



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0139924
(43) 공개일자 2022년10월17일

(51) 국제특허분류(Int. Cl.)
C12N 15/86 (2006.01) A61K 48/00 (2006.01)
A61P 27/16 (2006.01) C07K 14/005 (2006.01)
C07K 14/47 (2006.01) C12N 15/62 (2006.01)
(52) CPC특허분류
C12N 15/86 (2013.01)
A61K 48/005 (2013.01)
(21) 출원번호 10-2022-7030657
(22) 출원일자(국제) 2021년02월05일
심사청구일자 없음
(85) 번역문제출일자 2022년09월05일
(86) 국제출원번호 PCT/US2021/016720
(87) 국제공개번호 WO 2021/158854
국제공개일자 2021년08월12일
(30) 우선권주장
62/971,555 2020년02월07일 미국(US)

(71) 출원인
더 칠드런스 메디칼 센터 코포레이션
미국 매사추세츠주 02115 보스턴 웨스턴 스트리트 55
(72) 발명자
홀트, 제프리 알.
미국 02115 매사추세츠 보스턴 웨스턴 스트리트 55
더 칠드런스 메디칼 센터 코포레이션 내
슈비나-올레이닉, 올라
미국 02115 매사추세츠 보스턴 웨스턴 스트리트 55
더 칠드런스 메디칼 센터 코포레이션 내
(74) 대리인
양영준, 이상남

전체 청구항 수 : 총 47 항

(54) 발명의 명칭 큰 유전자 벡터 및 그의 전달 및 용도

(57) 요약

본 개시내용은 이중-벡터 인테인-매개 단백질 트랜스-스플라이싱 시스템, 세포, 조성물 및 유전자 요법을 위한 그의 사용 방법을 제공한다. 일부 실시양태에서, 본 개시내용은 본원에 기재된 이중-벡터 시스템을 사용하여 STRC 단백질을 코딩하는 STRC 유전자를 전달함으로써 상염색체 열성 유형의 비-증후군성 난청인 DFNB16을 치료하기 위한 방법 및 조성물을 제공한다.

(52) CPC특허분류

A61K 48/0075 (2013.01)

A61P 27/16 (2018.01)

C07K 14/005 (2013.01)

C07K 14/47 (2013.01)

C12N 15/625 (2013.01)

C12N 2740/15043 (2013.01)

C12N 2750/14143 (2013.01)

C12N 2800/40 (2013.01)

C12N 2830/50 (2013.01)

명세서

청구범위

청구항 1

세포에서 관심 단백질을 발현시키기 위한 이중-벡터 시스템이며,

a) 5'에서 3' 방향으로:

- 관심 단백질의 아미노 말단 (N-말단) 부분을 코딩하는 부분 코딩 서열의 5'-말단에 있는 신호 서열;
- 관심 단백질의 N-말단 부분을 코딩하는 부분 코딩 서열;
- 부분 코딩 서열에 인접하고 그의 하류인 인테인의 아미노 카르복시 말단 단편 (N-인테인)을 코딩하는 서열을 포함하는 제1 뉴클레오타이드 서열을 포함하는 제1 벡터; 및

b) 5'에서 3' 방향으로:

- 관심 단백질의 카르복시 말단 (C-말단) 부분을 코딩하는 부분 코딩 서열의 5'-말단에 있는 신호 서열;
- 신호 서열 및 관심 단백질의 C-말단 부분을 코딩하는 부분 코딩 서열이 플랭킹된 인테인의 카르복시 말단 단편 (C-인테인)을 코딩하는 서열;
- 관심 단백질의 C-말단 부분을 코딩하는 부분 코딩 서열을 포함하는 제2 뉴클레오타이드 서열을 포함하는 제2 벡터를 포함하는 이중-벡터 시스템.

청구항 2

a) 5'에서 3' 방향으로:

- 5'-역전된 말단 반복부 (5'-ITR) 서열;
 - 프로모터 서열;
 - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;
 - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 관심 단백질의 아미노 말단 (N-말단) 부분을 코딩하는 부분 코딩 서열;
 - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 스플릿 인테인-N을 코딩하는 서열;
 - 폴리-아데닐화 (폴리A) 신호 서열;
 - 3'-역전된 말단 반복부 (3'-ITR) 서열
- 을 포함하는 제1 뉴클레오타이드 서열을 포함하는 제1 벡터; 및

b) 5'에서 3' 방향으로:

- 5'-역전된 말단 반복부 (5'-ITR) 서열;
- 프로모터 서열;
- 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;
- 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 스플릿 인테인-C를 코딩하는 서열;
- 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 관심 단백질의 카르복시 말단 (C-말단) 부분을 코딩하는 부분 코딩 서열;
- 폴리-아데닐화 (폴리A) 신호 서열;
- 3'-역전된 말단 반복부 (3'-ITR) 서열

을 포함하는 제2 뉴클레오타이드 서열을 포함하는 제2 벡터
를 포함하는 이중-벡터 시스템.

청구항 3

제1항 또는 제2항에 있어서, 세포에서 제1 벡터 및 제2 벡터가 각각

a) N-말단에서 C-말단 방향으로:

- 신호 펩티드 서열, 이에 연결된 관심 단백질 서열의 N-말단 부분, 이의 C-말단 단부에서 융합된 N-인테인 단백질 서열

을 포함하는 제1 단백질 서열; 및

b) N-말단에서 C-말단 방향으로:

- 신호 펩티드 서열, 이에 연결된 C-인테인 단백질 서열, 이에 N-말단 단부가 융합된 관심 단백질 서열의 C-말단 부분

을 포함하는 제2 단백질 서열

을 발현하는 것인 이중-벡터 시스템.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 관심 단백질의 N-말단 부분 및 관심 단백질의 C-말단 부분이 전장 관심 단백질을 형성하도록 구성된 것인 이중-벡터 시스템.

청구항 5

제3항에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 동일한 것인 이중-벡터 시스템.

청구항 6

제3항에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 제1 단백질 서열 및 제2 단백질 서열을 세포의 동일한 세포 구획으로 수송하도록 구성된 것인 이중-벡터 시스템.

청구항 7

제3항에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 상이하고, 각각의 신호 펩티드 서열이 각각의 단백질 서열을 세포의 동일한 세포 구획으로 지시하는 것인 이중-벡터 시스템.

청구항 8

제1항 내지 제3항 중 어느 한 항에 있어서, 제1 벡터 및 제2 벡터가 각각 바이러스 벡터인 이중-벡터 시스템.

청구항 9

제8항에 있어서, 바이러스 벡터가 아데노-연관 바이러스 (AAV) 벡터 또는 렌티바이러스인 이중-벡터 시스템.

청구항 10

제1항 내지 제4항 중 어느 한 항에 있어서, 관심 단백질이 STRC 단백질인 이중-벡터 시스템.

청구항 11

제1항 내지 제4항 중 어느 한 항에 있어서, 신호 서열이 서열식별번호: 9 또는 서열식별번호: 11에 대해 적어도 80% 동일성의 핵산 서열을 포함하거나 또는 신호 서열이 서열식별번호: 10 또는 서열식별번호: 12에 대해 적어도 80% 동일성의 아미노산 서열을 갖는 신호 펩티드 서열을 코딩하는 것인 이중-벡터 시스템.

청구항 12

제1항 내지 제4항 중 어느 한 항에 있어서, 관심 단백질의 N-말단 부분이 서열식별번호: 15 또는 서열식별번호: 16을 코딩하는 핵산 서열에 대해 적어도 70% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 15 또는 서열식별번호: 16에 대해 적어도 70% 동일성의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.

청구항 13

제1항 내지 제4항 중 어느 한 항에 있어서, N-인테인 서열이 서열식별번호: 13에 대해 적어도 80% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 14에 대해 적어도 80% 동일성의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.

청구항 14

제1항 내지 제4항 중 어느 한 항에 있어서, 관심 단백질의 C-말단 부분이 서열식별번호: 23 또는 서열식별번호: 24를 코딩하는 핵산 서열에 대해 적어도 70% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 23 또는 서열식별번호: 24에 대해 적어도 70% 동일성의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.

청구항 15

제1항 내지 제4항 중 어느 한 항에 있어서, C-인테인 서열이 서열식별번호: 21 또는 서열식별번호: 46에 대해 적어도 80% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 22 또는 서열식별번호: 49에 대해 적어도 80% 동일성의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.

청구항 16

제1항 내지 제4항 중 어느 한 항에 있어서, 제1 뉴클레오타이드 서열이 서열식별번호: 5 또는 서열식별번호: 7에 대해 적어도 70% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 6 또는 서열식별번호: 8에 대해 적어도 70% 동일성의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.

청구항 17

제1항 내지 제4항 중 어느 한 항에 있어서, 제2 뉴클레오타이드 서열이 서열식별번호: 17 또는 서열식별번호: 19에 대해 적어도 70% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 18 또는 서열식별번호: 20에 대해 적어도 70% 동일성의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.

청구항 18

숙주 세포에서 STRC 유전자의 코딩 서열을 발현시키기 위한 벡터 시스템이며, 여기서 코딩 서열은 서열식별번호: 1, 서열식별번호: 3, 서열식별번호: 5, 서열식별번호: 7, 서열식별번호: 17, 서열식별번호: 19, 서열식별번호: 29, 서열식별번호: 31, 서열식별번호: 33 또는 서열식별번호: 38의 STRC 유전자, 서열식별번호: 30 또는 서열식별번호: 32의 mRNA 서열 또는 그의 단편을 포함하는 적어도 1개의 벡터를 포함하는 것인 벡터 시스템.

청구항 19

제18항에 있어서, STRC 유전자가 서열식별번호: 2, 서열식별번호: 4, 서열식별번호: 6, 서열식별번호: 8, 서열식별번호: 15, 서열식별번호: 16, 서열식별번호: 18, 서열식별번호: 20, 서열식별번호: 23, 서열식별번호: 24, 서열식별번호: 25, 서열식별번호: 26, 서열식별번호: 36, 서열식별번호: 39 또는 그의 조합의 STRC 단백질을 코딩하는 것인 벡터 시스템.

청구항 20

제18항 또는 제19항에 있어서, 숙주 세포에서 STRC 유전자의 코딩 서열을 발현시키기 위한 이중-벡터 시스템을 포함하며, 여기서 코딩 서열은 5' 말단 단편 및 3' 말단 단편을 포함하고, 이중-벡터 시스템은

a) 5'에서 3' 방향으로:

- 5'-역전된 말단 반복부 (5'-ITR) 서열;
- 프로모터 서열;

- 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;
 - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, STC 유전자 코딩 서열의 5' 말단 단편;
 - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 인테인의 아미노 말단 단편 (N-인테인)을 코딩하는 서열;
 - 폴리-아데닐화 (폴리A) 신호 서열; 및
 - 3'-역전된 말단 반복부 (3'-ITR) 서열
- 을 포함하는 제1 뉴클레오타이드 서열을 포함하는 제1 벡터; 및
- b) 5'에서 3' 방향으로:
- 5'-역전된 말단 반복부 (5'-ITR) 서열;
 - 프로모터 서열;
 - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;
 - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 인테인의 카르복시 말단 단편 (C-인테인)을 코딩하는 서열;
 - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, STC 유전자 코딩 서열의 3' 말단 단편;
 - 폴리-아데닐화 (폴리A) 신호 서열; 및
 - 3'-역전된 말단 반복부 (3'-ITR) 서열
- 을 포함하는 제2 뉴클레오타이드 서열을 포함하는 제2 벡터
- 를 포함하는 것인 벡터 시스템.

청구항 21

제20항에 있어서, 세포에서 제1 벡터 및 제2 벡터가 각각

a) N-말단에서 C-말단 방향으로:

- 신호 펩티드 서열, 이에 연결된 STC 단백질 서열의 N-말단 부분, 이의 C-말단 단부에서 융합된 N-인테인 단백질 서열
- 을 포함하는 제1 단백질 서열; 및

b) N-말단에서 C-말단 방향으로:

- 신호 펩티드 서열, 이에 연결된 C-인테인 단백질 서열, 이에 N-말단 단부가 융합된 STC 단백질 서열의 C-말단 부분
- 을 포함하는 제2 단백질 서열
- 을 발현하는 것인 이중-벡터 시스템.

청구항 22

제20항 또는 제21항에 있어서, STC 단백질의 N-말단 부분 및 STC 단백질의 C-말단 부분이 전장 STC 단백질을 형성하도록 구성된 것인 이중-벡터 시스템.

청구항 23

제21항에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 동일한 것인 이중-벡터 시스템.

청구항 24

제21항에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 제1 단백질

서열 및 제2 단백질 서열을 동일한 세포 구획으로 수송하도록 구성된 것인 이중-백터 시스템.

청구항 25

제21항에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 상이하고, 각각의 신호 펩티드 서열이 각각의 단백질 서열을 동일한 세포 구획으로 지시하는 것인 이중-백터 시스템.

청구항 26

제20항 또는 제21항에 있어서, 바이러스 벡터가 아데노-연관 바이러스 (AAV) 벡터인 이중-백터 시스템.

청구항 27

제20항 또는 제21항에 있어서, 신호 서열이 서열식별번호: 9 또는 서열식별번호: 11에 대해 적어도 80% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 10 또는 서열식별번호: 12에 대해 적어도 80% 동일성의 아미노산 서열을 코딩하는 것인 이중-백터 시스템.

청구항 28

제20항 또는 제21항에 있어서, STRC 단백질의 N-말단 부분이 서열식별번호: 15 또는 서열식별번호: 16을 코딩하는 핵산 서열에 대해 적어도 70% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 6, 서열식별번호: 8, 서열식별번호: 15 또는 서열식별번호: 16에 대해 적어도 70% 동일성의 아미노산 서열을 코딩하는 것인 이중-백터 시스템.

청구항 29

제20항 또는 제21항에 있어서, STRC 단백질의 N-말단 부분이 전장 STRC 단백질의 N-말단 단부 부분의 54% 미만을 포함하는 것인 이중-백터 시스템.

청구항 30

제20항 또는 제21항에 있어서, N-인테인 서열이 서열식별번호: 13에 대해 적어도 80% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 14에 대해 적어도 80% 동일성의 아미노산 서열을 코딩하는 것인 이중-백터 시스템.

청구항 31

제20항 또는 제21항에 있어서, STRC 단백질의 C-말단 부분이 서열식별번호: 18, 서열식별번호: 20, 서열식별번호: 23 또는 서열식별번호: 24를 코딩하는 핵산 서열에 대해 적어도 70% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 18, 서열식별번호: 20, 서열식별번호: 23 또는 서열식별번호: 24에 대해 적어도 70% 동일성의 아미노산 서열을 코딩하는 것인 이중-백터 시스템.

청구항 32

제20항 또는 제21항에 있어서, STRC 단백질의 C-말단 부분이 전장 STRC 단백질의 C-말단 단부 부분의 46% 이상을 포함하는 것인 이중-백터 시스템.

청구항 33

제20항 또는 제21항에 있어서, C-인테인 서열이 서열식별번호: 21에 대해 적어도 80% 동일한 핵산 서열을 포함하거나 또는 서열식별번호: 22에 대해 적어도 80% 동일한 아미노산 서열을 코딩하는 것인 이중-백터 시스템.

청구항 34

제20항 또는 제21항에 있어서, 제1 뉴클레오타이드 서열이 서열식별번호: 5 또는 서열식별번호: 7에 대해 적어도 70% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 6, 서열식별번호: 8, 서열식별번호: 15 또는 서열식별번호: 16에 대해 적어도 70% 동일성의 아미노산 서열을 코딩하는 것인 이중-백터 시스템.

청구항 35

제20항 또는 제21항에 있어서, 제2 뉴클레오타이드 서열이 서열식별번호: 17 또는 서열식별번호: 19에 대해 적어도 70% 동일성의 핵산 서열을 포함하거나 또는 서열식별번호: 18, 서열식별번호: 20, 서열식별번호: 23 또는 서열식별번호: 24에 대해 적어도 70% 동일성의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.

청구항 36

제1항 내지 제35항 중 어느 한 항의 벡터 시스템을 함유하는 세포.

청구항 37

제1항 내지 제35항 중 어느 한 항의 벡터 시스템 및 제약상 허용되는 비히클을 포함하는 제약 조성물.

청구항 38

상염색체 열성 청력 상실의 치료를 필요로 하는 대상체에게 유효량의 제1항 내지 제35항 중 어느 한 항의 이중-벡터 시스템을 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료하는 방법.

청구항 39

상염색체 열성 청력 상실의 치료를 필요로 하는 대상체에게 제36항의 세포를 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료하는 방법.

청구항 40

상염색체 열성 청력 상실의 치료를 필요로 하는 대상체에게 제37항의 제약 조성물을 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료하는 방법.

청구항 41

제38항 내지 제40항 중 어느 한 항에 있어서, 상염색체 열성 청력 상실이 DFNB16인 방법.

청구항 42

대상체의 적어도 1개의 세포를 제37항의 제약 조성물과 접촉시키는 것을 포함하는 방법이며, 여기서 접촉은 제1 뉴클레오타이드 서열 및 제2 뉴클레오타이드 서열을 포함하는 벡터 시스템을 대상체의 적어도 1개의 세포 내로 전달하고, 여기서 접촉된 적어도 1개의 세포는 펩티드 결합에 의해 연결된 단백질의 N-말단 부분 및 단백질의 C-말단 부분을 발현하여 전장 단백질을 형성하는 것인 방법.

청구항 43

청력 상실을 특징으로 하는 병리상태 또는 질환의 치료 및/또는 예방을 필요로 하는 대상체에게 유효량의 제1항 내지 제35항 중 어느 한 항에 따른 벡터 시스템, 제36항에 따른 적어도 1개의 세포 또는 제37항에 따른 제약 조성물을 투여하는 것을 포함하는, 청력 상실을 특징으로 하는 병리상태 또는 질환을 치료 및/또는 예방하는 방법.

청구항 44

제42항 또는 제43항에 있어서, 적어도 1개의 세포가 내이 세포인 방법.

청구항 45

제42항 또는 제43항에 있어서, 적어도 1개의 세포가 내유모 세포 또는 외유모 세포인 방법.

청구항 46

제42항 내지 제45항 중 어느 한 항에 있어서, 적어도 1개의 세포가 생체내 또는 시험관내 세포인 방법.

청구항 47

제42항 내지 제46항 중 어느 한 항에 있어서, 대상체에서 청각 기능을 개선 또는 회복시키는 방법.

발명의 설명

기술 분야

[0001] 관련 출원에 대한 상호-참조

[0002] 본 PCT 국제 출원은 2020년 2월 7일에 출원된 미국 가출원 번호 62/971,555를 우선권 주장하며, 이는 모든 목적을 위해 그 전문이 본원에 참조로 포함된다.

배경 기술

[0003] 청력 상실은 선진국 또는 산업화 국가에서 가장 흔한 신경계 장애 중 하나이고, 전세계적으로 4억 6천 6백만건 초과를 사례를 차지하는 가장 보편적인 감각신경성 장애이다. 비-중후군성 난청 또는 비-중후군성 유전적 난청은 임의의 다른 징후 또는 증상과 연관되지 않은 청력 상실이다. 4가지 유형의 비-중후군성 난청이 있다: DFNA (상염색체 우성), DFNB (상염색체 열성), DFNX (X-연관) 및 미토콘드리아 비-중후군성 난청. 비-중후군성 난청 중 16번째로 기재된 상염색체 열성 유형인 DFNB16은 스테레오실린으로 공지된 세포외 구조 단백질을 코딩하는 STRC 유전자 내의 돌연변이에 의해 유발된 단일유전자, 비-중후군성, 열성 청력 상실이다. 내이에서의 STRC의 정상 발현은 청각 기능에 필수적이다. 내이 내 감각 유모 세포의 정점에 있는 변형된 미세융모의 상단에서 발견되는 스테레오실린은 내이 내 특수 세포로부터 돌출된, 부동섬모로 공지된 유모-유사 구조와 회합된다. STRC 내의 돌연변이는 중등도 내지 중증 청력 상실을 유발하고, 미국에서 추정건대 ~50,000명의 환자에 영향을 미치며, 따라서 유전자 요법을 위한 매력적인 후보이다. 스테레오실린은 응집성 미세융모 다발을 유지하고, 다발을 내이의 와우 내에 있는 위에 놓인 덮개막에 커플링시키는 기능을 한다. 전세계적 통계는 DFNB16이, 특히 중등도 청력 장애를 갖는 대상체에서 유전적 난청의 유의한 비율을 구성한다는 것을 시사한다. 파트너스 래보러토리 오브 몰레큘라 메디신(Partners Laboratory of Molecular Medicine)으로부터의 데이터에 기초하면, 보스턴에서 시험된 유전적 청력 상실 환자의 19%는 STRC에 돌연변이를 가지며, 이에 따라 그것은 유전적 청력 상실의 두번째로 가장 흔한 형태이고, 내이의 감각 유모 세포에 영향을 미치는 가장 흔한 형태이다. 약 40개의 상이한 돌연변이 (주로 열성)가 STRC 유전자에서 확인되었으며, 대다수는 결합성 스테레오실린의 합성을 유도하거나 또는 그의 합성을 완전히 방지한다. 정상 STRC 단백질의 결여는 적절한 소리 유발 자극에 요구되는 위에 놓인 덮개막으로부터 감각 유모 다발을 탈커플링시킨다. DFNB16 환자는 중등도 내지 중증 청력 상실을 갖고, 전형적으로 보청기 또는 와우 이식물로 치료된다. 그러나, 현재 DFNB16 청력 상실에 대한 생물학적 치료는 없다.

[0004] AAV는 유전성 장애의 유전자 요법 치료를 위한 매력적인 벡터 시스템을 제공한다. 안전성 관점에서 이들 및 유전자 전달. 재조합 AAV (rAAV)는 비-병원성 및 복제-결합성 바이러스로부터 유래되고, 그의 숙주 세포에 대해 비-세포독성이다. 더욱이, rAAV는 또 다른 안전성 특색을 제시하는 역전된 말단 반복부 (ITR)를 제외하고 모든 바이러스 DNA 서열이 결여되어 있다. ITR은 AAV DNA 복제, 패키징, 염색체 통합 및 프로-바이러스 구출에 필요하다. AAV 벡터는 또한 효과적인 트랜스진 전달 및 예를 들어 내이 세포에서의 지속적인 발현을 위한 강력한 도구인 것으로 입증되었다. 그러나, 내이 기능에 중요한 많은 단백질은 STRC의 적재 용량 (~5.8 kB)을 비롯하여 AAV 벡터의 적재 용량 (~4.5 kB)을 초과하는 코딩 서열을 갖는다. 따라서, 상기 언급된 임의의 유전자 요법, 구축물 및 벡터의 효과적인 형태로서 큰 유전자 (예를 들어, 4 kB 초과)에 의해 코딩되는 단백질의 전달 및 발현 방법이 필요하다.

발명의 내용

[0005] 따라서, 본 개시내용의 목적은 유전자 돌연변이를 앓고 있는 대상체를 치료하기 위한 큰 유전자 (예를 들어, 4 kB 초과)에 대한 유전자 요법을 제공하는 것이다. 유전자 요법은, 예를 들어 DFNB16 청력 상실을 갖는 소아 및 성인에서 청력의 예방 및/또는 회복을 가능하게 할 수 있다.

[0006] 또 다른 목적은 큰 유전자 서열 (예를 들어, 4 kB 초과; STRC)을 전달하는 방법 및 큰 유전자 서열을 전달하기 위한 벡터 및 구축물을 제공하는 것이며, 여기서 방법은 벡터 크기 한계를 극복한다.

[0007] 한 측면은

[0008] a) 5'에서 3' 방향으로:

[0009] - 관심 단백질의 아미노 말단 (N-말단) 부분을 코딩하는 부분 코딩 서열의 5'-말단에 있는 신호 서열 (예를 들어, 서열식별번호: 9; 서열식별번호: 11);

- [0010] - 관심 단백질의 N-말단 부분 (예를 들어, N-STRC; 서열식별번호: 15; 서열식별번호: 16)을 코딩하는 부분 코딩 서열;
- [0011] - 부분 코딩 서열에 인접하고 그의 하류인 스플라이스 공여자 서열을 코딩하는 서열 (예를 들어, 인테인의 N-말단 단편 (N-인테인), 또한 스플릿 인테인-N으로도 공지됨; 서열식별번호: 14를 코딩하는 서열식별번호: 13)
- [0012] 을 포함하는 제1 뉴클레오티드 서열 (예를 들어, 서열식별번호: 5; 서열식별번호: 7)을 포함하는 제1 벡터; 및
- [0013] b) 5'에서 3' 방향으로:
- [0014] - 관심 단백질의 카르복시 말단 (C-말단) 부분을 코딩하는 부분 코딩 서열의 5'-말단에 있는 신호 서열 (예를 들어, 서열식별번호: 9; 서열식별번호: 11);
- [0015] - 신호 서열 및 관심 단백질의 C-말단 부분을 코딩하는 부분 코딩 서열이 플랭킹된 스플라이스 수용자 서열을 코딩하는 서열 (예를 들어, 인테인의 C-말단 단편 (C-인테인), 또한 스플릿 인테인-C로도 공지됨; 서열식별번호: 22를 코딩하는 서열식별번호: 21);
- [0016] - 관심 단백질의 C-말단 부분 (예를 들어, C-STRC; 서열식별번호: 23; 서열식별번호: 24)을 코딩하는 부분 코딩 서열
- [0017] 을 포함하는 제2 뉴클레오티드 서열 (예를 들어, 서열식별번호: 17; 서열식별번호: 19)을 포함하는 제2 벡터
- [0018] 를 포함하는, 세포에서 관심 단백질을 발현시키기 위한 벡터 시스템 (예를 들어, 이중-벡터 시스템)을 제공한다.
- [0019] 또 다른 측면은
- [0020] a) 5'에서 3' 방향으로:
- [0021] - 5'-역전된 말단 반복부 (5'-ITR) 서열;
- [0022] - 프로모터 서열;
- [0023] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;
- [0024] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 관심 단백질의 아미노 말단 (N-말단) 부분 (예를 들어, N-STRC)을 코딩하는 부분 코딩 서열;
- [0025] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 인테인의 아미노 말단 단편 (N-인테인)을 코딩하는 서열;
- [0026] - 폴리-아데닐화 (폴리A) 신호 서열;
- [0027] - 3'-역전된 말단 반복부 (3'-ITR) 서열
- [0028] 을 포함하는 제1 뉴클레오티드 서열을 포함하는 제1 벡터; 및
- [0029] b) 5'에서 3' 방향으로:
- [0030] - 5'-역전된 말단 반복부 (5'-ITR) 서열;
- [0031] - 프로모터 서열;
- [0032] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;
- [0033] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 인테인의 카르복시 말단 단편 (C-인테인)을 코딩하는 서열;
- [0034] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 관심 단백질의 카르복시 말단 (C-말단) 부분 (예를 들어, C-STRC)을 코딩하는 부분 코딩 서열;
- [0035] - 폴리-아데닐화 (폴리A) 신호 서열;
- [0036] - 3'-역전된 말단 반복부 (3'-ITR) 서열
- [0037] 을 포함하는 제2 뉴클레오티드 서열을 포함하는 제2 벡터

- [0038] 를 포함하는, 세포에서 관심 단백질을 발현하기 위한 이중-벡터 시스템을 제공한다.
- [0039] 이중-벡터 시스템의 또 다른 측면은 세포에서 제1 벡터 및 제2 벡터가 각각
- [0040] a) N-말단에서 C-말단 방향으로:
- [0041] - 신호 펩티드 서열, 이에 연결된 관심 단백질 (예를 들어, STRC) 서열의 N-말단 부분, 이의 C-말단 단부에서 융합된 N-인테인 단백질 서열
- [0042] 을 포함하는 제1 단백질 서열; 및
- [0043] b) N-말단에서 C-말단 방향으로:
- [0044] - 신호 펩티드 서열, 이에 연결된 C-인테인 단백질 서열, 이에 N-말단 단부가 융합된 관심 단백질 (예를 들어, STRC) 서열의 C-말단 부분
- [0045] 을 포함하는 제2 단백질 서열
- [0046] 을 발현하는 것을 제공한다.
- [0047] 또 다른 측면은 관심 단백질의 N-말단 부분 (예를 들어, N-STRC) 및 관심 단백질의 C-말단 부분 (예를 들어, C-STRC)이 전장 관심 단백질 (예를 들어, STRC)을 형성하는 것을 제공한다. 일부 측면에서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열은 동일하거나 상이하거나 또는 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열은 제1 단백질 서열 및 제2 단백질 서열을 동일한 세포 구획으로 수송하도록 구성된다. 추가 측면은 서열식별번호: 9 또는 서열식별번호: 11에 대해 적어도 80% 동일성을 갖는 핵산 서열을 포함하고 서열식별번호: 10 또는 서열식별번호: 12에 대해 적어도 80% 동일성을 갖는 아미노산 서열을 갖는 신호 펩티드 서열을 코딩하는 신호 서열에 관한 것일 수 있다. 다른 측면은 아데노-연관 바이러스 (AAV) 벡터 또는 렌티바이러스일 수 있는 바이러스 벡터일 수 있는 벡터 (예를 들어, 제1 벡터 및 제2 벡터)를 제공한다. 한 측면은 동일하거나 상이한 혈청형을 갖는 바이러스 벡터에 관한 것일 수 있다. 이중-벡터 시스템의 또 다른 측면은 관심 단백질의 인테인-매개 트랜스-스플라이싱을 제공하며, 여기서 관심 단백질의 N-말단 부분 (예를 들어, N-STRC) 및 관심 단백질의 C-말단 부분 (예를 들어, C-STRC)은 펩티드 결합을 통해 전장 관심 단백질 (예를 들어, STRC)을 형성할 수 있고, 여기서 관심 단백질은 STRC 유전자에 의해 코딩되는 STRC 단백질일 수 있다. 관심 단백질의 N-말단 부분 (예를 들어, N-STRC; 인간 서열식별번호: 15 또는 류린 서열식별번호: 16)을 코딩하는 뉴클레오타이드 서열은, 전장 관심 단백질 (예를 들어, STRC; 서열식별번호: 25 또는 서열식별번호: 26)의 N-말단 부분의 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 또는 그의 54% 미만 (예를 들어, 53%, 52%, 51%, 50%, 45%, 43%, 41%) 및/또는 그에 대해 54% 미만의 동일성 및/또는 그의 54% 미만의 길이의 관심 아미노산 서열 (예를 들어, STRC; 서열식별번호: 6; 서열식별번호: 8; 서열식별번호: 15 또는 서열식별번호: 16)을 코딩하는, 관심 뉴클레오타이드 서열 (예를 들어, STRC; 서열식별번호: 5 또는 서열식별번호: 7)에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함한다. 또 다른 측면은 전장 관심 단백질 (예를 들어, 서열식별번호: 25 또는 서열식별번호: 26)의 N-말단 부분의 41% 이상 (예를 들어, 42%, 43%, 44%, 45%, 50%, 51%, 52%, 53%) 및/또는 그에 대해 41% 이상의 동일성 및/또는 그의 41% 이상의 길이의 아미노산 서열을 포함하는 관심 단백질 (예를 들어, STRC)의 N-말단 부분을 제공한다.
- [0048] 추가 측면은 목적하는 신호 펩티드 서열 (예를 들어, 서열식별번호: 10; 서열식별번호: 12)에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 갖는 신호 펩티드 서열을 코딩하는, 목적하는 신호 서열 (예를 들어, 서열식별번호: 9; 서열식별번호: 11)에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 갖는 신호 서열을 포함하는 뉴클레오타이드 서열을 제공한다.
- [0049] 또 다른 측면은 목적하는 N-인테인 아미노산 서열 (예를 들어, 서열식별번호: 14)에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 목적하는 N-인테인 아미노산 서열을 코딩하는, 목적하는 N-인테인 뉴클레오타이드 서열 (예를 들어, 서열식별번호: 13)에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함하는 목적하는 N-인테인 서열을 제공할 수 있다. 추가 측면은 목적하는 C-인테인 아미노산 서열 (예를 들어, 서열식별번호: 22)에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 목적하는 C-인테인 아미노산 서열을 코딩하는, 목적하는 C-인테인 뉴클레오타이드 서열 (예를 들어, 서열식별번호: 21)에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산

서열을 포함하는 목적하는 C-인테인 서열에 관한 것일 수 있다.

[0050] 본 개시내용의 이중-벡터 시스템의 또 다른 측면에서, 관심 단백질 (예를 들어, STRC)의 C-말단 부분은, 전장 관심 단백질 (예를 들어, STRC; 서열식별번호: 25; 서열식별번호: 26)의 C-말단 부분의 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 또는 그의 46% 이상 및/또는 그에 대해 46% 이상의 동일성 및/또는 그의 46% 이상의 길이의 관심 아미노산 서열 (예를 들어, STRC; 서열식별번호: 18; 서열식별번호: 20; 서열식별번호: 23; 서열식별번호: 24)을 코딩하는, 뉴클레오타이드 서열 (예를 들어, STRC; 서열식별번호: 17; 서열식별번호: 19)에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함할 수 있다. 또 다른 측면은 전장 관심 단백질 (예를 들어, STRC; 서열식별번호: 25; 서열식별번호: 26)의 C-말단 부분에 대해 60% 이하의 동일성 및/또는 그의 60% 이하의 길이의 아미노산 서열을 포함하는 관심 단백질 (예를 들어, STRC)의 C-말단 부분을 제공할 수 있다.

[0051] 추가 측면은 숙주 세포에서 STRC 유전자의 코딩 서열을 발현시키기 위한 벡터 시스템이며, 여기서 코딩 서열은, 예를 들어 인간 STRC: 서열식별번호: 1 또는 서열식별번호: 30 또는 무린 STRC: 서열식별번호: 3 또는 서열식별번호: 32의 STRC 뉴클레오타이드 코딩 서열을 포함하는 적어도 1개의 벡터를 포함하고, 여기서 STRC 뉴클레오타이드 코딩 서열은, 예를 들어 서열식별번호: 2 또는 서열식별번호: 25 또는 서열식별번호: 4 또는 서열식별번호: 26의 STRC 단백질을 코딩하는 것인 벡터 시스템을 제공할 수 있다. 또 다른 측면은 목적하는 전장 단백질을 코딩하는 뉴클레오타이드 서열이며, 여기서 뉴클레오타이드 서열은, 예를 들어 인간 STRC: 서열식별번호: 1 또는 서열식별번호: 33 또는 무린 STRC: 서열식별번호: 3 또는 서열식별번호: 39를 포함하고, 이는 목적하는 단백질, 예를 들어 인간 STRC: 서열식별번호: 2 또는 서열식별번호: 25 또는 무린 STRC: 서열식별번호: 4 또는 서열식별번호: 26을 코딩하는 것인 뉴클레오타이드 서열에 관한 것일 수 있다.

[0052] 벡터 시스템의 한 측면은 본원에 기재된 바와 같은 숙주 세포에서 STRC 유전자의 코딩 서열을 발현시키기 위한 이중-벡터 시스템을 포함하며, 여기서 이중-벡터 시스템은 목적하는 관심 핵산 서열 (예를 들어, 서열식별번호: 5; 서열식별번호: 7)에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 목적하는 핵산 서열을 포함하는 제1 뉴클레오타이드 서열을 포함하는 제1 벡터를 제공한다. 한 측면에서, 제1 벡터 (예를 들어, 플라스미드, 트랜스플라이징 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 계통)는 5'에서 3' 방향으로: 부분 코딩 서열의 5'-말단에 있는 신호 서열 (예를 들어, 서열식별번호: 10을 코딩하는 서열식별번호: 9; 또는 서열식별번호: 12를 코딩하는 서열식별번호: 11); 관심 단백질 (예를 들어, STRC; 서열식별번호: 15; 서열식별번호: 16)의 아미노 말단 (N-말단) 부분을 코딩하는 부분 코딩 서열; 부분 코딩 서열에 플랭킹되거나 인접할 수 있는 스플라이스 공여자 서열을 코딩하는 하류 서열 (예를 들어, N-말단 인테인 (N-인테인, 또한 스플릿 인테인-N으로도 공지됨); 서열식별번호: 14를 코딩하는 서열식별번호: 13)을 포함하는 제1 뉴클레오타이드 서열 (예를 들어, 서열식별번호: 6을 코딩하는 서열식별번호: 5; 서열식별번호: 8을 코딩하는 서열식별번호: 7)을 포함한다. 또 다른 측면은 관심 단백질의 N-말단 부분을 포함하고, 또한 신호 서열 및 목적하는 N-인테인 단백질을 코딩하는 서열, 뿐만 아니라 역전된 말단 반복부 (ITR), 프로모터 및 폴리-아데닐화 (폴리A) 서열을 함유하는 제1 뉴클레오타이드 서열을 포함한다. 제1 뉴클레오타이드 서열은 목적하는 아미노산 서열 (예를 들어, 서열식별번호: 5; 서열식별번호: 16) 또는 전장 관심 아미노산 서열 (예를 들어, 서열식별번호: 25; 서열식별번호: 26)에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 관심 아미노산 서열을 코딩할 수 있다.

[0053] 본 개시내용의 이중-벡터 시스템의 또 다른 측면은 또한 목적하는 관심 핵산 서열 (예를 들어, 서열식별번호: 17; 서열식별번호: 19)에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 목적하는 핵산 서열의 나머지 부분을 포함하는 제2 뉴클레오타이드 서열을 제공한다. 한 측면에서, 제2 벡터 (예를 들어, 플라스미드, 트랜스플라이징 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 계통)는 5'에서 3' 방향으로: 신호 서열 (예를 들어, 서열식별번호: 10을 코딩하는 서열식별번호: 9; 또는 서열식별번호: 12를 코딩하는 서열식별번호: 11), 신호 서열이 그의 상류일 수 있는 스플라이스 수용자 서열 (예를 들어, C-말단 인테인 (C-인테인); 서열식별번호: 22를 코딩하는 서열식별번호: 21), 그에 바로 인접하여 위치하거나 플랭킹된 관심 단백질의 전장 코딩 서열의 나머지 부분, 즉 관심 단백질의 C-말단 부분 (예를 들어, STRC; 서열식별번호: 23; 서열식별번호: 24)을 코딩하는 하류 부분 코딩 서열을 포함하는 제2 뉴클레오타이드 서열 (예를 들어, 서열식별번호: 18을 코딩하는 서열식별번호: 17; 서열식별번호: 20을 코딩하는 서열식별번호: 19)을 포함한다. 또 다른 측면은 관심 단백질의 C-말단 부분, 신호 서

열을 코딩하는 서열 및 목적하는 C-인테인 단백질을 코딩하는 서열, 뿐만 아니라 역전된 말단 반복부 (ITR), 프로모터 및 폴리-아데닐화 (폴리A) 서열을 포함하는 제2 뉴클레오티드 서열을 포함한다. 일부 측면에서, 제2 뉴클레오티드 서열은 또한 링커 서열 및 myc 태그 서열을 함유할 수 있다. 제2 뉴클레오티드 서열은 목적하는 아미노산 서열 (예를 들어, 서열식별번호: 18; 서열식별번호: 20; 서열식별번호: 23; 서열식별번호: 24) 또는 전장 관심 아미노산 서열 (예를 들어, 서열식별번호: 25; 서열식별번호: 26)에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 코딩할 수 있다.

[0054] 다른 측면은 목적하는 유전자 또는 그의 목적하는 단백질 (예를 들어, STRC 단백질)을 전달하기 위한 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템)을 함유하는 세포(들) 또는 숙주 세포(들)를 제공할 수 있다.

[0055] 추가 측면은 목적하는 유전자 또는 그의 목적하는 단백질 (예를 들어, STRC 단백질)을 전달하기 위한 본 개시내용의 벡터 시스템 (예를 들어, 이중-벡터 시스템) 및 제약상 허용되는 비히클 (예를 들어, 희석제, 부형제)을 포함하는 제약 조성물에 관한 것일 수 있다.

[0056] 또 다른 측면에서, 질환 또는 상태의 유전자 돌연변이를 앓고 있는, 그 질환 또는 상태의 치료를 필요로 하는 대상체에게 유효량의 본 개시내용의 벡터 시스템 (예를 들어, 이중-벡터 시스템)을 투여하는 것을 포함하는, 상기 대상체에서 질환 또는 상태를 치료하는 방법이며, 여기서 방법은 동일한 유전자 내의 유전자 돌연변이에 의해 유발된 질환 또는 상태를 앓고 있는 대상체에게 목적하는 야생형 또는 교정된 유전자 또는 목적하는 야생형 또는 교정된 단백질 (예를 들어, STRC)을 전달하여, 그에 의해 대상체에서 질환 또는 상태를 치료한다. 또 다른 측면은 상염색체 열성 청력 상실을 앓고 있는, 그 질환 또는 상태의 치료를 필요로 하는 대상체에게 목적하는 야생형 또는 교정된 유전자 또는 목적하는 야생형 또는 교정된 단백질 (예를 들어, STRC)을 전달하는 유효량의 본원에 기재된 이중-벡터 시스템을 투여하는 것을 포함하는, 상기 대상체에서 질환 또는 상태를 치료하는 방법을 제공한다. 상염색체 열성 청력 상실의 치료를 필요로 하는 대상체에게 목적하는 유전자 또는 그의 목적하는 단백질 (예를 들어, STRC 단백질)을 전달하기 위한 본원에 기재된 이중-벡터 시스템을 함유하는 세포(들) 또는 숙주 세포(들) 또는 제약 조성물 (제약상 허용되는 비히클 (예를 들어, 희석제, 부형제) 함유)을 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료하는 방법을 제공한다. 또 다른 측면은 상염색체 열성 청력 상실, DFNB16을 제공할 수 있다.

[0057] 한 측면에서, 본 개시내용의 방법은 대상체의 세포를 목적하는 유전자 또는 그의 목적하는 단백질 (예를 들어, STRC 단백질)을 전달하기 위한 본 개시내용의 벡터 시스템 (예를 들어, 이중-벡터 시스템) 및 제약상 허용되는 비히클 (예를 들어, 희석제, 부형제)을 포함하는 조성물과 접촉시키는 것을 포함할 수 있으며, 여기서 접촉은 제1 뉴클레오티드 서열 및 제2 뉴클레오티드 서열의 세포 내로의 전달을 일으키고, 여기서 세포는 펩티드 결합에 의해 연결된 목적하는 단백질의 N-말단 부분 및 목적하는 단백질의 C-말단 부분을 발현하여 전장의 목적하는 단백질을 형성할 수 있다. 또 다른 측면은 청력 상실을 특징으로 하는 병리상태 또는 질환의 치료 및/또는 예방을 필요로 하는 대상체에게 유효량의 본원에 기재된 이중-벡터 시스템을 투여하거나, 또는 목적하는 유전자 또는 그의 목적하는 단백질 (예를 들어, STRC 단백질)을 전달하기 위한 본원에 기재된 이중-벡터 시스템을 함유하는 세포 또는 제약 조성물 (제약상 허용되는 비히클 (예를 들어, 희석제, 부형제) 함유)을 투여하는 것을 포함하는, 청력 상실을 특징으로 하는 병리상태 또는 질환을 치료 및/또는 예방하는 방법을 제공한다. 일부 측면에서, 세포는 귀의 내이 세포, 내유모 세포 또는 외유모 세포일 수 있으며, 여기서 세포 또는 세포에의 투여 방법은 생체내, 생체외 및/또는 시험관내에서 이루어질 수 있다. 본 개시내용의 추가 측면은 본원에 기재된 임의의 방법이 대상체에서 청각 기능의 개선 또는 회복을 가져오는 것이다.

[0058] 본 개시내용의 다른 특색 및 이점은 상세한 설명 및 청구범위로부터 분명할 것이다.

도면의 간단한 설명

[0059] 도 1은 AAV2 역전된 말단 반복부 (ITR), 프로모터 및 폴리-아데닐화 (폴리A) 서열을 갖는, 목적하는 전장 단백질 (예를 들어, STRC)의 발현을 위한 단일 벡터 시스템에 대한 구축물의 개략적 표현을 보여준다.

도 2a-2c는 인간 STRC 단백질을 코딩하는 뉴클레오티드 서열, 신호 펩티드 서열, 링커 서열, Myc 태그를 코딩하는 서열 및 출발 및 정지 코돈 (서열식별번호: 33)을 보여준다.

도 3은 도 2a-2c에 제시된 뉴클레오티드 서열에 의해 코딩되는 신호 펩티드 서열, 인간 STRC 단백질 서열, 링커 서열 및 myc 태그를 함유하는 아미노산 서열 (서열식별번호: 36)을 보여준다.

도 4a-4d는 뮤린 STRC 단백질을 코딩하는 뉴클레오타이드 서열, 신호 펩티드 서열, 링커 서열, Myc 태그를 코딩하는 서열 및 출발 및 정지 코돈 (서열식별번호: 38)을 보여준다.

도 5는 도 4a 내지 4d에 제시된 뉴클레오타이드 서열에 의해 코딩되는 신호 펩티드 서열, 뮤린 STRC 단백질 서열, 링커 서열 및 Myc 태그를 함유하는 아미노산 서열 (서열식별번호: 39)을 보여준다.

도 6은 AAV2 역전된 말단 반복부, 프로모터 및 폴리-아데닐화 서열을 사용한 이중 AAV 인테인-매개 스테레오실린 단백질 트랜스-스플라이싱의 개략적 표현을 보여준다. 인테인 단편은 단백질 재조합을 매개하고, 그 자체를 절제하고, 나머지 STRC 단편 (엑스테인)을 펩티드 결합으로 연결한다.

도 7a 및 7b는 인간 STRC 단백질의 N-말단 부분, 신호 펩티드 서열 및 스플라이스 공여자 서열 (예를 들어, N-인테인; CFS-N-Strc-N-Int, 구축물 2, N-부분)을 코딩하는 뉴클레오타이드 서열 (서열식별번호: 5)을 보여준다. 뉴클레오타이드 서열은 신호 서열, 5' Strc (야생형 Strc 코딩 서열의 5' 단편) 및 N-인테인 서열 (인테인 단백질의 N-말단 단편을 코딩함)을 함유한다.

도 8a 및 8b는 뮤린 STRC 단백질의 N-말단 부분, 신호 펩티드 서열 및 N-인테인 (예를 들어, CFS-N-Strc-N-Int, 구축물 2, N-부분)을 코딩하는 뉴클레오타이드 서열 (서열식별번호: 7)을 보여준다. 뉴클레오타이드 서열은 신호 서열, 5' Strc (야생형 Strc 코딩 서열의 5' 단편) 및 N-인테인 서열 (인테인 단백질의 N-말단 단편을 코딩함)을 함유한다.

도 9는 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱에 사용되는 신호 펩티드 서열, STRC 단백질의 N-말단 부분 및 N-인테인을 코딩하는 도 8a 및 8b의 뉴클레오타이드 서열을 함유하는 구축물의 개략적 표현을 보여준다.

도 10은 도 7a 및 7b에 제시된 뉴클레오타이드 서열에 의해 코딩되는 인간 STRC 단백질의 N-말단 부분, 신호 펩티드 서열 및 N-인테인을 함유하는 아미노산 서열 (서열식별번호: 6)을 보여준다.

도 11은 도 8a 및 8b에 제시된 뉴클레오타이드 서열에 의해 코딩되는 뮤린 STRC 단백질의 N-말단 부분, 신호 펩티드 서열 및 N-인테인을 함유하는 아미노산 서열 (서열식별번호: 8)을 보여준다.

도 12a 및 12b는 인간 STRC 단백질의 C-말단 부분, 신호 펩티드 서열 및 C-인테인 (예를 들어, CFS-C-Strc-C-Int, 구축물 2, C-부분)을 코딩하는 뉴클레오타이드 서열 (서열식별번호: 17)을 보여준다. 뉴클레오타이드 서열은 신호 서열, C-인테인 서열 (인테인 단백질의 C-말단 단편을 코딩함), 3' Strc (야생형 Strc 코딩 서열의 3' 단편), 링커 서열 및 myc 태그 서열을 함유한다.

도 13a 및 13b는 뮤린 STRC 단백질의 C-말단 부분, 신호 펩티드 서열 및 C-인테인 (예를 들어, CFS-C-Strc-C-Int, 구축물 2, C-부분)을 코딩하는 뉴클레오타이드 서열 (서열식별번호: 19)을 보여준다. 뉴클레오타이드 서열은 신호 서열, C-인테인 서열 (인테인 단백질의 C-말단 단편을 코딩함), 3' Strc (야생형 Strc 코딩 서열의 3' 단편), 링커 서열 및 myc 태그 서열을 함유한다.

도 14는 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱에 사용되는 신호 펩티드 서열, C-인테인 및 STRC 단백질의 C-말단 부분을 코딩하는 도 13a 및 13b의 뉴클레오타이드 서열을 함유하는 구축물의 개략적 표현을 보여준다.

도 15는 도 12a 및 12b에 제시된 뉴클레오타이드 서열에 의해 코딩되는 신호 펩티드 서열, C-인테인, 인간 STRC 단백질의 C-말단 부분, 링커 서열 및 myc 태그를 함유하는 아미노산 서열 (서열식별번호: 18)을 보여준다.

도 16은 도 13a 및 13b에 제시된 뉴클레오타이드 서열에 의해 코딩되는 신호 펩티드 서열, C-인테인, 뮤린 STRC 단백질의 C-말단 부분, 링커 서열 및 myc 태그를 함유하는 아미노산 서열 (서열식별번호: 20)을 보여준다.

도 17은 도 8a, 8b, 13a 및 13b의 서열에 의해 생성되고 도 9 및 14의 구축물에 의해 도시된 바와 같은 도 11 및 16의 시스테인 (C; Cys), 페닐알라닌 (F, Phe) 및 세린 (S; Ser)을 함유하는 명시된 CFS 분할 부위를 갖는 스테레오실린 (STRC) 단백질의 예측 구조를 보여준다.

도 18a 및 18b는 STRC 단백질의 N-말단 부분, 신호 펩티드 서열 및 N-인테인 (예를 들어, CFS-N-Strc-N-Int, 구축물 1, N-부분)을 코딩하는 뉴클레오타이드 서열 (서열식별번호: 51)을 보여준다. 뉴클레오타이드 서열은 신호 서열, 5' Strc (야생형 Strc 코딩 서열의 5' 단편) 및 N-인테인 서열 (인테인 단백질의 N-말단 단편을 코딩함)을 함유한다.

도 19는 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱에 사용된 신호 펩티드 서열, STRC 단백질의 N-말단 부분 및 N-인테인을 코딩하는 도 18a 및 18b의 뉴클레오타이드 서열을 함유하는 구축물의 개략적 표현을 보여준다.

도 20은 도 18a 및 18b에 제시된 뉴클레오타이드 서열에 의해 코딩되는 STRC 단백질의 N-말단 부분, 신호 펩티드 서열 및 N-인테인을 함유하는 아미노산 서열 (서열식별번호: 52)을 보여준다.

도 21a 및 21b는 STRC 단백질의 C-말단 부분, 신호 펩티드 서열 및 C-인테인 (예를 들어, CFS-C-Strc-C-Int, 구축물 1, C-부분)을 코딩하는 뉴클레오타이드 서열 (서열식별번호: 53)을 보여준다. 뉴클레오타이드 서열은 신호 서열, C-인테인 서열 (인테인 단백질의 C-말단 단편을 코딩함), 3' Strc (야생형 Strc 코딩 서열의 3' 단편), 링커 서열 및 myc 태그 서열을 함유한다.

도 22는 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱에 사용되는 신호 펩티드 서열, C-인테인 및 STRC 단백질의 C-말단 부분을 코딩하는 도 21a 및 21b의 뉴클레오타이드 서열을 함유하는 구축물의 개략적 표현을 보여준다.

도 23은 도 21a 및 21b에 제시된 뉴클레오타이드 서열에 의해 코딩되는 신호 펩티드 서열, C-인테인, STRC 단백질의 C-말단 부분, 링커 서열 및 myc 태그를 함유하는 아미노산 서열 (서열식별번호: 54)을 보여준다.

도 24는 도 8a, 8b, 13a 및 13b의 서열을 사용하여 라인 4에서 단리된 STRC 단백질의 웨스턴 블롯에 의해 입증된 바와 같이 및 도 9 및 14의 구축물 (AAV2/AAV9-Php.B-STRC-구축물 2)에 의해 도시된 바와 같이 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱 및 프로세싱을 확인시켜 준다.

도 25는 도 8a, 8b, 13a 및 13b의 서열을 사용하여 라인 6에서 단리된 STRC 단백질의 웨스턴 블롯에 의해 입증된 바와 같이 및 신호 서열이 결여된 라인 4와 대조적으로 신호 서열을 갖는 도 9 및 14의 구축물 (AAV2/AAV9-Php.B-STRC-구축물 2)에 의해 도시된 바와 같이 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱 및 프로세싱에서의 신호 서열의 유용성을 확인시켜 준다.

도 26은 야생형 (WT) 마우스 (Strc^{WT/WT}) 및 STRC 녹아웃 마우스 (Strc^{-/-})와 비교하여 도 9 및 14의 구축물로 감염된 STRC 녹아웃 마우스 (Strc^{-/-})에서의 음압 수준 (데시벨, dB)의 회복에 의해 입증된 바와 같이 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱 및 프로세싱을 사용한 청력 상실의 회복을 보여준다.

도 27은 Strc 녹아웃 마우스에서 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱 시스템 (구축물 2: AAV2/AAV9-PHP.B-CMV-Strc-N; AAV2/AAV9-PHP.B-CMV-Strc-C)으로의 처리에 의한 청각 기능의 회복을 입증하는 ABR 및 DPOAE 결과를 보여준다.

도 28은 도 9에 도시된 STRC 단백질의 N-말단 부분을 코딩하는 구축물만을 사용한 경우 Strc 녹아웃 마우스에서의 청각 기능 회복의 결여를 입증하는 ABR 및 DPOAE 결과를 보여준다.

도 29는 도 14에 도시된 STRC 단백질의 C-말단 부분을 코딩하는 구축물만을 사용한 경우 Strc 녹아웃 마우스에서의 청각 기능 회복의 결여를 입증하는 ABR 및 DPOAE 결과를 보여준다.

도 30은 본원에 기재된 이중-AAV 인테인-매개 단백질 트랜스-스플라이싱 시스템 (구축물 2: AAV2/AAV9-PHP.B-CMV-Strc-N; AAV2/AAV9-PHP.B-CMV-Strc-C)으로의 Strc 녹아웃 마우스의 처리 후 생체내에서의 시간 경과에 따른 ABR 및 DPOAE 결과를 보여준다.

도 31a-31c는 인테인-매개 단백질 재조합을 사용하는 이중 벡터 전략을 제공한다. 도 31a는 4개의 상이한 이중 벡터 변이체를 포함하는 생성된 8개의 AAV2 플라스미드를 제공한다. N-말단 및 C-말단 인테인을 4개의 변이체 각각에 대해 지시된 부위에서 인-프레임으로 융합시켰다. 변이체 1 및 2는 그의 분할 부위가 상이하며, 여기서 천연 시스테인은 위치 747 (변이체 1) 및 위치 970 (변이체 2)에 위치한다. 변이체 3 및 4는 각각 동일한 분할 부위 1 및 2를 가졌다. 추가로, 변이체 3 및 4는 C-인테인 서열의 상류에, C-말단 단편의 N-말단에 융합된 STRC의 N-말단에서 발견된 신호 서열을 가졌다. 인테인-매개 단백질 재조합은 전장 STRC 및 절제된 인테인 단편을 생성할 것으로 예측되었다. Myc 태그 (제시되지 않음)를 모든 C-말단 단편의 C-말단에 융합시켰다. 도 31b는 4개의 변이체에 대한 분할 부위 및 주위 아미노산 서열을 보여준다. 도 31c는 전장 STRC를 코딩하는 플라스미드 (레인 1), 비-형질감염된 대조군 (레인 2), 변이체 1 (레인 3) 및 변이체 3 (레인 5)의 C-말단 단편을 코딩하는 플라스미드 및 변이체 1 (레인 4) 및 변이체 3 (레인 6)에 대한 N- 및 C-단편 둘 다의 공동-형질감염으로 형질감염된 인간 배아 신장 (HEK) 293 세포로부터의 용해물의 대표적인 웨스턴 블롯 분석을 제공한다. 항

-Myc 항체를 사용하여 C-말단 단편 (120 kD) 및 전장 STRC (220 kD)를 확인하였다.

도 32a-32g는 Strc^{Δ/Δ} 마우스의 생성 및 특징화를 보여준다. 도 32a는 WT Strc의 파괴를 위한 CRISPR/Cas9 전략을 예시한다. 엑손 4를 표적화하도록 3개의 가이드 RNA (sgRNA)를 설계하였다. 유전자 파괴 전략은 249개의 뉴클레오타이드의 결실 및 2개의 전위 및 역위 (947-1139 - 자주색 및 1758-1835 - 황색)를 생성하였으며, 이는 돌연변이 Strc 대립유전자에 조기 정지 코돈을 도입하였다. 도 32b는 게놈 DNA를 증폭시키는 데 사용된 PCR의 결과를 제공하며, 이는 겔 상에 전개되었을 때 WT (1kB) 및 돌연변이 Strc (751 bp) 대립유전자에 대해 명확한 밴드를 생성하였다. 도 32c-32d는 유모 다발을 조명하기 위해 알렉사488 2차 및 팔로이딘-알렉사555, 항-STRC 항체로 염색된 조직으로부터 취한 WT 및 Strc^{Δ/Δ} 와우의 공초점 영상을 예시한다. 내유모 세포 (IHC) 및 3열의 외유모 세포 (OHC)가 가시적이다. 녹색 STRC 염색은 WT의 OHC 상에 존재하는 반면, 적색 액틴 염색은 WT 및 Strc^{Δ/Δ} 둘 다의 IHC 상에 존재한다. 축척 막대 = 10 mm. 도 32e는 Strc^{Δ/+} (Het-흑색 원형) 및 Strc^{Δ/Δ} (Homo-적색 마름모형) 마우스의 IHC 및 OHC로부터 측정된 평균 ± S.D. 감각 전달 전류 진폭을 보여준다. 도 32f는 Strc^{D/D} 마우스 (n=6; 적색) 및 WT 마우스 (n=5; 흑색)로부터 수득된 평균 ± S.D. DPOAE 역치를 보여준다. 도 32g는 Strc^{Δ/Δ} 마우스 (n=6; 적색) 및 WT 마우스 (n=5; 흑색)로부터 수득된 평균 ± S.D. ABR 역치를 예시한다.

도 33a-b는 이중 AAV 전달이 STRC 발현 및 유모 다발 형태를 회복시킨다는 것을 입증한다. 도 33a는 알렉사488 접합된 2차 (녹색) 및 알렉사546-팔로이딘 (적색)을 갖는 항-STRC 항체로 염색된 WT (좌측), Strc^{Δ/Δ} (중간) 및 이중 벡터 주사된 Strc^{Δ/Δ} (우측) 와우의 공초점 영상을 제공한다. 축척 막대 = 10 mm. 상부 열의 영상은 합병된 2개의 채널을 보여준다. 하부 열은 STRC 국제화 단독을 보여준다. 도 33b는 WT (상단), Strc^{Δ/Δ} (중간) 및 이중 벡터 주사된 Strc^{Δ/Δ} (하단) 외유모 세포 다발의 스캐닝 전자 현미경검사 영상을 보여준다. 조직을 수거하고, 고정시키고, 상기 나타낸 바와 같이 4 또는 12주에 영상화하였다. 축척 막대 = 5 mm (좌측 & 중간) 또는 2 mm (우측).

도 34a-34d는 이중 AAV 전달이 DPOAE 및 ABR 역치를 회복시킨다는 것을 보여준다. 도 34a는 WT 마우스 와우 (상부 트레이스)에서 자극 주파수 f1 (13.3 kHz) 및 f2 (16 kHz)에서의 2개의 주파수 성분 및 예측 주파수 2f1 - f2 (10.6 kHz)에서의 변조물을 나타낸 DPOAE 파형의 푸리에 분석을 제공한다. 하기 트레이스는 WT (좌측), Strc^{Δ/Δ} (중간) 및 이중 벡터 주사된 Strc^{Δ/Δ} (우측) 와우에 대한 확장된 주파수 및 진폭 스케일 상에서의 10 내지 50 dB의 음압 수준에 대한 변조물을 나타낸다. 볼드 트레이스는 DPOAE 역치를 나타낸다. 도 34b는 WT (흑색) 및 이중 벡터 주사된 Strc^{Δ/Δ} 마우스 중 회복이 있는 마우스 (자주색) 및 회복이 없는 마우스 (Strc^{Δ/Δ}; 적색; n=20; 상단 수평선)에 대한 (f2) 자극 주파수의 함수로서의 DPOAE 역치를 제공한다. 적색 및 자주색 선은 개체로부터의 데이터를 나타낸다. 평균 ± S.E.는 WT (흑색; n=5; 가장 하부의 선), 회복된 20마리의 이중 벡터 주사된 Strc^{Δ/Δ} 마우스 (자주색; n=20; 상단으로부터 두번째 선) 및 5마리의 최상의 회복 마우스 (녹색; n=5; 상단으로부터 세번째 선)에 대해 제시된다. 도 34c는 25 내지 110 dB의 음압 수준에 의해 유발된, WT (좌측), Strc^{Δ/Δ} (중간) 및 이중 벡터 주사된 Strc^{Δ/Δ} (우측) 와우로부터 기록된 ABR 트레이스의 패밀리를 예시한다. 볼드 트레이스는 ABR 역치를 나타낸다. 도 34d는 WT (흑색; n=5; 가장 하부의 선) 및 이중 벡터 주사된 Strc^{Δ/Δ} 마우스 중 회복이 있는 마우스 (자주색; n=20; 상단으로부터 두번째) 및 회복이 없는 마우스 (Strc^{Δ/Δ}; 적색; n=20; 상단 선)에 대한 자극 주파수의 함수로서 플롯팅된 ABR 역치를 제공한다. 자주색 선은 개체로부터의 데이터를 나타낸다. 평균 ± S.E.는 WT (흑색; n=5; 가장 하부의 선), 이중 벡터 주사된 Strc^{Δ/Δ} 마우스 중 회복이 있는 마우스 (자주색; n=20; 상단으로부터 두번째 선) 및 회복이 없는 마우스 (Strc^{Δ/Δ}; 적색; n=20; 상단 선) 및 가장 양호한 회복을 갖는 마우스 (녹색; n=5; 상단으로부터 세번째 선)에 대해 제시된다.

발명을 실시하기 위한 구체적인 내용

[0060]

본 개시내용의 상세한 실시양태가 본원에 개시되지만; 개시된 실시양태는 단지 다양한 형태로 구현될 수 있는 개시내용의 예시인 것으로 이해되어야 한다. 추가로, 본 개시내용의 다양한 실시양태와 관련하여 주어진 각각의 실시예는 예시적인 것으로 의도되고, 제한적인 것으로 의도되지 않는다.

- [0061] 스테레오실린 (STRC)의 발현을 통해 청력을 회복시키는 조성물 및 방법이 본원에 제공된다. 하나는 신호 서열, 유전자 코딩 서열의 5' 말단 단편 및 인테인의 아미노 말단 단편 (N-인테인, 또한 스플릿 인테인-N으로도 공지됨)을 코딩하는 서열을 포함하고, 하나는 신호 서열, 인테인의 카르복시 말단 단편 (C-인테인, 또한 스플릿 인테인-C로도 공지됨)을 코딩하는 서열 및 유전자 코딩 서열의 3' 말단 단편을 포함하는 2개의 별개의 AAV 입자 또는 벡터를 사용한 관심 유전자로의 처리가 제공된다.
- [0062] 달리 정의되지 않는 한, 본원에 사용된 모든 기술 과학 용어는 본 발명이 속하는 기술분야의 통상의 기술자에 의해 통상적으로 이해되는 의미를 갖는다. 하기 참고문헌은 관련 기술분야의 통상의 기술자에게 본 발명에 사용된 많은 용어의 일반적 정의를 제공한다: 문헌 [Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); 및 Hale & Marham, The Harper Collins Dictionary of Biology (1991)].
- [0063] 정의
- [0064] 본원에 사용된 모든 용어는 달리 제공되지 않는 한 관련 기술분야에서의 그의 통상의 의미를 갖는 것으로 의도된다. 모든 농도는, 달리 정의되지 않는 한, 국소 조성물의 전체 중량에 대한 명시된 성분의 중량 백분율에 대한 것이다.
- [0065] 본원에 사용된 단수형은 하나 이상을 의미할 것이다. 단어 "포함하는"과 함께 사용되는 경우에 본원에 사용된 단수형 단어는 하나 또는 하나 초과를 의미한다. 본원에 사용된 "또 다른"은 적어도 두번째 이상을 의미한다.
- [0066] "아데노-연관 바이러스" (AAV)는 인간 및 일부 다른 영장류 종을 감염시킬 수 있는 소형 바이러스이다. AAV를 사용하는 벡터는 숙주 세포의 게놈 내로 통합되지 않으면서 분열 및 휴지 세포를 감염시킬 수 있다. 이들 특색은 AAV를 유전자 요법 바이러스 벡터에 대한 매력적인 후보로 만든다.
- [0067] "AAV9-php.b 벡터"는 세포, 예를 들어 내이의 세포를 형질감염시킬 수 있는 AAV9-php.b 폴리뉴클레오티드 또는 그의 단편을 포함하는 바이러스 벡터인 아데노-연관 바이러스 혈청형 9를 의미한다. 한 실시양태에서, AAV9-php.b 벡터는 세포의 적어도 70% 이상 (예를 들어, 75%, 80%, 85%, 90%, 95%, 97%, 99%, 100%)을 형질감염시킨다. 또 다른 실시양태는 AAV9-php.b 벡터를 대상체의 내이에 투여하거나 AAV9-php.b 벡터를 시험관내에서 내이로부터 유래된 세포와 접촉시킨 후에 내유모 세포 및/또는 외유모 세포의 적어도 70% 이상 (예를 들어, 75%, 80%, 85%, 90%, 95%, 97%, 99%, 100%)을 형질감염시킬 수 있는 AAV9-php.b 폴리뉴클레오티드 또는 그의 단편을 포함하는 AAV9-php.b 벡터에 관한 것일 수 있다. 다른 실시양태에서, 내유모 세포의 적어도 85% (예를 들어, 90%, 95%, 100%) 및/또는 외유모 세포의 적어도 85% (예를 들어, 90%, 95%, 100%)가 AAV9-php.b 벡터로 형질감염된다. 형질감염 효율은 마우스 모델에서 표지 또는 태그 (예를 들어, 녹색 형광 단백질 (GFP)을 코딩하는 유전자)를 사용하여 평가될 수 있다.
- [0068] 본 개시내용의 한 실시양태는 목적하는 단백질을 코딩하는 뉴클레오티드 서열을 포함하는 적어도 1개의 벡터 (예를 들어, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터 (예를 들어, 렌티바이러스), 아데노바이러스, AAV, AAV 게놈)에 관한 것일 수 있다 (도 1). 도 2a-2c는 하기와 같은 5'에서 3' 방향으로 인간 STRC 유전자 코딩 서열 (서열식별번호: 1) (도 3에서 인간 STRC 단백질 서열 (대문자)인 서열식별번호: 2를 코딩함)을 함유하는 뉴클레오티드 서열 (서열식별번호: 33)을 보여준다:

5'-GTGACTCTGGCCCCTACTGGGCCTCATTCCCTGGACCCTGGTCTCTCCTTCCTGAAGTC
 ATTGCTCTCCACTCTGGACCAGGCTCCCCAGGGCTCCCTGAGCCGCTCACGGTTCTTTACAT
 TCCTGGCCAACATTTCTTCTTCTTTGAGCCTGGGAGAATGGGGGAAGGACCAGTAGGAGAG
 CCCCCACCTCTCCAGCCGCTGCTCTGCGGCTCCATGATTTTCTAGTGACACTGAGAGGTAG
 CCCCAGCTGGGAGCCAATGCTAGGGCTGCTAGGGGATATGCTGGCACTGCTGGGACAGGAGC
 AGACTCCCCGAGATTTCTTGGTGCACCAGGCAGGGGTGCTGGGTGGACTTGTGGAGGTGCTG
 CTGGGAGCCTTAGTTCCTGGGGGGCCCCCTACCCCAACTCGGCCCCCATGCACCCGTGATGG
 GCCGTCTGACTGTGTCTGGCTGCTGACTGGTTGCCTTCTCTGCTGCTGTTGTTAGAGGGCA
 CACGCTGGCAAGCTCTGGTGCAGGTGCAGCCCAGTGTGGACCCACCAATGCCACAGGCCTC
 GATGGGAGGGAGGCAGCTCCTCACTTTTTGCAGGGTCTGTTGGGTTTGCTTACCCCAACAGG
 GGAGCTAGGCTCCAAGGAGGCTCTTTGGGGCGGTCTGCTACGCACAGTGGGGGGCCCCCTCT
 ATGCTGCCTTTCAGGAGGGGCTGCTCCGTGTCACTCACTCCCTGCAGGATGAGGTCTTCTCC
 ATTTTGGGGCAGCCAGAGCCTGATACCAATGGGCAGTGCCAGGAGGTAACCTTCAACAGCT
 GCTCTTATGGGGCGTCCGGCACAACTTTCCTGGGATGTCCAGGCGCTGGGCTTTCTGTCTG

[0069]

GATCACCACCCCCACCCCTGCCCTCCTTCACTGCCTGAGCACGGGCGTGCCCTCTGCCAGA
GCTTCTCAGCCGTCAGCCCACATCAGCCCACGCCAACGGCGAGCCATCACTGTGGAGGCCCT
CTGTGAGAACCACCTTAGGCCCAGCACCACCCTACAGCATTTCCAACCTTCTCCATCCACTTGC
TCTGCCAGCACACCAAGCCTGCCACTCCACAGCCCCATCCCAGCACCCTGCCATCTGCCAG
ACAGCTGTGTGGTATGCAGTGTCTTGGGCACCAGGTGCCCAAGGCTGGCTACAGGCCTGCCA
CGACCAGTTTCTGATGAGTTTTTGGATGCGATCTGCAGTAACCTCTCCTTTTCAGCCCTGT
CTGGCTCCAACCGCCGCTGGTGAAGCGGCTCTGTGCTGGCCTGCTCCACCCCTACCAGC
TGCCCTGAAGGCCTGCCCCCTGTTCCCTCACCCAGACATCTTTGGGGCTGCTTCTTGGA
GAATGAGACTCTGTGGGCTGAGCGACTGTGTGGGGAGGCAAGTCTACAGGCTGTGCCCCCA
GCAACCAGGCTTGGGTCCAGCATGTGTGCCAGGGCCCCACCCAGATGTCACTGCCTCCCCA
CCATGCCACATTGGACCCTGTGGGGAACGCTGCCCGGATGGGGGCAGCTTCTGGTGATGGT
CTGTGCCAATGACACCATGTATGAGTCTTGGTGCCCTTCTGGCCTGGCTAGCAGGCCAAT
GCAGGATAAGTCGTGGGGCAATGACACTTGCTTCCTAGAAGGGCTGCTGGGCCCCCTTCTG
CCCTCTCTGCCACCACTGGGACCATCCCCACTCTGTCTGACCCCTGGCCCCCTCCTCCTGG
CATGCTATCCCAGTTGCCACGCTGTCACTCCTCTGTCCCAGCTCTTGTCTACCCACACGCC
TACACTATCTCCTCCGCTGCTGACCTTCTCTTGGGTCCAGGGGCTGGGGGCGCTGAGGCC
CAGGGGATGCTGGGTCGGGCCCTACTGCTCTCCAGTCTCCCAGACAACTGCTCCTTCTGGGA
TGCCTTTCGCCCAGAGGGCCGGCGCAGTGTGCTACGGACGATTGGGGAATACCTGGAACAAG
ATGAGGAGCAGCCAACCCCATCAGGCTTTGAACCCACTGTCAACCCAGCTCTGGTATAAGC
AAGATGGAGCTGCTGGCCTGCTTTAGTCCTGTGCTGTGGGATCTGCTCCAGAGGAAAAGAG
TGTTTGGGCCCTGCAGATTCTAGTGAGGCGTACCTGCATATGCCCCAGAAAACCTCCAGC
AGCTGGTGCTTTTCAGCAGAGAGGGAGGCTGCACAGGGCTTCTGACACTCATGCTGCAGGGG
AAGCTGCAGGGGAAGCTGCAGGTACCACCATCCGAGGAGCAGGCCCTGGGTGCGCTGACAGC
CCTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAGCTCTTCATTGACCTGTCACCACTCA
TCCCTTTCTTGGCTGTCTCTGACCTGATGCGCTTCCCACCATCCCTGTTAGCCAACGACAGT
GTCTTGGCTGCCATCCGGGATTACAGCCCAGGAATGAGGCCTGAACAGAAGGAGGCTCTGGC
AAAGCGACTGCTGGCCCCCTGAACGTGTTTGGGGAAGTGCTGCTGGCCCCAGGAGCTGCTGT
GGGCAGTGCTGCCCCCTGCTCCCCACCTCCCTCTGGAGAACTTTTTCAGCTCAGCCCTCAC
CAGATCCAGGCCCTGGAGGATAGCTGGCCAGCAGCAGGTCTGGGGCCAGGGCATGCCCGCCA
TGTGCTGCGCAGCCTGGTAAACCAGAGTGTCCAGGATGGTGAGGAGCAGGTACGCAGGCTTG
GGCCCCCTGCGCTGTTTCTGAGCCCTGAGGAGCTGCAGAGCCTAGTGCCCCTGAGTGATCCA
ACGGGGCCAGTAGAACGGGGGCTGCTGGAATGTGCAGCCAATGGGACCCTCAGCCCAGAAGG
ACGGGTGGCATATGAACTTCTGGGTGTGTTGCGCTCATCTGGAGGAGCGGTGCTGAGCCCC
GGGAGCTGCGGGTCTGGGCCCTCTCTTCTCTCAGCTGGGCCTCCGCTTCTTCAGGAGCTG

[0070]

TCAGAGCCCCAGCTTAGAGCCATGCTTCCTGTCCTGCAGGGAAGTAGTGTACACCTGCTCA
GGCTGTCTGCTGCTTGGACGGCTCCTTCCTAGGCACGATCTATCCCTGGAGGAACTCTGCT
CCTTGACACTTCTGCTACCAGGCCTCAGCCCCAGACACTCCAGGCCATCCCTAGGCGAGTC
CTGGTCGGGGCTTGTTCCTGCCTGGCCCCCTGAAGTGTACGCCTCTCAGCCTGCCAGACCGC
AGCACTGCTGCAGACCTTTCGGGTAAAGATGGTGTAAAAATATGGGTACAACAGGTGCTG
GTCCAGCTGTGTGTATCCCTGGTCAGCCTATCCCACCACCTGGCCAGACTGCCTGCTTCCC
CTGCTCCCATTAAAGCTGCTACAACCTGGATTCCCTTGCTCTTCTGGCAAATCGAAGACGCTA
CTGGGAGCTGCCCTGGTCTGAGCAGCAGGCACAGTTTCTCTGGAAGAAGATGCAAGTACCCA
CCAACCTTACCCTCAGGAATCTGCAGGCTCTGGGCACCCTGGCAGGAGGCATGTCCTGTGAG
TTTCTGCAGCAGATCAACTCCATGGTAGACTTCCTTGAAGTGGTGCACATGATCTATCAGCT
GCCCCTAGAGTTTCGAGGGAGCCTGAGGGCCTGTATCTGGGCAGAGCTACAGCGGAGGATGG
CAATGCCAGAACCAGAATGGACAACCTGTAGGGCCAGAAGTGAACGGGCTGGATAGCAAGCTA
CTCCTGGACTTACCGATCCAGTTGATGGACAGACTATCCAATGAATCCATTATGTTGGTGGT
GGAGCTGGTGCAAAGAGCTCCAGAGCAGCTGCTGGCACTGACCCCCCTCCACCAGGCAGCCC
TGCCAGAGAGGGCACTACAAAACCTGGCTCCAAAGGAGACTCCAGTCTCAGGGGAAGTGCTG
GAGACCTTAGGCCCTTTGGTTGGATTCTGGGGACAGAGAGCACACGACAGATCCCCCTACA
GATCCTGCTGTCCATCTCAGTCAGCTGCAAGGCTTCTGCCTAGGAGAGACATTTGCCACAG
AGCTGGGATGGCTGCTATTGCAGGAGTCTGTTCTTGGGAAACCAGAGTTGTGGAGCCAGGAT
GAAGTAGAGCAAGCTGGACGCCTAGTATTCACTCTGTCTACTGAGGCAATTTCTTGTATCCC
CAGGGAGGCCTTGGGTCCAGAGACCTGGAGCGGCTTCTAGAAAAGCAGCAGAGCTGGGAGC
AGAGCAGAGTTGGACAGCTGTGTAGGGAGCCACAGCTTGCTGCCAAGAAAGCAGCCCTGGTA
GCAGGGGTGGTGCGACCAGCTGCTGAGGATCTTCCAGAACCTGTGCCAAATTGTGCAGATGT
ACGAGGGACATTTCCAGCAGCCTGGTCTGCAACCCAGATTGCAGAGATGGAGCTCTCAGACT
TTGAGGACTGCCTGACATTATTTGCAGGAGACCCAGGACTTGGGCTGAGGAACTGCGGGCA
GCCATGGGCAAAGCAAAACAGTTGTGGGGTCCCCCGGGGATTTCTGCTCTGAGCAGATCCT
GCAGCTTGGTAGGCTCTTAATAGGTCTAGGAGATCGGGAACTACAGGAGCTGATCCTAGTGG
ACTGGGGAGTGTGAGCACCTGGGGCAGATAGATGGCTGGAGCACCCTCAGCTCCGCATT
GTGGTCTCCAGTTTCTACGGCAGAGTGGTCGGCATGTGAGCCACCTGGACTTCGTTTCATCT
GACAGCGCTGGGTATACTCTCTGTGGACTGCGGCCAGAGGAGCTCCAGCACATCAGCAGTT
GGGAGTTTCAGCCAAGCAGCTCTCTTCCTCGGCACCCTGCATCTCCAGTGCTCTGAGGAACAA
CTGGAGGTTCTGGCCACCTACTTGTACTGCCTGGTGGGTTTGGCCCAATCAGTAACTGGGG
GCCTGAGATCTTCACTGAAATTGGCACCATAGCAGCTGGGATCCCAGACCTGGCTCTTTCAG
CACTGCTGCGGGACAGATCCAGGGCGTTACTCCTCTTGCCATTTCTGTATCCCTCCTCCT
AAATTTGCTGTGGTGTTTAGTCCCATCCAACCTATCTAGTCTCACCAGTGCTCAGGCTGTGGC

[0071]

TGTCACCTCTGAGCAAATGGCCTTTCTGAGTCCTGAGCAGCGACGAGCAGTTGCATGGGCCC
AACATGAGGGAAAGGAGAGCCAGAACAGCAAGGTGGAAGTACAGCCTGGGGCCTCCAGGAC
TGGTCACGACCTTCCTGGTCCCTGGTATTGACTATCAGCTTCCTTGGCCACCTGCTA-3'

[0072]

(SEQ ID NO:1).

[0073]

추가 실시양태는 하기와 같은 5'에서 3' 방향으로 묶린 STRC 유전자 코딩 서열 (서열식별번호: 3) (도 5에서 묶린 STRC 단백질 서열인 서열식별번호: 4를 코딩함)을 포함하는 적어도 1개의 벡터 (예를 들어, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 계통) (예를 들어, 도 4a-4d; 서열식별번호: 38 참조)에 관한 것일 수 있다:

5'-GCCCCCTACTGGGCCTCAGTCTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTTCGTAGC
 CACTCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTCACGGTTCTCTGCGTTCCTGGCCA
 ACATTTCTTCATCCTTCCAGCTTGGGAGGATGGGGGAGGGACCGGTGGGAGAGCCCCACCT
 CTCCAGCCCCCTGCACTTCGACTTCATGATTTCCCTCGTGACACTGAGAGGTAGCCCAGACTG
 GGAGCCAATGCTAGGGCTTCTGGGAGATGTGCTGGCACTCCTGGGACAGGAACAGACTCCCC
 GGGACTTTTTTGGTGCACCAGGCAGGTGTACTGGGTGGACTTGTAGAGGCATTGTTGGGAGCG
 TTAGTTCTTGAGGCCCCCTGCCCCACTCGACCCCCATGCACCCGTGATGGCCCTTCTGA
 CTGTGTCTGGCTGCTGATTGGTTGCCTTCTCTGATGTTGTTATTAGAGGGTACACGCTGGC
 AGGCCCTGGTGCAGTTGCAGCCCAGTGTGGACCCAACCAATGCCACAGGTCTTGATGGTAGA
 GAGCCAGCTCCTCACTTTTTACAGGGTCTGCTGGGCTTGCTTACCCCAGCAGGAGAGTTGGG
 CTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGTGGGGGCCCCCTCTATGCTGCCT
 TCCAGGAGGGGCTACTGCGAGTCACTCATTTCTCTGCAAGATGAGGTCTTTTCTATTATGGGA
 CAGCCAGAGCCTGATGCCAGTGGGCAGTGCCAGGGAGGCAACCTTCAACAGCTGCTTTTATG
 GGGCATGCGGAACAACCTTTCTTGGGACGCCCAGCACTGGGTTTTCTATCTGGATCACCAC
 CTCCACCCCTGCTCTCCTGCACTGCCTGAGCAGAGGTGTGCCTCTGCCAGGGCTTCCCAG
 CCTGCGGCTCACATCAGCCCTCGACAGCGGCGAGCCATCTCTGTGGAGGCCCTCTGCGAGAA
 CCACTCAGGCCCAGAGCCACCCTACAGCATCTCCAATTCTCCATCTACTTGCTCTGCCAGC
 ACATCAAGCCTGCCACCCCGCGGCCCCCTCCTACCACCCACGGCCTCCTCCTACCACCCCA
 CAGCCCCCTCCTACCCTACACAGCCCATTCCTGACACTACACAGCCCCCTCCTGTACCCCC
 AAGGCCTCCTCCTACCACCCCAACCCCCCTCCTAGCACAGCTGTCATCTGCCAGACAGCTG
 TATGTTACGCAGTCTCGTGGGCACCAGGTGCCCCGAGGTTGGCTCCAAGCCTGCCATGATCAG
 TTTCTGATCAATTTCTGGATATGATCTGCGGCAACCTCTCATTTTCAGCCCTGTCTGGCCC
 CAGTCGTCCTTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCCACCCCCACTAGCTGTCCAC
 CAGGCCTGATCCCTGTGCCCTCACCCAGAAATATTCTGGGGCTGTTTCTGGAGAATGAG

[0074]

ACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCAGGCTGTGCCCCGAGGAACCA
GGCTTGGGTTTCAGCATGTGTGTGGGGCCCCACCTTGGACGCCACTGATTTCCACCGTGCC
GCGTTGGACCCTGTGGGAACGCTGCCAGATGGGGGCAGCTTCCTGCTCATGGTCTGTGCC
AATGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTGGCTAGCAGGCCAATGCAGAAT
TAGTCGTGGAGGAAATGATACTTGCTTTCTAGAAGGCATGCTGGGCCCCCTTGTGCCCTCTC
TGCCCCCTCTGGGACCATCCCCACTCTGTCTGGCTCCTGGTCCTTTTCTGCTTGGCATGTTA
TCCCAGTTGCCACGCTGTCAGTCCTCCGTGCCAGCCCTCGCCACCCACGCGCCTACATTA
CCTCTGCGCCTACTGACCTTCCCTTCTGGGTCCAGGGACTGGGGGTGCCGAGACGCAGGGGA
TGTTAGGTCAAGCCCTGCTGCTCTCTAGTCTCCCAGACAACTGTTCAATTCTGGGATGCCTTC
CGCCCAGAGGGCCGGAGAAGTGTACTGAGGACAGTCGGAGAGTACTTGACGCGGAAGAGCC
AACCCACCCAGGCTTAGACTCCTCCCTCAGCCTCGGCTCTGGTATGAGCAAGATGGAGCTTC
TGTCTGCTTCAGTCCTGTACTGTGGGATCTACTCCAGAGAGAGAAGAGCGTTTGGGCCCTG
AGGACCCTGGTGAAGGCCTACCTGCGCATGCCCTCCAGAAGACCTTCAGCAGCTTGTGCTTTC
AGCAGAGATGGAGGCTGCACAGGGCTTCCTGACGCTCATGCTTCGTTTCTGGGCTAAGCTGA
AGGTTCAACCATCCGAGGAGCAGGCCATGGGCCGCTGACAGCCTTGCTGCTCCAGCGGTAC
CCACGCCTCACCTCCCAACTCTTTATCGACATGTCACCGCTCATCCCCCTTCTGGCTGTCCC
TGACCTCATGCGCTTCCCACCGTCCCTTTTGGCCAACGACAGTGTCTGGCTGCCATCAGGG
ATCACAGCTCAGGAATGAAGCCTGAACAGAAGGAGGCCCTGGCAAAACGACTGCTGGCCCCT
GAGCTGTTTGGAGAAGTGCCTGATTGGCCCCAGGAGCTGCTGTGGGCAGCCCTGCCTCTGCT
TCCCCATCTGCCTCTGGAGAGCTTTCTCCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGG
ATAGCTGGCCAGTAGCAGATCTTGGGCCGGGACACGCCCGACATGTGCTTCGTAGCCTAGTA
AACCAGAGCATGGAGGATGGGGAGGAGCAGGTGCTCAGGCTTGGGTCCCTCGCCTGTTTCCT
GAGTCTGAGGAGCTACAGAGTCTGGTGCCCTTGAGTGATCCAATGGGGCCTGTAGAACAGG
GTCTGCTGGAATGTGCGGCCAATGGGACCTCAGCCCAGAAGGACGGGTGGCATATGAACTT
CTGGGAGTGTTCGTTTCATCTGGAGGAACTGTCTTAAGCCCCGAGAGCTGAGGGTCTGGGC
ACCTCTCTTTCCCCAGCTGGGCCCTCCGCTTCTGCAGGAGCTCTCAGAGACCCAGCTTAGAG
CCATGCTTCTCTGCCCTACAGGGAGCCAGTGTACACCTGCCAGGCTGTTCTGTTGTTTGGA
AGGCTCCTTCCTAAGCATGATCTGTCCCTGGAGGAACTCTGCTCCCTGCACCCTCTCCTGCC
AGGTCTCAGCCCCCAGACACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGGTGCTTGTTCCT
GCCTGGGCCCTGAACTGTCAAGGCTTTCAGCTTGCCAGATTGCAGCTCTGCTGCAGACCTTT
CGGGTAAAAGATGGTGTAAAAATATGGGTGCAGCAGGTGCCGGCTCAGCCGTGTGCATTCC
TGGGCAGCCCACCACTTGGCCAGACTGCCTGCTTCCCCTGCTCCCATTAAAGCTGCTACAGC
TGGACGCTGCAGCTCTTCTGGCAAACCGAAGACTCTATCGGCAGCTGCCTTGGTCTGAGCAA
CAGGCACAGTTTCTCTGGAAGAAAATGCAAGTGCCTACCAACCTGAGCCTGAGGAATCTGCA

[0075]

[0076]

[0077]

[0078]

부분 코딩 서열의 5'-말단에 있는 신호 서열, 부분 코딩 서열에 플랭킹되거나 인접할 수 있는 스플라이스 공여자 서열 (예를 들어, N-말단 인테인 (N-인테인))을 코딩하는 하류 서열, 스플라이스 공여자 서열에 플랭킹되거나 인접한 하류 3' ITR 서열을 함유하는 제1 뉴클레오타이드 서열을 포함하는 제1 벡터 (예를 들어, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 게놈)를 제공할 수 있다 (예를 들어, AAV9-php.B-Prot-trans/공여자). 관심 단백질 (예를 들어, STRC)의 나머지 C-말단 부분을 코딩하는 하류 부분 코딩 서열에 바로 인접하여 위치하거나 플랭킹된 스플라이스 수용자 서열 (예를 들어, C-말단 인테인 (C-인테인))의 상류일 수 있는 신호 서열의 상류의 5' ITR 서열을 함유하는 제2 뉴클레오타이드 서열이며, 추가로 관심 단백질의 C-말단 부분을 코딩하는 부분 코딩 서열의 하류에 C-말단 myc 태그를 함유할 수 있는 제2 뉴클레오타이드 서열을 포함하는 제2 벡터 (예를 들어, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 게놈)를 제공할 수 있다 (예를 들어, AAV9-phpB-Prot-trans/수용자). 전장 관심 mRNA (예를 들어, STRC mRNA)는 2개의 벡터 (예를 들어, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 게놈)로 공동-감염된 세포에서 2개의 트랜스스플라이싱 플라스미드 사이의 머리-대-꼬리 재조합 (5'에서 3' 말단 대 5'에서 3' 말단), 전사 및 역전된 말단 반복부 (ITR) 접합부에 걸친 후속 스플라이싱에 의해 형성될 수 있다.

[0079] 관심 단백질이, 예를 들어 스테레오실린 (STRC) 단백질인 "관심 단백질의 N-말단 부분 또는 단편을 코딩하는 서열"은 STRC의 N-말단 부분 또는 단편의 부분 코딩 서열을 의미하며, 여기서 일부 경우에, "STRC의 N-말단 부분 또는 단편을 코딩하는 서열"은 STRC의 N-말단 부분 또는 단편의 부분 코딩 서열 (대문자)의 상류의 신호 펩티드 코딩 서열을 코딩하는 서열 (소문자, 이탤릭체 및 밑줄표시)을 포함할 수 있다. 일부 실시양태에서, "스테레오실린 (STRC)의 N-말단 단편을 코딩하는 서열"은 신호 펩티드 코딩 서열을 코딩하는 뉴클레오타이드 서열을 포함하지 않을 수 있다. 또 다른 실시양태는 그의 C-말단 단부에서 인테인의 N-말단 부분 또는 단편 (N-인테인) (볼드체 및 밑줄표시)에 융합된 "STRC의 N-말단 단편 또는 부분을 코딩하는 서열"을 포함하는 뉴클레오타이드 서열을 제공할 수 있으며, 여기서 5' 말단의 뉴클레오타이드 서열은 "ATG" 개시 코돈 (볼드체)으로 시작하고 정지 코돈 (대문자, 이탤릭체 및 밑줄표시)으로 끝난다.

[0080] "STRC의 N-말단 부분 또는 단편을 코딩하는 서열"을 포함하는 예시적인 인간 뉴클레오타이드 서열은 하기와 같을 수 있다:

5' ATG*gctctcagcctctgcccctgctgctgctgctgctgctgctgctgctgctgctgctccttt*
*gca*GTGACTCTGGCCCTACTGGGCTCATTCCCTGGACCTGGTCTCTCCTTCCTGAAGTC
ATTGCTCTCCACTCTGGACCAGGCTCCCCAGGGCTCCCTGAGCCGCTCACGGTTCTTTACAT
TCCTGGCCAACATTTCTTCTTCTTTGAGCCTGGGAGAATGGGGGAAGGACCAGTAGGAGAG
CCCCACCTCTCCAGCCGCTGCTCTGCGGCTCCATGATTTTCTAGTGACACTGAGAGGTAG
CCCCGACTGGGAGCCAATGCTAGGGCTGCTAGGGGATATGCTGGCACTGCTGGGACAGGAGC
AGACTCCCCGAGATTTCTGCTGTCACAGGCAGGGGTGCTGGGTGGACTTGTGGAGGTGCTG
CTGGGAGCCTTAGTTCTGGGGGGCCCCCTACCCCAACTCGGCCCCATGCACCCGTGATGG
GCCGTCTGACTGTGTCTGGCTGCTGACTGGTTGCCTTCTCTGCTGCTGTTGTTAGAGGGCA
CACGCTGGCAAGCTCTGGTGCAGGTGCAGCCAGTGTGGACCCACCAATGCCACAGGCCTC
GATGGGAGGGAGGCAGCTCCTCACTTTTTCAGGGTCTGTTGGGTTTGCTTACCCCAACAGG
GGAGCTAGGCTCCAAGGAGGCTCTTTGGGGCGGTCTGCTACGCACAGTGGGGGGCCCCCTCT
ATGCTGCCTTTCAGGAGGGGTGCTCCGTGTCACTCACTCCCTGCAGGATGAGGTCTTCTCC
ATTTTGGGGCAGCCAGAGCCTGATACCAATGGGCAGTGCCAGGGAGGTAACCTTCAACAGCT
GCTCTTATGGGGCGTCCGGCACAACTTTCTGGGATGTCCAGGCGCTGGGCTTTCTGTCTG

[0081]

GATCACCACCCCCACCCCTGCCCTCCTTCACTGCCTGAGCACGGGCGTGCCCTCTGCCCAGA
 GCTTCTCAGCCGTCAGCCCACATCAGCCCACGCCAACGGCGAGCCATCACTGTGGAGGCCCT
 CTGTGAGAACCACCTTAGGCCCAGCACCCCTACAGCATTTCCAACCTTCTCCATCCACTTGC
 TCTGCCAGCACACCAAGCCTGCCACTCCACAGCCCCATCCCAGCACCACTGCCATCTGCCAG
 ACAGCTGTGTGGTATGCAGTGTCTTGGGCACCAGGTGCCCAAGGCTGGCTACAGGCCTGCCA
 CGACCACTTCTCTGATGAGTTTTTGGATGCGATCTGCAGTAACCTCTCCTTTTCAGCCCTGT
 CTGGCTCCAACCGCCGCTTGGTGAAGCGGCTCTGTGCTGGCCTGCTCCACCCCTACCAGC
 TGCCCTGAAGGCCTGCCCCCTGTTCCCTCACCCAGACATCTTTGGGGCTGCTTCTTGGA
 GAATGAGACTCTGTGGGCTGAGCGACTGTGTGGGGAGGCAAGTCTACAGGCTGTGCCCCCA
 GCAACCAGGCTTGGGTCCAGCATGTGTGCCAGGGCCCCACCCAGATGTCACCTGCCTCCCCA
 CCATGCCACATTGGACCTGTGGGGAACGCTGCCCGGATGGGGGAGCTTCTTGGTGATGGT
 CTGTGCCAATGACACCATGTATGAGTCTTGGTGGCCTTCTGGCCTGGCTAGCAGGCCAAT
 GCAGGATAAGTCGTGGGGGCAATGACACTTGCTTCCTAGAAGGGCTGCTGGGCCCCCTTCTG
 CCTCTCTGCCACCACTGGGACCATCCCCACTCTGTCTGACCCCTGGCCCCCTCTCTCTGG
 CATGCTATCCCACTTGGCCAGCTGTGAGTCTCTGTCCCAGCTCTTGTCTACCCACACGCC
 TACACTATCTCTCCGCTGCTGACCTTCTCTTGGGTCCAGGGGCTGGGGGCGCTGAGGCC
 CAGGGGATGCTGGGTGCGGCCCTACTGCTCTCCAGTCTCCCAGACAACTGCTCCTTCTGGGA
 TGCCTTTCGCCCAGAGGGCCGCGCAGTGTGCTACGGACGATTGGGGAATACCTGGAACAAG
 ATGAGGAGCAGCCAACCCCATCAGGCTTTGAACCCACTGTCAACCCAGCTCTGGTATAAGC
 AAGATGGAGCTGCTGGCCTTGCCTGTCTACGAAACCGAGATACTGACAGTAGAATATGGCCT
TCTGCCAATCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTGATAACA
ATGTTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTC
GAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGT
CGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTG
ACAACCTTCCTAATTAATAG 3' (SEQ ID NO:5).

