(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0073887
(43) 공개일자 2013년07월03일

(51) 국제특허분류(Int. Cl.)
C07K 19/00 (2006.01) *C12N 15/52* (2006.01)
C12N 15/63 (2006.01)
(21) 출원번호 10-2012-7028770
(22) 출원일자(국제) 2011년05월02일
심사청구일자 없음
(85) 번역문제출일자 2012년11월01일
(86) 국제출원번호 PCT/US2011/000758
(87) 국제공개번호 WO 2011/139349
국제공개일자 2011년11월10일
(30) 우선권주장
61/343,729 2010년05월03일 미국(US)

(71) 출원인
상가모 바이오사이언스 인코포레이티드
미국 캘리포니아 94804 리치몬드 스위트 에이100 커넬 불러바드 501
(72) 발명자
파스촌 데이비드
미국 캘리포니아 94804 리치몬드 스위트 에이100 커넬 불러바드 501 상가모 바이오사이언스 인코포레이티드 포인트 리치몬드 테크 센터 내
레바르 에드워드 제이.
미국 캘리포니아 94804 리치몬드 스위트 에이100 커넬 불러바드 501 상가모 바이오사이언스 인코포레이티드 포인트 리치몬드 테크 센터 내
(74) 대리인
송봉식, 정삼영

전체 청구항 수 : 총 13 항

(54) 발명의 명칭 **아연 핑거 모듈을 연결하기 위한 조성물**

(57) 요 약

본 명세서에서 1개 이상의 염기쌍에 의해 분리된 모듈 서브사이트에 DNA 결합 모듈을 특이적이며 선택적으로 결합시키는 연결을 위한 조성물이 개시된다. 또한 이들 링커를 포함하는 조성물의 제조 및 사용방법이 설명된다.

대 표 도 - 도7

하나의 염기쌍을 스킵하기 위해 선택한 링커에 의한 호환성 연구

A

링커		ELISA 스코어											
서열	지정	ZFP1	ZFP2	ZFP3	ZFP4	ZFP5	ZFP6	ZFP7	ZFP8	ZFP9	ZFP10	ZFP11	ZFP12
TGGGGSQKP	가요성	0.75	1.08	0.21	0.26	0.08	0.05	0.05	0.05	0.05	0.05	0.05	0.05
TVPRPTPPKP	1e	1.19	1.30	1.05	1.10	0.75	0.29	0.48	0.38	0.50	0.18	0.05	0.05
TYPRPIAAKP	1f	1.25	1.35	1.03	0.94	0.96	0.30	0.33	0.12	0.30	0.11	0.05	0.05
TPNRRPAPKP	1d	1.20	1.08	1.04	0.83	0.85	0.25	0.39	0.28	0.24	0.12	0.05	0.05
THPRAPIPKP	1c	0.92	1.18	0.66	0.56	0.91	0.23	0.21	0.26	0.18	0.05	0.05	0.05

B

링커		가요성 링커에 대해 정규화된 ELISA 스코어												
서열	지정	ZFP1	ZFP2	ZFP3	ZFP4	ZFP5	ZFP6	ZFP7	ZFP8	ZFP9	ZFP10	ZFP11	ZFP12	평균
TGGGGSQKP	가요성	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
TVPRPTPPKP	1e	1.6	1.2	4.9	4.1	9.2	5.9	9.6	7.6	10.1	3.6	1.0	1.0	5.0
TYPRPIAAKP	1f	1.7	1.2	4.9	3.5	11.7	5.9	6.5	2.3	6.0	2.2	1.0	1.0	4.0
TPNRRPAPKP	1d	1.6	1.0	4.9	3.1	10.4	4.9	7.8	5.7	4.8	2.4	1.0	1.0	4.0
THPRAPIPKP	1c	1.2	1.1	3.1	2.1	11.2	4.7	4.3	5.3	3.6	1.0	1.0	1.0	3.3

밑줄친 값은 4배 초과의 개선을 나타낸다

특허청구의 범위

청구항 1

5개 이상의 아미노산 잔기의 아미노산 링커 서열에 의해 접합된 2개의 아연 핑거 단백질을 포함하는 융합 단백질로서, 상기 링커는 N-말단 잔기, C-말단 잔기, 및 상기 말단 잔기들의 내부 잔기를 포함하며, 추가로 상기 N-말단 잔기 또는 내부 잔기는 적어도 하나의 프롤린 잔기를 포함하는 것인, 융합 단백질.

청구항 2

제1항에 있어서, 상기 링커는 적어도 2개의 염기성 아미노산 잔기를 더 포함하는 것인, 융합 단백질.

청구항 3

제2항에 있어서, 상기 염기성 아미노산 잔기는 아르기닌 잔기, 히스티딘 잔기, 리신 잔기 또는 이들의 조합으로 이루어진 군으로부터 선택된 것인, 융합 단백질.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 아연 핑거 단백질 중 하나 이상은 조절 도메인을 더 포함하는 것인, 융합 단백질.

청구항 5

제4항에 있어서, 상기 조절 도메인은 전사 조절 도메인인 것인, 융합 단백질.

청구항 6

제5항에 있어서, 상기 조절 도메인은 활성화 도메인 또는 억제 도메인인 것인, 융합 단백질.

청구항 7

제4항에 있어서, 상기 조절 도메인은 절단 도메인 또는 절단 절반 도메인인 것인, 융합 단백질.

청구항 8

제1항 내지 제7항 중 어느 한 항의 융합 단백질을 암호화하는 폴리뉴클레오타이드.

청구항 9

제1항 내지 제7항 중 어느 한 항의 융합 단백질 및/또는 제8항의 폴리뉴클레오타이드를 포함하는 세포.

청구항 10

세포 내 유전자의 발현을 조절하는 방법으로서, 상기 방법은 제1항 내지 제7항 중 어느 한 항의 융합 단백질 또는 제8항의 폴리뉴클레오타이드를 상기 세포에 투여하는 단계를 포함하는 것인, 세포 내 유전자의 발현 조절방법.

청구항 11

제10항에 있어서, 상기 조절은 상기 유전자의 상향 또는 하향 조절을 포함하는 것인, 세포 내 유전자의 발현 조절방법.

청구항 12

제10항에 있어서, 상기 조절은 상기 유전자의 절단을 포함하는 것인, 세포 내 유전자의 발현 조절방법.

청구항 13

제10항 내지 제13항 중 어느 한 항에 있어서, 상기 유전자는 내인성 세포 유전자인 것인, 세포 내 유전자의 발

현 조절방법.

명 세 서

기 술 분 야

[0001] **관련 출원과의 상호참조**

[0002] 본 출원은 2010년 5월 3일 출원된 미국 가특허출원 제61/343,729호의 우선권을 주장하며, 이 기초출원은 그의 전문이 본 명세서에 참조로 포함된다.

[0003] **연방정부 지원된 연구하에서 만들어진 발명에 대한 권리의 언급**

[0004] 해당 없음.

[0005] **기술분야**

[0006] 본 명세서는 게놈 및 단백질 유전자조작 분야에 관한 것이다.

배 경 기 술

[0007] 선택된 표적 부위에 결합하도록 유전자조작된 인식 영역을 가지는 아연 핑거 단백질은 다른 아연 핑거 단백질뿐만 아니라 조절 도메인에 규칙적으로 연결되며 유전자 발현 및 게놈 표적 부위를 변형하기 위해 사용된다. 예를 들어, 절단 도메인에 작동가능하게 연결된 DNA 결합 도메인을 포함하는 인공 뉴클레아제는 외인성 서열의 삽입, 하나 이상의 내인성 유전자의 불활성화, 유기체의 생성(예를 들어 동물 또는 작물) 및 변경된 유전자 발현 패턴을 가지는 세포주 등을 포함하는 게놈 서열의 표적화된 변경을 위해 사용되었다. 이에 대해서는, 예를 들어, 미국특허 제20050064474호; 제20060063231호; 제20070134796호; 제20080015164호 및 국제 특허 공개 제2007/139982호를 참조할 수 있다.

[0008] 아연 핑거 단백질 모듈(예를 들어 하나 이상의 핑거 중 유전자조작된 아연 핑거)은 전형적으로 TGEKP(서열번호 1)와 같은 5개 아미노산의 "정규(canonical)" 링커 서열 또는 더 긴 가요성 링커를 사용하여 서로 연결된다. 이에 대해서는, 예를 들어, 미국특허 제6,479,626호; 제6,903,185호; 제7,153,949호 및 미국특허 공개 제20030119023호를 참조할 수 있다. 그러나, 이들 정규 링커를 통해 연결된 아연 핑거 단백질 모듈은 표적 핵산 모듈 내 연결된 모듈 표적 서브사이트(subsite) 사이의 갭(gap)이 없을 때에만 가장 효과적으로 결합한다. 더 나아가, 상기 연결된 모듈이 1, 2 또는 3개의 염기쌍 갭이 있는 표적 부위에 결합하도록 설계된 앞서 설명한 길고, 가요성인 링커는 결합에 관하여 이들 상이한 염기쌍 갭 사이에서 차이를 나타내지 않는다. 이에 대해서는, 미국특허 제6,479,626호; 제6,903,185호; 제7,153,949호 및 미국특허 공개 제20030119023호를 참조할 수 있다. 따라서, 1, 2, 또는 3 bp 모듈간 갭을 포괄하는 단백질의 친화도를 개선시키고, 원하는 길이의 갭을 포괄하는 표적을 결합하기 위한 이들 단백질의 선택성을 개선시키며, 원하는 길이의 갭 없이 다른 표적에 비선택적으로 결합하지 않는 아연 핑거 모듈의 서로에 대한 연결을 위한 방법 및 조성물에 대한 필요가 존재한다. 인접한 모듈 서브사이트 사이의 0, 1, 2, 3 또는 훨씬 더 많은 염기쌍 갭 간에 차이를 나타내는 아연 핑거 모듈에 대한 링커는 아연 핑거 전사 인자(zinc-finger transcription factors, ZFP-TF) 및 아연 핑거 뉴클레아제(zinc finger nucleases, ZFN)를 포함하는 임의의 아연 핑거 융합 단백질의 더 큰 설계 능력을 허용한다.

발명의 내용

[0009] 본 명세서에서 서로에 대해 DNA 결합 모듈(예를 들어, 아연 핑거 모듈)을 연결하는 것에서 사용을 위한 링커가 개시된다. 또한 융합 단백질, 예를 들어 전사 조절 도메인과 같은 조절 도메인에 또는 뉴클레아제에 차례로 융합된 이들 링커를 포함하는 아연 핑거 단백질이 설명된다. 또한 본 명세서는 유전자 발현의 모듈화, 관심 영역 내 세포 DNA(예를 들어 내인성 세포 염색질)의 표적화된 절단 및/또는 세포 내 미리결정된 관심 영역에서 상동성 재조합을 위해 이들 융합 단백질 및 이들의 조성물을 사용하는 방법을 제공한다.

[0010] 따라서, 한 양태에서, 본 명세서에서 아미노 (N) 말단 핑거의 마지막 잔기(전형적으로 카복시 (C)-말단 아연 배위 잔기)와 C-말단 핑거의 첫 번째 잔기(전형적으로 첫 번째 (N-말단)-보존된 방향족 잔기) 사이에 5개 이상의 아미노산, 예를 들어 7 내지 17개의 아미노산을 포함하는 링커가 설명된다. 특정 실시형태에서, 해당 링커는 N-말단 잔기, C-말단 잔기, 및 말단 잔기의 내부 잔기를 포함하되, 추가로 N-말단 잔기 또는 내부 잔기는 적어도 하나의 프롤린 잔기, 예를 들어 상기 아미노산 서열 $X^{N\text{-말단}}$-X_n-$X^{C\text{-말단}}$을 포함하는 링커를 포함하며, 여기서 X는 임

의의 아미노산 잔기이고, X_n은 적어도 3개의 아미노산 잔기를 포함하며, $X^{N\text{-말단}}$ 및 X_n 중 적어도 하나는 프롤린 잔기를 포함한다. 특정 실시형태에서, 해당 링커는 적어도 2개의 프롤린 잔기(예를 들어 2, 3, 4 또는 그 이상)를 포함한다. 다른 실시형태에서, 상기 링커는 적어도 하나의 프롤린 잔기 및 적어도 하나의 염기성 잔기(예를 들어, Arg, His 또는 Lys)를 포함한다. 다른 실시형태에서, 상기 링커는 적어도 2개의 염기성 잔기(예를 들어 Arg, His 또는 Lys)를 포함한다. 특정 실시형태에서, 상기 링커는 표 4, 5, 6, 9, 10, 11 또는 13 중 어느 하나로 나타낸다.

[0011] 다른 양태에서, 본 명세서에서 설명되는 링커를 포함하는 융합 폴리펩타이드가 제공된다.

[0012] 다른 양태에서, 본 명세서에서 설명되는 링커 또는 융합 단백질을 암호화하는 폴리뉴클레오타이드가 제공된다.

[0013] 또 다른 양태에서, 본 명세서에서 설명되는 상기 폴리펩타이드(예를 들어 융합 폴리펩타이드) 및/또는 폴리뉴클레오타이드 중 어느 하나를 포함하는 세포가 또한 제공된다.

[0014] 추가 양태에서, 본 명세서에서 설명되는 폴리펩타이드(예를 들어, 융합 폴리펩타이드) 및/또는 폴리뉴클레오타이드를 포함하는 유기체(예를 들어, 포유류, 진균 및 식물)가 또한 제공된다.

[0015] 융합 단백질은, 예를 들어 세포에 융합 단백질을 전달함으로써 또는 세포에 융합 단백질을 암호화하는 폴리뉴클레오타이드를 전달함으로써 세포 내에서 발현될 수 있다. 폴리뉴클레오타이드가 DNA라면, 전사되고 번역되어 해당 융합 단백질을 만든다. RNA 분자로 전달된다면, 즉시 번역되며, 따라서 상기 융합 단백질을 만든다. 세포에 폴리뉴클레오타이드 및 폴리펩타이드를 전달하는 방법은 본 명세서의 다른 곳에서 제시된다.

[0016] 이들 및 다른 양태는 명세서에 비추어 전제로서 당업자에게 용이하게 명백할 것이다.

도면의 간단한 설명

[0017] **도 1, 패널 A 및 패널 B**는 예시적인 아연 단백질 및 링커의 아미노산 서열. **도 1a**는 이들 연구를 위해 사용한 각 숙주 ZFP의 아미노산 서열(서열번호 130; 서열번호 131; 서열번호 132 및 서열번호 133에서 나타내는 ZFP 8196의 F1-F4; 서열번호 134; 서열번호 135; 서열번호 136 및 서열번호 137에서 나타내는 ZFP 7263의 F1 내지 F4; 서열번호 138; 서열번호 139; 서열번호 140 및 서열번호 141에서 나타내는 ZFP 7264의 F1 내지 F4)을 나타낸다. 아미노산은 한 글자 코드로 명명된다. 각 서열은 아미노 말단 → 카복시 말단 방향으로 열거되며, 따라서 각 단백질의 아미노 말단은 핑거 1의 첫 번째 메티오닌이고, 카복시 말단은 핑거 4의 마지막 세린이다. "F1", "F2", "F3" 및 "F4"는 각각 각 단백질의 제1, 제2, 제3 및 제4 핑거를 명명한다. 밑줄은 링커 서열이 되는 것으로 통상적으로 고려되는 핑거 연접에 있는 아미노산 잔기를 의미한다. 인식 나선은 박스로 표시한다. **도 1b**는 중심 링커 내 2 또는 3개의 잔기에 대한 코돈을 2 내지 12개의 완전히 무작위화된 코돈의 혼합으로 치환함으로써 각 라이브러리가 만들어지는 링커 라이브러리 설계를 보여준다. 라이브러리 코돈은 $(NNS)_{2\text{-}12}$에 의해 나타낸다.

도 2, 패널 A 내지 패널 D는 표시한 아연 핑거 단백질 및 링커가 있는 파지 풀의 갭 선택성을 도시하는 그래프. 도 2A는 표 2C의 각 ZFP8196 표적에 결합을 위해 시험한 1bp 삽입 염기(ATAAACTGdCAAAAGGC(서열번호 33)(표 2A))를 가지는 표적을 사용하여 ZFP8196 라이브러리로부터 선택한 파지 풀을 나타낸다. 도 2B는 표 2C의 각 ZFP7263 표적에 결합을 위해 시험한 1bp 삽입 염기(CCACTCTGhTGGAAGTG(서열번호 43)(표 2A))를 가지는 표적을 사용하여 ZFP7263 라이브러리로부터 선택한 파지 풀을 나타낸다. 도 2C는 표 2C의 각 ZFP7264 표적에 결합을 위해 시험한 1bp 삽입 염기(TTAAAGCGhGCTCCGAA(서열번호 38)(표 2A))를 가지는 표적을 사용하여 ZFP7264 라이브러리로부터 선택한 파지 풀을 나타낸다. 도 2D는 표 2C의 각 ZFP8196 표적에 결합을 위해 시험한 2bp 삽입 염기(ATAAACTGdbCAAAAGGC(서열번호 34)(표 2A))를 가지는 표적을 사용하여 ZFP8196 라이브러리로부터 선택한 파지 풀을 나타낸다. 각 시험은 또한 DNA에 비특이적 결합을 배제하기 위한 다른 2개의 숙주 ZFP에 대한 2개의 대조군 표적뿐만 아니라 표적 부위를 포함하지 않은 음성 대조군 샘플을 포함하였다. 각 표적에 성공적으로 결합한 파지의 %가 표시된다. 각 파지 풀은 제5 라운드의 선택으로부터 나온다. 앞서 설명한 바와 같이 보유 효율을 근본적으로 결정하였다(Rebar, et al. Methods in Enzymology, 1996 (267):129-149).

도 3, 패널 A 및 패널 B는 표시한 갭을 함유하는 표적 부위에 대해 선택된 링커를 나타낸다. 도 3A는 ZFP8196, ZFP7263, 및 ZFP7264(서열 번호 142 내지 166)의 맥락에서 1 bp 갭을 스킵하기 위해 선택된 링커 서열을 나타낸다. 도 3B는 ZFP8196의 맥락에서 2 bp 갭을 스킵하기 위한 링커 서열(서열 번호 167 내지 174)을 나타낸다. 선택한 링커는 프롤린 및 아르기닌(음영 표시)에 대해 강화된다. 길이 선호도는 또한 명백하며 스킵한 염기의 수

에 의존한다.

도 4, 패널 A 내지 패널 E는 아연 핑거 단백질 명명 ZFP8196 내 1개의 염기쌍을 스킵하기 위해 선택된 링커에 대한 갭 선택성을 도시하는 그래프. 각 패널에 대해, ELISA 스코어는 그의 비갭화된 표적 부위 상에서 본래의 스킵하지 않는 링커로 정규화되었다. "갭 서열"은 (-)가 비갭화된 표적을 표시하는 경우 모듈 서브사이트 사이의 염기(들)의 동일성을 말한다. 도 4A 내지 도 4C는 갭 스킵 링커 중 3개(링커는 1f(서열번호 54), 1d(서열번호 56) 및 1c(서열번호 55)로 언급됨)로부터의 결과를 도시한다. 도 4D는 앞서 인간 세포 내 내인성 좌위의 변형을 가능하게 한 것으로 나타난 표준 가요성 링커에 의한 결과를 나타내며(TGGGGSQKP, 서열번호 2)(문헌[Hockemeyer et al. (2009) Nature Biotechnology 27:851-857] 참조) 도 4E는 앞서 공개된 가요성 링커에 대한 결과를 도시한다(LRQKDERP, 서열번호 3)(문헌[Kim JS & Pabo CO (1998) Proc Natl Acad Sci U S A 95(6):2812-2817] 참조). 선택된 링커 1c, 1d 및 1f(도 4A 내지 도 4C)는 모두 하나의 염기 쌍 갭과 함께 4개의 표적 부위에 대한 확실한 선호도를 보여주는 반면, 도 4D 및 도 4E의 대조군 링커는 덜 효과적인 전반적 결합 및 거의 없는 갭 선택성을 나타낸다.

도 5, 패널 A 내지 D는 ZFP7264 내 1개의 염기쌍을 스킵하기 위해 선택한 링커에 대한 갭 선택성을 도시하는 그래프. 도 5A 내지 도 5C는 ZFP7264 배경에서 1e 링커(서열번호 12)의 ELISA 시험으로부터의 결과를 도시한다. 도 5A는 모듈 서브사이트 사이의 1bp 갭을 스킵하기 위해 선택된 1e 링커에 대한 결과를 나타낸다. 도 5B는 표준 가요성 링커에 대한 결과(TGGGGSQKP, 서열번호 2)를 나타내며, 도 5C는 가요성 링커 LRQKDERP(서열번호 3)에 대한 결과를 나타낸다. ELISA 스코어는 그의 비갭 표적 상에서 본래의 스킵하지 않는 ZFP7264로 정규화되었다. "갭 서열"은 (-)가 갭화되지 않은 표적을 표시하는 경우 모듈 서브사이드 사이의 스킵되지 않은 염기(들)의 동일성이다. 도 5D는 다른 패널의 0-5와 비교하여 ELISA 스코어 범위가 0-0.6인 경우 도 5B로부터의 데이터의 확장된 버전을 나타낸다.

도 6, 패널 A 내지 F는 ZFP8196 내 2bp를 스킵하기 위해 선택된 링커에 대한 갭 선택성을 도시하는 그래프. 도 6A 내지 도 6E는 ZFP8196 배경에서 모듈 서브사이드 사이의 2 bp 갭을 스킵하기 위해 선택된 링커를 시험하는 ELISA로부터의 결과를 도시한다. 도 6A 내지 도 6C는 선택된 링커 2f(서열번호 69), 2d(서열번호 70) 및 2e(서열번호 71)에 대한 결과를 도시하는 한편, 도 6D는 앞서 공개된 가요성 링커(LRQKDGGGSERP(서열번호 68))에 대한 결과를 나타내고 도 6E는 표준 가요성 링커(TGGGGSGGSQKP(서열번호 14))에 대한 결과를 나타낸다. 도 6F는 다른 패널의 0-1과 비교하여 ELISA 스코어 범위가 0-0.1인 경우 도 6E에서 나타낸 데이터의 확장된 버전을 나타낸다. "갭 서열"은 (-)가 갭화되지 않은 표적을 표시하는 경우 모듈 서브사이드 사이의 스킵되지 않은 염기(들)의 동일성이다. 선택된 링커(도 6A 내지 도 6C)는 1 bp 갭과 비교하여 2 bp에 대해 확신한 선호도를 증명한 반면, 도 6D 및 도 6E의 대조군 링커는 덜 효과적인 전반적 결합 및 거의없는 갭 선택성을 나타낸다.

도 7, 패널 A 및 패널 B는 상이한 ZFP 배경에 대한 1 bp 스키핑 링커의 호환성을 분석하기 위해 설계된 연구로부터의 ELISA 데이터의 요약을 도시. 12개의 상이한 ZFP를 시험하였다(ZFP1, ZFP2 등으로 표시). 도 7A는 ZFN으로 사용될 때 내인성 IL2Rγ 좌위를 효율적으로 변형시키는 것으로 나타난 표준 양성 대조군 ZFP로 정규화된 ELISA 스코어를 나타낸다(Urnov et al. (2005) *Nature* 435(7042):646-651). 도 7B는 표준 가요성 링커 TGGGGSQKP(서열번호 2)를 함유하는 각각의 본래 ZFP로 추가로 정규화된 모든 스코어를 나타낸다. 도 7B의 밑줄친 값은 가요성 링커 TGGGGSQKP(서열번호 2)에 의한 동일한 숙주 ZFP와 비교하여 선택된 링커(1e(서열번호 12), 1f(서열번호 54), 1d(서열번호 56), 및 1c(서열번호 55))에 의한 ZFP에 대한 ELISA 스코어의 4배 초과의 개선을 나타낸다. 전반적인 링커 1e, 1f, 1d 및 1c는 가요성 링커에 걸쳐 3 내지 5배의 ELISA 스코어의 일반적 증가를 야기한다.

도 8, 패널 A 및 B는 상이한 ZFP 배경에 대해 2 bp 스키핑 링커의 호환성을 분석하도록 설계된 연구로부터의 ELISA 데이터의 요약을 도시. 6개의 상이한 ZFP를 시험하였다(ZFP13, ZFP14 등으로 표시). 도 8A는 ZFN으로 사용할 때 내인성 IL2Rγ 좌위를 효율적으로 변형시키는 것으로 나타난 표준 양성 대조군 ZFP으로 정규화된 ELISA 스코어를 나타낸다(Urnov *et al.* (2005) *Nature* 435(7042):646-651). 도 8B는 표준 가요성 링커 TGGGGSGGSQKP(서열번호 14)를 함유하는 각각의 본래 ZFP로 추가로 정규화된 모든 스코어를 나타낸다. 도 8B의 밑줄친 값은 가요성 링커 TGGGGSGGSQKP(서열번호 14)를 가지는 동일한 숙주 ZFP와 비교하여 선택된 링커(2f(서열번호 69), 2d(서열번호 70) 및 2e(서열번호 71))를 가지는 ZFP에 대한 ELISA 스코어의 2배 초과의 개선을 나타낸다. 전반적인 링커 2f(서열번호 69), 2d(서열번호 70) 및 2e(서열번호 71)는 가요성 링커보다 1.9-2.4배 이상의 ELISA 스코어의 일반적 증가를 야기한다.

