(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년01월12일
(11) 등록번호 10-2201867
(24) 등록일자 2021년01월06일

(51) 국제특허분류(Int. Cl.)
C07K 19/00 (2006.01) *A01K 67/027* (2006.01)
A61K 48/00 (2006.01) *C12N 15/62* (2006.01)
C12N 15/63 (2006.01)
(52) CPC특허분류
C07K 19/00 (2013.01)
A01K 67/027 (2013.01)
(21) 출원번호 10-2015-7008001
(22) 출원일자(국제) 2013년09월04일
심사청구일자 2018년08월31일
(85) 번역문제출일자 2015년03월27일
(65) 공개번호 10-2015-0070120
(43) 공개일자 2015년06월24일
(86) 국제출원번호 PCT/US2013/058100
(87) 국제공개번호 WO 2014/039585
국제공개일자 2014년03월13일
(30) 우선권주장
61/696,689 2012년09월04일 미국(US)
(뒷면에 계속)
(56) 선행기술조사문헌
WO2011146121 A1
WO2011072246 A2
WO2012033462 A1
KR1020090031938 A

(73) 특허권자
더 스크립스 리서치 인스티튜트
미국 캘리포니아주 92037 라 졸라 노스 토리 파인 로드 10550
(72) 발명자
바바스 3세 카를로스 에프.
미국, 캘리포니아 92037, 라 호야, 캔들라이트 드라이브 5416
메르세르 앤드류
미국, 메릴랜드 20837, 폴스빌, 웨스털리 에비뉴 19932
(뒷면에 계속)
(74) 대리인
특허법인인벤싱크

전체 청구항 수 : 총 15 항 심사관 : 최성호

(54) 발명의 명칭 **표적화된 바인딩 특이도를 갖는 키메라 폴리펩타이드들**

(57) 요 약

형질전환 세포들, 조직들, 식물들, 및 동물들의 생성을 위한, 키메라 폴리펩타이드들, 이의 조성물, 이의 발현 벡터 및 이의 사용 방법이 본 명세서에서 개시된다. 본 발명의 조성물, 벡터 및 방법들은 또한 유전체 치료 기술에서 유용하다.

(52) CPC특허분류
A61K 48/00 (2013.01)
C12N 15/62 (2013.01)
C12N 15/63 (2013.01)

(72) 발명자
램 브라이언 엠.
미국, 캘리포니아 92117, 샌 디에고, 카울리 웨이 3144, 에이피티. 3
가이 토마스
미국, 캘리포니아 92075, 솔라나 비치, 사우스 헬럭스 에비뉴 210, #비

(30) 우선권주장
61/753,763 2013년01월17일 미국(US)
61/818,364 2013년05월01일 미국(US)

명 세 서

청구범위

청구항 1

분리된 폴리펩타이드 (isolated polypeptide) 로서,

Xanthomonas로부터 유도된 합성 TALE (transcription activator-like effector) 단백질을 포함하며,

상기 Xanthomonas로부터 유도된 TALE 단백질은, N_{-1} 헤어핀 (hairpin) 에서,

Q는 Y로 변이, Q는 S로 변이, Q는 R로 변이, W는 R로 변이, W는 G로 변이, W 삭제, S는 R로 변이, S는 H로 변이, S는 A로 변이, S는 N으로 변이, 및 S는 T로 변이로부터 선택된 하나 이상의 변이들을 포함하는 SEQ ID NO: 3 (VGKQWSGARAL) 로 기재된 아미노산 서열을 포함하는 N-말단 도메인 (NTD) 을 갖는, 분리된 폴리펩타이드.

청구항 2

제 1 항에 있어서,

상기 하나 이상의 변이들을 포함하는 SEQ ID NO: 3 (VGKQWSGARAL) 로 기재된 아미노산 서열은 VGKYRGARAL (SEQ ID NO: 4), VGKSRSGARAL (SEQ ID NO: 5), VGKYHGARAL (SEQ ID NO: 6), 및 VGKRGAGARAL (SEQ ID NO: 7) 로부터 선택된 것인, 분리된 폴리펩타이드.

청구항 3

분리된 폴리펩타이드로서,

Ralstonia로부터 유도된 합성 TALE 단백질을 포함하며,

상기 Ralstonia로부터 유도된 TALE 단백질은, N_{-1} 헤어핀 (hairpin) 에서,

R_1는 K로 변이, Q는 Y로 변이, Q는 S로 변이, Q는 R로 변이, R_2는 W로 변이, R_2는 G로 변이, R_2 삭제, S는 R로 변이, S는 H로 변이, S는 A로 변이, S는 N으로 변이, 및 S는 T로 변이로부터 선택된 하나 이상의 변이들을 포함하는 SEQ ID NO: 8 (IVDIAR$_1$QR$_2$SGDLA) 로 기재된 아미노산 서열을 포함하는 NTD를 갖는, 분리된 폴리펩타이드.

청구항 4

제 3 항에 있어서,

상기 하나 이상의 변이들을 포함하는 SEQ ID NO: 8 (IVDIAR$_1$QR$_2$SGDLA) 로 기재된 아미노산 서열은, IVDIARQWSGDLA (SEQ ID NO: 9), IVDIARYRGDLA (SEQ ID NO: 10), IVDIARSRSGDLA (SEQ ID NO: 11), IVDIARYHGDLA (SEQ ID NO: 12), 및 IVDIARRGAGDLA (SEQ ID NO: 13) 로부터 선택된 것인, 분리된 폴리펩타이드.

청구항 5

제 1 항 내지 제 4 항 중 어느 한 항에 있어서,

리콤비나제 도메인 또는 뉴클레아제 도메인을 더 포함하는, 분리된 폴리펩타이드.

청구항 6

제 1 항 내지 제 4 항 중 어느 한 항에 기재된 분리된 펩타이드를 인코딩한 분리된 핵산 분자.

청구항 7

제 6 항에 기재된 분리된 핵산 분자를 포함하는 벡터.

청구항 8

제 6항에 기재된 분리된 핵산분자 또는, 상기 분리된 핵산분자를 포함하는 벡터로 형질전환되거나 형질주입된

숙주 세포.

청구항 9

TALE 단백질 N-말단 도메인 (NTD) 을 생성하는 방법으로서,

a) 상기 NTD의 N_{-1} 헤어핀 내의 하나 이상의 아미노산 잔기들을 변이 또는 삭제함으로써 상기 NTD의 아미노산 서열을 랜덤화하는 단계로서, 상기 아미노산 서열은 SEQ ID NO: 14 (VGKXXXGAR) 또는 SEQ ID NO: 15 (VDIAXXXXGDLA) 인, 상기 랜덤화하는 단계; 및

b) 상기 단계 a) 의 랜덤화된 TALE 단백질 NTD에 대하여 선택하는 단계로서, 상기 TALE 단백질 NTD는 목표된 뉴클레오타이드에 특이적으로 바인딩되거나 향상된 활성도를 보이는, 상기 선택하는 단계를 포함하는, TALE 단백질 NTD 생성 방법.

청구항 10

제 1 항 및 제 2 항 중 어느 한 항에 있어서,

상기 합성 TALE 단백질은 AcrXa7, Tal1c, 또는 PthXo1로부터 유도된, 분리된 폴리펩타이드.

청구항 11

제 1 항 및 제 2 항 중 어느 한 항에 있어서,

상기 합성 TALE 단백질은 합성 RVD 도메인 (domain) 을 포함하는, 분리된 폴리펩타이드.

청구항 12

제 11 항에 있어서,

상기 합성 TALE 단백질은 15 개의 RVD 반복들 (SEQ ID NO: 28의 아미노산 잔기들 311과 806간) 을 포함하는, 분리된 폴리펩타이드.

청구항 13

제 11 항에 있어서,

상기 TALE 단백질은 아미노산 잔기 120 또는 128에서 절단된 AvrXa7 N-말단, 및 아미노산 잔기 28에서 절단된 AvrXa7 C-말단을 포함하는, 분리된 폴리펩타이드.

청구항 14

제 5 항에 있어서,

상기 리콤비나제는 Gin, Hin, Tn3, Sin, Beta, Pin, Min, Din, 및 Cin 및 Gin의 돌연변이단백질들, Hin의 돌연변이단백질들, Sin의 돌연변이단백질들, Beta의 돌연변이단백질들, Pin의 돌연변이단백질들, Min의 돌연변이단백질들, Din의 돌연변이단백질들, Cin의 돌연변이단백질들 및 Tn3의 돌연변이단백질들로 구성된 그룹으로부터 선택된, 분리된 폴리펩타이드.

청구항 15

부위-특이적 재조합을 위한 방법으로서,

(a) 제 1 항에 기재된 폴리펩타이드와 특이적으로 상호작용하기 위한 적어도 2 개의 바인딩 부위들을 포함하는 DNA 서열을 제공하는 단계; 및

(b) 상기 폴리펩타이드와 상기 DNA 서열을 반응시키는 단계로서, 상기 폴리펩타이드는 상기 DNA 서열의 양 가닥들이 상기 폴리펩타이드와 특이적으로 상호작용하는 상기 2 개의 바인딩 부위들 간에서 절단되는 부위-특이적 재조합 이벤트에 촉매작용을 하는, 상기 반응시키는 단계를 포함하는, 부위-특이적 재조합을 위한 방법.

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

발명의 설명

기 술 분 야

[0001] 본 명세서는 2012년 9월 4일에 출원된 미국 가출원 번호 61/696,689, 2013년 1월 17일에 출원된 미국 가출원 번호 61/753,763, 2013년 5월 1일에 출원된 미국 가출원 번호 61/818,364에 대한 우선권을 35 U.S.C. §119(e) 하에서 주장하며, 이들 모든 문헌들은 그 전체가 본 명세서에서 참조로서 인용된다.

[0002] 본 발명은 전반적으로 바이오기술 분야 관한 것이며, 보다 구체적으로 특이적 DNA 서열들을 인식하는 키메라 리콤비나제들 (chimeric recombinases) 에 관한 것이다.

배 경 기 술

[0003] 다양한 단백질 도메인들이 서열-특이적 DNA 인식을 제공하도록 진화하였기 때문에, 서열-의존 방식으로 DNA를 인식하는 단백질들의 능력은 생명에 있어서 중심이 된다. 이러한 도메인들 중 선택된 몇몇에 의한 DNA 인식은 또한 다양한 바이오기술의 응용들의 기초이다. 특히, C_2H_2 ZFP들 (zinc-finger proteins:징크-핑거 단백질들) 은 사용자-규정된 DNA 서열들을 인식하도록 엔지니어링될 첫 번째 DNA-바인딩 단백질들 중에 있으며 많은 응용에 있어서 다양한 성공도로 사용되어 왔으며, 이러한 응용들은 전사 조절, 게놈 엔지니어링 (genome engineering) 및 후성적 개량 (epigenetic modification) 을 포함한다. ZFP들의 모듈식 어셈블리는 이러한 방식들을 실현하였다. 그러나, ZFP 기술의 진보 및 전망에도 불구하고, 특이적 서열들에 대한 특이적 고-친화도 ZFP들을 구성하는 것은 여전히 용이하지 않으며 선택된 경우들에서는 일반적인 연구소들에서는 쉽게 채택되지 않는 시간 소모적이고 노동 집약적인 선택 시스템들의 사용을 요구한다.

[0004] TALE (transcription activator-like effector) 도메인들은 ZFP 기술에 대한 잠재적 대안을 제공하는 자연적으로 발생하는 DNA-바인딩 도메인들 (DBD들) 의 부류이다. 식물 병원체 Xanthomonas에서 발견되는 TALE들은 표적 DNA 서열들에 선택적으로 바인딩하도록 기능하는 33 개 내지 35 개의 아미노산 리피트들 (repeats) 의 시리즈를 포함한다. 이러한 리피트들은 단일 뉴클레오티드로의 바인딩을 매개함으로써 DNA 특이도를 부여하는 2 개의 인접하는 RVD들 (repeat variable di-residues) 의 예외와 동일하다. 유사한 개수의 염기 쌍들 (bps) 의 DNA 부위들에 바인딩되는 30 개 이상의 리피트들의 어레이들이 기술되었다. 각 RVD 의 바인딩 시에 고유한 축중 (degeneracy) 이 존재하지만, 최근의 연구결과들은 합성 TALE 단백질들이 인간 게놈 내의 단일 좌위 (locus) 에 표적화하는데 충분하게 특이도가 있다고 알려왔다.

[0005] ZFN들 (zinc-finger nucleases: 징크-핑거 뉴클레아제들) 과 같은 키메라 뉴클레아제들에 의한 DNA DSB들 (double-strand breaks) 의 도입은 유전자 기능을 녹아웃 (knockout) 시키는데 사용되거나 외인적으로 부가된 DNA 존재 시에 표적화된 좌위에서의 카세트 통합을 구동시키는데 사용되었다. ZFN들은 지난 십 년에 걸쳐서 집

중적으로 연구되어 왔으며 일부 경우들에서는 유전자 치료를 위한 임상학적 사용으로 시도되었다. 최근에, 수많은 그룹들이 표적화된 게놈 에디팅 (editing) 을 위해서 TALEN들 (뉴클레아제들에 융합된 TALE DNA-바인딩 도메인들) 의 사용을 연구해오고 있다. 실제로, TALEN들이 DNA-바인딩 모듈성 (DNA-binding modularity) 측면에서 ZFN들에 비해서 이점을 가지고 있기 때문에 ZFN들로 하는 많은 연구들이 TALE 뉴클레아제들을 사용하여서 복제되었다. 그러나, ZFN들 및 TALEN들로 하는 인상적인 연구에도 불구하고, 이들의 안전성 및 특이도에 대해서는 여전히 의심이 남아 있다. 특히, 표적에서 벗어난 DSB의 가장 가능성이 있는 결과는 작은 삽입들 또는 삭제들의 도입이기 때문에 표적에서 벗어난 절단 이벤트들 (off-target cleavage events) 을 검출하는 것은 어렵다. 추가적으로, DSB들의 회복은 세포 타입에 따라서 상이한 세포 기구 (cell machinery) 에 의존한다.

[0006] 표적화된 유전자적 변이를 달성하기 위한 다른 방식은 SSR들 (site-specific recombinases:부위-특이적 리콤비나제들) 을 사용하는 것이다. SSR들, 예를 들어서 타이로신 리콤비나제들 Cre 및 Flp은 세포 내의 염색체 구조를 조작하는데 통상적으로 사용되는 중요한 분자적 생물학적 툴들이다. 이러한 효소들은 촉매작용을 조율하기 위해서 다수의 복잡한 단백질-단백질 상호작용 및 단백질-DNA 상호작용에 의존하기 때문에, SSR들은 우수한 표적 부위 특이도를 보인다. 그러나, 지금까지, 다수의 SSR들의 특이도를 변화시키는 것은 어렵다고 증명되었다. 레솔바제/인버타제 (resolvase/invertase) 타입의 세린 리콤비나제들은 게놈 엔지니어링에 있어서 타이로신 리콤비나제들을 여러모로 대신할 수 있다. 특성상, 이러한 효소들은 매우 높은 모듈 방식으로 재조합을 조율하는 다중-도메인 단백질 복합체들로서 기능한다. 그러나, 재조합을 위해서 보조 인자들 (accessory factors) 을 필요로 하지 않는 몇몇 세린 리콤비나제들의 변형체들이 확인되었다. 또한, 다수의 연구들이 세린 리콤비나제들의 원 (native) DBD들은 키메라 ZFR들 (징크-핑거 리콤비나제들) 을 생성하는데 맞춤형 ZFP들로 대체될 수 있다고 보고하였다. 원칙적으로는, 확장된 개수의 서열들을 인식할 수 있는 ZFR들이 생성될 수 있지만, 모든 가능한 DNA 트리플릿들을 인식할 수 있는 징크-핑거 도메인들의 부재는 이러한 효소들의 잠재적 모듈식 표적화 능력을 제한한다.

[0007] ZFR들은 세린 리콤비나제들의 레졸바제/인버타제 족으로부터 유도된 활성화된 촉매 도메인 및 거의 임의의 DNA 서열을 인식하도록 맞춤화될 수 있는 징크-핑거 DNA-바인딩 도메인으로 구성된다 (도 30a 참조). ZFR들은 리콤비나제 촉매 도메인에 의해서 인식된 중앙 20-bp 코어 서열을 측접하는 (flanking) 2 개의 역위된 ZFBS들 (징크-핑거 바인딩 부위들) 로 구성된 특이적 ZFR 표적 부위들 간의 재조합에 촉매 작용을 한다 (도 30b 참조). ZFN들 (징크-핑거 뉴클레아제들) 및 TALEN들 (TAL effector necleases) 과 대조하여서, ZFR들은 자율적으로 기능하며 세포상의 DNA 손상 반응 경로를 활성화시키지 않고서 인간 세포 및 마우스 세포 내의 전이 유전자들을 절개 및 통합시킬 수 있다. 그러나, 통상적인 부위-특이적 리콤비나제들에서와 같이, ZFR들의 응용은 리콤비나제 촉매 도메인에 의해서 부여된 서열 요건들에 의해서 제약을 받는데, 이러한 서열 요건들은 ZFR 표적 부위들이 원 세린 레졸바제/인버타제 재조합 부위로부터 유도된 20-bp 코어를 포함할 것을 요구한다.

[0008] 부위-특이적 DNA 재조합 시스템들, 예를 들어서 Cre-loxP, FLP-FRT 및 ^C31-att은 유전자 엔지니어링을 위한 강력한 툴들로서 출현하였다. 이러한 DNA 재배열들에 촉매작용을 하는 이러한 부위-특이적 리콤비나제들은 짧은 (30-bp 내지 40-bp) 서열들을 인식하고 DNA 분열, 가닥 교환 및 재-결찰 (re-ligation) 을 DNA 합성 또는 고에너지 코팩터 (high-energy cofactor) 를 요구하지 않는 메카니즘에 의해서 조율할 수 있다. 이러한 단순성은 연구자들로 하여금 놀라울 정도의 공간적 및 시간적 감도로 유전자 기능을 연구할 수 있게 한다. 그러나, 이러한 부위-특이적 리콤비나제들에 의해서 부여된 엄격한 서열 요건들은 인공적으로 도입된 재조합 부위들을 포함하는 세포들 및 유기체들에 이들을 이용하는 것을 제약하였다. 이러한 문제를 처리하기 위해서, 몇몇 리콤비나제들의 서열 특이도를 자연적으로 발생하는 DNA 서열들로 변경하는 연구가 계속되었다. 이러한 연구에도 불구하고, 복잡한 돌연변이유발 및 선택 전략들이 필요하고 재-엔지니어링된 리콤비나제 변형체들이 통상적으로 완화된 기질 특이도를 보인다는 사실로 인해서 이러한 기술의 광범위한 채택이 어려웠다.

[0009] 따라서, 특히 유전자 치료를 위해서 내생적 유전자의 표적화되고 부위-특이적인 재조합에 촉매작용을 하는 보다 일반화된 방법 및 이러한 표적화되고 부위-특이적인 재조합에 촉매작용을 하는 효소가 필요하다. 이는 특히 유전자 치료에 유용하지만, 분자 생물학에서의 수많은 다른 응용들을 찾을 수 있는데, 예를 들어서 산업용 미생물 및 농업용 식물 및 동물의 개량 시에 사용되고 유전자 복제에 사용될 수 있다.

발명의 내용

[0010] 형질전환 세포들, 조직들, 식물들, 및 동물들의 생성을 위한, 표적화된 키메라 폴리펩타이드들, 이의 조성물, 이의 발현 벡터 및 이의 사용 방법이 본 명세서에서 개시된다. 본 발명의 조성물, 벡터 및 방법들은 또한 유전

체 치료 기술에서 유용하다.

[0011] 일 양태에서, 본 발명은 키메라 폴리펩타이드를 제공한다. 키메라 폴리펩타이드는 a) 리콤비나제, 전사 인자 또는 뉴클레아제 또는 이들의 단편; 및 b) TALE (transcription activator-like effector) 단백질을 포함한다. 다양한 실시예들에서, TALE 단백질은 절단되고 C-말단 절단부 또는 N-말단 절단부를 포함한다. 실시예들에서, TALE 단백질은 AcrXa7, Tal1c, 또는 PthXo1이다. 실시예들에서, TALE 단백질은 SEQ ID NO: 2에 기재된 아미노산 서열을 일부 또는 전부 포함한다. 실시예들에서, TALE 단백질은 아미노산 잔기들 27 과 268 간에서, 아미노산 잔기들 92 와 134 간에서, 아미노산 잔기들 120 와 129 간에서, 아미노산 잔기들 74 와 147 간에서, 또는 아미노산 잔기들 87 과 120 간에서 절단된다. 일부 실시예들에서, TALE 단백질은 SEQ ID NO:2의 아미노산 잔기 28, 74, 87, 92, 95, 120, 124, 128, 129, 147 및 150에서 절단된다.

[0012] 다른 양태에서, 본 발명은 목표된 뉴클레오타이드와 특이적으로 바인딩하는 TALE 단백질 바인딩 도메인을 생성하는 방법을 제공한다. 이 방법은 a) RVD (repeat variable di-residue) 내의 또는 RVD의 1 내지 2 아미노산 잔기들 N-말단 또는 C-말단 내의 아미노산 잔기를 변이시킴으로써 TALE 단백질 바인딩 도메인의 아미노산 서열을 랜덤화하는 단계; 및 b) 단계 (a) 의 랜덤화된 TALE 단백질 바인딩 도메인에 대하여 선택하는 단계로서, TALE 단백질 바인딩 도메인은 목표된 뉴클레오타이드와 특이적으로 바인딩하는, 선택하는 단계를 포함한다.

[0013] 다른 양태에서, 본 발명은 단리된 폴리펩타이드를 제공하며 이 단리된 폴리펩타이드는 Xanthamonus 유도된 TALE 단백질을 포함하며, Xanthamonus 유도된 TALE 단백질은, Q는 Y로 변이, Q는 S로 변이, Q는 R로 변이, W는 R로 변이, W는 G로 변이, W 삭제, S는 R로 변이, S는 H로 변이, S는 A로 변이, S는 N으로 변이, 및 S는 T로 변이로부터 선택된 하나 이상의 변이들 또는 삭제들을 갖는 SEQ ID NO: 3 (VGKQWSGARAL) 에 기재된 아미노산 서열을 포함하는 N-말단 도메인 (NTD) 을 갖는다.

[0014] 다른 양태에서, 본 발명은 단리된 폴리펩타이드를 제공하며, 단리된 폴리펩타이드는 Ralstonia 유도된 TALE 단백질을 포함하며, Ralstonia 유도된 TALE 단백질은, R_1는 K로 변이, Q는 Y로 변이, Q는 S로 변이, Q는 R로 변이, R_2는 W로 변이, R_2는 G로 변이, R_2 삭제, S는 R로 변이, S는 H로 변이, S는 A로 변이, S는 N으로 변이, 및 S는 T로 변이로부터 선택된 하나 이상의 변이들 또는 삭제들을 갖는 SEQ ID NO: 8 ($IVDIAR_1QR_2SGDLA$) 에 기재된 아미노산 서열을 포함하는 NTD를 갖는다.

[0015] 다른 실시예에서, 본 발명은 TALE 단백질 N-말단 도메인 (NTD) 을 생성하는 방법을 제공한다. 이 방법은 a) NTD 내의 하나 이상의 아미노산 잔기들을 변이 또는 삭제함으로써 NTD의 아미노산 서열을 랜덤화하는 단계로서, 아미노산 서열은 SEQ ID NO: 14 (VGKXXXGAR) 또는 SEQ ID NO: 15 (VDIAXXXXGDLA) 인, 랜덤화하는 단계; 및 b) 단계 a) 의 랜덤화된 TALE 단백질 NTD에 대하여 선택하는 단계로서, TALE 단백질 NTD는 목표된 뉴클레오타이드에 특이적으로 바인딩되거나 향상된 활성도를 보이는, 선택하는 단계를 포함한다.

[0016] 또한, 세린 리콤비나제 및 하나 이상의 징크 핑거 바인딩 도메인들을 포함하는 키메라 단백질들, ZFR들을 생성하는 방법들, 이의 조성물들, 발현 벡터들, 및 이의 사용 방법들, 형질변환 세포들, 조직들, 식물들 및 동물들을 생성하기 위한 방법들이 본 명세서에서 개시된다. 본 발명의 조성물들, 벡터들 및 방법들은 또한 유전자 치료 기법에서도 유용하다.

[0017] 일 양태에서, 본 발명은 대응하는 와일드 타입 리콤비나제보다 큰 촉매성 특이도를 갖는 복수의 징크 핑거 리콤비나제 (ZFR) 단백질을 생성하는 방법을 제공한다. 이 방법은 a) Gin Ile120, Thr123, Leu127, Ile136 및 Gly137 또는 이들의 조합과 균등한 위치들에서 리콤비나제 촉매성 도메인에 대해서 랜덤 돌연변이유발을 수행하며 각 아미노산에 대해서 위치들 2 및 3에서 DNA를 변이시키는 단계; b) 단계 a)의 리콤비나제 촉매성 도메인을 복수의 징크 핑거 바인딩 도메인들과 융합시켜서 ZFR들을 형성하는 단계; 및 c) 대응하는 와일드 타입 리콤비나제보다 큰 촉매성 특이도를 갖게 단계 b)의 ZFR들에 대하여 농축을 시키는 단계를 포함한다. 실시예들에서, ZFR들은 GC, GT, CA, TT 및 AC로부터 선택된 DNA 표적들에 대하여 증가된 촉매적 활성도를 갖는다. 일 실시예에서, 리콤비나제 촉매성 도메인은 Ile136 및/또는 Gly137에서 돌이변이유발된다.

[0018] 다양한 양태들에서, 본 명세서에서 기술된 키메라 폴리펩타이드들은 본 명세서에서 개시된 바와 같이, 다음으로부터 유도되거나 다음으로부터 랜덤하게 돌연변이유발된 재조합 촉매성 도메인을 포함한다: a) EcoTn3로도 알려진 Tn3; StyHin로도 알려진 Hin; MuGin로도 알려진 Gin; Sin; Beta; Pin; Min; Din; Cin; EcoTn21; SfaTn917; BmeTn5083; Bme53; Cpe; SauSK1; SauSK41; SauTn552; Ran; Aac; L1a; pMER05; M1o92; M1o90; Rrh; Pje; Req; PpsTn5501; Pae; Xan; ISXc5; Spy; RhizY4cG; SarpNL1; SsolSC1904a; SsolSC1904b; SsoISC1913; Aam606; MjaM0014; Pab; HpylS607; MtulS_Y349; MtuRv2792c; MtuRv2979c; MtuRv3828c; MtuRv0921; MceRv0921; TnpX;

TndX; WwK; 락토콕칼 파지 TP901-1 세린 리콤비나제; S. 피오젠 파지 φ370.1 세린 리콤비나제; S. 피오젠 파지 φFC1 세린 리콤비나제; Listeria 파지 A1 18 세린 리콤비나제; S. 코엘리컬러 염색체 SC3C8.24 세린 리콤비나제; S. 코엘리컬러 염색체 SC2E1.37 세린 리콤비나제; S. 코엘리컬러 염색체 SCD78.04c 세린 리콤비나제; S. 코엘리컬러 염색체 SC8F4.15c 세린 리콤비나제; S. 코엘리컬러 염색체 SCD12A.23 세린 리콤비나제; S. 코엘리컬러 염색체 SCH10.38c 세린 리콤비나제; S. 코엘리컬러 염색체 SCC88.14 세린 리콤비나제; Streptomyces 파지 φC31 세린 리콤비나제; Streptomyces 파지 R4 세린 리콤비나제; Bacillus 파지 φ105 세린 리콤비나제; Bacillus 파지 SPBc2 세린 리콤비나제; Bacillus 프로파지 SKIN 세린 리콤비나제; S. aureus ccrA 세린 리콤비나제; S. aureus ccrB 세린 리콤비나제; M. tuberculosis 파지 Bxb1 세린 리콤비나제; M. tuberculosis 프로파지 φRV1 세린 리콤비나제; YBCK_ECOLI; Y4bA; Bja; Spn; Cac 1956; 및 Cac 1954; 또는 b) a)의 돌연변이단백질들.

[0019] 또 다른 양태에서, 본 발명은 본 명세서에서 기술된 키메라 폴리펩타이드를 인코딩한 단리된 핵산 분자를 제공한다.

[0020] 또 다른 양태에서, 본 발명은 본 명세서에서 기술된 단리된 핵산 분자를 포함하는 발현 카세트를 제공한다.

[0021] 또 다른 양태에서, 본 발명은 본 명세서에서 기술된 발현 카세트를 포함하는 벡터를 제공한다.

[0022] 또 다른 양태에서, 본 발명은 본 명세서에서 기술된 벡터를 함유하는 단리된 숙주 세포를 제공한다.

[0023] 또 다른 양태에서, 본 발명은 DNA 서열 내로의 부위-특이적 통합을 위한 방법을 제공한다. 이 방법은 DNA 서열을 본 발명의 키메라 폴리펩타이드와 접촉시키는 단계를 포함하며, 키메라 폴리펩타이드는 부위-특이적 통합에 촉매작용을 한다.

[0024] 또 다른 양태에서, 본 발명은 유전자 치료 방법을 제공한다. 이 방법은 본 명세서에서 개시된 키메라 폴리펩타이드를 인코딩한 핵산 분자를 포함하는 조성물을 대상에 투여하는 단계를 포함하며, 핵산 분자의 발현 시에, 대상의 게놈 내에 있는 유전자는 특이적으로 제거되거나 비활성화된다.

[0025] 또 다른 양태에서, 본 발명은 약학 조성물을 제공한다. 이 조성물은 본 명세서에서 기술된 키메라 폴리펩타이드; 및 약학적으로 허용가능한 캐리어를 포함한다. 다른 양태에서, 이 조성물은 본 명세서에서 기술된 키메라 폴리펩타이드를 인코딩한 핵산 분자; 및 약학적으로 허용가능한 캐리어를 포함한다.

[0026] 또 다른 양태에서, 본 발명은 본 발명의 키메라 폴리펩타이드에 의해서 촉매작용을 받는 재조합에 의해서 생성된 형질전환 유기체를 제공한다.

[0027] 또 다른 양태에서, 본 발명은 유전자 치료 방법을 제공한다. 이 방법은 본 명세서에서 기술된 부위-특이적 통합의 방법에 의해서 생성된 DNA 서열을 갖는 핵산 분자를 포함하는 세포를 대상에게 투여하는 단계를 포함한다.

[0028] 또 다른 양태에서, 본 발명은 본 명세서에서 기술된 카메라 단백질을 인코딩한 단리된 핵산 분자를 제공한다.

[0029] 또 다른 양태에서, 본 발명은 부위-특이적 재조합을 위한 방법을 제공한다. 이 방법은 (a) 본 명세서에서 기술된 키메라 단백질과 특이적으로 상호작용하기 위한 적어도 2 개의 바인딩 부위들을 포함하는 DNA 서열을 제공하는 단계; 및 (b) 키메라 단백질과 DNA 서열을 반응시키는 단계로서, 키메라 단백질은 DNA 서열의 양 가닥들이 키메라 단백질과 특이적으로 상호작용하는 2 개의 바인딩 부위들 간에서 절단되는 부위-특이적 재조합 이벤트에 촉매작용을 하는, 반응시키는 단계를 포함한다.

도면의 간단한 설명

[0030] 도 1은 TALER 융합 배향 및 활성도에 대한 일련의 그래픽적 및 도식적 표현들이다. A는 TALER 활성도를 평가하는데 사용되는 분할된 β-락타마아제 시스템을 예시한다. B는 각 TALER의 융합 배향 및 그의 대응하는 표적 부위 (1=SEQ ID NO: 288; 2=SEQ ID NO: 289; 3=SEQ ID NO: 290) 를 도시하는 개략도이다. C는 그의 의도된 DNA 표적에 대한 각 설계된 TALER 융합의 활성도를 도시한다. 재조합도가 백그라운드로 정규화되었다 (벡터 단독 제어). D는 코그네이트 (cognate) (Avr-20G) 및 비-코그네이트 (non-cognate) (Avr-20T, Avr-20GG, PthXo1-20G) DNA 표적들에 대한 Gin-Avr 활성도를 도시한다. 오차 막대들은 표준 편차 (s.d.) (n = 3) 를 나타낸다.

도 2는 선택된 TALER 절단부들의 재조합도 프로파일들에 대한 일련의 그래픽적 및 도식적 표현들이다. A는 20-멤버 TALER 절단부 라이브러리 (20-member TALER truncation library) 의 설계를 예시하는 개략도이다. B는

증가하는 길이 (14, 20, 26, 32 및 44-bp) 를 갖는 코어 서열을 포함하는 DNA 표적들에 대한 선택된 TALER 변형체들의 활성도를 도시한다. C는 증가하는 길이의 비-코그네이트 코어 서열들 또는 코어 부위들을 포함하는 기질들의 다양한 패널들에 대한 Gin-AvrXa7△120 활성도를 도시한다. 오차 막대들은 표준 편차 (s.d.) (n = 3) 를 나타낸다.

도 3은 증분된 절단부 라이브러리로부터 선택된 TALER 변형체들에 대한 일련의 그래픽적 표현들이다. A는 선택된 TALER 절단부 변형들의 빈도를 나타낸다. 3 회의 선택 후에, 증분적으로 절단된 Gin-AvrXa7 변형체들이 단리되고 DNA 서열화가 절단부 길이를 결정하는데 사용되었다. B는 Avr-32G DNA 표적에 대한 증분적으로 절단된 TALER 변형체들 (길이가 Δ92 내지 Δ 134로 변함) 의 활성도를 나타낸다. 참조를 위해서, 최단 (Δ145) 및 최장 (Δ74) 절단부 변형체들, 및 Δ87 절단부 변형체가 포함되었다. C는 코그네이트 및 비-코그네이트 DNA 표적들의 다양한 패널들에 대한 Gin-Avr△74, Gin-Avr△128 및 Gin-Avr△145의 활성도를 나타낸다. 오차 막대들은 표준 편차 (s.d.) (n = 3) 를 나타낸다.

도 4는 합성 TALER들의 활성도에 대한 일련의 그래픽적 표현들이다. A는 DNA 표적들 Avr-32G 또는 Pth-32G에 대한 합성 Gin-Avr15△128, Gin-Avr15△120 및 Gin-Pht15△120 변형체들의 활성도이다. B는 Avr-32G 및 Avr-32T에 대한 Gin-Avr△120에 기초한 길이가 15 내지 20 리피트로 되는 DBD들을 갖는 합성 TALER들의 활성도를 나타낸다. 오차 막대들은 표준 편차 (s.d.) (n = 3) 를 나타낸다.

도 5는 포유류의 세포 내에서의 TALER 활성도의 일련의 그래픽적 표현들이다. A 및 B는 리포터 플라스미드 (reporter plasmid) (Avr-32G, Avr-44G 및 C4-20G) 존재 시에 TALER 또는 ZFR 발현 벡터들 (Gin-Avr△120 및 GinC4) 로 (A의 경우) 또는 리포터 플라스미드 (Avr-G-ZF) 와 (Gin-Avr△120 + GinC4) 간의 결합 시에 TALER 및 ZFR 발현 벡터로 (B의 경우) 공동-형질주입된 (co-transfected) HEK293T 세포들 내에서의 루시페라제 발현도의 배-감소 (fold-reduction) 를 나타낸다. 오차 막대들은 표준 편차 (s.d.) (n = 3) 를 나타낸다.

도 6은 AvrXa7 (SEQ ID NO: 1 DNA 서열; SEQ ID NO: 2 아미노산 서열) 의 N-말단 설계된 절단부들에 대한 프라이머들의 위치의 도식적 표현이다. 별은 Δ120 융합 지점의 위치를 나타낸다.

도 7은 AvrXa7 표적 서열 (SEQ ID NO들: 16-18) 에 대한 원 와일드-타입 (native wild-type) 도메인 및 합성 RDV 도메인 간의 비교의 도식적 표현이다 .

도 8은 AvrXa7 단백질 TALE 및 TALER 아미노산 서열들 (SEQ ID NO: 19) 의 도식적 표현이다.

도 9는 AvrXa7 DNA 서열 (SEQ ID NO: 20) 의 구성의 도식적 표현이다.

도 10은 Gin-Avr△74 아미노산 서열 (SEQ ID NO: 21) 의 구성의 도식적 표현이다.

도 11은 Gin-Avr△87 아미노산 서열 (SEQ ID NO: 22) 의 구성의 도식적 표현이다.

도 12는 Gin-Avr△120 아미노산 서열 (SEQ ID NO: 23) 의 구성의 도식적 표현이다.

도 13은 Gin-Avr△120* 아미노산 서열 (SEQ ID NO: 24) 의 구성의 도식적 표현이다.

도 14는 Gin-Avr△147 아미노산 서열 (SEQ ID NO: 25) 의 구성의 도식적 표현이다.

도 15는 GinAvr15△128-합성 단백질 아미노산 서열 (SEQ ID NO: 26) 의 구성의 도식적 표현이다.

도 16은 GinAvr15A△128 -합성 단백질 DNA 서열 (SEQ ID NO: 27) 의 구성의 도식적 표현이다.

도 17은 GinAvr15△128-합성 단백질 아미노산 서열 (SEQ ID NO: 28) 의 구성의 도식적 표현이다.

도 18은 TALE N-말단 도메인의 특이도에 대한 일련의 그림 및 그래픽적 표현들이다. A는 그의 표적 DNA에 바인딩된 TALE (SEQ ID NO: 29) 를 예시한다. B의 구조적 분석은 N_{-1} 헤어핀 (hairpin) (N-0-SEQ ID NO: 30; N-1-SEQ ID NO: 31; 및 RVD - SEQ ID NO: 32) 의 W232에 의한 5'T의 접촉을 제시한다. 이러한 헤어핀은 RVD 헤어핀들과 상당한 서열 상동성 (homology) 을 공유한다. C 내지 F는 AvrXa7 TALE-R (C의 경우), AvrXa7 TALE-TF (D의 경우), AvrXa7 MBPTALE (E의 경우), 및 CCR5 표적화 TALEN (F의 경우) 의 상황에서의 NT-T (wt) NTD 의 분석결과들이다. (5'T와 비교될 시에 * = $p<0.05$, ** = $p< 0.01$, *** = $p<0.001$).

도 19는 리콤비나제 변형체들에 대한 일련의 그래픽적 및 도식적 표현들이다. A 내지 C는 A) 5'G, B) 5'A, 및 C) 5'C를 갖는 기질들에 대한 리콤비나제 선택 변형체들의 활동도들을 나타낸다. D는 N_{-1} 헤어핀 내의 서열 차들을 예시하는, 최적화된 TALE NTD들 (SEQ ID NO들: 33-36) 의 정렬이다. E는 MBP-TALE AvrXa7의 상황에서 최

적화된 NTD 활동도들 간의 종합적 비교이다. (와일드 타입 및 5'A/G/C와 비교될 시에 * = p<0.05, ** = p< 0.01, *** = p<0.001).

도 20은 TALE-TF들의 상황에서 선택된 NTD들의 분석의 일련의 도식적 및 그래픽적 표현들이다. A는 전사 활성화 실험들에서 사용된 루시페라제 리포터 플라스미드 상의 5xAvr 프로모터 영역 (SEQ ID NO: 37) 의 예시이다. B는 NT-T, NT-G, NT-áN 및 NT-âN 도메인들을 갖는 TALE-TF들에 의해 표시된 5' 잔기들을 갖는 기질들의 상대적 루시페라제 활성화를 나타낸다. (NT-T 및 각각의 5'A/G/C/T와 비교될 시에 * = p<0.05, ** = p< 0.01, *** = p<0.001).

도 21은 가변하는 5' 염기들을 갖는 와일드-타입 및 진화된 NTD들을 갖는 TALEN 쌍들의 설계 및 활성도의 일련의 도식적 및 그래픽적 표현들이다. A는 H32 돌연변이의 유도를 위한 표적 부위 (SEQ ID NO들: 40-47) 를 강조하도록 확장된 CCR5 유전자 (SEQ ID NO들: 38-39) 를 나타낸다. B는 와일트 타입 (NT-T) TALEN, 비-T 5' 잔기들에 대해서 최적화된 도메인들을 갖는 TALEN 및 dHax3 NTD를 갖는 TALEN의 유전자 에디팅 효율을 나타낸다. C는 5' 특이도를 갖는 TALEN들에 대한 최적화된 NTD를 갖는 TALEN 쌍들의 폴딩 강화 (fold enhancement) 를 나타낸다. 각 NTD의 활성도가 각 TALEN 쌍 기질 상에 도시된다.

도 22는 N- 및 C-말단 도메인들 (SEQ ID NO들: 48-53) 의 정렬을 도시하는 도식적 표현이다.

도 23은 TALE -리콤비나제 선택 프로토콜을 예시하는 개략적 표현이다. NTD의 라이브러리가 Not1/Stu1 제한 효소들 및 상보적 결찰을 사용하여서 Avr15 TALE-R로 복제된다. 활성 TALE-R는 항생제들 (카르베네실린) 으로 선택 및 증폭될 수 있는 보다 빈번한 재조합 이벤트들을 낳는다. 결과적인 출력 플라스미드는 다이제션된 Not1/Xba1이고 후속 선택 및 증폭을 위해서 TALE-R 백본 벡터 (backbone vector) 로 결찰된다.

도 24는 라이브러리 선택들 (라이브러리 XXXSGAR (SEQ ID NO: 39) 및 라이브러리 KXXGAR (SEQ ID NO: 291)) 로부터 발견된 변형체 집단들 (populations) 의 요약의 도식적 표현이다.

도 25는 NT-G (SEQ ID NO: 54) 의 NTD-Brg11 (SEQ ID NO: 55), Ralstonia TALE 도메인과의 정렬을 도시하는 도식적 표현이다. 이 정렬은 Brg11이 5'G 염기들에 대한 특이도를 보일 수 있음을 나타낸다.

도 26은 ELISA로 분석될 시에 MBP-TALE 단백질들의 표적 5' A/G/C/T Avr15 헤어핀 올리고뉴클레오타이드들에 대한 상대적 바인딩 친화도의 일련의 그래픽적 표현들이다. 단백질 농도는 ~ 75nM 였고 플레이트들은 120 분 동안 발현되었다.

도 27은 우측에 그려진 % 인델들 (indels) 및 인델 분포들을 사용한 TALEN 에디팅 후에 PCR 증폭된 CCR5의 세포 분석의 일련의 그림 및 그래픽적 표현들이다.

도 28은 도 27의 선택된 TALEN 실험들의 정렬 인델 서열화 (alignment indel sequencing) (SEQ ID NO들: 상단에서 하단으로 292-332) 의 도식적 표현이다.

도 29는 2 개의 개별 Goldy TALE-전사 인자 아키텍처들의 활성도들 간의 비교의 그래픽적 표현이며, 각 표적화하는 일치하는 (identical) 5x AvrXa7 프로모터들은 오직 5' 잔기 내에서 변한다.