[0082]

[0083]

"STRC의 N-말단 부분 또는 단편을 코딩하는 서열"을 포함하는 또 다른 예시적인 뮤린 뉴클레오타이드 서열은 하기와 같을 수 있다:

5' ATGgctctgagcctccagccccagctgctccttctcctgctcgtcctgcccaggaagtga
acttcaGCCCCCTACTGGGCCTCAGTCTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTCTGT
 AGCCACTCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTCACGGTTCTCTGCGTTCCTGG
 CCAACATTTCTTCATCCTTCCAGCTTGGGAGGATGGGGGAGGGACCGGTGGGAGAGCCCCCA
 CCTCTCCAGCCCCCTGCACTTCGACTTCATGATTTCTCTGCTGACACTGAGAGGTAGCCCAGA
 CTGGGAGCCAATGCTAGGGCTTCTGGGAGATGTGCTGGCACTCTGGGACAGGAACAGACTC

[0084]

CCCGGGACTTTTTGGTGCACCAGGCAGGTGTACTGGGTGGACTTGTAGAGGCATTGTTGGGA
GCGTTAGTTTCTGGAGGCCCCCTGCCCCACTCGACCCCATGCACCCGTGATGGCCCTTC
TGACTGTGTCTGGCTGCTGATTGGTTGCCTTCTCTGATGTTGTATTAGAGGGTACACGCT
GGCAGGCCCTGGTGCAGTTGCAGCCAGTGTGGACCAACCAATGCCACAGGTCTTGATGGT
AGAGAGCCAGCTCCTCACTTTTTACAGGGTCTGCTGGGCTTGCTTACCCAGCAGGAGAGTT
GGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGTGGGGGCCCCCTCTATGCTG
CCTTCCAGGAGGGGCTACTGCGAGTCACTCATTCTCTGCAAGATGAGGTCTTTCTATTATG
GGACAGCCAGAGCCTGATGCCAGTGGGCAGTGGCAGGGAGGCAACCTTCAACAGCTGCTTTT
ATGGGGCATGCGGAACAACCTTTCTTGGGACGCGCGAGCACTGGGTTTTCTATCTGGATCAC
CACCTCCACCCCTGCTCTCTGCACTGCCTGAGCAGAGGTGTGCCTCTGCCAGGGCTTCC
CAGCCTGCGGCTCACATCAGCCCTCGACAGCGCGAGCCATCTCTGTGGAGGCCCTCTGCGA
GAACCACTCAGGCCCAGAGCCACCCTACAGCATCTCCAATTCTCCATCTACTTGCTCTGCC
AGCACATCAAGCCTGCCACCCCGCGGCCCCCTCTACCACCCACGGCCTCCTCTACCACC
CCACAGCCCCCTCCTACCACTACACAGCCCATCTCTGACACTACACAGCCCCCTCCTGTCAC
CCCAAGGCCTCCTCCTACCACCCACAACCCCTCCTAGCACAGCTGTCTATCTGCCAGACAG
CTGTATGGTACGCAGTCTCGTGGGCACCAGGTGCCGAGGTGGCTCCAAGCCTGCCATGAT
CAGTTTCTGATCAATTTCTGGATATGATCTGCGGCAACCTCTCATTTTCAGCCCTGTCTGG
CCCCAGTCGTCCTTTGGTAAAGCAGCTCTGTGTGGCTTGCTCCACCCCCCACTAGCTGTC
CACCAGGCCTGATCCCTGTGCCCTCACCCAGAAATATTCTGGGGCTGTTTCTCTGGAGAAT
GAGACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCAGGCTGTGCCCCGAGGAA
CCAGGCTTGGGTTTACGATGTGTGTGCGGGCCCCACCTTGGACGCCACTGATTTTCCACCGT
GCCGCGTTGGACCTGTGGGGAACGCTGCCAGATGGGGGAGCTTCTCTGCTCATGGTCTGT
GCCAATGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTGGCTAGCAGGCCAATGCAG
AATTAGTCGTGGAGGAAATGATACTTGCTTTCTAGAAGGCATGCTGGGCCCCCTGTTGCCCT
CTCTGCCCCCTCTGGGACCATCCCCACTCTGTCTGGCTCCTGGTCTTTTCTGCTTGGCATG
TTATCCAGTTGCCACGCTGTCTAGTCCTCCGTGCCAGCCCTCGCCACCCACGCGCCTACA
TTACCTCTGCGCTACTGACCTTCTTCTGGGTCCAGGGACTGGGGGTGCCGAGACGCAGG
GGATGTTAGGTCAAGCCCTGCTGCTCTAGTCTCCAGACAACCTGTTTCTTCTGGGATGCC
TTCCGCCCAGAGGGCCGAGAAGTGTACTGAGGACAGTCGGAGAGTACTTGCAGCGGGAAGA
GCCAACCCACCAGGCTTAGACTCTCCCTCAGCCTCGGCTCTGGTATGAGCAAGATGGAGC
TTCTGTCTGCTGCTATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATC
GGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTGCGATAACAATGGTAACAT
TTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTGGAATACTGTC
TGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAG

ATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCC
TAATTAATAG 3' (SEQ ID NO:7).

관심 단백질이, 예를 들어 STRC 단백질인 "관심 단백질의 N-말단 부분 또는 단편"은 스테레오실린 (STRC) 폴리 펩티드의 N-말단 부분 또는 단편의 아미노산 서열을 의미한다. 예를 들어, STRC의 N-말단 단편의 아미노산 서 열은 ATG 개시 코돈에 의해 코딩되는 메티오닌 (M) (N-말단 단부에서 볼드체)으로 시작하는 N-말단 단부에 신호 펩티드 서열 (예를 들어, 22개 아미노산) (소문자, 이탤릭체 및 밑줄표시)을 포함할 수 있다. "STRC 단백질의 N-말단 부분 또는 단편"을 포함하는 아미노산 서열은 인테인의 N-말단 단편 (N-인테인) (볼드체 및 밑줄표시)의 하류 및/또는 그에 인접한 것을 추가로 포함할 수 있다. 또 다른 실시양태에서, "STRC 단백질의 N-말단 부분 또는 단편"은 신호 펩티드 서열이 없는 STRC의 N-말단 단편의 아미노산 서열을 의미한다.

[0088] 인간 "STRC 단백질의 N-말단 부분 또는 단편"을 포함하는 예시적인 아미노산 서열은 하기와 같을 수 있다:

MALSLWPLLLLLLLLLLSFAVTLAPTGP~~HS~~LD~~P~~GLS~~F~~LK~~S~~LL~~S~~TLDQAPQGSLSRSRFFTF
LANISSSFEPGRMGEGPVGEPPPLQPPALRLHDFLVTLRGSPDWEPMGLLGDMLALLGQEQ
TPRDFLVHQAGVLGGLVEVLLGALVPGGPPTPTRPPCTRDGSPDCVLAADWLPSLLLLLEGT
RWQALVQVQPSVDPTNATGLDGREAA~~P~~HFLQGLLGLLTPTGELGSKEALWGGLLRTVGAPLY
AAFQEGLLRVTHSLQDEVFSILGQPEPDTNGQCQGGNLQQLLLWGV~~R~~HNLSWDVQALGFLSG
SPPPPPALLHCLSTGVPLPRASQPSAHISPRQRRAITVEALCENHLGPAPPYSISNFSIHLL
CQHTKPATPQPHPTTAICQTAVWYAVSWAPGAQGWLQACHDQFPDEF~~L~~DAICSNLSFSALS
GSNRRLVKRLCAGLLPPPTSCPEGLPPVPLTPDIFWGCFL~~E~~NETLWAERLCGEASLQAVPPS
NQAWVQHVCQGPTPDVTASPPCHIGPCGERCPDGG~~S~~FLVMVCANDTMYEVLVPFWPWL~~A~~GQC
RISRGNDTCFLEGLLGP~~L~~LPSPPLG~~P~~SPCLCTPGPFLLGMLSQ~~L~~PRCQSSVPALAHPT~~R~~L
HYLLRLLTFLLGPGAGGA~~E~~AQGM~~L~~GRALLSSLPDNC~~S~~FWDAFRPEGRRSVLR~~T~~IGEYLEQD
EEQPTPSGFEPTVNPSSGISKMELLCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNN
GNIYTQPVAQWHD~~R~~GEQEVFEYCLEDGSLIRATKD~~H~~KFMTVDGQMLPIDEIFERE~~L~~DLMRVD
NL~~P~~N** (SEQ ID NO:6).

[0089]

[0090] 무린 "STRC 단백질의 N-말단 부분 또는 단편"을 포함하는 또 다른 예시적인 아미노산 서열은 하기와 같을 수 있다:

MALSLQPQLLLLLLSLLPQEVTSAPTGPQSLDAGLSLLKSFVATLDQAPQRSLSQSRFSAFLA
NISSSFQ~~L~~GRMGEGPVGEPPPLQPPALRLHDFLVTLRGSPDWEPMGLLGDVLALLGQEQT~~P~~
RDFLVHQAGVLGGLVEALLGALVPGGPAPTRPPCTRDGSPDCVLAADWLPSLM~~L~~LE~~G~~TRW
QALVQLQPSVDPTNATGLDGREPAPHFLQGLLGLLT~~P~~AGELGSEEALWGGLLRTVGAPLYAA
FQEGLLRVTHSLQDEVFSIMGQPEPDASGQCQGGNLQQLLLWGM~~R~~NNLSWDARALGFLSGSP
PPPPALLHCLSRGVPLPRASQPAAHISPRQRR~~A~~ISVEALCENHSGPEPPYSISNFSIYLLCQ
HIKPATPRPPPTTPRPPPTTPQPPPTTTQPIPD~~T~~TQPPPVTPRPPPTTPQPPSTAVICQTA
VWYAVSWAPGARGWLQACHDQFPDQFLDMICGNLSFSALSGPSRPLVKQLCAGLLPPPTSCP
PGLIPVPLTPEIFWGCFL~~E~~NETLWAERLCVEDSLQAVPPRNQAWVQHVC~~R~~GPTLDATDFPPC
RVGPCGERCPDGG~~S~~FLLMVCANDTLYEALVPFWAWLAGQCRISRGNDTCFLEGLMG~~L~~LP~~L~~PS
LPPLG~~P~~SPCLCLAPGPFLLGMLSQ~~L~~PRCQSSVPALAHPT~~R~~LHYLLRLLTFLLGPGTGAETQG
MLGQALLSSLPDNC~~S~~FWDAFRPEGRRSVLR~~T~~VGEYLQREEPTPPGLDSSLSLGSGMSKMEL
LSCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHD~~R~~GEQEVFEYCL
EDGSLIRATKD~~H~~KFMTVDGQMLPIDEIFERELDLMRVDNL~~P~~N** (SEQ ID NO:8).

[0091]

[0092] 예시적인 "N-말단 STRC 폴리펩티드" 서열이 각각 인간 및 무린에 대해 하기에 제공된다 (N-말단에서 C-말단 방향으로).

[0093] ATG 개시 코돈 또는 신호 펩티드 서열에 의해 코딩되는 메티오닌을 포함하지 않는 예시적인 인간 N-말단 STRC 폴리펩티드 서열 (N-말단에서 C-말단 방향)이 하기에 제공된다 (N-말단에서 C-말단 방향):

VTLAPTGP HSLDPGLSFLKSLSTLDQAPQGSLSRSRFFTF LANISSSFEPGRMGEGPVGEP
PPLQPPALRLHDFLVTLRGSPDWEPMLGLLGDM LALLGQEQT PRDFLVHQAGVLGGLVEVLL
GALVPGGPPTPTRPPCTRDGSPDCVLAADWLPSLLLLLEGT RWQALVQVQPSVDPTNATGLD
GREAAPHLQGLLGLLTPTGELGSKEALWGGLLR TVGAPLYAAAFQEGLLRVTHSLQDEVFSI
LGQPEPDTNGQCQGGNLQQLLLWGV RHNLSWDVQALGFLSGSPPPPALLHCLSTGVPLPRA
SQPSAHISPRQRRAITVEALCENHLGPAPPYSISNFSIHL LCQHTKPATPQPHPSTTAICQT
AVVYAVSWAPGAQGWLQACHDQFPDEF LDAICSNLSFSALSGSNRRLVKRLCAGLLPPPTSC
PEGLPPVPLTPDIFWGCFL ENETLWAERLCGEASLQAVPPSNQAWVQHVCQGP TP DVTASPP
CHIGPCGERCPDGGSF LVMVCANDTMYEVLVPFWP WLAGQCRISRGGNDTCFLEGLLG PLLP
SLPPLGPSPLCLTPGP FLLGMLSQ LPRCQSSVPALAHPTRLHYLLRLLT FLLGPGAGGAEAQ
GMLGRALLLSSLPDNC SFWD AFRPEGRRSVLRTIGEYLEQDEEQTPSGFEPTVNPSSGISK
MELLA (SEQ ID NO:15).

[0094]

[0095]

ATG 개시 코돈 또는 신호 펩티드 서열에 의해 코딩되는 메티오닌을 포함하지 않는 또 다른 예시적인 무린 N-말단 STRC 폴리펩티드 서열 (N-말단에서 C-말단 방향)이 하기에 제공된다 (17개-잔기 소수성 영역은 밑줄표시됨) (N-말단에서 C-말단 방향):

APTGPQSLDAGLSLLKSFVATLDQAPQRSLSQSRFSAFLANISSSFQLGRMGEGPVGEP PPL
QPPALRLHDFLVTLRGSPDWEPMLGLLGDV LALLGQEQT PRDFLVHQAGVLGGLVEALLGAL
VPGGPAPTRPPCTRDGSPDCVLAADWLPSLMLLLEGT RWQALVQLQPSVDPTNATGLD GRE
PAPHLQGLLGLLT PAGELGSEEALWGGLLR TVGAPLYAAAFQEGLLRVTHSLQDEVFSIMGQ
PEPDASGQCQGGNLQQLLLWGM RNNLSWDARALGFLSGSPPPPALLHCLSRGVPLPRASQP
AAHISPRQRR AISVEALCENHSGPEPPYSISNFSIYLLCQH IKPATPRPPPTTPRPPPTTPQ
PPPTTTQPIPD TTQPPVTPRPPPTTPQPPSTAVICQTAVVYAVSWAPGARGWLQACHDQF
PDQFLDMICGNLSFSALSGPSRPLVKQLCAGLLPPPTSCPPGLIPVPLTPEIFWGCFL ENET
LWAERLCVEDSLQAVPPRNQAWVQHVCRGPTLDATDFPCR VGPCGERCPDGGSFLLMVCAN
DTLYEALVPFWAWLAGQCRISRGGNDTCFLEGLLG PLLPSLPPLGPSPLCLAPGFLLGMLS
QLPRCQSSVPALAHPTRLHYLLRLLT FLLGPGTGGAETQGMLGQALLLSSLPDNC SFWD AFR
PEGRRSVLRTVGEYLQREEPTPPGLDSSLSLGSGMSKMELLS (SEQ ID NO:16).

[0096]

[0097]

"스테레오실린 (STRC)의 C-말단 부분 또는 단편을 코딩하는 서열"은 STRC의 C-말단 부분 또는 단편의 부분 코딩 서열을 의미한다. 한 실시양태는 5' 말단에서 "ATG" 개시 코돈 (5' 말단에서 볼드체), "STRC 단백질의 C-말단 부분 또는 단편을 코딩하는 뉴클레오티드 서열의 상류에 플랭킹된 인테인의 C-말단 단편 (C-인테인)을 코딩하는 서열 (볼드체 및 밑줄표시)의 상류에 플랭킹된 신호 펩티드 코딩 서열을 코딩하는 서열 (소문자, 이탤릭체 및 밑줄표시)을 포함하는 뉴클레오티드 서열을 제공할 수 있다. 또 다른 실시양태는 하류 링커 서열 (볼드체 및 이탤릭체), Myc 태그 (소문자) 및 정지 코돈 (대문자, 이탤릭체 및 밑줄표시)을 갖는 STRC의 C-말단 부분 또는 단편을 코딩하는 뉴클레오티드 서열을 추가로 제공할 수 있다.

[0098]

"STRC의 C-말단 부분 또는 단편을 코딩하는 서열"을 포함하는 예시적인 인간 뉴클레오티드 서열은 하기와 같을 수 있다:

5'ATGgctctcagcctctggcccctgctgctgctgctgctgctgctgctgctgctgctgctccttt
gcaATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGA
AAGAGATCACAACCTTGCTCTGAAGAACGGATT CATAGCTTCTAATTGCTTTAGTCCTGTGC
TGTGGGATCTGCTCCAGAGGGAAGAGTGT TGGGCCCTGCAGATTCTAGTGCAGGCGTAC
CTGCATATGCCCCAGAAAACCTCCAGCAGCTGGTGCTTTCAGCAGAGAGGGAGGCTGCACA
GGGCTTCTTGACACTCATGCTGCAGGGGAAGCTGCAGGGGAAGCTGCAGGTACCACTATCCG
AGGAGCAGGCCCTGGGTGCGCTGACAGCCCTGTGCTCCAGCGGTACCCACGCCTCACCTCC

[0099]

CAGCTCTTCATTGACCTGTCACTCATCCCTTTCTTGCTGTCTCTGACCTGATGCGCTT
 CCCACCATCCCTGTTAGCCAACGACAGTGTCTGGCTGCCATCCGGGATTACAGCCCAGGAA
 TGAGGCCTGAACAGAAGGAGGCTCTGGCAAAGCGACTGCTGGCCCTGAACGTTTGGGGAA
 GTGCCTGCCTGGCCCCAGGAGCTGTGTGGGCAGTGTGCCCTGCTCCCCACCTCCCTCT
 GGAGAACTTTTTGCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGGATAGCTGGCCAGCAG
 CAGGTCTGGGGCCAGGGCATGCCCCCATGTGCTGCGCAGCCTGGTAAACCAGAGTGTCCAG
 GATGGTGAGGAGCAGGTACGCAGGCTTGGGCCCTCGCCTGTTTCTGAGCCCTGAGGAGCT
 GCAGAGCCTAGTGCCCTGAGTGATCCAACGGGGCCAGTAGAACGGGGGCTGTGGAATGTG
 CAGCCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATATGAACTTCTGGGTGTGTTGCGC
 TCATCTGGAGGAGCGGTGTGAGCCCCGGGAGCTGCGGGTCTGGGCCCTCTCTTCTCTCA
 GCTGGGCCCTCCGCTTCCTTCAGGAGCTGTGAGAGCCAGCTTAGAGCCATGCTTCTGTCC
 TGCAGGGAAGTAGTGTTACACCTGTCTAGGCTGTCTGCTGCTTGGACGGCTCCTTCTTAGG
 CACGATCTATCCCTGGAGGAAGCTGTCTCCTTGACCTTCTGCTACCAGGCCTCAGCCCCCA
 GAACTCCAGGCCATCCCTAGGCGAGTCTGTGCGGGCTTGTCTGCTGGGCCCTGAAC
 TGTACGCCTCTCAGCCTGCCAGACCGCAGCACTGCTGCAGACCTTTCGGGTAAAGATGGT
 GTTAAAAATATGGGTACAACAGGTGTGGTCCAGCTGTGTGTATCCCTGGTCAGCCTATTCC
 CACCACCTGGCCAGACTGCCTGCTTCCCCTGCTCCCATTAAGCTGCTACAAGTGGATTCTCT
 TGGCTCTTCTGGCAAATCGAAGACGCTACTGGGAGCTGCCCTGGTCTGAGCAGCAGGCACAG
 TTTCTCTGGAAGAAGATGCAAGTACCCACCAACCTTACCCTCAGGAATCTGCAGGCTCTGGG
 CACCCTGGCAGGAGGCATGTCTGTGAGTTTCTGCAGCAGATCAACTCCATGGTAGACTTCC
 TTGAAGTGGTGACATGATCTATCAGCTGCCCCTAGAGTTCGAGGGAGCCTGAGGGCCTGT
 ATCTGGGCAGAGCTACAGCGGAGGATGGCAATGCCAGAACCAGAATGGACAAGTGTAGGGCC
 AGAACTGAACGGGCTGGATAGCAAGCTACTCCTGGACTTACCGATCCAGTTGATGGACAGAC
 TATCCAATGAATCCATTATGTTGGTGGTGGAGCTGGTGCAAAGAGCTCCAGAGCAGCTGCTG
 GCACTGACCCCCCTCCACCAGGCAGCCCTGGCAGAGAGGGCACTACAAAACCTGGCTCCAAA
 GGAGACTCCAGTCTCAGGGGAAGTGTGAGAGCCTTAGGCCCTTTGGTTGGATTCTGGGGA
 CAGAGAGCACACGACAGATCCCCCTACAGATCCTGCTGTCCCATCTCAGTCAGCTGCAAGGC
 TTCTGCCTAGGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTATTGCAGGAGTCTGTTCT
 TGGGAAACCAGAGTTGTGGAGCCAGGATGAAGTAGAGCAAGCTGGACGCCTAGTATTCACTC
 TGTCTACTGAGGCAATTTCTTGTATCCCCAGGGAGGCCTTGGGTCCAGAGACCCTGGAGCGG
 CTTCTAGAAAAGCAGCAGAGCTGGGAGCAGAGCAGAGTTGGACAGCTGTGTAGGGAGCCACA
 GCTTGCTGCCAAGAAAGCAGCCCTGGTAGCAGGGGTGGTGCGACCAGCTGCTGAGGATCTTC
 CAGAACCTGTGCCAAATTTGTGCAGATGTACGAGGGACATTCCCAGCAGCCTGGTCTGCAACC
 CAGATTGCAGAGATGGAGCTCTCAGACTTTGAGGACTGCCTGACATTATTTGCAGGAGACCC

[0100]

AGGACTTGGGCTGAGGAACTGCGGGCAGCCATGGGCAAAGCAAACAGTTGTGGGGTCCCC
 CCCGGGGATTTTCGTCTGAGCAGATCCTGCAGCTTGGTAGGCTCTTAATAGGTCTAGGAGAT
 CGGGAACCTACAGGAGCTGATCCTAGTGGACTGGGGAGTGCTGAGCACCTGGGGCAGATAGA
 TGGCTGGAGCACCCTCAGCTCCGCATTGTGGTCTCCAGTTTCCTACGGCAGAGTGGTCGGC
 ATGTGAGCCACCTGGACTTCGTTTCATCTGACAGCGCTGGGTTTACTCTCTGTGGACTGCGG
 CCAGAGGAGCTCCAGCACATCAGCAGTTGGGAGTTCAGCCAAGCAGCTCTCTTCCTCGGCAC
 CCTGCATCTCCAGTGCTCTGAGGAACAACTGGAGGTTCTGGCCACCTACTTGTACTGCCTG
 GTGGGTTTGGCCCAATCAGTAACTGGGGGCTGAGATCTTCACTGAAATTGGCACCATAGCA
 GCTGGGATCCCAGACCTGGCTCTTTCAGCACTGCTGCGGGGACAGATCCAGGGCGTTACTCC
 TCTTGCCATTTCTGTTCATCCCTCCTCCTAAATTTGCTGTGGTGTTTAGTCCCATCCAATAT
 CTAGTCTCACCAGTGCTCAGGCTGTGGCTGTCACTCCTGAGCAAATGGCCTTTCTGAGTCCT
 GAGCAGCGACGAGCAGTTGCATGGGCCAACATGAGGGAAAGGAGAGCCAGAACAGCAAGG
 TCGAAGTACAGCCTGGGGCCTCCAGGACTGGTCACGACCTTCCTGGTCCCTGGTATTGACTA
 TCAGCTTCCTTGGCCACCTGCTA**ACGCGTACGCGGCCGCTC**gagcagaaactcatctcagaa
 gaggatctg**TAATAG** (SEQ ID NO:17).

[0101]

[0102]

"STRC의 C-말단 부분 또는 단편을 코딩하는 서열"을 포함하는 예시적인 뮤린 뉴클레오티드 서열은 하기와 같을 수 있다:

5'**ATG**gctctgagcctccagccccagctgctccttctcctgtcgctcctgccgcaggaagtg
acttcaATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGT
CGAAAGAGATCACAACCTTGCTCTGAAGAACGGATTCATAGCTTCTAATTGCTTCAGTCCTG
 TACTGTGGGATCTACTCCAGAGAGAGAAGAGCGTTTGGGCCCTGAGGACCTGGTGAAGGCC
 TACCTGCGCATGCCTCCAGAAGACCTTCAGCAGCTTGTGCTTTCAGCAGAGATGGAGGCTGC
 ACAGGGCTTCCTGACGCTCATGCTTCGTTCCCTGGGCTAAGCTGAAGGTTCAACCATCCGAGG
 AGCAGGCCATGGGCCGCTGACAGCCTTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAA
 CTCTTTATCGACATGTCACCGCTCATCCCTTCCTGGCTGTCCCTGACCTCATGCGCTTCCC
 ACCGTCCCTTTTGGCCAACGACAGTGTCTGGCTGCCATCAGGGATCACAGCTCAGGAATGA
 AGCCTGAACAGAAGGAGGCCCTGGCAAACGACTGCTGGGCCCTGAGCTGTTTGGAGAAGTG
 CCTGATTGGCCCCAGGAGCTGCTGTGGGCAGCCCTGCCTCTGCTTCCCCATCTGCCTCTGGA
 GAGCTTTCTCCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGGATAGCTGGCCAGTAGCAG
 ATCTTGGGCCGGGACACGCCCACATGTGCTTCGTAGCCTAGTAAACCAGAGCATGGAGGAT
 GGGGAGGAGCAGGTGCTCAGGCTTGGGTCCCTCGCCTGTTTCTGAGTCCTGAGGAGCTACA
 GAGTCTGGTGCCCTTGAGTGATCCAATGGGGCCTGTAGAACAGGGTCTGCTGGAATGTGCGG
 CCAATGGGACCTCAGCCCAGAAGGACGGGTGGCATATGAACTTCTGGGAGTGTTGCGTTCA

[0103]

TCTGGAGGAACTGTCTTAAGCCCCGAGAGCTGAGGGTCTGGGCACCTCTCTTTCCCCAGCT
GGGCTCCGCTTCCTGCAGGAGCTCTCAGAGACCCAGCTTAGAGCCATGCTTCCTGCCCTAC
AGGGAGCCAGTGTACACCTGCCAGGCTGTTCTGTTGTTTGAAGGCTCCTTCCTAAGCAT
GATCTGTCCCTGGAGGAACTCTGCTCCCTGCACCTCTCCTGCCAGGTCTCAGCCCCAGAC
ACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGGTGCTTGTTCCTGCCTGGGCCCTGAACTGT
CAAGGCTTTCAGCTTGCCAGATTGCAGCTCTGCTGCAGACCTTTCGGGTAAAAGATGGTGT
AAAAATATGGGTGCAGCAGGTGCCGGCTCAGCCGTGTGCATTCTTGGGCAGCCCACCACTTG
GCCAGACTGCCTGCTTCCCCTGCTCCCATTAAGCTGCTACAGCTGGACGCTGCAGCTCTTC
TGGCAAACCGAAGACTCTATCGGCAGCTGCCTTGGTCTGAGCAACAGGCACAGTTTCTCTGG
AAGAAAATGCAAGTGCTACCAACCTGAGCCTGAGGAATCTGCAGGCTCTGGGCAACTTGGC
AGGAGGCATGACCTGCGAGTTTCTGCAGCAGATCAGCTCAATGGTTGACTTTCTTGATGTGG
TACACATGCTCTACCAGCTGCCACTGGTGTTCGAGAGAGCCTGCGGGCCTGTATCTGGACA
GAGCTACAGCGGAGGATGACAATGCCAGAGCCAGAGCTGACCACCTAGGGCCAGAACTGAG
TGAATTGACACAAAGCTACTCCTGGACTTGCCGATCCAGCTGATGGACAGATTGTCCAATG
ATTCCATTATGTTGGTGGTGGAGATGGTCCAAGGCGCTCCAGAGCAGCTGCTGGCACTGACC
CCACTCCACCAGACAGCCTTGGCAGAGCGAGCACTTAAAAACCTGGCTCCAAAGGAGACCCC
AATCTCCAAAGAAGTGCTGGAGACACTGGGCCCTTGGTTGGATTCTTGGGAATAGAGAGCA
CGCGACGGATCCCTTTACCCATTCTACTGTCTCATCTCAGTCAGCTGCAGGGCTTCTGCCTA
GGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTGTTGCAGGAGCCTGTTCTTGGAAAACC
AGAATTGTGGAGCCAGGATGAAATAGAGCAAGCTGGACGCCTAGTATTCACTCTGTCTGCTG
AGGCTATTTCTCGATCCCCAGGGAGGCTTTGGGCCAGAGACACTGGAGAGGCTTCTGGGA
AAGCATCAAAGCTGGGAGCAGAGCAGAGTGGGCCATCTGTGTGGGGAGTCACAGCTTGCCCA
CAAGAAAGCAGCTCTGGTAGCTGGGATTGTGCATCCAGCTGCTGAGGGTCTCCAAGAGCCTG
TACCAAACCTGTGCAGACATACGGGGAACCTTCCCAGCGGCCTGGTCTGCGACACAAATCTCA
GAGATGGAACCTCTCAGACTTTGAAGACTGCCTGTCACTATTTGCTGGAGATCCAGGACTTGG
TCCTGAGGAACTACGGGCAGCCATGGGCAAGGCCAAGCAGTTGTGGGGTCCCCCTCGAGGAT
TCCGTCTCTGAGCAGATCTTGCAGCTGGGCCGTCTCCTGATAGGTCTAGGAGAACGGGAAGTG
CAGGAGCTTACCTTGGTGGACTGGGGTGTGCTGAGCAGCCTGGGGCAAATAGATGGCTGGAG
TTCCATGCAGCTCCGAGCCGTGGTCTCCAGTTTCTTAAGGCAGAGTGGTCGGCATGTGAGCC
ACCTGGACTTCATTTATCTGACAGCACTGGGTTACACAGTCTGTGGATTGCGACCAGAGGAG
TTACAGCACATCAGCAGTTGGGAGTTTAGCCAAGCAGCTCTCTTCTGGGTAGCTTGCATCT
CCCGTGTCTGAGGAACAGCTGGAAGTTCTGGCCTATCTCCTTGTGTTGCCTGGTGGCTTTG
GCCCAGTCAGTAACGGGGGCTGAGATCTTCACTGAAATTGGCACAATAGCAGCTGGCATC
CCAGACCTGGCTCTTTTCAGCATTACTGCGGGGACAGATCCAAGGCCTGACTCCTCTTGCCAT

[0104]

TTCTGTCAATCCTGCTCCCAAGTTTGCAGTGGTCTTCAACCCCATCCAGTTATCTAGTCTCA
CCAGGGGTGAGGCCGTAGCTGTTACTCCTGAACAGCTGGCCTATCTGAGTCCTGAGCAGCGG
CGAGCAGTTGCATGGGCCAACACGAAGGGAAGGAGATCCCAGAGCAGCTGGGTGCAAACTC
AGCCTGGGGTCTCTACGACTGGTTCCAAGCCTCCTGGGCCCTGGCATTGCCCGTCAGATTT
TTGGCCACCTATTA**ACGCGTACGCGGCCGCTC**gagcagaaactcatctcagaagaggatctg
TAATAG-3' (SEQ ID NO:19).

[0105]

[0106]

"STRC 단백질의 C-말단 부분 또는 단편"은 스테레오실린 (STRC) 폴리펩티드의 C-말단 부분 또는 단편의 아미노산 서열을 의미한다. "STRC 단백질의 C-말단 부분 또는 단편"을 포함하는 아미노산 서열은 그 앞에, N-말단에서 C-말단 방향으로, ATG 개시 코돈에 의해 코딩되는 메티오닌 (M) (N-말단 단부에서 볼드체), 신호 펩티드 서열 (소문자, 이탤릭체 및 밑줄표시) 및 인테인의 C-말단 단편 (C-인테인) (볼드체 및 밑줄표시)이 선행할 수 있다. 아미노산 서열은 "STRC 단백질의 C-말단 부분 또는 단편"의 하류에 링커 서열 (볼드체 및 이탤릭체) 및 Myc 태그 (소문자)를 추가로 포함할 수 있다.

[0107] 인간 "STRC 단백질의 C-말단 부분 또는 단편"을 포함하는 예시적인 아미노산 서열은 하기와 같을 수 있다:

MaslslwplllllllllllllsfaIKIATRKYLGKQNVYDIGVERDHNFKNGFIASNCFSPVL
WDLLQREKSVWALQILVQAYLHMPPENLQQLVLSAEREAQGF¹TLMLQGKLQGKLQVPPSE
EQALGRLTALLLQRYPR²LTSQLFIDLSPLIPFLAVSDLMRFPPSLLANDSVLAAIRDYSPGM
RPEQKEALAKRL³LAPELFGEVPAWPQELLWAVLPL⁴PHLPLEN⁵FLQLSPHQIQALED⁶SWPAA
GLGPGHARHVL⁷RLSNQSVQDGEQVRLG⁸PLACFLSPEELQSLVPLSDPTGPVERGLLECA
ANGT⁹LSPEGRVAYELLGVL¹⁰RSSGGAVLSPREL¹¹RVWAPL¹²FSQGLR¹³FLQELSEPQLRAMLPVL
QGT¹⁴SVTPAQAVLL¹⁵GRLLPRHDL¹⁶SLLEELCSLHLL¹⁷LPGLSPQTLQAI¹⁸PRRVLVGACSC¹⁹LAPEL
SRLSACQTAALLQ²⁰TRFKDGVKNMGTTGAGPAVCIPGQPIPTTWPDC²¹LLPL²²LPKLLQ²³LDSL
ALLANRRRYWELPWSEQQAQFLWKKMQVPTNLT²⁴LRNLQALGTLAGGMSCEFLQ²⁵QINSMVDFL
EVVHMIYQLPTRV²⁶RGSLRACIWAELQRRMAMPEPEWTTVGPEL²⁷NGLDSKLLLDLP²⁸IQ²⁹LMDR³⁰L
SNESIMLVVELVQRAPEQLLALTPLHQAALAE³¹RALQNLAPKETPVSGEVLET³²LGPLVGFLGT
ESTRQIPLQILLSHLSQLQGFC³³LGETFATELGWLL³⁴LQESVLGKPELWSQDEVEQAGRLVFTL
STEAISLIPREALGPETLERLLEKQ³⁵QSWEQSRVGQLCREPQLAAKKAALVAGVVRPAAEDLP
EPVPNCADVRGTFPAAWSATQIAEMELSD³⁶FEDCLTLFAGDPGLGPEELRAAMGKAKQLWGPP
RGFRPEQILQLGRLLIGLGDRELQELILVDWGVLS³⁷TLGQIDGWSTTQLRIVVSSFLRQSGRH
VSHLDFVHLTALGYTLCGLRPEELQHIS³⁸SWEFSQAALFLGTLHLQCSEEQLEVL³⁹LAHLLVLPG

[0108] GFGPISNWGPEIFTEIGTIAAGIPDLALSALLRGQIQGV⁴⁰TPLAISVIPPPKFAVVFSP⁴¹IQLS
SLTSAQAVAVTPEQMAFLSPEQRR⁴²AWAQHEGKESPEQQGRSTAWGLQDWSRPSWSLVLT⁴³I
SFLGHL⁴⁴L**TRTRP**L⁴⁵eqkliseed1** (SEQ ID NO:18).

[0109]

[0110] 뮤린 "STRC 단백질의 C-말단 부분 또는 단편"을 포함하는 또 다른 예시적인 아미노산 서열은 하기와 같을 수 있다:

MaslslqpqllllllsllpqevtsIKIATRKYLGKQNVYDIGVERDHNFKNGFIASNCFSPV
LWDL¹LQREKSVWALRTLVKAYLRMPEDLQQLVLSAEMEAQGF²TLMLRSWAKLKVQPSEE
QAMGR³LTALLLQRYPR⁴LTSQLFIDMSPLIPFLAVPDL⁵MRFP⁶PSLLANDSVLAAIRDHSSGMK
PEQKEALAKRL⁷LAPELFGEVDPWPQELLWAALPL⁸PHLPLESFLQLSPHQIQALED⁹SWPVAD
LGPGHARHVL¹⁰RLSNQSMEDGEEQVLR¹¹LGSL¹²ACFLSPEELQSLVPLSDPMGPVEQGLLECAA
NGT¹³LSPEGRVAYELLGVL¹⁴RSSGGTVLSPREL¹⁵RVWAPL¹⁶FPQLGLR¹⁷FLQELSETQLRAMLPALQ
GASVT¹⁸PAQAVLLFGRLLPKHDL¹⁹SLLEELCSLHPL²⁰LPGLSPQTLQAI²¹PKRVLVGACSC²²LGPELS
RLSACQIAALLQ²³TRFKDGVKNMGAAGAGSAVCIPGQPTTWPDC²⁴LLPL²⁵LPKLLQ²⁶LDAALL
ANRRLYRQLPWSEQQAQFLWKKMQVPTNLSLRNLQALGNLAGGMTCEFLQ²⁷QISSMVDFLDV²⁸V
HMLYQLPTGVRESLRACIWTELQRRMTMPEPELT²⁹TLGPELSELDTKLLLDLP³⁰IQ³¹LMDR³²LSND
SIMLVVEMVQGAPEQLLALTPLHQTALAERALKNLAPKETPISKEVLET³³LGPLVGFLG³⁴IEST
RRIPLPILLSHLSQLQGFC³⁵LGETFATELGWLL³⁶LQEPVLGKPELWSQDEIEQAGRLVFTLSAE
AISSIPREALGPETLERLLGKHQSWEQSRVGHL³⁷CGESQLAHKKAALVAGIVHPAAEGLQE³⁸EPV
PNCADIRGTFPAAWSATQISEMELSD³⁹FEDCLSLFAGDPGLGPEELRAAMGKAKQLWGPPRGF
RPEQILQLGRLLIGLGERELQELTLVDWGVLS⁴⁰SLGQIDGWSSMQLRAVVSSFLRQSGRHVSH
LDFIYLTALGYTVCGLRPEELQHIS⁴¹SWEFSQAALFLGSLHLP⁴²CSEEQLEVLAYLLVLPGGFG
PVSNWGPEIFTEIGTIAAGIPDLALSALLRGQIQGLT⁴³PLAISVIPAPKFAVVFNP⁴⁴IQ⁴⁵LSST⁴⁶T
RGQAVAVTPEQLAYLSPEQRR⁴⁷AWAQHEGKEIPEQLGRNSAWGLYDW⁴⁸FQASWALALPVSIF
GHLL**TRTRP**L⁴⁹eqkliseed1* (SEQ ID NO:20).

[0111]

[0112] 예시적인 "C-말단 STRC 폴리펩티드" 서열이 각각 인간 및 뮤린에 대해 하기에 (N-말단에서 C-말단 방향으로) 제
공된다.

[0113] 신호 펩티드 서열을 포함하지 않는 STRC 단백질의 예시적인 인간 C-말단 부분은 하기와 같을 수 있다:

CFSPVLWDLQREKSVWALQILVQAYLHMPDENLQQLVLSAEREAAGFLTLMLQKGLQKGL
QVPPSEEQALGRLTALLQRYPRLTSQLFIDLSPLIPFLAVSDLMRFPPSLLANDSVLAAIR
DYSPGMRPEQKEALAKRLLAPELFGFVPAWPQELLWAVLPLLPPLPLENFLQLSPHQIQALE
DSWPAAGLPGHARHVLRLSLVNQSVQDGEEQVRLGPLACFLSPEELQSLVPLSDPTGPVER
GLLECAANGTLSPEGRVAYELLGVLRSSGGAVLSPRELRVWAPLFSQLGLRFLQELSEPQLR
AMLPVLQGTSVTPAQAVLLLRLLPRHDLSEELCSLHLLPLGLSPQTLQAIPRRVLVGACS
CLAPELSRLSACQTAALLQTFRVKDGVMNGTTGAGPAVCIPGQPIPTTWPDCLLPLPLKL
LQLDSLALLANRRRYWELPWSEQQAQFLWKKMQVPTNLTLRNLQALGTLAGGMSCEFLQQIN
SMVDFLEVVMHIYQLPTRVRGSLRACIWAELQRRMAMPEPEWTTVGPELNGLDSKLLLDLPI
QLMDRLSNESIMLVVELVQRAPEQLLALTPLHQAAALAEALQNLAPKETPVSGEVLETLGPL
VGFLGTESTRQIPLQILLSHLSQLQGFCGETFATELGWLLQLQESVLGKPELWSQDEVEQAG
RLVFTLSTEAISLIPREALGPETLERLLEKQSQWEQSRVGLCREPQLAAKKAALVAGVVRP
AAEDLPEPVPNCADVRGTFFAAWSATQIAEMELSDFFEDCLTLFAGDPGLGPEELRAAMGKAK
QLWGPPRGFRPEQILQLGRLLIGLGDRELQELILVDWGVLSLTLGQIDGWSTTQLRIVVSSFL
RQSGRHVSHLDFVHLTALGYTLGCLRPPEELQHISSEWFSQAALFLGTLHLQCSEEQLEVLAH
LLVLPGGFGPISNWGPFIETIGTIAAGIPDLALSALLRGQIQGLTPLAISVIPPKFAVVF
SPIQLSSLTSAQAVAVTPEQMAFLSPEQRRAVAWAQHEGKESPEQQGRSTAWGLQDWSRPSW
SLVLTISFLGHLL (SEQ ID NO:23).

[0114]

[0115] 신호 펩티드 서열을 포함하지 않지만 밑줄표시된 바와 같은 적어도 16개 잔기의 소수성 영역을 포함하는 STRC 단백질의 예시적인 류린 C-말단 부분은 하기와 같을 수 있다:

CFSPVLWDLQREKSVWALRTLVKAYLRMPPELQQLVLSAEMEAAGFLTLMLRSWAKLKV
QPSEEQAMGRLTALLQRYPRLTSQLFIDMSPLIPFLAVPDLMRFPSPSLLANDSVLAAIRDH
SSGMKPEQKEALAKRLLAPELFGFVDPWPQELLWAALPLLPPLPLESFLQLSPHQIQALEDS
WPVADLPGHARHVLRLSLVNQSMEDGEEQVRLGSLACFLSPEELQSLVPLSDPMGPVEQGL
LECAANGTLSPEGRVAYELLGVLRSSGGTVLSPRELRVWAPLFPQLGLRFLQELSETQLRAM
LPALQGASVTTPAQAVLLFGRLLPKHDLSEELCSLHPLLPGLSPQTLQAIPKRVLVGACSCS
GPELSRLSACQIAALLQTFRVKDGVMNGAAGAGSAVCIPGQPTTWPDCLLPLPLKLLQLD
AAALLANRRLYRQLPWSEQQAQFLWKKMQVPTNLSLRNLQALGNLAGGMTCEFLQQISSMVD
FLDVVHMLYQLPTGVRESLRACIWTELQRRMTMPEPELTTLGPELSELDTKLLLDLPIQLMD
RLSNDSIMLVVEMVQGAPEQLLALTPLHQATALAERALKNLAPKETPISKEVLETLGPLVGFL
GIESTRRIPLPILLSHLSQLQGFCGETFATELGWLLQLQEPVLGKPELWSQDEIEQAGRLVF
TLSEAEAISSIPREALGPETLERLLGKHQSWEQSRVGHLCGESQLAHKKAALVAGIVHPAAEG
LQEPVPNCADIRGTFFAAWSATQISEMELSDFFEDCLSLFAGDPGLGPEELRAAMGKAKQLWG
PPRGFRPEQILQLGRLLIGLGERELQELTLVDWGVLSLTLGQIDGWSSMQLRAVVSSFLRQSG
RHVSHLDFIYLTALGYTVCGLRPEELQHISSEWFSQAALFLGSLHLPCSEEQLEVLAYLLVL
PGGFGPVSNWGPFIETIGTIAAGIPDLALSALLRGQIQGLTPLAISVIPAPKFAVVFNPIQ
LSSLTRGQAVAVTPEQLAYLSPEQRRAVAWAQHEGKEIPEQLGRNSAWGLYDWFQASWALAL
PVSIFGHLL (SEQ ID NO:24).

[0116]

[0117] STRC 단백질의 N-말단 부분 및 STRC 단백질의 C-말단 부분의 라이게이션은, 예컨대 펩티드 결합을 통해 발생하여, 전장 STRC 단백질을 생성할 수 있다.

[0118]

"인테인"은 단백질 스플라이싱으로 공지된 과정에서 그 자체를 절제하고 나머지 단편 (엑스테인)을 펩티드 결합으로 연결시킬 수 있는 단백질의 단편이다. 인테인은 또한 "단백질 인트론"으로 지칭된다. 인테인이 그 자체를 절제하고 단백질의 나머지 부분을 연결하는 과정은 본원에서 "단백질 스플라이싱" 또는 "인테인-매개 단백질 스플라이싱"으로 지칭된다. 일부 실시양태에서, 전구체 단백질의 인테인 (인테인-매개 단백질 스플라이싱 전의

인테인 함유 단백질)은 2개의 유전자로부터 유래한다. 이러한 인테인은 본원에서 스플릿 인테인 (예를 들어, 스플릿 인테인-N 및 스플릿 인테인-C)으로 지칭된다. 예를 들어, 시아노박테리아에서, DNA 폴리머라제 III의 촉매 서브유닛인 DnaE는 2개의 별개의 유전자 dnaE-n 및 dnaE-c에 의해 코딩된다. dnaE-n 유전자에 의해 코딩되는 인테인은 본원에서 "인테인-N"으로 지칭될 수 있다. dnaE-c 유전자에 의해 코딩되는 인테인은 본원에서 "인테인-C"로 지칭될 수 있다.

[0119] 다른 인테인 시스템이 또한 사용될 수 있다. 예를 들어, dnaE 인테인, Cfa-N (예를 들어, 스플릿 인테인-N) 및 Cfa-C (예를 들어, 스플릿 인테인-C) 인테인 쌍에 기초한 합성 인테인이 기재되었다 (예를 들어, 본원에 참조로 포함된 문헌 [Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5]). 본 개시내용에 따라 사용될 수 있는 인테인 쌍의 비제한적 예는 Cfa DnaE 인테인, Ssp GyrB 인테인, Ssp DnaX 인테인, Ter DnaE3 인테인, Ter ThyX 인테인, Rma DnaB 인테인 및 Cne Prp8 인테인 (예를 들어, 본원에 참조로 포함된 미국 특허 번호 8,394,604에 기재된 바와 같음)을 포함한다.

[0120] "myc 태그"는 c-myc 유전자로부터 유래된 폴리펩티드 단백질을 의미하며, 여기서 합성 펩티드 서열 (즉, EQKLISEEDL (서열식별번호: 27))은 인간 c-myc 단백질의 C-말단 아미노산 (410-419)에 상응한다. 이러한 태그는 단백질 단리 (예를 들어, 웨스턴 블롯팅, 면역형광, 면역침전)와 같은, 그러나 이에 제한되지는 않는 추가의 연구를 가능하게 한다.

[0121] 예시적인 AAV9-php.b 벡터의 서열이 하기에 제공된다. 일부 실시양태에서, AAV9-php.b 벡터 (5'에서 3')는 하기에 대해 적어도 70% 이상 (예를 들어, 75%, 80%, 85%, 90%, 95%, 97%, 100%) 동일성의 뉴클레오티드 서열을 제공할 수 있다:

[0122] 5'-CCAATGATACGCGTCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGAT
GTGCTGCAAGGCGATTAAAGTTGGGTAAACGCCAGGGTTTTCCAGTCACGACGTTGTAAAACG

ACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACATCGACGGTATC
GGGGGAGCTCGCAGGGTCTCCATTTTGAAGCGGGAGGTTTGAACGCGCAGCCGCCATGCCGG
GGTTTTACGAGATTGTGATTAAAGTCCCCAGCGACCTTGACGAGCATCTGCCCGGCATTTCT
GACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGATTCTGACATGGA
TCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTTCTGA
CGGAATGGCGCCGTGTGAGTAAGGCCCCGAGGCTCTTTTCTTTGTGCAATTTGAGAAGGGA
GAGAGCTACTTCCACATGCACGTGCTCGTGGAACCAACCGGGGTGAAATCCATGGTTTTGGG
ACGTTTCTGAGTCAGATTTCGCGAAAACTGATTTCAGAGAATTTACCGCGGGATCGAGCCGA
CTTTGCCAACTGGTTCGCGGTACAAAAGACCAGAAATGGCGCCGAGGCGGGAACAAGGTG
GTGGATGAGTGCTACATCCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGC
GTGGACTAATATGGAACAGTATTTAAGCGCCTGTTTGAATCTCACGGAGCGTAAACGGTTGG
TGGCGCAGCATCTGACGCACGTGTGCGAGCGCAGGAGCAGAACAAAGAGAATCAGAATCCC
AATTCTGATGCGCCGGTGATCAGATCAAAAACCTCAGCCAGGTACATGGAGCTGGTCGGGTG
GCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCATACA
TCTCCTTCAATGCGGCCTCCAACTCGCGGTCCCAAATCAAGGCTGCCTTGGACAATGCGGGA
AAGATTATGAGCCTGACTAAAACCGCCCCGACTACCTGGTGGGCCAGCAGCCCGTGGAGGA
CATTTCCAGCAATCGGATTTATAAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCGG
CTTCGTCTTTCTGGGATGGGCCACGAAAAAGTTCCGGCAAGAGGAACACCATCTGGCTGTTT
GGGCTGCAACTACCGGAAGACCAACATCGCGGAGGCCATAGCCACACTGTGCCCTTCTA
CGGGTGCCTAACTGGACCAATGAGAATTTCCCTTCAACGACTGTGTGGACAAGATGGTGA
TCTGGTGGGAGGAGGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCCAAAGCCATTCTCGGA
GGAAGCAAGGTGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGT
GATCGTCACCTCCAACACCAATATGTGCGCCGTGATTGACGGGAACTCAACGACCTTCGAAC
ACCAGCAGCCGTGCAAGACCGGATGTTCAAATTTGAACTACCCGCCGTCTGGATCATGAC
TTTGGGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGGT
TGAGGTGGAGCATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGACCCGCCCCCAGTG
ACGCAGATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAGAC
GCGGAAGCTTCGATCAACTACGCGGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGCAT
GAATCTGATGCTGTTTCCCTGCAGACAATGCGAGAGACTGAATCAGAATTCAAATATCTGCT
TCACTCACGGTGTCAAAGACTGTTTAGAGTGCTTTCCCGTGTGAGAATCTCAACCCGTTTCT
GTCGTCAAAAAGGCGTATCAGAACTGTGCTACATTCATCACATCATGGGAAAGGTGCCAGA
CGCTTGCACTGCTTGCGACCTGGTCAATGTGGACTTGGATGACTGTGTTTCTGAACAATAAA
TGACTTAAACCAGGTATGAGTCGGCTGGATAAATCTAAAGTCATAAACGGCGCTCTGGAATT
ACTCAATGAAGTCGGTATCGAAGGCTGACGACAAGGAACTCGCTCAAAAGCTGGGAGTTG

[0123]

AGCAGCCTACCCTGTACTGGCACGTGAAGAACAAGCGGGCCCTGCTCGATGCCCTGGCCATC
GAGATGCTGGACAGGCATCATACCCACTTCTGCCCCCTGGAAGGCGAGTCATGGCAAGACTT
TCTGCGGAACAACGCCAAGTCATTCCGCTGTGCTCTCCTCTCACATCGCGACGGGGCTAAAG
TGCATCTCGGCACCCGCCAACAGAGAAACAGTACGAAACCCTGGAAAATCAGCTCGCGTTC
CTGTGTGAGCAAGGCTTCTCCCTGGAGAACGCACTGTACGCTCTGTCCGCCGTGGGCCACTT
TACACTGGGCTGCGTATTGGAGGAACAGGAGCATCAAGTAGCAAAAGAGGAAAGAGAGACAC
CTACCACCGATTCTATGCCCCACTTCTGAGACAAGCAATTGAGCTGTTTCGACCGGCAGGGA
GCCGAACCTGCCTTCTTTTCGGCCTGGAATAATCATATGTGGCCTGGAGAAACAGCTAAA
GTGCGAAAGCGGCGGGCCGGCCGACGCCCTTGACGATTTTGAAGTAGACATGCTCCAGCCG
ATGCCCCTGACGACTTTGACCTTGATATGCTGCCTGCTGACGCTCTTGACGATTTTGACCTT
GACATGCTCCCCGGGTAATATGCATGAATTCGATCTAGAGGGCCCTATTCTATAGTGTACCT
AAATGTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCTCCCCCGTGCCTTCTTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCTTAATAA
AATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGG
GCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT
CTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGAATCAAGCTATCAAGTGCCACCTG
ACGTCTCCCTATCAGTGATAGAGAAGTCGACACGTCTCGAGCTCCCTATCAGTGATAGAGAA
GGTACGTCTAGAACGTCTCCCTATCAGTGATAGAGAAGTCGACACGTCTCGAGCTCCCTATC
AGTGATAGAGAAGGTACGTCTAGAACGTCTCCCTATCAGTGATAGAGAAGTCGACACGTCTC
GAGCTCCCTATCAGTGATAGAGAAGGTACGTCTAGAACGTCTCCCTATCAGTGATAGAGAAG
TCGACACGTCTCGAGCTCCCTATCAGTGATAGAGAAGGTACCCCTATATAAGCAGAGAGAT
CTGTTCAAATTTGAACTGACTAAGCGGCTCCCGCCAGATTTTGGCAAGATTACTAAGCAGGA
AGTCAAGGACTTTTTTGCTTGGGCAAAGGTCAATCAGGTGCCGGTGACTCACGAGTTTAAAG
TTCCAGGGAATTGGCGGGAATAAAGGGGCGGAGAAATCTCTAAAACGCCCACTGGGTGAC
GTCACCAATACTAGCTATAAAAGTCTGGAGAAGCGGGCCAGGCTCTCATTTGTTCCCGAGAC
GCCTCGCAGTTTCAGACGTGACTGTTGATCCCGCTCCTCTGCGACCGCTAGCTTCGATCAACT
ACGCAGACAGGTACCAAAACAAGTGTCTCGTCACGTGGGCATTAATCTGATTCTGTTTCCC
TGCAGACAATGCGAGAGAATGAATCAGAACTCAAATATCTGCTTCACTCACGGACAGAAAGA
CTGTTTAGAGTGCTTTCCCGTGTGAGAATCTCAACCCGTTTCTGTCGTCAAAAAGGCGTATC
AGAACTGTGCTACATTCATCATATCATGGGAAAGGTGCCAGACGCTTGCACTGCCTGCGAT
CTGGTCAATGTGGATTTGGATGACTGCATCTTTGAACAATAAATGACTTAAGCCAGGTATGG
CTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTTAGTGAAGGAATTCGCGAGTGG
TGGGCTTTGAAACCTGGAGCCCCCAACCCAAGGCAAATCAACAACATCAAGACAACGCTAG
AGGTCTTGTGCTTCCGGGTTACAAATACCTTGACCCGGCAACGGACTCGACAAGGGGGAGC

[0124]

CGGTCAACGCAGCAGACGCGGCGGCCCTCGAGCACGACAAAGCCTACGACCAGCAGCTCAAG
 GCGGGAGACAACCCGTACCTCAAGTACAACCACGCCGACGCCGAGTTCCAGGAGCGGCTCAA
 AGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAAAAGAGGCTTC
 TTGAACCTCTTGGTCTGGTTGAGGAAGCGGCTAAGACGGCTCCTGGAAAGAAGAGGCCTGTA
 GAGCAGTCTCCTCAGGAACCGGACTCCTCCGCGGGTATTGGCAAATCGGGTGACAGCCCGC
 TAAAAAGAGACTCAATTTCCGTCAGACTGGCGACACAGAGTCAGTCCCAGACCCTCAACCAA
 TCGGAGAACCTCCCGCAGCCCCCTCAGGTGTGGGATCTCTTACAATGGCTTCAGGTGGTGGC
 GCACCACTGGCAGACAATAACGAAGGTGCCGATGGAGTGGGTAGTTCTCGGGAAATTGGCA
 TTGCGATTCCCAATGGCTGGGGGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGC
 CCACCTACAACAATCACCTCTACAAGCAAATCTCCAACAGCACATCTGGAGGATCTTCAAAT
 GACAACGCCTACTTCGGCTACAGCACCCCTGGGGGTATTTTGACTTCAACAGATTCCACTG
 CCACTTCTCACCACGTGACTGGCAGCGACTCATCAACAACAACCTGGGGATTCCGGCCTAAGC
 GACTCAACTTCAAGCTCTTTAACATTTCAGGTCAAAGAGGTTACGGACAACAATGGAGTCAAG
 ACCATCGCCAATAACCTTACCAGCACGGTCCAGGTCTTCACGGACTCAGACTATCAGTCCC
 GTACGTGCTCGGGTTCGGCTCACGAGGGCTGCCCTCCCGCCGTTCACGCGGACGTTTTTCATGA
 TTCCTCAGTACGGGTATCTGACGCTTAATGATGGAAGCCAGGCCGTGGGTGCTTCGTCCTTT
 TACTGCCTGGAATATTTCCCGTCGCAAATGCTAAGAACGGGTAAACAATTCAGTTTCAGCTA
 CGAGTTTGAGAACGTACCTTTCCATAGCAGCTACGCTCACAGCCAAAGCCTGGACCGACTAA
 TGAATCCACTCATCGACCAATACTTGTACTATCTCTCTAGAACTATTAACGGTTCTGGACAG
 AATCAACAACGCTAAAATTTCAGTGTGGCCGGACCCAGCAACATGGCTGTCCAGGGAAGAAA
 CTACATACCTGGACCCAGCTACCGACAACAACGTGTCTCAACCACTGTGACTCAAAACAACA
 ACAGCGAATTTGCTTGGCCTGGAGCTTCTTCTGGGCTCTCAATGGACGTAATAGCTTGATG
 AATCCTGGACCTGCTATGGCCTCTCACAAGAAGGAGAGGACCGTTTCTTTCTTTGTCTGG
 ATCTTTAATTTTGGCAAACAAGGTACTGGCAGAGACAACGTGGATGCGGACAAAGTCATGA
 TAACCAACGAAGAAGAAATTAAACTACTAACCCGGTAGCAACGGAGTCTATGGACAAGTG
 GCCACAAACCACAGAGTGCCCAAATTTGGCGGTGCCTTTTAAGGCACAGGCGCAGACCGG
 TTGGGTTCAAACCAAGGAATACTTCGGGTATGGTTTGGCAGGACAGAGATGTGTACCTGC
 AAGGACCCATTTGGGCCAAAATTCCTCACACGGACGGCAACTTTCACCCTTCTCCGCTGATG
 GGAGGGTTTGAATGAAGCACCCGCCTCCTCAGATCCTCATCAAAAACACACCTGTACCTGC
 GGATCCTCCAACGGCCTTCAACAAGGACAAGCTGAACTCTTTCATCACCCAGTATTCTACTG
 GTCAAGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAGCGCTGGAACCCG
 GAGATCCAGTACACTTCCAATATTACAAGTCTAATAATGTTGAATTTGCTGTTAATACTGA
 AGGTGTATATAGTGAACCCCGCCCATTTGGCACCAGATACCTGACTCGTAATCTGTAAGTCG
 ACTTGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTTGAACTTTGGTCTCTGCGAAGG

[0125]

GCAATTCGTTTAAACCTGCAGGACTAGAGGTCCTGTATTAGAGGTCACGTGAGTGTTTTGCG
ACATTTTTCGACACCATGTGGTCACGCTGGGTATTTAAGCCCGAGTGAGCACGCAGGGTCTC
CATTTTGAAGCGGGAGGTTTGAACGCGCAGCCGCAAGCCGAATTCGTCAGATATCACATGT
CCTAGGAACTATCGATCCATCACACTGGCGGCCGCTCGACTAGAGCGGCCGCCACCGCGGTG
GAGCTCCAGCTTTTTCGCGACCGAATCGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTCGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCA
TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTTCCCTTGGAAGCTCCCTCGTGGCGCTCTCCTGTTCCGACCCTGCGCGTTACCGGATAC
CTGTCCGCCTTTCTCCCTTCGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCT
CAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCAGAACCCCGCTTCAGCCCG
ACCGCTGCGCCTTATCCGGTAACATCGTCTTGAGTCCAACCCGCTAAGACACGACTTATCG
CCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTC
TGCTGAAGCCAGTTACCTTCGGAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
GCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCA
AGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAACTCACGTTAAG
GGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAAT
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTGTTTCATCCATAGTTGCCTGACTCCCCG
TCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGA
GCGCAGAAGTGGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAG
CTAGAGTAAGTAGTTCGCCAGTTAATAGTTTTCGCGAACGTTGTTGCCATTGCTACAGGCATC
GTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTG
TCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTT
ACTGTCTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTG
AGAATAGTGTATGCGCGCAGCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC
CACATAGCAGAACTTTAAAAGTGCTCATCATTTGAAAACGTTCTTCGGGGCGAAACTCTCA
AGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCAACTGATCTTC
AGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAATGTTGAATACTCATACTCTTCTTTTCAATATTAT
TGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAA

[0126]

TAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC-3' (SEQ ID NO:28).

[0127]

[0128]

"투여하다"는 대상체에게 본원에 기재된 1개 이상의 조성물, 구축물 또는 바이러스 벡터를 제공하는 것을 의미한다. 예로서 및 비제한적으로, 투여는, 예를 들어 와우 내로의 주사에 의해 수행될 수 있다. 돌연변이에 의해 영향을 받는 세포에 조성물을 전달하는 다른 경로가 사용될 수 있다 (예를 들어, 정맥내, 직접 주사, 피하, 혈관 및/또는 비-혈관 정맥내). 투여는, 예를 들어 볼루스 주사에 의해 또는 시간 경과에 따른 점진적 관류에 의해 이루어질 수 있다.

[0129]

"작용제"는 임의의 소분자 화학적 화합물, 항체, 핵산 분자 또는 폴리펩티드 또는 그의 단편을 의미한다.

[0130]

"변경"은 표준 공지된 방법, 예컨대 본원에 기재된 방법에 의해 검출된 바와 같은 유전자 또는 폴리펩티드의 발현 수준 또는 활성의 변화 (증가 또는 감소)를 의미한다. 본원에 사용된 변경은 10% 이상 (예를 들어, 20%, 25%, 30%, 40%, 50%)의 발현 수준의 변화를 포함할 수 있다.

[0131]

"호전시키다"는 질환 또는 상태의 발생 또는 진행을 감소, 저하, 억제, 약화, 경감, 정지 또는 안정화시키는 것

을 의미한다. 예시적인 질환 또는 상태는 임의의 질환, 예컨대 유전자 돌연변이와 연관된 것을 포함할 수 있다. 한 실시양태에서, 질환은 우성 돌연변이 또는 열성 돌연변이, 예를 들어 비제한적으로, 난청, 상염색체 열성 1A (DFNB1A); 난청, 상염색체 열성 1B (DFNB1B); 난청, 상염색체 열성 2 (DFNB2); 난청, 상염색체 열성 4 (DFNB4); 난청, 상염색체 열성 6 (DFNB6); 난청, 상염색체 열성 7 (DFNB7); 난청, 상염색체 열성 8 (DFNB8); 난청, 상염색체 열성 9 (DFNB9); 난청, 상염색체 열성 10 (DFNB10); 난청, 상염색체 열성 16 (DFNB16); 난청, 상염색체 열성 24 (DFNB24); 난청, 상염색체 열성 25 (DFNB25); 난청, 상염색체 열성 59 (DFNB59); 난청, 상염색체 열성 98 (DFNB98); 난청, 상염색체 열성 110 (DFNB110); 선천성 이상각화증, 상염색체 열성 5 (DKCB5); 난청, 상염색체 우성 2B (DFNA2B); 난청, 상염색체 우성 3A (DFNA3A); 난청, 상염색체 우성 9 (DFNA9); 난청, 상염색체 우성 30 (DFNA30); 난청, 상염색체 우성 36 (DFNA36); 상염색체 우성 고립성 신경감각 난청 유형 (DFNA); 선천성 이상각화증, 상염색체 우성 4; 난청, 상염색체 우성 비증후군성 감각신경 (NSSL); 볼프람-유사 증후군, 상염색체 우성 (WFSL)과 연관될 수 있다.

[0132] 본원에 기재된 이중-벡터 시스템이 치료할 수 있는 추가의 질환 또는 상태는 불완전 상아질발생증 (DGI) 1; 난청, 상염색체 열성 16, 난청-불임 증후군 (DIS), CATSPER-관련 남성 불임증; 정자발생 실패 7 (SPGF7); 어서 증후군, 유형 I (USH1); 블루 증후군 (BLM); 총배설장 외변증; 펜드레드 증후군 (PDS); 뇌회형 맥락막 망막 위축 (GACR); 백내장 41 (CTRCT41); 전립선암; 및 유방암을 포함할 수 있으나 이에 제한되지는 않는다.

[0133] 본 개시내용은 관심 물질을 보유하거나 보유할 수 있는 임의의 세포인 세포 (예를 들어, 숙주 세포)를 제공할 수 있다. 종종 숙주 세포는 포유동물 세포 (예를 들어, 인간, 개, 고양이, 말, 무린)이다. 숙주 세포는, 예를 들어 AAV 구축물, AAV 플라스미드, 헬퍼 구축물, 보조 기능 벡터 등을 수용할 수 있다. 본원에 사용될 수 있는 숙주 세포는 형질감염된 원래 세포의 자손을 포함한다. 본 개시내용의 "숙주 세포"는 또한 외인성 DNA 서열로 형질감염된 세포를 지칭할 수 있다. 단일 모 세포의 자손은 자연적, 비의도적 또는 고의적 돌연변이의 관점에서, 원래 모 세포와 형태학 또는 계통 또는 총 DNA 상보체가 완전히 동일하지 않을 수 있다. 본원에 사용된 용어 "형질감염"은 세포에 의한 외래 DNA의 흡수를 지칭할 수 있고, 세포는 외인성 DNA가 세포 막 내부에 도입된 경우에 "형질감염된" 것이다. 다수의 형질감염 기술이 일반적으로 관련 기술분야에 공지되어 있다. 예를 들어, 문헌 [Graham et al. (1973) Virology, 52:456, Sambrook et al. (1989) Molecular Cloning, a laboratory manual, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) Basic Methods in Molecular Biology, Elsevier, and Chu et al. (1981) Gene 13:197]을 참조한다. 이러한 기술을 사용하여 1개 이상의 외인성 핵산, 예컨대 뉴클레오티드 통합 벡터 및 다른 핵산 분자를 적합한 숙주 세포 내로 도입할 수 있다.

[0134] "핵산", "뉴클레오티드 서열" 및 "폴리뉴클레오티드 서열"은 단일- 또는 이중-가닥 형태의 데옥시리보뉴클레오티드 또는 리보뉴클레오티드 중합체를 의미하며, 달리 제한되지 않는 한, 자연 발생 뉴클레오티드와 유사한 방식으로 기능할 수 있는 자연 뉴클레오티드의 공지된 유사체를 포괄한다.

[0135] 용어 "실질적으로 상동인", "실질적으로 동일한" 또는 "실질적으로 상응하는"은 선택된 핵산 또는 아미노산 서열이 선택된 참조 핵산 또는 아미노산 서열과 비교하여 적어도 70% 서열 동일성을 갖는 것인 핵산 또는 아미노산 서열의 특징을 지칭한다. 선택된 서열 및 참조 서열은 적어도 75% 이상 (예를 들어, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%)의 서열 동일성을 가질 수 있다.

[0136] 서열 동일성 또는 상동성은 비교되는 서열의 전체 길이에 걸쳐 결정될 수 있거나 또는 선택된 참조 서열의 길이보다 총 25% 이하 (예를 들어, 20%, 15%, 10%, 5%)일 수 있는 서열의 단편에 의해 결정될 수 있다. 참조 서열은 보다 큰 서열의 부분, 예를 들어 유전자 또는 플랜킹 서열의 부분, 또는 염색체의 반복 부분일 수 있다. 2개 이상의 폴리뉴클레오티드 서열은 전형적으로 적어도 18-25개의 뉴클레오티드, 적어도 26-35개의 뉴클레오티드 또는 적어도 40개 (예를 들어, 50, 60, 70, 80, 90, 100, 500, 1000, 1500, 2000)의 뉴클레오티드를 갖는 참조 서열과 비교될 수 있다. 서열 동일성 또는 상동성은 널리 공지된 서열 비교 알고리즘, 예컨대 FASTA 생물학적 서열 정렬/비교 소프트웨어 프로그램을 사용하여 결정될 수 있다 (예를 들어, 문헌 [Pearson and Lipman, 1985, 1988] 참조).

[0137] 관심 유전자 코딩 서열의 부분, 예를 들어 스테레오실린 단백질을 코딩하는 STRC 유전자를 세포로 전달하기 위한 벡터, 플라스미드 등을 포함한 폴리뉴클레오티드가 본원에 제공된다. 일부 실시양태는 19 kb의 29개 엑손을 함유하는 인간 STRC 유전자로부터 유래된 코딩 서열을 제공할 수 있다 (예를 들어, NCBI 진 ID: 161497; AC016135.3; NG_011636.1; NCBI 뉴클레오티드 ID: NM_153700.2; AF375594 참조). 다른 실시양태는 무스

무스쿨루스) STRC 유전자를 제공할 수 있다 (예를 들어, NCBI 진 ID: 140476; AL845466; AF375593; AK144985; NM_080459 참조). 1,809개 아미노산을 갖는 스테레오실린 (STRC) 단백질 또는 STRC 전구체 (예를 들어, NCBI 단백질 ID: NP_714544; AAL35321; BAE26168; NP_536707; 유니프랏KB/스위스 프랏: Q7RTU9 참조)는 내이 내 외유모 세포의 부동섬모에서 발견되는 큰 세포외 구조 단백질이다.

STRC 유전자 (예를 들어, NCBI 수탁 번호 NG_011636; OMIM # 603720; MIM 606440)는 부동섬모 단부의 적절한 응집 및 위치설정에 중요한 수평 상단 연결자 및 덮개막 연결관과 회합되어 있는, 내이의 외유모 세포의 부동섬모에서 발견되는 큰 세포외 구조 단백질인 스테레오실린 (STRC) 단백질을 코딩한다. STRC 유전자는 염색체 15q15 상에 위치하고, 상염색체 열성 DFNB16 난청 유전자좌를 규정하고, 19 kb를 포괄하는 29개의 엑손을 함유한다. "STRC 폴리뉴클레오티드"는 STRC 폴리펩티드 또는 그의 단편을 코딩하는 핵산 분자를 의미한다.

상염색체 열성 비-증후군성 난청인 DFNB16 청력 상실의 기저 원인은 돌연변이된 STRC 유전자와 연관되었다. 인간 스테레오실린 (STRC) 단백질 (NCBI RefSeq: NP_714544)을 코딩하는 인간 뉴클레오타이드 STRC 코딩 서열을 함유하는 인간 STRC 유전자 서열 (유전자 ID: 161497)이며, 신호 서열이 없는 인간 STRC 코딩 서열 (예를 들어, 서열식별번호: 29)은 하기와 같다:

5'-GTGACTCTGGCCCCCTACTGGGCCCTCATTTCCCTGGACCTGGTCTCTCCTTTCCTGAAGTC
ATTGCTCTCCACTCTGGACCAGGCTCCCCAGGGCTCCCTGAGCCGCTCACGGTTCTTTACAT
TCCTGGCCAACATTTCTTCTTCTTTGAGCCTGGGAGAATGGGGGAAGGACCAGTAGGAGAG
CCCCCACCTCTCCAGCCGCTGCTCTGCGGCTCCATGATTTTCTAGTGACACTGAGAGGTAG
CCCCGACTGGGAGCCAATGCTAGGGCTGCTAGGGGATATGCTGGCACTGCTGGGACAGGAGC
AGACTCCCCGAGATTTCTGGTGCACCAGGCAGGGGTGCTGGGTGGACTTGTGGAGGTGCTG
CTGGGAGCCTTAGTTCTTGGGGGCCCCCTACCCCAACTCGGCCCCCATGCACCCGTGATGG
GCCGTCTGACTGTGTCTTGGCTGCTGACTGGTTGCCTTCTCTGCTGCTGTTGTTAGAGGGCA
CACGCTGGCAAGCTCTGGTGCAGGTGCAGCCAGTGTGGACCCACCAATGCCACAGGCCTC
GATGGGAGGGAGGCAGCTCCTCACTTTTTGCAGGGTCTGTTGGGTTTGCTTACCCCAACAGG
GGAGCTAGGCTCCAAGGAGGCTCTTTGGGGCGGTCTGCTACGCACAGTGGGGGGCCCCCTCT
ATGCTGCCTTTTCAGGAGGGGCTGCTCCGTGTCACTCACTCCCTGCAGGATGAGGTCTTCTCC
ATTTTGGGGCAGCCAGAGCCTGATACCAATGGGCAGTGCCAGGGAGGTAACCTTCAACAGCT
GCTCTTATGGGGCGTCCGGCACAACCTTTCTTGGGATGTCCAGGCCTGGGCTTTCTGTCTG
GATCACCACCCCCACCCCTGCCCTCCTTCACTGCCTGAGCACGGGCGTGCTCTGCCCAGA
GCTTCTCAGCCGTGAGCCACATCAGCCCACGCCAACGGCGAGCCATCACTGTGGAGGCCCT
CTGTGAGAACCACCTTAGGCCACGACCACCTACAGCATTTCCAACCTTCTCCATCCACTTGC
TCTGCCAGCACACCAAGCCTGCCACTCCACAGCCCCATCCCAGCACCCTGCCATCTGCCAG
ACAGCTGTGTGGTATGCAGTGTCTTGGGCACCAGGTGCCCCAAGGCTGGCTACAGGCCTGCCA
CGACCAGTTTCTGATGAGTTTTTGGATGCGATCTGCAGTAACCTCTCCTTTTCAGCCCTGT
CTGGCTCCAACCGCCGCTGGTGAAGCGGCTCTGTGCTGGCCTGCTCCACCCCCCTACCAGC
TGCCCTGAAGGCCTGCCCCCTGTTCCCTCACCCAGACATCTTTTGGGGCTGCTTCTTGA
GAATGAGACTCTGTGGGCTGAGCGACTGTGTGGGGAGGCAAGTCTACAGGCTGTGCCCCCA
GCAACCAGGCTTGGGTCCAGCATGTGTGCCAGGGCCCCACCCAGATGTCACTGCCTCCCCA
CCATGCCACATTGGACCTGTGGGGAACGCTGCCCGGATGGGGGCAGCTTCTGGTGATGGT
CTGTGCCAATGACACCATGTATGAGGTCTTGGTGCCCTTCTGGCCTTGGCTAGCAGGCCAAT
GCAGGATAAGTCGTGGGGCAATGACACTTGCTTCTAGAAAGGGCTGCTGGGCCCCCTTCTG
CCCTCTCTGCCACCCTGGGACCATCCCCACTCTGTCTGACCCCTGGCCCCCTTCTCTTGG
CATGCTATCCCAAGTTGCCACGCTGTGAGTCTCTGTCCAGCTCTTGTCACCCCACAGCGC

TACACTATCTCCTCCGCTGCTGACCTTCCTCTTGGGTCCAGGGGCTGGGGGCGCTGAGGCC
CAGGGGATGCTGGGTGCGGCCCTACTGCTCTCCAGTCTCCCAGACAACTGCTCCTTCTGGGA
TGCCTTTCGCCCAGAGGGCCGGCGCAGTGTGCTACGGACGATTGGGGAATACCTGGAACAAG
ATGAGGAGCAGCCAACCCCATCAGGCTTTGAACCCACTGTCAACCCAGCTCTGGTATAAGC
AAGATGGAGCTGCTGGCCTGCTTTAGTCCTGTGCTGTGGGATCTGCTCCAGAGGAAAAGAG
TGTTTGGGCCCTGCAGATTCTAGTGCAGGCGTACCTGCATATGCCCCAGAAAACCTCCAGC
AGCTGGTGCTTTTCAGCAGAGAGGGAGGCTGCACAGGGCTTCCTGACACTCATGCTGCAGGGG
AAGCTGCAGGGGAAGCTGCAGGTACCACCATCCGAGGAGCAGGCCCTGGGTGCGCTGACAGC
CCTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAGCTCTTCATTGACCTGTCAACACTCA
TCCCTTTCTTGGCTGTCTCTGACCTGATGCGCTTCCCACCATCCCTGTTAGCCAACGACAGT
GTCTTGGCTGCCATCCGGGATTACAGCCCAGGAATGAGGCCTGAACAGAAGGAGGCTCTGGC
AAAGCGACTGCTGGCCCCCTGAACTGTTTGGGGAAGTGCTGCGCTGGCCCCAGGAGCTGCTGT
GGGCAGTGCTGCCCCCTGCTCCCCACCTCCCTCTGGAGAACTTTTTCAGCTCAGCCCTCAC
CAGATCCAGGGCCCTGGAGGATAGCTGGCCAGCAGCAGGTCTGGGGCCAGGGCATGCCCCCCA
TGTGCTGCGCAGCCTGGTAAACCAGAGTGTCCAGGATGGTGAGGAGCAGGTACGCAGGCTTG
GGCCCCTCGCTGTTTCTGAGCCCTGAGGAGCTGCAGAGCCTAGTGCCCCTGAGTGATCCA
ACGGGGCCAGTAGAACGGGGGCTGCTGGAATGTGCAGCCAATGGGACCCTCAGCCCAGAAGG
ACGGGTGGCATATGAACTTCTGGGTGTGTTGCGCTCATCTGGAGGAGCGGTGCTGAGCCCCC
GGGAGCTGCGGGTCTGGGCCCTCTCTTCTCTCAGCTGGGCCTCCGCTTCCTTCAGGAGCTG
TCAGAGCCCCAGCTTAGAGCCATGCTTCCTGTCTGCAGGGAAGTAGTGTTACACCTGCTCA
GGCTGTCTGCTGCTTGGACGGCTCCTTCTAGGCACGATCTATCCCTGGAGGAACTCTGCT
CCTTGACCTTCTGTACACAGGCTCAGCCCCCAGACACTCCAGGCCATCCCTAGGCGAGTC
CTGGTGGGGCTTGTTCCTGCCTGGCCCCCTGAACTGTACGCCTCTCAGCCTGCCAGACCGC
AGCACTGCTGCAGACCTTTCGGGTAAAGATGGTGTTAAAAATATGGGTACAACAGGTGCTG
GTCCAGCTGTGTGTATCCCTGGTCAGCCTATTCCCACCACCTGGCCAGACTGCCTGCTTCCC
CTGCTCCCATTAAGCTGCTACAACCTGGATTCCCTTGGCTCTTCTGGCAAATCGAAGACGCTA
CTGGGAGCTGCCCTGGTCTGAGCAGCAGGCACAGTTTCTCTGGAAGAAGATGCAAGTACCCA
CCAACCTTACCCTCAGGAATCTGCAGGCTCTGGGCACCCTGGCAGGAGGCATGTCCTGTGAG
TTTCTGCAGCAGATCAACTCCATGGTAGACTTCCTTGAAGTGGTGACATGATCTATCAGCT
GCCCCACTAGAGTTTCGAGGGAGCCTGAGGGCCTGTATCTGGGCAGAGCTACAGCGGAGGATGG
CAATGCCAGAACCAGAATGGACAACCTGTAGGGCCAGAAGTGAACGGGCTGGATAGCAAGCTA
CTCCTGGACTTACCAGTCCAGTTGATGGACAGACTATCCAATGAATCCATTATGTTGGTGGT
GGAGCTGGTGCAAAGAGCTCCAGAGCAGCTGCTGGCACTGACCCCCCTCCACCAGGCAGCCC
TGGCAGAGAGGGCACTACAAAACCTGGCTCCAAAGGAGACTCCAGTCTCAGGGGAAGTGCTG

[0141]

GAGACCTTAGGCCCTTTGGTTGGATTCTCTGGGGACAGAGACACACGACAGATCCCCCTACA
GATCCTGCTGTCCCATCTCAGTCAGCTGCAAGGCTTCTGCCTAGGAGAGACATTTGCCACAG
AGCTGGGATGGCTGCTATTGCAGGAGTCTGTTCTTGGGAAACCAGAGTTGTGGAGCCAGGAT
GAAGTAGAGCAAGCTGGACGCCCTAGTATTCACTCTGTCTACTGAGGCAATTTCTTGTATCCC
CAGGGAGGCCCTTGGGTCCAGAGACCCCTGGAGCGGCTTCTAGAAAAGCAGCAGAGCTGGGAGC
AGAGCAGAGTTGGACAGCTGTGTAGGGAGCCACAGCTTGTCTGCCAAGAAAGCAGCCCTGGTA
GCAGGGGTGGTGCGACCAGCTGCTGAGGATCTTCCAGAACCCTGTGCCAAATTGTGCAGATGT
ACGAGGGACATTTCCAGCAGCCTGGTCTGCAACCCAGATTGCAGAGATGGAGCTCTCAGACT
TTGAGGACTGCCTGACATTATTTGCAGGAGACCCAGGACTTGGGCCTGAGGAACTGCGGGCA
GCCATGGGCAAAGCAAACAGTTGTGGGGTCCCCCCCCGGGATTTCGTCTGAGCAGATCCT
GCAGCTTGGTAGGCTCTTAATAGGTCTAGGAGATCGGGAACCTACAGGAGCTGATCCTAGTGG
ACTGGGGAGTGCTGAGCACCCCTGGGGCAGATAGATGGCTGGAGCACCACTCAGCTCCGCATT
GTGGTCTCCAGTTTCTACGGCAGAGTGGTCTGGCATGTGAGCCACCTGGACTTCGTTTCATCT
GACAGCGCTGGGTTATACTCTCTGTGGACTGCGGCCAGAGGAGCTCCAGCACATCAGCAGTT
GGGAGTTCAGCCAAGCAGCTCTCTTCTCGGCACCCTGCATCTCCAGTGCTCTGAGGAACAA
CTGGAGGTTCTGGCCACCTACTTGTACTGCCTGGTGGGTTTGGCCCAATCAGTAACTGGGG
GCCTGAGATCTTCACTGAAATTGGCACCATAGCAGCTGGGATCCCAGACCTGGCTCTTTCAG
CACTGCTGCGGGGACAGATCCAGGGCGTTACTCCTCTTGCCATTTCTGTTCATCCCTCCCTCCT
AAATTTGCTGTGGTGTTTAGTCCCATCCAACCTATCTAGTCTCACCAGTGCTCAGGCTGTGGC
TGTCACCTCTGAGCAAATGGCCCTTCTGAGTCTGAGCAGCGACGAGCAGTTGCATGGGCCC
AACATGAGGGAAAGGAGAGCCAGAACAGCAAGGTCTGAAGTACAGCCTGGGGCCTCCAGGAC
TGGTCACGACCTTCTGTTCCCTGGTATTGACTATCAGCTTCTTGGCCACCTGCTA-3'.

[0142]

[0143] 인간 STRC 코딩 서열은 도 2a-2c의 구조물에서 찾아볼 수 있고, 인간 STRC 코딩 서열의 부분은 도 7a-7b 및 도 12a-12b의 구조물에서 찾아볼 수 있다.

[0144] 인간 STRC 단백질을 코딩하는 인간 mRNA 서열 (NCBI RefSeq: NM153700) (서열식별번호: 30)은 하기와 같다:

5'-GCCCTGCCCTCACCTGGCTATCCCACACAGGTGAGAATAACCAGAACTCACCTCCGGTA
CCAGTGTTCACTTGGAACATGGCTCTCAGCCTCTGGCCCCTGCTGCTGCTGCTGCTGCTGC
TGCTGCTGCTGTCTTTGCAGTGACTCTGGCCCCCTACTGGGCCTCATTCCCTGGACCCTGGT
CTCTCCTTCCTGAAGTCATTGCTCTCCACTCTGGACCAGGCTCCCCAGGGCTCCCTGAGCCG
CTCACGGTTCTTTACATTCTGGCCAACATTCTTCTTCTCTTTGAGCCTGGGAGAAATGGGGG
AAGGACCAGTAGGAGAGCCCCACCTCTCCAGCCGCTGCTCTGCGGCTCCATGATTTTCTA

[0145]

GTGACACTGAGAGGTAGCCCCGACTGGGAGCCAATGCTAGGGCTGCTAGGGGATATGCTGGC
 ACTGCTGGGACAGGAGCAGACTCCCCGAGATTTCCTGGTGCACCAGGCAGGGGTGCTGGGTG
 GACTTGTGGAGGTGCTGCTGGGAGCCTTAGTTCCCTGGGGGCCCCCTACCCCAACTCGGCCC
 CCATGCACCCGTGATGGGCCGTCTGACTGTGTCCCTGGCTGCTGACTGGTTGCCTTCTCTGCT
 GCTGTTGTTAGAGGGCACACGCTGGCAAGCTCTGGTGCAGGTGCAGCCAGTGTGGACCCCA
 CCAATGCCACAGGCCTCGATGGGAGGGAGGCAGCTCCTCACTTTTTCAGGGTCTGTTGGGT
 TTGCTTACCCCAACAGGGGAGCTAGGCTCCAAGGAGGCTCTTTGGGGCGGTCTGTACGCAC
 AGTGGGGGCCCCCTCTATGCTGCCCTTTCAGGAGGGGCTGCTCCGTGTCACTCACTCCCTGC
 AGGATGAGGTCTTCTCCATTTTGGGGCAGCCAGAGCCTGATACCAATGGGCAGTGCCAGGGA
 GGTAACCTTCAACAGCTGCTCTTATGGGGCGTCCGGCACAACCTTTCCTGGGATGTCCAGGC
 GCTGGGCTTTCTGTCTGGATCACCACCCCCACCCCTGCCCTCCTTCACTGCCTGAGCACGG
 GCGTGCTCTGCCCAGAGCTTCTCAGCCGTGAGCCACATCAGCCACGCCAACGGCGAGCC
 ATCACTGTGGAGGCCCTCTGTGAGAACCACTTAGGCCCAGCACCACCCTACAGCATTTCCAA
 CTTCTCCATCCACTTGTCTGCCAGCACACCAAGCCTGCCACTCCACAGCCCCATCCCAGCA
 CCACTGCCATCTGCCAGACAGCTGTGTGGTATGCAGTGTCTGGGCACCAGGTGCCCCAAGGC
 TGGCTACAGGCCTGCCACGACCAGTTTCCTGATGAGTTTTTGGATGCGATCTGCAGTAACCT
 CTCCTTTTCAGCCCTGTCTGGCTCCAACCGCCGCTGGTGAAGCGGCTCTGTGCTGGCCTGC
 TCCACCCCTACAGCTGCCCTGAAGGCCTGCCCCCTGTTCCCTCAGCCAGACATCTTT
 TGGGGCTGCTTCTTGGAGAATGAGACTCTGTGGGCTGAGCGACTGTGTGGGAGGCAAGTCT
 ACAGGTGTGCCCCCAGCAACCAGGCTTGGGTCCAGCATGTGTGCCAGGGCCCCACCCAG
 ATGTCACTGCCTCCCCACCATGCCACATTGGACCCTGTGGGGAACGCTGCCCGGATGGGGGC
 AGCTTCCTGGTGTGGTCTGTGCCAATGACACCATGTATGAGGTCTGGTGCCCTTCTGGCC
 TTGGCTAGCAGGCCAATGCAGGATAAGTCGTGGGGCAATGACACTTGCTTCTAGAAAGGGC
 TGCTGGGGCCCCCTTCTGCCCTCTCTGCCACCACTGGGACCATCCCCACTCTGTCTGACCCCT
 GGCCCCCTCCTCCTTGGCATGCTATCCCAGTTGCCACGCTGTCACTCCTCTGTCCCAGCTCT
 TGCTCACCACACAGCCTACACTATCTCCTCCGCTGCTGACCTTCCTCTTGGGTCCAGGGG
 CTGGGGGCGCTGAGGCCAGGGGATGCTGGGTGCGGCCCTACTGCTCTCCAGTCTCCCAGAC
 AACTGCTCCTTCTGGGATGCCTTTCGCCAGAGGGCGGCGCAGTGTGCTACGGACGATTGG
 GGAATACCTGGAACAAGATGAGGAGCAGCCAACCCCATCAGGCTTTGAACCCACTGTCAACC
 CCAGCTCTGGTATAAGCAAGATGGAGCTGCTGGCCTGCTTTAGTCTGTGCTGTGGGATCTG
 CTCCAGAGGGAAAAGAGTGTGGGGCCTGCAGATTCTAGTGCAGGCGTACCTGCATATGCC
 CCCAGAAAACCTCCAGCAGCTGGTGCTTTCAGCAGAGAGGGAGGCTGCACAGGGCTTCTGA
 CACTCATGCTGCAGGGGAAGCTGCAGGGGAAGCTGCAGGTACCACCATCCGAGGAGCAGGCC
 CTGGGTGCGCTGACAGCCCTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAGCTCTTCAT

[0146]

TGACCTGTCACCACTCATCCCTTTCTTGGCTGTCTCTGACCTGATGCGCTTCCCACCATCCC
 TGTTAGCCAACGACAGTGTCCTGGCTGCCATCCGGGATTACAGCCCAGGAATGAGGCCTGAA
 CAGAAGGAGGCTCTGGCAAAGCGACTGCTGGCCCCTGAAGTGTGGGGGAAGTGCCTGCCTG
 GCCCCAGGAGCTGCTGTGGGCAGTGCTGCCCCGTCTCCCCACCTCCCTCTGGAGAACTTTT
 TGCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGGATAGCTGGCCAGCAGCAGGTCTGGGG
 CCAGGGCATGCCCGCCATGTGCTGCGCAGCCTGGTAAACCAGAGTGTCCAGGATGGTGAGGA
 GCAGGTACGCAGGCTTGGGCCCCCTCGCCTGTTTCTGAGCCCTGAGGAGCTGCAGAGCCTAG
 TGCCCCTGAGTGATCCAACGGGGCCAGTAGAACGGGGGCTGCTGGAATGTGCAGCCAATGGG
 ACCCTCAGCCCAGAAGGACGGGTGGCATATGAACTTCTGGGTGTGTTGCGCTCATCTGGAGG
 AGCGGTGCTGAGCCCCCGGGAGCTGCGGGTCTGGGCCCCCTCTTCTCTCAGCTGGGCCTCC
 GCTTCCTTCAGGAGCTGTCAGAGCCCCAGCTTAGAGCCATGCTTCTGTCTGCAGGGAAGT
 AGTGTACACCTGCTCAGGCTGTCTGCTGCTGGACGGCTCCTTCTAGGCACGATCTATC
 CCTGGAGGAACTCTGCTCCTTGCACCTTCTGCTACCAGGCCTCAGCCCCAGACACTCCAGG
 CCATCCCTAGGCGAGTCTGGTGGGGCTTGTTCCTGCCTGGCCCCTGAAGTGTACGCCTC
 TCAGCCTGCCAGACCGCAGCACTGCTGCAGACCTTTCGGGTAAAGATGGTGTAAAAATAT
 GGGTACAACAGGTGCTGGTCCAGCTGTGTGTATCCCTGGTCAGCCTATTCCCACCACCTGGC
 CAGACTGCCTGCTTCCCCTGCTCCCATTAAGCTGCTACAACGGATTCTTGGCTCTTCTG
 GCAAATCGAAGACGCTACTGGGAGCTGCCCTGGTCTGAGCAGCAGGCACAGTTTCTCTGGAA
 GAAGATGCAAGTACCCACCAACCTTACCCTCAGGAATCTGCAGGCTCTGGGCACCTGGCAG
 GAGGCATGTCTGTGAGTTTCTGCAGCAGATCAACTCCATGGTAGACTTCCTTGAAGTGGTG
 CACATGATCTATCAGCTGCCACTAGAGTTCGAGGGAGCCTGAGGGCCTGTATCTGGGCAGA
 GCTACAGCGGAGGATGGCAATGCCAGAACCAGAATGGACAACGTAGGGCCAGAACTGAACG
 GGCTGGATAGCAAGCTACTCCTGGACTTACCGATCCAGTTGATGGACAGACTATCCAATGAA
 TCCATTATGTTGGTGGTGGAGCTGGTGCAAAGAGCTCCAGAGCAGCTGCTGGCACTGACCCC
 CCTCCACCAGGCAGCCCTGGCAGAGAGGGCACTACAAAACCTGGCTCCAAAGGAGACTCCAG
 TCTCAGGGGAAGTGCTGGAGACCTTAGGCCCTTTGGTTGGATTCTGGGGACAGAGAGCACA
 CGACAGATCCCCCTACAGATCCTGTGTCCCATCTCAGTCAGCTGCAAGGCTTCTGCCTAGG
 AGAGACATTTGCCACAGAGCTGGGATGGCTGCTATTGCAGGAGTCTGTTCTTGGGAAACCAG
 AGTTGTGGAGCCAGGATGAAGTAGAGCAAGCTGGACGCCTAGTATTCACCTGTCTACTGAG
 GCAATTTCTTGATCCCCAGGGAGGCCTTGGGTCCAGAGACCCTGGAGCGGCTTCTAGAAAA
 GCAGCAGAGCTGGGAGCAGAGCAGAGTTGGACAGCTGTGTAGGGAGCCACAGCTTGCTGCCA
 AGAAAGCAGCCCTGGTAGCAGGGTGGTGCGACCAGCTGCTGAGGATCTTCCAGAACCTGTG
 CCAAATTGTGCAGATGTACGAGGGACATTCCAGCAGCCTGGTCTGCAACCCAGATTGCAGA
 GATGGAGCTCTCAGACTTTGAGGACTGCCTGACATTATTTGCAGGAGACCCAGGACTTGGGC

[0147]

CTGAGGAAGTGCAGGAGCCATGGGCAAAGCAAAACAGTTGTGGGGTCCCCCGGGGATTT
CGTCCTGAGCAGATCCTGCAGCTTGGTAGGCTCTTAATAGGTCTAGGAGATCGGGAAGTACA
GGAGCTGATCCTAGTGGACTGGGGAGTGCTGAGCACCTGGGGCAGATAGATGGCTGGAGCA
CCACTCAGCTCCGCATTGTGGTCTCCAGTTTCTACGGCAGAGTGGTCGGCATGTGAGCCAC
CTGGACTTCGTTTCATCTGACAGCGCTGGGTTTATACTCTCTGTGGACTGCGGCCAGAGGAGCT
CCAGCACATCAGCAGTTGGGAGTTCAGCCAAGCAGCTCTCTTCCTCGGCACCTGCATCTCC
AGTGCTCTGAGGAACAAGTGGAGGTTCTGGCCACCTACTTGTACTGCCTGGTGGGTTTGGC
CCAATCAGTAAGTGGGGGCTGAGATCTTCACTGAAATTGGCACCATAGCAGCTGGGATCCC
AGACCTGGCTCTTTTCAGCACTGCTGCGGGGACAGATCCAGGGCGTTACTCCTCTTGCCATTT
CTGTCTATCCCTCCTCTAAATTTGCTGTGGTGTCTTAGTCCCATCCAAGTATCTAGTCTCACC
AGTGCTCAGGCTGTGGCTGTCACTCTGAGCAAATGGCCTTTCTGAGTCTGAGCAGCGACG
AGCAGTTGCATGGGCCCCAATGAGGGAAAGGAGAGCCAGAACAGCAAGGTCGAAGTACAG
CCTGGGGCCTCCAGGACTGGTCACGACCTTCCTGGTCCCTGGTATTGACTATCAGCTTCCTT
GGCCACCTGCTATGAGCCTGTCTCTACAGTAGAAGGAGATTGTGGGGAGAGAAATCTTAAGT
CATAATGAATAAAGTGCAGAACAGAAAGTGCATCCTGATTATTTTCAGAAAGCTGATGAGGAATA
-3'.

