLNSAQANSYKDP**ENLYFQSSYND**
DFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~
SGLQNNSTDFHHHHHHHHHH**
- [1119]
- [1120] 서열번호:62 Pro-CsgF-Eco-(WT-N30-TEV-S31)-H10
- MRVKHAVVLLMLISPLSWAGTMTFQFRNP**FGGNPNNGAFLNSAQAN**ENLYFQSSYKDPSYND**
DFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~
SGLQNNSTDFHHHHHHHHHH**
- [1121]
- [1122] 서열번호:63 Pro-CsgF-Eco-(WT-T45-TEV-F51)-H10
- MRVKHAVVLLMLISPLSWAGTMTFQFRNP**FGGNPNNGAFLNSAQANSYKDPSYNDDFGIET**E**
NLYFQSFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~SGLQ
NNSTDFHHHHHHHHHH**
- [1123]
- [1124] 서열번호:64 Pro-CsgF-Eco-(WT-N30-TEV-Y37)-H10
- MRVKHAVVLLMLISPLSWAGTMTFQFRNP**FGGNPNNGAFLNSAQAN**ENLYFQSYNDDFGIET**
PSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~SGLQ
NSTDFHHHHHHHHHH**
- [1125]
- [1126] 서열번호:65 Pro-CsgF-Eco-(WT-D34-[C3]-S36)
- MRVKHAVVLLMLISPLSWACTMTFQFRNP**FGGNPNNGAFLNSAQANSYKD**LEVL**FQGPSYND
DFGIETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~
SGLQNNSTDFSAWSHPQFEK**
- [1127]
- [1128] 서열번호:66 Pro-CsgF-Eco-(WT-I42-[C3]-E43)
- MRVKHAVVLLMLISPLSWACTMTFQFRNP**FGGNPNNGAFLNSAQANSYKDPSYNDDFGI**LEVL**
FQGPETPSALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~
VSGLQNNSTDFSAWSHPQFEK**
- [1129]
- [1130] 서열번호:67 Pro-CsgF-Eco-(WT-N38-[C3]-S47)
- MRVKHAVVLLMLISPLSWACTMTFQFRNP**FGGNPNNGAFLNSAQANSYKDPSYN**LEVL**FQGPS
ALDNFTQAIQSQILGGLLSNINTGKPGRMVTNDYIVDIANRDGQLQLNVTD~~DRKTGQTSTIQV~~SGLQNN
TDFSAWSHPQFEK**
- [1131]

- [1132] 서열번호: 68
- MPRAQSYKDLTHLPMPTGKIFVSVYNIQDETQFKPYPASNFSTAVPQSATAMLVLTALKDSRWFIPLERQ
GLQNLLNERKIIIRAAQENGTVAINNRIPQLSLTAANIMVEGSIIGYESNVKSGGVGARYFGIGADTQYQL
DQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEIGYTSNEPVMCLCMSAIETGVI
FLINDGIDRGLWDLQNKAEQNDILVKYRHMSVPPES
- [1133]
- [1134] 서열번호: 69
- CLTAPPKQAAKPTLMPRAQSYKDLTHLPAPTGKIFVSVYNIQDETQFKPYPASNFSTAVPQSATAMLVTA
ALKDSRWFIPLERQGLQNLLNERKIIIRAAQENGTVAMNNRIPQLSLTAANIMVEGSIIGYESNVKSGGVG
ARYFGIGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEIGYTSNEP
VMLCMSAIETG
- [1135]
- [1136] 서열번호: 70
- CLTAPPKEAAKPTLMPRAQSYKDLTHLPIPTGKIFVSVYNIQDETQFKPYPASNFSTAVPQSATAMLVTA
LKDSRWFIPLERQGLQNLLNERKIIIRAAQENGTVAINNRIPQLSLTAANIMVEGSIIGYESNVKSGGVGA
RYFGIGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEIGYTSNEPV
MLCMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRHMSVPPES
- [1137]
- [1138] 서열번호: 71
- CLTPPKAAKPTLMPRAQSYKDLTHLPVPTGKIFVSVYNIQDETQFKPYPASNFSTAVPQSATAMLVTA
LKDSRWFIPLERQGLQNLLNERKIIIRAAQENGTVAINNRIPQLSLTAANIMVEGSIIGYESNVKSGGAGA
RYFGIGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEIGYTSNEPV
MLCMSAIETGVIFLINDGIDRGLWDLQNKADRQNDILVKYRQMSVPPES
- [1139]
- [1140] 서열번호: 72
- CLTAPPKEAAKPTLMPRAQSYRDLTHLPAPTGKIFVSVYNIQDETQFKPYPASNFSTAVPQSATAMLVTA
ALKDSHWFIPLERQGLQNLLNERKIIIRAAQENGTVANNRMPLQLSLAAANVMIEGSIIGYESNVKSGGV
GARYFGIGADTQYQLDQIAVNLRVVNVSTGEVLSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEIGYTSN
EPVMMCLCMSAIETGVIFLINDGIDRGLWDLQNKADQNPVLVKYRDMSVPPES
- [1141]
- [1142] 서열번호: 73
- CLTAPPKEAAKPTLMPRAQSYRDLTHLPLPSGKVFVSVYNIQDETQFKPYPASNFSTAVPQSATAMLVTA
ALKDSRWFIPLERQGLQNLLNERKIIIRAAQENGTVADNNRIPQLSLTAANVMIEGSIIGYESNVKSGGV
GARYFGIGADTQYQLDQIAVNLRVVNVSTGEVLSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEIGYTSN
EPVMCLCMSAIETGVIIYINDGIERGLWDLQKADVDNPILARYRNMSAPPES
- [1143]
- [1144] 서열번호: 74
- CLTAPPKEAAKPTLMPRAQSYRDLTNLPDPKGKLFVSVYNIQDETQFKPYPASNFSTAVPQSATSMLVT
ALKDSRWFIPLERQGLQNLLNERKIIIRAAQENGTVAEENRMPLQLSLVAANVMIEGSIIGYESNVKSGGV
GARYFGIGGDTQYQLDQIAVNLRVVNVSTGEVLSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEIGYTAN
EPVMCLCMSAIETGVIIHINDGINRGLWELKNKGDAKNTILAKYRSMVPPES
- [1145]
- [1146] 서열번호: 75
- CLTAAPKEAARPTLLPRAPSYTDLTHLPSPQGRIFVSVYNIQDETQFKPYPACNFSTAVPQSATAMLVSA
LKDSKWFPIPLERQGLQNLLNERKIIIRAAQENGSAINNQRPLSSLVAANILIEGSIIGYESNVKSGGVGA
RYFGIGASTQYQLDQIAVNLRAVDVNTGEVLSSVNTSKTILSYEVQAGVFRFIDYQRLLEGELGYTTNEP
VMLCMSAIESGVIIYLVNDGIERNLWQLQNPSEINSPILQRYKNNIVPAES
- [1147]

- [1148] 서열번호: 76
- CITSPPKQAAKPTLLPRSQSYQDLTHLPEPQGRLFVSVYNISDETQGFQYPASNFSSTVPQSATAMLV
ALKDSNWFIPLERQGLQNLNLERKIIRAAQENGTVAVNNRTQLPSLVAANILIEGSIIGYESNVKSGGAG
ARYFGIGASTQYQLDQIAVNLRVVNVSTGEVLSSVNTSKTILSYEFQAGVFRYIDYQRLLEGEVGYTVNE
PVMCLMSAIETGVIYLVNDGISRNWQLKNASDINSPVLEKYKSIIVP
- [1149]
- [1150] 서열번호: 77
- CLTAPPKQAAKPTLMPRAQSYQDLTHLPEPAGKLFVSVYNIQDETQGFQYPASNFSSTAVPQSATAMLV
ALKDSGWFIPLERQGLQNLNLERKIIRAAQENGTAAVNNQHQLSSLVAANVLVEGSIIGYESNVKSGGA
GARFFGIGASTQYQLDQIAVNLRVVDVNTGQVLLSSVNTSKTILSYEVQAGVFRYIDYQRLLEGEIGYTTN
EPVMCLCMSAIETGVIYLVNDGINRNWTLKNPQDAKSSVLERYKSTIVP
- [1151]
- [1152] 서열번호: 78
- CITTPQEAAKPTLLPRDATYKDLVSLPQPRGKIYVAVYNIQDETQGFQYPASNFSSTVPQSATAMLVSS
LKDSRVFVPLERQGLNLLNERKIIRAAQQNGTVGDNNASPLPSLYSANVIVEGSIIGYASNVKTGGFG
ARYFGIGGSTQYQLDQVAVNLRVNVHTGEVLSSVNTSKTILSYEQAGVFRFIDYQRLLEGEAGFTTNEP
VMTCLMSAIEEGVIHLINDGINKKLWALSNAADINSEVLTRYRK
- [1153]
- [1154] 서열번호: 79
- ITEVPKEAAKPTLMPRASTYKDLVALPKPNGKIIVSVYSVQDETQGFQPLPASNFSTAVPQSGNAMLTSA
KDSGWVFWPLERQGLNLLNERKIIRAAQENGTVAAANNQQLPSLLSANVVIEGAIIGYDSDIKTGGAGAR
YFGIGADGKYRVDQVAVNLRAVDVRTGEVLLSVNTSKTILSSELSAGVFRFIEYQRLLELEAGYTTNEPV
MMCMMSALEAGVAHLIVEGIRQNLWSLQNPSDINNPIIQRYMKEDVP
- [1155]
- [1156] 서열번호: 80
- PETSEPTLMQRGANYIDLISLPKQKGIFVSVYDFRDQTGQYKPQPNFSSTAVPQGGTALLTMALLDS
EWFYPLERQGLQNLNTERKIIRAAQKKQESISNHGSTLPSSANVMIEGGIVAYDSNIKTGGAGARYLG
IGSGSQYRADQVTNIRAVDVRSKILTSTVTSKILSYEVSAGAFRVDYKELLELEVELGYTNNEPVNIAL
MSAIDSAVIHLIVKGVQQGLWRPANLDRNNPIFKKY
- [1157]
- [1158] 서열번호: 81
- PDASEPTLMQRGATYLDLISLPKQGGKIYVSVYDFRDQTGQYKPQPNFSSTAVPQGGTALLTMALLD
SEWFYPLERQGLQNLNTERKIIRAAQKKQESISNHGSTLPSSANVMIEGGIVAYDSNIKTGGAGARYL
GIGSGSQYRADQVTNIRAVDVRSKILTSTVTSKILSYELSGAFRVDYKELLELEVELGYTNNEPVNIA
LMSAIDSAVIHLIVKGIEEGLWRPENQNGKENPIFRKY
- [1159]
- [1160] 서열번호: 82
- PETSKEPTLMARGTAYQDLVSLPLPKGKVYVSVYDFRDQTGQYKPQPNFSSTAVPQGGALLTTALLD
SRWFMPLERQGLQNLNTERKIIRAAQKKDEIPTNHGVHLPSLASANIMVEGGIVAYDTNIQTGGAGARYL
GVGASGQYRTDQVTNIRAVDVRTGRILLSVTTSKILSKELQTVFKFVDYKDLLEAELGYTTNEPVNL
AVMSAIDAAVVHVIVDGIKTGLWEPLRGEDLQHPPIQEYMNRSKP
- [1161]
- [1162] 서열번호: 83
- CATHIGSPVADEKATLMPRSVSYKELISLPKPKGKIVAAYVDFRDQTGQYLPAPASNFSSTAVTQGGVAML
STALWDSQWFVPLERQGLQNLNTERKIVRAAQKNPNVPGNNANQLPSLVAANILIEGGIVAYDSNVRTG
GAGAKYFGIGASGEYRVDQVTNLRVAVDIRSGRILNSVTTSKTVMSQQVQAGVFRFVEYKRLLEAEAGF
STNEPVQMCVMSAIESGVIRLIANGVRDNLWQLADQRDIDNPILQEYLDQDNAP
- [1163]