도 30은 DNA에 바인딩된 징크-핑거 리콤비나제 다이머 (dimer) 의 구조에 대한 일련의 도식적 표현들이다. 도면 A에서, 각 징크-핑거 리콤비나제 (ZFR) 모노머 (블루 또는 오렌지) 는 맞춤화된 징크-핑거 DNA-바인딩 도메인에 링크된 활성화된 세린 리콤비나제 촉매성 도메인으로 구성된다. γδ 레졸바제 및 Aart 징크-핑거 단백질 (각기 PDB ID들: 1GDT 및 2I13) 의 결정질 구조들로부터 모델이 생성되었다. B는 DNA (SEQ ID NO들: 333-334) 에 바인딩된 ZFR 다이머를 도시한다. ZFR 표적 부위들은 ZFR 촉매성 도메인에 의해서 인식된 중앙의 20-bp 코어 서열을 측접하는 2 개의 역위된 징크-핑거 바인딩 부위들 (ZFBS) 로 구성된다. 징크-핑거 단백질들 (ZFP들) 은 '좌측' 또는 '우측' 절반-부위들 (각기 블루 박스 및 오렌지 박스) 을 인식하도록 설계될 수 있다. 축약형은 다음과 같다: N은 A, T, C, 또는 G를 나타내며; R은 G 또는 A를 나타내며; 및 Y는 C 또는 T를 나타낸다.

도 31은 Gin 리콤비나제 촉매성 도메인의 특이도의 일련의 그래픽적 및 도식적 표현들이다. A 내지 D의 도면에서, 재조합도가 디뉴클레오타이드 코어에서 각 가능한 2-염기 조합을 포함한 DNA 표적 (A, SEQ ID NO: 335) 상에서, 위치들 3 및 2에서 각 가능한 2-염기 조합을 포함하는 DNA 표적 (B, SEQ ID NO: 336) 상에서, 위치들 6, 5, 및 4에서 각 가능한 단일-염기 치환을 포함하는 DNA 표적 (C, SEQ ID NO: 337) 상에서 그리고 위치들 10, 9, 8, 및 7에서 각 가능한 단일-염기 치환을 포함하는 DNA 표적 (D, SEQ ID NO: 338) 상에서 측정되었다. 치환된 염기들은 각 패널 위에 박스처리된다. 재조합도는 분할된 유전자 재어셈블리에 의해서 평가되고 카르베

니실린-내성 형질 전환체의 클로람페니콜(chloramphenicol)-내성 형질 전환체 에 대한 비로서 측정된다 ("재료들 및 방법들" 섹션 참조). 오차 막대들은 표준 편차 (n = 3) 를 나타낸다. E는 γδ 레졸바제 다이머와 (좌측) 디뉴클레오타이드 코어, (중간) 위치들 6, 5, 및 4, 및 (우측) 위치들 10, 9, 8, 및 7 (PDB ID: 1GDT) 에서의 DNA와의 상호작용들을 나타낸다. 상호작용 잔기들이 자주색 막대들로 도시된다. 염기들이 다음과 같이 컬러로 표시된다: A, 엘로우; T, 블루; C, 브라운; 및 G, 핑크.

도 32는 Gin 리콤비나제 촉매성 특이도를 재엔지니어링하는 일련의 그래픽적 및 도식적 표현들이다. A는 Gin 촉매성 도메인에 의해 인식된 카노니컬 (canonical) 20-bp 코어를 나타낸다. 위치들 3 및 2은 박스처리된다 (SEQ ID NO: 339). B (상단) 은 DNA (PDB ID: 1GDT) 와 결합된 (in complex with) γδ 레졸바제의 구조를 도시한다. 돌연변이유발을 위해서 선택된 암 영역 (arm region) 잔기들이 자주색 막대로 표시된다. B (하단) 은 γδ 레졸바제 (SEQ ID NO: 341) 및 Gin 리콤비나제 (SEQ ID NO: 342) 촉매성 도메인들의 서열 정렬을 나타낸다. 보존된 잔기들 (conserved residues) 은 음영처리된 오렌지 색이다. 흑색 화살표들은 돌연변이유발을 위해 선택된 암 영역 위치들을 표시한다. C는 분할된 유전자 재어셈블리 선택 시스템의 개략적 표현이다. 활성 ZFR 변형체들의 발현은 β-락타마아제 판독 프레임 회복 및 암피실린에 대한 숙주-세포 내성 (host-cell resistance) 으로 이어진다. 실선들은 ZFR 표적 부위들의 위치 및 일치성을 나타낸다. 위치들 3 및 2 에는 밑줄이 그어져 있다 (SEQ ID NO: 340). D는 위치들 3 및 2에서 GC, GT, CA, TT, 및 AC 염기 조합들을 포함하는 코어 부위들을 재조합하는 Gin 돌연변이체들의 선택을 나타낸다. *는 인큐베이션 시간이 16 시간에서 6 시간으로 감소한 선택 단계들을 나타낸다 ("재료들 및 방법들" 섹션, 실례 5). E는 위치들 3 및 2에서의 각 가능한 2-염기 조합을 위해 선택된 촉매성 도메인들 (β, γ, δ, ε, 및 ζ, a로 표시된 와일드-타입 Gin) 의 재조합 특이도를 나타낸다. 의도된 DNA 표적들은 밑줄이 그어져 있다. 재조합은 분할된 유전자 재어셈블리에 의해서 결정되고 3중으로 수행되었다.

도 33은 포유류 세포들 내에서의 사용자-규정된 서열들을 재조합하는 ZFR들의 능력을 예시하는 일련의 그래픽적 및 도식적 표현들이다. A는 포유류 세포들 내에서의 ZFR 활성도를 평가하는데 사용된 루시페라제 리포터 시스템의 개략적 표현이다. ZFR 표적 부위들은 루시페라제 발현을 일으키는 SV40 프로모터를 측접한다. 실선들은 잠재적 ZFR 표적 부위들을 식별하는데 사용된 44-bp 컨센서스 표적 서열을 나타낸다. 밑줄친 염기들은 징크-핑거 표적들 및 위치들 3 및 2 (SEQ ID NO: 343) 을 나타낸다. B는 설계된 ZFR 쌍들 및 그들의 코그네이트 리포터 플라스미드로 공동-형질주입된 (co-transfected) HEK293T 세포들 내에서의 루시페라제 발현의 배-감소를 나타낸다. 배-감소는 공 벡터 (empty vector) 및 리포터 플라스미드를 사용한 형질주입으로 정규화되었다. 각 ZFR 표적 부위 (SEQ ID NO들: 상단에서 하단으로 344-362) 의 서열 일치 및 염색체 위치 및 각 ZFR 쌍의 촉매성 도메인 조성이 도시된다. 밑줄친 염기들은 위치들 3 및 2을 나타낸다. 표준 오차들은 3 개의 독립적인 실험들로부터 계산되었다. ZFR 아미노산 서열들이 표 2에 제공된다. C는 ZFR 쌍들의 특이도를 나타낸다. 루시페라제 발현의 배-감소가 각 비-코그네이트 리포터 플라스미드에 대하여 ZFR 쌍들 1 내지 9 및 GinC4에 대하여 측정되었다. 재조합도는 그의 코그네이트 리포터 플라스미드를 사용하여 각 ZFR 쌍의 배-감소로 정규화되었다. 분석들이 3중으로 수행되었다.

도 34는 인간 게놈으로의 통합를 표적화할 수 있는 ZFR들 능력을 예시하는 일련의 그래픽적 및 도식적 표현들이다. A는 ZFR들 1 (SEQ ID NO: 363), 2 (SEQ ID NO: 364), 및 3 (SEQ ID NO: 365) 에 의해서 표적화된 도너 플라스미드 (상단) 및 게놈상의 좌위 (genomic loci) 의 개략적 표현이다. 열린 상자들은 이웃하는 엑손들을 나타낸다. 화살표들은 전사 방향을 표시한다. 각 ZFR 표적의 서열 및 위치가 도시된다. 밑줄친 염기들은 징크-핑거 표적들 및 위치들 3 및 2을 나타낸다. B는 ZFR-매개된 통합의 효율을 나타낸다. 데이터는 오직 도너 플라스미드를 사용하여서 형질주입된 세포들로부터의 데이터로 정규화되었다. 오차 막대들은 표준 편차 (n = 3) 를 예시한다. C는 ZFR-매개된 통합의 PCR 분석을 나타낸다. PCR 프라이머 조합들은 비수정된 좌위를 증폭하거나 (상단 참조), 플라스미드를 순방향 배향으로 (중간 참조) 또는 역방향 배향으로 (하단 참조) 통합하였다. D는 ZFR들 1 (SEQ ID NO: 366) 및 3 (SEQ ID NO: 367) 에 대한 PCR-증폭된 통합된 도너의 대표적인 크로마토그램들 (chromato grams) 을 나타낸다. 화살표들은 프라이머 배향을 서열화하는 것을 나타낸다. 음영진 박스들은 게놈상의 표적 서열들을 나타낸다.

도 35은 다이머 인터페이스 외측에 위치한 리콤비나제 DNA-바인딩 잔기들의 도식적 표현이다. γδ 레졸바제가 표적 DNA와 결합된다. 촉매성 도메인 다이머은 시안 색으로 표시된다. DNA는 그레이 색으로 표현된다. 암 영역 잔기들은 적색 막대로 표시된다. 다이머 인터페이스에서의 잔기들은 자주색 막대들로서 표시된다 (PDB ID: 1GDT).

도 36은 선택된 리콤비나제들의 서열 분석의 도식적 표현이다. 파이형 차트들은 각 표적화된 암 위치에서 아미

노산 치환들의 퍼센티지를 표시한다. 4 회의 선택 후에, >20 클론들이 각 라이브러리로부터 서열화되었다. TT를 재조합하는 클론들의 서열 분석은 다른 곳에서 기술된다 (1).

도 37은 단리된 촉매성 도메인들의 코어 특이도를 도시하는 표이다. 4 회의 선택 후에, 위치들 3 및 2에서의 치환들로 코어 서열들을 재조합하는 선택된 촉매성 도메인들의 능력이 평가되었다. 할당된 DNA 표적들에 밑줄이 그어진다. 리콤비나제 돌연변이들이 도시된다. *는 후속 분석을 위해 선택된 촉매성 도메인들을 나타낸다. 위치들 3 및 2에서의 와일드-타입 염기 조합은 CC이다. 재조합은 분할된 유전자 재어셈블리 (2) 에서 결정되고 3중으로 수행되었다. TT 치환들을 재조합하는 촉매성 도메인들은 (1) 아닌 다른 곳에서 기술된다.

도 38은 선택된 촉매성 도메인들의 위치 특이도의 일련의 그래픽적 표현들이다. α, β, ɣ, δ 및 ζ 촉매성 도메인들과 대칭적으로 치환된 표적 부위들 간의 재조합 분석들이 수행된다. 재조합도는 위치들 6, 5 및 4에서의 (A (SEQ ID NO: 368)) 4,000 개 초과의 랜덤한 강 염기 (S: G 또는 C) 치환들을 포함한 라이브러리 DNA 표적 상에서 그리고 위치들 10, 9, 8 및 7에서의 (B (SEQ ID NO: 369)) (가능한 4.29 x 10^9 개 중에서) 10^6 개 초과의 고유한 염기 조합들 (N: A, T, C 또는 G) 을 포함한 라이브러리 DNA 표적 상에서 측정되었다. 재조합도는 분할된 유전자 재어셈블리 (2) 에 의해서 측정되었다 (n = 3).

도 39는 ZFR 호모다이머 활성도의 일련의 그래픽적 표현들이다. HEK293T 세포들이 대응하는 2.5 ng의 pGL3 ZFR 리포터 플라스미드와 함께 150 ng ZFR-L 또는 150 ng ZFR-R 으로 공동-형질주입되었다. 재조합도는 2.5 ng pGL3 ZFR 리포터 플라스미드와 함께 150 ng ZFR-L 및 150 ng ZFR-R으로의 공동-형질주입으로 정규화되었다.

도 40은 ZFR-변형된 세포들의 클론상의 분석을 도시하는 일련의 그림상의 표현들이다. PCR 프라이머 조합들이 비수정된 게놈상의 표적을 증폭하거나 순방향 또는 역방향 배향으로 플라스미드를 통합시켰다.

발명을 실시하기 위한 구체적인 내용

[0031] 본 발명은 TALE 리콤비나제 (TALER) 를 먼저 개시한다. 증분적으로 절단된 TALE 도메인들의 라이브러리를 사용하여서, 박테리아 세포 및 포유류 세포 내의 DNA를 재조합하는데 사용될 수 있는 최적화된 TALER 아키텍처가 식별되었다. 임의의 맞춤화된 TALE 리피트 어레이는 본 명세서에서 기술된 TALER 아키텍처 내로 삽입될 수 있으며, 이로써 바이오기술 및 의약 분야에서 응용되도록 엔지니어링된 리콤비나제들의 표적화 역량을 크게 확대시킬 수 있다.

[0032] TALE (transcription activator-like effector) 단백질들은 실제로 임의의 DNA 서열을 바인딩하도록 설계될 수 있다. TALE DNA-바인딩 도메인들의 설계를 위한 일반적인 가이드라인들은 TALE 에 의해서 바인딩된 DNA 서열의 5'에 가장 가까운 염기 (5'-most base) (N_0 염기) 는 티민이어야 함을 제시하고 있다. 이 위치에서의 각 DNA 염기와의 TALE 전사 인자들 (TALE-TF), TALE 리콤비나제들 (TALE-R) 및 TALE 뉴클레아제들 (TALEN들) 의 활성도들의 분석에 의해서 N_0 염기 요건이 정량화되었다. 5'T 부재 시에, 5'T를 포함하는 표적 서열들과 비교할 시에, TALE 활성도는 감소하는데, TALE-TF 활성도는 1000 배 이상 감소하고, TALE-R 활성도는 100 배까지 감소하고, TALEN 활성도는 10 배까지 감소하였음이 관측되었다. 모든 가능한 N_0 염기들을 인식하는 TALE 아키텍처들을 발현하기 위해서, TALE-R 활성도 선택들과 결합된 구조-가이드된 라이브러리 설계가 사용되어서 임의의 N_0 염기들을 수용하는 신규한 TALE N-말단 도메인들을 진화시켰다. G-선택적 도메인 및 광범위하게 반응성인 도메인들이 단리되고 특성화되었다. TALE-R 포맷 내의 선택된 엔지니어링된 TALE 도메인들은 모듈성 (modularity) 을 보였으며 TALE-TF 및 TALEN 아키텍처들에서 활성적이었다. 이 진화된 N-말단 도메인들은 TALE 바인딩 단백질들 및 설계자 효소들로서 임의의 DNA 서열의 효과적이면서 제약되지 않은 TALE-기반 표적화를 제공한다.

[0033] 또한, 서열 요건 제한사항들을 처리하기 위해서, 지식-기반 방식이 세린 리콤비나제 촉매적 특이도를 재엔지니어링하기 위해서 기술되었다. 특이도-결정 DNA-바인딩 잔기들의 포화 돌연변이유발에 기초한 이러한 전략은 특이도에 있어서 10,000 배 이상의 시프트를 보였던 리콤비나제 변형체들을 생성하는데 사용되었다. 중요하게는, 이 방식은 리콤비나제 다이머 인터페이스 외측에 위치한 아미노산 잔기들에만 집중하였다 (도 35). 따라서, 재-엔지니어링된 촉매성 도메인들은 ZFR 헤테로다이머들을 형성하기 위해서 연관될 수 있으며 이러한 설계된 ZFR 쌍들은 예외적인 특이도로 사전결정된 DNA 서열들을 재조합시킨다는 것이 결정되었다. 이러한 결과들은 함께, 이 방법에 의해서 개발된 특정화된 촉매성 도메인들의 확장된 카탈로그가 맞춤 특이도를 갖는 ZFR들을 생성하는데 사용될 수 있다는 가정으로 이끌었다. 여기에서, 기질 특이도 분석 및 유도된 진화의 조합이 사용되어서 추정된 4 x 10^8 고유 20-bp 코어 서열들을 인식할 수 있는 Gin 리콤비나제 촉매성 도메인들의 다양한 집합을 발현

할 수 있다. 이러한 재-엔지니어링된 촉매성 도메인들로부터 어셈블리된 ZFR들은 높은 특이도로 사용자-규정된 서열들을 재조합시키고 DNA를 인간의 세포들 내의 표적화된 내생성 좌위 (endogenous loci) 내로 통합시킨다는 것이 보였다. 이러한 결과들은 게놈 엔지니어링 및 유전자 치료를 포함하여서 다양한 응용들에서의 ZFR 기술의 잠재성을 입증하였다.

[0034] 본 조성들 및 방법들이 기술되기 이전에, 본 발명은 기술된 특정 조성들, 방법들 및 실험 조건들로 한정되지 않는다는 것이 이해되어야 하며 그 이유는 이러한 디바이스들, 방법들 및 조건들은 변할 수 있기 때문이다. 또한, 본 명세서에서 기술된 용어들은 오직 특정 실시예들만을 기술하기 위한 것이며 한정적으로 해석되지 말아야 하며 본 발명의 범위는 오직 첨부된 청구항들에서만 한정될 것이라는 것도 이해되어야 한다.

[0035] 본 명세서 및 첨부된 청구항들에서 사용되는 바와 같이, 명사의 단수 형태는 해당 문맥이 달리 명시적으로 지시하지 않는 이상 그 해당 명사의 복수형을 포함한다. 따라서, 예를 들어서, "조성" 또는 "방법"이 참조되면 이는 본 개시 등을 독해하면 본 기술 분야의 당업자에게 명백할 본 명세서에서 기술된 타입의 하나 이상의 조성들 및 방법들 및/또는 단계들을 포함한다.

[0036] 달리 규정되지 않는 이상, 본 명세서에서 사용된 모든 테크니컬 및 과학적 용어들은 본 발명이 속하는 기술의 당업자에 의해서 통상적으로 이해될 바와 동일한 의미를 갖는다. 본 명세서에서 기술된 것들과 유사하거나 또는 균등한 임의의 방법들 및 재료들이 본 발명의 실시 또는 테스트 시에 사용될 수 있지만, 바람직한 방법들 및 재료들이 이제 기술된다.

[0037] "리콤비나제들"은 리콤비나제에 의해서 인식된 특이적 DNA 서열들 간의 부위-특이적 재조합을 매개하는 효소들의 족이다 (Esposito, D., and Scocca, J. J., Nucleic Acids Research 25, 3605-3614 (1997); Nunes-Duby, S. E., et al, Nucleic Acids Research 26, 391-406 (1998); Stark, W. M., et al, Trends in Genetics 8, 432-439 (1992)).

[0038] 본 명세서에서 사용되는 바와 같이, 용어 "키메라 TALE 리콤비나제"는 다음으로 한정되지 않지만 서열-특이적 바인딩 활성도를 갖는 합성하여 유도된 TALE 단백질 또는 도메인 또는 자연적으로 발생한 TALE 단백질로부터 유도된 TALE 도메인을 갖는 리콤비나제들을 포함한다.

[0039] 본 명세서에서 사용되는 바와 같이, 용어 "키메라 징크 핑거 리콤비나제"는 다음으로 한정되지 않지만 서열-특이적 바인딩 활성도를 갖는 합성하여 유도된 징크 핑거 바인딩 단백질 또는 도메인 또는 자연적으로 발생한 징크 핑거 DNA 바인딩 단백질로부터 유도된 징크 핑거 바인딩 도메인을 갖는 리콤비나제들을 포함한다.

[0040] 본 명세서에서 사용되는 바와 같이, 용어 "징크 핑거 (zinc finger)", "징크 핑거 뉴클레오타이드 바인딩 도메인," 또는 이와 유사한 용어는 자연적으로 발생한 징크 핑거 및 인공적으로 생성된 징크 핑거 모두를 말한다. 이러한 징크 핑거들은 예를 들어서 다음으로 한정되지 않지만 C2H2, C4, H4, H3C, C3X, H3X, C2X2, 및 H2X2와 같은 다양한 골격 구조들을 가질 수 있으며, 여기서 X는 징크 결찰 (ligating) 아미노산이다. 이러한 골격 구조들에서, 징크 핑거 구조들을 인용할 시에 통상적인 바와 같이, "C"는 시스테인 잔기를 나타내며 "H"는 히스티딘 잔기를 나타낸다. 골격 C2H2을 갖는 징크 핑거들은 다음으로 한정되지 않지만 예를 들어서 Barbas 등에게로의 국제 공개 번호 WO2008/006028, Barbas에게로의 미국 특허 번호 7,101,972, Barbas 등에게로의 미국 특허 번호 7,067,617, Barbas 등에게로의 미국 특허 번호 6,790,941, Barbas에게로의 미국 특허 번호 6,610,512, Barbas 등에게로의 미국 특허 번호 6,242,568, Barbas 등에게로의 미국 특허 번호 6,140,466, Barbas에게로의 미국 특허 번호 6,140,081, Barbas에 의한 미국 특허 출원 공개 번호 20060223757, Barbas 등에 의한 미국 특허 출원 공개 번호 20060211846, Barbas 등에 의한 미국 특허 출원 공개 번호 20060078880, Barbas에 의한 미국 특허 출원 공개 번호 20050148075, Barbas 등에 의한 미국 특허 출원 공개 번호 20050084885, Barbas 등에 의한 미국 특허 출원 공개 번호 20040224385, Barbas 등에 의한 미국 특허 출원 공개 번호 20030059767 및 Barbas 등에 의한 미국 특허 출원 공개 번호 20020165356에 기술되어 있는 징크 핑거들을 포함하며, 이 문헌들 모두는 본 명세서에서 참조로서 인용된다. 다른 징크 핑거들이 Rebar 등에게로의 미국 특허 번호 7,067,317; Liu 등에게로의 미국 특허 번호 7,030,215; Rebar 등에로의 미국 특허 번호 7,026,462; Case 등에게로의 미국 특허 번호 7,013,219; Cox III 등에게로의 미국 특허 번호 6,979,539; Case 등에게로의 미국 특허 번호 6,933,113; Cox III 등에게로의 미국 특허 번호 6,824,978; Eisenberg 등에게로의 미국 특허 번호 6,794,136; Eisenberg 등에게로의 미국 특허 번호 6,785,613; Case 등에게로의 미국 특허 번호 6,777,185; Choo 등에게로의 미국 특허 번호 6,706,470; Cox IM 등에게로의 미국 특허 번호 6,607,882; Case 등에게로의 미국 특허 번호 6,599,692; Cox II 등에게로의 미국 특허 번호 6,534,261; Case 등에게로의 미국 특허 번호 6,503,717; Eisenberg 등에게로의 미국 특허 번호 6,453,242; Rebar 등에게로의 미국 특허 출원 공개 번호 2006/0246588; Rebar 등에게로의 미국

특허 출원 공개 번호 2006/0246567; Case 등에게로의 미국 특허 출원 공개 번호 2006/0166263; Cox HI 등에게로의 미국 특허 출원 공개 번호 2006/0078878; Rebar 등에게로의 미국 특허 출원 공개 번호 2005/0257062; Cox III 등에게로의 미국 특허 출원 공개 번호 2005/0215502; Cox MI 등에게로의 미국 특허 출원 공개 번호 2005/0130304; Case 등에게로의 미국 특허 출원 공개 번호 2004/0203064; Case 등에게로의 미국 특허 출원 공개 번호 2003/0166141; Case 등에게로의 미국 특허 출원 공개 번호 2003/0134318; Eisenberg 등에게로의 미국 특허 출원 공개 번호 2003/0105593; Cox IM 등에게로의 미국 특허 출원 공개 번호 2003/0087817; Rebar 등에게로의 미국 특허 출원 공개 번호 2003/0021776; 및 Case 등에게로의 미국 특허 출원 공개 번호 2002/0081614에 기술되어 있으며, 이 모든 문헌들은 본 명세서에서 참조로서 인용된다. 예를 들어서, 이러한 미국 특허들 및 미국 특허 출원 공개문들에서 기술된 하나의 대안은 이른바 "D-가능 부위들" 및 이러한 부위들에 바인딩될 수 있는 징크 핑거 모듈들 또는 징크 핑거 DNA 바인딩 도메인들의 사용을 수반한다. "D-가능" 부위는 적합하게 설계된 징크 핑거 모듈 또는 징크 핑거 DNA 바인딩 도메인이 표적 가닥의 3 개의 염기보다는 4 개의 염기에 바인딩되게 하는 표적 부위의 영역이다. 이러한 징크 핑거 모듈 또는 징크 핑거 DNA 바인딩 도메인은 이중-가닥 DNA 표적 분절 (표적 가닥) 중 일 가닥 상의 3 개의 염기들의 트리플릿에 그리고 다른 상보적 가닥 상의 4 번째 염기에 바인딩된다. 단일 징크 핑거의 4 염기 표적 분절로의 바인딩은 징크 핑거의 아미노산 서열 및 표적 가닥의 서열 상에 제약사항들을 부여한다.

[0041] 본 명세서에서 사용되는 바와 같이, 본 명세서에서 나타나는 다양한 아미노산 서열들에서 발생한 아미노산들은 그들의 잘 알려진 3-글자 또는 1-글자 축약형들에 따라서 식별된다. 다양한 DNA 단편들에서 발생하는 뉴클레오타이드들은 본 기술 분야에서 통상적으로 사용되는 표준 단일-글자 지정사항들로 지정된다.

[0042] 펩타이드 또는 단백질에서, 아미노산들의 적합한 보존적 치환들은 본 기술 분야의 당업자에게 알려져 있으며 일반적으로 결과적인 분자의 생물학적 활성도를 변화시키지 않으면서 이루어질 수 있다. 본 기술 분야의 당업자는 일반적으로 폴리펩타이드의 비본질적 영역들 내에서의 단일 아미노산 치환들은 실질적으로 생물학적 활성도를 변화시키지 않는다 (예를 들어서, Watson et al. Molecular Biology of Gene, 4th Edition, 1987, Benjamin/Cummings, p. 224 참조). 특히, 이러한 보존적 변형체는 수정된 아미노산 서열을 가지며 이로써 변화(들)는 단백질의 (보존적 변형체의) 구조 및/또는 활성도, 예를 들어서 항체 활성도, 효소 활성도, 또는 리셉터 활성도를 실질적으로 변화시키지 않는다. 이러한 것은 아미노산 서열의 보존적으로 수정된 변형체들, 즉 아미노산 치환들, 단백질 활성도를 위해서 중요하지 않는 이러한 잔기들의 첨가 또는 제거, 또는 심지어 중요한 아미노산들의 치환들이 구조 및/또는 활성도를 실질적으로 변화시키지 않도록 아미노산들을 유사한 특성들 (예를 들어서, 산성, 염기성, 양으로 대전 또는 음으로 대전, 극성 또는 비극성 등) 을 갖는 잔기들로 치환하는 것을 포함한다. 작용적으로 유사한 아미노산들을 제공하는 보존적 치환 표들은 본 기술 분야에서 잘 알려져 있다. 예를 들어서, 보존적 치환들을 선택하는 하나의 예시적인 가이드라인은 (최초 잔기 및 이에 후행하는 예시적인 치환): Ala/Gly 또는 Ser; Arg/Lys; Asn/Gln 또는 His; Asp/Glu; Cys/Ser; Gln/Asn; Gly/Asp; Gly/Ala 또는 Pro; His/Asn 또는 Gin; Ile/Leu 또는 Val; Leu/Ile 또는 Val; Lys/Arg 또는 Gin 또는 Glu; Met/Leu 또는 Tyr 또는 Ile; Phe/Met 또는 Leu 또는 Tyr; Ser/Thr; Thr/Ser; Trp/Tyr; Tyr/Trp 또는 Phe; Val/Ile 또는 Leu를 포함한다. 다른 예시적인 가이드라인들은 각각이 서로에 대해서 보존적 치환들인 아미노산들을 포함하는 다음의 6 개의 그룹들을 사용한다: (1) 알라닌 (A 또는 Ala), 세린 (S 또는 Ser), 트레오닌 (T 또는 Thr); (2) 아스파르트산 (D 또는 Asp), 글루탐산 (E 또는 Glu); (3) 아스파라긴 (N 또는 Asn), 글루타민 (Q 또는 Gln); (4) 아르기닌 (R 또는 Arg), 리신 (K 또는 Lys); (5) 이소류신 (I 또는 Ile), 류신 (L 또는 Leu), 메티오닌 (M 또는 Met), 발린 (V 또는 Val); 및 (6) 페닐알라닌 (F 또는 Phe), 티로신 (Y 또는 Tyr), 트립토판 (W 또는 Trp); (또한 예를 들어서 Creighton (1984) Proteins, W. H. Freeman and Company; Schulz and Schimer (1979) Principles of Protein Structure, Springer-Verlag 참조). 본 기술 분야의 당업자는 상기 식별된 치환들만이 가능한 보존적 치환들이 아님을 이해할 것이다. 예를 들어서, 일부 목적들을 위해서, 당업자는 포지티브이든 네거티브이든 서로에 대한 보존적 치환들로서 모든 하전된 (charged) 아미노산들을 간주할 수 있다. 또한, 인코딩된 서열로 단일 아미노산 또는 아미노산들의 소량 분율을 변경, 추가 또는 제거하는 개별적 치환들, 제거들 또는 추가들도 또한 전달될 단백질의 3 차원 구조 및 기능이 이러한 변형체에 의해서 보존되면 "보존되게 수정된 변형체들"로서 고려될 수 있다.

[0043] 본 명세서에서 사용된 바와 같이, 용어 "발현 벡터"는 비상동 DNA, 예를 들어 본 명세서에서 제공된 핵산 인코딩 융합 단백질들 또는 본 명세서에서 제공된 발현 카세트들의 삽입 또는 통합에 의해서 조작된, 본 기술 분야에서 알려진 플라스미드, 바이러스, 파지미드 (phagemid), 또는 다른 운반체를 말한다. 이러한 발현 벡터들은 통상적으로 세포 내에 삽입된 핵산의 효율적인 전사를 위한 프로모터 서열을 포함한다. 발현 벡터는 통상적으로 복제의 근원물 (origin), 프로모터, 및 변형된 세포들의 표현형 선택을 가능하게 하는 특이적 유전자들을 포

함한다.

[0044] 본 명세서에서 사용된 바와 같이, 용어 "숙주 세포들"은 벡터가 전파될 수 있고 그의 DNA가 발현될 수 있는 세포들을 말한다. 이 용어는 또한 대상의 숙주 세포의 임의의 자세포 (progeny) 를 포함한다. 모든 자세포가 모 세포와 동일할 수 없는데 그 이유는 복제 동안 변이가 발생할 수 있기 때문이다. 이러한 자세포는 용어 "숙주 세포"가 사용되는 때에 포함된다. 외래성 DNA가 연속적으로 유지되는 안정적인 전사의 방법들은 본 기술 분야에서 알려져 있다.

[0045] 본 명세서에서 사용된 바와 같이, 유전자 치료는 이러한 치료가 필요한 질병 또는 상태들을 갖는 포유류, 특히 인간의 특정 세포들, 표적 세포들로 비상동 DNA를 전사하는 것을 수반한다. 이 DNA는 비상동 DNA가 발현되는 방식으로 선택된 표적 세포들 내로 도입되고, 이로써 인코딩된 치료제가 생성된다. 이와 달리, 비상동 DNA는 일부 방식에서는 치료제를 인코딩한 DNA의 발현을 매개하거나 일부 방식에서 치료제의 발현을 직접적으로 또는 간접적으로 매개하는 산물, 예를 들어 펩타이드 또는 RNA를 인코딩할 수 있다. 유전자 치료는 또한 핵산이 도입되는 포유류 또는 세포에 의해서 생성된 유전자 산물을 보충하거나 결함이 있는 유전자를 대체하는 유전자 산물을 인코딩한 핵산을 전달하는데 사용될 수 있다. 도입된 핵산은 통상적으로 포유류 숙주 내에서 생성되지 않거나 치료학적으로 유효한 양으로 또는 치료학적으로 유용한 시간에 생성되지 않는 치료적 화합물, 예를 들어 그의 성장 인자 억제제 또는 그의 종양 괴사 인자 또는 억제제, 예를 들어서 그에 대한 리셉터를 인코딩할 수 있다. 비상동 DNA 인코딩 치료제는, 이 치료제 또는 이의 발현을 강화하거나 이와 달리 변형하기 위해서, 질병 상태에 있는 숙주의 세포들 내에 도입되기 이전에 수정될 수 있다. 유전자 치료는 또한 유전자 발현의 억제제 또는 리프레셔 (repressor) 또는 다른 조절인자 (modulator) 의 전달을 수반할 수 있다.

[0046] 본 명세서에서 사용된 바와 같이, 비상동 DNA는 전사, 번역 (translation) 또는 다른 조절가능한 생화학적 프로세스들에 의해서, 정상적으로는 자신이 발현되는 세포에 의해서 생체 내에서 생성되지 않는 RNA 및 단백질들을 인코딩하거나, 외인성 DNA의 발현을 변형시키는 매개체들을 매개시키거나 인코딩한 DNA이다. 비상동 DNA는 또한 외인성 DNA로서 지칭될 수 있다. 본 기술 분야의 당업자가 자신이 발현되는 세포에 대해서 비상동이거나 외인성으로 인식하거나 간주할 임의의 DNA가 본 명세서에서 비상동 DNA 내에 포함된다. 비상동 DNA의 실례들은 다음으로 한정되지 않지만 추적가능한 마커 단백질들, 예를 들어서 약제 내성을 부여하는 단백질을 인코딩한 DNA, 치료학적으로 유효한 물질들, 예를 들어서 항암 작용제들, 효소들 및 호르몬들을 인코딩한 DNA, 및 다른 타입의 단백질들, 예를 들어 항체들을 인코딩한 DNA를 포함한다. 비상동 DNA에 의해서 인코딩되는 항체들은 비상동 DNA가 도입된 세포의 표면 상에서 분비되거나 발현될 수 있다.

[0047] 따라서, 본 명세서에서 비상동 DNA 또는 외인성 DNA는 게놈에서 발견된 대응하는 DNA 분자와 같이 정확한 배향 및 위치로 존재하지 않는 DNA 분자를 포함한다. 또한, 비상동 DNA 또는 외인성 DNA는 다른 유기체 또는 종들로부터의 (즉, 외생의) DNA 분자를 말할 수도 있다.

[0048] 본 명세서에서 사용되는 바와 같이, 치료학적으로 유효한 산물은 비상동 핵산, 통상적으로 DNA에 의해서 인코딩된 산물이며, 핵산이 숙주 내에 도입되면, 이 산물이 발현되어서 선천적 또는 후천적 질병의 증상들 또는 징후들을 경감 또는 제거하거나 질병을 치료한다. 통상적으로, 표적화된 유전자 산물을 인코딩한 DNA는 통상적인 방법들, 예를 들어 칼슘-포스페이트 매개된 DNA 흡수 ((1981) Somat. Cell. Mol. Genet. 7:603-616 참조) 에 의해서 또는 프로듀셔 (producer) 세포들, 예를 들어서 팩키징 세포들 내로의 미세주입에 의해서 도입된 플라스미드 벡터로 복제된다 (cloned). 프로듀셔 세포들 내에서 증폭된 후에, 비상동 DNA를 포함하는 벡터들이 선택된 표적 세포들 내로 도입된다.

[0049] 본 명세서에서 사용되는 바와 같이, 발현 벡터 또는 전달 벡터는 외인성 또는 비상동 DNA가 적합한 숙주 세포 내에서 발현되기 위해서 삽입될 수 있는, 즉 이 DNA에 의해서 인코딩된 단백질 또는 폴리펩타이드가 숙주 세포의 시스템 내에서 합성되는, 임의의 플라스미드 또는 바이러스를 말한다. 하나 이상의 단백질들을 인코딩한 DNA 분절들 (유전자들) 의 발현을 제어할 수 있는 벡터들이 본 명세서에서 "발현 벡터들"로서 지칭된다. 또한, 역전사 효소를 사용하여서 생성된 mRNA들로부터 cDNA (상보적 DNA) 를 복제하게 하는 벡터들이 포함된다.

[0050] 본 명세서에서 사용되는 바와 같이, 유전자는 RNA 또는 폴리펩타이드를 인코딩한 뉴클레오타이드 서열을 갖는 핵산 분자를 말한다. 유전자는 RNA 또는 DNA일 수 있다. 유전자들은 코딩 영역을 선행하는 영역 (리더 (leader)) 및 후행하는 영역 (트레일러 (trailer)) 및 개별 코딩 분절들 (엑손들) 간에 개재된 서열들 (인트론들) 을 포함할 수 있다.

[0051] 본 명세서에서 사용되는 바와 같이, 핵산 분자 또는 폴리펩타이드 또는 다른 바이오분자를 참조할 때에 용어 "

단리된"은 핵산 또는 폴리펩타이드가 이 핵산 분자 또는 폴리펩타이드가 획득된 유전자적 환경으로부터 분리되었음을 의미한다. 또한, 바이오분자가 자연 상태로부터 변형되었음을 의미하기도 한다. 예를 들어서, 살아있는 동물에서 자연적으로 존재하는 폴리뉴클레오타이드 또는 폴리펩타이드는 "단리되지" 않지만, 그의 자연 상태의 공존하는 재료들로부터 분리된 동일한 폴리뉴클레오타이드 또는 폴리펩타이드는 이 용어가 본 명세서에서 사용되는 바와 같이 "단리된" 것이다. 따라서, 재조합형 숙주 세포 내에서 생성 및/또는 포함된 폴리펩타이드 또는 폴리뉴클레오타이드는 단리된 것으로 간주된다. 또한, 재결합형 숙주 세포로부터 또는 원 소스로부터 부분적으로 또는 실질적으로 정제된 폴리펩타이드들 또는 폴리뉴클레오타이드들도 역시 "단리된 폴리펩타이드" 또는 "단리된 폴리뉴클레오타이드"로서 간주된다. 예를 들어서, 일 화합물의 재조합 방식으로 생성된 버전은 Smith et al. (1988) Gene 67:3140에서 기술된 1-단계 방법에 의해서 실질적으로 정제될 수 있다. 용어들 "단리된" 및 "정제된"은 때로 상호교환가능하게 사용된다.

[0052] 따라서, "단리된"은 자연적으로 발생한 게놈에서 관심 핵산을 인코딩한 유전자를 바로 옆에서 측접하는 이러한 유전자들의 코딩 서열들이 핵산에 없다는 것을 의미한다. 단리된 DNA는 단일-가닥 또는 이중-가닥을 가질 수 있으며 게놈상의 DNA, cDNA, 재조합형 혼성 DNA, 또는 합성 DNA일 수 있다. 이는 원 DNA 서열과 동일할 수 있거나, 하나 이상의 뉴클레오타이드들의 삭제, 첨가 또는 치환에 의해서 원 DNA 서열과 상이할 수도 있다.

[0053] 용어들 "단리된" 또는 "정제된"이 생물학적 세포들 또는 숙주들로부터 만들어지는 제제들을 참조하여서 사용되는 경우에 이러한 용어들은 관심 DNA 또는 단백질의 조추출물 (crude extract) 을 포함하여 언급된 DNA 또는 단백질을 포함하는 임의의 세포 추출물을 의미한다. 예를 들어서, 단백질의 경우에, 정제된 제제 (purified preparation) 이 개별 기법 또는 일련의 조제상의 기법들 또는 생화학적 기법들을 따라서 획득될 수 있으며, 관심 DNA 또는 단백질은 이러한 제제들 내에서 다양한 순도로 존재할 수 있다. 특히 단백질들에 있어서, 이러한 절차들은 예를 들어서, 다음으로 한정되지 않지만, 암모늄 설파이트 분별 (ammonium sulfate fractionation), 겔 여과 (gel filtration), 이온 교환 크로마토그래피 (ion exchange chromatography), 친화도 크로마토그래피, 밀도 구배 원심분리법 (density gradient centrifugation), 일렉트로포커싱 (electrofocusing), 크로마토포커싱 (chromatofocusing) 및 전기영동 (electrophoresis) 을 포함할 수 있다.

[0054] "실질적으로 순수한" 또는 "단리된" DNA 또는 단백질의 제제는 이러한 DNA 또는 단백질이 정상적으로는 특성상 연관되는 자연적으로 발생하는 재료들이 없는 제제를 의미하는 것으로 이해되어야 한다. "본질적으로 순수한"은 관심 DNA 또는 단백질의 적어도 95 퍼센트를 포함하는 "고도로" 정제된 제제를 의미하는 것으로 이해되어야 한다.

[0055] 관심 DNA 또는 단백질을 함유하는 세포 추출물은 관심 단백질을 발현하거나 관심 DNA를 포함하는 세포들부터 획득된 균질 (homogenate) 제제 또는 무세포 제제를 의미하는 것으로 이해되어야 한다. 용어 "세포 추출물"은 세포들이 제거된 배지, 특히 소모된 배지 (spent culture media) 를 포함하는 것으로 의도된다.

[0056] 본 명세서에서 사용되는 바와 같이, 유전자의 프로모터 영역은 통상적으로 구조적 유전자까지 5' 만큼 놓여 있는 조절 요소 또는 요소들을 포함하며; 뉴클레오타이드 서열들을 개재시킴으로써 분리된 다수의 조절 요소들이 존재할 수 있다. 유전자가 활성화될 예정이면, 전사 인자들로서 알려진 단백질들이 유전자의 프로모터 영역에 부착된다. 이러한 어셈블리는 효소가 제 2 유전자적 분절를 DNA에서 RNA로 전사시키게 함으로써 "온 스위치 (on switch)" 와 유사하다. 대부분의 경우들에서, 결과적인 RNA 분자는 특이적 단백질 합성을 위한 템플릿 역할을 하며; 때로 RNA 자체가 최종 산물이다. 프로모터 영역은 정상적인 세포상의 프로모터 또는 예를 들어서, 종양 (onco)-프로모터일 수 있다. 종양-프로모터는 일반적으로 바이러스-유도된 프로모터이다. 징크 핑거 바인딩 폴리펩타이드들이 표적화될 수 있는 바이러스성 프로모터들은 다음으로 한정되지 않지만, 리트로바이러스성 LTR들 (long terminal repeats) 및 렌트바이러스 프로모터들, 예를 들어 HTLV (human T-Cell lymphotrophic virus) 1 및 2로부터의 프로모터들 또는 HIV (human immunodeficiency virus) 1 또는 2로부터의 프로모터들을 포함한다.

[0057] 본 명세서에서 사용되는 바와 같이, 용어 "절단된" 또는 이와 유사한 용어들은 원 단백질, 예를 들어 ZFP, TALE 또는 세린 리콤비나제의 전체 아미노산 서열보다 작은 서열을 포함하는 폴리펩타이드 유도체를 말한다.

[0058] 본 명세서에서 사용되는 바와 같이, 폴리펩타이드 "변형체" 또는 "유도체"는 폴리펩타이드의 돌연변이된 형태 또는 재조합을 통해서 생성된 것이지만 여전히 표적화된 활성도, 예를 들어서 리간드 또는 핵산 분자에 바인딩되거나 전사를 조절할 수 있는 능력을 보유한 폴리펩타이드를 말한다.

[0059] 본 명세서에서 사용될 때, 용어 "약학적으로 허용 가능한", "생리학적으로 허용 가능한" 및 이의 문법적 파생어

들은, 이것이 조성물, 캐리어, 희석제 및 시약을 의미할 때, 상호 교체적으로 사용되며 물질이 조성물의 투여를 방해하는 정도의 메스꺼움, 현기증, 위장 장애 등과 같은 바람직하지 않은 생리학적 영향의 발생 없이 인간에게 투여될 수 있음을 나타낸다.