[0148]

[0149]

스테레오실린 발현은 내이의 감각 유모 세포에서만 발견되었고, 부동섬모, 즉 소리 자극의 기계적감수를 위한 구조를 형성하는 강성 미세용모와 회합된다. 신호 펩티드 서열 (아미노산 1-21에서; 밑줄표시)을 포함하고, 링커 서열 또는 Myc 태그 서열을 포함하지 않으며, 스플라이스 부위가 Ala708과 Cys709 또는 Ala933과 Cys934 (볼드체, 밑줄표시) 사이에 존재할 수 있는 인간 STRC 단백질 (1,775개 아미노산; 서열식별번호: 25)은 하기와 같다:

MALSLWPLLLLLLLLLLLLLLSFAVTLPAPTGPDSLDPGLSFLKSLLSTLDQAPQGSLSRSRFFTF
LANISSSFEPGRMGEGPVGEPPPLQPPALRLHDFLVTLRGSPDWEPMGLLGDMLALLGQEQ
TPRDFLVHQAGVLGGLEVEVLLGALVPGGPPTPTRPPCTRDGPSDCVLAADWLPSLLLLLEGT
RWQALVQVQPSVDPTNATGLDGREAPHFLQGLLGLLTPTGELGSKEALWGGLLRVTGAPLY
AAFQEGLLRVTHSLQDEVFSILGQPEPDTNGQCQGGNLQQLLLWGVHRNLSWDVQALGFLSG
SPPPPPALLHCLSTGVPLPRASQPSAHISPRQRRAITVEALCENHLGPAPPYSISNFSIHLL
CQHTKPATPQPHPSTTAICQTAVWYAVSWAPGAQGWLQACHDQFPDEFDAICSNLSFSALS
GSNRRLVKRLCAGLLPPPTSCPEGLPPVPLTPDIFWGCFLNETLWAERLCGEASLQAVPPS
NQAWVQHVCQGPTPDVTASPPCHIGPCGERCPDGGSLVMVCANDTMYEVLVPFWPWLQAGQC
RISRGNDTCFLEGLLGPPLPSPLCLTPGPFLLGMLSQLPRCQSSVPALAHPTRL
HYLLRLLTFLLGPGAGGAEAQGMLGRALLSSLPDNCSEWDAFRPEGRRSVLRITIGEYLEQD

[0150]

EEQPTPSGFTEPTVNPSSGISKMELIACFSPVLWDLLQREKSVWALQILVQAYLHMPPENLQQ
 LVLSAEREAAGQFLLTMLQGLQGLQVPPSEEQALGRLTALLQRYPRLTSQLFIDLSPLI
 PFLAVSDLMRFPPSLLANDSVLAAIRDYSPGMRPEQKEALAKRLLAPELFGVPAWPQELLW
 AVLPLLPPLPLENFLQLSPHQIQALEDSWPAAGLPGHARHVLRLSVNQSVQDGEQVRRLG
 PLACFLSPEELQSLVPLSDPTGPVERGLLECAANGTLSPEGRVAYELLGVLRSSGGAVLSPR
 ELRVWAPLFSQLGLRFLQELSEPQLRAMLPVLQGTSTVTPAQAVLLGRLLPRHDLSEELCS
 LHLLLPGLSPQTLQAIPIRRVLVGACSCLAPELSRLSACQTAALLQTFRVKDGVMNGTTGAG
 PAVCIPGQPIPTTWPDCLLPLLPKLLQLDSLALLANRRRYWELPWSEQQAQFLWKMQVPT
 NLTLRNLQALGTLAGGMSCEFLQQINSMVDFLEVVMHIYQLPTRVRGSLRACIWAELQRRMA
 MPEPEWTTVGPELNLGLSKLLLDLPQLMDRLSNESIMLVVELVQRAPEQLLALTPHQAAL
 AERALQNLAPKETPVSSEVLETLGPLVGFLGTSTRQIPLQIILSHLSQLQGFCLETGFATE
 LGWLLQLQESVLGKPELWSQDEVEQAGRLVFTLSTEAIISLIPREALGPETLERLLEKQQSWEQ
 SRVQQLCREPQLAAKKAALVAGVVRPAAEDLPEPVPNCADVGRGTFPAAWSATQIAEMELSDF
 EDCLTLFAGDPGLGPEELRAAMGKAKQLWGPPRGFRPEQILQLGRLLIGLGDRELQELILVD
 WGVLSLTLGQIDGWSTTQLRIVVSSFLRQSGRHVSHLDFVHLTALGYTLCLGRPEELQHISW
 EFSQAALFLGTLHLQCSEEQLEVLHLVLPGGFGPISNWGPEIFTEIGTIAAGIPDLALSA
 LLRGQIQGVTPLAISVIPPPKFAVVFSPQLSSLSAQAVAVTPEQMAFLSPEQRRAVAWAQ
 HEGKESPEQQGRSTAWGLQDWSRPSWSLVLTISFLGHLL**.

[0151]

[0152]

추정 신호 펩티드 및 여러 소수성 부분을 포함한 1,809개 아미노산을 포함하는 뮤린 STRC 단백질 (NCBI RefSeq: NP_536707)을 코딩하는 뮤린 (무스 무스쿨루스) STRC 유전자 (유전자 ID: 140476; 염기 쌍 79-5508에서의 CD S)이며, 신호 서열이 없는 뮤린 STRC 코딩 서열 (예를 들어, 서열식별번호: 31)은 하기와 같다:

5'-GCCCTACTGGGCCTCAGTCTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTCGTAGC
 CACTCTGGACCAAGCTCCTCAGCGTTCCTCAGCCAGTCACGGTCTCTGCGTTCCTGGCCA
 ACATTTCTTCATCCTTCAGCTTGGGAGGATGGGGGAGGGACCGGTGGGAGAGCCCCACCT
 CTCCAGCCCCCTGCACTTCGACTTCATGATTTCTCGTGACACTGAGAGGTAGCCAGACTG
 GGAGCCAATGCTAGGGCTTCTGGGAGATGTGCTGGCACTCCTGGGACAGGAACAGACTCCCC
 GGGACTTTTGGTGACACAGGCAGGTGTACTGGGTGGACTTGTAGAGGCATTGTTGGGAGCG
 TTAGTTCTGGAGGCCCCCTGCCCCACTCGACCCCATGCACCCGTGATGGCCCTTCTGA
 CTGTGTCTGGCTGCTGATTGGTTGCCTTCTCTGATGTTGTTATTAGAGGGTACACGCTGGC
 AGGCCCTGGTGCAGTTGCAGCCAGTGTGGACCAACCAATGCCACAGGTCTTGATGGTAGA
 GAGCCAGCTCCTCACTTTTACAGGGTCTGCTGGGCTTGCTTACCCAGCAGGAGAGTTGGG
 CTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGTGGGGGCCCCCTCTATGCTGCCT

[0153]

TCCAGGAGGGGCTACTGCGAGTCACTCATTCTCTGCAAGATGAGGTCTTTTCTATTATGGGA
CAGCCAGAGCCTGATGCCAGTGGGCACTGCCAGGGAGGCAACCTTCAACAGCTGCTTTTATG
GGGCATGCGGAACAACCTTTCTTGGGACGCCCCGAGCACTGGGTTTCTATCTGGATCACCAC
CTCCACCCCTGCTCTCCTGCACTGCCTGAGCAGAGGTGTGCCTCTGCCCAGGGCTTCCCAG
CCTGCGGCTCACATCAGCCCTCGACAGCGGCGAGCCATCTCTGTGGAGGCCCTCTGCGAGAA
CCACTCAGGCCCAGAGCCACCCTACAGCATCTCCAACCTTCTCCATCTACTTGCTCTGCCAGC
ACATCAAGCCTGCCACCCCGCGGCCCTCCTACCACCCACGGCCTCCTCCTACCACCCCA
CAGCCCCCTCCTACCCTACACAGCCCATTCCTGACACTACACAGCCCCCTCCTGTCACCCC
AAGGCCTCCTCCTACCACCCCAACAACCCCTCCTAGCACAGCTGTCATCTGCCAGACAGCTG
TATGGTACGCAGTCTCGTGGGCACCAGGTGCCCCGAGGTGGCTCCAAGCCTGCCATGATCAG
TTTCTGATCAATTTCTGGATATGATCTGCGGCAACCTCTCATTTTCAGCCCTGTCTGGCCC
CAGTCGTCCTTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCCACCCCCACTAGCTGTCCAC
CAGGCCTGATCCCTGTGCCCCCTACCCCAGAAATATTCTGGGGCTGTTTCTGGAGAATGAG
ACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCAGGCTGTGCCCCGAGGAACCA
GGCTTGGGTTTCAGCATGTGTGTGCGGGCCCCACCTTGGACGCCACTGATTTTCCACCGTGCC
GCGTTGGACCCTGTGGGAACGCTGCCCAGATGGGGCAGCTTCTGCTCATGGTCTGTGCC
AATGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTGGCTAGCAGGCCAATGCAGAAT
TAGTCGTGGAGGAAATGATACTTGCTTTCTAGAAGGCATGCTGGGCCCCCTGTTGCCCTCTC
TGCCCCCTCTGGGACCCTCCCCACTCTGTCTGGCTCCTGGTCCCTTTCTGCTTGGCATGTTA
TCCCAGTTGCCACGCTGTCACTCCTCCGTGCCAGCCCTCGCCCCACCCACGCGCTACATTA
CCTCCTGCGCCTACTGACCTTCCCTTCTGGGTCCAGGGACTGGGGTGCCGAGACGCAGGGGA
TGTTAGGTCAAGCCCTGCTGCTCTCTAGTCTCCCAGACAACTGTTTCTGTTGATGCCTTC
CGCCCAGAGGGCCGAGAGAGTGTACTGAGGACAGTCGGAGAGTACTTGCAGCGGAAGAGCC
AACCCACAGGCTTAGACTCCTCCCTCAGCCTCGGCTCTGGTATGAGCAAGATGGAGCTTC
TGTCTGCTTCAGTCTGTACTGTGGGATCTACTCCAGAGAGAGAAGAGCGTTTGGGCCCTG
AGGACCTTGGTGAAGGCCTACCTGCGCATGCCCTCCAGAAGACCTTCCAGCAGCTTGTGCTTTC
AGCAGAGATGGAGGCTGCACAGGGCTTCCCTGACGCTCATGCTTCGTTTCTGGGCTAAGCTGA
AGGTTCAACCATCCGAGGAGCAGGCCATGGGCCGCTGACAGCCTTGCTGCTCCAGCGGTAC
CCACGCCTCACCTCCCAACTCTTTATCGACATGTCACCGCTCATCCCCCTTCTGGCTGTCCC
TGACCTCATGCGCTTCCCACCGTCCCTTTTGGCCAACGACAGTGTCTGGCTGCCATCAGGG
ATCACAGCTCAGGAATGAAGCCTGAACAGAAGGAGGCCCTGGCAAAACGACTGCTGGCCCCCT
GAGCTGTTTGGAGAAGTGCTGATTGGCCCCAGGAGCTGCTGTGGGCAGCCCTGCCTCTGCT
TCCCCATCTGCCTCTGGAGAGCTTTCTCCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGG
ATAGCTGGCCAGTAGCAGATCTTGGGCCGGGACACGCCCACATGTGCTTCGTAGCCTAGTA

[0154]

AACCAGAGCATGGAGGATGGGGAGGAGCAGGTGCTCAGGCTTGGGTCCCTCGCCTGTTTCCT
GAGTCCTGAGGAGCTACAGAGTCTGGTGCCCTTGAGTGATCCAATGGGGCCTGTAGAACAGG
GTCTGCTGGAATGTGCGGCCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATATGAACCT
CTGGGAGTGTTGCGTTCATCTGGAGGAACTGTCTTAAGCCCCGAGAGCTGAGGGTCTGGGC
ACCTCTCTTTCCCCAGCTGGGCCTCCGCTTCTGCAGGAGCTCTCAGAGACCCAGCTTAGAG
CCATGCTTCTGCCCACAGGGAGCCAGTGTCACACCTGCCCAGGCTGTTCTGTTGTTTGGGA
AGGCTCCTTCCTAAGCATGATCTGTCCCTGGAGGAACTCTGCTCCCTGCACCTCTCCTGCC
AGGTCTCAGCCCCCAGACACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGGTGCTTGTTCCT
GCCTGGGCCCTGAACTGTCAAGGCTTTCAGCTTGCCAGATTGCAGCTCTGCTGCAGACCTTT
CGGGTAAAAGATGGTGTTAAAAATATGGGTGCAGCAGGTGCCGGCTCAGCCGTGTGCATTCC
TGGGCAGCCCACCCTTGCCAGACTGCCTGCTTCCCCTGCTCCCATTAAGCTGCTACAGC
TGGACGCTGCAGCTCTTCTGGCAAACCGAAGACTCTATCGGCAGCTGCCTTGGTCTGAGCAA
CAGGCACAGTTTCTCTGGAAGAAAATGCAAGTGCCTACCAACCTGAGCCTGAGGAATCTGCA
GGCTCTGGGCAACTTGCCAGGAGGCATGACCTGCGAGTTTCTGCAGCAGATCAGCTCAATGG
TTGACTTCTTGTGATGTGGTACACATGCTCTACCAGCTGCCCACTGGTGTTGAGAGAGCCTG
CGGGCTGTATCTGGACAGAGCTACAGCGGAGGATGACAATGCCAGAGCCAGAGCTGACCAC
CCTAGGGCCAGAAGTGAAGTGAACCTGACACAAAGCTACTCCTGGACTTGCCGATCCAGCTGA
TGGACAGATTGTCCAATGATTCCATTATGTTGGTGGTGGAGATGGTCCAAGGCGCTCCAGAG
CAGCTGCTGGCACTGACCCCACTCCACCAGACAGCCTTGGCAGAGCGAGCACTAAAAACCT
GGCTCCAAAGGAGACCCCAATCTCCAAAGAAGTGTGGAGACACTGGGCCCCCTTGGTTGGAT
TCCTGGGAATAGAGAGCACGCGACGGATCCCTTTACCCATTCTACTGTCTCATCTCAGTCAG
CTGCAGGGCTTCTGCCTAGGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTGTTGCAGGA
GCCTGTTCTTGGAAAACCAGAATTGTGGAGCCAGGATGAAATAGAGCAAGCTGGACGCCTAG
TATTAAGTCTGTCTGCTGAGGCTATTTCTCGATCCCCAGGGAGGCTTTGGGCCCAGAGACA
CTGGAGAGGCTTCTGGGAAAGCATCAAAGCTGGGAGCAGAGCAGAGTGGGCCATCTGTGTGG
GGAGTCACAGCTTGCCACAAGAAAGCAGCTCTGGTAGCTGGGATTGTGCATCCAGCTGCTG
AGGGTCTCCAAGAGCCTGTACCAACTGTGCAGACATACGGGGAACCTTCCCAGCGGCCTGG
TCTGCGACACAAATCTCAGAGATGGAACCTCTCAGACTTTGAAGACTGCCTGTCACTATTTGC
TGGAGATCCAGGACTTGGTCTGAGGAACTACGGGCAGCCATGGGCAAGGCCAAGCAGTTGT
GGGGTCCCCCTCGAGGATTCCGTCTGAGCAGATCTTGACAGCTGGGCCGTCTCCTGATAGGT
CTAGGAGAACGGGAAGTGCAGGAGCTTACCTTGGTGGACTGGGGTGTGCTGAGCAGCCTGGG
GCAATAGATGGCTGGAGTTCCATGCAGCTCCGAGCCGTGGTCTCCAGTTTCTAAGGCAGA
GTGGTCGGCATGTGAGCCACCTGGACTTCATTTATCTGACAGCACTGGGTTACACAGTCTGT
GGATTGCGACCAGAGGAGTTACAGCACATCAGCAGTTGGGAGTTTAGCCAAGCAGCTCTCTT

[0155]

CCTGGGTAGCTTGCATCTCCCGTGCTCTGAGGAACAGCTGGAAGTTCTGGCCTATCTCCTTG
TGTTGCCTGGTGGCTTTGGCCAGTCAGTAAGTGGGGCCTGAGATCTTCACTGAAATTGGC
ACAATAGCAGCTGGCATCCAGACCTGGCTCTTTCAGCATTACTGCGGGGACAGATCCAAGG
CCTGACTCCTCTTGCCATTTCTGTCACTTCTGCTCCCAAGTTTGCAGTGGTCTTCAACCCCA
TCCAGTTATCTAGTCTCACCAGGGGTGAGGCCGTAGCTGTTACTCTGAACAGCTGGCCTAT
CTGAGTCTGAGCAGCGCGAGCAGTTGCATGGGCCCCAACGAAGGAAGGAGATCCAGA
GCAGCTGGGTGAAACTCAGCCTGGGGTCTCTACGACTGGTTCCAAGCCTCCTGGGCCCTGG
CATTGCCCCGTGAGCATTTTTGGCCACCTATTA-3'.

[0156]

[0157]

뮤린 STRC 코딩 서열은 도 4a-4d의 구축물에서 찾아볼 수 있고, 뮤린 STRC 코딩 서열의 부분은 도 8a-8b 및 도 13a-13b의 구축물에서 찾아볼 수 있다.

[0158] 뮤린 STRC 단백질을 코딩하는 뮤린 mRNA 서열 (NCBI RefSeq: NM_080459; 서열식별번호: 32)은 하기와 같다:

5'-GCCCCGTCTTCACCTGGCTATCCCTTCATGGTGAGCATAGCCAGAACTCACCTCTAGGC
CCAGTGTGCACCTGGAAATATGGCTCTGAGCCTCCAGCCCCAGCTGCTCCTTCTCCTGTCGC
TCCTGCCGCAGGAAGTGACTTCAGCCCCTACTGGGCCTCAGTCTTTGGATGCTGGTCTCTCC
CTTCTGAAGTCATTCTGAGCCACTCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTCACG
GTTCTCTGCGTTCCCTGGCCAACATTTCTTCATCCTTCCAGCTTGGGAGGATGGGGGAGGGAC
CGGTGGGAGAGCCCCCACCTCTCCAGCCCCCTGCACCTTCGACTTCATGATTTCTCGTGACA
CTGAGAGGTAGCCCAGACTGGGAGCCAATGCTAGGGCTTCTGGGAGATGTGCTGGCACTCCT
GGGACAGGAACAGACTCCCCGGGACTTTTTGGTGCACCAGGCAGGTGTACTGGGTGGACTTG
TAGAGGCATTGTTGGGAGCGTTAGTTCTTGAGGCCCCCTGCCCCACTCGACCCCCATGC
ACCCGTGATGGCCCTTCTGACTGTGTCTGGCTGCTGATTGGTTGCCTTCTCTGATGTTGTT
ATTAGAGGGTACACGCTGGCAGGCCCTGGTGCAGTTGCAGCCCAGTGTGGACCCAACCAATG
CCACAGGTCTTGATGGTAGAGAGCCAGCTCCTCACTTTTTACAGGGTCTGCTGGGCTTGCTT
ACCCCAGCAGGAGAGTTGGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGTGGG
GGCCCCCTCTATGCTGCCTTCCAGGAGGGGCTACTGCGAGTCACTCATTCTCTGCAAGATG
AGGTCTTTTCTATTATGGGACAGCCAGAGCCTGATGCCAGTGGGCAGTGCCAGGGAGGCAAC
CTTCAACAGCTGCTTTTATGGGGCATGCGGAACAACCTTTCTTGGGACGCCCAGCACTGGG
TTTTCTATCTGGATCACACCTCCACCCCCTGCTCTCCTGCACTGCCTGAGCAGAGGTGTGC
CTCTGCCCAGGGCTTCCAGCCTGCGGCTCACATCAGCCCTCGACAGCGGCGAGCCATCTCT
GTGGAGGCCCTCTGCGAGAACCCTCAGGCCCAGAGCCACCTACAGCATCTCCAACCTTCTC
CATCTACTTGCTCTGCCAGCACATCAAGCCTGCCACCCCGCGGCCCCCTCCTACCACCCAC
GGCCTCCTCTACCACCCACAGCCCCCTCCTACCCTACACAGCCCATTCTGACACTACA

[0159]

CAGCCCCCTCCTGTCACCCCAAGGCCTCCTCCTACCACCCCAACCCCTCCTAGCACAGC
 TGTCACTGTCCAGACAGCTGTATGGTACGCAGTCTCGTGGGCACCAGGTGCCCAGGTTGGC
 TCCAAGCCTGCCATGATCAGTTTCTGTATCAATTTCTGGATATGATCTGCGGCAACCTCTCA
 TTTTCAGCCCTGTCTGGCCCCAGTCGTCCTTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCC
 ACCCCCCACTAGCTGTCCACCAGGCCTGATCCCTGTGCCCCCACCAGAAATATTCTGGG
 GCTGTTTCTGGAGAATGAGACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCAG
 GCTGTGCCCCCGAGGAACCAGGCTTGGGTTTTCAGCATGTGTGTGCGGGCCCCACCTTGACGC
 CACTGATTTTCCACCGTGCCGCGTTGGACCCTGTGGGGAACGCTGCCAGATGGGGGCAGCT
 TCCTGCTCATGGTCTGTGCCAATGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTGG
 CTAGCAGGCCAATGCAGAATTAGTCGTGGAGGAAATGATACTTGCTTTCTAGAAGGCATGCT
 GGGCCCCCTTGTGGCCCTCTCTGCCCCCTCTGGGACCATCCCCACTCTGTCTGGCTCCTGGTC
 CTTTTCTGCTTGGCATGTTATCCAGTTGCCACGCTGTGAGTCCCGTGCCAGCCCTCGCC
 CACCCACGCGCCTACATTACCTCCTGCGCCTACTGACCTTCCTTCTGGGTCCAGGGACTGG
 GGGTGCCGAGACGCAGGGGATGTTAGGTCAAGCCCTGCTGCTCTCTAGTCTCCAGACAAC
 GTTCATTCTGGGATGCCTTCCGCCCAGAGGGCCGAGAAGTGTACTGAGGACAGTCGGAGAG
 TACTTGACGCGGGAAGAGCCAACCCACCAGGCTTAGACTCCTCCCTCAGCCTCGGCTCTGG
 TATGAGCAAGATGGAGCTTCTGTCTGCTTCAGTCTGTACTGTGGGATCTACTCCAGAGAG
 AGAAGAGCGTTTGGGCCCTGAGGACCCTGGTGAAGGCCTACCTGCGCATGCCCTCCAGAAGAC
 CTTGAGCAGCTTGTGCTTTGAGCAGAGATGGAGGCTGCACAGGGCTTCCTGACGCTCATGCT
 TCGTTCCTGGGCTAAGCTGAAGGTTCAACCATCCGAGGAGCAGGCCATGGGCCGCTGACAG
 CCTTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAACTCTTTATCGACATGTCACCGCTC
 ATCCCCCTCCTGGCTGTCCCTGACCTCATGCGCTTCCACCGTCCCTTTTGGCCAACGACAG
 TGTCTGGCTGCCATCAGGGATCACAGCTCAGGAATGAAGCCTGAACAGAAGGAGGCCCTGG
 CAAAACGACTGCTGGCCCCGAGCTGTTTGGAGAAGTGCTGATTGGCCCCAGGAGCTGCTG
 TGGGCAGCCCTGCCCTCTGCTTCCCCATCTGCCTCTGGAGAGCTTCTCCAGCTCAGCCCTCA
 CCAGATCCAGGCCCTGGAGGATAGCTGGCCAGTAGCAGATCTTGGGCCGGGACACGCCCGAC
 ATGTGCTTCGTAGCCTAGTAAACCAGAGCATGGAGGATGGGGAGGAGCAGGTGCTCAGGCTT
 GGGTCCCTCGCCTGTTTCTGAGTCTGAGGAGCTACAGAGTCTGGTGGCCTTGAGTGATCC
 AATGGGGCCTGTAGAACAGGGTCTGCTGGAATGTGCGGCCAATGGGACCTCAGCCCAGAAG
 GACGGGTGGCATATGAACTTCTGGGAGTGTTGCGTTCATCTGGAGGAACTGTCTTAAGCCCC
 CGAGAGCTGAGGGTCTGGGCACCTCTCTTTCCCCAGCTGGGCCTCCGCTTCTGCAGGAGCT
 CTCAGAGACCCAGCTTAGAGCCATGCTTCTGCCCCTACAGGGAGCCAGTGTACACCTGCCC
 AGGCTGTTCTGTTGTTTGAAGGCTCCTTCTAAGCATGATCTGTCCCTGGAGGAACTCTGC
 TCCCTGCACCCCTCTCCTGCCAGGTCTCAGCCCCCAGACACTCCAGGCCATCCCTAAGAGAGT

[0160]

TCTGGTTGGTGCTTGTTCCTGCCTGGGCCCTGAACTGTCAAGGCTTTCAGCTTGCCAGATTG
CAGCTCTGCTGCAGACCTTTCGGGTAAAAGATGGTGTAAAAATATGGGTGCAGCAGGTGCC
GGCTCAGCCGTGTGCATTCTTGGGCAGCCACCCTTGGCCAGACTGCCTGCTTCCCCTGCT
CCCATTAAAGCTGTACAGCTGGACGCTGCAGCTCTTCTGGCAAACCGAAGACTCTATCGGC
AGCTGCCTTGGTCTGAGCAACAGGCACAGTTTCTCTGGAAGAAAATGCAAGTGCCTACCAAC
CTGAGCCTGAGGAATCTGCAGGCTCTGGGCAACTTGGCAGGAGGCATGACCTGCGAGTTTCT
GCAGCAGATCAGCTCAATGGTTGACTTTCTTGATGTGGTACACATGCTCTACCAGCTGCCCA
CTGGTGTTCGAGAGAGCCTGCGGGCTGTATCTGGACAGAGCTACAGCGGAGGATGACAATG
CCAGAGCCAGAGCTGACCACCCTAGGGCCAGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT
GGACTTGCCGATCCAGCTGATGGACAGATTGTCCAATGATTCCATTATGTTGGTGGTGGAGA
TGGTCCAAGGCGCTCCAGAGCAGCTGCTGGCACTGACCCCACTCCACCAGACAGCCTTGGCA
GAGCGAGCACTTAAAACTGGCTCCAAAGGAGACCCCAATCTCCAAAGAAGTGTGGAGAC
ACTGGGCCCTTGGTTGGATTCTTGGGAATAGAGAGCACGCGACGGATCCCTTTACCCATTCT
TACTGTCTCATCTCAGTCAGCTGCAGGGCTTCTGCCTAGGAGAGACATTTGCCACAGAGCTG
GGATGGCTGCTGTTGCAGGAGCCTGTCTTGGAAAACCAGAATTGTGGAGCCAGGATGAAAT
AGAGCAAGCTGGACGCCCTAGTATTCACTCTGTCTGCTGAGGCTATTTCTCGATCCCCAGGG
AGGCTTTGGGCCAGAGACACTGGAGAGGCTTCTGGGAAAGCATCAAAGCTGGGAGCAGAGC
AGAGTGGGCCATCTGTGTGGGGAGTCACAGCTTGCCCAAGAAAGCAGCTCTGGTAGCTGG
GATTGTGCATCCAGCTGCTGAGGGTCTCCAAGAGCCTGTACCAACTGTGCAGACATACGGG
GAACCTTCCCAGCGGCTGGTCTGCGACACAAATCTCAGAGATGGAAGTCTCAGACTTTGAA
GACTGCCTGTCACTATTTGCTGGAGATCCAGGACTTGGTCTGAGGAACTACGGGCAGCCAT
GGGCAAGGCCAAGCAGTTGTGGGGTCCCCCTCGAGGATTCCTGCTGAGCAGATCTTGCAGC
TGGGCCGTCTCTGATAGGTCTAGGAGAACGGGAACTGCAGGAGCTTACCTTGGTGGACTGG
GGTGTGCTGAGCAGCCTGGGGCAAATAGATGGCTGGAGTTCCATGCAGCTCCGAGCCGTGGT
CTCCAGTTTCCTAAGGCAGAGTGGTCGGCATGTGAGCCACCTGGACTTCATTTATCTGACAG
CACTGGGTTACACAGTCTGTGGATTGCGACCAGAGGAGTTACAGCACATCAGCAGTTGGGAG
TTTAGCCAAGCAGCTCTCTTCTGGGTAGCTTGCATCTCCCGTCTCTGAGGAACAGCTGGA
AGTTCTGGCCTATCTCCTTGTGTGCTGGTGGCTTTGGCCCAGTCAGTAAGTGGGGGCCTG
AGATCTTCACTGAAATTGGCACAATAGCAGCTGGCATCCCAGACCTGGCTCTTTCAGCATTA
CTGCGGGGACAGATCCAAGGCCTGACTCCTCTTGCCATTTCTGTCAATCCTGCTCCCAAGTT
TGCAGTGGTCTTCAACCCCATCCAGTTATCTAGTCTCACCAGGGGTGAGCCGTAGCTGTTA
CTCCTGAACAGCTGGCCTATCTGAGTCTGAGCAGCGCGAGCAGTTGCATGGGCCCCAACAC
GAAGGGAAGGAGATCCCAGAGCAGCTGGGTGAAACTCAGCCTGGGGTCTCTACGACTGGTT
CCAAGCCTCCTGGGCCCTGGCATTGCCCCGTAGCATTTTTGGCCACCTATTATGATAACTGT

[0161]

TCCTTCAGTTGAGGGAGAAAATTTACATCATACTGAACAACTTGTAATGGAAGTGCATACT
AATTATTCTCAGTAAGTGGATGAGGATTGTGGGTAAAATTCGAATGCATTCCACCCACCTGA
GAAGTGTGCTCCTGGCATATACGCCTCTTGCCATCATGAATAACCTCACTGTTTCTTTTCAT
TTCCTACTTCTCCTCACCACACCAATAGAAATAACACAGACAGCTGCAACATAAT-3'.

[0162]

[0163]

신호 펩티드 서열 (아미노산 1-22에서; 밑줄표시)을 포함하고, 링커 서열 또는 Myc 태그 서열을 포함하지 않는
뮤린 STRC 단백질 (1,809개 아미노산; 서열식별번호: 26) (여기서 22개-아미노산 신호 펩티드 서열의 절단은 예
측 분자 질량이 194 kD인 1,787개 아미노산을 갖는 단백질을 남김)은 하기 서열을 가지며, 여기서 Ser746 및
Cys747 및 Ala969 및 Cys970 스플라이스 부위는 볼드체, 밑줄표시된 문자이다:

MALSLQPQLLLLLLSLLPQEVTSAPTGPQSLDAGLSLLKSFVATLDQAPQRSLSQSRFSAFLA
 NISSSFQLGRMGEGPVGEPPLQPPALRLHDFLVTLRGSPDWEPMGLGLGDVLALLGQEQT
 RDFLVHQAGVLGGLVEALLGALVPGPPAPTRPPCTRDGPSDCVLAADWLPSLMLLLEGTRW
 QALVQLQPSVDPTNATGLDGREPAPHFLQGLLGLLTPAGELGSEEALWGGLLRTVGAPLYAA
 FQEGLLRVTHSLQDEVFSIMGQPEPDASGQCQGGNLQQLLLWGMRRNLSWDARALGFLSGSP
 PPPALLHCLSRGVPLPRASQPAAHISPRQRRASVEALCENHSGPEPPYSISNFSIYLLCQ
 HIKPATPRPPPTTPRPPPTTPQPPPTTQPIPDTTQPPPVTTPRPPPTTPQPPPSTAVICQTA
 VWYAVSWAPGARGWLQACHDQFPDQFLDMICGNLSFSALSGPSRPLVKQLCAGLLPPPTSCP
 PGLIPVPLTPEIFWGCFLNETLWAEERLCVEDSLQAVPPRNQAWVQHVCRGPTLDATDFPPC
 RVGPCGERCPDGGSFLLMVCANDTLYEALVPFWAWLAGQCRISRGGNDTCFLEGMLGPLLPS
 LPPLGSPSLCLAPGPFLLGMLSQLPQCSSVPALAHPTRLHYLLRLLTFLGPGTGGAETQG
 MLGQALLSSLPDNCSFWDAFRPEGRRSVLRTVGEYLQREEPTPPGLDSSLSLGSMSKME
 LSCFSPVLWDLQREKSVWALRTLKVAYLRMPPELQQLVLSAEMEAQGFLLMLRSWAKL
 KVQPSEEQAMGRLTALLQRYPRLTSQLFIDMSPLIPFLAVPDLMRFPSSLANDSVLAAIR
 DHSSGMKEQKEALAKRLLAPELFGVDPWPQELLWAALPLPHLPLESFLQLSPHQIQALE
 DSWPVADLGPGRHARHVLRSVLNQSMEDGEQVLRRLGSLACFLSPEELQSLVPLSDPMGPVEQ
 GLLECAANGTLSPGGRVAYELLGVLRRSSGGTVLSPRELVRWAPLFPQLGLRFLQELSETQLR
 AMLPALQGASVTPAQAVLLFGRLLPKHDLSEELCSLHPLLPGLSPQTLQAIKRVLVGACS
 CLGPELSRLSACQIAALLQTFRVKDGVKNMGAAGAGSAVCIPGQPTTWPDCLLPLPLKLLQ
 LDAAALLANRRLYRQLPWSEQQAQFLWKKMQVPTNLSLRNLQALGNLAGGMTCEFLQQISSM
 VDFLDVHMLYQLPTGVRESLRACIWTQLRRMTPEPELTTLGPELSELDTKLLLDLPIQL
 MDRLSNDSIMLVVEMVQGAPEQLLALTPLHQTALAEALKNLAPKETPISKEVLETGLPLVG
 FLGIESTRRIPLPILLSHLSQLQGFCLETGFWLLQEPVLGKPELWSQDEIEQAGRL
 VFTLSAEAISSIPREALGPETLERLLGKHQSWEQSRVGHLCGESQLAHKKAALVAGIVHPAA

[0164]

EGLQEPVPCADIRGTFFPAAWSATQISEMELSDFEDCLSLFAGDPGLGPEELRAAMGKAKQL
 WGPPRGFRPEQILQLGRLLIGLGERELQELTLVDWGVLSLQIDGWSSMQLRAVVSSFLRQ
 SGRHVSHLDFIYLTALGYTVCGLRPEELQHISSEWFSQAALFLGSLHLPCEEQLEVLAYLL
 VLPGGFGFVSNWGPETFEIGTIAAGIPDLALSALLRGQIQGLTPLAISVIPAPKFAVFN
 IQLSSLTRGQAVAVTPEQLAYLSPEQRRAVAWAQHEGKEIPEQLGRNSAWGLYDWFQASWAL
 ALPVSI FGHLL.

[0165]

[0166] "스테레오실린 (STRC) 단백질"은, 예를 들어 NCBI 수탁 번호 NP_714544 또는 NP_536707; 진뱅크 번호 AAL35321에 대해 적어도 약 80% 아미노산 동일성 (예를 들어, 82%, 85%, 88%, 90%, 95%, 97%, 98%, 99%, 100%)을 갖는 폴리펩티드 또는 그의 단편을 의미한다.

[0167] "검출하다"는 검출될 분석물의 존재, 부재 또는 양을 확인하는 것을 지칭한다.

[0168] "검출가능한 표지"는 관심 분자에 연결되는 경우에 분광학적, 광화학적, 생화학적, 면역화학적 또는 화학적 수단을 통해 후자의 분자를 검출가능하게 하는 조성물을 의미한다. 예를 들어, 유용한 표지는 방사성 동위원소, 자기 비드, 금속성 비드, 콜로이드성 입자, 형광 염료, 예를 들어 녹색 형광 단백질, 전자-고밀도 시약, 효소 (예를 들어, ELISA에 통상적으로 사용되는 바와 같음), 비오틴, 디코시게닌 또는 합텐을 포함한다.

[0169] "질환"은 세포, 조직 또는 기관의 정상적인 기능을 손상시키거나 방해하는 임의의 상태 또는 장애를 의미한다. 질환의 예는 임의의 병리상태, 예컨대 열성 돌연변이와 연관된 청력 장애, 예를 들어 DFNB16을 포함하나 이에 제한되지는 않는 청력 장애를 포함한다.

[0170] "유효량"은 비치료되지 않은 환자에 비해 질환의 증상을 호전시키는 데 요구되는 작용제의 양을 의미한다. 질환의 치유적 치료를 실시하는 데 사용되는 활성 화합물(들)의 유효량은 투여 방식, 대상체의 연령, 체중 및 전반적 건강에 따라 달라진다. 궁극적으로, 담당 의사 또는 수의사는 적절한 양 및 투여 요법을 결정할 것이다. 이러한 양은 "유효량"으로 지칭된다.

- [0171] "단편"은 폴리펩티드 또는 핵산 서열 또는 분자의 부분을 의미한다. 이러한 부분은 참조 핵산 분자 또는 폴리펩티드의 전체 길이의 적어도 10% 이상 (예를 들어, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%)을 함유할 수 있다. 단편은 10개 이상 (예를 들어, 20, 30, 40, 50, 60, 70, 80, 90 또는 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500개)의 뉴클레오티드 또는 아미노산을 함유할 수 있다.
- [0172] "혼성화"는 상보적 핵염기 사이의 왓슨-크릭, 후그스틴 또는 역 후그스틴 수소 결합일 수 있는 수소 결합을 의미한다. 예를 들어, 아데닌 및 티민은 수소 결합의 형성을 통해 쌍형성하는 상보적 핵염기이다.
- [0173] "동일성"은 관심 서열과 참조 서열 사이의 아미노산 또는 핵산 서열 동일성을 의미한다. "실질적으로 동일한"은 참조 아미노산 서열 (예를 들어, 본원에 기재된 아미노산 서열 중 어느 하나) 또는 핵산 서열 (예를 들어, 본원에 기재된 핵산 서열 중 어느 하나)에 대해 적어도 50% 동일성을 나타내는 폴리펩티드 또는 핵산 분자를 의미한다. 이러한 서열은 비교를 위해 사용된 서열 또는 참조 서열에 대해 아미노산 수준 또는 핵산 수준에서 적어도 60% 이상 (예를 들어, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 100%)의 동일성을 가질 수 있다.
- [0174] 서열 동일성은 서열 분석 소프트웨어 (예를 들어, 제네틱스 컴퓨터 그룹 (53705 위스콘신주 매디슨 유니버시티 에비뉴 1710 소재 유니버시티 오브 위스콘신 바이오테크놀로지 센터)의 서열 분석 소프트웨어 패키지, BLAST, BESTFIT, GAP 또는 PILEUP/PRETTYBOX 프로그램)를 사용하여 측정될 수 있다. 이러한 소프트웨어는 다양한 치환, 결실 및/또는 다른 변형에 상동성 정도를 매칭함으로써 동일하거나 유사한 서열을 매칭시킨다. 보존적 치환은 전형적으로 하기 군 내에서의 치환을 포함한다: 글리신, 알라닌; 발린, 이소류신, 류신; 아스파르트산, 글루탐산, 아스파라긴, 글루타민; 세린, 트레오닌; 리신, 아르기닌; 및 페닐알라닌, 티로신. 동일성 정도를 결정하기 위한 예시적인 접근법에서, BLAST 프로그램이 사용될 수 있으며, 여기서 e^{-3} 내지 e^{-100} 의 확률 점수는 밀접하게 관련된 서열을 나타낸다.
- [0175] 용어 "단리된", "정제된" 또는 "생물학적으로 순수한"은 그의 천연 상태에서 발견되는 바와 같은 정상적으로 동반되는 성분이 다양한 정도로 없는 물질을 지칭한다. "단리하다"는 원래의 공급원 또는 주변으로부터의 분리 정도를 나타낸다. "정제하다"는 단리보다 더 높은 분리 정도를 나타낸다. "정제된" 또는 "생물학적으로 순수한" 단백질은 임의의 불순물이 단백질의 생물학적 특성에 실질적으로 영향을 미치거나 다른 유해 결과를 유발하지 않도록 다른 물질이 충분히 없다. 즉, 본 개시내용의 핵산 또는 펩티드는 재조합 DNA 기술에 의해 생산되는 경우에 정제되거나 (예를 들어, 세포 물질, 바이러스 물질, 배양 배지가 실질적으로 없음) 또는 화학적으로 합성되는 경우에 화학적 전구체 또는 다른 화학물질이 정제된다. 순도 및 균질성은 전형적으로 분석 화학 기술, 예를 들어 폴리아크릴아미드 겔 전기영동 또는 고성능 액체 크로마토그래피를 사용하여 결정된다. 용어 "정제된"은 핵산 또는 단백질이 전기영동 겔에서 본질적으로 1개의 밴드를 생성함을 나타낼 수 있다. 변형, 예를 들어 인산화 또는 글리코실화에 적용될 수 있는 단백질의 경우, 상이한 변형은 상이한 단리된 단백질을 생성할 수 있으며, 이는 개별적으로 정제될 수 있다.
- [0176] "단리된 폴리뉴클레오티드"는 본 발명의 핵산 분자가 유래된 유기체의 자연-발생 게놈에서 유전자에 플랭킹된 유전자들이 없는 본 개시내용의 핵산 (예를 들어, DNA)을 의미한다. 따라서, 상기 용어는, 예를 들어 벡터 내로; 자율 복제 플라스미드 또는 바이러스 내로; 또는 원핵생물 또는 진핵생물의 게놈 DNA 내로 혼입되거나; 또는 다른 서열과 독립적으로 별개의 분자 (예를 들어, PCR 또는 제한 엔도뉴클레아제 소화에 의해 생산된 cDNA 또는 게놈 또는 cDNA 단편)로서 존재하는 재조합 DNA를 포함한다. 추가로, 상기 용어는 DNA 분자로부터 전사되는 RNA 분자, 뿐만 아니라 추가의 폴리펩티드 서열을 코딩하는 하이브리드 유전자의 일부인 재조합 DNA를 포함한다.
- [0177] "단리된 폴리펩티드"는 자연적으로 동반되는 성분들로부터 분리된 본 개시내용의 폴리펩티드를 의미한다. 전형적으로, 폴리펩티드는 자연적으로 회합되는 단백질 및 자연 발생 유기 분자가 적어도 60 중량% 이상 (예를 들어, 75%, 80%, 90%, 95%, 99 중량%) 없는 경우에 단리된 것이다. 본 개시내용의 단리된 폴리펩티드는, 예를 들어 천연 공급원으로부터의 추출에 의해, 이러한 폴리펩티드를 코딩하는 재조합 핵산의 발현에 의해; 또는 단백질을 화학적으로 합성함으로써 수득될 수 있다. 순도는 임의의 적절한 방법, 예를 들어 칼럼 크로마토그래피, 폴리아크릴아미드 겔 전기영동 또는 HPLC 분석에 의해 측정될 수 있다.
- [0178] "마커"는 질환 또는 장애와 연관된 발현 수준 또는 활성의 변경을 갖는 임의의 단백질 또는 폴리뉴클레오티드를 의미한다.
- [0179] "기계적감각"은 기계적 자극에 대한 반응을 의미한다. 기계적 자극의 뉴런 신호로의 전환 예의 촉각, 청력 및 균형. 기계적감각 입력은 "기계적전달"이라 불리는 과정을 통해 기계적 자극에 대한 반응으로 전환된다.

- [0180] 본원에 사용된 "작용제를 수득하는"에서와 같은 "수득하는"은 작용제를 합성하거나, 구입하거나 또는 달리 획득하는 것을 포함한다.
- [0181] 본원에 사용된 용어 "예방하다", "예방하는", "예방", "예방적 치료" 등은 장애 또는 상태를 갖지 않지만, 그의 발생 위험이 있거나 또는 그에 걸리기 쉬운 대상체에서 장애 또는 상태가 발생할 확률을 감소시키는 것을 지칭한다.
- [0182] "프로모터"는 하류 폴리뉴클레오타이드의 전사를 지시하기에 충분한 폴리뉴클레오타이드를 의미한다. 일부 실시양태에서, 본원에 기재된 폴리뉴클레오타이드는 1개 이상의 조절 요소를 포함할 수 있다. 관련 기술분야의 통상의 기술자는 세포, 예를 들어 포유동물 또는 인간 숙주 세포에서 사용하기에 적절한 조절 요소를 선택할 수 있다. 조절 요소의 비제한적 예는 프로모터, 전사 종결 서열, 번역 종결 서열, 인핸서 및 폴리아데닐화 요소를 포함한다. 본원에 기재된 폴리뉴클레오타이드는 목적하는 폴리펩티드, 예컨대 스테레오실린을 코딩하는 뉴클레오타이드 서열에 작동가능하게 연결된 프로모터 서열을 포함할 수 있다. 본 발명에서 사용하기 위해 고려되는 프로모터는 시토메갈로바이러스 (CMV) 프로모터, SV40 프로모터, 라우스 육종 바이러스 (RSV) 프로모터, 키메라 CMV/닭 β 액틴 프로모터 (CBA) 및 말단절단된 형태의 CBA (smCBA)를 포함하나 이에 제한되지는 않는다. 일부 실시양태에서, 프로모터는 CMV 프로모터이다.
- [0183] 어구 "제약상 허용되는 부형제"는 대상 화합물을 신체의 한 기관 또는 부분으로부터 신체의 또 다른 기관 또는 부분으로 운반 또는 수송하는 데 수반되는 제약상 허용되는 물질, 조성물 또는 비히클, 예컨대 액체 또는 고체 충전제, 희석제, 담체, 용매 또는 캡슐화 물질을 포함할 수 있다. 각각의 담체는 제제의 다른 성분과 상용성이 고 환자에게 유해하지 않다는 의미에서 "허용되는" 것이어야 한다. 제약상 허용되는 담체로서의 역할을 할 수 있는 물질의 일부 예는 다음을 포함한다: (1) 당, 예컨대 락토스, 글루코스 및 수크로스; (2) 전분, 예컨대 옥수수 전분 및 감자 전분; (3) 셀룰로스 및 그의 유도체, 예컨대 소듐 카르복시메틸 셀룰로스, 에틸 셀룰로스 및 셀룰로스 아세테이트; (4) 분말화 트라가칸트; (5) 맥아; (6) 젤라틴; (7) 활석; (8) 부형제, 예컨대 코코아 버터 및 좌제 왁스; (9) 오일, 예컨대 땅콩 오일, 목화씨 오일, 홍화 오일, 참깨 오일, 올리브 오일, 옥수수 오일 및 대두 오일; (10) 글리콜, 예컨대 프로필렌 글리콜; (11) 폴리올, 예컨대 글리세린, 소르비톨, 만니톨 및 폴리에틸렌 글리콜; (12) 에스테르, 예컨대 에틸 올레아이트 및 에틸 라우레이트; (13) 한천; (14) 완충제, 예컨대 수산화마그네슘 및 수산화알루미늄; (15) 알긴산; (16) 발열원-무함유 물; (17) 등장성 염수; (18) 링거액; (19) 에틸 알콜; (20) 포스페이트 완충제 용액; 및 (21) 제약 제제에 사용되는 다른 비-독성 상용성 물질.
- [0184] 추가의 적합한 담체 및 그의 제제는, 예를 들어 문헌 [Remington's Pharmaceutical Sciences by E. W. Martin]의 가장 최신판에 기재되어 있다. 투여될 치료제의 양은 투여 방식 및 모드, 연령 및 질환 상태 (예를 들어, 치료 전에 존재하는 청력 상실의 정도)에 따라 달라진다.
- [0185] "스테레오실린 프로모터"는 유모 다발 내의 인접한 부동섬모의 정단 영역을 연결하는 수평 상단 연결자 및 가장 높은 부동섬모를 위에 놓인 덮개막 (TM)에 부착시키는 연결관과 회합되어 있는, 성숙 와우 내 내유모 세포 (IHC) 또는 외유모 세포 (OHC)에서의 하류 폴리뉴클레오타이드의 발현을 지시하기에 충분한 NCBI 참조 서열: NG_011636.1로부터 유래된 조절 폴리뉴클레오타이드 서열을 의미한다. 스테레오실린은 또한 전정 유모 세포 및 미성숙 OHC의 운동섬모 주위에서 발현될 수 있다. 본 개시내용의 한 실시양태는 스테레오실린 코딩 서열의 상류의 적어도 350개 이상 (예를 들어, 500, 1000, 2000, 3000, 4000, 5000개)의 염기 쌍을 포함하거나 또는 그로 이루어진 스테레오실린 프로모터를 제공한다.
- [0186] "감소시키다"는 적어도 5% 이상 (예를 들어, 10%, 15%, 20%, 25%, 50%, 75%, 100%)의 음성 변경을 의미한다.
- [0187] "참조"는 표준 또는 대조군 조건을 의미한다.
- [0188] "참조 서열"은 서열 비교를 위한 기초로서 정의되고 사용될 수 있는 서열이다. 참조 서열은 특정한 서열의 부분 또는 전체, 예를 들어 전장 cDNA 또는 유전자 서열의 단편, 또는 완전한 cDNA 또는 유전자 서열일 수 있다. 참조 폴리펩티드 서열의 길이는 폴리펩티드에 대해 적어도 10개 아미노산 이상 (예를 들어, 15, 20, 25, 30, 35, 50, 100개) 또는 그 부근 또는 그 사이의 임의의 정수일 수 있다. 참조 핵산 서열의 길이는 핵산에 대해 적어도 50개 뉴클레오타이드 이상 (예를 들어, 55, 60, 75, 90, 100, 200, 300개) 또는 그 부근 또는 그 사이의 임의의 정수일 수 있다.
- [0189] "특이적으로 결합한다"는 본 개시내용의 폴리펩티드를 인식하고 그에 결합하지만, 자연적으로 본 개시내용의 폴리펩티드를 포함하는 샘플, 예를 들어 생물학적 샘플 내의 다른 분자를 실질적으로 인식하고 그에 결합하지 않는 화합물 또는 항체를 의미한다.

- [0190] 본 개시내용의 방법에 유용한 핵산 분자는 본 개시내용의 폴리펩티드 또는 그의 단편 또는 부분을 코딩하는 임의의 핵산 분자를 포함한다. 이러한 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 전형적으로 실질적인 동일성을 나타낼 것이다. 내인성 서열에 대해 "실질적 동일성"을 갖는 폴리뉴클레오티드는 전형적으로 이중-가닥 핵산 분자의 적어도 1개의 가닥과 혼성화할 수 있다. 본 발명의 방법에 유용한 핵산 분자는 본 발명의 폴리펩티드 또는 그의 단편을 코딩하는 임의의 핵산 분자를 포함한다. 이러한 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 전형적으로 실질적인 동일성을 나타낼 것이다. 내인성 서열에 대해 "실질적 동일성"을 갖는 폴리뉴클레오티드는 전형적으로 이중-가닥 핵산 분자의 적어도 1개의 가닥과 혼성화할 수 있다. "혼성화하다"는 다양한 엄격도 조건 하에 상보적 폴리뉴클레오티드 서열 (예를 들어, 본원에 기재된 유전자) 또는 그의 부분 사이에 이중-가닥 분자를 형성하기 위해 쌍을 형성하는 것을 의미한다. (예를 들어, 문헌 [Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507] 참조).
- [0191] 혼성화는, 예를 들어 통상적으로 750 mM 미만의 NaCl (예를 들어, 500 mM; 250 mM) 및 75 mM 미만의 시트르산삼나트륨 (예를 들어, 50 mM; 25 mM)일 수 있는 엄격한 염 농도 하에 이루어질 수 있다. 낮은 엄격도 혼성화는 유기 용매, 예를 들어 포름아미드의 부재 하에 수득될 수 있는 반면, 높은 엄격도 혼성화는 적어도 35% 포름아미드 (예를 들어, 50% 포름아미드)의 존재 하에 수득될 수 있다. 엄격한 온도 조건은 통상적으로 적어도 30°C (예를 들어, 37°C, 42°C)의 온도를 포함할 수 있다. 추가의 파라미터, 예컨대 혼성화 시간, 세제, 예를 들어 소듐 도데실 술페이트 (SDS)의 농도 및 캐리어 DNA의 포함 또는 배제를 변동시키는 것은 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다. 다양한 수준의 엄격도는 필요에 따라 이들 다양한 조건을 조합함으로써 달성될 수 있다. 한 실시양태에서, 혼성화는 30°C에서 750 mM NaCl, 75 mM 시트르산삼나트륨 및 1% SDS 중에서 이루어질 수 있다. 또 다른 실시양태에서, 혼성화는 37°C에서 500 mM NaCl, 50 mM 시트르산삼나트륨, 1% SDS, 35% 포름아미드 및 100 µg/ml 변성 연어 정자 DNA (ssDNA) 중에서 이루어질 수 있다. 추가 실시양태에서, 혼성화는 42°C에서 250 mM NaCl, 25 mM 시트르산삼나트륨, 1% SDS, 50% 포름아미드 및 200 µg/ml ssDNA 중에서 이루어질 수 있다. 이들 조건에 대한 유용한 변동은 관련 기술분야의 통상의 기술자에게 자명할 것이다.
- [0192] 대부분의 적용에 있어서, 혼성화에 이어지는 세척 단계는 또한 엄격도가 달라질 수 있다. 세척 엄격도 조건은 염 농도 및 온도에 의해 정의될 수 있다. 상기와 같이, 세척 엄격도는 염 농도를 감소시킴으로써 또는 온도를 증가시킴으로써 증가될 수 있다. 예를 들어, 세척 단계를 위한 엄격한 염 농도는 30 mM 미만의 NaCl (예를 들어, 15 mM) 및 3 mM 미만의 시트르산삼나트륨 (예를 들어, 1.5 mM)일 수 있다. 세척 단계에 대한 엄격한 온도 조건은 통상적으로 적어도 25°C (예를 들어, 42°C, 68°C)의 온도를 포함할 수 있다. 또 다른 실시양태에서, 세척 단계는 25°C에서 30 mM NaCl, 3 mM 시트르산삼나트륨 및 0.1% SDS 중에서 이루어질 수 있다. 또 다른 실시양태는 42°C에서 15 mM NaCl, 1.5 mM 시트르산삼나트륨 및 0.1% SDS 중에서 이루어지는 세척 단계를 제공할 수 있다. 추가 실시양태는 68°C에서 15 mM NaCl, 1.5 mM 시트르산삼나트륨 및 0.1% SDS 중에서 이루어지는 세척 단계를 제공할 수 있다. 이들 조건에 대한 추가의 변동은 관련 기술분야의 통상의 기술자에게 자명할 것이다. 혼성화 기술은 관련 기술분야의 통상의 기술자에게 널리 공지되어 있고, 예를 들어 문헌 [Benton and Davis (*Science* 196:180, 1977); Grunstein and Hogness (*Proc. Natl. Acad. Sci., USA* 72:3961, 1975); Ausubel et al. (*Current Protocols in Molecular Biology*, Wiley Interscience, New York, 2001); Berger and Kimmel (*Guide to Molecular Cloning Techniques*, 1987, Academic Press, New York); 및 Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York]에 기재되어 있다.
- [0193] "대상체"는 인간 또는 비-인간 포유동물, 예컨대 소, 말, 개, 양, 고양이 또는 무린을 포함하나 이에 제한되지 않는 포유동물을 의미한다.
- [0194] "STRC 단백질 또는 폴리펩티드"는 NCBI 수탁 번호 NP_714544 또는 진뱅크 번호 AAL35321에 대해 적어도 약 85% 이상의 아미노산 서열 동일성을 갖는 폴리펩티드, 또는 청각 기능에 필수적인 STRC를 발현하기에 충분한 활성을 갖는 그의 단편을 의미한다.
- [0195] 본원에 사용된 용어 "치료하다", "치료하는", "치료" 등은 그와 연관된 일부 장애 및/또는 증상의 감소 또는 호전을 지칭한다. 배제되는 것은 아니지만, 장애 또는 상태를 치료하는 것이 장애, 상태 또는 그와 연관된 증상을 완전히 제거하는 것을 요구하지는 않는다는 것이 인지될 것이다.
- [0196] 구체적으로 언급되거나 문맥으로부터 명백하지 않는 한, 본원에 사용된 용어 "또는"은 포괄적인 것으로 이해된다. 구체적으로 언급되거나 문맥으로부터 명백하지 않는 한, 본원에 사용된 단수 용어는 단수 또는 복수인 것으로 이해된다.

- [0197] 본 개시내용에서, "포함한다", "포함하는", "함유하는" 및 "갖는" 등은 미국 특허법에서 그에 부여된 의미를 가질 수 있으며 "포함한다", "포함하는" 등을 의미할 수 있고; "본질적으로 이루어진" 또는 "본질적으로 이루어진 다"는 마찬가지로 미국 특허법에서 부여된 의미를 가지며 이 용어는 개방형이어서, 인용된 것의 기초 또는 신규 특징이 인용된 것 초과 존재에 의해 변화되지 않는 한, 인용된 것 초과 존재를 허용하지만, 선행 기술 실시양태는 배제한다. "본질적으로 이루어진다"는 것은, 예를 들어 구성성분이 상업적 물질에 존재하는 통상의 불순물과 함께 및 본원에 개시된 실시양태의 작업에 영향을 미치지 않는 수준으로, 예를 들어 5 중량% 미만 또는 1 중량% 미만 또는 심지어 0.5 중량%의 수준으로 존재하는 임의의 다른 첨가제와 함께, 단지 열거된 성분만을 포함하는 것을 의미한다.
- [0198] 본원에 사용된 모든 수치 범위는 종점 및 개시된 값 사이에 개시된 모든 가능한 값을 포함한다. 모든 절반 정수 수치 값의 정확한 값은 또한 구체적으로 개시된 바와 같이 및 개시된 범위의 모든 하위세트에 대한 한계로서 고려된다. 예를 들어, 1 내지 50의 범위는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50으로 이루어진 군으로부터의 임의의 수, 수의 조합 또는 하위-범위를 포함하는 것으로 이해된다. 추가적으로, 0.1% 내지 3%의 범위는 구체적으로, 예를 들어 0.1%, 1%, 1.5%, 2.0%, 2.5% 및 3% 또는 그 사이의 임의의 다른 수치 값의 백분율을 개시한다. 추가적으로, 0.1 내지 3%의 범위는 0.5% 내지 2.5%, 1% 내지 3%, 0.1% 내지 2.5% 등을 포함한 원래 범위의 하위세트를 포함한다. 개별 성분의 모든 중량%의 합계는 100%를 초과하지 않을 것으로 이해될 것이다. 본원에 제공된 범위는 범위 내의 모든 값에 대한 약칭인 것으로 이해된다.
- [0199] 본원의 변수의 임의의 정의에서 화학적 기의 목록의 언급은 임의의 단일 기 또는 열거된 기의 조합으로서의 그 변수의 정의를 포함한다. 본원의 변수 또는 측면에 대한 실시양태의 언급은 임의의 단일 실시양태로서의 또는 임의의 다른 실시양태 또는 그의 부분과 조합된 그 실시양태를 포함한다.
- [0200] 본원에 제공된 임의의 조성물 또는 방법은 본원에 제공된 임의의 다른 조성물 및 방법 중 하나 이상과 조합될 수 있다.
- [0201] 본 개시내용은 시스템 (예를 들어, 벡터, 재조합 바이러스), 분할 관심 단백질, 예컨대 STRC 단백질을 코딩하는 핵산 서열 또는 구축물, 및 본원에 기재된 핵산 서열 또는 구축물을 사용하여 관심 단백질 (예를 들어, STRC)을 숙주, 숙주 세포 또는 세포 내로 전달하는 방법에 관한 것이다. 단백질의 아미노 말단 (N-말단) 단편 및 단백질의 카르복시 말단 (C-말단) 단편을 포함하는 분할 단백질은, 예를 들어 벡터 (예를 들어, 바이러스 벡터, 예컨대 아데노-연관 바이러스 (AAV) 또는 렌티바이러스)를 사용하는 세포 내로의 전달을 위해 상이한 및 별개의 핵산 서열 또는 구축물 상에 각각 코딩된다. 관심 단백질의 N-말단 및 C-말단 폴리펩티드 단편은, 예를 들어 인테인-매개 단백질 스플라이싱을 사용하여 함께 연결되어 전장 관심 단백질을 형성할 수 있으며, 여기서 관심 단백질의 N-말단 및 C-말단 폴리펩티드 단편은 펩티드 결합에 의해 연결된다. 상이한 종들 사이의 분할 부위는 유사하거나 상동인 영역에 위치할 것이다.
- [0202] 벡터 시스템
- [0203] 일부 실시양태는 목적하는 유전자 구축물을 함유하는 적어도 1개의 벡터를 포함하는, 목적하는 전장 단백질의 코딩 서열을 전달하기 위한 벡터 (예를 들어, 캡시드, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 게놈, 렌티바이러스) 시스템을 제공할 수 있다. 도 1은 이러한 예시적인 AAV STRC 구축물의 개략적 표현을 보여준다. 다른 실시양태에서, 벡터 시스템은 전체 인간 STRC 코딩 서열 (서열식별번호: 1)을 포함할 수 있으며, 여기서 5' 말단에서의 뉴클레오티드 서열 (도 2a-2c; 서열식별번호: 33)은 "ATG" 개시 코돈 (볼드체)으로 시작하고 정지 코돈 (대문자, 이탤릭체 및 밑줄표시)으로 끝나며, 신호 펩티드 코딩 서열 (소문자, 이탤릭체 및 밑줄표시; 서열식별번호: 9)은 STRC 코딩 서열의 상류에 위치한다. 임의로, 3' 말단에 링커 서열 (볼드체 및 이탤릭체; 서열식별번호: 34) 및 myc 태그를 코딩하는 서열 (소문자; 서열식별번호: 35)이 포함될 수 있으며, 이는 단백질 단리 (예를 들어, 웨스턴 블롯팅, 면역형광, 면역침전)를 포함하나 이에 제한되지는 않는 후속 연구에 사용될 수 있다. 도 3은 도 2의 인간 STRC 뉴클레오티드 서열에 의해 코딩되는 아미노산 서열 (서열식별번호: 36)을 제시하며, 인간 STRC 단백질 서열은 ATG 코돈에 의해 코딩되는 메티오닌 (M) (볼드체), 신호 펩티드 서열 (소문자, 이탤릭체 및 밑줄표시; 서열식별번호: 10), 임의적인 링커 서열 (볼드체 및 이탤릭체) (N-term-TRTRPL-C-term; 서열식별번호: 37) 및 임의적인 Myc 태그 (소문자; 서열식별번호: 27)를 포함한다.
- [0204] 다른 실시양태에서, 벡터 시스템은 전체 유인 STRC 코딩 서열 (서열식별번호: 3)을 포함할 수 있으며, 여기서

5' 말단에서의 뉴클레오타이드 서열 (도 4a-4d; 서열식별번호: 38)은 "ATG" 개시 코돈 (볼드체)으로 시작하고 정지 코돈 (대문자, 이탤릭체 및 밑줄표시)으로 끝나며, 신호 펩티드 코딩 서열 (소문자, 이탤릭체 및 밑줄표시; 서열식별번호: 11)은 STRC 코딩 서열의 상류에 위치한다. 임의로, 3' 말단에 링커 서열 (볼드체 및 이탤릭체; 서열식별번호: 34) 및 myc 태그를 코딩하는 서열 (소문자; 서열식별번호: 35)이 포함될 수 있으며, 이는 단백질 단리 (예를 들어, 웨스턴 블롯팅, 면역형광, 면역침전)를 포함하나 이에 제한되지는 않는 후속 연구에 사용될 수 있다. 도 5는 도 4의 뮤린 STRC 뉴클레오타이드 서열에 의해 코딩되는 아미노산 서열을 제시하며, 뮤린 STRC 단백질 서열 (서열식별번호: 39)은 ATG 코돈에 의해 코딩되는 메티오닌 (M) (볼드체), 신호 펩티드 서열 (소문자, 이탤릭체 및 밑줄표시), 임의적인 링커 서열 (볼드체 및 이탤릭체) (서열식별번호: 37) 및 임의적인 Myc 태그 (소문자) (서열식별번호: 27)를 포함한다.

[0205] 관련 기술분야의 통상의 기술자는 문헌 [MR Green and J Sambrook, Molecular Cloning: A Laboratory Manual (2012), 4th Ed.; Ausubel et al., Current Protocols in Molecular Biology (2003)] (이들 둘 다는 본원에 참조로 포함됨)에 기재된 표준 재조합 기술을 통해 벡터 (예를 들어, 바이러스 (박테리오파지, 동물 바이러스 및 식물 바이러스, 캡시드), 플라스미드, 코스미드 및 인공 염색체 (예를 들어, YAC) 포함)를 구축하기에 충분히 준비가 갖추어져 있을 것으로 이해된다. 추가적으로, DNA를 세포로 도입 또는 전달하는 방법은, 예를 들어 문헌 [Sung, Y., Kim, S. "Recent advances in the development of gene delivery systems." Biomater Res 23, 8 (2019); Jin, Lian et al. "Current progress in gene delivery technology based on chemical methods and nano-carriers." Theranostics 4(3):240-255, 2014; Nayerossadat N, Maedeh T, Ali PA. "Viral and nonviral delivery systems for gene delivery." Adv Biomed Res. 1:27, 2012; Machida, Curtis A. Viral Vectors for Gene Therapy Methods and Protocols. Humana Press, 2003; Heiser, William C. Gene Delivery to Mammalian Cells. Humana Press, 2004] (이들은 각각 본원에 참조로 포함됨)에 개시되어 있다.

[0206] 인테인-매개 단백질 트랜스-스플라이싱

[0207] 다른 실시양태는 인테인을 사용하여 동일한 목적하는 관심 단백질의 상이한 부분을 각각 전달하기 위한 2개의 벡터를 포함하는 이중-벡터 시스템을 포함하는 벡터 시스템을 제공할 수 있다. 인테인은 단백질 인트론인 것으로 간주될 수 있으며, 여기서 이들은 아미노산 서열로부터 그 자체를 절제하고 남아있는 플랭킹 영역 (엑스테인)을 단백질 스플라이싱으로 공지된 과정에 의해 펩티드 결합과 연결시킬 수 있는 단백질의 일부이다. 예를 들어, 문헌 [Mills et al. "Protein Splicing: How Inteins Escape from Precursor Proteins" JBC. 289(21):14498-14505, 2014]을 참조하며, 이는 인테인-매개 단백질 트랜스-스플라이싱 과정에 대해 그 전문이 본원에 참조로 포함된다. 인테인-매개 단백질 스플라이싱은 인테인을 함유하는 mRNA가 단백질로 번역된 후에 발생한다. 도 6을 참조한다.

[0208] 인테인은 숙주 단백질-인테인 융합체로부터 그 자체를 절제하는 반응을 촉매하여, 성숙 숙주 단백질 (엑스테인) 및 분리된 인테인을 생성하는 효소의 부류이며, 여기서 펩티드 결합은 공여자 인테인과 수용자 인테인 사이의 스플라이스 접합부를 라이게이션한다. 문헌 [Perler, F.B. (InBase. The intein database. Nucleic Acids Res. 30:383-384, 2002)] (이들 모두는 그 전문이 본원에 참조로 포함될 수 있음)에 의해 확인된 것 (예를 들어, 노스톡 폰크티포르메(Nostoc punctiforme) PCC 73102 (ATCC® 29133™)로부터의 Npu-PCC73102 (DnaE-c 인테인 (수탁 번호 ZP_00108882); DnaE-n 인테인 (수탁 번호 ZP_00111398)) (이들 모두는 그 전문이 본원에 참조로 포함될 수 있음)을 포함하나 이에 제한되지는 않는 임의의 공지된 인테인이 본 개시내용의 실시양태에 사용될 수 있다.

[0209] 본 개시내용의 일부 실시양태에서, 시아노박테리아 노스톡 폰크티포르메 (Npu)로부터의 DNA 폴리머라제 III DnaE의 촉매 서브유닛이 본원에 기재된 스플릿-인테인-STRC 이중-벡터 시스템에 사용될 수 있다. 예를 들어, 시아노박테리아 노스톡 폰크티포르메 (Npu)로부터의 DNA 폴리머라제 III dnaE의 알파 서브유닛이 존재하며, 이는 자연적으로 분할되고 2개의 유전자 dnaE-n 및 dnaE-c에 위치한다. dnaE의 N- 및 C-말단 부분은 게놈 내 및 반대편 DNA 가닥 상의 2개의 별개의 유전자에 의해 코딩된다. dnaE-n 코딩되는 단백질은 아미노-말단 (N-말단) dnaE 단편 (예를 들어, N-엑스테인) 및 아미노 말단 인테인 (N-인테인)을 함유하는 반면, dnaE-c는 카르복시-말단 인테인 (C-인테인) 엔티티가 선행하는 카르복시-말단 (C-말단) dnaE 단편 (예를 들어, C-엑스테인)을 함유하는 단백질을 코딩한다. N-인테인 및 C-인테인은 서로 인식하고, 아미노산 서열로부터 그 자체를 스플라이싱하며, 동시에 펩티드 결합을 통해 관심 플랭킹 N- 및 C-말단 엑스테인을 라이게이션 또는 융합시켜, 융합체를 생성함으로써 전장 관심 단백질을 형성한다.

[0210] 인테인 활성화는 컨텍스트 의존성일 수 있으며, 라이게이션 또는 융합 접합부 주위의 특정 펩티드 서열 (N- 및 C-

엑스테인으로 불림)이 효율적인 스플라이싱이 일어나는 데 요구될 수 있다. 예를 들어, C-엑스테인 내의 제1 잔기로서 티올 또는 히드록실 기 (예를 들어, 시스테인 (Cys), 세린 (Ser), 트레오닌 (Thr))를 함유하는 아미노산이 효율적인 스플라이싱에 유용할 수 있다. 천연 시스테인은 스플라이스 부위로서 사용될 수 있다. 예를 들어, AAV 벡터는, 예를 들어 서열식별번호: 26의 위치 747 (Cys747; 변이체 1 & 3) 및 970 (Cys970; 변이체 2 & 4)에서 천연 시스테인을 이용할 수 있고, 이는 Ser746과 Cys747 또는 Ala969와 Cys970 사이의 스플라이스 부위를 제공할 수 있으며 (도 31a-31b), 여기서 분할 부위는 아미노산 서열: ELLSCFSPV (서열식별번호: 60) 또는 GPLACFLSP (서열식별번호: 61) 또는 그의 부분에서 발생할 수 있다. 또 다른 예는 서열식별번호: 25의 Ala708과 Cys709 또는 Ala933과 Cys934 사이의 스플라이스 부위를 제공하며, 여기서 분할 부위는 아미노산 서열: ELLACFSPV (서열식별번호: 62) 또는 GPLACFLSP (서열식별번호: 63) 또는 그의 부분에서 발생할 수 있다.

[0211] 본 개시내용의 한 실시양태는 관심 단백질, 예를 들어 스테레오실린의 각각의 절반을 갖는 엑스테인 영역을 제공한다. 예를 들어, 엑스테인 영역은 N-말단 스테레오실린 및 C-말단 스테레오실린을 포함할 수 있고, 이는 펩티드 결합을 통해 융합되는 경우에 전장 스테레오실린을 구성한다 (예를 들어, NCBI: NM_153700; NP_714544; NM_080459; NP_536707; 진뱅크: BK000138; AF375594; DAA00085; AF375593; AAL35321). 도 6을 참조한다. 또 다른 실시양태는 관심 단백질 (예를 들어, 스테레오실린)의 N-말단 부분 및 C-말단 부분의 합계가 100%인, 스테레오실린에 대한 상이한 분할-부위를 제공한다. 추가 실시양태에서, 분할 부위는 단백질 2차 구조의 나선 (예를 들어, 알파) 내에, 그 후에, 그 전에 또는 그에 인접하여 발생할 수 있다. 한 실시양태는 베타 가닥 또는 가교 내에 있지 않은 분할 부위를 제공할 수 있다. 또 다른 실시양태는 코일 영역의 바로 상류의 나선 (예를 들어, 알파) 내의 분할 부위를 제공한다. 또 다른 실시양태는 관심 단백질 (예를 들어, 스테레오실린) 및 막횡단 도메인의 외부의, 즉 막횡단 부분이 아닌 관심 단백질의 단편을 제공한다.

[0212] 예를 들어, 분할은 관심 단백질의 N-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 15 또는 류린 서열식별번호: 16)이 전장 관심 단백질 (예를 들어, 전장 인간 서열식별번호: 2 또는 서열식별번호: 25 또는 류린 서열식별번호: 4 또는 서열식별번호: 26)의 N-말단 단부의 적어도 10% 이상 (예를 들어, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%)의 길이를 포함하도록 이루어질 수 있다. 추가 실시양태는 전장 관심 단백질 (예를 들어, 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 N-말단 단부의 100% 이하 (예를 들어, 99%, 97%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%)의 길이를 포함하는 관심 단백질의 N-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 15 또는 류린 서열식별번호: 16)을 제공한다. 하나의 다른 실시양태는 전장 관심 단백질 (예를 들어, 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 N-말단 단부의 10% - 100% (예를 들어, 15%-90%, 20%-80%, 30%-70%, 40%-60%, 50%)의 길이를 포함하는 관심 단백질의 N-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 15 또는 류린 서열식별번호: 16)을 제공할 수 있다. 추가 실시양태는 전장 관심 단백질 (예를 들어, 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 N-말단 단부의 54% 미만 (예를 들어, 53%, 52%, 51%, 50%, 45%, 43%, 41%, 40%) 및/또는 그에 대해 54% 미만의 동일성 및/또는 그의 54% 미만의 길이의 관심 단백질의 N-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 15 또는 류린 서열식별번호: 16)을 제공한다. 추가 실시양태는 전장 관심 단백질 (예를 들어, 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 N-말단 단부의 54% 미만인 길이를 포함하는 관심 단백질의 N-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 15 또는 류린 서열식별번호: 16)에 관한 것일 수 있다. 또 다른 실시양태는 전장 관심 단백질 (예를 들어, 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 N-말단 단부의 40% 이상 (예를 들어, 41%, 42%, 43%, 44%, 45%, 50%, 51%, 52%, 53%) 및/또는 그에 대해 적어도 41% 이상의 동일성 및/또는 그의 41% 이상의 길이를 포함하는 관심 단백질의 N-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 15 또는 류린 서열식별번호: 16)을 제공한다.

[0213] 또 다른 실시양태에서, 관심 단백질의 C-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 23 또는 류린 서열식별번호: 24)은 전장 관심 단백질 (예를 들어, 스테레오실린; 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 C-말단 단부의 적어도 10% 이상 (예를 들어, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%; 100%)의 길이를 포함한다. 추가 실시양태는 전장 관심 단백질 (예를 들어, 스테레오실린; 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 C-말단 단부의 100% 이하 (예를 들어, 99%, 97%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%)의 길이를 포함하는 관심 단백질의 C-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 23 또는 류린 서열식별번호: 24)을 제공한다. 하나의 다른 실시양태는 전장 관심 단백질 (예를 들어, 스테레오실린; 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 C-말단 단부의 10% - 99% (예를 들어, 15%-90%, 20%-80%, 30%-70%, 40%-60%, 50%)의 길이를 포함하는 관심 단백질의 C-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 23 또는 류린 서열

식별번호: 24)을 제공할 수 있다. 추가 실시양태는 전장 관심 단백질 (예를 들어, 스테레오실린; 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 C-말단 단부의 46% 이상을 포함하는 관심 단백질의 C-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 23 또는 류린 서열식별번호: 24)을 제공한다. 추가 실시양태는 전장 관심 단백질 (예를 들어, 스테레오실린; 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 C-말단 단부의 46% 이상 (예를 들어, 47%, 48%, 49%, 50%, 55%, 60%)인 길이 및/또는 그에 대해 46% 이상의 동일성 및/또는 그의 46% 이상의 길이를 포함하는 관심 단백질의 C-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 23 또는 류린 서열식별번호: 24)에 관한 것일 수 있다. 추가 실시양태는 전장 관심 단백질 (예를 들어, 스테레오실린; 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 N-말단 단부의 46% 이상 (예를 들어, 47%, 48%, 49%, 50%, 55%, 60%)인 길이를 포함하는 관심 단백질의 C-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 23 또는 류린 서열식별번호: 24)에 관한 것일 수 있다. 또 다른 실시양태는 전장 관심 단백질 (예를 들어, 스테레오실린; 전장 인간 서열식별번호: 25 또는 류린 서열식별번호: 26)의 C-말단 단부의 60% 이하 (예를 들어, 59%, 58%, 57%, 56%, 55%, 50%, 45%) 및/또는 그에 대해 60% 이하의 동일성 및/또는 그의 60% 이하의 길이를 포함하는 관심 단백질의 C-말단 부분 (예를 들어, 스테레오실린; 인간 서열식별번호: 23 또는 류린 서열식별번호: 24)을 제공한다.

- [0214] 추가 실시양태는, 예를 들어 스테레오실린의 N-말단 및 C-말단 단편을 형성하기 위한, 인간 서열식별번호: 25의 Ala708과 Cys709 또는 Ala933과 Cys934 사이의 전장 야생형 스테레오실린 단백질의 분할, 또는 류린 서열식별번호: 26의 Ser746과 Cys747 또는 Ala969와 Cys970 사이의 분할을 제공할 수 있다.
- [0215] 이중-벡터 시스템
- [0216] 본 개시내용의 한 실시양태는 세포에서 관심 단백질을 발현시키기 위한 이중-벡터 시스템을 제공하며, 여기서 이중-벡터 시스템은 하기를 포함한다:
- [0217] a) 5'에서 3' 방향으로:
 - [0218] - 부분 코딩 서열의 5'-말단에 있는 신호 서열 (예를 들어, 서열식별번호: 11; 서열식별번호: 12를 코딩함);
 - [0219] - 관심 단백질의 아미노 말단 (N-말단) 부분 (예를 들어, STRC; 서열식별번호: 15; 서열식별번호: 16)을 코딩하는 부분 코딩 서열;
 - [0220] - 스플라이스 공여자 서열을 코딩하는 서열 (예를 들어, 인테인의 아미노 말단 단편 (N-인테인); 서열식별번호: 13; 서열식별번호: 14를 코딩함)
- [0221] 을 포함하는 제1 뉴클레오티드 서열을 포함하는 제1 벡터 (예를 들어, 캡시드, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 계통, 렌티바이러스); 및
- [0222] b) 5'에서 3' 방향으로:
 - [0223] - 스플라이스 수용자 서열의 상류일 수 있는 신호 서열 (예를 들어, 서열식별번호: 11; 서열식별번호: 12를 코딩함);
 - [0224] - 스플라이스 수용자 서열을 코딩하는 서열 (예를 들어, 인테인의 카르복시 말단 단편 (C-인테인); 서열식별번호: 21; 서열식별번호: 22를 코딩함);
 - [0225] - 관심 단백질의 카르복시 말단 (C-말단) 부분 (예를 들어, STRC; 서열식별번호: 23; 서열식별번호: 24)을 코딩하는 부분 코딩 서열
- [0226] 을 포함하는 제2 뉴클레오티드 서열을 포함하는 제2 벡터 (예를 들어, 캡시드, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 계통, 렌티바이러스).
- [0227] 추가 실시양태는 세포에서 관심 단백질을 발현시키기 위한 이중-벡터 시스템을 제공하며, 여기서 이중-벡터 시스템은 하기를 포함한다 (예를 들어, 도 6 참조):
- [0228] a) 5'에서 3' 방향으로:
 - [0229] - 5'-역전된 말단 반복부 (5'-ITR) 서열;
 - [0230] - 프로모터 서열;
 - [0231] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열 (예를 들어, 서열식별번호: 11; 서

도 13b 및 도 14를 참조한다. 도 12a-12b (서열식별번호: 17) 및 도 13a-13b (서열식별번호: 19)는 5'에서 3' 방향으로, 개시 코돈 (볼드체 ATG), 신호 서열 (소문자, 이탤릭체 및 밑줄표시; 각각 서열식별번호: 9; 서열식별번호: 11), C-인테인 서열 (볼드체 및 밑줄표시) (5'-ATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGAAAGAGATCACAACCTTGTCTGAAGAACGGATTTCATAGCTTCTAAT-3'; 서열식별번호: 46), STRC 유전자의 C-말단 부분의 코딩 서열 (후색) (여기서 코딩 서열은 스테레오실린 (STRC) 단백질을 코딩함), 링커 서열 (볼드체 및 이탤릭체) (5'-ACGCGTACGCGGCCGCTC-3'; 서열식별번호: 47) 및 Myc 태그 서열 (소문자) (5'-GAGCAGAACTCATCTCAGAAGAGGATCTGTAATAG-3'; 서열식별번호: 48) 및 정지 코돈 (이탤릭체 및 밑줄표시)을 포함하는 뉴클레오티드 서열을 도시한다. 도 14는 ITR, 프로모터 및 폴리A 테일을 포함한 추가의 요소를 보여주며, 여기서 유린 3'-STRC는 전장 스테레오실린 (STRC) 단백질의 C-말단 부분 (747 (Cys)-1,810 아미노산; 116.7 kDa)을 코딩한다. 도 15 및 16은 제2 뉴클레오티드 서열에 의해 코딩되는 아미노산 서열을 보여주며, 여기서 C-말단 부분의 아미노산 서열 (서열식별번호: 18; 서열식별번호: 20)은 ATG 코돈에 의해 코딩되는 메티오닌 (M) (볼드체), 신호 펩티드 서열 (소문자, 이탤릭체 및 밑줄표시) (서열식별번호: 10; 서열식별번호: 44), C-인테인 (볼드체 및 밑줄표시) (N-IKIATRKYLKGQNVYDIGVERDHNFKNGFIASN-C; 서열식별번호: 49), 스테레오실린 단백질의 C-말단 부분 (후색), 링커 서열 (볼드체 및 이탤릭체) (N-TRTRPL-C; 서열식별번호: 50) 및 Myc 태그 (소문자) (서열식별번호: 27)를 포함한다. 한 실시양태는 스테레오실린 단백질의 C-말단 부분이 시스테인 (C; Cys), 페닐알라닌 (F; Phe) 및 세린 (S; Ser)으로 시작되는 전장 STRC 단백질을 제공할 수 있다 (도 17).

[0248] 추가 실시양태에서, 이중-벡터 시스템은 5'에서 3' 방향으로 서열식별번호: 51의 서열을 포함하는 제1 뉴클레오티드 서열을 포함하는 제1 벡터를 제공하지 않을 수 있다. 예를 들어, 도 18a, 도 18b 및 도 19를 참조한다. 도 18a-18b (서열식별번호: 51)는 5'에서 3' 방향으로, 개시 코돈 (ATG, 볼드체), 신호 서열 (소문자, 이탤릭체 및 밑줄표시) (서열식별번호: 11), STRC 유전자의 N-말단 부분의 코딩 서열 (후색) 및 N-인테인 (볼드체 및 밑줄표시) (서열식별번호: 42) (여기서 코딩 서열은 스테레오실린 (STRC) 단백질을 코딩함) 및 정지 코돈 (이탤릭체 및 밑줄표시)을 포함하는 뉴클레오티드 서열을 도시한다. 도 19는 ITR, 프로모터 및 폴리A 테일을 포함한 추가의 요소를 보여주며, 여기서 5'-STRC는 전장 스테레오실린 (STRC) 단백질의 N-말단 부분 (1-969 (Ala) 아미노산; 104.8 kDa)을 코딩한다. 도 20은 제1 뉴클레오티드 서열에 의해 코딩되는 아미노산 서열을 보여주며, 여기서 N-말단 부분의 아미노산 서열 (서열식별번호: 52)은 ATG 코돈에 의해 코딩되는 메티오닌 (M) (볼드체), 신호 펩티드 서열 (소문자, 이탤릭체 및 밑줄표시) (서열식별번호: 44), 스테레오실린 단백질의 N-말단 부분 (후색) 및 N-인테인 (볼드체 및 밑줄표시) (서열식별번호: 45)을 포함한다.

[0249] 추가 실시양태는 5'에서 3' 방향으로 서열식별번호: 53의 서열을 포함하는 제2 뉴클레오티드 서열을 포함하는 제2 벡터를 갖는 이중-벡터 시스템을 제공하지 않을 수 있다. 예를 들어, 도 21a, 도 21b 및 도 22를 참조한다. 도 21a-21b (서열식별번호: 53)는 5'에서 3' 방향으로, 개시 코돈 (볼드체 ATG), 신호 서열 (소문자, 이탤릭체 및 밑줄표시) (서열식별번호: 11), C-인테인 서열 (볼드체 및 밑줄표시) (서열식별번호: 21), STRC 유전자의 C-말단 부분의 코딩 서열 (후색) (여기서 코딩 서열은 스테레오실린 (STRC) 단백질을 코딩함), 링커 서열 (볼드체 및 이탤릭체) (서열식별번호: 47), Myc 태그 서열 (소문자) (서열식별번호: 48) 및 정지 코돈 (이탤릭체 및 밑줄표시)을 포함하는 뉴클레오티드 서열을 도시한다. 도 22는 ITR, 프로모터 및 폴리A 테일을 포함한 추가의 요소를 보여주며, 여기서 3'-STRC는 전장 스테레오실린 (STRC) 단백질의 C-말단 부분 (970 (Cys)-1,810 아미노산; 91.6 kDa)을 코딩한다. 도 23은 제2 뉴클레오티드 서열에 의해 코딩되는 아미노산 서열을 보여주며, 여기서 C-말단 부분의 아미노산 서열 (서열식별번호: 54)은 ATG 코돈에 의해 코딩되는 메티오닌 (M) (볼드체), 신호 펩티드 서열 (소문자, 이탤릭체 및 밑줄표시) (서열식별번호: 44), C-인테인 (볼드체 및 밑줄표시) (서열식별번호: 49), 스테레오실린 단백질의 C-말단 부분 (후색), 링커 서열 (볼드체 및 이탤릭체) (서열식별번호: 50) 및 Myc 태그 (소문자) (서열식별번호: 27)를 포함한다.

[0250] 이중-벡터 시스템의 한 실시양태는 관심 단백질 (예를 들어, 스테레오실린)의 각각의 분할된 부분, 즉 N-말단 부분 및 C-말단 부분에 대한 벡터를 제공한다. 관심 단백질 (예를 들어, 스테레오실린)의 분할된 부분 및 관심 단백질을 조절, 생산 또는 발현시키는 데 필요한 임의의 추가의 영역은 각각의 부분 및 연관된 영역이 그의 각각의 벡터 (예를 들어, 바이러스 (예를 들어, 바이러스 벡터, 박테리오파지, 파지, 레트로바이러스), 플라스미드, 코스미드, 박테리아 인공 염색체, 효모 인공 염색체, 인간 인공 염색체)의 적재 용량을 초과하지 않도록 하는 것이다. 동일한 관심 단백질의 분할된 부분 및 추가의 영역은 선택된 벡터의 적재 용량을 초과하는 크기여서는 안된다.

[0251] 본 개시내용의 한 실시양태는 4 kB 이상 (예를 들어, 4.5 kB, 5 kB, 5.5 kB, 5.8 kB, 6 kB, 6.5 kB, 7 kB, 7.5

kB, 8 kB, 8.5 kB, 9 kB, 9.5 kB, 10 kB, 11 kB, 12 kB)을 포함한 큰 코딩 서열을 갖는 유전자를 전달하기 위한 벡터 시스템 (예를 들어, 이중-벡터 시스템)을 제공한다. 본 개시내용의 벡터 (예를 들어, 제1 벡터 및 제2 벡터)는 각각 바이러스 벡터 (예를 들어, 아데노바이러스, 아데노-연관 바이러스 (AAV), 렌티바이러스, 단순 포진 바이러스 I, 백시니아 바이러스)일 수 있고, 일부 실시양태에서, 바이러스 벡터는 AAV 벡터 또는 재조합 AAV 벡터일 수 있다. 또 다른 실시양태는 동일하거나 상이한 혈청형 (예를 들어, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, 합성 혈청형)의 바이러스 벡터를 제공할 수 있다. 본원에 기재된 개시내용에 유용한 AAV 혈청형은 AAV1, AAV2, AAV5, AAV6, AAV8 및 AAV9를 포함할 수 있으나 이에 제한되지는 않는다.

[0252] 한 실시양태에서, 제1 및 제2 벡터는 바이러스 DNA가 결여된, 본원에서 상호교환가능하게 사용되는 바이러스 벡터, 예를 들어 아데노-연관 바이러스 (AAV) 또는 재조합 AAV (rAAV)이다. 이중-벡터 시스템의 제1 벡터는 5'에서 3' 방향으로: 5'-역전된 말단 반복부 (5'-ITR) 서열; 하류 관심 폴리뉴클레오티드 (예를 들어, STRC)의 전사를 구동시킬 수 있는 프로모터 서열; 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열; 관심 단백질 (예를 들어, STRC)의 아미노 말단 (N-말단) 부분을 코딩하는 부분 코딩 서열인, 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 부분 코딩 서열; 인테인의 아미노 말단 단편 (N-인테인)을 코딩하는 서열인, 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 N-인테인을 코딩하는 서열; 폴리-아데닐화 (폴리A) 신호 서열; 및 3'-ITR 서열을 함유하는 뉴클레오티드 서열을 포함한다.

[0253] 또 다른 실시양태에서, 제2 벡터는 5'에서 3' 방향으로: 5'-역전된 말단 반복부 (5'-ITR) 서열; 하류 관심 폴리뉴클레오티드 (예를 들어, STRC)의 전사를 구동시킬 수 있는 프로모터 서열; 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열; 인테인의 카르복시 말단 단편 (C-인테인)을 코딩하는 서열인, 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 C-인테인을 코딩하는 서열; 관심 단백질 (예를 들어, STRC)의 카르복시 말단 (C-말단) 부분을 코딩하는 부분 코딩 서열인, 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 부분 코딩 서열; 폴리-아데닐화 (폴리A) 신호 서열; 및 3'-ITR 서열을 포함한다.

[0254] 추가 실시양태에서, 본 개시내용의 제1 벡터 및 제2 벡터가 바이러스 형질도입, 염화칼슘을 사용한 박테리아 형질전환, 박테리아 교배 또는 접합에 의한 박테리아 형질전환 또는 형질도입, 형질감염 (예를 들어, 전기천공, 인산칼슘, 리포솜-기반 형질감염), 유전자 총 등을 포함하나 이에 제한되지는 않는 임의의 수단에 의해 세포(들) (예를 들어, 숙주 세포, 포유동물 세포, 인간 세포, 박테리아 세포) 내로 삽입되는 경우에, 벡터는 각각 N-말단에서 C-말단 방향으로, 신호 펩티드 서열, 이에 연결된 관심 단백질 서열 (예를 들어, STRC)의 N-말단 부분, 이의 C-말단 단부에서 융합된 N-인테인 단백질 서열을 포함하는 제1 단백질 서열; 및 N-말단에서 C-말단 방향으로, 신호 펩티드 서열, 이에 연결된 C-인테인 단백질 서열, 이에 N-말단 단부에서 융합된 관심 단백질 서열 (예를 들어, STRC)의 C-말단 부분을 포함하는 제2 단백질 서열을 발현한다. 관심 단백질 (예를 들어, STRC)의 N-말단 부분 및 동일한 관심 단백질 (예를 들어, STRC)의 C-말단 부분의 인테인-매개 단백질 스플라이싱은 전장 관심 단백질 (예를 들어, STRC)의 발현을 일으킨다.

[0255] 또 다른 실시양태는 이중-벡터 시스템의 신호 펩티드 서열을 제공하며, 여기서 신호 펩티드 서열은 이중-벡터 시스템에 의해 코딩되는 각각의 단백질 서열의 N-말단 단부에 위치할 수 있다. 관심 단백질 (예를 들어, STRC)의 N-말단 부분을 포함하는 제1 단백질 서열에서, 신호 펩티드 서열은 관심 단백질의 N-말단 부분의 코딩 영역뿐만 아니라 N-인테인의 상류일 수 있다. 관심 단백질 (예를 들어, STRC)의 C-말단 부분을 포함하는 제2 단백질 서열에서, 신호 펩티드 서열은 C-인테인뿐만 아니라 관심 단백질 (예를 들어, STRC)의 C-말단 부분의 코딩 영역의 상류일 수 있다.

[0256] 본 개시내용의 또 다른 실시양태는 세포 내의 제1 벡터 및 제2 벡터가, 각각 동일한 신호 펩티드 서열을 함유하는, 각각 제1 단백질 서열 및 제2 단백질 서열을 발현하는 이중-벡터 시스템을 제공한다. 따라서, 동일한 신호 펩티드 서열은 제1 단백질 서열 및 제2 단백질 서열이 동일한 세포 구획으로 수송되도록 한다. 상이한 실시양태에서, 제1 및 제2 단백질 서열의 신호 펩티드 서열은 상이할 수 있지만, 이들 신호 펩티드 서열은 각각의 단백질 서열을 동일한 세포 구획으로 지시한다. 제1 및 제2 단백질 서열의 신호 펩티드 서열은 제1 단백질 서열 및 제2 단백질 서열을 동일한 세포 구획으로 수송하도록 구성될 수 있다. 그렇게 함에 있어서, 각각의 단백질 서열은 인테인-매개 단백질 융합이 발생하기에 충분히 근접할 수 있으며, 그에 의해 전장 관심 단백질 (예를 들어, STRC)을 형성한다. 신호 펩티드 서열은 관심 단백질 (예를 들어, STRC)과 회합될 수 있다. 추가 실시양태는 관심 단백질 이외의 다른 단백질과 회합된 신호 펩티드 서열을 코딩하는 신호 서열을 제공할 수 있으며, 여기서 제1 뉴클레오티드 서열의 신호 서열 및 제2 뉴클레오티드 서열의 신호 서열은 상이하고, 또한 신호 서열들이 상이한 신호 펩티드 서열들을 코딩하지만, 신호 서열들은 제1 단백질 서열 및 제2 단백질 서열을 동일한 세

또 구체적으로 수송하도록 구성된다. 본 개시내용의 또 다른 실시양태는 신호 서열이 인테인-매개 트랜스-스플라이싱을 파괴하지 않으면서 2개의 단편을 동일한 세포 또는 세포내 구체적으로 지시하도록 하는 신호 서열 또는 신호 서열들을 제공한다. 신호 서열은 제1 단백질 서열 및 제2 단백질 서열이 서로에 대해 충분히 근접하여 관심 단백질 (예를 들어, STRC)의 N-말단 부분 및 관심 단백질 (예를 들어, STRC)의 C-말단 부분이 펩티드 결합을 통해 전장 관심 단백질 (예를 들어, STRC)을 형성하게 하는 것을 보장하는 데 특히 유용할 수 있다.

[0257] 본 개시내용의 한 실시양태에서, 신호 서열은 관심 단백질 (예를 들어, STRC)의 신호 펩티드 서열을 코딩하는 신호 서열에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함할 수 있다. 예를 들어, 신호 서열은 STRC 단백질 신호 펩티드를 코딩하는 유전자 서열의 서열식별번호: 11에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함할 수 있거나, 또는 신호 서열은, 예를 들어 관심 단백질 및 관심 단백질의 각각의 부분을 동일한 세포 구체적으로 지시하는 임의의 신호 서열 (예를 들어, 서열식별번호: 9 또는 서열식별번호: 11)로 이루어진 핵산 서열을 포함할 수 있다. 또 다른 실시양태는 관심 단백질 (예를 들어, STRC; 서열식별번호: 10; 서열식별번호: 12)의 신호 펩티드 서열에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 갖는 신호 펩티드 서열을 코딩하는 신호 서열을 제공할 수 있다. 예를 들어, 신호 펩티드 서열은 STRC 단백질 신호 펩티드의 서열식별번호: 10 또는 서열식별번호: 12에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 포함할 수 있거나, 또는 신호 펩티드 서열은 서열식별번호: 10 또는 서열식별번호: 12로 이루어진 아미노산 서열을 포함할 수 있다.

[0258] 추가 실시양태는 관심 단백질 (예를 들어, STRC)의 N-말단 부분을 코딩하는 부분 코딩 서열을 제공할 수 있으며, 여기서 부분 코딩 서열은 관심 단백질의 N-말단 부분을 코딩하는 코딩 서열 (예를 들어, STRC; 서열식별번호: 6, 8, 15, 16, 25 또는 26)에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함한다. 예를 들어, STRC 단백질의 N-말단 부분을 코딩하는 부분 코딩 서열 (상이한 신호 서열과 교환가능할 수 있는 신호 서열을 포함함)은 하기 서열에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함할 수 있다.