- [1164] 서열번호: 84
- ASSSLMPKGESYYDLINLPAPQGVMLAAVYDFRDQTGQYKPISSNFSTAVPQSGTAFLAQALNDSSWF
IPVEREGLQNLLTERKIVRAGLKGDKANKLPQLNSAQILMEGGIVAYDTNVRTGGAGARYLGIGAATQFRV
DTVTNLRRAVDIRTGRLLSSVTTTKSILSKEITAGVFKFIDAQELLESELGYTSNEPVSLCVAIAESAVVH
MIADGIWKGAWNLADQASGLRSPVLQKY
- [1165]
- [1166] 서열번호: 85
- QDSEPTLTTPRASYDYLINMPRPKGRLMAVVGFRDQTGQYKTPASSFSTSVTQGAASMLMDALSAS
GWFVVLEREGLQNLLTERKIIRASQKKPDVAENIMGELPPLQAANLMLEGGIIAYDTNVRSGGEGARYLG
IDISREYRVDQVTNLRRAVDVRTGQVLANVMTSKTIYSVGRSAGVFKFIEFKLLEAEVGYTTNEPAQLC
VLSAIESAVGHLLAQGIEQRLWQV
- [1167]
- [1168] 서열번호: 86
- MPKSDTYDYLIGLPHPGQSMMLAAVYDFRDQTGQYKAIPSSNFSTAVPQSGTAFLAQALNDSSWFVPVER
EGLQNLLTERKIVRAGLKGANQLPQLSSAQILMEGGIVAYDTNIKTGGAGARYLGIGVNSKFRVDTVTV
NLRRAVDIRTGRLLSSVTTTKSILSKEVSAGVFKFIDAQDLLESELGYTSNEPVSLCVAQAIESAVVHMIAD
GIWKRAWNLADTASGLNNPVLQKY
- [1169]
- [1170] 서열번호: 87
- LTRRMSTYQDLIDMPAPRGKIVTAVYSFRDQSGQYKPAPSSSFSTAVTQGAAAMLVNVLNDSGWFIPLE
REGLQNILTERKIIRAAALKKDNVPVNNASAGLPSSLAANIMLEGGIVGYDSNIHTGGAGARYFGIGASEKY
RVDEVTNLRRAIDIRTGRILHSVLTSKKILSREIRSDVYRFIEFKHLEMEAGITTNDPAQLCVLSAIESAV
AHLIVDGVIKKSWSLADPNELNSPVIQAYQQQRI
- [1171]
- [1172] 서열번호: 88
- PSDPERSTMGELTPSTAELRNLPNEKIVIGVYKFRDQTGQYKPSSENGNNWSTAVPQGTITILIKALD
SRWFIPIERENIANLLNERQIIRSTRQEYMKDADKNSQSLPPLLYAGILLEGGVISYDSNTMTGGFGARYF
GIGASTQYRQDRITIYLRAVSTLNGEILKTVYTSKITLSTSVNGSFFRYIDTERLLEAEVGLTQNEPVQLAV
TEAIEKAVRSLIIEGTRDKIW
- [1173]
- [1174] 서열번호: 89 내지 116은 표 3에 제시된 CsgF 펩티드이다.
- [1175] 서열번호: 117은 위치 97에 W가 있는 서열번호: 3이다.
- CLTAPPKEAARPTLMPRAQSYKDLTHLPAPTGKIFVSVYNIQDETGGQFKYPASNSTAVPQSATAMLVT
ALKDSRWFIPLERQGLQNLLNERKIIWAAQENGTVAINNRIPQLSLTAANIMVEGSIIGYESNVKSGGVG
ARYFGIGADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEVGYTSNEP
VMLCLMSAIETGVIFLINDGIDRGLWDLQNKARQNDILVKYRHMSVPPES
- [1176]
- [1177] 서열번호: 118(>P32270; 엔테로박테리아 파지 T4)
- MTFDDLTEGQKNFNIVMKAIKEKKHHVTINGPAGTGKTTLTKEIIEALISTGGTGIIA
APTHAAKKILSKLSGKEASTIHSILKINPVTYEENVLFEQKEVPDLAKCRVLICDEVSMY
DRKLFKILLSTIPPWCTIIGIGDNKQIRPVEPGENTAYISPFTHKDFYQCELTEVKRSN
APIIDVATDVRNGKWNYDKVVDGHGVRGFTGDTALRDFMVNYFSIVKSLDDL FENRVMFAF
TNKSVDKLSNIIRKKIFETDKDFIVGEIIVMQEPLFKTYKIDGKPVSEIIFNNGQLVRII
EAETSTFVKARGVPGEYLIRHWDLTVETYGDDEYREKIKIISDEELYKFNFLAKTA
ETYNWKNKGKAPWSDFWDAKSQFSKV KALPASTFHKAQGM SVDRAFIYTPCIHYADVEL
AQQLLYVGVTGRGRYDVFFV
- [1178]

[1179] 서열번호: 119(>D0MKQ2; 로도써머스 마리누스)

MEELSNEQQRVLDPVHVLAWLERNDAPPFILTGSAGTGKTLIRHLVRALQDRRIHYALAA
PTGRAARILSERTGDHARTLHSLIYIFDRYQLVEEADRQTDEPLSLQLHFALRSAEHDAR
LIIVDEASMVSDTAGEEELYRFGSGRLLNDLLTFARLIPKRDRPPTTRLLFVGDPAPQLPP
VGQSVSPALSAQYLRDTFGLSAETAHLRSVYRQRKGHPILETATALRNALEKGHYHTFRL
PEQPPDLRPVGLLEEAIETTATDFRRQNPSVLLCRTNALARKLNAAVRARLWGREGLPQP
GDLLLVNRNAPLHGLFNGDLVLVETVGPLEHRRVGRRGRPPVDLYFRDVELLYPHEKPRN
RIRCKLLENLLESPDGLSPDIIQALLIDFYRRHPSLKHGSSEFRLMLANDAYFNALHVR
YGYAMTVHKAQGGWEKRVFVNDWRHFRHAEFFRWAYTAITRAREELLTIGAPSFEALS
DMRWQPPASVPAPEQAAENATRFPLKALETYHQRLSEALTAAGIETTGVELLQYAVRYHL
ARADRTTRIQQYYRGDGGQISRVITLGGADDPELTQQAYALFERILSEPPADSGELPENPL
LREFLEAHLRLLEGSGIRIVHWKEMPYALRLYFSADGENVTIDFYNNRRGVWTHAQEVGR
SSSGALFARIQSLQADS

[1180]

[1181] 서열번호: 120(>B1X365; 시아노테세 ATCC51142)