도 9, 패널 A 및 B는 CEL-I 분석에 의해 결정한 바와 같이 선택된 링커를 함유하는 ZFN에 의한 내인성 유전자

변화 연구의 결과를 도시한다. 도 9A 및 도 9B는, 본 명세서에서 설명한 링커가 상이한 ZFN의 맥락에서 사용될 수 있는지 여부를 결정하기 위하여, CEL-I 분석에 의해 내인성 좌위에서 ZFN 뉴클레아제 활성(비상동성 말단 접합(non-homologous end joining, NHEJ) 활성, Surveyor(상표명), Transkaryotic을 측정)을 결정하도록 사용한 예시적 겔을 도시한다. 도 9A에서 나타낸 겔은 ZFN3 및 ZFN4 배경에서 1e(서열번호 12), 1f(서열번호 54), 1d(서열번호 56), 및 1c(서열번호 55) 링커로부터의 결과를 도시한다. 도 9B에서 나타내는 겔은 ZFN14 배경에서 2f, 2d 및 2e 링커로부터의 결과를 도시한다. NHEJ에 의한 유전자 변형%, "Gene mod. (%)"는 레인의 바닥부에 표시한다. 음성 대조군"neg"는 GFP 함유 플라스미드로 트랜스펙션한 샘플이다. 표준 가요성 링커(도 9A에 대해 TGGGGSQKP(서열번호 2) 및 도 9B에 대해 TGGGGSGGSQKP(서열번호 14))를 사용하여 ZFN으로부터의 결과를 "C"로 라벨 표시한 레인에 나타낸다. 라벨 표시하지 않은 레인은 이 연구에서 추가로 발생하지 않은 다른 링커를 함유하는 ZFN의 샘플을 함유한다. 겔의 데이터는 본 명세서에서 설명한 링커는 가요성 링커와 비교하여 유전자 변형 수준이 상당히 증가한다는 것을 증명한다.

도 10, 패널 A 및 패널 B는 1bp 스킵을 위해 선택된 표시 링커를 함유하는 도 9에 대해 상기 설명한 바와 같은 ZFN에 대한 유전자 변형 연구의 요약을 도시한다. 도 10A는 시험한 5개의 링커(가요성, 1e(서열번호 12), 1f(서열번호 54), 1d(서열번호 56), 및 1c(서열번호 55))의 세트를 가지는 각 ZFN에 대한 유전자 변형%의 정량화이다. 도 10B는 가요성 링커(TGGGGSQKP, 서열번호 2)에 대해 정규화된 이런 동일 데이터를 나타내며 또한 모든 활성 ZFN 쌍을 가로지르는 유전자 변형의 평균 증가를 나타낸다. 높은 발현 조건을 사용하여 생성된 샘플(실시예 3 참조)은 회색으로 강조된다. 2배 초과로 유전자 변형 수준을 개선한 ZFN 함유 예시적 링커는 도 10B에서 밑줄 표시된다. 전반적인 ZFN 함유 링커 1e(서열번호 12), 1f(서열번호 54), 1d(서열번호 56), 및 1c(서열번호 55)는 가요성 링커를 함유하는 각각의 숙주 ZFN을 그의 1.8 내지 2.8배의 유전자 변형으로 평균적 증가를 야기한다.

도 11, 패널 A 및 B는 숙주 ZFN의 모듈 서브사이트 사이의 2 bp 갭을 스킵하기 위해 선택된 표시 링커를 함유하는 ZFN에 대해 도 9에서 설명한 것과 같은 유전자 변형 연구의 요약을 도시한다. 도 11A는 시험한 4개의 링커(가요성, 2f(서열번호 69), 2d(서열번호 70) 및 2e(서열번호 71))의 세트를 가지는 각 ZFN에 대한 유전자 변형%의 정량화이다. 도 11B는 가요성 링커(TGGGGSGGSQKP, 서열번호 14)로 정규화된 이런 동일 데이터를 나타내며 또한 모든 활성 ZFN 쌍에 걸친 평균적 증가를 나타낸다. 높은 발현 조건을 사용하여 생성된 샘플(실시예 3을 참조)은 회색으로 강조된다. 2배 초과로 유전자 변형 수준을 개선시킨 ZFN 함유 예시적 링커는 도 11B에서 밑줄 표시된다. ZFN 함유 링커 2f(서열번호 69), 2d(서열번호 70) 및 2e(서열번호 71)은 가요성 링커를 함유하는 각각의 숙주 ZFN을 그에 대해 1.5 내지 2.0배의 유전자 변형으로 평균 증가를 야기하였다.

도 12는 2-bp 갭을 포괄하는 링커에 대한 두 번째 선택을 위해 사용한 숙주 ZFP8196의 아미노산 서열을 나타낸다. 아미노산은 한 글자 코드로 명명된다. 서열은 아미노 말단 → 카복시 말단 방향으로 열거되며, 따라서 단백질의 아미노 말단은 핑거 1의 제1 메티오닌이고, 카복시 말단은 핑거 4의 마지막 세린이다. "F1"(서열번호 130), "F2"(서열번호 131), "F3"(서열번호 132) 및 "F4"(서열번호 133)는 각각 단백질의 제1, 제2, 제3 및 제4 핑거를 명명한다. 인식 나선은 박스로 표시한다. 중심 링커 내 2개의 잔기를 5 내지 7개의 완전히 무작위화된 코돈의 혼합으로 치환한 다음, 하나의 코돈을 무작위화하여 페닐알라닌(F), 리신(L), 또는 티로신(Y) 잔기 중 하나를 얻음으로써 링커 라이브러리가 만들어지며, 마지막 3개의 코돈은 아르기닌(R), 프롤린(P), 및 프롤린(P)로 고정된다. 라이브러리 코돈은 $(NNS)_{5-7}$ 및 (F/L/Y)으로 나타난다.

도 13, 패널 A 및 패널 B는 2-bp 갭을 포괄하는 링커에 대한 두 번째 선택으로부터 파지 풀의 갭 선택성 및 선택에서 얻어진 클론의 결과 아미노산 서열을 도시한다. 도 13A는 표 2C에서 각 ZFP8196 표적에 대한 결합을 위해 시험한 2bp 삽입 갭(ATAAACTGdbCAAAAGGC(서열번호 34)(표 2A))을 가지는 표적을 사용하여 ZFP8196 라이브러리로부터 선택된 파지 풀을 나타낸다. 각 시험은 또한 DNA에 비특이적 결합을 배제하기 위한 하나의 다른 숙주 ZFP에 대한 대조군 표적뿐만 아니라 표적 부위를 포함하지 않은 음성 대조군 샘플을 포함하였다. 각 표적과 성공적으로 결합된 파지의 %가 표시된다. 파지 풀은 제6 라운드의 선택으로부터 나온다. 앞서 설명한 바와 같이 보유 효율을 근본적으로 결정하였다(Rebar, *et al. Methods in Enzymology*, 1996 (267):129-149). 도 13B는 ZFP8196의 맥락에서 두 번째 선택으로부터 2 bp 갭을 스킵하기 위해 선택한 링커의 아미노산 서열(서열번호 175 내지 210)을 나타낸다. 선택된 링커는 프롤린 및 아르기닌이 풍부하다(음영처리).

발명을 실시하기 위한 구체적인 내용

[0018] 본 명세서에서 DNA-결합 도메인, 특히 아연 핑거 모듈을 다른 아연 핑거 모듈에 연결하기 위한 조성물이 설명된다. 앞서 설명한 링커와 달리, 본 명세서에서 설명하는 링커는 1 또는 2 bp의 모듈 서브사이트 사이에 갭을 함

유하는 표적을 우선적으로 및/또는 선택적으로 결합시킨다. 링커는 또한 현존하는 링커 설계보다 더 높은 친화도로 1 또는 2 bp 갭을 함유하는 표적을 결합할 수 있다. 예시적인 링커는 표 11 및 표 13에 나타낸다. 따라서 본 명세서에 설명된 특정 링커는 특이적 표적 부위에 결합하는 아연-핑거 단백질을 설계하는 능력을 상당히 증가시키며, 이에 따라 이들 링커를 포함하는 융합 단백질(예를 들어 ZFP-TF 또는 ZFN)의 활성을 증가시킨다.

[0019] **일반**

[0020] 방법의 실시뿐만 아니라 본 발명에 개시된 조성물의 제조와 사용은, 달리 언급되지 않는 한, 분자생물학, 생화학, 염색질 구조 및 분석, 컴퓨터 계산 화학, 세포 배양, 재조합 DNA 및 당업계에 내의 관련 분야의 통상적인 기술이 사용된다. 이러한 기술은 문헌에서 상세하게 설명된다. 이에 대해서는, 예를 들어, 문헌[Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987] 및 정기 업데이트물; 연속 간행물[METHODS IN ENZYMOLOGY, Academic Press, San Diego]; [Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol.304, "Chromatin" (P.M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999]을 참조할 수 있다.

[0021] **정의**

[0022] 용어 "핵산", "폴리뉴클레오타이드" 및 "올리고뉴클레오타이드"는 상호호환적으로로 사용되며, 선형 또는 고리형 입체구조, 단일- 또는 이중-가닥 형태인 데옥시리보뉴클레오타이드 또는 리보뉴클레오타이드 중합체를 말한다. 본 발명의 개시의 목적을 위하여, 이러한 용어는 중합체의 길이에 대한 한정으로서 해석되어서는 않는다. 상기 용어는 천연 뉴클레오타이드의 공지된 유사체뿐만 아니라, 염기, 당 및/또는 인산염 부분(예를 들어, 포스포로티오에이트 백본)이 변형된 뉴클레오타이드를 포괄할 수 있다. 일반적으로, 특정 뉴클레오타이드의 유사체는 동일한 염기쌍 특이성을 가지며; 즉, A의 유사체는 T와 염기쌍을 형성할 것이다.

[0023] 용어 "폴리펩타이드", "펩타이드", 및 "단백질"은 상호호환적으로로 사용되며, 아미노산 잔기의 중합체를 말한다. 상기 용어는 또한 하나 이상의 아미노산이 대응하는 자연적으로 발생하는 아미노산의 화학적 유사체 또는 변형된 유도체인 아미노산 중합체에 적용된다.

[0024] 폴리펩타이드는 전형적으로 2개의 펩타이드가 단지 보존적 치환에 의해 다른 경우, 제2 폴리펩타이드와 실질적으로 동일하다. 단백질을 설명할 때 "보존적 치환"은 단백질의 활성을 실질적으로 변경하지 않는 단백질의 아미노산 조성물 내 변화를 말한다. 따라서, 특정 아미노산 서열의 "보존적으로 변형된 변형"은 단백질 활성에 중요하지 않은 아미노산의 아미노산 치환 또는 아미노산의 유사한 특성(예를 들어, 산성, 염기성, 양으로 또는 음으로 하전, 극성 또는 비극성 등)을 가지는 다른 아미노산으로 치환을 말하며, 동등하게 중요한 아미노산의 치환은 활성을 실질적으로 변경하지 않는다. 기능적으로 유사한 아미노산을 제공하는 보존적 치환 표는 당업계에 잘 알려져 있다. 이에 대해서는, 예를 들어 문헌[Creighton (1984) Proteins, W. H. Freeman and Company]을 참조할 수 있다. 추가로, 암호화된 서열 내 하나의 아미노산 또는 적은 %의 아미노산을 변경하고, 첨가하거나 결실시키는 개개의 치환, 결실 또는 첨가는 또한 "보존적으로 변형된 변형"이다.

[0025] "결합"은 거대분자 사이(예를 들어, 단백질과 핵산 사이)의 서열 특이적인 비공유적 상호작용을 말한다. 상호작용이 전체로서 서열 특이적이라면, 결합 상호작용의 모든 성분들이 서열 특이적(예를 들어, DNA 백본의 인산염 잔기와의 접촉)일 필요는 없다. 일반적으로, 이런 상호작용은 $10^{-6}M^{-1}$ 또는 그 이하의 해리 상수(K_d)를 특징으로 한다. "친화도"는 것은 결합 강도를 의미하는데, 결합 친화도의 증가는 더 낮은 K_d와 상관 관계가 있다.

[0026] "결합 단백질"은 다른 분자와 비공유적으로 결합할 수 있는 단백질이다. 결합 단백질은, 예를 들어, DNA 분자(DNA-결합 단백질), RNA 분자(RNA-결합 단백질) 및/또는 단백질 분자(단백질-결합 단백질)와 결합할 수 있다. 단백질-결합 단백질의 경우에, 이것은 자기 자신과 결합할 수 있고(동종이합체, 동종삼합체 등을 형성) 및/또는 상이한 단백질 또는 단백질들의 하나 이상의 분자와 결합할 수 있다. 결합 단백질은 한 가지 이상의 유형의 결합 활성을 가질 수 있다. 예를 들어, 아연 핑거 단백질은 DNA-결합 활성, RNA-결합 활성 및 단백질-결합 활성을 가진다.

[0027] "아연 핑거 DNA 결합 단백질"(또는 결합 도메인)은, 아연 이온의 배위 결합을 통하여 구조가 안정화되는 결합 도메인 내의 아미노산 서열의 영역인 하나 이상의 아연 핑거를 통하여 서열 특이적인 방식으로 DNA와 결합하는

단백질 또는 더 큰 단백질 내의 도메인이다. 용어 아연 핑거 DNA 결합 단백질은 흔히 아연 핑거 단백질 또는 ZFP로 약칭된다.

[0028] 아연 핑거 결합 도메인은 미리 결정해 놓은 뉴클레오타이드 서열과 결합하도록 "유전자 조작"될 수 있다. 아연 핑거 단백질을 조작하는 방법의 비제한적인 예는 설계 및 선택이다. 설계된 아연 핑거 단백질은 그 설계/조성이 주로 합리적인 기준으로부터 정해지는 것으로 자연에서 발생하지 않는 단백질이다. 설계를 위한 합리적인 기준에는 치환 규칙 및 기존의 ZFP 설계와 결합 데이터의 정보를 저장하고 있는 데이터베이스 정보 처리를 위한 컴퓨터를 이용한 알고리즘의 적용이 포함된다. 이에 대해서는, 예를 들어, 미국등록특허 제6,140,081호; 제6,453,242호; 제6,534,261호; 및 제WO 98/53058호; 제WO 98/53059호; 제WO 98/53060호; 제WO 02/016536호; 및 제WO 03/016496호를 참조할 수 있다.

[0029] "선택된" 아연 핑거 단백질은 파지 디스플레이, 상호작용 트랩 또는 혼성 선택과 같이 주로 실험적 과정의 결과로서 생산되는 것으로 자연에서 발견되지 않는 단백질이다. 이에 대해서는, 예를 들어, 미국등록특허 제5,789,538호; 제5,925,523호; 제6,007,988호; 제6,013,453호; 제6,200,759호; WO 95/19431; 제WO 96/06166호; 제WO 98/53057호; 제WO 98/54311호; 제WO 00/27878호; 제WO 01/60970호; 제WO 01/88197호; 및 제WO 02/099084호를 참조할 수 있다.

[0030] "조절 도메인" 또는 "기능적 도메인"은 DNA 결합 도메인, 즉 ZFP에 묶일 때 전사 조절 활성을 가지는 단백질 또는 단백질 도메인을 말한다. 전형적으로, 조절 도메인은 ZFP에 공유적으로 또는 비공유적으로 연결되어(예를 들어 융합 분자를 형성) 전사 조절을 달성한다. 조절 도메인은 활성화 도메인 또는 억제 도메인일 수 있다. 활성화 도메인은, 이에 제한되는 것은 아니지만, VP16, VP64 및 핵인자 카파-B의 p65 서브유닛을 포함한다. 억제 도메인은, 이에 제한되는 것은 아니지만, KOX, KRAB MBD2B 및 v-ErbA을 포함한다. 추가적인 조절 도메인은, 예를 들어 전사 인자 및 공동인자(예를 들어, MAD, ERD, SID, 조기 생장 반응 인자 1, 및 핵 호르몬 수용체), 엔도뉴클레아제, 인테그라제, 재조합효소, 메틸트랜스퍼라제, 히스톤 아세틸트랜스퍼라제, 히스톤 데아세틸라제 등을 포함한다. 활성자 및 억제자는 보조 활성자 및 보조 억제자를 포함한다(예를 들어, 문헌[Utley *et al.*, *Nature* 394:498-502 (1998)] 참조). 대안으로, ZFP는 조절 도메인 없이 단독으로 작용하여 전사 조절을 달성할 수 있다. 조절 도메인은 또한 뉴클레아제, 예컨대 절단 도메인 또는 절단 절반 도메인일 수 있다.

[0031] "절단"은 DNA 분자의 공유 백본의 파손을 말한다. 절단은, 이에 제한되는 것은 아니지만, 포스포다이에스터 결합의 효소적 또는 화학적 가수분해를 포함하는 다양한 방법에 의해 개시될 수 있다. 단일 가닥 절단과 이중 가닥 절단이 둘다 가능하며, 이중 가닥 절단은 2가지의 별개의 단일 가닥 절단 현상의 결과로써 일어날 수 있다. DNA 절단은 블런트(blunt) 말단 또는 엇갈린(staggered) 말단 중 하나의 생성을 초래할 수 있다. 특정 실시형태에서, 융합 폴리펩타이드는 표적화된 이중 가닥 DNA 절단을 위해 사용된다.

[0032] "절단 절반 도메인"은 제2 폴리펩타이드(동일 또는 상이함)와 함께 절단 활성(바람직하게 이중 가닥 절단 활성)을 갖는 복합체를 형성하는 폴리펩타이드 서열이다. 용어 "제1 및 제2 절단 절반 도메인", "+ 및 - 절단 절반 도메인" 및 "우측 및 좌측 절단 절반 도메인"은 이합체화되는 절단 절반 도메인 쌍을 말하며, 상호 호환적으로 사용된다.

[0033] "유전자조작된 절단 절반 도메인"은 다른 절단 절반 도메인(예를 들어, 유전자조작된 다른 절단 절반 도메인)과 절대 이종 이합체를 형성하도록 변형된 절단 절반 도메인이다. 또한, 이에 대해서는, 미국특허공보 제2005/0064474호; 제2007/13989호를 참조할 수 있으며, 이들은 그 전문이 본 명세서에 참조로 포함된다.

[0034] "염색질"은 세포 게놈을 포함하는 핵단백질 구조물이다. 세포 염색질은 핵산, 주로 DNA, 및 히스톤과 비-히스톤 염색체 단백질을 포함하는 단백질을 포함한다. 대부분의 진핵 세포 염색질은 뉴클레오솜의 형태로 존재하되, 뉴클레오솜 코어는 히스톤 H2A, H2B, H3 및 H4를 각각 2개씩 포함하는 옥타머와 결합된 대략 150개 염기쌍의 DNA를 포함하며; 링커 DNA(유기체에 따라 길이가 가변적)는 뉴클레오솜 코어 사이에서 연장된다. 히스톤 H1의 분자는 일반적으로 링커 DNA에 결합된다. 본 명세서의 목적을 위해, 용어 "염색질"은 모든 유형의 세포성 핵단백질(원핵생물과 진핵생물 모두)을 포함하는 것을 의미한다. 세포 염색질은 염색체 염색질과 에피솜 염색질을 둘다 포함한다.

[0035] "염색체"는 세포 게놈의 전체 또는 일부를 포함하는 염색질 복합체이다. 세포의 게놈은 종종 세포의 게놈을 포함하는 모든 염색체의 집합인, 그의 핵형(karyotype)을 특징으로 한다. 세포의 게놈은 하나 이상의 염색체를 포함할 수 있다.

[0036] "에피솜"은 세포의 염색체 핵형의 부분이 아닌 핵산을 포함하는 복제성 핵산, 핵단백질 복합체 또는 기타 구조

물이다. 에피솜의 예는 플라스미드 및 특정 바이러스 게놈을 포함한다.

[0037] "접근가능한 영역"은 표적 부위를 인식하는 외인성 분자에 의해 결합될 수 있는 핵산 중에서 표적 부위가 존재하는 세포 염색질 내 부위이다. 임의의 특정 이론에 구속되는 것을 바라지는 않지만, 접근가능한 영역은 뉴클레오솜 구조로 포장되지 않는 것으로 믿어진다. 접근가능한 영역의 별개의 구조는 종종 화학적 및 효소적 프로브, 예를 들어 뉴클레아제에 대한 그의 선택성에 의해 검출될 수 있다.

[0038] "표적 부위" 또는 "표적 서열"은 결합 분자가 결합할 핵산의 일부를 한정하는 핵산 서열이며, 단, 결합을 위한 충분 조건이 존재한다. 부분을 나타내는 핵산 서열이다. 예를 들어, 서열 5'-GAATTC-3'은 Eco RI 제한 엔도뉴클레아제에 대한 표적 부위이다.

[0039] "모듈 서브사이트"는 더 큰 아연 핑거 DNA 결합 단백질 내 아연 핑거 모듈(예를 들어, 1, 2, 3 또는 그 이상의 아연 핑거)이 결합될 핵산의 일부를 한정하는 핵산 서열이며, 단, 결합에 충분한 조건이 존재한다.

[0040] "외인성" 분자는 보통은 세포에 존재하지 않지만 하나 이상의 유전학적, 생화학적 또는 기타 방법에 의해 세포 내로 도입될 수 있는 분자이다. "세포 내에서 보통 존재"는 세포의 특정 발달 단계 및 환경 조건에 대하여 결정된다. 따라서, 예를 들어, 단지 근육의 배아 발달 동안 존재하는 분자는 성인 근육 세포에 대해서는 외인성 분자이다. 유사하게, 열 충격에 의해 유발된 분자는 열 충격을 받지 않은 세포에 대해서는 외인성 분자이다. 외인성 분자는 예를 들어, 기능 부전 내인성 분자의 기능적 형태 또는 정상적으로 기능하는 내인성 분자의 기능 부전성 형태를 포함할 수 있다. 또한, 외인성 분자는 다른 종에서 보통 발견되는 분자, 예를 들어, 동물의 게놈 내에 도입된 인간 서열일 수 있다.

[0041] 특히, 외인성 분자는 소분자, 예컨대, 조합 화학 공정에 의해 만들어진 소분자, 또는 거대 분자, 예컨대, 단백질, 핵산, 탄수화물, 지질, 당단백질, 지단백질, 다당류, 상기 분자들의 임의 변형 유도체, 또는 상기 분자들 중 하나 이상을 포함하는 임의의 복합체일 수 있다. 핵산은 DNA 및 RNA를 포함하며, 이는 단일- 또는 이중 가닥일 수 있고, 선형, 분지형 또는 고리형일 수 있으며, 임의의 길이를 가질 수 있다. 핵산은 듀플렉스를 형성할 수 있는 것뿐만 아니라 트리플렉스(triplex)-형성 핵산을 포함한다. 이에 대해서는, 예를 들어, 미국특허 제5,176,996호 및 제5,422,251호를 참조할 수 있다. 단백질은, 이에 제한되는 것은 아니지만, DNA-결합 단백질, 전사 인자, 염색질 리모델링 인자, 메틸화 DNA 결합 단백질, 폴리머라제, 메틸라제, 데메틸라제, 아세틸라제, 데아세틸라제, 키나제, 포스파타아제, 인테그라제, 리콤비나제, 리가제, 토포이소머라제, 자이라제 및 헬리카제를 포함한다.

[0042] 외인성 분자는 내인성 분자와 동일한 유형의 분자, 예를 들어, 외인성 단백질 또는 핵산일 수 있다. 예를 들어, 외인성 핵산은 감염성 바이러스 게놈, 세포 내로 도입된 플라스미드 또는 에피솜, 또는 세포 내에 보통은 존재하지 않는 염색체를 포함할 수 있다. 외인성 분자를 세포 내로 도입하는 방법은 당업자에게 공지되어 있는데, 이에 제한되는 것은 아니지만, 지질 매개 전달(즉, 중성 및 양이온성 지질을 포함하는 리포솜), 전기천공, 직접 주입, 세포 융합, 유전자총(particle bombardment), 인산칼슘 동시침전, DEAE-덱스트란 매개 전달 및 바이러스 벡터 매개형 전달을 포함한다.

[0043] 대조적으로, "내인성" 분자는 특정 환경 조건 하의 특정 성장 단계에서 특정 세포 내에 정상적으로 존재하는 분자이다. 예를 들어, 내인성 핵산은 염색체, 미토콘드리아, 엽록체 또는 다른 세포 소기관의 게놈, 또는 자연적으로 발생하는 에피솜 핵산을 포함할 수 있다. 추가의 내인성 분자는 단백질, 예를 들어 전사 인자 및 효소를 포함할 수 있다.

[0044] "융합" 분자는 2개 이상의 소단위 분자가, 바람직하게는 공유적으로 연결된 분자이다. 소단위 분자는 동일한 화학적 유형의 분자이거나, 또는 상이한 화학적 유형의 분자일 수 있다. 첫 번째 유형의 융합 분자의 예는, 이에 제한되는 것은 아니지만, 융합 단백질(예를 들어, ZFP DNA-결합 도메인과 절단 도메인 사이의 융합체) 및 융합 핵산(예를 들어, 상기에서 기술된 융합 단백질을 암호화하는 핵산)을 포함한다. 두 번째 유형의 융합 분자의 예는, 이에 제한되는 것은 아니지만, 트리플렉스-형성 핵산과 폴리펩타이드 사이의 융합체, 및 좁은 그루브(minor groove) 결합제와 핵산 사이의 융합체를 포함한다.