[0259] STRC 단백질의 인간 N-말단 부분을 코딩하는 부분 코딩 서열 (개시 코돈 ATG (볼드체) 및 신호 서열 (소문자, 이탤릭체 및 밑줄표시) 포함)은 하기와 같을 수 있다:

5' ATGgctctcagcctctggccctgctgctgctgctgctgctgctgctgctgctgctgctccttt
gcaGTGACTCTGGCCCCCTACTGGGCTCATTCCCTGGACCCTGGTCTCTCCTTCCTGAAGTC
ATTGCTCTCCACTCTGGACCAGGCTCCCCAGGGCTCCCTGAGCCGCTCACGGTCTTTTACAT
TCCTGGCCAACATTTCTTCTTCTTTGAGCCTGGGAGAATGGGGGAAGGACCAGTAGGAGAG
CCCCACCTCTCCAGCCGCTGCTCTGCGGCTCCATGATTTTCTAGTGACACTGAGAGGTAG
CCCCGACTGGGAGCCAATGCTAGGGCTGCTAGGGGATATGCTGGCACTGCTGGGACAGGAGC
AGACTCCCCGAGATTTCTTGGTGCACCAGGCAGGGGTGCTGGGTGGACTTGTGGAGGTGCTG
CTGGGAGCCTTAGTTCTTGGGGGCCCCCTACCCCAACTCGGCCCCCATGCACCCGTGATGG
GCCGTCTGACTGTGTCTGGCTGCTGACTGGTTGCCTTCTCTGCTGCTGTTGTAGAGGGCA
CACGCTGGCAAGCTCTGGTGCAGGTGCAGCCAGTGTGGACCCACCAATGCCACAGGCCTC
GATGGGAGGGAGGCAGCTCCTCATTTTGTCAGGGTCTGTTGGGTTTGCTTACCCCAACAGG
GGAGCTAGGCTCCAAGGAGGCTCTTTGGGGCGGTCTGCTACGCACAGTGGGGGCCCCCTCT
ATGCTGCCTTTTCTAGGAGGGGTGCTCCGTGTCACTCACTCCCTGCAGGATGAGGTCTTCTCC
ATTTTGGGGCAGCCAGAGCCTGATACCAATGGGCAGTGCCAGGGAGGTAACCTTCAACAGCT
GCTCTTATGGGGCGTCCGGCACAACCTTTCTTGGGATGTCCAGGCGCTGGGCTTTCTGTCTG
GATCACCAACCCCCACCCCTGCCCTCCTTCACTGCCTGAGCACGGGCGTGCCTCTGCCAGA
GCTTCTCAGCCGTCAGCCACATCAGCCACGCCAACGGCGAGCCATCACTGTGGAGGCCCT
CTGTGAGAACCATTAGGCCCAGCACACCTACAGCATTTCCAACCTCTCCATCCACTTGC
TCTGCCAGCACACCAAGCCTGCCACTCCACAGCCCCATCCAGCACCCTGCCATCTGCCAG
ACAGCTGTGTGGTATGCAGTGTCTGGGCACCAGGTGCCAAGGCTGGCTACAGGCCTGCCA
CGACCAGTTTCTGATGAGTTTTTGGATGCGATCTGCAGTAACCTCTCCTTTTCTAGCCCTGT
CTGGCTCCAACCGCCGCTGGTGAAGCGGTCTGTGCTGGCCTGCTCCACCCCTACCAGC
TGCCCTGAAGGCCTGCCCTGTTCCTTCACCCAGACATCTTTTGGGGCTGCTTCTTGGA
GAATGAGACTCTGTGGGCTGAGCGACTGTGTGGGGAGGCAAGTCTACAGGCTGTGCCCCCA
GCAACCAGGCTTGGGTCCAGCATGTGTGCCAGGGCCCCACCCAGATGTCACTGCCTCCCCA
CCATGCCACATTGGACCCTGTGGGGAACGCTGCCCGGATGGGGGCGAGCTTCTGGTGATGGT
CTGTGCCAATGACACCATGTATGAGGTCTTGGTGCCTTCTGGCCTTGGCTAGCAGGCCAAT
GCAGGATAAGTCGTGGGGCAATGACACTTGCTTCCTAGAAGGGCTGCTGGGGCCCCCTTCTG
CCCTCTCTGCCACCACTGGGACCATCCCCACTCTGTCTGACCCCTGGCCCCCTTCTCCTTGG
CATGCTATCCCAGTTGCCACGCTGTCTAGTCTCTGTCCCAGCTCTTGTCTACCCACACGCC
TACACTATCTCTCCGCTGCTGACCTTCTCTTGGGTCCAGGGGCTGGGGGCGCTGAGGCC
CAGGGGATGCTGGGTGGGGCCCTACTGCTCTCCAGTCTCCAGACAACTGCTCCTTCTGGGA
TGCCTTTGCCCCAGAGGGCCGGCGCAGTGTGTACGGACGATTGGGGAATACCTGGAACAAG
ATGAGGAGCAGCCAACCCCATCAGGCTTTGAACCCACTGTCAACCCAGCTCTGGTATAAGC

[0260]

AAGATGGAGCTGCTGGCC (SEQ ID NO:55).

[0261]

[0262]

STRC 단백질의 류린 N-말단 부분을 코딩하는 부분 코딩 서열 (개시 코돈 ATG (볼드체) 및 신호 서열 (소문자, 이탤릭체 및 밑줄표시) 포함)은 하기와 같을 수 있거나:

5'-ATGgctctgagcctccagccccagctgctccttctcctgtcgctcctgccgcaggaagt
gacttcaGCCCTACTGGGCTCAGTCTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTCTG
TAGCCACTCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTCACGGTTCTCTGCGTTCTCTG
GCCAACATTTCTTCATCCTTCCAGCTTGGGAGGATGGGGGAGGGACCGGTGGGAGAGCCCCC
ACCTCTCCAGCCCCCTGCACTTCGACTTCATGATTTCCTCGTGACACTGAGAGGTAGCCCAG
ACTGGGAGCCAATGCTAGGGCTTCTGGGAGATGTGCTGGCACCTCTGGGACAGGAACAGACT
CCCCGGGACTTTTTTGGTGCACAGGAGGTGTACTGGGTGGACTTGTAGAGGCATTGTTGGG
AGCGTTAGTTCTGGAGGCCCCCTGCCCCCACTCGACCCCCATGCACCCGTGATGGCCCTT
CTGACTGTGTCTGGCTGTGATTGGTTGCCTTCTCTGATGTTGTTATTAGAGGGTACACGC
TGGCAGGCCCTGGTGCAGTTGCAGCCAGTGTGGACCCAACCAATGCCACAGGTCTTGATGG
TAGAGAGCCAGCTCCTCACTTTTTACAGGGTCTGTCTGGGCTTGCTTACCCAGCAGGAGAGT
TGGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGTGGGGGCCCCCTCTATGCT
GCCTTCCAGGAGGGGCTACTGCGAGTCACTCATTCTCTGCAAGATGAGGTCTTTTCTATTAT
GGGACAGCCAGAGCCTGATGCCAGTGGGAGTGGCAGGGAGGCAACCTTCAACAGCTGCTTT
TATGGGGCATGCGGAACAACCTTTCTTGGGACGCCCCGAGCACTGGGTTTTCTATCTGGATCA
CCACCTCCACCCCTGCTCTCCTGCACTGCCTGAGCAGAGGTGTGCCTCTGCCAGGGCTTC
CCAGCTGCGGCTCACATCAGCCCTCGACAGCGGCGAGCCATCTCTGTGGAGGCCCTCTGCG
AGAACCACTCAGGCCAGAGCCACCTACAGCATCTCCAACCTCTCCATCTACTTGTCTGCTGC
CAGCACATCAAGCCTGCCACCCCGCGGCCCCCTCTACCACCCACGGCTCCTCCTACCAC
CCCACAGCCCCCTCTACCCTACACAGCCCATTCCTGACACTACACAGCCCCCTCCTGTCA
CCCCAAGGCTCCTCCTACCACCCCAACCCCTCCTAGCACAGCTGTCTATCTGCCAGACA
GCTGTATGGTACGCAGTCTCGTGGGACACAGGTGCCCCGAGGTGGCTCCAAGCCTGCCATGA
TCAGTTTCTGATCAATTTCTGGATATGATCTGCGGCAACCTCTCATTTTTCAGCCCTGTCTG
GCCCCAGTCGTCCTTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCACCCCCCACTAGCTGT
CCACAGGCTGATCCTGTGCCCCCTACCCCAAGAAATATCTGGGGCTGTTTCTGGAGAA
TGAGACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCAGGCTGTGCCCCGAGGA
ACCAGGCTTGGGTTTCTGATGTGTGTGCGGGCCCCACCTTGGACGCCACTGATTTTCCACCG
TGCCGCGTTGGACCCTGTGGGGAACGCTGCCAGATGGGGGAGCTTCTGCTCATGGTCTG
TGCCAATGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTGGCTAGCAGGCCAATGCA

GAATTAGTCGTGGAGGAAATGATACTTGCTTTCTAGAAGGCATGCTGGGCCCCCTTGTTGCCC
TCTCTGCCCCCTCTGGGACCATCCCCACTCTGTCTGGCTCCTGGTCTCTTTCTGCTTGGCAT
GTTATCCAGTTGCCACGCTGTGAGTCTCCGTGCCAGCCCTCGCCACCCACGCGCCTAC
ATTACCTCCTGCGCCTACTGACCTTCCTTCTGGGTCCAGGGACTGGGGGTGCCGAGACGCAG
GGGATGTTAGGTCAAGCCCTGCTGCTCTCTAGTCTCCAGACAACTGTTTCTGTTGGATGC
CTTCCGCCAGAGGGCCGAGAAGTGTACTGAGGACAGTCCGAGAGTACTTGCAGCGGGAAG
AGCCAACCCACAGGCTTAGACTCCTCCCTCAGCCTCGGCTCTGGTATGAGCAAGATGGAG
CTTCTGTCC-3' (SEQ ID NO:56)

또는 STRC 단백질의 N-말단 부
분을 코딩하는 부분 코딩 서열은 서열식별번호: 55 또는 56으로 이루어진 핵산 서열을 포함할 수 있다. 또 다
른 실시양태는 관심 단백질 (예를 들어, STRC; 서열식별번호: 25 또는 26)에 대해 적어도 5% (예를 들어, 6%,
7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산
서열을 갖는 관심 단백질 (예를 들어, STRC)의 N-말단 부분 (개시 코돈 ATG에 상응하는 메티오닌 및 신호 펩티
드 서열을 포함함)을 제공할 수 있다. 예를 들어, STRC 단백질의 N-말단 부분 (개시 코돈 ATG에 상응하는 메티
오닌 및 STRC 신호 펩티드 서열 포함, 이는 교환가능할 수 있음)은 서열식별번호: 15 또는 16의 서열에 대해 적
어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%,
100%) 동일성의 아미노산 서열을 포함할 수 있거나, 또는 STRC 단백질의 N-말단 부분은 서열식별번호: 15 또는
16으로 이루어진 아미노산 서열을 포함할 수 있다. 제1 벡터의 제1 뉴클레오타이드 서열은 자신의 신호 서열 (예
를 들어, 서열식별번호: 10 또는 12) 또는 대안적으로 상이한 신호 서열을 포함하는 관심 단백질의 N-말단 부분

(예를 들어, STRC; 서열식별번호: 15 또는 16)을 코딩하는 부분 코딩 서열 및 스플라이스 공여자 서열 (예를 들어, N-말단 인테인 (N-인테인); 서열식별번호: 14)을 코딩하는 서열을 포함할 수 있다.

[0265] 한 실시양태에서, 관심 단백질 (예를 들어, STRC)의 C-말단 부분을 코딩하는 부분 코딩 서열 (개시 코돈 ATG에 상응하는 메티오닌 및 신호 서열 포함, 이는 교환가능할 수 있음; 임의로 링커 서열 및 Myc 태그 서열 포함)이며, 여기서 부분 코딩 서열은 관심 단백질 (예를 들어, STRC; 서열식별번호: 18, 20, 23, 24, 25 또는 26)에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함한다. 예를 들어, STRC 단백질의 C-말단 부분을 코딩하는 부분 코딩 서열은 하기 서열에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함할 수 있다.

[0266] STRC 단백질의 인간 C-말단 부분을 코딩하는 부분 코딩 서열은 하기와 같을 수 있다 (즉, ATG 개시 코돈, 신호 서열 또는 스플라이스 수용자 서열이 없음):

```
5'-TGCTTTAGTCCTGTGCTGTGGGATCTGCTCCAGAGGAAAAGAGTGTTTGGGCCCTGCA
GATTCTAGTGCAGGCGTACCTGCATATGCCCCAGAAAACCTCCAGCAGCTGGTGCTTTCAG
CAGAGAGGGAGGCTGCACAGGGCTTCCTGACACTCATGCTGCAGGGGAAGCTGCAGGGGAAG
CTGCAGGTACCACCATCCGAGGAGCAGGCCCTGGGTGCGCTGACAGCCCTGCTGCTCCAGCG
GTACCCACGCCTCACCTCCCAGCTCTTCATTGACCTGTCACTCATCCCTTCTTGGCTG
TCTCTGACCTGATGCGCTTCCACCATCCCTGTTAGCCAACGACAGTGTCTGGCTGCCATC
CGGGATTACAGCCCAGGAATGAGGCTGAACAGAAGGAGGCTCTGGCAAAGCGACTGCTGGC
CCCTGAACTGTTTGGGGAAGTGCCTGCCTGGCCCCAGGAGCTGCTGTGGGCAGTGCTGCCCC
TGCTCCCCACCTCCCTCTGGAGAACTTTTTGCAGCTCAGCCCTCACCAGATCCAGGCCCTG
GAGGATAGCTGGCCAGCAGCAGGTCTGGGGCCAGGGCATGCCCGCCATGTGCTGCGCAGCCT
GGTAAACCAGAGTGTCCAGGATGGTGAGGAGCAGGTACGCAGGCTTGGGCCCTCGCCTGTT
TCCTGAGCCCTGAGGAGCTGCAGAGCCTAGTGCCCTGAGTGATCCAACGGGGCCAGTAGAA
CGGGGGCTGCTGGAATGTGCAGCCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATATGA
ACTTCTGGGTGTGTTGCGCTCATCTGGAGGAGCGGTGCTGAGCCCCGGGAGCTGCGGGTCT
GGGCCCTCTCTTCTCTCAGCTGGGCCTCCGCTTCCTTCAGGAGCTGTGAGAGCCCCAGCTT
AGAGCCATGCTTCTGTCTGCTGCAGGGAAGTAGTGTACACCTGCTCAGGCTGTCTGTGCT
TGGACGGCTCCTTCCTAGGCACGATCTATCCCTGGAGGAAGCTGCTCCTTGACCTTCTGC
TACCAGGCCTCAGCCCCAGACACTCCAGGCCATCCCTAGGCGAGTCTGGTCGGGGCTTGT
TCCTGCCTGGCCCCGAACTGTACGCCTCTCAGCCTGCCAGACCGCAGCACTGCTGCAGAC
CTTTCGGGTAAAGATGGTGTTAAAAATATGGGTACAACAGGTGCTGGTCCAGCTGTGTGTA
TCCCTGGTCAGCCTATTCACCACCTGGCCAGACTGCCTGCTTCCCTGCTCCCATTAAG
CTGCTACAACCTGGATTCTTGGCTCTTCTGGCAAATCGAAGACGCTACTGGGAGCTGCCCTG
GTCTGAGCAGCAGGCACAGTTCCTCTGGAAGAAGATGCAAGTACCCACCAACCTTACCCTCA
GGAATCTGCAGGCTCTGGGCACCCTGGCAGGAGCATGTCTGTGAGTTTCTGCAGCAGATC
AACTCCATGGTAGACTTCCTTGAAGTGGTGCACATGATCTATCAGCTGCCCCACTAGAGTTCTG
AGGGAGCCTGAGGGCCTGTATCTGGGCAGAGCTACAGCGGAGGATGGCAATGCCAGAACCAG
AATGGACAACGTAGGGCCAGAAGTGAACGGGCTGGATAGCAAGCTACTCCTGGACTTACCG
ATCCAGTTGATGGACAGACTATCCAATGAATCCATTATGTTGGTGGTGGAGCTGGTGCAAAG
```

[0267]

AGCTCCAGAGCAGCTGCTGGCACTGACCCCCCTCCACCAGGCAGCCCTGGCAGAGAGGGCAC
TACAAAACCTGGCTCCAAAGGAGACTCCAGTCTCAGGGGAAGTGTGGAGACCTTAGGCCCT
TTGGTTGGATTCTTGGGGACAGAGAGCACACGACAGATCCCCCTACAGATCCTGCTGTCCCA
TCTCAGTCAGCTGCAAGGCTTCTGCCTAGGAGAGACATTTGCCACAGAGCTGGGATGGCTGC
TATTGCAGGAGTCTGTTCTTGGGAAACCAGAGTTGTGGAGCCAGGATGAAGTAGAGCAAGCT
GGACGCCCTAGTATTCACTCTGTCTACTGAGGCAATTTCTTGATCCCCAGGGAGGCCTTGGG
TCCAGAGACCCTGGAGCGGCTTCTAGAAAAGCAGCAGAGCTGGGAGCAGAGCAGAGTTGGAC
AGCTGTGTAGGGAGCCACAGCTTGCTGCCAAGAAAGCAGCCCTGGTAGCAGGGGTGGTGCGA
CCAGCTGCTGAGGATCTTCCAGAACCTGTGCCAAATTGTGCAGATGTACGAGGGACATTCCC
AGCAGCCTGGTCTGCAACCCAGATTGCAGAGATGGAGCTCTCAGACTTTGAGGACTGCCTGA
CATTATTTGCAGGAGACCCAGGACTTGGGCCTGAGGAACTGCGGGCAGCCATGGGCAAAGCA
AAACAGTTGTGGGGTCCCCCGGGGATTTCTGTCTGAGCAGATCCTGCAGCTTGGTAGGCT
CTTAATAGGTCTAGGAGATCGGGAACCTACAGGAGCTGATCCTAGTGGACTGGGGAGTGCTGA
GCACCCCTGGGGCAGATAGATGGCTGGAGCACCCTCAGCTCCGCATTGTGGTCTCCAGTTTC
CTACGGCAGAGTGGTCGGCATGTGAGCCACCTGGACTTCGTTTCATCTGACAGCGCTGGGTTA
TACTCTCTGTGGACTGCGGCCAGAGGAGCTCCAGCACATCAGCAGTTGGGAGTTCAGCCAAG
CAGCTCTCTTCTCGGCACCCTGCATCTCCAGTGCTCTGAGGAACAACTGGAGGTTCTGGCC
CACCTACTTGTACTGCCTGGTGGGTTTGGCCCAATCAGTAACGGGGGCCTGAGATCTTCAC
TGAAATTGGCACCATAGCAGCTGGGATCCCAGACCTGGCTCTTTCAGCACTGTGCGGGGAC
AGATCCAGGGCGTTACTCCTCTTGCCATTTCTGTCTATCCCTCCTCTAAATTTGCTGTGGTG
TTTAGTCCCATCCAATATCTAGTCTCACCAGTGCTCAGGCTGTGGCTGTCACTCCTGAGCA
AATGGCCTTTCTGAGTCCTGAGCAGCGACGAGCAGTTGCATGGGCCCAACATGAGGGAAAGG
AGAGCCCAGAACAGCAAGGTCAAGTACAGCCTGGGGCCTCCAGGACTGGTCCAGACCTTCC
TGGTCCCTGGTATTGACTATCAGCTTCCTTGGCCACCTGCTA-3' (SEQ ID NO:57).

[0268]

[0269]

STRC 단백질의 류린 C-말단 부분을 코딩하는 부분 코딩 서열은 하기와 같을 수 있거나:

5'-TGCTTCAGTCCTGTACTGTGGGATCTACTCCAGAGAGAGAAGAGCGTTTGGGCCCTGAG
GACCCCTGGTGAAGGCCTACCTGCGCATGCCTCCAGAAGACCTTCAGCAGCTTGTGCTTTCAG
CAGAGATGGAGGCTGCACAGGGCTTCCTGACGCTCATGCTTCGTTCTGGGCTAAGCTGAAG
GTTCAACCATCCGAGGAGCAGGCCATGGGCCGCTGACAGCCTTGCTGCTCCAGCGGTACCC
ACGCCTCACCTCCCACTCTTTATCGACATGTACCGCTCATCCCTTCTTGCTGTCCCTG
ACCTCATGCGCTTCCCACCGTCCCTTTTGGCCAACGACAGTGTCTGGCTGCCATCAGGGAT
CACAGCTCAGGAATGAAGCCTGAACAGAAGGAGGCCCTGGCAAAACGACTGCTGGCCCCTGA

[0270]

GCTGTTTGGAGAAGTGCCTGATTGGCCCCAGGAGCTGCTGTGGGCAGCCCTGCCTCTGCTTC
 CCCATCTGCCTCTGGAGAGCTTTCTCCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGGAT
 AGCTGGCCAGTAGCAGATCTTGGGCCGGGACACGCCCAGCATGTGCTTCGTAGCCTAGTAAA
 CCAGAGCATGGAGGATGGGGAGGAGCAGGTGCTCAGGCTTGGGTCCCTCGCCTGTTTCCTGA
 GTCCTGAGGAGCTACAGAGTCTGGTGCCCTTGAGTGATCCAATGGGGCCTGTAGAACAGGGT
 CTGCTGGAATGTGCGGCCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATATGAACTTCT
 GGGAGTGTGCGTTCATCTGGAGGAACTGTCTTAAGCCCCGAGAGCTGAGGGTCTGGGCAC
 CTCTCTTTCCCCAGCTGGGCTCCGCTTCTGCAGGAGCTCTCAGAGACCCAGCTTAGAGCC
 ATGCTTCTCTGCCCTACAGGGAGCCAGTGTACACCTGCCAGGCTGTTCTGTTGTTTGAAG
 GCTCCTTCCTAAGCATGATCTGTCCCTGGAGGAACTCTGCTCCCTGCACCCTCTCCTGCCAG
 GTCTCAGCCCCCAGACACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGGTGCTTGTTCCTGC
 CTGGGCCCTGAACTGTCAAGGCTTTTCTGAGCTTGCCAGATTGCAGCTCTGCTGCAGACCTTTTCG
 GGTAAGATGGTGTAAAAATATGGGTGCAGCAGGTGCCGGCTCAGCCGTGTGCATTCTTG
 GGCAGCCCACCACTTGCCAGACTGCCTGCTTCCCTGCTCCCATTAAGCTGCTACAGCTG
 GACGCTGCAGCTCTTCTGGCAAACCGAAGACTCTATCGGCAGCTGCCTTGGTCTGAGCAACA
 GGCACAGTTTCTCTGGAAGAAAATGCAAGTGCTTACCAACCTGAGCCTGAGGAATCTGCAGG
 CTCTGGGCAACTTGCCAGGAGGCATGACCTGCGAGTTTCTGCAGCAGATCAGCTCAATGGTT
 GACTTTCTTGATGTGGTACACATGCTCTACCAGCTGCCCACTGGTGTTCGAGAGAGCCTGCG
 GGCCTGTATCTGGACAGAGCTACAGCGGAGGATGACAATGCCAGAGCCAGAGCTGACCACCC
 TAGGGCCAGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG
 GACAGATTGTCCAATGATTCCATTATGTTGGTGGTGGAGATGGTCCAAGGCGCTCCAGAGCA
 GCTGCTGGCACTGACCCCACTCCACCAGACAGCCTTGGCAGAGCGAGCACTAAAAACCTGG
 CTCCAAAGGAGACCCCAATCTCCAAAGAAGTGCTGGAGACACTGGGCCCTTGGTTGGATTCT
 CTGGGAATAGAGAGCACGCGACGGATCCCTTTACCCATTCTACTGTCTCATCTCAGTCAGCT
 GCAGGGCTTCTGCCTAGGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTGTTGCAGGAGC
 CTGTTCTTGGAAAACCAAGAATTGTGGAGCCAGGATGAAATAGAGCAAGCTGGACGCCTAGTA
 TTCCTCTGTCTGCTGAGGCTATTTCTCGATCCCCAGGGAGGCTTTGGGCCAGAGACACT
 GGAGAGGCTTCTGGGAAAGCATCAAAGCTGGGAGCAGAGCAGAGTGGGCCATCTGTGTGGGG
 AGTCACAGCTTGCCACAAGAAAGCAGCTCTGGTAGCTGGGATTGTGCATCCAGCTGCTGAG
 GGTCTCCAAGAGCCTGTACCAAAGTGTGCAGACATACGGGGAACCTTCCAGCGGCCTGGTC
 TGCGACACAAATCTCAGAGATGGAATCTCAGACTTTGAAGACTGCCTGTCACTATTTGCTG
 GAGATCCAGGACTTGGTCTGAGGAATACGGGCAGCCATGGGCAAGGCCAAGCAGTTGTGG
 GGTCCCCCTCGAGGATTCCGTCTGAGCAGATCTTGCAGCTGGGCCGTCTCCTGATAGGTCT
 AGGAGAACGGGAAGTGCAGGAGCTTACCTTGGTGGACTGGGGTGTGCTGAGCAGCCTGGGGC

[0271]

AAATAGATGGCTGGAGTTCCATGCAGCTCCGAGCCGTGGTCTCCAGTTTCCTAAGGCAGAGT
 GGTCGGCATGTGAGCCACCTGGACTTCATTTATCTGACAGCACTGGGTACACAGTCTGTGG
 ATTGCGACCAGAGGAGTTACAGCACATCAGCAGTTGGGAGTTTAGCCAAGCAGCTCTCTTCC
 TGGGTAGCTTGCATCTCCCGTGTCTGAGGAACAGCTGGAAGTTCTGGCCTATCTCCTTGTG
 TTGCCTGGTGGCTTTGGCCAGTCAGTAAGTGGGGCCTGAGATCTTCACTGAAATTGGCAC
 AATAGCAGCTGGCATCCAGACCTGGCTCTTTTACGATTACTGCGGGGACAGATCCAAGGCC
 TGACTCCTCTTGCCATTTCTGTATTCCTGCTCCCAAGTTTGAGTGGTCTTCAACCCCATC
 CAGTTATCTAGTCTCACCAGGGGTGAGGCCGTAGCTGTTACTCTGAACAGCTGGCCTATCT
 GAGTCTGAGCAGCGGCAGCAGTTGCATGGGCCAACACGAAGGAAGGAGATCCAGAGC
 AGCTGGGTGGAAGTGCAGCCTGGGGTCTCTACGACTGGTTCCAAGCCTCCTGGGCCCTGGCA
 TTGCCCCGTCAGCATTTTGGCCACCTATTA-3' (SEQ ID NO:58)

[0272]

또는 STRC 단백질의 C-말단 부
 분을 코딩하는 부분 코딩 서열은 서열식별번호: 57 또는 58로 이루어진 핵산 서열을 포함할 수 있다. 또 다른
 실시양태는 관심 단백질 (예를 들어, STRC; 서열식별번호: 25 또는 26)에 대해 적어도 5% (예를 들어, 6%, 7%,

8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 갖는 관심 단백질 (예를 들어, STRC)의 C-말단 부분 (개시 코돈 ATG에 상응하는 메티오닌, 신호 펩티드 서열; 임의로 링커 서열 및 Myc 태그 서열 포함)을 제공할 수 있다. 예를 들어, STRC 단백질의 C-말단 부분 (개시 코돈 ATG에 상응하는 메티오닌, STRC 신호 펩티드 서열; 임의로 링커 서열 및 Myc 태그 서열 포함)은 서열식별번호: 23 또는 24의 서열에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 포함할 수 있거나, 또는 STRC 단백질의 C-말단 부분은 서열식별번호: 23 또는 24로 이루어진 아미노산 서열을 포함할 수 있다. 제2 벡터의 제2 뉴클레오티드 서열은 자신의 신호 서열 (예를 들어, 서열식별번호: 10 또는 12) 또는 대안적으로 상이한 신호 서열을 포함하는 관심 단백질의 C-말단 부분 (예를 들어, STRC; 서열식별번호: 23 또는 24)을 코딩하는 부분 코딩 서열 및 스플라이스 수용자 서열 (예를 들어, C-말단 인테인 (C-인테인); 서열식별번호: 22)을 코딩하는 서열을 포함할 수 있다.

[0273] 한 실시양태는 서열식별번호: 42에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함하는 N-인테인 서열 또는 서열식별번호: 45에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 코딩하는 N-인테인 서열을 제공할 수 있다. 다른 실시양태는 서열식별번호: 42의 핵산 서열로 이루어진 N-인테인 서열 또는 서열식별번호: 45로 이루어진 아미노산 서열을 코딩하는 N-인테인 서열에 관한 것일 수 있다.

[0274] 또 다른 실시양태는 서열식별번호: 46에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함하는 C-인테인 서열 또는 서열식별번호: 49에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 코딩하는 C-인테인 서열을 제공한다. 다른 실시양태는 서열식별번호: 46의 핵산 서열로 이루어진 C-인테인 서열 또는 서열식별번호: 49로 이루어진 아미노산 서열을 코딩하는 C-인테인 서열에 관한 것일 수 있다.

[0275] 추가 실시양태에서, 본 개시내용의 이중-벡터 시스템은 관심 단백질의 N-말단 부분 (예를 들어, STRC; 서열식별번호: 15 또는 16)을 코딩하는 제1 뉴클레오티드 서열을 제공하며, 여기서 제1 뉴클레오티드 서열은 관심 단백질 (예를 들어, STRC)의 N-말단 부분의 부분 코딩 서열 (5')의 일부를 형성하거나 또는 그와 별개일 수 있는 관심 단백질의 신호 서열, 및 관심 단백질 (예를 들어, STRC)의 N-말단 부분의 부분 코딩 서열의 일부를 형성하거나 또는 그와 별개일 수 있는 스플라이스 공여자 서열 (예를 들어, N-인테인 서열)을 포함하나 이에 제한되지는 않고, 여기서 제1 뉴클레오티드 서열은 관심 단백질 (예를 들어, STRC)의 N-말단 부분을 코딩하는 뉴클레오티드 서열에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함할 수 있고, 관심 단백질의 내인성 신호 서열 또는 외인성 신호 서열, 관심 단백질의 N-말단 부분의 부분 코딩 서열 (5') 및 스플라이스 공여자 서열 (예를 들어, N-인테인 서열)을 포함하나 이에 제한되지는 않는다 (서열식별번호: 5 또는 7). 한 실시양태는, 예를 들어 서열식별번호: 5 또는 7에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함하는 제1 뉴클레오티드 서열을 제공할 수 있거나, 또는 제1 뉴클레오티드 서열은 서열식별번호: 5 또는 7로 이루어진 핵산 서열을 포함할 수 있다.

[0276] 또 다른 실시양태는 관심 단백질 (예를 들어, STRC)의 N-말단 부분을 포함하는 서열에 대해 적어도 80% (예를 들어, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 코딩하는 제1 뉴클레오티드 서열을 제공하며, 여기서 제1 뉴클레오티드 서열은 관심 단백질 (예를 들어, STRC)의 N-말단 부분의 부분 코딩 서열 (5')의 일부를 형성하거나 또는 그와 별개일 수 있는 관심 단백질의 내인성 신호 서열 또는 외인성 신호 서열, 및 관심 단백질 (예를 들어, STRC)의 N-말단 부분의 부분 코딩 서열의 일부를 형성하거나 또는 그와 별개일 수 있는 스플라이스 공여자 서열 (예를 들어, N-인테인 서열)을 포함하나 이에 제한되지는 않는다. 추가 실시양태는, 예를 들어 서열식별번호: 6 또는 8에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 코딩하는 제1 뉴클레오티드 서열을 제공할 수 있거나, 또는 제1 뉴클레오티드 서열은 서열식별번호: 6 또는 8로 이루어진 아미노산 서열을 코딩할 수 있다.

[0277] 본 개시내용의 이중-벡터 시스템의 추가 실시양태는 관심 단백질의 C-말단 부분 (예를 들어, STRC; 서열식별번호: 23 또는 24)을 코딩하는 제2 뉴클레오티드 서열을 제공하며, 여기서 제2 뉴클레오티드 서열은 관심 단백질 (예를 들어, STRC)의 C-말단 부분의 부분 코딩 서열 (3')의 일부를 형성하거나 또는 그와 별개일 수 있는 관심 단백질의 내인성 신호 서열 또는 외인성 신호 서열, 및 관심 단백질 (예를 들어, STRC)의 C-말단 부분의 부분 코딩 서열의 일부를 형성하거나 또는 그와 별개일 수 있는 스플라이스 수용자 서열 (예를 들어, C-인테인 서

열)을 포함하나 이에 제한되지는 않고, (여기서 제2 뉴클레오타이드 서열은 일부 실시양태에서 링커 서열 및 Myc-태그 서열을 임의로 포함할 수 있음), 여기서 제2 뉴클레오타이드 서열은 관심 단백질 (예를 들어, STRC)의 C-말단 부분을 코딩하는 뉴클레오타이드 서열에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함할 수 있고, 관심 단백질의 내인성 신호 서열 또는 외인성 신호 서열, 관심 단백질의 C-말단 부분의 부분 코딩 서열 (3') 및 스플라이스 수용자 서열 (예를 들어, C-인테인 서열 (및 임의로 링커 서열 및 Myc-태그 서열을 포함함))을 포함하나 이에 제한되지는 않는다 (서열식별번호: 17 또는 19). 한 실시양태는, 예를 들어 서열식별번호: 17 또는 19에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 핵산 서열을 포함하는 제2 뉴클레오타이드 서열을 제공할 수 있거나, 또는 제2 뉴클레오타이드 서열은 서열식별번호: 17 또는 19로 이루어진 핵산 서열을 포함할 수 있다.

[0278] 또 다른 실시양태는 관심 단백질의 신호 서열, C-인테인 서열 및 관심 단백질의 C-말단 부분의 부분 코딩 서열 (3') (및 임의로 링커 서열 및 Myc-태그 서열을 포함함)을 포함하나 이에 제한되지는 않는 관심 단백질 (예를 들어, STRC)의 C-말단 부분을 포함하는 서열에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 코딩하는 제2 뉴클레오타이드 서열을 제공한다. 추가 실시양태는 서열식별번호: 18 또는 20에 대해 적어도 5% (예를 들어, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%) 동일성의 아미노산 서열을 코딩하는 제2 뉴클레오타이드 서열을 제공할 수 있거나, 또는 제2 뉴클레오타이드 서열은 서열식별번호: 18 또는 20으로 이루어진 아미노산 서열을 코딩할 수 있다.

[0279] 추가 실시양태에서, 본 개시내용의 이중-벡터 시스템은 5'에서 3' 방향으로, ITR, 프로모터 (예를 들어, CMV 프로모터), 관심 부분 코딩 서열 (예를 들어, 5' Strc), 스플라이스 공여자 서열 (예를 들어, 5' 인테인) 및 ITR을 갖는 제1 뉴클레오타이드 서열 (예를 들어, AAV 벡터 1); 및 ITR, 프로모터 (예를 들어, CMV 프로모터), 스플라이스 수용자 서열 (예를 들어, 3' 인테인), 관심 부분 코딩 서열 (예를 들어, 3' Strc) 및 ITR을 갖는 제2 뉴클레오타이드 서열 (예를 들어, AAV 벡터 2)을 제공한다 (도 31a). 번역을 겪으면, 제1 및 제2 뉴클레오타이드 서열은 다수의 변이체를 생성할 수 있으며, 이들은 서열식별번호: 26의 위치 747 및 970에서의 천연 시스테인, 즉 Cys747 및 Cys970 (또는 서열식별번호: 25의 위치 709 및 934에서의 천연 시스테인, 즉 Cys709 및 Cys934)을 이용하여 스플라이싱되는 경우에, 전장 STRC 단백질 및 n-인테인 및 c-인테인을 포함하는 절제된 인테인을 형성한다. 도 31a는 4개의 상이한 이중 벡터 변이체를 포함하는 8개의 AAV2 플라스미드를 예시한다.

[0280] 예를 들어, 변이체 1은 N-말단에서 C-말단 방향으로, 신호 펩티드 서열, 관심 단백질의 N-말단 부분 (예를 들어, ~80 kD의 n-STRC), 스플라이스 부위 (예를 들어, Ser746) 및 스플라이스 공여자 서열 (n-인테인)을 포함하는 제1 단백질 서열; 및 N-말단에서 C-말단 방향으로, 스플라이스 수용자 서열 (c-인테인), 스플라이스 부위 (예를 들어, Cys747) 및 관심 단백질의 C-말단 부분 (예를 들어, ~117 kD의 c-STRC)을 포함하는 제2 단백질 서열을 포함한다. 변이체 2는, 예를 들어 N-말단에서 C-말단 방향으로, 신호 펩티드 서열, 관심 단백질의 N-말단 부분 (예를 들어, ~105 kD의 n-STRC), 스플라이스 부위 (예를 들어, Ala969) 및 스플라이스 공여자 서열 (n-인테인)을 포함하는 제1 단백질 서열; 및 N-말단에서 C-말단 방향으로, 스플라이스 수용자 서열 (c-인테인), 스플라이스 부위 (예를 들어, Cys970) 및 관심 단백질의 C-말단 부분 (예를 들어, ~92 kD의 c-STRC)을 포함하는 제2 단백질 서열을 포함한다. 예를 들어, 변이체 3은 N-말단에서 C-말단 방향으로, 신호 펩티드 서열, 관심 단백질의 N-말단 부분 (예를 들어, ~80 kD의 n-STRC), 스플라이스 부위 (예를 들어, Ser746) 및 스플라이스 공여자 서열 (n-인테인)을 포함하는 제1 단백질 서열; 및 N-말단에서 C-말단 방향으로, 신호 펩티드 서열, 스플라이스 수용자 서열 (c-인테인), 스플라이스 부위 (예를 들어, Cys747) 및 관심 단백질의 C-말단 부분 (예를 들어, ~117 kD의 c-STRC)을 포함하는 제2 단백질 서열을 포함한다. 변이체 4는, 예를 들어 N-말단에서 C-말단 방향으로, 신호 펩티드 서열, 관심 단백질의 N-말단 부분 (예를 들어, ~105 kD의 n-STRC), 스플라이스 부위 (예를 들어, Ala969) 및 스플라이스 공여자 서열 (n-인테인)을 포함하는 제1 단백질 서열; 및 N-말단에서 C-말단 방향으로, 신호 펩티드 서열, 스플라이스 수용자 서열 (c-인테인), 스플라이스 부위 (예를 들어, Cys970) 및 관심 단백질의 C-말단 부분 (예를 들어, ~92 kD의 c-STRC)을 포함하는 제2 단백질 서열을 포함한다.

[0281] 번역된 변이체 서열의 단백질 스플라이싱 시, 전장 관심 단백질 (예를 들어, STRC)이 형성된다. 예를 들어, N-말단에서 C-말단 방향으로, 관심 단백질 (n-STRC)의 N-말단 부분은 관심 단백질 (예를 들어, c-STRC)의 C-말단 부분에 연결될 수 있다. N-말단 및 C-말단 단부의 스플라이싱은 스플라이스 공여자 서열 및 스플라이스 수용자 서열의 절제를 유발한다. 예를 들어, 절제된 스플라이스 서열은 전장 스플라이스 서열 (예를 들어, n-인테인 및 c-인테인의 절제된 인테인)을 형성할 수 있다 (도 31a).

- [0282] 도 31c는 변이체 3의 N-말단 부분 및 C-말단 부분 (c+n) 둘 다로 형질감염된 HEK 세포가, 단지 C-말단 부분 (c)만을 갖는 변이체 3 또는 C-말단 부분 단독 (c) 또는 이와 함께 N-말단 부분 (c+n)을 갖는 변이체 1과는 대조적으로, 전장 STRC의 발현을 일으켰다는 것을 입증한다. 변이체 1과 변이체 3 사이의 유일한 차이는 C-말단 부분에 신호 서열이 존재한다는 것이다. 따라서, 도 31c는 N-말단 부분 및 C-말단 부분 둘 다에서의 신호 서열이 각각의 부분을 세포의 동일한 세포 구획으로 지시하여, 단백질 스플라이싱이 전장 STRC를 형성할 수 있게 한다는 것을 입증한다.
- [0283] 추가 실시양태는 제1 벡터 및 제2 벡터를 포함하는, 본원에 기재된 이중-벡터 시스템을 함유하는 세포 (예를 들어, 숙주 세포, 포유동물 세포, 인간 세포, 박테리아 세포)를 제공한다. 한 실시양태에서, 세포는 내이 세포, 내유모 세포 또는 외유모 세포일 수 있다. 일부 실시양태는 포유동물 세포 (예를 들어, 인간, 개, 고양이, 말, 뮌)인 세포에 관한 것일 수 있다. 다른 실시양태는 귀 세포 (예를 들어, 내이 세포, 외이 세포, 내유모 세포, 외유모 세포)를 제공할 수 있다. 본 개시내용의 세포는 생체내 또는 시험관내일 수 있다. 일부 실시양태에서, 세포는 관련 기술분야에 일반적으로 공지된 임의의 다수의 공지된 형질감염 및 형질전환 기술을 사용하여 본 개시내용의 이중-벡터 시스템의 제1 벡터 및 제2 벡터로 형질감염 또는 형질전환될 수 있다. 예를 들어, 문헌 [Graham et al. (1973) Virology, 52:456, Sambrook et al. (1989) Molecular Cloning, a laboratory manual, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) Basic Methods in Molecular Biology, Elsevier, and Chu et al. (1981) Gene 13:197]을 참조하며, 이들 모두는 그 전문이 본원에 참조로 포함된다. 본 개시내용의 이중-벡터 시스템의 제1 벡터 및 제2 벡터는 바이러스 형질도입, 염화칼슘을 사용한 박테리아 형질전환, 박테리아 교배 또는 접합에 의한 박테리아 형질전환 또는 형질도입, 형질감염 (예를 들어, 전기천공, 인산칼슘, 리포솜-기반 형질감염), 유전자 총 등을 포함하나 이에 제한되지는 않는 임의의 수단에 의해 본원에 기재된 세포(들) 내로 삽입될 수 있다.
- [0284] 또 다른 실시양태는 본원에 기재된 이중-벡터 시스템 및 제약상- 또는 생리학상-허용되는 비히클 (예를 들어, 희석제, 담체, 부형제)을 포함하는 조성물 또는 제약 조성물에 관한 것일 수 있다. 돌연변이와 연관된 질환 또는 상태의 치료를 위해 본원에서 고려되는 조성물은 본원에 기재된 이중-벡터 시스템을 포함할 수 있다. 치료 목적을 위해, 본원에 개시된 방법에 따라 적절하게 프로세싱되는 경우에, 본원에 기재된 바와 같은 질환 또는 상태 (예를 들어, 상염색체 열성 DFNB16 청력 상실)를 유발하거나 또는 이에 기여하는 돌연변이를 포함하는 게놈에서 전장 관심 단백질 (예를 들어, STRC 단백질)의 발현을 일으키는 관심 폴리뉴클레오티드 (예를 들어, STRC) 또는 그의 단편을 포함하는 조성물이 질환 또는 상태에 의해 영향을 받는 신체의 영역 (예를 들어, 와우, 내이)에 직접 투여될 수 있다. 일부 실시양태에서, 조성물은 제약상 허용되는 완충제, 예컨대 생리 염수 중에 제제화된다. 비제한적 투여 방법은 귀, 내이, 와우관 또는 와우관 주위의 외림프-충전된 공간 (예를 들어, 고실계 및 전정계) 내로 주사하는 것을 포함한다. 고칼륨 내림프액으로 충전된 와우관 내로의 주사는 유모 세포에 대한 직접적 접근을 제공할 수 있다. 그러나, 이러한 섬세한 유체 환경에 대한 변경은 와우내 잠재력을 파괴하여, 주사-관련 독성에 대한 위험을 높일 수 있다. 와우관을 둘러싸는 외림프-충전된 공간, 고실계 및 전정계는 난원창 또는 원형창 막을 통해 중이로부터 접근될 수 있다. 내이 내로의 유일한 비-골성 개구부인 원형창 막은 많은 동물 모델에서 비교적 용이하게 접근가능하고, 이 경로를 사용한 바이러스 벡터의 투여는 잘 허용된다. 인간에서, 와우 이식물 배치는 상용적으로 원형창 막을 통한 외과적 전극 삽입에 의존한다.
- [0285] 사용 방법
- [0286] 한 실시양태는 치료를 필요로 하는 대상체에게 유효량의 본원에 기재된 이중-벡터 시스템을 투여하는 것을 포함하는, 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템; 캡시드, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 계통, 렌티바이러스)을 사용하는 방법을 제공할 수 있으며, 여기서 방법은 결합성 또는 돌연변이된 유전자로부터 유발된 질환, 상태 또는 그의 증상을 치료 및/또는 감소 및/또는 예방할 수 있다.
- [0287] 또 다른 실시양태는 (예를 들어, 대상체의) 세포를, 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템) 및 제약상- 또는 생리학상-허용되는 비히클 (예를 들어, 담체, 희석제, 부형제)을 포함하는 조성물과 접촉시키는 것을 포함하는 방법에 관한 것일 수 있다. 세포 (예를 들어, 대상체의 세포)와의 접촉 단계는, 예를 들어 서열식별번호: 33 (도 2a-2c) 또는 서열식별번호: 38 (도 4a-4d)을 포함하는 벡터 (예를 들어, 플라스미드, 트랜스스플라이싱 플라스미드, 바이러스 벡터, 아데노바이러스, AAV, AAV 계통)의 뉴클레오티드 서열의 전달을 일으킬 수 있으며, 여기서 세포는 전장 관심 단백질 (예를 들어, STRC; 인간 서열식별번호: 2 또는 25; 뮌 서열식별번호: 4 또는 26)을 발현할 수 있다. 대상체의 세포와의 접촉 단계는 (각각 제1 벡터 및 제2 벡터의) 제1 뉴클레오티드 서열 및 제2 뉴클레오티드 서열의 전달을 일으킬 수 있으며, 여기서 세포는 관심 단백질 (예를 들어,

어, STRC)의 N-말단 부분 및 관심 단백질의 C-말단 부분을 발현할 수 있고, 관심 단백질 (예를 들어, STRC)의 N-말단 부분 및 관심 단백질의 C-말단 부분은 펩티드 결합에 의해 연결되어 전장 관심 단백질 (예를 들어, STRC; 인간 서열식별번호: 2 또는 25; 뮤린 서열식별번호: 4 또는 26)을 형성한다.

[0288] 한 실시양태는 상염색체 열성 청력 상실의 치료를 필요로 하는 대상체에게 유효량의 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템), 또는 본원에 기재된 조성물, 또는 본 개시내용의 벡터 시스템 (예를 들어, 이중-벡터 시스템) 또는 본 개시내용의 조성물을 함유하는 세포, 또는 관심 단백질 (예를 들어, STRC; 서열식별번호: 25; 서열식별번호: 26)을 코딩하는 적어도 1개의 뉴클레오타이드 서열 (예를 들어, STRC; 서열식별번호: 5 서열식별번호: 7; 서열식별번호: 17; 서열식별번호: 19; 서열식별번호: 30; 서열식별번호: 32) 또는 관심 단백질 (예를 들어, STRC; 서열식별번호: 25; 서열식별번호: 26) 자체 또는 관심 단백질의 부분 (예를 들어, 서열식별번호: 6 - 서열식별번호: 16; 서열식별번호: 18 - 서열식별번호: 24)을 포함하는 벡터 또는 조성물을 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료하는 방법을 제공할 수 있으며, 여기서 투여는 상염색체 열성 청력 상실 또는 그의 증상의 감소 또는 회복을 가져올 수 있다. 일부 실시양태에서, 치료를 필요로 하는 대상체는 치료 후 대상체가 69 dB 이하 (예를 들어, 60 dB, 55 dB, 50 dB, 45 dB, 40 dB, 35 dB, 30 dB, 26 dB, 25 dB, 20 dB, 15 dB, 10 dB, 5 dB, 0 dB)의 청력 수준을 갖는 경우에 성공적으로 치료되었을 것이다. 일반적으로, 극심한 청력 상실을 갖는 대상체는 95 dB 미만의 소리를 들을 수 없고; 중증 청력 상실을 갖는 대상체는 70 dB 내지 94 dB 미만의 소리를 들을 수 없고; 중등도 청력 상실 대상체는 40 dB 내지 69 dB 미만의 소리를 들을 수 없고; 경도 청력 상실을 갖는 대상체는 26 dB 내지 40 dB 미만의 소리를 들을 수 없다. 그러나, 예를 들어 상염색체 열성 청력 상실을 포함한 청력 상실을 앓고 있는 대상체는 본원에 기재된 임의의 방법에 의해 치료될 수 있으며, 따라서 청력 상실 또는 그의 증상의 감소 및/또는 청력 (또는 대상체에서의 청각 기능)의 회복 또는 개선 및/또는 청력의 유지를 유발할 수 있다. 정상 청력을 갖는 대상체는 25 dB 이하 (예를 들어, 20 dB, 15 dB, 10 dB, 5 dB, 0 dB)의 청력 소리를 특징으로 할 수 있다. 일부 실시양태에서, 상염색체 열성 청력 상실은 DFNB16이다.

[0289] 추가 실시양태는 청력 상실을 특징으로 하는 병리상태 또는 질환의 치료 및/또는 예방을 필요로 하는 대상체에게 유효량의 본원에 기재된 이중-벡터 시스템, 본 개시내용에 따른 세포 또는 본원에 기재된 조성물 또는 제약 조성물을 투여하는 것을 포함하는, 청력 상실을 특징으로 하는 병리상태 또는 질환을 치료 및/또는 예방하는 방법을 제공할 수 있다. 본 개시내용의 세포는 내이 세포, 내유모 세포, 외유모 세포, 생체내 세포, 시험관내 세포 등 또는 상기 중 임의의 것의 조합일 수 있다.

[0290] 폴리뉴클레오타이드 전달

[0291] 이들 접근법에서의 치료 성공은 와우 내 코르티 기관에서의 관련 치료 세포 표적으로의 외인성 유전자 구축물의 안전하고 효율적인 전달에 유의하게 의존한다. 코르티 기관은 2가지 부류의 감각 유모 세포를 포함한다: 소리에 의해 운반되는 기계적 정보를 뉴런 구조로 전송되는 전기적 신호로 전환시키는 내유모 세포 및 복잡한 청각 기능에 요구되는 과정인 와우 반응을 증폭시키고 조정하는 역할을 하는 외유모 세포.

[0292] 핵산을 세포로 전달하는 방법은 일반적으로 관련 기술분야에 공지되어 있고, 트랜스젠을 함유하는 바이러스 (이는 또한 바이러스 입자로도 지칭될 수 있음)를 생체내 내이 세포로 전달하는 방법은 본원에 기재되어 있다. 본원에 기재된 바와 같이, 약 10^8 내지 약 10^{12} 개 바이러스 입자가 대상체에게 투여될 수 있고, 바이러스는 적합한 부피 (예를 들어, 10 μ L, 50 μ L, 100 μ L, 500 μ L 또는 1000 μ L)의, 예를 들어 인공 외림프 용액 중에 현탁될 수 있다.

[0293] 본원에 기재된 바와 같은 역전된 말단 반복부 (ITR), 프로모터 (예를 들어, 에스핀 프로모터, PCDH15 프로모터, PTPRQ 프로모터, Myo6 프로모터, KCNQ4 프로모터, Myo7a 프로모터, 시냅신 프로모터, GFAP 프로모터, CMV 프로모터, CAG 프로모터, CBH 프로모터, CBA 프로모터, U6 프로모터 및 TMHS (LHFPL5) 프로모터), 신호 서열, 관심 단백질 (예를 들어, STRC 단백질)을 코딩하는 폴리뉴클레오타이드 및 폴리아데닐화 (폴리A) 서열, 및 일부 실시양태에서, c-myc 태그를 연결하기 위한 링커 서열을 함유하는 바이러스는 임의의 수의 수단을 사용하여 내이 세포 (예를 들어, 와우 내의 세포)에 전달될 수 있다. 예를 들어, 본원에 기재된 바와 같은 이중-벡터 인테인-매개 단백질 트랜스-스플라이싱 시스템을 함유하는 바이러스 입자를 포함하는 조성물의 치료 유효량은 전형적으로 비교적 간단한 (예를 들어, 외래환자) 절차로 원형창 또는 난원창 또는 난형창을 통해 주사될 수 있다. 일부 실시양태에서, 본원에 기재된 바와 같은 이중-벡터 인테인-매개 단백질 트랜스-스플라이싱 시스템 (예를 들어, 이중-AAV 인테인-매개 STRC 단백질 시스템)을 함유하거나 또는 상이한 바이러스 입자의 1개 이상의 세트를 함유하는 바이러스 입자의 치료 유효량을 포함하는 조성물은 수술 (예를 들어, 와우개창술 또는 반고리관절개술) 동안

귀 안의 적절한 위치로 전달될 수 있다.

- [0294] 추가로, 고막을 가로지른 및/또는 원형창 또는 난형낭을 통한 작용제의 전달을 용이하게 하는 전달 비히클 (예를 들어, 중합체)이 이용가능하고, 임의의 이러한 전달 비히클은 본원에 기재된 바이러스를 전달하는 데 사용될 수 있다. 예를 들어, 전달 비히클에 대해 문헌 [Arnold et al., 2005, *Audiol. Neurotol.*, 10:53-63] (그 전문이 본원에 참조로 포함됨)을 참조한다.
- [0295] 본원에 기재된 조성물 및 방법은 내이 세포, 예를 들어 와우 세포로의 핵산의 고도로 효율적인 전달을 가능하게 한다. 예를 들어, 관심 단백질 (예를 들어, STRC 단백질) 또는 그의 단편을 코딩하는 폴리뉴클레오티드는 바이러스 벡터 내로 클로닝될 수 있고, 발현은 그의 내인성 프로모터로부터, 바이러스 역전된 말단 반복부로부터 또는 관심 표적 세포 유형에 특이적인 프로모터로부터 구동될 수 있다. 사용될 수 있는 다른 바이러스 벡터는, 예를 들어 백시니아 바이러스, 소 유두종 바이러스 또는 포진 바이러스, 예컨대 엡스타인-바르 바이러스를 포함한다. 바이러스 벡터는 임상 환경에서 사용되었다. 일부 실시양태에서, 바이러스 벡터 (예를 들어, rAAV)는 큰 (예를 들어, Strc 유전자) 폴리뉴클레오티드를 단편들로 투여하는 데 사용될 수 있다. 일부 실시양태에서, 바이러스 벡터는 신체의 특정 영역에 Strc 폴리뉴클레오티드 단편을 투여하는 데 사용될 수 있다.
- [0296] 예를 들어, 본원에 기재된 조성물 및 방법은 내유모 및/또는 외유모 세포의 적어도 65% 이상 (예를 들어, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%)으로의 관심 폴리뉴클레오티드 (예를 들어, Strc)의 전달 및 그에서의 발현, 또는 외유모 세포의 적어도 65% 이상 (예를 들어, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%)으로의 전달 및 그에서의 발현을 가능하게 한다.
- [0297] 본원에 기재된 이중-벡터 인테인-매개 시스템을 사용하여 전달된 STRC 폴리뉴클레오티드의 발현은 내유모 및 외유모 세포의 개선된 구조 및 기능을 가져와, 청력이 연장된 기간 (예를 들어, 수일, 수주, 수개월, 수년, 수십 년, 평생) 동안 회복되도록 할 수 있다. 한 실시양태에서, 청력 상실은 STRC 유전자 내의 돌연변이에 의해 유발된 상염색체 열성 유형의 비-중후군성 난청인 DFNB16을 앓고 있는 대상체에서 회복될 수 있다. STRC 및 스테레오실린 (STRC) 단백질의 정상 발현은 청각 기능에 필수적이다.
- [0298] 본원에 기재된 바와 같이, 아데노-연관 바이러스 (AAV)는 핵산 (예를 들어, STRC 폴리펩티드를 코딩하는 폴리뉴클레오티드)을 내이 세포로 전달하는 데 특히 효율적이다. Anc80 벡터는 60% 이상 (예를 들어, 70%, 80%, 90%, 95%, 100%)의 내유모 또는 외유모 세포를 유리하게 형질도입시킨 내이 유모 세포 표적화 AAV의 예이다. 한 실시양태는 Anc80 조상 캡시드 단백질의 부류 내에 속하는 조상 캡시드 단백질, 예를 들어 국제 공개 번호 WO 2018/145111 (PCT/US2018/017104)에 기재된 Anc80-0065 (이는 Anc80에 관하여 그 전문이 본원에 참조로 포함됨)를 이용할 수 있다. WO 2015/054653 (이는 또한 그 전문이 본원에 참조로 포함됨)에는 Anc80 조상 캡시드 단백질의 부류 내에 속하는 다수의 추가의 조상 캡시드 단백질이 기재되어 있다.
- [0299] 특정한 실시양태에서, 아데노-연관 바이러스 (AAV)는 유모 세포에 대한 천연 또는 조작된 향성을 갖는 조상 AAV 캡시드 단백질을 함유한다. 일부 실시양태에서, 바이러스는 관심 폴리펩티드 (예를 들어, STRC 단백질)를 코딩하는 관심 폴리뉴클레오티드를 대상체 (예를 들어, DFNB16 및/또는 STRC 유전자 내의 돌연변이를 앓고 있는 대상체)의 내이로 전달하는 내이 유모 세포 표적화 AAV이다. 일부 실시양태에서, 바이러스는 정제된 캡시드 폴리펩티드를 포함하는 AAV이다. 일부 실시양태에서, 바이러스는 인공이다. 일부 실시양태에서, 바이러스는 AAV2 보다 더 낮은 혈청유병률을 갖는 AAV이다. 일부 실시양태에서, 바이러스는 엑솜-연관 AAV이다. 일부 실시양태에서, 바이러스는 엑솜-연관 AAV1이다. 일부 실시양태에서, 바이러스는 Anc80 캡시드 단백질에 대해 적어도 95% 아미노산 서열 동일성 또는 상동성을 갖는 캡시드 단백질을 포함한다.
- [0300] 관심 폴리뉴클레오티드 (예를 들어, STRC)의 발현은 이중 프로모터 (예를 들어, CMV 프로모터, Espin 프로모터, PCDH15 프로모터, PTPRQ 프로모터, TMHS (LHFPL5) 프로모터)에 의해 지시될 수 있다. 본원에 사용된 "이중 프로모터"는 그 서열의 발현을 자연적으로 지시하지 않는 (즉, 자연에서 그 서열과 함께 발견되지 않는) 프로모터를 지칭한다.
- [0301] 트랜스진을, 예를 들어 Anc80 캡시드 단백질을 함유하는 바이러스 내로 패키징하는 방법은 관련 기술분야에 공지되어 있고, 통상적인 분자 생물학 및 재조합 핵산 기술을 이용한다. 한 실시양태에서, Anc80 캡시드 단백질을 코딩하는 핵산 서열을 포함하는 구축물, 및 Anc80 캡시드 단백질 내에 패키징되도록 하는 적합한 역전된 말단 반복부 (ITR)가 플랜킹된 STRC 단백질의 N-말단 및 C-말단 부분을 코딩하는 폴리뉴클레오티드의 단편을 보유하는 구축물이 제공된다.
- [0302] 관심 폴리뉴클레오티드 (예를 들어, STRC)는, 예를 들어 패키징 숙주 세포를 사용하여 Anc80 캡시드 단백질을

함유하는 AAV 내로 패키징될 수 있다. 바이러스 입자의 성분 (예를 들어, rep 서열, cap 서열, 역전된 말단 반복부 (ITR) 서열)은 본원에 기재된 바와 같은 1개 이상의 구축물을 사용하여 패키징 숙주 세포 내로 일시적으로 또는 안정하게 도입될 수 있다. 관심 폴리뉴클레오티드는 일반적으로 큰 유전자 (예를 들어, 4 kB 이상)일 수 있으며, 이는 1개 초과 AAV 내로 분할 및 패키징되는 것이 요구될 수 있다.

[0303] 일부 실시양태에서, AAV9-php.b 벡터를 함유하는 AAV는 내이 세포를 효율적으로 표적화하는 데 사용될 수 있다. AAV9-php.b는 국제 공개 번호 WO 2019/173367 (PCT/US2019/020794)에 기재되어 있으며, 이의 내용은 그 전문이 본원에 참조로 포함된다. AAV-PHP.B는 7-량체 서열 TLAVPFK (서열식별번호: 59)를 코딩하고, 트랜스진을 와우로 효율적으로 전달하며, 여기서 이는 내유모 및 외유모 세포에서 현저하게 특이적이고 강건한 발현을 나타냈다. AAV-PHP.B 벡터는 본원에 기재된 프로모터 중 임의의 것을 포함할 수 있지만 이에 제한되지는 않는다.

[0304] 폴리뉴클레오티드 요법 방법에 사용하기 위한 cDNA 발현은 임의의 적합한 프로모터 (예를 들어, 인간 시토크로마 바이러스 (CMV), 원숭이 바이러스 40 (SV40) 또는 메탈로티오네인 프로모터)로부터 지시되고, 임의의 적절한 포유동물 조절 요소에 의해 조절될 수 있다. 예를 들어, 원하는 경우에, 특정 세포 유형에서 유전자 발현을 우선적으로 지시하는 것으로 공지된 인핸서가 핵산의 발현을 지시하는 데 사용될 수 있다. 사용되는 인핸서는 비제한적으로 조직- 또는 세포-특이적 인핸서로서 특징화되는 것을 포함할 수 있다. 대안적으로, 게놈 클론이 치료 구축물로서 사용되는 경우에, 조절은 동족 조절 서열에 의해 매개될 수 있거나, 또는 원하는 경우에, 상기 기재된 임의의 프로모터 또는 조절 요소를 포함한, 이중 공급원으로부터 유래된 조절 서열에 의해 매개될 수 있다.

[0305] 치료 방법

[0306] 본 개시내용에 포함된 또 다른 치료 접근법은 재조합 치료제 (예를 들어, 재조합 STRC 단백질, 그의 변이체 또는 단편)를 잠재적 또는 실제 질환-이환 조직의 부위에 직접적으로 또는 전신으로 (예를 들어, 임의의 통상적인 재조합 단백질 투여 기술에 의해) 투여하는 것을 수반할 수 있다. 투여되는 단백질의 투여량은 개별 환자의 크기 및 건강을 포함한 다수의 인자에 좌우된다. 임의의 특정한 대상체의 경우, 구체적 투여 요법은 개체 필요에 따라 및 조성물을 투여하거나 투여를 감독하는 사람의 전문적 판단에 따라 시간 경과에 따라 조정되어야 한다.

[0307] 본 개시내용의 일부 실시양태는 세포 (예를 들어, 대상체의 세포) (여기서 세포의 게놈은 돌연변이를 포함할 수 있음)에서 전장 관심 단백질 (예를 들어, STRC)을 코딩하는 뉴클레오티드 서열을 함유하는 벡터 시스템 (예를 들어, 이중-벡터 인테인-매개 시스템)을 포함하는 제약 조성물의 치료 유효량을 투여하는 것을 포함하는, 질환 및/또는 장애 또는 그의 증상의 치료 또는 예방 또는 감소를 필요로 하는 대상체에서 질환 및/또는 장애 또는 그의 증상을 치료 또는 예방 또는 감소시키는 방법을 제공할 수 있다. 본 개시내용의 이중-벡터 인테인-매개 시스템을 사용하는 실시양태에서, 질환 및/또는 장애 또는 그의 증상의 치료 또는 예방 또는 감소를 필요로 하는 대상체에서 질환 및/또는 장애 또는 그의 증상을 치료 또는 예방 또는 감소시키는 방법은, 그를 필요로 하는 대상체 (여기서 대상체는, 예를 들어 포유동물, 예컨대 인간임)에서 돌연변이를 포함하는 게놈에 관심 단백질의 부분 (예를 들어, N-STRC)을 코딩하는 제1 뉴클레오티드 서열 및 관심 단백질의 나머지 부분 (예를 들어, C-STRC)을 코딩하는 제2 뉴클레오티드 서열을 함유하는 이중-벡터 인테인-매개 시스템을 포함하는 제약 조성물의 치료 유효량을 투여하는 것을 포함할 수 있다. 따라서, 한 실시양태는 돌연변이와 연관된 질환 또는 장애 또는 그의 증상을 앓고 있거나 그에 걸리기 쉬운 대상체를 치료하는 방법이다. 방법은 대상체에게 질환 또는 장애 또는 증상을 치료 또는 예방 또는 감소시키기에 충분한 치료량의 본원의 조성물을 투여하는 것을 포함한다. 일부 실시양태에서, 돌연변이는 열성 돌연변이이다.

[0308] 본 발명의 치료 방법 (이는 예방적 치료를 포함함)은 일반적으로 그를 필요로 하는 포유동물, 예를 들어 인간을 포함한 대상체 (예를 들어, 동물, 인간)에게 치료 유효량의 본원의 화합물 또는 조성물, 예컨대 본원의 화학식의 화합물을 투여하는 것을 포함한다. 이러한 치료는 질환, 장애 또는 그의 증상을 앓고 있거나, 그를 갖거나, 그에 걸리기 쉽거나 또는 그의 위험이 있는 대상체, 특히 인간에게 적합하게 투여될 것이다. "위험이 있는" 대상체의 결정은 대상체 또는 건강 관리 제공자의 진단 검사 또는 의견 (예를 들어, 유전자 검사, 효소 또는 단백질 마커, 마커 (본원에 정의된 바와 같음), 가족력 등)에 의한 임의의 객관적 또는 주관적 결정에 의해 이루어질 수 있다.

[0309] 인간 환자 또는 비-인간 동물의 치료는 생리학상 허용되는 담체 중 치료 유효량의 조합 치료제를 사용하여 수행될 수 있다. 어구 "제약상 허용되는"은 타당한 의학적 판단의 범주 내에서, 합리적인 이익/위험 비에 상응하게, 과도한 독성, 자극, 알레르기 반응 또는 다른 문제 또는 합병증 없이 인간 및 동물의 조직과 접촉하

여 사용하기에 적합한 본 개시내용의 화합물, 이러한 화합물을 함유하는 조성물 및/또는 투여 형태를 지칭한다.

- [0310] 조성물은 편리하게는 단위 투여 형태로 제공될 수 있고, 관련 기술분야에 널리 공지된 임의의 방법에 의해 제조될 수 있다. 단일 투여 형태를 생산하기 위해 담체 물질과 조합될 수 있는 조성물 (예를 들어, 관심 단백질 (예를 들어, STRC)의 N-말단 또는 C-말단 부분을 코딩하는 서열을 함유하는 벡터)의 양은 치료될 숙주, 특정한 투여 방식에 따라 달라질 것이다. 담체 물질과 조합되어 단일 투여 형태를 생성할 수 있는 조성물의 양은 일반적으로 치료 효과를 생성하는 조성물의 양일 것이다. 일반적으로, 100 퍼센트 중, 이 양은 조성물의 1% 내지 99% (예를 들어, 5%-70%, 10%-30%)의 범위일 것이다.
- [0311] 조성물은 질환 또는 상태의 임상 또는 생리학적 증상을 제어하는 투여량으로 투여될 수 있으며, 이는 일부 경우에 관련 기술분야의 통상의 기술자에게 공지된 진단 방법에 의해 결정될 수 있다.
- [0312] 치료 조성물 및 치료 조합물은 유효량으로 투여된다. 예를 들어, 약 10^8 내지 약 10^{12} 바이러스 입자가 대상체에게 투여될 수 있고, 바이러스는 적합한 부피 (예를 들어, 10 μ L, 50 μ L, 100 μ L, 500 μ L 또는 1000 μ L)의, 예를 들어 인공 외림프 용액 중에 현탁될 수 있다.
- [0313] 상염색체 열성 청력 상실을 치료하는 방법
- [0314] 상염색체 열성 청력 상실 (예를 들어, 난청, 상염색체 열성 16의 비-중후군성 난청 (DFNB16))을 치료하기 위한 조성물 및 방법이 제공된다.
- [0315] 간략하게, DFNB16은 이환된 개체의 STRC 유전자 내의 돌연변이와 연관된다. 내이 내, 스테레오실린 (STRC) 세포의 구조 단백질을 코딩하는 STRC의 정상 발현은 청각 기능에 필수적이다. 청력 상실의 회복을 유도하기 위해, 야생형 STRC 유전자는 본원에 기재된 바와 같은 벡터 시스템 (예를 들어, 이중-AAV 인테인-매개 STRC 단백질 트랜스-스플라이싱 시스템)을 사용하여, 즉 전장 mRNA 및 전장 STRC 단백질이 발현되도록 야생형 STRC 유전자 서열 또는 그의 단편을 패키징함으로써 대상체에게 투여된다.
- [0316] 일부 실시양태에서, 예를 들어 이중-AAV 인테인-매개 STRC 단백질 트랜스-스플라이싱 벡터를 포함한 STRC 단백질을 코딩하는 벡터 시스템은 스테레오실린 (STRC) 단백질을 코딩하는 적어도 1개의 벡터 (예를 들어, 5' STRC 및 3' STRC 벡터)를 대상체의 와우 내로 직접 주사함으로써 DFNB16 청력 상실을 갖는 대상체에게 투여될 수 있다. 일부 실시양태에서, 1개의 벡터는 N-STRC 단백질만을 코딩하고, 1개의 벡터는 C-STRC 단백질만을 코딩한다. 치료 목적을 위해, STRC 폴리펩티드 또는 STRC 폴리펩티드를 코딩하는 STRC 폴리뉴클레오티드를 포함하는 조성물은 질환 또는 상태에 의해 이환된 신체 영역 (예를 들어, 와우)에 직접 투여될 수 있으며, 여기서 대상체의 계층은 본원에 기재된 바와 같은 청력 상실 (예를 들어, DFNB16)을 유발하거나 또는 이에 기여하는 STRC 돌연변이를 포함한다.
- [0317] 한 실시양태는 상염색체 열성 청력 상실의 치료를 필요로 하는 대상체에게 유효량의 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템); 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템)을 함유하는 세포; 또는 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템) 및 제약상 허용되는 비히클을 포함하는 제약 조성물을 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료하는 방법을 제공할 수 있다. 일부 실시양태는 DFNB16인 상염색체 열성 청력 상실을 치료하는 방법에 관한 것일 수 있다.
- [0318] 본 개시내용의 또 다른 실시양태는 대상체의 세포를 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템) 및 제약상 허용되는 비히클을 포함하는 조성물과 접촉시키는 것을 포함하며, 여기서 접촉시키는 것은 단백질의 N-말단 부분을 발현하는 제1 뉴클레오티드 서열 및 단백질의 C-말단 부분을 발현하는 제2 뉴클레오티드 서열의 세포 내로의 전달을 일으키고, 여기서 세포는 펩티드 결합에 의해 연결된 단백질의 N-말단 부분 및 단백질의 C-말단 부분을 발현하여 전장 단백질을 형성하는 것인 방법을 제공한다.
- [0319] 본 개시내용의 추가 실시양태는 청력 상실을 특징으로 하는 병리상태 또는 질환의 치료 및/또는 예방을 필요로 하는 대상체에게 유효량의 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템); 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템)을 함유하는 세포; 또는 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템) 및 제약상 허용되는 비히클을 포함하는 제약 조성물을 투여하는 것을 포함하며, 여기서 투여 단계는 대상체의 적어도 1개의 세포 (예를 들어, 내이 세포, 내유모 세포, 외유모 세포)에서 이루어지는 것인, 청력 상실을 특징으로 하는 병리상태 또는 질환을 치료 및/또는 예방하는 방법을 제공한다. 한 실시양태에서, 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템); 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템)을 함유하는 세포; 또는 본원에 기재된 벡터 시스템 (예를 들어, 이중-벡터 시스템) 및 제약상 허용되는 비히클

을 포함하는 제약 조성물의 유효량을 세포와 접촉시키거나 또는 세포에 투여하는 방법은 생체내, 생체외 및/또는 시험관내에서 이루어진다. 또 다른 실시양태는 본원에 기재된 임의의 방법을 제공하며, 여기서 방법은 대상체에서 청각 기능을 개선 또는 회복시킨다.

[0320] 비제한적 투여 방법은 와우관 또는 와우관 주위의 외림프-충진된 공간 (예를 들어, 고실계 및 전정계) 내로 주사하는 것을 포함할 수 있다. 고칼륨 내림프액으로 충전된 와우관 내로의 주사는 유모 세포에 대한 직접적 접근을 제공할 수 있다. 그러나, 이러한 섬세한 유체 환경에 대한 변경은 와우내 잠재력을 파괴하여, 주사-관련 독성에 대한 위험을 높일 수 있다. 와우관을 둘러싸는 외림프-충진된 공간, 고실계 및 전정계는 난원창 또는 원형창 막을 통해 중이로부터 접근될 수 있다. 내이 내로의 유일한 비-골성 개구부인 원형창 막은 많은 동물 모델에서 비교적 용이하게 접근가능하고, 이 경로를 사용한 바이러스 벡터의 투여는 잘 허용된다. 인간에서, 와우 이식물 배치는 상용적으로 원형창 막을 통한 외과적 전극 삽입에 의존한다.

[0321] 일부 실시양태에서, 관심 단백질 (예를 들어, 야생형 STRC)을 발현하는 것은 대상체에서 청각 기능을 회복시킬 수 있다. 일부 실시양태에서, 대상체에게 회복된 청각 기능은 10% 이상 (예를 들어, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%); 100% 이하 (예를 들어, 95%, 85%, 75%, 65%, 55%, 45%, 35%, 25%, 15%)일 수 있다.

[0322] 실시예

[0323] 하기 실시예는 본 설명의 구체적 측면을 예시한다. 실시예는 단지 실시양태 및 그의 다양한 측면의 구체적 이해 및 실시를 제공하기 때문에, 실시예는 제한적인 것으로 해석되어서는 안된다.

[0324] 실시예 1: 이중-벡터 시스템은 시험관내에서 큰 유전자를 형질도입한다

[0325] AAV 벡터 생산

[0326] 본원에 개시된 관심 단백질의 전달을 위한 이중-벡터 또는 이중-트랜스-스플라이싱 시스템을, 예를 들어 내이 내로의 2개의 독립적 아테노-연관 바이러스, 혈청형 2 (AAV2)/아테노-연관 바이러스, 혈청형 9 (AAV9)-Php.B 벡터에서의 STRC 코딩 서열을 사용하여 입증하였다. STRC 코딩 서열의 처음 2,247개 염기 쌍 (즉, STRC의 N-말단 부분) 바로 뒤에 위치하고 3'-ITR이 이어지는 스플라이스 공여자 서열 (예를 들어, N-인테인)을 갖는 제1 AAV 게놈을 생산하였다 (예를 들어, AAV2/AAV9-Php.B-STRC-trans/공여자). 5'-ITR 뒤에 및 나머지 STRC 코딩 서열 (즉, STRC의 C-말단 부분) (C-말단 myc 태그를 가짐) 바로 앞에 위치하는 상응하는 스플라이스 수용자 서열을 갖는 제2 AAV 게놈을 생산하였다 (예를 들어, AAV2/AAV9-Php.B-STRC-trans/수용자). AAV2 게놈은 바이러스 게놈의 각각의 말단에 역전된 말단 반복부 (ITR)를 갖는 콘카테머를 형성하는 것으로 공지되어 있기 때문에, STRC 코딩 서열을 2개의 단편으로 나누고, 개별 AAV 캡시드 (예를 들어, 합성 AAV: 효율적으로 내유모 및 외유모 세포를 형질도입시키는 것으로 제시된 Anc80) 내로 패키징하였다.

[0327] 인테인-매개 트랜스-스플라이싱 및 프로세싱을 확인하기 위해, 전장 STRC의 부분을 코딩하는 AAV2/AAV9-Php.B 벡터로 감염된 HEK293 세포 (감염 5일 후)를 웨스턴 블롯에 의해 분석하였다 (도 24). 구체적으로, 도 24는 하기를 보여준다: 레인 1: 대조군 또는 형질감염되지 않은 HEK293T 세포; 레인 2: 전장 스테레오실린 (STRC) (196.4 kDa); 레인 3: pCFS-#2, C 부분 (C-부분 (116.7 kDa)만을 갖는 구축물 #2가 있는 벡터); 레인 4: pCFS-#2, N+C 부분 (N- 및 C-부분을 갖는 구축물 #2가 있는 벡터); 레인 5: pCFS-#1, C 부분 (C-부분 (91.6 kDa)만을 갖는 구축물 #1이 있는 벡터); 레인 6: pCFS-#1, N+C 부분 (N- 및 C-부분을 갖는 구축물 #1이 있는 벡터). 웨스턴 블롯의 우측 화살표는 레인 4에서의 전장 STRC 단백질을 가리키며, 이는 STRC의 N-말단 부분 및 C-말단 부분을 코딩하는 서열을 각각 함유하는 2개의 AAV2/AAV9-Php.B 벡터가 HEK293 세포 내로 형질감염되었을 때 STRC의 N-말단 부분 및 C-말단 부분 둘 다가 형성되었다는 것을 입증한다.

[0328] 신호 서열의 유용성을 확인하기 위해, 추가의 웨스턴 블롯을 수행하였다. 도 25는 하기를 보여준다: 레인 1: 전장 스테레오실린 (STRC) (196.4 kDa); 레인 2: 대조군 또는 형질감염되지 않은 HEK293T 세포; 레인 3: pCFS-#2, C 부분, (-) 신호 (신호 서열이 없는 C-부분 (116.7 kDa)만을 갖는 구축물 #2가 있는 벡터); 레인 4: pCFS-#2, N+C 부분, (-) 신호 (신호 서열이 없는 N- 및 C-부분을 갖는 구축물 #2가 있는 벡터); 레인 5: pCFS-#2, C 부분, (+) 신호 (신호 서열이 있는 C-부분 (91.6 kDa)만을 갖는 구축물 #2가 있는 벡터); 레인 6: pCFS-#2, N+C 부분, (+) 신호 (신호 서열이 있는 N- 및 C-부분을 갖는 구축물 #2가 있는 벡터). 웨스턴 블롯의 우측의 화살표는 레인 6에서의 전장 STRC 단백질을 가리키며, 이는 신호 서열이 전장 STRC 단백질의 형성을 일으키는 데 필요하였다는 것을 입증한다.

[0329] AAV 벡터는 보스턴 아동 병원 바이러스 코어(Boston Children's Hospital Viral Core) (미국 매사추세츠주 보스턴)에 의해 생산되었다. STRC 및 인테인을 함유하는 플라스미드를 AAV9-php.b-cmv 내로의 패키징 (MGH DNA

코어, 완전 플라스미드 서열분석) 전에 서열분석하였다. 벡터 역가는 바이러스의 역전된 말단 반복부 (AAV2)에 특이적인 qPCR에 의해 결정 시 4.8×10^{14} gc/ml였다.

[0330] 실시예 2: 생체내 Strc 녹아웃 마우스에서의 이중-벡터 시스템의 분석

[0331] 동물

[0332] 모든 동물을 사육하고, 시설에 수용하였다. 동물을 수반하는 모든 연구는 동물에 대한 HMS 상설 위원회 (프로토콜 번호 03524) 및 보스턴 아동 병원 동물 실험 윤리 위원회 (프로토콜 번호 2878 및 3396)에 의해 승인되었다. 모든 실험을 동물 프로토콜에 따라 수행하였다.

[0333] 스테레오실린 결핍인 널 대립유전자 ("녹아웃") 마우스 ($STRC^{-/-}$; $Strc^{\Delta/\Delta}$)를 생성하고, 부재하는 또는 비-기능적 스테레오실린 단백질을 유도하는 STRC 돌연변이에 의해 유발된 인간 청력 상실 DFN16 표현형에 대한 마우스 모델로서 제공하였다. Strc 동형접합 돌연변이 마우스 ($STRC\#16$ homo)는 청각 뇌간 반응 (ABR)에 의해 결정 시 4주령까지 중증 청력 상실을 나타냈고, 6주령까지 돌연변이 마우스는 완전히 사망하였다. Strc 동형접합 돌연변이 마우스는 또한 최대 80 dB 음압 수준의 검출가능한 변조 이음향방사 (DPOAE)가 결여되었으며, 이는 정상 외유모 세포 (OHC) 기능의 부재를 반영한다.

[0334] $Strc^{\Delta/\Delta}$ 마우스를 생성하고, 도 32a-32g에서 특징화하였다. 야생형 (WT) 관심 단백질, 예를 들어 Strc를, Strc 유전자의 엑손 4를 표적화하도록 3개의 가이드 RNA (sgRNA)를 설계함으로써 CRISPR/Cas9 전략을 사용하여 파괴하였다. 파괴는 249개의 뉴클레오타이드 결실 (위치 1509-1758) 및 2개의 전위 및 역위 (3' 말단에 보다 가까운 위치 947-1139; 5' 말단에 보다 가까운 위치 1758-1835)를 생성하였다.

[0335] 내이 주사

[0336] 생후 제1일 (P1)에 $Strc^{-/-}$ 또는 $Strc^{WT/WT}$ 마우스 새끼의 내이에 1 μ l의 AAV9-php.b-cmv-STRC 인테인 바이러스를 60 nl/분의 속도로 주사하였다. 새끼를 2-3분 동안 빙수에서 저체온 노출을 사용하여 마취시켰다. 마취 시, 귀-후방부 절개를 수행하여 귀 수포를 노출시키고 와우를 가시화하였다. 유리 마이크로피펫에 의해 수동으로 주사하였다. 주사 후, 봉합사를 사용하여 피부 절단부를 봉합하였다. 이어서, 주사된 마우스를 회복시키기 위해 42°C 가열 패드 상에 두었다. 새끼가 ~10분 내에 완전히 회복된 후에 어미에게 돌려보냈다. 수술 후에 표준 수술 후 관리를 적용하였다. 생체내 연구를 위한 샘플 크기를 계속적으로 결정하여 샘플 크기를 최적화하고 분산을 감소시켰다. P5 내지 P7에, 주사된 귀로부터 코르티 기관을 절제하였다. 코르티 기관 조직을 37°C, 5% CO₂에서 8-10일 동안 인큐베이션하고, 전기생리학 기록 직전에 덮개막을 제거하였다.

[0337] 청력 검사

[0338] 각각 N-인테인을 갖는 N-STRC 및 C-인테인을 갖는 C-STRC를 코딩하는 서열을 포함하는 AAV 벡터를 사용한 이중-벡터 시스템이 청력 상실을 회복시킬 수 있는지 여부 및 회복시키는 정도를 결정하기 위해, 스테레오실린 녹아웃 ($STRC^{-/-}$) 마우스에서 청각 뇌간 반응 (ABR) 및 변조 이음향방사 (DPOAE)를 측정하였다. EPL 음향 시스템 (매사추세츠 아이 앤드 이어(Massachusetts Eye and Ear), 보스턴)을 사용하여 ABR 및 DPOAE 측정을 기록하였다. 음향 자극을 PXI-1042Q 새시에서 24-비트 디지털 입력/출력 카드 (내셔널 인스트루먼트(National Instruments) PXI-4461)로 생성하고, SA-1 스피커 드라이버 (터커-데이비스 테크놀로지스, 인크.(Tucker-Davis Technologies, Inc.))에 의해 증폭시키고, 맞춤 음향 시스템에서 2개의 정전기적 드라이버 (CUI CDMG15008-03A)로부터 전달하였다. 작은 프로브 튜브 말단의 일렉트렛 마이크로폰 (놀스(Knowles) FG-23329-P07)을 사용하여 이도 음압을 모니터링하였다. 동일한 세션 동안 마우스로부터 ABR 및 DPOAE를 기록하였다. 컷바퀴 (활동 전극), 두정 (기준 전극) 및 둔부 (접지 전극)에 삽입된 피하 바늘 전극을 사용하여 ABR 신호를 수집하였다. ABR 전위를 증폭시키고 (10,000x), 통과-여과시키고 (0.3-10 kHz), 이튼-피바디 레보러토리즈(Eaton-Peabody Laboratories)의 와우 기능 검사 모음으로부터의 맞춤 데이터 획득 소프트웨어 (랩뷰(LabVIEW))를 사용하여 디지털화하였다. 소리 자극 및 전극 전압을 디지털 I-O 보드 (내셔널 인스트루먼트)를 사용하여 40- μ s 간격으로 샘플링하고, 오프라인 분석을 위해 저장하였다. 역치는 시각적으로 피크 1이 검출될 수 있고 소리 강도를 증가 시킴에 따라 재현될 수 있는 최저 데시벨 수준으로 정의하였다. ABR 역치를 각각의 실험군 내에서 평균내고, 통계적 분석에 사용하였다. 유전자형에 대해 맹검인 조사자에 의해 ABR 및 DPOAE 측정을 수행하였다.

[0339] 마우스를 크실라진 (5-10 mg/kg) 및 케타민 (60-100 mg/kg)의 복강내 (i.p.) 주사로 마취시키고, 컷바퀴의 기저부를 잘라내어 이도를 노출시켰다. a) 양쪽 귀 사이의 등쪽 (기준 전극); b) 왼쪽 컷바퀴 뒤쪽 (기록 전극);

및 c) 동물의 둔부의 등쪽 (접지 전극)을 비롯하여, 피부 내로 3개의 피하 바늘 전극을 삽입하였다. 세션 전반에 걸쳐 필요한 경우 마취를 유지하기 위해 케타민의 추가의 분취물 (60-100 mg/kg i.p.)을 제공하였다. ABR 시험 전에, 모든 자극 주파수에서 각각의 개별 시험 대상체에 대해 이도의 입구에서의 음압을 보정하였다. ABR 및 DPOAE 데이터를 동일한 조건 하에 및 동일한 기록 세션 동안 수집하였다.

[0340] DPOAE를 먼저 기록하였다. 2f1-f2에서의 DPOAE를 생성하기 위해 1.2의 주파수 비 (f1 및 f2 1차 음의 주파수 비 ($f2/f1 = 1.2$))로 1차 음을 생성하였으며, 여기서 f2 수준은 각각의 f2/f1 쌍에 대해 f1 수준 미만의 10 dB 음압 수준이었다. 음을 1/2-옥타브 단계에서 5.6 내지 32.0 kHz로 달라지는 f2 및 L1-L2 = 10 데시벨 음압 수준 (dB SPL)으로 제시하였다. 각각의 f2에서, L2를 10 dB 증분으로 10 내지 80 dB로 변동시켰다. DPOAE 역치는 평균 스펙트럼으로부터 잡음 플로어보다 5 dB 높은 크기의 DPOAE를 도출하는 L2-수준으로 정의하였다. 평균 잡음 플로어 수준은 모든 주파수에 걸쳐 0 dB 미만이었다. 각각의 수준에서, 기록된 이도 음압의 신호-대-잡음 (s/n) 비를 증가시키기 위해 파형 및 스펙트럼 평균을 사용하였다. 2f1-f2에서의 DPOAE는 평균낸 스펙트럼으로부터 추출된 진폭, 뿐만 아니라 스펙트럼 내의 이웃 지점에서의 잡음 플로어를 가졌다. DPOAE 진폭 대 소리 수준의 플롯으로부터의 내삽은 이소-반응 곡선을 생성하였다. 역치는 0 dB 초과 DPOAE를 생성하는 데 요구되는 f2 수준으로 정의하였다.

[0341] 이어서, 방음 챔버 내 32°C에서 ABR 실험을 수행하였다. 청각 기능을 시험하기 위해, 마우스에게 반-옥타브 단계로 5.6 내지 32.0 kHz의 광대역 "클릭" 음뿐만 아니라 순수 음의 자극을 제시하였고, 모두 5-ms 음 뺄소리로 제시하였다. 반응을 증폭시키고 (10,000회), 여과하고 (0.1-3 kHz), PC-기반 데이터-획득 시스템 (EPL, 와우 기능 검사 모듈, MEE, 보스턴)에서 아날로그-대-디지털 보드로 평균내었다. 다양한 시험에 걸쳐, 소리 수준을 5 내지 10 dB 단계로 0에서 110 dB SPL로 상승시켰다. 각각의 수준에서, 512개의 반응을 수집하고, "인공물 거부" 후에 각각의 음압 수준 (자극 극성이 교대됨)에 대해 평균내었다. 배경 잡음 대비 피크 1의 출현의 육안 검사에 의해 역치를 결정하였다. 오리진-2015 (오리진랩 코포레이션(OriginLab Corporation), 매사추세츠주)를 사용하여 데이터를 분석하고 플롯팅하였다. 달리 언급되지 않는 한 역치 평균 $\pm$ 표준 편차가 제시된다. 이들 실험의 대다수를 맹검 조건 하에 수행하지 않았다.

[0342] NHEJ-매개 Cas9-생성된 절단에 의해 Strc 유전자 코딩 서열을 파괴함으로써 STRC가 결여된 녹아웃 마우스를 생성하였다. STRC 유전자의 엑손 4 내에 ~200개 염기 쌍 결실을 생성하였으며, 이는 기능적 단백질의 합성을 방해하였다. 이러한 마우스 모델은 부재하는 또는 비-기능적 스테레오실린 단백질을 생성하는 STRC 돌연변이에 의해 유발된 DFNB16 표현형의 인간 청력 상실을 정확하게 재현하는 것으로 밝혀졌다. 제4주 연령의 STRC 동형접합 돌연변이 마우스는 청각 뇌간 반응 (ABR)에 기초하여 중증 청력 상실을 나타냈고, 제6주까지 이들 마우스는 완전히 사망하였다. ABR 역치는 명확한 반응 (CR)이 존재하는 최저 수준일 수 있다. 변조 이음향방사 (DPOAE)는 외유모 세포 완전성 및 와우 기능을 반영한다. 이들 STRC 동형접합 돌연변이 마우스는 또한 최대 80 데시벨 (dB) 음압 수준의 검출가능한 DPOAE를 가졌으며, 이는 정상 외유모 세포 (OHC) 기능의 부재를 반영한다.

[0343] 도 3은 ABR 파형 결과로부터의 음압 수준을 보여준다. 중앙 및 좌측 파형은 각각 STRC KO 마우스, STRC^{-/-} 마우스, (Strc#16 homo) 단독 (n=5) 및 AAV 벡터를 포함하는 본원에 기재된 이중-벡터 시스템 (여기서 각각의 AAV 벡터는 신호 서열, N-인테인 또는 C-인테인 및 STRC 단백질의 N-말단 또는 C-말단 부분을 코딩하는 서열을 함유함) (예를 들어, AAV2/AAV9-Php.B-Cmv-Strc-N; AAV2/AAV9-Php.B-Cmv-Strc-C)이 주사된 마우스 (n=9)의 결과를 보여주며, 우측 파형은 무손상 STRC 유전자를 갖는 야생형 (WT) 마우스 (STRC^{WT/WT}) (n=6)를 나타낸다. WT 마우스는 30 dB 내지 100 dB 범위의 음압 수준을 나타냈다. 중앙의 STRC KO 마우스 (Strc#16 homo)는 70 dB 내지 120 dB 범위의 제한된 음압 수준을 나타냈으며, 이는 70 dB 미만의 청력 상실을 입증한다. 그러나, 좌측 파형은 AAV2/AAV9-Php.B-Cmv-Strc-N; AAV2/AAV9-Php.B-Cmv-Strc-C가 주사된 STRC KO 마우스가 WT STRC 마우스의 것에 상응하는 수준으로 청력 상실을 회복할 수 있었다는 것을 입증한다.

[0344] 평균 청력 역치 또는 ABR을 4-주령 마우스에서 모든 주파수에 걸쳐 시험하였다. 도 4a의 ABR 결과는 AAV 벡터를 포함하는 본원에 기재된 이중-벡터 시스템 (여기서 각각의 AAV 벡터는 각각 신호 서열, N-인테인 또는 C-인테인 및 STRC 단백질의 N-말단 또는 C-말단 부분을 코딩하는 서열을 함유함) (예를 들어, AAV2/AAV9-Php.B-Cmv-Strc-N; AAV2/AAV9-Php.B-Cmv-Strc-C)이 각각 주사된 STRC 녹아웃 (KO) 마우스 (Strc^{-/-}) (n=9; 중간선)가 최저 ABR 역치가 일부 주파수에서 30 dB만큼 낮아져 청력 상실의 회복을 입증하였다는 것을 보여주었다. 야생형 마우스 (n=6; 가장 하부의 선)로부터의 ABR 반응은 최저 ABR 역치가 일부 주파수에서 20 dB만큼 낮아졌다. 그러나, STRC KO 마우스 (n=5; 상부 선)는 역치가 80 dB 초과로 중증 청력 상실을 가졌다.

- [0345] 도 5a 및 도 6a에서의 ABR 결과는 역치가 80 dB 초과인 중증 청력 상실을 갖는 STRC KO 마우스 (n=5; 상부 선과 오차 막대)를 보여준다. N-STRC (n=4; AAV2/AAV9-Php.B-Cmv-Strc-N) 또는 C-STRC (n=3; AAV2/AAV9-Php.B-Cmv-Strc-C)를 코딩하는 야생형 STRC의 부분을 코딩하는 인테인 AAV 중 어느 하나가 주사된 STRC KO 마우스는 STRC KO 마우스 결과와 유사한, 즉 80 dB 초과인 청력 역치를 나타냈다. 그러나, 야생형 마우스 (n=5; 가장 하부의 선)로부터의 ABR 반응은 최저 ABR 역치가 일부 주파수에서 20 dB만큼 낮아졌다.
- [0346] 도 4b의 DPOAE 결과 (도 45a에 사용된 동일한 STRC KO 마우스를 사용함)는 AAV 벡터를 포함하는 본원에 기재된 이중-벡터 시스템 (여기서 각각의 AAV 벡터는 신호 서열, N-인테인 또는 C-인테인 및 STRC 단백질의 N-말단 또는 C-말단 부분을 코딩하는 서열을 함유함) (예를 들어, 각각 AAV2/AAV9-Php.B-Cmv-Strc-N; AAV2/AAV9-Php.B-Cmv-Strc-C; n=9; 중간선)이 주사된 STRC KO 마우스가 일부 주파수에서 30 dB만큼 낮아진 DPOAE 반응에 의해 입증된 바와 같이 청력 상실의 회복을 나타냈다는 것을 입증하였다. 유사하게, 야생형 마우스 (n=6; 가장 하부의 선)는 최저 DPOAE 역치가 일부 주파수에서 30 dB만큼 낮아진 DPOAE 반응을 나타냈다. 그러나, STRC KO 마우스 (n=5; 상부 선)는 시험된 조건 (최대 80 dB) 하에 어떠한 DPOAE 반응도 나타내지 않았다.
- [0347] 도 5b 및 도 6b의 DPOAE 결과 (도 5a 및 6a에 사용된 동일한 STRC KO 마우스를 사용함)는 STRC KO 마우스 (n=5; 상부 선)가 시험된 조건 (최대 80 dB) 하에 어떠한 DPOAE 반응도 나타내지 않았다는 것을 입증하였다. 유사하게, N-STRC (n=4; AAV2/AAV9-Php.B-Cmv-Strc-N) 또는 C-STRC (n=3; AAV2/AAV9-Php.B-Cmv-Strc-C)를 코딩하는 야생형 STRC의 부분을 코딩하는 인테인 AAV 중 어느 하나가 주사된 STRC KO 마우스는 어떠한 DPOAE 반응도 나타내지 않았다. 그러나, 야생형 마우스 (n=5; 가장 하부의 선)는 최저 DPOAE 역치가 일부 주파수에서 30 dB만큼 낮아진 DPOAE 반응을 나타냈다.
- [0348] 도 7a 및 7b는 AAV 벡터를 포함하는 본원에 기재된 이중-벡터 시스템 (여기서 각각의 AAV 벡터는 신호 서열, N-인테인 또는 C-인테인 및 STRC 단백질의 N-말단 또는 C-말단 부분을 코딩하는 서열을 함유함) (예를 들어, 각각 AAV2/AAV9-Php.B-Cmv-Strc-N; AAV2/AAV9-Php.B-Cmv-Strc-C)이 주사된 3마리의 STRC KO 마우스의 시간 경과에 따른 모니터링 결과를 입증한다. 각각의 마우스에 대한 ABR 반응은 일반적으로 4주에서 마우스에 대해 일부 주파수에서 최저 역치를 보여주었다 (실선). 일반적으로, 각각의 마우스에 대해 시간 경과에 따라, 역치는 4주에서 달성된 것과 비교하여 6주 (파선) 및 8주 (파선/점선)에서 증가하였다. 예를 들어, 도 7a는 4주 (실선)에서, 마우스 #3이 10 kHz 미만의 주파수에서 40 dB 내지 55 dB 범위의 ABR 역치를 가졌고, 6주 (파선)에서, 마우스 #3이 10 kHz 미만의 주파수에서 50 dB 내지 70 dB 범위의 ABR 역치를 가졌으며, 이는 4주에서 관찰된 것보다 더 큰 데시벨이었다는 것을 보여준다. DPOAE 반응의 경우, 시간 경과에 따라 더 낮은 주파수로부터 더 높은 주파수로의 이동이 관찰되었다. 도 7b에서, 가난 낮은 DPOAE 역치 반응 (50 dB)은 마우스 #3에 대해 4주에서 11 kHz의 주파수 및 6주에서 16 kHz에서 발생하였다.
- [0349] 실시예 3: 이중-벡터 시스템을 사용한 형태의 회복
- [0350] 이중 AAV 전달 시스템은 알렉사488 접합된 2차 항체 (녹색) 및 알렉사546-팔로이딘 (적색)을 갖는 항-STRC 항체로 염색된 와우의 시각적 관찰에 의해 입증된 바와 같이 STRC 발현 및 유모 다발 형태를 회복시켰다. 예를 들어, 도 33a는 야생형 (WT) Strc 또는 Strc^{Δ/Δ}가 주사된 와우 및 이중 AAV 벡터 주사된 Strc^{Δ/Δ} 와우의 공초점 영상을 제시한다. STRC 및 액틴을 염색하였으며, WT (상부 좌측)에서는 둘 다 관찰되었고, Strc^{Δ/Δ} + 이중 AAV 벡터 샘플 (상부 우측)에서는 부분적으로 관찰된 반면, Strc^{Δ/Δ} (상부 중간)에서는 파괴된 액틴 외유모 세포 (OHC) 다발이 관찰되었다. WT에서의 녹색 염색 및 역전된 V 유모 다발의 형성 (하부 좌측) 및 Strc^{Δ/Δ} + 이중 AAV 벡터 샘플에서의 부분적 제시 또는 회복 (하부 우측)에 의해 STRC 국재화가 제시되었고, Strc^{Δ/Δ}는 어떠한 녹색 염색된 유모 다발 (하부 중앙)도 나타내지 않았다.
- [0351] 스캐닝 전자 현미경검사 영상에서 본 개시내용의 이중 AAV 벡터 전달 시스템이 도 33b의 유모 다발 형태 (하단 패널)를 거의 WT 수준 (상단 패널)으로 회복시키는 것으로 관찰되었다. Strc^{Δ/Δ} 주사된 외유모 세포 다발은 야생형의 조직화된 OHC 다발과 대조적으로 비배열 또는 비조직화된 OHC 다발 (중간 패널)을 생성한다.
- [0352] 실시예 4: 이중-벡터 시스템을 사용한 청각 기능의 회복
- [0353] 이중 AAV 벡터 시스템은 또한 음압 수준 10 dB 내지 50 dB 범위의 DPOAE 파형의 푸리에 분석에 의해 입증된 바와 같이 DPOAE 및 ABR 역치를 회복시켰으며, 여기서 Strc^{Δ/Δ} + 이중 AAV 벡터 샘플 (도 34a, 우측)은 야생형 (도 34a, 좌측)의 경우와 유사한 청각 기능 패턴을 갖는 것으로 관찰되었고, DPOAE 역치는 이중 AAV 벡터 주사

된 $Strc^{\Delta/\Delta}$ 마우스가 청각 기능을 회복시켰음을 나타낸다 (도 34b). 야생형 (도 34c, 좌측) 및 $Strc^{\Delta/\Delta} +$ 이중 AAV 벡터 (도 34c, 우측) 주사된 와우에 대해 기록된 ABR 트레이스는 25 dB 내지 110 dB 범위의 유사한 음압 수준을 유발한 반면; $Strc^{\Delta/\Delta}$ 주사된 와우의 경우 70 dB 내지 120 dB의 음압 수준을 유발하였다 (도 34c, 중간). ABR 역치는 마우스에게 $Strc^{\Delta/\Delta}$ 단독과 비교하여 $Strc^{\Delta/\Delta} +$ 이중 AAV 벡터를 주사한 경우에 5 kHz 내지 30 kHz 범위의 주파수에서 회복을 입증하는 것으로 나타났다 (도 34d).