[0060] 본 명세서에서 사용될 때, 용어 "벡터"는 이것이 작용하게 연결된 또 다른 핵산을 상이한 유전 환경 간에 운반할 수 있는 핵산 분자를 의미한다. 바람직한 벡터는 이들이 작용하게 연결되는 DNA 절편에 존재하는 구조적 유전자 생성물의 발현 및 자율 복제가 가능한 것들이다. 따라서, 벡터는 앞에서 개시한 선택적 마커 및 레플리콘을 포함하는 것이 바람직하다. 벡터는 발현 벡터를 포함하지만 반드시 이에 한정되지는 않는다.

[0061] DNA 단편을 포함하는 핵산 분자와 관련하여 본원에서 사용될 때, "작용하게 연결된"이란 구는 작용하게 연결된 부분이 의도하는 대로 기능하도록 단일 가닥 형태이든 또는 이중 가닥 형태이든 서열 또는 절편이 바람직하게는 종래의 인산이에스테르 결합에 의하여 DNA의 한 가닥에 공유 결합되었음을 의미한다. 본원에서 제공되는 전사 단위 또는 카세트가 작용하게 연결된 벡터의 선택은, 당업계에 널리 공지된 바와 같이, 목표된 작용적 특성, 예컨대, 벡터 복제 및 단백질 발현 및 형질전환되는 숙주 세포에 직접적으로 의존하는데, 이러한 것들은 재조합 DNA 분자구축 분야에서의 고유한 제약사항들이다.

[0062] 본원에서 사용될 때, 치료 조성물의 투여는 임의의 운반체로 실시될 수 있으며 경구 투여, 피하 투여, 정맥내 투여, 근내 투여, 흉골내 투여, 주입 기술, 복강내 투여 및 비경구 투여를 포함하나 이에 한정되지 않는다.

[0063] 세포들을 형질변환하는 방법들은 본 기술 분야에서 잘 알려져 있다. "형질변환된"은 외인성 DNA의 흡수로부터 기인되는 세포 내에서의 유전적인 변형을 의미한다. 적합한 방법들은 바이러스 감염, 형질주입, 접합, 원형질체 융합, 전기천공법, 입자 건 기술 (particle gun technology), 칼슘 포스페이트 석출 (calcium phosphate precipitation), 직접 미세주입 (direct microinjection) 등을 포함한다. 방법 선택은 일반적으로 형질변환되는 세포의 타입 및 형질변환이 발생하는 상황들 (즉, 인 비트로 (in vitro), 엑스 비보 (ex vivo) 또는 인 비보 (in vivo)) 에 의존한다. 이러한 방법들의 일반적인 논의는 Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995에서 찾을 수 있다.

[0064] 용어들 "핵산 분자" 및 "폴리뉴클레오타이드"는 상호교환가능하게 사용되며 디옥시리보뉴클레오타이드들 또는 리보뉴클레오타이드들 중 하나이면서 임의의 길이를 갖는 폴리머 형태의 뉴클레오타이드들 또는 이의 아날로그들을 말한다. 폴리뉴클레오타이드들은 임의의 3 차원 구조를 가질 수 있으며 공지되었거나 되지 않은 임의의 기능을 수행할 수 있다. 폴리뉴클레오타이드들의 비한정 실례들은 유전자, 유전자 단편, 엑손들, 인트론들, 메신저 RNA (mRNA), 전달 RNA, 리보조말 (ribosomal) RNA, 리보자임들, cDNA, 재조합 폴리뉴클레오타이드들, 분기된 폴리뉴클레오타이드들, 플라스미드들, 벡터들, 임의의 서열의 단리된 DNA, 임의의 서열의 단리된 RNA, 핵산 프로브들, 및 프라이머들을 포함한다.

[0065] "발현 카세트"는 관심 유전자/코딩 서열의 발현을 유도할 수 있는 임의의 핵산 구조체를 포함한다. 이러한 카세트들은, 발현 카세트를 표적 세포들 내로 전달하기 위해서, "벡터," "벡터 구조체," "발현 벡터," 또는 "유전자 전달 벡터" 내로 구축될 수 있다. 따라서, 이 용어는 복제 운반체 및 발현 운반체 및 바이러스 벡터들을 포함한다.

[0066] 핵산 및 아미노산 "서열 일치"를 결정하기 위한 기법들이 또한 본 기술 분야에서 알려져 있다. 통상적으로, 이러한 기법들은 유전자에 대한 mRNA의 뉴클레오타이드 서열을 결정하는 것 및/또는 이로써 인코딩된 아미노산 서열을 결정하는 것 및 이러한 서열들을 제 2 뉴클레오타이드 또는 아미노산 서열과 비교하는 것을 포함한다. 일반적으로, "일치"는 각기 2 개의 폴리뉴클레오타이드들 또는 폴리펩타이드 서열들의 정확한 뉴클레오타이드-대-뉴클레오타이드 또는 아미노산-대-아미노산 대응을 말한다. 2 개 이상의 서열들 (폴리뉴클레오타이드 또는 아미노산) 이 그들의 "퍼센트 일치도"를 결정함으로써 비교될 수 있다. 핵산 서열이든 아미노산 서열이든, 2 개의 서열들의 퍼센트 일치도는 2 개의 정렬된 서열들 간의 정확한 매칭 수가 보다 짧은 서열들의 길이에 의해서 나누어지고 100을 곱한 값이다. 핵산 서열들에 대한 근사 정렬도가 Advances in Applied Mathematics 2:482-489 (1981) 에서 Smith 및 Waterman의 국부적 상동성 알고리즘 (local homology algorithm) 에 의해서 제공된다. 이 알고리즘은 Atlas of Protein Sequences and Structure, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA의 Dayhoff에 의해서 개발된 측정 행렬 (scoring matrix) 을 사용하여서 아미노산 서열들에 적용될 수 있으며, Nucl. Acids Res. 14(6):6745-6763 (1986) 의 Gribskov 에 의해서 정규화될 수 있다. 서열의 퍼센트 일치도를 결정하기 위한 이러한 알고리즘의 예시적인 구현예는 Genetics Computer Group (Madison, Wis.) 에 의해서 "BestFit" 유틸리티 애플리케이션으로 제공된다. 이 방법에 대한 디폴트 파라미터들은 Wisconsin Sequence Analysis Package Program Manual,

Version 8 (1995) (Genetics Computer Group, Madison, Wis.으로부터 입수가능함) 에서 기술되어 있다. 본 발명의 상황에서 퍼센트 일치도를 확정하는 바람직한 방법은 University of Edinburgh에게 저작권이 있으며 John F. Collins 및 Shane S. Sturrok에 의해서 개발되었고 IntelliGenetics, Inc. (Mountain View, Calif.) 에 의해서 배포된 MPSRCE 프로그램 패키지를 사용하는 것이다. 이러한 세트의 패키지들로부터, Smith-Waterman 알고리즘이 채용될 수 있으며 여기서 디폴트 파라미터들은 측정 표에 대해서 사용되었다 (예를 들어서, 12의 갭 개방 페널티, 1의 갭 확장 패널티, 및 6의 갭 (gap)). 생성된 데이터로부터, "매칭" 값은 "서열 일치도" 를 반영한다. 서열들 간의 퍼센트 일치도 또는 유사도를 계산하기 위한 다른 적합한 프로그램들이 일반적으로 본 기술 분야에서 알려져 있으며, 예를 들어서, 다른 정렬 프로그램은 디폴트 파라미터들과 함께 사용되는 BLAST이다. 예를 들어서, BLASTN 및 BLASTP 이 다음과 같은 디폴트 파라미터들을 사용하여서 사용될 수 있다: 유전자적 코드 = 표준; 필터 = 없음; 가닥 = 양 가닥; 컷오프 = 60; 예상치 = 10; 행렬 = BLOSUM62; 디스크립션들 (Descriptions) = 50 서열들; 분류 (sort by) = HIGH SCORE; 데이터베이스들 = 비-리던던트 (non-redundant), GenBank+EMBL+DDBJ+PDB+GenBank CDS 번역들+Swiss 단백질+Spupdate+PIR.

[0067] 이와 달리, 상동성은 상동적인 영역들 간의 안정한 듀플렉스들 (duplexes) 을 형성하는 조건들 하에서 폴리뉴클레오타이드들의 혼성화 및 이어서 단일-가닥-특이적 뉴클레아제(들) 로의 다이제션 (다이제션) 및 다이제션된 단편들의 크기 결정에 의해서 결정될 수 있다. 2 개의 DNA, 또는 2 개의 폴리펩타이드 서열들이 위의 방법들을 사용하여서 결정될 시에 분자들의 규정된 길이에 걸쳐서 적어도 약 80% 내지 85%, 바람직하게는 적어도 약 85% 내지 90%, 보다 바람직하게는 적어도 약 90% 내지 95%, 및 가장 바람직하게는 적어도 약 95% 내지 98% 서열 일치도를 보일 때에 이 서열들이 서로 "실질적으로 상동적"이다. 본 명세서에서 사용되는 바와 같이, "실질적으로 상동적인"은 또한 특정된 DNA 또는 폴리펩타이드 서열에 대한 완벽한 일치를 보이는 서열들을 말하기도 한다. 실질적으로 상동적인 DNA 서열들은 예를 들어서 이러한 특정 시스템에 대해서 규정된 바와 같은 엄격한 조건들 하에서 Southern 혼성화 실험에서 확인될 수 있다. 적합한 혼성화 조건들을 규정하는 것은 본 기술 분야에서 알려져 있다. 예를 들어서, Sambrook et al, supra; DNA Cloning, supra; Nucleic Acid Hybridization, supra 참조.

[0068] 이로써, 본 발명은 실질적으로 상동적이고 균등한 생물학적 활성도를 보유한 폴리펩타이드들을 인코딩한 본 발명의 키메라 폴리펩타이드들을 인코딩한 핵산 및 아미노산 서열들을 제공한다.

[0069] 2 개의 핵산 단편들은 본 명세서에서 기술된 바와 같이 "선택적으로 혼성화"하도록 고려된다. 2 개의 핵산 분자들 간의 서열 일치 정도는 이러한 분자들 간의 혼성화 이벤트들의 효율 및 강도에 영향을 준다. 부분적으로 일치하는 핵산 서열은 완전하게 일치하는 서열이 표적 분자로 혼성화되는 것을 적어도 부분적으로 억제할 것이다. 완전하게 일치하는 서열의 혼성화의 억제는 본 기술 분야에서 잘 알려진 혼성화 분석법을 사용하여서 평가될 수 있다 (예를 들어서, Southern 블롯법, Northern 블롯법, 용액 혼성화, 등이며, Sambrook, et al., Molecular Cloning: A Laboratory Manual, 제 2 Edition, (1989) Cold Spring Harbor, N.Y. 참조). 이러한 분석법은 가변하는 선택도를 사용하여서, 예를 들어서 저 충실도로부터 고 충실도로 변하는 조건들을 사용하여서 수행될 수 있다. 저 충실도의 조건들이 사용되면, 비-특이적 바인딩의 부재는 부분적인 서열 일치도도 없는 2 차 프로브 (예를 들어서, 표적 분자와의 약 30 퍼센트 미만의 서열 일치도를 갖는 프로브) 사용하여서 평가될 수 있는데, 이러한 비-특이적 바인딩 이벤트 부재 시에는, 2 차 프로브는 표적으로 혼성화되지 않을 것이다.

[0070] 혼성화-기반 검출 시스템을 사용할 때에, 표적 핵산 서열에 상보적인 핵산 프로브가 선택되고, 이어서 적합한 조건들의 선택에 의해서, 이 프로브 및 표적 서열이 서로 "선택적으로 혼성화되거나" 또는 바인딩되어서 혼성 분자를 형성한다. "중간 정도의 충실도" 하에서 표적 서열에 선택적으로 혼성화될 수 있는 핵산 분자는 통상적으로 선택된 핵산 프로브의 서열과 적어도 대략 70% 서열 일치도를 갖는, 길이가 적어도 약 10-14 뉴클레오타이드들인 표적 핵산 서열의 검출을 가능하게 하는 조건들 하에서 혼성화된다. 충실한 혼성화 조건들은 통상적으로 선택된 핵산 프로브의 서열과 약 90 내지 95% 이상의 서열 일치도를 갖는, 길이가 적어도 약 10-14 뉴클레오타이드들인 표적 핵산 서열들의 검출을 가능하게 한다. 프로브 및 표적이 특정 서열 일치도를 갖는 프로브/표적 혼성화에서 유용한 혼성화 조건들은 본 기술 분야에서 알려진 바와 같이 결정될 수 있다 (예를 들어서, Nucleic Acid Hybridization: A Practical Approach, editors B. D. Hames and S. J. Higgins, (1985) Oxford; Washington, D.C.; IRL Press 참조).

[0071] 혼성화에 대한 충실도 조건들과 관련하여서, 다수의 균등한 조건들이 예를 들어서 다음과 같은 인자들을 변화시킴으로써 특정 충실도를 확정하는데 채용될 수 있다는 것이 본 기술 분야에서 잘 알려져 있다: 프로브 및 표적 서열들의 길이 및 성질, 다양한 서열들의 염기 조성, 염들 및 다른 혼성화 용액 성분들의 농도들, 혼성화 용액들 내에서의 차단제들 (예를 들어서, 포르마미드, 덱스트란 설파이트, 및 폴리에틸렌 글리콜) 의 존재 또는 부

재, 혼성화 반응 온도 및 시간 파라미터들, 및 가변하는 워싱 (wash) 조건들. 혼성화 조건들의 특정 세트의 선택은 본 기술 분야에서의 표준 방법들을 따라서 이루어진다 (예를 들어서, Sambrook, et al., Molecular Cloning: A Laboratory Manual, 제 2 Edition, (1989) Cold Spring Harbor, N.Y. 참조).

[0072] 제 1 폴리뉴클레오타이드가 제 2 폴리뉴클레오타이드의 영역, 그의 cDNA, 그의 보체들과 동일한 또는 실질적으로 동일한 염기쌍 서열을 갖는다면, 또는 제 1 폴리뉴클레오타이드가 상술한 바와 같이 서열 일치를 보인다면, 제 1 폴리뉴클레오타이드는 제 2 폴리뉴클레오타이드로부터 "유도된다".

[0073] (i) 제 1 폴리펩타이드가 제 2 폴리뉴클레오타이드로부터 유도된 제 1 폴리뉴클레오타이드에 의해서 인코딩되면, 또는 (ii) 제 1 폴리펩타이드가 상술한 바와 같이 제 2 폴리펩타이드들과 서열 일치를 보인다면, 제 1 폴리펩타이드는 제 2 폴리펩타이드로부터 "유도된다".

[0074] 부위-특이적 리콤비나제들은 게놈 엔지니어링을 위한 강력한 툴들이다. 세린 리콤비나제들의 레졸바제/인버타제 족의 과잉활성화된 (hyperactivated) 변형체들은 보조 인자들 없이 기능하며, 따라서 엔지니어링된 징크-핑거 단백질들 (ZFP들) 으로 원 DNA-바인딩 도메인들을 대체함으로써 관심 서열들로 재-표적화될 수 있다.

[0075] 본 명세서에서 기술된 징크 핑거 리콤비나제들은 세린 리콤비나제들의 레졸바제/인버타제 족으로부터 유도된 활성화된 촉매성 도메인 및 맞춤 설계된 징크-핑거 DNA-바인딩 도메인으로 구성된 키메라 효소들이 본 명세서에서 기술된다. 엔지니어링된 촉매성 도메인들으로부터 어셈블리된 ZFR들은 높은 특이도로 사용자-규정된 DNA 표적들을 효율적으로 재조합시키고 설계된 ZFR들은 인간 세포들 내의 표적화된 내생성 좌위 내로 DNA를 통합시킨다.

[0076] 일 양태에서, 본 발명은 대응하는 와일드 타입 리콤비나제보다 큰 촉매적 특이도를 갖는 복수의 징크 핑거 리콤비나제 (ZFR들) 단백질들을 생성하는 방법을 제공한다. 이 방법은 와일드-타입 Gin 촉매성 도메인에 있어서의 Gin Ile120, Thr123, Leu127, Ile136 및 Gly137 또는 이들의 조합과 균등한 위치들에서 리콤비나제 촉매성 도메인에 대한 랜덤한 돌연변이유발을 수행하며, 각 아미노산에 대해서 위치들 2 및 3에서 DNA를 변이시키는 단계; 및 이 리콤비나제 촉매성 도메인을 복수의 징크 핑거 바인딩 도메인들과 융합시켜서 ZFR들을 형성하는 단계; 및 대응하는 와일드 타입 리콤비나제보다 큰 촉매성 특이도를 갖게 이 ZFR들에 대하여 농축을 시키는 단계를 포함한다. 일 실시예에서, ZFR들은 GC, GT, CA, TT 및 AC으로부터 선택된 DNA 표적들에 대하여 증가된 촉매적 활성도를 갖는다. 일 실시예에서, 리콤비나제 촉매성 도메인은 Ile136 및/또는 Gly137에서 돌연변이유발될 수 있다.

[0077] 본 명세서에서 사용되는 바와 같이, 와일드-타입 Gin 촉매성 도메인은 다음과 같이 SEQ ID NO: 56으로서 기재된 아미노산 서열을 갖는 폴리펩타이드의 일부 또는 전부를 포함하는 Gin 촉매성 도메인을 말한다.

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD
TLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAE
MERELIIERTMAGLAAARNKGRIGGRPPKLTKAEWEQAGRLLAQGIPRKQVALIYDV
ALSTLYKKHP

[0078]

[0079] 다양한 실시예들에서, 본 발명의 키메라 폴리펩타이드들은 본 발명의 방법에 의해서 생성된 것들과 같은 Gin 촉매성 도메인을 포함한다. 특정 Gin 촉매성 도메인들은 표 1에서 제시된 것들을 포함한다.

표 1

[0080]

Gin 촉매성 도메인		
변형	SEQ ID NO:	서열
Gin α	57	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELIIERTMAGLAAARNKGRQGRPPKSG
Gin β	58	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELIIERTMAGLAAARNKGRRGRPPKS
Gin γ	59	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELIIERTMAGLAAARNKGRPWGRPPKSG
Gin δ	60	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELIIERTMAGLAAARNKGRPLGRPPKSG
Gin ε	61	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERLSIIERTMAGLAAARNKGRRGRPPKSG
Gin ζ	62	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELIIERTSAGRAAAANKGRMGRPKSG
표적화된 암 영역 위치들이 강조되었다. 랜덤 치환들이 볼드체로 밑줄이 그어진다. 과잉활성화 H106Y 돌연변이에 밑줄이 그어진다.		

[0081]

[0082] 다양한 실시예들에서, 본 발명의 방법에 의해서 생성된 ZFR들은 복수의 징크 핑거 바인딩 도메인들에 작용하게 연결된 Gin 촉매성 도메인을 포함한다. 본 발명에 의해서 생성된 예시적인 ZFR들은 표 2에서 제시된 것들을 포함한다.

표 2

예시적인 ZFR들의 아미노산 서열들		
ZFR-1 Left	SEQ ID NO: 63	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFSTSGNLVRHQRTHTGEKPYKCPECG KSFSQSGDLRRHQRTHTGEKPYKCPECGKSFSTSGNLVRHQRTHTG EKPYKCPECGKSFSTSGELVRHQRTHTGKKTSGQAGQ
ZFR-1 Right	SEQ ID NO: 64	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFSHRTTLTNHQRTHTGEKPYKCPECG KSFSQSGDLRRHQRTHTGEKPYKCPECGKSFSQSGDLRRHQRTHTG EKPYKCPECGKSFSQSGDLRRHQRTHTGKKTSGQAGQ
ZFR-2 Left	SEQ ID NO: 65	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFSQSGDLRRHQRTHTGEKPYKCPECG KSFSQRAHLERHQRTHTGEKPYKCPECGKSFSTSGNLVRHQRTHTG EKPYKCPECGKSFSRSDELVRHQRTHTGKKTSGQAGQ

[0083]

ZFR-2 Right	SEQ ID NO: 66	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFSRSDKLVRHQRTHTGEKPYKCPECG KSFSRKDNLKNHQRTHTGEKPYKCPECGKSFSTSGELVRHQRTHTG EKPYKCPECGKSFSRSDKLVRHQRTHTGKKTSGQAGQ
ZFR-3 Left	SEQ ID NO: 67	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGG

[0084]

		RPPKSGTGEKPYKCPECGKSFSTTGNLTVHQRTHTGEKPYKCPECG KSFSDPGALVRHQRTHTGEKPYKCPECGKSFSQSSNLVRHQRTHTG EKPYKCPECGKSFSRSDHLTNHQRTHTGKKTSGQAGQ
ZFR-3 Right	SEQ ID NO: 68	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFSRKDNLKNHQRTHTGEKPYKCPECG KSFSRSDHLTNHQRTHTGEKPYKCPECGKSFSDPGNLVRHQRTHTG EKPYKCPECGKSFSRKDNLKNHQRTHTGKKTSGQAGQ
ZFR-4 Left	SEQ ID NO: 69	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIAV GRPPKSGTGEKPYKCPECGKSFSQRANLRAHQRTHTGEKPYKCPEC GKSFSQSSSLVRHQRTHTGEKPYKCPECGKSFSTTGNLTVHQRTHT GEKPYKCPECGKSFSQRAHLERHQRTHTGKKTSGQAGQ
ZFR-4 Right	SEQ ID NO: 70	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIAV GRPPKSGTGEKPYKCPECGKSFSQRANLRAHQRTHTGEKPYKCPEC GKSFSRRDELNVHQRTHTGEKPYKCPECGKSFSQLAHLRAHQRTHT GEKPYKCPECGKSFSQRAHLERHQRTHTGKKTSGQAGQ

[0085]

ZFR-5 Left	SEQ ID NO: 71	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIAV GRPPKSGTGEKPYKCPECGKSFSRRDELNVHQRTHTGEKPYKCPEC GKSFSRSDHLTNHQRTHTGEKPYKCPECGKSFSQLAHLRAHQRTHT GEKPYKCPECGKSFSQRAHLERHQRTHTGKKTSGQAGQ
ZFR-5 Right	SEQ ID NO: 72	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFSTSGSLVRHQRTHTGEKPYKCPECG KSFSRSDKLVRHQRTHTGEKPYKCPECGKSFSQSGDLRRHQRTHTG EKPYKCPECGKSFSTSGELVRHQRTHTGKKTSGQAGQ

[0086]

ZFR-6 Left	SEQ ID NO: 73	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIIERTSAGRAAAINKGRIKGR PRKSGTGEKPYKCPECGKSFSQLAHLRAHQRTHTGEKPYKCPECGK SFSQLAHLRAHQRTHTGEKPYKCPECGKSFSDPGHLVRHQRTHTGE KPYKCPECGKSFSDSGNLRVHQRTHTGKKTSGQAGQ
ZFR-6 Right	SEQ ID NO: 74	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIIERTMAGRAAARNKGRIKGR GRPPKSGTGEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPEC GKSFSTTGNLTVHQRTHTGEKPYKCPECGKSFSDSGNLRVHQRTHT GEKPYKCPECGKSFSQSSNLVRHQRTHTGKKTSGQAGQ
ZFR-7 Left	SEQ ID NO: 75	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIIERTMAGRAAARNKGRIKGR PPKSGTGEKPYKCPECGKSFSTHLDLIRHQRTHTGEKPYKCPECGKS FSTTGNLTVHQRTHTGEKPYKCPECGKSFSQSSSLVRHQRTHTGEKP YKCPECGKSFSRSDNLVRHQRTHTGKKTSGQAGQ
ZFR-7 Right	SEQ ID NO: 76	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIIERTMAGRAAARNKGRIKGR PPKSGTGEKPYKCPECGKSFSRSDKLVRHQRTHTGEKPYKCPECGK SFSRRDELNVHQRTHTGEKPYKCPECGKSFSQSSSLVRHQRTHTGE KPYKCPECGKSFSRSDHLTNHQRTHTGKKTSGQAGQ
ZFR-8 Left	SEQ ID NO: 77	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIIERTMAGRAAARNKGRIKGG RPPKSGTGEKPYKCPECGKSFSQRAHLERHQRTHTGEKPYKCPECG KSFSTSGNLVRHQRTHTGEKPYKCPECGKSFSRSDELVRHQRTHTG EKPYKCPECGKSFSHKNALQNHQRTHTGKKTSGQAGQ
ZFR-8 Right	SEQ ID NO: 78	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD

[0087]

[0088]

		SIDTSSPMGRFFFYVMGALAEMERELIIERTSMAGIAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFSRRDELNVHQRTHTGEKPYKCPECG KSFSQSSNLVRHQRTHTGEKPYKCPECGKSFSQSSSLVRHQRTHTGE KPYKCPECGKSFSTTGNLTVHQRTHTGKKTSGQAGQ
ZFR-9 Left	SEQ ID NO: 79	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIIERTSMAGIAAARNKGRIVG GRPPKSGTGEKPYKCPECGKSFSTTGNLTVHQRTHTGEKPYKCPEC GKSFSQSSNLVRHQRTHTGEKPYKCPECGKSFSQRAHLERHQRTHT GEKPYKCPECGKSFSQKSSLIAHQRTHTGKKTSGQAGQ
ZFR-9 Right	SEQ ID NO: 80	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIIERTSAGIAAAINKGRIVGR PIKSGTGEKPYKCPECGKSFSDPGALVRHQRTHTGEKPYKCPECGK SFSQSSSLVRHQRTHTGEKPYKCPECGKSFSQLAHLRAHQRTHTGE KPYKCPECGKSFSQRANLRAHQRTHTGKKTSGQAGQ
암 영역 돌연변이들이 강조되었다. 특이도-결정 알파-헬리컬 징크-핑거 잔기들에 밑줄이 그어진다.		

[0089]

[0090] 실례들은 Gin 촉매성 도메인을 갖는 ZFR들의 생성을 예시하지만, 이 방법들은 다수의 다른 리콤비나제들의 촉매성 도메인들에 적용될 수 있다. 이러한 리콤비나제들은 a) EcoTn3로도 알려진 Tn3; StyHin로도 알려진 Hin; MuGin로도 알려진 Gin; Sin; Beta; Pin; Min; Din; Cin; EcoTn21; SfaTn917; BmeTn5083; Bme53; Cpe; SauSK1; SauSK41; SauTn552; Ran; Aac; Lla; pMER05; Mlo92; Mlo90; Rrh; Pje; Req; PpsTn5501; Pae; Xan; ISXc5; Spy; RhizY4cG; SarpNL1; SsolSC1904a; SsolSC1904b; SsoISC1913; Aam606; MjaMO014; Pab; HpylS607; MtulS_Y349; MtuRv2792c; MtuRv2979c; MtuRv3828c; MtuRv0921; MceRv0921; TnpX; TndX; WwK; 락토콕칼 파지 (lactococcal phage) TP901-1 세린 리콤비나제; S. 피오젠 파지 (pyogenes phage) ϕ370.1 세린 리콤비나제; S. 피오젠 파지 ϕFC1 세린 리콤비나제; Listeria 파지 A118 세린 리콤비나제; S. 코엘리컬러 염색체 (coelicolor chromosome) SC3C8.24 세린 리콤비나제; S. 코엘리컬러 염색체 SC2E1.37 세린 리콤비나제; S. 코엘리컬러 염색체 SCD78.04c 세린 리콤비나제; S. 코엘리컬러 염색체 SC8F4.15c 세린 리콤비나제; S. 코엘리컬러 염색체 SCD12A.23 세린 리콤비나제; S. 코엘리컬러 염색체 SCH10.38c 세린 리콤비나제; S. 코엘리컬러 염색체 SCC88.14 세린 리콤비나제; Streptomyces 파지 ϕC31 세린 리콤비나제; Streptomyces 파지 R4 세린 리콤비나제; Bacillus 파지 ϕ105 세린 리콤비나제; Bacillus 파지 SPBc2 세린 리콤비나제; Bacillus 프로파지 (prophage) SKIN 세린 리콤비나제; S. aureus ccrA 세린 리콤비나제; S. aureus ccrB 세린 리콤비나제; M. tuberculosis 파지 Bxb1 세린 리콤비나제; M. tuberculosis 프로파지 ϕRV1 세린 리콤비나제; YBCK_ECOLI; Y4bA; Bja; Spn; Cac 1956; 및 Cac 1954; 및 b) 상기 a)의 리콤비나제들의 돌연변이단백질들 (muteins) 을 포함한다.

[0091] 특정 도메인들과의 불완전한 모듈성, 모든 DNA 트리플릿들로의 고-친화도 바인딩의 부재, 및 구축의 비용이성은 통상적인 연구소들에서 ZFP들을 광범위하게 채택하는 것을 못하게 하고 있었다. Xanthomonas로부터의 TALE 단백질들 내에서의 새로운 타입의 DNA-바인딩 도메인의 발견은 ZFP들에 대한 대안을 제공한다. 키메라 TALE 리콤비나제들 (TALER들): DNA 인버타제 Gin로부터의 과잉활성화된 촉매성 도메인과 최적화된 TALE 아키텍처 간의 엔지니어링된 융합체들이 본 명세서에서 개시된다. 증분적으로 절단된 TALE 변형체들의 라이브러리가 박테리아 세포들 내의 징크-핑거 리콤비나제들과 비교할만한 효율성 및 특이도로 DNA를 수정하는 TALER 융합체들을 식별한다고 확인되었다. 실례들에서 도시된 바와 같이, TALER들은 포유류 세포들 내의 DNA과 재조합한다. 본 명세서에서 기술된 TALER 아키텍처는 맞춤된 TALE 도메인들의 삽입을 위한 플랫폼을 제공하며, 이로써 엔지니어링된 리콤비나제들의 표적화 능력을 크게 늘리며 바이오기술 및 의약 분야에서의 그들의 잠재적 응용들을 크게 할 수 있다.

[0092] TALE 단백질들은 사실상 임의의 DNA 서열에 바인딩되게 설계될 수 있다. TALE DNA-바인딩 도메인들의 설계에 대한 일반적인 가이드라인들은 TALE 에 의해서 바인딩된 DNA 서열의 5'에 가장 가까운 염기 (5' most base) (N_0

염기) 는 티민이어야 함을 제시하고 있다. 이 위치에서의 각 DNA 염기와의 TALE 전사 인자들 (TALE-TF), TALE 리콤비나제들 (TALE-R) 및 TALE 뉴클레아제들 (TALEN들) 의 활성도들의 분석에 의해서 이러한 N_0 염기 요건을 본 발명자들은 정량화하였다. 5'T 부재 시에, 5'T를 포함하는 표적 서열들과 비교할 시에, TALE 활성도는 감소하는데, TALE-TF 활성도는 1000 배 이상 감소하고, TALE-R 활성도는 100 배까지 감소하고, TALEN 활성도는 10 배까지 감소하였음이 관측되었다. 모든 가능한 N_0 염기들을 인식하는 TALE 아키텍처들을 개발하기 위해서, TALE-R 활성도 선택들과 결합된 구조-가이드된 라이브러리 설계가 사용되어서 임의의 N_0 염기들을 수용하는 신규한 TALE N-말단 도메인들을 진화시켰다. G-선택적 도메인 및 광범위하게 반응성인 도메인들이 단리되고 특성화되었다. TALE-R 포맷 내의 선택된 엔지니어링된 TALE 도메인들은 모듈성 (modularity) 을 보였으며 TALE-TF 및 TALEN 아키텍처들에서 활성적이었다. 이 진화된 N-말단 도메인들은 TALE 바인딩 단백질들 및 설계자 효소들로서 임의의 DNA 서열의 효과적이면서 제약되지 않은 TALE-기반 표적화를 제공한다.

[0093] 일 양태에서, 본 발명은 목표된 뉴클레오타이드에 특이적으로 바인딩되는 TALE (transcription activator-like effector) 단백질 바인딩 도메인을 생성하는 방법을 제공한다. 실례들에서 도시된 바와 같이, 이 방법은 a) RVD (repeat variable di-residue) 내의 또는 상기 RVD의 1 내지 2 아미노산 잔기들 N-말단 또는 C-말단 내의 아미노산 잔기를 변이시킴으로써 상기 TALE 단백질 바인딩 도메인의 아미노산 서열을 랜덤화하는 단계; 및 b) 상기 단계 a) 의 랜덤화된 TALE 단백질 바인딩 도메인에 대하여 선택하는 단계로서, 상기 TALE 단백질 바인딩 도메인은 상기 목표된 뉴클레오타이드와 특이적으로 바인딩하는, 상기 선택하는 단계를 포함한다.

[0094] 서열-특이적 뉴클레아제들, 리콤비나제들, 뉴클레아제들 및 전사 인자들이 본 명세서에서 제공된다. 서열 특이적리 콤비나제들은 맞춤화된 TALE DNA 바인딩 도메인을 포함한다. 이로써, 다른 양태에서, 본 발명은 키메라 폴리펩타이드를 제공한다. 이 폴리펩타이드는 a) 리콤비나제, 전사 인자 또는 뉴클레아제; 및 b) TALE 단백질을 포함한다.

[0095] TALE들은 병원체에 의해서 식물 세포 내로 주입된 식물 병원체성 박테리아의 단백질들이며, 세포 내에서 이들은 세포핵으로 이동하여서 전자 인자들로서 기능하여서 특이적 유전자들을 활성화시킨다 (turn on). TALE의 주 (primary) 아미노산 서열은 자신이 바인딩할 뉴클레오타이드 서열을 지시한다. 따라서, 표적 부위들이 TALE에 대해서 예측될 수 있으며, 또한 TALE는 본 명세서에서 기술된 바와 같이, 특정 뉴클레오타이드 서열들에 바인딩되도록 엔지니어링 및 생성될 수 있다.

[0096] 뉴클레아제, 전사 인자 또는 리콤비나제, 또는 이들의 일부를 인코딩한 서열들이 TALE-인코딩 핵산 서열들에 융합된다. 본 발명에서 사용될 수 있는 다수의 이러한 단백질들이 본 기술 분야에서 알려져 있다.

[0097] 다양한 실시예들에서, 키메라 폴리펩타이드는 리콤비나제의 촉매성 도메인을포함한다. 상술한 바와 같이, 다수의 리콤비나제들의 촉매성 도메인들이 사용될 수 있다. 이러한 리콤비나제들은 a) EcoTn3로도 알려진 Tn3; StyHin로도 알려진 Hin; MuGin로도 알려진 Gin; Sin; Beta; Pin; Min; Din; Cin; EcoTn21; SfaTn917; BmeTn5083; Bme53; Cpe; SauSK1; SauSK41; SauTn552; Ran; Aac; Lla; pMER05; Mlo92; Mlo90; Rrh; Pje; Req; PpsTn5501; Pae; Xan; ISXc5; Spy; RhizY4cG; SarpNL1; SsolSC1904a; SsolSC1904b; SsoISC1913; Aam606; MjaM0014; Pab; HpylS607; MtulS_Y349; MtuRv2792c; MtuRv2979c; MtuRv3828c; MtuRv0921; MceRv0921; TnpX; TndX; WwK; 락토콕칼 파지 (lactococcal phage) TP901-1 세린 리콤비나제; S. 피오젠 파지 (pyogenes phage) ϕ370.1 세린 리콤비나제; S. 피오젠 파지 ϕFC1 세린 리콤비나제; Listeria 파지 A118 세린 리콤비나제; S. 코엘리컬러 염색체 (coelicolor chromosome) SC3C8.24 세린 리콤비나제; S. 코엘리컬러 염색체 SC2E1.37 세린 리콤비나제; S. 코엘리컬러 염색체 SCD78.04c 세린 리콤비나제; S. 코엘리컬러 염색체 SC8F4.15c 세린 리콤비나제; S. 코엘리컬러 염색체 SCD12A.23 세린 리콤비나제; S. 코엘리컬러 염색체 SCH10.38c 세린 리콤비나제; S. 코엘리컬러 염색체 SCC88.14 세린 리콤비나제; Streptomyces 파지 ϕC31 세린 리콤비나제; Streptomyces 파지 R4 세린 리콤비나제; Bacillus 파지 ϕ105 세린 리콤비나제; Bacillus 파지 SPBc2 세린 리콤비나제; Bacillus 프로파지 (prophage) SKIN 세린 리콤비나제; S. aureus ccrA 세린 리콤비나제; S. aureus ccrB 세린 리콤비나제; M. tuberculosis 파지 Bxb1 세린 리콤비나제; M. tuberculosis 프로파지 ϕRV1 세린 리콤비나제; YBCK_ECOLI; Y4bA; Bja; Spn; Cac 1956; 및 Cac 1954; 및 b) 상기 a)의 리콤비나제들의 돌연변이단백질들 (muteins) 을 포함한다. 바람직한 실시예들에서, 강한 활성의 Gin 촉매성 도메인이 사용된다. 이러한 도메인은 본 명세서에서 기술된 바와 같이 본 발명의 방법들을 사용하여서 생성될 수 있다.

[0098] 본 명세서에서 기술된 바와 같이, TALE들은 이들이 DNA와 상호작용하는 특이도를 결정하는 다수의 불완전한 리피트들을 포함할 수 있다. 각 리피트는 이 리피트의 잔기들 12 및 13에서의 특정 디-아미노산 서열에 따라서

단일 염기에 바인딩된다. 따라서, TALE 내의 리피트들을 엔지니어링함으로써, 특정 DNA 부위들이 표적화될 수 있다. 이러한 엔지니어링된 TALE들은 예를 들어서, 특정 DNA 서열들에 표적화된 전사 인자들로서 사용될 수 있다.

[0099] 실례들에서 예시된 바와 같이, 본 발명의 키메라 단백질들은 표 3에서 제시된 바와 같은 변형체들 및 이들의 부분들 (예를 들어서, RVD들 및 NTD들) 에 의해서 예시된다.

표 3

[0100]

변형	SEQ ID NO:	서열
TALEN (Goldy) NT-T T1 Protein Sequence	81	MRSPKKKRKVQVDLRTLGYSQQQQEKIKPKVRSTVAQHH EALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEAT HEDIVGVGKQWSGARALEALLTDAGELRGPPLQLDTGQLV KIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNG GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA LETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQR LLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC QDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLT PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI
		ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALET VQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIVSNGGGKQALETVQRLLPVLCQD HGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPD QVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIA SNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGK QALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALESI VAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLP HAPELIRRVNRRIGERTSHRVAGSQLVKSELEEKKSELRHK LKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGK HLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQA DEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSG HFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGT LTLEEVRRKFNNGEINF

= N-Terminal Domain (NTD) *-varied as shown below*

[0101]

TALEN RVD Sequences		
G1	82	NG-NN-HD-NG-HD-NI-NG-NG-NI-HD-NI-HD-HD-NG-NN-HD-NI targeting (TCTTCATTACACCTGCA; SEQ ID NO: 280)
G2	83	HD-NI-NN-NG-HD-NI-NN-NG-NI-NG-HD-NI-NI-NG-NG targeting (CAGTCAGTATCAATT; SEQ ID NO: 281)
A1	84	HD-HD-NG-NN-HD-NI-NN-HD-NG-HD-NG-HD-NI-NG-NG-NG-NG targeting (CCTGCAGCTCTCATTTT; SEQ ID NO: 282)
A2	85	NI-NG-NG-HD-NG-NG-HD-HD-NI-NN-NI-NG-NG-NN-NI targeting (ATTCTTCCAGAATTGA; SEQ ID NO: 283)

[0102]

[0103]

A2	85	NI-NG-NG-HD-NG-NG-HD-HD-NI-NN-NI-NG-NG-NN-NI targeting (ATTCTTCCAGAATTGA; SEQ ID NO: 283)
C2	86	HD-NI-NN-NI-NI-NG-NG-NN-NI-NG-NI-HD-NG-NN-NI-HD-NG targeting (CAGAATTGATACTGACT; SEQ ID NO: 284)
T1	87	NG-HD-NI-NG-NG-NI-HD-NI-HD-HD-NG-NN-HD-NI-NN-HD targeting (TCATTACACCTGCAGC; SEQ ID NO: 285)

[0104]

T2	88	HD-NG-NG-HD-HD-NI-NN-NI-NI-NG-NG-NN-NI-NG-NI-HD-NG-NN targeting (CTTCCAGAATTGATACTG; SEQ ID NO: 286)
N-Terminal Domains		
NTD = dHax3-TALEN DNA Sequence	89	ATGAGATCTCCTAAGAAAAAGAGGAAGATGGTGGACTTGA GGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAA GCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCG CTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTT TCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATA CCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGG CAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGAGC ACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGC CTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCG AAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCATGCAT CGCGCAATGCACTGACGGGTGCCCCC
NTD = dHax3-TALEN	90	MRSPKKKRKMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAI VGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRG GVTAVEAVHASRNALTGAP… repeat variable diresidues
NTD = NT-βN TALEN DNA Sequence	91	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACCACGAG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAATATCATGGGGCACGCGC TCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGAGGTC CGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGATTGCA AAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCATGCAT CGCGCAATGCACTGACGGGTGCCCCC

[0105]

NTD = NT-βN TALEN	92	MRSPKKKRKVQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIV GVGKYHGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGV TAMEAVHASRNALTGAP…. repeat variable diresidues

[0106]

[0107]

NTD NT-G TALEN DNA Sequence	93	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACCACGAG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAATCGCGGTCGGGGGCAC GCGCTCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC
NTD NT-G TALEN Protein Sequence	94	MRSPKKKRKVQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIV GVGKSRSGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGG VTAMEAVHASRNALTGAP… repeat variable diresidues

[0108]

NTD = NT-αN TALEN DNA Sequence	95	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACCACGGG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAACGGGGGGCTGGTGCAC GCGCTCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC
NTD = NT-αN TALEN Protein	96	MRSPKKKRKVQVDLRTLGYSQQQQEKIKPKVRSTVAQHHGA LVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDI VGVGKRGAGARALEALLTDAGELRGPPLQLDTGQLVKIAKRG GVTAMEAVHASRNALTGAP…repeat variable diresidues
NTD = NT-T T-1 TALEN DNA	97	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACCACGAG

[0109]

		GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGGCGCAC GCGCCCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC

[0110]

MBP-TALE Protein Sequence	98	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPD KLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKA FQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPK TWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAF KYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDY SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKG QPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNK DKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMS AFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNN NNLGIEGRISEFGSPARPPRAKPAPRRRSAQPSDASPAAQVDLR TLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQ HPAALGTVAVTYQHIITALPEATHEDIVGVGK[**XXX**]GARALE ALLTDAGELLRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRN ALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGL TPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA LETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGL TPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQ ALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQRL LPVLCQDHGLTPDQVVAIVSNGGGKQALETVQRLLPVLCQDH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVV AIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK QALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVAQ LSRPDPALAALTNDHLVALACLG XXX: NT-T = QWS NT-G = SRS; NT-αN = RGA; NT-βN = Y-H

[0111]

TALE-R Protein Sequence	99	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDR PGLKRALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERG INFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLA AARNKGRIGGRPPKSGSPRPPRAKPAPRRRAAQPSDASPAAQV DLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVA LSQHPAALGTVAVTYQHIITALPEATHEDIVGVGK[**XXX**]GARA LEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASR NALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVA IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ ALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLL PVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA IASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGK QALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQR LLPVLCQDHGLTPDQVVAIVSNGGGKQALETVQRLLPVLCQD HGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVA QLSRPDPALAALTNDHLVALACLG XXX: NT-T = QWS NT-G = SRS; NT-αN = RGA; NT-βN = Y-H

[0112]

Avr15 TALE-TF Protein Sequence	100	MAQAASGSPRPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYS QQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAAL GTVAVTYQHIITALPEATHEDIVGVGK[**XXX**]GARALEALLTDA GELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPL NLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVV AIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK QALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRL LPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVV AIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGG KQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQ RLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQRLLPVLCQ DHGLTPDQVVAIVSNGGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALET

[0113]

		VQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVAQLSRPD PALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNR RIGERTSHRVADYAQVVRVLEFFQCHSHPAYAFDEAMTQFGM SGQAGQASPKKKRKVGRADALDDFDLDMLGSDALDDFDLDM LGSDALDDFDLDMLGSDALDDFDLDMLINYPYDVPDYAS XXX: NT-T = QWS NT-G = SRS; NT-αN = RGA; NT-βN = Y-H
Avr15 RVD Sequence (for TALE-R, TALE-TF, MBP-TALE)	101	NI-NG-NI-NI-NI-HD-HD-HD-HD-HD-NI-HD-HD-NI-NI targeting (ATAAACCCCCTCCAA; SEQ ID NO: 287)

[0114] 다양한 실시예들에서, 키메라 단백질은 C-말단 또는 N-말단 절단부를 갖는 TALE 단백질을 포함한다. 예를 들어서, TALE 단백질은 SEQ ID NO: 2의 일부 또는 전부를 포함할 수 있다. 실시예들에서, TALE 단백질은 아미노산 잔기들 27 과 268 간에서, 아미노산 잔기들 92 와 134 간에서, 아미노산 잔기들 120 와 129 간에서, 아미노산 잔기들 74 와 147 간에서, 또는 아미노산 잔기들 87 과 120 간에서 절단되고, 예를 들어서 상기 TALE 단백질은 아미노산 잔기 28, 74, 87, 92, 95, 120, 124, 128, 129, 147 및 150에서 절단된다.