[0045] 세포 내에서 융합 단백질의 발현은, 해당 융합 단백질을 해당 세포에 전달함으로써, 또는 이러한 융합 단백질을 암호화하는 폴리뉴클레오타이드를 세포에 전달함으로써 생길 수 있되, 상기 폴리뉴클레오타이드는 전사되고, 해당 전사체가 번역되어 융합 단백질이 생성된다. 트랜스-스플라이싱(trans-splicing), 폴리펩타이드 절단 및 폴리펩타이드 결찰이 또한 세포 내 단백질의 발현에 수반될 수 있다. 폴리뉴클레오타이드 및 폴리펩타이드를 세포에 전달하는 방법은 본 명세서의 다른 곳에서 제시된다.

[0046] 본 명세서의 목적을 위해, "유전자"는 유전자 산물(하기 참조)을 암호화하는 DNA 영역뿐만 아니라, 유전자 산물의 생성을 조절하는 모든 DNA 영역을, 이러한 조절 서열이 암호화 서열 및/또는 전사된 서열에 인접하는지 여부에 상관없이 포함한다. 따라서, 유전자는, 이에 반드시 제한되는 것은 아니지만, 프로모터 서열, 종결자, 번역 조절 서열, 예컨대, 리보솜 결합 부위 및 내부 리보솜 도입 부위, 인핸서, 사일런서(silencer), 격리자(insulator), 경계 성분, 복제 원점, 매트릭스 부착 부위 및 유전자 좌위 제어 영역이 포함된다.

[0047] "유전자 발현"은 유전자에 포함된 정보가 유전자 산물로 전환되는 것을 말한다. 유전자 산물은 유전자의 직접 전사 산물(예를 들어, mRNA, tRNA, rRNA, 안티센스 RNA, 리보자임, 구조 RNA 또는 다른 임의 유형의 RNA), 또는 mRNA의 번역에 의해 생성된 단백질일 수 있다. 유전자 산물은 또한 캡핑, 폴리아데닐화, 메틸화 및 편집과 같은 과정에 의해 변형된 RNA, 및 예를 들어 메틸화, 아세틸화, 인산화, 유비퀴틴화, ADP-리보실화, 미리스틸화 및 글라이코실화에 의해 변형된 단백질을 포함한다.

[0048] 유전자 발현의 "조절"은 유전자 활성의 변화를 지칭한다. 발현의 조절은, 이에 제한되는 것은 아니지만, 유전자 활성화 및 유전자 억제를 포함할 수 있다. 게놈 편집(예를 들어, 절단, 변경, 불활성화, 무작위 돌연변이 유발)이 발현을 조절하는데 사용될 수 있다. 유전자 불활성화는 본 명세서에서 설명되는 것과 같은 ZFP를 포함하지 않는 세포에 비하여 유전자 발현이 조금이라도 감소된 것을 말한다. 따라서, 유전자 불활성화는 부분적일 수도 있고 또는 완전할 수도 있다.

[0049] "진핵" 세포는, 이에 제한되는 것은 아니지만, 진균 세포(예컨대 효모), 식물 세포, 동물 세포, 포유류 세포 및 인간 세포(예를 들어 T 세포)를 포함한다.

[0050] "관심 영역"은 외인성 분자에 결합하는 것이 바람직한 세포 염색질의 임의의 영역, 예를 들어, 유전자 또는 유전자 내부 혹은 유전자에 인접한 비암호화 서열이다. 결합은 표적화된 DNA 절단 및/또는 표적화된 재조합을 목적으로 할 수 있다. 관심 영역은, 예를 들어 염색체, 에피솜, 세포 소기관 게놈(예를 들어, 미토콘드리아, 엽록체), 또는 감염성 바이러스 게놈 내에 존재할 수 있다. 관심의 영역은 유전자의 암호화 영역 내에, 전사된 비암호화 영역, 예를 들면, 리더 서열, 트레일러 서열 또는 인트론 내에, 또는 암호화 영역의 상류 또는 하류의 비전사 영역 내에 존재할 수 있다. 관심 영역은 그 길이가 단일 뉴클레오타이드 쌍만큼 짧거나, 또는 20,000개까지의 뉴클레오타이드 쌍이거나, 또는 임의 정수 값의 뉴클레오타이드 쌍일 수 있다.

[0051] 용어 "작동가능한 연결" 및 "작동가능하게 연결된" (또는 "작동할 수 있도록 연결된")은 2개 이상의 성분(예를 들어, 서열 구성요소)의 병렬 배치에 관하여 상호호환적으로 사용되며, 여기서 해당 성분은 성분들이 둘다 정상적으로 작용하고 적어도 하나의 성분이 적어도 하나의 다른 성분에 의해 발휘되는 기능을 매개할 수 있도록 배열된다. 예로써, 전사 조절 서열이 하나 이상의 전사 조절 인자의 존재 또는 부재에 응하여 암호화 서열의 전사 수준을 제어한다면, 프로모터와 같은 전사 조절 서열은 암호화 서열에 작동가능하게 연결된다. 전사 조절 서열은 일반적으로 암호화 서열과 시스(cis) 형태로 작동가능하게 연결되지만, 그에 직접적으로 인접할 필요는 없다. 예를 들어, 인핸서는 암호화 서열과 근접하지 않는 경우에도, 암호화 서열에 작동가능하게 연결되는 전사 조절 서열이다.

[0052] 융합 폴리펩타이드에 관하여, 용어 "작동가능하게 연결된"은, 각 성분이 서로 다른 성분에 연결되어 이렇게 연결되지 않았다면 수행할 기능과 동일한 기능을 수행한다는 사실을 의미할 수 있다. 예를 들어, ZFP DNA-결합 도메인이 절단 도메인에 융합된 융합 폴리펩타이드에 관하여, 융합 폴리펩타이드 내에서 ZFP DNA-결합 도메인 일부가 이의 표적 부위 및/또는 이의 결합 부위에 결합할 수 있는 한편 절단 도메인이 표적 부위의 근처에서 DNA를 절단할 수 있다고 한다면, ZFP DNA-결합 도메인과 절단 도메인은 작동가능하게 연결되어 있다.

[0053] 단백질, 폴리펩타이드 또는 핵산의 "기능적 단편"은 서열이 전장 단백질, 폴리펩타이드 또는 핵산과 동일하지는 않지만 전장 단백질, 폴리펩타이드 또는 핵산과 동일한 기능을 보유하는 단백질, 폴리펩타이드 또는 핵산을 말한다. 기능적 단편은 그에 상응하는 천연 분자보다 많거나, 적거나 또는 그와 동일한 수의 잔기를 보유할 수 있고 및/또는 하나 이상의 아미노산 또는 뉴클레오타이드 치환을 함유할 수 있다. 핵산의 기능(예를 들어, 암호화 기능, 다른 핵산에 혼성화하는 능력)을 측정하는 방법은 당업계에 공지되어 있다. 유사하게, 단백질의 기능을 결정하는 방법도 잘 알려져 있다. 예를 들어, 폴리펩타이드의 DNA-결합 기능은 예컨대, 필터-결합, 전기영동 이동도-변화, 또는 면역침전 분석법에 의해 측정할 수 있다. DNA 절단은 겔 전기영동으로 분석할 수 있다. 이에 대해서는, 문헌[Ausubel et al., 상기와 동일]을 참조할 수 있다. 다른 단백질과 상호작용하는 한 단백질의 능력은 예를 들어, 공동-면역침전, 2-혼성 분석법 또는 상보성 분석법, 유전학적 및 생화학적 둘 다로 측정할 수 있다. 이에 대해서는, 예를 들어, 문헌[Fields et al., (1989) Nature 340:245-246]; 미국특허 제5,585,245호;

및 PCT 제WO 98/44350호를 참조할 수 있다.

[0054] **링커**

[0055] 본 명세서에서 서로에 대해 DNA-결합 모듈(예를 들어, 아연 핑거 모듈)을 융합하는(연결하는) 아미노산 서열이 설명된다. 본 명세서에서 설명된 링커를 사용하여 융합된 아연 핑거 모듈은 1, 2, 3, 4 또는 훨씬 더 많은 아연 핑거를 함유할 수 있다. 특정 실시형태에서, 아연 핑거 모듈은 1, 2, 또는 3개의 아연 핑거를 함유하는데, 이는 함께 연결될 때 3개 이상의 핑거 아연 핑거 단백질을 형성한다.

[0056] 본 명세서에서 설명된 링커 서열은 아연 핑거 내 α-나선의 마지막 잔기와 다음 아연 핑거 내 β-시트의 첫 번째 잔기 사이에서 연장된다. 따라서 링커 서열은 2개의 아연 핑거와 함께 접합한다. 전형적으로 아연 핑거 내 마지막 (C-말단) 아미노산은 C-말단의 아연 배위 잔기인 반면, 방향족 잔기(예를 들어, Phe)는 다음의 아연 핑거의 첫 번째 아미노산이다. 따라서 "야생형" 아연 핑거 내 트레오닌은 링커 내 제1 잔기이며, 프롤린은 링커의 마지막 잔기이다. 따라서, 예를 들어 Zif268에 대한 정규 링커 서열은 TG(E/Q)(K/R)P(서열번호 129)이다. 이에 대해서는, 예를 들어, 미국특허 제6,479,626호; 제6,903,185호 및 제7,153,949호를 참조할 수 있다.

[0057] 추가적인 링커는, 예를 들어 미국 특허 공개 제20030119023호에서 설명되며, 이는 다수의 글라이신 잔기(예를 들어, TGGGGSQKP(서열번호 2), TGGGGSGGSQKP(서열번호 14) 및 TGGGGSGGSGGSQKP(서열번호 15), TGGEKP(서열번호 16), TGGQKP(서열번호 17), TGGSGEKP(서열번호 18), TGGSGQKP(서열번호 19), TGGSGGSGEKP(서열번호 20), 및 TGGSGGSGQKP(서열번호 21)를 포함하는 링커를 설명한다.

[0058] 전형적으로, 링커는 링커 아미노산 서열을 통해 융합된 링커 및 링커 핵산 결합 모듈을 암호화하는 재조합 핵산을 사용하여 만들어진다. 링커는 또한 펩타이드 합성을 사용하여 만들어진 다음, 핵산 결합 모듈에 연결될 수 있다. 핵산 및 펩타이드 합성 방법의 조작 방법은 당업계에 알려져 있다(예를 들어, 문헌[Maniatis, et al., 1991. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press] 참조).

[0059] 본 명세서에서 설명된 링커는 앞서 설명한 링커보다 더 뻣뻣하며, 서브사이트가 특정 수의 염기쌍에 의해 분리될 때만 각 아연 핑거 모듈을 그의 표적 부위에 효율적으로 결합시킨다.

[0060] 따라서, 이전의 링커와 달리, 본 명세서에서 설명하는 링커는 적어도 하나의 내부 또는 N-말단 프롤린 잔기, 즉 링커의 C-말단에 있지 않은 프롤린 잔기를 포함한다. 본 명세서에서 설명하는 링커는 다음의 일반적인 아미노산 구조를 가진다:

[0061] $X^{N\text{-말단}}\text{-}X_n\text{-}X^{C\text{-말단}}$

[0062] 여기서 X는 임의의 아미노산 잔기이며, X_n은 적어도 3개의 아미노산 잔기이고 적어도 하나의 $X^{N\text{-말단}}$을 포함하며, X_n은 프롤린 잔기를 포함한다. 이러한 링커의 비제한적 예는 표 4, 5, 6, 9, 10, 11 또는 13에서 나타낸다. 더 나아가, 본 명세서에서 설명하는 링커는 또한 적어도 2개의 염기성 잔기, 예를 들어 아르기닌 잔기, 하나 이상의 히스티딘 잔기, 하나 이상의 리신 잔기 또는 이들의 조합을 포함한다.

[0063] 본 발명의 링커는 임의의 길이, 전형적으로 5개 이상의 아미노산 길이일 수 있다. 특정 실시형태에서, 링커는 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 훨씬 더 많은 아미노산 길이이다.

[0064] **DNA-결합 모듈**

[0065] 본 명세서에서 설명하는 링커 서열은 DNA-결합 모듈을 연결하기 위해 유리하게 사용된다.

[0066] 임의의 DNA-결합 도메인은 본 명세서에서 개시된 방법으로 사용될 수 있다. 특정 실시형태에서, DNA 결합 도메인은 아연 핑거 단백질을 포함한다. 바람직하게는, 아연 핑거 단백질은 비천연적으로 생기며, 즉 유전자조작되어 선택의 표적 부위에 결합한다. 이에 대해서는, 예를 들어 문헌[Beerli *et al.* (2002) *Nature Biotechnol.* 20:135-141; Pabo *et al.* (2001) *Ann. Rev. Biochem.* 70:313-340; Isalan *et al.* (2001) *Nature Biotechnol.* 19:656-660; Segal *et al.* (2001) *Curr. Opin. Biotechnol.* 12:632-637; Choo *et al.* (2000) *Curr. Opin. Struct. Biol.* 10:411-416]을 참조할 수 있다. 유전자조작된 아연 핑거 결합 도메인은 자연적으로 발생하는 아연 핑거 단백질과 비교하여 신규한 결합 특이성을 가질 수 있다. 유전자조작된 방법은, 이에 제한되는 것은 아니지만, 합리적인 설계 및 다양한 유형의 선택을 포함한다. 합리적인 설계는, 예를 들어 삼중선(또는 사중선) 뉴클레오타이드 서열 및 개개의 아연 핑거 아미노산 서열을 포함하는 데이터베이스를 사용하는 것을

포함하는데, 이때 각각의 삼중선 또는 사중선 뉴클레오타이드 서열은 특정 삼중선 또는 사중선 서열을 결합하는 아연 핑거의 하나 이상의 아미노산 서열과 관련된다. 이에 대해서는, 예를 들어, 본 명세서에 전문이 참조로 포함된 공동 소유된 미국 특허 제6,453,242호 및 제6,534,261호를 참조할 수 있다.

[0067] 파지 디스플레이 및 2혼성 시스템을 포함하는 예시적인 선택 방법은 미국 특허 제5,789,538호; 제5,925,523호; 제6,007,988호; 제6,013,453호; 제6,410,248호; 제6,140,466호; 제6,200,759호; 및 제6,242,568호뿐만 아니라 WO 98/37186호; WO 98/53057호; WO 00/27878호; WO 01/88197호 및 영국특허 제2,338,237호에 개시되어 있다. 추가로, 아연 핑거 결합 도메인의 결합 특이성의 증대는, 예를 들어 공동 소유된 WO 02/077227호에서 설명되었다.

[0068] 표적 부위의 선택; 융합 단백질(및 그것을 암호화하는 폴리뉴클레오타이드)의 설계 및 구성을 위한 ZFP 및 방법은 당업자에게 알려져 있고, 미국출원공개 제20050064474호 및 제20060188987호에서 상세하게 설명되며, 본 명세서에 그의 전문이 참조로 포함된다.

[0069] 게다가, 이들 및 다른 참고문헌에서 개시되는 바와 같이, 아연 핑거 도메인 및/또는 멀티 핑거 아연 핑거 단백질은, 예를 들어 5개 이상의 아미노산 길이의 링커를 포함하는 임의의 적합한 링커 서열을 사용하여 함께 연결될 수 있다. 또한 6개 이상의 아미노산 길이의 예시적인 링커 서열에 대해서는 미국특허 제6,479,626호; 제6,903,185호; 및 제7,153,949호를 참조할 수 있다. 본 명세서에 설명된 단백질은 단백질의 개개의 아연 핑거 사이의 적합한 링커의 임의의 조합을 포함할 수 있다.

[0070] 대안으로, DNA-결합 도메인은 뉴클레아제로부터 유래될 수 있다. 예를 들어, 호밍 엔도뉴크렐아제 및 메가뉴클레아제, 예컨대 I-*Sce*I, I-*Ceu*I, PI-*Psp*I, PI-*Sce*, I-*Sce*IV, I-*Csm*I, I-*Pan*I, I-*Sce*II, I-*Ppo*I, I-*Sce*III, I-*Cre*I, I-*Tev*I, I-*Tev*II 및 I-*Tev*III의 인식 서열이 알려져 있다. 또한 이에 대해서는, 미국특허 제5,420,032호; 제6,833,252호; 문헌[Belfort et al., (1997) *Nucleic Acids Res.* 25:3379-3388; Dujon et al., (1989) Gene 82:115-118; Perler et al., (1994) Nucleic Acids Res. 22, 1125-1127; Jasin (1996) Trends Genet. 12:224-228; Gimble et al., (1996) J. Mol. Biol. 263:163-180; Argast et al., (1998) J. Mol. Biol. 280:345-353] 및 뉴 잉글랜드 바이오랩스 카탈로그를 참조할 수 있다. 게다가, 호밍 엔도뉴크레아제 및 메가뉴클레아제의 DNA 결합 특이성은 유전자조작되어 비천연 표적 부위에 결합될 수 있다. 이에 대해서는, 예를 들어, 문헌[Chevalier *et al.* (2002) *Molec. Cell* 10:895-905; Epinat *et al.* (2003) *Nucleic Acids Res.* 31:2952-2962; Ashworth *et al.* (2006) Nature 441:656-659; Paques *et al.* (2007) *Current Gene Therapy* 7:49-66; 미국특허 공개 제20070117128호]을 참조할 수 있다.

[0071] 일부 실시형태에서, DNA 결합 도메인은 식물 병원체인 산토모나스 속 (Xanthomonas)에서 유래된 TAL 효과기의 유전자조작된 도메인이다(문헌[Miller *et al.* (2010) *Nature Biotechnology*, Dec 22[Epub ahead of print]; Boch *et al*, (2009) *Science* 29 Oct 2009 (10.1126/science.117881) 및 Moscou and Bogdanove, (2009) *Science* 29 Oct 2009 (10.1126/science.1178817)]; 또한 2010년 3월 17일 출원된 미국 가특허출원 제61/395,836호; 2010년 8월 21일 출원된 제61/409,421호; 2010년 10월 13일 출원된 제61/45,121호; 2010년 12월 20일 출원된 제61/459,891호; 2011년 2월 2일 출원된 출원 제61/462,482호; 2011년 3월 24일 출원된 출원 제61/465,869호를 참조하며, 이들의 개시는 본 명세서에 그의 전문이 참조로 포함되어 있다.

[0072] **조절 도메인**

[0073] 본 명세서에서 설명한 바와 같이 연결된 아연 핑거 모듈은 종종 융합 단백질로서 외인성 도메인(또는 그의 기능적 단편)과 함께 발현된다. ZFP에 첨가를 위한 보통의 조절 도메인은, 예를 들어 전사 인자 도메인(활성자, 억제자, 보조 활성자, 보조 억제자), 사일런서, 발암유전자(예를 들어, myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos 패밀리 멤버 등); DNA 복구 효소 및 그의 관련 인자 및 변경유전자; DNA 복구 효소 및 그의 관련 인자 및 변경유전자; DNA 재배열 효소 및 그것의 관련 인자 및 변경유전자; 염색질 관련 단백질 및 그의 변경유전자(예를 들어 키나제, 아세틸라제 및 데아세틸라제); 및 DNA 변형 효소(예를 들어, 메틸트랜스퍼라제, 토포아이소머라제, 헬리카제, 리가제, 키나제, 포스파타제, 폴리머라제, 엔도뉴클레아제) 및 그의 관련 인자 및 변경유전자를 포함한다.

[0074] 예를 들어 유전자의 발현을 억제하기 위해 사용되는 ZFP와 같은 DNA 결합 도메인을 융합하기 위한 예시적인 기능적 도메인은 인간 KOX-1 단백질로부터의 KRAB 억제 도메인이다(예를 들어, 문헌[Thiesen *et al.*, New Biologist 2, 363-374 (1990); Margolin *et al.*, *Proc. Natl. Acad. Sci.* USA 91, 4509-4513 (1994); Pengue *et al.*, Nucl. Acids Res. 22:2908-2914 (1994); Witzgall et al., Proc. *Natl. Acad. Sci.* USA 91, 4514-

4518 (1994)] 참조). 다른 적합한 억제 도메인은 메틸 결합 도메인 단백질 2B(MBD-2B)이다(또한 MBD 단백질의 설명에 대해서는 문헌[Hendrich et al. (1999) Mamm. Genome 10:906-912]을 참조할 수 있다). 다른 유용한 억제 도메인은 v-ErbA 단백질과 관련된 것이다. 이에 대해서는, 예를 들어, 문헌[Damm, et al. (1989) Nature 339:593-597; Evans (1989) Int. J. Cancer Suppl. 4:26-28; Pain et al. (1990) New Biol. 2:284-294; Sap et al. (1989) Nature 340:242-244; Zenke et al. (1988) Cell 52:107-119; 및 Zenke et al. (1990) Cell 61:1035-1049]을 참조할 수 있다.

[0075] 추가적인 예시적 억제 도메인은, 이에 제한되는 것은 아니지만, KRAB(또한 "KOX"로 언급됨), SID, MBD2, MBD3, DNMT 패밀리의 멤버(예를 들어, DNMT1, DNMT3A, DNMT3B), Rb, 및 MeCP2를 포함한다. 이에 대해서는, 예를 들어, 문헌[Bird *et al.* (1999) Cell 99:451-454; Tyler *et al.* (1999) Cell 99:443-446; Knoepfler *et al.* (1999) Cell 99:447-450; 및 Robertson *et al.* (2000) Nature Genet. 25:338-342]을 참조할 수 있다. 추가적인 예시적인 억제 도메인은, 이에 제한되는 것은 아니지만, ROM2 및 AtHD2A을 포함한다. 이에 대해서는, 예를 들어, 문헌[Chem et al. (1996) Plant Cell 8:305-321; 및 Wu et al. (2000) Plant J. 22:19-27]을 참조할 수 있다.

[0076] 활성화를 이루기 위한 적합한 도메인은 HSV VP16 활성화 도메인(예를 들어 문헌[Hagmann et al., *J. Virol.* 71, 5952-5962 (1997)] 참조) 핵 호르몬 수용체(예를 들어, 문헌[Torchia et al., *Curr. Opin. Cell. Biol.* 10:373-383 (1998)]); 핵 인자 카파 B의 p65 서브유닛(Bitko and Barik, *J. Virol.* 72:5610-5618 (1998) 및 Doyle and Hunt, Neuroreport 8:2937-2942 (1997)); Liu et *al., Cancer Gene Ther.* 5:3-28 (1998) 참조), 또는 인공 키메라 기능적 도메인, 예컨대 VP64(Seifpal et al., EMBO J. 11, 4961-4968 (1992))를 포함한다. 추가적인 예시적 활성화 도메인은, 이에 제한되는 것은 아니지만, VP16, VP64, p300, CBP, PCAF, SRC1 PvALF, AtHD2A 및 ERF-2를 포함한다. 이에 대해서는, 예를 들어, 문헌[Robyr et al. (2000) Mol. Endocrinol. 14:329-347; Collingwood et al. (1999) J. Mol. Endocrinol. 23:255-275; Leo et al. (2000) Gene 245:1-11; Manteuffel-Cymborowska (1999) Acta Biochim. Pol. 46:77-89; McKenna et al. (1999) *J. Steroid Biochem. Mol.* Biol. 69:3-12; Malik *et al.* (2000) *Trends Biochem. Sci.* 25:277-283; 및 Lemon et al. (1999) *Curr. Opin. Genet. Dev.* 9:499-504]을 참조할 수 있다. 추가적인 예시적 활성화 도메인은, 이에 제한되는 것은 아니지만, OsGAI, HALF-1, C1, AP1, ARF-5, -6, -7 및 -8, CPRF1, CPRF4, MYC-RP/GP, 및 TRAB1을 포함한다. 이에 대해서는, 예를 들어 문헌[Ogawa et al. (2000) Gene 245:21-29; Okanami et al. (1996) *Genes Cells* 1:87-99; Goff et al. (1991) *Genes Dev.* 5:298-309; Cho et al. (1999) Plant Mol. Biol. 40:419-429; Ulmason et al. (1999) *Proc. Natl. Acad. Sci.* USA 96:5844-5849; Sprenger-Haussels et al. (2000) Plant J. 22:1-8; Gong et al. (1999) Plant Mol. Biol. 41:33-44; 및 Hobo *et al.* (1999) Proc. Natl. Acad. Sci. USA 96:15, 348-15, 353]을 참조할 수 있다.

[0077] 특정 실시형태에서, 조절 도메인은 뉴클레아제(예를 들어 절단 도메인)를 포함한다. 이러한 유전자조작된 뉴클레아제는 표적 뉴클레오타이드 서열 내 이중 가닥 파손(double-strand break, DSB)을 만들기 위해 사용될 수 있는데, 이는 1000배 초과로 표적화된 좌위(표적화된 통합)에서 상동성 재조합을 통해 공여자 핵산 도입의 빈도를 증가시킨다. 게다가, 비상동성 말단 접합(NHEJ)에 의한 부위 특이적 DSB의 부정확한 복구는 또한 유전자 붕괴를 초래할 수 있다. 뉴클레아제는 예컨대 세포주 유전자조작뿐만 아니라 치료적 용도를 위한 매우 다양한 목적을 위해 사용될 수 있다.