[0354] 구체적 실시양태

[0355] 비제한적 구체적 실시양태가 하기 기재되며, 이들 각각은 본 개시내용 내에 있는 것으로 간주된다.

[0356] 구체적 실시양태 1. 세포에서 관심 단백질을 발현시키기 위한 이중-벡터 시스템이며,

[0357] a) 5'에서 3' 방향으로:

[0358] - 관심 단백질의 아미노 말단 (N-말단) 부분을 코딩하는 부분 코딩 서열의 5'-말단에 있는 신호 서열;

[0359] - 관심 단백질의 N-말단 부분을 코딩하는 부분 코딩 서열;

[0360] - 부분 코딩 서열에 인접하고 그의 하류인 스플라이스 공여자 서열을 코딩하는 서열

[0361] 을 포함하는 제1 뉴클레오티드 서열을 포함하는 제1 벡터; 및

[0362] b) 5'에서 3' 방향으로:

[0363] - 관심 단백질의 카르복시 말단 (C-말단) 부분을 코딩하는 부분 코딩 서열의 5'-말단에 있는 신호 서열;

[0364] - 신호 서열 및 관심 단백질의 C-말단 부분을 코딩하는 부분 코딩 서열이 플랭킹된 스플라이스 수용자 서열을 코딩하는 서열;

[0365] - 관심 단백질의 C-말단 부분을 코딩하는 부분 코딩 서열

[0366] 을 포함하는 제2 뉴클레오티드 서열을 포함하는 제2 벡터

[0367] 를 포함하는 이중-벡터 시스템.

[0368] 구체적 실시양태 2. 구체적 실시양태 1에 있어서,

[0369] a) 5'에서 3' 방향으로:

[0370] - 5'-역전된 말단 반복부 (5'-ITR) 서열;

[0371] - 프로모터 서열;

[0372] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;

[0373] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 관심 단백질의 아미노 말단 (N-말단) 부분을 코딩하는 부분 코딩 서열;

[0374] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 인테인의 아미노 말단 단편 (N-인테인)을 코딩하는 서열;

[0375] - 폴리-아데닐화 (폴리A) 신호 서열;

[0376] - 3'-역전된 말단 반복부 (3'-ITR) 서열

[0377] 을 포함하는 제1 뉴클레오티드 서열을 포함하는 제1 벡터; 및

[0378] b) 5'에서 3' 방향으로:

[0379] - 5'-역전된 말단 반복부 (5'-ITR) 서열;

[0380] - 프로모터 서열;

[0381] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;

[0382] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 인테인의 카르복시 말단 단편 (C-인테인)을

코딩하는 서열;

- [0383] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 관심 단백질의 카르복시 말단 (C-말단) 부분을 코딩하는 부분 코딩 서열;
- [0384] - 폴리-아데닐화 (폴리A) 신호 서열;
- [0385] - 3'-역전된 말단 반복부 (3'-ITR) 서열
- [0386] 을 포함하는 제2 뉴클레오타이드 서열을 포함하는 제2 벡터
- [0387] 를 포함하는 이중-벡터 시스템.
- [0388] 구체적 실시양태 3. 구체적 실시양태 1 또는 2에 있어서, 세포에서 제1 벡터 및 제2 벡터가 각각
- [0389] a) N-말단에서 C-말단 방향으로:
- [0390] - 신호 펩티드 서열, 이에 연결된 관심 단백질 서열의 N-말단 부분, 이의 C-말단 단부에서 융합된 N-인테인 단백질 서열
- [0391] 을 포함하는 제1 단백질 서열; 및
- [0392] b) N-말단에서 C-말단 방향으로:
- [0393] - 신호 펩티드 서열, 이에 연결된 C-인테인 단백질 서열, 이에 N-말단 단부가 융합된 관심 단백질 서열의 C-말단 부분
- [0394] 을 포함하는 제2 단백질 서열
- [0395] 을 발현하는 것인 이중-벡터 시스템.
- [0396] 구체적 실시양태 4. 구체적 실시양태 1-3 중 어느 하나에 있어서, 관심 단백질의 N-말단 부분 및 관심 단백질의 C-말단 부분이 전장 관심 단백질을 형성하도록 구성된 것인 이중-벡터 시스템.
- [0397] 구체적 실시양태 5. 구체적 실시양태 1-4 중 어느 하나에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 동일한 것인 이중-벡터 시스템.
- [0398] 구체적 실시양태 6. 구체적 실시양태 1-4 중 어느 하나에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 제1 단백질 서열 및 제2 단백질 서열을 동일한 세포 구획으로 수송하도록 구성된 것인 이중-벡터 시스템.
- [0399] 구체적 실시양태 7. 구체적 실시양태 1-4 중 어느 하나에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 상이하고, 각각의 신호 펩티드 서열이 각각의 단백질 서열을 동일한 세포 구획으로 지시하는 것인 이중-벡터 시스템.
- [0400] 구체적 실시양태 8. 구체적 실시양태 1-7 중 어느 하나에 있어서, 제1 벡터 및 제2 벡터가 각각 바이러스 벡터인 이중-벡터 시스템.
- [0401] 구체적 실시양태 9. 구체적 실시양태 8에 있어서, 바이러스 벡터가 아데노-연관 바이러스 (AAV) 벡터인 이중-벡터 시스템.
- [0402] 구체적 실시양태 10. 구체적 실시양태 8 또는 구체적 실시양태 9에 있어서, 바이러스 벡터가 동일하거나 상이한 혈청형인 이중-벡터 시스템.
- [0403] 구체적 실시양태 11. 구체적 실시양태 1-10 중 어느 하나에 있어서, N-말단 부분 및 C-말단 부분이 펩티드 결합을 통해 전장 관심 단백질을 형성하도록 구성된 것인 이중-벡터 시스템.
- [0404] 구체적 실시양태 12. 구체적 실시양태 1-11 중 어느 하나에 있어서, 관심 단백질이 STRC 단백질인 이중-벡터 시스템.
- [0405] 구체적 실시양태 13. 구체적 실시양태 12에 있어서, STRC 단백질이 STRC 유전자에 의해 코딩되는 것인 이중-벡터 시스템.
- [0406] 구체적 실시양태 14. 구체적 실시양태 1-13 중 어느 하나에 있어서, 신호 서열이 서열식별번호: 9 또는 서열식별번호: 11에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하

는 것인 이중-벡터 시스템.

- [0407] 구체적 실시양태 15. 구체적 실시양태 1-14 중 어느 하나에 있어서, 신호 서열이 서열식별번호: 10 또는 서열식별번호: 12에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 갖는 신호 펩티드 서열을 코딩하는 것인 이중-벡터 시스템.
- [0408] 구체적 실시양태 16. 구체적 실시양태 1-15 중 어느 하나에 있어서, 관심 단백질의 N-말단 부분이 서열식별번호: 5 또는 서열식별번호: 7에 대해 또는 서열식별번호: 15 또는 서열식별번호: 16을 코딩하는 핵산 서열에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0409] 구체적 실시양태 17. 구체적 실시양태 1-16 중 어느 하나에 있어서, 관심 단백질의 N-말단 부분이 서열식별번호: 15 또는 서열식별번호: 16에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0410] 구체적 실시양태 18. 구체적 실시양태 1-17 중 어느 하나에 있어서, N-인테인 서열이 서열식별번호: 13에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0411] 구체적 실시양태 19. 구체적 실시양태 1-18 중 어느 하나에 있어서, N-인테인 서열이 서열식별번호: 14에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0412] 구체적 실시양태 20. 구체적 실시양태 1-19 중 어느 하나에 있어서, 관심 단백질의 C-말단 부분이 서열식별번호: 23 또는 서열식별번호: 24를 코딩하는 핵산 서열에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0413] 구체적 실시양태 21. 구체적 실시양태 1-20 중 어느 하나에 있어서, 관심 단백질의 C-말단 부분이 서열식별번호: 23 또는 서열식별번호: 24에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0414] 구체적 실시양태 22. 구체적 실시양태 1-21 중 어느 하나에 있어서, C-인테인 서열이 서열식별번호: 21 또는 서열식별번호: 46에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0415] 구체적 실시양태 23. 구체적 실시양태 1-22 중 어느 하나에 있어서, C-인테인 서열이 서열식별번호: 22 또는 서열식별번호: 49에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0416] 구체적 실시양태 24. 구체적 실시양태 1-23 중 어느 하나에 있어서, 제1 뉴클레오티드 서열이 서열식별번호: 5 또는 서열식별번호: 7에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0417] 구체적 실시양태 25. 구체적 실시양태 1-24 중 어느 하나에 있어서, 제1 뉴클레오티드 서열이 서열식별번호: 6 또는 서열식별번호: 8에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0418] 구체적 실시양태 26. 구체적 실시양태 1-25 중 어느 하나에 있어서, 제2 뉴클레오티드 서열이 서열식별번호: 17 또는 서열식별번호: 19에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0419] 구체적 실시양태 27. 구체적 실시양태 1-26 중 어느 하나에 있어서, 제2 뉴클레오티드 서열이 서열식별번호: 18 또는 서열식별번호: 20에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0420] 구체적 실시양태 28. 숙주 세포에서 STRC 유전자의 코딩 서열을 발현시키기 위한 벡터 시스템이며, 여기서 코딩 서열은 서열식별번호: 1, 서열식별번호: 3, 서열식별번호: 5, 서열식별번호: 7, 서열식별번호: 17, 서열식별번호: 19, 서열식별번호: 29, 서열식별번호: 31, 서열식별번호: 33 또는 서열식별번호: 38의 STRC 유전자, 서열식별번호: 30 또는 서열식별번호: 32의 mRNA 서열 또는 그의 단편을 포함하는 적어도 1개의 벡터를 포함하는 것

인 벡터 시스템.

- [0421] 구체적 실시양태 29. 구체적 실시양태 28에 있어서, STRC 유전자가 서열식별번호: 2, 서열식별번호: 4, 서열식별번호: 6, 서열식별번호: 8, 서열식별번호: 15, 서열식별번호: 16, 서열식별번호: 18, 서열식별번호: 20, 서열식별번호: 23, 서열식별번호: 24, 서열식별번호: 25, 서열식별번호: 26, 서열식별번호: 36 또는 서열식별번호: 39 또는 그의 조합의 STRC 단백질을 코딩하는 것인 벡터 시스템.
- [0422] 구체적 실시양태 30. 구체적 실시양태 28에 있어서, 숙주 세포에서 STRC 유전자의 코딩 서열을 발현시키기 위한 이중-벡터 시스템을 포함하며, 여기서 코딩 서열은 5' 말단 단편 및 3' 말단 단편을 포함하고, 이중-벡터 시스템은
- [0423] a) 5'에서 3' 방향으로:
- [0424] - 5'-역전된 말단 반복부 (5'-ITR) 서열;
- [0425] - 프로모터 서열;
- [0426] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;
- [0427] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, STRC 유전자 코딩 서열의 5' 말단 단편;
- [0428] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 인테인의 아미노 말단 단편 (N-인테인)을 코딩하는 서열;
- [0429] - 폴리-아데닐화 (폴리A) 신호 서열; 및
- [0430] - 3'-역전된 말단 반복부 (3'-ITR) 서열
- [0431] 을 포함하는 제1 뉴클레오티드 서열을 포함하는 제1 벡터; 및
- [0432] b) 5'에서 3' 방향으로:
- [0433] - 5'-역전된 말단 반복부 (5'-ITR) 서열;
- [0434] - 프로모터 서열;
- [0435] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는 신호 서열;
- [0436] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, 인테인의 카복시 말단 단편 (C-인테인)을 코딩하는 서열;
- [0437] - 프로모터에 작동가능하게 연결되고 프로모터의 제어 하에 있는, STRC 유전자 코딩 서열의 3' 말단 단편;
- [0438] - 폴리-아데닐화 (폴리A) 신호 서열; 및
- [0439] - 3'-역전된 말단 반복부 (3'-ITR) 서열
- [0440] 을 포함하는 제2 뉴클레오티드 서열을 포함하는 제2 벡터
- [0441] 를 포함하는 것인 벡터 시스템.
- [0442] 구체적 실시양태 31. 구체적 실시양태 30에 있어서, 세포에서 제1 벡터 및 제2 벡터가 각각
- [0443] a) N-말단에서 C-말단 방향으로:
- [0444] - 신호 펩티드 서열, 이에 연결된 STRC 단백질 서열의 N-말단 부분, 이의 C-말단 단부에서 융합된 N-인테인 단백질 서열
- [0445] 을 포함하는 제1 단백질 서열; 및
- [0446] b) N-말단에서 C-말단 방향으로:
- [0447] - 신호 펩티드 서열, 이에 연결된 C-인테인 단백질 서열, 이에 N-말단 단부가 융합된 STRC 단백질 서열의 C-말단 부분
- [0448] 을 포함하는 제2 단백질 서열

- [0449] 을 발현하는 것인 이중-벡터 시스템.
- [0450] 구체적 실시양태 32. 구체적 실시양태 30-31 중 어느 하나에 있어서, STRC 단백질의 N-말단 부분 및 STRC 단백질의 C-말단 부분이 전장 STRC 단백질을 형성하는 것인 이중-벡터 시스템.
- [0451] 구체적 실시양태 33. 구체적 실시양태 30-32 중 어느 하나에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 동일한 것인 이중-벡터 시스템.
- [0452] 구체적 실시양태 34. 구체적 실시양태 30-33 중 어느 하나에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 제1 단백질 서열 및 제2 단백질 서열을 동일한 세포 구획으로 수송하도록 구성된 것인 이중-벡터 시스템.
- [0453] 구체적 실시양태 35. 구체적 실시양태 30-34 중 어느 하나에 있어서, 제1 단백질 서열의 신호 펩티드 서열 및 제2 단백질 서열의 신호 펩티드 서열이 상이하고, 각각의 신호 펩티드 서열이 각각의 단백질 서열을 동일한 세포 구획으로 지시하는 것인 이중-벡터 시스템.
- [0454] 구체적 실시양태 36. 구체적 실시양태 30-35 중 어느 하나에 있어서, 제1 벡터 및 제2 벡터가 각각 바이러스 벡터인 이중-벡터 시스템.
- [0455] 구체적 실시양태 37. 구체적 실시양태 36에 있어서, 바이러스 벡터가 아데노-연관 바이러스 (AAV) 벡터인 이중-벡터 시스템.
- [0456] 구체적 실시양태 38. 구체적 실시양태 36 또는 구체적 실시양태 37에 있어서, 바이러스 벡터가 동일한 혈청형을 갖는 것인 이중-벡터 시스템.
- [0457] 구체적 실시양태 39. 구체적 실시양태 36 또는 구체적 실시양태 37에 있어서, 바이러스 벡터가 상이한 혈청형을 갖는 것인 이중-벡터 시스템.
- [0458] 구체적 실시양태 40. 구체적 실시양태 30-39 중 어느 하나에 있어서, N-말단 부분 및 C-말단 부분이 펩티드 결합을 통해 전장 STRC 단백질을 형성하는 것인 이중-벡터 시스템.
- [0459] 구체적 실시양태 41. 구체적 실시양태 30-40 중 어느 하나에 있어서, STRC 단백질이 STRC 유전자에 의해 코딩되는 것인 이중-벡터 시스템.
- [0460] 구체적 실시양태 42. 구체적 실시양태 30-41 중 어느 하나에 있어서, 신호 서열이 서열식별번호: 9 또는 서열식별번호: 11에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0461] 구체적 실시양태 43. 구체적 실시양태 30-42 중 어느 하나에 있어서, 신호 서열이 서열식별번호: 10 또는 서열식별번호: 12에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0462] 구체적 실시양태 44. 구체적 실시양태 30-43 중 어느 하나에 있어서, STRC 단백질의 N-말단 부분이 서열식별번호: 5 또는 서열식별번호: 7에 대해 또는 서열식별번호: 15 또는 서열식별번호: 16을 코딩하는 핵산 서열에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0463] 구체적 실시양태 45. 구체적 실시양태 30-44 중 어느 하나에 있어서, STRC 단백질의 N-말단 부분이 서열식별번호: 6, 서열식별번호: 8, 서열식별번호: 15 또는 서열식별번호: 16에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0464] 구체적 실시양태 46. 구체적 실시양태 30-45 중 어느 하나에 있어서, STRC 단백질의 N-말단 부분이 전장 STRC 단백질의 N-말단 단부 부분의 54% 미만 (예를 들어, 53.8%, 53.6%, 53.4%, 53.2%, 53%, 52%, 50%, 45%)을 포함하는 것인 이중-벡터 시스템.
- [0465] 구체적 실시양태 47. 구체적 실시양태 30-46 중 어느 하나에 있어서, N-인테인 서열이 서열식별번호: 13에 대해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0466] 구체적 실시양태 48. 구체적 실시양태 30-47 중 어느 하나에 있어서, N-인테인 서열이 서열식별번호: 14에 대

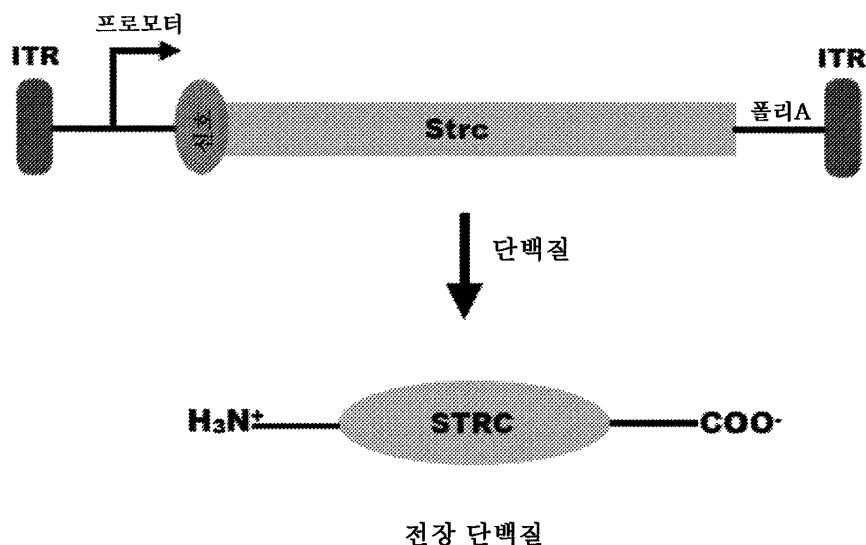
해 적어도 80% 동일성 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.

- [0467] 구체적 실시양태 49. 구체적 실시양태 30-48 중 어느 하나에 있어서, STRC 단백질의 C-말단 부분이 서열식별번호: 17, 서열식별번호: 19에 대해 또는 서열식별번호: 18, 서열식별번호: 20, 서열식별번호: 23 또는 서열식별번호: 24를 코딩하는 핵산 서열에 대해 적어도 70% (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%) 동일성의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0468] 구체적 실시양태 50. 구체적 실시양태 30-49 중 어느 하나에 있어서, STRC 단백질의 C-말단 부분이 서열식별번호: 18, 서열식별번호: 20, 서열식별번호: 23 또는 서열식별번호: 24에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0469] 구체적 실시양태 51. 구체적 실시양태 30-50 중 어느 하나에 있어서, STRC 단백질의 C-말단 단부 부분의 46% 이상 (예를 들어, 46.2%, 46.4%, 46.6%, 46.8%, 47%, 48%, 50%, 55%)을 포함하는 것인 이중-벡터 시스템.
- [0470] 구체적 실시양태 52. 구체적 실시양태 30-51 중 어느 하나에 있어서, C-인테인 서열이 서열식별번호: 21에 대해 적어도 80% 동일한 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%) 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0471] 구체적 실시양태 53. 구체적 실시양태 30-52 중 어느 하나에 있어서, C-인테인 서열이 서열식별번호: 22에 대해 적어도 80% 동일한 (예를 들어, 85%, 90%, 95%, 96%, 97%, 98%, 99%) 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0472] 구체적 실시양태 54. 구체적 실시양태 30-53 중 어느 하나에 있어서, 제1 뉴클레오티드 서열이 서열식별번호: 5 또는 서열식별번호: 7에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0473] 구체적 실시양태 55. 구체적 실시양태 30-54 중 어느 하나에 있어서, 제1 뉴클레오티드 서열이 서열식별번호: 6, 서열식별번호: 8, 서열식별번호: 15 또는 서열식별번호: 16에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0474] 구체적 실시양태 56. 구체적 실시양태 30-55 중 어느 하나에 있어서, 제2 뉴클레오티드 서열이 서열식별번호: 17 또는 서열식별번호: 19에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 핵산 서열을 포함하는 것인 이중-벡터 시스템.
- [0475] 구체적 실시양태 57. 구체적 실시양태 30-56 중 어느 하나에 있어서, 제2 뉴클레오티드 서열이 서열식별번호: 18, 서열식별번호: 20, 서열식별번호: 23 또는 서열식별번호: 24에 대해 적어도 70% 동일성 (예를 들어, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%)의 아미노산 서열을 코딩하는 것인 이중-벡터 시스템.
- [0476] 구체적 실시양태 58. 구체적 실시양태 1-57 중 어느 하나의 벡터 시스템을 함유하는 적어도 1개의 세포 (예를 들어, 10, 20, 50, 100, 200, 500, 1000개 또는 생물학적 활성을 갖는 큰 단백질을 성공적으로 발현하기에 충분한 임의의 개수의 세포)이며, 대상체에서 상염색체 열성 청력 상실일 수 있는 청력 상실을 치료, 억제 또는 감소시키기 위한 것일 수 있는 적어도 1개의 세포.
- [0477] 구체적 실시양태 59. 대상체에서 상염색체 열성 청력 상실일 수 있는 청력 상실을 치료, 억제 또는 감소시키기 위한, 구체적 실시양태 1-57 중 어느 하나의 벡터 시스템 및 제약상 허용되는 비히클을 포함하는 제약 조성물.
- [0478] 구체적 실시양태 60. 상염색체 열성 청력 상실의 치료, 억제 또는 감소를 필요로 하는 대상체에게 유효량의 구체적 실시양태 1-57 중 어느 하나의 이중-벡터 시스템을 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료, 억제 또는 감소시키는 방법.
- [0479] 구체적 실시양태 61. 상염색체 열성 청력 상실의 치료, 억제 또는 감소를 필요로 하는 대상체에게 유효량의 구체적 실시양태 58의 적어도 1개의 세포를 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료, 억제 또는 감소시키는 방법.
- [0480] 구체적 실시양태 62. 상염색체 열성 청력 상실의 치료, 억제 또는 감소를 필요로 하는 대상체에게 유효량의 구체적 실시양태 59의 제약 조성물을 투여하는 것을 포함하는, 대상체에서 상염색체 열성 청력 상실을 치료, 억제 또는 감소시키는 방법.

- [0481] 구체적 실시양태 63. 구체적 실시양태 60-62 중 어느 하나에 있어서, 상염색체 열성 청력 상실이 DFNB16인 방법.
- [0482] 구체적 실시양태 64. 대상체의 적어도 1개의 세포를 구체적 실시양태 59의 제약 조성물과 접촉시키는 것을 포함하는 방법이며, 여기서 접촉은 제1 뉴클레오티드 서열 및 제2 뉴클레오티드 서열을 포함하는 벡터 시스템을 대상체의 적어도 1개의 세포 내로 전달하고, 여기서 접촉된 적어도 1개의 세포는 펩티드 결합에 의해 연결된 단백질의 N-말단 부분 및 단백질의 C-말단 부분을 발현하여 전장 단백질을 형성하는 것인 방법.
- [0483] 구체적 실시양태 65. 청력 상실을 특징으로 하는 병리상태 또는 질환의 치료 및/또는 예방을 필요로 하는 대상체에게 유효량의 구체적 실시양태 1-57 중 어느 하나에 따른 벡터 시스템, 구체적 실시양태 58에 따른 적어도 1개의 세포 또는 구체적 실시양태 59에 따른 제약 조성물을 투여하는 것을 포함하는, 청력 상실을 특징으로 하는 병리상태 또는 질환을 치료 및/또는 예방하는 방법.
- [0484] 구체적 실시양태 66. 구체적 실시양태 64-65 중 어느 하나에 있어서, 적어도 1개의 세포가 내이 세포인 방법.
- [0485] 구체적 실시양태 67. 구체적 실시양태 64-66 중 어느 하나에 있어서, 적어도 1개의 세포가 내유모 세포 또는 외유모 세포인 방법.
- [0486] 구체적 실시양태 68. 구체적 실시양태 60-67 중 어느 하나에 있어서, 적어도 1개의 세포가 생체내 또는 시험관내 세포인 방법.
- [0487] 구체적 실시양태 69. 구체적 실시양태 60-68 중 어느 하나에 있어서, 대상체에서 청각 기능을 개선 또는 회복시키는 방법.
- [0488] 상기 기재로부터, 본원에 기재된 본 발명을 다양한 용법 및 조건에 채택하기 위해 그에 대해 변경 및 변형이 이루어질 수 있다는 것이 분명할 것이다. 이러한 실시양태는 또한 하기 청구범위의 범주 내에 있다.
- [0489] 본원의 변수의 임의의 정의에서 요소의 목록의 언급은 열거된 요소의 임의의 단일 요소 또는 조합 (또는 하위조합)으로서의 그 변수의 정의를 포함한다. 본원의 실시양태의 언급은 그 실시양태를 임의의 단일 실시양태로서 또는 임의의 다른 실시양태 또는 그의 부분과 조합하여 포함한다.
- [0490] 본 명세서에서 언급된 모든 특허 및 간행물은 각각의 독립적인 특허 및 간행물이 구체적으로 및 개별적으로 참조로 포함되는 것으로 나타내어진 것과 동일한 정도로 그 전문이 본원에 참조로 포함된다.

도면

도면1



도면2a

5' ATGgctctcagcctctgccccctgctgctgctgctgctgctgctgctgctgctgctgctccttt
gcaGTGACTCTGGCCCCCTACTGGGCCTCATTCCCTGGACCCTGGTCTCTCCTTCCTGAAGTC
ATTGCTCTCCACTCTGGACCAGGCTCCCCAGGGCTCCCTGAGCCGCTCACGGTTCTTTACAT
TCCTGGCCAAACATTTCTTCTTCTTTGAGCCTGGGAGAATGGGGGAAGGACCAGTAGGAGAG
CCCCACCTCTCCAGCCGCTGCTCTGCGGCTCCATGATTTTCTAGTGACACTGAGAGGTAG
CCCCGACTGGGAGCCAATGCTAGGGCTGCTAGGGGATATGCTGGCACTGCTGGGACAGGAGC
AGACTCCCCGAGATTTCTTGGTGCACCAGGCAGGGGTGCTGGGTGGACTTGTGGAGGTGCTG
CTGGGAGCCTTAGTTCTTGGGGGCCCCCTACCCAACTCGGCCCCATGCACCCGTGATGG
GCCGTCTGACTGTGTCTGGCTGCTGACTGGTTGCCTTCTCTGCTGCTGTTGTTAGAGGGCA
CACGCTGGCAAGCTCTGGTGCAGGTGCAGCCCAGTGTGGACCCACCAATGCCACAGGCCTC
GATGGGAGGGAGGCAGCTCCTCACTTTTGCAGGGTCTGTTGGGTTTGCTTACCCCAACAGG
GGAGCTAGGCTCCAAGGAGGCTCTTTGGGGCGGTCTGCTACGCACAGTGGGGGCCCCCTCT
ATGCTGCCTTTCAGGAGGGGTGCTCCGTGTCACTCACTCCCTGCAGGATGAGGTCTTCTCC
ATTTTGGGGCAGCCAGAGCCTGATACCAATGGGCAGTGCCAGGGAGGTAACCTTCAACAGCT
GCTCTTATGGGGCGTCCGGCACAACCTTTCCTGGGATGTCCAGGCGCTGGGCTTTCTGTCTG
GATCACCACCCCCACCCCTGCCCTCCTTCACTGCCTGAGCACGGGCGTGCTCTGCCCAGA
GCTTCTCAGCCGTGAGCCACATCAGCCACGCCAACGGCGAGCCATCACTGTGGAGGCCCT
CTGTGAGAACCCTTAGGCCCAGCACCCCTACAGCATTTCCAACCTTCTCCATCCACTTGC
TCTGCCAGCACACCAAGCCTGCCACTCCACAGCCCCATCCCAGCACCCTGCCATCTGCCAG
ACAGCTGTGTGGTATGCAGTGTCTTGGGCACCAGGTGCCAAGGCTGGCTACAGGCCTGCCA
CGACCAGTTTCTGATGAGTTTTTGGATGCGATCTGCAGTAACCTCTCCTTTTCAGCCCTGT
CTGGCTCCAACCGCCGCTGCTGAAGCGGCTCTGTGCTGGCCTGCTCCCACCCCTACCAGC
TGCCCTGAAGGCCTGCCCCCTGTTCCCTCACCCCAGACATCTTTTGGGGCTGCTTCTTGGA
GAATGAGACTCTGTGGGCTGAGCGACTGTGTGGGAGGCAAGTCTACAGGCTGTGCCCCCA
GCAACCAGGCTTGGGTCCAGCATGTGTGCCAGGGCCCCACCCAGATGTCACTGCCTCCCCA
CCATGCCACATTGGACCCTGTGGGGAACGCTGCCCCGATGGGGGCAGCTTCTGGTGATGGT
CTGTGCCAATGACACCATGTATGAGGTCCTGGTGCCTTCTGGCCTTGGCTAGCAGGCCAAT
GCAGGATAAGTCGTGGGGCAATGACACTTGCTTCTAGAGGGCTGCTGGGGCCCTTCTG
CCCTCTCTGCCACCACTGGGACCATCCCCACTCTGTCTGACCCCTGGCCCCTTCTCCTTGG

도면2b

CATGCTATCCCAGTTGCCACGCTGTGCTCCTCTGTCCCAGCTCTTGCTCACCCACACGCC
TACACTATCTCTCCGCCTGCTGACCTTCTCTTGGGTCCAGGGGCTGGGGCGCTGAGGCC
CAGGGGATGCTGGGTGCGGCCCTACTGCTCTCCAGTCTCCCAGACAACTGCTCCTTCTGGGA
TGCCTTTGCGCCAGAGGGCCGGCGCAGTGTGCTACGGACGATTGGGGAATACCTGGAACAAG
ATGAGGAGCAGCCAACCCCATCAGGCTTTGAACCCACTGTCAACCCAGCTCTGGTATAAGC
AAGATGGAGCTGCTGGCCTGCTTTAGTCCTGTGCTGTGGGATCTGCTCCAGAGGAAAAGAG
TGTTTGGGCCCTGCAGATTCTAGTGCAGGCGTACCTGCATATGCCCCAGAAAACCTCCAGC
AGCTGGTGCTTTTCAGCAGAGAGGGAGGCTGCACAGGGCTTCTGACACTCATGCTGCAGGGG
AAGCTGCAGGGGAAGCTGCAGGTACCAACCATCCGAGGAGCAGGCCCTGGGTGCGCTGACAGC
CCTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAGCTCTTCATTGACCTGTCAACACTCA
TCCCTTTCTTGGCTGTCTCTGACCTGATGCGCTTCCCACCATCCCTGTTAGCCAACGACAGT
GTCCTGGCTGCCATCCGGGATTACAGCCCAGGAATGAGGCCTGAACAGAAGGAGGCTCTGGC
AAAGCGACTGCTGGCCCTGAACTGTTTGGGGAAGTGCTGCTGGCCCCAGGAGCTGCTGT
GGGCAGTGCTGCCCCTGCTCCCCACCTCCCTCTGGAGAACTTTTTCAGCTCAGCCCTCAC
CAGATCCAGGCCCTGGAGGATAGCTGGCCAGCAGCAGGTCTGGGGCCAGGGCATGCCCCCCA
TGTGCTGCGCAGCCTGGTAAACCAGAGTGTCCAGGATGGTGAGGAGCAGGTACGCAGGCTTG
GGCCCCCTGCGCTGTTTCTGAGCCCTGAGGAGCTGCAGAGCCTAGTGCCCCCTGAGTGATCCA
ACGGGGCCAGTAGAACGGGGGCTGCTGGAATGTGCAGCCAATGGGACCCTCAGCCCAGAAGG
ACGGGTGGCATATGAACTTCTGGGTGTGTTGCGCTCATCTGGAGGAGCGGTGCTGAGCCCCC
GGGAGCTGCGGGTCTGGGCCCCCTCTCTTCTCTCAGCTGGGCCTCCGCTTCCTTCAGGAGCTG
TCAGAGCCCCAGCTTAGAGCCATGCTTCCTGTCTGCAGGGAAGTAGTGTTACACCTGCTCA
GGCTGTCTGCTGCTTGACGGCTCCTTCTTAGGCACGATCTATCCCTGGAGGAACTCTGCT
CCTTGACCTTCTGCTACCAGGCCTCAGCCCCCAGACACTCCAGGCCATCCCTAGGCGAGTC
CTGGTCGGGGCTTGTTCTGCTGCGGCCCTGAACTGTACGCCTCTCAGCCTGCCAGACCGC
AGCACTGCTGCAGACCTTTGCGGTTAAAGATGGTGTTAAAAATATGGGTACAACAGGTGCTG
GTCCAGCTGTGTGTATCCCTGGTCAGCCTATTCCCACCACCTGGCCAGACTGCCTGCTTCCC
CTGCTCCCATTAAGCTGCTACAACCTGGATTCTTGGCTCTTCTGGCAAATCGAAGACGCTA
CTGGGAGCTGCCCTGGTCTGAGCAGCAGGCACAGTTTCTCTGGAAGAAGATGCAAGTACCCA
CCAACCTTACCCTCAGGAATCTGCAGGCTCTGGGCACCCTGGCAGGAGGCATGTCCTGTGAG

도면2c

TTTCTGCAGCAGATCAACTCCATGGTAGACTTCCTTGAAGTGGTGACATGATCTATCAGCT
 GCCCCTAGAGTTTCGAGGGAGCCTGAGGGCCTGTATCTGGGCAGAGCTACAGCGGAGGATGG
 CAATGCCAGAACCAGAATGGACAACCTGTAGGGCCAGAAGTGAACGGGCTGGATAGCAAGCTA
 CTCCTGGACTTACCGATCCAGTTGATGGACAGACTATCCAATGAATCCATTATGTTGGTGGT
 GGAGCTGGTGCAAAGAGCTCCAGAGCAGCTGCTGGCACTGACCCCCCTCCACCAGGCAGCCC
 TGGCAGAGAGGGCACTACAAAACCTGGCTCCAAAGGAGACTCCAGTCTCAGGGGAAGTGCTG
 GAGACCTTAGGCCCTTTGGTTGGATTCTTGGGGACAGAGAGCACACGACAGATCCCCCTACA
 GATCCTGCTGTCCCATCTCAGTCAGCTGCAAGGCTTCTGCCTAGGAGAGACATTTGCCACAG
 AGCTGGGATGGCTGCTATTGCAGGAGTCTGTTCTTGGGAAACCAGAGTTGTGGAGCCAGGAT
 GAAGTAGAGCAAGCTGGACGCCTAGTATTCACTCTGTCTACTGAGGCAATTTCTTGATCCC
 CAGGGAGGCCTTGGGTCCAGAGACCCCTGGAGCGGCTTCTAGAAAAGCAGCAGAGCTGGGAGC
 AGAGCAGAGTTGGACAGCTGTGTAGGGAGCCACAGCTTGCTGCCAAGAAAGCAGCCCTGGTA
 GCAGGGGTGGTGCGACCAGCTGCTGAGGATCTTCCAGAACCTGTGCCAAATTGTGCAGATGT
 ACGAGGGACATTTCCAGCAGCCTGGTCTGCAACCCAGATTGCAGAGATGGAGCTCTCAGACT
 TTGAGGACTGCCTGACATTATTTGCAGGAGACCCAGGACTTGGGCCTGAGGAACTGCGGGCA
 GCCATGGGCAAAGCAAACAGTTGTGGGGTCCCCCGGGGATTTTCGTCTGAGCAGATCCT
 GCAGCTTGGTAGGCTCTTAATAGGTCTAGGAGATCGGGAACTACAGGAGCTGATCCTAGTGG
 ACTGGGGAGTGCTGAGCACCCCTGGGGCAGATAGATGGCTGGAGCACCACTCAGCTCCGCATT
 GTGGTCTCCAGTTTCTACGGCAGAGTGGTCGGCATGTGAGCCACCTGGACTTCGTTCATCT
 GACAGCGCTGGGTTATACTCTCTGTGGACTGCGGCCAGAGGAGCTCCAGCACATCAGCAGTT
 GGGAGTTTCAGCCAAGCAGCTCTCTTCTCGGCACCCTGCATCTCCAGTGCTCTGAGGAACAA
 CTGGAGGTTCTGGCCACCTACTTGTACTGCCTGGTGGGTTTGGCCCAATCAGTAACTGGGG
 GCCTGAGATCTTCACTGAAATTGGCACCATAGCAGCTGGGATCCCAGACCTGGCTCTTTCAG
 CACTGCTGCGGGGACAGATCCAGGGCGTTACTCCTCTTGCCATTTCTGTATCCCTCCTCCT
 AAATTTGCTGTGGTGTTTAGTCCCATCCAATATCTAGTCTCACCAGTGCTCAGGCTGTGGC
 TGTCATCCTGAGCAAATGGCCTTTCTGAGTCCTGAGCAGCGACGAGCAGTTGCATGGGCCC
 AACATGAGGGAAAGGAGAGCCCAGAACAGCAAGGTCGAAGTACAGCCTGGGGCCTCCAGGAC
 TGGTCACGACCTTCCTGGTCCCTGTTATTGACTATCAGCTTCCCTGGCCACCTGCTA**ACGCG**
TACGCGGCCGCTCgagcagaaactcatctcagaagaggatctg**TAATAG** 3'

(SEQ ID NO:33)

도면3

MalslwpllllllllllllsfaVTLAPTGPHS LDPGLSFLKSLLSTLDQAPQGSLSRSRFFTF
LANISSSFEPGRMGEGPVGEPPPLQPPALRLHDFLVTLRGSPDWEPMGLLGDM LALLGQEQ
TPRDFLVHQAGVLGGLVEVLLGALVPGGPPTPTRPPCTRDGPSDCVLAADWLPSLLLLLEGT
RWQALVQVQPSVDPTNATGLDGREAAHF LQGLLGLLTPTGELGSKEALWGGLLR TVGAPLY
AAFQEGLLRVTHSLQDEVFSILGQPEPDTNGQCQGGNLQQLLLWGV RHNLSWDVQALGFLSG
SPPPPPALLHCLSTGVPLPRASQPSAHISPRQRRAITVEALCENHLGPAPPY S ISNFSIHLL
CQHTK PATPQPH PSTTAICQTAVWYAVSWAPGAQGWLQACHDQFPDEF L DAICSNLSFSALS
GSNRRLVKRLCAGLLPPPTSCPEGLPPVPLTPDI FWGCFLENETLWAERLCGEASLQAVPPS
NQAWVQHVCQGP TPDVTASPPCHIGPCGERCPDGGSF LVMVCANDTMYEVLVPFWPWL AGQC
RISRGNDTCFLEGLLGPLLPSLPPLGPSPLCLTPGPFLLGMLSQ LPRCQSSVPALAHPTRL
HYLLRLLTFLLGPGAGGAEAQGMLGRALLSSLPDNC SFWDAFRPEGRRSVLRTIGEYLEQD
EEQPTPSGFEPTVNPSSGISKMELLACFS PVLWDLQREKSVWALQILVQAYLHMP PENLQQ
LVLSAEREAAGFLTLM LQGLQGLQVPPSEEQALGRLTALLQ RYPRLTSQLFIDLSPLI
PFLAVSDLMRFPPSLLANDSVLAAIRDYSPGMRPEQKEALAKRLLA PELFGEVPAWPQELLW
AVLPLLPHLPLENFLQLSPHQIQALED SWPAAGLGPGHARHVLRS L VNQSVQDGEEQVRR LG
PLACFLSPEELQSLVPLSDPTGPVERGLECAANGT L SPEGRVAYELLGVLRSSGGAVLS PR
ELRVWAPLFSQLGLRFLQELSE PQLRAMLPVLQGT SVTPAQAVLLLGRLLPRHDL SLEELCS
LHLLLPGLS PQTLQAI PRRLVVGACSC LAPELSRLSACQTAALLQTFRVKDGVKNM GTTGAG
PAVCIPGQPIPTTWPDCLLPLLPLKLLQLDSLALLANRRRYWELPWSEQQAQFLWKMQVPT
NLTLRNLQALGTLAGGMSCEFLQQINSMVDFLEV VHM IYQLPTRVRGSLRACIWAELQRRMA
MPEPEWTTVGPELNGLDSKLLLDLP IQLMDRLSNESIMLVVELVQRAPEQLLAL TPLHQAAL
AERALQNLAPKETPVSGEVLET LGPLVGFLGT ESTRQIPLQILLSHLSQLQGFC LGETFATE
LGWLLLQESVLGKPELWSQDEVEQAGRLVFTLSTE AISLI PREALGPETLERLLEKQQSWEQ
SRVGQLCREPQLAAKKAALVAGVVRPA AEDLPEPVPNCADV RGTFFPAAWSATQIAEMELSDF
EDCLTLFAGDPGLGPEELRAAMGKAKQLWGPPRGFRPEQILQLGRLLIGLGDRELQELILVD
WGVLSTLGQIDGWSTTQLRIVVSSFLRQSGRHVSHLDFVHLTALGYTLCGLRPEELQH ISSW
EFSQAALFLGTLHLQCSEEQLEVL A HLLVLPGGFGPISNWGPEIFTEIGTIAAGIPDLALSA
LLRGQIQGVTPLAISVIPPKFAVVFSP IQLSSLTSAQAVAVTPEQMAFLSPEQRR A VAWAQ
HEGKESPEQQGRSTAWGLQDWSRPSWSLVLTISFLGHLL**TRTRPL**eqkliseedl**

(SEQ ID NO:36)

도면4a

5' ATGgctctgagcctccagccccagctgctccttctcctgtcgctcctgccgcaggaagtg
acttcaGCCCCCTACTGGGCCTCAGTCTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTTCGT
AGCCACTCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTCACGGTTCTCTGCGTTCTTGG
CCAACATTTCTTCATCCTTCCAGCTTGGGAGGATGGGGGAGGGACCGGTGGGAGAGCCCCCA
CCTCTCCAGCCCCCTGCACTTCGACTTCATGATTTCCCTCGTGACACTGAGAGGTAGCCCAGA
CTGGGAGCCAATGCTAGGGCTTCTGGGAGATGTGCTGGCACTCCTGGGACAGGAACAGACTC
CCCCGGACTTTTTGGTGCAACAGGCAGGTGTACTGGGTGGACTTGTAGAGGCATTGTTGGGA
GCGTTAGTTCTTGGAGCCCCCTGCCCCACTCGACCCCCATGCACCCGTGATGGCCCTTC
TGA CTGTGCTTGGCTGCTGATTGGTTGCCTTCTCTGATGTTGTTATTAGAGGGTACACGCT
GGCAGGCCCTGGTGAGTTGCAGCCAGTGTGGACCAACCAATGCCACAGGTCTTGATGGT
AGAGAGCCAGCTCCTCACTTTTTACAGGGTCTGCTGGGCTTGCTTACCCAGCAGGAGAGTT
GGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGTGGGGGCCCCCTCTATGCTG
CCTTCCAGGAGGGGCTACTGCGAGTCACTCATTCTCTGCAAGATGAGGTCTTTTCTATTATG
GGACAGCCAGAGCCTGATGCCAGTGGGCAGTGCCAGGGAGGCAACCTTCAACAGCTGCTTTT
ATGGGGCATGCGGAACAACCTTTCTTGGGACGCCCAGCACTGGGTTTTCTATCTGGATCAC
CACCTCCACCCCTGCTCTCTGCACTGCCTGAGCAGAGGTGTGCCTCTGCCCAGGGCTTCC
CAGCCTGCGGCTCACATCAGCCCTCGACAGCGGCGAGCCATCTCTGTGGAGGCCCTCTGCGA
GAACCACTCAGGCCCAGAGCCACCCTACAGCATCTCCAATTCTCCATCTACTTGCTCTGCC
AGCACATCAAGCCTGCCACCCCGCGGCCCCCTCCTACCACCCACGGCCTCCTCCTACCACC
CCACAGCCCCCTCCTACCCTACACAGCCCATTCCTGACACTACACAGCCCCCTCCTGTCTAC
CCCCAGGCCCTCCTCCTACCACCCACAACCCCTCCTAGCACAGCTGTCTCTGCCAGACAG
CTGTATGGTACGCAGTCTCGTGGGCACCAGGTGCCCAGGTTGGCTCCAAGCCTGCCATGAT
CAGTTTCTGATCAATTTCTGGATATGATCTGCGGCAACCTCTCATTTTCAGCCCTGTCTGG
CCCCAGTCGTCCTTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCACCCCCCACTAGCTGTC
CACCAGGCCTGATCCCTGTGCCCCCTCACCCAGAAATATTCTGGGGCTGTTTCTGGAGAAT
GAGACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCAGGCTGTGCCCCGAGGAA
CCAGGCTTGGGTTTCAAGTGTGTGTGCGGGCCCCACCTTGGACGCCACTGATTTTCCACCGT
GCCGCTTGGACCTGTGGGGAACGCTGCCAGATGGGGGAGCTTCTGCTCATGGTCTGT
GCCAATGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTGCTAGCAGGCCAATGCAG

도면4b

AATTAGTCGTGGAGGAAATGATACTTGCTTTCTAGAAGGCATGCTGGGCCCCCTGTTGCCCT
 CTCTGCCCCCTCTGGGACCATCCCCACTCTGTCTGGCTCCTGGTCCTTTTCTGCTTGGCATG
 TTATCCCAGTTGCCACGCTGTCACTCCTCCGTGCCAGCCCTCGCCCCACCCACGCGCCTACA
 TTACCTCCTGCGCCTACTGACCTTCCTTCTGGGTCCAGGGACTGGGGGTGCCGAGACGCAGG
 GGATGTTAGGTCAAGCCCTGCTGCTCTCTAGTCTCCAGACAACGTTCATTCTGGGATGCC
 TTCCGCCCAGAGGGCCGAGAAGTGTACTGAGGACAGTCGGAGAGTACTTGCAGCGGAAGA
 GCCAACCCACCAGGCTTAGACTCCTCCCTCAGCCTCGGCTCTGGTATGAGCAAGATGGAGC
 TTCTGTCTGCTTCAGTCTGTACTGTGGGATCTACTCCAGAGAGAGAAGAGCGTTTGGGCC
 CTGAGGACCTGGTGAAGGCTACCTGCGCATGCCTCCAGAAGACCTTCAGCAGCTTGTGCT
 TTCAGCAGAGATGGAGGCTGCACAGGGCTTCCTGACGCTCATGCTTCGTTCTGGGCTAAGC
 TGAAGGTTCAACCATCCGAGGAGCAGGCCATGGGCCCGCTGACAGCCTTGCTGCTCCAGCGG
 TACCCACGCTCACCTCCCAACTCTTTATCGACATGTACCGCTCATCCCCCTTCCTGGCTGT
 CCCTGACCTCATGCGCTTCCACCGTCCCTTTTGGCCAACGACAGTGTCTGGCTGCCATCA
 GGGATCACAGCTCAGGAATGAAGCCTGAACAGAAGGAGGCCCTGGCAAAACGACTGCTGGCC
 CCTGAGCTGTTTGGAGAAGTGCCTGATTGGCCCCAGGAGCTGCTGTGGGCAGCCCTGCCTCT
 GCTTCCCCATCTGCCTCTGGAGAGCTTTCTCCAGCTCAGCCCTCACCAGATCCAGGCCCTGG
 AGGATAGCTGGCCAGTAGCAGATCTTGGGCCGGGACACGCCCCGACATGTGCTTCGTAGCCTA
 GTAAACCAGAGCATGGAGGATGGGGAGGAGCAGGTGCTCAGGCTTGGGTCCCTCGCCTGTTT
 CCTGAGTCTTGAGGAGCTACAGAGTCTGGTGCCCTTGAGTGATCCAATGGGGCCTGTAGAAC
 AGGGTCTGCTGGAATGTGCGGCCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATATGAA
 CTTCTGGGAGTGTTGCGTTCATCTGGAGGAAGTGTCTTAAGCCCCGAGAGCTGAGGGTCTG
 GGCACCTCTCTTTCCCCAGCTGGGCCTCCGCTTCCTGCAGGAGCTCTCAGAGACCCAGCTTA
 GAGCCATGCTTCCTGCCCTACAGGGAGCCAGTGTACACCTGCCCAGGCTGTTCTGTTGTTT
 GGAAGGCTCCTTCCTAAGCATGATCTGTCCCTGGAGGAAGTCTGCTCCCTGCACCTCTCCT
 GCCAGGTCTCAGCCCCAGACACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGGTGTGTT
 CCTGCCTGGGCCCTGAACTGTCAAGGCTTTCAGCTTGCCAGATTGCAGCTCTGCTGCAGACC
 TTTTCGGGTAAAAGATGGTGTTAAAAATATGGGTGCAGCAGGTGCCGGCTCAGCCGTGTGCAT
 TCCTGGGCAGCCCACCACTTGGCCAGACTGCCTGCTTCCCTGCTCCCATTAAGCTGCTAC
 AGCTGGACGCTGCAGCTCTTCTGGCAAACCGAAGACTCTATCGGCAGCTGCCTTGGTCTGAG

도면4c

CAACAGGCACAGTTTCTCTGGAAGAAAATGCAAGTGCCTACCAACCTGAGCCTGAGGAATCT
GCAGGCTCTGGGCAACTTGGCAGGAGGCATGACCTGCGAGTTTCTGCAGCAGATCAGCTCAA
TGGTTGACTTTCTTGATGTGGTACACATGCTCTACCAGCTGCCCCACTGGTGTTCGAGAGAGC
CTGCGGGCCTGTATCTGGACAGAGCTACAGCGGAGGATGACAATGCCAGAGCCAGAGCTGAC
CACCTTAGGGCCAGAACTGAGTGAACCTTGACACAAAGCTACTCCTGGACTTGCCGATCCAGC
TGATGGACAGATTGTCCAATGATTCCATTATGTTGGTGGTGGAGATGGTCCAAGGCGCTCCA
GAGCAGCTGCTGGCACTGACCCCACTCCACCAGACAGCCTTGGCAGAGCGAGCACTTAAAAA
CCTGGCTCCAAAGGAGACCCCCAATCTCCAAAGAAGTGTGGAGACACTGGGCCCCCTTGGTTG
GATTCTGGGAATAGAGAGCAGCGGACGGATCCCTTTACCCATTCTACTGTCTCATCTCAGT
CAGCTGCAGGGCTTCTGCCTAGGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTGTTGCA
GGAGCCTGTTCTTGGAAAACAGAATTGTGGAGCCAGGATGAAATAGAGCAAGCTGGACGCC
TAGTATTCACTCTGTCTGCTGAGGCTATTTCTCGATCCCCAGGGAGGCTTTGGGCCCAGAG
ACACTGGAGAGGCTTCTGGGAAAGCATCAAAGCTGGGAGCAGAGCAGAGTGGGCCATCTGTG
TGGGGAGTCACAGCTTGCCACAAGAAAGCAGCTCTGGTAGCTGGGATTGTGCATCCAGCTG
CTGAGGGTCTCCAAGAGCCTGTACCAAACCTGTGCAGACATACGGGGAACCTTCCCAGCGGCC
TGGTCTGCACACAAATCTCAGAGATGGAACCTCTCAGACTTTGAAGACTGCCTGTCACTATT
TGCTGGAGATCCAGGACTTGGTCCTGAGGAACCTACGGGCAGCCATGGGCAAGGCCAAGCAGT
TGTGGGGTCCCCCTCGAGGATTCCGTCCTGAGCAGATCTTGACAGCTGGGCCGTCTCCTGATA
GGTCTAGGAGAACGGGAACTGCAGGAGCTTACCTTGGTGGACTGGGGTGTGCTGAGCAGCCT
GGGGCAAATAGATGGCTGGAGTTCCATGCAGCTCCGAGCCGTGGTCTCCAGTTTCCTAAGGC
AGAGTGGTCGGCATGTGAGCCACCTGGACTTCATTTATCTGACAGCACTGGGTTACACAGTC
TGTGGATTGCGACCAGAGGAGTTACAGCACATCAGCAGTTGGGAGTTTAGCCAAGCAGCTCT
CTTCTGGGTAGCTTGCATCTCCCGTCTCTGAGGAACAGCTGGAAGTTCTGGCCTATCTCC
TTGTGTTGCCTGGTGGCTTTGGCCCAGTCAGTAACTGGGGGCCTGAGATCTTCACTGAAATT
GGCACAATAGCAGCTGGCATCCCAGACCTGGCTCTTTCAGCATTACTGCGGGGACAGATCCA
AGGCCTGACTCCTCTTGCCATTTCTGTCAATTCCTGCTCCCAAGTTTGCAGTGGTCTTCAACC
CCATCCAGTTATCTAGTCTCACCAGGGGTGAGGCCGTAGCTGTTACTCCTGAACAGCTGGCC
TATCTGAGTCCTGAGCAGCGGCGAGCAGTTGCATGGGCCCCAACACGAAGGGAAGGAGATCCC

도면4d

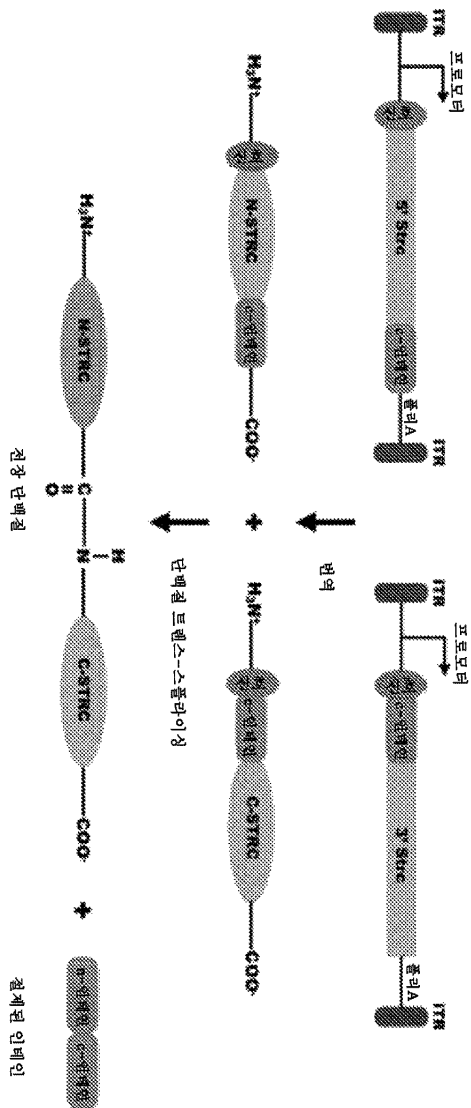
AGAGCAGCTGGGTGCAAACTCAGCCTGGGGTCTCTACGACTGGTTCCAAGCCTCCTGGGCCC
TGGCATTGCCGTCAGCATTTTGGCCACCTATTAA**ACGCGTACGCGGCCGCTC**gagcagaaa
ctcatctcagaagaggatctgTAATAG -3'

(SEQ ID NO:38)

도면5

MalslqpqlllllslpgevtsAPTGPQSLDAGLSLLKSFVATLDQAPQRSLSQSRFSAFLA
 NISSSFQIGRMGEGPVGEPPPLQPPALRLHDFLVTLRGSPDWEPMGLLGDVLALLGQEQT
 RDFLVHQAGVLGGLVEALLGALVPGGPPAPTRPPCTRDGPSDCVLAADWLPSIMLLLEGTRW
 QALVQLQPSVDPTNATGLDGREPAPHFLQGLLGLLTPAGELGSEALWGGLLRTVGAPLYAA
 FQEGLLRVTHSLQDEVFSIMQOPEPDASGQCQGGNLQQLLLWGMNNLSWDARALGFLSGSP
 PPPALLHCLSRGVPLPRASQPAAHISPRQRRASVEALCENHSGPEPPYSISNFSIYLLCQ
 HIKPATPRPPPTTPRPPPTTPQPPPTTTQPIPDTTQPPPVTPRPPPTTPQPPSTAVICQTA
 VWYAVSWAPGARGWLQACHDQFPDQFLDMICGNLSFSALSGPSRPLVKQLCAGLLPPPTSCP
 PGLIPVPLTPEIFWGCFLNETLWAERLCVEDSLQAVPPRNQAWQHVCRGPTLDATDFPPC
 RVGPCGERCPDGGSFLLMVCANDTLYEALVPFWAWLAGQCRISRGGNDTCFLEGMLGPLLPS
 LPPLGSPSLCLAPGFLLGMLSQLPQCQSSVPALAHPTRLHYLLRLLTFLLGPGTGGAETQG
 MLGQALLSSLPDNCSFWDAFRPEGRRSVLRTVGEYLQREEPTPPGLDSSSLSGSGMSKMEL
LSCFSPVLWDLQREKSVWALRTLKAYLRMPPEDLQQLVLSAEMEAAGFLTIMLRSWAKL
 KVQPSEEQAMGRLTALLQRYPRLTSQLFIDMSPLIPFLAVPDLMRFPSSLANDSVLAAIR
 DHSSGMKPEQKEALAKRLLAPELFGVEPDWPQELLWAALPLLPHLPLESFLQLSPHQIQALE
 DSWPVADLGPGHARHVLRLSVNQSMEDGEQVLRRLGSLACFLSPEELQSLVPLSDPMGPVEQ
 GLLECAANGTLSPEGRVAYELLGVLRSSGGTVLSPRELRVWAPLFPQLGLRFLQELSETQLR
 AMLPALQGASVTPAQAVLLFGRLLPKHDLSEELCSLHPLLPGLSPTLQAIPKRVLVGACS
 CLGPELSRLSACQIAALLQTFRVKDGVKNMGAAGAGSAVCIPGQPTTWPDCLLPLLPLKLLQ
 LDAAALLANRRLYRQLPWSEQQAQFLWKKMQVPTNLSLRNLQALGNLAGGMTCEFLQQISSM
 VDFLDVHMLYQLPTGVRESLRACIWTELQRRMTPEPELTTLGPELSELDTKLLDLPIQL
 MDRLSNDSIMLVVEMVQGAPEQLLALTPLHQTALAERALKNLAPKETPISKEVLETGPLVG
 FLGIESTRRIPLPILLSHLSQLQGFCGETFATELGWLLLQEPVLGKPELWSQDEIEQAGRL
 VFTLSAEAISSIPREALGPETLERLLGKHQSWEQSRVGHLCGESQLAHKKAALVAGIVHPAA
 EGLQEPVPNCADIRGTFFAAWSATQISEMELSDFEDCLSLFAGDPGLGPEELRAAMGKAKQL
 WGPPRGFRPEQILQLGRLLIGLGERELQELTLVDWGVLSLQIDGWSSMQLRAVVSSFLRQ
 SGRHVSHLDFIYLTALGYTVCGLRPEELQHISSEWFSQAALFLGSLHLPCSEEQLEVLAYLL
 VLPGGFGPVSNWGPETFEIGTIAAGIPDLALSALLRGQIQGLTPLAISVIPAPKFAVFN
 IQLSSLTRGQAVAVTPEQLAYLSPEQRRAVAWAQHEGKEIPEQLGRNSAWGLYDWFQASWAL
 ALPVSI FGHLL **TRTRPL**eqkliseedl** (SEQ ID NO:39)

도면6



도면7a

5' **ATG***gctctcagccctctggccccctgctgctgctgctgctgctgctgctgctgctgctgctt*
*gca*GTGACTCTGGCCCCCTACTGGGCCCTCATTCCTTGACCCTGGTCTCTCCTTCCTGAAGTC
ATTGCTCTCCACTCTGGACCAGGCTCCCCAGGGCTCCCTGAGCCGCTCACGGTTCTTTACAT
TCCTGGCCAACATTTCTTCTTCCTTTGAGCCTGGGAGAATGGGGGAAGGACCAGTAGGAGAG
CCCCACCTCTCCAGCCGCTGCTCTGCGGCTCCATGATTTTCTAGTGACACTGAGAGGTAG
CCCCGACTGGGAGCCAATGCTAGGGCTGCTAGGGGATATGCTGGCACTGCTGGGACAGGAGC
AGACTCCCCGAGATTTCTGGTGCACCAGGCAGGGGTGCTGGGTGGACTTGTGGAGGTGCTG
CTGGGAGCCTTAGTTCTGGGGGCCCCCTACCCCAACTCGGCCCCCATGCACCCGTGATGG
GCCGTCTGACTGTGTCTGGCTGCTGACTGGTTGCCTTCTCTGCTGCTGTTGTTAGAGGGCA
CACGCTGGCAAGCTCTGGTGCAGGTGCAGCCAGTGTGGACCCCACCAATGCCACAGGCCTC
GATGGGAGGGAGGCAGCTCCTCACTTTTTTGACGGGTCTGTTGGGTTTGTCTACCCCAACAGG
GGAGCTAGGCTCCAAGGAGGCTCTTTGGGGCGGTCTGCTACGCACAGTGGGGGGCCCCCTCT
ATGCTGCCTTTTCAGGAGGGGCTGCTCCGTGTCACTCACTCCCTGCAGGATGAGGTCTTCTCC
ATTTTGGGGCAGCCAGAGCCTGATACCAATGGGCAGTGCCAGGGAGGTAACCTTCAACAGCT
GCTCTTATGGGGCGTCCGGCACAACCTTTCTGGGATGTCCAGGCGCTGGGCTTTCTGTCTG
GATCACCACCCCCACCCCTGCCCTCCTTCACTGCCTGAGCACGGGCGTGCTCTGCCCAGA
GCTTCTCAGCCGTCAGCCACATCAGCCACGCCAACGGCGAGCCATCACTGTGGAGGCCCT
CTGTGAGAACCACTTAGGCCAGCACCACCCTACAGCATTTCCAACCTTCTCCATCCACTTGC
TCTGCCAGCACACCAAGCCTGCCACTCCACAGCCCCATCCCAGCACCCTGCCATCTGCCAG
ACAGCTGTGTGGTATGCAGTGTCTGGGCACCAGGTGCCCAAGGCTGGCTACAGGCCTGCCA
CGACCAGTTTCTGATGAGTTTTTGGATGCGATCTGCAGTAACCTCTCCTTTTCAGCCCTGT
CTGGCTCCAACCGCCGCCCTGGTGAAGCGGCTCTGTGCTGGCCTGCTCCCACCCCTTACCAG
TGCCCTGAAGGCCTGCCCTTGTTCCTTCACCCCAGACATCTTTTGGGGCTGCTTCTTGGA
GAATGAGACTCTGTGGGCTGAGCGACTGTGTGGGGAGGCAAGTCTACAGGCTGTGCCCCCA
GCAACCAGGCTTGGGTCCAGCATGTGTGCCAGGGCCCCACCCAGATGTCACTGCCTCCCCA
CCATGCCACATTGGACCTGTGGGGAACGCTGCCCGGATGGGGGCAGCTTCTGGTGATGGT
CTGTGCCAATGACACCATGTATGAGGTCTGGTGCCCTTCTGGCCTTGGCTAGCAGGCCAAT
GCAGGATAAGTCGTGGGGGCAATGACACTTGCTTCTTAGAAGGGCTGCTGGGGCCCCCTTCT

도면 7b

CCCTCTCTGCCACCACCTGGGACCATCCCCACTCTGTCTGACCCCTGGCCCTTCTCTCTTGG
CATGCTATCCCAGTTGCCACGCTGTAGTCCTCTGTCCCAGCTCTTGCTCACCCACACGCC
TACACTATCTCCTCCGCCTGCTGACCTTCCTCTTGGGTCCAGGGGCTGGGGGCGCTGAGGCC
CAGGGGATGCTGGGTGCGGCCCTACTGCTCTCCAGTCTCCAGACAACTGCTCCTTCTGGGA
TGCCTTTTCGCCCAGAGGGCCGGCGCAGTGTGCTACGGACGATTGGGGAATACCTGGAACAAG
ATGAGGAGCAGCCAACCCCATCAGGCTTTGAACCCACTGTCAACCCAGCTCTGGTATAAGC
AAGATGGAGCTGCTGGCCTGCTGTGCATACGAAACCGAGATACTGACAGTAGAATATGGCCT
TCTGCCAATCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTGATAACA
ATGGTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTC
GAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGT
CGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTG
ACAACCTTCCTAATTAATAG 3'

(SEQ ID NO:5)

도면8a

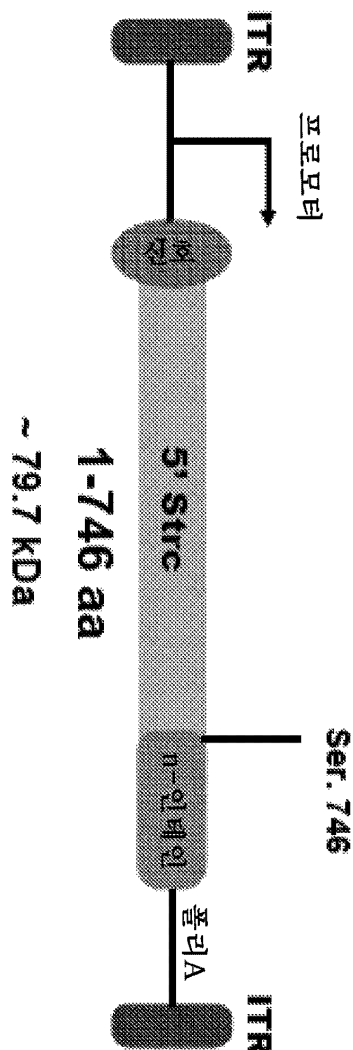
5' ATGgctctgagcctccagccccagctgctccttctcctgtcgctcctgccgcaggaagtg
acttcaGCCCCTACTGGGCCTCAGTCTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTTCGT
AGCCACTCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTCACGGTTCTCTGCGTTCCTGG
CCAACATTTCCTTCATCCTTCCAGCTTGGGAGGATGGGGAGGGACCGGTGGGAGAGCCCCCA
CCTCTCCAGCCCCCTGCACTTCGACTTCATGATTTCCCTCGTGACACTGAGAGGTAGCCCAGA
CTGGGAGCCAATGCTAGGGCTTCTGGGAGATGTGCTGGCACTCCTGGGACAGGAACAGACTC
CCCGGGACTTTTTGGTGCAACAGGCAGGTGTACTGGGTGGACTTGTAGAGGCATTGTTGGGA
GCGTTAGTTTCCTGGAGGCCCCCTGCCCCACTCGACCCCCATGCACCCGTGATGGCCCTTC
TGA CTGTGCTGCTGGCTGCTGATTGGTTGCCCTTCTCTGATGTTGTTATTAGAGGGTACACGCT
GGCAGGCCCTGGTGAGTTGCAGCCAGTGTGGACCCAACCAATGCCACAGGTCTTGATGGT
AGAGAGCCAGCTCCTCACTTTTTACAGGGTCTGCTGGGCTTGCTTACCCAGCAGGAGAGTT
GGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGTGGGGGCCCCCTCTATGCTG
CCTTCCAGGAGGGGCTACTGCGAGTCACTCATTCTCTGCAAGATGAGGTCTTTTCTATTATG
GGACAGCCAGAGCCTGATGCCAGTGGGCAGTGCCAGGGAGGCAACCTTCAACAGCTGCTTTT
ATGGGGCATGCGGAACAACCTTTCTTGGGACGCCCCGAGCACTGGGTTTTCTATCTGGATCAC
CACCTCCACCCCCTGCTCTCCTGCACTGCCTGAGCAGAGGTGTGCCTCTGCCAGGGCTTCC
CAGCCTGCGGCTCACATCAGCCCTCGACAGCGGCGAGCCATCTCTGTGGAGGCCCTCTGCGA
GAACCACTCAGGCCCAGAGCCACCCTACAGCATCTCCAATTCTCCATCTACTTGCTCTGCC
AGCACATCAAGCCTGCCACCCCGCGGCCCCCTCCTACCACCCACGGCCTCCTCCTACCACC
CCACAGCCCCCTCCTACCCTACACAGCCATTCTGACACTACACAGCCCCCTCCTGTCAC
CCCAAGGCCTCCTCCTACCACCCACAACCCCTCCTAGCACAGCTGTCATCTGCCAGACAG
CTGTATGGTACGCAGTCTCGTGGGCACCAGGTGCCCAGGTTGGCTCCAAGCCTGCCATGAT
CAGTTTCTGATCAATTTCTGGATATGATCTGCGGCAACCTCTCATTTTCAGCCCTGTCTGG
CCCCAGTCGTCCTTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCACCCCCCACTAGCTGTC
CACCAGGCCTGATCCCTGTGCCCCCTACCCCAGAAATATTCTGGGGCTGTTTCTGGAGAAT
GAGACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCAGGCTGTGCCCCGAGGAA
CCAGGCTTGGGTTTCAGCATGTGTGTGCGGGCCCCACCTTGGACGCCACTGATTTTCCACCGT
GCCGCGTTGGACCCTGTGGGGAACGCTGCCAGATGGGGGCAGCTTCTGCTCATGGTCTGT
GCCAATGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTGCTAGCAGGCCAATGCAG

도면8b

AATTAGTCGTGGAGGAAATGATACTTGCTTTCTAGAAGGCATGCTGGGCCCCCTTGTTGCCCT
 CTCTGCCCCCTCTGGGACCATCCCCACTCTGTCTGGCTCCTGGTCCTTTTCTGCTTGGCATG
 TTATCCCAGTTGCCACGCTGTCTAGTCCTCCGTGCCAGCCCTCGCCCACCCACGCGCCTACA
 TTACCTCCTGCGCCTACTGACCTTCCTTCTGGGTCCAGGGACTGGGGGTGCCGAGACGCAGG
 GGATGTTAGGTCAAGCCCTGTCTGCTCTCTAGTCTCCCAGACAACGTTCATTCTGGGATGCC
 TTCCGCCCAGAGGGCCGAGAAGTGTACTGAGGACAGTCGGAGAGTACTTGCAGCGGGAAGA
 GCCAACCCACCAGGCTTAGACTCCTCCCTCAGCCTCGGCTCTGGTATGAGCAAGATGGAGC
 TTCTGTCCTGCCTGTCTACGAAACCGAGATCTGACAGTAGAATATGGCCTTCTGCCAATC
GGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTGATAACAATGGTAACAT
TTATACTCAGCCAGTTGCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTGAATACTGTC
TGGAGGATGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAG
ATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCC
TAATTAATAG 3'

(SEQ ID NO:7)

도면9



도면10

Ma1slwp111111111111sfaVTLAPTGHSLDPGLSFLKSLSTLDQAPQGSLSRSRFFTF
LANISSSFEPGRMGEGPVGEPPPLQPPALRLHDFLVTLRGSPDWEPMLGGLGDMALLGQEQ
TPRDFLVHQAGVLGGLVEVLLGALVPGGPPTPTRPPCTRDGPSDCVLAADWLPSLLLLLEGT
RWQALVQVQPSVDPTNATGLDGREAAHFLQGLLGLLTPTGELGSKEALWGGLLRTVGAPLY
AAFQEGLLRVTHSLQDEVFSILGQPEPDTNGQCQGGNLQQLLLWGVNRHNSWDVQALGFLSG
SPPPPPALLHCLSTGVPLPRASQPSAHISPRQRRAITVEALCENHLGPAPPYSISNFSIHLL
CQHTKPATPQHPSTTAICQTAVWYAVSWAPGAQGWLQACHDQFPDEFDAICSNLSFSALS
GSNRRLVKRLCAGLLPPPTSCPEGLPPVPLTPDIFWGCFLNETLWAERLCGEASLQAVPPS
NQAWVQHVCQGPTPDVTASPPCHIGPCGERCPDGGSFVMVCANDTMYEVLVPFWPWLQAGQC
RISRGNDTCFLEGLLGPLLPSLPPLGPSPLCLTPGPFLLGMLSQLPQCQSSVPALAHPTRL
HYLLRLLTFLGPGAGGAEAQGMLGRALLSSLPDNCSEFWDAFRPEGRRSVLRRTIGEYLEQD
EEQPTPSGFEPVTNPSSGISKMELLACLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNN
GNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVD
NLPN**

(SEQ ID NO:6)

도면11

Ma1slqpq111111111111sllpgevtSAPTGPQSLDAGLSLLKSFVATLDQAPQRSLSQSRFSAFLA
NISSSFQGRMGEGPVGEPPPLQPPALRLHDFLVTLRGSPDWEPMLGGLGDLVALLGQEQTP
RDFLVHQAGVLGGLVEALLGALVPGGPPAPTRPPCTRDGPSDCVLAADWLPSLMMLLEGTRW
QALVQLQPSVDPTNATGLDGREPAPHFLQGLLGLLTAPAGELGSEEALWGGLLRTVGAPLYAA
FQEGLLRVTHSLQDEVFSIMGQPEPDASGQCQGGNLQQLLLWGMNRNLSWDARALGFLSGSP
PPPPALLHCLSRGVPLPRASQPAAHISPRQRRASVEALCENHSGPEPPYSISNFSIYLLCQ
HIKPATPRPPPTTRPPPTTPQPPPTTTQPIPDTTQPPPVTTRPPPTTPQPPSTAVICQTA
VWYAVSWAPGARGWLQACHDQFPDQFLDMICGNLSFSALSGPSRPLVKQLCAGLLPPPTSCP
PGLIPVPLTPEIFWGCFLNETLWAERLCVEDSLQAVPPRNQAWVQHVCRGPTLDATDFPPC
RVGPCGERCPDGGSFLLMVCANDTLYEALVPFWAWLAGQCRISRGNDTCFLEGLMLGPLLPS
LPPLGPSPLCLAPGFLLGMLSQLPQCQSSVPALAHPTRLHYLLRLLTFLGPGTGGAETQG
MLGQALLSSLPDNCSEFWDAFRPEGRRSVLRRTVGEYLQREEPTPPGLDSSLSLGSMSKMEL
LSCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL
EDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN**

(SEQ ID NO:8)

도면 12a

5' ATGgctctcagccctctgcccctgctgctgctgctgctgctgctgctgctgctccttt
gcaATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGA
AAGAGATCACAACTTTGCTCTGAAGAACGGATTATAGCTTCTAATTGCTTTAGTCCCTGTGC
TGTTGGGATCTGCTCCAGAGGGAAAAGAGTGTTTGGGCCCTGCAGATTCTAGTGCAGGCGTAC
CTGCATATGCCCCAGAAAACCTCCAGCAGCTGGTGCTTTCAGCAGAGAGGGAGGCTGCACA
GGGCTTCCTGACACTCATGTCTGCAGGGGAAGCTGCAGGGGAAGCTGCAGGTACCACCATCCG
AGGAGCAGGCCCTGGGTGCCTGACAGCCCTGCTGTCTCCAGCGGTACCCACGCCTCACCTCC
CAGCTCTTCATTGACCTGTCAACCACTCATCCCTTTCCTTGCTGTCTCTGACCTGATGCGCTT
CCCACCATCCCTGTTAGCCAACGACAGTGTCTGGCTGCCATCCGGGATTACAGCCCAGGAA
TGAGGCCTGAACAGAAGGAGGCTCTGGCAAAGCGACTGCTGGCCCCTGAAC TGTTGGGGAA
GTGCCTGCCTGGCCCCAGGAGCTGCTGTGGGCAGTGCTGCCCTGCTCCCCACCTCCCTCT
GGAGAACTTTTTCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGGATAGCTGGCCAGCAG
CAGGTCTGGGGCCAGGGCATGCCGCCATGTGCTGCGCAGCCTGGTAAACCAGAGTGTCAG
GATGGTGAGGAGCAGGTACGCAGGCTTGGGCCCCCTGCCTGTTTCTTGAGCCCTGAGGAGCT
GCAGAGCCCTAGTGCCCTGAGTGATCCAACGGGGCCAGTAGAACGGGGGCTGCTGGAATGTG
CAGCCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATATGAAC TTCTGGGTGTGTGCGC
TCATCTGGAGGAGCGGTGCTGAGCCCCCGGGAGCTGCGGGTCTGGGCCCCCTCTCTTCTCTCA
GCTGGGCCTCCGCTTCCTTCAGGAGCTGTGAGAGCCCCAGCTTAGAGCCATGCTTCCTGTCC
TGCAGGGAAC TAGTGTTACACCTGCTCAGGCTGTCTGCTGCTTGACGGCTCCTTCCTAGG
CACGATCTATCCCTGGAGGAACTCTGCTCCTTGACCTTCTGCTACCAGGCCTCAGCCCCCA
GACACTCCAGGCCATCCCTAGGCGAGTCTGGTCGGGGCTTGTTCTGCTGCTGGCCCCCTGAAC
TGTCACGCCTCTCAGCCTGCCAGACCGCAGCACTGCTGCAGACCTTTCGGGTAAAGATGGT
GTTAAAAATATGGGTACAACAGGTGCTGGTCCAGCTGTGTGTATCCCTGGTCAGCCTATTCC
CACCACCTGGCCAGACTGCCTGCTTCCCCCTGCTCCCATTAAGCTGTACAAC TGGATTCCCT
TGGCTCTTCTGGCAAATCGAAGACGCTACTGGGAGCTGCCCTGGTCTGAGCAGCAGGCACAG
TTTCTCTGGAAGAAGATGCAAGTACCCACCAACCTTACCCTCAGGAATCTGCAGGCTCTGGG
CACCCTGGCAGGAGGCATGTCTGTGAGTTTCTGCAGCAGATCAACTCCATGGTAGACTTCC
TTGAAGTGGTGCACATGATCTATCAGCTGCCCACTAGAGTTTCAGGGGAGCCTGAGGGCCTGT
ATCTGGGCAGAGCTACAGCGGAGGATGGCAATGCCAGAACCAGAATGGACAAC TGTAGGGCC

도면 12b

AGAACTGAACGGGCTGGATAGCAAGCTACTCCTGGACTTACCGATCCAGTTGATGGACAGAC
TATCCAATGAATCCATTATGTTGGTGGTGGAGCTGGTGCAAAGAGCTCCAGAGCAGCTGCTG
GCACTGACCCCCCTCCACCAGGCAGCCCTGGCAGAGAGGGGCACTACAAAACCTGGCTCCAAA
GGAGACTCCAGTCTCAGGGGAAGTGCTGGAGACCTTAGGCCCTTTGGTTGGATTCTTGGGGA
CAGAGAGCACACGACAGATCCCCCTACAGATCCTGCTGTCCCATCTCAGTCAGCTGCAAGGC
TTCTGCCTAGGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTATTGCAGGAGTCTGTTCT
TGGGAAACCAGAGTTGTGGAGCCAGGATGAAGTAGAGCAAGCTGGACGCCTAGTATTCACCTC
TGTCTACTGAGGCAATTTCTTGGATCCCCAGGGAGGCCTTGGGTCCAGAGACCCTGGAGCGG
CTTCTAGAAAAGCAGCAGAGCTGGGAGCAGAGCAGAGTTGGACAGCTGTGTAGGGAGCCACA
GCTTGCTGCCAAGAAAGCAGCCCTGGTAGCAGGGGTGGTGCGACCAGCTGCTGAGGATCTTC
CAGAACCTGTGCCAAATTTGTGCAGATGTACGAGGGACATTCCCAGCAGCCTGGTCTGCAACC
CAGATTGCAGAGATGGAGCTCTCAGACTTTGAGGACTGCCTGACATTATTTGCAGGAGACCC
AGGACTTGGGCCTGAGGAACTGCGGGCAGCCATGGGCAAAGCAAAACAGTTGTGGGGTCCCC
CCCGGGGATTTGCTCCTGAGCAGATCCTGCAGCTTGGTAGGCTCTTAATAGGTCTAGGAGAT
CGGGAACCTACAGGAGCTGATCCTAGTGGACTGGGGAGTGCTGAGCACCCCTGGGGCAGATAGA
TGGCTGGAGCACCACCTCAGCTCCGCATTGTGGTCTCCAGTTTCTACGGCAGAGTGGTCGGC
ATGTGAGCCACCTGGACTTCGTTTCATCTGACAGCGCTGGGTTATACTCTCTGTGGACTGCGG
CCAGAGGAGCTCCAGCACATCAGCAGTTGGGAGTTCAGCCAAGCAGCTCTCTTCCTCGGCAC
CCTGCATCTCCAGTGCTCTGAGGAACAACCTGGAGGTTCTGGCCACCTACTTGTACTGCCTG
GTGGGTTTGGCCCAATCAGTAACTGGGGGCCTGAGATCTTCACTGAAATTGGCACCATAGCA
GCTGGGATCCCAGACCTGGCTCTTTTCAGCACTGCTGCGGGGACAGATCCAGGGCGTTACTCC
TCTTGCCATTTCTGTATCCCTCCTCCTAAATTTGCTGTGGTGTTTAGTCCCATCCAATAT
CTAGTCTCACCAGTGCTCAGGCTGTGGCTGTCACTCCTGAGCAAATGGCCTTTCTGAGTCCT
GAGCAGCGACGAGCAGTTGCATGGGCCCCAATGAGGGAAAGGAGAGCCCAGAACAGCAAGG
TCGAAGTACAGCCTGGGGCCTCCAGGACTGGTCACGACCTTCCTGGTCCCTGGTATTGACTA
TCAGCTTCCTTGCCACCTGCTA**ACGCGTACGCGGCCGCTC**gagcagaaactcatctcagaa
gaggatctg**TAATAG**

(SEQ ID NO:17)

도면13a

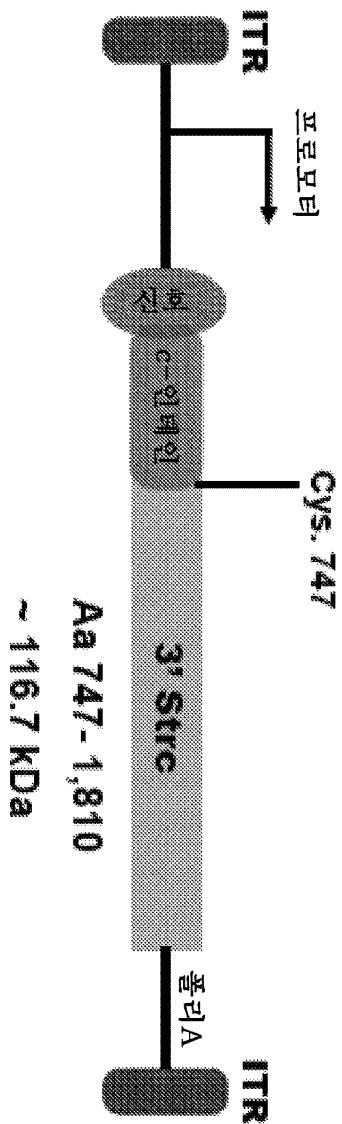
5' ATGgctctgagcctccagccccagctgctccttctcctgtcgctcctgccgcaggaagtg
acttcaATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGT
CGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTTCATAGCTTCTAATTGCTTCAGTCCTG
TACTGTGGGATCTACTCCAGAGAGAGAAGAGCGTTTGGGCCCTGAGGACCCTGGTGAAGGCC
TACCTGCGCATGCCTCCAGAAGACCTTCAGCAGCTTGTGCTTTTTCAGCAGAGATGGAGGCTGC
ACAGGGCTTCCTGACGCTCATGCTTCGTTCTGGGCTAAGCTGAAGGTTCAACCATCCGAGG
AGCAGGCCATGGGCCGCTGACAGCCTTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAA
CTCTTTATCGACATGTACCGCTCATCCCCCTTCTGGCTGTCCCTGACCTCATGCGCTTCCC
ACCGTCCCTTTTGGCCAACGACAGTGCCTGGCTGCCATCAGGGATCACAGCTCAGGAATGA
AGCCTGAACAGAAGGAGGCCCTGGCAAAACGACTGCTGGCCCCCTGAGCTGTTTGGAGAAGTG
CCTGATTGGCCCCAGGAGCTGCTGTGGGCAGCCCTGCCTCTGCTTCCCCATCTGCCTCTGGA
GAGCTTTCTCCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGGATAGCTGGCCAGTAGCAG
ATCTTGGGCCGGGACACGCCCACATGTGCTTCGTAGCCTAGTAAACCAGAGCATGGAGGAT
GGGAGGAGCAGGTGCTCAGGCTTGGGTCCCTCGCCTGTTTCTGAGTCTGAGGAGCTACA
GAGTCTGGTGCCCTTGAGTGATCCAATGGGGCCTGTAGAACAGGGTCTGCTGGAATGTGCGG
CCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATATGAACTTCTGGGAGTGTGCGTTCA
TCTGGAGGAAGTGTCTTAAGCCCCGAGAGCTGAGGGTCTGGGCACCTCTCTTTCCCCAGCT
GGGCCTCCGCTTCTGTCAGGAGCTCTCAGAGACCCAGCTTAGAGCCATGCTTCCTGCCCTAC
AGGGAGCCAGTGTACACCTGCCCAGGCTGTTCTGTTGTTTGAAGGCTCCTTCCTAAGCAT
GATCTGTCCCTGGAGGAAGTCTGCTCCCTGCACCTCTCCTGCCAGGTCTCAGCCCCCAGAC
ACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGGTGCTTGTTCCTGCCTGGGCCCTGAACTGT
CAAGGCTTTCAGCTTGCCAGATTGCAGCTCTGCTGCAGACCTTTCGGGTAAAAGATGGTGTT
AAAAATATGGGTGCAGCAGGTGCCGGCTCAGCCGTGTGCATTCTGGGCAGCCCACCACTTG
GCCAGACTGCCTGCTTCCCCTGCTCCCATTAAGCTGCTACAGCTGGACGCTGCAGCTCTTC
TGGCAAACCGAAGACTCTATCGGCAGCTGCCTTGGTCTGAGCAACAGGCACAGTTTCTCTGG
AAGAAAATGCAAGTGCTTACCAACCTGAGCCTGAGGAATCTGCAGGCTCTGGGCAACTTGGC
AGGAGGCATGACCTGCGAGTTTCTGCAGCAGATCAGCTCAATGGTTGACTTTCTTGATGTGG
TACACATGCTCTACCAGCTGCCCCTGGTGTTTCGAGAGAGCCTGCGGGCCTGTATCTGGACA
GAGCTACAGCGGAGGATGACAATGCCAGAGCCAGAGCTGACCACCCTAGGGCCAGAACTGAG

도면13b

TGAAC TTGACACAAAGCTACTCCTGGACTTGCCGATCCAGCTGATGGACAGATTGTCCAATG
ATTCCATTATGTTGGTGGTGGAGATGGTCCAAGGCGCTCCAGAGCAGCTGCTGGCACTGACC
CCACTCCACCAGACAGCCTTGGCAGAGCGAGCACTTAAAAACCTGGCTCCAAAGGAGACCCC
AATCTCCAAAGAAGTGCTGGAGACACTGGGCCCCCTGGTTGGATTCTGGGAATAGAGAGCA
CGCGACGGATCCCTTTACCCATTCTACTGTCTCATCTCAGTCAGCTGCAGGGCTTCTGCCTA
GGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTGTTGCAGGAGCCTGTTCTTGAAAACC
AGAATTGTGGAGCCAGGATGAAATAGAGCAAGCTGGACGCCTAGTATTCACCTCTGTCTGCTG
AGGCTATTTCTCGATCCCCAGGGAGGCTTTGGGCCCCAGAGACACTGGAGAGGCTTCTGGGA
AAGCATCAAAGCTGGGAGCAGAGCAGAGTGGGCCATCTGTGTGGGGAGTCACAGCTTGCCCA
CAAGAAAGCAGCTCTGGTAGCTGGGATTGTGCATCCAGCTGCTGAGGGTCTCCAAGAGCCTG
TACCAAACTGTGCAGACATACGGGGAACCTTCCCAGCGGCCTGGTCTGCGACACAAATCTCA
GAGATGGAACCTCTCAGACTTTGAAGACTGCCTGTCACTATTTGCTGGAGATCCAGGACTTGG
TCCTGAGGAACCTACGGGCAGCCATGGGCAAGGCCAAGCAGTTGTGGGGTCCCCCTCGAGGAT
TCCGTCTCGAGCAGATCTTGCAGCTGGGCCGTCTCCTGATAGGTCTAGGAGAACGGGAACTG
CAGGAGCTTACCTTGGTGGACTGGGGTGTGCTGAGCAGCCTGGGGCAAATAGATGGCTGGAG
TTCCATGCAGCTCCGAGCCGTGGTCTCCAGTTTCCCTAAGGCAGAGTGGTCGGCATGTGAGCC
ACCTGGACTTCATTTATCTGACAGCACTGGGTTACACAGTCTGTGGATTGCGACCAGAGGAG
TTACAGCACATCAGCAGTTGGGAGTTTAGCCAAGCAGCTCTCTTCTGGGTAGCTTGCATCT
CCCGTGCTCTGAGGAACAGCTGGAAGTTCTGGCCTATCTCCTTGTGTTGCCTGGTGGCTTTG
GCCCAGTCAGTAACTGGGGGCCTGAGATCTTCACTGAAATTGGCACAATAGCAGCTGGCATC
CCAGACCTGGCTCTTTTACGATTACTGCGGGGACAGATCCAAGGCCTGACTCCTCTTGCCAT
TTCTGTCAATTCCTGCTCCCAAGTTTGCAGTGGTCTTCAACCCCATCCAGTTATCTAGTCTCA
CCAGGGGTCAGGCCGTAGCTGTTACTCCTGAACAGCTGGCCTATCTGAGTCCTGAGCAGCGG
CGAGCAGTTGCATGGGGCCCAACACGAAGGGAAGGAGATCCCAGAGCAGCTGGGTCGAAACTC
AGCCTGGGGTCTCTACGACTGGTTCCAAGCCTCCTGGGCCCTGGCATTGCCCGTCAGCATTT
TTGGCCACCTATTA**ACGCGTACGCGGCCGCTC**gagcagaaactcatctcagaagaggatctg
TAATAG 3'

(SEQ ID NO:19)

도면14



도면 15

[illegible]

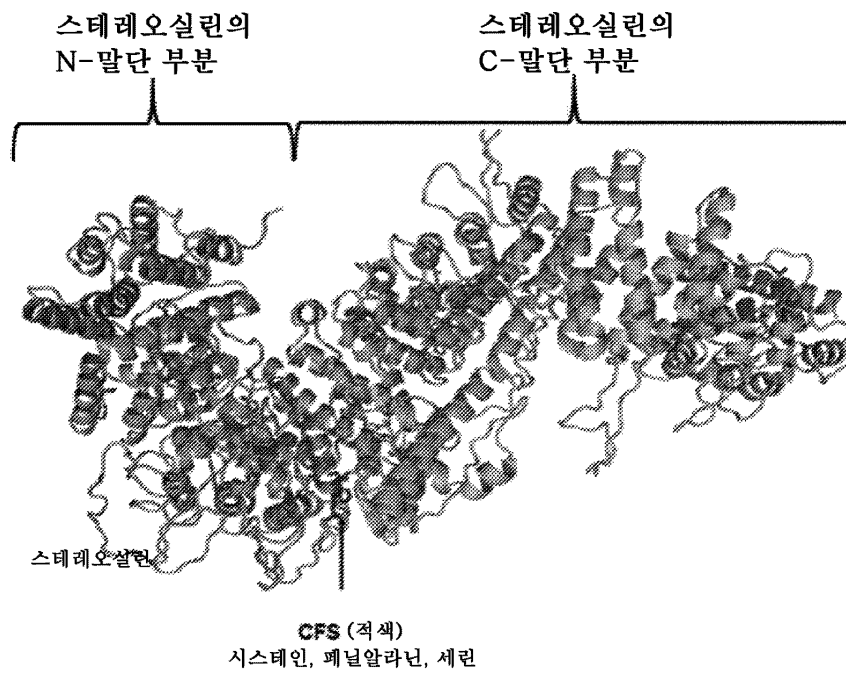
(SEQ ID NO:18)

도면 16

MalslqpqlllllslpgevtsIKIATRYKYLKGQNVYDIGVERDHNFKLNGFIASNCFCSP
LWDLQLREKSVWALRTLKAYLRMPPEDLQQLVLSAEMEAQGFLLTMLRSWAKLKVPQSEE
QAMGRLTALLLQRYPRLTSQLFIDMSPLIPFLAVPDLMRFPSSLANDSVLAAIRDHSSGMK
PEQKEALAKRLLAPELFGVEVPDWPQELLWAALPLLPHLPLESFLQLSPHQIQALEDSWPVAD
LGPGRHARHVLRSVLVNQSMEDGEEQVLRRLGSLACFLSPEELQSLVPLSDPMGPVEQGLLECAA
NGTLSPEGRVAYELLGVLRRSSGGTVLSPRELRVWAPLFPQLGLRFLQELSETQLRAMLPALQ
GASVTPAQAVLLFGRLLPKHDLSEELCSLHPLLPGLSPTQLQAIPKRVLVGACSCLGPELS
RLSACQIAALLQTRVKDGVKNMGAAGAGSAVCIPGQPTTWPDCLLPLLPLKLLQLDAAAL
ANRRLYRQLPWSEQQAQFLWKKMQVPTNLSLRNLQALGNLAGGMTCEFLQQISSMVDFLDVV
HMLYQLPTGVRESLRACIWTQLQRRMTMPEPELTTLGPELSELDTKLLLDLPIQLMDRLSND
SIMLVVEMVQGAPEQLLALTPHLQTALAERALKNLAPKETPISKEVLETGLPLVGLFGIEST
RRIPLPILLSHLSQLQGFCFLGETFATELGWLLLQEPVLGKPELWSQDEIEQAGRLVFTLSAE
AISSIPREALGPETLERLLGKHQSWEQSRVGHLCGESQLAHKKAALVAGIVHPAAEGLQEPV
PNCADIRGTFPAAWSATQISEMELSDFEDCLSLFAGDPGLGPEELRAAMGKAKQLWGPPRGF
RPEQILQLGRLLIGLGERELQELTLVDWGVLSLGLQIDGWSSMQLRAVVSSFLRQSGRHVSH
LDFIYLTALGYTVCGLRPEELQHISSWEFSQAALFLGSLHLPCSEEQLEVLAYLLVLPGGFG
PVSNWGPEIFTEIGTIAAGIPDLALSALLRGQIQGLTPLAISVIPAPKFAVVFNPIQLSSLT
RGQAVAVTPEQLAYLSPEQRRAVAWAQHEGKEIPEQLGRNSAWGLYDWFQASWALALPVSIF
GHLL**TRTRPL**eqkliseedl*