[1182] MSQSVVVPDELGEITAVIEFYQDAVDKIEPKIVFLELRKNVVDWVSRTQLKIEEKEIQA

TGLTRQQQTAYKEMINFIENSSEQYFRLSGYAGTGKSFLMAKVIEWLWKQEDYKYSVAAPT
NKAANKLTQIARSQGIKEATTVAKLLKLQPTIDVDTGQSFENSEKELELKDYDVIII
DEYSMLNKDNFRDLQQAVKGGESKFIFVGDSSQLPPVKEKEPIVANHPDIRKSANLTQIV
RYDGEIVKVAESIRRNPRWNHQTYPFETVADGTIIKLNTEWLQQALSHFEKEDWLSNPD
YVRMITWRNKTADKYNQAIREALYGENVEQLVVGDRLIAKKPVFRSLPGGKKKEKKIILN
NSEECKVIETPKINYNEKYKWEFYQVKVRTDEGGMIELILTSESEEKQKKLAKRA
REEENYSEKKKQWAIYYELDFDNMAYAYALTCHKAQGSSIDNVFLLVSDMHYCRDKTK
MIYTGLTRAKKCCYVG

[1183]

[1184] 서열번호: 121(>Q2S429; 살리니박터 루버)

MSTFADAPFTEDQEEAYDHVYDRLAQGERFTGLRGYAGTGKTYLVSRLEQLLEDCTVT
VCAPTHKAVQVLSDELGDAPVQMQLHSLGLRLQPKQDGEYELVAEEERNFAEGVVIVD
EASMIGREEWSHIQDAPFWVQWLFVGDPAQLPPVNEDPSPALDVPGPTLETIHRQAADNP
ILELATKIRTGADGRFGSTFEDGKGVAVTRNREEFLDSILRAFDADAFAEDATHARVLAY
RNKTVRRYNREIRAERYGADADRFEVEGEWLVGTETWYDGVQRLTNSEEVVKKAQVETF
EADDQSEWTVWELKIRTGRLTRTIHVLHEEERERYENALERRRGKAEDDPKWDREFE
LRERFARVDYAYATTVHRAQGSTYDTVFVDHRDLRVCRGEERGALLYAVTRPSRRLALLV

[1185]

[1186] 서열번호: 122(>B6BJ43(UPI000183B2F5); 술푸리모나스 고틀란드카(*Sulfurimonas gotlandica*) GD1)

MKILNKETYKLSLHQEEVFTQIVSQLDTKVSSILKSTNIEDYLLSLTGPA GTGKFTLTQ
IAKYLVEKRKESEYPMSSDFDTITAPTHKAVGVLSKLLRENNIQSSCKTIHSFLGIKPF
IDYTTGEEKFVVDKTNKRKDRTSILIVDESSMIGNTYEYILEAIEDKRVNVVLFIDPY
QLLPIENSKNEIYDLNPRFFLSEVVRQAENSYIIRVATKLREIKNQDFISLQQFFQENM
EDEITFFHNKEAFLEDFYKEEEWYKENKILATYKNKDVAFNKIIRNKFEQKGNTPST
LLAGDMIRFKDAYTVGDITIHNGQELQLGSTEVKYHDSLHIEYWECKSIYALEQQVFRV
VNPDSEAVFNQKLQSLATKAKQAKFPDNKKLWKLYYETRNMFANVQYIHASTIHKLQGST
YDVSYIDIFSLVHNHYMSDEEKYRLLYVAITRASKDIKIFMSAFDRTSDEKVIINNQNSE
TMNTLKLHLDIDILKDLDL

[1187]

- [1188] 서열번호: 123(>M4MBC3; 비브리오 파지 헨리에테 12B8)
- MADFELTLGQKTVLGEVISTILKPVNLNDSRFHTMHGPAGSGKTTVLQRIISQIPAYKT
IGFCSPTHKSVKVIIRMAREAGISHRVDIRTIHSALGLVMKPVRGDEVLVKEPFAEERIY
DVLIIIDEAGMLNDELIMYILESQSSKVFVGDMCQIGPIQSNLPEEDGYTPTSTDDVSKV
FTEVEMMSALTEVVRQAEGSPIIQLATEFRLAQDDIYADLPRIVTNTTPDGNGIITMPNG
NWVDSAVARFQSDQFKEDPDHCRIVCYTNAMVDLCNDLVRKRLFGADVPEWLEDEILVAQ
EMGSTWNNADELRIVSIDDHFDQQYEVPCWRMQLESVEDHKLHNALVVKGDYIEDFKFRL
NAIAERANTDKNMSGMHWKEFWGMRKKFNTFKNVYAGTAHKSQGSTFDYTYVFTPDFYKF
GATMTIKRLLYTAITRSRYTTYFAMNTGAQ
- [1189]
- [1190] 서열번호: 124(>I6XGX8; 비브리오 파지 파이-pp2)
- MGLTNCQQGAMDAFLES DGHTISGPAGSGKTFMKLSILEALESKGKNVTMVTPTHQAKN
VLHKATGQEVSTIHSLLKIHPDTYEDQKHFTQSGEVEGLDEIDLVLVEEASMVDEELFQI
TGRTMPRKCRILAVGDKYQLQPVKHDPGVISPFFTKFTTFEMNEVVRQAKDNPLIQVATE
VRNGQWLRTNWSKERRQGVLVHPNVNKMMLDTYLSKVNSPEDLLDYRILAYTND CVDTFNG
IIREHVYNTSEFPFIPGEYLVQMPVMVSNKGYPVCVIENGEVVKILDVRQKTIDGMLPKV
DNEAFDVAVLTVKEKEDGNVYEFTVLWDDLQKERFARYLSVAAGTYKSMRGNTKRYWRAFW
GLKEQMIETKSLGASTVHKSQGSTTVKGVCLYTQDMGYAEPEILQQLVYVGLTRPTDWALY
N
- [1191]
- [1192] 서열번호: 125(>E5DRP6; 아에로모나스 파지 65)
- MSESEITLTPSQNMAVNEVKNGTGHITISGPPGSGKTFVLKYLIKMLGDELGTVLAAPTH
QAKIVLTEM SGIEACTIHSMLKIHPETLEDIQIFDQSKLPDL SNIRYLIVEEASMH SKTL
FKITMKSIPPTCRIIAIGDKDQIQPEEHAQGELSPYFTDPRFSQIRLTDIMRQSLDNPII
QVATKIREGGWIEPNWNRDTKTGVYKVSIGITDLVNSYLRAVKTPEDLTKYRFLAYTNKVV
NKVNSIVREHVYKTKLPFIEGEKIVLQEPVMVEHEDDTIETIFTNGEVVTINEIEVFDRT
IRIDGSPFEKVNAAKLSVSSDYSIGIEHDFCVLYGSESRLFEYQLSESAGNIQMGKGGN
QRSAWKSFWAAKKMFIETKSLGASTIHKSGSTVKGVWLALHDIHYADEELKQQLVYVGV
TRPTDFCLYFDGTK
- [1193]
- [1194] 서열번호: 126(>I6XH64; 아에로모나스 파지 CC2)
- MAVDAVQSGTGHITISGPPGSGKTFVLKYIIMKMLGDELGTVLAAPTHQAKIVLTEM SGIE
ACTIHSMLKIHPETLEDIQIFDQSKMPDLSTVRYLIIIEEASMH SKALFNITMKSIPPTCR
IIAIGDKDQIQPVHDAPGELSPYFTDSRFTQIRMTDIMRQSLDNPIIQVATTIREGGWIY
QNWNKEKKS GYKVSITDLINSYL RVVKT PEDLTKYRFLAFTNKVVDKVN SIVRKHVYK
TDLPFIEGEKLV LQEPVMVEYDDDTIETIFTNGEVVTVDEIEVSDMNIRIDGSPAFSISV
AKLKVTSDFSGVTHDIMS VYGEDSKAEFNYQLSEAAAVIKQMQRGQTKAAWASFWDAKKT
FTETKSLGACTIHKSGSTVKGVWLGLHDISYADTDLQQLVYVGVTRPTDFCLYFDGSK
- [1195]
- [1196] 서열번호: 127(>K4FBD0; 크로노박터 파지 vB CsaM GAP161)
- MSELTFDLSDDDQKSAHDRVIHNIQNAIHTTITGGPGVGKTTLVKFVNTLKGLGISGIW
LTAPTHQAKNVLAAATGMDATTIHSALKISPVTNEELRVFEQQKGKKAPDLSTCRVFVVE
EVSMDMDLFIIRRSIPSNAILGLGDKDQIRPVNADGRVELSPFFDEEIFDVIRMDKI
MRQAEGNPIIQVSRVARDGKMLKPM SVGDLGVFQHANA VDFLRQYFRRVKT PDDL IENRM
FAYTNDNVNKL NATIRKHYLKTTEPFILDEVIMQEPLVQEMRLNGQIFTEIVYNNNEKI
RVLEIIPRREVIAEKCKDEKIEIEFYLLKTVSLEEETEAQIQVVDPVMKDR LGNYLAYV
ASTYKRIKQQTGYKAPWHSFWAIKNKFQDVKPLPVCTYHKSQGSTYDHAYMYTRDAYAFA
DYDLCKQLIYVGVTRARYTVDYV
- [1197]