[0115] 다른 실시예에서, TALE 단백질을 포함하는 단리된 폴리펩타이드가 제공되며, 여기서 이 TALE 단백질은 Q는 Y로 변이, Q는 S로 변이, Q는 R로 변이, W는 R로 변이, W는 G로 변이, W 삭제, S는 R로 변이, S는 H로 변이, S는 A로 변이, S는 N으로 변이, 및 S는 T로 변이로부터 선택된 하나 이상의 변이들 또는 삭제들을 갖는 SEQ ID NO: 3 (VGKQWSGARAL) 에 기재된 아미노산 서열을 포함하는 N-말단 도메인 (NTD) 을 갖는다.

[0116] 일부 실시예들에서, NTD는 VGKYRGARAL (SEQ ID NO: 4), VGKSRSGARAL (SEQ ID NO: 5), VGKYHGARAL (SEQ ID NO: 6), 및 VGKRGAGARAL (SEQ ID NO: 7) 으로부터 선택된 아미노산 서열을 포함한다.

[0117] 다른 실시예에서, TALE 단백질을 포함하는 단리된 폴리펩타이드가 제공되며, 여기서 이 TALE 단백질은 R_1는 K로 변이, Q는 Y로 변이, Q는 S로 변이, Q는 R로 변이, R_2는 W로 변이, R_2는 G로 변이, R_2 삭제, S는 R로 변이, S는 H로 변이, S는 A로 변이, S는 N으로 변이, 및 S는 T로 변이로부터 선택된 하나 이상의 변이들 또는 삭제들을 갖는 SEQ ID NO: 8 ($IVDIAR_1QR_2SGDLA$) 에 기재된 아미노산 서열을 포함하는 NTD를 갖는다.

[0118] 일부 실시예들에서, 이 NTD는 IVDIARQWSGDLA (SEQ ID NO: 9), IVDIARYRGDLA (SEQ ID NO: 10), IVDIARSRSGDLA

(SEQ ID NO: 11), IVDIARYHGDLA (SEQ ID NO: 12), 및 IVDIARRGAGDLA (SEQ ID NO: 13) 로부터 선택된 아미노산 서열을 포함한다.

[0119] 다른 실시예에서, TALE 단백질은 다음과 같이 제시된 아미노산 서열을 갖는 수정된 N_0 도메인을 포함한다:

[0120] LTPDQLVKIAKRGGTAMEAVHASRNALTGAPLN (SEQ ID NO: 102).

[0121] 다양한 실시예들에서, TALE 단백질은 SEQ ID NO: 102의 KRGG (SEQ ID NO: 103) 가 다음으로부터 선택된 돌연변이된 변형체를 포함한다:LDYE (SEQ ID NO: 104), INLV (SEQ ID NO: 105), YSKK (SEQ ID NO: 106), NMAH (SEQ ID NO: 107), SPTN (SEQ ID NO: 108), SNTR (SEQ ID NO: 109), LTTT (SEQ ID NO: 110), VADL (SEQ ID NO: 111), MVLS (SEQ ID NO: 112), YNGR (SEQ ID NO: 113), RIPR (SEQ ID NO: 114), YSKI (SEQ ID NO: 115), LTQY (SEQ ID NO: 116), YLSK (SEQ ID NO: 117), LRPN (SEQ ID NO: 118), LFTN (SEQ ID NO: 119), LLTN (SEQ ID NO: 120), EEDK (SEQ ID NO: 121), VTAM (SEQ ID NO: 122), CPSR (SEQ ID NO: 123), LTRV (SEQ ID NO: 124), KGDL (SEQ ID NO: 125), QKAL (SEQ ID NO: 126), LYLL (SEQ ID NO: 127), WISV (SEQ ID NO: 128), GDQV (SEQ ID NO: 129) 및 CPSR (SEQ ID NO: 130).

[0122] 다른 실시예에서, TALE 단백질은 다음과 같이 제시된 아미노산 서열을 갖는 수정된 N_{-1} 도메인을 포함한다:

[0123] MRSPKKKRKVQVDLRTLGYSQQQQEKIKPKVRSTVAQHH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGXXXXXARA
LEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAP (SEQ ID
NO: 131).

[0124] 다양한 실시예들에서, SEQ ID NO: 131의 XXXXX은 KRPAG (SEQ ID NO: 132) 또는 KRPSG (SEQ ID NO: 133) 이다. 또한, 이 단백질은 향상된 활성도를 보이는 (SEQ ID NO: 131에 있어서의) E40G 돌연변이를 포함할 수 있다.

[0125] 다른 실시예에서, TALE 단백질은 다음과 같이 제시된 아미노산 서열을 갖는 리피트 도메인을 포함한다:

[0126] LTPDVVAISNNGGKQALETVQRLLPVLCQDGH (SEQ ID NO: 134)

[0127] 다양한 실시예들에서, 이 TALE 단백질은 SEQ ID NO: 134의 SNNG (SEQ ID NO: 135) 이 다음으로부터 선택된 돌연변이된 변형체를 포함한다: RGGG (SEQ ID NO: 136), RGGR (SEQ ID NO: 137), RGVR (SEQ ID NO: 138), KGGG (SEQ ID NO: 139), SGGG (SEQ ID NO: 140), GGRG (SEQ ID NO: 141), LGGS (SEQ ID NO: 142), MDNI (SEQ ID NO: 143), RVMA (SEQ ID NO: 144), LASV (SEQ ID NO: 145), VGTG (SEQ ID NO: 146) 및 QGGG (SEQ ID NO: 147).

[0128] 다음의 실례들은 본 발명의 특징 및 장점을 더 예시하는데 제공되지만 본 발명의 범위를 한정하는 것은 아니다. 이들은 사용될 수 있는 것들의 전형이며, 본 기술 분야의 당업자에게 알려진 다른 절차들, 방법들 및 기법들이 이와 달리 사용될 수 있다.

[0129] **실례 1**

[0130] **키메라 TALE 리콤비나제들**

[0131] **실험 개요**

[0132] 본 연구는 TALE 리콤비나제 (TALER) 의 제 1 실례를 제공한다. 증분적으로 절단된 TALE 도메인들의 라이브러리를 사용하여서, 박테리아 및 포유류 세포들 내의 DNA를 재조합하는데 사용될 수 있는 최적화된 TALER 아키텍처가 식별되었다. 임의의 맞춤화된 TALE 리피트 어레이가 본 명세서에서 기술된 TALER 아키텍처 내로 삽입될 수 있으며, 이로써 엔지니어링된 리콤비나제들의 표적화 능력 및 의약 및 바이오기술에서의 응용도를 크게 높였다.

[0133] 다음의 재료들 및 방법들이 본 실례에서 사용되었다.

[0134] 시약들 (reagents)

[0135] 모든 효소들이 이와 달리 표시되지 않은 이상 New England BioLabs로부터 입수되었다. 프라이머 서열들이 표 4에서 제공된다.

표 4

[0136]

본 연구에서 사용된 프라이머들		
pBLA 기질 구축을 위한 프라이머들		
AvrXa7 lac target F	SEQ ID NO: 148	TTAATTAAGAGTCTAGAAATATAAACCCCCTCC AACCAGGTGCTAACTGTAAACCATGGTTTTGGA TTAGCACCTGGTTGGAGGGGGTTTATAAGATCT AGGAGGAATTTAAAATGAG
AvrXa7 lac target R	SEQ ID NO: 149	ACTGACCTAGAGAAGCTTATATAAACCCCCTCC AACCAGGTGCTAATCCAAAACCATGGTTTAC AGTTAGCACCTGGTTGGAGGGGGTTTATACTG CAGTTATTTGTACAGTTCATC
AvrXa7 N F	SEQ ID NO: 150	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAAGGTTTTGGTACCAAATGTC TATAAACCCCCTCCAACCAGGTGCTAAAGATC TAGGAGGAATTTAAAATGAG
AvrXa7 N R	SEQ ID NO: 152	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGACATTTGGTACCAAAACC

		TTATAAACCCCCTCCAACCAGGTGCTAACTGC AGTTATTTGTACAGTTCATC
AvrXa7 N RC F	SEQ ID NO: 153	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATATCCAAAACCATGGTTTACAG TATAAACCCCCTCCAACCAGGTGCTAAAGATC TAGGAGGAATTTAAAATGAG
AvrXa7 N RC R	SEQ ID NO: 154	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATATCCAAAACCATGGTTTACA GTATAAACCCCCTCCAACCAGGTGCTAACTGC AGTTATTTGTACAGTTCATC
AvrXa7 N RC +3 F	SEQ ID NO: 155	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAGCTTCCAAAACCATGGTTTA CAGGGTTATAAACCCCCTCCAACCAGGTGCTA AAGATCTAGGAGGAATTTAAAATGAG

[0137]

AvrXa7 N RC +3 R	SEQ ID NO: 277	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAACCCTGTAAACCATGGTTT TGGAAGCTATAAACCCCCTCCAACCAGGTGCT AACTGCAGTTATTTGTACAGTTCATC
AvrXa7 N RC +6 F	SEQ ID NO: 156	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAGCTTCATCCAAAACCATGGT TTACAGGGTTCCTATAAACCCCCTCCAACCAG GTGCTAAAGATCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC +6 R	SEQ ID NO: 157	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGCAACCCTGTAAACCATGG TTTTGGATGAAGCTATAAACCCCCTCCAACCA GGTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 N RC +12 F	SEQ ID NO: 158	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAGCTTCAGCTTCATCCAAAAC CATGGTTTACAGGGTTCCGGTTCCTATAAACC CCCTCCAACCAGGTGCTAAAGATCTAGGAGGA ATTTAAAATGAG

[0138]

AvrXa7 N RC +12 R	SEQ ID NO: 278	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGCAACCGCAACCCTGTAAA CCATGGTTTTGGATGAAGCTGAAGCTATAAAC CCCCTCCAACCAGGTGCTAACTGCAGTTATTT GTACAGTTCATC
AvrXa7 N RC -3 F	SEQ ID NO: 160	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG

[0139]

		AGGGGGTTTATAAAAACCATGGTTTATATAAA CCCCCTCCAACCAGGTGCTAAAGATCTAGGAG GAATTTAAAATGAG
AvrXa7 N RC -3 R	SEQ ID NO: 161	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATATAAACCATGGTTTTTATAA ACCCCCTCCAACCAGGTGCTAACTGCAGTTAT TTGTACAGTTCATC
AvrXa7 N RC GG F	SEQ ID NO: 162	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATATCCAAAACCGGGGTTTACA GTATAAACCCCCTCCAACCAGGTGCTAAAGA TCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC GG R	SEQ ID NO: 163	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATACTGTAAACCCCGGTTTT GGATATAAACCCCCTCCAACCAGGTGCTAAC TGCAGTTATTTGTACAGTTCATC

[0140]

[0141]

AvrXa7 N 20t F	SEQ ID NO: 164	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATACGAAATATTATAAATTA TCATATAAACCCCCTCCAACCAGGTGCTAA AGATCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC 20t R	SEQ ID NO: 165	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATATGATAATTTATAATATT TCGTATAAACCCCCTCCAACCAGGTGCTAAC TGCAGTTATTTGTACAGTTCATC
AvrXa7 32 GG F	SEQ ID NO: 166	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATAGCTTCATCCAAAACCGG GGTTTACAGGGTTCCTATAAACCCCCTCCAA CCAGGTGCTAAAGATCTAGGAGGAATTTAA AATGAG
AvrXa7 32 GG R	SEQ ID NO: 167	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATAGCAACCCTGTAAACCGG GGTTTTGGATGAAGCTATAAACCCCCTCCAA CCAGGTGCTAACTGCAGTTATTTGTACAGTTC ATC
AvrXa7 32t F	SEQ ID NO: 168	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATAGCTTCACGAAATATTATA AATTATCAGGTTCCTATAAACCCCCTCCAAC CAGGTGCTAAAGATCTAGGAGGAATTTAAA
		ATGAG
AvrXa7 32t R	SEQ ID NO: 169	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATAGCAACCTGATAATTTAT AATATTTCGTGAAGCTATAAACCCCCTCCAA CCAGGTGCTAACTGCAGTTATTTGTACAGTT CATC

[0142]

[0143]

pGL3Pro 표적 부위 구축을 위한 프라이머들

5' pGL3 SV40 Avr.32G BglII	SEQ ID NO: 170	TTAATTAAGAGAGATCTTTAGCACCTGGTTG GAGGGGGTTTATAGCTTCATCCAAAACCATG GTTTACAGGGTTCCTATAAACCCCCTCCAAC CAGGTGCTAAGCGATCTGCATCTCAATTAGT CAGC
3' pGL3 SV40 Avr.20G HindIII	SEQ ID NO: 171	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTT ATAGCAACC CTG TAA ACC ATG GTT TTG GATGAAGCT ATA AAC CCC CTC CAA CCA GGT GCT AAT TTG CAA AAG CCT AGG CCT CCA AA
5' pGL3 SV40 PH4.20G6Avr BglII	SEQ ID NO: 172	TTAATTAAGAGAGATCTGCGGGAGGCGTGTC CAAAACCATGGTTTACAGGGTTCCTATAAAC CCCCTCCAACCAGGTGCTAAGCGATCTGCAT CTCAATTAGTCAGC

[0144]

[0145]

3' pGL3 SV40 PH4.20G6Avr HindIII	SEQ ID NO: 173	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTT ATAGCAACCCTGTA AACCATGGTTTTGGACACGCCTCCCGCTTTG CAAAAGCCTAGGCCTCCAAA
5' pGL3 SV40 Avr.44G BglII	SEQ ID NO: 174	TTAATTAAGAGAGATCTTTAGCACCTGGTTG GAGGGGGTTTATAGCTTCAGCTTCATCCAAA ACCATGGTTTACAGGGTTCCGGTTCCTATAA ACCCCCTCCAACCAGGTGCTAAGCGATCTGC ATCTCAATTAGTCAGC
3' pGL3 SV40 Avr.44G HindIII	SEQ ID NO: 175	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTTATAGCAACCGCAA CCCTG TAA ACC ATG GTT TTG GATGAAGC TGAAGCT ATA AAC CCC CTC CAA CCA GGT GCT AAT TTG CAA AAG CCT AGG CCT CCA AA

[0146]

BamHI 융합을 위한 프라이머들

Gin_N-term F	SEQ ID NO: 176	AGTCAGTCGAGAGCTCATGGATCCCGGCTCTA TGCTGATTGGCTATGTAAGG
Gin_N-term R	SEQ ID NO: 177	ATGCTGATATCTAGACTATCCCGATTTAGGTGG GCGACC
Gin_C-term F	SEQ ID NO: 178	AGTCAGTCGAGAGCTCATGCTGATTGGCTATGT AAGG
Gin_C-term R	SEQ ID NO: 179	TCTAGACTACGGATCCACCGATTTACGCGGGC

[0147]

설계된 절단부들을 위한 프라이머들

TalR+28 Xba	SEQ ID NO: 180	ATCGCGTATCTAGACTAGCCGAGGCAGGCCAA GGCGACG
TalR+95 Xba AvrX	SEQ ID NO: 181	ATCGCGTATCTAGACTAGCTCATCTCGAACTGC GTCATG
avr n 1	SEQ ID NO: 182	GTCGCCCGCGTAAATCGGGATCCACTGCAGAT CGGGGGGGGGC
avr n 2	SEQ ID NO: 183	GTCGCCCGCGTAAATCGGGATCCCCCTCGCCTG CGTTCTCGGC
avr n 3	SEQ ID NO: 184	GTCGCCCGCGTAAATCGGGATCCGATTCGATGC CTGCCGTCGG
avr n 4	SEQ ID NO: 185	GTCGCCCGCGTAAATCGGGATCCACCGTGCGT GTCGCTGTCACTG
avr n 5	SEQ ID NO: 186	GTCGCCCGCGTAAATCGGGATCCGTGGATCTAC GCACGCTCGGC
avr n 6	SEQ ID NO: 187	GTCGCCCGCGTAAATCGGGATCCACACACGCG CACATCGTTGC

[0148]

avr n 7	SEQ ID NO: 188	GTCGCCCGCGTAAATCGGGATCCCACGAAGAC ATCGTTGGCGTCG
avr n 8	SEQ ID NO: 189	GTCGCCCGCGTAAATCGGGATCCAGCGCtCTGG AGGCCTTGCTC
avr n 9	SEQ ID NO: 190	GTCGCCCGCGTAAATCGGGATCCTTGGACACA GGCCAACTTCTC
avr n 10	SEQ ID NO: 191	GTCGCCCGCGTAAATCGGGATCCAGCGGCGTG ACCGCAgTGGA
GinNTALPCRfusR	SEQ ID NO: 192	GGATCCCGATTTACGCGGGC

pcDNA 복제를 위해서 사용된 프라이머들		
Nhe-SD-Gin F	SEQ ID NO: 193	ATCGTAGCAGCTAGCGCCACCATGCTGATTGGC TATGTAAG
GinGS R	SEQ ID NO: 194	GGATCCAGACCCCGATTTACGCGGGC

[0149]

[0150] **플라스미드 구축**

[0151] BamH1 제한 부위 5' 또는 3' 중 하나를 Gin 코딩 서열로 도입하기 위해서, Gin 촉매성 도메인이 각기 프라이머들 5' Gin_N-말단 및 3' Gin_N-말단 또는 5' Gin_C-말단 및 3' Gin_C-말단을 사용하여서 PCR 증폭되었다. 이 PCR 산물들은 pBluescriptII (Fermentas) 의 SacI 및 XbaI 제한 부위들 내로 결찰되어서 pB-Bam-Gin 및 pB-Gin-Bam을 생성하였다. C-말단 및 N-말단 TALER 융합체들을 생성하기 위해서, AvrXa7 유전자 (Dr. B. Yang, Iowa State University에 의해서 제공됨) 가 BamH1로 pWAvrXa7 로부터 분리되고 pB-Bam-Gin 및 pB-Gin-Bam (41) 의 BamH1 부위들 내로 결찰되어서 각기 pB-Avr-Bam-Gin 및 pB-Gin-Bam-Avr을 확립하였다. 각 TALER의 정확한 구축은 서열 분석에 의해서 입증되었다 (도 6 내지 도 16).

[0152] AvrXa7의 N-말단 절단부들을 생성하기 위해서, AvrXa7이 5' Avr-n-(1-10) 및 3' Avr +28 또는 3' Avr +95 프라이머들을 사용하여서 Expand High Fidelity PCR 시스템 (Roche) 을 통해서 다음의 같은 프로그램에 의해서 PCR 증폭되었다: 94 ℃에서의 3 분의 1 사이클, 94 ℃에서의 1 분의 16 사이클들, 52 ℃에서 1 분, 68 ℃에서 6 분; 및 68 ℃에서의 1 시간의 최종 사이클. Gin 촉매성 도메인은 5' Gin_C-단말 및 3' GinNTalPCRFus을 사용하여서 표준 PCR 조건들 하에서 PCR 증폭되고 위에서 기술된 PCR 조건들을 사용한 오버랩 PCR에 의해서 절단된 AvrXa7 변형체들로 융합되었다. 정제된 Gin-Avr PCR 산물들은 등몰 비 (equimolar ratio) 로 혼합되었고 SacI 및 XbaI를 사용하여서 다이제션되었다.

[0153] 설계자 TALE들을 생성하기 위해서, 본 발명자들은 다음과 같은 수정과 함께 TALEN 키트 (Addgene) 를 사용하였다: pTAL1이 Δ120, Δ128, 또는 +28에서 절단부들을 포함하도록 수정됨. 이를 달성하기 위해서, AvrXa7△120 및 AvrXa7△128 단편들이 5' Avr n4 또는 Avr n128 및 3' TalR Xba+28을 사용하여서 PCR 증폭되고 pTAL1의 BamH1 제한 부위 내로 결찰되어서 pTALA120 및 pTALA128를 생성하였다. 이 플라스미드들 pTALA120 및 pTALA128은 골든 게이트 복제를 위한 Esp31 제한 부위들을 보유하였다. pTAL△120 및 pTAL△128 내로 복제된 TALE 어레이들이 pB-Gin-Bam 내로 결찰되기 위해서 BamH1 및 XbaI을 사용하여서 다이제션되었다.

[0154] 포유류 TALER 발현 벡터들을 생성하기 위해서, Gin 촉매성 도메인이 5' Nhe-SD-Gin F 및 3' GinGS R을 사용하여서 pB-Gin-Avr로부터 PCR 증폭되고 pcDNA 3.1 (Invitrogen) 의 NheI 및 BamH1 제한 부위들로 결찰되었다 (ligated). Avr 15은 BamH1 및 XbaI을 사용하여서 pTAL△120 또는 pTAL△128로부터 다이제션되고 pcDNA-Gin-Bam 내로 결찰되어서 pcDNA-Gin-Avr 발현 벡터들을 생성하였다.

[0155] pBLA 기질 플라스미드들이 전술한 바와 같이 구축되었다.

[0156] pGL3 리포터 플라스미드들을 생성하기 위해서, SV40 프로모터가 재조합 부위-함유 프라이머들 5' pGL3 SV40 BglII 및 3' pGL3 SV40 HindIII을 사용하여서 pGL3-프로모터 (Promega) 로부터 PCR 증폭되고, pGL3-프로모터의 BglII 및 HindIII 제한 부위들 내로 결찰되었다.

[0157] **박테리아 재조합 분석**

[0158] 박테리아 재조합 분석이 전술한 바와 같이 수행되었다.

[0159] **증분적 절단부 라이브러리**

[0160] 증분적 절단부 라이브러리가 전술한 수정된 프로토콜을 사용하여서 생성되었다. 간략하게 말하자면, 엑소뉴클레아제 다이제션으로부터 Gin 코딩 서열을 보호하기 위해서, SmaI 제한 부위를 갖는 스터퍼 단편 (stuffer fragment) 이 BamHI 내로 삽입되어서 pB-Gin-SmaI-Bam-Avr를 생성하였다. 이 플라스미드는 NheI로 선형화되고 37 ℃에서 2.5 분 동안 엑소뉴클레아제 III를 사용하여서 인큐베이션되고 이후에 75 ℃에서 25 분 동안 열 비활성화되었다. 이어서, pB-Gin-Bam-Avr이 200 μM dNTP들 및 5 μM [알파]-S-dNTP들을 갖는 Klenow 단편 (3' 내지 5' Exo) 를 사용하여서 37 ℃에서 30 분 동안 인큐베이션되고, 이후에 80 ℃에서 25 분 동안 열 비활성화

되었다. 절단부 라이브러리를 생성하기 위해서, pB-Gin-Bam-Avr가 엑소뉴클레아제 III를 사용하여서 37 ℃에서 2.5 분 동안 인큐베이션되고, 이후에 열 비활성화되고, 이어서 30 ℃에서 1 시간 동안 Mung Bean 뉴클레아제를 사용하여서 블런트-말단 (blunt-ending) 처리되었다. SmaI으로 다이제션된 후에, 이 리콤비나제 코딩 서열의 블런트 3' 말단이 TALE 단편들의 블런트-말단처리된 라이브러리에 결찰되었다. 형질변환 및 정제 이후에, 이 플라스미드들은 SacI 및 XbaI로 다이제션되어서 Gin-△Avr를 릴리스하였다.

[0161] **포유류 리포터 분석들**

[0162] HEK293T 세포들이 웰 (well) 당 4 x 10^4 세포들의 밀도로 96-웰 플레이트들 상으로 시딩되고 (seeded) 37 ℃에서 습화된 (humidified) 5% CO_2 분위기에서 성장하였다. 시딩 후 24 시간에서, 세포들은 Renilla 루시페라제의 발현을 위해서 150 ng pcDNA TALER 발현 벡터, 2.5 ng pGL3 리포터 플라스미드, 및 1 ng pRL-CMV로 Lipofectamine 2000 (Invitrogen) 을 사용하여서 제조자의 지침들에 따라서 형질주입되었다. 형질주입 후 48 시간에서, 이 세포들은 Passive Lysis Buffer (Promega) 로 파괴되고 (lysed) 루시페라제 발현도가 Dual-Luciferase Reporter Assay System (Promega) 을 사용하여서 제조자의 지침들에 따라서 결정되었다. 발광도가 Veritas Microplate Luminometer (Turner Biosystems) 를 사용하여서 측정되었다.

[0163] **결과들**

[0164] **TALER 아키텍처**

[0165] 리콤비나제 활성도의 평가 및 유도된 진화를 위한 정량적 시스템이 기술되었다. 이 시스템 (도 1a) 에서, 재조합 부위들에 의해서 측접된 GFPuv 전이유전자는 유전자 인코딩 TEM-1 β-락타마아제 내로 삽입된다. 이러한 변형은 β-락타마아제 발현을 방해하며 이 플라스미드 (pBLA) 를 품고 있는 (harboring) *Escherichia coli* 세포들이 ampicillin에 민감하게 한다. 그러나, 기질-함유 플라스미드로부터의 활성 리콤비나제의 발현은 표적 부위들 간의 재조합 및 β-락타마아제 판독 프레임의 회복으로 이어진다. 이러한 수정은 ampicillin에 대한 숙주-세포 내성을 확립하며, 활성 리콤비나제 변형체들이 기질-함유 플라스미드로부터 단리되게 한다. 플라스미드 정제 및 재-변형 후에 ampicillin-내성 형질전환체들의 수를 측정함으로써, 리콤비나제 활성도가 또한 직접적으로 평가될 수 있다. 키메라 리콤비나제의 활성도는 촉매성 도메인 및 DBD 모두에 의존하기 때문에, 이러한 분할된 유전자 재어셈블리 선택 시스템은 또한 개별 DBD들의 유효성을 평가하는데 사용될 수도 있다. 따라서, 이 시스템은 최적의 TALER 아키텍처를 결정하도록 구성되었다.

[0166] 중요하게는, DNA 인버타제 Gin 및 이와 관련된 세린 리콤비나제들의 촉매성 도메인이 사전 규정된 촉매성 특이도들을 갖기 때문에, TALER 융합 단백질들은 TALEN들에 대해서 기술된 설계를 사용하여서 구축될 수 없다. γδ 레졸바제 및 설계된 효소들에 대한 구조적 및 작용적 연구들은 C-말단 E-헬릭스가 세린 리콤비나제 DNA 인식을 매개한다는 것을 나타내었다. ZFR들에서, 이러한 헬릭스는 DNA를 C-말단으로부터 N-말단으로, 5'로부터 3'으로 바인딩한다. 따라서, TALE들이 DNA을 5'에서 3'으로의 방향으로 바인딩하기 때문에, TALE 바인딩 부위가 20-bp 코어의 반대편 가닥 상에 위치하는 때에만 재조합이 발생할 것이라고 예상되었다 (도 1b).

[0167] AvrXa7을 사용하여서 TALER들을 생성하는 것이 선택되었는데, 이는 이러한 TALE 단백질이 TALE 뉴클레아제들 및 전사 인자들을 생성하는데 이전에 사용되었기 때문이다. 편리하게는, BamH1 제한 부위들이 AvrXa7을 포함하여서 다수의 TALE들에 측접하며, 다수의 그룹들이 이러한 제한 부위를 사용하여서 합성 TALE 융합체들을 생성하였다. 주목하게는, 이러한 BamH1 단편은 TALE의 N-말단이 그대로 유지되게 하지만 원 이펙터 도메인 (native effector domain) 을 C-말단으로부터 제거한다. 이러한 전략이 채택되어서 BamH1 제한 다이제션에 의해서 Gin-AvrXa7 융합체를 생성하였다.

[0168] Gin-AvrXa7은 Gin 촉매성 도메인에 의해서 인식된 중앙 20-bp 코어 서열 및 2 개의 측접 26-bp AvrXa7 바인딩 부위들로 구성된 pBLA 선택 벡터 함유 재조합 부위들로 복제되었다. 예상된 바와 같이, Gin-AvrXa7 융합체는, AvrXa7 바인딩 부위들이 20-bp 코어에 인접하여 위치될 때에는 DNA를 재조합할 수 없다 (도 1c). 그러나, AvrXa7 바인딩 부위들이 20-bp 코어의 반대편 가닥 상에 위치하는 경우에는, 재조합이 자명하며 (도 1c), 이는 재조합 부위 배향이 TALE N-말단으로의 촉매성 도메인 융합을 위해서 중요한 성분임을 말한다. N-말단 융합이 재조합을 위해서 필요하다는 것을 더 확립하기 위해서, 촉매성 도메인 활성도를 제약하는 것으로 예상된 비-카노니컬 (canonical) 융합 배향을 포함하는 C-말단 AvrXa7-Gin 변형체가 구축되었다 (도 1b 및 표 5). 예상된 바와 같이, 이러한 C-말단 AvrXa7 융합체는 박테리아 세포들 내에서의 무시할만한 활성도를 보였음이 결정되었다 (도 1c).

표 5

[0169]

변형체	SEQ ID NO:	서열
Gin-Avr (#1) /Avr20G	195	TTAGCACCTGGTTGGAGGGGGTTTATA TCCAAAACCATGGTTTACAG TATAAACCCCCTCCAACCAGGTGCTAA
Gin-Avr (#2)	196	TTAGCACCTGGTTGGAGGGGGTTTATA AGGTTTTGGTACCAAATGTC TATAAACCCCCTCCAACCAGGTGCTAA
Avr-Gin (#3)	197	TATAAACCCCCTCCAACCAGGTGCTAA CTGTAAACCATGGTTTTGGA TTAGCACCTGGTTGGAGGGGGTTTATA
Avr14G	198	TTAGCACCTGGTTGGAGGGGGTTTATA AAAACCATGGTTTA TATAAACCCCCTCCAACCAGGTGCTAA
Avr26G	199	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCCAAAACCATGGTTTACAGGGT TATAAACCCCCTCCAACCAGGTGCTAA

[0170]

Avr32G	200	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCATCCAAAACCATGGTTTACAGGGTTCC TATAAACCCCCTCCAACCAGGTGCTAA
Avr44G	201	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCAGCTTCATCCAAAACCATGGTTTACAGGGTTCCGGTTCC TATAAACCCCCTCCAACCAGGTGCTAA
Avr20GG	202	TTAGCACCTGGTTGGAGGGGGTTTATA TCCAAAACCGGGGTTTACAG TATAAACCCCCTCCAACCAGGTGCTAA
Avr20T	203	TTAGCACCTGGTTGGAGGGGGTTTATA

[0171]

		CGAAATATTATAAATTATCA TATAAACCCCCTCCAACCAGGTGCTAA
Avr32GG	204	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCATCCAAAACCGGGGTTTACAGGGTTCC TATAAACCCCCTCCAACCAGGTGCTAA
Avr32T	205	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCACGAAATATTATAAATTATCAGGTTCC TATAAACCCCCTCCAACCAGGTGCTAA
Avr-G-ZF	206	GCGGGAGGCGTG TCCAAAACCATGGTTTACAGGGTTCC TATAAACCCCCTCCAACCAGGTGCTAA
PthXo1-20G	207	GTGGTGTACAGTAGGGGGAGATGCA TCCAAAACCATGGTTTACAG TGCATCTCCCCCTACTGTACACCAC
PthXo1-32G	208	GTGGTGTACAGTAGGGGGAGATGCA GCTGCTTCCAAAACCATGGTTTACAGGGTGGT TGCATCTCCCCCTACTGTACACCAC

[0172] **설계된 절단부들**

[0173] 상술된 Gin-AvrXa7 융합체가 재조합에 대한 촉매작용을 할지라도, 이러한 변형체의 활성도는 엔지니어링된 ZFR들의 활성도보다 매우 낮았다. 또한, 특이도 분석은 Gin-AvrXa7 융합체가 인식 부위들 함유 비-코그네이트 DBD 부위들과 비-원 20-bp 코어 서열들 간을 신뢰할만하게 식별할 수 없다는 것을 보였으며, 이는 재조합이 Gin-매개되지 않을 수도 있다는 것을 말한다 (도 1d). 최근 보고들에 따르면, TALEN 활성도는 융합 단백질의 TALE 부분이 절단될 때에 개선되었다. 따라서, TALER 활성도를 높이기 위해서, 일련의 N 및 C-말단 AvrXa7 절단부들이 생성되었다 (도 2a).

[0174] 10 개의 N-말단 절단부들이 AvrXa7 Thr 27 (Δ27) 에서 시작하여서 AvrXa7 Gly 268 (Δ268) 에서 끝나게 대략 동일한 간격들로 어셈블리되었다 (도 6). TALEN들에 대한 N-말단 절단부 변형체로서 보고된 AvrXa7 Δ150이 또한 생성되었다. 2 개의 C-말단 AvrXa7 절단부들이 위치들 28 (+28) 및 95 (+95) 에서 생성되었다. 양 (both) +28 및 +95 은 TALEN들 내에서 안정된 융합 지점들로서 보고되었다. 각 TALE 절단부 변형체는 Gin 촉매성 도메인에 융합되었고, 이러한 20-멤버 TALER 라이브러리가 pBLA 선택 벡터 함유 Avr-20G 인식 부위들로 복제되었다. 박테리아 세포들 내에서의 1 회 선택 후에 ("재료들 및 방법들" 섹션), 개별 ampicillin-내성 클론들이 서열들이었으며, 모든 선택된 TALER들이 2 개의 N-말단 절단부들: Δ87 및 Δ120 중 어느 하나를 함유한 것으로 드러났다. 각 선택된 클론은 또한 C-말단 상의 +28 이었다. 융합 지점 (Δ120*) 근처에서의 자발적 12 아미노산 삭제를 갖는 단일 Δ120 클론을 제외하면, 이러한 클론들의 활성도는 매우 낮았다 (도 2b). 이러한 분석 시에, Gin-기반 ZFR들은 통상적으로 20 내지 40% 재조합도를 보였지만, 선택된 TALER 융합체들 중에 관측된 가장 높은 활성도는 ~7% 재조합도 (Gin-AvrXa7△120*) 이었다. TALE DBD는 ZF 도메인 (요구된 측접 펩타이드 서열을 포함하지 않음) 보다 3 배 더 크기 때문에, 본 발명자들은 이러한 TALER 구축들을 위해서 사용된 20-bp 스페이서는 재조합을 위한 최적의 길이가 아닐 수 있다고 사료하였다.

[0175] **코어 서열 길이**

[0176] 이어서 코어 서열 길이가 재조합에 대해서 갖는 효과가, 14 (Avr-14G), 26 (Avr-26G) 및 32-bp (Avr-32G) 코어 부위들을 함유하는 DNA 표적들이 선택된 TALER들에 의해서 재조합될 수 있는지의 여부를 평가함으로써 조사되었다. 리콤비나제-매개된 재어셈블리 후에 β-락타마아제 유전자의 판독 프레임을 유지하기 위해서, 코어 절반-부위들이 ± 3-bps 만큼 수정되었다 (표 1). 상술된 20-멤버 TALER 라이브러리가 각 표적 부위 변형체에 대해서 1 회의 선택을 받았다. 최단 표적을 재조합할 수 있는 TALER 변형체들의 식별은 가능하지 않았지만, Avr-26G 및 Avr-32G을 재조합한, Avr-14G (미도시된 데이터), 2 개의 Gin-△AvrXa7 변형체들이 (N-말단 TALE 절단부들 Δ87 및 Δ120 및 C-말단 절단부 +28 에 기초하여서) 식별되었다. 특히, 클론 분석은 이 선택된 TALER들 (Gin-AvrXa7△87 및 Gin-AvrXa7△120) 은 보다 짧은 코어들 (예를 들어서, 14 및 20-bps) 보다는 보다 긴 코어들 (예를 들어서, 26 및 32-bps) 과 함께 DNA를 적어도 100 배로 보다 효율적으로 재조합하였음이 드러났다 (도 2b). 또한, Gin-AvrXa7△120이 비-코그네이트 코어 (Avr-20T, Avr-20GG, Avr-32T 및 Avr-32GG) 보다는 코그네이트 코어 서열 (Avr-26G 및 Avr-32G) 을 함유하는 표적들을 100 배 이상 보다 효율적으로 재조합하였음이 드러났다 (도 2c). 흥미롭게는, Gin-AvrXa7△120 융합은 44-bp 코어들 (Avr-44G) 에 대해서는 그렇게 활성적이지 않았으며 (그 재조합도는 Avr-32보다 약 3 배 정도 낮았음) (도 2c), 이는 26-bp와 44-bp 간의 코어 길이들이 E. coli 내에서의 Gin-AvrXa7△120에 의한 재조합에 있어서는 최적일 가능성이 높다는 것을 말한다.

[0177] **증분적 절단부 라이브러리**

[0178] Gin-AvrXa7△120가 Gin-AvrXa7에 비해서 증가된 재조합 정도를 나타내었지만, Gin-AvrXa7△120이 다음의 이유들로 인해서 최적의 TALE 융합 아키텍처가 아닐 수도 있다고 사료되었다: (i) Gin 촉매성 도메인을 함유하는 ZFR들은 Gin-AvrXa7△120보다 2 배 이상으로 보다 효율적으로 DNA을 재조합하였으며, (ii) Gin-AvrXa7△120는 TALE 절단부 변형체들의 포괄적 라이브러리로부터 식별되지 않았기 때문이다. 따라서, 보다 양호한 융합 아키텍처들을 식별하기 위해서, 증분적으로 절단된 TALE DBD들의 라이브러리의 생성에 기초한 선별 (screen) 이 고안되었다.

[0179] 이를 위해서, 프로토콜이 절단된 C-말단 단편들 (AvrXa7) 의 라이브러리로의 비수정된 N-말단 도메인 (Gin) 의 융합을 가능하게 하도록 이전에 기술된 바와 같이 구성되었다 ("재료들 및 방법들" 섹션). AvrXa7 N-말단 (Met 1) 과 제 1 AvrXa7 리피트 (Leu 298) 간의 영역에 걸쳐 있는 N-말단 AvrXa7 절단부들이 엑소뉴클레아제 다이제션에 의해서 생성되고 Gin 촉매성 도메인의 비수정된 카피 (unmodified copy) (단백질 변형체들의 이론적 개수: 약 300) 로 융합되었다. 이전의 결과들이 +28이 최적의 C-말단 절단부이라고 나타내었기 때문에, 본 발명자들은 이러한 아키텍처를 절단부 라이브러리 내에 포함시켰다. TALER들은 pBLA 선택 벡터 함유 Avr-32G 표

적 부위들로 복제되고 E. coli 로 형질변환되었다 (>1 x 10^5 형질전환체들). 서열 분석은 관심 영역에 걸쳐 있는 (spanning) 절단부들의 균등한 분포 (데이터 미도시) 를 확인하였다.

[0180] 3 회의 선택 후에, 개별 ampicillin-내성 클론들이 서열들이었으며 및 다수의 고유한 절단부 변형체들이 확인되었다 (도 3a). 최적의 N-말단 TALER 융합 지점들이 위치들 87 및 120 근처에 위치할 가능성이 높다는 것을 암시한, 20-멤버 TALE 절단부 라이브러리를 사용하여서 수행된 선택들과 일관되게, 모든 선택된 Gin-AvrXa7 변형체들은 위치들 74 (Δ74) 과 147 (Δ147) 간의 절단부를 함유하는 것으로 드러났다. 특히, 73 개의 클론들 중 26 개의 클론들 (35.6%, $p < .001$) 이 위치들 124 (Δ124) 과 129 (Δ129) 간의 절단부들을 포함하였다. 이러한 분포로부터, 위치 128 (Δ128) 에서의 절단부들이 가장 대표적인 것이었다.

[0181] 선택된 AvrXa7 도메인들이 TALER 활성도를 증가시켰는지의 여부를 체계적으로 결정하기 위해서, 본 발명자들은 E. coli 내에서의 DNA 기질s 함유 Avr-32G 표적 부위들에 대한 단리된 Gin-AvrXa7 변형체들의 성능을 평가하였다. 본 발명자들은 AvrXa7 위치 92 (Δ92) 과 134 (Δ134) 간의 N-말단 삭제들을 포함하는 클론들에 대하여 분석을 집중하였다. 서열 분석과 일관되게, Δ120 와 Δ129 간의 N-말단 절단부들을 포함하는 TALER들이 비교상 보다 긴 또는 보다 짧은 절단부들에 기초한 변형체들보다 효율적으로 DNA를 재조합하였음이 드러났지만, Δ92 융합도 또한 매우 활성적이었다 (도 3b). 3 개의 클론들이 다음과 같이 더 특성화되었다: Δ74 및 Δ145가 가능한 융합 지점들의 경계들을 나타내기 때문에 선택되었으며, Δ128 이 선택들 시에 가장 우세한 (prevalent) 클론으로 드러났기 때문에 분석되었다. 스페이서 길이들 14-bp 내지 44-bp을 갖는 5 개의 표적들이 3 개의 네거티브 콘트롤 (three negative controls) (Avr32T, Avr32GG, 및 PthXo1-32G) 과 함께 분석되었다. Gin-Avr32G△74 및 Gin-Avr32G△145이 20-bp보다 긴 스페이서들에 대해서 적정한 활성도를 가진 한편, Gin-Avr32G△128가 ZFR GinC4에 필적할만한 효율로 DNA를 재조합하였음이 결정되었다 (도 3c). 또한, 특이도 분석에 의하면, Gin-Avr32G△74, Gin-Avr32G△128, 및 Gin-Avr32G△145이 비-코그네이트 코어들 (Avr-32T, Avr-32GG 및 PthXo1-32G) 보다 코그네이트 코어들이 살고 있는 기질들을 100 배 이상 효율적으로 조합할 수 있다 (도 3c). 이러한 결과들은 함께 Δ120 과 Δ129 간의 N-말단 삭제들을 포함하는 TALE 단백질들이 리콤비나제로의 융합을 위한 최적의 절단부를 나타내었음을 제시하였다.