[0078] 본 명세서에 개시된 융합 단백질의 절단 도메인은 임의의 엔도뉴클레아제 또는 엑소뉴클레아제로부터 얻어질 수 있다. 절단 도메인이 유래될 수 있는 예시적인 엔도뉴클레아제는, 이에 제한되는 것은 아니지만, 제한 엔도뉴클레아제 및 호밍 엔도뉴클레아제로부터 유래될 수 있다. 이에 대해서는, 예를 들어, 문헌[미국 메사추세츠주 비벌리에 소재한 뉴잉글랜드 바이오랩스 2002-2003 카탈로그; 및 Belfort *et al.* (1997) Nucleic Acids Res. 25:3379-3388]을 참조할 수 있다. DNA를 절단하는 추가적인 효소는 알려져 있다(예를 들어, S1 뉴클레아제; 녹두 뉴클레아제; 췌장 DNase I; 마이크로코칼 뉴클레아제; 효모 HO 엔도뉴클레아제; 또한 문헌[Linn *et al.* (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press,1993] 참조). 하나 이상의 이들 효소(또는 그의 기능적 단편)는 절단 도메인 및 절단 절반 도메인의 공급원으로 사용될 수 있다.

[0079] 유사하게, 절단 절반 도메인은 절단 활성을 위해 이합체화를 필요로하는 상기 설명한 바와 같은 임의의 뉴클레아제 또는 그의 일부로부터 유래될 수 있다. 일반적으로, 융합 단백질이 절단 절반 도메인을 포함한다면 2개의 융합 도메인은 절단을 필요로한다. 대안으로, 2개의 절단 절반 도메인을 포함하는 단일 단백질이 사용될 수 있다. 2개의 절단 절반 도메인은 동일한 엔도뉴클레아제(또는 그의 기능적 단편)으로부터 유래될 수 있으며, 각각

의 절단 절반 도메인은 상이한 엔도뉴클레아제(또는 그의 기능적 단편)로부터 유래될 수 있다.

[0080] 추가로, 2개의 융합 단백질에 대한 표적 부위는 서로에 대해 바람직하게 배치되며, 그의 각 표적 부위에 대해 2개의 융합 단백질의 결합은 절단 절반 도메인이, 예를 들어 이합체화에 의해 기능적 절단 도메인을 형성하도록 하는 서로에 대한 공간 배향으로 절단 절반 도메인이 위치한다. 따라서, 특정 실시형태에서, 표적 부위의 근처 가장자리는 5 내지 8개의 뉴클레오타이드 또는 15 내지 18개의 뉴클레오타이드에 의해 분리된다. 그러나 뉴클레오타이드 또는 뉴클레오타이드 쌍의 임의의 정수의 수는 2개의 표적 부위 사이에(2 내지 50개의 뉴클레오타이드 쌍 또는 그 이상) 있을 수 있다. 일반적으로 절단 부위는 표적 부위 사이에 놓여있다.

[0081] 제한 엔도뉴클레아제(제한 효소)는 다수의 종에 존재하며, (인식 부위에서) DNA에 서열 특이적 결합을 가능하게 하고, 결합 부위에서 또는 결합 부위 근처에서 DNA를 절단할 수 있다. 특정 제한 효소(예를 들어 IIS형)는 인식 부위로부터 제거된 부위에서 DNA를 절단하며, 분리가능한 결합 및 절단 도메인을 가진다. 예를 들어, IIS형 효소 *Fok* I은 하나의 가닥 상의 그의 인식 부위로부터 9개 뉴클레오타이드 및 나머지 상에서 그의 인식 부위로부터 13개의 뉴클레오타이드에서 DNA의 이중 가닥 절단을 촉매한다. 이에 대해서는, 예를 들어, 미국특허 제5,356,802호; 제5,436,150호 및 제5,487,994호뿐만 아니라 문헌[Li et al. (1992) *Proc. Natl. Acad. Sci.* USA 89:4275-4279; Li et al. (1993) Proc. Natl. Acad. Sci. USA 90:2764-2768; Kim et al. (1994a) Proc. Natl. Acad. Sci. USA 91:883-887; Kim et al. (1994b) J. Biol. Chem. 269:31, 978-31, 982]을 참조할 수 있다. 따라서, 한 실시형태에서, 융합 단백질은 적어도 하나의 IIS형 제한 효소 및 하나 이상의 아연 핑거 결합 도메인으로부터 절단 도메인(또는 절단 절반 도메인)을 포함하는데, 이는 유전자조작될 수도 있고 유전자조작되지 않을 수도 있다.

[0082] 절단 도메인이 결합 도메인으로부터 분리될 수 있는 예시적인 IIS형 제한 효소는 *Fok* I이다. 이 특정 효소는 이합체로서 활성이다. 문헌[Bitinaite et al. (1998) *Proc. Natl. Acad. Sci.* USA 95: 10,570-10,575]. 따라서, 본 명세서의 목적을 위해, 개시된 융합 단백질에서 사용된 *Fok* I 효소의 일부는 절단 절반 도메인으로 생각된다. 따라서, 아연 핑거 *Fok* I 융합을 사용하여 세포 서열의 표적화된 이중 가닥 절단 및/또는 표적화된 치환을 위하여, 2개의 융합 단백질은, 각각 *Fok* I 절단 절반 도메인을 포함하며 재구성되어 촉매적으로 활성인 절단 도메인을 재구성하기 위해 사용될 수 있다. 대안으로, 아연 핑거 결합 도메인 및 2개의 *Fok* I 절단 절반 도메인을 함유하는 단일 폴리펩타이드 분자가 또한 사용될 수 있다. 아연 핑거 *Fok* I 융합을 사용하여 표적화된 절단 및 표적화된 서열 변경을 위한 변수는 본 명세서의 다른 곳에서 제공된다.

[0083] 절단 도메인 또는 절단 절반 도메인은 절단 활성을 보유하는 단백질의 임의의 일부일 수 있고, 또는 다합체화(예를 들어 이합체화)하기 위한 능력을 보유하여 기능적 절단 보메인을 형성한다.

[0084] 예시적인 IIS 제한 효소는 국제특허 공개 WO 07/014275에서 설명되며, 본 명세서에 그의 전문이 참조로 포함된다. 또한 추가적인 제한 효소는 분리가능한 결합 및 절단 도메인을 함유하며, 이들은 본 명세서에 의해 고려된다. 이에 대해서는, 예를 들어, 문헌[Roberts *et al.* (2003) Nucleic Acids Res. 31:418-420]을 참조할 수 있다.

[0085] 특정 실시형태에서, 절단 도메인은, 예를 들어 모든 목적을 위하여 본 명세서에 전문이 참조로 포함된 미국특허 공개 제20050064474호 및 제20060188987호 및 미국축원 제11/805,850호(2007년 5월 23일 출원됨)에서 설명한 바와 같은 동질이합체화를 최소화하거나 방지하는 하나 이상의 유전자조작된 절단 절반 도메인(또한 이합체화 도메인 돌연변이체로 언급됨)을 포함한다. *Fok* I의 위치 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, 및 538에서 아미노산 잔기는 *Fok* I 절단 절반 도메인의 이합체화에 영향을 미치기 위한 모든 표적이다.

[0086] 절대 이종이합체를 형성하는 *Fok* I의 예시적인 유전자조작된 절단 절반 도메인은 *Fok* I의 위치 490 및 538에서 아미노산 잔기에서 돌연변이를 포함하며, 제2 절단 절반 도메인은 아미노산 잔기 486 및 499에서 돌연변이를 포함한다.

[0087] 따라서, 한 실시형태에서, 490번에서 돌연변이는 Glu(E)을 Lys(K)로 치환하고; 538에서 돌연변이는 Iso(I)을 Lys(K)으로 치환하며; 486번에서 돌연변이는 Gln(Q)을 Glu(E)으로 치환하고; 499번에서 돌연변이는 Iso(I)을 Lys(K)으로 치환한다. 구체적으로, 본 명세서에 설명된 유전자조작된 절단 절반 도메인은, 한쪽의 절단 절반 도메인에서는 490번 (E→K)과 538번 (I→K) 위치에 돌연변이를 유발시켜 "E490K:I538K"로 명명된 유전자조작된 절단 절반 도메인을 생성하고, 다른 한쪽의 절단 절반 도메인에서는 위치 486번(Q→E)과 499번(I→L) 위치에 돌연변이를 유발시켜 "Q486E:I499L"로 명명된 유전자조작된 절단 절반 도메인을 생성시켜 제조하였다. 본 명세서에

설명된 유전자조작된 절단 절반 도메인은 비정상적인 절단이 최소화되거나 제거되는 절대 이종이합체 돌연변이체이다. 이에 대해서는, 예를 들어, 제WO 07/139898호의 실시예 1을 참조할 수 있다. 특정 실시형태에서, 유전자조작된 절단 절반 도메인은 위치 486, 499 및 496(야생형 *Fok*I에 대해 넘버링됨)에서, 예를 들어 위치 486에서 야생형 Gln(Q) 잔기가 Glu(E) 잔기로, 위치 499에서 야생형 Iso(I) 잔기가 Leu(L) 잔기로, 위치 496에서 야생형 Asn(N) 잔기가 Asp(D) 또는 Glu(E) 잔기로 치환되는(또한 각각 "ELD" 및 "ELE"도메인으로 언급됨) 돌연변이를 포함한다. 다른 실시형태에서, 유전자조작된 절단 절반 도메인은 위치 490, 538 및 537(야생형 *Fok*I에 대해 넘버링됨)에서, 예를 들어 위치 490에서 야생형 Glu(E) 잔기가 Lys(K) 잔기로, 위치 538에서 야생형 Iso(I) 잔기가 Lys(K) 잔기로, 위치 537에서 야생형 His(H) 잔기가 Lys(K) 또는 Arg(R) 잔기로 치환되는(또한 각각 "KKK" 및 "KKR" 도메인으로 언급됨) 돌연변이를 포함한다. 다른 실시형태에서, 유전자조작된 절단 절반 도메인은 위치 490 및 537(야생형 *Fok*I에 대해 넘버링됨)에서, 예를 들어 위치 490에서 야생형 Glu(E) 잔기가 Lys(K) 잔기로, 위치 537에서 야생형 His(H) 잔기가 Lys(K) 잔기 또는 Arg(R) 잔기로 치환되는(또한 각각 "KIK" 및 "KIR"도메인으로 언급됨) 돌연변이를 포함한다(미국출원 제12/931,660호 참조).

[0088] 본 명세서에 설명된 유전자조작된 절단 절반 도메인은 임의의 적합한 방법, 예를 들어 미국특허 공개 제20050064474호(예를 들어, 실시예 5 참조); 및 WO 07/139898에서 설명된 바와 같은 야생형 절단 절반 도메인(Fok I)의 부위 지정 돌연변이를 사용하여 제조될 수 있다.

[0089] 대안으로, 뉴클레아제는 소위 "분할 효소(split-enzyme)" 기술을 사용하여 핵산 표적 부위에서 생체내에 조립될 수 있다(예를 들어, 미국특허 공개 제20090068164호 참조). 이러한 분할 효소의 성분은 개별 발현 구성체 상에서 발현될 수 있고, 또는 예를 들어, 자기 절단 2A 펩타이드 또는 IRES 서열에 의해 개개의 성분이 분리되는 하나의 오픈 리딩 프레임 내에 연결될 수 있다. 성분은 개개의 아연 핑거 결합 도메인 또는 메가뉴클레아제 핵산 결합 도메인의 도메인일 수 있다.

[0090] 융합 분자는 당업자에게 잘 알려진 클로닝 및 생화학적 컨쥬게이션 방법에 의해 구성된다. 융합 분자는 DNA 결합 도메인 및 기능적 도메인(예를 들어, 전사 활성화 또는 억제 도메인)을 포함한다. 융합 분자는 또한 (예컨대 SV40 매개 T-항원으로부터의)핵 이행신호(nuclear localization signal) 및 에피토프 태그(예컨대 FLAG 및 혈구응집소)를 선택적으로 포함한다. 융합 단백질(및 그것을 암호화하는 핵산)은 번역 리딩 프레임이 융합의 성분 중에서 보존되도록 설계된다.

[0091] 이러한 적용을 위해, 융합 분자는 당업자에게 알려진 것과 같은 약제학적으로 허용가능한 담체와 함께 조제된다. 이에 대해서는, 예를 들어, 문헌[Remington's Pharmaceutical Sciences, 17th ed., 1985]; 및 공동소유된 WO 00/42219호를 참조할 수 있다.

[0092] **키트**

[0093] 또한 본 명세서에서 설명한 링커 중 어느 하나를 포함하며 및/또는 상기 설명한 방법 중 어느 하나를 수행하기 위한 키트가 제공된다. 키트는 본 명세서에서 설명하는 링커 서열(또는 본 명세서에서 설명하는 링커를 암호화하는 폴리뉴클레오타이드)을 전형적으로 함유한다. 키트는 링커를 단독으로 공급할 수 있고 또는 DNA 결합 도메인 및/또는 선택의 뉴클레아제가 용이하게 삽입될 수 있는 벡터를 제공할 수 있다. 키트는 또한 세포, 세포의 형질전환을 위한 완충제, 세포에 대한 배양물 배지, 및/또는 분석을 수행하기 위한 완충제를 함유할 수 있다. 전형적으로, 키트는 다른 성분의 키트에 부착되거나 또는 달리 다른 성분의 키트를 수반하는 설명서, 포장 또는 광고성 리플렛과 같은 임의의 재료를 포함하는 라벨을 함유한다.

[0094] **적용**

[0095] 개시된 링커는 유전자조작된 아연 핑거 단백질에 대한 표적 부위의 레퍼토리를 향상시키기 위해 유리하게 사용된다. 예를 들어, 본 명세서에 설명된 링커는 모듈 서브사이트가 인접하지 않았을 때 원하는 표적 부위에 결합을 용이하게 한다. 따라서, 주어진 레퍼토리 크기에 대해 주어진 핵산 서열을 표적화하도록 구성될 수 있었던 ZFP의 수가 효과적으로 증가된다. 더 나아가, 설명된 링커가 다양한 모듈 서브사이트 분리(예를 들어, 0, 1 및 2개의 염기쌍 갭) 사이를 구별하기 때문에, 그들은 부적절한 표적 부위에 ZFP의 결합을 감소시킨다. 예를 들어, 2개의 염기쌍을 스킵하도록 설계된 가요성 링커가 있는 ZFP(예를 들어, TGGGGSGGSQKP(서열번호 14))는 모듈 서브사이트 사이에 0, 1, 또는 2개의 염기쌍 중 하나와 함께 표적 부위에 결합할 수 있다. 본 명세서에서 설명한 것과 같은 2bp-스킵 링커를 가지는 이 동일한 ZFP는 모듈 서브사이트 사이에 2개 염기쌍을 가지는 표적에 잘 결합해야 하지만, 모듈 서브사이트 사이에 0 또는 1개의 염기쌍을 가지는 표적(부적절한 또는 의도되지 않은 표적 부위)에 효율적으로 결합할 수 없어야 한다.

[0096] 따라서, 본 명세서에서 설명한 링커는, 이에 제한되는 것은 아니지만, 유전자 발현의 조절을 위한 아연 핑거 전사 인자(ZFP-TF) 및/또는 절단을 위한 아연-핑거 뉴클레아제(ZFN) 내에서 포함하여, 아연 핑거 단백질이 현재 사용되기 위한 임의의 적용에 사용될 수 있다. 이에 대해서는, 예를 들어, 미국특허 제6,534,261호; 제6,599,692호; 제6,689,558호; 제7,067,317호; 제7,262,054호 및 제7,253,273호; 미국특허 공개 제20050064474호; 제2006/0063231호; 제2007/0134796호; 제2007/0218528호; 제2008/0015164호; 제2008/0188000호; 제2008/0299580호 및 제2008/0159996호를 참조할 수 있으며, 이들의 전문은 본 명세서에 참조로 포함된다.

[0097] 따라서, 개시된 링커는 특이적으로 표적화된 조절 또는 절단이 바람직한 임의의 방법에 대해 임의의 ZFP 또는 ZFN으로 사용될 수 있다. 예를 들어, ZFP-TF 및 ZFN은 유전적 질병, 감염(바이러스 또는 박테리아)을 치료하기 위해, 상동 재조합 및/또는 녹-인(knocked-in) 또는 녹아웃(knocked-out)에 의해 원하는 유전자가 활성화되고, 억제되거나 표적화되는 세포주, 동물 및 또는 식물을 만들기 위해 사용될 수 있다. 따라서, 본 명세서에서 설명되는 링커는 또한 DNA를 더 효율적으로 클로닝하기 위해 사용될 수 있고 게놈 내 변형은 ZFN에 의해 용이하게 되는데, 이는 생명공학 및 기초 과학의 많은 영역에서 널리 적용가능하다.

[0098] 실시예

[0099] **실시예 1: 링커의 선택**

[0100] 각각 4개의 핑거를 함유한 3가지의 상이한 숙주 ZFP: "ZFP7263", "ZFP7264" 및 "ZFP8196"(7263 및 7264에 대해 미국특허 공개 제20050064474호 및 8196에 대해 제20080159996호)의 맥락에서 링커 선택을 수행하였다. 각 숙주 ZFP에 대한 각 핑거의 인식 나선을 표 1에 제공하는 한편, 각 숙주 ZFP의 전체 서열을 도 1a에서 제공한다. 선택을 다음과 같이 수행하였다: (i) 우선, 중심 링커 내 코돈을 2 내지 12개의 완전히 무작위화된 코돈의 혼합으로 치환한 각 숙주 ZFP 내에서 라이브러리를 만들었고(도 1b); 나이브(naive) 라이브러리의 서열은 1회 이상 클론을 나타내지 않은 서열의 양호한 다양성을 나타내었으며; (ii) 다음으로, 섬질 박테리오파지의 표면상에서 라이브러리를 발현시켰고; (iii) 다음에 숙주 단백질 결합 부위의 중심에서(즉, 무작위화된 링커에 의해 걸쳐진 영역 내) 1- 또는 2-bp 삽입을 함유한 바이오틴화된 표적 변이체에 결합을 위해 파지 발현된 ZFP 라이브러리를 선택하였다(표 2A). 각각의 삽입은 효율적인 결합을 가능하게 하도록 더 긴 링커에 의해 브릿지되어야 하는 숙주 단백질의 제2와 제3 핑거에 대한 결합 서열 사이의 갭을 포함하였다(표 2A). 삽입은 본질적 염기 특이성이 없는 링커의 선택을 유리하게 하기 위한 염기의 혼합물로 이루어졌다. 5개의 선택 주기를 수행하였다. 최종 4주기 동안, 바이오틴화되지 않고 비표적화된 갭 길이를 함유한 1000배 몰 과량의 결합 부위로 역선택을 사용하였다(즉, 파지가 1bp 갭 길이를 가지는 표적 서열을 사용하여 선택된다면, 역선택은 0, 2, 3 및 4 bp 갭을 가지는 표적을 포함하였고; 파지가 2bp 갭 길이를 가지는 표적을 사용하여 선택되었다면, 역선택은 0, 1 및 3 및 4 bp 갭을 가지는 표적을 포함하였다. 표 2B 참조).

[0101] 제5 라운드의 선택으로부터의 파지 풀을 표적화된 갭 길이를 함유하는 서열을 선택적으로 결합하기 위한 능력을 위해 스크리닝하였고, 이 연구들은 갭 선택적인 결합을 나타내었다(도 2). 특히 ZFP8196의 맥락에서 1 bp 갭을 스킵하도록 선택한 파지 풀은 갭이 없을 때와 비교하여 1bp 갭을 함유하는 표적에 대해 25배 선호도를 나타내었다. ZFP7263의 맥락에서 1 bp 갭을 스킵하도록 선택된 파지 풀은 갭이 없을 때와 비교하여 1bp 갭을 함유하는 표적에 대해 26배 선호도를 나타내었다. ZFP7264의 맥락에서 1 bp 갭을 스킵하도록 선택된 파지 풀은 갭이 없을 때와 비교하여 1bp 갭을 함유하는 5.5배 선호도 표적을 나타내었다. 각각의 이들 풀은 또한 더 긴 갭 길이(2, 3 또는 4 bp)를 표적화하기 위해 결합을 거의 또는 전혀 나타내지 않았다.

[0102] ZFP8196의 맥락에서 2 bp 갭을 스킵하도록 선택한 파지 풀은 1bp 갭과 비교하여 2 bp 갭을 함유하는 표적에 대해 7배 선호도뿐만 아니라 0, 3 및 4 bp 갭을 함유하는 표적에 대해 30배 선호도를 나타내었다.

표 1

[0103]

숙주 ZFP 인식 나선				
ZFP	핑거 1	핑거 2	핑거 3	핑거 4
8196	RSDNLSV (서열번호22)	QKINLQV (서열번호23)	RSDVLSE (서열번호24)	QRNHRTT (서열번호25
7264	RSDTLSE (서열번호26)	ARSTRTT (서열번호27)	RSDSLSK (서열번호28)	QRSNLKV (서열번호29)
7263	RSDNLSV (서열번호22)	RNAHRIN (서열번호30)	RSDTLSE (서열번호26)	ARSTRTN (서열번호31)

[0104] (표 2A)

선택을 위해 사용한 표적 부위	
ZFP w/무작위화된 링커	**표적 부위**
8196	ATAAACTGdCAAAAGGC (서열번호 33) ATAAACTGdbCAAAAGGC (서열번호 34)
7264	TTAAAGCGhGCTCCGAA (서열번호 38) TTAAAGCGhdGCTCCGAA (서열번호 39)
7263	CCACTCTGhTGGAAGTG (서열번호 43) CCACTCTGhhTGGAAGTG (서열번호 44)

표 2A. 선택을 위해 사용한 표적 부위. 파지 연구에서 사용한 듀플렉스 DNA 표적 부위는 다음의 일반 형태를 가졌으며: TATAAT$(X)_{17-18}$TTCACAGTCAGTCCACACGTC, (서열번호 67), 여기서 $(X)_{17-18}$은 표에서 열거한 서열로 치환되었다. DNA 듀플렉스는 이탤릭체 서열로 어닐링된 프라이머를 연장함으로써 만들어졌고 그것의 5'말단에서 바이오틴화되었다. 밑줄친 염기는 각 숙주 ZFP의 4개의 핑거에 대해 결합 서열을 표시하는 한편, 소문자는 선택된 링커에 의해 포괄되어야 하는 삽입된 뉴클레오타이드(또는 "갭" 염기)를 표시한다. 갭 염기에 대한 축퇴 코드는 다음과 같다: "d"는 A, G, 및 T의 혼합을 의미하며; "b"는 C, G, 및 T의 혼합을 의미하고; "h"는 A, C, 및 T의 혼합을 의미하며; "v"는 A, C, 및 G의 혼합을 의미한다.

[0105]

[0106] (표 2B)

선택동안 사용한 경쟁자 부위	
ZFP w/무작위화된 링커	**경쟁자 부위**
8196	ATAAACTGCAAAAGGC(서열번호 32) ATAAACTGdCAAAAGGC(서열번호 33) ATAAACTGdbCAAAAGGC(서열번호 34) ATAAACTGdbbCAAAAGGC(서열번호 35) ATAAACTGdbbbCAAAAGGC(서열번호 36)
7264	TTAAAGCGGCTCCGAA(서열번호 37) TTAAAGCGhGCTCCGAA(서열번호 38) TTAAAGCGhdGCTCCGAA(서열번호 39) TTAAAGCGhdvGCTCCGAA(서열번호 40) TTAAAGCGhdvdGCTCCGAA(서열번호 41)
7263	CCACTCTGTGGAAGTG(서열번호 42) CCACTCTGhTGGAAGTG(서열번호 43) CCACTCTGhhTGGAAGTG(서열번호 44) CCACTCTGhhhTGGAAGTG(서열번호 45) CCACTCTGhhhbTGGAAGTG(서열번호 46)

표 2B. 선택동안 사용한 경쟁자 부위. 듀플렉스 DNA 경쟁자 부위는 다음의 일반 형태를 가졌으며: TATAAT$(X)_{16-20}$TTCACAGTCAGTCCACACGTC, (서열번호 67), 여기서 $(X)_{16-20}$은 표에서 열거한 서열로 치환되었다. DNA 듀플렉스는 이탤릭체 서열로 어닐링된 (비-바이오틴화된) 프라이머를 연장함으로써 만들어졌다. 밑줄친 염기는 각 숙주 ZFP의 4개의 핑거에 대해 결합 서열을 표시하는 한편, 소문자는 선택된 링커에 의해 포괄되어야 하는 삽입된 뉴클레오타이드(또는 "갭" 염기)를 표시한다. 갭 염기에 대한 축퇴 코드는 다음과 같다: "d"는 A, G, 및 T의 혼합을 의미하며; "b"는 C, G, 및 T의 혼합을 의미하고; "h"는 A, C, 및 T의 혼합을 의미하며; "v"는 A, C, 및 G의 혼합을 의미한다.