- [1198] 서열번호: 128(>D5JF67; 클램시엘라 파지 KP15)
- MSELTFFDDLSEDQKNAHDRVIKIRNKIHTTITGGPGVGKTTLVKFVFETLKKLGISGIW
LTAPTHQAKNVLSEAVGMDATTIHSALKISPVTNEELRVFEQQKGKKAADLSECRVVFVE
EVSMDVKELFRIKRTIPSCAVILGLGDKDQIRPVNTEGITELSPFFDEEIFDVIRMDKI
MRQAEGNPQQVSRRAIRDGKPLMPLMNGELGVMKHENASDFLRRYFSRVKTPDDLNNNRM
FAYTNANVDKLNNAVIRKHLKYTDQPFIVGEVVVMQEPLVTEGRVNGVSFVEVIYNNNEQI
KILEIIPRSDTIKADRCDPVQIDYFLMKTESMFEDTKADIQVIADPVMQERLGDYLNVA
FQYKKMKQETGYKAPWYSFWQIKNKFQTVKALPVCTYHKQGSGTYDHSYMYTRDAYAYAD
YELCKQLLYVGTTRARFTVDYV
- [1199]
- [1200] 서열번호: 129(>J7HXT5; 스테노트로포모나스 파지 IME13)
- MVYDDLTVGQKDAIEKALQAMRTKRHITIRGPAGSGKTTMTRFLLERLFTGQQGIVLT
APTHQAKKELSKHALRKSQYIQLSVLKINPSTLEENQIFEQKGTPDFSKTRVLICDEVSYF
TRKLFIDILMRNVPSHCVVIGIGDKAQIRGVSEDDTHELSPFFTDRNFEQVELTEVKRHQ
PIIEVATDIRNGKWIYEKLDDSGNGVKQFHTVKDFLSKYFERTKTPNDLLENRIMAYTNN
SVDKLSNVIRKQLYGANAAPFLPDEILVMQEPLMFDIDIGGQTLKEVIFNNGQNVRVINV
KPSRKTLKAKGVGEIEVECTMLECESYEEDEDDYRRRAWFTVVHDQNTQYAINFLSIIAE
KYRSREVFNPWKDFWAIIRNTFVKVRPLGAMTFHKSQGSTFDNAYLFTPCLHQYCRDPDVA
QELIYVGNTRARKNVCFV
- [1201]
- [1202] 서열번호: 130(>E5EYE6; 아시네토박터 파지 Ac42)
- MNFEDLTEGQKNAYTAAIKAIETVPSSSAEKRLHTINGPAGTGKTTLTFLIAELIRRG
RGVYLAAPTHQAKKVLSQAHGMEASTIHSLLKINPTTYEDSTTFEQKDVPDMSECRVLIC
DEASMYDLKLFQILMSSIPLCCTVIALGDIQIRPVEPGAFEGQVSPFFTYEKFQVSLT
EVMRSNAPIIDVATSIRTGNWIYENVIDGAGVHNLTSESVKSFMEKYFSIVKTPEDLFE
NRLAFTNKSVDLKNIVRKKIYNTLEPFIDGEVLVMQEPLIKSYTYEGKKVSEIVFNNG
EMVKVLCCSQTSDEISVRGCSTKYMVRWQLDLQSLDDPDLTGSINIVDEAEINKNLV
LGKSAEQFKSGAVKAAWADWWKLKRNFKVKALPCSTIHKSQGTSVDNVFLYTPCIHKAD
SQLAQQLLYVGATRARNVYYI
- [1203]
- [1204] 서열번호: 131(>E3SFA5; 시겔라 파지 SP18)
- MIKFEDLNTGQKEAFDYITEAIQRRSGECITLNGPAGTGKTTLTKFVIDHLVRNGVMGIV
LAAPTHQAKKVLSKLSGQTANTIHSILKINPTTYEDQNIFEQREMPDMSKCNVLCDEAS
MYDGSFLFKIICNSVPEWCTILGIGDMHQLQVPDPGSTQQKISPFFTHPKFKQIHLTEVMR
SNAPIIEVATEIRNGGWFRDCMYDGHGVQGFTSQTALKDFMVNYFGIVKADMLMENRMY
AYTNKSVEKLNNIIRKLYETDKAFLPYEVLVMQEPHMKELFEFGKKFSETIFNNGQLVR
IKDCKYSTILRCKGESHLVINYWDLEVESIDEDEEYQVDRIKVLPEQQPKFQAYLAK
VADTYKQMKAAAGKRPEWKDFWKARRFTLKVRLPVSTIHKAGVSVDKAFIYTPCIHMAE
ASLASQLAYVGITRARYDAYV
- [1205]
- [1206] 서열번호: 132(>I7J3V8; 예르시니아 파지 phiR1-RT)
- MITYDDLTGQKSAFNDTMEAIKKNKGHITINGPAGTGKTTLTKFIIHDLIKTGEAGIIL
CAPTHQAKKVLSKLSGMDASTIHSVLKINPTTYEENQIFEQREVPLAACRVLICDEASF
YDRKLFGIILATVPSWCTVIALGDKDQLRPVTPGESEQQLSPFFSHAKFKQVHLTEIKRS
NGPIIQVATDIRNGGWLSENIVDGEGVHAFNSNTALKDFMIRYFDVVKTADDLIESRMLA
YTNKSVDKLNIIIRKLYETDKPFINGEVLVMQEPLMKELFEFGKKFHEIVFNNGQLVKI
LYASETSTFISARNVPGEMYIRYWNLEVTADSDDDYATSIQVICDPAEMTKFQMFLAK
TADTYKNSGVKAYWKDFWSVKNKFKVKALPVSTIHKSQGCTVNNTFLYTPCIHMADAQL
AKQLLYVGATRARTNLYI
- [1207]

[1208] 서열번호: 133(>M1EA88; 살모넬라 파지 S16)

```
MITFEQLTSGQKLAFDETIIRAIKEKKNHVTINGPAGTGKTTLTKFIMEHLVSTGETGIIL
TAPTHAAKKVLTKLSGMEANTIHKILKINPTTYEESMLFEQKEVPDLASCRVLICDEASM
WDRKLFKILMASIPKWCTIVAIGDVAQIRPVDPGETEAHISPPFIHKDFKQLNLTEVMRS
NAPIIDVATDIRNGSWIYEKTVDPGHGVHGFSTTALKDFMMQYFSIVKSPEDLFENRMLA
FTNKSVDKLNIIIRRLYQTEEFVVGIVVMQEPLMRELVFEGKKFHETLFTNGQYVRI
LSADYTSSFLGAKGVSGEHLIRHWLDVETYDDEEYAREKINVISDEQEMNKFQFFLAKT
ADTYKNWNKGGKAPWSEFWDAKRKFHKVKALPCSTFHKAQGISVDSSFIYTPCIHVSSDN
KFKLELLYVGATRGRHVDFFV
```

[1209]

[1210] 다음 실시예는 본 발명을 예시한다.

[1211] 실시예

[1212] 돌연변이체 CsgG 기공 및 돌연변이체 CsgG/CsgF 복합체를 제조하고 테스트하는 상세한 방법은 W02016/034591, W02017/149316, W02017/149317, W02017/149318, W02018/211241 및 W02019/002893(모두 그 전문이 본원에 참조로 포함됨)에 기재되어 있다. 변형된 Dda 헬리카제를 제조하는 상세한 방법은 W02015/055981, W02015/166276 및 W02016/055777(모두 그 전문이 본원에 참조로 포함됨)에 기재되어 있다.

[1213] 실시예 1

[1214] 이 실시예는 폴리뉴클레오티드의 이동을 제어하는 폴리뉴클레오티드 결합 단백질을 사용하여 변이체 나노기공의 속도 특성을 대조군 변이체 나노기공에 대해 비교하는 방법을 기재한다.

[1215] 3.6 kb 분석물 제조

[1216] 특이적 프라이머 및 PCR을 사용하여 이중 가닥 3.6 kb DNA 분석물을 제조하였다. PCR 생성물을 NEBNext 단부 복구 및 NEBNext dA-테일링 모듈(New England Biolabs(NEB))에 적용하여, 3' dA 돌출부를 생성하였다.

[1217] 변이체 및 대조군 나노기공 제조

[1218] WO 2019/002893(이의 전문이 본원에 참조로 포함됨)에 기재된 바와 같은 변이체 나노기공 또는 대조군 나노기공을 암호화하는 재조합 발현 벡터를 화학적으로 적격한 이.콜라이 세포로 형질전환하였다. 작제물은 C-말단 Strep 친화성 태그를 포함하였다. 세포를 선택을 위한 적절한 항생제를 함유하는 LB 아가 플레이트에 플레이트팅하였다. 아가 플레이트로부터의 콜로니를 항생제를 함유한 LB 배지에 접종하고 밤새 성장시킨 후 자가유도 배지로 희석하고 18 C에서 68시간 동안 인큐베이션하였다. 세포를 원심분리를 통해 수확한 후 용해하고 머그부스터(Merck 70921) 및 0.1% DDM에서 추출하였다. 상청액을 친화성 크로마토그래피, 60°C에서 30분 동안 열 처리 및 크기 배제 크로마토그래피를 사용하여 정제하여 SDS-PAGE에 의해 판단된 바와 같이 올리고머 나노기공을 선택하였다.

[1219] CsgG-CsgF 복합체를 위에서와 같이 정제된 나노기공 및 화학적으로 합성된 CsgF 펩티드로부터 제조하였다. 나노기공을 50mM Tris, 150mM NaCl, 2mM EDTA, 0.1% SDS, 0.1% Brij58, pH7.0을 포함하는 완충액으로 완충액 교환한 다음 25°C에서 1시간 동안 CsgG 단량체에 대한 4x 몰 과량의 펩티드에서 인큐베이션하였다. 반응은 60°C에서 15분 동안 가열하여 중단한 후 원심분리하여 임의의 침전물을 제거하였다.