[0182] **합성 TALE 리피트 어레이들의 혼입**

[0183] 위에서 기술된 연구들은 자연적으로 발생한 AvrXa7 TALE 단백질의 원 DBD들을 사용하였다. 설계된 TALE 리피트 어레이들이 선택된 Gin-△AvrXa7 골격들 내로 혼입될 수 있는지의 여부를 결정하기 위해서, 일련의 합성 TALE 단백질들 (길이가 15 리피트 내지 20 리피트임) 이 AvrXa7 바인딩 부위를 표적화하도록 생성 및 설계되었다 (도 7). 이 TALE 단백질들은 공중에서 입수가능한 TALEN 플라스미드 세트 (Addgene) 를 사용하여서 구축되었다. 복제 (cloning) 플라스미드는 +28 C-말단 절단부 및 Δ120 N-말단 절단부 또는 Δ128 N-말단 절단부 중 하나를 포함하도록 수정되었다. 설계된 TALE들이 Gin 촉매성 도메인 (Gin-Avr15△120 및 Gin-Avr15△128으로서 표시됨) 에 융합되었고 pBLA 선택 벡터 함유 Avr-32G 또는 Avr-32T 표적 부위들로 복제되었다.

[0184] E. coli에서의 활성도 분석은 Gin-Avr15△120 및 Gin-Avr15△128 모두가 활성 촉매성 도메인에 융합되는 때에 DNA를 재조합하는데 사용될 수 있으며 합성 리피트들의 혼입은 활성도 증가를 제공하였음을 드러내었다 (도 4a). 중요하게는, 각 TALER는 충실한 선택도 (stringent selectivity) 를 나타내었으며, 비-코그네이트 코어들보다 코그네이트 코어들을 포함하는 표적 부위들을 1000 배 이상으로 보다 효율적으로 재조합하였다 (도 4b). 놀랍게도, Δ120 절단부에 기초한 TALER들은 또한 Δ128 아키텍처에 기초한 TALE들만큼 효과적으로 DNA를 재조합하는 것으로 드러났으며 (도 4a), 이는 설계된 TALE들이 원 AvrXa7 DBD을 포함하는 것들보다 N-말단 절단부에 덜 민감할 수 있다는 것을 나타낸다.

[0185] 본 명세서에서 기술된 TALER 아키텍처가 임의의 DNA 서열을 표적화하도록 재프로그램될 수 있다는 것을 더 입증하기 위해서, 합성 효소가 자연적으로 발생한 TALE 단백질 PthXo1 (Gin-Pth15△120) 에 의해서 인식된 서열을 표적화하도록 생성 및 설계되었다. Gin-Pth15△120가 그의 코그네이트 기질 상에서 고도로 활성적임이 드러났으며, Gin-Pth15△120 및 Gin-Avr15△120 모두가 그들의 코그네이트 바인딩 부위들을 갖는 표적들에 대한 재조합에 있어서 600 배 이상의 증가를 보였다 (도 4a). 길이가 15 리피트 내지 20 리피트를 갖는 DBD들을 포함하는 일련의 설계된 TALER들의 활성도가 또한 평가되었으며, 각 융합체 (fusion) 가 유사하게 높은 효율 및 특이도로 재조합에 촉매 작용을 하였다고 드러났으며 (도 4b), 이는 합성 TALE 리피트 어레이들을 혼입한 키메라 리콤비나제들이 부위-특이적 재조합을 위해서 사용될 수 있음을 입증한다.

[0186] **포유류 세포들에서의 TALER 활성도**

[0187] TALER들이 포유류 세포들 내의 DNA를 수정할 수 있는지의 여부가 또한 결정되었다. 이를 위해서, 본 발명자들은 세포 배양 시에 리콤비나제 활성도의 신속한 평가를 가능하게 하는 에피솜 리포터 분석 (episomal reporter assay) 을 사용하였다. 이 분석 시에, HEK (human embryonic kidney) 293T 세포들이 재조합 부위들에 의해서 측접된 SV40 프로모터의 제어 하에서 루시페라제 유전자를 포함하는 리포터 플라스미드 (pGL3) 및 리콤비나제 발현 벡터를 사용하여 공동-형질주입되었다 (co-transfected). 적합한 리콤비나제의 과도적 (transient) 발현은 세포들 내에서의 감소된 루시페라제 발현 및 SV40 프로모터의 제거 (excision) 로 이어진다. 따라서, 리콤비나제 활성도는 루시페라제 발현에서의 배-감소에 정비례한다.

[0188] Avr-44G 인식 부위들 (pGL3-Avr-44G) 이 살고 있는 리포터 플라스미드를 사용하여서 Gin-Avr15△120을 공동-형질주입하는 것은 pGL3-Avr-44G만을 사용하여서 형질주입하는 것에 비해서 루시페라제 발현에서의 약 20-배 감소로 이어졌다 (도 5a). Gin-Avr15△120가 E. coli 내에서 ZFR GinC4와 유사한 활성도를 보였다는 사실에도 불구하고, 본 발명자들은 GinC4은 그의 코그네이트 표적 플라스미드, pGL3-C4-20G를 사용하는 공동-형질주입 후에 80 배 이상만큼 루시페라제 발현을 감소시켰다는 것을 알게 되었다 (도 5a). 이러한 차이는 pBLA보다 pGL3 내에서의 리콤비나제 표적 부위들 간의 비교적으로 보다 짧은 개재된 DNA 서열 또는 포유류 세포들 내에서의 TALER들과 ZFR들 간의 발현 차로 인한 것일 수 있다. 그러나, 이러한 차이에 대한 기본적인 이유는 불명확하게 남아있다. 마지막으로, 32-bp가 E. coli 내에서의 TALER들에 대한 최적의 코어 서열 길이로 결정되었지만, pGL3-Avr-32G과 함께 Gin-Avr15△120을 공동-형질주입하는 것은 루시페라제 발현에서의 단지 6-배 감소로 이어졌다는 것이 결정되었다 (도 5a). 또한, 이러한 차이에 대한 기본적인 이유는 불명확하게 남아 있다.

[0189] 이어서, ZFR (GinC4) 및 TALER (Gin-Avr15△120) 가 포유류 세포들 내에서의 양립가능한 헤테로다이머 (compatible heterodimer) 를 형성할 수 있는지의 여부가 조사되었다. 이러한 가능성을 평가하기 위해서, AvrXa7 바인딩 부위 및 C4 징크-핑거 바인딩 부위 (GCG GGA GGC GTG; SEQ ID NO: 279) 가 Gin 촉매성 도메인 (pGL3-Avr-G-ZF) 에 의해서 인식된 코어 서열을 측접 (flank) 하는 혼성 재조합 부위가 생성되었다 (표 2 참조). 놀랍게도, GinC4 및 Gin-Avr15△120과 함께 pGL3-Avr-G-ZF을 공동-형질주입하는 것은 pGL3-Avr-G-ZF에 비해서 루시페라제 발현에서의 140 배 이상의 감소로 이어졌지만 (도 5b), pGL3-Avr-G-ZF과 함께 GinC4 또는 Gin-Avr15△120 중 어느 하나를 사용하여서 형질주입하는 것은 리포터 유전자 발현에서의 무시할만한 감소로 이어졌다. 이러한 결과들은 ZF-TALE 헤테로다이머들을 생성하는 것이 키메라 리콤비나제들의 표적화 능력을 개선하기 위한 잠재적으로 효과적인 방식을 나타내는 것을 입증한다.

[0190] **<u>논의사항</u>**

[0191] 매우 최소의 융합 아키텍처를 포함하는 ZFP들과 달리, TALE DBD들은 DBD 어레이의 어느 일 측 상의 원 단백질 골격이 기능할 것을 요구한다. 위치 0에서 티미딘 잔기의 바인딩을 매개하고 거의 모든 알려진 TALE 인식 부위들에 발견되는 이른바 0번째 리피트 및 첫번째 리피트는 그러한 N-말단 골격을 나타낸다. 최근의 결정 구조는 위치 0 티민의 바인딩을 설명하였지만, 아직 최소 TALE 아키텍처를 결정하는데 있어서 데이터가 불충분하다. 실제로, 지금까지의 모든 연구들은 위치 0에서의 바인딩을 매개하는데 요구되는 것들보다 매우 많은 잔기들을 포함하는 N-말단 절단부를 사용하였다. 단백질의 이 부분이 적합한 DNA 바인딩 형태 (conformation) 를 이룰 시에 어떠한 역할을 하거나, 무엇이 최소 TALE 도메인을 구성할 수 있는지는 불명확하게 남아 있다. 작용적 TALE 키메라들을 생성하고자 한 초기 시도들은 전체-길이 TALE 단백질들로의 융합에 기초하였지만, 보다 최근의 연구들은 Δ150 N-말단 아키텍처의 맥락에서 이펙터 도메인 기능을 개선하는 고유한 C-말단 절단부들의 식별에 초점을 두었다. 이전의 보고에 따르면, AvrBs3 TALE의 N-말단 잔기들 2-153 (Δ 150) 의 삭제는 TALE의 전좌 (translocation) 를 위해서 요구된 도메인을 그의 원 박테리아로부터 표적 식물 세포로 이동시켰지만 전사 인자 활성도에서는 타협이 없었다.

[0192] 그러나, 활성 TALER을 개발하는 것은 고유한 N-말단 TALE 변형체들이 식별될 것을 요구하였다. 광범위하고 체계적인 조사가 초기에는 C-말단 절단부들 +28 및 +95을 갖는 N-말단 TALE들에 대해서 수행되었으며 오직 2 개의 도메인들 (+28에서의 Δ87 및 +28에서의 Δ120) 이 추후 분석을 위해서 충분하게 높은 활성도를 보였다고 드러났다. AvrXa7 N-말단의 증분적 절단부에 기초한 2 차 분석은 AvrXa7 위치 74 (Δ74) 와 위치 145 (Δ145) 간에서 센터링된 (centered) 절단부 변형체들의 광범위한 클러스터의 식별로 이어졌다. 이 실험에서 회복된 클론들 중에서, 38 퍼센트가 위치들 Δ119와 Δ128 간의 절단부들을 포함하였으며, 이 영역에서의 융합들을 사용하여서 TALER들에 대해서 획득된 데이터의 조사는 높은 활성도를 보였다. 특히, 이 영역 (Δ128 및 Δ120) 으로부터의 N-말단 절단부들에 기초한 TALER들이 박테리아 및 포유류 세포들 내의 DNA를 재조합하는데 사용될 수 있음이 결정되었다. Δ119 와 Δ128 간의 절단부 변형체들의 클러스터링 (clustering) 은 또한 이 영역의 내재적

안정성을 나타낼 수 있다.

[0193] ZFR들은 통상적으로 길이가 44 내지 50-bp인 표적 부위들 간의 재조합에 촉매작용을 한다. 각 표적 부위는 리콤비나제 촉매성 도메인에 의해서 인식된 중앙 20-bp 코어 서열, 및 2 개의 인접하는 ZFP 바인딩 부위들을 포함한다. 그러나, TALER들의 융합 배향은 TALE 바인딩 부위들이 중앙 코어 서열에 대해서 반대편의 가닥 상에 있을 것을 요구한다. 이러한 고유한 기하구조는 본 발명자들로 하여금 재조합을 위한 최소의 코어 서열 요건들을 조사하게 했다. TALE DBD들의 길이 (TALE 리피트들은 ZFP들보다 길이가 3 내지 4 배 더 김) 및 촉매성 도메인과 TALE 도메인 간의 확장된 N-말단 링커 (linker) 로 인해서, 본 발명자들은 보다 긴 코어 서열들 (32 또는 44-bp) 이 재조합을 위해서 필요할 것이라고 사료하였다. 실제로, 자발적 삭제 (Δ120*) 를 품고 있는 TALE 변형체의 예외를 사용하여서, 본 연구에서 확인된 최 (most) N-말단 절단부 변형체들이 32-bp 코어들에 비해서 최적의 성능을 보였다. 이러한 결과들은 TALEN들에서 보고된 것들과 일관되는데, TALEN들은 ZFN들과는 달리, DNA를 효율적으로 분해하기 위해서 상당히 긴 스페이서 서열들 (예를 들어서 TALEN들: 17 내지 20-bp, ZFNs: 5 내지 6-bp) 을 요구한다. 이러한 관찰사항들을 지지하면서, 짧은 코어 서열 (14-bp) 에 비해서 고유 N-말단 절단부 변형체들을 선택하는 것은 어떠한 클론들도 산출하지 않았음이 드러났다.

[0194] Gin-AvrXa7△128이 최적의 TALE 융합체로서 식별되었지만, 공중에서 입수가능한 TALE 어셈블리 키트를 사용하여서 생성된 합성 TALE 단백질들을 사용하는 후속 연구들은 Δ128 및 △120-기반 TALER들이 E. coli 내에서의 유사한 활성도를 보였다고 나타내었다. 이러한 설계된 TALE들은 서로 밀접하게 관련되고 자연적으로 발생한 Tal1c 및 PthXo1 TALE 단백질들로부터 유도된 키메라 단백질에 기초하였다. TALE들은 높은 상동성 (high homology) 을 공유하지만, 이들은 서로 일치하지 않는다. 잔기들 12 및 13 외측의 RVD 리피트들에서의 다형성들 (polymorphisms) 이 TALE 융합 활동도에 어떠한 영향도 주지 않는다고 드러났으며, 본 발명자들의 지식으로서는 DBD들 외측의 TALE 골격 내에서의 차이들을 체계적으로 평가하는 것은 지금까지 보고된 바가 없다. 증분적 절단부 라이브러리의 분석에 의해서 입증된 바와 같이, 작은 (minor) 아미노산 변형들은 특정 융합체의 활성도에 크게 영향을 줄 수 있다. 따라서, 본 발명자들이 관측한, Gin-AvrXa7△120과 합성 Gin-Avr15△120 간에서의 활성도에서의 어느 정도의 차이는 이전에 사용된 TALE 골격 아키텍처와 AvrXa 7 골격 간의 서열 변화들로부터 기인될 수 있다.

[0195] 합성 TALE들의 구축을 위해서 선호된 4 개의 RVD들 (NI: A, HD: C, NG: T, 및 NN: G) 은 성질상 가장 보편적이다; 그러나, 이러한 리피트들이 가장 가장 특이적인 RVD 모듈들을 나타내는지를 결정할 필요가 남아 있다. 26-리피트 AvrXa7 TALE의 경우에, 동일한 서열을 표적화하는 합성 버전 (version) 은 RVD 조성에서의 16 개의 변화들을 가질 것이다 (도 7). 이들이 성질상 보다 더 보편적이라고 발견되었기 때문에, 합성 사용을 위해서 선택된 이 4 개의 RVD들은 다른 RVD들보다 그들의 코그네이트 염기들에 대하여 보다 높은 친화도를 가질 것이라고 가설화되었다. 그렇다면, 합성 RVD 리피트들을 사용하여 생성된 TALE가 원 도메인들을 사용한 TALE보다 높은 DNA-바인딩 친화도를 가질 것이라고 가정하는 것은 합리적일 수 있다. RVD 친화도 이슈 (issue) 가 직접적으로 다루어지지 않았지만, 합성 리피트 어레이들을 포함한 TALER들이 원 AvrXa7 DBD을 포함한 구축체들보다 보다 활성적이라고 결정되었다. 합성 DBD들을 갖는 TALER들은 매우 적은 개수의 DBD들을 포함할지라도, 원 리피트들을 포함하는 구축체들보다 대략 2 배 높은 활성도를 보였다. 또한, 합성 어레이들에서 관측된 활성도에서의 이득은 표적을 벗어난 재조합에서의 임의의 증가와는 상관되지 않았다.

[0196] 몇몇 연구들은 TALE들이 그들의 표적 서열에서의 어느 정도의 비일치를 허용할 수 있다고 보였다. 이러한 사실들은 놀랍지 않은데, 그 이유는 특정 염기들과 긍정적으로 연관된 RVD들은 성질상 비-코그네이트 염기들을 허용하는 것으로 드러났기 때문이다. TALER들에 의해서 부여된 협력적 특이도는 잠재적 한계점들을 극복하는데 사용될 수 있다. 그러나, 촉매성 도메인이 재조합에 대한 특이도에 기여하기 때문에, 고도로 상동적인 게놈상의 서열들을 선택적으로 수정할 수 있는 설계자 TALER들이 역시 생성될 수 있다고 사료된다. 실제로, 리콤비나제 촉매성 특이도가 비자연적인 코어 부위들을 표적화하도록 효과적으로 재프로그램될 수 있다고 최근에 입증되었다.

[0197] **실례 2**

[0198] **신규한 0 번째 잔기 특이도의 선택**

[0199] 새로운 부류의 Tal-기반 DNA 바인딩 단백질들이 엔지니어링되었다. TAL (전사 활성제-유사) 이펙터들 (TALE들) 은 예측가능한 특이도를 갖는 새로운 부류의 DNA-바인딩 단백질들을 구성한다. TAL 이펙터들은 후술하는 박테리아가 병원성 결정인자들 (virulence determinants) 로서 역할을 하는 경우에, 상이한 이펙터 단백질들의 칵테일을 식물 세포들 내로 T3SS (type III secretion system) 을 통해서 전좌 (translocate) 시키는 속 (genus)

Xanthomonas의 Gram-네거티브 식물-병원성 박테리아에 의해서 사용된다. TAL들의 DNA-바인딩 특이도는 탠덤(tandem) 리피트들의 중앙 도메인에 의해서 결정된다. 각 리피트는 DNA 내의 하나의 염기 쌍 (bp) 의 인식을 제공한다. 리피트 모듈들의 재배열은 소정의 중요한 제약사항들로 목표된 DNA-바인딩 특이도들을 갖는 단백질들의 설계를 가능하게 한다. 예를 들어서, Tal 도메인으로 DNA 서열을 표적화하는 것의 가장 제약적인 특징은 Tal DNA 부위가 염기 T 및 때로 C로 시작해야 한다는 요건이다. G 또는 A 염기로 시작하는 바인딩 부위를 표적화하는 것은 -1 위치에서 가능하지 않았다. Tal-리콤비나제 활성도 선택들이 사용되어서 -1 번째 및 0 번째 RVD 영역들로 돌연변이들을 표적화시킴으로써 이러한 제한사항들이 없는 Tal DNA 바인딩 도메인들을 선택하였다. 이러한 발견의 실제적 결과들은 대단한데 그 이유는 이제 매 DNA 서열이 새로운 Tal 도메인들로 표적화될 수 있으며, 이로써 전사를 턴 온/업 또는 턴 오프/다운시키기 위한, TAL 전사 인자들에 대한 새로운 제약이 없는 접근법들이 가능하고, 유전자 기능을 녹아웃 (knock out) 시키기 위하거나 또한 상동적 재조합을 제어하기 위하거나 또한 우리 자신의 TAL 리콤비나제들 또는 다른 TAL 효소들을 표적화하기 위한, 표적 TAL 뉴클레아제들로의 새로운 제약이 없는 접근법들이 가능하기 때문이다.

[0200] (-1) 위치에서의 G 특이도에 대해서, 아미노산들 QWSG (SEQ ID NO: 209) 은 GmAvr15△128-합성 단백질의 (-1) 도메인 내에서 NNK 코돈 (codon) 전략을 사용하여서 먼저 랜덤화되었다. 3 회의 결과적인 라이브러리의 tal 리콤비나제 활성도 선택 후에, 표적화된 영역 내의 선택된 서열들 RSNG (SEQ ID NO: 210) 및 SRSG (SEQ ID NO: 211) 을 갖는 신규한 tal 바인딩 도메인들이 선택되었다. 이어서, 이들은 시작 클론에 의해서 인식된 모 (parental) T에 비해서 표적 서열의 0 번째 위치에서 G에 바인딩하는 것으로 드러났다. 초기에 선택된 QWSG (SEQ ID NO: 212) 와 중첩하는 이하에서 적색 (red) 으로 도시된 KQW 영역을 랜덤화시킴으로써 선택이 반복되었다. 이제, 선택된 SSR, SRA, SRC, 및 KRC 서열들을 갖는 클론들이 선택되었다. 모든 선택된 Tal 바인딩 도메인들이 G 치환을 보유하는 올리고들을 규정하기 위한 바인딩 연구들 시에 분석되었으며 이제 서열 G-ATAAACCCCCTCCAA (SEQ ID NO: 213) 에 우선적으로 바인딩한다고 드러났다. Tal 리콤비나제 활성도 선택은 이러한 동일한 서열을 사용하여서 수행되었다는 것이 주목된다. 시작 Tal 바인딩 단백질 GinAvr15△128은 T-ATAAACCCCCTCCAA (SEQ ID NO: 214) 을 바인딩한다. 선택된 돌연변이들을 보유한 Tal 뉴클레아제들의 후속 테스트는 이러한 서열들의 G 특이성을 입증하였으며 이로써 이러한 신규한 부류의 Tal들이 처음으로 개발되게 할 수 있었다. 선택된 서열들은 다른 종들로부터 유도된 Tal들로 이식가능하다 (portable).

표 6

선택들	
SEQ ID NO: 215	ATHEDIVGVGKQWSGARALEALLTDAGELRGPPLQ (-1 domain)
SEQ ID NO: 216	ATHEDIVGVGK**QWSG**ARALEALLTDAGELRGPPLQ (randomized AA in bold)
SEQ ID NO: 217	KQWSG-starting clone sequence
SEQ ID NO: 218	K*RSNG-selected to bind G*
SEQ ID NO: 219	K*SRSG-selected to bind G*

[0201]

SEQ ID NO: 220	ATHEDIVGVG**KQW**SGARALEALLTDAGELRGPPLQ
SEQ ID NO: 221	**KQW**SG- WT
	SSR-*selected to bind G*
	SRA-*selected to bind G*
	SRC-*selected to bind G*
	KRC-*selected to bind G*

[0202]

[0203] 선택들은 또한 표적 A에 대한 동일한 라이브러리를 사용하여 수행되었다. 본 연구에서, 서열들 PRG, PTR, 및 PKD이 선택되었다. 모든 선택된 Tal 바인딩 도메인들은 A 치환을 보유한 올리고들을 규정하는 바인딩 연구들에서 분석되었고 이제는 우선적으로 서열 A-ATAAACCCCCTCCAA (SEQ ID NO: 222) 을 바인딩하는 것으로 드러났다. Tal 리콤비나제 활성도 선택이 이러한 동일한 서열을 사용하여서 수행되었음이 주목된다. 시작 Tal 바인딩 단백질 GinAvr15△128은 T-ATAAACCCCCTCCAA (SEQ ID NO: 223) 을 바인딩한다. 선택된 돌연변이들을 보유한 Tal 뉴클레아제들의 후속 테스트는 이러한 서열들의 A 특이성을 입증하였으며 이로써 이러한 신규한 부류의 Tal들이 처음으로 개발되게 할 수 있었다. 바인딩 활동도에서의 후속하는 세분화들 (refinements) 은 N-말단 도메인의 랜덤 돌연변이유발 또는 리콤비나제 시스템 내에서의 0 번째 도메인 및 재선택 내에서의 KRGG (SEQ ID NO: 224) 서열의 표적 돌연변이유발에 의해서 달성될 수 있다.

[0204] **실례 3**

[0205] **선택들**

[0206] 상황 의존적인 RVD 선택들 및 새로운 특이도들을 갖는 RVD들의 선택들을 위해서, 이하에서 볼드체로 된 HD 서열을 랜덤화한 라이브러리들이 생성되었다: LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (프로토타입 RVD 서열; SEQ ID NO: 225)

[0207] 통상적으로 라이브러리는 이러한 2 개의 위치들에서 모든 아미노산들을 허용하지만, N, D, H, K, 및 Q 아미노산들로 한정된 라이브러리들이 때로 H 잔기에 대한 성공적인 치환물들이다. 이와 달리, S**HD**G (SEQ ID NO: 226) 및 AS**HD**GG (SEQ ID NO: 227) 영역들을 랜덤화한 보다 큰 라이브러리들이 상황 의존적인 특성들과 함께 고유한 RVD 특이도들의 선택을 가능하게 한다.

[0208] 이어서, Tal 리콤비나제 활성도 선택들이 표적화된 RVD 도메인 내에서의 새로운 특이도들의 선택을 신속하게 가능하게 한다. 최종 RVD들은 그들의 서열 인식에 있어서 고도로 모듈성이거나 상황 의존적이며 Tal 뉴클레아제들 및 전사 인자들을 생성하는데 사용될 수 있다.

[0209] 본 기술의 활용은 전사를 턴 온/업 또는 턴 오프/다운시키기 위한 TAL 전사 인자들로의 제한되지 않은 접근법, 유전자 기능을 녹아웃시키기 위하거나 상동적 재조합을 제어하기 위하거나, 툴들 및 치료학적으로 사용될 우리 자신의 TAL 리콤비나제들 또는 다른 TAL 효소들을 표적화하기 위한, 표적 TAL 뉴클레아제들로의 제한되지 않은 접근법을 포함한다.

[0210] 이러한 발견의 장점들 및 실제적 결과들은 대단한데 그 이유는 이제 매 DNA 서열이 본 발명의 새로운 Tal 도메인들로 표적화될 수 있으며 그들의 특이도들이 용이하게 최적화될 수 있기 때문이다.

[0211] **실례 4**

[0212] **티민이 아닌 5' 염기들을 수용하기 위한 TALE N-말단 도메인의 유도된 진화**

[0213] TALE 단백질들은 실제로 해당 임의의 DNA 서열과 바인딩되게 설계될 수 있다. 식물 비병원성 유전자들을 표적화하는 자연적 TALE 전사 인자들 (TALE-TF들) 에 대한 DNA 바인딩 부위들은 5' 티미딘을 갖는다. 합성 TALE-TF들 또한 이러한 요건을 갖는다. 최근의 구조적 데이터는 N-말단 도메인 (NTD) 과 표적 서열의 5 'T 간에 상호작용이 존재한다고 알리고 있다. 최근 TALE 뉴클레아제 (TALEN) 문헌 조사에 따르면, 표적 서열의 첫번째 염

기, N_0 잔기의 중요성에 대한 모순되는 데이터들이 있었다. 또한, TALE 리콤비나제들 (TALE-R들) 의 활성도들에 대한 N_0 염기의 영향에 대한 어떠한 연구도 없었다. 본원에서는, N_0 염기의 영향이 TALE-R들, TALE-TF들, MBP (maltose binding protein) 와의 융합들로서 발현된 TALE DNA-바인딩 도메인들 (MBP-TALE들) 및 TALEN들의 바인딩 영역들 내에서 정량화된다. 이러한 TALE 플랫폼들 각각은 개별적인 N-말단 아키텍처 및 C-말단 아키텍처를 가지지만, N_0 잔기가 티미딘일 때에 모두가 최고의 활성도를 보였다. 이러한 플랫폼들에서 효과적인 TALE들을 구축하기 위한 규칙들을 단순화시키기 위해서, 그리고 임의의 랜덤한 DNA 서열에서 정밀 게놈 엔지니어링 애플리케이션을 가능하게 하기 위해서, 본 발명자들은 본 발명자들의 최근에 개발된 TALE-R 시스템을 사용하여서 구조-가이드된 (structure-guided) 활성도 선택을 고안하였다. 5'G을 갖는 TALE 바인딩 부위들 상에서 고도로 활성적이며 선택적인 TALE-R 활성도를 제공한 신규한 NTD 서열들이 식별되었으며, 임의의 5' N_0 잔기의 일반적인 표적화를 허용하는 추가적인 도메인 서열들이 선택되었다. 이러한 도메인들은 TALE-TF, MBP-TALE 및 TALEN 아키텍처들 내로 입력되며 비-T 5' 잔기들을 갖는 표적 서열들 상에서 와일드-타입 NTD가 보인 것보다 큰 활성도를 일관되게 보였다. 신규한 NTD들은 골든 게이트 TALEN 어셈블리 프로토콜과 양립하며, 이제 임의의 DNA 서열을 인식하는 DNA-바인딩 단백질들, TALE 전사 인자들, 리콤비나제들, 및 뉴클레아제들의 효율적인 구축이 가능하게 하며, 이로써 가장 자연적인 TALE 단백질들을 제약하는 5'T 규칙에 대하지 않고서 DNA 상의 TALE-기반 단백질들의 정밀하고 비제약적인 포지셔닝을 가능하게 한다.

[0214] 다음과 같은 "재료 및 방법들"이 본 실례에서 사용되었다:

[0215] **올리고뉴클레오타이드들**

[0216] 프라이머들 및 다른 올리고뉴클레오타이드들 (이하의 표 7) 이 Integrated DNA Technologies (San Diego, CA) 로부터 주문되었다.

표 7

프라이머	SEQ ID NO:	서열
KXXG Lib Rev	228	TCTCAACTCCCCCGCCTCCGTGAGCAAGGCCTCCAGAGCGCGTGCC CCMNNMNNTTTGCCGACGCCAACGATGTCTTCGTG
KXXXX Lib Rev	229	TCT CAA CTC CCC CGC CTC CGT GAG CAA GGC CTC CAG AGC GCG TGC MNN MNN MNN MNN TTT GCC GAC GCC AAC GAT GTC TTC GTG
XXXSG Lib Rev	230	CCCGCCTCCGTGAGCAAGGCCTCCAGGGCGCGTGCGCCGGAMNNM NNMNNGCCGACGCCAACGATGTCTTCGTGTGTCGC
KRGG Lib Rev	231	GGC ACC CGT CAG TGC ATT GCG CCA TGC ATG CAC TGC CTC CAC TGC GGT CAC MNN MNN MNN MNN TGC AAT CTT GAG AAG TTG GCC TGT GTC

[0217]

[0218]

Goldy TALEN fwd	232	AGAGAGAGAAGAAAATGAGATCTCCTAAGAAAAAGAGGAAGGTGC AGGTGGATCTACGCACGCTCGGCTAC
NTD-dHax3 Fwd	233	AGGAAGAAGAGAAGCATGAGATCTCCTAAGAAAAAGAGGAAGGTG ATGGTGGACTTGAGGACACTCGGTTA
NTD-dHax3 Rev	234	AAGAGAAGAAGAAGAAGCATTGCGCCATGCATGCACTGCCTCTA
pTal127 Not1 fwd	235	CCC GCC ACC CAC CGT GC
N-Term Sph1	236	TGC TCT ATG CAT GCA CTG CCT CC
pTAL127-SFI Fwd	237	AGA GAA GAG AAG AGA AGG CGC CCG CGG CCC AGG CGG CCT CGG GAT CCC CTC GGC CTC CGC GCG CCA AG
pTAL127-SFI +95 Rev	238	AGA GAG AGA GAG AGA GTC TAG AGG CCG GCC TGG CCG CTC ATC CCG AAC TGC GTC ATG GCC TCA TC

[0219]

pTAL127 Xba +28 Rev	239	GCC CCA GAT CCT GGT ACG CTC TAG AGG
Avr 5'A biotin hairpin	240	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATTGG GTT TTC CCAAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'T biotin hairpin	241	5'Biosg/ATC TTA GCA CCT GGT TGG AGG GGG TTT ATAGG GTT TTC CCTAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'G biotin hairpin	242	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATCGG GTT TTC CCGAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'C biotin hairpin	243	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATGGG GTT TTC CCCAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
CCR5-inner fwd	244	TTAAAAGCCAGGACGGTCAC
CCR5-inner rev	245	TGTAGGGAGCCCAGAAGAGA
CCR5-outer fwd	246	ACAGTTTGCATTCATGGAGGGC
CCR5-outer rev	247	CCGAGCGAGCAAGCTCAGTT
CCR5-indel fwd	248	CGCGGATCCCCGCCCAGTGGGACTTTG

CCR5-indel rev2	249	CCGGAATTCACCTGTTAGAGCTACTGC
pGL3 NTD stuffer fwd	250	AGA GAG AGA GAG AGG CGG CCG CCC TAC CAG GGA TTT CAG TCG ATG TAC ACG TTC
pGL3 NTD stuffer rev	251	AAG AAG AAG AAG GAA GAG AAG TAG GCC TGT CAT CGT CGG GAA GAC CTG CGA CAC CTG C
pgl3 5X Avr Xho1	252	ACTGCTATCCGAGTATAAACCCCCTCCAACCAGGTATAAACCCCCT CCAACCAGGTATAAACCCCCTCCAACCAGGTATAAACCCCCTCCAA CCAGGTATAAACCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
AvrXa7 32G A F	253	TTAATTAAGAGTCTAGAttagcacctggttggaggggggtttatTgcttcaTCCAAAACC ATGGTTTACAGggttccAATAAACCCCCTCCAACCAGGTGCTAAAGAT CTAGGAGGAATTTAAAATGAG
AvrXa7 32G A R	254	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATTgc aaccCTGTAAACCATGGTTTTGGAtgaagcAATAAACCCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 32G G F	255	TTAATTAAGAGTCTAGAttagcacctggttggaggggggtttatCgcttcaTCCAAAACC ATGGTTTACAGggttccGATAAACCCCCTCCAACCAGGTGCTAAAGAT

[0220]

		CTAGGAGGAATTTAAAATGAG
AvrXa7 32G G R	256	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATCgc aaccCTGTAAACCATGGTTTTGGAtgaagcGATAAACCCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 32G C F	257	TTAATTAAGAGTCTAGAttagcacctggttggaggggggtttatGgcttcaTCCAAAACC ATGGTTTACAGggttccCATAAACCCCCTCCAACCAGGTGCTAAAGAT CTAGGAGGAATTTAAAATGAG
AvrXa7 32G C R	258	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATGgc aaccCTGTAAACCATGGTTTTGGAtgaagcCATAAACCCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC

[0221]

루시페라제. 벡터 = pgl3 basic. XhoI/SphI		
포함하는 순방향 표적	SEQ ID NO:	
5x Avr15 n-1c xhoF:	259	actgctatctcgagcTATAAACCCCCTCCAACCAGGcTATAAACCCCCTCCAACC AGGcTATAAACCCCCTCCAACCAGGcTATAAACCCCCTCCAACCAGGcTA TAAACCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=A n-1c	260	actgctatctcgagcAATAAACCCCCTCCAACCAGGcAATAAACCCCCTCCAACC AGGcAATAAACCCCCTCCAACCAGGcAATAAACCCCCTCCAACCAGGcA ATAAACCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=C n-1c	261	actgctatctcgagcCATAAACCCCCTCCAACCAGGcCATAAACCCCCTCCAACC AGGcCATAAACCCCCTCCAACCAGGcCATAAACCCCCTCCAACCAGGcC ATAAACCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=G n-1c	262	actgctatctcgagcGATAAACCCCCTCCAACCAGGcGATAAACCCCCTCCAACC AGGcGATAAACCCCCTCCAACCAGGcGATAAACCCCCTCCAACCAGGcG ATAAACCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
루시페라제 역방향 프라이머:	263	TCAGAAACAGCTCTTCTTCAAATCT

[0222]

[0223] **TALE-R NTD 진화 플라스미드들의 생성**

[0224] 이전에 보고된 TALE-R 시스템이 본 연구에서 적응되었다. 간략하게, pBCS (chloramphenicol 및 carbenicillin 내성 유전자들을 포함함) 가 HindIII/Spe1을 사용하여서 다이제션되었다. 트윈 (twin) 리콤비나제 부위들을 포함하는 스터퍼 (Avr X, 여기서 X는 N_0 염기) 가 HindIII/XbaI로 다이제션되고 벡터 내로 결찰되어서 분할된 베타-락타마아제 유전자를 생성하였다. 이어서, pBCS AvrX가 BamHI/Sac1로 다이제션되고 Gin127-N-스터퍼-Avr15가 BamHI/Sac1로 다이제션되어서 벡터 내로 결찰되어서 Gin127-N-스터퍼-Avr15-X를 생성하였다. 이 스터퍼는 N_{-1} TALE 헤어핀에서의 진화들을 위한 Not1/Stu1 및 N_0 TALE 헤어핀에서의 진화들을 위한 Not1/Sph1로 다이제션되었다.

[0225] **TALE NTD 진화 라이브러리들의 생성**

[0226] 프라이머 ptal127 Not1 fwd 및 역방향 프라이머들 KXXG lib rev 또는 KXXXX lib rev가 사용되어서 N_{-1} TALE 헤어핀에서 N-말단 변형체들을 생성하였으며 이후에 Not1/Stu1로 다이제션되고 이어서 다이제션된 Gin127-AvrX 내로 결찰되었다. 순방향 프라이머 ptal127 Not1 fwd 및 역방향 프라이머 KRGG Lib Rev이 사용되어서 N_0 TALE 헤어핀에서 돌연변이들을 갖는 라이브러리를 PCR 증폭하였다. 이어서 Not1/Sph1로 다이제션되고 Not1/Sph1로-다이제션된 Gin127-AvrX 내로 결찰되었다.

[0227] **TALE-R NTD 진화 분석**

[0228] 1 회 결찰물들 (round 1 ligations) 이 에탄올 석출되어서 일렉트로컴피턴트 (electrocompetent) Top10F' 세포들로 형질변환되고 이어서 1 시간 동안 SOC 내에서 회복되었다. 이 세포들은 100 mg/ml chloramphenicol을 함유한 100 ml Super Broth (SB) 배지 내에서 하룻밤 동안 성장되었다. DNA가 표준 절차를 통해서 단리되었다. 결과적인 플라스미드 DNA (Rd 1 인풋 (input)) 가 일렉트로컴피턴트 Top10F' 세포들로 형질변환되었고; 이 세포들은 100 mg/ml carbenicillin 및 100 mg/ml chloramphenicol를 함유한 100 ml SB 내에서 하룻밤 동안 성장되었다. 플라스미드 DNA가 표준 절차를 통해서 단리되었다. 1 회 결과물 (round 1 output) 이 Not1/Xba1로 다이제션되고 상보적 스티키 단부들 (sticky ends) 을 갖는 Gin127-AvrX 벡터 내로 결찰되었다. 이러한 프로토콜이, 컨센서스 서열이 관찰되고 클론들이 특성화될 때에, 3 회 내지 4 회 반복되었다.

[0229] **N-말단 TALEN 활성도의 측정**

[0230] 각각이 가능한 염기를 포함하는 4 개의 TALEN 쌍들이 골든 게이트 프로토콜을 사용하여서 생성되었다. 융합 A 및 B 플라스미드들이 제 2 골든 게이트 반응을 통해서 Goldy TALEN (N △152/C +63) 골격 내로 직접적으로 결찰되었다. NTD가 pCAG 벡터를 Bgl11/Nsil로 다이제션하고 이를 Bgl11/Nsil로 다이제션된 PCR 증폭된 NTD 내로 결찰시킴으로써 수정되었다. TALEN 쌍들 (각 TALEN/웰 50-75 ng) 이 1.5 x 10^4 세포들/웰의 밀도로 96-웰 플레이트들의 웰들 내에서 HeLa 세포들 내로 형질주입되었다. 형질주입 후에, 세포들이 24 시간 동안 37 ℃ 인큐베이터 내에 배치되고 이어서 2 일 동안에 30 ℃ 로 이동되었으며 이어서 24 시간 동안 37 ℃ 로 이동되었다. 게놈 DNA가 공개된 프로토콜을 따라서 단리되고 DNA 돌연변이 레이트들이 Cell Surveyor 분석 및 서열화를 통해서 정량화되었다. 세포 분석을 위해서, 게놈 DNA는 네스트된 (nested) PCR에 의해서, 먼저는 프라이머들 CCR5 외측 fwd/CCR5 외측 rev로 그리고 이어서 CCR5 내측 fwd/CCR5 내측 rev로 증폭되었다. 인델들의 서열화를 위해서, 두 번째 PCR이 CCR5 인델 fwd/CCR5 인델 rev로 수행되었다. 이어서, 단편들이 BamHI/EcoR1로 다이제션되고 상보적 다이제션을 갖는 pUC19 내로 결찰되었다.

[0231] **TALE-TF들 및 루시페라제 분석**

[0232] 리콤비나제 선택으로부터의 변형체 NTD들이 프라이머들 ptal127 SFI fwd 및 N-말단 Sph1를 사용하여서 PCR 증폭되었다. 이러한 PCR 산물은 증폭되고 Not1/Stu1로 다이제션되고 이어서 pTAL127-SFI Avr15 내로 결찰되었으며, pTAL127-SFI Avr15은 N-말단-수정된 TALE의 pTAL127-SFI Avr15로부터 pcDNA 3.0 VP64 내로의 전달을 실현하는 트윈 SFI-1 다이제션 부위들을 포함한다. 대응하는 TALE 바인딩 부위들이 루시페라제 유전자의 업스트림에 있는 pGL3 Basic 벡터 (Promega) 내로 복제되었다. 각 분석을 위해서, 100 ng pcDNA가 5 ng pGL3 벡터 및 1 ng pRL Renilla 루시페라제 제어 벡터와 함께 96-웰 플레이트의 웰 내의 HEK293t 세포들 내로 Lipofectimine 2000 (Life Technology) 를 사용하여서 제조자의 사양에 따라서 공동-형질주입되었다. 48 시간 후에, 세포들이 워싱되고 파괴되고 루시페라제 활성도가 Veritas Microplate 루미노미터 (luminometer) (Turner Biosystems) 상에

서 이중-루시페라제 리포터 시스템 (Promega) 을 사용하여서 분석되었다. 형질주입들은 삼중으로 수행되고 그 결과들이 평균화되었다.

[0233] **MBP-TALE 분석**

[0234] 바이오티닐화된 (biotinylated) 올리고뉴클레오타이드들에 바인딩한 MBP-TALE의 친화도 분석이 이전에 기술된 프로토콜을 사용하여서 수행되었다. 간략하게는, AvrXa7 TALE 도메인들이 XL1-Blue 세포들 내의 pMAL MBP-AvrXa7 플라스미드로부터 발현되었고 아밀로우즈 수지 (amylose resin) 상에서 정제되었다. 수정된 잔기들을 갖는 표적 AvrXa7 표적 부위를 포함하는 바이오티닐화된 올리고뉴클레오타이드들이 사용되어서 TALE-바인딩 활성도가 샌드위치 효소-링크된 (linked) 면역흡착 (immunosorbent) 분석 포맷 내에서 결정되었다. MBP 치환체를 표적화하는 항체들이 분석법 개발을 위해서 사용되었다.

[0235] **결과들**

[0236] **5'T 규칙의 예비 분석**

[0237] PthXo7 DNA 서열에 바인딩된 TALE 단백질의 최근 결정 구조는 DNA 기질 (N_0 염기) 의 접촉된 영역의 5' 단부에서의 티미딘과 N_{-1} 헤어핀에서의 W232 간의 고유한 상호작용을 보였다. 이러한 연구는 TALE 코드가 먼저 암호해독될 때에 보고된 이전에 확립된 5'T 규칙에 대한 구조적 기초를 제공하였다 (도 18a 및 도 18b). TALEN들의 표적 서열의 첫 번째 염기의 중요성에 대한 상충되는 데이터가 있다. 표적 DNA 내에서의 5'T에 대한 요건이 Gin32G 코어를 측접하는 모든 가능한 5' 잔기들을 갖는 4 개의 AvrXa7 바인딩 부위들을 포함하는 4 개의 분할된 베타 락타마아제 TALE 리콤비나제 선택 벡터들을 사용하여서 먼저 TALE-R들의 맥락에서 평가되었다 (도 18c). 이어서, TALE-TF들에 의한 N_0 잔기의 인식이 각 가능한 5' 잔기를 포함하는 인식 부위들을 갖는 pentamer AvrXa7 프로모터 영역을 갖는 4 개의 루시페라제 리포터 벡터들을 사용하여서 평가되었다 (도 18d). 5'T 이외의 염기에 있어서, 본 발명자들은 5'T를 갖는 서열에 비해서 활성도가 TALE-R들에서는 100 배 이상이 감소하고 TALE-TF들에서는 1000 배 정도로 감소한다는 것을 알게 되었다. 이러한 감소는 특히 크게 단축된 C-말단 도메인 (CTD) 의 존재 시에, 5'T 바이어스를 제거한다고 보고된 이러한 키메라들의 C-말단 아키텍처들에서의 변형들에도 불구하고 관찰되었다. 효소-링크된 면역흡착 분석 또한 비-T5' 잔기들을 갖는 표적 올리고뉴클레오타이드들을 향한 MBP-TALE DNA-바인딩 단백질들의 친화도가 감소하였음을 말하였다 (도 18e). 마지막으로, 비-T5' 뉴클레오타이드들을 갖는 표적들에 대한 와일드-타입 NTD들로 설계된 TALEN들의 활성도 조사는 5'T을 갖는 것들에 비해서 활성도가 10 배까지 감소하였음을 보였다 (도 18f). 이 결과들은 5'T가 리콤비나제들, 전사 인자들, 뉴클레아제들 및 간단한 DNA-바인딩 단백질들의 맥락에서 최대로 효과적인 TALE 도메인들에 대한 중요한 설계 파라미터임을 나타낸다.