(SEQ ID NO:20)

도면17



도면18a

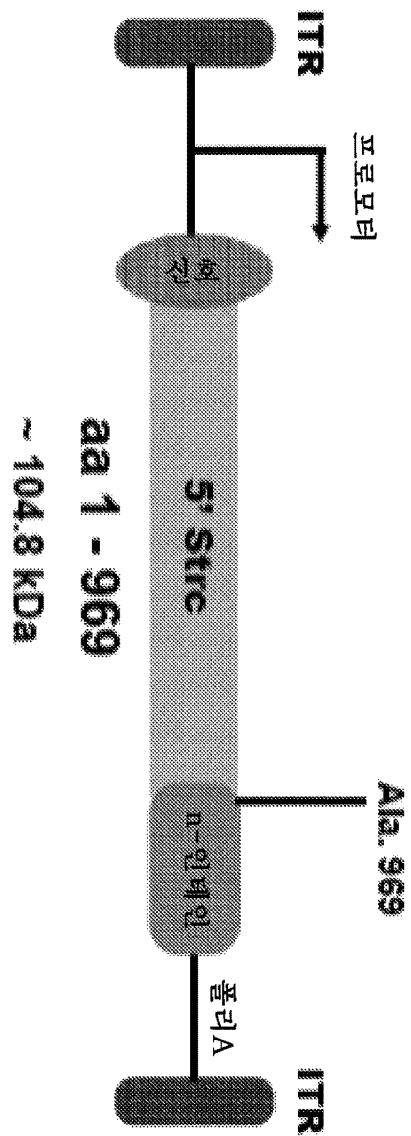
5' ATGgctctgagcctccagccccagctgctccttctcctgtcgctcctgccgcaggaagtga
acttcaGCCCCCTACTGGGCCTCAGTCTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTCGT
AGCCACTCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTCACGGTTCTCTGCGTTCCTGG
CCAACATTTCTTCATCCTTCCAGCTTGGGAGGATGGGGGAGGGACCGGTGGGAGAGCCCCCA
CCTCTCCAGCCCCCTGCACTTCGACTTCATGATTTCTCGTGACACTGAGAGGTAGCCCAGA
CTGGGAGCCAATGCTAGGGCTTCTGGGAGATGTGCTGGCACTCCTGGGACAGGAACAGACTC
CCCCGGACTTTTTGGTGCAACCAGGCAGGTGTACTGGGTGGACTTGTAGAGGCATTGTTGGGA
GCGTTAGTTCTTGAGGCCCCCTGCCCCACTCGACCCCCATGCACCCGTGATGGCCCTTC
TGA CTGTGCTCTGGCTGCTGATTGGTTGCCTTCTCTGATGTTGTTATTAGAGGGTACACGCT
GGCAGGCCCTGGTGAGTTGCAGCCAGTGTGGACCCAACCAATGCCACAGGTCTTGATGGT
AGAGAGCCAGCTCCTCACTTTTTACAGGGTCTGCTGGGCTTGCTTACCCAGCAGGAGAGTT
GGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGTGGGGGCCCCCTCTATGCTG
CCTTCCAGGAGGGGCTACTGCGAGTCACTCATTCTCTGCAAGATGAGGTCTTTTCTATTATG
GGACAGCCAGAGCCTGATGCCAGTGGGCAGTGCCAGGGAGGCAACCTTCAACAGCTGCTTTT
ATGGGGCATGCGGAACAACCTTTCTTGGGACGCCCCGAGCACTGGGTTTTCTATCTGGATCAC
CACCTCCACCCCTGCTCTCCTGCACTGCCTGAGCAGAGGTGTGCCTCTGCCAGGGCTTCC
CAGCCTGCGGCTCACATCAGCCCTCGACAGCGGCGAGCCATCTCTGTGGAGGCCCTCTGCGA
GAACCACTCAGGCCAGAGCCACCTACAGCATCTCCAATTCTCCATCTACTTGCTCTGCC
AGCACATCAAGCCTGCCACCCCGCGGCCCCCTCCTACCACCCACGGCCTCCTCCTACCACC
CCACAGCCCCCTCCTACCCTACACAGCCCATTCCTGACACTACACAGCCCCCTCCTGTAC
CCCCAGGCCTCCTCCTACCACCCACCAACCCCTCCTAGCACAGCTGTCATCTGCCAGACAG
CTGTATGGTACGCAGTCTCGTGGGCACCAGGTGCCCAGGTTGGCTCCAAGCCTGCCATGAT
CAGTTTCTGATCAATTTCTGGATATGATCTGCGGCAACCTCTCATTTTACGCCCTGTCTGG
CCCCAGTCGTCCTTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCACCCCCCACTAGCTGTC
CACCAGGCCTGATCCCTGTGCCCCTCACCCAGAAATATTCTGGGGCTGTTTCTGGAGAAT
GAGACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCAGGCTGTGCCCCGAGGAA
CCAGGCTTGGGTTACAGCATGTGTGTGCGGGCCCCACCTTGGACGCCACTGATTTTCCACCGT
GCCGCTTGGACCTGTGGGGAACGCTGCCAGATGGGGGAGCTTCTGCTCATGGTCTGT
GCCAATGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTGGCTAGCAGGCCAATGCAG

도면18b

AATTAGTCGTGGAGGAAATGATACTTGCTTTCTAGAAGGCATGCTGGGCCCCTTGTTGCCCT
 CTCTGCCCCCTCTGGGACCATCCCCACTCTGTCTGGCTCCTGGTCCTTTTCTGCTTGGCATG
 TTATCCCAGTTGCCACGCTGTCAGTCCTCCGTGCCAGCCCTCGCCCACCCACGCGCCTACA
 TTACCTCTGCGCCTACTGACCTTCTTCTGGGTCCAGGGACTGGGGTGCCGAGACGCAGG
 GGATGTTAGGTCAAGCCCTGCTGCTCTCTAGTCTCCAGACAACCTGTTTATTCTGGGATGCC
 TTCCGCCCAGAGGGCCGAGAAGTGTACTGAGGACAGTCGGAGAGTACTTGCAGCGGGAAGA
 GCCAACCCACCAGGCTTAGACTCCTCCCTCAGCCTCGGCTCTGGTATGAGCAAGATGGAGC
 TTCTGTCTGCTTCAGTCCTGTACTGTGGGATCTACTCCAGAGAGAGAAGAGCGTTTGGGCC
 CTGAGGACCCTGGTGAAGGCCTACCTGCGCATGCCTCCAGAAGACCTTCAGCAGCTTGTGCT
 TTCAGCAGAGATGGAGGCTGCACAGGGCTTCTTGACGCTCATGCTTCGTTCTGGGCTAAGC
 TGAAGGTTCAACCATCCGAGGAGCAGGCCATGGGCCGCTGACAGCCTTGCTGCTCCAGCGG
 TACCCACGCTCACCTCCCAACTCTTTATCGACATGTACCGCTCATCCCCCTCCTGGCTGT
 CCCTGACCTCATGCGCTTCCACCGTCCCTTTTGGCCAACGACAGTGTCTGGCTGCCATCA
 GGGATCACAGCTCAGGAATGAAGCCTGAACAGAAGGAGGCCCTGGCAAAACGACTGCTGGCC
 CCTGAGCTGTTTGGAGAAGTGCCTGATTGGCCCCAGGAGCTGCTGTGGGCAGCCCTGCCTCT
 GCTTCCCCATCTGCCTCTGGAGAGCTTCTCCAGCTCAGCCCTCACCAGATCCAGGCCCTGG
 AGGATAGCTGGCCAGTAGCAGATCTTGGGCCGGGACACGCCCGACATGTGCTTCGTAGCCTA
 GTAAACCAGAGCATGGAGGATGGGGAGGAGCAGGTGCTCAGGCTTGGGTCCCTCGCC**TGCCT**
GTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGTGG
AGAAACGGATAGAATGCACAGTTTACTCTGTGCGATAACAATGGTAACATTTATACTCAGCCA
GTTGCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTGCAATACTGTCTGGAGGATGGAAG
TCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAG
ACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAATTAATAG 3'

(SEQ ID NO:51)

도면19



도면20

MalslqpqlllllslpgevtsAPTGPQSLDAGLSLLKS FVATLDQAPQRSLSQSRFSAFLA
 NISSSFQLGRMGEGPVGEPPPLQPPALRLHDFLVTLRGSPDWEPMGLLGDVLALLGQEQT
 RDFLVHQAGVLGGLVEALLGALVPGGPPAPTRPPCTRDGSPDCVLAADWLPSLMLLLEGTRW
 QALVQLQPSVDPTNATGLDGREPAPHFLQGLLGLLTPAGELGSEEALWGGLLRVTGAPLYAA
 FQEGLLRVTHSLQDEVFSIMGQPEPDASGQCQGGNLQQLLLWGMNRNLSWDARALGFLSGSP
 PPPPALLHCLSRGVPLPRASQPAAHISPRQRAISVEALCENHSGPEPPYSISNFSIYLLCQ
 HIKPATPRPPPTTPRPPPTTPQPPPTTTPQIPDTPQPPVTTPRPPPTTPQPPSTAVICQTA
 VWYAVSWAPGARGWLQACHDQFPDQFLDMICGNLSFSALSGPSRPLVKQLCAGLLPPPTSCP
 PGLIPVPLTPEIFWGCFLNETLWAERLCVEDSLQAVPPRNQAWVQHVCRGPTLDATDFPPC
 RVGPCGERCPDGGSFLLMVCANDTLYEALVPFWAWLAGQCRISRGNDTCFLEGMLGPLLPS
 LPPLGPSPLCLAPGPFLLGMLSQLPQCQSSVPALAHPTRLHYLLRLLTFLLGPGTGGAETQG
 MLGQALLSSSLPDNCSFWDAFRPEGRRSVLRVTGEYLQREEPTPPGLDSSLSLGSMSKME
 LSCFSPVLWDLQREKSVWALRTLKAYLRMPPELQQLVLSAEMEAAGFLTLMLRSWAKL
 KVQPSEEQAMGRLTALLQRYPRLTSQLFIDMSPLIPFLAVPDLMRFPSSLANDSVLAIR
 DHSSGMKPEQKEALAKRLLAPELFGVDPWPQELLWAALPLLPHLPLESFLQLSPHQIQA
 DSWPVADLGPGHARHVLRLSVNQSMEDGEEQVLRGLSLACLSYTEILTVEYGLLPIGKIVE
KRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFM TVDGQMLPID
EIFERELDLMRVDNLPN**

(SEQ ID NO:52)

도면21a

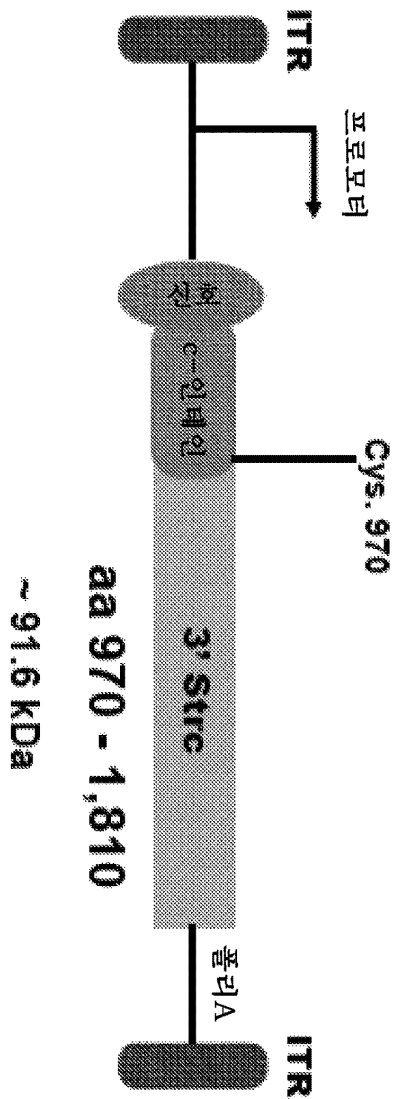
5'ATG~~gctctgagcctccagccccagctgctccttctcctgtcgctcctgccgcaggaagtga~~
~~acttcaATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGT~~
~~CGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTTCATAGCTTCTAAT~~TGTTTCCTGAGTC
CTGAGGAGCTACAGAGTCTGGTGCCCTTGAGTGATCCAATGGGGCCTGTAGAACAGGGTCTG
CTGGAATGTGCGGCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATATGAACTTCTGGG
AGTGTTCGCTTCATCTGGAGGAACGTCTTAAGCCCCGAGAGCTGAGGGTCTGGGCACCTC
TCTTTCCCCAGCTGGGCCTCCGCTTCCTGCAGGAGCTCTCAGAGACCCAGCTTAGAGCCATG
CTTCCTGCCCTACAGGGAGCCAGTGTACACCTGCCAGGCTGTTCTGTTGTTTGGAAAGGCT
CCTTCCTAAGCATGATCTGTCCCTGGAGGAACCTCTGCTCCCTGCACCTCTCCTGCCAGGTC
TCAGCCCCCAGACACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGGTGCTTGTTCCTGCCTG
GGCCCTGAACTGTCAAGGCTTTCAGCTTGCCAGATTGCAGCTCTGCTGCAGACCTTTCGGGT
AAAAGATGGTGTTAAAAATATGGGTGCAGCAGGTGCCGGCTCAGCCGTGTGCATTCTGGGC
AGCCCACCACTTGGCCAGACTGCCTGCTTCCCCTGCTCCCATTAAGCTGTACAGCTGGAC
GCTGCAGCTCTTCTGGCAAACCGAAGACTCTATCGGCAGCTGCCTTGGTCTGAGCAACAGGC
ACAGTTTCTCTGGAAGAAAATGCAAGTGCCTACCAACCTGAGCCTGAGGAATCTGCAGGCTC
TGGGCAACTTGGCAGGAGGCATGACCTGCGAGTTTCTGCAGCAGATCAGCTCAATGGTTGAC
TTTCTTGATGTGGTACACATGCTCTACCAGCTGCCCACTGGTGTTCGAGAGAGCCTGCGGGC
CTGTATCTGGACAGAGCTACAGCGGAGGATGACAATGCCAGAGCCAGAGCTGACCACCCTAG
GGCCAGAAGTGAAGTGAACCTTGACACAAAGCTACTCCTGGACTTGCCGATCCAGCTGATGGAC
AGATTGTCCAATGATTCCATTATGTTGGTGGTGGAGATGGTCCAAGGCGCTCCAGAGCAGCT
GCTGGCACTGACCCCACTCCACCAGACAGCCTTGGCAGAGCGAGCACTTAAAAACCTGGCTC
CAAAGGAGACCCCAATCTCCAAAGAAGTGCTGGAGACACTGGGCCCCTTGGTTGGATTCTCTG
GGAATAGAGAGCACGCGACGGATCCCTTTACCCATTCTACTGTCTCATCTCAGTCAGCTGCA
GGGCTTCTGCCTAGGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTGTTGCAGGAGCCTG
TTCTTGGAAAACCGAATTGTGGAGCCAGGATGAAATAGAGCAAGCTGGACGCCTAGTATTC
ACTCTGTCTGCTGAGGCTATTTCTCGATCCCCAGGGAGGCTTTGGGCCCAGAGACACTGGA
GAGGCTTCTGGGAAAGCATCAAAGCTGGGAGCAGAGCAGAGTGGGCCATCTGTGTGGGGAGT
CACAGCTTGCCACAAAGAAAGCAGCTCTGGTAGCTGGGATTGTGCATCCAGCTGCTGAGGGT
CTCCAAGAGCCTGTACCAAACCTGTGCAGACATACGGGGAACCTTCCAGCGGCCTGGTCTGC

도면21b

GACACAAATCTCAGAGATGGAACCTCTCAGACTTTGAAGACTGCCTGTCACTATTTGCTGGAG
 ATCCAGGACTTGGTCCTGAGGAACTACGGGCAGCCATGGGCAAGGCCAAGCAGTTGTGGGGT
 CCCCCTCGAGGATTCCTGCTGAGCAGATCTTGAGCTGGGCCGTCTCCTGATAGGTCTAGG
 AGAACGGGAAGTGCAGGAGCTTACCTTGGTGGACTGGGGTGTGCTGAGCAGCCTGGGGCAAA
 TAGATGGCTGGAGTTCCATGCAGCTCCGAGCCGTGGTCTCCAGTTTCCTAAGGCAGAGTGGT
 CGGCATGTGAGCCACCTGGACTTCATTTATCTGACAGCAGTGGGTTACACAGTCTGTGGATT
 GCGACCAGAGGAGTTACAGCACATCAGCAGTTGGGAGTTTAGCCAAGCAGCTCTCTCCTGG
 GTAGCTTGCATCTCCCGTGCTCTGAGGAACAGCTGGAAGTTCTGGCCTATCTCCTTGTGTTG
 CCTGGTGGCTTTGGCCAGTCAGTAACTGGGGCCTGAGATCTTCACTGAAATTGGCACAAAT
 AGCAGCTGGCATCCAGACCTGGCTCTTTCAGCATTACTGCGGGACAGATCCAAGGCCTGA
 CTCCTCTTGCCATTTCTGTCAATCCTGCTCCCAAGTTTGCAGTGGTCTTCAACCCCATCCAG
 TTATCTAGTCTCACCAGGGGTCAGGCCGTAGCTGTTACTCCTGAACAGCTGGCCTATCTGAG
 TCCTGAGCAGCGGCAGCAGTTGCATGGGCCCAACACGAAGGGAAGGAGATCCCAGAGCAGC
 TGGGTCGAAACTCAGCCTGGGGTCTCTACGACTGGTTCCAAGCCTCCTGGGCCCTGGCATTG
 CCCGTCAGCATTTTTGGCCACCTATTAA**ACGCGTACGCGGCCGCTC**gagcagaaactcatctc
 agaagaggatctgTAATAG 3'

(SEQ ID NO:53)

도면22

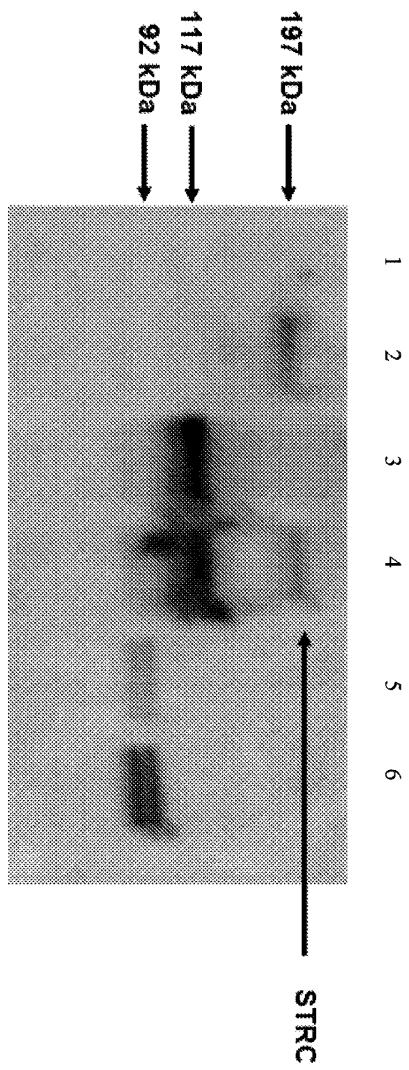


도면23

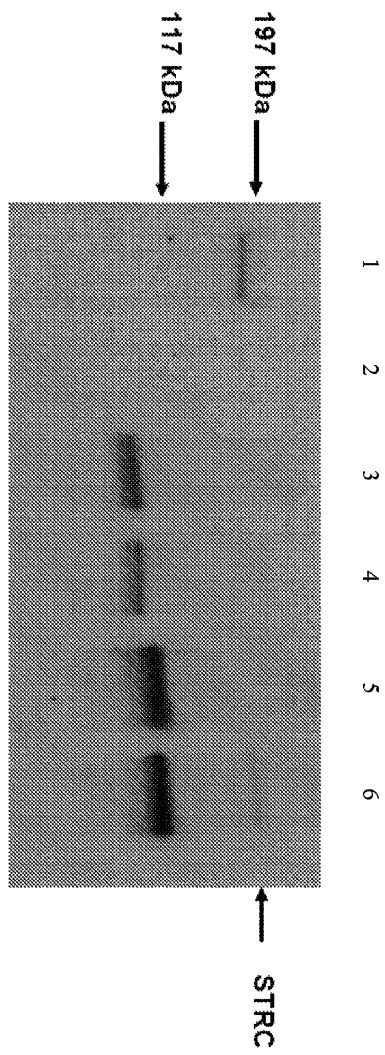
MalslqpqlllllslpgevtsIKIATRKYLGKQNVYDIGVERDHNFAKNGFIASNCFLSP
 EELQSLVPLSDPMGPVEQGLLECAANGTLSPEGRVAYELLGVLRSSGGTVLSPRELRVWAPL
 FPQLGLRFLQELSETQLRAMLPALQGASVTPAQAVLLFGRLLPKHDLSEELCSLHPLLPG
 SPQTLQAIKRVLVGACSCLGPELSRLSACQIAALLQTFRVKDGVKNMGAAGAGSAVCIPGQ
 PTTWPDCLLPLPLKLLQLDAAALLANRRLYRQLPWSEQQAQFLWKKMQVPTNLSLRNLQAL
 GNLAGGMTCEFLQQISSMVDFLDVVHMLYQLPTGVRESLRACIWTTELQRRMTMPEPELTTLG
 PELSELDTKLLLDLPIQLMDRLSNDSIMLVVEMVQGAPEQLLALTPLHQTALAERALKNLAP
 KETPISKEVLETLGPLYVGLGIESTRRIPLPILLSHLSQLQGFCLETTFATELGWLLLQEPV
 LGKPELWSQDEIEQAGRLVFTLSAEAISSIPREALGPETLERLLGKHQSWEQSRVGHL CGES
 QLAHKAALVAGIVHPAAEGLQEVPVNCADIRGTFPAAWSATQISEMELSDFEDCLSLFAGD
 PGLGPEELRAAMGKAKQLWGPPRGFRPEQILQLGRLLIGLGERELQELTLVDWGVLSLGGI
 DGWSSMQLRAVVSSFLRQSGRHVSHLDFIYLTALGYTVCGLRPEELQHISSWEFSQAALFLG
 SLHLPCSEEQLEVLAYLLVLPGGFGPVSNWGPEIFTEIGTIAAGIPDLALSALLRGQIQGLT
 PLAISVIPAPKFAVVFNPIQLSSLTRGQAVAVTPEQLAYLSPEQRRAVAWAQHEGKEIPEQL
 GRNSAWGLYDWFQASWALALPVSIFGHLL**TRTRPL**eqkliseedl**

(SEQ ID NO:54)

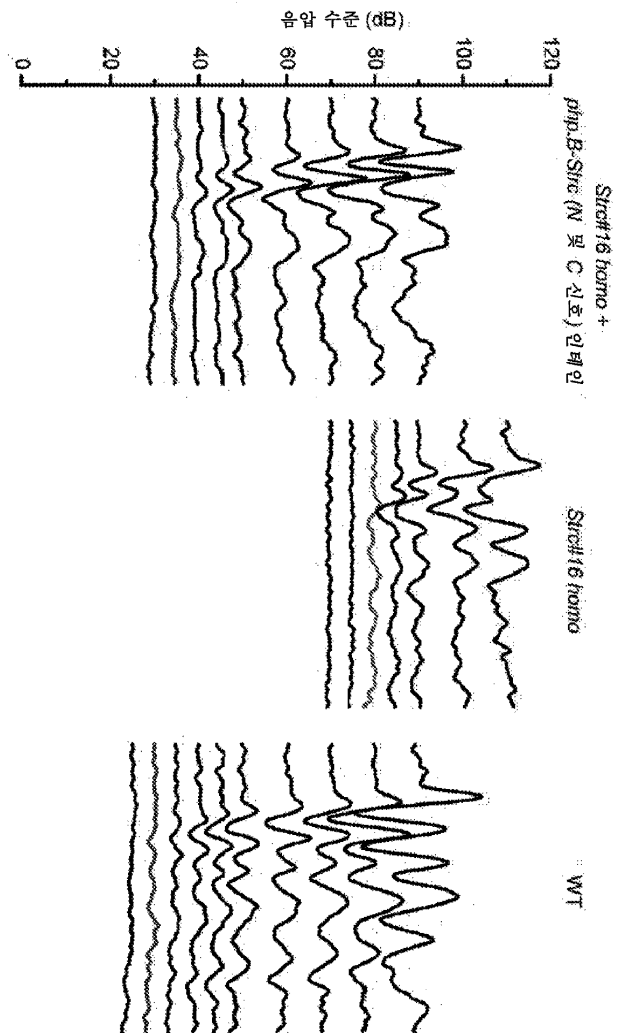
도면24



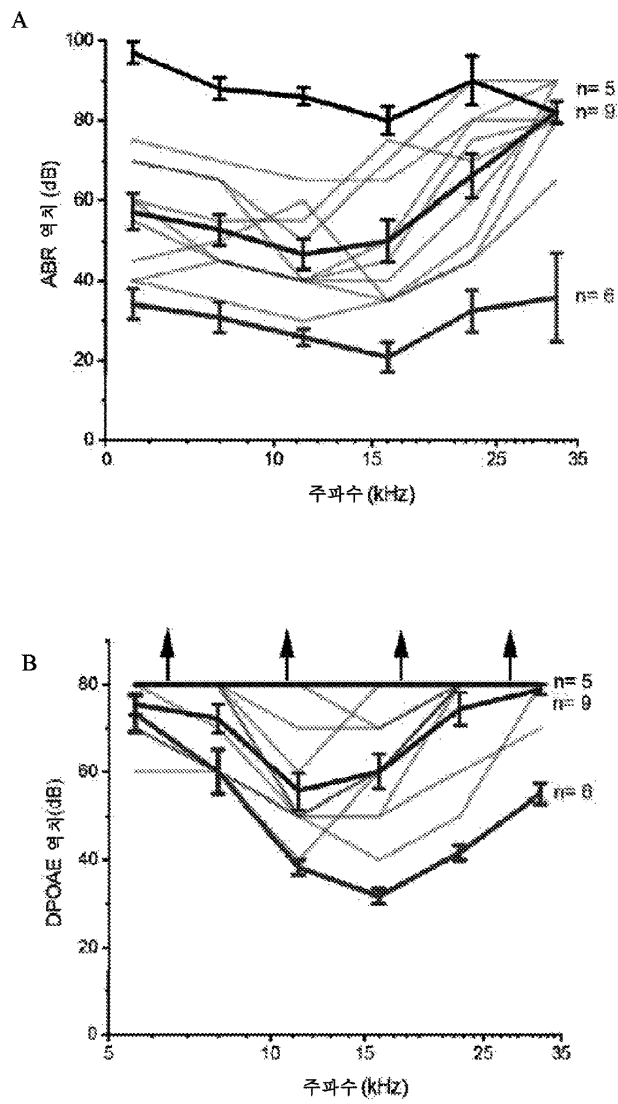
도면25



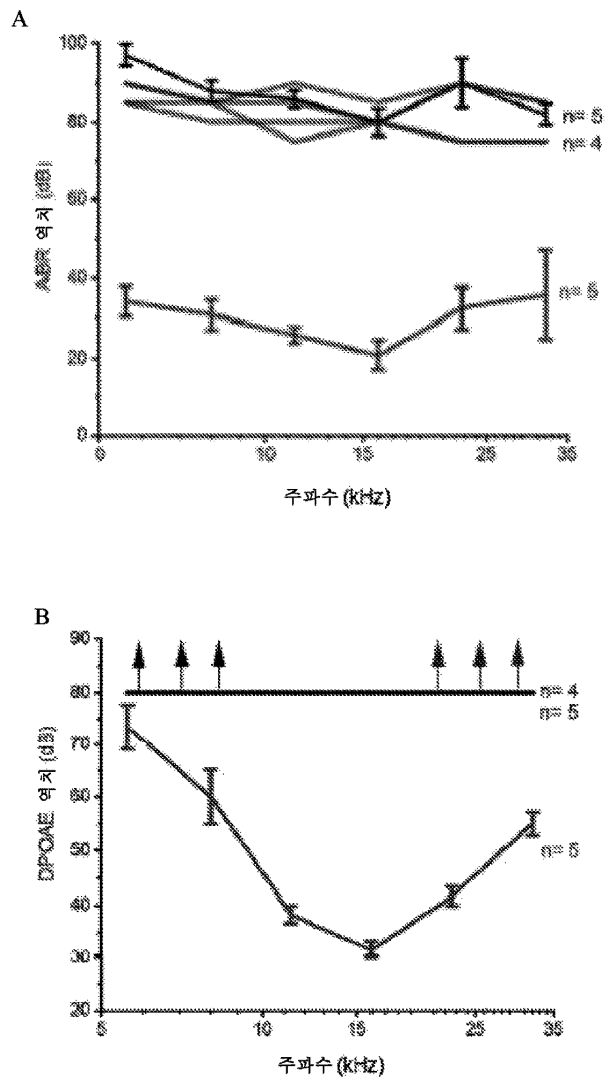
도면26



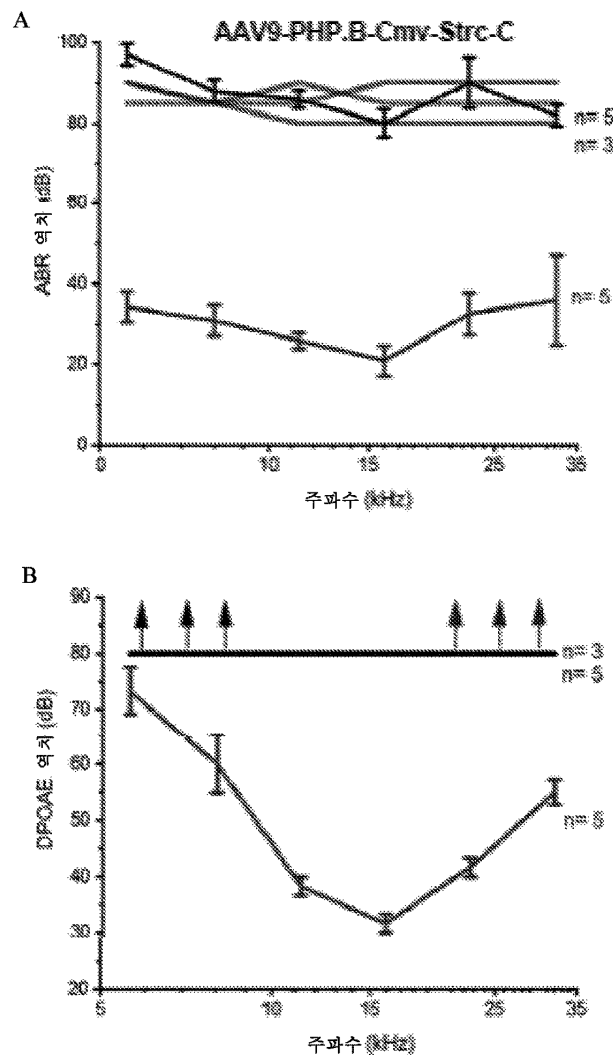
도면27



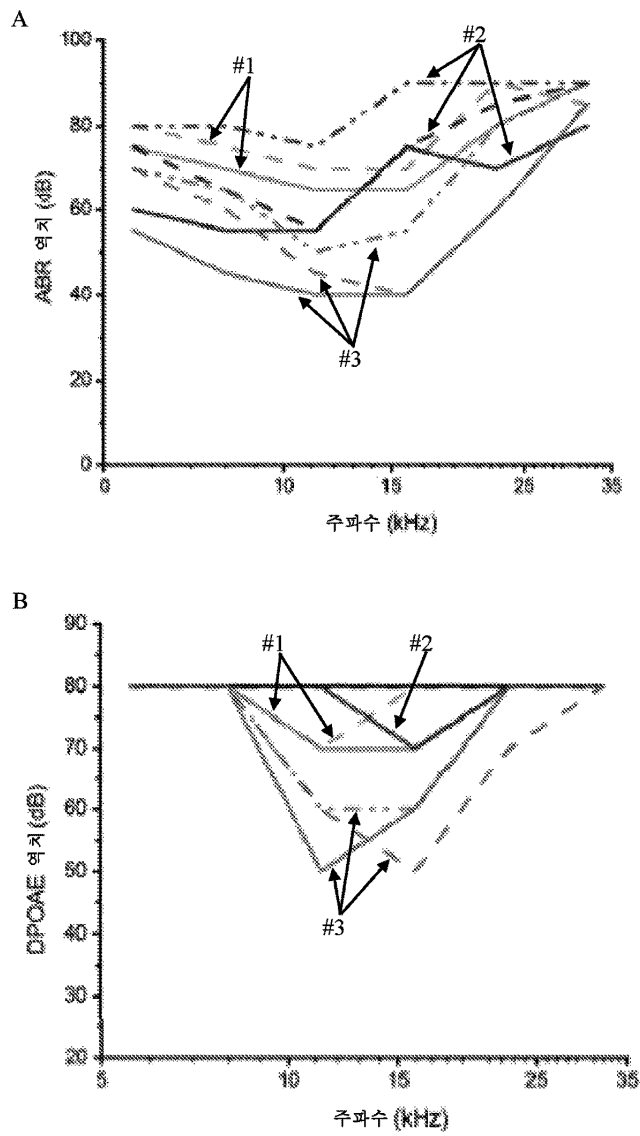
도면28



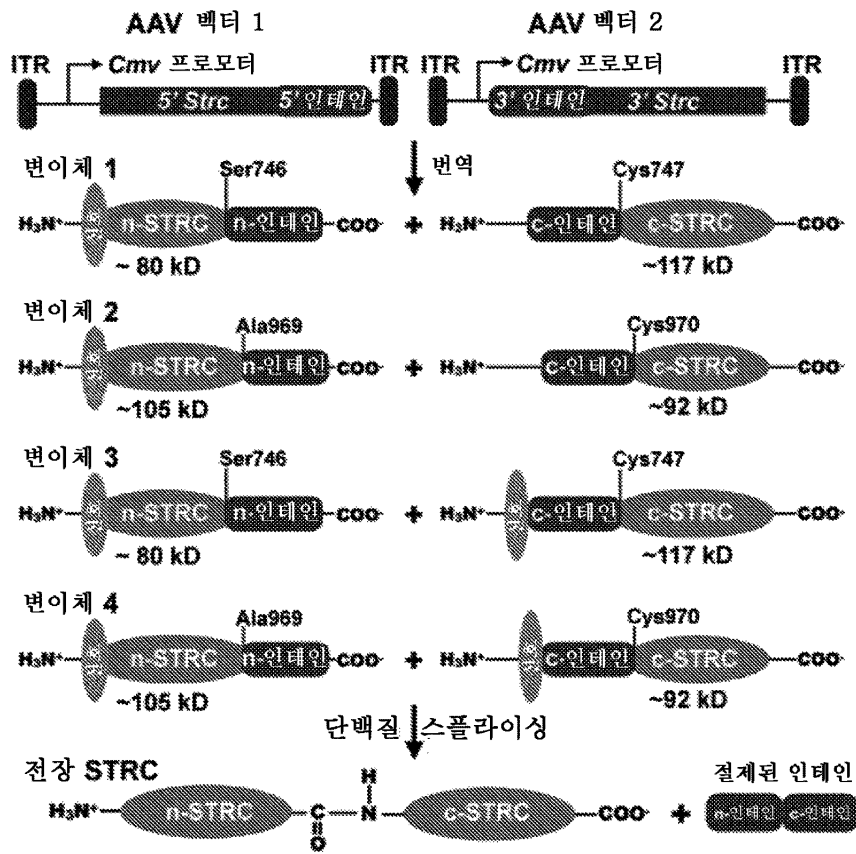
도면29



도면30



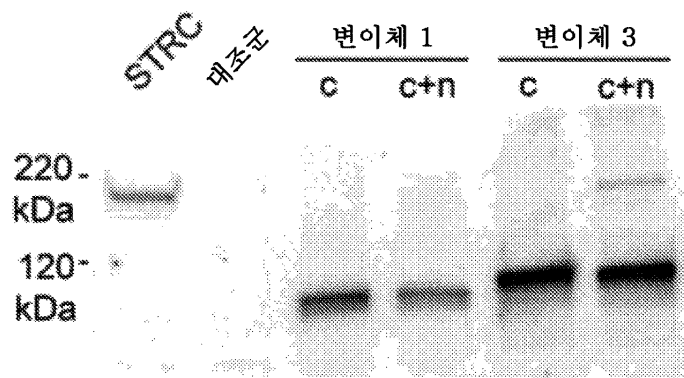
도면31a



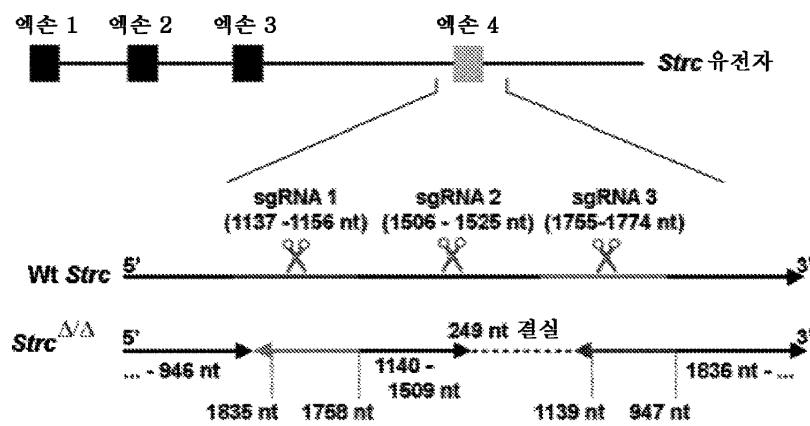
도면31b

분할 부위	AA 서열	SEQ ID NO:
변이체 1 & 3 (SEQ ID NO:26의 Cys747)	ELLSC <u>F</u> SPV	60
변이체 2 & 4 (SEQ ID NO:26의 Cys970)	GSLAC <u>F</u> LSP	61
SEQ ID NO:25의 Cys709	ELLAC <u>F</u> SPV	62
SEQ ID NO:25의 Cys934	GPLAC <u>F</u> LSP	63

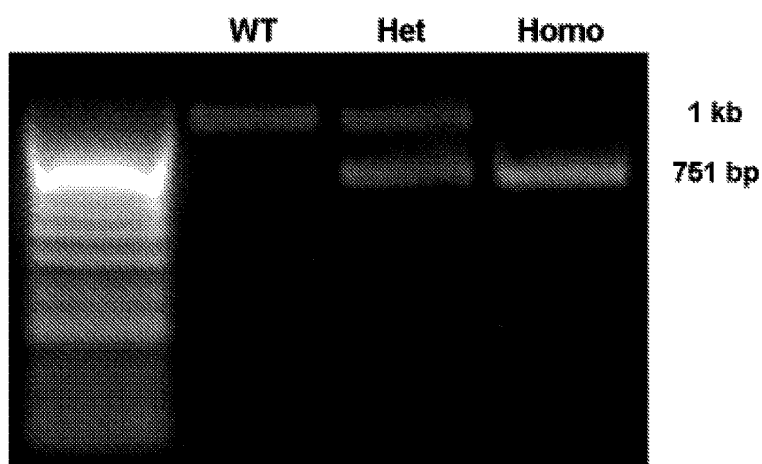
도면31c



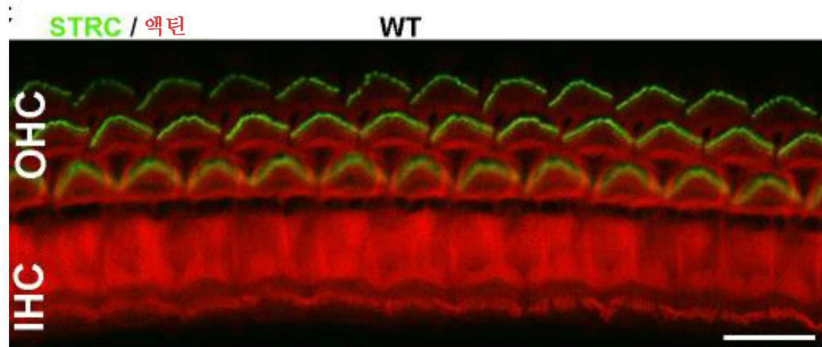
도면32a



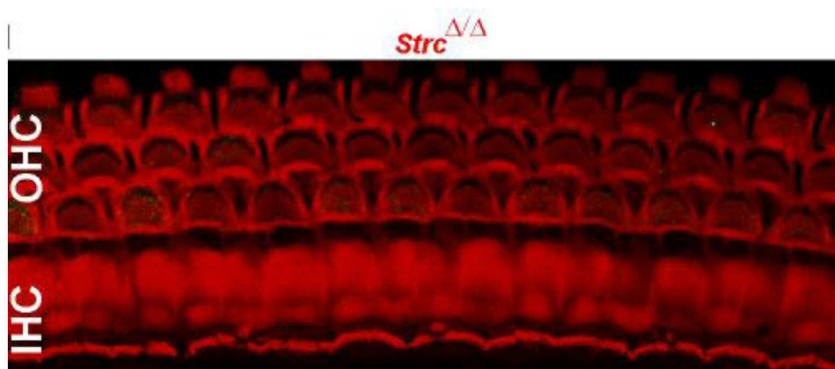
도면32b



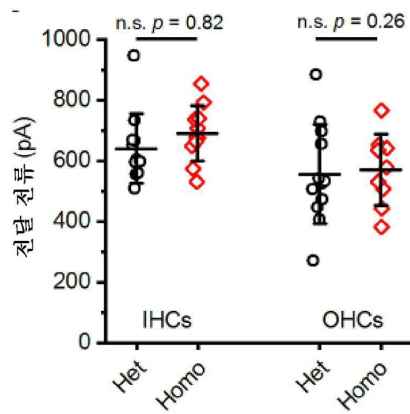
도면32c



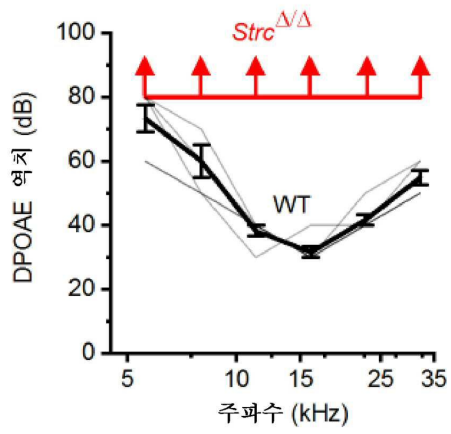
도면32d



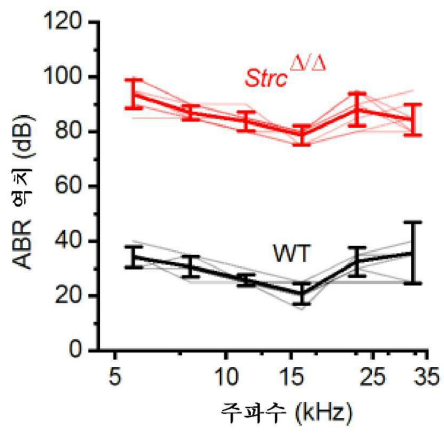
도면32e



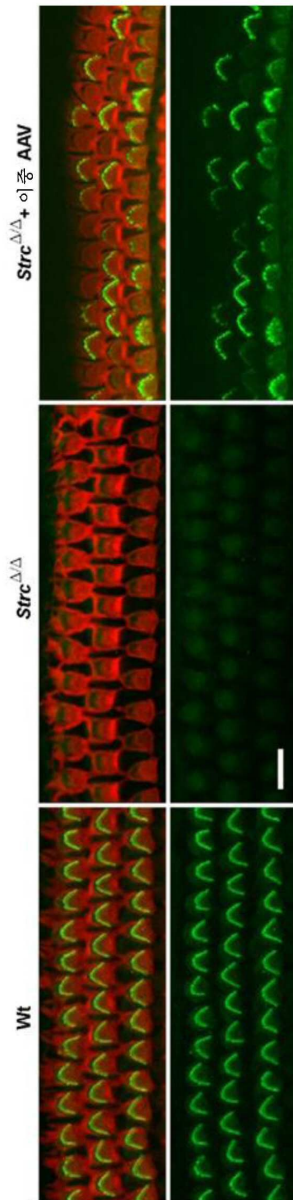
도면32f



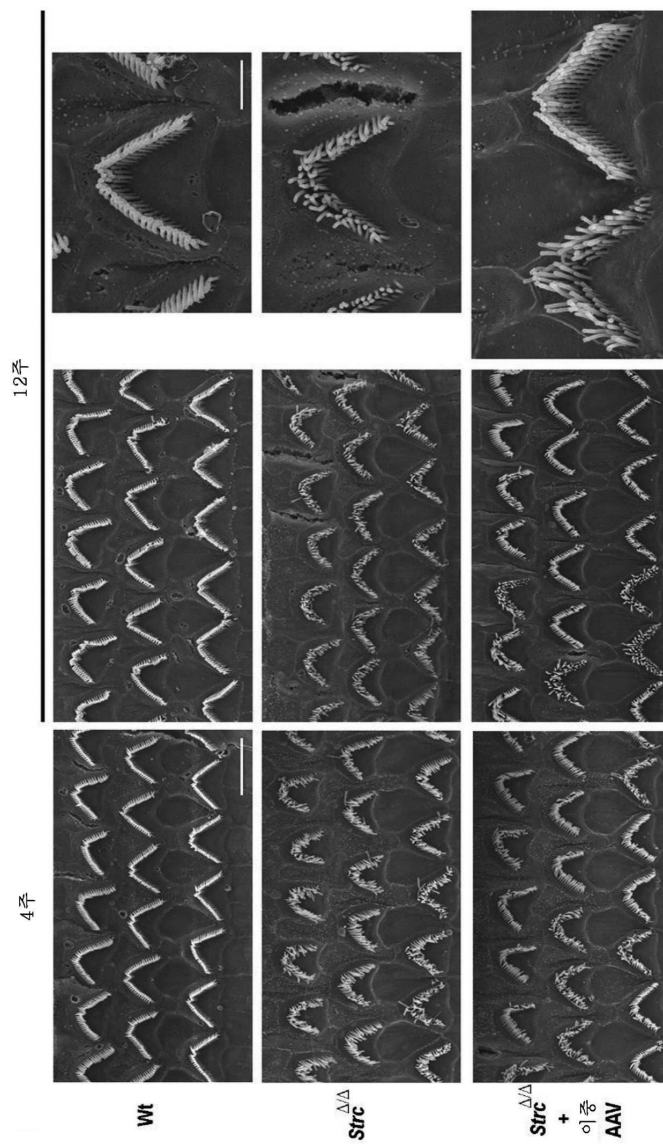
도면32g



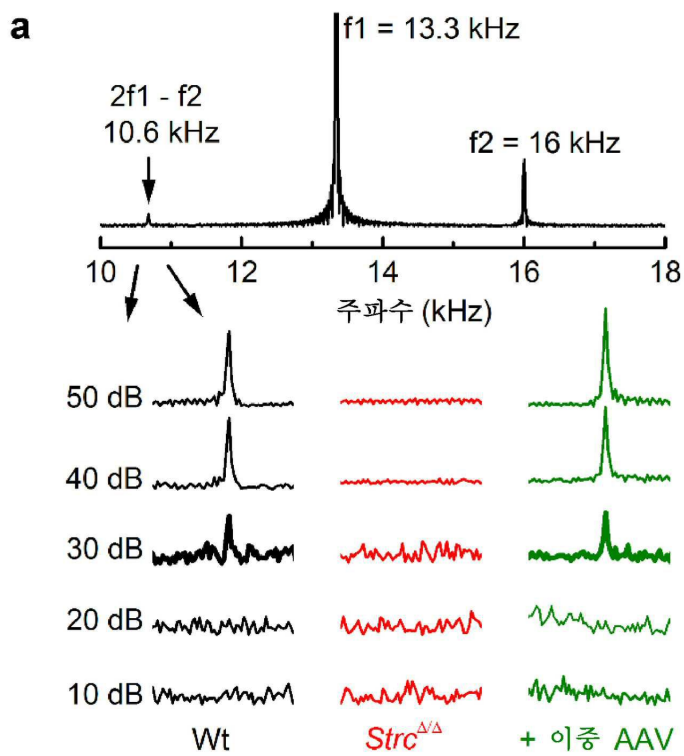
도면33a



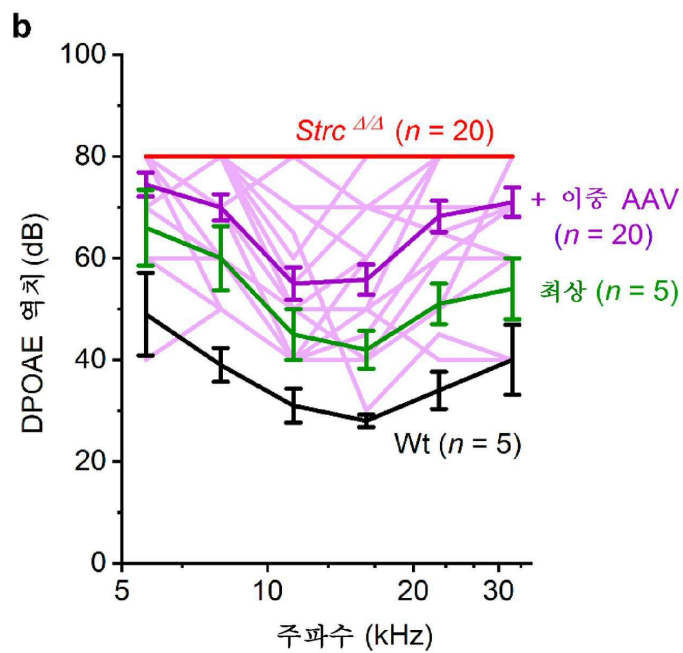
도면33b



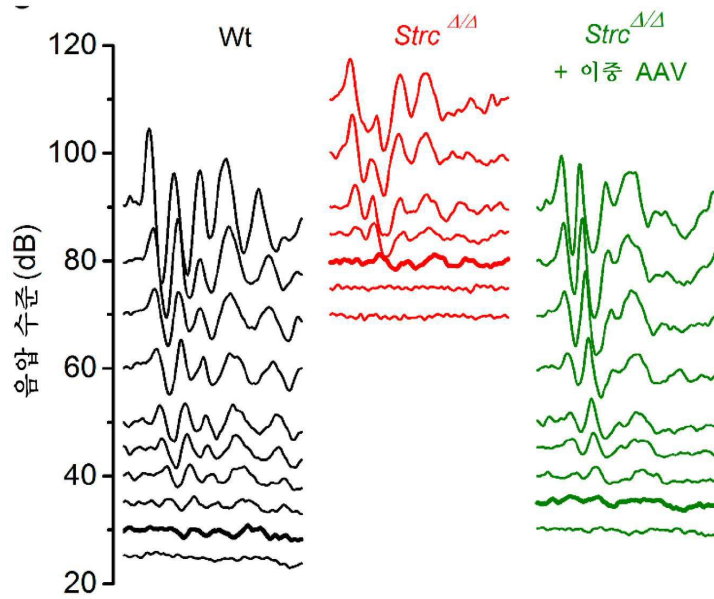
도면34a



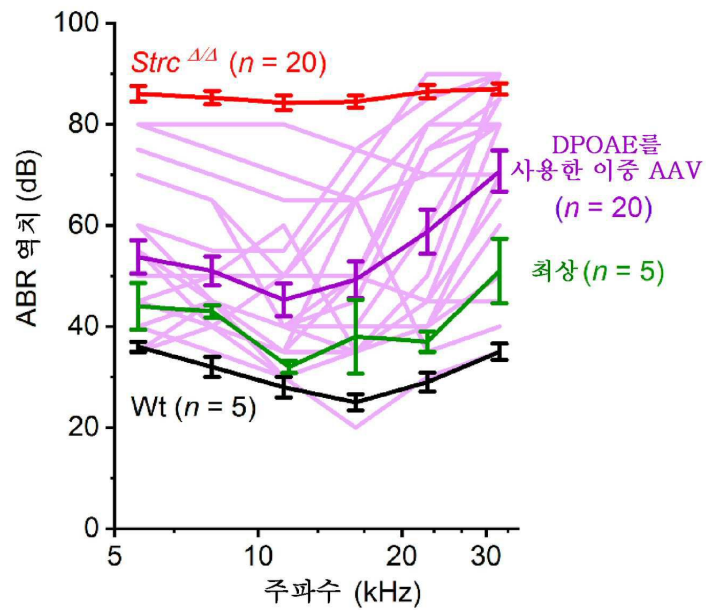
도면34b



도면34c



도면34d



서열 목록

SEQUENCE LISTING

<110> THE CHILDREN'S MEDICAL CENTER CORPORATION

<120> LARGE GENE VECTORS AND DELIVERY AND USES THEREOF

<130> 167705.020601/PCT

<140> PCT/US2021/016720

<141> 2021-02-05

<150> 62/971,555

<151> 2020-02-07

<160> 63

<170> PatentIn version 3.5

<210> 1

<211> 5262

<212> DNA

<213> Homo sapiens

<400> 1

gtgactctgg cccctactgg gcctcattcc ctggaccctg gtctctcctt cctgaagtca	60
ttgtcttcca ctctggacca ggctccccag ggctccctga gccgctcacg gtctcttaca	120
ttctctggcca acattttctt ttcttttgag cctgggagaa tgggggaagg accagtagga	180
gagccccac ctctccagcc gcctgctctg cggtccatg attttctagt gacactgaga	240
ggtagccccg actgggagcc aatgctaggg ctgctagggg atatgctggc actgctggga	300
caggagcaga ctccccaga ttctctggtg caccaggcag gggtgctggg tggacttgtg	360
gaggtgctgc tgggagcctt agttctctgg ggcccccta cccaactcg gccccatgc	420
accctgatg ggccgtctga ctgtgtcctg gctgtgact ggttgccctt tctgtgctg	480
ttgttagagg gcacacgtg gcaagctctg gtgcaggtgc agcccagtgt ggacccacc	540
aatgccacag gcctcgatgg gagggaggca gctcctcact ttttgaggg tctgttgggt	600
ttgcttacc caacagggga gctaggctcc aaggaggctc tttggggcgg tctgtacgc	660
acagtggggg cccccctta tgcctcctt caggaggggc tgcctcgtgt cactcactcc	720
ctgcaggatg aggtcttctc cattttgggg cagccagagc ctgatacaca tgggcagtgc	780
caggagggtt accttcaaca gctgtcttta tggggcgtcc ggacacact ttctgggat	840
gtccaggcgc tgggctttct gtctggatca ccacccccac cccctgcct ccttactgc	900
ctgagcacgg gcgtgcctct gccagagct tctcagcgt cagccacat cagccacgc	960
caacggcgag ccatcactgt ggaggccctc tgtgagaacc acttaggcc agcaccacc	1020
tacagcattt ccaacttctc cactcacttg ctctgccagc acaccaagcc tgccactcca	1080
cagccccatc ccagcaccac tgccatctgc cagacagctg tgtggtatgc agtgtcctgg	1140
gcaccaggtg cccaaggctg gctacaggcc tgccacgacc agtttctga tgagttttg	1200
gatgcgatct gcagtaacct ctctttttca gccctgtctg gctccaaccg ccgctgggtg	1260
aagcggtctt gtgctggcct gctccaccc cctaccagct gccctgaagg cctgccccct	1320
gttccccctc cccagacat cttttggggc tgcttcttgg agaagtagac tctgtgggct	1380

gagcgactgt gtggggaggc aagtctacag gctgtgcccc ccagcaacca ggcttgggtc 1440

cagcatgtgt gccagggccc caccacagat gtcactgcct cccaccatg ccacattgga 1500

ccctgtgggg aacgctgccc ggatgggggc agcttccctg tgatggctctg tgccaatgac 1560

accatgtatg aggtcctggt gcccttctgg ccttggctag caggccaatg caggataagt 1620

cgtgggggca atgacacttg ctctctagaa gggtgctgg gcccccttct gccctctctg 1680

ccaccactgg gaccatcccc actctgtctg acccctggcc ctttctctct tggcatgcta 1740

tcccagttgc cacgtgtgca gtcctctgtc ccagctcttg ctacccccac acgcctacac 1800

tatctctcc gccctgtgac ctctctcttg ggtccagggg ctgggggcgc tgaggcccag 1860

gggatgctgg gtcgggcctt actgtctctc agtctcccag acaactgtct cttctgggat 1920

gcctttcgcc cagagggccg gcgcagtgtg ctacggacga ttggggaata cctggaacaa 1980

gatgaggagc agccaacccc atcaggcttt gaaccactg tcaaccccag ctctggtata 2040

agcaagatgg agctgctggc ctgctttagt cctgtgctgt gggatctgct ccagagggaa 2100

aagagtgttt ggccctgca gattctagt caggcgtacc tgcatatgcc ccagaaaac 2160

ctccagcagc tgggtcttcc agcagagagg gaggtgcac agggcttctt gacactcatg 2220

ctgcagggga agctgcaggg gaagctgcag gtaccacat ccgaggagca ggccctgggt 2280

cgctgacag cctgtctgt ccagcggtag ccacgcctca cctcccagct ctctattgac 2340

ctgtcaccac tcatcccttt cttggctgtc tctgacctga tgcgttccc accatccctg 2400

ttagccaacg acagtgtctt ggctgccatc cgggattaca gccaggaat gaggcctgaa 2460

cagaaggagg ctctggcaaa gcgactgtg gccctgaac tgtttgggga agtgctgcc 2520

tggccccagg agctgctgtg ggcagtgtg ccctgtctc ccacctccc tctggagaac 2580

tttttgacg tcagccctca ccagatccag gccttgagg atagctggcc agcagcaggt 2640

ctggggccag ggcattgccc ccatgtgtg cgcagcctgg taaaccagag tgtccaggat 2700

ggtgaggagc aggtacgcag gcttgggccc ctgcctgtt tcctgagccc tgaggagctg 2760

cagagcctag tgccctgag tgatccaacg ggccagtag aacgggggct gctggaatgt 2820

gcagccaatg ggaccctcag ccagaaagga cgggtggcat atgaacttct ggggtgtgtg 2880

cgctcatctg gaggagcggg gctgagcccc cgggagctgc gggcttgggc cctctcttc 2940

tctcagctgg gcttccgtt ccttcaggag ctgtcagagc ccagcttag agccatgctt 3000

cctgtcctgc agggaactag tgttacacct gctcaggctg tcctgtgtgt tggacggctc 3060

cttctaggc acgatctatc cctggaggaa ctctgtctct tgcaccttct gctaccaggc 3120

ctcagccccc agacactcca ggccatccct aggcgagtc tggtcggggc ttgttcctgc	3180
ctggccctg aactgtcacg cctctcagcc tgccagaccg cagcactgct gcagaccttt	3240
cgggttaaag atgggtgtaa aaatatgggt acaacagggt ctgggtccagc tgtgtgtatc	3300
cctgggtcagc ctattcccac caccitggcca gactgcctgc ttccctgct cccattaaag	3360
ctgctacaac tggattcctt ggctcttctg gcaaatcgaa gacgctactg ggagctgccc	3420
tggctgagc agcaggcaca gtttctctgg aagaagatgc aagtaccac caaccttacc	3480
ctcaggaatc tgcaggctct gggcacccctg gcaggaggca tgtcctgtga gtttctgcag	3540
cagatcaact ccatggtaga ctctcttgaa gtggtgcaca tgatctatca gctgcccact	3600
agagttcgag ggagcctgag ggccctgtatc tgggcagagc tacagcggag gatggcaatg	3660
ccagaaccag aatggacaac tgtagggcca gaactgaacg ggctggatag caagctactc	3720
ctggacttac cgatccagtt gatggacaga ctatccaatg aatccattat gttggtggtg	3780
gagctggtgc aaagagctcc agagcagctg ctggcactga ccccccctca ccaggcagcc	3840
ctggcagaga gggcactaca aaacctggct ccaaaggaga ctccagtctc aggggaagtg	3900
ctggagacct taggcccttt gggttgattc ctggggacag agagcacacg acagatcccc	3960
ctacagatcc tgctgtccca tctcagtcag ctgcaaggct tctgcctagg agagacattt	4020
gccacagagc tgggatggct gctattgcag gagtctgttc ttgggaaacc agagttgttg	4080
agccaggatg aagtagagca agctggacgc ctagtattca ctctgtctac tgaggcaatt	4140
tccttgatcc ccagggaggc ctgggtcca gagacctgg agcggttctc agaaaagcag	4200
cagagctggg agcagagcag agttggacag ctgtgtaggg agccacagct tgctgccaag	4260
aaagcagccc tggtagcagg ggtggtgcga ccagctgctg aggatcttcc agaacctgtg	4320
ccaaattgtg catagtacg agggacattc ccagcagcct ggtctgcaac ccagattgca	4380
gagatggagc tctcagactt tgaggactgc ctgacattat ttgcaggaga cccaggactt	4440
gggcctgagg aactgcgggc agccatgggc aaagcaaac agttgtgggg tccccccgg	4500
ggatttcgtc ctgagcagat cctgcagctt ggtaggctct taataggtct aggagatcgg	4560
gaactacagg agctgatcct agtggactgg ggagtgtga gcaccctggg gcagatagat	4620
ggctggagca ccactcagct ccgcattgtg gtctccagtt tcctacggca gagtggtcgg	4680
catgtgagcc acctggactt cgttcatctg acagcgtgg gttatactct ctgtggactg	4740
cggccagagg agctccagca catcagcagt tgggagtca gccaaagcagc tctcttctc	4800
ggcaccctgc atctccagtg ctctgaggaa caactggagg ttctggcca cctacttgta	4860
ctgcctgggtg ggtttggccc aatcagtaac tgggggcctg agatcttcac tgaaattggc	4920
accatagcag ctgggatccc agacctggct ctttcagcac tgctgcgggg acagatccag	4980

ggcgttactc ctcttgccat ttctgtcatc cctcctccta aatttgctgt ggtgtttagt 5040
 cccatccaac tatctagtct caccagtgtc caggctgtgg ctgtcactcc tgagcaaag 5100
 gcctttctga gtcctgagca gcgacgagca gttgcatggg cccaacatga gggaaaggag 5160
 agcccagaac agcaaggtcg aagtacagcc tggggcctcc aggactggtc acgaccttcc 5220

tggtccttgg tattgactat cagcttcctt ggccacctgc ta 5262

<210> 2

<211> 1754

<212> PRT

<213> Homo sapiens

<400> 2

Val Thr Leu Ala Pro Thr Gly Pro His Ser Leu Asp Pro Gly Leu Ser

1 5 10 15

Phe Leu Lys Ser Leu Leu Ser Thr Leu Asp Gln Ala Pro Gln Gly Ser

20 25 30

Leu Ser Arg Ser Arg Phe Phe Thr Phe Leu Ala Asn Ile Ser Ser Ser

35 40 45

Phe Glu Pro Gly Arg Met Gly Glu Gly Pro Val Gly Glu Pro Pro Pro

50 55 60

Leu Gln Pro Pro Ala Leu Arg Leu His Asp Phe Leu Val Thr Leu Arg

65 70 75 80

Gly Ser Pro Asp Trp Glu Pro Met Leu Gly Leu Leu Gly Asp Met Leu

85 90 95

Ala Leu Leu Gly Gln Glu Gln Thr Pro Arg Asp Phe Leu Val His Gln

100 105 110

Ala Gly Val Leu Gly Gly Leu Val Glu Val Leu Leu Gly Ala Leu Val

115 120 125

Pro Gly Gly Pro Pro Thr Pro Thr Arg Pro Pro Cys Thr Arg Asp Gly

130 135 140

Pro Ser Asp Cys Val Leu Ala Ala Asp Trp Leu Pro Ser Leu Leu Leu

145 150 155 160

Leu Leu Glu Gly Thr Arg Trp Gln Ala Leu Val Gln Val Gln Pro Ser

165	170	175
Val Asp Pro Thr Asn Ala Thr Gly Leu Asp Gly Arg Glu Ala Ala Pro		
180	185	190
His Phe Leu Gln Gly Leu Leu Gly Leu Leu Thr Pro Thr Gly Glu Leu		
195	200	205
Gly Ser Lys Glu Ala Leu Trp Gly Gly Leu Leu Arg Thr Val Gly Ala		
210	215	220
Pro Leu Tyr Ala Ala Phe Gln Glu Gly Leu Leu Arg Val Thr His Ser		
225	230	235
240		
Leu Gln Asp Glu Val Phe Ser Ile Leu Gly Gln Pro Glu Pro Asp Thr		
245	250	255
Asn Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln Leu Leu Leu Trp Gly		
260	265	270
Val Arg His Asn Leu Ser Trp Asp Val Gln Ala Leu Gly Phe Leu Ser		
275	280	285
Gly Ser Pro Pro Pro Pro Pro Ala Leu Leu His Cys Leu Ser Thr Gly		
290	295	300
Val Pro Leu Pro Arg Ala Ser Gln Pro Ser Ala His Ile Ser Pro Arg		
305	310	315
320		
Gln Arg Arg Ala Ile Thr Val Glu Ala Leu Cys Glu Asn His Leu Gly		
325	330	335
Pro Ala Pro Pro Tyr Ser Ile Ser Asn Phe Ser Ile His Leu Leu Cys		
340	345	350
Gln His Thr Lys Pro Ala Thr Pro Gln Pro His Pro Ser Thr Thr Ala		
355	360	365
Ile Cys Gln Thr Ala Val Trp Tyr Ala Val Ser Trp Ala Pro Gly Ala		
370	375	380
Gln Gly Trp Leu Gln Ala Cys His Asp Gln Phe Pro Asp Glu Phe Leu		
385	390	395
400		
Asp Ala Ile Cys Ser Asn Leu Ser Phe Ser Ala Leu Ser Gly Ser Asn		
405	410	415

Arg Arg Leu Val Lys Arg Leu Cys Ala Gly Leu Leu Pro Pro Pro Thr
420 425 430

Ser Cys Pro Glu Gly Leu Pro Pro Val Pro Leu Thr Pro Asp Ile Phe
435 440 445

Trp Gly Cys Phe Leu Glu Asn Glu Thr Leu Trp Ala Glu Arg Leu Cys
450 455 460

Gly Glu Ala Ser Leu Gln Ala Val Pro Pro Ser Asn Gln Ala Trp Val
465 470 475 480

Gln His Val Cys Gln Gly Pro Thr Pro Asp Val Thr Ala Ser Pro Pro
485 490 495

Cys His Ile Gly Pro Cys Gly Glu Arg Cys Pro Asp Gly Gly Ser Phe
500 505 510

Leu Val Met Val Cys Ala Asn Asp Thr Met Tyr Glu Val Leu Val Pro
515 520 525

Phe Trp Pro Trp Leu Ala Gly Gln Cys Arg Ile Ser Arg Gly Gly Asn
530 535 540

Asp Thr Cys Phe Leu Glu Gly Leu Leu Gly Pro Leu Leu Pro Ser Leu
545 550 555 560

Pro Pro Leu Gly Pro Ser Pro Leu Cys Leu Thr Pro Gly Pro Phe Leu
565 570 575

Leu Gly Met Leu Ser Gln Leu Pro Arg Cys Gln Ser Ser Val Pro Ala
580 585 590

Leu Ala His Pro Thr Arg Leu His Tyr Leu Leu Arg Leu Leu Thr Phe
595 600 605

Leu Leu Gly Pro Gly Ala Gly Gly Ala Glu Ala Gln Gly Met Leu Gly
610 615 620

Arg Ala Leu Leu Leu Ser Ser Leu Pro Asp Asn Cys Ser Phe Trp Asp
625 630 635 640

Ala Phe Arg Pro Glu Gly Arg Arg Ser Val Leu Arg Thr Ile Gly Glu
645 650 655

Tyr Leu Glu Gln Asp Glu Glu Gln Pro Thr Pro Ser Gly Phe Glu Pro

660	665	670
Thr Val Asn Pro Ser Ser Gly Ile Ser Lys Met Glu Leu Leu Ala Cys		
675	680	685
Phe Ser Pro Val Leu Trp Asp Leu Leu Gln Arg Glu Lys Ser Val Trp		
690	695	700
Ala Leu Gln Ile Leu Val Gln Ala Tyr Leu His Met Pro Pro Glu Asn		
705	710	715
Leu Gln Gln Leu Val Leu Ser Ala Glu Arg Glu Ala Ala Gln Gly Phe		
725	730	735
Leu Thr Leu Met Leu Gln Gly Lys Leu Gln Gly Lys Leu Gln Val Pro		
740	745	750
Pro Ser Glu Glu Gln Ala Leu Gly Arg Leu Thr Ala Leu Leu Leu Gln		
755	760	765
Arg Tyr Pro Arg Leu Thr Ser Gln Leu Phe Ile Asp Leu Ser Pro Leu		
770	775	780
Ile Pro Phe Leu Ala Val Ser Asp Leu Met Arg Phe Pro Pro Ser Leu		
785	790	795
Leu Ala Asn Asp Ser Val Leu Ala Ala Ile Arg Asp Tyr Ser Pro Gly		
805	810	815
Met Arg Pro Glu Gln Lys Glu Ala Leu Ala Lys Arg Leu Leu Ala Pro		
820	825	830
Glu Leu Phe Gly Glu Val Pro Ala Trp Pro Gln Glu Leu Leu Trp Ala		
835	840	845
Val Leu Pro Leu Leu Pro His Leu Pro Leu Glu Asn Phe Leu Gln Leu		
850	855	860
Ser Pro His Gln Ile Gln Ala Leu Glu Asp Ser Trp Pro Ala Ala Gly		
865	870	875
880		
Leu Gly Pro Gly His Ala Arg His Val Leu Arg Ser Leu Val Asn Gln		
885	890	895
Ser Val Gln Asp Gly Glu Glu Gln Val Arg Arg Leu Gly Pro Leu Ala		
900	905	910

Cys Phe Leu Ser Pro Glu Glu Leu Gln Ser Leu Val Pro Leu Ser Asp

915 920 925

Pro Thr Gly Pro Val Glu Arg Gly Leu Leu Glu Cys Ala Ala Asn Gly

930 935 940

Thr Leu Ser Pro Glu Gly Arg Val Ala Tyr Glu Leu Leu Gly Val Leu

945 950 955 960

Arg Ser Ser Gly Gly Ala Val Leu Ser Pro Arg Glu Leu Arg Val Trp

965 970 975

Ala Pro Leu Phe Ser Gln Leu Gly Leu Arg Phe Leu Gln Glu Leu Ser

980 985 990

Glu Pro Gln Leu Arg Ala Met Leu Pro Val Leu Gln Gly Thr Ser Val

995 1000 1005

Thr Pro Ala Gln Ala Val Leu Leu Leu Gly Arg Leu Leu Pro Arg

1010 1015 1020

His Asp Leu Ser Leu Glu Glu Leu Cys Ser Leu His Leu Leu Leu

1025 1030 1035

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

1040 1045 1050

Leu Val Gly Ala Cys Ser Cys Leu Ala Pro Glu Leu Ser Arg Leu

1055 1060 1065

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

1070 1075 1080

Asp Gly Val Lys Asn Met Gly Thr Thr Gly Ala Gly Pro Ala Val

1085 1090 1095

Cys Ile Pro Gly Gln Pro Ile Pro Thr Thr Trp Pro Asp Cys Leu

1100 1105 1110

Leu Pro Leu Leu Pro Leu Lys Leu Leu Gln Leu Asp Ser Leu Ala

1115 1120 1125

Leu Leu Ala Asn Arg Arg Arg Tyr Trp Glu Leu Pro Trp Ser Glu

1130 1135 1140

Gln Gln Ala Gln Phe Leu Trp Lys Lys Met Gln Val Pro Thr Asn

1145	1150	1155
Leu Thr Leu Arg Asn Leu Gln	Ala Leu Gly Thr Leu	Ala Gly Gly
1160	1165	1170
Met Ser Cys Glu Phe Leu Gln	Gln Ile Asn Ser Met	Val Asp Phe
1175	1180	1185
Leu Glu Val Val His Met Ile	Tyr Gln Leu Pro Thr	Arg Val Arg
1190	1195	1200
Gly Ser Leu Arg Ala Cys Ile	Trp Ala Glu Leu Gln	Arg Arg Met
1205	1210	1215
Ala Met Pro Glu Pro Glu Trp	Thr Thr Val Gly Pro	Glu Leu Asn
1220	1225	1230
Gly Leu Asp Ser Lys Leu Leu	Leu Asp Leu Pro Ile	Gln Leu Met
1235	1240	1245
Asp Arg Leu Ser Asn Glu Ser	Ile Met Leu Val Val	Glu Leu Val
1250	1255	1260
Gln Arg Ala Pro Glu Gln Leu	Leu Ala Leu Thr Pro	Leu His Gln
1265	1270	1275
Ala Ala Leu Ala Glu Arg Ala	Leu Gln Asn Leu Ala	Pro Lys Glu
1280	1285	1290
Thr Pro Val Ser Gly Glu Val	Leu Glu Thr Leu Gly	Pro Leu Val
1295	1300	1305
Gly Phe Leu Gly Thr Glu Ser	Thr Arg Gln Ile Pro	Leu Gln Ile
1310	1315	1320
Leu Leu Ser His Leu Ser Gln	Leu Gln Gly Phe Cys	Leu Gly Glu
1325	1330	1335
Thr Phe Ala Thr Glu Leu Gly	Trp Leu Leu Leu Gln	Glu Ser Val
1340	1345	1350
Leu Gly Lys Pro Glu Leu Trp	Ser Gln Asp Glu Val	Glu Gln Ala
1355	1360	1365
Gly Arg Leu Val Phe Thr Leu	Ser Thr Glu Ala Ile	Ser Leu Ile
1370	1375	1380

Pro Arg	Glu Ala Leu Gly Pro	Glu Thr Leu Glu Arg	Leu Leu Glu
1385	1390	1395	
Lys Gln	Gln Ser Trp Glu Gln	Ser Arg Val Gly Gln	Leu Cys Arg
1400	1405	1410	
Glu Pro	Gln Leu Ala Ala Lys	Lys Ala Ala Leu Val	Ala Gly Val
1415	1420	1425	
Val Arg	Pro Ala Ala Glu Asp	Leu Pro Glu Pro Val	Pro Asn Cys
1430	1435	1440	
Ala Asp	Val Arg Gly Thr Phe	Pro Ala Ala Trp Ser	Ala Thr Gln
1445	1450	1455	
Ile Ala	Glu Met Glu Leu Ser	Asp Phe Glu Asp Cys	Leu Thr Leu
1460	1465	1470	
Phe Ala	Gly Asp Pro Gly Leu	Gly Pro Glu Glu Leu	Arg Ala Ala
1475	1480	1485	
Met Gly	Lys Ala Lys Gln Leu	Trp Gly Pro Pro Arg	Gly Phe Arg
1490	1495	1500	
Pro Glu	Gln Ile Leu Gln Leu	Gly Arg Leu Leu Ile	Gly Leu Gly
1505	1510	1515	
Asp Arg	Glu Leu Gln Glu Leu	Ile Leu Val Asp Trp	Gly Val Leu
1520	1525	1530	
Ser Thr	Leu Gly Gln Ile Asp	Gly Trp Ser Thr Thr	Gln Leu Arg
1535	1540	1545	
Ile Val	Val Ser Ser Phe Leu	Arg Gln Ser Gly Arg	His Val Ser
1550	1555	1560	
His Leu	Asp Phe Val His Leu	Thr Ala Leu Gly Tyr	Thr Leu Cys
1565	1570	1575	
Gly Leu	Arg Pro Glu Glu Leu	Gln His Ile Ser Ser	Trp Glu Phe
1580	1585	1590	
Ser Gln	Ala Ala Leu Phe Leu	Gly Thr Leu His Leu	Gln Cys Ser
1595	1600	1605	
Glu Glu	Gln Leu Glu Val Leu	Ala His Leu Leu Val	Leu Pro Gly

1610 1615 1620
 Gly Phe Gly Pro Ile Ser Asn Trp Gly Pro Glu Ile Phe Thr Glu
 1625 1630 1635
 Ile Gly Thr Ile Ala Ala Gly Ile Pro Asp Leu Ala Leu Ser Ala
 1640 1645 1650
 Leu Leu Arg Gly Gln Ile Gln Gly Val Thr Pro Leu Ala Ile Ser
 1655 1660 1665

Val Ile Pro Pro Pro Lys Phe Ala Val Val Phe Ser Pro Ile Gln
 1670 1675 1680
 Leu Ser Ser Leu Thr Ser Ala Gln Ala Val Ala Val Thr Pro Glu
 1685 1690 1695
 Gln Met Ala Phe Leu Ser Pro Glu Gln Arg Arg Ala Val Ala Trp
 1700 1705 1710
 Ala Gln His Glu Gly Lys Glu Ser Pro Glu Gln Gln Gly Arg Ser
 1715 1720 1725

Thr Ala Trp Gly Leu Gln Asp Trp Ser Arg Pro Ser Trp Ser Leu
 1730 1735 1740
 Val Leu Thr Ile Ser Phe Leu Gly His Leu Leu
 1745 1750

<210> 3

<211> 5361

<212> DNA

<213> Mus musculus

<400> 3

gcccctactg ggcctcagtc tttaggatgct ggtctctccc ttctgaagtc attcgtagcc	60
actctggacc aagctctctca gcgttccttc agccagtcac ggttctctgc gttcctggcc	120
aacatttctt catccttcca gcttggggagg atgggggagg gaccggtggg agagcccca	180
cctctccagc cccctgcaact tcgaattcat gatttctctg tgacactgag aggtagccca	240
gactggggagc caatgctagg gcttctggga gatgtgctgg cactcctggg acaggaacag	300
actccccggg actttttggt gcaccaggca ggtgtactgg gtggacttgt agaggcattg	360
ttgggagcgt tagttctctgg aggccccct gccccactc gacccccatg caccctgat	420
ggcccttctg actgtgtcct ggctgctgat tggttgcctt ctctgatgtt gttattagag	480

ggtacacgct ggcaggccct ggtgcagttg cagcccagtg tggaccaac caatgccaca 540
 ggtcttgatg gtagagagcc agctcctcac tttttacagg gtctgctggg ctgtcttacc 600

 ccagcaggag agttgggctc tgaggaggct ctttgggggtg gtctgctgcg cacagtgggg 660
 gccccctct atgtgcctt ccaggagggg ctactgcgag tcactcattc tctgcaagat 720
 gaggtctttt ctattatggg acagccagag cctgatgcca gtgggcagtg ccagggaggc 780
 aaccttaac agctgctttt atggggcatg cggaacaacc tttcttggga cggccgagca 840
 ctgggttttc tatctggatc accacctcca cccctgctc tcctgcactg cctgagcaga 900
 ggtgtgcctc tgcccagggc ttccagcct gcggctcaca tcagccctcg acagcggcga 960
 gccatctctg tggaggccct ctgcgagaac cactcaggcc cagagccacc ctacagcatc 1020

 tccaacttct ccatctactt gctctgccag cacatcaagc ctgccacccc gcggccccct 1080
 cctaccaccc cagggcctcc tcctaccacc ccacagcccc ctctaccac tacacagccc 1140
 attcctgaca ctacacagcc cctcctgtc accccaaggc ctctcctac caccaccaca 1200
 cccctccta gcacagctgt catctgccag acagctgtat ggtacgcagt ctctgggca 1260
 ccaggtgccc gaggttggct ccaagcctgc catgatcagt ttctgatca atttctggat 1320
 atgatctgcg gcaacctctc atttcagcc ctgtctggcc ccagtcgtcc tttggtaaag 1380
 cagctctgtg ctggcttgtc cccaccccc actagctgtc caccaggcct gatccctgtg 1440

 cccctacccc cagaaatatt ctggggctgt ttcttgaga atgagacact gtgggctgaa 1500
 cggttgtgtg tggaggacag tctgcaggct gtcccccca ggaaccaggc ttgggttcag 1560
 catgtgtgtc ggggccccac ctgggacgcc actgatttc caccgtgccg cgttggaccc 1620
 tgtggggaac gctgccaga tgggggcagc ttctgtctca tggctgtgtc caatgacact 1680
 ctgtatgaag ctttggttcc cttctgggct tgcttagcag gccaatgcag aattagtcgt 1740
 ggaggaaatg atacttgcct tctagaagc atgctgggcc cttgttgcc ctctctgccc 1800
 cctctgggac catccccact ctgtctggct cctggctctt ttctgcttgg catgttatcc 1860

 cagttgccac gctgtcagtc ctccgtgcca gccctgccc acccacgcg cctacattac 1920
 ctctgcgcc tactgacctt ctttctgggt ccagggactg ggggtgccga gacgcagggg 1980
 atgttaggtc aagccctgct gctctctagt ctcccagaca actgttcatt ctgggatgcc 2040
 ttccgccag agggccggag aagtgtactg aggacagtcg gagagtactt gcagcgggaa 2100
 gagccaaccc caccagctt agactcctcc ctacagcctg gctctggat gagcaagatg 2160
 gagcttctgt cctgcttcag tcctgtactg tgggatctac tccagagaga gaagagcgtt 2220

tgggccctga ggaccctggt gaaggcctac ctgcgcatgc ctccagaaga ccttcagcag	2280
cttgtgcttt cagcagagat ggaggtgca cagggtcttc tgacgtcat gcttcgttcc	2340
tgggctaagc tgaaggttca accatccgag gagcaggcca tgggccgcct gacagccttg	2400
ctgtccagc ggiaccacg cctcacctcc caactcttta tcgacatgtc accgtcatc	2460
cccttcttg ctgtccctga cctcatgcgc ttcccaccgt cctttttggc caacgacagt	2520
gtcctggctg ccatcaggga tcacagctca ggaatgaagc ctgaacagaa ggaggccctg	2580
gcaaaacgac tgctggcccc tgagctgttt ggagaagtgc ctgattggcc ccaggagctg	2640
ctgtgggcag cctgcctct gcttcccat ctgcctctgg agagctttct ccagctcagc	2700
cctcaccaga tccaggccct ggagatagc tggccagtag cagatcttgg gccgggacac	2760
gcccacatg tgcttcgtag cctagtaaac cagagcatgg aggatgggga ggagcaggtg	2820
ctcaggcttg ggtccctgc ctgtttctg agtctgagg agctacagag tctggtgccc	2880
ttgagtgatc caatggggcc tgtagaacag ggtctgctgg aatgtgcggc caatgggacc	2940
ctcagcccag aaggacgggt ggcatatgaa ctctctgggag tgttgcgttc atctggagga	3000
actgtcttaa gccccgaga gctgagggtc tgggcacctc tctttccca gctgggcctc	3060
cgcttctgc aggagctctc agagaccag cttagagcca tgcttctgc cctacaggga	3120
gccagtgtca cacctgccc ggtgttctg ttgtttggaa ggctccttc taagcatgat	3180
ctgtccctgg aggaactctg ctccctgcac cctctctgc caggtctcag ccccagaca	3240
ctccaggcca tccctaagag agttctggtt ggtgcttgtt cctgcctggg ccctgaactg	3300
tcaaggcttt cagcttgcca gattgcagct ctgctgcaga cctttcgggt aaaagatggt	3360
gttaaaaata tgggtgcagc aggtgccggc tcagccgtgt gcattcctgg gcagcccacc	3420
acttgccag actgcctgct tccctgctc ccattaaagc tgctacagct ggacgtgca	3480
gctcttctgg caaacgaag actctatcgg cagctgcctt ggtctgagca acaggcacag	3540
tttctctgga agaaaatgca agtgctacc aacctgagcc tgaggaatct gcaggctctg	3600
ggcaacttgg caggaggcat gacctgcgag tttctgcagc agatcagctc aatggttgac	3660
tttcttgatg tggtaacat gctctaccag ctgcccactg gtgttcgaga gagcctgcgg	3720
gcctgtatct ggacagagct acagcggagg atgacaatgc cagagccaga gctgaccacc	3780
ctagggccag aactgagtga acttgacaca aagctactcc tggacttgcc gatccagctg	3840
atggacagat tgtccaatga ttccattatg ttggtggtgg agatggtcca aggcgtcca	3900
gagcagctgc tggcactgac cccactccac cagacagcct tggcagagcg agcacttaaa	3960

aacctggctc caaaggagac cccaatctcc aaagaagtgc tggagacact gggccccttg 4020
gttggattcc tgggaataga gagcacgcga cggatccctt tacccattct actgtctcat 4080
ctcagtcagc tgcagggtt ctgcctagga gagacatttg ccacagagct gggatggctg 4140
ctgttgacgg agcctgttct tggaaaacca gaattgtgga gccaggatga aatagagcaa 4200
gctggacgcc tagtattcac tctgtctgct gaggtatatt cctcgatccc caggaggct 4260
ttgggccag agacactgga gaggttctg ggaaagcatc aaagctggga gcagagcaga 4320
gtgggccatc tgtgtggga gtcacagctt gccacaaga aagcagctct ggtagctggg 4380

attgtgcatc cagctgctga gggctctcaa gagcctgtac caaactgtgc agacatacgg 4440
ggaaccttcc cagcggcctg gtctcgcaca caaatctcag agatggaact ctcagacttt 4500
gaagactgcc tgtcactatt tgctggagat ccaggacttg gtcctgagga actacgggca 4560
gccatgggca aggccaagca gttgtgggt cccctcgag gattccgtcc tgagcagatc 4620
ttgcagctgg gccgtctcct gataggtcta ggagaacggg aactgcagga gcttaccttg 4680
gtggactggg gtgtgctgag cagcctgggg caaatagatg gctggagtgc catgcagctc 4740
cgagccgtgg tctccagttt cctaaggcag agtggtcggc atgtgagcca cctggacttc 4800

atttatctga cagcactggg ttacacagtc tgtggattgc gaccagagga gttacagcac 4860
atcagcagtt gggagtttag ccaagcagct ctcttcctgg gtagcttgca tctccgtgc 4920
tctgaggaac agctggaagt tctggcctat ctcttctgtg tgctgggtgg ctttgccca 4980
gtcagtaact gggggcctga gatcttact gaaattggca caatagcagc tggcatcca 5040
gacctggctc tttcagcatt actgcgggga cagatccaag gcctgactcc tcttgccatt 5100
tctgtcattc ctgtcccaa gtttgagtg gtcttcaacc ccatccagtt atctagtctc 5160
accaggggtc aggccgtagc tgttactcct gaacagctgg cctatctgag tcttgagcag 5220

cggcgagcag ttgcatgggc ccaacacgaa gggaaggaga tcccagagca gctgggtcga 5280
aactcagcct ggggtctcta cgactgggtc caagcctcct gggccctggc attgcccgtc 5340
agcatttttg gccacctatt a 5361

<210> 4

<211> 1787

<212> PRT

<213> Mus musculus

<400> 4

Ala Pro Thr Gly Pro Gln Ser Leu Asp Ala Gly Leu Ser Leu Leu Lys

1 5 10 15

Ser Phe Val Ala Thr Leu Asp Gln Ala Pro Gln Arg Ser Leu Ser Gln

20				25				30							
Ser	Arg	Phe	Ser	Ala	Phe	Leu	Ala	Asn	Ile	Ser	Ser	Ser	Phe	Gln	Leu
35				40				45							
Gly	Arg	Met	Gly	Glu	Gly	Pro	Val	Gly	Glu	Pro	Pro	Pro	Leu	Gln	Pro
50				55				60							
Pro	Ala	Leu	Arg	Leu	His	Asp	Phe	Leu	Val	Thr	Leu	Arg	Gly	Ser	Pro
65				70				75				80			
Asp	Trp	Glu	Pro	Met	Leu	Gly	Leu	Leu	Gly	Asp	Val	Leu	Ala	Leu	Leu
85				90				95							
Gly	Gln	Glu	Gln	Thr	Pro	Arg	Asp	Phe	Leu	Val	His	Gln	Ala	Gly	Val
100				105				110							
Leu	Gly	Gly	Leu	Val	Glu	Ala	Leu	Leu	Gly	Ala	Leu	Val	Pro	Gly	Gly
115				120				125							
Pro	Pro	Ala	Pro	Thr	Arg	Pro	Pro	Cys	Thr	Arg	Asp	Gly	Pro	Ser	Asp
130				135				140							
Cys	Val	Leu	Ala	Ala	Asp	Trp	Leu	Pro	Ser	Leu	Met	Leu	Leu	Leu	Glu
145				150				155				160			
Gly	Thr	Arg	Trp	Gln	Ala	Leu	Val	Gln	Leu	Gln	Pro	Ser	Val	Asp	Pro
165				170				175							
Thr	Asn	Ala	Thr	Gly	Leu	Asp	Gly	Arg	Glu	Pro	Ala	Pro	His	Phe	Leu
180				185				190							
Gln	Gly	Leu	Leu	Gly	Leu	Leu	Thr	Pro	Ala	Gly	Glu	Leu	Gly	Ser	Glu
195				200				205							
Glu	Ala	Leu	Trp	Gly	Gly	Leu	Leu	Arg	Thr	Val	Gly	Ala	Pro	Leu	Tyr
210				215				220							
Ala	Ala	Phe	Gln	Glu	Gly	Leu	Leu	Arg	Val	Thr	His	Ser	Leu	Gln	Asp
225				230				235				240			
Glu	Val	Phe	Ser	Ile	Met	Gly	Gln	Pro	Glu	Pro	Asp	Ala	Ser	Gly	Gln
245				250				255							
Cys	Gln	Gly	Gly	Asn	Leu	Gln	Gln	Leu	Leu	Leu	Trp	Gly	Met	Arg	Asn
260				265				270							

Asn Leu Ser Trp Asp Ala Arg Ala Leu Gly Phe Leu Ser Gly Ser Pro
275 280 285

Pro Pro Pro Pro Ala Leu Leu His Cys Leu Ser Arg Gly Val Pro Leu
290 295 300

Pro Arg Ala Ser Gln Pro Ala Ala His Ile Ser Pro Arg Gln Arg Arg
305 310 315 320

Ala Ile Ser Val Glu Ala Leu Cys Glu Asn His Ser Gly Pro Glu Pro
325 330 335

Pro Tyr Ser Ile Ser Asn Phe Ser Ile Tyr Leu Leu Cys Gln His Ile
340 345 350

Lys Pro Ala Thr Pro Arg Pro Pro Pro Thr Thr Pro Arg Pro Pro Pro
355 360 365

Thr Thr Pro Gln Pro Pro Pro Thr Thr Thr Gln Pro Ile Pro Asp Thr
370 375 380

Thr Gln Pro Pro Pro Val Thr Pro Arg Pro Pro Pro Thr Thr Pro Gln
385 390 395 400

Pro Pro Pro Ser Thr Ala Val Ile Cys Gln Thr Ala Val Trp Tyr Ala
405 410 415

Val Ser Trp Ala Pro Gly Ala Arg Gly Trp Leu Gln Ala Cys His Asp
420 425 430

Gln Phe Pro Asp Gln Phe Leu Asp Met Ile Cys Gly Asn Leu Ser Phe
435 440 445

Ser Ala Leu Ser Gly Pro Ser Arg Pro Leu Val Lys Gln Leu Cys Ala
450 455 460

Gly Leu Leu Pro Pro Pro Thr Ser Cys Pro Pro Gly Leu Ile Pro Val
465 470 475 480

Pro Leu Thr Pro Glu Ile Phe Trp Gly Cys Phe Leu Glu Asn Glu Thr
485 490 495

Leu Trp Ala Glu Arg Leu Cys Val Glu Asp Ser Leu Gln Ala Val Pro
500 505 510

Pro Arg Asn Gln Ala Trp Val Gln His Val Cys Arg Gly Pro Thr Leu

515	520	525	
Asp Ala Thr Asp Phe Pro Pro Cys Arg Val Gly Pro Cys Gly Glu Arg			
530	535	540	
Cys Pro Asp Gly Gly Ser Phe Leu Leu Met Val Cys Ala Asn Asp Thr			
545	550	555	560
Leu Tyr Glu Ala Leu Val Pro Phe Trp Ala Trp Leu Ala Gly Gln Cys			
565	570	575	
Arg Ile Ser Arg Gly Gly Asn Asp Thr Cys Phe Leu Glu Gly Met Leu			
580	585	590	
Gly Pro Leu Leu Pro Ser Leu Pro Pro Leu Gly Pro Ser Pro Leu Cys			
595	600	605	
Leu Ala Pro Gly Pro Phe Leu Leu Gly Met Leu Ser Gln Leu Pro Arg			
610	615	620	
Cys Gln Ser Ser Val Pro Ala Leu Ala His Pro Thr Arg Leu His Tyr			
625	630	635	640
Leu Leu Arg Leu Leu Thr Phe Leu Leu Gly Pro Gly Thr Gly Gly Ala			
645	650	655	
Glu Thr Gln Gly Met Leu Gly Gln Ala Leu Leu Leu Ser Ser Leu Pro			
660	665	670	
Asp Asn Cys Ser Phe Trp Asp Ala Phe Arg Pro Glu Gly Arg Arg Ser			
675	680	685	
Val Leu Arg Thr Val Gly Glu Tyr Leu Gln Arg Glu Glu Pro Thr Pro			
690	695	700	
Pro Gly Leu Asp Ser Ser Leu Ser Leu Gly Ser Gly Met Ser Lys Met			
705	710	715	720
Glu Leu Leu Ser Cys Phe Ser Pro Val Leu Trp Asp Leu Leu Gln Arg			
725	730	735	
Glu Lys Ser Val Trp Ala Leu Arg Thr Leu Val Lys Ala Tyr Leu Arg			
740	745	750	
Met Pro Pro Glu Asp Leu Gln Gln Leu Val Leu Ser Ala Glu Met Glu			
755	760	765	

Ala Ala Gln Gly Phe Leu Thr Leu Met Leu Arg Ser Trp Ala Lys Leu
770 775 780
Lys Val Gln Pro Ser Glu Glu Gln Ala Met Gly Arg Leu Thr Ala Leu
785 790 795 800

Leu Leu Gln Arg Tyr Pro Arg Leu Thr Ser Gln Leu Phe Ile Asp Met
805 810 815
Ser Pro Leu Ile Pro Phe Leu Ala Val Pro Asp Leu Met Arg Phe Pro
820 825 830
Pro Ser Leu Leu Ala Asn Asp Ser Val Leu Ala Ala Ile Arg Asp His
835 840 845
Ser Ser Gly Met Lys Pro Glu Gln Lys Glu Ala Leu Ala Lys Arg Leu
850 855 860

Leu Ala Pro Glu Leu Phe Gly Glu Val Pro Asp Trp Pro Gln Glu Leu
865 870 875 880
Leu Trp Ala Ala Leu Pro Leu Leu Pro His Leu Pro Leu Glu Ser Phe
885 890 895
Leu Gln Leu Ser Pro His Gln Ile Gln Ala Leu Glu Asp Ser Trp Pro
900 905 910
Val Ala Asp Leu Gly Pro Gly His Ala Arg His Val Leu Arg Ser Leu
915 920 925

Val Asn Gln Ser Met Glu Asp Gly Glu Glu Gln Val Leu Arg Leu Gly
930 935 940
Ser Leu Ala Cys Phe Leu Ser Pro Glu Glu Leu Gln Ser Leu Val Pro
945 950 955 960
Leu Ser Asp Pro Met Gly Pro Val Glu Gln Gly Leu Leu Glu Cys Ala
965 970 975
Ala Asn Gly Thr Leu Ser Pro Glu Gly Arg Val Ala Tyr Glu Leu Leu
980 985 990

Gly Val Leu Arg Ser Ser Gly Gly Thr Val Leu Ser Pro Arg Glu Leu
995 1000 1005
Arg Val Trp Ala Pro Leu Phe Pro Gln Leu Gly Leu Arg Phe Leu