[1220] 서열분석 어댑터 제조

[1221] N-말단 친화성 및 용해성 태그가 있는 W02016/055777에 기재된 폴리뉴클레오티드 결합 단백질의 변이체를 암호화하는 재조합 발현 벡터를 화학적으로 적격한 이.콜라이 세포로 형질전환하였다. 세포를 항생제를 함유하는 LB 아가 플레이트에 플레이트팅하였다. 아가 플레이트로부터의 콜로니는 LB 성장 배지에 접종하고, OD 0.400 - 0.800으로 성장시키고 유도한 다음, 18°C에서 추가 16시간 동안 성장시켰다. 세포를 벤조나제 및 프로테아제 억제제의 존재 하에 초음파처리로 용해시켰다. 상청액을 친화성 크로마토그래피를 사용하여 추가로 정제하였다. 정제된 대조군 폴리뉴클레오티드 결합 단백질을 주위 온도에서 10분 동안 50 mM Hepes pH8.0, 100mM 칼륨 아세테이트, 1mM EDTA에서 WO 2015/110813A1(그 전문이 본원에 참조로 포함됨)에 기재된 서열분석 Y 어댑터에 결합시켰다. TMAD(SIGMA)를 100 μM의 최종 농도가 되도록 첨가하고 34°C에서 1시간 동안 인큐베이션하였다. 10 mM ATP, 10 mM MgCl₂ 및 0.5 M NaCl을 첨가하고 주위 온도에서 10분 동안 인큐베이션하였다. 폴리뉴클레오티드 결합 단백질 결합된 서열분석 어댑터를 음이온 교환으로 정제하였다.

[1222] 라이브러리 제조

[1223] 서열분석 어댑터를 3.6 kb 분석물에 결합시켰다. 서열분석을 위해 라이브러리를 제조하고 제조업체의 지침(Oxford Nanopore Technologies)에 따라 MinION 유동 셀에서 실행하였다.

[1224] 전기적 측정

[1225] 전기적 측정은 MinION 유동 셀에 삽입된 나노기공에서 획득하였다. 단일 기공을 각 채널 내의 블록 공중합체 막에 삽입한 후, 25 mM 칼륨 포스페이트, 150 mM 칼륨 페로시아나이드(II), 150 mM 칼륨 페로시아나이드(III), pH 8.0을 포함하는 완충액 1 mL를 시스템을 통해 흘려보내 임의의 과도한 단백질 나노기공을 제거하였다.

[1226] 표준 서열분석 스트립트를 실행하고 원시 데이터를 MinKNOW 소프트웨어(Oxford Nanopore Technologies)를 사용하여 대량 FAST5 파일에 수집하였다.

[1227] 데이터 분석

[1228] DNA 라이브러리를 Guppy(Oxford Nanopore Technologies)의 표준 염기호출 알고리즘을 사용하여 서열분석하였다. 각 나노기공 변이체에 대해 145개 이상의 판독을 3.6 kb 분석물 참조 서열에 매핑하였다.

[1229] 개별 DNA 가닥이 기공을 통해 전위할 때의 속도는 참조에 매핑된 염기의 수를 판독 지속시간(초 당 염기로 측정됨)으로 나누어 계산하였다. 중간 속도(초 당 염기)는 다중 개별 DNA 가닥이 기공을 통해 전위할 때의 중간 속도였다. 대조군 나노기공 유동 셀의 중간 속도는 변이체 나노기공 유동 셀의 중간 속도에서 뺐다. 이는 속도 델타였다. 따라서, 양의 속도 델타는 DNA가 대조군 나노기공보다 변이체 나노기공을 통해 더 빠르게 전위됨을 나타내었다.

[1230] 다중 개별 DNA 가닥이 기공을 통해 전위할 때의 속도의 변동은 사분위 범위를 계산하여 측정하였다. 따라서, 작은 사분위 범위는 좁은 속도 분포를 의미하고, 큰 사분위 범위는 광범위한 속도 분포를 의미한다. 정규화된 속도 분포는 속도의 사분위 범위를 중간 속도로 나누고 100을 곱하여 계산하였다. 대조군 나노기공 유동 셀의 정규화된 속도 분포를 변이체 나노기공 유동 셀의 정규화된 속도 분포에서 뺐으며, 이는 정규화된 속도 분포 델타였다. 따라서, 음의 정규화된 속도 분포 델타는 변이체 나노기공이 대조군 나노기공보다 더 좁은 속도 분포를 가짐을 나타내었다.

[1231]

변이체 위치	변이체 AA	속도 Δ
97	D	58.23
97	G	28.47
97	N	23.22
97	R	44.64
97	S	64.24
100	A	77.22
100	K	76.49
100	S	33.38
101	A	-3.93
101	결실	43.00
101	G	58.77
101	S	44.82
102	A	33.83
102	D	21.27
102	R	35.98
102	S	14.59
102	W	44.70

[1232] 표 11. 속도 Δ 는 변이체 나노기공 유동 셀과 대조군 나노기공 유동 셀 사이의 중간 속도(bps) 차이이다.

변이체 위치	변이체 AA	정규화된 속도 분포 Δ
97	D	3.72
97	G	-0.06
97	N	-0.33
97	R	0.56
97	S	-3.18
100	A	-1.04
100	K	0.03
100	S	-0.66
101	A	-0.44
101	결합	3.99
101	G	-0.13
101	S	1.13
102	A	-0.64
102	D	11.75
102	R	-1.37
102	S	-2.37
102	W	27.66

[1233]

[1234] 표 12. 정규화된 속도 분포 Δ 는 변이체 나노기공 유동 셀과 대조군 나노기공 유동 셀 사이의 정규화된 속도 분포의 차이이다.

[1235] 실시예 2

[1236] 이 실시예는 W02017/149316에 기재된 나노기공을 사용하여 폴리뉴클레오티드의 이동을 제어하는 대조군 변이체 폴리뉴클레오티드 결합 단백질에 대한 폴리뉴클레오티드의 이동을 제어하는 변이체 폴리뉴클레오티드 결합 단백질의 속도 및 정확도 특성을 비교하는 방법을 기재한다.

[1237] 바코드가 있는 3.6 kb 분석물 제조

[1238] 이중 가닥 3.6 kb DNA 분석물을 특이적 프라이머 및 PCR을 사용하여 제조하였다. PCR 생성물을 NEBNext 단부 복구 및 NEBNext dA-테일링 모듈(New England Biolabs(NEB))에 적용하여 3' dA 돌출부를 생성하였다. 바코드를 제조업체의 지침에 따라 EXP-PBC001(Oxford Nanopore Technologies)을 사용하여 분석물에 도입하였다.

[1239] 변이체 서열분석 어댑터 제조

[1240] N-말단 친화성 및 용해성 태그가 있는 W02016/055777에 기재된 폴리뉴클레오티드 결합 단백질의 변이체를 암호화하는 제조할 발현 벡터를 화학적으로 적격한 이.콜라이 세포로 형질전환하였다. 세포를 항생제를 함유하는 LB 아가 플레이트에 플레이트하였다. 아가 플레이트로부터의 콜로니를 LB 성장 배지에 접종하고, OD 0.400 - 0.800 으로 성장시키고 유도한 다음, 18℃에서 추가 16시간 동안 성장시켰다. 세포를 리소자임, 벤조나제 및 프로테아제 억제제의 존재 하에 버그부스터 추출 시약(Merck 70921)으로 용해시키거나 벤조나제 및 프로테아제 억제제의 존재 하에 초음파처리로 용해시켰다. 상청액을 친화성 크로마토그래피를 사용하여 추가로 정제하였다. 물 초과량의 정제된 변이체 폴리뉴클레오티드 결합 단백질을 주위 온도에서 10분 동안 50 mM Hepes pH8.0, 100mM 칼륨 아세테이트, 1mM EDTA에서 WO 2015/110813A1(그 전문이 본원에 참조로 포함됨)에 기재된 서열분석 Y 어댑터에 결합시켰다. TMAD(SIGMA)를 100 μ M의 최종 농도가 되도록 첨가하고 34℃에서 1시간 동안 인큐베이션하였다. 10 mM ATP, 10 mM $MgCl_2$ 및 0.5 M NaCl을 첨가하고 주위 온도에서 10분 동안 인큐베이션하였다. 변이체 폴리뉴클레오티드 결합 단백질 결합된 서열분석 어댑터를 Sera-Mag SpeedBeads(Thermo Scientific)를 사용하여 정제하였다. 이들은 변이체 서열분석 어댑터였다.

[1241] 대조군 서열분석 어댑터 제조

[1242] W02016/055777에 기재된 폴리뉴클레오티드 결합 단백질의 변이체를 변이체 위치 각각에 대한 대조군으로 사용하였다. 대조군 폴리뉴클레오티드 결합 단백질을 정제하고 위에 기재된 바와 같이 서열분석 어댑터에 로딩하였다.

대조군 폴리뉴클레오티드 결합 단백질 결합된 서열분석 어댑터를 음이온 교환 컬럼에서 또는 Sera-Mag SpeedBeads(Thermo Scientific)를 사용하여 정제하였다. 이들은 대조군 서열분석 어댑터였다.

[1243] 라이브러리 제조

[1244] 변이체 서열분석 어댑터 및 대조군 서열분석 어댑터를 바코드가 있는 3.6 kb 분석물에 결합시키고, 각 변이체를 상이한 바코드에 결합시키고, 각 대조군을 상이한 바코드에 결합시켰다. 변이체 라이브러리 및 대조군 라이브러리를 풀링하였다. 풀링된 라이브러리를 서열분석을 위해 제조하고 제조업체의 지침(Oxford Nanopore Technologies)에 따라 MinION 유동 셀에서 실행하였다. 최대 6개의 변이체를 그들의 대조군과 함께 단일 MinION 유동 셀에서 실행하였으며, 이 대조군은 내부 유동 셀 대조군이였다.