[0238] **비-T 5' 잔기들을 수용하기 위한 TALE NTD의 진화**

[0239] DNA 인식을 위한 보다 유연한 시스템을 생성하기 위해서, 최근에 개발된 TALE-R 선택 시스템이 5'T 제약사항을 제거하도록 TALE의 NTD을 진화시키는데 사용될 수 있다고 가정되었다 (도 23). 라이브러리들이 랜덤화된 잔기들 K230 내지 G234을 사용하여서 생성되었고, 각 가능한 5' 염기에 대한 (against) 활성도를 갖는 TALE-R들이 몇 회의 선택 후에 단리되었다 (도 19a 내지 도 19c). 가장 활성적이라고 선택된 클론들은 K230 및 G234의 강한 보존을 보였으며; 전자 (former) 는 DNA 포스페이트 백본 (phosphate backbone) 과 접촉할 수 있으며, 후자는 헤어핀 루프 형성에 영향을 줄 수 있다 (도 24). 라이브러리 K230 내지 W232의 경우에, K230S가 빈번하게 관측되었지만 개별적으로 분석된 거의 모든 변형체들 내에서 K230R 또는 K230 변형체들보다 매우 낮은 활성도를 보였다. W232에서 R232로의 돌연변이로 관측된 몇몇 클론들 중 일 클론 (NT-G) 이 5'T로부터 5'G로의 선택도의 큰 이동을 보였으며; 이 서열은 이 영역에서 최근에 기술된 Ralstonia TALE 단백질의 NTD의 것과 유사하였다. Ralstonia NTD은 식물 전사 인자 리포터 유전자 조절의 맥락에서는, 그의 기질 내에서 5'G을 선호한다고 보고되었다 (단백질 정렬에 대해서는 도 25 참조). 잔기 R232는 5'G에 대한 NT-G의 충실도에 의해서 표시된 바와 같이 G 염기와 특이적으로 접촉한다. 이러한 5'G에 대한 NT-G의 선호도는 5'T에 대한 와일드-타입 도메인의 특이도와 비교할만하다. 5'A 또는 5'C에 대해 특이적인 NTD 변형체들은 유도될 수 없지만, 임의의 5' 잔기를 갖는 기질들을 허용하고 높은 활성도를 유지하는 K265-G268 N_0 헤어핀을 닮은 허용가능한 (permissive) NTD, NT-αN가 획득되었다. 이러한 변형체는 와일드-타입 NTD에 비해서 DNA 포스페이트 백본과의 개선된 비-특이적 접촉을 하였으며 이로써 특이적 5' 잔기와 접촉하지 않고서도 TALE-DNA 복합체의 전반적인 바인딩을 향상시켰다고 가정되었다. 단축된 헤어핀 구조가 5'A 또는 5'C 잔기들에 대한 특이도를 갖는 변형체들의 선택을 가능하게 할 것이라고 가정되었다. Q231-W232에서 랜덤화되고 잔기 233가 삭제된 라이브러리는 추정상의 (putative) DNA-바인

딩 루프를 단축시키도록 설계되었다. 리콤비나제 선택은 다수의 클론들에서 높은 활성도를 갖는 고도로 보존된 Q231Y 돌연변이를 보였다 (도 19d). 특히, NT-βN이 5'A, C 및 G을 갖는 기질들에 대한 개선된 활성도를 보였지만, 와일드-타입 NTD을 갖는 TALE들에 비해서 5'T 기질들에 대한 활성도가 감소되었다 (도 19e).

[0240] **진화된 TALE NTD들의 애플리케이션들**

[0241] 설계자 TALE 융합 단백질 애플리케이션들 내에서의 진화된 NTD들의 이식성을 평가하기 위해서, 최적화된 NTD들이 TALE-TF들, MBP-TALE들 및 TALEN들 내에 혼입되었다. NT-G, NT-αN 및 NT-βN 도메인들을 갖는 TALE-TF들은, NT-T 도메인을 갖는 TALE-TF와 비교하였을 때에, 5'T 잔기가 없는 루시페라제 표적 유전자 보유 오퍼레이터 (operator) 부위들의 전사적 활성도가 400 내지 1500 배 증가하였음을 보였다. NT-G-기반 TF는 TALE-R 선택 시스템에서 관측하였을 때에 5'G 선택성을 보유하였다. 모든 5' 뉴클레오타이드들에 대한 (against) NT-αN-기반 TF 및 NT-βN-기반 TF의 활성도들은 리콤비나제 포맷에서 관측된 상대적 활성도를 추적하였다 (도 20). MBP-TALE들 또한 와일드-타입 MBP-TALE가 보인 것보다, 5'T을 구비하지 않은 부위들을 갖는 표적 올리고뉴클레오타이드들에 대한 보다 큰 상대적 바인딩 친화도를 보였으며 (도 26), 이는 이 선택된 도메인들이 비-티민 5' 염기들에 대한 허용오차 (tolerance) 또는 이 염기들의 인식을 개선하였다는 추가 입증을 제공한다.

[0242] 최적화된 NTD들 중 4 개가 이어서 Goldy TALEN 골격 내로 들어갔다. 이러한 실험들을 위해서, 4 개의 기질들이 CCR5 유전자의 Δ32 좌위의 상황 내에서 구축되었다 (도 21a). 각 기질은 상이한 5' 잔기를 포함하였다. 실험들은 유전자 에디팅 활성도를 벤치마킹하기 위해서, 5'T에 대한 특이도를 갖는 와일드-타입 (NT-T) 및 dHax3 NTD들 (dHax3는 Xanthomonas campestris로부터 단리된 통상적으로 사용된 NTD 변형체임) 을 갖는 TALEN들을 포함하였다. 기질 TALEN 쌍들은 변형체 NTD들의 활성도 향상 기여사항들을 결정하는 RVD 상동성을 가능한 한 많이 (50 내지 90%) 보유하도록 설계되었다 (도 21a).

[0243] TALEN들의 활성도들이 서열화에 의해서 그리고 세포 분석을 사용하여서 양자에 의해서 분석되었다. 선택된 도메인들은 TALEN 함유 와일드-타입 도메인의 활성도들에 비교하여서, 비-T5' 잔기들에 대한, 2 내지 9 배의 유전자 에디팅 활성도 증가를 보였다 (도 21 및 도 27). 이 활성도는 와일드-타입 또는 dHax3 NTD을 갖는 TALEN 쌍 T1/T2 에 대해서 최고였다. TALEN 쌍 기질은 NT-αN, NT-βN 및 NT-G을 갖는 TALEN들에 의해서 가장 효과적으로 처리되었는데, NT-T에 비해서 이들은 2.0 내지 3.5 배의 향상을 보였다. NT-αN은 TALEN 쌍들 A1/A2 및 C1/C2에 대해서 와일드-타입 NT-T보다 각기 9 배 및 2 배 높은 활성도를 보였다. 5' 잔기에서의 불일치의 영향이 TALE-TF 및 TALE-R 골격들에서보다 TALEN들에서 보다 온화하였지만, 최적화된 NTD들은 유전자 에디팅 실험들에서 사용되는 경우에 TALEN 활성도를 크게 향상시켰다.

[0244] **논의사항**

[0245] 모두는 아니지만 대부분의 이전의 연구들은 티미딘이 최적의 TALE DNA-바인딩 도메인들의 설계 시에 5'에 가장 가까운 (5'-most) 잔기로서 요구된다고 제시하였다. 본 명세서에서 기술된 분석들은 작용적 TALE 융합 단백질들을 구축하기 위해서 티미딘이 최적이며 일부 경우들에서는 매우 중요하다는 것을 드러낸다. 따라서, 이러한 요건은 TALE 전사 인자, 뉴클레아제 및 리콤비나제 키메라들로 효과적으로 표적화될 수 있는 서열들에 대한 제약사항들을 부여한다. 이러한 요건이 이론상 유전자 녹아웃 (knockout) 을 유도하기 위한 TALEN들의 사용에 작은 제약사항들을 부여하지만, 그들의 넓은 스페이서 영역 허용오차 (tolerance) 가 주어진다면, 임의의 5' 잔기를 수용할 수 있는 NTD들은 효과적인 TALE 구축을 위한 규칙들을 더 단순화시키고 게놈 엔지니어링 및 조사를 위한 정밀한 TALE 배치를 요구하는 애플리케이션들 (예를 들어서, TALEN들을 사용하는, 규정된 염기 쌍에서의 DNA의 정밀한 절단 (precise cleavage), TALE-리콤비나제들을 통한 심리스 (seamless) 유전자 삽입 및 교환, 그들의 작용적 역할을 조사하기 위해서 특이적 내생성 DNA 서열들로부터 자연적 DNA-바인딩 단백질들을 제거하는 것, 경로 엔지니어링을 위한 직교 (orthogonal) 전사 인자들의 개발, 전사 인자 배치가 핵심인 자연적 유전자 및 합성 유전자의 상생적 활성화, 및 임의의 다른 애플리케이션들) 을 개선시킬 것이다. DNA-기반 나노기술에서의 다른 사용들은 특이적 DNA-바인딩 단백질들로 DNA 나조구조물들/오리가미 (DNA nanostructures/origami) 를 데코레이션 (decoration) 하는 것을 포함한다. 여기서, 특이적 부위들로의 표적화는 DNA 폴딩/구조에 기초하여서 제약되며 따라서 임의의 부위로 바인딩할 수 있는 것이 중요하다. DNA-바인딩 단백질들로 이러한 구조물들 및 디바이스들을 정교화하는 것 (elaboration) 은 기능을 확장시키는 것에 대한 흥미로운 접근법일 것이다. 실제로, 모든 표적화 제약사항들이 제거되면 DNA 바인딩 단백질들 및 그들의 융합들을 위한 수많은 애플리케이션들을 상상하는 것은 어렵지 않다. 이러한 잠재적 애플리케이션들에 의해서 고무되어서, 본 발명자들은 임의의 염기에서 개시되는 부위들의 표적화를 가능하게 하는 NTD들을 개발하고자 하였다.

[0246] 최근 개발된 TALE-R 시스템이 사용되어서 TALE의 NTD를 진화시켜서 5'-T 제약사항을 제거하였다. 3 회

선택으로, 5'G에 대한 특이도를 갖는 NTD가 획득되었다. 다수의 선택들이 5'A 또는 5'C 중 어느 하나를 인식한 변형체들을 획득하기 위한 시도들에서 수행되었다. G230-K234 헤어핀이 역위되며 (inverted), K230-G234/ins232 헤어핀이 확장되고, K265-G268 N_0 헤어핀의 수정이 시도되었으며, 랜덤 돌연변이유발 라이브러리들이 평가되었다. 이러한 전략들 중 어느 것도 5'A 또는 5'C을 갖는 표적 서열들에 대한 친화도를 갖는 NTD들을 산출하지 않았지만, 본 발명자들은 허용가능한 친화도로 5'A 및 5'C 잔기들 모두를 갖는 기질들을 인식한 삭제를 갖는, NTD, ＮＴ-βＮ을 식별하였다. NTD들 NT-T 및 NT-G이 보인 강한 선택 선호도 및 NT-G 내에서의 R232 및 NT-T 내에서의 W232의 중요성은 DNA 인식 서열의 5' 말단 잔기와 이러한 아미노산들의 특이적 상호작용들 때문일 가능성이 높다. Ralstonia solanacearum TALE가 5'G을 엄격하게 요구한다고 최근에 보고되었으며, NT-G과의 서열 정렬은 NT-G 내의 232와 유사한 위치에서 아르지닌 (arginine) 을 함유한 비교할만한 N_{-1} 헤어핀일 것처럼 보이는 것을 나타낸다고 보고되었다 (도 25). NTD들 Brg11과 NT-T 간의 높은 구조적 상동성으로 인해서, 티민에 대한 Ralstonia TALE NTD의 선호도를 간단한 아르지닌에서의 트립토판으로의 변이에 의해서 수정할 수 있거나, NT-αN 또는 ＮＴ-βＮ 도메인들을 이러한 관련된 단백질 내로 그래프팅 (grafting) 함으로써 특이도를 제거할 수 있다. 또한 아르지닌-구아닌 상호작용들 (arginine-guanine interactions) 이 진화된 징크 핑거 도메인들에서 흔하다는 것을 주목하는 것이 흥미롭다.

[0247] 선택된 변형체 NTD들이 TALE-TF들, MBP-TALE들 및 TALEN들 내로 성공적으로 들어갔으며 대체적으로 리콤비나제 진화 시스템으로부터의 데이터에 기초하여서 예상된 활성도 및 특이도를 부여하였다. 최적화된 NTD들을 갖는 TALE-TF들은 비-T 5' 잔기들을 갖는 AvrXa7 프로모터 부위들에 대한 (against) NT-T의 활성도에 비해서 400 내지 1500 배로 TALE 활성화를 증가시켰다. TALEN들 내로 혼입된 때에, 비-T 선택성을 갖는 본 발명의 NTD은 5'A, C 또는 G을 갖는 기질들에 대해서 NT-T 도메인의 활성도에 비해서 활성도를 2 내지 9 배 증가시켰다. TALEN 유전자 에디팅에서의 증가는 대체적으로 TALE-R 및 TALE-TF 구축체들 내에서 관측된 활성도에서의 증가와 상관되었다. NT-G의 특이도 및 높은 활성도는, TALEN 쌍들 A1/A2, C1/C2, 및 T1/T2을 사용하는 분석 시에서의 낮은 활성도에 의해서 입증되는 바와 같이, 유지되며, NT-αN 및 ＮＴ-βＮ의 대체적으로 높은 활성도가 또한 TALEN Δ152/+63 아키텍처 내로 부여되었다.

[0248] 합성 TALE RVD 도메인들을 갖는 교번적으로 (alternatively) 절단된 TALE들은 DNA 기질 내의 5'T을 요구하지 않는다고 최근에 보고되었다. 보고된 Δ143, +47 절단부는 Goldy TALE-TF로서 구성되며, Δ127, +95 절단부에서보다 AvrXa7 기질 상에서의 실질적으로 보다 낮은 활성도가 관측되었으며, 이 Δ127, +95 절단부는 다른 연구들에 의해서 가장 통상적으로 사용되어 왔으며, 본 연구에서 사용된 절단부 세트이다 (도 29). 따라서, 보고된 결과들에서의 차이는 사용된 절단된 아키텍처들로 인한 것일 수 있다.

[0249] 요약하면, 설계된 TALE들의 활성도 및 바인딩을 위한 DNA 기질 내의 5' 티미딘의 중요성은 TALE-R, TALE-TF, MBP-TALE들 및 TALEN 키메라들의 맥락에서 결정되었다. 표적화된 돌연변이유발 및 TALE-R 선택은 기질 DNA의 5'에 가장 가까운 염기로서 티민이 아닌 다른 염기들을 인식하는 TALE NTD들을 엔지니어링하기 위해서 적용되었다. 본 발명에서 개발된 엔지니어링된 TALE 도메인들은 모듈성을 보였으며 TALE-TF 및 TALEN 아키텍처들에서 고도로 활성적이었다. 이러한 신규한 NTD들은 현 TALE-R들에 의해서 표적화될 수 있는 부위들의 개수를 약 15 배 정도만큼 증가시켰는데, 이 현 TALE-R들은 그들의 바인딩 부위들에 대한 엄격한 기하학적 요건들을 가지며 N_0 염기의 일치에 고도로 민감하다. 또한, 이들은 이제 임의의 DNA 서열에서의 TALE DBD들 및 TALE-TF들의 정밀한 배치를 가능하게 하여서 유전자 조절, 내생성 DNA-바인딩 단백질들의 분리 및 정밀한 바인딩이 핵심일 수 있는 합성 생물학 애플리케이션들을 실현할 수 있다. 원 NTD에 기초한 TALEN들이 N_0 염기 치환들의 가변하는 정도의 허용오차를 보이지만, 데이터는 본 발명에서 보고된 신규한 NTD들도 또한 자연적 NTD-기반 TALEN들에 비교하여서 임의의 N_0 염기를 사용하는, 보다 높은 효율적인 유전자 에디팅을 실현한다고 나타내었다.

[0250] **실례 5**

[0251] **키메라 징크 핑거 리콤비나제들**

[0252] 다음의 재료들 및 방법이 사용되었다.

[0253] 분할된 유전자 재어셈블리 벡터 (pBLA) 가 pBluescriptII SK (-) (Stratagene) 으로부터 유도되고 lac 프로모터의 제어 하에서 chloramphenicol 내성 유전자 및 인터럽트된 (interrupted) TEM-1 p 락타마아제 유전자를 포함하도록 수정되었다. ZFR 표적 부위들이 전술한 바와 같이 도입되었다. 간략하게는, GFPuv (Clontech) 가 프라이머들 GFP-ZFR-XbaI-Fwd 및 GFP-ZFR-HindIII-Rev로 PCR 증폭되고 pBLA의 SpeI 및 HindIII 제한 부위들 내로

복제되어서 pBLA-ZFR 기질들을 생성하였다. 모든 프라이머 서열들이 표 8에서 제공된다.

표 8

[0254]

프라이머	SEQ ID NO:	서열
GFP-ZFR-20G-XbaI-Fwd	264	TTAATTAAGAGTCTAGAGGAGGCGTGTCCAAAACCATGGTTTACAGCAC GCCTCCAGATCTAGGAGGAATTTAAAATGAG

[0255]

GFP-ZFR-20G-HindIII-Rev	265	ACTGACCTAGAGAAGCTTGGAGGCGTGCTGTAAACCATGGTTTTGGACA CGCCTCCCTGCAGTTATTTGTACAGTTCATC
SV40-ZFR-1-BglII-Fwd	266	TTAATTAAGAGAGATCTGCTGATGCAGATACAGAAACCAAGGTTTTCTT ACTTGCTGCTGCGCGATCTGCATCTCAATTAGTCAGC
CMV-PstI-ZFR-1 Rev	267	CACCACCACGGATCCGCAGCAGCAAGTAAGAAAACCTTGGTTTCTGTAT CTGCATCAGCAATTTCGATAAGCCAGTAAGCAG
5' Gin-HBS-Koz	268	CACCACCACGCGCGCAAGCTTAGATCTGGCCCAGGCGGCCACCATGCT GATTGGCTATGTAAGGG
3' Gin-AgeI-Rev	269	CACCACCACACCGGTTCCCGATTTAGGTGGGCGAC

[0256]

ZFR-Target-1-Fwd	270	GTTCCTGCCAGGATCCACTAG
ZFR-Target-1-Rev	271	GCATGTGTCCAGATGCATAGG
ZFR-Target-2-Fwd	272	CACCTTCTCCCAGGATAAGG
ZFR-Target-2-Rev	273	GTTGGCCTGTATTCCTCTGG
ZFR-Target-3-Fwd	274	AATGAAGTTCCCTTGGCACTTC
ZFR-Target-3-Rev	275	CTGAAGGGTTTTAAGTGCAGAAG
CMV-Mid Prim-1	276	TGACGTCAATGACGGTAAATGG
ZFR 표적들에 밑줄인 그어짐		

[0257] 루시페라제 리포터 플라스미드들을 생성하기 위해서, SV40 프로모터가 pGL3- Prm (Promega) 로부터 프라이머들 SV40-ZFR-BglIII-Fwd 및 SV40-ZFR-HindIII-Rev을 사용하여서 PCR 증폭되었다. PCR 산물들이 Bgl11 및 HindIII

으로 다이제션되고 pGL3-Prm의 동일한 제한 부위들 내로 결찰되어서 pGL3-ZFR-1, 2, 3...18을 생성하였다. pBPS-ZFR 도너 플라스미드가 다음과 같은 예외를 두면서 전술한 바와 같이 구축되었다: ZFR-1, 2 및 3 재조합 부위들이 프라이머들 3' CMV-PstI-ZFR-1, 2 또는 3-Rev에 의해서 인코딩되었다. 각 플라스미드의 올바른 구축이 서열 분석에 의해서 확인되었다.

[0258] **재조합 분석들**

[0259] ZFR들이 전술한 바와 같이 PCR에 의해서 어셈블리되었다. PCR 산물들이 SacI 및 XbaI로 다이제션되고 pBLA의 동일한 제한 부위들 내로 결찰되었다. 이 결찰물들이 Escherichia coli TOP10F' (Invitrogen) 내로 전기천공에 의해서 형질변환되었다. SOC 배지 내에서의 1 시간 회복 후에, 세포들은 5 mL SB 배지에서 30 ^gm L^{-1} chloramphenicol을 사용하여 인큐베이션되고 37℃에서 배양되었다. 16 시간 후에, 세포들이 수확되었고; 플라스미드 DNA는 Mini-prep (Invitrogen) 에 의해서 단리되고, 200 ng pBLA가 사용되어서 E. coli TOP10F'을 형질변환하였다. SOC 배지 내에서의 1 시간 회복 후에, 세포들이 고체 LB 배지 상에 30 ^gm L^{-1} chloramphenicol 또는 30 ^gm L^{-1} chloramphenicol 및 100 ^gm L^{-1} carbenicillin, ampicillin 유사체와 함께 플레이팅되었다 (plated). 재조합도가 chloramphenicol을 함유하는 LB 배지 상의 클론들의 개수에 의해서 나누어진, chloramphenicol 및 carbenicillin을 함유하는 LB 배지 상의 클론들의 개수로서 결정되었다. 클론 개수는 GelDoc XR Imaging System (Bio-Rad) 을 사용하여서 자동화된 계수에 의해서 결정되었다.

[0260] **선택들**

[0261] ZFR 라이브러리가 전술한 바와 같이 중첩 확장 PCR에 의해서 구축되었다. 돌연변이들이 축중 (degenerate) 코돈 NNK (N: A, T, C 또는 G 및 K: G 또는 T) 와 함께 위치들 120, 123, 127, 136 및 137 로 도입되는데, 이 코돈 NNK는 모든 20 아미노산들을 인코딩한다. PCR 산물들이 SacI 및 XbaI로 다이제션되고 pBLA의 동일한 제한 부위들 내로 결찰되었다. 결찰물들이 에탄올 석출되어서 E. coli TOP10F'을 형질변환시키는데 사용되었다. 라이브러리 크기는 ~ 5 x 10^7으로 통상적으로 결정되었다. SOC 배지 내에서의 1 시간 회복 후에, 세포들이 37℃에서 100 mL SB 배지 내에서 30 ^gm L^{-1} chloramphenicol와 함께 인큐베이션되었다. 16 시간 후에, 30 mL 세포들이 수확되었다; 플라스미드 DNA가 Mini-prep에 의해서 단리되고 3 ^g 플라스미드 DNA가 사용되어서 E. coli TOP10F'을 형질변환시켰다. SOC 배지 내에서의 1 시간 회복 후에, 세포들이 37℃에서 100 mL SB 배지 내에서 30 ^gm L^{-1} chloramphenicol 및 100 ^gm L^{-1} carbenicillin와 함께 인큐베이션되었다. 16 시간 후에, 세포들이 수확되었고, 플라스미드 DNA가 Maxi-prep (Invitrogen) 에 의해서 단리되었다. 농축된 ZFR들이 SacI 및 XbaI 다이제션에 의해서 단리되고 후속 선택을 위해서 프레시한 (fresh) pBLA 내로 결찰되었다. 4 회의 선택 후에, 서열 분석이 개별 carbenicillin-내성 클론들에 대해서 수행되었다. 재조합 분석들이 상술한 바와 같이 수행되었다.

[0262] **ZFR 구축**

[0263] 리콤비나제 촉매성 도메인들이 그들 각각의 pBLA 선택 벡터로부터 프라이머들 5' Gin-HBS-Koz 및 3' Gin-AgeI-Rev을 사용하여 PCR 증폭되었다. PCR 산물들이 HindIII 및 AgeI로 다이제션되고 pBH의 동일한 제한 부위들 내로 결찰되어서 SuperZiF-양립성 서브클론닝 (subcloning) 플라스미드들: pBH-Gin-a, P, y, 5, s 또는 Z을 생성하였다. 징크-핑거들이 SuperZiF에 의해서 어셈블리되고 pBH-Gin-a, P, y, 5, s 또는 Z의 AgeI 및 SpeI 제한 부위들 내로 결찰되어서 pBH-ZFR-L/R-1, 2, 3.18 (L: 좌측 ZFR; R: 우측 ZFR) 을 생성하였다. ZFR 유전자들이 SfiI 다이제션에 의해서 pBH로부터 분리되고 pcDNA 3.1 (Invitrogen) 내로 결찰되어서 pcDNA-ZFR-L/R-1, 2, 3.18을 생성하였다. 각 ZFR의 올바른 구축이 서열 분석 (표 9) 에 의해서 확인되었다.

표 9

촉매성 도메인	표적	위치들				
		120	123	127	136	137
A	CC[a]	Ile	Thr	Leu	Ile	Gly
B	GC	Ile	Thr	Leu	Arg	Phe
Γ	GT	Leu	Val	Ile	Arg	Trp
Δ	CA	Ile	Val	Leu	Arg	Phe
ε[b]	AC	Leu	Pro	His	Arg	Phe
ζ[c]	TT	Ile	Thr	Arg	Ile	Phe

[0264]

[0265] **촉매성 도메인 치환물들 및 의도된 DNA 표적들**

[0266] 위첨자 a는 와일드-타입 DNA 표적을 표시한다.

[0267] ε^{b} 촉매성 도메인은 또한 치환물들 E117L 및 LI18S을 포함한다.

[0268] ζ^{c} 촉매성 도메인은 또한 M124S, R131I 및 P141R을 포함한다.

[0269] **루시페라제 분석들**

[0270] HEK (Human embryonic kidney) 293 및 293T 세포들 (ATCC) 이 10% (vol/vol) FBS 및 1% (vol/vol) Antibiotic-αNtimycotic (Anti-αNti; Gibco) 을 포함하는 DMEM 내에서 유지되었다. HEK293 세포들이 웰당 4 x 10^4 세포들의 밀도로 96-웰 플레이트들 상에 시딩되었고 37℃에서 습화된 5% CO_2 분위기 내에서 확립되었다. 시딩 후 24 시간에서, 세포들이 150 ng pcDNA-ZFR-L 1-18, 150 ng pcDNA-ZFR-R 1-18, 2.5 ng pGL3-ZFR-1, 2, 3. 또는 18 및 1 ng pRL-CMV 로 Lipofectamine 2000 (Invitrogen) 을 사용하여서 제조자의 지침들에 따라서 형질주입되었다. 형질주입 후 48 시간에, 세포들이 Passive Lysis Buffer (Promega) 으로 파괴되고 루시페라제 발현도가 Veritas Microplate Luminometer (Turner Biosystems) 을 사용하는 이중-루시페라제 리포터 분석 시스템 (Promega) 으로 결정되었다.

[0271] **통합 분석들**

[0272] HEK293 세포들이 웰당 5 x 10^5 세포들의 밀도로 6-웰 플레이트들 상으로 시딩되고 37℃에서 습화된 5% CO_2 분위기 내에서 혈청-함유 배지 내에서 유지되었다. 시딩 후 24 시간 후에, 세포들이 1 ^g pcDNA-ZFR-L-1, 2 또는 3 및 1 ^g pcDNA-ZFR- R-1, 2 또는 3 및 200 ng pBPS-ZFR-1, 2 또는 3로 Lipofectamine 2000을 사용하여서 제조자의 지침들에 따라서 형질주입되었다. 형질주입 후 48 시간 후에, 세포들이 웰당 5 x 10^4 세포들의 밀도로 6-웰 플레이트들 상으로 분할되고 혈청-함유 배지 내에서 2 ^gm L^{-1} puromycin와 함께 유지되었다. 100 퍼센트 콘플루런스 (confluence) 에 도달하면, 세포들이 수확되었고, 게놈 DNA가 Quick Extract DNA Extraction Solution (Epicentre) 로 단리되었다. ZFR 표적들이 Expand High Fidelity Taq System (Roche) 을 사용하여서 다음과 같은 프라이머 조합들을 사용하여서 PCR 증폭되었다:ZFR-표적-1, 2 또는 3-Fwd 및 ZFR-표적-1 , 2 또는 3-Rev (비수정된 표적); ZFR-표적-1, 2 또는 3-Fwd 및 CMV-Mid-Prim-1 (순방향 통합); 및 CMV-Mid-Prim-1 및 ZFR-표적-1, 2 또는 3-Rev (역방향 통합). 클론 분석을 위해서, 형질주입 후 2일 후에, 1 x 10^5 세포들이 100

mm 디시 (dish) 상으로 분할되고,혈청-함유 배지 내에서 2 ^gm L^{-1} puromycin과 함께 유지되었다. 개별 클론들이 살균 실리콘 그리스 (sterile silicone grease) (Millipore) 와 함께 10 mm X 10 mm 개방-단부형 복제 실린더들 (open-ended cloning cylinders) 을 사용하여서 단리되고 배양 시에 확장되었다. 세포들이 100% 콘플루언스에 도달하면 수확되고, 게놈 DNA가 단리되고, 상술한 바와 같이 PCR을 위한 템플릿으로서 사용되었다. 클론 계수 분석들을 위해서, 형질주입 후 2일 후에, 세포들이 웰당 1 x 10^4 세포들의 밀도로 6-웰 플레이트들 내로 분할되고, 2 ^gm L^{-1} puromycin과 함께 또는 없이 혈청-함유 배지 내에서 유지되었다. 16 일 후에, 세포들이 0.2%> 크리스탈 바이올렛 용액 (crystal violet solution) 으로 착색되고 (stained) 및 통합 효율이 puromycin의 부재 시에 형성된 클론들의 개수에 의해서 나누어진, puromycin-함유 배지 내에서 형성된 클론들의 개수를 계수함으로써 결정되었다. 클론 개수는 GelDoc XR Imaging System (Bio-Rad) 을 사용하는 자동화된 계수를 사용하여서 결정되었다.

[0273] **결과들**

[0274] **Gin 리콤비나제의 특이도 프로파일**

[0275] 세린 리콤비나제 촉매성 특이도를 재-엔지니어링하기 위해서, 이러한 효소들의 족에 의한 기질 인식의 기반이 되는 인자들의 상세한 이해가 이루어졌다. 이를 위해서, 대칭적으로 치환된 (symmetrically substituted) 표적 부위들의 포괄적 세트를 재조합할 수 있는 DNA 인버타제 Gin의 촉매성 도메인의 활성화된 돌연변이체 (activated mutant) 의 능력이 평가되었다. Gin 촉매성 도메인은 2 개의 10-bp 절반-부위 영역들로 구성된 의사-대칭적 (pseudo-symmetric) 20-bp 코어를 재조합한다. 따라서, 이러한 재조합 부위들의 집합은 디뉴클레오타이드 코어 내에서 위치들 10, 9, 8, 7, 6, 5, 및 4에서의 각 가능한 단일-염기 치환 및 위치들 3 및 2에서의 각 가능한 2-염기 조합을 포함하였다. 재조합도가 분할된 유전자 재어셈블리, 리콤비나제 활성도를 항생제 내성에 연결하는 전술한 방법에 의해서 결정되었다.

[0276] 일반적으로, Gin은 (i) 디뉴클레오타이드 코어에서의 16 개의 가능한 2-염기 조합들 중 12 개 (AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, GA, GT); (ii) 위치들 3 및 2에서의 16 개의 가능한 2-염기 조합들 중 4 개 (CC, CG, GG 및 TG); (iii) 위치들 6, 5, 또는 4에서의 단일 A에서 T로의 치환; 및 (iv) 위치들 10, 9, 8, 및 7에서의 모든 12 개의 가능한 단일-염기 치환들을 허용한다고 알려졌다 (도 31a 내지 도 31d). 또한, Gin은 위치들 10, 9, 8, 및 7에서의 (가능한 4.29 x 10^9 개 중의) 적어도 10^6 개의 고유한 염기 조합들을 포함하는 표적 부위 라이브러리를 재조합시킬 수 있다고 알려졌다 (도 31d).

[0277] 이러한 사실들은 yS 레졸바제의 결정 구조들로부터 수행된 관측사항들과 일관되며, 이는 (i) 디뉴클레오타이드 코어에 걸쳐 리콤비나제 다이머에 의해서 이루어진 상호작용들이 비대칭이며 지배적으로 비-특이적이고; (ii) 리콤비나제 암 영역 내의 진화적으로 보존된 Gly-Arg motif와 DNA 마이너 그루브 (minor groove) 간의 상호작용들이 위치들 6, 5, 및 4에서의 아데닌 또는 티민을 위한 요건을 부여하며; 및 (iii) 위치들 10, 9, 8, 또는 7에서는 암 영역과 마이너 그루브 간에 어떠한 서열-특이적 상호작용들도 없음을 알려주고 있다 (도 31e). 이러한 결과들은 밀접하게 연관된 Hin 리콤비나제의 DNA-바인딩 특성들의 결정에 초점을 둔 연구들과 또한 일관된다.

[0278] **Gin 리콤비나제 촉매성 특이도 재-엔지니어링 (re-engineering)**

[0279] 위치들 3 및 2에서의 치환들 (즉, CC, CG, GG, 및 TG) 을 Gin이 허용한다는 사실에 기초하여서, Gin 촉매성 특이도가 원 효소에 의해서 허용되지 않은 12 개의 염기 조합들 각각을 포함한 코어 서열들을 특이적으로 인식하기 위해서 재-엔지니어링될 수 있는지의 여부 (도 32a) 가 조사되었다. Gin에 의한 DNA 인식에 참여한 특이적 아미노산 잔기들을 식별하기 위해서, 2 개의 관련된 세린 리콤비나제들, y6 레졸바제 및 Sin 리콤비나제의 결정 구조들이 그들 각각의 DNA 표적들과 결합되어서 검사되었다. 이러한 모델들에 기초하여서, 위치들 3 및 2에서서 DNA와 접촉하는 5 개의 잔기들:(y5 레졸바제에 따라서 넘버링된) Leu 123, Thr 126, Arg 130, Val 139, 및 Phe 140이 식별되었다 (도 32b). 랜덤 돌연변이유발이 중첩 확장 PCR에 의해서 Gin 촉매성 도메인 (Ile 120, Thr 123, Leu 127, Ile 136, 및 Gly 137) 내의 균등한 잔기들에 대해서 수행되고, 이러한 촉매성 도메인 변형체들을 'H1' ZFP의 비수정된 카피 (unmodified copy) 로 융합시킴으로써 ZFR 돌연변이체들의 라이브러리를 구축하였다. 이 라이브러리의 이론적 크기는 3.3 x 10^7 변형체들이다.

[0280] ZFR 라이브러리가 원 효소 (GC, GT, CA, AC, 또는 TT) 에 의해서 허용되지 않은 5 개의 염기 조합들 중 하나를 포함하는 기질 플라스미드들로 복제되고 분할된 유전자 재어셈블리에 의해서 활성 ZFR들을 위해서 농축되었다

(도 32c). 4 회 선택 후에, 각 ZFR 집단의 활성도가 GC, GT, CA, 및 TT 치환들을 포함하는 DNA 표적들에 대해서는 1,000 배 이상 증가하였고 AC 치환들을 포함하는 DNA 표적에 대해서는 100 배 이상 증가하였다고 나타났다 (도 32d).

[0281] 개별 리콤비나제 변형체들이 각 집단으로부터 서열화되고, 높은 레벨의 아미노산 다양성 (diversity) 이 위치들 120, 123, 및 127 에서 존재하였고, 선택된 클론들 중 >80%가 위치 136에서 Arg을 포함하고 위치 137에서 Trp 또는 Phe을 포함한다고 드러났다 (도 36). 이러한 결과들은 위치들 136 및 137이 비자연적 코어 서열들을 인식하는데 있어서 중요한 역할을 한다는 것을 제시한다. 그의 표적 DNA을 재조합할 수 있는 각 선택된 효소의 능력이 평가되었고, 거의 모든 리콤비나제들이 활성도 (>10% 재조합도) 를 보였고 그들의 의도된 코어 서열을 향한 특이도에서의 1,000 배 이상의 시프트 (shift) 를 보였음이 드러났다 (도 37). 모 (parental) Gin에서와 같이, 몇몇 리콤비나제들은 위치들 3 및 2에서의 보존적 치환들을 허용하였음이 드러났으며 (즉, GT 및 CT 또는 AC 및 AG에 대한 (against) 교차-반응), 이는 단일 재-엔지니어링된 촉매성 도메인이 다수의 코어 부위들을 표적화하는데 사용될 수 있음을 말한다 (도 37).

[0282] 리콤비나제 특이도를 더 조사하기 위해서, 모 효소 (GC, TC, GT, CT, GA, CA, AG, AC, 및 TT) (표 1) 에 의해서 허용되지 않은 12 개의 가능한 2-염기 조합들 중 9 개를 인식하도록 도시된 5 개의 Gin 변형체들 (이후부터는 Gin p, y, 6, e 및 Z로서 지정됨) 의 재조합 프로파일들이 결정되었다. Gin p, 6, 및 e은 모 효소 (이후부터는 Gin a로 지칭됨) 의 것과 비교할만한 활성도 및 특이도로 그들의 의도된 코어 서열들을 재조합하였으며, Gin y 및 Z은 Gin a의 것을 뛰어넘은 특이도로 그들의 의도된 코어 서열들을 재조합할 수 있었음이 드러났다 (도 32e). 각 리콤비나제는 위치들 6, 5, 및 4에서 아데닌 또는 티민에 대한 1000 배 이상의 선호도를 보였으며, 위치들 10, 9, 8, 및 7에서는 염기 선호도를 보이지 않았다 (도 38). 이러한 결과들은 DNA-바인딩 암의 돌연변이유발이 리콤비나제 특이도를 떨어뜨리지 않는다는 것을 말한다. 위치들 3 및 2에서 AA, AT, 또는 TA 치환들을 허용할 수 있는 Gin 변형체들을 선택할 수 없었다. 이러한 결과에 대한 일 가능성은 5 개 이상의 연속하는 A-T bp들을 포함한 DNA 표적들은 리콤비나제 바인딩 및/또는 촉매를 방해하는 벤딩된 (bent) DNA 형체들 (conformations) 을 나타낼 수 있다는 것이다.

[0283] <u>사용자 규정된 서열들을 재조합하는 ZFR들을 엔지니어링하는 것</u>

[0284] 재-엔지니어링된 촉매성 도메인들로 구성된 ZFR들이 사전결정된 서열들을 재조합할 수 있는지의 여부가 조사되었다. 이 가능성을 테스트하기 위해서, 인간 게놈 (GRCh37 주 기준 (primary reference) 어셈블리) 가 랜덤 DNA의 대략 매 400,000 bp 마다 한번 발생한다고 예측된 44-bp 컨센서스 재조합 부위를 사용하여서 잠재적 ZFR 표적 부위들에 대해서 탐색되었다 (도 4a). 선택된 Gin 변형체들의 코어 서열 프로파일들로부터 유도된, 이 ZFR 컨센서스 표적 부위는 21 개의 가능한 촉매성 도메인 조합들 및 각 ZFBS 내의 5'-CNN-3' 및 5-TNN-3' 트리플릿들 (triplets) 을 배제한, 모듈식 징크 핑거 도메인들의 보존적 선택에 의해서 허용될 것으로 예측된 (가능한 1.0955×10^{12} 개 중의) 대략 7×10^{8} 개의 고유한 20-bp 코어 조합들을 포함한다. 선택을 주 결정인자로서 ZFP 특이도를 사용하여서, 비-단백질 코딩 좌위에서의 8 개인 인간 염색체들 (Chr. 1, 2, 4, 6, 7, 11, 13 및 X) 에 걸친 18 개의 가능한 ZFR 표적 부위들이 확인되었다. 평균적으로, 각 20-bp 코어는 원 Gin 촉매성 도메인에 의해서 인식된 코어 서열과 ~ 46% 서열 일치를 보였다 (도 33b). 각 대응하는 ZFR이 모듈식 어셈블리에 의해서 구축되었다 ("재료들 및 방법들" 섹션 참조).

[0285] 각 ZFR 쌍은 그의 의도된 DNA 표적을 재조합할 수 있는지의 여부를 결정하기 위해서, ZFR-매개된 재조합을 감소된 루시페라제 발현과 상관시키는 과도적 리포터 분석이 개발되었다 (도 33a 및 도 39). 이를 달성하기 위해서, ZFR 표적 부위들이 루시페라제 리포터 유전자의 발현을 구동하는 SV40 프로모터 업스트림에 그리고 다운스트림에 도입되었다. HEK (Human embryonic kidney) 293T 세포들이 각 ZFR 쌍 및 그의 대응하는 리포터 플라스미드에 대한 발현 벡터들로 공동-형질주입되었다. 루시페라제 발현도가 형질주입 후 48 시간 후에 측정되었다. 분석된 18 개의 ZFR 쌍들 중에서, 38% (18 개 중 7 개의) 가 루시페라제 발현도를 75 배 이상만큼 감소시켰으며, 22% (18 개 중 4 개) 가 루시페라제 발현도를 140 배 이상만큼 감소시켰다 (도 33b). 비교 시에, GinC4, 원 Gin 촉매성 도메인에 의해서 인식된 코어 서열을 표적화하도록 설계된 포지티브 ZFR 제어는 루시페라제 발현도 107 배 만큼 감소시켰다. 대체적으로, 평가된 ZFR 쌍들 중 50% (18 개 중 9 개) 가 루시페라제 발현도를 적어도 20 배만큼 감소시켰다. 중요하게는, 박테리아 세포들에서의 상당한 활성도 (>20% 재조합도) 를 보인 실제로 모든 촉매성 도메인이 포유류 세포들 내에서 적어도 하나의 자연적으로 발생한 서열을 재조합하는데 성공적으로 사용되었다.