1010	1015	1020
Gln Glu Leu Ser Glu Thr	Gln Leu Arg Ala Met Leu	Pro Ala Leu
1025	1030	1035
Gln Gly Ala Ser Val Thr	Pro Ala Gln Ala Val Leu	Leu Phe Gly
1040	1045	1050
Arg Leu Leu Pro Lys His Asp	Leu Ser Leu Glu Glu	Leu Cys Ser
1055	1060	1065
Leu His Pro Leu Leu Pro Gly	Leu Ser Pro Gln Thr	Leu Gln Ala
1070	1075	1080
Ile Pro Lys Arg Val Leu Val	Gly Ala Cys Ser Cys	Leu Gly Pro
1085	1090	1095
Glu Leu Ser Arg Leu Ser Ala	Cys Gln Ile Ala Ala	Leu Leu Gln
1100	1105	1110
Thr Phe Arg Val Lys Asp Gly	Val Lys Asn Met Gly	Ala Ala Gly
1115	1120	1125
Ala Gly Ser Ala Val Cys Ile	Pro Gly Gln Pro Thr	Thr Trp Pro
1130	1135	1140
Asp Cys Leu Leu Pro Leu Leu	Pro Leu Lys Leu Leu	Gln Leu Asp
1145	1150	1155
Ala Ala Ala Leu Leu Ala Asn	Arg Arg Leu Tyr Arg	Gln Leu Pro
1160	1165	1170
Trp Ser Glu Gln Gln Ala Gln	Phe Leu Trp Lys Lys	Met Gln Val
1175	1180	1185
Pro Thr Asn Leu Ser Leu Arg	Asn Leu Gln Ala Leu	Gly Asn Leu
1190	1195	1200
Ala Gly Gly Met Thr Cys Glu	Phe Leu Gln Gln Ile	Ser Ser Met
1205	1210	1215
Val Asp Phe Leu Asp Val Val	His Met Leu Tyr Gln	Leu Pro Thr
1220	1225	1230
Gly Val Arg Glu Ser Leu Arg	Ala Cys Ile Trp Thr	Glu Leu Gln
1235	1240	1245

Arg Arg Met Thr Met Pro Glu	Pro Glu Leu Thr Thr	Leu Gly Pro
1250	1255	1260
Glu Leu Ser Glu Leu Asp Thr	Lys Leu Leu Leu Asp	Leu Pro Ile
1265	1270	1275
Gln Leu Met Asp Arg Leu Ser	Asn Asp Ser Ile Met	Leu Val Val
1280	1285	1290
Glu Met Val Gln Gly Ala Pro	Glu Gln Leu Leu Ala	Leu Thr Pro
1295	1300	1305
Leu His Gln Thr Ala Leu Ala	Glu Arg Ala Leu Lys	Asn Leu Ala
1310	1315	1320
Pro Lys Glu Thr Pro Ile Ser	Lys Glu Val Leu Glu	Thr Leu Gly
1325	1330	1335
Pro Leu Val Gly Phe Leu Gly	Ile Glu Ser Thr Arg	Arg Ile Pro
1340	1345	1350
Leu Pro Ile Leu Leu Ser His	Leu Ser Gln Leu Gln	Gly Phe Cys
1355	1360	1365
Leu Gly Glu Thr Phe Ala Thr	Glu Leu Gly Trp Leu	Leu Leu Gln
1370	1375	1380
Glu Pro Val Leu Gly Lys Pro	Glu Leu Trp Ser Gln	Asp Glu Ile
1385	1390	1395
Glu Gln Ala Gly Arg Leu Val	Phe Thr Leu Ser Ala	Glu Ala Ile
1400	1405	1410
Ser Ser Ile Pro Arg Glu Ala	Leu Gly Pro Glu Thr	Leu Glu Arg
1415	1420	1425
Leu Leu Gly Lys His Gln Ser	Trp Glu Gln Ser Arg	Val Gly His
1430	1435	1440
Leu Cys Gly Glu Ser Gln Leu	Ala His Lys Lys Ala	Ala Leu Val
1445	1450	1455
Ala Gly Ile Val His Pro Ala	Ala Glu Gly Leu Gln	Glu Pro Val
1460	1465	1470

Pro Asn Cys Ala Asp Ile Arg Gly Thr Phe Pro Ala Ala Trp Ser

1475	1480	1485
Ala Thr Gln Ile Ser Glu Met	Glu Leu Ser Asp Phe	Glu Asp Cys
1490	1495	1500
Leu Ser Leu Phe Ala Gly Asp	Pro Gly Leu Gly Pro	Glu Glu Leu
1505	1510	1515
Arg Ala Ala Met Gly Lys Ala	Lys Gln Leu Trp Gly	Pro Pro Arg
1520	1525	1530
Gly Phe Arg Pro Glu Gln Ile	Leu Gln Leu Gly Arg	Leu Leu Ile
1535	1540	1545
Gly Leu Gly Glu Arg Glu Leu	Gln Glu Leu Thr Leu	Val Asp Trp
1550	1555	1560
Gly Val Leu Ser Ser Leu Gly	Gln Ile Asp Gly Trp	Ser Ser Met
1565	1570	1575
Gln Leu Arg Ala Val Val Ser	Ser Phe Leu Arg Gln	Ser Gly Arg
1580	1585	1590
His Val Ser His Leu Asp Phe	Ile Tyr Leu Thr Ala	Leu Gly Tyr
1595	1600	1605
Thr Val Cys Gly Leu Arg Pro	Glu Glu Leu Gln His	Ile Ser Ser
1610	1615	1620
Trp Glu Phe Ser Gln Ala Ala	Leu Phe Leu Gly Ser	Leu His Leu
1625	1630	1635
Pro Cys Ser Glu Glu Gln Leu	Glu Val Leu Ala Tyr	Leu Leu Val
1640	1645	1650
Leu Pro Gly Gly Phe Gly Pro	Val Ser Asn Trp Gly	Pro Glu Ile
1655	1660	1665
Phe Thr Glu Ile Gly Thr Ile	Ala Ala Gly Ile Pro	Asp Leu Ala
1670	1675	1680
Leu Ser Ala Leu Leu Arg Gly	Gln Ile Gln Gly Leu	Thr Pro Leu
1685	1690	1695
Ala Ile Ser Val Ile Pro Ala	Pro Lys Phe Ala Val	Val Phe Asn
1700	1705	1710

Pro Ile Gln Leu Ser Ser Leu Thr Arg Gly Gln Ala Val Ala Val
 1715 1720 1725
 Thr Pro Glu Gln Leu Ala Tyr Leu Ser Pro Glu Gln Arg Arg Ala
 1730 1735 1740
 Val Ala Trp Ala Gln His Glu Gly Lys Glu Ile Pro Glu Gln Leu
 1745 1750 1755
 Gly Arg Asn Ser Ala Trp Gly Leu Tyr Asp Trp Phe Gln Ala Ser
 1760 1765 1770

Trp Ala Leu Ala Leu Pro Val Ser Ile Phe Gly His Leu Leu
 1775 1780 1785

<210> 5

<211> 2436

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 5

atggctctca gcctctggcc cctgctgctg ctgctgctgc tgcctgctgt gctgtccttt	60
gcagtgactc tggcccctac tgggcctcat tccttgacc ctggtctctc cttcctgaag	120
tcattgctct ccactctgga ccaggctccc cagggtctccc tgagccgctc acggttcttt	180
acattcctgg ccaacatttc ttcttccttt gacacctggga gaatggggga aggaccagta	240
ggagagcccc cacctctcca gccgcctgct ctgcggctcc atgattttct agtgacactg	300
agaggttagcc ccgactggga gccaatgcta gggctgctag gggatatgct ggcactgctg	360
ggacaggagc agactcccc agatttctct gtgcaccagg caggggtgct ggggtggactt	420
gtggaggtgc tgcctgggagc cttagtctct gggggccccc ctaccccaac tcggccccc	480
tgcacccgtg atgggccgtc tgactgtgtc ctggctgctg actggttgcc ttctctgctg	540
ctgttgtag agggcacacg ctggcaagct ctggtgcagg tgcagcccag tgtggacccc	600
accaatgcca caggcctcga tgggaggag gcagctctc actttttgca gggctctgtt	660
ggtttgctta cccaacagg ggagctaggc tccaaggagg ctctttgggg cggctctgcta	720
cgcacagtgg gggccccct ctatctgcc ttccaggagg ggctgctccg tgtcactcac	780
tccttcgagg atgaggtctt ctccattttg gggcagccag agcctgatac caatgggcag	840
tgccaggag gtaaccttca acagctgtc ttatggggcg tccggcaca cctttcctgg	900

gatgtccagg cgctgggctt tctgtctgga tcaccacccc caccctctgc cctccttcac 960
 tgcctgagca cgggcgtgcc tctgcccaga gcttctcagc cgtcagccca catcagccca 1020
 cgccaacggc gagccatcac tgtggaggcc ctctgtgaga accacttagg cccagcacca 1080

ccctacagca tticcaactt ctccatccac ttgctctgcc agcacaccaa gcctgccact 1140
 ccacagcccc atcccagcac cactgccatc tgccagacag ctgtgtggta tgcagtgtcc 1200
 tgggcaccag gtgcccgaagg ctggtacag gcctgccacg accagtttcc tgatgagttt 1260
 ttggatgcga tctgcagtaa cctctccttt tcagccctgt ctggctccaa ccgccgcctg 1320
 gtgaagcggc tctgtgctgg cctgtccca cccctacca gctgccctga aggctgccc 1380
 cctgttcccc tcacccaga catcttttgg ggctgttct tggagaatga gactctgtgg 1440
 gctgagcgac tgtgtgggga ggcaagtcta caggctgtgc ccccagcaa ccaggcttgg 1500

gtccagcatg tgtgccaggg cccacccca gatgtcactg cctccccacc atgccacatt 1560
 ggacctgtg gggaacgctg cccgatggg ggagcttcc tggatggt ctgtgccaat 1620
 gacacatgt atgaggtcct ggtgcccttc tggccttggc tagcaggcca atgcaggata 1680
 agtcgtgggg gcaatgacac ttgcttcta gaagggtgc tgggccccct tctgccctct 1740
 ctgccaccac tgggaccatc cccactctgt ctgacctg gcccttctc cttggcatg 1800
 ctatcccagt tgccacgtg tcagtctct gtcccagctc ttgtcaccc cacacgcta 1860
 cactatctc tcgcctgt gacctctc ttgggtccag gggtgggg cgctgagggc 1920

caggggatgc tgggtcggc cctactgtc tccagtctc cagacaactg ctcttcttg 1980
 gatgccttc gccagagg cggcgagct gtgctacga cgattgggga atacctggaa 2040
 caagatgagg agcagccaac cccatcaggc ttgaacca ctgtcaacc cagctctggt 2100
 ataagcaaga tggagctgt ggcctgctg tcatacga cagagatact gacagtagaa 2160
 tatggccttc tgccaatcg gaagattgt gagaaacga tagaatgcac agtttactct 2220
 gtcgataaca atggtaacat ttatactcag ccagttgcc agtggcacga ccggggagag 2280
 caggaagtat tcgaatactg tctggaggat ggaagtctca ttagggccac taaggaccac 2340

aaatttatga cagtcgatgg ccagatgctg cctatagacg aaatcttga gcgagagttg 2400
 gacctcatgc gattgacaa ccttctaataa taatag 2436

<210> 6

<211> 810

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 6

Met Ala Leu Ser Leu Trp Pro Leu Leu Leu Leu Leu Leu Leu Leu

1 5 10 15

Leu Leu Ser Phe Ala Val Thr Leu Ala Pro Thr Gly Pro His Ser Leu

20 25 30

Asp Pro Gly Leu Ser Phe Leu Lys Ser Leu Leu Ser Thr Leu Asp Gln

35 40 45

Ala Pro Gln Gly Ser Leu Ser Arg Ser Arg Phe Phe Thr Phe Leu Ala

50 55 60

Asn Ile Ser Ser Ser Phe Glu Pro Gly Arg Met Gly Glu Gly Pro Val

65 70 75 80

Gly Glu Pro Pro Pro Leu Gln Pro Pro Ala Leu Arg Leu His Asp Phe

85 90 95

Leu Val Thr Leu Arg Gly Ser Pro Asp Trp Glu Pro Met Leu Gly Leu

100 105 110

Leu Gly Asp Met Leu Ala Leu Leu Gly Gln Glu Gln Thr Pro Arg Asp

115 120 125

Phe Leu Val His Gln Ala Gly Val Leu Gly Gly Leu Val Glu Val Leu

130 135 140

Leu Gly Ala Leu Val Pro Gly Gly Pro Pro Thr Pro Thr Arg Pro Pro

145 150 155 160

Cys Thr Arg Asp Gly Pro Ser Asp Cys Val Leu Ala Ala Asp Trp Leu

165 170 175

Pro Ser Leu Leu Leu Leu Leu Glu Gly Thr Arg Trp Gln Ala Leu Val

180 185 190

Gln Val Gln Pro Ser Val Asp Pro Thr Asn Ala Thr Gly Leu Asp Gly

195 200 205

Arg Glu Ala Ala Pro His Phe Leu Gln Gly Leu Leu Gly Leu Leu Thr

210 215 220

Pro Thr Gly Glu Leu Gly Ser Lys Glu Ala Leu Trp Gly Gly Leu Leu

225 230 235 240
 Arg Thr Val Gly Ala Pro Leu Tyr Ala Ala Phe Gln Glu Gly Leu Leu
 245 250 255
 Arg Val Thr His Ser Leu Gln Asp Glu Val Phe Ser Ile Leu Gly Gln
 260 265 270
 Pro Glu Pro Asp Thr Asn Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln
 275 280 285

 Leu Leu Leu Trp Gly Val Arg His Asn Leu Ser Trp Asp Val Gln Ala
 290 295 300
 Leu Gly Phe Leu Ser Gly Ser Pro Pro Pro Pro Pro Ala Leu Leu His
 305 310 315 320
 Cys Leu Ser Thr Gly Val Pro Leu Pro Arg Ala Ser Gln Pro Ser Ala
 325 330 335
 His Ile Ser Pro Arg Gln Arg Arg Ala Ile Thr Val Glu Ala Leu Cys
 340 345 350

 Glu Asn His Leu Gly Pro Ala Pro Pro Tyr Ser Ile Ser Asn Phe Ser
 355 360 365
 Ile His Leu Leu Cys Gln His Thr Lys Pro Ala Thr Pro Gln Pro His
 370 375 380
 Pro Ser Thr Thr Ala Ile Cys Gln Thr Ala Val Trp Tyr Ala Val Ser
 385 390 395 400
 Trp Ala Pro Gly Ala Gln Gly Trp Leu Gln Ala Cys His Asp Gln Phe
 405 410 415

 Pro Asp Glu Phe Leu Asp Ala Ile Cys Ser Asn Leu Ser Phe Ser Ala
 420 425 430
 Leu Ser Gly Ser Asn Arg Arg Leu Val Lys Arg Leu Cys Ala Gly Leu
 435 440 445
 Leu Pro Pro Pro Thr Ser Cys Pro Glu Gly Leu Pro Pro Val Pro Leu
 450 455 460
 Thr Pro Asp Ile Phe Trp Gly Cys Phe Leu Glu Asn Glu Thr Leu Trp
 465 470 475 480

Ala Glu Arg Leu Cys Gly Glu Ala Ser Leu Gln Ala Val Pro Pro Ser
485 490 495

Asn Gln Ala Trp Val Gln His Val Cys Gln Gly Pro Thr Pro Asp Val
500 505 510

Thr Ala Ser Pro Pro Cys His Ile Gly Pro Cys Gly Glu Arg Cys Pro
515 520 525

Asp Gly Gly Ser Phe Leu Val Met Val Cys Ala Asn Asp Thr Met Tyr
530 535 540

Glu Val Leu Val Pro Phe Trp Pro Trp Leu Ala Gly Gln Cys Arg Ile
545 550 555 560

Ser Arg Gly Gly Asn Asp Thr Cys Phe Leu Glu Gly Leu Leu Gly Pro
565 570 575

Leu Leu Pro Ser Leu Pro Pro Leu Gly Pro Ser Pro Leu Cys Leu Thr
580 585 590

Pro Gly Pro Phe Leu Leu Gly Met Leu Ser Gln Leu Pro Arg Cys Gln
595 600 605

Ser Ser Val Pro Ala Leu Ala His Pro Thr Arg Leu His Tyr Leu Leu
610 615 620

Arg Leu Leu Thr Phe Leu Leu Gly Pro Gly Ala Gly Gly Ala Glu Ala
625 630 635 640

Gln Gly Met Leu Gly Arg Ala Leu Leu Leu Ser Ser Leu Pro Asp Asn
645 650 655

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

Arg Thr Ile Gly Glu Tyr Leu Glu Gln Asp Glu Glu Gln Pro Thr Pro
675 680 685

Ser Gly Phe Glu Pro Thr Val Asn Pro Ser Ser Gly Ile Ser Lys Met
690 695 700

Glu Leu Leu Ala Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu
705 710 715 720

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

725

730

735

Thr Val Tyr Ser Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val

740

745

750

Ala Gln Trp His Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu

755

760

765

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

770

775

780

Val Asp Gly Gln Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu

785

790

795

800

Asp Leu Met Arg Val Asp Asn Leu Pro Asn

805

810

<210> 7

<211> 2550

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 7

atggctctga gcctccagcc ccagctgctc cttctcctgt cgctcctgcc gcaggaagtg 60

acttcagccc ctactgggcc tcagctcttg gatgctggtc tctcccttct gaagtcattc 120

gtagccactc tggaccaagc tcctcagcgt tcctcagcc agtcacggtt cttgcgttc 180

ctggccaaca tttcttcac cttccagctt gggaggatgg gggagggacc ggtgggagag 240

ccccacctc tccagccccc tgcacttga cttcatgatt tcctcgtgac actgagaggt 300

agcccagact gggagccaat gctagggctt ctgggagatg tgctggcact cctgggacag 360

gaacagactc cccgggactt tttggtgcac caggcagggtg tactgggtgg acttgtagag 420

gcattgttgg gagcgtagt tcctggaggc cccctgccc cactcgacc ccatgcacc 480

cgtgatggcc cttctgactg tgtctggct gctgatttgt tgccttctct gatgttgta 540

ttagagggtg cagctggca ggccctggtg cagttgcagc ccagtgtgga cccaaccaat 600

gccacaggtc ttgatggtag agagccagct cctcactttt tacagggtct gctgggcttg 660

cttaccag caggagagtt gggctctgag gaggtctttt ggggtgtgtc gctgcgcaca 720

gtgggggccc ccctctatgc tgccttcag gaggggctac tgcgagtcac tcattctctg 780

caagatgagg tcttttctat tatgggacag ccagagcctg atgccagtgg gcagtgccag 840
ggaggcaacc ttcaacagct gcttttatgg ggcatgcgga acaacctttc ttgggacgcc 900
cgagcactgg gttttctatc tggatcacca cctccacccc ctgctctcct gcactgcctg 960
agcagagggtg tgcctctgcc cagggccttc cagcctgcgg ctacatcag ccctcgacag 1020
cggcgagcca tctctgtgga ggccctctgc gagaaccact caggcccaga gccaccctac 1080

agcatctcca acttctccat ctacttctc tgccagcaca tcaagcctgc caccgccggg 1140
ccccctcta ccacccacg gcctctcct accacccac agccccctcc taccactaca 1200
cagccattc ctgacactac acagccccct cctgtcacc caaggcctcc tctaccacc 1260
ccacaacccc ctctagcac agctgtcatc tgccagacag ctgtatggtc cgcagtctcg 1320
tgggcaccag gtgcccagg ttggctcaa gcctgccatg atcagtttcc tgatcaattt 1380
ctggatatga tctgcggcaa cctctcattt tcagccctgt ctggccccag tctcctttg 1440
gtaaagcagc tctgtgctgg ctgtctcca cccccacta gctgtccacc aggcctgac 1500

cctgtgcccc tcacccaga aatatcttg ggctgtttcc tggagaatga gacactgtgg 1560
gctgaacggt tgtgtgtgga ggacagtctg caggctgtgc ccccgaggaa ccagcttgg 1620
gttcagcatg tgtgtcgggg cccaccttg gacgccactg attttccacc gtccgcggtt 1680
ggacctgtg gggaaacgtg cccagatggg ggcagcttcc tgtcatggt ctgtgccaat 1740
gacactctgt atgaagcctt gggtcccttc tgggcttggc tagcaggcca atgcagaatt 1800
agtctggag gaaatgatac ttgctttcta gaaggcatgc tgggccccct gttgccctct 1860
ctgccccctc tgggaccatc cccactctgt ctggctcctg gtccttttct gcttggcatg 1920

ttatccagct tgccacgtg tcagtctcc gtgccagccc tcgcccacc caccgccta 1980
cattacctc tgcgcctact gaccttctt ctgggtccag ggactggggg tgccgagacg 2040
caggggatgt taggtcaagc cctgtctctc tctagtctcc cagacaactg ttcattctgg 2100
gatgccttcc gccagaggg ccggagaagt gtactgagga cagtgcgaga gtacttgcag 2160
cgggaagagc caacccacc aggtcttagac tctcctca gcctcggtc tggatgagc 2220
aagatggagc ttctgtcctg cctgtcatc gaaaccgaga tactgacagt agaatatggc 2280
cttctgcaa tcgggaagat tgtggagaaa cggatagaat gcacagtta ctctgtcat 2340

aacaatggt acaattatc tcagccagt gcccagtggc acgaccgggg agagcaggaa 2400
gtattcgaat acigtctgga ggatggaagt ctcataggg ccactaagga ccacaaattt 2460
atgacagtcg atggccagat gctgcctata gacgaaatct ttgagcgaga gttggacctc 2520
atgcgagttg acaaccttcc taattaatag 2550

<210> 8

<211> 848

<212> PRT

<213> Artificial Sequence

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

<400> 8

Met Ala Leu Ser Leu Gln Pro Gln Leu Leu Leu Leu Ser Leu Leu

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

65 70 75 80
Pro Pro Pro Leu Gln Pro Pro Ala Leu Arg Leu His Asp Phe Leu Val
 85 90 95
Thr Leu Arg Gly Ser Pro Asp Trp Glu Pro Met Leu Gly Leu Leu Gly
 100 105 110
Asp Val Leu Ala Leu Leu Gly Gln Glu Gln Thr Pro Arg Asp Phe Leu
 115 120 125
Val His Gln Ala Gly Val Leu Gly Gly Leu Val Glu Ala Leu Leu Gly

130 135 140
Ala Leu Val Pro Gly Gly Pro Pro Ala Pro Thr Arg Pro Pro Cys Thr
145 150 155 160
Arg Asp Gly Pro Ser Asp Cys Val Leu Ala Ala Asp Trp Leu Pro Ser
 165 170 175
Leu Met Leu Leu Leu Glu Gly Thr Arg Trp Gln Ala Leu Val Gln Leu
 180 185 190
Gln Pro Ser Val Asp Pro Thr Asn Ala Thr Gly Leu Asp Gly Arg Glu

195 200 205
 Pro Ala Pro His Phe Leu Gln Gly Leu Leu Gly Leu Leu Thr Pro Ala
 210 215 220
 Gly Glu Leu Gly Ser Glu Glu Ala Leu Trp Gly Gly Leu Leu Arg Thr
 225 230 235 240
 Val Gly Ala Pro Leu Tyr Ala Ala Phe Gln Glu Gly Leu Leu Arg Val
 245 250 255
 Thr His Ser Leu Gln Asp Glu Val Phe Ser Ile Met Gly Gln Pro Glu

 260 265 270
 Pro Asp Ala Ser Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln Leu Leu
 275 280 285
 Leu Trp Gly Met Arg Asn Asn Leu Ser Trp Asp Ala Arg Ala Leu Gly
 290 295 300
 Phe Leu Ser Gly Ser Pro Pro Pro Pro Pro Ala Leu Leu His Cys Leu
 305 310 315 320
 Ser Arg Gly Val Pro Leu Pro Arg Ala Ser Gln Pro Ala Ala His Ile

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

 385 390 395 400
 Gln Pro Ile Pro Asp Thr Thr Gln Pro Pro Pro Val Thr Pro Arg Pro
 405 410 415
 Pro Pro Thr Thr Pro Gln Pro Pro Pro Ser Thr Ala Val Ile Cys Gln
 420 425 430
 Thr Ala Val Trp Tyr Ala Val Ser Trp Ala Pro Gly Ala Arg Gly Trp
 435 440 445

Leu Gln Ala Cys His Asp Gln Phe Pro Asp Gln Phe Leu Asp Met Ile

450 455 460

Cys Gly Asn Leu Ser Phe Ser Ala Leu Ser Gly Pro Ser Arg Pro Leu

465 470 475 480

Val Lys Gln Leu Cys Ala Gly Leu Leu Pro Pro Pro Thr Ser Cys Pro

485 490 495

Pro Gly Leu Ile Pro Val Pro Leu Thr Pro Glu Ile Phe Trp Gly Cys

500 505 510

Phe Leu Glu Asn Glu Thr Leu Trp Ala Glu Arg Leu Cys Val Glu Asp

515 520 525

Ser Leu Gln Ala Val Pro Pro Arg Asn Gln Ala Trp Val Gln His Val

530 535 540

Cys Arg Gly Pro Thr Leu Asp Ala Thr Asp Phe Pro Pro Cys Arg Val

545 550 555 560

Gly Pro Cys Gly Glu Arg Cys Pro Asp Gly Gly Ser Phe Leu Leu Met

565 570 575

Val Cys Ala Asn Asp Thr Leu Tyr Glu Ala Leu Val Pro Phe Trp Ala

580 585 590

Trp Leu Ala Gly Gln Cys Arg Ile Ser Arg Gly Gly Asn Asp Thr Cys

595 600 605

Phe Leu Glu Gly Met Leu Gly Pro Leu Leu Pro Ser Leu Pro Pro Leu

610 615 620

Gly Pro Ser Pro Leu Cys Leu Ala Pro Gly Pro Phe Leu Leu Gly Met

625 630 635 640

Leu Ser Gln Leu Pro Arg Cys Gln Ser Ser Val Pro Ala Leu Ala His

645 650 655

Pro Thr Arg Leu His Tyr Leu Leu Arg Leu Leu Thr Phe Leu Leu Gly

660 665 670

Pro Gly Thr Gly Gly Ala Glu Thr Gln Gly Met Leu Gly Gln Ala Leu

675 680 685

Leu Leu Ser Ser Leu Pro Asp Asn Cys Ser Phe Trp Asp Ala Phe Arg

690 695 700
Pro Glu Gly Arg Arg Ser Val Leu Arg Thr Val Gly Glu Tyr Leu Gln

705 710 715 720
Arg Glu Glu Pro Thr Pro Pro Gly Leu Asp Ser Ser Leu Ser Leu Gly

725 730 735
Ser Gly Met Ser Lys Met Glu Leu Leu Ser Cys Leu Ser Tyr Glu Thr

740 745 750
Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu Pro Ile Gly Lys Ile Val

755 760 765
Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser Val Asp Asn Asn Gly Asn

770 775 780
Ile Tyr Thr Gln Pro Val Ala Gln Trp His Asp Arg Gly Glu Gln Glu

785 790 795 800
Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser Leu Ile Arg Ala Thr Lys

805 810 815
Asp His Lys Phe Met Thr Val Asp Gly Gln Met Leu Pro Ile Asp Glu

820 825 830
Ile Phe Glu Arg Glu Leu Asp Leu Met Arg Val Asp Asn Leu Pro Asn

835 840 845

<210> 9

<211> 60

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 9

gctctcagcc tctggccctt gctgtgtgtg ctgtgtgtgc tgctgtgtgt gtcctttgca 60

<210> 10

<211> 20

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 10

Ala Leu Ser Leu Trp Pro Leu Leu Leu Leu Leu Leu Leu Leu Leu

1 5 10 15

Leu Ser Phe Ala

20

<210> 11

<211> 63

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 11

gctctgagcc tccagcccca gctgctcctt ctctgtgcgc tctgcccga ggaagtgact 60

tca 63

<210> 12

<211> 21

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 12

Ala Leu Ser Leu Gln Pro Gln Leu Leu Leu Leu Ser Leu Leu Pro

1 5 10 15

Gln Glu Val Thr Ser

20

<210> 13

<211> 306

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 13

tgcctgtcat acgaaaccga gatactgaca gtagaatatg gccttctgcc aatcggaag 60
 attgtggaga aacggataga atgcacagtt tactctgtcg ataacaatgg taacatttat 120
 actcagccag ttgccagtgc gcacgaccgg ggagagcagg aagtattcga atactgtctg 180
 gaggatggaa gtctcattag ggccactaag gaccacaaat ttatgacagt cgatggccag 240

atgtgccta tagacgaaat ctttgagcga gagttggacc tcatgcgagt tgacaacctt 300
 cctaatt 306

<210> 14

<211> 102

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 14

Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu

1 5 10 15

Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser

20 25 30

Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His

35 40 45

Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser

50 55 60

Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln

65 70 75 80

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

85 90 95

Val Asp Asn Leu Pro Asn

100

<210> 15

<211> 687

<212> PRT

<213> Homo sapiens

<400> 15

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

245 250 255
 Asn Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln Leu Leu Leu Trp Gly
 260 265 270
 Val Arg His Asn Leu Ser Trp Asp Val Gln Ala Leu Gly Phe Leu Ser
 275 280 285
 Gly Ser Pro Pro Pro Pro Pro Ala Leu Leu His Cys Leu Ser Thr Gly
 290 295 300

 Val Pro Leu Pro Arg Ala Ser Gln Pro Ser Ala His Ile Ser Pro Arg
 305 310 315 320
 Gln Arg Arg Ala Ile Thr Val Glu Ala Leu Cys Glu Asn His Leu Gly
 325 330 335
 Pro Ala Pro Pro Tyr Ser Ile Ser Asn Phe Ser Ile His Leu Leu Cys
 340 345 350
 Gln His Thr Lys Pro Ala Thr Pro Gln Pro His Pro Ser Thr Thr Ala
 355 360 365

 Ile Cys Gln Thr Ala Val Trp Tyr Ala Val Ser Trp Ala Pro Gly Ala
 370 375 380
 Gln Gly Trp Leu Gln Ala Cys His Asp Gln Phe Pro Asp Glu Phe Leu
 385 390 395 400
 Asp Ala Ile Cys Ser Asn Leu Ser Phe Ser Ala Leu Ser Gly Ser Asn
 405 410 415
 Arg Arg Leu Val Lys Arg Leu Cys Ala Gly Leu Leu Pro Pro Pro Thr
 420 425 430

 Ser Cys Pro Glu Gly Leu Pro Pro Val Pro Leu Thr Pro Asp Ile Phe
 435 440 445
 Trp Gly Cys Phe Leu Glu Asn Glu Thr Leu Trp Ala Glu Arg Leu Cys
 450 455 460
 Gly Glu Ala Ser Leu Gln Ala Val Pro Pro Ser Asn Gln Ala Trp Val
 465 470 475 480
 Gln His Val Cys Gln Gly Pro Thr Pro Asp Val Thr Ala Ser Pro Pro
 485 490 495

Cys His Ile Gly Pro Cys Gly Glu Arg Cys Pro Asp Gly Gly Ser Phe
500 505 510
Leu Val Met Val Cys Ala Asn Asp Thr Met Tyr Glu Val Leu Val Pro
515 520 525
Phe Trp Pro Trp Leu Ala Gly Gln Cys Arg Ile Ser Arg Gly Gly Asn
530 535 540
Asp Thr Cys Phe Leu Glu Gly Leu Leu Gly Pro Leu Leu Pro Ser Leu
545 550 555 560

Pro Pro Leu Gly Pro Ser Pro Leu Cys Leu Thr Pro Gly Pro Phe Leu
565 570 575
Leu Gly Met Leu Ser Gln Leu Pro Arg Cys Gln Ser Ser Val Pro Ala
580 585 590
Leu Ala His Pro Thr Arg Leu His Tyr Leu Leu Arg Leu Leu Thr Phe
595 600 605
Leu Leu Gly Pro Gly Ala Gly Gly Ala Glu Ala Gln Gly Met Leu Gly
610 615 620

Arg Ala Leu Leu Leu Ser Ser Leu Pro Asp Asn Cys Ser Phe Trp Asp
625 630 635 640
Ala Phe Arg Pro Glu Gly Arg Arg Ser Val Leu Arg Thr Ile Gly Glu
645 650 655
Tyr Leu Glu Gln Asp Glu Glu Gln Pro Thr Pro Ser Gly Phe Glu Pro
660 665 670
Thr Val Asn Pro Ser Ser Gly Ile Ser Lys Met Glu Leu Leu Ala
675 680 685

<210> 16

<211> 724

<212> PRT

<213> Mus musculus

<400> 16

Ala Pro Thr Gly Pro Gln Ser Leu Asp Ala Gly Leu Ser Leu Leu Lys
1 5 10 15
Ser Phe Val Ala Thr Leu Asp Gln Ala Pro Gln Arg Ser Leu Ser Gln

20 25 30
 Ser Arg Phe Ser Ala Phe Leu Ala Asn Ile Ser Ser Ser Phe Gln Leu
 35 40 45
 Gly Arg Met Gly Glu Gly Pro Val Gly Glu Pro Pro Pro Leu Gln Pro
 50 55 60
 Pro Ala Leu Arg Leu His Asp Phe Leu Val Thr Leu Arg Gly Ser Pro
 65 70 75 80
 Asp Trp Glu Pro Met Leu Gly Leu Leu Gly Asp Val Leu Ala Leu Leu
 85 90 95
 Gly Gln Glu Gln Thr Pro Arg Asp Phe Leu Val His Gln Ala Gly Val
 100 105 110
 Leu Gly Gly Leu Val Glu Ala Leu Leu Gly Ala Leu Val Pro Gly Gly
 115 120 125
 Pro Pro Ala Pro Thr Arg Pro Pro Cys Thr Arg Asp Gly Pro Ser Asp
 130 135 140
 Cys Val Leu Ala Ala Asp Trp Leu Pro Ser Leu Met Leu Leu Leu Glu
 145 150 155 160
 Gly Thr Arg Trp Gln Ala Leu Val Gln Leu Gln Pro Ser Val Asp Pro
 165 170 175
 Thr Asn Ala Thr Gly Leu Asp Gly Arg Glu Pro Ala Pro His Phe Leu
 180 185 190
 Gln Gly Leu Leu Gly Leu Leu Thr Pro Ala Gly Glu Leu Gly Ser Glu
 195 200 205
 Glu Ala Leu Trp Gly Gly Leu Leu Arg Thr Val Gly Ala Pro Leu Tyr
 210 215 220
 Ala Ala Phe Gln Glu Gly Leu Leu Arg Val Thr His Ser Leu Gln Asp
 225 230 235 240
 Glu Val Phe Ser Ile Met Gly Gln Pro Glu Pro Asp Ala Ser Gly Gln
 245 250 255
 Cys Gln Gly Gly Asn Leu Gln Gln Leu Leu Leu Trp Gly Met Arg Asn
 260 265 270

Asn Leu Ser Trp Asp Ala Arg Ala Leu Gly Phe Leu Ser Gly Ser Pro
 275 280 285
 Pro Pro Pro Pro Ala Leu Leu His Cys Leu Ser Arg Gly Val Pro Leu
 290 295 300
 Pro Arg Ala Ser Gln Pro Ala Ala His Ile Ser Pro Arg Gln Arg Arg

 305 310 315 320
 Ala Ile Ser Val Glu Ala Leu Cys Glu Asn His Ser Gly Pro Glu Pro
 325 330 335
 Pro Tyr Ser Ile Ser Asn Phe Ser Ile Tyr Leu Leu Cys Gln His Ile
 340 345 350
 Lys Pro Ala Thr Pro Arg Pro Pro Pro Thr Thr Pro Arg Pro Pro Pro
 355 360 365
 Thr Thr Pro Gln Pro Pro Pro Thr Thr Thr Gln Pro Ile Pro Asp Thr

 370 375 380
 Thr Gln Pro Pro Pro Val Thr Pro Arg Pro Pro Pro Thr Thr Pro Gln
 385 390 395 400
 Pro Pro Pro Ser Thr Ala Val Ile Cys Gln Thr Ala Val Trp Tyr Ala
 405 410 415
 Val Ser Trp Ala Pro Gly Ala Arg Gly Trp Leu Gln Ala Cys His Asp
 420 425 430
 Gln Phe Pro Asp Gln Phe Leu Asp Met Ile Cys Gly Asn Leu Ser Phe

 435 440 445
 Ser Ala Leu Ser Gly Pro Ser Arg Pro Leu Val Lys Gln Leu Cys Ala
 450 455 460
 Gly Leu Leu Pro Pro Pro Thr Ser Cys Pro Pro Gly Leu Ile Pro Val
 465 470 475 480
 Pro Leu Thr Pro Glu Ile Phe Trp Gly Cys Phe Leu Glu Asn Glu Thr
 485 490 495
 Leu Trp Ala Glu Arg Leu Cys Val Glu Asp Ser Leu Gln Ala Val Pro

 500 505 510
 Pro Arg Asn Gln Ala Trp Val Gln His Val Cys Arg Gly Pro Thr Leu

515 520 525
 Asp Ala Thr Asp Phe Pro Pro Cys Arg Val Gly Pro Cys Gly Glu Arg
 530 535 540
 Cys Pro Asp Gly Gly Ser Phe Leu Leu Met Val Cys Ala Asn Asp Thr
 545 550 555 560
 Leu Tyr Glu Ala Leu Val Pro Phe Trp Ala Trp Leu Ala Gly Gln Cys

 565 570 575
 Arg Ile Ser Arg Gly Gly Asn Asp Thr Cys Phe Leu Glu Gly Met Leu
 580 585 590
 Gly Pro Leu Leu Pro Ser Leu Pro Pro Leu Gly Pro Ser Pro Leu Cys
 595 600 605
 Leu Ala Pro Gly Pro Phe Leu Leu Gly Met Leu Ser Gln Leu Pro Arg
 610 615 620
 Cys Gln Ser Ser Val Pro Ala Leu Ala His Pro Thr Arg Leu His Tyr

 625 630 635 640
 Leu Leu Arg Leu Leu Thr Phe Leu Leu Gly Pro Gly Thr Gly Gly Ala
 645 650 655
 Glu Thr Gln Gly Met Leu Gly Gln Ala Leu Leu Leu Ser Ser Leu Pro
 660 665 670
 Asp Asn Cys Ser Phe Trp Asp Ala Phe Arg Pro Glu Gly Arg Arg Ser
 675 680 685
 Val Leu Arg Thr Val Gly Glu Tyr Leu Gln Arg Glu Glu Pro Thr Pro

 690 695 700
 Pro Gly Leu Asp Ser Ser Leu Ser Leu Gly Ser Gly Met Ser Lys Met
 705 710 715 720
 Glu Leu Leu Ser

<210> 17

<211> 3423

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 17

atggctctca gcctctggcc cctgctgctg ctgctgctgc tgctgctgct gctgtccttt	60
gcaatcaaga tagctacaag gaagtatctt ggcaaacaaa acgtttatga tattggagtc	120
gaaagagatc acaactttgc tctgaagaac ggattcatag cttctaattg ctttagtcct	180
gtgctgtggg atctgctcca gagggaaaag agtgtttggg ccctgcagat tctagtgcag	240
gcgtacctgc atatgcccc agaaaacctc cagcagctgg tgctttcagc agagaggagag	300
gctgcacagg gcttctgac actcatgctg caggggaagc tgcaggggaa gctgcaggtg	360
ccaccatccg aggagcaggc cctgggtcgc ctgacagccc tgctgctcca gcggtaccca	420
cgcctcacct ccagctctt cattgacctg tcaccactca tccctttctt ggctgtctct	480
gacctgatgc gcttcccacc atccctgtta gccaacgaca gtgtcctggc tgccatccgg	540
gattacagcc caggaatgag gcctgaacag aaggaggctc tggcaaagcg actgctggcc	600
cctgaactgt ttggggaagt gcctgcctgg ccccaggagc tgctgtgggc agtgctgccc	660
ctgtccccc acctccctct ggagaacttt ttgcagctca gccctacca gatccaggcc	720
ctggaggata gctggccagc agcaggctcg gggccagggc atgcccgcca tgtgctgcgc	780
agcctggtaa accagagtgt ccaggatggt gaggagcagg tacgcaggct tgggccctc	840
gcctgtttcc tgagccctga ggagctgcag agcctagtgc ccctgagtga tccaacgggg	900
ccagtagaac gggggctgct ggaatgtgca gccaatggga ccctcagccc agaaggacgg	960
gtggcatatg aacttctggg tgtgttgccg tcatctggag gagcgggtgt gagccccgg	1020
gagctgcggg tcigggcccc tctcttctct cagctgggcc tccgcttctt tcaggagctg	1080
tcagagcccc agcttagagc catgttctt gtctgcagg gaactagtgt tacacctgct	1140
caggtgttcc tgctgcttgg acggtctctt cctaggcacg atctatccct ggaggaaactc	1200
tgctccttgc accttctgct accaggctc agccccaga cactccaggc catccctagg	1260
cgagtcttgg tcggggcttg ttctgcctg gccctgaac tgtcacgct ctcagcctgc	1320
cagaccgcag cactgctgca gaccttctgg gttaaagatg gtgttaaaaa tatgggtaca	1380
acagggtgctg gtccagctgt gtgtatccct ggtcagccta ttcccaccac ctggccagac	1440
tgctgtctc cctgtctccc attaaagctg ctacaactgg attccttggc tcttctggca	1500
aatcgaagac gctactggga gctgccctgg tctgagcagc aggcacagtt tctctggaag	1560
aagatgcaag taccaccaa ccttacctc aggaatctgc aggtcttggg caccctggca	1620
ggaggcatgt cctgtgagtt tctgcagcag atcaactcca tggtagactt ccttgaagtg	1680
gtgcacatga tctatcagct gccactaga gttcgaggga gccaggggc ctgtatctgg	1740

gcagagctac agcggaggat ggcaatgcc aaccagaat ggacaactgt agggccagaa 1800

ctgaacgggc tggatagcaa gctactcctg gacttaccga tccagttgat ggacagacta 1860

tccaatgaat ccattatgtt ggtggtggag ctggtgcaaa gagctccaga gcagctgctg 1920

gcactgacct ccctccacca ggcagccctg gcagagaggg cactacaaaa cctggctcca 1980

aaggagactc cagtctcagg ggaagtgtg gagaccttag gcccttttgt tggattcctg 2040

gggacagaga gcacacgaca gatecccccta cagatcctgc tgtcccatct cagtcagctg 2100

caaggcttct gcctaggaga gacatttgcc acagagctgg gatggctgct attgcaggag 2160

tctgttcttg ggaaccaga gttgtggagc caggatgaag tagagcaagc tggacgccta 2220

gtattcactc tgictactga ggcaatttcc ttgatcccca gggaggcctt gggccagag 2280

accctggagc ggcttctaga aaagcagcag agctgggagc agagcagagt tggacagctg 2340

tgtaggagc cacagcttgc tgccaagaaa ggcacctgg tagcagggtt ggtgcgacca 2400

gctgctgagg atcttccaga acctgtgcca aattgtgcag atgtacagg gacattccca 2460

gcagcctggt ctgcaacca gattgcagag atggagctct cagactttga ggactgcctg 2520

acattatttg caggagacc aggacttggg cctgaggaac tgcgggcagc catgggcaaa 2580

gcaaaacagt tgtgggggcc ccccgggga tttcgtcctg agcagatcct gcagcttggt 2640

aggtcttaa taggtctagg agatcgggaa ctacaggagc tgatcctagt ggactgggga 2700

gtgctgagca ccctggggca gatagatggc tggagcacca ctgagctccg cattgtggtc 2760

tccagtttcc tacggcagag tggctggcat gtgagccacc tggacttcgt tcattgaca 2820

gcgtgggtt atactctctg tggactgcgg ccagaggagc tccagcacat cagcagttgg 2880

gagttcagcc aagcagctct ctctctggc acctgcatc tccagtgtc tgaggaacaa 2940

ctggagggtc tgcccacct acttgtactg cctggtgggt ttggcccaat cagtaactgg 3000

gggcctgaga tcttcaactga aattggcacc atagcagctg ggatcccaga cctggtctt 3060

tcagcactgc tgcggggaca gatccagggc gttactctc ttgccatttc tgcatccct 3120

cctctaaat ttgtgtggt gtttagtccc atccaactat ctagtctcac cagtgtcag 3180

gctgtggctg tcaactctga gcaaatggcc tttctgagtc ctgagcagcg acgagcagtt 3240

gcatgggccc aacatgaggg aaaggagagc ccagaacagc aaggctgaag tacagcctgg 3300

ggcctccagg actggtcacg accttcttgg tccttggtat tgactatcag ctctcttggc 3360

cacctgctaa cgcgtacgcg gccgctcgag cagaaactca tctcagaaga ggatctgtaa 3420

tag 3423

<210> 18

<211> 1139

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 18

Met Ala Leu Ser Leu Trp Pro Leu Leu Leu Leu Leu Leu Leu Leu

1 5 10 15

Leu Leu Ser Phe Ala Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys

20 25 30

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

35 40 45

Lys Asn Gly Phe Ile Ala Ser Asn Cys Phe Ser Pro Val Leu Trp Asp

50 55 60

Leu Leu Gln Arg Glu Lys Ser Val Trp Ala Leu Gln Ile Leu Val Gln

65 70 75 80

Ala Tyr Leu His Met Pro Pro Glu Asn Leu Gln Gln Leu Val Leu Ser

85 90 95

Ala Glu Arg Glu Ala Ala Gln Gly Phe Leu Thr Leu Met Leu Gln Gly

100 105 110

Lys Leu Gln Gly Lys Leu Gln Val Pro Pro Ser Glu Glu Gln Ala Leu

115 120 125

Gly Arg Leu Thr Ala Leu Leu Leu Gln Arg Tyr Pro Arg Leu Thr Ser

130 135 140

Gln Leu Phe Ile Asp Leu Ser Pro Leu Ile Pro Phe Leu Ala Val Ser

145 150 155 160

Asp Leu Met Arg Phe Pro Pro Ser Leu Leu Ala Asn Asp Ser Val Leu

165 170 175

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

180 185 190

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

195 200 205
 Ala Trp Pro Gln Glu Leu Leu Trp Ala Val Leu Pro Leu Leu Pro His
 210 215 220
 Leu Pro Leu Glu Asn Phe Leu Gln Leu Ser Pro His Gln Ile Gln Ala
 225 230 235 240

 Leu Glu Asp Ser Trp Pro Ala Ala Gly Leu Gly Pro Gly His Ala Arg
 245 250 255
 His Val Leu Arg Ser Leu Val Asn Gln Ser Val Gln Asp Gly Glu Glu
 260 265 270
 Gln Val Arg Arg Leu Gly Pro Leu Ala Cys Phe Leu Ser Pro Glu Glu
 275 280 285
 Leu Gln Ser Leu Val Pro Leu Ser Asp Pro Thr Gly Pro Val Glu Arg
 290 295 300

 Gly Leu Leu Glu Cys Ala Ala Asn Gly Thr Leu Ser Pro Glu Gly Arg
 305 310 315 320
 Val Ala Tyr Glu Leu Leu Gly Val Leu Arg Ser Ser Gly Gly Ala Val
 325 330 335
 Leu Ser Pro Arg Glu Leu Arg Val Trp Ala Pro Leu Phe Ser Gln Leu
 340 345 350
 Gly Leu Arg Phe Leu Gln Glu Leu Ser Glu Pro Gln Leu Arg Ala Met
 355 360 365

 Leu Pro Val Leu Gln Gly Thr Ser Val Thr Pro Ala Gln Ala Val Leu
 370 375 380
 Leu Leu Gly Arg Leu Leu Pro Arg His Asp Leu Ser Leu Glu Glu Leu
 385 390 395 400
 Cys Ser Leu His Leu Leu Leu Pro Gly Leu Ser Pro Gln Thr Leu Gln
 405 410 415
 Ala Ile Pro Arg Arg Val Leu Val Gly Ala Cys Ser Cys Leu Ala Pro
 420 425 430

 Glu Leu Ser Arg Leu Ser Ala Cys Gln Thr Ala Ala Leu Leu Gln Thr
 435 440 445

Phe Arg Val Lys Asp Gly Val Lys Asn Met Gly Thr Thr Gly Ala Gly
 450 455 460
 Pro Ala Val Cys Ile Pro Gly Gln Pro Ile Pro Thr Thr Trp Pro Asp
 465 470 475 480
 Cys Leu Leu Pro Leu Leu Pro Leu Lys Leu Leu Gln Leu Asp Ser Leu
 485 490 495
 Ala Leu Leu Ala Asn Arg Arg Arg Tyr Trp Glu Leu Pro Trp Ser Glu
 500 505 510
 Gln Gln Ala Gln Phe Leu Trp Lys Lys Met Gln Val Pro Thr Asn Leu
 515 520 525
 Thr Leu Arg Asn Leu Gln Ala Leu Gly Thr Leu Ala Gly Gly Met Ser
 530 535 540
 Cys Glu Phe Leu Gln Gln Ile Asn Ser Met Val Asp Phe Leu Glu Val
 545 550 555 560
 Val His Met Ile Tyr Gln Leu Pro Thr Arg Val Arg Gly Ser Leu Arg
 565 570 575
 Ala Cys Ile Trp Ala Glu Leu Gln Arg Arg Met Ala Met Pro Glu Pro
 580 585 590
 Glu Trp Thr Thr Val Gly Pro Glu Leu Asn Gly Leu Asp Ser Lys Leu
 595 600 605
 Leu Leu Asp Leu Pro Ile Gln Leu Met Asp Arg Leu Ser Asn Glu Ser
 610 615 620
 Ile Met Leu Val Val Glu Leu Val Gln Arg Ala Pro Glu Gln Leu Leu
 625 630 635 640
 Ala Leu Thr Pro Leu His Gln Ala Ala Leu Ala Glu Arg Ala Leu Gln
 645 650 655
 Asn Leu Ala Pro Lys Glu Thr Pro Val Ser Gly Glu Val Leu Glu Thr
 660 665 670
 Leu Gly Pro Leu Val Gly Phe Leu Gly Thr Glu Ser Thr Arg Gln Ile
 675 680 685
 Pro Leu Gln Ile Leu Leu Ser His Leu Ser Gln Leu Gln Gly Phe Cys

690 695 700
 Leu Gly Glu Thr Phe Ala Thr Glu Leu Gly Trp Leu Leu Leu Gln Glu
 705 710 715 720
 Ser Val Leu Gly Lys Pro Glu Leu Trp Ser Gln Asp Glu Val Glu Gln
 725 730 735
 Ala Gly Arg Leu Val Phe Thr Leu Ser Thr Glu Ala Ile Ser Leu Ile
 740 745 750

 Pro Arg Glu Ala Leu Gly Pro Glu Thr Leu Glu Arg Leu Leu Glu Lys
 755 760 765
 Gln Gln Ser Trp Glu Gln Ser Arg Val Gly Gln Leu Cys Arg Glu Pro
 770 775 780
 Gln Leu Ala Ala Lys Lys Ala Ala Leu Val Ala Gly Val Val Arg Pro
 785 790 795 800
 Ala Ala Glu Asp Leu Pro Glu Pro Val Pro Asn Cys Ala Asp Val Arg
 805 810 815

 Gly Thr Phe Pro Ala Ala Trp Ser Ala Thr Gln Ile Ala Glu Met Glu
 820 825 830
 Leu Ser Asp Phe Glu Asp Cys Leu Thr Leu Phe Ala Gly Asp Pro Gly
 835 840 845
 Leu Gly Pro Glu Glu Leu Arg Ala Ala Met Gly Lys Ala Lys Gln Leu
 850 855 860
 Trp Gly Pro Pro Arg Gly Phe Arg Pro Glu Gln Ile Leu Gln Leu Gly
 865 870 875 880

 Arg Leu Leu Ile Gly Leu Gly Asp Arg Glu Leu Gln Glu Leu Ile Leu
 885 890 895
 Val Asp Trp Gly Val Leu Ser Thr Leu Gly Gln Ile Asp Gly Trp Ser
 900 905 910
 Thr Thr Gln Leu Arg Ile Val Val Ser Ser Phe Leu Arg Gln Ser Gly
 915 920 925
 Arg His Val Ser His Leu Asp Phe Val His Leu Thr Ala Leu Gly Tyr
 930 935 940

Thr Leu Cys Gly Leu Arg Pro Glu Glu Leu Gln His Ile Ser Ser Trp
 945 950 955 960
 Glu Phe Ser Gln Ala Ala Leu Phe Leu Gly Thr Leu His Leu Gln Cys
 965 970 975
 Ser Glu Glu Gln Leu Glu Val Leu Ala His Leu Leu Val Leu Pro Gly
 980 985 990
 Gly Phe Gly Pro Ile Ser Asn Trp Gly Pro Glu Ile Phe Thr Glu Ile
 995 1000 1005

Gly Thr Ile Ala Ala Gly Ile Pro Asp Leu Ala Leu Ser Ala Leu
 1010 1015 1020
 Leu Arg Gly Gln Ile Gln Gly Val Thr Pro Leu Ala Ile Ser Val
 1025 1030 1035
 Ile Pro Pro Pro Lys Phe Ala Val Val Phe Ser Pro Ile Gln Leu
 1040 1045 1050
 Ser Ser Leu Thr Ser Ala Gln Ala Val Ala Val Thr Pro Glu Gln
 1055 1060 1065

Met Ala Phe Leu Ser Pro Glu Gln Arg Arg Ala Val Ala Trp Ala
 1070 1075 1080
 Gln His Glu Gly Lys Glu Ser Pro Glu Gln Gln Gly Arg Ser Thr
 1085 1090 1095
 Ala Trp Gly Leu Gln Asp Trp Ser Arg Pro Ser Trp Ser Leu Val
 1100 1105 1110
 Leu Thr Ile Ser Phe Leu Gly His Leu Leu Thr Arg Thr Arg Pro
 1115 1120 1125

Leu Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
 1130 1135

<210> 19

<211> 3414

<212> DNA

<213> Artificial Sequence

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

<400> 19

atggctctga gcctccagcc ccagctgctc cttctcctgt cgctcctgcc gcaggaagtg	60
acttcaatca agatagctac aaggaagtat cttggcaaac aaaacgttta tgatattgga	120
gtcgaagag atcacaactt tgctctgaag aacggattca tagcttctaa ttgcttcagt	180
cctgtactgt gggatctact ccagagagag aagagcgttt gggccctgag gaccctggtg	240
aaggcctacc tgcgcatgcc tccagaagac cttcagcagc ttgtgctttc agcagagatg	300
gaggctgcac agggcttctt gacgctcatg cttcgttcct gggctaagct gaaggttcaa	360
ccatccgagg agcaggccat gggccgctg acagccttgc tgctccagcg gtaccacgc	420
ctcacctccc aactctttat cgacatgtca ccgctcatcc ctttctggc tgtccctgac	480
ctcatgcgt tcccaccgtc ctttttgcc aacgacagtg tcctggctgc catcagggat	540
cacagctcag gaatgaagcc tgaacagaag gaggcctgg caaaacgact gctggccct	600
gagctgtttg gagaagtgcc tgattggccc caggagctgc tgtgggcagc cctgcctctg	660
cttccccatc tgcctctgga gagctttctc cagctcagcc ctcaccagat ccaggccctg	720
gaggatagct ggccagtagc agatcttggg ccgggacacg cccgacatgt gcttcgtagc	780
ctagtaaacc agagcatgga ggatggggag gagcaggtgc tcaggcttgg gtccctcgcc	840
tgtttcctga gtctgagga gctacagagt ctggtgccct tgagtgatcc aatggggcct	900
gtagaacagg gtctgctgga atgtcgggcc aatgggaccc tcagcccaga aggacgggtg	960
gcataatgaac ttctgggagt gttgcgttca tctggaggaa ctgtcttaag cccccgagag	1020
ctgagggtct gggcacctct ctttccccag ctgggcctcc gcttctgca ggagctctca	1080
gagaccagc ttagagccat gcttctgcc ctacaggag ccagtgctac acctgcccag	1140
gctgttctgt tgtttggaag gctccttctt aagcatgac tgtccctgga ggaactctgc	1200
tccctgcacc ctctcctgcc aggtctcagc cccagacac tccaggccat ccctaagaga	1260
gttctggttg gtgcttgctc ctgcctgggc cctgaactgt caaggcttct agcttgccag	1320
attgcagctc tgctgcagac ctttcgggta aaagatggtg ttaaaaatat gggcgcagca	1380
ggcgcggct cagccgtgtg cattcctggg cagcccacca cttggccaga ctgcctgctt	1440
cccctgctcc cattaaagct gctacagctg gacgctgcag ctcttctggc aaaccgaaga	1500
ctctatcggc agctgccttg gtctgagcaa caggcacagt ttctctggaa gaaaatgcaa	1560
gtgcctacca acctgagcct gaggaatctg caggctctgg gcaacttggc aggaggcatg	1620
acctgcgagt ttctgcagca gatcagctca atggttgact ttcttgatgt ggtacacatg	1680
ctctaccagc tgcccactgg tgttcgagag agcctgcggg cctgtatctg gacagagcta	1740
cagcggagga tgacaatgcc agagccagag ctgaccaccc tagggccaga actgagtga	1800

cttgacacaa agctactcct ggacttgccg atccagctga tggacagatt gtccaatgat	1860
tccattatgt tgggtggtgga gatggtccaa ggcgctccag agcagctgct ggcactgacc	1920
ccactccacc agacagcctt ggcagagcga gcacttaaaa acctggctcc aaaggagacc	1980
ccaatctcca aagaagtgct ggagacactg ggccccttgg ttggattcct gggaatagag	2040
agcacgcgac ggatcccttt acccattcta ctgtctcatc tcagtcagct gcagggttc	2100
tgcctaggag agacatttgc cacagagctg ggatggctgc tgttcagga gcctgttctt	2160
ggaaaaccag aattgtggag ccaggatgaa atagagcaag ctggacgcct agtattcact	2220
ctgtctgctg aggtatttct ctcgatcccc agggaggctt tgggccaga gacactggag	2280
aggttcttgg gaaagcatca aagctgggag cagagcagag tgggccatct gtgtggggag	2340
tcacagcttg cccacaagaa agcagctctg gtagctggga ttgtgcatcc agctgctgag	2400
ggtctccaag agcctgtacc aaactgtgca gacatacggg gaaccttccc agcggcctgg	2460
tctgcgacac aaatctcaga gatggaactc tcagactttg aagactgcct gtcactattt	2520
gctggagatc caggacttgg tcttgaggaa ctacgggcag ccatgggcaa ggccaagcag	2580
ttgtggggtc cccctcgagg attccgtcct gaggagatct tgcagctggg ccgtctctctg	2640
ataggcttag gagaacggga actgcaggag ctacacttgg tggactgggg tgtgctgagc	2700
agcctggggc aaatagatgg ctggagtcc atgcagctcc gagccgtggt ctccagtctc	2760
ctaaggcaga gtggtcggca tgtgagccac ctggacttca tttatctgac agcactgggt	2820
tacacagtct gtggattgcg accagaggag ttacagcaca tcagcagttg ggagtttagc	2880
caagcagctc tcttcttggg tagcttgcct ctcccgtgct ctgaggaaca gctggaagtt	2940
ctggcctatc tcttctgttt gcctgggtggc ttggcccag tcagtaactg ggggcctgag	3000
atcttactg aaattggcac aatagcagct ggcatcccag acctggctct ttcagcatta	3060
ctgcggggac agatccaagg cctgactcct cttgccattt ctgtcattcc tgctccaag	3120
tttgagtgg tcttcaacc catccagtta tctagtctca ccagggttca ggccgtagct	3180
gttactcctg aacagctggc ctatctgagt cctgagcagc ggcgagcagt tgcattgggc	3240
caacacgaag ggaaggagat cccagagcag ctgggtcgaa actcagcctg gggctcttac	3300
gactggttcc aagcctcctg ggccctggca ttccccgtca gcatttttgg ccacctatta	3360
acgctgacgc ggccgctcga gcagaaactc atctcagaag aggatctgta atag	3414
<210> 20	
<211> 1136	
<212> PRT	
<213> Artificial Sequence	

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

polypeptide

<400> 20

Met Ala Leu Ser Leu Gln Pro Gln Leu Leu Leu Leu Ser Leu Leu

1 5 10 15
Pro Gln Glu Val Thr Ser Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly

20 25 30
Lys Gln Asn Val Tyr Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala

35 40 45
Leu Lys Asn Gly Phe Ile Ala Ser Asn Cys Phe Ser Pro Val Leu Trp

50 55 60
Asp Leu Leu Gln Arg Glu Lys Ser Val Trp Ala Leu Arg Thr Leu Val

65 70 75 80
Lys Ala Tyr Leu Arg Met Pro Pro Glu Asp Leu Gln Gln Leu Val Leu

85 90 95
Ser Ala Glu Met Glu Ala Ala Gln Gly Phe Leu Thr Leu Met Leu Arg

100 105 110
Ser Trp Ala Lys Leu Lys Val Gln Pro Ser Glu Glu Gln Ala Met Gly

115 120 125
Arg Leu Thr Ala Leu Leu Leu Gln Arg Tyr Pro Arg Leu Thr Ser Gln

130 135 140
Leu Phe Ile Asp Met Ser Pro Leu Ile Pro Phe Leu Ala Val Pro Asp

145 150 155 160
Leu Met Arg Phe Pro Pro Ser Leu Leu Ala Asn Asp Ser Val Leu Ala

165 170 175
Ala Ile Arg Asp His Ser Ser Gly Met Lys Pro Glu Gln Lys Glu Ala

180 185 190
Leu Ala Lys Arg Leu Leu Ala Pro Glu Leu Phe Gly Glu Val Pro Asp

195 200 205
Trp Pro Gln Glu Leu Leu Trp Ala Ala Leu Pro Leu Leu Pro His Leu

210 215 220

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

465 470 475 480
 Pro Leu Leu Pro Leu Lys Leu Leu Gln Leu Asp Ala Ala Ala Leu Leu
 485 490 495
 Ala Asn Arg Arg Leu Tyr Arg Gln Leu Pro Trp Ser Glu Gln Gln Ala
 500 505 510
 Gln Phe Leu Trp Lys Lys Met Gln Val Pro Thr Asn Leu Ser Leu Arg

 515 520 525
 Asn Leu Gln Ala Leu Gly Asn Leu Ala Gly Gly Met Thr Cys Glu Phe
 530 535 540
 Leu Gln Gln Ile Ser Ser Met Val Asp Phe Leu Asp Val Val His Met
 545 550 555 560
 Leu Tyr Gln Leu Pro Thr Gly Val Arg Glu Ser Leu Arg Ala Cys Ile
 565 570 575
 Trp Thr Glu Leu Gln Arg Arg Met Thr Met Pro Glu Pro Glu Leu Thr

 580 585 590
 Thr Leu Gly Pro Glu Leu Ser Glu Leu Asp Thr Lys Leu Leu Leu Asp
 595 600 605
 Leu Pro Ile Gln Leu Met Asp Arg Leu Ser Asn Asp Ser Ile Met Leu
 610 615 620
 Val Val Glu Met Val Gln Gly Ala Pro Glu Gln Leu Leu Ala Leu Thr
 625 630 635 640
 Pro Leu His Gln Thr Ala Leu Ala Glu Arg Ala Leu Lys Asn Leu Ala

 645 650 655
 Pro Lys Glu Thr Pro Ile Ser Lys Glu Val Leu Glu Thr Leu Gly Pro
 660 665 670
 Leu Val Gly Phe Leu Gly Ile Glu Ser Thr Arg Arg Ile Pro Leu Pro
 675 680 685
 Ile Leu Leu Ser His Leu Ser Gln Leu Gln Gly Phe Cys Leu Gly Glu
 690 695 700
 Thr Phe Ala Thr Glu Leu Gly Trp Leu Leu Leu Gln Glu Pro Val Leu

 705 710 715 720

Gly Lys Pro Glu Leu Trp Ser Gln Asp Glu Ile Glu Gln Ala Gly Arg
 725 730 735
 Leu Val Phe Thr Leu Ser Ala Glu Ala Ile Ser Ser Ile Pro Arg Glu
 740 745 750
 Ala Leu Gly Pro Glu Thr Leu Glu Arg Leu Leu Gly Lys His Gln Ser
 755 760 765
 Trp Glu Gln Ser Arg Val Gly His Leu Cys Gly Glu Ser Gln Leu Ala
 770 775 780
 His Lys Lys Ala Ala Leu Val Ala Gly Ile Val His Pro Ala Ala Glu
 785 790 795 800
 Gly Leu Gln Glu Pro Val Pro Asn Cys Ala Asp Ile Arg Gly Thr Phe
 805 810 815
 Pro Ala Ala Trp Ser Ala Thr Gln Ile Ser Glu Met Glu Leu Ser Asp
 820 825 830
 Phe Glu Asp Cys Leu Ser Leu Phe Ala Gly Asp Pro Gly Leu Gly Pro
 835 840 845
 Glu Glu Leu Arg Ala Ala Met Gly Lys Ala Lys Gln Leu Trp Gly Pro
 850 855 860
 Pro Arg Gly Phe Arg Pro Glu Gln Ile Leu Gln Leu Gly Arg Leu Leu
 865 870 875 880
 Ile Gly Leu Gly Glu Arg Glu Leu Gln Glu Leu Thr Leu Val Asp Trp
 885 890 895
 Gly Val Leu Ser Ser Leu Gly Gln Ile Asp Gly Trp Ser Ser Met Gln
 900 905 910
 Leu Arg Ala Val Val Ser Ser Phe Leu Arg Gln Ser Gly Arg His Val
 915 920 925
 Ser His Leu Asp Phe Ile Tyr Leu Thr Ala Leu Gly Tyr Thr Val Cys
 930 935 940
 Gly Leu Arg Pro Glu Glu Leu Gln His Ile Ser Ser Trp Glu Phe Ser
 945 950 955 960
 Gln Ala Ala Leu Phe Leu Gly Ser Leu His Leu Pro Cys Ser Glu Glu

965 970 975
Gln Leu Glu Val Leu Ala Tyr Leu Leu Val Leu Pro Gly Gly Phe Gly
980 985 990
Pro Val Ser Asn Trp Gly Pro Glu Ile Phe Thr Glu Ile Gly Thr Ile
995 1000 1005
Ala Ala Gly Ile Pro Asp Leu Ala Leu Ser Ala Leu Leu Arg Gly
1010 1015 1020
Gln Ile Gln Gly Leu Thr Pro Leu Ala Ile Ser Val Ile Pro Ala

1025 1030 1035
Pro Lys Phe Ala Val Val Phe Asn Pro Ile Gln Leu Ser Ser Leu
1040 1045 1050
Thr Arg Gly Gln Ala Val Ala Val Thr Pro Glu Gln Leu Ala Tyr
1055 1060 1065
Leu Ser Pro Glu Gln Arg Arg Ala Val Ala Trp Ala Gln His Glu
1070 1075 1080
Gly Lys Glu Ile Pro Glu Gln Leu Gly Arg Asn Ser Ala Trp Gly

1085 1090 1095
Leu Tyr Asp Trp Phe Gln Ala Ser Trp Ala Leu Ala Leu Pro Val
1100 1105 1110
Ser Ile Phe Gly His Leu Leu Thr Arg Thr Arg Pro Leu Glu Gln
1115 1120 1125
Lys Leu Ile Ser Glu Glu Asp Leu
1130 1135

<210> 21

<211> 105

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 21

atcaagatag ctacaaggaa gtatcttggc aaacaaaacg tttatgatat tggagtcgaa 60
agagatcaca actttgctct gaagaacgga ttcatagctt ctaat 105

<210> 22

<211> 35

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 22

Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn Val Tyr Asp

1 5 10 15

Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn Gly Phe Ile

20 25 30

Ala Ser Asn

35

<210> 23

<211> 1067

<212> PRT

<213> Homo sapiens

<400> 23

Cys Phe Ser Pro Val Leu Trp Asp Leu Leu Gln Arg Glu Lys Ser Val

1 5 10 15

Trp Ala Leu Gln Ile Leu Val Gln Ala Tyr Leu His Met Pro Pro Glu

20 25 30

Asn Leu Gln Gln Leu Val Leu Ser Ala Glu Arg Glu Ala Ala Gln Gly

35 40 45

Phe Leu Thr Leu Met Leu Gln Gly Lys Leu Gln Gly Lys Leu Gln Val

50 55 60

Pro Pro Ser Glu Glu Gln Ala Leu Gly Arg Leu Thr Ala Leu Leu Leu

65 70 75 80

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

85 90 95

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

100 105 110

Leu Leu Ala Asn Asp Ser Val Leu Ala Ala Ile Arg Asp Tyr Ser Pro
 115 120 125
 Gly Met Arg Pro Glu Gln Lys Glu Ala Leu Ala Lys Arg Leu Leu Ala
 130 135 140
 Pro Glu Leu Phe Gly Glu Val Pro Ala Trp Pro Gln Glu Leu Leu Trp
 145 150 155 160
 Ala Val Leu Pro Leu Leu Pro His Leu Pro Leu Glu Asn Phe Leu Gln
 165 170 175

 Leu Ser Pro His Gln Ile Gln Ala Leu Glu Asp Ser Trp Pro Ala Ala
 180 185 190
 Gly Leu Gly Pro Gly His Ala Arg His Val Leu Arg Ser Leu Val Asn
 195 200 205
 Gln Ser Val Gln Asp Gly Glu Glu Gln Val Arg Arg Leu Gly Pro Leu
 210 215 220
 Ala Cys Phe Leu Ser Pro Glu Glu Leu Gln Ser Leu Val Pro Leu Ser
 225 230 235 240

 Asp Pro Thr Gly Pro Val Glu Arg Gly Leu Leu Glu Cys Ala Ala Asn
 245 250 255
 Gly Thr Leu Ser Pro Glu Gly Arg Val Ala Tyr Glu Leu Leu Gly Val
 260 265 270
 Leu Arg Ser Ser Gly Gly Ala Val Leu Ser Pro Arg Glu Leu Arg Val
 275 280 285
 Trp Ala Pro Leu Phe Ser Gln Leu Gly Leu Arg Phe Leu Gln Glu Leu
 290 295 300

 Ser Glu Pro Gln Leu Arg Ala Met Leu Pro Val Leu Gln Gly Thr Ser
 305 310 315 320
 Val Thr Pro Ala Gln Ala Val Leu Leu Leu Gly Arg Leu Leu Pro Arg
 325 330 335
 His Asp Leu Ser Leu Glu Glu Leu Cys Ser Leu His Leu Leu Leu Pro
 340 345 350
 Gly Leu Ser Pro Gln Thr Leu Gln Ala Ile Pro Arg Arg Val Leu Val

355 Gly Ala Cys Ser Cys Leu Ala Pro Glu Leu Ser Arg Leu Ser Ala Cys
 370 375 380
 Gln Thr Ala Ala Leu Leu Gln Thr Phe Arg Val Lys Asp Gly Val Lys
 385 390 395 400
 Asn Met Gly Thr Thr Gly Ala Gly Pro Ala Val Cys Ile Pro Gly Gln
 405 410 415
 Pro Ile Pro Thr Thr Trp Pro Asp Cys Leu Leu Pro Leu Leu Pro Leu
 420 425 430

 Lys Leu Leu Gln Leu Asp Ser Leu Ala Leu Leu Ala Asn Arg Arg Arg
 435 440 445
 Tyr Trp Glu Leu Pro Trp Ser Glu Gln Gln Ala Gln Phe Leu Trp Lys
 450 455 460
 Lys Met Gln Val Pro Thr Asn Leu Thr Leu Arg Asn Leu Gln Ala Leu
 465 470 475 480
 Gly Thr Leu Ala Gly Gly Met Ser Cys Glu Phe Leu Gln Gln Ile Asn
 485 490 495

 Ser Met Val Asp Phe Leu Glu Val Val His Met Ile Tyr Gln Leu Pro
 500 505 510
 Thr Arg Val Arg Gly Ser Leu Arg Ala Cys Ile Trp Ala Glu Leu Gln
 515 520 525
 Arg Arg Met Ala Met Pro Glu Pro Glu Trp Thr Thr Val Gly Pro Glu
 530 535 540
 Leu Asn Gly Leu Asp Ser Lys Leu Leu Leu Asp Leu Pro Ile Gln Leu
 545 550 555 560

 Met Asp Arg Leu Ser Asn Glu Ser Ile Met Leu Val Val Glu Leu Val
 565 570 575
 Gln Arg Ala Pro Glu Gln Leu Leu Ala Leu Thr Pro Leu His Gln Ala
 580 585 590
 Ala Leu Ala Glu Arg Ala Leu Gln Asn Leu Ala Pro Lys Glu Thr Pro
 595 600 605

Val Ser Gly Glu Val Leu Glu Thr Leu Gly Pro Leu Val Gly Phe Leu
610 615 620

Gly Thr Glu Ser Thr Arg Gln Ile Pro Leu Gln Ile Leu Leu Ser His
625 630 635 640

Leu Ser Gln Leu Gln Gly Phe Cys Leu Gly Glu Thr Phe Ala Thr Glu
645 650 655

Leu Gly Trp Leu Leu Leu Gln Glu Ser Val Leu Gly Lys Pro Glu Leu
660 665 670

Trp Ser Gln Asp Glu Val Glu Gln Ala Gly Arg Leu Val Phe Thr Leu
675 680 685

Ser Thr Glu Ala Ile Ser Leu Ile Pro Arg Glu Ala Leu Gly Pro Glu
690 695 700

Thr Leu Glu Arg Leu Leu Glu Lys Gln Gln Ser Trp Glu Gln Ser Arg
705 710 715 720

Val Gly Gln Leu Cys Arg Glu Pro Gln Leu Ala Ala Lys Lys Ala Ala
725 730 735

Leu Val Ala Gly Val Val Arg Pro Ala Ala Glu Asp Leu Pro Glu Pro
740 745 750

Val Pro Asn Cys Ala Asp Val Arg Gly Thr Phe Pro Ala Ala Trp Ser
755 760 765

Ala Thr Gln Ile Ala Glu Met Glu Leu Ser Asp Phe Glu Asp Cys Leu
770 775 780

Thr Leu Phe Ala Gly Asp Pro Gly Leu Gly Pro Glu Glu Leu Arg Ala
785 790 795 800

Ala Met Gly Lys Ala Lys Gln Leu Trp Gly Pro Pro Arg Gly Phe Arg
805 810 815

Pro Glu Gln Ile Leu Gln Leu Gly Arg Leu Leu Ile Gly Leu Gly Asp
820 825 830

Arg Glu Leu Gln Glu Leu Ile Leu Val Asp Trp Gly Val Leu Ser Thr
835 840 845

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

850 855 860
 Ser Ser Phe Leu Arg Gln Ser Gly Arg His Val Ser His Leu Asp Phe
 865 870 875 880

 Val His Leu Thr Ala Leu Gly Tyr Thr Leu Cys Gly Leu Arg Pro Glu
 885 890 895
 Glu Leu Gln His Ile Ser Ser Trp Glu Phe Ser Gln Ala Ala Leu Phe
 900 905 910
 Leu Gly Thr Leu His Leu Gln Cys Ser Glu Glu Gln Leu Glu Val Leu
 915 920 925
 Ala His Leu Leu Val Leu Pro Gly Gly Phe Gly Pro Ile Ser Asn Trp
 930 935 940

 Gly Pro Glu Ile Phe Thr Glu Ile Gly Thr Ile Ala Ala Gly Ile Pro
 945 950 955 960
 Asp Leu Ala Leu Ser Ala Leu Leu Arg Gly Gln Ile Gln Gly Val Thr
 965 970 975
 Pro Leu Ala Ile Ser Val Ile Pro Pro Pro Lys Phe Ala Val Val Phe
 980 985 990
 Ser Pro Ile Gln Leu Ser Ser Leu Thr Ser Ala Gln Ala Val Ala Val
 995 1000 1005

 Thr Pro Glu Gln Met Ala Phe Leu Ser Pro Glu Gln Arg Arg Ala
 1010 1015 1020
 Val Ala Trp Ala Gln His Glu Gly Lys Glu Ser Pro Glu Gln Gln
 1025 1030 1035
 Gly Arg Ser Thr Ala Trp Gly Leu Gln Asp Trp Ser Arg Pro Ser
 1040 1045 1050
 Trp Ser Leu Val Leu Thr Ile Ser Phe Leu Gly His Leu Leu
 1055 1060 1065

<210> 24

<211> 1063

<212> PRT

<213> Mus musculus

<400> 24

Cys Phe Ser Pro Val Leu Trp Asp Leu Leu Gln Arg Glu Lys Ser Val

1 5 10 15

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

20 25 30

Asp Leu Gln Gln Leu Val Leu Ser Ala Glu Met Glu Ala Ala Gln Gly

35 40 45

Phe Leu Thr Leu Met Leu Arg Ser Trp Ala Lys Leu Lys Val Gln Pro

50 55 60

Ser Glu Glu Gln Ala Met Gly Arg Leu Thr Ala Leu Leu Leu Gln Arg

65 70 75 80

Tyr Pro Arg Leu Thr Ser Gln Leu Phe Ile Asp Met Ser Pro Leu Ile

85 90 95

Pro Phe Leu Ala Val Pro Asp Leu Met Arg Phe Pro Pro Ser Leu Leu

100 105 110

Ala Asn Asp Ser Val Leu Ala Ala Ile Arg Asp His Ser Ser Gly Met

115 120 125

Lys Pro Glu Gln Lys Glu Ala Leu Ala Lys Arg Leu Leu Ala Pro Glu

130 135 140

Leu Phe Gly Glu Val Pro Asp Trp Pro Gln Glu Leu Leu Trp Ala Ala

145 150 155 160

Leu Pro Leu Leu Pro His Leu Pro Leu Glu Ser Phe Leu Gln Leu Ser

165 170 175

Pro His Gln Ile Gln Ala Leu Glu Asp Ser Trp Pro Val Ala Asp Leu

180 185 190

Gly Pro Gly His Ala Arg His Val Leu Arg Ser Leu Val Asn Gln Ser

195 200 205

Met Glu Asp Gly Glu Glu Gln Val Leu Arg Leu Gly Ser Leu Ala Cys

210 215 220

Phe Leu Ser Pro Glu Glu Leu Gln Ser Leu Val Pro Leu Ser Asp Pro

225 230 235 240

Met Gly Pro Val Glu Gln Gly Leu Leu Glu Cys Ala Ala Asn Gly Thr

245 250 255

Leu Ser Pro Glu Gly Arg Val Ala Tyr Glu Leu Leu Gly Val Leu Arg

260 265 270

Ser Ser Gly Gly Thr Val Leu Ser Pro Arg Glu Leu Arg Val Trp Ala

275 280 285

Pro Leu Phe Pro Gln Leu Gly Leu Arg Phe Leu Gln Glu Leu Ser Glu

290 295 300

Thr Gln Leu Arg Ala Met Leu Pro Ala Leu Gln Gly Ala Ser Val Thr

305 310 315 320

Pro Ala Gln Ala Val Leu Leu Phe Gly Arg Leu Leu Pro Lys His Asp

325 330 335

Leu Ser Leu Glu Glu Leu Cys Ser Leu His Pro Leu Leu Pro Gly Leu

340 345 350

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

355 360 365

Cys Ser Cys Leu Gly Pro Glu Leu Ser Arg Leu Ser Ala Cys Gln Ile

370 375 380

Ala Ala Leu Leu Gln Thr Phe Arg Val Lys Asp Gly Val Lys Asn Met

385 390 395 400

Gly Ala Ala Gly Ala Gly Ser Ala Val Cys Ile Pro Gly Gln Pro Thr

405 410 415

Thr Trp Pro Asp Cys Leu Leu Pro Leu Leu Pro Leu Lys Leu Leu Gln

420 425 430

Leu Asp Ala Ala Ala Leu Leu Ala Asn Arg Arg Leu Tyr Arg Gln Leu

435 440 445

Pro Trp Ser Glu Gln Gln Ala Gln Phe Leu Trp Lys Lys Met Gln Val

450 455 460

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

465 470 475 480

Gly Gly Met Thr Cys Glu Phe Leu Gln Gln Ile Ser Ser Met Val Asp

485 490 495
Phe Leu Asp Val Val His Met Leu Tyr Gln Leu Pro Thr Gly Val Arg

500 505 510
Glu Ser Leu Arg Ala Cys Ile Trp Thr Glu Leu Gln Arg Arg Met Thr

515 520 525
Met Pro Glu Pro Glu Leu Thr Thr Leu Gly Pro Glu Leu Ser Glu Leu

530 535 540
Asp Thr Lys Leu Leu Leu Asp Leu Pro Ile Gln Leu Met Asp Arg Leu

545 550 555 560
Ser Asn Asp Ser Ile Met Leu Val Val Glu Met Val Gln Gly Ala Pro

565 570 575
Glu Gln Leu Leu Ala Leu Thr Pro Leu His Gln Thr Ala Leu Ala Glu

580 585 590
Arg Ala Leu Lys Asn Leu Ala Pro Lys Glu Thr Pro Ile Ser Lys Glu

595 600 605
Val Leu Glu Thr Leu Gly Pro Leu Val Gly Phe Leu Gly Ile Glu Ser

610 615 620
Thr Arg Arg Ile Pro Leu Pro Ile Leu Leu Ser His Leu Ser Gln Leu

625 630 635 640
Gln Gly Phe Cys Leu Gly Glu Thr Phe Ala Thr Glu Leu Gly Trp Leu

645 650 655
Leu Leu Gln Glu Pro Val Leu Gly Lys Pro Glu Leu Trp Ser Gln Asp

660 665 670
Glu Ile Glu Gln Ala Gly Arg Leu Val Phe Thr Leu Ser Ala Glu Ala

675 680 685
Ile Ser Ser Ile Pro Arg Glu Ala Leu Gly Pro Glu Thr Leu Glu Arg

690 695 700
Leu Leu Gly Lys His Gln Ser Trp Glu Gln Ser Arg Val Gly His Leu

705 710 715 720
Cys Gly Glu Ser Gln Leu Ala His Lys Lys Ala Ala Leu Val Ala Gly

725 730 735

Ile Val His Pro Ala Ala Glu Gly Leu Gln Glu Pro Val Pro Asn Cys
740 745 750
Ala Asp Ile Arg Gly Thr Phe Pro Ala Ala Trp Ser Ala Thr Gln Ile
755 760 765
Ser Glu Met Glu Leu Ser Asp Phe Glu Asp Cys Leu Ser Leu Phe Ala
770 775 780
Gly Asp Pro Gly Leu Gly Pro Glu Glu Leu Arg Ala Ala Met Gly Lys
785 790 795 800
Ala Lys Gln Leu Trp Gly Pro Pro Arg Gly Phe Arg Pro Glu Gln Ile
805 810 815
Leu Gln Leu Gly Arg Leu Leu Ile Gly Leu Gly Glu Arg Glu Leu Gln
820 825 830
Glu Leu Thr Leu Val Asp Trp Gly Val Leu Ser Ser Leu Gly Gln Ile
835 840 845
Asp Gly Trp Ser Ser Met Gln Leu Arg Ala Val Val Ser Ser Phe Leu
850 855 860
Arg Gln Ser Gly Arg His Val Ser His Leu Asp Phe Ile Tyr Leu Thr
865 870 875 880
Ala Leu Gly Tyr Thr Val Cys Gly Leu Arg Pro Glu Glu Leu Gln His
885 890 895
Ile Ser Ser Trp Glu Phe Ser Gln Ala Ala Leu Phe Leu Gly Ser Leu
900 905 910
His Leu Pro Cys Ser Glu Glu Gln Leu Glu Val Leu Ala Tyr Leu Leu
915 920 925
Val Leu Pro Gly Gly Phe Gly Pro Val Ser Asn Trp Gly Pro Glu Ile
930 935 940
Phe Thr Glu Ile Gly Thr Ile Ala Ala Gly Ile Pro Asp Leu Ala Leu
945 950 955 960
Ser Ala Leu Leu Arg Gly Gln Ile Gln Gly Leu Thr Pro Leu Ala Ile
965 970 975
Ser Val Ile Pro Ala Pro Lys Phe Ala Val Val Phe Asn Pro Ile Gln

980 985 990
 Leu Ser Ser Leu Thr Arg Gly Gln Ala Val Ala Val Thr Pro Glu Gln
 995 1000 1005
 Leu Ala Tyr Leu Ser Pro Glu Gln Arg Arg Ala Val Ala Trp Ala

1010 1015 1020
 Gln His Glu Gly Lys Glu Ile Pro Glu Gln Leu Gly Arg Asn Ser
 1025 1030 1035
 Ala Trp Gly Leu Tyr Asp Trp Phe Gln Ala Ser Trp Ala Leu Ala
 1040 1045 1050
 Leu Pro Val Ser Ile Phe Gly His Leu Leu
 1055 1060

<210> 25

<211> 1775

<212> PRT

<213> Homo sapiens

<400> 25

Met Ala Leu Ser Leu Trp Pro Leu Leu Leu Leu Leu Leu Leu Leu

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

65 70 75 80
 Gly Glu Pro Pro Pro Leu Gln Pro Pro Ala Leu Arg Leu His Asp Phe
 85 90 95
 Leu Val Thr Leu Arg Gly Ser Pro Asp Trp Glu Pro Met Leu Gly Leu
 100 105 110
 Leu Gly Asp Met Leu Ala Leu Leu Gly Gln Glu Gln Thr Pro Arg Asp
 115 120 125

Phe Leu Val His Gln Ala Gly Val Leu Gly Gly Leu Val Glu Val Leu

130

135

140

Leu Gly Ala Leu Val Pro Gly Gly Pro Pro Thr Pro Thr Arg Pro Pro

145

150

155

160

Cys Thr Arg Asp Gly Pro Ser Asp Cys Val Leu Ala Ala Asp Trp Leu

165

170

175

Pro Ser Leu Leu Leu Leu Glu Gly Thr Arg Trp Gln Ala Leu Val

180

185

190

Gln Val Gln Pro Ser Val Asp Pro Thr Asn Ala Thr Gly Leu Asp Gly

195

200

205

Arg Glu Ala Ala Pro His Phe Leu Gln Gly Leu Leu Gly Leu Leu Thr

210

215

220

Pro Thr Gly Glu Leu Gly Ser Lys Glu Ala Leu Trp Gly Gly Leu Leu

225

230

235

240

Arg Thr Val Gly Ala Pro Leu Tyr Ala Ala Phe Gln Glu Gly Leu Leu

245

250

255

Arg Val Thr His Ser Leu Gln Asp Glu Val Phe Ser Ile Leu Gly Gln

260

265

270

Pro Glu Pro Asp Thr Asn Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln

275

280

285

Leu Leu Leu Trp Gly Val Arg His Asn Leu Ser Trp Asp Val Gln Ala

290

295

300

Leu Gly Phe Leu Ser Gly Ser Pro Pro Pro Pro Pro Ala Leu Leu His

305

310

315

320

Cys Leu Ser Thr Gly Val Pro Leu Pro Arg Ala Ser Gln Pro Ser Ala

325

330

335

His Ile Ser Pro Arg Gln Arg Arg Ala Ile Thr Val Glu Ala Leu Cys

340

345

350

Glu Asn His Leu Gly Pro Ala Pro Pro Tyr Ser Ile Ser Asn Phe Ser

355

360

365

Ile His Leu Leu Cys Gln His Thr Lys Pro Ala Thr Pro Gln Pro His

370 375 380
 Pro Ser Thr Thr Ala Ile Cys Gln Thr Ala Val Trp Tyr Ala Val Ser

 385 390 395 400
 Trp Ala Pro Gly Ala Gln Gly Trp Leu Gln Ala Cys His Asp Gln Phe
 405 410 415
 Pro Asp Glu Phe Leu Asp Ala Ile Cys Ser Asn Leu Ser Phe Ser Ala
 420 425 430
 Leu Ser Gly Ser Asn Arg Arg Leu Val Lys Arg Leu Cys Ala Gly Leu
 435 440 445
 Leu Pro Pro Pro Thr Ser Cys Pro Glu Gly Leu Pro Pro Val Pro Leu

 450 455 460
 Thr Pro Asp Ile Phe Trp Gly Cys Phe Leu Glu Asn Glu Thr Leu Trp
 465 470 475 480
 Ala Glu Arg Leu Cys Gly Glu Ala Ser Leu Gln Ala Val Pro Pro Ser
 485 490 495
 Asn Gln Ala Trp Val Gln His Val Cys Gln Gly Pro Thr Pro Asp Val
 500 505 510
 Thr Ala Ser Pro Pro Cys His Ile Gly Pro Cys Gly Glu Arg Cys Pro

 515 520 525
 Asp Gly Gly Ser Phe Leu Val Met Val Cys Ala Asn Asp Thr Met Tyr
 530 535 540
 Glu Val Leu Val Pro Phe Trp Pro Trp Leu Ala Gly Gln Cys Arg Ile
 545 550 555 560
 Ser Arg Gly Gly Asn Asp Thr Cys Phe Leu Glu Gly Leu Leu Gly Pro
 565 570 575
 Leu Leu Pro Ser Leu Pro Pro Leu Gly Pro Ser Pro Leu Cys Leu Thr

 580 585 590
 Pro Gly Pro Phe Leu Leu Gly Met Leu Ser Gln Leu Pro Arg Cys Gln
 595 600 605
 Ser Ser Val Pro Ala Leu Ala His Pro Thr Arg Leu His Tyr Leu Leu
 610 615 620

Arg Leu Leu Thr Phe Leu Leu Gly Pro Gly Ala Gly Gly Ala Glu Ala
 625 630 635 640
 Gln Gly Met Leu Gly Arg Ala Leu Leu Leu Ser Ser Leu Pro Asp Asn

 645 650 655
 Cys Ser Phe Trp Asp Ala Phe Arg Pro Glu Gly Arg Arg Ser Val Leu
 660 665 670
 Arg Thr Ile Gly Glu Tyr Leu Glu Gln Asp Glu Glu Gln Pro Thr Pro
 675 680 685
 Ser Gly Phe Glu Pro Thr Val Asn Pro Ser Ser Gly Ile Ser Lys Met
 690 695 700
 Glu Leu Leu Ala Cys Phe Ser Pro Val Leu Trp Asp Leu Leu Gln Arg

 705 710 715 720
 Glu Lys Ser Val Trp Ala Leu Gln Ile Leu Val Gln Ala Tyr Leu His
 725 730 735
 Met Pro Pro Glu Asn Leu Gln Gln Leu Val Leu Ser Ala Glu Arg Glu
 740 745 750
 Ala Ala Gln Gly Phe Leu Thr Leu Met Leu Gln Gly Lys Leu Gln Gly
 755 760 765
 Lys Leu Gln Val Pro Pro Ser Glu Glu Gln Ala Leu Gly Arg Leu Thr

 770 775 780
 Ala Leu Leu Leu Gln Arg Tyr Pro Arg Leu Thr Ser Gln Leu Phe Ile
 785 790 795 800
 Asp Leu Ser Pro Leu Ile Pro Phe Leu Ala Val Ser Asp Leu Met Arg
 805 810 815
 Phe Pro Pro Ser Leu Leu Ala Asn Asp Ser Val Leu Ala Ala Ile Arg
 820 825 830
 Asp Tyr Ser Pro Gly Met Arg Pro Glu Gln Lys Glu Ala Leu Ala Lys

 835 840 845
 Arg Leu Leu Ala Pro Glu Leu Phe Gly Glu Val Pro Ala Trp Pro Gln
 850 855 860
 Glu Leu Leu Trp Ala Val Leu Pro Leu Leu Pro His Leu Pro Leu Glu

865 870 875 880
 Asn Phe Leu Gln Leu Ser Pro His Gln Ile Gln Ala Leu Glu Asp Ser
 885 890 895
 Trp Pro Ala Ala Gly Leu Gly Pro Gly His Ala Arg His Val Leu Arg

 900 905 910
 Ser Leu Val Asn Gln Ser Val Gln Asp Gly Glu Glu Gln Val Arg Arg
 915 920 925
 Leu Gly Pro Leu Ala Cys Phe Leu Ser Pro Glu Glu Leu Gln Ser Leu
 930 935 940
 Val Pro Leu Ser Asp Pro Thr Gly Pro Val Glu Arg Gly Leu Leu Glu
 945 950 955 960
 Cys Ala Ala Asn Gly Thr Leu Ser Pro Glu Gly Arg Val Ala Tyr Glu

 965 970 975
 Leu Leu Gly Val Leu Arg Ser Ser Gly Gly Ala Val Leu Ser Pro Arg
 980 985 990
 Glu Leu Arg Val Trp Ala Pro Leu Phe Ser Gln Leu Gly Leu Arg Phe
 995 1000 1005
 Leu Gln Glu Leu Ser Glu Pro Gln Leu Arg Ala Met Leu Pro Val
 1010 1015 1020
 Leu Gln Gly Thr Ser Val Thr Pro Ala Gln Ala Val Leu Leu Leu

 1025 1030 1035
 Gly Arg Leu Leu Pro Arg His Asp Leu Ser Leu Glu Glu Leu Cys
 1040 1045 1050
 Ser Leu His Leu Leu Leu Pro Gly Leu Ser Pro Gln Thr Leu Gln
 1055 1060 1065
 Ala Ile Pro Arg Arg Val Leu Val Gly Ala Cys Ser Cys Leu Ala
 1070 1075 1080
 Pro Glu Leu Ser Arg Leu Ser Ala Cys Gln Thr Ala Ala Leu Leu

 1085 1090 1095
 Gln Thr Phe Arg Val Lys Asp Gly Val Lys Asn Met Gly Thr Thr
 1100 1105 1110

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

1340	1345	1350
Gly Phe Cys Leu Gly Glu Thr	Phe Ala Thr Glu Leu	Gly Trp Leu
1355	1360	1365
Leu Leu Gln Glu Ser Val Leu	Gly Lys Pro Glu Leu	Trp Ser Gln
1370	1375	1380
Asp Glu Val Glu Gln Ala Gly	Arg Leu Val Phe Thr	Leu Ser Thr

1385	1390	1395
Glu Ala Ile Ser Leu Ile Pro	Arg Glu Ala Leu Gly	Pro Glu Thr
1400	1405	1410
Leu Glu Arg Leu Leu Glu Lys	Gln Gln Ser Trp Glu	Gln Ser Arg
1415	1420	1425
Val Gly Gln Leu Cys Arg Glu	Pro Gln Leu Ala Ala	Lys Lys Ala
1430	1435	1440
Ala Leu Val Ala Gly Val Val	Arg Pro Ala Ala Glu	Asp Leu Pro

1445	1450	1455
Glu Pro Val Pro Asn Cys Ala	Asp Val Arg Gly Thr	Phe Pro Ala
1460	1465	1470
Ala Trp Ser Ala Thr Gln Ile	Ala Glu Met Glu Leu	Ser Asp Phe
1475	1480	1485
Glu Asp Cys Leu Thr Leu Phe	Ala Gly Asp Pro Gly	Leu Gly Pro
1490	1495	1500
Glu Glu Leu Arg Ala Ala Met	Gly Lys Ala Lys Gln	Leu Trp Gly

1505	1510	1515
Pro Pro Arg Gly Phe Arg Pro	Glu Gln Ile Leu Gln	Leu Gly Arg
1520	1525	1530
Leu Leu Ile Gly Leu Gly Asp	Arg Glu Leu Gln Glu	Leu Ile Leu
1535	1540	1545
Val Asp Trp Gly Val Leu Ser	Thr Leu Gly Gln Ile	Asp Gly Trp
1550	1555	1560
Ser Thr Thr Gln Leu Arg Ile	Val Val Ser Ser Phe	Leu Arg Gln