[1245] 전기적 측정

[1246] 전기적 측정은 FLO-MIN106 MinION 유동 셀 및 GridION(Oxford Nanopore Technologies)에서 획득하였다. 표준 서열분석 스트림을 실행하고 원시 데이터를 MinKNOW 소프트웨어(Oxford Nanopore Technologies)를 사용하여 대량 FAST5 파일에 수집하였다.

[1247] 데이터 분석

[1248] DNA 라이브러리를 Guppy(Oxford Nanopore Technologies)의 표준 염기호출 알고리즘을 사용하여 서열분석하였다. 서열분석된 판독을 Guppy(Oxford Nanopore Technologies)를 사용하여 디멀티플렉싱하였다. 각 변이체 및 대조군에 대해 500개 이상의 판독을 3.6 kb 분석물 참조 서열에 맵핑하였다. 속도 및 정규화된 속도 분포 델타는 실시예 1에 기재된 바와 같이 계산하였다.

변이체 위치	변이체 AA	속도 Δ
177	D	3.07
177	E	58.70
177	F	-33.41
177	G	-27.62
177	H	10.64
177	I	42.98
177	L	51.39
177	M	48.31
177	N	2.40
177	Q	59.68
177	R	21.69
177	S	-31.10
177	T	22.62
177	V	-16.00
177	W	-23.42
177	Y	-43.93

[1249]

[1250] 표 13. 속도 Δ 는 위치 177에서의 변이체와 내부 유동 셀 대조군 사이의 중간 속도(bps) 차이이다.

변이체 위치	변이체 AA	속도 Δ
114	A	13.80
114	G	-11.30
114	I	93.69
114	L	-42.01
114	M	81.76
114	P	84.49
114	S	-7.28
114	T	-35.67
114	V	26.04

[1251]

[1252] 표 14. 속도 Δ 는 위치 114에서의 변이체와 내부 유동 셀 대조군 사이의 중간 속도(bps) 차이이다.

변이체 위치	변이체 AA	속도 Δ
358	E	-53.89
358	I	-40.19
358	L	-42.66
358	M	-48.58

[1253]

[1254] 표 15. 속도 Δ 는 위치 358에서의 변이체와 내부 유동 셀 대조군 사이의 중간 속도(bps) 차이이다.

변이체 위치	변이체 AA	정확도 Δ
177	F	-0.08
177	H	-0.2
177	I	-0.14
177	L	-0.08
177	M	-0.09
177	N	-0.66
177	W	-0.46

[1255]

[1256] 표 16. 정확도 Δ 는 177 위치에서의 변이체와 및 내부 유동 셀 대조군 사이의 중간 정확도(%) 차이이다.

변이체 위치	변이체 AA	정확도 Δ
114	G	0.13
114	I	-0.76
114	P	-0.70

[1257]

[1258] 표 17. 정확도 Δ 는 114 위치에서의 변이체 및 내부 유동 셀 대조군 사이의 중간 정확도(%) 차이이다.

변이체 위치	변이체 AA	정확도 Δ
358	I	0.96
358	M	0.94

[1259]

[1260] 표 18. 정확도 Δ 는 358 위치에서의 변이체와 내부 유동 셀 대조군 사이의 중간 정확도(%) 차이이다.

변이체 위치	변이체 AA	정규화된 속도 분포 Δ
177	F	-8.19
177	H	-5.38
177	I	-4.87
177	L	-4.44
177	M	-1.06
177	N	0.85
177	W	-4.42

[1261]

[1262] 표 19. 정규화된 속도 분포 Δ 는 177 위치에서의 변이체와 내부 유동 셀 대조군 사이의 정규화된 속도 분포 차이이다.

변이체 위치	변이체 AA	정규화된 속도 분포 Δ
114	G	-4.99
114	I	-8.23
114	P	-7.45

[1263]

[1264] 표 20. 정규화된 속도 분포 Δ 는 114 위치에서의 변이체와 내부 유동 셀 대조군 사이의 정규화된 속도 분포 차이이다.

변이체 위치	변이체 AA	정규화된 속도 분포 Δ
358	I	-0.86
358	M	0.45

[1265]

[1266] 표 21. 정규화된 속도 분포 Δ 는 358 위치에서의 변이체와 내부 유동 셀 대조군 사이의 정규화된 속도 분포 차이이다.

[1267] 실시예 3

[1268] 이 실시예는 W02019/002893에 기재된 나노기공을 사용하여 폴리뉴클레오티드의 이동을 제어하는 대조군 변이체 폴리뉴클레오티드 결합 단백질에 대한 폴리뉴클레오티드의 이동을 제어하는 변이체 폴리뉴클레오티드 결합 단백질의 속도 및 정확도 특성을 비교하는 방법을 기재한다.

[1269] 바코드가 있는 3.6 kb 분석물 제조

[1270] 이중 가닥 3.6 kb DNA 분석물을 특이적 프라이머 및 PCR을 사용하여 제조하였다. PCR 생성물을 NEBNext 단부 복구 및 NEBNext dA-테일링 모듈(New England Biolabs(NEB))에 적용하여 3' dA 돌출부를 생성하였다. 바코드는 제조업체의 지침에 따라 EXP-PBC001(Oxford Nanopore Technologies)을 사용하여 분석물에 도입하였다.

[1271] 변이체 서열분석 어댑터 제조

[1272] N-말단 친화성 및 용해성 태그를 갖는 W02016/055777에 기재된 폴리뉴클레오티드 결합 단백질의 변이체를 암호화하는 제조합 발현 벡터를 화학적으로 적격한 이.콜라이 세포로 형질전환하였다. 세포를 항생제를 함유하는 LB 아가 플레이트에 플레이트하였다. 아가 플레이트로부터의 콜로니를 LB 성장 배지에 접종하고, OD 0.400 - 0.800 으로 성장시키고 유도한 다음, 18℃에서 추가 16시간 동안 성장시켰다. 세포를 리소자임, 벤조나제 및 프로테아제 억제제의 존재 하에 버그부스터 추출 시약(Merck 70921)으로 용해시키거나 벤조나제 및 프로테아제 억제제의 존재 하에 초음파처리로 용해시켰다. 상청액을 친화성 크로마토그래피를 사용하여 추가로 정제하였다. 물 초과량의 정제된 변이체 폴리뉴클레오티드 결합 단백질을 주위 온도에서 10분 동안 50 mM Hepes pH8.0, 100mM 칼륨 아세테이트, 1mM EDTA에서 WO 2015/110813A1(그 전문이 본원에 참조로 포함됨)에 기재된 서열분석 Y 어댑터에 결합시켰다. TMAH(SIGMA)를 100 μ M의 최종 농도가 되도록 첨가하고 34℃에서 1시간 동안 인큐베이션하였다. 10 mM ATP, 10 mM MgCl₂ 및 0.5 M NaCl을 첨가하고 주위 온도에서 10분 동안 인큐베이션하였다. 변이체 폴리뉴클레오티드 결합 단백질 결합된 서열분석 어댑터를 Sera-Mag SpeedBeads(Thermo Scientific)를 사용하여 정제하였다. 이들은 변이체 서열분석 어댑터였다.

[1273] 대조군 서열분석 어댑터 제조

[1274] W02016/055777에 기재된 폴리뉴클레오티드 결합 단백질의 변이체를 변이체 위치 각각에 대한 대조군으로 사용하였다. 대조군 폴리뉴클레오티드 결합 단백질을 정제하고 위에 기재된 바와 같이 서열분석 어댑터에 로딩하였다. 이들은 대조군 서열분석 어댑터였다.

[1275] 라이브러리 제조

[1276] 변이체 서열분석 어댑터 및 대조군 서열분석 어댑터를 바코드가 있는 3.6 kb 분석물에 결합시키고, 각 변이체를 상이한 바코드에 결합시키고, 각 대조군을 상이한 바코드에 결합시켰다. 변이체 라이브러리 및 대조군 라이브러리를 풀링하였다. 풀링된 라이브러리를 서열분석을 위해 제조하고 제조업체의 지침(Oxford Nanopore Technologies)에 따라 MinION 유동 셀에서 실행하였다. 최대 6개의 변이체를 그들의 변이체와 함께 단일 MinION 유동 셀에서 실행하였으며, 이 대조군은 내부 유동 셀 대조군이었다.

[1277] 전기적 측정

[1278] 전기적 측정은 FLO-MIN112 MinION 유동 셀 및 GridION(Oxford Nanopore Technologies)에서 획득하였다. 표준 서열분석 스트립트를 실행하고 원시 데이터를 MinKNOW 소프트웨어(Oxford Nanopore Technologies)를 사용하여

대량 FAST5 파일에 수집하였다.

[1279] 데이터 분석

[1280] DNA 라이브러리를 Guppy(Oxford Nanopore Technologies)의 표준 염기호출 알고리즘을 사용하여 서열분석하였다. 서열분석된 판독을 Guppy(Oxford Nanopore Technologies)를 사용하여 디멀티플렉싱하였다. 각 변이체 및 대조군에 대해 500개 이상의 판독을 3.6 kb 분석물 참조 서열에 맵핑하였다. 속도 및 정규화된 속도 분포 델타는 실시예 1에 기재된 바와 같이 계산하였다.

변이체 위치	변이체 AA	속도 Δ
350	I	2.22
351	I	-9.76
351	Q	-5.61
354	A	11.26
354	Q	0.09
358	A	-0.99
358	E	-7.28
358	F	-10.93
358	I	-5.51
358	M	-11.75
358	S	1.60
361	I	21.60
361	Q	13.79

[1281]

[1282] 표 22. 속도 Δ 는 변이체 위치와 내부 유동 셀 대조군 사이의 중간 속도(bps) 차이이다.