[0286] ZFR 특이도를 평가하기 위해서, 별도로 HEK293T 세포들이 각 비-코그네이트 리포터 플라스미드와 함께 9 개의

가장 활성인 ZFR들에 대한 발현 플라스미드들로 공동-형질주입되었다. 각 ZFR 쌍은 그의 의도된 DNA 표적에 대한 높은 특이도를 보였으며, 평가된 ZFR들 중 77% (9 중에서 7 개) 가 포지티브 제어 GinC4의 것과 거의 일치하는 전체적인 재조합 특이도를 보였다 (도 4c). 감소된 루시페라제 발현이 의도된 ZFR 헤테로다이머의 산물이지 재조합-컴피턴트 (competent) ZFR 호모다이머들의 부산물이 아님을 확정하기 위해서, 재조합에 대한 각 ZFR 모노머의 기여도가 측정되었다. 그의 대응하는 리포터 플라스미드와 함께 ZFR 1 '좌측' 모노머의 공동-형질주입은 루시페라제 발현도에서의 중간 정도의 감소로 이어졌지만 (재조합에 대한 총 기여도: ~ 22%), 개별 ZFR 모노머들의 대부분 (18 개 중의 16 개) 은 재조합도에 대하여 크게 기여하지 않았으며 (<10% 재조합도), 및 다수 (18 개 중의 7 개) 가 어떠한 활성도도 보이지 않았다 (도 39). 이러한 연구들이 함께 취해지면 이러한 연구들은 ZFR들이 높은 특이도로 사용자-규정된 서열들을 재조합하도록 엔지니어링될 수 있다는 것을 말한다.

[0287] **엔지니어링된 ZFR들이 인간 게놈으로의 표적화된 통합을 매개함**

[0288] ZFR들이 인간 세포들 내의 내생성 좌위 내로 DNA를 통합시킬 수 있는지의 여부가 이어서 평가되었다. 이를 위해서, HEK293 세포들이 ZFR 발현 벡터들 및 이에 대응하는 DNA 도너 플라스미드 (특이적 ZFR 표적 부위 및 puromycin-내성 유전자를 포함함) 로 SV40 프로모터의 제어 하에서 공동-형질주입되었다. 이러한 분석을 위해서, 각기 인간 염색체들 4, X, 및 4 상의 비-단백질 코딩 좌위를 표적화하도록 설계된 ZFR 쌍들 1 , 2, 및 3 이 사용되었다 (도 34a). 형질주입 후 2 일 후에, 세포들이 puromycin-함유 배지로 인큐베이션되고 puromycin-내성 ($puro^R$) 클론들의 개수를 결정함으로써 통합 효율을 측정하였다. (i) 도너 플라스미드 및 이에 대응하는 ZFR 쌍의 공동-형질주입은 오직 도너 플라스미드로만 형질주입한 것에 비해서 $puro^R$ 클론들의 개수가 12 배 이상으로 증가시켰으며, (ii) 양쪽 ZFR들로의 공동-형질주입은 개별 ZFR 모노머들로의 형질주입에 비해서 puroR 클론들의 개수를 6 배 내지 9 배 이상으로 증가시켰음이 드러났다 (도 34b). ZFR 쌍들이 통합을 올바르게 표적화하였는지의 여부를 평가하기 위해서, 게놈 DNA가 $puro^R$ 집단들로부터 단리되고 각 표적화된 좌위를 PCR 증폭하였다. 순방향 및/또는 역방향 배향들에서 통합에 대응하는 PCR 산물들이 이러한 ZFR 쌍들에 의해서 표적화된 각 좌위에서 관측되었다 (도 34c). 이어서, ZFR-매개된 통합의 전체적 특이도를 결정하기 위해서, 게놈 DNA가 클론 세포 집단들로부터 단리되고 PCR에 의해 플라스미드 삽입을 평가하였다. 이 분석은 각기 ZFR 쌍들 1 , 2, 및 3에 대하여 8.3% (12 클론들 중 1개) 의 표적화 효율, 14.2% (35 개의 클론들 중 5 개) 의 표적화 효율, 및 9.1% (11 개의 클론들 중 1 개) 의 표적화 효율을 드러냈다 (도 36). 각 PCR 산물의 서열 분석은 ZFR-매개된 통합을 확정하였다 (도 34d). 이러한 결과들이 함께 취해지면, 이는 ZFR들이 내생성 좌위 내로 DNA를 정확하게 통합시키도록 설계될 수 있음을 드러냈다.

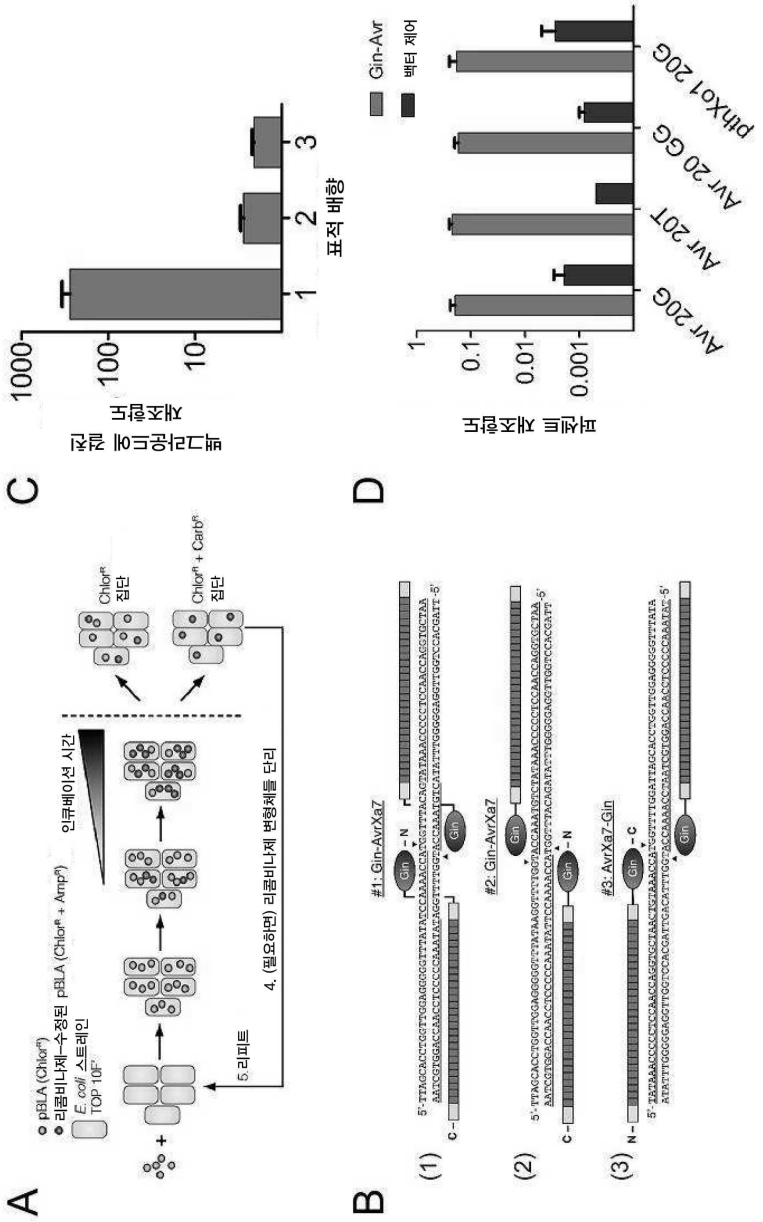
[0289] 마지막으로, ZFR-1 '좌측' 모노머가 ZFR-1 좌위 내로의 통합을 표적화하는 것으로 드러났다는 것이 주목된다 (도 34c). 상술한 루시페라제 리포터 연구들 (도 39) 과 일관된 이러한 결과는 재조합-컴피턴트 ZFR 호모다이머들이 표적을 벗어난 통합을 매개하는 능력을 가짐을 말한다. 최적화된 헤테로다이머성 (heterodimeric) ZFR 아키텍처의 후속 개발 및 표적을 벗어난 통합의 포괄적인 평가는 보다 큰 표적화 효율을 나타내는 ZFR들을 설계하는 것으로 이끌어져야 한다.

[0290] ZFR들이 높은 특이도로 사용자-규정된 서열들을 재조합하도록 설계될 수 있으며 ZFR들이 인간 세포들 내의 사전결정된 내생성 좌위 내로 DNA를 통합시킬 수 있다는 것이 본 명세서에서 드러났다. 기질 특이도 분석과 유도된 진화를 조합함으로써, ZFR 촉매성 도메인에 의해서 부여된 사실상 모든 서열 요건들이 제거되었다. 45 개의 사전-선택된 징크-핑거 모듈들의 라이브러리를 사용하여서, ZFR들이 $>1 \times 10^{22}$ 개의 고유한 44-bp DNA 서열들을 인식하도록 설계될 수 있다고 추정되며, 이러한 값은 랜덤 서열의 매 4,000 bp에 대한 대략 하나의 잠재적 ZFR 표적 부위에 대응한다. 선택에 의해서 맞춤화된 징크-핑거 도메인들의 구축은 표적화를 더 확장할 것이다. 본 명세서에서 기술된 재-엔지니어링된 촉매성 도메인들은 최근에 기술된 TAL 이펙터 리콤비나제들과 양립할 것이다. 이러한 성과는 맞춤 특이도로 ZFR들을 생성할 수 있는 가능성을 나타내며, 게놈 엔지니어링, 합성 생물학, 및 유전자 치료를 포함하여서 다양한 범위의 애플리케이션들에서 ZFR들의 잠재적 활용을 예시한다.

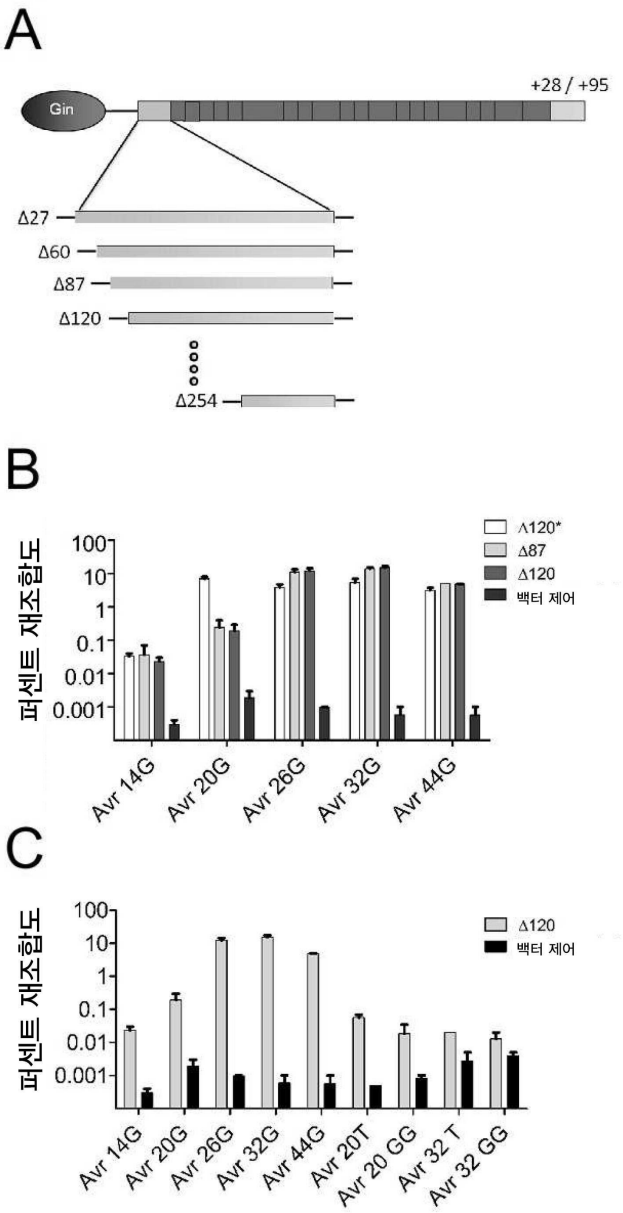
[0291] 본 발명이 위의 실례를 참조하여서 기술되었지만, 수정 및 변경이 본 발명의 사상 및 범위 내에서 포함됨이 이해될 것이다. 따라서, 본 발명은 다음의 청구항들에 의해서만 한정된다.

도면

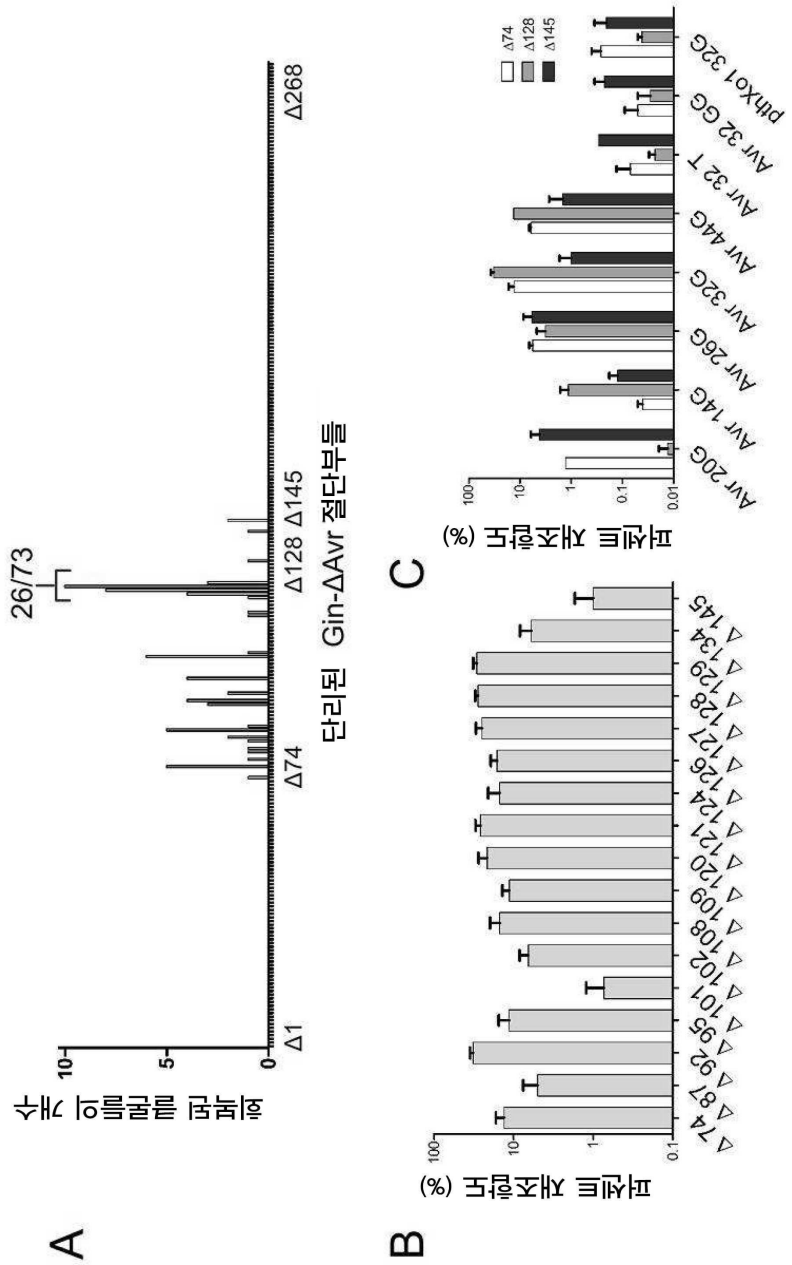
도면1


A
pBLA (Chlor^R)
리콤비니제-수정된 pBLA (Chlor^R + Amp^R)
E. coli 스트레인
TOP 10F'
인큐베이션 시간
Chlor^R 집단
Chlor^R + Carb^R 집단
5. 리피트
4. (필요하면) 리콤비니제 변형체들 단리
B
#1: Gin-AvrXa7
#2: Gin-AvrXa7
#3: AvrXa7-Gin
(1)
(2)
(3)
C
벡그라운드에 걸친 재조합도
표적 배향
D
Gin-Avr
베타 제어
퍼센트 재조합도
AVR 20G
AVR 20T
AVR 20 GG
pthXo1 20G

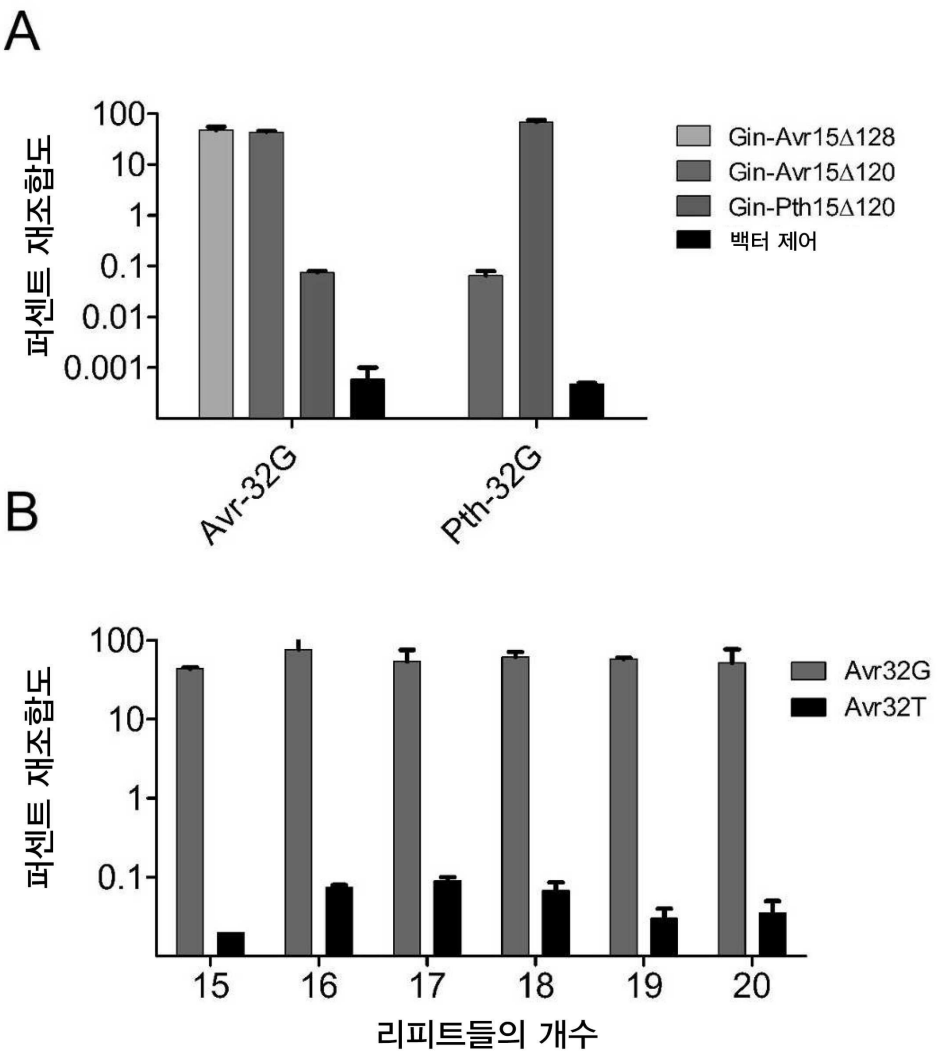
도면2


A
+28 / +95
Gin
Δ27
Δ60
Δ87
Δ120
Δ254
B
퍼센트 재조합도
100
10
1
0.1
0.01
0.001
Δ120*
Δ87
Δ120
백터 제어
Avr 14G
Avr 20G
Avr 26G
Avr 32G
Avr 44G
C
퍼센트 재조합도
100
10
1
0.1
0.01
0.001
Δ120
백터 제어
Avr 14G
Avr 20G
Avr 26G
Avr 32G
Avr 44G
Avr 20T
Avr 20 GG
Avr 32 T
Avr 32 GG

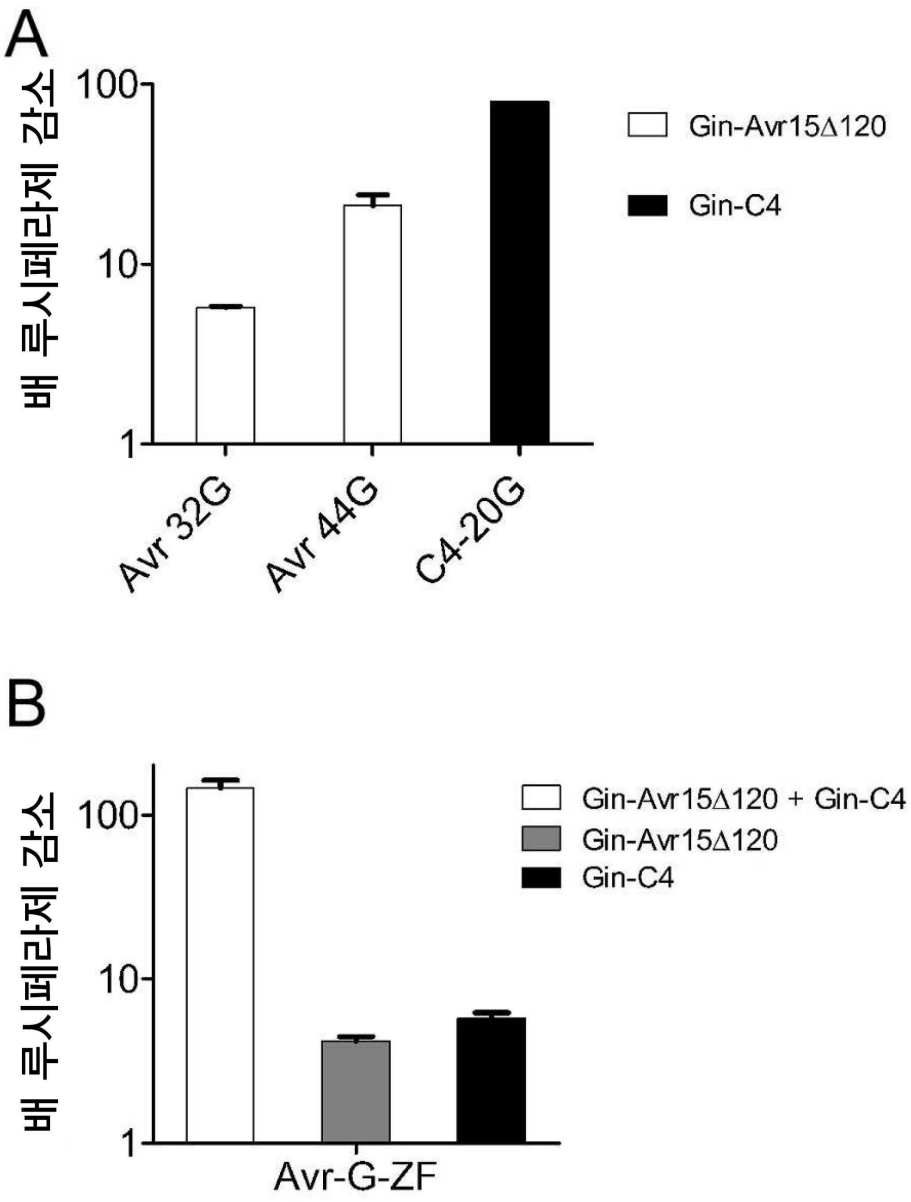
도면3


A
회복된 클론들의 개수
10
5
0
Δ1
Δ74
Δ128 Δ145
Δ268
26/73
단리된 Gln-ΔAvr 절단부들
B
피셀트 자조합도 (%)
100
10
1
0.1
Δ 74
Δ 87
Δ 92
Δ 95
Δ 101
Δ 102
Δ 108
Δ 109
Δ 120
Δ 121
Δ 124
Δ 126
Δ 127
Δ 128
Δ 129
Δ 134
Δ 145
C
피셀트 자조합도 (%)
100
10
1
0.1
0.01
Δ74
Δ128
Δ145
Avr 20G
Avr 14G
Avr 26G
Avr 32G
Avr 44G
Avr 32 T
Avr 32 GG
pHx01 32G

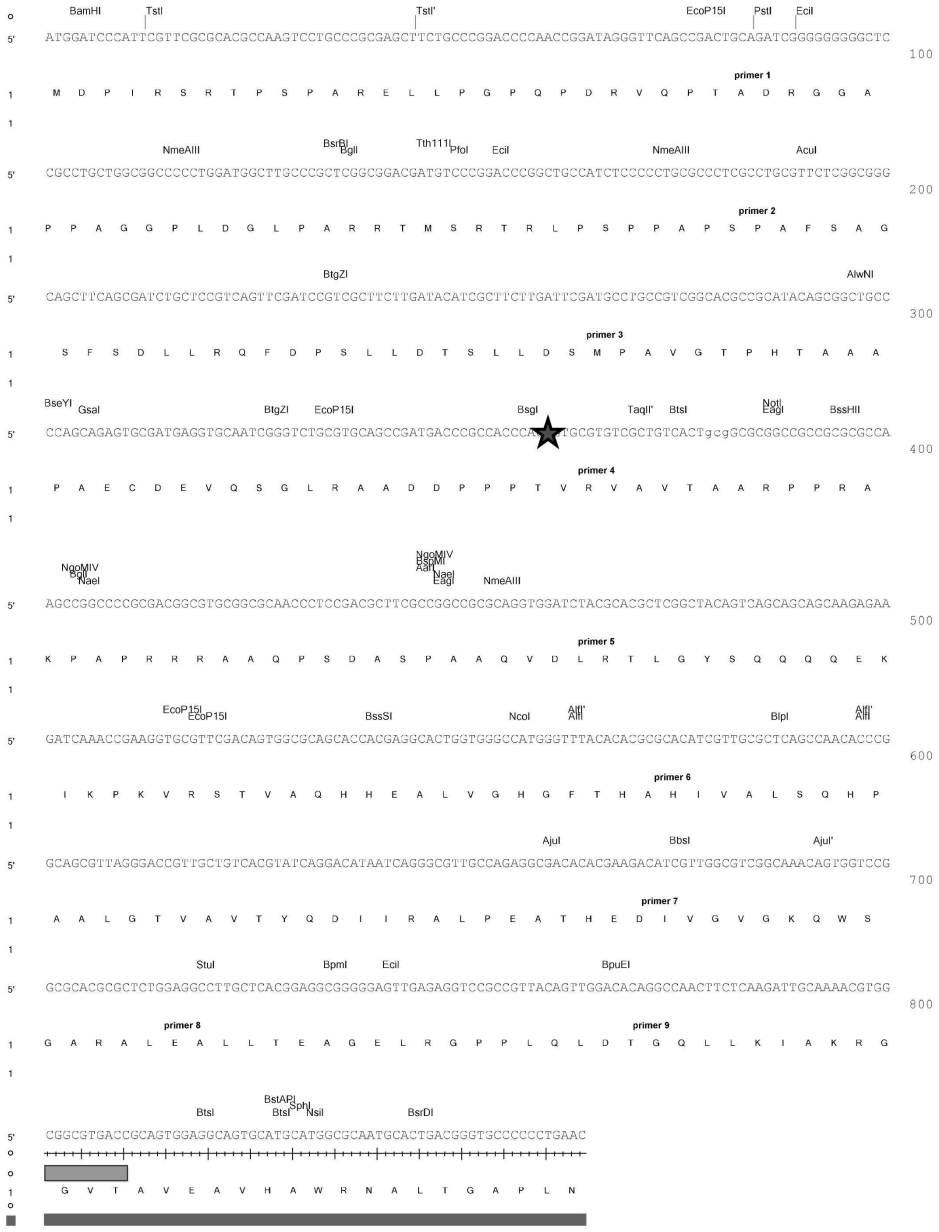
도면4


A
퍼센트 재조합도
100
10
1
0.1
0.01
0.001
Gin-Avr15Δ128
Gin-Avr15Δ120
Gin-Pth15Δ120
백터 제어
Avr-32G
Pth-32G
B
퍼센트 재조합도
100
10
1
0.1
Avr32G
Avr32T
15
16
17
18
19
20
리피트들의 개수

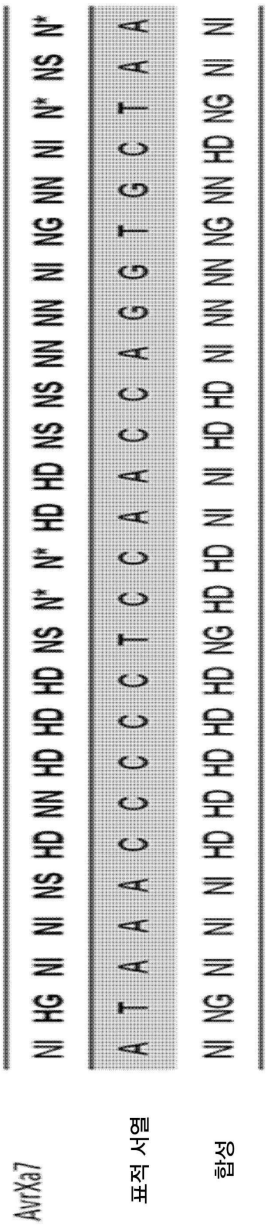
도면5


A
배 루시페라제 감소
100
10
1
Avr 32G
Avr 44G
C4-20G
Gin-Avr15Δ120
Gin-C4
B
배 루시페라제 감소
100
10
1
Avr-G-ZF
Gin-Avr15Δ120 + Gin-C4
Gin-Avr15Δ120
Gin-C4

도면6


BamHI
TstI
TstI'
EcoP15I
PstI
EciI
primer 1
NmeAIII
BsrBI
BglI
Tth111I
PfoI
EciI
NmeAIII
AcuI
primer 2
BtgZI
AlwNI
primer 3
BseYI
GsaI
BtgZI
EcoP15I
BsgI
TaqII'
BtsI
NotI
EagI
BssHII
primer 4
NgoMIV
BglI
NaeI
NgoMIV
BspMI
AarI
NaeI
EagI
NmeAIII
primer 5
EcoP15I
EcoP15I
BssSI
NcoI
AflI'
AflII
BlpI
AflII'
primer 6
AjuI
BbsI
AjuI'
primer 7
StuI
BpmI
EciI
BpuEI
primer 8
primer 9
BtsI
BstAPI
BtsI
SphI
NsiI
BsrDI

도면7


AvrXa7
표적 서열
합성
NI HG NI NI NS HD NN HD HD HD NS N* N* HD NS NS NN NN NI NG NN NI N* NS N*
A T A A A C C C C C T C C C A A C C A G G T G C T A A
NI NG NI NI NI HD HD HD HD HD NG HD HD NI NI HD HD NI NN NN NG NN HD NG NI NI

도면8

TALE 및 TALER 서열들

AvrXa7단백질 :

MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPPAGGPLDGLPARRTMSRTRLPSPPAPSPA
FSAGSFSDLLRQFDPSLLDTSLLDSMPAVGTPHTAAAPAECDEVQSGLRAADDPPPTVRVAV
TAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGH
GFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALLTEAGEL
RGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNIGGKQALET
VQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIG
GKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPAQVV
AIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETLQRLLPVLCQDHG
LTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHDGGKQAL
ETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVQRLLPVLCQDHG
LTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALE
TVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANN
NGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQ
VVAIASNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNGGKQALETVQRL
LPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNGGKQAL
ESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPELIRRINRRIPERTSHRV
PDLAHVVRVLGFFQSHSHPAQAFDDAMTQFEMSRHGLVQLFRRVGVTEFEARYGTLPPASQ
RWDRILQASGMKRAKPSPTSAQTPDQASLHAFADSLERDLDAPSPMHEGDQTRASSRKRSR
SDRAVTGPSTQQSFEVRVPEQQDALHLPLSWRVKRPRTRIGGGLPDPGTPIAADLAASSTVM
WEQDAAPFAGAADDFPAFNEEELAWLMELLPQSGSVGGTI

도면9

AvrXa7 유전자 :

ATGGATCCCATTCGTTCGCGCACGCCAAGTCCTGCCCGCGAGCTTCTGCCCGGACCCCA
ACCGGATAGGGTTCAGCCGACTGCAGATCGGGGGGGGGCTCCGCCTGCTGGCGGCCCC
CTGGATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCGGCTGCCATCTCCCCCTGC
GCCCTCGCCTGCGTTCTCGGCGGGCAGCTTCAGCGATCTGCTCCGTCAGTTCGATCCGTC
GCTTCTTGATACATCGCTTCTTGATTCGATGCCTGCCGTCGGCACGCCGCATACAGCGGC
TGCCCCAGCAGAGTGCGATGAGGTGCAATCGGGTCTGCGTGCAGCCGATGACCCGCCAC
CCACCGTGCGTGTCGCTGTCACTgcgGCGCGGCCGCCGCGCGCCAAGCCGGCCCCGCGAC
GGCGTGCGGCGCAACCCTCCGACGCTTCGCCGGCCGCGCAGGTGGATCTACGCACGCTC
GGCTACAGTCAGCAGCAGCAAGAGAAGATCAAACCGAAGGTGCGTTCGACAGTGGCGC
AGCACCACGAGGCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGCGCTCAGC
CAACACCCGGCAGCGTTAGGGACCGTTGCTGTCACGTATCAGGACATAATCAGGGCGTT
GCCAGAGGCGACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGGCGCACGC
GCTCTGGAGGCCTTGCTCACGGAGGCGGGGGAGTTGAGAGGTCCGCCGTTACAGTTGGA
CACAGGCCAACTTCTCAAGATTGCAAAACGTGGCGGCGTGACCGCAGTGGAGGCAGTG
CATGCATGGCGCAATGCACTGACGGGTGCCCCCCTGAACCTGACCCCGGACCAAGTGGT
GGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTG
CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTCGTGGCCATCGCCAGCCA
TGGCGGCGGCAAGCAGGCGCTGGAGACGGTGC
AGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCC
ATCGCCAGCAATATTGGCGGCAAGCAGGCGCTAGAGACGGTGCAGCGGCTGTTGCCGG
TGCTGTGCCAGGCCCATGGCCTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAATATT
GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACC
ATGGCCTGACCCCGGCCCAGGTGGTGGCCATCGCCAGCAATAGTGGCGGCAAGCAGGC
GCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGG
ACCAAGTCGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGCTGCA
GCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTCGTGGCCA
TCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGCTGCAGCGGCTGTTGCCGGT
GCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCCATCGCCAGCCACGATG
GCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCA
TGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCG
CTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGC
CCAAGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAG
CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCCAT
CGCCAGCAATAGCGGCGGCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTGCCGGTG
CTGTGCCAGGACCATGGACTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAATGGCG
GCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACCCCGGACCAGGTCGTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGGAG
ACGGTGCAGCGGCTGTTGCCGGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACCCAGGACCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTG
GAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCA
AGTGGTGGCCATCGCCAGCCACGATGGCGGCAAACAGGCGCTGGAGACGGTGCAGCGG
CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCCATCGC
CAGCAATAGTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTG
TGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCCATCGCCAGCAATAGTGGCG
GCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACCCCAGGACCAGGTGGTGGCCATCGCCAGCAACAATAACGGCGGCAAGCAGGCGCTG
GAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCA
GGTCGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGG
CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGCGCAGGTGGTGGCCATCGC
CAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCCAGGACCATGGCCTGACCCTGGACCAGGTGGTGGCCATTGCCAGCAATGGCGGCAG
CAAACAGGCGCTAGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
CTGACCCCGGACCAAGTGGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGG
AGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAG
GTCGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCTGGACCAGGTGGTGGCCATCGCC
AGCAATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCC
AGGACCATGGCCTGACCCCGAACCAGGTGGTGGCCATCGCCAGCAATAGTGGCGGCAA
GCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGA
CCCCGAACCAGGTGGTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGGAGAGCAT
TGTTGCCCAGTTATCTCGCCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCACCTCGT
CGCCTTGGCCTGCCTCGGCGGACGTCCTGCCCTGGATGCAGTGAAAAAGGGATTGCCGC
ACGCGCCGGAATTGATCAGAAGAATCAATCGCCGCATTCCCGAACGCACGTCCCATCGC
GTTCCCGACCTCGCGCACGTGGTTCGCGTGCTTGGTTTTTTCCAGAGCCACTCCCACCCA
GCGCAAGCATTCGATGACGCCATGACGCAGTTCGAGATGAGCAGGCACGGCTTGGTAC
AGCTCTTTCGCAGAGTGGGCGTCACCGAATTCGAAGCCCGCTACGGAACGCTCCCCCCA
GCCTCGCAGCGTTGGGACCGTATCCTCCAGGCATCAGGG
ATGAAAAGGGCCAAACCGTCCCCTACTTCAGCTCAAACACCGGATCAGGCGTCTTTGCA
TGCATTCGCCGATTCGCTGGAGCGTGACCTTGATGCGCCCAGCCCAATGCACGAGGGAG
ATCAGACGCGGGCAAGCAGCCGTAAACGGTCCCGATCGGATCGTGCTGTCACCGGCCCC
TCCACACAGCAATCTTTCGAGGTGCGCGTTCCCGAACAGCAAGATGCGCTGCATTTGCC
CCTCAGCTGGAGGGTAAAACGCCCGCGTACCAGGATCGGGGGCGGCCTCCCGGATCCT
GGTACGCCCATCGCTGCCGACCTGGCAGCGTCCAGCACCGTGATGTGGGAACAAGATGC
GGCCCCCTTCGCAGGGGCAGCGGATGATTTCCCGGCATTCAACGAAGAGGAGCTCGCAT
GGTTGATGGAGCTATTGCCTCAGTCAGGCTCAGTCGGAGGGACGATCTGA

도면10

Gin-AvrΔ74

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVV
WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERT
MAGLAAARNKGRIGGRPRKSGSGSPRQFDPSLLDTSLLDSMPAVGTPHTAAAPAECDEVQS
GLRAADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQEKIKP
KVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQ
WSGARALEALLTEAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTP
DQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQ
RLLPVLCQDHGLTPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
QALETLQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQVVAI
ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGL
TPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALETV
QRLLPVQRLLPVLCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI
ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGL
TPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALET
VQRLLPVLCQDHGLTLDQVVAIASNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIANNN
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQV
VAIASNGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHG
LTPNQVVAIASNGGKQALESIVAQLSRPDPALAALTNDHLVALACLG

도면11

Gin-AvrΔ87

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVV
WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERT
MAGLAAARNKGRIGGRPRKSGSGSPDSMPAVGTPHTAAAPAECDEVQSGLRAADDPPPTV
RVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHE
ALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALL
TEAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQVV
AIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGL
TPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETLQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE
TVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHD
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVQRLLPV
LCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE
TVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNS
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQD
HGLTLDQVVAIASNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQR
LLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNGGKQA
LETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIAS
NGGKQALESIVAQLSRPDPALAALTNDHLVALACLG

도면12

Gin-AvrΔ120

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVV
WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERT
MAGLAAARNKGRIGGRPRKSGSGSTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDL
RTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRA
LPEATHEDIVGVGKQWSGARALEALLTEAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHA
WRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGK
QALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAI
ASNIGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASHDGGKQALETLQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVL
CQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALET
VQRLLPVLCQDHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSG
GKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVLCQDHGLTPDQVV
AIASNGGKQALETVQRLLPVQRLLPVLCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVL
CQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALET
VQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNG
GKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQV
VAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNGGSKQALETVQRLLPVLCQDH
GLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLL
PVLCQDHGLTLDQVVAIASNGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALE
TVQRLLPVLCQDHGLTPNQVVAIASNGGKQALESIVAQLSRPDPALAALTNDHLVALACLG

도면13

Gin-AvrΔ120*

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVV
WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERT
MAGLAAARNKGRIGGRPRKSGSGSTVRVAVTAARPPHAVAGPAAQVDLRTLGYSQQQQE
KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGV
GKQWSGARALEALLTEAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLN
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALE
TVQRLLPVLCQDHGLTPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD
GGKQALETLQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQ
DHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQ
RLLPVLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQA
LETVQRLLPVQRLLPVLCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQD
HGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRL
LPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQA
LETVQRLLPVLCQDHGLTLDQVVAIASNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIAN
NNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLD
QVVAIASNGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQD
HGLTPNQVVAIASNGGKQALESIVAQLSRPDPALAALTNDHLVALACLG

도면14

Gin-AvrΔ147

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVV
WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERT
MAGLAAARNKGRIGGRPRKSGSGSPASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHE
ALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALL
TEAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQVV
AIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGL
TPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETLQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE
TVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHD
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVQRLLPV
LCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE
TVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNS
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQD
HGLTLDQVVAIASNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQR
LLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNGGKQA
LETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIAS
NGGKQALESIVAQLSRPDPALAALTNDHLVALACLG

도면15

GinAvr15Δ128-합성 단백질

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVV
WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERT
MAGLAAARNKGRIGGRPRKSGSGSPALRPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQ
QQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHED
IVGVGKQWSGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTG
APLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQ
RLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ
ALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTP
DQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVL
CQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALET
VQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDG
GKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVV
AIASNIGGKQALESIVAQLSRPDPALAALTNDHLVALACLGPKKKRKV

도면16

Gin-Avr15Δ128-합성 DNA:

ATGCTGATTGGCTATGTAAGGGTATCAACAAATGACCAGAATACAGACCTGCAACGAA
ACGCTCTTGTTTGTGCAGGATGTGAACAAATATTTGAAGATAAATTAAGCGGAACAAGG
ACAGACCGACCGGGATTAAAACGCGCTTTAAAGCGCCTTCAAAAAGGTGACACACTGG
TTGTCTGGAAACTGGATCGCCTCGGGCGAAGCATGAAACATTTGATTTCTCTCGTAGGG
GAATTACGAGAGCGAGGGATTAATTTTCGCAGTCTTACTGACAGTATTGATACGTCATC
TCCAATGGGGCGTTTTTTCTTCTACGTTATGGGTGCCCTGGCTGAAATGGAACGAGAACT
AATTATCGAGCGAACGATGGCTGGACTTGCTGCCGCCAGAAATAAAGGCCGTATTGGAG
GTCGCCCGCGTAAATCGGGGTCTGGATCCCCCGCGCGGCCGCCGCGCGCCAAGCCGGCC
CCGCGACGGCGTGCTGCGCAACCCTCCGACGCTTCGCCGGCCGCGCAGGTGGATCTACG
CACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGATCAAACCGAAGGTGCGTTCGACA
GTGGCGCAGCACCACGAGGCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC
GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCACGTATCAGCACATAATCA
CGGCGTTGCCAGAGGCGACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGG
CGCACGCGCCCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGAGGTCCGCCGTTAC
AGTTGGACACAGGCCAACTTGTGAAGATTGCAAAACGTGGCGGCGTGACCGCAATGGA
GGCAGTGCATGCATCGCGCAATGCACTGACGGGTGCCCCCCTGGAGCTGACTCCGGACC
AAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCG
GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCG
CCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCT
GTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCG
GCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
GTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGG
CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
CTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGG
CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
CTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCC
AGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGG
CAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCTGATCCGGCGTTGGCCG
CGTTGACCAACGACCACCTCGTCGCCTTGGCCTGCCTCGGCCCCAAGAAGAAGCGCAAG
GTGTAG

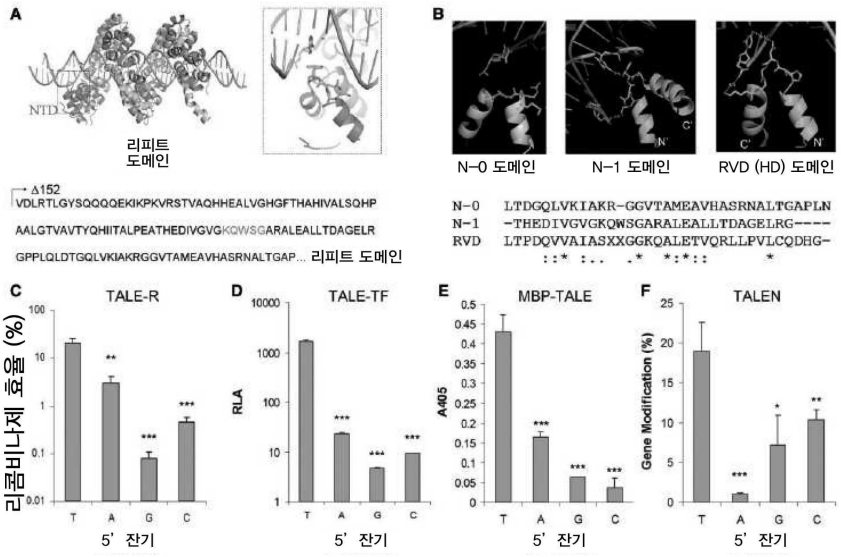
도면17

GinAvr15Δ128-합성 단백질 :