[0107]

[0108] (표 2C)

파지 풀 갭 선택성 연구를 위해 사용한 표적		
ZFP w/무작위화된 링커	**갭**	**표적 부위**
8196	0 갭	ATAAACTGCAAAAGGC(서열번호 32)
	1 갭	ATAAACTGdCAAAAGGC(서열번호 33)
	2 갭	ATAAACTGdbCAAAAGGC(서열번호 34)
	3 갭	ATAAACTGdbbCAAAAGGC(서열번호 35)
	4 갭	ATAAACTGdbbbCAAAAGGC(서열번호 36)
7264	0 갭	TTAAAGCGGCTCCGAA(서열번호 37)
	1 갭	TTAAAGCGhGCTCCGAA(서열번호 38)
	2 갭	TTAAAGCGhdGCTCCGAA(서열번호 39)
	3 갭	TTAAAGCGhdvGCTCCGAA(서열번호 40)
	4 갭	TTAAAGCGhdvdGCTCCGAA(서열번호 41)
7263	0 갭	CCACTCTGTGGAAGTG(서열번호 42)
	1 갭	CCACTCTGhTGGAAGTG(서열번호 43)
	2 갭	CCACTCTGhhTGGAAGTG(서열번호 44)
	3 갭	CCACTCTGhhhTGGAAGTG(서열번호 45)
	4 갭	CCACTCTGhhhbTGGAAGTG(서열번호 46)

표 2C. 파지 풀 갭 선택성 연구를 위해 사용한 표적. 파지 풀 갭 선택성에서 사용한 듀플렉스 DNA 부위는 다음의 일반 형태를 가졌으며: TATAAT$(X)_{16-20}$TTCACAGTCACACGTC, (서열번호 67), 여기서 $(X)_{16-20}$는 표에서 열거한 서열로 치환되었다. DNA 듀플렉스는 이탤릭체 서열로 어닐링된 프라이머를 연장함으로써 만들어졌다. 밑줄친 염기는 각 숙주 ZFP의 4개의 핑거에 대해 결합 서열을 표시하는 한편, 소문자는 선택된 링커에 의해 포괄되어야 하는 삽입된 뉴클레오타이드(또는 "갭" 염기)를 표시한다. 갭 염기에 대한 축퇴 코드는 다음과 같다: "d"는 A, G, 및 T의 혼합을 의미하며; "b"는 C, G, 및 T의 혼합을 의미하고; "h"는 A, C, 및 T의 혼합을 의미하며; "v"는 A, C, 및 G의 혼합을 의미한다.

[0109]

[0110] **시퀀싱**

[0111] 선택된 ZFP를 암호화하는 유전자를 서브클로닝하고 시퀀싱하였다. 도 3A는 모두 3개의 숙주 단백질의 맥락에서 1 bp 갭을 스킵하기 위해 선택된 링커를 제시하는 한편, 도 3B는 "ZFP8196" 숙주의 맥락에서 2 bp를 스킵하기 위해 선택된 링커를 나타낸다. 시퀀싱 결과는 프롤린- 및 아르기닌-풍부 서열에 대해 선택된 링커 내 강한 조성의 바이어스를 나타내었다. 분명한 링커 길이 경향은 또한 명백하였으며: 출발 라이브러리가 대략적으로 동일한 비율의 11개의 상이한 링커 길이(2 내지 12개의 잔기)로 암호화되었지만, 선택된 링커는 5 내지 8개 잔기(1bp 갭에 대해) 또는 9 내지 11개의 잔기(2bp 갭에 대해)로부터의 더 좁은 분포를 특징으로 하였다.

[0112] **실시예 2: 선택된 ZFP의 초기 특징화**

[0113] 1 bp를 스킵하도록 선택한 링커의 초기 기능적 평가로서, 도 3A에서 열거한 링커를 함유하는 ZFP를 서브클로닝하였고, 시험관내 전사-번역 키트를 사용하여 유리 단백질로 발현시켰으며, 선택한 링커 맞은 편의 0, 1 또는 2 bp의 삽입을 함유하는 표적에 결합을 위해 ELISA로 평가하였다. 이 연구를 위한 표적을 표 3에서 열거한다. 9개의 추가적인 대조군 단백질을 각 숙주 ZFP의 중심 링커를 3개의 또 다른 앞서 특징화한, 1bp를 포괄하기 위한 기술의 언급을 수집적으로 나타낸 링커 서열로 치환함으로써 만들었다. 이 대조군 링커의 서열은 LRQKDERP(서열번호 3)(미국특허 제6,479,626호 참조), TGEGGKP(서열번호 48), TGGGGSQKP(서열번호 2)이었다. 이 대조군 단백질뿐만 아니라 숙주 ZFP를 또한 ELISA 연구에 포함시켰다.

[0114] 표 3은 1bp 갭을 스킵하도록 선택한 ZFP의 ELISA 연구를 위해 사용한 표적을 나타낸다. 이 연구를 사용한 듀플렉스 DNA 부위는 일반적 형태 TTAG$(X)_{16-18}$TATC(서열번호 94)를 가지며, 여기서 $(X)_{16-18}$는 표에 열거한 서열로 치환하였다. 각 듀플렉스 DNA 표적은 그의 5' 말단에서 바이오틴을 함유하는 상보적 올리고뉴클레오타이드를 어닐링함으로써 만들어졌다. 밑줄은 각각의 숙주 ZFP의 4개의 핑거에 대한 결합 서열을 나타내는 한편, 소문자는 삽입 뉴클레오타이드(또는 "갭" 염기)를 나타낸다.

[0115] 이 연구 결과를 표 4, 5 및 6에서 제공하며, 각 테이블은 상이한 숙주 ZFP로부터 유래된 단백질에 대한 데이터를 열거한다. 표 4는 ZFP8196-유래 단백질에 대한 데이터를 제공하며; 표 5는 ZFP7263-유래 단백질에 대한 데이터를 제공하고; 표 6은 ZFP7264-유래 단백질에 대한 데이터를 제공한다. 각 표에서, 숙주 ZFP에 대한 결합 데이터는 최상단의 열에 열거한 다음, 열 2 내지 4에서 3개의 대조군 단백질에 대한 결합 데이터를 열거한다. 본래

의 ZFP의 그의 비변형 표적에 대한 결합으로부터 얻은 ELISA 신호로 값을 정규화한다.

[0116] 각 세트의 단백질은 3가지 중요한 사항에서 유사한 패턴의 결합 거동을 나타내었다: 우선, 각각의 본래 ZFP는 그의 비변형 표적에 잘 결합하였지만(표 4, 5 및 6에서 "0-bp 갭" 표적) 1 또는 2 bp의 삽입을 함유하는 임의의 변이체와 결합하지 않았다. 이는 본래의 링커(TGEKP(서열번호 1)(ZFP8196에 대해) 또는 TGSQKP(서열번호 72)(ZFP7263 및 ZFP7264에 대해))가 어떤 추가적인 삽입 염기를 포괄하기에 너무 짧기 때문인 것으로 예상되었다.

[0117] 둘째로, 거의 모든 경우에, 대조군 단백질은 1bp 삽입으로 표적화하기에 매우 불량하게 결합되었다(정규화된 ELISA 값은 36개의 이러한 측정 중 31개에 대해 0.10 또는 그 미만이었다). 이는 이들 연구 전 이용가능한 링커의 불량한 수행을 나타낸다. 게다가, 이들 단백질에 의해 사용된 링커는 1bp 삽입(대 0bp 삽입)을 함유하는 표적에 대해 일관된 선호도를 나타내지 않았다.

[0118] 셋째로, 대조군 단백질의 거동에 대해, 파지-선택 ZFP는 1 bp 삽입을 함유하는 표적에 대해 훨씬 더 높은 친화도로뿐만 아니라 삽입 염기를 함유하는 결합 표적에 대해 훨씬 더 높은 수준의 차별로 결합되었다. 이 단백질은 또한 1bp 삽입을 가지는 결합 표적 대 2 bp 삽입을 함유하는 표적을 결합하는 것에 대해 매우 선택적이었다.

표 3

[0119]

1bp 갭을 스킵하기 위해 선택한 ZFP의 ELISA 연구를 위해 사용한 표적		
ZFP w/랜덤화된 링커	갭 서열	표적 부위
8196	-	ATAAACTGCAAAAGGC (서열번호 32)
	A	ATAAACTGaCAAAAGGC (서열번호 73)
	C	ATAAACTGcCAAAAGGC (서열번호 74)
	G	ATAAACTGgCAAAAGGC (서열번호 75)
	T	ATAAACTGtCAAAAGGC (서열번호 76)
	TC	ATAAACTGtcCAAAAGGC (서열번호 77)
	AC	ATAAACTGacCAAAAGGC (서열번호 78)
	TG	ATAAACTGtgCAAAAGGC (서열번호 79)
7264	-	TTAAAGCGGCTCCGAA (서열번호 37)
	A	TTAAAGCGaGCTCCGAA (서열번호 80)
	C	TTAAAGCGcGCTCCGAA (서열번호 81)
	G	TTAAAGCGgGCTCCGAA (서열번호 82)
	T	TTAAAGCGtGCTCCGAA (서열번호 83)
	TT	TTAAAGCGttGCTCCGAA (서열번호 84)
	TA	TTAAAGCGtaGCTCCGAA (서열번호 85)
	CT	TTAAAGCGctGCTCCGAA (서열번호 86)
7263	-	CCACTCTGTGGAAGTG (서열번호 42)
	A	CCACTCTGaTGGAAGTG (서열번호 87)
	C	CCACTCTGcTGGAAGTG (서열번호 88)
	G	CCACTCTGgTGGAAGTG (서열번호 89)
	T	CCACTCTGtTGGAAGTG (서열번호 90)
	AC	CCACTCTGacTGGAAGTG (서열번호 91)
	AT	CCACTCTGatTGGAAGTG (서열번호 92)
	CT	CCACTCTGctTGGAAGTG (서열번호 93)

표 4

[0120]

상이한 중심 링커를 가지는 ZFP "8196"의 변이체에 대한 ELISA 결과									
중심 링커의 서열	표시한 갭을 가지는 표적에 결합을 위해 사용한 ELISA 스코어 [스코어는 그의 비 갭화된 표적(밑줄친 항목)에 결합된 8196으로 정규화된다]								
	0-bp 갭	1-bp 갭					2-bp 갭		
	-	A	C	G	T	1bp:0bp 스코어의 평균 비	TC	AC	TG
TGEKP (서열번호 1)	1.00	0.01	0.01	0.03	0.01	0.02	0.00	0.00	0.00
TGGGGSQKP (서열번호 2)	0.00	0.00	0.00	0.00	0.00	1.20	0.00	0.00	0.00
LRQKDERP (서열번호 3)	0.01	0.01	0.01	0.04	0.08	3.49	0.00	0.00	0.00
TGEGGKP (서열번호 48)	0.10	0.00	0.00	0.03	0.03	0.15	0.00	0.00	0.00
TPDAPKPKP (서열번호 49)	0.02	0.16	0.13	0.68	0.95	23.75	0.01	0.00	0.01
TPGLHRPKP (서열번호 50)	0.04	0.19	0.10	0.65	0.81	10.94	0.01	0.00	0.01
TEPRAKPPKP (서열번호 51)	0.01	0.39	0.17	0.78	0.93	70.72	0.02	0.01	0.01
TPSHTPRPKP (서열번호 52)	0.02	0.30	0.13	0.84	0.80	25.10	0.02	0.01	0.01
TGYSIPRPKP (서열번호 53)	0.01	0.13	0.06	0.43	0.55	44.57	0.01	0.00	0.01
TYPRPIAAKP (서열번호 54) (명명된 1f)	0.01	0.41	0.14	0.65	0.64	82.25	0.01	0.00	0.01
THPRAPIPKP (서열번호 55) (명명된 1c)	0.00	0.20	0.09	0.57	0.60	78.86	0.01	0.00	0.00
TPNRRPAPKP (서열번호 56) (명명된 1d)	0.00	0.23	0.09	0.52	0.52	90.27	0.01	0.01	0.01
TSPRLPAPKP (서열번호 57)	0.01	0.26	0.14	0.62	0.81	67.95	0.01	0.00	0.01
TCPRPPTRKP (서열번호 58)	0.00	0.18	0.05	0.48	0.62	70.16	0.01	0.00	0.01
TSSPRSNAKP (서열번호 59)	0.01	0.05	0.02	0.20	0.25	20.85	0.01	0.00	0.01
TVSPAPCRSKP (서열번호 60)	0.01	0.03	0.01	0.14	0.19	11.52	0.02	0.00	0.01
TPDRPISTCKP (서열번호 61)	0.01	0.11	0.05	0.29	0.47	15.41	0.03	0.01	0.02

표 5

[0121]

상이한 중심 링커를 가지는 ZFP "7263"의 변이체에 대한 ELISA 결과

중심 링커의 서열	표시한 갭을 가지는 표적에 대한 결합에 대한 ELISA 스코어 [스코어는 그의 비갭화된 표적(밑줄친 항목)에 결합된 7263으로 정규화된다]								
	0-bp 갭	1-bp 갭					2-bp 갭		
	-	A	C	G	T	1bp:0bp 스코어의 평균 비	AC	AT	CT
TGSQKP (서열번호 72)	1.00	0.01	0.01	0.03	0.02	0.02	0.01	0.01	0.00
TGGGGSQKP (서열번호 2)	0.51	0.06	0.05	0.41	0.39	0.44	0.01	0.01	0.02
LRQKDERP (서열번호 3)	0.25	0.03	0.02	0.18	0.13	0.36	0.01	0.01	0.01
TGEGGKP (서열번호 48)	1.30	0.02	0.02	0.05	0.04	0.03	0.01	0.01	0.01
TPRPPIPKP (서열번호 4)	0.14	0.97	0.67	1.85	2.09	10.20	0.02	0.01	0.01
TQRPQIPPKP (서열번호 62)	0.15	1.66	1.00	2.86	3.05	14.68	0.03	0.02	0.01
TPNRCPPTKP (서열번호 63)	0.31	1.68	1.13	2.62	3.16	7.53	0.03	0.02	0.01
TYPRPLLAKP (서열번호 7)	0.29	1.95	1.27	3.88	3.97	10.08	0.03	0.01	0.01
TPLCQRPMKQKP (서열번호 8)	0.28	1.82	1.28	3.44	4.00	10.88	0.08	0.05	0.02

표 6

[0122]

상이한 중심 링커를 가지는 ZFP "7264"의 변이체에 대한 ELISA 결과									
중심 링커의 서열	표시한 갭을 가지는 표적에 대한 결합에 대한 ELISA 스코어 [스코어는 그의 비갭화된 표적(밑줄친 항목)에 결합된 7264로 정규화된다]								
	0-bp 갭	1-bp 갭					2-bp 갭		
	-	A	C	G	T	1bp:0bp 스코어의 평균 비	TT	TA	CT
TGSQKP (서열번호 72)	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TGGGGSQKP (서열번호 2)	0.46	0.07	0.04	0.08	0.17	0.19	0.03	0.03	0.07
LRQKDERP (서열번호 3)	0.26	0.05	0.03	0.06	0.10	0.22	0.02	0.02	0.02
TGEGGKP (서열번호 48)	1.39	0.02	0.03	0.05	0.08	0.03	0.03	0.02	0.03
TGLPKPKP (서열번호 64)	0.14	0.19	0.11	0.38	0.86	2.84	0.03	0.02	0.02
TSRPRPKP (서열번호 11)	0.18	0.52	0.22	0.77	2.07	4.93	0.03	0.03	0.03
TLPLPRPKP (서열번호 65)	0.25	0.58	0.25	0.85	1.36	3.01	0.04	0.03	0.03
TVPRPTPPKP (서열번호 12) (명명된 1e)	0.16	2.35	1.02	1.58	2.55	11.71	0.05	0.05	0.06
TLPPCFRPKP (서열번호 66)	0.36	0.72	0.25	0.77	2.72	3.11	0.06	0.06	0.05
TKHGTPKHREDKP (서열번호 13)	0.01	0.01	0.01	0.01	0.01	0.79	0.00	0.00	0.00

[0123] ELISA 연구에서 얻은 결과를 추가로 지지하고 확장하기 위하여, 선택한 링커를 가지는 ZFP를 국제특허 공개 WO 2009/042163에서 설명한 효모 스크리닝 분석을 사용하여 다양한 표적 부위에서 생체내 절단 활성에 대해 평가하였다. 이 실험은 더욱 노동 집약적이기 때문에, 그들을 거의 없는 ZFP 상에서 수행하였다. 이 연구를 위해, 표 5로부터의 6개 ZFP 및 표 6으로부터의 5개 ZFP를 WO 2007/139982에서 설명한 것과 같은 아연 핑거 뉴클레아제(ZFN)로서 발현되도록 하는 구성체 내로 조립하였다. 다음에 생체내 활성을 다양한 표적 부위를 가지는 효모 균주로부터 MEL-1 선택을 평가함으로써 측정하였다. 이들 연구를 위해 사용한 표적 서열을 표 7 및 표 8에서 제공하며, 0, 1- 또는 2-bp의 중심 삽입과 함께 7263 및 7264 결합 부위의 변형을 포함하였다.

표 7

[0124]

ZFP7263 내 1bp 갭을 스킵하기 위해 사용한 ZFP의 효모 스크리닝 분석을 위해 사용한 표적		
ZFP w/무작위화된 링커	**갭 서열**	**표적 부위**
7263	-	ACTCTGTGGAAG (서열번호 95)
	A	ACTCTGaTGGAAG (서열번호 96)
	C	ACTCTGcTGGAAG (서열번호 97)
	G	ACTCTGgTGGAAG (서열번호 98)
	T	ACTCTGtTGGAAG (서열번호 99)
	AC	ACTCTGacTGGAAG (서열번호 100)
	AT	ACTCTGatTGGAAG (서열번호 101)
	CT	ACTCTGctTGGAAG (서열번호 102)

[0125] 뉴클레아제 표적 카세트가 GATCTGTT*CGGAGCCGCTTT*AACCC$(X)_{12-14}$TGCTCGCG(서열번호 103)의 일반적 형태를 가지는 것을 제외하고, 뉴클레아제 표적 부위를 함유하는 리포터 플라스미드를 국제특허 공개 WO 2009/042163에서 설명한 것과 같이 본질적으로 구성하였으며, 여기서 (1) 말단 중 하나에서 4개의 밑줄친 염기는 BamHI/BssHII 분해된 리포터 플라스미드 내로 클로닝을 위해 사용한 돌출부를 나타내며, (2) 이탤릭체 서열은 안티센스 가닥에 결합하며 이 스크린에 대한 변이체인 7264 ZFN에 대한 결합 부위를 나타내고, (3) $(X)_{12-14}$는 표에서 열거한 서열로 치환시켰다. 대문자 염기는 각 숙주 ZFP의 4개의 핑거에 대한 결합 서열을 나타내는 한편, 소문자는 삽입된 뉴클레오타이드(또는 "갭" 염기)를 나타낸다.

표 8

[0126]

ZFP7264 내 1bp 갭을 스킵하기 위해 선택한 ZFP의 효모 스크리닝 분석을 위해 사용한 표적		
ZFP w/무작위화된 링커	**갭 서열**	**표적 부위**
7264	-	AAAGCGGCTCCG (서열번호 104)
	A	AAAGCGaGCTCCG (서열번호 105)
	C	AAAGCGcGCTCCG (서열번호 106)
	G	AAAGCGgGCTCCG (서열번호 107)
	T	AAAGCGtGCTCCG (서열번호 108)
	TT	AAAGCGttGCTCCG (서열번호 109)
	TA	AAAGCGtaGCTCCG (서열번호 110)
	CT	AAAGCGctGCTCCG (서열번호 111)

[0127] 뉴클레아제 표적 카세트가 GATCTGTT$(X)_{12-14}$AACCC*ACTCT*GTGGAAGTGCTCGCG(서열번호 112)의 일반적 형태를 가지는 것을 제외하고, 뉴클레아제 표적 부위를 함유하는 리포터 플라스미드는 국제특허 공개 WO 2009/042163에서 설명한 것과 같이 본질적으로 구성되며, 여기서 (1) 말단 중 하나에서 밑줄친 염기는 BamHI/BssHII 분해된 리포터 플라스미드 내로 클로닝을 위해 사용된 돌출부를 나타내며, (2) 이탤릭체 서열은 이 스크린에 대한 변이체인 7263 ZFN에 대한 결합 부위를 나타내고, (3) $(X)_{12-14}$는 표에서 열거한 서열로 치환된다. 대문자 염기는 각 숙주 ZFP의 4개 핑거에 대한 결합 서열을 나타내는 한편, 소문자는 삽입된 뉴클레오타이드(또는 "갭" 염기)를 나타낸

다. 7264 ZFN이 안티센스 가닥에 결합하기 때문에 표에서 열거한 표적 부위는 표적 카세트 내 존재하는 것의 역보체라는 것에 주의한다.

[0128] 이 실험에 대한 데이터는 표 9 및 표 10에 나타내며, 각 표는 상이한 숙주 ZFN으로부터 유래된 단백질에 대한 데이터를 열거한다. 표 9는 7263-유래 ZFN에 대한 데이터를 제공하며, 표 10은 7264-유래 ZFN에 대한 데이터를 제공한다. 각 표에서, 숙주 ZFN에 대한 뉴클레아제 활성 데이터는 최상부 열에서 열거한 다음, 2열에서 하나의 대조군 단백질에 대한 뉴클레아제 활성 데이터를 열거한 후, 파지 디스플레이 라이브러리로부터 선택한 ZFP에 대한 데이터를 열거한다. ZFP7263 및ZFP7264가 동일한 아연 핑거 뉴클레아제 다이머의 2개의 절반이기 때문에, 숙주 ZFN에 대한 데이터는 각 표에서 동일하다. 이 연구 결과를 ELISA 연구에서 관찰한 패턴과 매치시켰고, 즉 ZFN 함유 파지 선택 링커는 ZFP 함유 대조군 링커보더 1bp 삽입을 함유하는 표적에 대해 더 높은 활성과 더 양호한 선호도를 나타내었다.

표 9

[0129]

상이한 중심 링커를 가지는 ZFP7263의 변이체에 대한 효모 스크리닝 결과

중심 링커의 서열	표시한 갭을 가지는 표적에 대한 결합에 대한 MEL-1 스코어 [스코어는 그의 비갭화된 표적(밑줄친 항목)에 결합된 ZFP7263으로 정규화된다]								
	0-bp 갭	1-bp 갭					2-bp 갭		
	-	A	C	G	T	1bp:0bp 스코어의 평균 비	AC	AT	CT
TGSQKP (서열번호 72)	1.00	0.05	0.05	0.09	0.06	0.06	0.09	0.02	0.08
TGGGGSQKP (서열번호 2)	0.11	0.14	0.04	0.60	0.31	2.57	0.04	0.02	0.04
TPRPPIPKP (서열번호 4)	0.04	1.30	1.41	2.18	1.40	35.90	0.05	0.03	0.04
TQRPQIPPKP (서열번호 62)	0.04	1.05	0.43	2.18	1.19	34.18	0.03	0.02	0.03
TPNRCPPTKP (서열번호 63)	0.05	1.30	0.34	2.85	1.59	33.22	0.04	0.03	0.06
TYPRPLLAKP (서열번호 7)	0.05	0.69	0.37	2.21	1.06	20.15	0.10	0.05	0.05
TPLCQRPMKQKP (서열번호 8)	0.04	0.97	0.30	1.52	1.19	27.52	1.05	0.01	0.03

표 10

[0130]

상이한 중심 링커를 가지는 ZFP7264의 변이체에 대한 효모 스크리닝 결과

중심 링커의 서열	표시한 갭을 가지는 표적에 대한 결합에 대한 MEL-1 스코어 [스코어는 그의 비갭화된 표적(밑줄친 항목)에 결합된 ZFP7264로 정규화된다]								
	0-bp 갭	1-bp 갭					2-bp 갭		
	-	A	C	G	T	1bp:0bp스코어의 평균 비	TT	TA	CT
TGSQKP (서열번호 72)	1.00	0.05	0.05	0.09	0.06	0.06	0.09	0.02	0.08
TGGGGSQKP (서열번호 2)	0.28	0.08	0.07	0.14	0.34	0.57	0.12	0.09	0.15
TGLPKPKP (서열번호 64)	0.04	0.08	0.07	0.42	0.79	8.03	0.11	0.08	0.09

TSRPRPKP (서열번호 11)	0.08	0.18	0.08	0.61	3.38	13.33	0.03	0.07	0.10
TVPRPTPPKP (서열번호 12) (명명된 1e)	0.08	1.24	0.14	1.40	2.34	20.28	0.12	0.09	0.11

[0131] ELISA가 효모 스크리닝 데이터에 밀접하게 일치하기 때문에, 본 발명자들은 ELISA 분석 중에서 최고로 수행된 예시적인 1bp-스킵 링커의 세트를 선택하였다. 이것들을 표 11에서 열거하며 "1c", "1d", "1e", 및 "1f"로 명명한다. 이 명명들은 또한 표 4, 표 6 및 표 10에서 제시한 데이터 내에 포함시켰다. 이 예시적인 링커에 대한 ELISA 데이터를 또한 도 4 및 도 5에서 나타낸다.