변이체 위치	변이체 AA	정확도 Δ
350	I	-0.50
351	I	-3.36
351	Q	-0.89
354	A	-1.96
354	Q	-1.43
358	A	1.07
358	E	0.84
358	F	1.36
358	I	1.05
358	M	0.81
358	S	0.86
361	I	-4.62
361	Q	-1.81

[1283]

[1284] 표 23. 정확도 Δ 는 변이체 위치와 내부 유동 셀 대조군 사이의 중간 정확도(%) 차이이다.

변이체 위치	변이체 AA	정규화된 속도 분포 Δ
350	I	-0.52
351	I	2.25
351	Q	0.54
354	A	-0.44
354	Q	-0.53
358	A	-29.70
358	E	-26.60
358	F	0.73
358	I	-5.31
358	M	-0.56
358	S	-1.56
361	I	0.51
361	Q	0.07

[1285]

[1286] 표 24. 정규화된 속도 분포 Δ 는 변이체 위치와 내부 유동 셀 대조군 사이의 정규화된 속도 분포 차이이다.

[1287] 실시예 4

[1288] 이 실시예는 변이체 나노기공을 사용하여 폴리뉴클레오타이드의 이동을 제어하는 대조군 변이체 폴리뉴클레오타이드 결합 단백질에 대한 폴리뉴클레오타이드의 이동을 제어하는 변이체 폴리뉴클레오타이드 결합 단백질의 속도 및 정확도 특성을 비교하는 방법을 기재한다.

[1289] 바코드가 있는 3.6 kb 분석물 제조

[1290] 이중 가닥 3.6 kb DNA 분석물을 특이적 프라이머 및 PCR을 사용하여 제조하였다. PCR 생성물을 NEBNext 단부 복구 및 NEBNext dA-테일링 모듈(New England Biolabs(NEB))에 적용하여 3' dA 돌출부를 생성하였다. 바코드를 제조업체의 지침에 따라 EXP-PBC001(Oxford Nanopore Technologies)을 사용하여 분석물에 도입하였다.

[1291] 변이체 나노기공 제조

[1292] W02019002893에 기재된 바와 같이 변이체 나노기공 또는 대조군 나노기공을 암호화하는 재조합 발현 벡터를 화학적으로 적격한 이. 콜라이 세포로 형질전환하였다. 작제물은 C-말단 Strep 친화성 태그를 포함하였다. 세포를 선택을 위한 적절한 항생제를 함유하는 LB 아가 플레이트에 플레이팅하였다. 아가 플레이트로부터의 콜로니를 항생제를 함유한 LB 배지에 접종하고 밤새 성장시킨 후 자가유도 배지로 희석하고 18 C에서 68시간 동안 인큐베이션하였다. 세포를 원심분리를 통해 수확한 후 용해하고 버그부스터(Merck 70921) 및 0.1% DDM에서 추출하였다. 상청액을 친화성 크로마토그래피, 60°C에서 30분 동안 열 처리 및 크기 배제 크로마토그래피를 사용하여 정제하여 SDS-PAGE에 의해 판단된 바와 같이 올리고머 나노기공을 선택하였다.

[1293] CsgG-CsgF 복합체를 위에서와 같이 정제된 나노기공 및 화학적으로 합성된 CsgF 펩티드로부터 제조하였다. 나노기공을 50mM Tris, 150mM NaCl, 2mM EDTA, 0.1% SDS, 0.1% Brij58, pH7.0을 포함하는 완충액으로 완충액 교환한 다음 25°C에서 1시간 동안 CsgG 단량체에 대한 4x 물 과량의 펩티드에서 인큐베이션하고, 반응은 60°C에서 15분 동안 가열하여 중단한 후 원심분리하여 임의의 침전물을 제거하였다.

[1294] 변이체 서열분석 어댑터 제조

[1295] N-말단 친화성 및 용해성 태그를 갖는 W02016/055777에 기재된 폴리뉴클레오타이드 결합 단백질의 변이체를 암호화하는 재조합 발현 벡터를 화학적으로 적격한 이.콜라이 세포로 형질전환하였다. 세포를 항생제를 함유하는 LB 아가 플레이트에 플레이팅하였다. 아가 플레이트로부터의 콜로니를 LB 성장 배지에 접종하고, OD 0.400 - 0.800 으로 성장시키고 유도한 다음, 18°C에서 추가 16시간 동안 성장시켰다. 세포를 리소자임, 벤조나제 및 프로테아제 억제제의 존재 하에 버그부스터 추출 시약(Merck 70921)으로 용해시키거나 벤조나제 및 프로테아제 억제제의 존재 하에 초음파처리로 용해시켰다. 상청액을 친화성 크로마토그래피를 사용하여 추가로 정제하였다. 물 초과량의 정제된 변이체 폴리뉴클레오타이드 결합 단백질을 주위 온도에서 10분 동안 50 mM Hepes pH8.0, 100mM 칼륨

아세테이트, 1mM EDTA에서 WO 2015/110813(그 전문이 본원에 참조로 포함됨)에 개시된 바와 같은 서열분석 Y 어댑터에 결합시켰다. TMAD(SIGMA)를 100 μ M의 최종 농도가 되도록 첨가하고 34℃에서 1시간 동안 인큐베이션하였다. 10 mM ATP, 10 mM MgCl₂ 및 0.5 M NaCl을 첨가하고 주위 온도에서 10분 동안 인큐베이션하였다. 변이체 폴리뉴클레오티드 결합 단백질 결합된 서열분석 어댑터를 Sera-Mag SpeedBeads(Thermo Scientific)를 사용하여 정제하였다. 이들은 변이체 서열분석 어댑터였다.

[1296] 대조군 서열분석 어댑터 제조

[1297] WO2016/055777에 기재된 폴리뉴클레오티드 결합 단백질의 변이체를 변이체 위치 각각에 대한 대조군으로 사용하였다. 대조군 폴리뉴클레오티드 결합 단백질을 정제하고 위에 기재된 바와 같이 서열분석 어댑터에 로딩하였다. 대조군 폴리뉴클레오티드 결합 단백질 결합된 서열분석 어댑터를 음이온 교환 컬럼에서 또는 Sera-Mag SpeedBeads(Thermo Scientific)를 사용하여 정제하였다. 이들은 대조군 서열분석 어댑터였다.

[1298] 라이브러리 제조

[1299] 변이체 서열분석 어댑터 및 대조군 서열분석 어댑터를 바코드가 있는 3.6 kb 분석물에 결합시키고, 각 변이체를 상이한 바코드에 결합시키고, 각 대조군은 상이한 바코드에 결합시켰다. 변이체 라이브러리 및 대조군 라이브러리를 풀링하였다. 풀링된 라이브러리를 서열분석을 위해 제조하고 제조업체의 지침(Oxford Nanopore Technologies)에 따라 MinION 유동 셀에서 실행하였다. 최대 6개의 변이체를 그들의 대조군과 함께 단일 MinION 유동 셀에서 실행하였으며, 이 대조군은 내부 유동 셀 대조군이었다.

[1300] 전기적 측정

[1301] 전기적 측정은 MinION 유동 셀에 삽입된 나노기공에서 획득하였다. 단일 기공을 각 채널 내의 블록 공중합체 막에 삽입한 후, 25 mM 칼륨 포스페이트, 150 mM 칼륨 페로시아나이드(II), 150 mM 칼륨 페로시아나이드(III), pH 8.0을 함유하는 완충액 1 mL를 시스템을 통해 흘려보내 임의의 과도한 단백질 나노기공을 제거하였다.

[1302] 표준 서열분석 스트립트를 실행하고 원시 데이터를 MinKNOW 소프트웨어(Oxford Nanopore Technologies)를 사용하여 대량 FAST5 파일에 수집하였다.

[1303] 데이터 분석

[1304] DNA 라이브러리를 Guppy(Oxford Nanopore Technologies)의 표준 염기호출 알고리즘을 사용하여 서열분석하였다. 각 나노기공 변이체에 대해 145개 이상의 관독을 3.6 kb 분석물 참조 서열에 맵핑하였다. 속도 및 정규화된 속도 분포 텔타는 실시예 1에 기재된 바와 같이 계산하였다.

나노기공 변이체 위치	나노기공 변이체 AA	변이체 폴리뉴클레오티드 결합 단백질 위치	변이체 폴리뉴클레오티드 결합 단백질 변이체 AA	속도 Δ
100	A	350	I	-0.75
100	A	351	I	-11.90
100	A	351	Q	-3.05
100	A	354	A	-3.11
100	A	354	Q	-4.27
100	A	358	I	-9.09
100	A	361	I	8.41
100	A	361	Q	1.36
102	A	350	I	-7.29
102	A	350	S	1.75
102	A	351	I	-12.19
102	A	351	Q	-8.71
102	A	354	A	7.82
102	A	354	Q	1.54
102	A	358	I	-8.99
102	A	358	L	-5.86
102	A	358	Q	-1.89
102	A	361	I	14.99
102	A	361	Q	8.76
102	S	350	I	4.49
102	S	351	I	-9.90
102	S	351	Q	-6.19
102	S	354	A	7.74
102	S	354	Q	3.20
102	S	358	I	-12.91
102	S	361	I	13.93
102	S	361	Q	7.43

[1305]

[1306]

표 25. 속도 Δ 는 변이체 폴리뉴클레오티드 결합 단백질과 변이체 나노기공의 내부 유동 셀 대조군 사이의 중간 속도(bps) 차이이다.