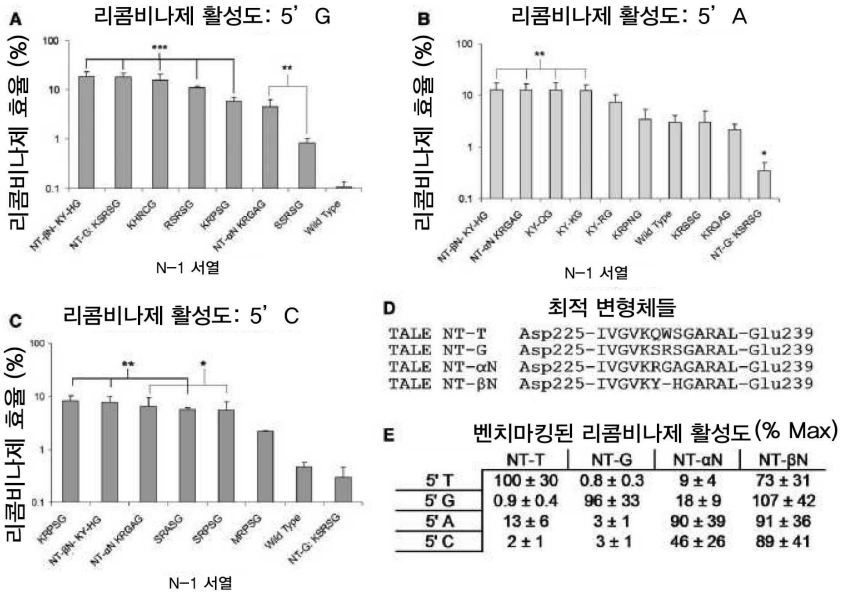
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKR
ALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSI
DTSSPMGRFFFYVMGALAEMERELIIERTMAGLAAARNKGRIGGRPR
*KSGSGS*PALRPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQE
KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHI
ITALPE
ATHEDIVGVGKQWSGARALEALLTDAGELRGPPLQ (-1)
LDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLN (0)T
LTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHG (1)A
LTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHG (2)T
LTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHG (3)A
LTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHG (4)A
LTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHG (5)A
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (6)C
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (7)C
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (8)C
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (9)C
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (10)C
LTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHG (11)T
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (12)C
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (13)C
LTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHG (14)A
LTPDQVVAIAS**NI**GGKQALE (15)A
SIVAQLSRPDPALAALTNDHLVALACLGPKKKRKV

키 :
Gin 리콤비나제 (이텔릭체)
TALE N-말단 골격 (이중 밑줄)
RVD (DNA 바인딩 도메인) (볼드체)
TALE C-말단 골격 (점선 밑줄)

도면18

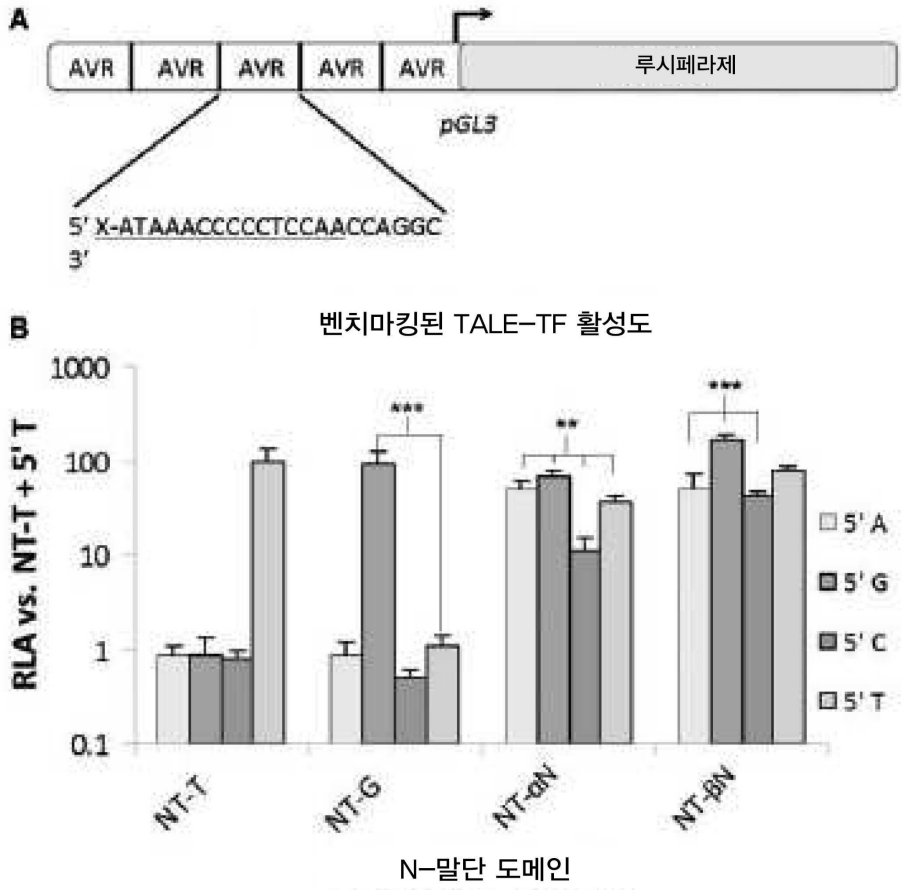


도면19

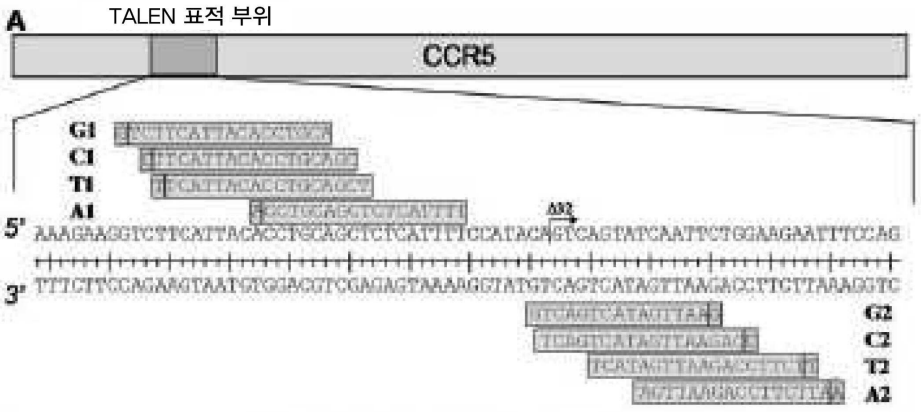


	NT-T	NT-G	NT-αN	NT-βN
5' T	100 ± 30	0.8 ± 0.3	9 ± 4	73 ± 31
5' G	0.9 ± 0.4	96 ± 33	18 ± 9	107 ± 42
5' A	13 ± 6	3 ± 1	90 ± 39	91 ± 36
5' C	2 ± 1	3 ± 1	46 ± 26	89 ± 41

도면20



도면21



B

TALEN 유전자 에디팅 효율

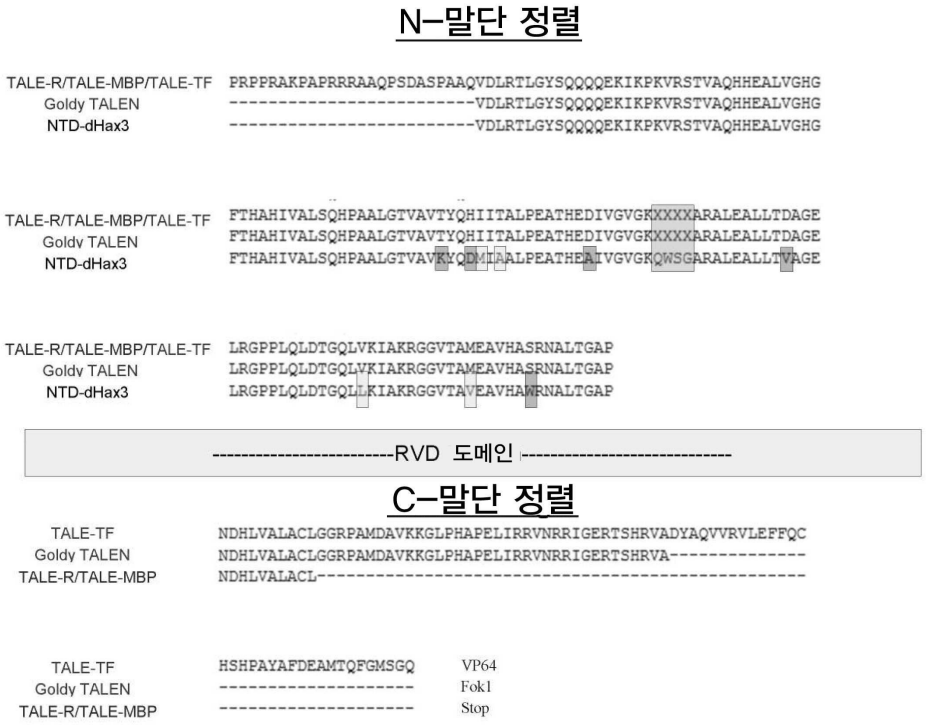
TALEN 쌍	NT-T	NT-G	NT-αN	NT-βN	dHax3
T1/T2	18.9 ± 3.7	9.5 ± 2.7	14.5 ± 2.3	11.2 ± 0.3	22.1 ± 2.2
G1/G2	7.1 ± 3.8	17.2 ± 2.8	14.4 ± 1.7	25.1 ± 3.2	8.25 ± 2.9
A1/A2	1 ± 0.2	0.7 ± 0.4	9.1 ± 1.1	0.4 ± 0.2	1.2 ± 0.5
C1/C2	10.4 ± 1.2	9.1 ± 0.4	21.1 ± 4.5	15.1 ± 2.3	11.3 ± 3.1

C

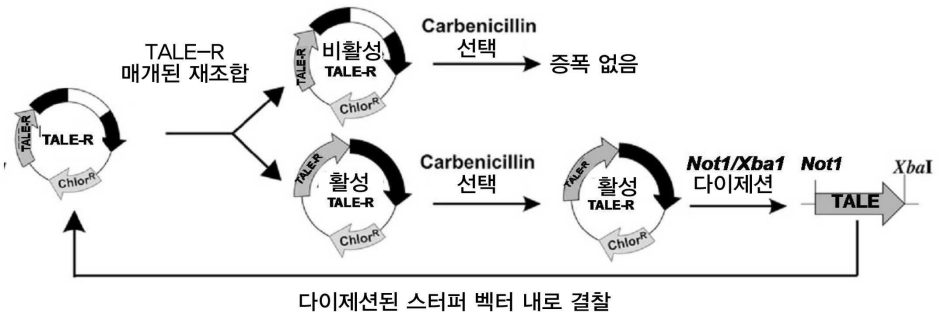
TALEN NTD 활성도 증가 vs. NT-T

TALEN 쌍	NT-T	NT-G	NT-αN	NT-βN	dHax3
T1/T2	100 ± 19	50 ± 14	77 ± 12	59 ± 2	117 ± 12
G1/G2	100 ± 51	242 ± 15	203 ± 23	354 ± 45	116 ± 41
A1/A2	100 ± 20	70 ± 40	910 ± 110	40 ± 20	120 ± 50
C1/C2	100 ± 12	88 ± 4	203 ± 43	145 ± 20	109 ± 30.

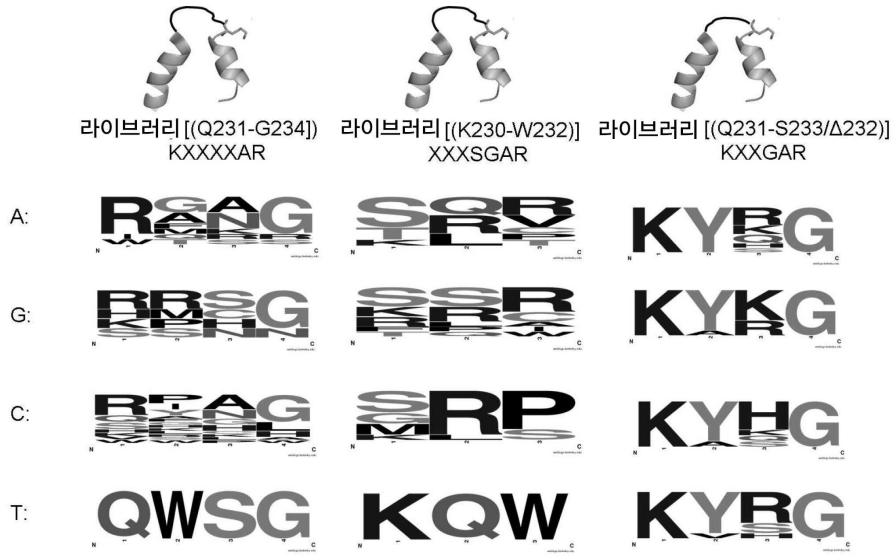
도면22



도면23



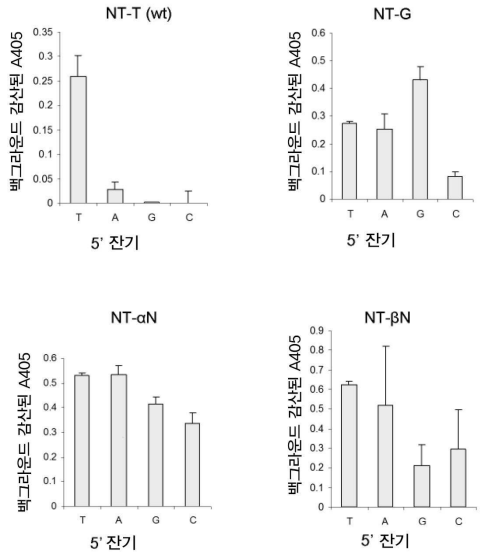
도면24



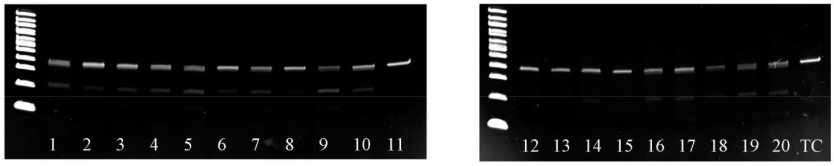
도면25

```
TALE-G          QVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQ 60

Brg11           ---MAALGYSREQIRKLKQESLSGVAKYHAPLTRHGFTHTDICRISRRWQSLRMVAKNYP 57
                   : :****::* .*:* :  * **::* .*. *****:.*  :*::  :*  ** .*

TALE-G          HIITALPEATHEDIVGVGKSRSGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTA 120

Brg11           KLIAALPDLTRTHIVDIARQRSGDLALEALLPVATALAAAPLRLRASQIAIIAQCGERPA 117
                ::*:***: *: .**.:.: : *  ******. *  * ..**:* :.*:. **: *  .*

TALE-G          MEAVHASRNALTGAP 135

Brg11           ILALHRLRRKLTGAP 132
                : *:*  *. *****
```

도면26



도면27



TALEN 쌍	NTD	에디팅 %	인델
1: T1/T2	NT-T	16.6%	
2: T1/T2	NT-G	7.5%	
3: T1/T2	NT-βN	11.3%	
4: T1/T2	NT-αN	12.9%	
5: T1/T2	NT-dHax3	20.1%	
6: G1/G2	NT-T	4.4%	
7: G1/G2	NT-G	15.2%	(9/30)
8: G1/G2	NT- αN	4.9%	3x Ave = 14.4
9: G1/G2	NT- βN	28.9%	(10/30)
10: G1/G2	NT-dHax3	6.2%	
11: A1/A2	NT-T	1%	

TALEN 쌍	NTD	에디팅 %	인델
12: A1/A2	NT-G	0.7%	
13: A1/A2	NT- βN	0.4%	
14: A1/A2	NT- αN	9.1%	(3/30)
15: A1/A2	NT-dHax3	1.2%	
16: C1/C2	NT-T	8.6%	
17: C1/C2	NT-G	8.9%	
18: C1/C2	NT- βN	13.4%	(9/30)
19: C1/C2	NT- αN	17.9%	(8/30)
20: C1/C2	NT-dHax3	9.1%	

도면28

TALEN 쌍 7 (9/30)

CCR5 (x20) TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-6 (x3) TGGAAATTCTTCCATAATTGATATTGACTGTATGGA------AGGCTGCGGGTGTAATGAATACCTTCTTTTTGAGATCTGGT
-6 TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAAAA------CTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-6 TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAA------AGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-7 TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAA-------GCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-10 TGGAAATTCTTCCAGAATTGATACTGACTGTAG----------AGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-11 TGGAAATTCTTCCAGAATTGATACTGACTGTAT----------GCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-13 TGGAAATTCTTCCAGAATTGATACTGACTGT----------GCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT

TALEN 쌍 9 (10/30)
(CCR5 WT X 21)

TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-4 TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAAA----GAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-5 TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAA-----GAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-6 TGGAAATTCTTCCAGAATTGATACTGACTGTATGGA------GAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-9 TGGAAATTCTTCCAGAATTGATACTGACT---------ATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-10 (x2) TGGAAATTCTTCCAGAATTGATACTGACTGTA---------GAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-11 TGGAAATTCTTCCAGAATTGATACTGAC----------TGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-11 TGGAAATTCTTCCAGGATTGATACTGACT----------GAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-11 TGGAAATTCTTCCAGAATTGATACTGACT----------GAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-11 TGGAAATTCTTCCAGAATTGATACTGACT----------AGGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT

TALEN 쌍 14 (3/30)
(CCR5 WT X27)

TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-2 TGGAAATTCTTCCAGAATTGATACTGACT--ATGGAAAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-3 TGGAAATTCTTCCTCAATTGATACTGA---TATGGAAAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-9 TGGAAATTCTTCCAGAATTGATA---------TGGAAAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT

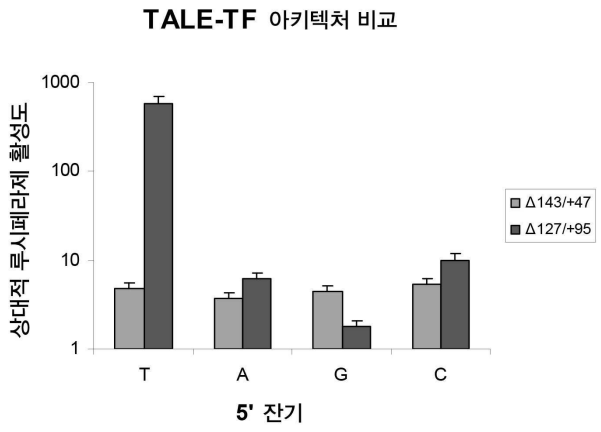
TALEN 쌍 18 (9/30)
(CCR5 WT X 20)

TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-4 TGGAAATTCTTCCAGAATTGATACTGACTGTATGA----TGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-4 TGGAAATTCTTCCAGAATTGATACTGACTGTATGG----TGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-5 TGGAAATTCTTCCAGAATTGATACTGACTGTA-----AATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-5 TGGAAATTCTTCCAGAATTGATACTGACTGTATG-----TGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-9 TGGAAATTCTTCCAGAATTGATACTGA---------AAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-9 TGGAAATTCTTCCAGAATTGATACTGACTGTATGGA---------CTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-10 TGGAAATTCTTCCAGAATTGATACTGACTG---------TGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
+3/-10 TGGAAATTCTTCCAGAATTGATACTGgta---------TGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-12 TGGAAATTCTTCCAGAATTGATACTGACTGT----------AGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-

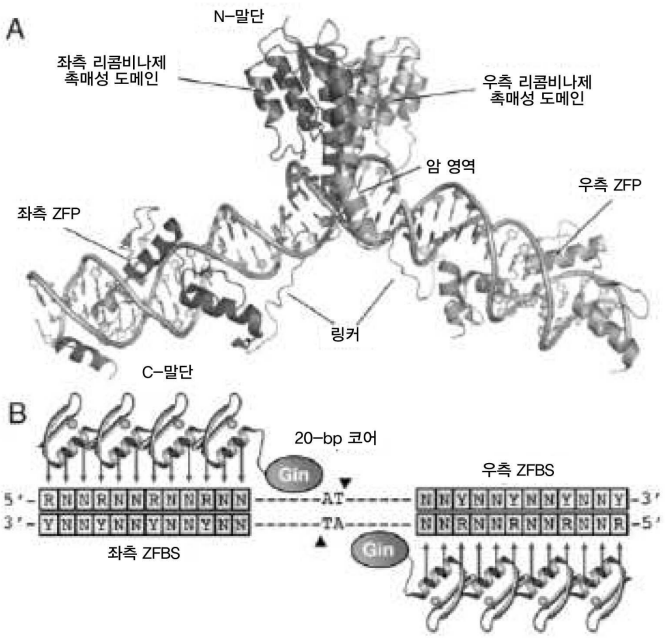
TALEN 쌍 19 (8/30)
(CCR5 WT (x 23)

TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAAAATGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-7 TGGAAATTCTTCCAGAATTGATACTGACTGTATGGAAA-------CTGC-GGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-7 TGGAAATTCTTCCAGAATTGATACTGACTGTA-------TGAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-9 TGGAAATTCTTCCAGAATTAATACTGACTGT---------GAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-10 TGGAAATTCTTCCAGAATTGATACTGACTG----------GAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-11 TGGAAATTCTTCCAGAATTGATACTGACT----------GAGAGCTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-15 TGGAAATTCTTCCAGAATTGATACTGACTG----------CTGCAGGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-18 TGGAAATTCTTCCAGAATTGATACTGACTGT----------ATGTGTAATGAAGACCTTCTTTTTGAGATCTGGT
-44 TGGAAATTCTTCCAGAATTGATA----------CTTTTTGAGATCTGGT

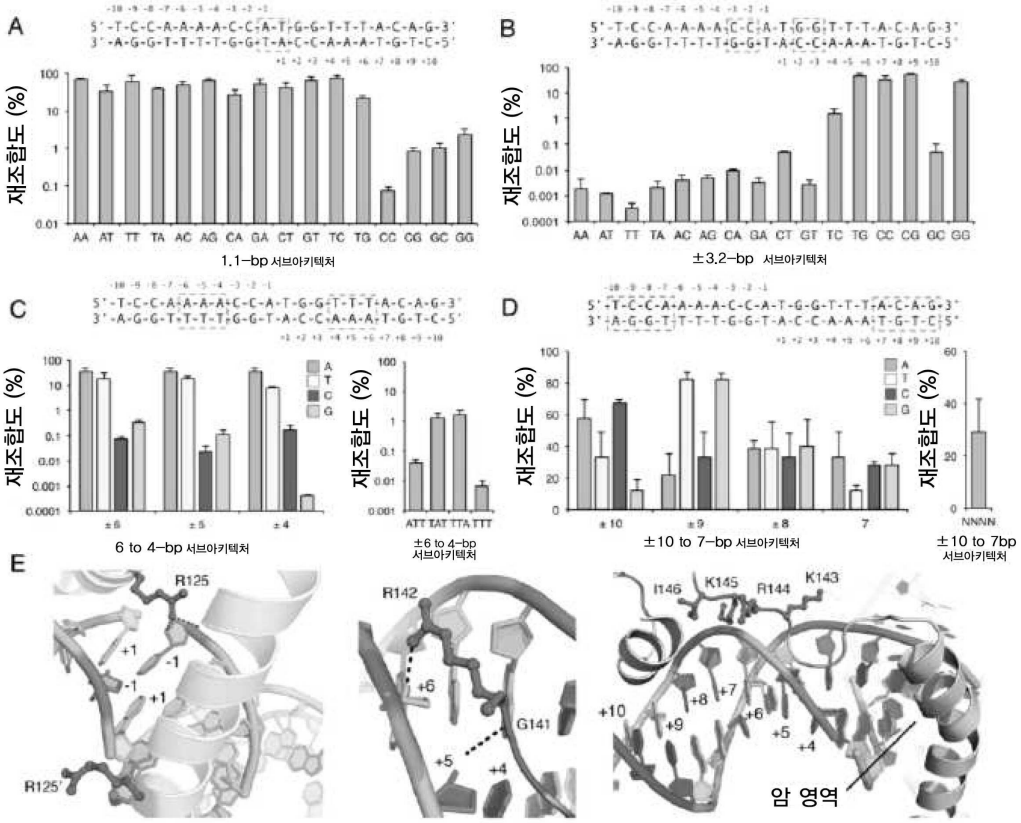
도면29



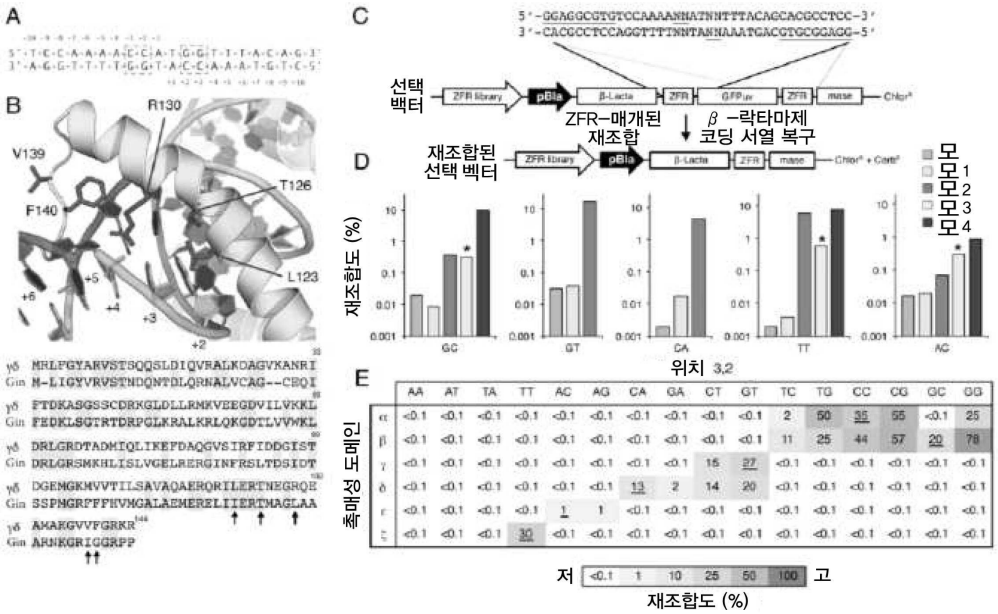
도면30


A
N-말단
좌측 리콤비나제
촉매성 도메인
우측 리콤비나제
촉매성 도메인
암 영역
우측 ZFP
좌측 ZFP
링커
C-말단
B
20-bp 코어
Gin
우측 ZFBS
5'-
3'-
AT
TA
Gin
좌측 ZFBS

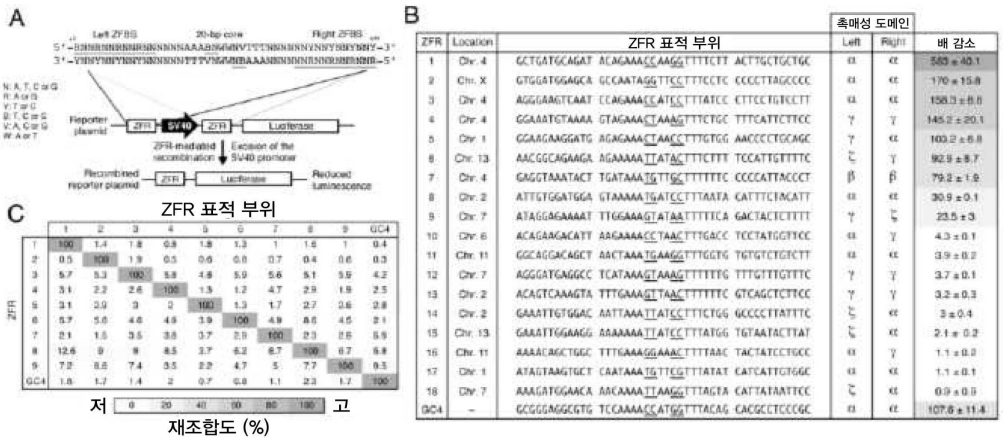
도면31


A
재조합도 (%)
AA AT TT TA AC AG CA GA CT GT TC TG CC CG GC GG
1.1-bp 서브아키텍처
B
재조합도 (%)
±3.2-bp 서브아키텍처
C
재조합도 (%)
6 to 4-bp 서브아키텍처
ATT TAT TTA TTT
±6 to 4-bp
서브아키텍처
D
재조합도 (%)
±10 to 7-bp 서브아키텍처
NNNN
±10 to 7bp
서브아키텍처
E
R125
R125'
R142
G141
I146
K145
R144
K143
암 영역

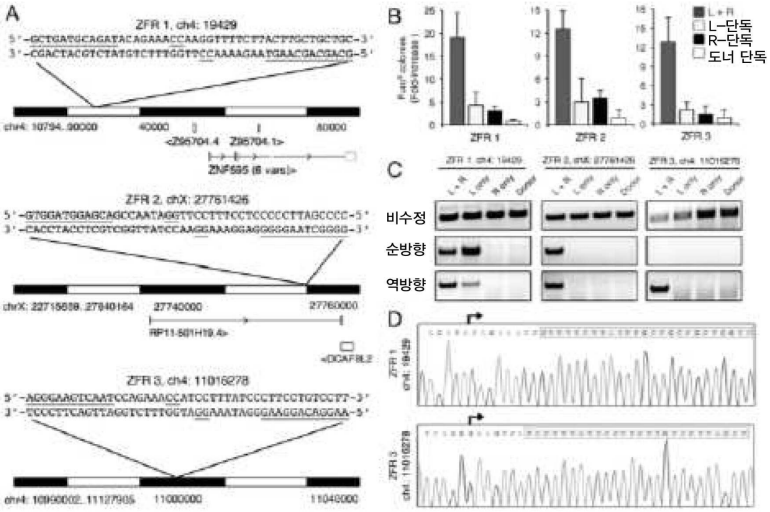
도면32


A
B
R130
V139
T126
F140
L123
+5
+6
+4
+3
+2
C
선택 백터
ZFR-매개된 재조합
β -락타마제 코딩 서열 복구
재조합된 선택 벡터
D
재조합도 (%)
위치 3,2
모
모1
모2
모3
모4
GC
GT
CA
TT
AC
E
축퇴성 도메인
저
고
재조합도 (%)

도면33

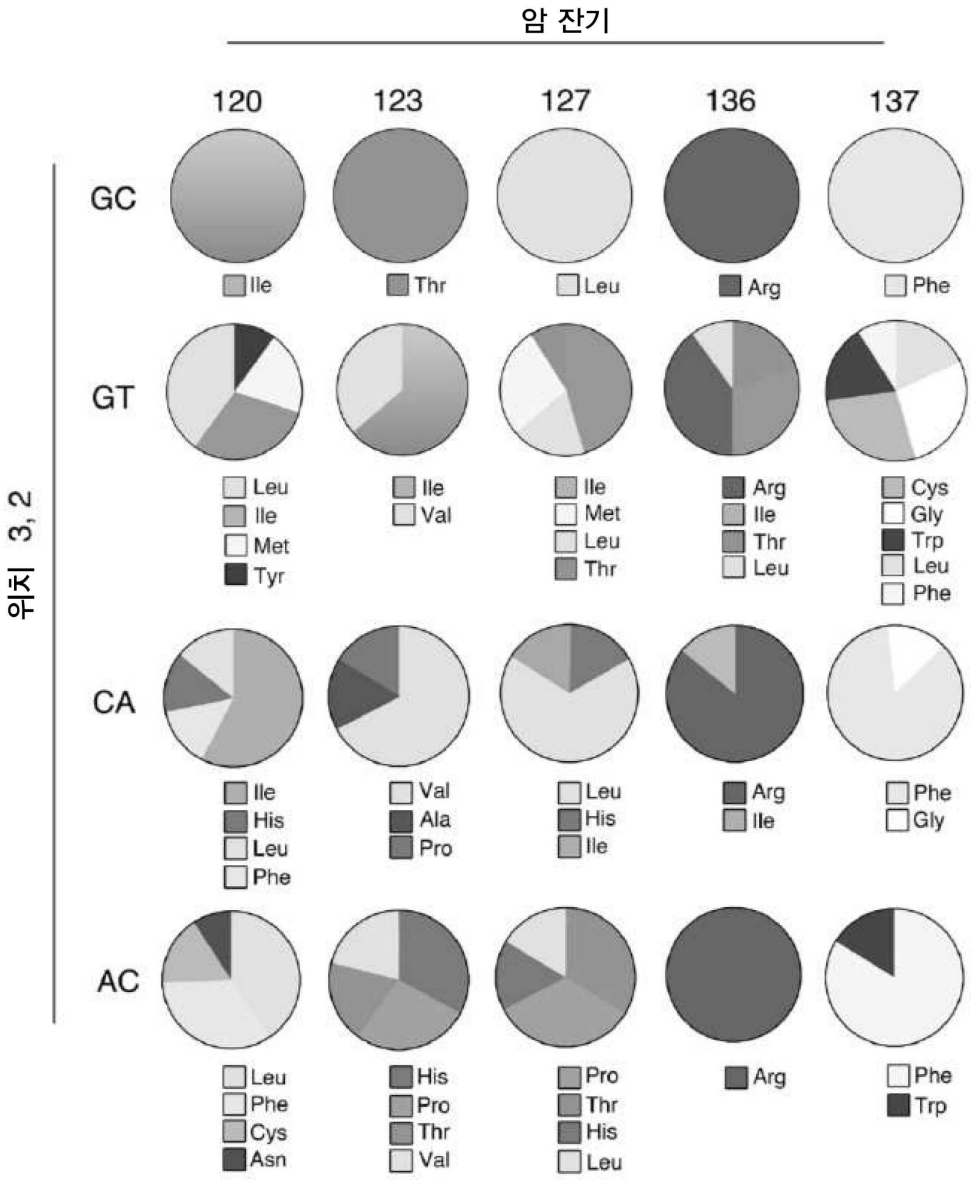

A
Left ZFBS
20-bp core
Right ZFBS
Reporter plasmid
ZFR
SV40
Luciferase
ZFR-mediated recombination
Excision of the SV40 promoter
Recombined reporter plasmid
Reduced Luminescence
C
ZFR 표적 부위
저
고
재조합도 (%)
B
촉매성 도메인
ZFR
Location
ZFR 표적 부위
Left
Right
배 감소

도면34


A
ZFR 1, ch4: 19429
ZFR 2, chX: 27761426
ZFR 3, ch4: 11016278
B
L+R
L-단독
R-단독
도너 단독
ZFR 1
ZFR 2
ZFR 3
C
비수정
순방향
역방향
D

도면35

도면36


암 잔기
120
123
127
136
137
위치 3, 2
GC
Ile
Thr
Leu
Arg
Phe
GT
Leu
Ile
Met
Tyr
Ile
Val
Ile
Met
Leu
Thr
Arg
Ile
Thr
Leu
Cys
Gly
Trp
Leu
Phe
CA
Ile
His
Leu
Phe
Val
Ala
Pro
Leu
His
Ile
Arg
Ile
Phe
Gly
AC
Leu
Phe
Cys
Asn
His
Pro
Thr
Val
Pro
Thr
His
Leu
Arg
Phe
Trp

도면37

3,2-bp

	클론	돌연변이들	CC	GG	CG	GC	GT	TT
	모		36	25	55	0.01	0.003	0.08
★	GC-1	L127I, I136R, G137F	44	78	57	20	0.09	0.06

	클론	돌연변이들	CC	TT	GA	CA	GT	CT
	모		36	0.002	0.008	0.004	0.003	0.086
★	GT-1	I120L, T123V, L127T, I136R, G137W	0.009	0.053	0.08	3.1	27	13.4
	GT-2	L127M, I136R, G137C	0.016	0.0003	0.092	0.207	39	6
	GT-3	L127I	0.0209	0.0009	0.41	0.77	39	13.5
	GT-4	D12A*, I120Y, T123I, L127I, I136T, G137L	0.0318	0.034	-	-	15	-
	GT-5	I120M, T123V	0.0068	0.013	-	-	6	-
	GT-6	I120L, T123I, L127I, G137F	0.0125	0.022	-	-	3	-

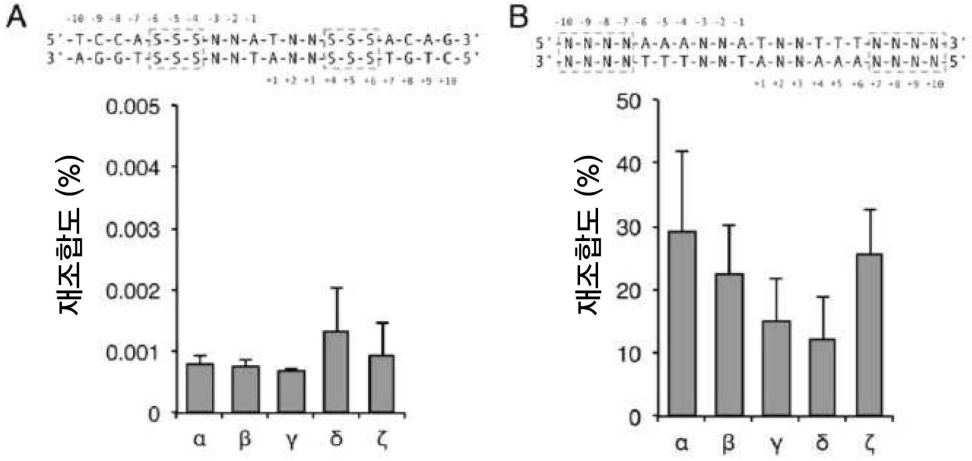
	클론	돌연변이들	CC	TT	GA	CA	GT	CT
	모		36	0.002	0.01	0.004	0.003	0.09
★	CA-1	T123V, I136R, G137F	0.018	0.003	1.2	13	23	14
	CA-2	T123V	0.02	0.06	-	11	27	-
	CA-3	T123V, N132D	0.02	0.6	-	8.2	11.2	-
	CA-4	L52I*, L118S*, I120F, L127H, I136R, G137F	-	-	-	0.87	-	-
	CA-5	I120L, T123V, I136R, G137F	-	-	-	0.43	-	-
	CA-6	E117Q, L118C*, I120H, T123P, L127I	-	-	-	0.5	-	-

	클론	돌연변이들	TT	AC	CA	GT	AG	GC
	모		0.002	0.005	0.004	0.003	0.008	0.01
★	AC-1	E117L, L118S*, I120L, T123P, L127H, I136R, G137F	0.09	1.7	0.05	6	1	0.008
	AC-2	E117L, I120N, I136R, G137F	0.0002	0.7	-	0.12	-	0.03
	AC-3	R116Q*, E117L, L118S*, I120F, T123H, L127T, I136P, G127F	0.0007	0.1	0.0005	0.0014	0.2	-
	AC-4	E117Y, I120F, T123P, L127P, I136R, G137F	0.0005	0.01	0.056	0.001	0.016	-

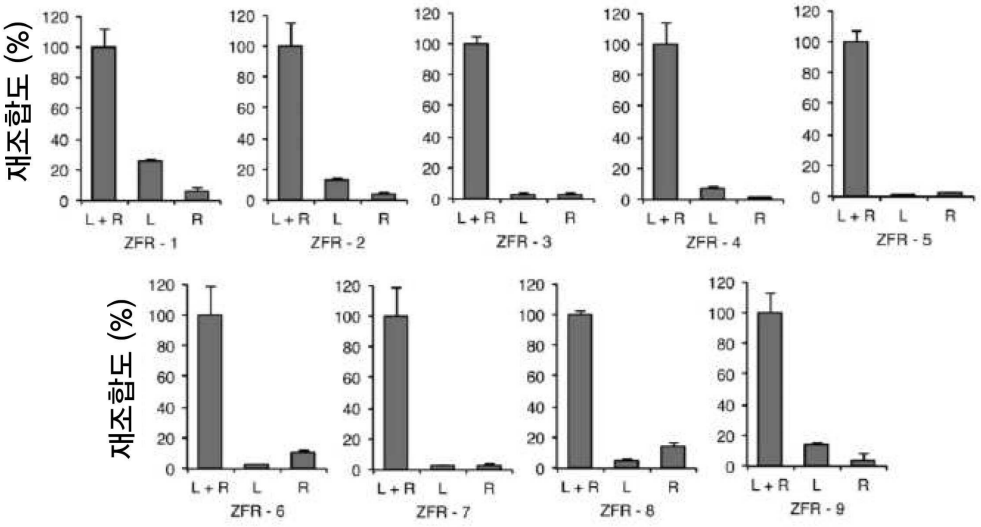
저 <0.1 1 10 100 고

재조합도 (%)

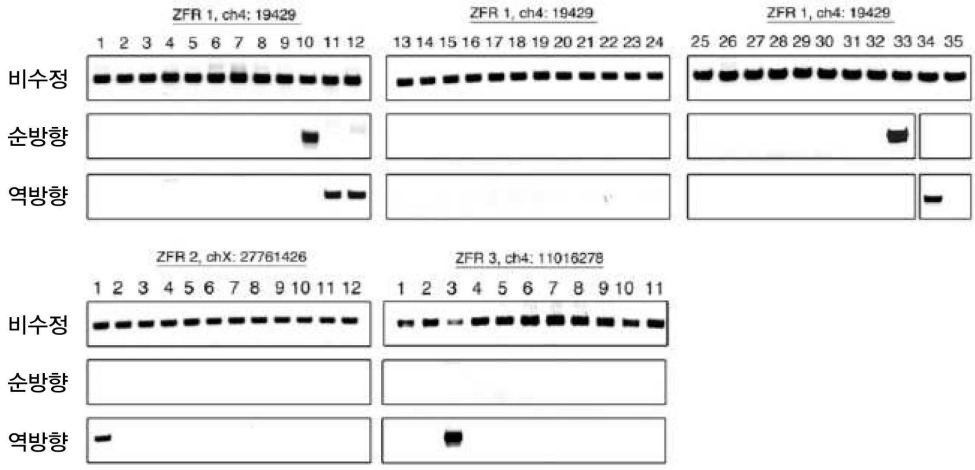
도면38



도면39


재조합도 (%)
120
100
80
60
40
20
0
L+R
L
R
ZFR - 1
ZFR - 2
ZFR - 3
ZFR - 4
ZFR - 5
ZFR - 6
ZFR - 7
ZFR - 8
ZFR - 9

도면40


ZFR 1, ch4: 19429
1 2 3 4 5 6 7 8 9 10 11 12
13 14 15 16 17 18 19 20 21 22 23 24
25 26 27 28 29 30 31 32 33 34 35
비수정
순방향
역방향
ZFR 2, chX: 27761426
1 2 3 4 5 6 7 8 9 10 11 12
ZFR 3, ch4: 11016278
1 2 3 4 5 6 7 8 9 10 11