표 11

[0132]

예시적 링커 설계		
링커 서열	링커 명명	스킵한 염기의 수
THPRAPIPKP (서열번호 55)	1c	1
TPNRRPAPKP (서열번호 56)	1d	1
TVPRPTPPKP (서열번호 12)	1e	1
TYPRPIAAKP (서열번호 54)	1f	1

[0133] 실시예 1에서 앞서 언급한 바와 같이, 선택을 위해 사용한 표적 부위(표 2A)는 특정 갭 서열에 대해 본래의 선호도를 나타내지 않은 링커의 선택을 유리하게 하기 위해서 갭 내에 축퇴 염기를 함유하였다. 도 4 및 도 5에서 나타낸 데이터는 이 선택 전략이 성공적이었음을 시사한다: 예시적인 링커를 함유하는 단백질은 A, G, C 또는 T의 갭 염기를 가지는 표적 중에서 결합 내 변형을 거의 나타내지 않는다. 게다가, 관찰된 변형의 미량은 대조군 가요성 라이너에 의해 얻어진 결과에서 반영되며(예를 들어 가요성 링커(TGGGGSQKP)(서열번호 2)에 대해 도 5D), 변형은 옆에 있는 핑거의 특성을 나타낸다.

[0134] ZFP8196의 맥락에서 2 염기쌍 갭을 스킵하도록 선택된 링커에 대해 유사한 ELISA 실험을 수행하였다. 앞서 2bp를 포괄하는 것에 대한 기술의 언급을 수집적으로 나타낸 링커 서열을 특징으로 한 2개의 대체물로 각 숙주 ZFP의 중심 링커를 치환함으로써 2개의 추가적인 대조군 단백질을 만들었다. 이 대조군 링커의 서열은 TGGGGSGGSQKP(서열번호 14) 및 LRQKDGGGSERP(서열번호 68)이었다. 이 대조군 단백질뿐만 아닐 숙주 ZFP를 또한 ELISA 연구에 포함시켰다. 갭을 함유하지 않거나, 각각 4개 중 1개의 가능한 염기쌍 갭, 및 각각 16개 중 2개의 가능한 염기쌍 갭을 함유하는 표적 부위에 결합을 위해 ZFP를 시험하였다. 표적 부위를 표 12에 열거한다. 8196 ZFP의 그의 비갭화 표적 부위에 결합된 스코어로 ELISA 스코어를 정규화하였다. 1bp-스킵 링커에 유사한 방식으로 그의 ELISA 거동을 기준으로 예시적인 링커를 선택하였다. 이 예시적인 링커는 "2d", "2e", 및 "2f"를 명명하였다. 8196 ZFP 및 대조군 ZFP에서 예시적 링커에 대한 ELISA 결과를 도 6에서 나타낸다. 도 6에서 나타낸 바와 같이, 대조군 링커를 함유하는 ZFP 중 어떤 것도 2염기쌍 갭을 가지는 표적 부위에 대해 선호도를 나타내지 않았다. 그러나 선택 링커를 가지는 ZFP는 1개의 염기쌍 및 비갭화된 표적 부위 둘 다에 걸쳐 2개의 염기쌍 갭에 대해 확실한 선호도를 나타내었다.

표 12

[0135]

2bp 스킵 링커의 ELISA 특징화를 위해 사용한 표적 부위		
ZFP w/무작위화된 링커	**갭 서열**	**표적 부위**

8196	-	ATAAACTGCAAAAGGC (서열번호 32)
	A	ATAAACTGaCAAAAGGC (서열번호 73)
	C	ATAAACTGcCAAAAGGC (서열번호 74)
	G	ATAAACTGgCAAAAGGC (서열번호 75)
	T	ATAAACTGtCAAAAGGC (서열번호 76)
	AA	ATAAACTGaaCAAAAGGC (서열번호 113)
	AC	ATAAACTGacCAAAAGGC (서열번호 114)
	AG	ATAAACTGagCAAAAGGC (서열번호 115)
	AT	ATAAACTGatCAAAAGGC (서열번호 116)
	CA	ATAAACTGcaCAAAAGGC (서열번호 117)
	CC	ATAAACTGccCAAAAGGC (서열번호 118)
	CG	ATAAACTGcgCAAAAGGC (서열번호 119)
	CT	ATAAACTGctCAAAAGGC (서열번호 120)
	GA	ATAAACTGgaCAAAAGGC (서열번호 121)
	GC	ATAAACTGgcCAAAAGGC (서열번호 122)
	GG	ATAAACTGggCAAAAGGC (서열번호 123)
	GT	ATAAACTGgtCAAAAGGC (서열번호 124)
	TA	ATAAACTGtaCAAAAGGC (서열번호 125)
	TC	ATAAACTGtcCAAAAGGC (서열번호 126)
	TG	ATAAACTGtgCAAAAGGC (서열번호 127)
	TT	ATAAACTGttCAAAAGGC (서열번호 128)

표 12. ELISA 특징화 연구를 위해 사용한 듀플렉스 DNA 표적 부위는 다음의 일반적 형태를 가지며: TTAG$(X)_{16-18}$TATC, (서열번호 94), 여기서 $(X)_{16-18}$은 표에서 열거한 서열로 치환되었다. DNA 듀플렉스는 상보적 올리고뉴클레오타이드를 어닐링함으로써 만들어졌다. 표에서 열거한 서열에 상보적인 올리고뉴클레오타이드는 5' 바이오틴을 함유하였다. 밑줄친 염기는 각 숙주 ZFP의 4개 핑거에 대한 결합 서열을 표시하는 한편, 소문자 염기는 삽입된 뉴클레오타이드(또는 "갭" 염기)를 표시한다.

[0136] 도 1에서 앞서 언급한 바와 같이, 선택을 위해 사용한 표적 부위(표 2A)는 특정 갭 서열에 대해 고유의 선호도가 없다는 것을 나타낸 링커의 선택을 돕기 위하여 갭 내에 축퇴 염기를 함유하였다. 가요성 링커(TGGGGSGGSQKP (서열번호 14)) 중 하나에 대한 스케일의 확대는 도 6에서 나타낸다. 이 가요성 링커는 표적 부위와 임의의 상호작용을 가지지 않아야 하며, 따라서 보여진 패턴은 아연 핑거 단백질의 결합에 기인할 가능성이 있다. 예시적인 링커가 2개의 염기쌍 갭 표적 부위에 대해 유사한 결합 패턴을 나타낸다는 사실은 선택 링커가 또한 ZFP 결합 내 임의의 갭 조성 바이어스를 부과하기 않아야 한다는 것을 시사한다.

[0137] 이 데이터의 더 간결한 요약은 표 13에서 제시하며, ELISA 스코어는 1 또는 2개의 염기쌍 갭 조성의 모두에 따라 평균화하였다. 또한 1개의 염기쌍 갭 및 비갭화된 표적에 따라 2개의 염기쌍 갭에 대한 배수 선호도를 보고한다. 가장 선택적인 링커(TPNPHRRTDPSHKP(서열번호 69), "2f")는 대조군 링커와 비교하여 0개 염기쌍 갭에 걸쳐 100배 초과, 1개 염기쌍 갭에 걸쳐 20배 초과의 2개 염기쌍 갭 선택성에서 개선을 나타낸다.

표 13

[0138]

2bp 스킵 링커에 대한 ELISA 데이터의 요약						
		평균 정규화 ELISA 스코어(갭)			2bp-갭 선택성 대:	
링커 서열	명명	0bp	1bp	2bp	0bp	1bp
TGGGGSGGSQKP (서열번호 14)	가요성	0.019	0.041	0.010	0.6	0.2
LRQKDGGGSERP (서열번호 68)	Kim,Pabo	0.010	0.047	0.003	0.4	0.1
TPNPHRRTDPSHKP (서열번호 69)	2f	0.003	0.046	0.219	64.6	4.7
TLAPRPYRPPKP (서열번호 70)	2d	0.005	0.035	0.127	24.4	3.6
TPGGKSSRTDRNKP (서열번호 71)	2e	0.005	0.099	0.100	22.0	1.0

[0139] **실시예 3: 다양한 숙주 ZFP에서 링커의 ELISA 특징화**

[0140] 예시적인 링커의 일반성을 증명하기 위하여, 표 11(1c-1f)에서 열거한 4개의 1bp-스킵 링커를 12개의 상이한 숙주 ZFP 내로 클로닝하였다. 숙주 ZFP는 ZFP1, ZFP2 등을 명명하였다. 결과 단백질을 시험관 내 전사 및 번역을 통해 발현시켰고, 상기 설명한 것과 같이 ELISA를 통해 시험하였다. 비교를 위해, 본 발명자들은 또한 가요성 링커(TGGGGSQKP(서열번호 2))를 가지는 숙주 ZFP를 시험하였고, 결과를 도 7에서 제시한다. 이 데이터는 표준 가요성 링커에 대해, 분석(ZFP1 및 ZFP2)을 포화시키거나 결합이 검출가능하지 않게 낮은(ZFP 11 및 ZFP 12) ZFP가 되는 것의 유일한 예외로, 새로운 링커가 대부분의 숙주 ZFP의 ELISA 스코어를 상당히 증가시켰다는 것을 증명한다. 모든 숙주 ZFP에 걸친 ELISA 스코어에서 평균 배수의 증가는 3 내지 5이었다.

[0141] 유사한 연구에서, 표 13(2d-2f)에서 열거한 2bp 갭을 스킵하기 위해 선택한 3개의 예시적 링커는 상기 설명한 것과 같은 6개의 상이한 숙주 ZFP(ZFP13, ZFP14 등)에서 시험하였고, 이 결과를 도 8에서 제시한다. 이 실험에서, 모든 숙주 ZFP에 걸친 ELISA 스코어의 평균 배수 개선은 1.9 내지 2.4의 범위에 있었다.

[0142] **실시예 4: 포유류 세포 내 내인성 좌위에서 예시적인 링커를 가지는 ZFN의 특징화**

[0143] 다음에 내인성 좌위에서 이중 가닥 파손을 유발하는 ZFN의 능력에 대해 ZFN을 시험하였다. 간략하게, 상기 설명한(실시예 3) 18개의 ZFN을 암호화하는 플라스미드(ZFP-*Fok*I 융합)를 그의 적절한 파트너 ZFN과 짝지었고, 제조업자에 의해 특정된 바와 같은 Amaxa(상표명) Nucleofection 키트를 사용하여 트랜스펙션에 의해 K562 세포 내로 도입하였다. 비상동성 말단 접합(NHEJ)의 수준에 의해 측정한 바와 같은 표적 좌위에서 ZFN 활성을 결정하기 위하여, 제조업자의 설명서(Transgenomic SURVEYORTM)에 따라 CEL-I 미스매치 분석을 본질적으로 수행하였다. 세포를 채취하였고 제조업자의 지시에 따라서 Quickextract(상표명) 키트를 사용하여 염색체 DNA를 제조하였다(Epicentre(등록상표)). 표적 좌위의 적절한 영역을 Accuprime(상표명) Taq High-fidelity DNA 폴리머라제(Invitrogen)를 사용하여 PCR 증폭한 후 CEL-I 효소로 처리하였다.

[0144] CEL-I 분석을 위해 생성한 예시적 겔을 도 9에 나타낸다. 도 9A는 숙주 ZFN으로서 ZFN3 및 ZFN4(1개 염기쌍을 스킵하는 ZFN)에 대한 스크리닝 데이터를 나타내는 반면, 도 9B는 숙주 ZFN으로서 ZFN14(2개 염기쌍을 스킵하는 ZFN)에 대한 스크리닝 데이터를 나타낸다. 모든 1 bp 스킵의 예시적인 링커에 대한 데이터를 도 10에 요약한다(ZFN1-ZFN12). 일부 ZFN을 높은 발현 조건을 사용하여 발현시켰다. 24시간 동안 37℃에서 세포를 인큐베이션한 다음 48시간 동안 30℃에서 인큐베이션함으로써 트랜스펙션 후 높은 발현을 얻은 다음, 게놈 DNA를 분리하였다. 이 조건을 이용하는 ZFN은 도 10에서 강조 표시하였다. 특히 TGGGGSQKP(서열번호 2) 링커("가요성 링커")(ZFN 4, 9 및 10)을 가지는 불활성인 3개의 ZFN은 본 명세서에서 설명하는 링커를 사용할 때 활성이 되었다. 이 경우를 위해, 1.0% 변형 값을 정규화 목적을 위한 가요성 링커에 부과하였다(분석의 검출 제한). 새로운 링커로 시험한 ZFN의 85%에서, 유전자 변형의 수준 증가를 관찰하였고, 9개의 활성 ZFN 쌍에 걸쳐 대략 1.8배 내지 2.8배의 평균 증가가 있었다.

[0145] 유사하게, 2bp 스킵 예시적인 링커(ZFN13-ZFN18)를 함유하는 상기 설명한 ZFN(실시예 3)을 내인성 좌위에서 시험하였고, 결과를 도 11에서 요약한다. 이 연구에서, 본 명세서에서 설명한 링커의 치환은 4개의 활성 ZFN 중 3개에 대한 가요성 링커와 비교하여 활성을 개선시켰고, 평균 개선은 모든 활성 ZFN 쌍에 걸쳐 1.5 내지 2배이었다.

[0146] **실시예 5: 2 bp 스킵 링커에 대한 2차적 선택**

[0147] 라이브러리의 2차적 세트를 2-bp 스킵 링커에 대한 초기 선택으로부터 얻은 정보(실시예 3 및 도 3)를 기반으로 구성하였다. 이들 라이브러리는 RPP로서 링커의 3개 카복시 말단 잔기(리신, 프롤린, 프롤린)를 고정시켰고, 남아있는 아미노 말단 잔기를 무작위화하였다. 라이브러리 설계를 도 12에 나타낸다.

[0148] 숙주 단백질로서 ZFP8196을 사용하여 실시예 1에서와 동일한 방식으로 선택을 수행하였다. 선택한 파지 풀의 갭 선택성은 도 13A에서 나타내며, 개개의 클론으로부터의 링커 서열은 도 13B에 나타낸다.

[0149] 실시예 2와 유사하게, 2차적 선택으로부터 개개의 클론 각각에서 ELISA 실험을 수행하였다(도 13B). 갭 없음, 4개 중 1개의 가능한 염기쌍 갭의 풀, 및 16개 중 2개의 가능한 염기쌍 갭의 풀을 함유하는 표적 부위에 결합을 위해 ZFP를 시험하였다. 표적 부위를 표 12에 열거한다. ELISA 스코어를 그의 비갭화된 표적 부위에 결합된 숙주 ZFP8196의 스코어로 정규화하였다. 2-bp 갭 표적 부위의 풀 상에서 양호한 정규화된 ELISA 스코어와 양호한 갭 선택성을 둘 다 나타낸 ZFP 함유 링커에 대한 ELISA 결과를 표 14에 나타낸다.

표 14

[0150]

2bp 스킵 링커에 대한 ELISA 데이터의 요약					
	평균 정규화 ELISA 스코어(갭)			2bp-갭 선택성 대:	
링커 서열	0bp	1bp	2bp	0bp	1bp
TETTRPFRPPKP (서열번호 183)	0.001	0.001	0.570	570.0	570.0
TGSLRPYRRPKP (서열번호 177)	0.001	0.010	0.310	310.0	31.0
TSINRPFRRPKP (서열번호 184)	0.010	0.020	0.570	57.0	28.5
TNTTRPYRPPKP (서열번호 175)	0.001	0.010	0.410	410.0	41.0
TASCPRPFRPPKP (서열번호 194)	0.010	0.020	0.370	37.0	18.5
TGEARPYRPPKP (서열번호 178)	0.001	0.010	0.610	610.0	61.0

[0151] 나타낸 바와 같이, 선택한 링커를 가지는 ZFP는 1개의 염기쌍과 비갭화된 표적 부위 둘 다에 걸쳐 2개의 염기쌍 갭에 대해 명확한 선호도를 나타내었다.

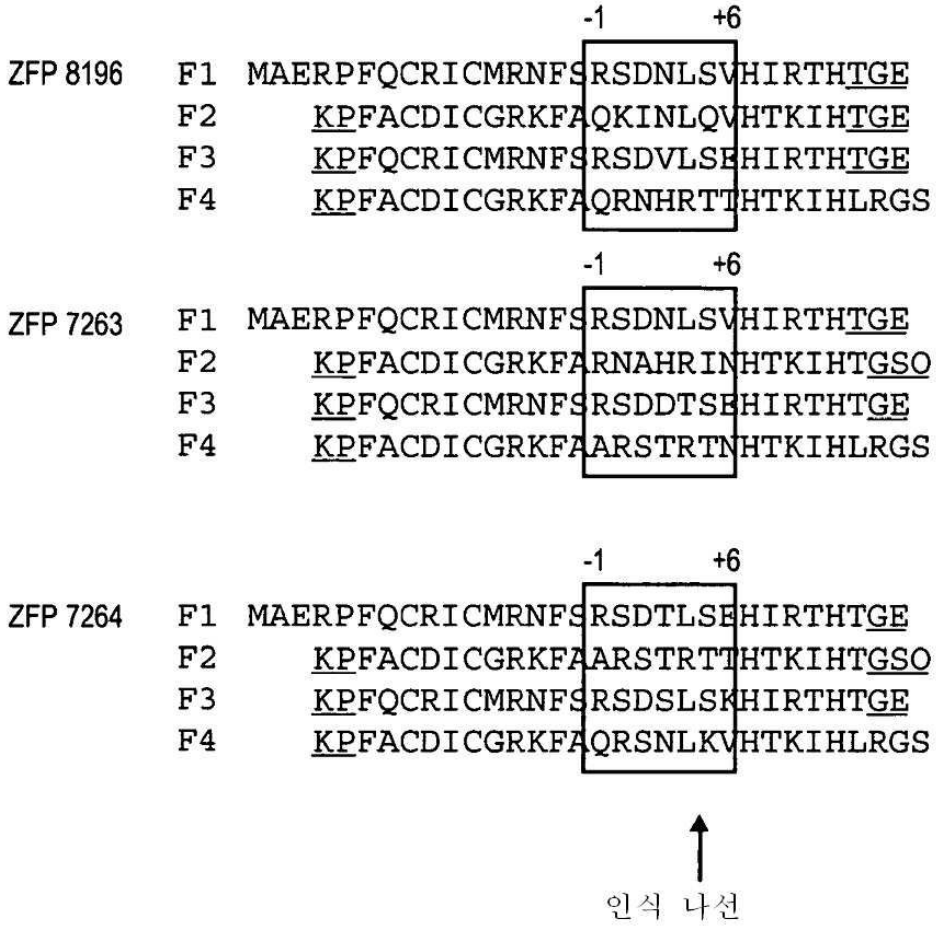
[0152] 본 명세서에서 언급한 모든 특허, 특허 출원 및 간행물은 본 명세서에 그의 전문이 참조로 포함된다.

[0153] 명세서가 이해의 명확함의 목적을 위해 예증 및 예시의 방법으로 일부 상세하게 제공되었지만, 당업자는 본 명세서의 정신 또는 범주로부터 벗어나지 않고 다양한 변화 및 변형이 실행될 수 있다는 것을 인식할 것이다. 따라서, 앞서 설명한 설명 및 실시예는 제한으로서 해석되어서는 안 된다.