나노기공 변이체 위치	나노기공 변이체 AA	변이체 폴리뉴클레오티드 결합 단백질 위치	변이체 폴리뉴클레오티드 결합 단백질 변이체 AA	정확도 Δ
100	A	350	I	-0.03
100	A	351	I	-2.65
100	A	351	Q	-1.32
100	A	354	A	-1.06
100	A	354	Q	-1.24
100	A	358	I	0.89
100	A	361	I	-2.99
100	A	361	Q	-1.73
102	A	350	I	-0.81
102	A	350	S	-0.34
102	A	351	I	-4.42
102	A	351	Q	-1.60
102	A	354	A	-3.22
102	A	354	Q	-1.59
102	A	358	I	1.00
102	A	358	L	0.97
102	A	358	Q	0.94
102	A	361	I	-4.12
102	A	361	Q	-2.47
102	S	350	I	-0.36
102	S	351	I	-4.40
102	S	351	Q	-1.24
102	S	354	A	-2.33
102	S	354	Q	-1.43
102	S	358	I	1.17
102	S	361	I	-4.00
102	S	361	Q	-1.93

표 26. 정확도 Δ 는 변이체 폴리뉴클레오티드 결합 단백질과 변이체 나노기공의 내부 유동 셀 대조군 사이의 중간 정확도(%) 차이이다.

나노기공 변이체 위치	나노기공 변이체 AA	변이체 폴리뉴클레오티드 결합 단백질 위치	변이체 폴리뉴클레오티드 결합 단백질 변이체 AA	정규화된 속도 분포 Δ
100	A	350	I	0.27
100	A	351	I	1.96
100	A	351	Q	0.57
100	A	354	A	0.54
100	A	354	Q	0.75
100	A	358	I	0.30
100	A	361	I	0.32
100	A	361	Q	0.35
102	A	350	I	3.65
102	A	350	S	1.25
102	A	351	I	2.25
102	A	351	Q	1.29
102	A	354	A	0.17
102	A	354	Q	-0.04
102	A	358	I	0.67
102	A	358	L	0.82
102	A	358	Q	1.35
102	A	361	I	1.04
102	A	361	Q	0.75
102	S	350	I	-0.05
102	S	351	I	2.78
102	S	351	Q	1.41
102	S	354	A	-0.11
102	S	354	Q	0.04
102	S	358	I	-0.55
102	S	361	I	1.05
102	S	361	Q	0.96

[1309]

[1310] 표 27. 정규화된 속도 분포 Δ 는 변이체 폴리뉴클레오티드 결합 단백질과 변이체 나노기공의 내부 유동 셀 대조군 사이의 정규화된 속도 분포 차이이다.

[1311]

실시예 5

[1312]

이 실시예는 나노기공을 사용하여 폴리뉴클레오티드의 이동을 제어하는 대조군 변이체 폴리뉴클레오티드 결합 단백질에 대한 폴리뉴클레오티드의 이동을 제어하는 변이체 폴리뉴클레오티드 결합 단백질의 속도 및 정확도 특성을 비교하는 방법을 기재한다.

[1313]

바코드가 있는 3.6 kb 분석물 제조

[1314]

이중 가닥 3.6 kb DNA 분석물을 특이적 프라이머 및 PCR을 사용하여 제조하였다. PCR 생성물을 NEBNext 단부 복구 및 NEBNext dA-테일링 모듈(New England Biolabs(NEB))에 적용하여 3' dA 돌출부를 생성하였다. 바코드는 제조업체의 지침에 따라 EXP-PBC001(Oxford Nanopore Technologies)을 사용하여 분석물에 도입하였다.

[1315]

나노기공 제조

[1316]

나노기공은 실시예 4에 기재된 바와 같이 제조하였다.

[1317]

변이체 서열분석 어댑터 제조

[1318]

N-말단 친화성 및 용해성 태그가 있는 W02016/055777에 기재된 폴리뉴클레오티드 결합 단백질의 변이체를 암호화하는 재조합 발현 벡터를 화학적으로 적격한 이.콜라이 세포로 형질전환하였다. 세포를 항생제를 함유하는 LB 아가 플레이트에 플레이트하였다. 아가 플레이트로부터의 콜로니를 LB 성장 배지에 접종하고, OD 0.400 - 0.800 으로 성장시키고 유도한 다음, 18℃에서 추가 16시간 동안 성장시켰다. 세포를 리소자임, 벤조나제 및 프로테아제 억제제의 존재 하에 버그부스터 추출 시약(Merck 70921)으로 용해시켰다. 상청액을 친화성 크로마토그래피를 사용하여 추가로 정제하였다. 변이체 폴리뉴클레오티드 결합 단백질을 주위 온도에서 10분 동안 1x 완충 BXT(IBA Lifesciences GmbH)에서 WO 2015/110813A1(그 전문이 본원에 참조로 포함됨)에 기재된 서열분석 Y 어

댁터에 결합시켰다. TMAD(SIGMA)를 100 μ M의 최종 농도가 되도록 첨가하고 34℃에서 1시간 동안 인큐베이션하였다. 10 mM ATP, 10 mM MgCl₂ 및 0.5 M NaCl을 첨가하고 주위 온도에서 10분 동안 인큐베이션하였다. 변이체 폴리뉴클레오타이드 결합 단백질 결합된 서열분석 어댑터를 Sera-Mag SpeedBeads(Thermo Scientific)를 사용하여 정제하였다. 이들은 변이체 서열분석 어댑터였다.

[1319] 대조군 서열분석 어댑터 제조

[1320] WO2016/055777에 기재된 폴리뉴클레오타이드 결합 단백질의 변이체를 변이체 위치 각각에 대한 대조군으로 사용하였다. 대조군 폴리뉴클레오타이드 결합 단백질을 정제하고 위에 기재된 바와 같이 서열분석 어댑터에 로딩하였다. 이들은 대조군 서열분석 어댑터였다

[1321] 라이브러리 제조

[1322] 변이체 서열분석 어댑터 및 대조군 서열분석 어댑터를 바코드가 있는 3.6 kb 분석물에 결합시키고, 각 변이체를 상이한 바코드에 결합시키고, 각 대조군을 상이한 바코드에 결합시켰다. 변이체 라이브러리 및 대조군 라이브러리를 풀링하였다. 풀링된 라이브러리를 서열분석을 위해 제조하고 제조업체의 지침(Oxford Nanopore Technologies)에 따라 MinION 유동 셀에서 실행하였다. 최대 46개의 변이체를 그들의 대조군과 함께 단일 PromethION 유동 셀에서 실행하였으며, 이 대조군은 내부 유동 셀 대조군이다. 라이브러리 당 적어도 2개의 유동셀을 실행하였다.

[1323] 전기적 측정

[1324] 전기적 측정은 PromethION 유동 셀에 삽입된 나노기공에서 획득하였다. 단일 기공을 각 채널 내의 블록 공중합체 막에 삽입한 후, 25 mM 칼륨 포스페이트, 150 mM 칼륨 페로시아나이드(II), 150 mM 칼륨 페리시아나이드(III), pH 8.0를 포함하는 완충액 1 mL를 시스템을 통해 흘려보내 임의의 과도한 단백질 나노기공을 제거하였다.

[1325] 표준 서열분석 스트립트를 실행하고 원시 데이터를 MinKNOW 소프트웨어(Oxford Nanopore Technologies)를 사용하여 대량 FAST5 파일에 수집하였다.

[1326] 데이터 분석

[1327] DNA 라이브러리를 Guppy(Oxford Nanopore Technologies)의 표준 염기호출 알고리즘을 사용하여 서열분석하였다. 서열분석된 판독을 Guppy(Oxford Nanopore Technologies)를 사용하여 디멀티플렉싱하였다. 각 변이체 및 대조군에 대해 20,000개 이상의 판독을 3.6 kb 분석물 참조 서열에 맵핑하였다. 속도 및 정규화된 속도 분포 델타는 실시예 1에 기재된 바와 같이 계산하였다.

변이체 위치	변이체 AA	속도 Δ
55	K	8.29
156	F	13.84
210	K	0.52
221	E	4.37
350	E	6.06

[1328]

[1329] 표 28. 속도 Δ 는 변이체 위치와 내부 유동 셀 대조군 사이의 중간 속도(bps) 차이이다.

변이체 위치	변이체 AA	정확도 Δ
55	K	0.15
156	F	0.26
210	K	0.45
221	E	0.23
350	E	0.48

[1330]

[1331] 표 29. 정확도 Δ 는 변이체 위치와 내부 유동 셀 대조군 사이의 중간 정확도(%) 차이이다.

변이체 위치	변이체 AA	정규화된 속도 분포 Δ
55	K	-0.20
156	F	-0.67
210	K	-0.85
221	E	-0.94
350	E	-1.02

[1332]

[1333] 표 30. 정규화된 속도 분포 Δ 는 변이체 위치와 내부 유동 셀 대조군 사이의 정규화된 속도 분포 차이이다.

서 열 목 록 (첨부)



아이콘을 클릭하시면 서열목록 파일이 열립니다.

본 공보 PDF는 첨부파일을 가지고 있습니다. Acrobat Reader PDF뷰어를 제공하지 않는 브라우저(크롬, 파이어폭스, 사파리 등)의 경우 첨부파일 열기가 제한되어 있으므로 Acrobat Reader PDF뷰어 설치 후 공보 PDF를 다운로드 받아 해당 뷰어에서 조회해주시기 바랍니다.