도면

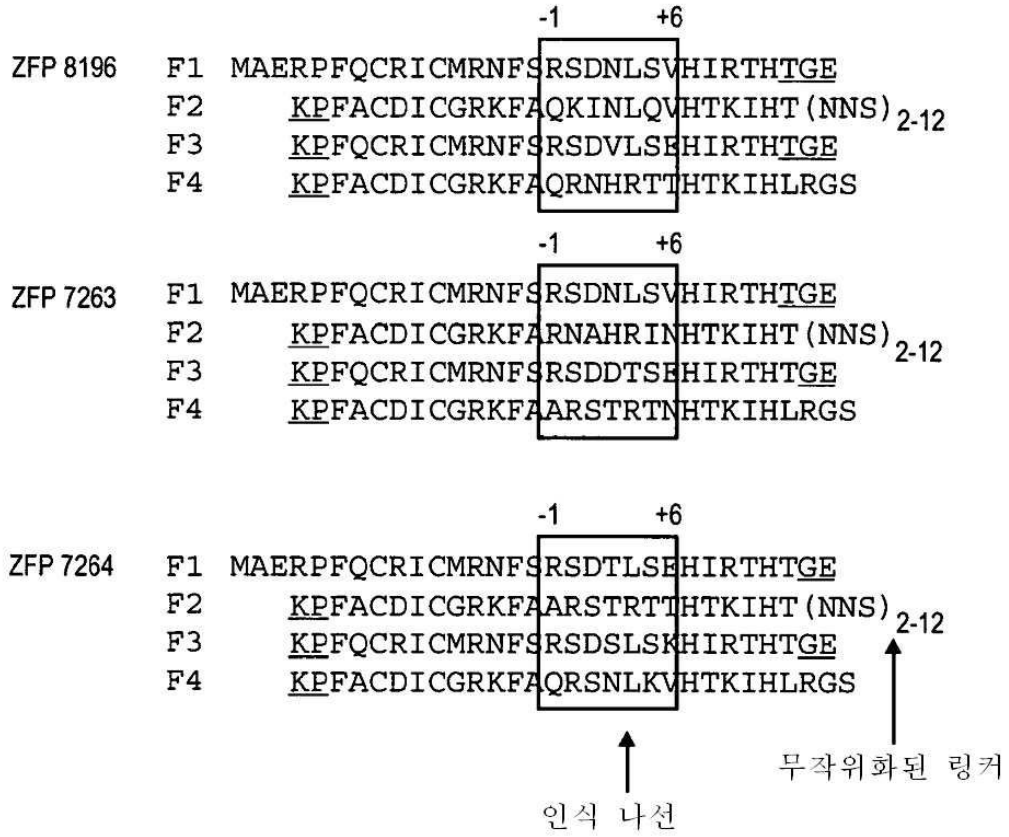
도면1a

선택을 위해 사용한 숙주 ZFP

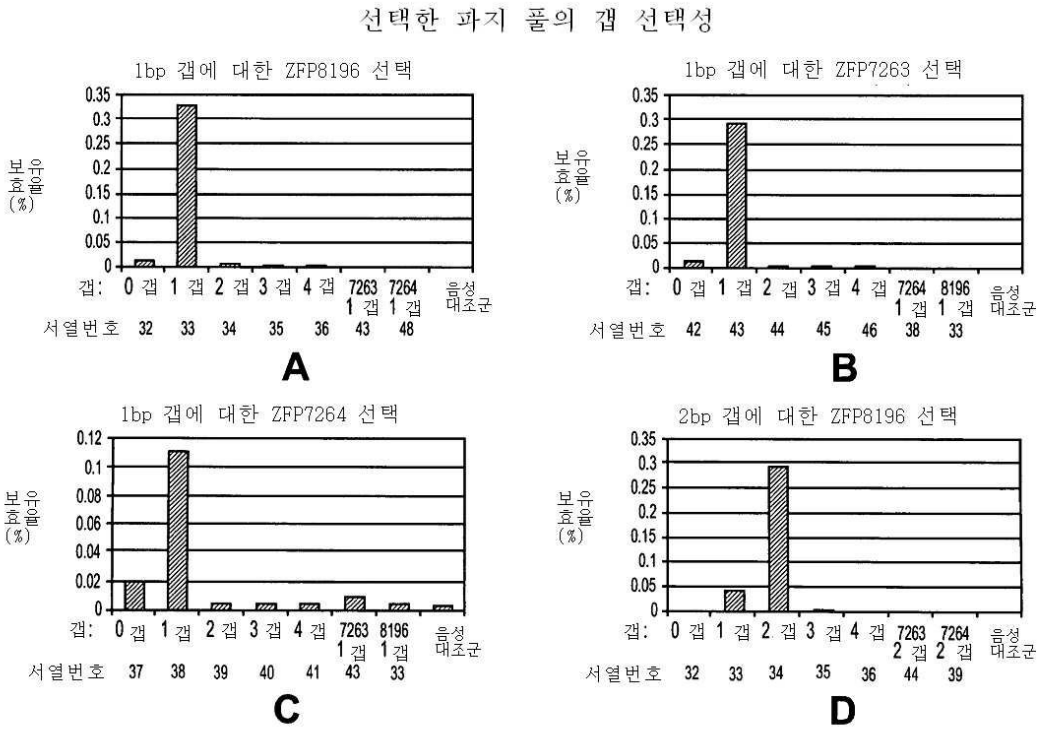


도면1b

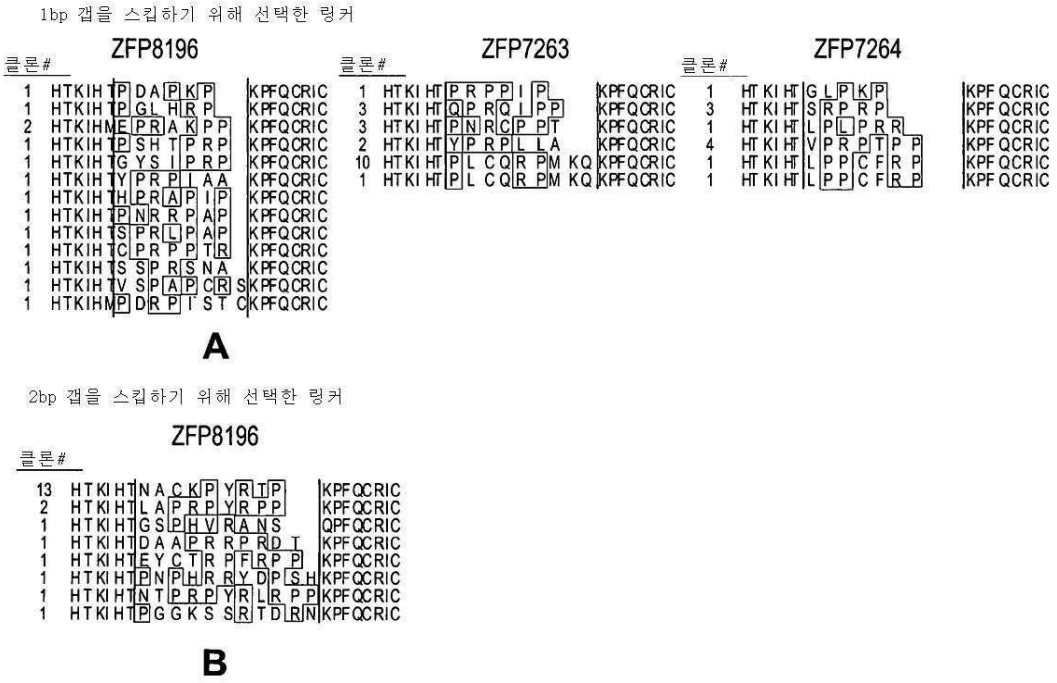
선택을 위해 사용한 링커 라이브러리 설계



도면2

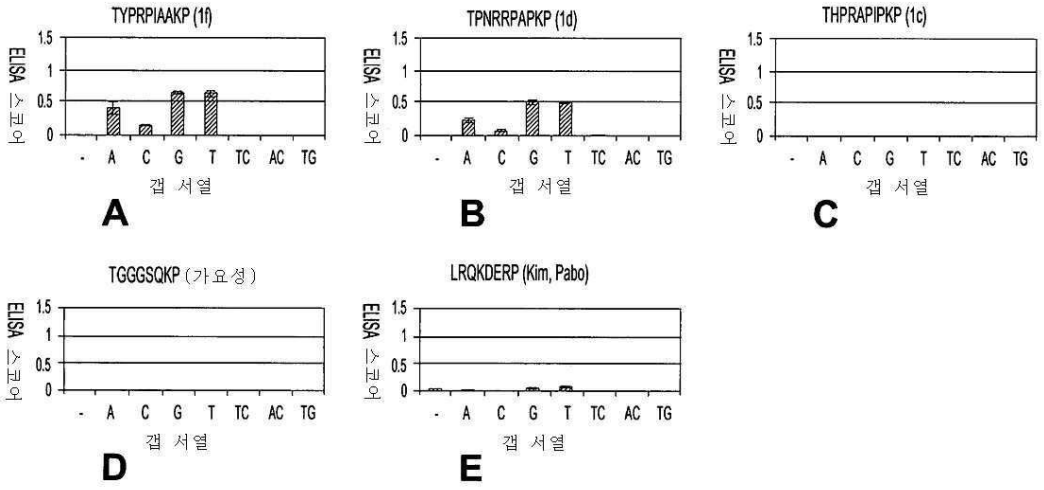

선택한 파지 풀의 갭 선택성
1bp 갭에 대한 ZFP8196 선택
보유율 (%)
갭: 0 갭 1 갭 2 갭 3 갭 4 갭 7263 1 갭 7264 1 갭 음성 대조군
서열번호 32 33 34 35 36 43 48
A
1bp 갭에 대한 ZFP7263 선택
보유율 (%)
갭: 0 갭 1 갭 2 갭 3 갭 4 갭 7264 1 갭 8196 1 갭 음성 대조군
서열번호 42 43 44 45 46 38 33
B
1bp 갭에 대한 ZFP7264 선택
보유율 (%)
갭: 0 갭 1 갭 2 갭 3 갭 4 갭 7263 1 갭 8196 1 갭 음성 대조군
서열번호 37 38 39 40 41 43 33
C
2bp 갭에 대한 ZFP8196 선택
보유율 (%)
갭: 0 갭 1 갭 2 갭 3 갭 4 갭 7263 2 갭 7264 2 갭 음성 대조군
서열번호 32 33 34 35 36 44 39
D

도면3


1bp 갭을 스킵하기 위해 선택한 링커
ZFP8196
클론#
1 HTKIHTPDAPKP KPFQCRIC
1 HTKIHTPGLHRP KPFQCRIC
2 HTKIHMEPRAKPP KPFQCRIC
1 HTKIHTPSHTPRP KPFQCRIC
1 HTKIHTGYSIPRP KPFQCRIC
1 HTKIHTYPRPIAA KPFQCRIC
1 HTKIHTHLRAPIP KPFQCRIC
1 HTKIHTPNRRPAP KPFQCRIC
1 HTKIHTSPRLPAP KPFQCRIC
1 HTKIHTCPRPPTR KPFQCRIC
1 HTKIHTSSPRSNA KPFQCRIC
1 HTKIHTVSPAPCRS KPFQCRIC
1 HTKIHMPDRPTSTC KPFQCRIC
A
2bp 갭을 스킵하기 위해 선택한 링커
ZFP8196
클론#
13 HTKIHTNACKPYRTP KPFQCRIC
2 HTKIHTLAPRPYRPP KPFQCRIC
1 HTKIHTGSPHVRANS QPFQCRIC
1 HTKIHTDAAPRRPRDI KPFQCRIC
1 HTKIHTEYCTRPFRPP KPFQCRIC
1 HTKIHTPNPHRRYDPSH KPFQCRIC
1 HTKIHTNTPRPYRLRPP KPFQCRIC
1 HTKIHTPGGKSSRTDLRN KPFQCRIC
B
ZFP7263
클론#
1 HTKIHTPRPPIP KPFQCRIC
3 HTKIHTQPRQIPP KPFQCRIC
3 HTKIHTPNRCPPT KPFQCRIC
2 HTKIHTYPRPLLA KPFQCRIC
10 HTKIHTPLCQRPMKQ KPFQCRIC
1 HTKIHTPLCQRPMKQ KPFQCRIC
ZFP7264
클론#
1 HTKIHTGLPKP KPFQCRIC
3 HTKIHTSRPRP KPFQCRIC
1 HTKIHTLPLPRR KPFQCRIC
4 HTKIHTVPRPTPP KPFQCRIC
1 HTKIHTLPPCFRP KPFQCRIC
1 HTKIHTLPPCFRP KPFQCRIC

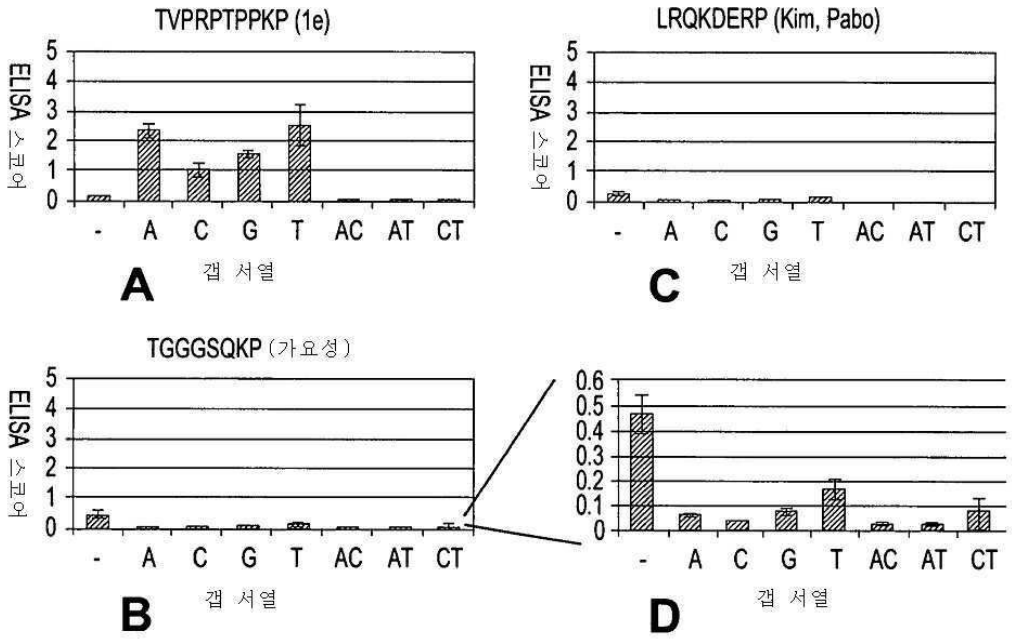
도면4

ZFP8196 내 하나의 염기쌍을 스킵하기 위해 선택한 링커의 갭 선택성



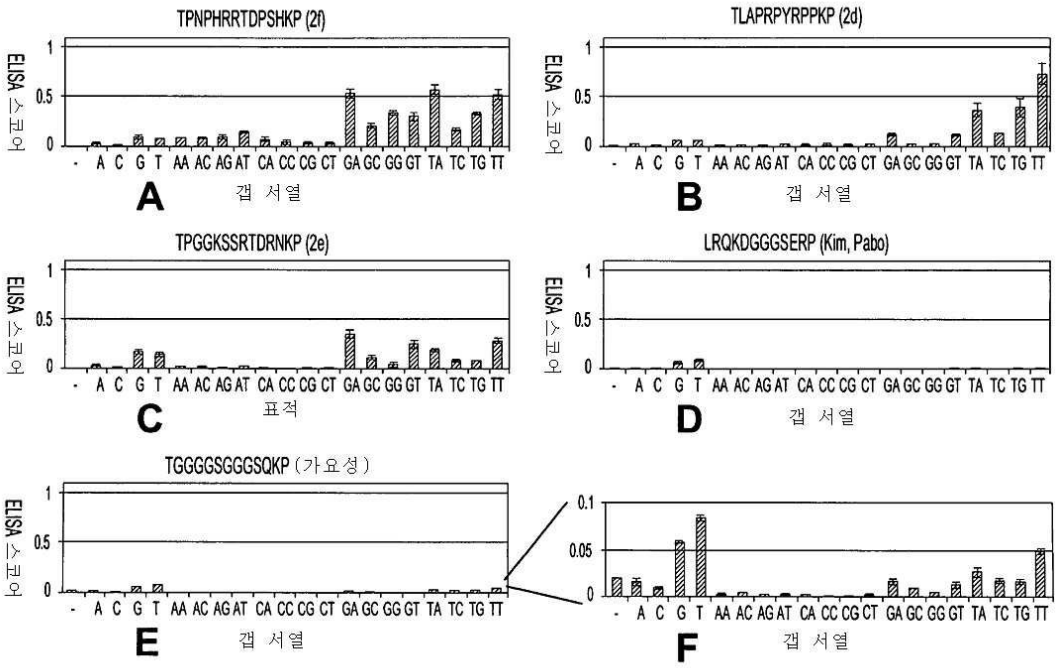
도면5

ZFP7264 내 하나의 염기쌍을 스킵하기 위해 선택한 링커의 갭 선택성



도면6

ZFP8196 내 2개 염기쌍을 스킵하기 위하여 선택한 링커의 갭 선택성



도면7

하나의 염기쌍을 스킵하기 위해 선택한 링커에 의한 호환성 연구

A

링커		ELISA 스코어											
서열	지정	ZFP1	ZFP2	ZFP3	ZFP4	ZFP5	ZFP6	ZFP7	ZFP8	ZFP9	ZFP10	ZFP11	ZFP12
TGGGGSQKP	가요성	0.75	1.08	0.21	0.26	0.08	0.05	0.05	0.05	0.05	0.05	0.05	0.05
TVPRPTPPKP	1e	1.19	1.30	1.05	1.10	0.75	0.29	0.48	0.38	0.50	0.18	0.05	0.05
TYPRPIAAKP	1f	1.25	1.35	1.03	0.94	0.96	0.30	0.33	0.12	0.30	0.11	0.05	0.05
TPNRRPAPKP	1d	1.20	1.08	1.04	0.83	0.85	0.25	0.39	0.28	0.24	0.12	0.05	0.05
THPRAPIPKP	1c	0.92	1.18	0.66	0.56	0.91	0.23	0.21	0.26	0.18	0.05	0.05	0.05

B

링커		가요성 링커에 대해 정규화된 ELISA 스코어												
서열	지정	ZFP1	ZFP2	ZFP3	ZFP4	ZFP5	ZFP6	ZFP7	ZFP8	ZFP9	ZFP10	ZFP11	ZFP12	평균
TGGGGSQKP	가요성	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
TVPRPTPPKP	1e	1.6	1.2	4.9	4.1	9.2	5.9	9.6	7.6	10.1	3.6	1.0	1.0	5.0
TYPRPIAAKP	1f	1.7	1.2	4.9	3.5	11.7	5.9	6.5	2.3	6.0	2.2	1.0	1.0	4.0
TPNRRPAPKP	1d	1.6	1.0	4.9	3.1	10.4	4.9	7.8	5.7	4.8	2.4	1.0	1.0	4.0
THPRAPIPKP	1c	1.2	1.1	3.1	2.1	11.2	4.7	4.3	5.3	3.6	1.0	1.0	1.0	3.3

밑줄친 값은 4배 초과의 개선을 나타낸다

도면8

두 개의 염기쌍을 스킵하기 위해 선택한 링커에 의한 호환성 연구

A

링커		ELISA 스코어					
서열	지정	ZFP13	ZFP14	ZFP15	ZFP16	ZFP17	ZFP18
TGGGGSGGSQKP	가요성	0.25	0.61	0.05	0.13	0.35	0.51
TPNPHRRTDPSHKP	2f	0.48	1.13	0.27	0.31	0.79	0.47
TLAPRPYRPPKP	2d	0.44	1.13	0.11	0.34	0.82	0.43
TPGGKSSRTDRNKP	2e	0.42	0.75	0.17	0.27	0.68	0.52

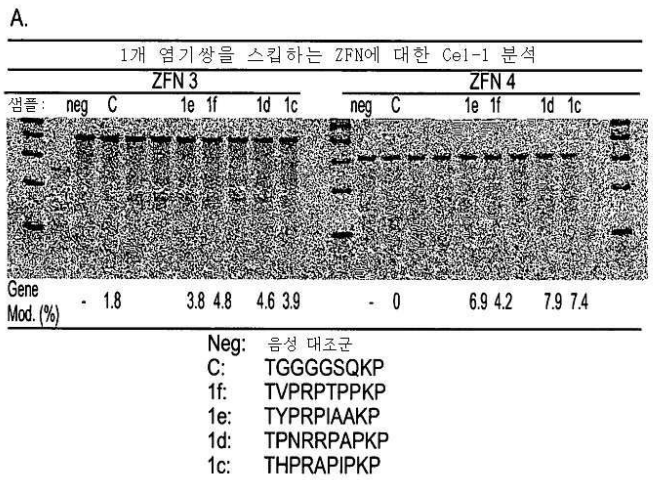
B

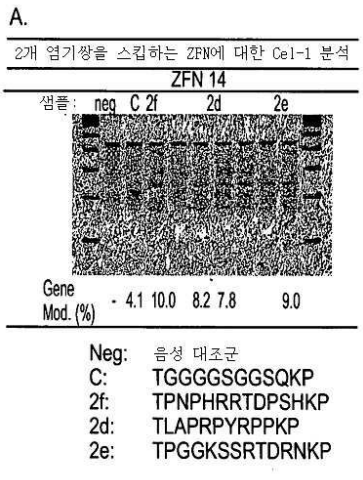
링커		가요성 링커에 대해 정규화된 ELISA 스코어						
서열	지정	ZFP13	ZFP14	ZFP15	ZFP16	ZFP17	ZFP18	평균
TGGGGSGGSQKP	가요성	1.0	1.0	1.0	1.0	1.0	1.0	
TPNPHRRTDPSHKP	2f	1.9	1.9	5.1	2.3	2.3	0.9	2.4
TLAPRPYRPPKP	2d	1.7	1.9	2.1	2.5	2.4	0.8	1.9
TPGGKSSRTDRNKP	2e	1.7	1.2	3.3	2.0	2.0	1.0	1.9

밑줄친 값은 2배 초과의 개선을 나타낸다

도면9

예시적 분석 겔





도면10

하나의 염기쌍을 스킵하는 새로운 링커를 함유하는 ZFN에 대한 유전자 변형

A

링커		유전자 변형(%)											
서열	지정	ZFN1	ZFN2	ZFN3	ZFN4	ZFN5	ZFN6	ZFN7	ZFN8	ZFN9	ZFN10	ZFN11	ZFN12
TGGGGSQKP	가요성	6.1	1.8	5.9	0.0	3.5	-	1.1	-	0.0	0.0	-	1.7
TVPRPTPPKP	1e	9.5	3.8	9.3	1.0	4.9	-	2.6	-	3.7	6.9	-	1.9
TYPRPIAAKP	1f	9.6	4.8	12.0	1.0	2.5	-	1.5	-	1.5	4.2	-	2.8
TPNRRPAPKP	1d	6.4	4.6	12.5	1.5	6.2	-	2.6	-	4.5	7.9	-	1.9
THPRAPIPKP	1c	9.9	3.9	12.8	1.4	5.1	-	2.7	-	NA	7.4	-	NA

B

링커		가요성 링커에 대해 정규화된 유전자 변형												
서열	지정	ZFN1	ZFN2	ZFN3	ZFN4	ZFN5	ZFN6	ZFN7	ZFN8	ZFN9	ZFN10	ZFN11	ZFN12	평균
TGGGGSQKP	가요성	1.0	1.0	1.0	-	1.0	-	1.0	-	-	-	-	1.0	
TVPRPTPPKP	1e	1.6	2.1	1.6	>1	1.4	-	2.4	-	>3.5	>7.0	-	1.1	2.4
TYPRPIAAKP	1f	1.6	2.7	2.0	>1	0.7	-	1.4	-	>1.5	>4.0	-	1.6	1.8
TPNRRPAPKP	1d	1.0	2.6	2.1	>1.5	1.8	-	2.4	-	>4.5	>8.0	-	1.1	2.8
THPRAPIPKP	1c	1.6	2.2	2.2	>1.5	1.5	-	2.5	-	NA	>7.0	-	NA	2.6

고발현 조건

밑줄친 값은 2배초과의 개선을 나타낸다

- 정량적인 신호 없음
NA 시험하지 않음

도면11

두 개의 염기쌍을 스킵하는 새로운 링커를 함유하는 ZFN에 대한 유전자 변형

A

링커		유전자 변형(%)					
서열	지정	ZFN13	ZFN14	ZFN15	ZFN16	ZFN17	ZFN18
TGGGGSGGSQKP	가요성	-	7.5	2.5	2.9	5.3	-
TPNPHRRTDPSHKP	2f	-	17.0	5.6	7.3	4.7	-
TLAPRPYRPPKP	2d	-	11.3	5.0	4.8	4.9	-
TPGGKSSRTDRNKP	2e	-	11.7	5.0	4.3	6.3	-

B

링커		가요성 링커에 대해 정규화된 유전자 변형						
서열	지정	ZFN13	ZFN14	ZFN15	ZFN16	ZFN17	ZFN18	평균
TGGGGSGGSQKP	가요성	-	1.0	1.0	1.0	1.0	-	
TPNPHRRTDPSHKP	2f	-	2.3	2.2	2.5	0.9	-	2.0
TLAPRPYRPPKP	2d	-	1.5	2.0	1.7	0.9	-	1.5
TPGGKSSRTDRNKP	2e	-	1.6	2.0	1.5	1.2	-	1.6

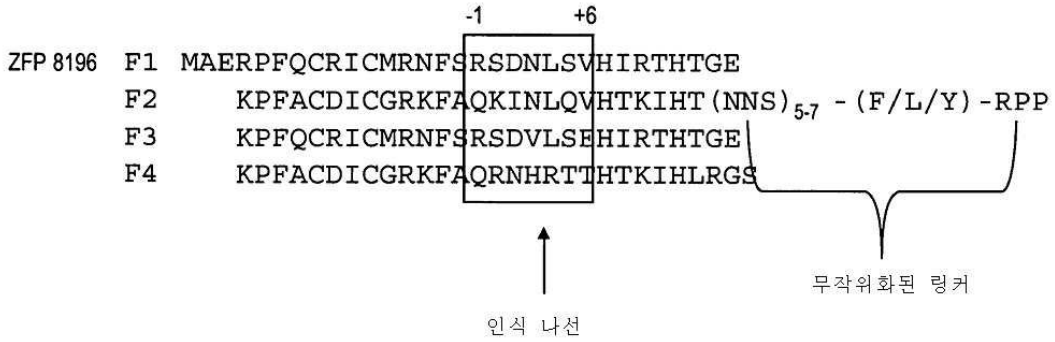
고발현 조건

밑줄친 값은 2배초과의 개선을 나타낸다

-- 정량적인 신호 없음

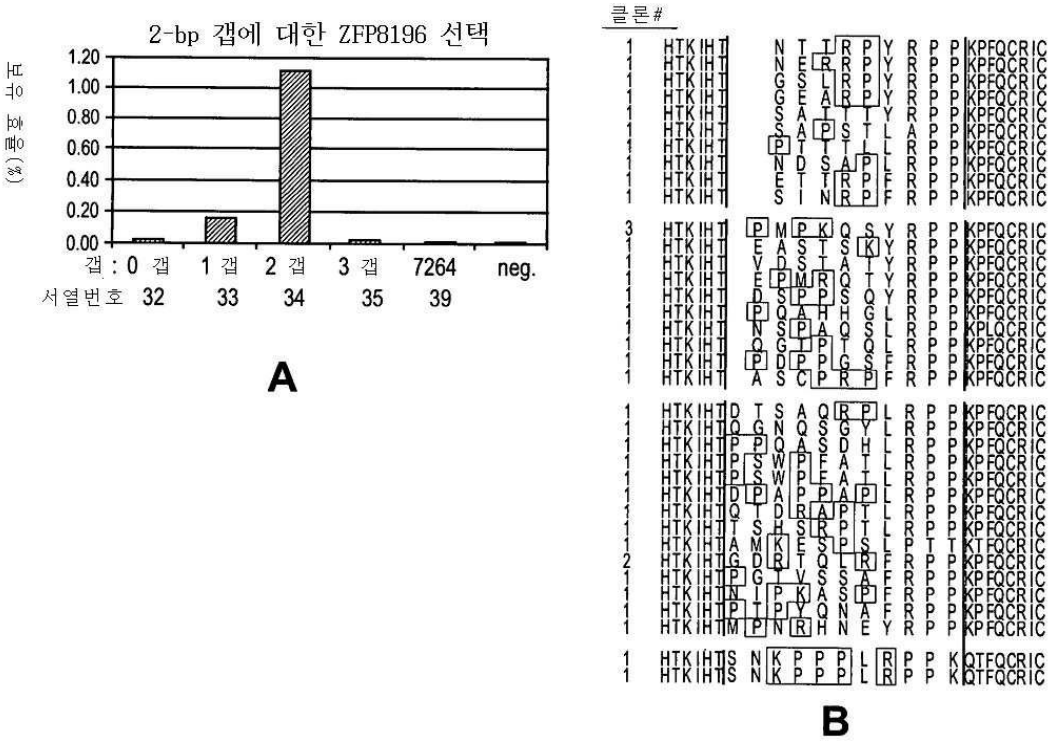
도면12

2-bp 갭에 대한 2차적 선택을 위해
사용한 링커 라이브러리 설계



도면13

2-bp 갭 및 링커 서열에 대한 2차적 선택으로부터
선택한 파지 풀의 갭 선택성



서 열 목 록

SEQUENCE LISTING

<110> SANGAMO BIOSCIENCES, INC.

<120> COMPOSITIONS FOR LINKING ZINC FINGER MODULES

<130> 8325-0071.40

<140> PCT/US2011/000758

<141> 2011-05-02

<150> 61/343,729

<151> 2010-05-03

<160> 213

<170> PatentIn version 3.5

<210> 1

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 1

Thr Gly Glu Lys Pro
1 5

<210> 2

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 2

Thr Gly Gly Gly Gly Ser Gln Lys Pro
1 5

<210> 3

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 3

Leu Arg Gln Lys Asp Glu Arg Pro
1 5

<210> 4

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 4

Thr Pro Arg Pro Pro Ile Pro Lys Pro

1 5

<210> 5

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> misc_feature

<222> (7)..(26)

<223> This region may encompass any of SEQ ID NOS 32-46

<400> 5

tataatnnnn nnnnnnnnnn nnnnnnttca cagtcagtcc acacgtc 47

<210> 6

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> misc_feature

<222> (5)..(22)

<223> This region may encompass any of SEQ ID NOS 32, 73-76, or 113-128

<400> 6

ttagnnnnnn nnnnnnnnnn nntatc 26

<210> 7

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 7

Thr Tyr Pro Arg Pro Leu Leu Ala Lys Pro

1 5 10

<210> 8

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 8

Thr Pro Leu Cys Gln Arg Pro Met Lys Gln Lys Pro

1 5 10

<210> 9

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 9

Thr Gly Ser Leu Arg Pro Tyr Arg Arg Pro Lys Pro

1 5 10

<210> 10

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 10

Thr Ser Ile Asn Arg Pro Phe Arg Arg Pro Lys Pro

1 5 10

<210> 11

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 11

Thr Ser Arg Pro Arg Pro Lys Pro

1 5

<210> 12

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 12

Thr Val Pro Arg Pro Thr Pro Pro Lys Pro

1 5 10

<210> 13

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 13

Thr Lys His Gly Thr Pro Lys His Arg Glu Asp Lys Pro

1 5 10

<210> 14

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 14

Thr Gly Gly Gly Gly Ser Gly Gly Gly Ser Gln Lys Pro

1 5 10

<210> 15

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 15

Thr Gly Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Gln Lys Pro

1 5 10 15

<210> 16

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 16

Thr Gly Gly Glu Lys Pro

1 5

<210> 17

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 17

Thr Gly Gly Gln Lys Pro

1 5

<210> 18

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 18

Thr Gly Gly Ser Gly Glu Lys Pro

1 5

<210> 19

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 19

Thr Gly Gly Ser Gly Gln Lys Pro

1 5

<210> 20

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 20

Thr Gly Gly Ser Gly Gly Ser Gly Glu Lys Pro

1 5 10

<210> 21

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 21

Thr Gly Gly Ser Gly Gly Ser Gly Gln Lys Pro

1 5 10

<210> 22

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 22

Arg Ser Asp Asn Leu Ser Val

1 5

<210> 23

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 23

Gln Lys Ile Asn Leu Gln Val

1 5

<210> 24

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 24

Arg Ser Asp Val Leu Ser Glu

1 5

<210> 25

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 25

Gln Arg Asn His Arg Thr Thr

1 5

<210> 26

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 26

Arg Ser Asp Thr Leu Ser Glu

1 5

<210> 27

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 27

Ala Arg Ser Thr Arg Thr Thr

1 5

<210> 28

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 28

Arg Ser Asp Ser Leu Ser Lys

1 5

<210> 29

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 29

Gln Arg Ser Asn Leu Lys Val

1 5

<210> 30

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 30

Arg Asn Ala His Arg Ile Asn

1 5

<210> 31

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 31

Ala Arg Ser Thr Arg Thr Asn

1 5

<210> 32

<211> 16

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 32

ataaactgca aaaggc 16

<210> 33

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 33

ataaactgdc aaaaggc 17

<210> 34

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 34

ataaactgdb caaaaggc 18

<210> 35

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 35

ataaactgdb bcaaaaggc 19

<210> 36

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 36

ataaactgdb bbcaaaaggc 20

<210> 37

<211> 16

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 37

ttaaagcggc tccgaa 16

<210> 38

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 38

ttaaagcghg ctccgaa 17

<210> 39

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 39

ttaaagcghd gctccgaa 18

<210> 40

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 40

ttaaagcghd vgctccgaa 19

<210> 41

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 41

ttaaagcghd vdgctccgaa 20

<210> 42

<211> 16

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 42

ccactctgtg gaagtg 16

<210> 43

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 43

ccactctght ggaagtg 17

<210> 44

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 44

ccactctghh tggaagtg 18

<210> 45

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 45

ccactctghh htggaagtg 19

<210> 46

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 46

ccactctghh hbtggaagtg 20

<210> 47

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MOD_RES

<222> (27)..(38)

<223> Any amino acid and this region may encompass 2-12 residues

<400> 47

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Gln Lys Ile

1 5 10 15

Asn Leu Gln Val His Thr Lys Ile His Thr Xaa Xaa Xaa Xaa Xaa Xaa

20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa

35

<210> 48

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 48

Thr Gly Glu Gly Gly Lys Pro

1 5

<210> 49

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 49

Thr Pro Asp Ala Pro Lys Pro Lys Pro

1 5

<210> 50

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 50

Thr Pro Gly Leu His Arg Pro Lys Pro

1 5

<210> 51

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 51

Thr Glu Pro Arg Ala Lys Pro Pro Lys Pro

1 5 10

<210> 52

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 52

Thr Pro Ser His Thr Pro Arg Pro Lys Pro

1 5 10

<210> 53

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 53

Thr Gly Tyr Ser Ile Pro Arg Pro Lys Pro

1 5 10

<210> 54

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 54

Thr Tyr Pro Arg Pro Ile Ala Ala Lys Pro

1 5 10

<210> 55

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 55

Thr His Pro Arg Ala Pro Ile Pro Lys Pro

1 5 10

<210

> 56

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 56

Thr Pro Asn Arg Arg Pro Ala Pro Lys Pro
1 5 10

<210> 57

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 57

Thr Ser Pro Arg Leu Pro Ala Pro Lys Pro
1 5 10

<210> 58

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 58

Thr Cys Pro Arg Pro Pro Thr Arg Lys Pro
1 5 10

<210> 59

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 59

Thr Ser Ser Pro Arg Ser Asn Ala Lys Pro
1 5 10

<210> 60

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 60

Thr Val Ser Pro Ala Pro Cys Arg Ser Lys Pro

1 5 10

<210> 61

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 61

Thr Pro Asp Arg Pro Ile Ser Thr Cys Lys Pro

1 5 10

<210> 62

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 62

Thr Gln Arg Pro Gln Ile Pro Pro Lys Pro

1 5 10

<210> 63

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 63

Thr Pro Asn Arg Cys Pro Pro Thr Lys Pro

1 5 10

<210> 64

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 64

Thr Gly Leu Pro Lys Pro Lys Pro

1 5

<210> 65

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 65

Thr Leu Pro Leu Pro Arg Pro Lys Pro

1 5

<210> 66

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 66

Thr Leu Pro Pro Cys Phe Arg Pro Lys Pro

1 5 10

<210> 67

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> misc_feature

<222> (7)..(24)

<223> This region may encompass any of SEQ ID NOS 33, 34, 38, 39, 43, or 44

<400> 67

tataatnnnn nnnnnnnnnn nnnnttcaca gtcagtccac acgtc 45

<210> 68

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 68

Leu Arg Gln Lys Asp Gly Gly Gly Ser Glu Arg Pro
1 5 10

<210> 69

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 69

Thr Pro Asn Pro His Arg Arg Thr Asp Pro Ser His Lys Pro
1 5 10

<210> 70

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 70

Thr Leu Ala Pro Arg Pro Tyr Arg Pro Pro Lys Pro

1 5 10

<210> 71

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 71

Thr Pro Gly Gly Lys Ser Ser Arg Thr Asp Arg Asn Lys Pro

1 5 10

<210> 72

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 72

Thr Gly Ser Gln Lys Pro

1 5

<210> 73

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 73

ataaactgac aaaaggc 17

<210> 74

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 74

ataaactgcc aaaaggc 17

<210> 75

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 75

ataaactggc aaaaggc 17

<210> 76

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 76

ataaactgtc aaaaggc 17

<210> 77

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 77

ataaactgtc caaaaggc 18

<210> 78

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 78

ataaactgac caaaaggc 18

<210> 79

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 79

ataaactgtg caaaaggc 18

<210> 80

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 80

ttaaagcgag ctccgaa 17

<210> 81

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 81

ttaaagcgcg ctccgaa 17

<210> 82

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 82

ttaaagcggg ctccgaa 17

<210> 83

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 83

ttaaagcgtg ctccgaa 17

<210> 84

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 84

ttaaagcgtt gctccgaa 18

<210> 85

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 85

ttaaagcgta gctccgaa 18

<210> 86

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 86

ttaaagcgct gctccgaa 18

<210> 87

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 87

ccactctgat ggaagtg 17

<210> 88

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 88

ccactctgct ggaagtg 17

<210> 89

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 89

ccactctggt ggaagtg 17

<210> 90

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 90

ccactctgtt ggaagtg 17

<210> 91

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 91

ccactctgac tggaagtg 18

<210> 92

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 92

ccactctgat tggaagtg 18

<210> 93

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 93

ccactctgct tggaagtg 18

<210> 94

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> misc_feature

<222> (5)..(22)

<223> This region may encompass any of SEQ ID NOS 32, 37, 42,
or 73-93

<400> 94

ttagnnnnnn nnnnnnnnnn nntatc 26

<210> 95

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 95

actctgtgga ag 12

<210> 96

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 96

actctgatgg aag 13

<210> 97

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 97

actctgctgg aag 13

<210> 98

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 98

actctggtgg aag 13

<210> 99

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 99

actctgttgg aag 13

<210> 100

<211> 14

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 100

actctgactg gaag 14

<210> 101

<211> 14

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 101

actctgattg gaag 14

<210> 102

<211> 14

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 102

actctgcttg gaag 14

<210> 103

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> misc_feature

<222> (26)..(39)

<223> This region may encompass any of SEQ ID NOS 95-102

<400> 103

gatctgttcg gagccgcttt aacccnnnnn nnnnnnnnnt gctcgcg 47

<210> 104

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 104

aaagcggctc cg 12

<210> 105

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 105

aaagcgagct ccg 13

<210> 106

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 106

aaagcgcgct ccg 13

<210> 107

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 107

aaagcgggct ccg 13

<210> 108

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 108

aaagcgtgct ccg 13

<210> 109

<211> 14

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 109

aaagcgttgc tccg 14

<210> 110

<211> 14

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 110

aaagcgtagc tccg 14

<210> 111

<211> 14

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 111

aaagcgctgc tccg 14

<210> 112

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> misc_feature

<222> (9)..(22)

<223> This region may encompass any of SEQ ID NOS 104-111

<400> 112

gatctgttnn nnnnnnnnnn nnaacccact ctgtggaagt gctcgcg 47

<210> 113

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 113

ataaactgaa caaaaggc 18

<210> 114

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 114

ataaactgac caaaaggc 18

<210> 115

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 115

ataaactgag caaaaggc 18

<210> 116

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 116

ataaactgat caaaaggc 18

<210> 117

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 117

ataaactgca caaaaggc 18

<210> 118

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 118

ataaactgcc caaaaggc 18

<210> 119

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 119

ataaactgcg caaaaggc 18

<210> 120

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 120

ataaactgct caaaaggc 18

<210> 121

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 121

ataaactgga caaaaggc 18

<210> 122

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 122

ataaactggc caaaaggc 18

<210> 123

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 123

ataaactggg caaaaggc 18

<210> 124

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 124

ataaactggt caaaaggc 18

<210> 125

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 125

ataaactgta caaaaggc 18

<210> 126

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 126

ataaactgtc caaaaggc 18

<210> 127

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 127

ataaactgtg caaaaggc 18

<210> 128

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 128

ataaactgtt caaaaggc 18

<210> 129

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<220><221> MOD_RES

<222> (3)..(3)

<223> Glu or Gln

<220><221> MOD_RES

<222> (4)..(4)

<223> Lys or Arg

<400> 129

Thr Gly Xaa Xaa Pro

1 5

<210> 130

<211> 31

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 130

Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
1 5 10 15

Arg Ser Asp Asn Leu Ser Val His Ile Arg Thr His Thr Gly Glu
20 25 30

<210> 131

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 131

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Gln Lys Ile
1 5 10 15

Asn Leu Gln Val His Thr Lys Ile His Thr Gly Glu
20 25

<210> 132

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 132

Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp
1 5 10 15

Val Leu Ser Glu His Ile Arg Thr His Thr Gly Glu
20 25

<210> 133

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 133

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Gln Arg Asn
1 5 10 15

His Arg Thr Thr His Thr Lys Ile His Leu Arg Gly Ser
20 25

<210> 134

<211> 31

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 134

Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
1 5 10 15

Arg Ser Asp Asn Leu Ser Val His Ile Arg Thr His Thr Gly Glu
20 25 30

<210> 135

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 135

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Asn Ala
1 5 10 15

His Arg Ile Asn His Thr Lys Ile His Thr Gly Ser Gln
20 25

<210> 136

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 136

Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp
1 5 10 15

Asp Thr Ser Glu His Ile Arg Thr His Thr Gly Glu
20 25

<210> 137

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 137

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ala Arg Ser
1 5 10 15

Thr Arg Thr Asn His Thr Lys Ile His Leu Arg Gly Ser
20 25

<210> 138

<211> 31

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 138

Met Ala Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
1 5 10 15

Arg Ser Asp Thr Leu Ser Glu His Ile Arg Thr His Thr Gly Glu
20 25 30

<210> 139

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 139

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ala Arg Ser
1 5 10 15

Thr Arg Thr Thr His Thr Lys Ile His Thr Gly Ser Gln
20 25

<210> 140

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 140

Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp
1 5 10 15

Ser Leu Ser Lys His Ile Arg Thr His Thr Gly Glu
20 25

<210> 141

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 141

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Gln Arg Ser
1 5 10 15

Asn Leu Lys Val His Thr Lys Ile His Leu Arg Gly Ser
20 25

<210> 142

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 142

His Thr Lys Ile His Thr Pro Asp Ala Pro Lys Pro Lys Pro Phe Gln
1 5 10 15

Cys Arg Ile Cys
20

<210> 143

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 143

His Thr Lys Ile His Thr Pro Gly Leu His Arg Pro Lys Pro Phe Gln
1 5 10 15

Cys Arg Ile Cys
20

<210> 144

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 144

His Thr Lys Ile His Met Glu Pro Arg Ala Lys Pro Pro Lys Pro Phe
1 5 10 15

Gln Cys Arg Ile Cys
20

<210> 145

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 145

His Thr Lys Ile His Thr Pro Ser His Thr Pro Arg Pro Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 146

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 146

His Thr Lys Ile His Thr Gly Tyr Ser Ile Pro Arg Pro Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 147

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 147

His Thr Lys Ile His Thr Tyr Pro Arg Pro Ile Ala Ala Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 148

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 148

His Thr Lys Ile His Thr His Pro Arg Ala Pro Ile Pro Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 149

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 149

His Thr Lys Ile His Thr Pro Asn Arg Arg Pro Ala Pro Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 150

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223>

Description of Artificial Sequence: Synthetic

peptide

<400> 150

His Thr Lys Ile His Thr Ser Pro Arg Leu Pro Ala Pro Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 151

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 151

His Thr Lys Ile His Thr Cys Pro Arg Pro Pro Thr Arg Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 152

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 152

His Thr Lys Ile His Thr Ser Ser Pro Arg Ser Asn Ala Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 153

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400

> 153

His Thr Lys Ile His Thr Val Ser Pro Ala Pro Cys Arg Ser Lys Pro

1 5 10 15

Phe Gln Cys Arg Ile Cys

20

<210> 154

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 154

His Thr Lys Ile His Met Pro Asp Arg Pro Ile Ser Thr Cys Lys Pro

1 5 10 15

Phe Gln Cys Arg Ile Cys

20

<210> 155

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 155

His Thr Lys Ile His Thr Pro Arg Pro Pro Ile Pro Lys Pro Phe Gln

1 5 10 15

Cys Arg Ile Cys

20

<210> 156

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 156

His Thr Lys Ile His Thr Gln Pro Arg Gln Ile Pro Pro Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 157

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 157

His Thr Lys Ile His Thr Pro Asn Arg Cys Pro Pro Thr Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 158

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223>

Description of Artificial Sequence: Synthetic

peptide

<400> 158

His Thr Lys Ile His Thr Tyr Pro Arg Pro Leu Leu Ala Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 159

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 159

His Thr Lys Ile His Thr Pro Leu Cys Gln Arg Pro Met Lys Gln Lys

1 5 10 15

Pro Phe Gln Cys Arg Ile Cys

20

<210> 160

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 160

His Thr Lys Ile His Thr Pro Leu Cys Gln Arg Pro Met Lys Gln Lys

1 5 10 15

Pro Phe Gln Cys Arg Ile Cys

20

<210> 161

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 161

His Thr Lys Ile His Thr Gly Leu Pro Lys Pro Lys Pro Phe Gln Cys

1 5 10 15

Arg Ile Cys

<210> 162

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 162

His Thr Lys Ile His Thr Ser Arg Pro Arg Pro Lys Pro Phe Gln Cys

1 5 10 15

Arg Ile Cys

<210> 163

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 163

His Thr Lys Ile His Thr Leu Pro Leu Pro Arg Pro Lys Pro Phe Gln
1 5 10 15

Cys Arg Ile Cys
20

<210> 164

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 164

His Thr Lys Ile His Thr Val Pro Arg Pro Thr Pro Pro Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys
20

<210> 165

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 165

His Thr Lys Ile His Thr Leu Pro Pro Cys Phe Arg Pro Lys Pro Phe
1 5 10 15

Gln Cys Arg Ile Cys
20

<210> 166

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223>

Description of Artificial Sequence: Synthetic

peptide

<400> 166

His Thr Lys Ile His Thr Leu Pro Pro Cys Phe Arg Pro Lys Pro Phe

1 5 10 15

Gln Cys Arg Ile Cys

20

<210> 167

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 167

His Thr Lys Ile His Thr Asn Ala Cys Lys Pro Tyr Arg Thr Pro Lys

1 5 10 15

Pro Phe Gln Cys Arg Ile Cys

20

<210> 168

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 168

His Thr Lys Ile His Thr Leu Ala Pro Arg Pro Tyr Arg Pro Pro Lys

1 5 10 15

Pro Phe Gln Cys Arg Ile Cys

20

<210> 169

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 169

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

1 5 10 15

Pro Phe Gln Cys Arg Ile Cys

20

<210> 170

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 170

His Thr Lys Ile His Thr Asp Ala Ala Pro Arg Arg Pro Arg Asp Thr

1 5 10 15

Lys Pro Phe Gln Cys Arg Ile Cys

20

<210> 171

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 171

His Thr Lys Ile His Thr Glu Tyr Cys Thr Arg Pro Phe Arg Pro Pro

1 5 10 15

Lys Pro Phe Gln Cys Arg Ile Cys

20

<210> 172

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 172

His Thr Lys Ile His Thr Pro Asn Pro His Arg Arg Tyr Asp Pro Ser
1 5 10 15

His Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 173

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 173

His Thr Lys Ile His Thr Asn Thr Pro Arg Pro Tyr Arg Leu Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 174

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 174

His Thr Lys Ile His Thr Pro Gly Gly Lys Ser Ser Arg Thr Asp Arg
1 5 10 15

Asn Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 175

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 175

His Thr Lys Ile His Thr Asn Thr Thr Arg Pro Tyr Arg Pro Pro Lys

1 5 10 15

Pro Phe Gln Cys Arg Ile Cys

20

<210> 176

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 176

His Thr Lys Ile His Thr Asn Glu Arg Arg Pro Tyr Arg Pro Pro Lys

1 5 10 15

Pro Phe Gln Cys Arg Ile Cys

20

<210> 177

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 177

His Thr Lys Ile His Thr Gly Ser Leu Arg Pro Tyr Arg Pro Pro Lys

1 5 10 15

Pro Phe Gln Cys Arg Ile Cys

20

<210> 178

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 178
His Thr Lys Ile His Thr Gly Glu Ala Arg Pro Tyr Arg Pro Pro Lys
1 5 10 15
Pro Phe Gln Cys Arg Ile Cys
20

<210> 179
<211> 23
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 179
His Thr Lys Ile His Thr Ser Ala Thr Thr Thr Tyr Arg Pro Pro Lys
1 5 10 15
Pro Phe Gln Cys Arg Ile Cys

20

<210> 180
<211> 23
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 180
His Thr Lys Ile His Thr Ser Ala Pro Ser Thr Leu Ala Pro Pro Lys
1 5 10 15
Pro Phe Gln Cys Arg Ile Cys
20

<210> 181
<211> 23
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 181
His Thr Lys Ile His Thr Pro Thr Thr Thr Leu Leu Arg Pro Pro Lys
1 5 10 15
Pro Phe Gln Cys Arg Ile Cys
20

<210> 182
<211> 23
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 182
His Thr Lys Ile His Thr Asn Asp Ser Ala Pro Leu Arg Pro Pro Lys
1 5 10 15
Pro Phe Gln Cys Arg Ile Cys

20

<210> 183
<211> 23
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 183
His Thr Lys Ile His Thr Glu Thr Thr Arg Pro Phe Arg Pro Pro Lys
1 5 10 15
Pro Phe Gln Cys Arg Ile Cys
20

<210> 184
<211> 23
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 184
His Thr Lys Ile His Thr Ser Ile Asn Arg Pro Phe Arg Pro Pro Lys
1 5 10 15
Pro Phe Gln Cys Arg Ile Cys
20

<210> 185
<211> 24
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 185
His Thr Lys Ile His Thr Pro Met Pro Lys Gln Ser Tyr Arg Pro Pro
1 5 10 15
Lys Pro Phe Gln Cys Arg Ile Cys
20

<210> 186
<211> 24
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 186
His Thr Lys Ile His Thr Glu Ala Ser Thr Ser Lys Tyr Arg Pro Pro
1 5 10 15
Lys Pro Phe Gln Cys Arg Ile Cys
20

<210> 187
<211> 24
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 187

His Thr Lys Ile His Thr Val Asp Ser Thr Ala Thr Tyr Arg Pro Pro
1 5 10 15

Lys Pro Phe Gln Cys Arg Ile Cys
20

<210> 188

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 188

His Thr Lys Ile His Thr Glu Pro Met Arg Gln Thr Tyr Arg Pro Pro
1 5 10 15

Lys Pro Phe Gln Cys Arg Ile Cys
20

<210> 189

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 189

His Thr Lys Ile His Thr Asp Ser Pro Pro Ser Gln Tyr Arg Pro Pro
1 5 10 15

Lys Pro Phe Gln Cys Arg Ile Cys
20

<210> 190

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 190
His Thr Lys Ile His Thr Pro Gln Ala His His Gly Leu Arg Pro Pro
1 5 10 15
Lys Pro Phe Gln Cys Arg Ile Cys
20

<210> 191
<211> 24
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 191
His Thr Lys Ile His Thr Asn Ser Pro Ala Gln Ser Leu Arg Pro Pro
1 5 10 15
Lys Pro Leu Gln Cys Arg Ile Cys

20

<210> 192
<211> 24
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 192
His Thr Lys Ile His Thr Gln Gly Thr Pro Thr Gln Leu Arg Pro Pro
1 5 10 15
Lys Pro Phe Gln Cys Arg Ile Cys
20

<210> 193
<211> 24
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 193
His Thr Lys Ile His Thr Pro Asp Pro Pro Gly Ser Phe Arg Pro Pro
1 5 10 15
Lys Pro Phe Gln Cys Arg Ile Cys
20

<210> 194
<211> 24
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 194
His Thr Lys Ile His Thr Ala Ser Cys Pro Arg Pro Phe Arg Pro Pro
1 5 10 15
Lys Pro Phe Gln Cys Arg Ile Cys

20

<210> 195
<211> 25
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 195
His Thr Lys Ile His Thr Asp Thr Ser Ala Gln Arg Pro Leu Arg Pro
1 5 10 15
Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 196
<211> 25
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 196

His Thr Lys Ile His Thr Gln Gly Asn Gln Ser Gly Tyr Leu Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 197

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 197

His Thr Lys Ile His Thr Pro Pro Gln Ala Ser Asp His Leu Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 198

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 198

His Thr Lys Ile His Thr Pro Ser Trp Pro Phe Ala Thr Leu Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 199

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 199

His Thr Lys Ile His Thr Pro Ser Trp Pro Phe Ala Thr Leu Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 200

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 200

His Thr Lys Ile His Thr Asp Pro Ala Pro Pro Ala Pro Leu Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 201

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 201

His Thr Lys Ile His Thr Gln Thr Asp Arg Ala Pro Thr Leu Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 202

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 202

His Thr Lys Ile His Thr Thr Ser His Ser Arg Pro Thr Leu Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 203

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 203

His Thr Lys Ile His Thr Ala Met Lys Glu Ser Pro Ser Leu Pro Pro
1 5 10 15

Pro Lys Thr Phe Gln Cys Arg Ile Cys
20 25

<210> 204

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 204

His Thr Lys Ile His Thr Gly Asp Arg Thr Gln Leu Arg Phe Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 205

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 205

His Thr Lys Ile His Thr Pro Gly Thr Val Ser Ser Ala Phe Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 206

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 206

His Thr Lys Ile His Thr Asn Ile Pro Lys Ala Ser Pro Phe Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 207

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 207

His Thr Lys Ile His Thr Pro Thr Pro Tyr Gln Asn Ala Phe Arg Pro
1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys
20 25

<210> 208

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 208

His Thr Lys Ile His Thr Met Pro Asn Arg His Asn Glu Tyr Arg Pro

1 5 10 15

Pro Lys Pro Phe Gln Cys Arg Ile Cys

20 25

<210> 209

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 209

His Thr Lys Ile His Thr Ser Asn Lys Pro Pro Pro Leu Arg Pro Pro

1 5 10 15

Lys Gln Thr Phe Gln Cys Arg Ile Cys

20 25

<210> 210

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 210

His Thr Lys Ile His Thr Ser Asn Lys Pro Pro Pro Leu Arg Pro Pro

1 5 10 15

Lys Gln Thr Phe Gln Cys Arg Ile Cys

20 25

<210> 211

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<220><221> MOD_RES

<222> (27)..(38)

<223> Any amino acid and this region may encompass 2-12 residues

<400> 211

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Asn Ala
1 5 10 15

His Arg Ile Asn His Thr Lys Ile His Thr Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa
35

<210> 212

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<220><221> MOD_RES

<222> (27)..(38)

<223> Any amino acid and this region may encompass 2-12 residues

<400> 212

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ala Arg Ser
1 5 10 15

Thr Arg Thr Thr His Thr Lys Ile His Thr Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa
35

<210> 213

<211> 37

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<220><221> MOD_RES

<222> (27)..(33)

<223> Any amino acid and this region may encompass 5-7 residues

<220><221> MOD_RES

<222> (34)..(34)

<223> Phe, Leu, or Tyr

<400> 213

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Gln Lys Ile
1 5 10 15

Asn Leu Gln Val His Thr Lys Ile His Thr Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Xaa Arg Pro Pro
35