1565	1570	1575
------	------	------

Ser Gly Arg His Val Ser His Leu Asp Phe Val His Leu Thr Ala
1580 1585 1590

Leu Gly Tyr Thr Leu Cys Gly Leu Arg Pro Glu Glu Leu Gln His
1595 1600 1605

Ile Ser Ser Trp Glu Phe Ser Gln Ala Ala Leu Phe Leu Gly Thr
1610 1615 1620

Leu His Leu Gln Cys Ser Glu Glu Gln Leu Glu Val Leu Ala His
1625 1630 1635

Leu Leu Val Leu Pro Gly Gly Phe Gly Pro Ile Ser Asn Trp Gly
1640 1645 1650

Pro Glu Ile Phe Thr Glu Ile Gly Thr Ile Ala Ala Gly Ile Pro
1655 1660 1665

Asp Leu Ala Leu Ser Ala Leu Leu Arg Gly Gln Ile Gln Gly Val
1670 1675 1680

Thr Pro Leu Ala Ile Ser Val Ile Pro Pro Pro Lys Phe Ala Val
1685 1690 1695

Val Phe Ser Pro Ile Gln Leu Ser Ser Leu Thr Ser Ala Gln Ala
1700 1705 1710

Val Ala Val Thr Pro Glu Gln Met Ala Phe Leu Ser Pro Glu Gln
1715 1720 1725

Arg Arg Ala Val Ala Trp Ala Gln His Glu Gly Lys Glu Ser Pro
1730 1735 1740

Glu Gln Gln Gly Arg Ser Thr Ala Trp Gly Leu Gln Asp Trp Ser
1745 1750 1755

Arg Pro Ser Trp Ser Leu Val Leu Thr Ile Ser Phe Leu Gly His
1760 1765 1770

Leu Leu
1775

<210> 26

<211> 1809

<212> PRT

<213> Mus musculus

<400> 26

Met Ala Leu Ser Leu Gln Pro Gln Leu Leu Leu Leu Ser Leu Leu

1 5 10 15

Pro Gln Glu Val Thr Ser Ala Pro Thr Gly Pro Gln Ser Leu Asp Ala

20 25 30

Gly Leu Ser Leu Leu Lys Ser Phe Val Ala Thr Leu Asp Gln Ala Pro

35 40 45

Gln Arg Ser Leu Ser Gln Ser Arg Phe Ser Ala Phe Leu Ala Asn Ile

50 55 60

Ser Ser Ser Phe Gln Leu Gly Arg Met Gly Glu Gly Pro Val Gly Glu

65 70 75 80

Pro Pro Pro Leu Gln Pro Pro Ala Leu Arg Leu His Asp Phe Leu Val

85 90 95

Thr Leu Arg Gly Ser Pro Asp Trp Glu Pro Met Leu Gly Leu Leu Gly

100 105 110

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

115 120 125

Val His Gln Ala Gly Val Leu Gly Gly Leu Val Glu Ala Leu Leu Gly

130 135 140

Ala Leu Val Pro Gly Gly Pro Pro Ala Pro Thr Arg Pro Pro Cys Thr

145 150 155 160

Arg Asp Gly Pro Ser Asp Cys Val Leu Ala Ala Asp Trp Leu Pro Ser

165 170 175

Leu Met Leu Leu Leu Glu Gly Thr Arg Trp Gln Ala Leu Val Gln Leu

180 185 190

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

195 200 205

Pro Ala Pro His Phe Leu Gln Gly Leu Leu Gly Leu Leu Thr Pro Ala

210 215 220

Gly Glu Leu Gly Ser Glu Glu Ala Leu Trp Gly Gly Leu Leu Arg Thr

225 230 235 240

Val Gly Ala Pro Leu Tyr Ala Ala Phe Gln Glu Gly Leu Leu Arg Val
245 250 255

Thr His Ser Leu Gln Asp Glu Val Phe Ser Ile Met Gly Gln Pro Glu
260 265 270

Pro Asp Ala Ser Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln Leu Leu
275 280 285

Leu Trp Gly Met Arg Asn Asn Leu Ser Trp Asp Ala Arg Ala Leu Gly
290 295 300

Phe Leu Ser Gly Ser Pro Pro Pro Pro Pro Ala Leu Leu His Cys Leu
305 310 315 320

Ser Arg Gly Val Pro Leu Pro Arg Ala Ser Gln Pro Ala Ala His Ile
325 330 335

Ser Pro Arg Gln Arg Arg Ala Ile Ser Val Glu Ala Leu Cys Glu Asn
340 345 350

His Ser Gly Pro Glu Pro Pro Tyr Ser Ile Ser Asn Phe Ser Ile Tyr
355 360 365

Leu Leu Cys Gln His Ile Lys Pro Ala Thr Pro Arg Pro Pro Pro Thr
370 375 380

Thr Pro Arg Pro Pro Pro Thr Thr Pro Gln Pro Pro Pro Thr Thr Thr
385 390 395 400

Gln Pro Ile Pro Asp Thr Thr Gln Pro Pro Pro Val Thr Pro Arg Pro
405 410 415

Pro Pro Thr Thr Pro Gln Pro Pro Pro Ser Thr Ala Val Ile Cys Gln
420 425 430

Thr Ala Val Trp Tyr Ala Val Ser Trp Ala Pro Gly Ala Arg Gly Trp
435 440 445

Leu Gln Ala Cys His Asp Gln Phe Pro Asp Gln Phe Leu Asp Met Ile
450 455 460

Cys Gly Asn Leu Ser Phe Ser Ala Leu Ser Gly Pro Ser Arg Pro Leu
465 470 475 480

Val Lys Gln Leu Cys Ala Gly Leu Leu Pro Pro Pro Thr Ser Cys Pro

485 490 495
 Pro Gly Leu Ile Pro Val Pro Leu Thr Pro Glu Ile Phe Trp Gly Cys
 500 505 510
 Phe Leu Glu Asn Glu Thr Leu Trp Ala Glu Arg Leu Cys Val Glu Asp
 515 520 525
 Ser Leu Gln Ala Val Pro Pro Arg Asn Gln Ala Trp Val Gln His Val
 530 535 540

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

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

 Pro Gly Thr Gly Gly Ala Glu Thr Gln Gly Met Leu Gly Gln Ala Leu
 675 680 685
 Leu Leu Ser Ser Leu Pro Asp Asn Cys Ser Phe Trp Asp Ala Phe Arg
 690 695 700
 Pro Glu Gly Arg Arg Ser Val Leu Arg Thr Val Gly Glu Tyr Leu Gln
 705 710 715 720
 Arg Glu Glu Pro Thr Pro Pro Gly Leu Asp Ser Ser Leu Ser Leu Gly
 725 730 735

Ser Gly Met Ser Lys Met Glu Leu Leu Ser Cys Phe Ser Pro Val Leu
740 745 750

Trp Asp Leu Leu Gln Arg Glu Lys Ser Val Trp Ala Leu Arg Thr Leu
755 760 765

Val Lys Ala Tyr Leu Arg Met Pro Pro Glu Asp Leu Gln Gln Leu Val
770 775 780

Leu Ser Ala Glu Met Glu Ala Ala Gln Gly Phe Leu Thr Leu Met Leu
785 790 795 800

Arg Ser Trp Ala Lys Leu Lys Val Gln Pro Ser Glu Glu Gln Ala Met
805 810 815

Gly Arg Leu Thr Ala Leu Leu Leu Gln Arg Tyr Pro Arg Leu Thr Ser
820 825 830

Gln Leu Phe Ile Asp Met Ser Pro Leu Ile Pro Phe Leu Ala Val Pro
835 840 845

Asp Leu Met Arg Phe Pro Pro Ser Leu Leu Ala Asn Asp Ser Val Leu
850 855 860

Ala Ala Ile Arg Asp His Ser Ser Gly Met Lys Pro Glu Gln Lys Glu
865 870 875 880

Ala Leu Ala Lys Arg Leu Leu Ala Pro Glu Leu Phe Gly Glu Val Pro
885 890 895

Asp Trp Pro Gln Glu Leu Leu Trp Ala Ala Leu Pro Leu Leu Pro His
900 905 910

Leu Pro Leu Glu Ser Phe Leu Gln Leu Ser Pro His Gln Ile Gln Ala
915 920 925

Leu Glu Asp Ser Trp Pro Val Ala Asp Leu Gly Pro Gly His Ala Arg
930 935 940

His Val Leu Arg Ser Leu Val Asn Gln Ser Met Glu Asp Gly Glu Glu
945 950 955 960

Gln Val Leu Arg Leu Gly Ser Leu Ala Cys Phe Leu Ser Pro Glu Glu
965 970 975

Leu Gln Ser Leu Val Pro Leu Ser Asp Pro Met Gly Pro Val Glu Gln

980	985	990
Gly Leu Leu Glu Cys Ala Ala Asn	Gly Thr Leu Ser Pro	Glu Gly Arg
995	1000	1005
Val Ala Tyr Glu Leu Leu Gly	Val Leu Arg Ser Ser	Gly Gly Thr
1010	1015	1020
Val Leu Ser Pro Arg Glu Leu	Arg Val Trp Ala Pro	Leu Phe Pro
1025	1030	1035
Gln Leu Gly Leu Arg Phe Leu	Gln Glu Leu Ser Glu	Thr Gln Leu
1040	1045	1050
Arg Ala Met Leu Pro Ala Leu	Gln Gly Ala Ser Val	Thr Pro Ala
1055	1060	1065
Gln Ala Val Leu Leu Phe Gly	Arg Leu Leu Pro Lys	His Asp Leu
1070	1075	1080
Ser Leu Glu Glu Leu Cys Ser	Leu His Pro Leu Leu	Pro Gly Leu
1085	1090	1095
Ser Pro Gln Thr Leu Gln Ala	Ile Pro Lys Arg Val	Leu Val Gly
1100	1105	1110
Ala Cys Ser Cys Leu Gly Pro	Glu Leu Ser Arg Leu	Ser Ala Cys
1115	1120	1125
Gln Ile Ala Ala Leu Leu Gln	Thr Phe Arg Val Lys	Asp Gly Val
1130	1135	1140
Lys Asn Met Gly Ala Ala Gly	Ala Gly Ser Ala Val	Cys Ile Pro
1145	1150	1155
Gly Gln Pro Thr Thr Trp Pro	Asp Cys Leu Leu Pro	Leu Leu Pro
1160	1165	1170
Leu Lys Leu Leu Gln Leu Asp	Ala Ala Ala Leu Leu	Ala Asn Arg
1175	1180	1185
Arg Leu Tyr Arg Gln Leu Pro	Trp Ser Glu Gln Gln	Ala Gln Phe
1190	1195	1200
Leu Trp Lys Lys Met Gln Val	Pro Thr Asn Leu Ser	Leu Arg Asn
1205	1210	1215

Leu Gln	Ala Leu Gly Asn Leu	Ala Gly Gly Met Thr	Cys Glu Phe
1220	1225	1230	
Leu Gln	Gln Ile Ser Ser Met	Val Asp Phe Leu Asp	Val Val His
1235	1240	1245	
Met Leu	Tyr Gln Leu Pro Thr	Gly Val Arg Glu Ser	Leu Arg Ala
1250	1255	1260	
Cys Ile	Trp Thr Glu Leu Gln	Arg Arg Met Thr Met	Pro Glu Pro
1265	1270	1275	
Glu Leu	Thr Thr Leu Gly Pro	Glu Leu Ser Glu Leu	Asp Thr Lys
1280	1285	1290	
Leu Leu	Leu Asp Leu Pro Ile	Gln Leu Met Asp Arg	Leu Ser Asn
1295	1300	1305	
Asp Ser	Ile Met Leu Val Val	Glu Met Val Gln Gly	Ala Pro Glu
1310	1315	1320	
Gln Leu	Leu Ala Leu Thr Pro	Leu His Gln Thr Ala	Leu Ala Glu
1325	1330	1335	
Arg Ala	Leu Lys Asn Leu Ala	Pro Lys Glu Thr Pro	Ile Ser Lys
1340	1345	1350	
Glu Val	Leu Glu Thr Leu Gly	Pro Leu Val Gly Phe	Leu Gly Ile
1355	1360	1365	
Glu Ser	Thr Arg Arg Ile Pro	Leu Pro Ile Leu Leu	Ser His Leu
1370	1375	1380	
Ser Gln	Leu Gln Gly Phe Cys	Leu Gly Glu Thr Phe	Ala Thr Glu
1385	1390	1395	
Leu Gly	Trp Leu Leu Leu Gln	Glu Pro Val Leu Gly	Lys Pro Glu
1400	1405	1410	
Leu Trp	Ser Gln Asp Glu Ile	Glu Gln Ala Gly Arg	Leu Val Phe
1415	1420	1425	
Thr Leu	Ser Ala Glu Ala Ile	Ser Ser Ile Pro Arg	Glu Ala Leu
1430	1435	1440	
Gly Pro	Glu Thr Leu Glu Arg	Leu Leu Gly Lys His	Gln Ser Trp

1445	1450	1455
Glu Gln Ser Arg Val Gly His	Leu Cys Gly Glu Ser	Gln Leu Ala
1460	1465	1470
His Lys Lys Ala Ala Leu Val	Ala Gly Ile Val His	Pro Ala Ala
1475	1480	1485
Glu Gly Leu Gln Glu Pro Val	Pro Asn Cys Ala Asp	Ile Arg Gly
1490	1495	1500
Thr Phe Pro Ala Ala Trp Ser	Ala Thr Gln Ile Ser	Glu Met Glu
1505	1510	1515
Leu Ser Asp Phe Glu Asp Cys	Leu Ser Leu Phe Ala	Gly Asp Pro
1520	1525	1530
Gly Leu Gly Pro Glu Glu Leu	Arg Ala Ala Met Gly	Lys Ala Lys
1535	1540	1545
Gln Leu Trp Gly Pro Pro Arg	Gly Phe Arg Pro Glu	Gln Ile Leu
1550	1555	1560
Gln Leu Gly Arg Leu Leu Ile	Gly Leu Gly Glu Arg	Glu Leu Gln
1565	1570	1575
Glu Leu Thr Leu Val Asp Trp	Gly Val Leu Ser Ser	Leu Gly Gln
1580	1585	1590
Ile Asp Gly Trp Ser Ser Met	Gln Leu Arg Ala Val	Val Ser Ser
1595	1600	1605
Phe Leu Arg Gln Ser Gly Arg	His Val Ser His Leu	Asp Phe Ile
1610	1615	1620
Tyr Leu Thr Ala Leu Gly Tyr	Thr Val Cys Gly Leu	Arg Pro Glu
1625	1630	1635
Glu Leu Gln His Ile Ser Ser	Trp Glu Phe Ser Gln	Ala Ala Leu
1640	1645	1650
Phe Leu Gly Ser Leu His Leu	Pro Cys Ser Glu Glu	Gln Leu Glu
1655	1660	1665
Val Leu Ala Tyr Leu Leu Val	Leu Pro Gly Gly Phe	Gly Pro Val
1670	1675	1680

Ser Asn Trp Gly Pro Glu Ile Phe Thr Glu Ile Gly Thr Ile Ala
1685 1690 1695

Ala Gly Ile Pro Asp Leu Ala Leu Ser Ala Leu Leu Arg Gly Gln
1700 1705 1710

Ile Gln Gly Leu Thr Pro Leu Ala Ile Ser Val Ile Pro Ala Pro
1715 1720 1725

Lys Phe Ala Val Val Phe Asn Pro Ile Gln Leu Ser Ser Leu Thr
1730 1735 1740

Arg Gly Gln Ala Val Ala Val Thr Pro Glu Gln Leu Ala Tyr Leu
1745 1750 1755

Ser Pro Glu Gln Arg Arg Ala Val Ala Trp Ala Gln His Glu Gly
1760 1765 1770

Lys Glu Ile Pro Glu Gln Leu Gly Arg Asn Ser Ala Trp Gly Leu
1775 1780 1785

Tyr Asp Trp Phe Gln Ala Ser Trp Ala Leu Ala Leu Pro Val Ser
1790 1795 1800

Ile Phe Gly His Leu Leu
1805

<210> 27

<211> 10

<212> PRT

<213> Artificial Sequence

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

<400> 27

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu

1 5 10

<210> 28

<211> 8543

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 28

ccaatgatac gcgtcgggtgc gggcctcttc gctattacgc cagctggcga aagggggatg	60
tgctgcaagg cgattaagtt gggtaacgcc aggggtttcc cagtcacgac gttgtaaaac	120
gacggccagt gagcgcgcgt aatacagctc actatagggc gaattgggta catcgacggt	180
atcgggggag ctgcaggggt ctccattttg aagcgggagg tttgaacgcg cagccgccat	240
gccgggggtt tacgagattg tgattaaaggt cccacgcgac cttgacgagc atctgcccg	300
catttctgac agctttgtga actgggtggc cgagaaggaa tgggagttgc cgccagattc	360
tgacatggat ctgaatctga ttgagcaggc acccctgacc gtggccgaga agctgcagcg	420
cgactttctg acggaatggc gccgtgtgag taaggcccg gaggtcttt tctttgtgca	480
atttgagaag ggagagagct acttcacat gcacgtgctc gtggaaacca cgggggtgaa	540
atccatggtt ttgggacgtt tcctgagtca gattcgcgaa aaactgattc agagaattta	600
ccgcgggatac gagccgactt tgccaaactg gttcgcggtc acaaagacca gaaatggcgc	660
cggaggcggg aacaaggtgg tggatgagtg ctacatcccc aattacttgc tccccaaac	720
ccagcctgag ctccagtggg cgtggactaa tatggaacag tatttaagcg cctgtttgaa	780
tctcacggag cgtaaacggt tggtagcgca gcatctgacg cacgtgtcgc agacgcagga	840
gcagaacaaa gagaatcaga atcccaattc tgatgcgccg gtgatcagat caaaaacttc	900
agccaggtag atggagctgg tcgggtggct cgtggacaag gggattacct cggagaagca	960
gtggatccag gaggaccagg cctcatacat ctcttcaat gcggcctcca actcgcggtc	1020
ccaaatcaag gctgccttgg acaatgcggg aaagattatg agcctgacta aaaccgcccc	1080
cgactacctg gtgggccagc agcccgtgga ggacatttcc agcaatcgga tttataaaat	1140
tttggaacta aacgggtacg atccccaata tgcggcttcc gtctttcttg gatgggccac	1200
gaaaaagtgc ggcaagagga acaccatctg gctgtttggg cctgcaacta cggggaagac	1260
caacatcgcg gaggccatag cccacactgt gcccttctac ggggtcgtaa actggaccaa	1320
tgagaacttt cccttcaacg actgtgtgga caagatggtg atctgggtggg aggaggggaa	1380
gatgaccgcc aaggtcgtgg agtcggccaa agccattctc ggaggaagca aggtgcgcgt	1440
ggaccagaaa tgcaagtctt cggcccagat agaccgact cccgtgatcg tcacctcaa	1500
caccaatatg tgcgccgtga ttgacgggaa ctcaacgacc ttcgaacacc agcagccgtt	1560
gcaagaccgg atgttcaat ttgaactcac ccgccgtctg gatcatgact ttgggaaggt	1620
caccaagcag gaagtcaaag actttttccg gtgggcaaag gatcacgtgg ttgaggtgga	1680
gcatgaattc tacgtcaaaa aggggtggagc caagaaaaga cccgccccca gtgacgcaga	1740

tataagtgag cccaaacggg tgcgcgagtc agttgcgcag ccatcgacgt cagacgcgga	1800
agcttcgacg aactacgcgg acaggtacca aaacaaatgt tctcgtcacg tgggcatgaa	1860
tctgatgctg tttccctgca gacaatgcga gagactgaat cagaattcaa atatctgctt	1920
cactcacggt gtcaaagact gtttagagtg ctttcccggtg tcagaatctc aaccctttc	1980
tgctgtcaaa aaggcgtatc agaaactgtg ctacattcat cacatcatgg gaaaggtgcc	2040
agacgcttgc actgcttgcg acctgggtcaa tgtggacttg gatgactgtg tttctgaaca	2100
ataaatgact taaaccaggt atgagtcggc tggataaatc taaagtcata aacggcgctc	2160
tggaattact caatgaagtc ggtatcgaag gcctgacgac aaggaaactc gctcaaaagc	2220
tgggagttag gcagcctacc ctgtactggc acgtgaagaa caagcgggcc ctgctcgatg	2280
ccctggccat cgagatgctg gacaggcatc ataccactt ctgccccctg gaaggcgagt	2340
catggcaaga ctttctgcgg aacaacgcca agtcattccg ctgtgctctc ctctcacatc	2400
gcgacggggc taaagtgcac ctgcgcaccc gcccaacaga gaaacagtac gaaacctgg	2460
aaaatcagct cgcgttcctg tgctagcaag gcttctccct ggagaacgca ctgtacgctc	2520
tgcccgccgt gggccacttt acactgggct gcgtattgga ggaacaggag catcaagtag	2580
caaaagagga aagagagaca cctaccaccg attctatgcc cccacttctg agacaagcaa	2640
ttgagctgtt cgaccggcag ggagccgaac ctgccttctt tttcgccctg gaactaatca	2700
tatgtggcct ggagaaacag ctaaagtgcg aaagcggcgg gccggccgac gcccttgacg	2760
atattgactt agacatgctc ccagccgatg cccttgacga ctttgacctt gatatgctgc	2820
ctgctgacgc tcttgacgat tttgaccttg acatgctccc cgggtaaatg catgaattcg	2880
atctagaggg ccctattcta tagtgtcacc taaatgctag agctcgctga tcagcctcga	2940
ctgtgccttc tagttgccag ccatctgttg tttgccccct ccccgctcct tccttgacct	3000
tggaagggtg cactcccatc gtcctttcct aataaaatga ggaaattgca tcgcattgtc	3060
tgagtagggt tcattctatt ctgggggggtg ggggtggggca ggacagcaag ggggaggatt	3120
gggaagacaa tagcaggcat gctggggatg cgggtgggctc tatggcttct gaggcggaaa	3180
gaaccagctg gggctcgaat caagetatca agtgccacct gacgtctccc tatcagtgat	3240
agagaagtcg acacgtctcg agctccctat cagtgataga gaaggtacgt ctagaacgtc	3300
tcctatcag tgatagagaa gtcgacacgt ctcgagctcc ctatcagtga tagagaaggt	3360
acgtctagaa cgctcccta tcagtgatag agaagtcgac acgtctcgag ctccctatca	3420
gtgatagaga aggtacgtct agaactgtct cctatcagtg atagagaagt cgacacgtct	3480
cgagctccct atcagtgata gagaaggtag ccctatata agcagagaga tctgttcaaa	3540
tttgaactga ctaagcggct cccgccgat tttggcaaga ttactaagca ggaagtcaag	3600

gacttttttg cttgggcaaa ggtcaatcag gtgccggtga ctcacgagtt taaagttccc	3660
agggaattgg cggaactaa aggggcggag aaatctctaa aacgccact gggtgacgtc	3720
accaatacta gctataaaag tctggagaag cgggccaggc tctcatttgt tcccagagacg	3780
cctcgagatt cagacgtgac tgttgatccc gctcctctgc gaccgctagc ttcgatcaac	3840
tacgcagaca ggtacaaaa caagtgttct cgtcacgtgg gcattaatct gattctgttt	3900
ccctgcagac aatgcgagag aatgaatcag aactcaaata tctgcttcac tcacggacag	3960
aaagactgtt tagagtgtt tcccggtgca gaatctcaac ccgtttctgt cgtcaaaaag	4020
gcgtatcaga aactgtgcta cattcatcat atcatgggaa aggtgccaga cgcttgact	4080
gcctgcgac tggtcaatgt ggatttggat gactgcatct ttgaacaata aatgacttaa	4140
gccaggtatg gctgccgatg gttatcttcc agattggctc gaggacaacc ttagtgaagg	4200
aattcgagag tgggtggctt tgaaacctgg agccctcaa cccaaggcaa atcaacaaca	4260
tcaagacaac gctagagtc ttgtgcttcc gggttacaaa taccttgac cggcaacgg	4320
actcgacaag ggggagccgg tcaacgcagc agacgcggcg gccctcgagc acgacaaagc	4380
ctacgaccag cagctcaagg ccggagacaa cccgtacctc aagtacaacc acgccgacgc	4440
cgagttccag gagcggctca aagaagatac gtcttttggg ggcaacctcg ggcgagcagt	4500
cttcaggcc aaaaagaggc ttcttgaacc tcttggctg gttgaggaag cggctaagac	4560
ggctcctgga aagaagaggc ctgtagagca gtctctcag gaaccggact cctccgcggg	4620
tattggcaaa tcgggtgac agcccgttaa aaagagactc aatttcggtc agactggcga	4680
cacagagtca gtcccagacc ctcaaccaat cggagaacct cccgcagccc cctcaggtgt	4740
gggatctctt acaatggctt cagggtgttg cgcaccagtg gcagacaata acgaaggtgc	4800
cgatggagtg ggtagtctt cgggaaattg gcattgcgat tccaatggc tgggggacag	4860
agtcacacc accagaccc gaacctgggc cctgccacc tacaacaac accctacaa	4920
gcaaatctcc aacagcacat ctggaggatc ttcaaatgac aacgcctact tcggctacag	4980
cacccctgg gggtattttg acttcaacag attccactgc cacttctcac cactgactg	5040
gcagcgactc atcaacaaca actggggatt ccggcctaag cgactcaact tcaagctctt	5100
taacattcag gtcaaagagg ttacggacaa caatggagtc aagaccatcg ccaataacct	5160
taccagcac gtccaggtct tcacggactc agactatcag ctcccgtagc tgcctgggtc	5220
ggctcacgag ggctgcctcc cgccgttccc agcggacgtt ttcatgattc ctcagtacgg	5280
glatctgacg cttaatgatg gaagccaggc cgtgggtcgt tcgtcctttt actgcctgga	5340

atatttcccg tcgcaaatgc taagaacggg taacaacttc cagttcagct acgagtttga	5400
gaacgtacct ttccatagca gctacgtca cagccaaagc ctggaccgac taatgaatcc	5460
actcatcgac caatacttgt actatctctc tagaactatt aacggttctg gacagaatca	5520
acaaacgcta aaattcagtg tggccggacc cagcaacatg gctgtccagg gaagaaacta	5580
catactgga cccagctacc gacaacaacg tgtctcaacc actgtgactc aaaacaacaa	5640
cagcgaattt gcttggcctg gagcttcttc ttgggctctc aatggacgta atagcttgat	5700
gaatcctgga cctgctatgg cctctcacia agaaggagag gaccgtttct ttcctttgtc	5760
tggatcttta atttttggca aacaaggtag tggcagagac aacgtggatg cggacaaagt	5820
catgataacc aacgaagaag aaattaaaaac tactaaccg gtagcaacgg agtcctatgg	5880
acaagtggcc acaaacacc agagtgccca aactttggcg gtgccttta aggcacaggc	5940
gcagaccggt tgggttcaaa accaaggaat acttccgggt atggtttggc aggacagaga	6000
tgtgtacctg caaggacca tttgggcaa aattcctcac acggacggca actttcacc	6060
ttctccgctg atgggagggt ttggaatgaa gcaccgcct cctcagatcc tcatcaaaaa	6120
cacacctgta cctgcggatc ctccaacggc ctcaacaag gacaagctga actctttcat	6180
caccagtat tctactggtc aagtcagcgt ggagatcgag tgggagctgc agaaggaaaa	6240
cagcaagcgc tggaaccgg agatccagta cacttccaac tattacaagt ctaataatgt	6300
tgaatttgct gttactactg aaggtgtata tagtgaacce cgccccattg gcaccagata	6360
cctgactcgt aatctgtaag tgcacttgct tgttaatcaa taaaccgttt aattcgtttc	6420
agttgaactt tggctctctc gaagggaat tcgttttaac ctgcaggact agaggctctg	6480
tattagaggt cactgagtg ttttgcgaca ttttgcgaca ccatgtggtc acgctgggta	6540
tttaagcccc agtgagcacg cagggtctcc attttgaagc gggaggtttg aacgcgcagc	6600
cgccaagccg aattctgcag atatcacatg tcttaggaac tatcgatcca tcacactggc	6660
ggccgctcga ctagagcggc cgccaccgcg gtggagctcc agcttttgcg gaccgaatcg	6720
gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc	6780
tggcgttttt ccatagctc cgccccctg acgagcatca caaaaatcga cgctcaagtc	6840
agagggtggcg aaaccgaca ggactataaa gataccaggc gtttccccct ggaagctccc	6900
tcgtgcgtc tctgttccg accctgccgc ttaccgata cctgtccgcc tttctccctt	6960
cgggaaagct ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg	7020
ttcgtccaa gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat	7080
ccggtaaacta tcgtcttgag tccaaccggg taagacacga cttatcgcca ctggcagcag	7140
ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt	7200

ggtaggcctaa ctacggctac actagaagaa cagtatttgg tatctgcgct ctgctgaagc 7260
 cagttacctt cggaataaga gttgtagct cttgatccgg caaacaacc accgctggtta 7320
 gcggtggttt ttttgttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag 7380
 atcctttgat cttttctacg gggctctgacg ctacgtggaa cgaaaactca cgtaaggga 7440

 ttttggcat gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa 7500
 gttttaaatc aatctaaagt atatatgagt aaacttggc tgacagttac caatgcttaa 7560
 tcagttaggc acctatctca gcatctgtc tatttcgttc atccatagtt gcctgactcc 7620
 ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga 7680
 taccgcgaga cccacgtca ccggctccag atttatcagc aataaaccag ccagccggaa 7740
 gggccgagcg cagaagtggc cctgcaactt tatccgctc catccagtct attaattgtt 7800
 gccgggaagc tagagtaagt agttcgccag ttaatatgtt gcgcaacgtt gttgccattg 7860

 ctacaggcat cgtgggtgca cgctcgtcgt ttggtatggc ttatttcagc tccggttccc 7920
 aacgatcaag gcgagttaca tgatcccca tgttgtgcaa aaaagcgggt agctccttcg 7980
 gtctccgat cgttgtcaga agtaagtgg ccgcagtgtt atcactcatg gttatggcag 8040
 cactgcataa ttctcttact gtcatgcat ccgtaagatg cttttctgtg actggtgagt 8100
 actcaacca gtcattctga gaatagtga tgcggcgacc gatttgcctc tgcccggtcgt 8160
 caatacggga taataccgag ccacatagca gaactttaaa agtgcctatc attggaaaac 8220
 gttcttcggg gcgaaaactc tcaaggatct taccgtgtt gagatccagt tcgatgtaac 8280

 ccaactcgtc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag 8340
 caaaaacagg aaggcaaat gccgcaaaa aggaataag ggcgacacgg aatgttgaa 8400
 tactcatact ctctctttt caatattatt gaagcattha tcagggttat tgtctcatga 8460
 gcggatacat atttgaatgt atttagaaaa ataacaaat aggggttcg cgcacatttc 8520
 cccgaaaagt gccacctgac gtc 8543

 <210> 29
 <211> 5262
 <212> DNA
 <213> Homo sapiens
 <400> 29

 gtgactctgg cccctactgg gcctcattcc ctggaccctg gtctctcctt cctgaagtca 60

 ttgctctcca ctctggacca ggctccccag ggtccctga gccgtcagc gttctttaca 120
 ttcctggcca acattttctt ttcctttgag cctgggagaa tgggggaagg accagtagga 180

gagccccac ctctccagcc gcctgctctg cggtccatg attttctagt gacactgaga	240
ggtagcccc actgggagcc aatgctaggg ctgctagggg atatgctggc actgctggga	300
caggagcaga ctccccgaga tttcctggtg caccaggcag gggtgctggg tggacttgtg	360
gaggtgctgc tgggagcctt agttcctggg gggcccccta cccaactcg gccccatgc	420
accctgatg ggcctctga ctgtgtcctg gctgctgact ggttgccttc tctgctgctg	480
ttgttagagg gcacacgctg gcaagctctg gtgcaggtgc agccagtggt ggaccccacc	540
aatgccacag gcctcgatgg gagggaggca gctcctcact ttttgcaggg tctgttgggt	600
ttgcttacc caacaggga gctaggctcc aaggaggctc tttggggcgg tctgctacgc	660
acagtggggg cccccctta tgcctccttt caggaggggc tgctccgtgt cactcactcc	720
ctgcaggatg aggtcttctc cattttgggg cagccagagc ctgataccaa tgggcagtgc	780
cagggaggtg accttcaaca gctgctctta tggggcgtcc ggcaaacct ttctgggat	840
gtccaggcgc tgggctttct gtctggatca ccacccccac cccctgcct ctttactgc	900
ctgagcacgg gcgtgcctct gccagagct tctcagccgt cagcccacat cagcccacgc	960
caacggcgag ccatcactgt ggaggccctc tgtgagaacc acttaggcc agcaccacc	1020
tacagcattt ccaacttctc catccacttg ctctgccagc acaccaagcc tgccactcca	1080
cagccccatc ccagcaccac tgccatctgc cagacagctg tgtggtatgc agtgtcctgg	1140
gcaccaggtg cccaaggctg gctacaggcc tgccacgacc agtttctga tgagttttt	1200
gatgcgatct gcagtaacct ctcttttca gccctgtctg gctccaaccg ccgctgggtg	1260
aagcgctct gtgctggcct gctccaccc cctaccagct gccctgaagg cctgccccct	1320
gttccctca cccagacat ctttggggc tgcttcttgg agaattgagac tctgtgggt	1380
gagcgactgt gtggggaggg aagtctacag gctgtgcccc ccagcaacca ggcttgggtc	1440
cagcatgtgt gccaggcccc caccacagat gtactgcct cccaccatg ccacattgga	1500
ccctgtgggg aacgtgccc ggatgggggc agcttcttgg tgatggtctg tgccaatgac	1560
accatgtatg aggtcctggt gcccttctgg ccttggctag caggccaatg caggataagt	1620
cgtgggggca atgacacttg ctctctagaa gggctgctgg gcccccttct gccctctctg	1680
ccaccactgg gaccatcccc actctgtctg acccctggcc ctttctctct tggcatgcta	1740
tcccagtgc cacgtgtca gtctctgtc ccagctcttg ctacccccac acgctacac	1800
tatctctcc gcctgtgac ctctctcttg ggtccagggg ctgggggcgc tgaggcccag	1860
gggatgctgg gtcgggcct actgtctcc agtctcccag acaactgtc cttctgggat	1920
gcctttcgcc cagagggcc gcgcagtgtg ctacggacga ttggggaata cctggaacaa	1980
gatgaggagc agccaacccc atcaggcttt gaaccactg tcaaccccag ctctgttata	2040

agcaagatgg agctgctggc ctgctttagt cctgtgctgt gggatctgct ccagagggaa 2100
 aagagtgttt gggccctgca gattctagtg caggcgtacc tgcatatgcc cccagaaaac 2160

 ctccagcagc tgggtgcttc agcagagagg gaggtgcac agggcttcct gacactcatg 2220
 ctgcagggga agctgcaggg gaagctgcag gtaccacat cagaggagca ggccctgggt 2280
 cgctgacag cctgctgct ccagcggtac ccacgctca cctcccagct cttcattgac 2340
 ctgtcaccac tcattccctt cttggctgtc tctgacctga tgcgttccc accatccctg 2400
 ttagccaacg acagtgtcct ggctgccatc cgggattaca gccaggaat gaggcctgaa 2460
 cagaaggagg ctctggcaaa gcgactgctg gccctgaac tgtttgggga agtgcctgcc 2520
 tggccccagg agctgctgtg ggcagtgtg cccctgctcc cccacctccc tctggagaac 2580

 tttttgcagc tcagccctca ccagatccag gccttgagg atagctggcc agcagcaggt 2640
 ctggggccag ggcattcccg ccatgtgctg cgcagcctgg taaaccagag tgtccaggat 2700
 ggtgaggagc aggtacgcag gcttgggccc ctgcctgtt tcctgagccc tgaggagctg 2760
 cagagcctag tgccctgag tgatccaacg gggccagtag aacgggggct gctggaatgt 2820
 gcagccaatg ggacctcag cccagaagga cgggtggcat atgaacttct ggggtgtgtg 2880
 cgctcatctg gaggagcgtt gctgagcccc cgggagctgc gggcttgggc cctctcttc 2940
 tctcagctgg gcctccgctt ccttcaggag ctgtcagagc cccagcttag agccatgctt 3000

 cctgtcctgc agggaaactag tgttacacct gctcaggctg tcctgctgct tggacggctc 3060
 cttcttaggc acgatctatc cctggaggaa ctctgtcct tgcaccttct gctaccaggc 3120
 ctcagcccc agacactcca ggccatccct aggcgagtc tggtcggggc ttgttcctgc 3180
 ctggccctg aactgtcacg cctctcagcc tgccagaccg cagcactgct gcagaccttt 3240
 cgggttaaag atggtgttaa aaatatgggt acaacaggtg ctggtccagc tgtgtgtatc 3300
 cctggtcagc ctattccac cacctggcca gactgcctgc tteccctgct cccattaaag 3360
 ctgtacaac tggattcctt ggctcttctg gcaaatcgaa gacgctactg ggagctgccc 3420

 tggctgagc agcaggcaca gtttctctgg aagaagatgc aagtaccac caaccttacc 3480
 ctcaggaatc tgcaggctct gggcaccttg gcaggaggca tgtcctgtga gtttctgcag 3540
 cagatcaact ccatggtaga cttccttgaa gtggtgcaca tgatctatca gctgccact 3600
 agagttagg ggagcctgag ggcctgtatc tgggcagagc tacagcggag gatggcaatg 3660
 ccagaaccag aatggacaac tgtagggcca gaactgaacg ggctggatag caagctactc 3720
 ctggacttac cgatccagtt gatggacaga ctatccaatg aatccattat gttggtggtg 3780

gagctggtgc aaagagctcc agagcagctg ctggcactga cccccctcca ccaggcagcc 3840

ctggcagaga gggcactaca aaacctggct ccaaaggaga ctccagtctc aggggaagtg 3900

ctggagacct taggcccttt gggttgattc ctggggacag agagcacacg acagatcccc 3960

ctacagatcc tgctgtccca tctcagtcag ctgcaaggct tctgcctagg agagacattt 4020

gccacagagc tgggatggct gctattgcag gagtctgttc ttgggaaacc agagtgtgtg 4080

agccaggatg aagtagagca agctggacgc ctagtattca ctctgtctac tgaggcaatt 4140

tccttgatcc ccaggagggc cttgggtcca gagaccctgg agcggcttct agaaaagcag 4200

cagagctggg agcagagcag agttggacag ctgtgtaggg agccacagct tgctgccaag 4260

aaagcagccc tggtagcagg ggtggtgcga ccagctgctg aggatcttcc agaacctgtg 4320

ccaaattgtg cagatgtacg agggacattc ccagcagcct ggtctgcaac ccagattgca 4380

gagatggagc tctcagactt tgaggactgc ctgacattat ttgcaggaga cccaggactt 4440

gggcctgagg aactgcgggc agccatgggc aaagcaaac agttgtgggg tccccccgg 4500

ggatttcgtc ctgagcagat cctgcagctt ggtaggtctt taataggtct aggagatcgg 4560

gaactacagg agctgatcct agtggactgg ggagtgtga gcaccctggg gcagatagat 4620

ggctggagca ccactcagct ccgcattgtg gtctccagtt tcctacggca gagtggtcgg 4680

catgtgagcc acctggactt cgttcatctg acagcgtgg gttatactct ctgtggactg 4740

cggccagagg agctccagca catcagcagt tgggagtca gccaagcagc tctcttctc 4800

ggcaccctgc atctccagtg ctctgaggaa caactggagg ttctggccca cctacttgta 4860

ctgcctggtg gggttggccc aatcagtaac tgggggcctg agatcttcac tgaaattggc 4920

accatagcag ctgggatccc agacctggct ctttcagcac tgctgcgggg acagatccag 4980

ggcgttactc ctcttgccat ttctgtcacc cctctctcta aatttgctgt ggtgtttagt 5040

cccatccaac tatctagtct caccagtgtc caggctgtgg ctgtcactcc tgagcaaatg 5100

gcctttctga gtctgagca gcgacgagca gttgcatggg cccaacatga gggaaaggag 5160

agcccagaac agcaaggtcg aagtacagcc tggggcctcc aggactggtc acgaccttcc 5220

tggtccctgg tattgactat cagcttcctt ggccacctgc ta 5262

<210> 30

<211> 5515

<212> DNA

<213> Homo sapiens

<400> 30

gccctgccct cacctggcta tccacacag gtgagaataa ccagaactca cctccggtac 60

cagtgttcac ttggaacat ggctctcagc ctctggcccc tgctgctgct gctgctgctg	120
ctgctgctgc tgcctttgc agtgactctg gccctactg ggctcattc cctggaccct	180
ggctctctct tectgaagtc attgctctcc actctggacc aggetcccca gggtccctg	240
agccgctcac ggttctttac attcctggcc aacatttctt cttcctttga gcctgggaga	300
atgggggaag gaccagtagg agagcccca cctctccagc cgctgctct gcggtccat	360
gattttctag tgacactgag aggtagcccc gactgggagc caatgctagg gctgctaggg	420
gatatgctgg cactgctggg acaggagcag actccccgag atttcttggt gcaccaggca	480
ggggtgctgg gtggacttgt ggaggtgctg ctgggagcct tagttcttg ggccccct	540
acccaactc ggccccatg caccgtgat ggccgtctg actgtgtct ggctgctgac	600
tggttgctt ctctgctgct gttgttagag ggcacacgt ggcaagctct ggtgcaggtg	660
cagccagtg tggacccca caatgccaca ggctcgtat ggaggagagc agctctcac	720
tttttgagg gtctgttggg ttgtcttacc ccaacagggg agctaggctc caaggaggct	780
ctttggggcg gtctgtacg cacagtgggg gccccctct atgtgctt tcaggagggg	840
ctgtccgtg tcaactact cctgcaggat gaggtcttct ccattttggg gcagccagag	900
cctgatacca atgggcagt ccaggagggt aaccttaac agctgctctt atggggcgtc	960
cggcacaacc tttctggga tgtccaggcg ctgggcttct tgtctggatc accacccca	1020
ccccctgcc tcttactg cctgagcacg ggcgtgcctc tgcccagagc ttctcagccg	1080
tcagcccaca tcagcccag ccaacggcg gccatcactg tggaggccct ctgtgagaac	1140
cacttaggcc cagcaccacc ctacagcatt tccaacttct ccatccatt gctctgccag	1200
cacaccaagc ctgccactcc acagcccat ccagcacca ctgccatctg ccagacagct	1260
gtgtggtatg cagtgtcctg ggcaccaggt gcccaaggct ggctacagc ctgccacgac	1320
cagtttctg atgagtttt ggatcgatc tgcagtaacc tctcttttc agccctgtct	1380
ggctccaacc gccgcctggt gaagcggctc tgtgtggcc tgtcccacc cctaccagc	1440
tgccctgaag gctgcccc tgttcccc accccagaca tctttgggg ctgcttcttg	1500
gagaatgaga ctctgtgggc tgagcgactg tgtggggagg caagtctaca ggctgtgcc	1560
cccagcaacc agcttgggt ccagcatgtg tgccagggcc ccacccaga tgtactgcc	1620
tccccacat gccacattgg acctgtggg gaacgtgcc cggatggggg cagcttctg	1680
gtgatggtct gtgccaatga caccatgtat gaggtcctgg tgcccttctg gccttgcta	1740
gcaggccaat gcaggataag tcgtgggggc aatgacatt gcttctaga agggctgctg	1800

ggcccccttc tgccctctct gccaccactg ggaccatccc cactctgtct gaccctggc	1860
cccttcctcc ttggcatgct atcccagttg ccacgtgtc agtcctctgt cccagctctt	1920
gtcacccca cagcctaca ctatctctc cgctgtga ccttcctctt gggtccagg	1980
gtggggcg ctgaggcca ggggatgtg ggtcggccc tactgtctc cagtctcca	2040
gacaactgct cttctggga tgccttcgc ccagaggcc ggcgagtg gctacggac	2100
attggggaat acctggaaca agatgaggag cagccaacce catcaggctt tgaaccact	2160
gtcaaccca gctctggtat aagcaagat gagctgttg cctgcttag tctgtgtg	2220
tgggatctgc tccagaggga aaagagtgt tggccctgc agattctagt gcaggcgta	2280
ctgcatatgc cccagaaaa cctccagcag ctggtgctt cagcagagag ggaggctgca	2340
cagggttcc tgacactcat gctgcaggg aagctgcagg ggaagctgca ggtaccacca	2400
tccgaggagc agccctggg tgcctgaca gcctgtgc tccagcgta cccacgctc	2460
acctccagc tctcatga cctgtacca ctcatcctt tcttggtgt cctgacctg	2520
atgcgttcc caccatcct gttagccaac gacagtgtc tggctgcat ccgggattac	2580
agcccaggaa tgaggcctga acagaaggag gctctggca agcgactgt ggccctgaa	2640
ctgtttggg aagtgcctgc ctggccccag gagctgtgt gggcagtgct gcccctgctc	2700
ccccacctc cctggagaa cttttgcag ctgagccctc accagatcca ggccctggag	2760
gatagctggc cagcagcagg tctgggcca gggcatgcc gccatgtgt gcgcagcctg	2820
gtaaaccaga gtgtccagga tggtagagg caggtacga ggcttgggc cctcgctgt	2880
ttctgagcc ctgaggagct gcagagccta gtgcccga gtgatccaac ggggccagta	2940
gaacggggc tgctggaatg tgcagccaat gggacctca gccagaagg acgggtggca	3000
tatgaactc tgggtgtgt gcgctcatc ggaggagcgg tgctgagccc ccgggagctg	3060
cgggtctggg cccctctctt ctctcagctg ggcctcgtt tcttcagga gctgtcagag	3120
ccccagctta gagccatgct tctgtcctg cagggaacta gtgttacacc tgctcaggct	3180
gtctgtgc ttggacggt ccttcctagg cagatctat cctggagga actctgtcc	3240
ttgcacctc tgctaccagg cctcagccc cagacctcc aggccatccc taggcgagtc	3300
ctggtcggg cttgttctg cctggccct gaactgtac gcctctcagc ctgccagacc	3360
gcagactgc tgcagacct tcgggttaaa gatggtgtta aaaatatggg tacaacaggt	3420
gtggtccag ctgtgtgtat ccctgtcag cctattcca ccacctggc agactgcctg	3480
cttccctgc tccattaaa gctgtacaa ctggattcct tggctcttct ggcaaatcga	3540

agacgtact gggagctgcc ctggctgag cagcaggcac agtttctctg gaagaagatg	3600
caagtaccca ccaaccttac cctcaggaat ctgcaggctc tgggcaccct ggcaggaggc	3660
atgtcctgtg agtttctgca gcagatcaac tccatggtag acttccttga agtggtgcac	3720
atgatctatc agctgcccac tagagttcga gggagcctga gggcctgtat ctgggcagag	3780
ctacagcggg ggaatggcaat gccagaacca gaatggacaa ctgtagggcc agaactgaac	3840
gggctggata gcaagctact cctggactta ccgatccagt tgatggacag actatccaat	3900
gaatccatta tgttgggtgt ggagctgggt caaagagctc cagagcagct gctggcactg	3960
acccccctcc accaggcagc cctggcagag agggcactac aaaacctggc tccaaaggag	4020
actccagtct cagggaagt gctggagacc ttaggcctt tggttggatt cctggggaca	4080
gagagcacac gacagatccc cctacagatc ctgctgtccc atctcagtca gctgcaaggc	4140
ttctgcttag gagagacatt tgccacagag ctgggatggc tgctattgca ggagtctgtt	4200
cttgggaaac cagagttgtg gagccaggat gaagtagagc aagctggacg cctagtattc	4260
actctgtcta ctgaggcaat ttccttgatc ccaggggagg ccttgggtcc agagaccctg	4320
gagcggcttc tagaaaagca gcagagctgg gagcagagca gatttggaca gctgtgtagg	4380
gagccacagc ttctgcca gaaagcagcc ctggtagcag gggtggtgcg accagctgct	4440
gaggatcttc cagaacctgt gccaaattgt gcagatgtac gagggacatt ccagcagcc	4500
tggctctgca ccagattgc agagatggag ctctcagact ttgaggactg cctgacatta	4560
tttgaggag acccaggact tgggcctgag gaactgcggg cagccatggg caaagcaaaa	4620
cagttgtggg gtccccccg gggatttctg cctgagcaga tctgcagct tggtaggctc	4680
ttaataggtc taggagatcg ggaactacag gagctgatcc tagtggactg gggagtctg	4740
agcacctgg ggcagataga tggctggagc accactcagc tccgattgt ggtctccagt	4800
ttctacggc agagtggctg gcatgtgagc cacctggact tcgttcatct gacagcgtg	4860
ggttatactc tctgtggact gcggccagag gagctccagc acatcagcag ttgggagttc	4920
agccaagcag ctctcttctt cggcaccctg catctccagt gctctgagga acaactggag	4980
gttctggccc acctacttgt actgcctggt gggtttggcc caatcagtaa ctgggggcct	5040
gagatcttca ctgaaattgg caccatagca gctgggatcc cagacctggc tctttcagca	5100
ctgctcgggg gacagatcca gggcgttact cctcttgcca tttctgtcat cctctctct	5160
aaatttgctg tgggtgttag tcccatccaa ctatctagtc tcaccagtgc tcaggetgtg	5220
gctgtcactc ctgagcaaat ggcctttctg agtctgagc agcgacgagc agttgcatgg	5280
gccaacatg agggaaagga gagcccagaa cagcaaggtc gaagtacagc ctggggcctc	5340
caggactggt cagaccttc ctggctccctg gtattgacta tcagcttctt tggccacctg	5400

ctatgagcct gtctctacag tagaaggaga ttgtggggag agaaatctta agtcataatg 5460
aataaagtgc aaacagaagt gcatcctgat tattttcaga agctgatgag gaata 5515
<210> 31
<211> 5361
<212> DNA
<213> Mus musculus
<400> 31
gcccctactg ggcctcagtc tttggatgct ggtctctccc ttctgaagtc attcgtagcc 60

actctggacc aagctcctca gcgttccctc agccagtcac ggttctctgc gttcctggcc 120
aacatttctt catccttcca gcttgggagg atgggggagg gaccggtggg agagccccc 180
cctctccagc cccctgcaat tgcattcat gatctcctcg tgacactgag aggtagccca 240
gactgggagc caatgctagg gcttctggga gatgtgctgg cactcctggg acaggaacag 300
actccccggg acttttttgt gcaccaggca ggtgtactgg gtggacttgt agaggcattg 360
ttgggagcgt tagttctgg aggccccct gccccactc gacccccatg caccctgat 420
ggcccttctg actgtgtcct ggctgctgat tggttgcctt ctctgatgtt gttattagag 480

ggtacacgct ggcaggccct ggtgcagttg cagcccagtg tggacccaac caatgccaca 540
ggtcttgatg gtagagagcc agctcctcac tttttacagg gtctgctggg ctgtcttacc 600
ccagcaggag agttgggctc tgaggaggct ctttggggtg gtctgctgcg cacagtgggg 660
gccccctct atgctgcctt ccaggagggg ctactgcgag tcaactatc tctgcaagat 720
gaggtctttt ctattatggg acagccagag cctgatgcca gtgggcagtg ccaggaggc 780
aaccttcaac agctgctttt atggggcatg cggaacaacc tttcttggga cggccagca 840
ctgggttttc tatctggatc accacctcca cccctgctc tctgcaactg cctgagcaga 900

ggtgtgcctc tgcccagggc ttcccagcct gcggctcaca tcagccctcg acagcggcga 960
gccatctctg tggaggccct ctgcgagaac cactcaggcc cagagccacc ctacagcatc 1020
tccaacttct ccatctactt gctctgccag cacatcaage ctgccacccc gcggccccct 1080
cctaccacc cagggcctcc tctaccacc ccacagcccc ctctaccac tacacagccc 1140
attctgaca ctacacagcc cctctctgtc accccaaggc ctctctctac caccacaaa 1200
ccccctcta gcacagctgt catctgccag acagctgtat ggtacgcagt ctctgggca 1260
ccaggtgccc gaggttggt ccaagcctgc catgatcagt ttcctgatca atttctggat 1320

atgatctgcg gcaacctctc attttcagcc ctgtctggcc ccagtcgtcc tttggtaaag 1380
cagctctgtg ctggcttct cccaccccc actagctgtc caccaggcct gatccctgtg 1440

cccctcacc cagaaatatt ctggggctgt ttcttgaga atgagacact gtgggctgaa 1500
 cggttgtgtg tggaggacag tctgcaggct gtccccca ggaaccaggc ttgggttcag 1560
 catgtgtgtc ggggccccac cttggacgcc actgattttc caccgtgccg cgttggaccc 1620
 tgtggggaac gctgccaga tgggggcagc ttctgtctca tggctgtgtc caatgacact 1680
 ctgtatgaag ccttggttcc cttctgggct tggctagcag gccaatgcag aattagtcgt 1740

 ggaggaaatg atacttgcct tctagaaggc atgctgggcc ccttgttgcc ctctctgccc 1800
 cctctgggac catccccact ctgtctggct cctggctcct ttctgcttgg catgttatcc 1860
 cagttgccac gctgtcagtc ctccgtgcc aacctcgccc accccacgcg cctacattac 1920
 ctctcgccc tactgacctt cttctgggt ccagggactg ggggtgccga gacgcagggg 1980
 atgttaggtc aagccctgct gctctctagt ctcccagaca actgttcatt ctgggatgcc 2040
 ttccgccag agggccggag aagtgtactg aggacagtcg gagagtactt gcagcgggaa 2100
 gagccaaccc caccaggctt agactcctcc ctccagctcg gctctggtat gagcaagatg 2160

 gagcttctgt cctgcttcag tctgtactg tgggatctac tccagagaga gaagagcgtt 2220
 tgggccctga ggaccttgt gaaggcctac ctgcgcatgc ctccagaaga ccttcagcag 2280
 cttgtgcttt cagcagagat ggaggctgca cagggtctcc tgacgctcat gcttcgttcc 2340
 tgggctaagc tgaaggttca accatccgag gacgagcca tgggccctc gacagccttg 2400
 ctgtccagc ggtaccacag cctcacctcc caactcttta tgcacatgtc accgtcatc 2460
 ccttctctgg ctgtccctga cctcatgcgc ttcccaccgt ccttttggc caacgacagt 2520
 gtcttgctg ccatcagga tcacagctca ggaatgaagc ctgaacagaa ggaggccctg 2580

 gcaaacgac tgctggcccc tgagctgttt ggagaagtgc ctgattggcc ccaggagctg 2640
 ctgtgggag cctgacctt ctttcccat ctgctctgg agagctttct ccagctcagc 2700
 cctcaccaga tccaggccct ggaggatagc tggccagtag cagatcttgg gccgggacac 2760
 gcccacatg tgcttcgtag cctagtaaac cagagcatgg aggatgggga ggagcaggtg 2820
 ctgagcttg ggtccctgc ctgtttctg agtctgagg agctacagag tctggtgcc 2880
 ttgagtgate caatggggcc tgtagaacag ggtctgtgg aatgtgcgc caatgggacc 2940
 ctgagcccag aaggacgggt ggcatatga cttctgggag tgttgcgttc atctggagga 3000

 actgtcttaa gccccgaga gctgagggtc tgggcacctc ttttccca gctgggcctc 3060
 cgcttctgc aggagctctc agagaccag cttagagcca tgcttctgc cctacaggga 3120
 gccagtgtca cacctgcccc ggctgttctg ttgtttggaa ggctccttc taagcatgat 3180
 ctgtccctgg aggaactctg ctccctgcac cctctctgc caggtctcag ccccagaca 3240
 ctccaggcca tcctaagag agttctggtt ggtgcttgtt cctgcctggg cctgaactg 3300

tcaaggcttt cagcttgcca gattgcagct ctgctgcaga cctttcgggt aaaagatggt	3360
gttaaaaata tgggtgcagc aggtgccggc tcagccgtgt gcattcctgg gcagcccacc	3420
acttgccag actgcctgct tccctgctc ccattaaagc tgctacagct ggacgctgca	3480
gctcttctgg caaacggaag actctatcgg cagctgcctt ggtctgagca acaggcacag	3540
tttctctgga agaaaatgca agtgccctacc aacctgagcc tgaggaatct gcaggctctg	3600
ggcaacttgg caggaggcat gacctgcgag tttctgcagc agatcagctc aatggttgac	3660
tttcttgatg tggtaacacat gctctaccag ctgccactg gtgttcgaga gacctgcgg	3720
gcctgtatct ggacagagct acagcggagg atgacaatgc cagagccaga gctgaccacc	3780
ctagggccag aactgagtga acttgacaca aagctactcc tggacttgcc gatccagctg	3840
atggacagat tgtccaatga ttccattatg ttggtggtgg agatggtcca aggcgctcca	3900
gagcagctgc tggcactgac cccactccac cagacagcct tggcagagcg agcacttaaa	3960
aacctggctc caaaggagac cccaatctcc aaagaagtgc tggagacact gggccccctg	4020
gttggattcc tgggaataga gagcacgcga cggatccctt taccattct actgtctcat	4080
ctcagtcagc tgcagggtct ctgcctagga gagacatttg ccacagagct gggatggctg	4140
ctgttgaggc agcctgttct tggaaaacca gaattgtgga gccaggatga aatagagcaa	4200
gctggacgcc tagtattcac tctgtctgct gaggtatatt cctcgatccc caggagggt	4260
ttgggccag agacactgga gaggtttctg ggaaagcatc aaagctggga gcagagcaga	4320
gtgggccatc tgttgggga gtcacagctt gccacaaga aagcagctct ggtagctggg	4380
attgtgcatc cagctgctga gggcttccaa gagcctgtac caaactgtgc agacatacgg	4440
ggaaccttcc cagcggcctg gtctgcgaca caaatctcag agatggaact ctgagcttt	4500
gaagactgcc tgtcactatt tgctggagat ccaggacttg gtcctgagga actacgggca	4560
gccatgggca agccaagca gttgtggggc cccctcgag gattccgtcc tgagcagatc	4620
ttgcagctgg gccgtctcct gataggtcta ggagaacggg aactgcagga gcttaccttg	4680
gtggactggg gtgtgctgag cagcctgggg caaatagatg gctggagtgc catgcagctc	4740
cgagccgtgg tctccagttt cctaaggcag agtggtcggc atgtgagcca cctggacttc	4800
atttatctga cagcactggg ttacacagtc tgtggattgc gaccagagga gttacagcac	4860
atcagcagtt gggagttag ccaagcagct ctcttcttgg gtagcttgca tctccgtgc	4920
tctgaggaa agctggaagt tctggcctat ctcttctgtg tgcctggtgg ctttgccca	4980
gtcagtaact gggggcctga gatcttact gaaattggca caatagcagc tggcatccca	5040

gacctggctc tticagcatt actgcgggga cagatccaag gcctgactcc tcttgccatt 5100

tctgtcattc ctgctcccaa gtttgacgtg gtcttcaacc ccatccagtt atctagtctc 5160

accaggggtc aggccgtagc tggtactcct gaacagctgg cctatctgag tectgagcag 5220

cggcgagcag ttgcatgggc ccaacacgaa gggaaggaga tcccagagca gctgggtcga 5280

aactcagcct ggggtctcta cgactgggtc caagcctcct gggccctggc attgcccgtc 5340

agcatttttg gccacctatt a 5361

<210> 32

<211> 5757

<212> DNA

<213> Mus musculus

<400> 32

gccccgtctt cacttgcta tcccttcatg gtgagcatag ccagaactca cctctaggcc 60

cagtgtgcac ctggaaatat ggctctgagc ctccagcccc agctgtcctt tctcctgtcg 120

ctcctgccgc aggaagtgac ttacacccct actgggcctc agtctttgga tgctggtctc 180

tcccttctga agtcattcgt agccactctg gaccaagctc ctacagcttc cctcagccag 240

tcacggttct ctgcgttctt ggccaacatt tcttcactct tccagcttgg gaggatgggg 300

gaggagaccg tgggagagcc cccacctctc cagccccctg cacttcgact tcatgatctc 360

ctcgtgacac tgagaggtag cccagactgg gagccaatgc tagggcttct gggagatgtg 420

ctggcactcc tgggacagga acagactccc cgggactttt tgggtgcacca ggcaggtgta 480

ctgggtggac ttgtagaggc attgttggga gcgttagttc ctggaggccc cctgcccc 540

actcgacccc catgcacccg tgatggccct tctgactgtg tcttggtgct tgattggttg 600

ccttctctga tgttgttatt agagggtaca cgttggcagg ccttgggtga gttgcagccc 660

agtgtggacc caaccaatgc cacaggtctt gatggttagag agccagctcc tcaactttta 720

cagggtctgc tgggcttctt taccacagca ggagagtgg gctctgagga ggctctttgg 780

ggtggtctgc tgcgcacagt gggggccccc ctctatgctg ccttcagga ggggtactg 840

cgagtcactc attctctgca agatgaggtc ttttctatta tgggacagcc agagcctgat 900

gccagtgggc agtgcacagg aggcaacctt caacagctgc ttttatgggg catgcggaac 960

aacctttctt gggacgcccc agcactgggt tttctatctg gatcaccacc tccacccct 1020

gctctcctgc actgcctgag cagaggtgtg cctctgccc gggcttccca gcctgcggct 1080

cacatcagcc ctgcacagcg gcgagccatc tctgtggagg cctctgcga gaaccactca 1140

ggcccagagc caccctacag catctccaac ttctccatct acttgctctg ccagcacatc 1200

aagcctgcc a ccccgccg ccctcctacc accccacggc ctctcctac caccacacag	1260
ccccctcta cactacaca gccatttct gacactacac agccccctcc tgtcaccca	1320
aggcctctc ctaccacccc acaacccct cctagcacag ctgtcatctg ccagacagct	1380
glatggtag cagtctctg ggcaccaggt gcccgaggtt ggctccaagc ctgcatgat	1440
cagtttctg atcaatttct ggatatgac tgcggcaacc tctcatttct agcctgtct	1500
ggccccagtc gtcttttgg aaagcagctc tgtgtggct tgtcccacc cccactagc	1560
tgtccaccag gcctgatccc tgtgcccctc accccagaaa tattctgggg ctgtttctg	1620
gagaatgaga cactgtgggc tgaacggtt tgtgtggagg acagtctgca ggctgtgccc	1680
ccgaggaacc agccttgggt tcagcatgtg tgcggggcc ccaccttga cgcactgat	1740
tttccaccgt gccgcgttgg accctgtggg gaacgctgcc cagatggggg cagcttctg	1800
ctcatggtct gtgccaatga cactctgtat gaagccttgg ttcccttctg ggcttggcta	1860
gcaggccaat gcagaattag tcgtggagga aatgatactt gctttctaga aggcctgctg	1920
ggccccctgt tgcctctct gccccctctg ggaccatccc cactctgtct ggctcctggt	1980
ccttttctgc ttggcatgtt atcccagttg ccacgtgtc agtcctccgt gccagccctc	2040
ggccacccca cgcgcctaca ttacctctg cgcctactga ccttcttct gggtccaggg	2100
actgggggtg ccgagacgca ggggatgtta ggtcaagccc tgcgtctctc tagtctcca	2160
gacaactgtt cattctggga tgccttccgc ccagagggcc ggagaagtgt actgaggaca	2220
gtcggagagt acttgacgag ggaagagcca accccaccag gcttagactc ctccctcagc	2280
ctcgctctg gtatgagcaa gatggagctt ctgtcctgt tcagtcctgt actgtgggat	2340
ctactccaga gagagaagag cgtttgggcc ctgaggaccc tggagaaggc ctacctgcgc	2400
atgcctccag aagaccttca gcagcttctg ctttcagcag agatggaggc tgcacagggc	2460
ttctgacgc tcatgcttcg ttctgggct aagctgaagg ttcaaccatc caggagcag	2520
gccatgggcc gcctgacagc cttgtgtctc cagcggtacc cagcctcac ctcccaactc	2580
tttatcgaca tgtcacctg cateccctc ctggctgtcc ctgacctcat gcgttccca	2640
ccgtcccttt tggccaacga cagtgtcctg gctgcatca gggatcacag ctgaggaatg	2700
aagcctgaac agaaggaggc cctggcaaaa cgactgctgg cccctgagct gtttgagaa	2760
gtgctgatt ggccccagga gctgtgtgg gcagccctgc ctctgttcc ccattgctt	2820
ctggagagct ttctccagct cagccctcac cagatccagg cctggagga tagctggcca	2880
gtagcagatc ttgggccggg acacgccga catgtgttc gtagcctagt aaaccagagc	2940

atggaggatg gggaggagca ggtgctcagg cttgggtccc tcgcctgttt cctgagtcct	3000
gaggagctac agagtctggt gcccttgagt gatccaatgg ggctctgtaga acagggtctg	3060
ctggaatgtg cgccaatgg gacctcagc ccagaaggac gggtagcata tgaacttctg	3120
ggagtgttgc gtcatctgg aggaactgtc ttaagcccc gagagctgag ggtctgggca	3180
cctctctttc cccagctggg cctccgcttc ctgcaggagc tctcagagac ccagcttaga	3240
gccatgcttc ctgccctaca gggagccagt gtcacacctg cccaggctgt tctgttgttt	3300
ggaaggctcc ttctaagca tgatctgtcc ctggaggaac tctgtccct gcacctctc	3360
ctgccaggtc tcagccccc gacactccag gccatcccta agagagttct ggttggtgct	3420
tgttcctgcc tgggccctga actgtcaagg ctttcagctt gccagattgc agctctgctg	3480
cagacctttc gggtaaaaga tgggtttaaa aatatgggtg cagcaggtgc cggctcagcc	3540
gtgtgcattc ctgggcagcc caccacttgg ccagactgcc tgcttccct gctccatta	3600
aagctgctac agctggacgc tgcagctctt ctggcaaacc gaagactcta tcggcagctg	3660
ccttggtctg agcaacaggc acagtttctc tggaagaaaa tgcaagtgcc taccaacctg	3720
agcctgagga atctgcagc tctgggcaac ttggcaggag gcatgacctg cgagtttctg	3780
cagcagatca gctcaatggt tgactttctt gatgtggtac acatgctcta ccagtgccc	3840
actggtgttc gagagagcct gcgggcctgt atctggacag agctacagcg gaggatgaca	3900
atgccagagc cagagctgac caccctaggg ccagaactga gtgaacttga caciaagcta	3960
ctcctggact tgccgatcca gctgatggac agattgtcca atgattccat tatgttggtg	4020
gtggagatgg tccaaggcgc tccagagcag ctgctggcac tgacccact ccaccagaca	4080
gccttggcag agcgagcact taaaaacctg gctccaaagg agacccaat ctccaaagaa	4140
gtgctggaga cactgggccc cttggttga ttcttgggaa tagagagcac gcgacggatc	4200
cctttaccca ttctactgtc tcactcagt cagctgcagg gcttctgcct aggagagaca	4260
tttgccacag agctgggatg gctgctgttg caggagcctg ttcttggaaa accagaattg	4320
tggagccagg atgaaataga gcaagctgga cgctagtat tcactctgtc tgctgaggct	4380
atttcctcga tccccaggga ggctttgggc ccagagacac tggagaggct tctgggaaag	4440
catcaaagct gggagcagag cagagtgggc catctgtgtg gggagtcaca gcttgccac	4500
aagaaagcag ctctggtagc tgggattgtg catccagctg ctgagggtct ccaagagcct	4560
gtaccaaact gtgcagacat acggggaacc ttcccagcgg cctggtctgc gacacaaatc	4620
tcagagatgg aactctcaga ctttgaagac tgctgtcac tatttgctgg agatccagga	4680

cttggctcctg aggaactacg ggcagccatg ggcaaggcca agcagttgtg gggccccct 4740

cgaggattcc gtcttagca gatcttgac ctgggccgtc tcctgatagg tctaggagaa 4800

cgggaactgc aggagcttac cttggtggac tggggtgtgc tgagcagcct ggggcaaata 4860

gatggctgga gticcatgca gctccgagcc gtggtctcca gtttcctaag gcagagtgg 4920

cggcatgtga gccacctgga cttcatttat ctgacagcac tgggttacac agtctgtgga 4980

ttgcgaccag aggagttaca gcacatcagc agtgggagt ttagccaagc agctctcttc 5040

ctgggtagct tgcatctccc gtgctctgag gaacagctgg aagtcttggc ctatctcctt 5100

gtgttgctg gtggctttgg cccagtcagt aactgggggc ctgagatctt cactgaaatt 5160

ggcacaatag cagctggcat cccagacctg gctctttcag cattactgcg gggacagatc 5220

caaggcctga ctctcttgcc catttctgtc attctgtct ccaagtttgc agtggctctc 5280

aaccccatcc agttatctag tctcaccagg ggtcaggccg tagctgttac tcctgaacag 5340

ctggcctatc tgagtctga gcagcggcga gcagttgcat gggccaaca cgaagggaag 5400

gagatcccag agcagctggg tcgaaactca gcctggggtc tctacgactg gttccaagcc 5460

tcctggggcc tggcattgcc cgtcagcatt tttggccacc tattatgata actgttcctt 5520

cagttgaggg agaaaattta catcactatg aacaacttgt aaatggaagt gcataactat 5580

tattctcagt aagtggatga ggattgtggg taaaattcca atgcattcca cccacctgag 5640

aactgtgctc ctggcatata cgctcttgc catcatgaat aacctcactg tttcttttca 5700

tttctactt cctctcacc acaccaatag aaataacaca gacagctgca acataat 5757

<210> 33

<211> 5379

<212> DNA

<213> Artificial Sequence

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

<400> 33

atggctctca gcctctggcc cctgctgctg ctgctgctgc tgctgctgct gctgtccttt 60

gcagtgaactc tggcccttac tgggcctcat tccctggacc ctggtctctc cttctgaag 120

tcattgctct ccaactctga ccaggctccc cagggtctcc tgagccgctc acggttcttt 180

acattctctg ccaacatttc ttcttcttt gagcctggga gaatggggga aggaccagta 240

ggagagcccc cacctctcca gccgctgct ctgcggctcc atgattttct agtgacactg 300

agaggtagcc ccgactggga gccaatgcta gggctgctag gggatatgct ggcactgctg 360

ggacaggagc agactcccc agatttctg gtgcaccagg caggggtgct ggggtgactt 420

gtggagggtgc tgctgggagc cttagtctct gggggccccc ctaccccaac tcggccccc 480

tgcaccctg atgggcccgc tgactgtgtc ctggctgtg actggttgcc ttctctgtg 540

ctgttttag agggcacacg ctggcaagct ctggcagcagg tgcagcccag tgtggacccc 600

accaatgcca caggccctga tgggaggag gcagctctc actttttgca gggctctgtg 660

ggtttgccta cccaacagg ggagctaggc tccaaggagg ctctttgggg cggctctgta 720

cgcacagtgg gggccccct ctatctgcc ttccaggagg ggctgtccg tgtcactcac 780

tccttcagc atgaggtctt ctccattttg gggcagccag agcctgatac caatgggcag 840

tgccaggag gtaacctca acagctgtc ttatggggcg tccggcaca cctttcttg 900

gatgtccagg cgctgggctt tctgtctgga tcaccacccc caccctgc cctcctcac 960

tgcttgagca cgggcgtgcc tctgccaga gcttctcagc cgtcagcca catcagcca 1020

cgccaacggc gagccatcac tgtggaggcc ctctgtgaga accacttagg cccagcacca 1080

ccctacagca ttccaactt ctccatccac ttgctctgcc agcacacaa gcctgccact 1140

ccacagcccc atcccagcac cactgccatc tgccagacag ctgtgtgta tgcagtgtc 1200

tgggcaccag gtgccaagg ctggctacag gcctgccag accagtttcc tgatgagttt 1260

ttggtgcca tctcagtaa cctctcttt tcagccctgt ctggctcaa cggccgctg 1320

gtgaagcggc tctgtctgg cctgtccca cccctacca gctgccctga aggctgccc 1380

cctgttcccc tcacccaga catcttttg ggctgttct tggagaatga gactctgtg 1440

gctgagcgc tgtgtggga ggcaagtcta caggctgtgc ccccagcaa ccaggcttg 1500

gtccagcatg tgtgccagg cccacccca gatgtcactg cctccccacc atgccacatt 1560

ggacctgtg gggaacgctg cccgatggg gcagcttcc tggatggt ctgtgccaat 1620

gacacatgt atgaggtctt ggtgccctt tgcccttggc tagcaggcca atgcaggata 1680

agtcgtggg gcaatgacac ttgcttcta gaagggtgc tgggccccct tctgccctt 1740

ctgccaccac tgggaccatc cccactctgt ctgaccctg gcccttctt ccttgcatg 1800

ctatcccagt tgccacgtg teagctctt gtccagctc ttgtcaccc cacacgcta 1860

cactatctc tccgctgtg gacctctc ttgggtccag gggctggggg cgtgaggcc 1920

caggggatgc tgggtcgggc cctactgtc tccagtctc cagacaactg ctcttcttg 1980

gatgccttt gccagaggg ccggcgagc gtgtacgga cgattgggga atacctgga 2040

caagatgagg agcagccaac cccatcaggc ttgaacca ctgtcaacc cagctctggt 2100

ataagcaaga tggagctgt ggcctgctt agtctgtgc tgtgggatct gctccagg 2160

gaaaagagtg tttagggccct gcagattctia gtgcaggcgt acctgcatat gccccagaa 2220
 aacctccagc agctggtgct ttcagcagag agggaggctg cacagggtt cctgacactc 2280
 atgctgcagg ggaagctgca ggggaagctg caggtaccac catccgagga gcaggccctg 2340
 ggtcgccctga cagccctgct gctccagcgg taccacgcc tcacctcca gctcttcatt 2400
 gacctgtcac cactcatccc tttcttggct gtctctgacc tgatgcgctt cccaccatcc 2460
 ctgttagcca acgacagtgt cctggctgcc atccgggatt acagcccagg aatgaggcct 2520
 gaacagaagg aggtctctggc aaagcgactg ctggccctg aactgtttgg ggaagtcct 2580

 gcctggcccc aggagctgct gtgggcagtg ctgcccctgc tccccacct cctctggag 2640
 aactttttgc agctcagccc tcaccagatc caggccctgg aggatagctg gccagcagca 2700
 ggtctggggc cagggcattg ccgccatgtg ctgcgcagcc tggtaaacca gagtgtccag 2760
 gatggtgagg agcaggtacg caggcttggg cccctgcct gtttctgag cctgaggag 2820
 ctgcagagcc tagtgccct gagtgatcca acggggccag tagaacgggg gctgctggaa 2880
 tgtgcagcca atgggacct cagcccagaa ggacgggtgg catatgaact tctgggtgtg 2940
 ttgcgctcat ctggaggagc ggtgctgagc cccgggagc tgcgggtctg ggccctctc 3000

 ttctctcagc tgggcctccg ctctctcag gagctgtcag agccccagct tagagccatg 3060
 ctctctgtcc tgcagggaac tagtgttaca cctgtcagg ctgtcctgct gcttggacgg 3120
 ctcttctcta ggcacgatct atccctggag gaactctgct ccttgacct tctgtacca 3180
 ggccctcagc cccagacact ccaggccatc cctaggcgag tctgggtcgg ggcttgttcc 3240
 tgccctggccc ctgaactgtc acgctctca gcctgccaga ccgcagcact gctgcagacc 3300
 ttctgggtta aagatggtgt taaaaatatg ggtacaacag gtgctggtcc agctgtgtgt 3360
 atccctggtc agcctattcc caccacctgg ccagactgcc tgcttccct gctccatta 3420

 aagctgctac aactggattc cttagctctt ctggcaaac gaagacgcta ctgggagctg 3480
 ccctggctctg agcagcaggc acagtcttc tggaagaaga tgcaagtacc caccaacctt 3540
 acctcagga atctgcaggc tctgggcacc ctggcaggag gcatgtcctg tgagtttctg 3600
 cagcagatca actccatggt agacttctt gaagtgtgac acatgatcta tcagctgccc 3660
 actagagttc gagggagcct gagggcctgt atctgggcag agctacagcg gaggatggca 3720
 atgccagaac cagaatggac aactgtaggg ccagaactga acgggctgga tagcaagcta 3780
 ctctggact taccgatcca gttgatggac agactatcca atgaatccat tatgttggtg 3840

 gtaggagctg tgcaaagagc tccagagcag ctgctggcac tgacccccct ccaccaggca 3900
 gccctggcag agagggcact acaaaacctg gctccaaagg agaactcagt ctcaggggaa 3960
 gtgctggaga ccttaggccc ttgtgttga ttctgggga cagagagcac acgacagatc 4020

cccctacaga tctgtctgtc ccatctcagt cagctgcaag gcttctgcct aggagagaca 4080
 tttgccacag agctgggatg gctgtctattg caggagtctg ttcttgggaa accagagttg 4140
 tggagccagg atgaagtaga gcaagctgga cgcctagtat tcactctgtc tactgaggca 4200
 atttcttga tccccaggga ggccttgggt ccagagaccc tggagcggct tctagaaaag 4260

cagcagagct gggagcagag cagagttgga cagctgtgta gggagccaca gcttctgcc 4320
 aagaaagcag ccttggtagc aggggtggtg cgaccagctg ctgaggatct tccagaacct 4380
 gtgccaaatt gtgcagatgt acgagggaca ttcccagcag cctggtctgc aaccagatt 4440
 gcagagatgg agctctcaga cttttaggac tgcctgacat tatttgcagg agaccagga 4500
 cttaggcctg aggaactcgc ggcagccatg ggcaaagcaa aacagttgtg gggcccccc 4560
 cggggatttc gtcttgagca gatcctgcag cttaggtaggc tcttaatagg tctaggagat 4620
 cgggaactac aggagctgat cctagtggac tggggagtgc tgagcacct ggggcagata 4680

gatggctgga gcaccactca gctccgcat gtggtctcca gtttctacg gcagagtgg 4740
 cggcatgtga gccacctgga ctctgttcat ctgacagcgc tgggttatac tctctgtgga 4800
 ctgcggccag aggagctcca gcacatcagc agttgggagt tcagccaagc agctctcttc 4860
 ctcgccacc tgcatctcca gtgctctgag gaacaactgg aggttctggc ccacctactt 4920
 gtactgcctg gtgggtttgg cccaatcagt aactgggggc ctgagatctt cactgaaatt 4980
 ggcaccatag cagctgggat ccagacctg gctctttcag cactgctgcg gggacagatc 5040
 cagggcgta ctctcttgc catttctgtc atccctctc ctaaatttgc tgtggtgttt 5100

agtccatcc aactatctag tctcaccagt gctcaggctg tggtgtcac tctgagcaa 5160
 atggcctttc tgagtctga gcagcgacga gcagttgcat gggcccaaca tgagggaag 5220
 gagagcccag aacagcaagg tcgaagtaca gcctggggcc tccaggactg gtcacgacct 5280
 tcttggtccc tggattgac tatcagcttc cttagccacc tgctaacgcg tacgcggcgc 5340
 ctcgagcaga aactcatctc agaagaggat ctgtaatag 5379

<210> 34

<211> 18

<212> DNA

<213> Artificial Sequence

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

<400> 34

acgcgtacgc ggccgctc

18

<210> 35

<211> 30

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 35

gagcagaaac tcatctcaga agaggatctg

30

<210> 36

<211> 1791

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 36

Met Ala Leu Ser Leu Trp Pro Leu Leu Leu Leu Leu Leu Leu Leu

1 5 10 15
Leu Leu Ser Phe Ala Val Thr Leu Ala Pro Thr Gly Pro His Ser Leu

20 25 30
Asp Pro Gly Leu Ser Phe Leu Lys Ser Leu Leu Ser Thr Leu Asp Gln

35 40 45
Ala Pro Gln Gly Ser Leu Ser Arg Ser Arg Phe Phe Thr Phe Leu Ala

50 55 60
Asn Ile Ser Ser Ser Phe Glu Pro Gly Arg Met Gly Glu Gly Pro Val

65 70 75 80
Gly Glu Pro Pro Pro Leu Gln Pro Pro Ala Leu Arg Leu His Asp Phe

85 90 95
Leu Val Thr Leu Arg Gly Ser Pro Asp Trp Glu Pro Met Leu Gly Leu

100 105 110
Leu Gly Asp Met Leu Ala Leu Leu Gly Gln Glu Gln Thr Pro Arg Asp

115 120 125
Phe Leu Val His Gln Ala Gly Val Leu Gly Gly Leu Val Glu Val Leu

130 135 140
 Leu Gly Ala Leu Val Pro Gly Gly Pro Pro Thr Pro Thr Arg Pro Pro
 145 150 155 160
 Cys Thr Arg Asp Gly Pro Ser Asp Cys Val Leu Ala Ala Asp Trp Leu
 165 170 175
 Pro Ser Leu Leu Leu Leu Leu Glu Gly Thr Arg Trp Gln Ala Leu Val
 180 185 190
 Gln Val Gln Pro Ser Val Asp Pro Thr Asn Ala Thr Gly Leu Asp Gly

 195 200 205
 Arg Glu Ala Ala Pro His Phe Leu Gln Gly Leu Leu Gly Leu Leu Thr
 210 215 220
 Pro Thr Gly Glu Leu Gly Ser Lys Glu Ala Leu Trp Gly Gly Leu Leu
 225 230 235 240
 Arg Thr Val Gly Ala Pro Leu Tyr Ala Ala Phe Gln Glu Gly Leu Leu
 245 250 255
 Arg Val Thr His Ser Leu Gln Asp Glu Val Phe Ser Ile Leu Gly Gln

 260 265 270
 Pro Glu Pro Asp Thr Asn Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln
 275 280 285
 Leu Leu Leu Trp Gly Val Arg His Asn Leu Ser Trp Asp Val Gln Ala
 290 295 300
 Leu Gly Phe Leu Ser Gly Ser Pro Pro Pro Pro Pro Ala Leu Leu His
 305 310 315 320
 Cys Leu Ser Thr Gly Val Pro Leu Pro Arg Ala Ser Gln Pro Ser Ala

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

Pro Ser Thr Thr Ala Ile Cys Gln Thr Ala Val Trp Tyr Ala Val Ser

385 390 395 400

Trp Ala Pro Gly Ala Gln Gly Trp Leu Gln Ala Cys His Asp Gln Phe

405 410 415

Pro Asp Glu Phe Leu Asp Ala Ile Cys Ser Asn Leu Ser Phe Ser Ala

420 425 430

Leu Ser Gly Ser Asn Arg Arg Leu Val Lys Arg Leu Cys Ala Gly Leu

435 440 445

Leu Pro Pro Pro Thr Ser Cys Pro Glu Gly Leu Pro Pro Val Pro Leu

450 455 460

Thr Pro Asp Ile Phe Trp Gly Cys Phe Leu Glu Asn Glu Thr Leu Trp

465 470 475 480

Ala Glu Arg Leu Cys Gly Glu Ala Ser Leu Gln Ala Val Pro Pro Ser

485 490 495

Asn Gln Ala Trp Val Gln His Val Cys Gln Gly Pro Thr Pro Asp Val

500 505 510

Thr Ala Ser Pro Pro Cys His Ile Gly Pro Cys Gly Glu Arg Cys Pro

515 520 525

Asp Gly Gly Ser Phe Leu Val Met Val Cys Ala Asn Asp Thr Met Tyr

530 535 540

Glu Val Leu Val Pro Phe Trp Pro Trp Leu Ala Gly Gln Cys Arg Ile

545 550 555 560

Ser Arg Gly Gly Asn Asp Thr Cys Phe Leu Glu Gly Leu Leu Gly Pro

565 570 575

Leu Leu Pro Ser Leu Pro Pro Leu Gly Pro Ser Pro Leu Cys Leu Thr

580 585 590

Pro Gly Pro Phe Leu Leu Gly Met Leu Ser Gln Leu Pro Arg Cys Gln

595 600 605

Ser Ser Val Pro Ala Leu Ala His Pro Thr Arg Leu His Tyr Leu Leu

610 615 620

Arg Leu Leu Thr Phe Leu Leu Gly Pro Gly Ala Gly Gly Ala Glu Ala

625					630				635				640			
Gln	Gly	Met	Leu	Gly	Arg	Ala	Leu	Leu	Leu	Ser	Ser	Leu	Pro	Asp	Asn	
				645				650				655				
Cys	Ser	Phe	Trp	Asp	Ala	Phe	Arg	Pro	Glu	Gly	Arg	Arg	Ser	Val	Leu	
				660				665				670				
Arg	Thr	Ile	Gly	Glu	Tyr	Leu	Glu	Gln	Asp	Glu	Glu	Gln	Pro	Thr	Pro	
				675				680				685				
Ser	Gly	Phe	Glu	Pro	Thr	Val	Asn	Pro	Ser	Ser	Gly	Ile	Ser	Lys	Met	
				690				695				700				
Glu	Leu	Leu	Ala	Cys	Phe	Ser	Pro	Val	Leu	Trp	Asp	Leu	Leu	Gln	Arg	
705				710				715				720				
Glu	Lys	Ser	Val	Trp	Ala	Leu	Gln	Ile	Leu	Val	Gln	Ala	Tyr	Leu	His	
				725				730				735				
Met	Pro	Pro	Glu	Asn	Leu	Gln	Gln	Leu	Val	Leu	Ser	Ala	Glu	Arg	Glu	
				740				745				750				
Ala	Ala	Gln	Gly	Phe	Leu	Thr	Leu	Met	Leu	Gln	Gly	Lys	Leu	Gln	Gly	
				755				760				765				
Lys	Leu	Gln	Val	Pro	Pro	Ser	Glu	Glu	Gln	Ala	Leu	Gly	Arg	Leu	Thr	
770				775				780								
Ala	Leu	Leu	Leu	Gln	Arg	Tyr	Pro	Arg	Leu	Thr	Ser	Gln	Leu	Phe	Ile	
785				790				795				800				
Asp	Leu	Ser	Pro	Leu	Ile	Pro	Phe	Leu	Ala	Val	Ser	Asp	Leu	Met	Arg	
				805				810				815				
Phe	Pro	Pro	Ser	Leu	Leu	Ala	Asn	Asp	Ser	Val	Leu	Ala	Ala	Ile	Arg	
				820				825				830				
Asp	Tyr	Ser	Pro	Gly	Met	Arg	Pro	Glu	Gln	Lys	Glu	Ala	Leu	Ala	Lys	
835				840				845								
Arg	Leu	Leu	Ala	Pro	Glu	Leu	Phe	Gly	Glu	Val	Pro	Ala	Trp	Pro	Gln	
850				855				860								
Glu	Leu	Leu	Trp	Ala	Val	Leu	Pro	Leu	Leu	Pro	His	Leu	Pro	Leu	Glu	
865				870				875				880				

Asn Phe Leu Gln Leu Ser Pro His Gln Ile Gln Ala Leu Glu Asp Ser
885 890 895

Trp Pro Ala Ala Gly Leu Gly Pro Gly His Ala Arg His Val Leu Arg
900 905 910

Ser Leu Val Asn Gln Ser Val Gln Asp Gly Glu Glu Gln Val Arg Arg
915 920 925

Leu Gly Pro Leu Ala Cys Phe Leu Ser Pro Glu Glu Leu Gln Ser Leu
930 935 940

Val Pro Leu Ser Asp Pro Thr Gly Pro Val Glu Arg Gly Leu Leu Glu
945 950 955 960

Cys Ala Ala Asn Gly Thr Leu Ser Pro Glu Gly Arg Val Ala Tyr Glu
965 970 975

Leu Leu Gly Val Leu Arg Ser Ser Gly Gly Ala Val Leu Ser Pro Arg
980 985 990

Glu Leu Arg Val Trp Ala Pro Leu Phe Ser Gln Leu Gly Leu Arg Phe
995 1000 1005

Leu Gln Glu Leu Ser Glu Pro Gln Leu Arg Ala Met Leu Pro Val
1010 1015 1020

Leu Gln Gly Thr Ser Val Thr Pro Ala Gln Ala Val Leu Leu Leu
1025 1030 1035

Gly Arg Leu Leu Pro Arg His Asp Leu Ser Leu Glu Glu Leu Cys
1040 1045 1050

Ser Leu His Leu Leu Leu Pro Gly Leu Ser Pro Gln Thr Leu Gln
1055 1060 1065

Ala Ile Pro Arg Arg Val Leu Val Gly Ala Cys Ser Cys Leu Ala
1070 1075 1080

Pro Glu Leu Ser Arg Leu Ser Ala Cys Gln Thr Ala Ala Leu Leu
1085 1090 1095

Gln Thr Phe Arg Val Lys Asp Gly Val Lys Asn Met Gly Thr Thr
1100 1105 1110

Gly Ala Gly Pro Ala Val Cys Ile Pro Gly Gln Pro Ile Pro Thr

1115 1120 1125
 Thr Trp Pro Asp Cys Leu Leu Pro Leu Leu Pro Leu Lys Leu Leu
 1130 1135 1140
 Gln Leu Asp Ser Leu Ala Leu Leu Ala Asn Arg Arg Arg Tyr Trp

1145 1150 1155
 Glu Leu Pro Trp Ser Glu Gln Gln Ala Gln Phe Leu Trp Lys Lys
 1160 1165 1170
 Met Gln Val Pro Thr Asn Leu Thr Leu Arg Asn Leu Gln Ala Leu
 1175 1180 1185
 Gly Thr Leu Ala Gly Gly Met Ser Cys Glu Phe Leu Gln Gln Ile
 1190 1195 1200
 Asn Ser Met Val Asp Phe Leu Glu Val Val His Met Ile Tyr Gln

1205 1210 1215
 Leu Pro Thr Arg Val Arg Gly Ser Leu Arg Ala Cys Ile Trp Ala
 1220 1225 1230
 Glu Leu Gln Arg Arg Met Ala Met Pro Glu Pro Glu Trp Thr Thr
 1235 1240 1245
 Val Gly Pro Glu Leu Asn Gly Leu Asp Ser Lys Leu Leu Leu Asp
 1250 1255 1260
 Leu Pro Ile Gln Leu Met Asp Arg Leu Ser Asn Glu Ser Ile Met

1265 1270 1275
 Leu Val Val Glu Leu Val Gln Arg Ala Pro Glu Gln Leu Leu Ala
 1280 1285 1290
 Leu Thr Pro Leu His Gln Ala Ala Leu Ala Glu Arg Ala Leu Gln
 1295 1300 1305
 Asn Leu Ala Pro Lys Glu Thr Pro Val Ser Gly Glu Val Leu Glu
 1310 1315 1320
 Thr Leu Gly Pro Leu Val Gly Phe Leu Gly Thr Glu Ser Thr Arg

1325 1330 1335
 Gln Ile Pro Leu Gln Ile Leu Leu Ser His Leu Ser Gln Leu Gln
 1340 1345 1350

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

1580	1585	1590
Leu Gly Tyr Thr Leu Cys Gly	Leu Arg Pro Glu Glu	Leu Gln His
1595	1600	1605
Ile Ser Ser Trp Glu Phe Ser	Gln Ala Ala Leu Phe	Leu Gly Thr
1610	1615	1620
Leu His Leu Gln Cys Ser Glu	Glu Gln Leu Glu Val	Leu Ala His
1625	1630	1635
Leu Leu Val Leu Pro Gly Gly	Phe Gly Pro Ile Ser	Asn Trp Gly
1640	1645	1650
Pro Glu Ile Phe Thr Glu Ile	Gly Thr Ile Ala Ala	Gly Ile Pro
1655	1660	1665
Asp Leu Ala Leu Ser Ala Leu	Leu Arg Gly Gln Ile	Gln Gly Val
1670	1675	1680
Thr Pro Leu Ala Ile Ser Val	Ile Pro Pro Pro Lys	Phe Ala Val
1685	1690	1695
Val Phe Ser Pro Ile Gln Leu	Ser Ser Leu Thr Ser	Ala Gln Ala
1700	1705	1710
Val Ala Val Thr Pro Glu Gln	Met Ala Phe Leu Ser	Pro Glu Gln
1715	1720	1725
Arg Arg Ala Val Ala Trp Ala	Gln His Glu Gly Lys	Glu Ser Pro
1730	1735	1740
Glu Gln Gln Gly Arg Ser Thr	Ala Trp Gly Leu Gln	Asp Trp Ser
1745	1750	1755
Arg Pro Ser Trp Ser Leu Val	Leu Thr Ile Ser Phe	Leu Gly His
1760	1765	1770
Leu Leu Thr Arg Thr Arg Pro	Leu Glu Gln Lys Leu	Ile Ser Glu
1775	1780	1785
Glu Asp Leu		
1790		
<210> 37		
<211> 6		
<212> PRT		

<213> Artificial Sequence

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

<400> 37

Thr Arg Thr Arg Pro Leu

1 5

<210> 38

<211> 5481

<212> DNA

<213> Artificial Sequence

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

<400> 38

atggctctga gcctccagcc ccagctgctc cttctcctgt cgctcctgcc gcaggaagtg	60
acttcagccc ctactgggcc tcagctcttg gatgctggtc tctcccttct gaagtcattc	120
gtagccactc tggaccaagc tcctcagcgt tcctcagcc agtcacggtt ctctgcgttc	180
ctggccaaca tttcttcac cttccagctt gggaggatgg gggagggacc ggtgggagag	240
ccccacctc tccagcccc tgcatttga cttcatgatt tcctcgtgac actgagaggt	300
agcccagact gggagccaat gctagggctt ctgggagatg tgctggcact cctgggacag	360
gaacagactc cccgggactt tttggtgcac caggcaggtg tactgggtgg acttgtagag	420
gcattgttgg gagcgttagt tcctggaggc cccctgccc ccactcgacc cccatgcacc	480
cgtgatggcc cttctgactg tgcctggct gctgattggt tgccttctct gatgttgta	540
ttagagggtg cacgctggca ggccctgggt cagttgcagc ccagtgtgga cccaaccaat	600
gccacaggtc ttgatgtag agagccagct ctcactttt tacagggtct gctgggcttg	660
cttaccacag caggagagtt gggctctgag gaggtcttt ggggtggtct gctgcgcaca	720
gtggggggccc cctctatgc tgccttcag gaggggtac tgcgagtcac tcattctctg	780
caagatgagg tcttttctat tatgggacag ccagagcctg atgccagtgg gcagtgccag	840
ggaggcaacc ttcaacagct gcttttatgg ggcatgcgga acaacctttc ttgggacgcc	900
cgagcactgg gttttctatc tggatcacca cctccacccc ctgctctcct gcactgcctg	960
agcagaggtg tgcctctgcc cagggttcc cagcctgcgg ctcacatcag cctcgacag	1020
cggcgagcca tctctgtgga ggccctctgc gagaaccact caggcccaga gccaccctac	1080

agcatctcca acttctccat ctacttgctc tgccagcaca tcaagcctgc caccgccgg 1140

ccccctcta ccacccacg gctctctct accacccac agccccctcc taccactaca 1200

cagcccatc ctgacactac acagccccc cctgtcacc caaggcctcc tctaccacc 1260

ccacaacccc ctctagcac agctgtcatc tgccagacag ctgtatggt cgcagtctcg 1320

tgggcaccag gtgcccagg ttggctccaa gcctgccatg atcagtttcc tgatcaattt 1380

ctggatatga tctgcgcaa cctctcattt tcagccctgt ctggccccag tegtctttg 1440

glaaagcagc tctgtgctgg cttgtctcca cccccacta gctgtccacc aggcctgac 1500

cctgtgcccc tcacccaga aatattctgg ggtgtttcc tggagaatga gacactgtg 1560

gctgaacggt tgtgtgtgga ggacagtctg caggctgtgc ccccgaggaa ccaggcttgg 1620

gttcagcatg tgtgtcgggg cccaccttg gacgccactg attttccacc gtgccgcgtt 1680

ggacctgtg gggaacgtg ccagatggg ggagcttcc tgcctatggt ctgtgccaat 1740

gacactctgt atgaagcctt ggttccctc tgggcttggc tagcaggcca atgcagaatt 1800

agtcgtggag gaaatgatac ttgctttcta gaaggcatgc tgggccccctt gttgccctct 1860

ctgccccctc tgggaccatc cccactctgt ctggctctcg gtccttttct gcttggcatg 1920

ttatccaggt tgccacgtg tcagtcctcc gtgccagccc tcgccaccc caccgccta 1980

cattacctcc tgcgcctact gaccttctt ctgggtccag ggactggggg tgccgagacg 2040

caggggatgt taggtcaagc cctgtgtctc tctagtctcc cagacaactg ttcattctgg 2100

gatgccttcc gccagaggg ccggagaagt gtactagga cagtcggaga gtacttgacg 2160

cgggaagagc caacccacc aggccttagac tcttccctca gcctcggtc tggatgagc 2220

aagatggagc ttctgtcctg cttcagtcct gtactgtggg atctactcca gagagagaag 2280

agcgtttggg cctgaggac cctggtgaag gcctacctgc gcatgcctcc agaagacctt 2340

cagcagcttg tgccttcagc agagatggag gctgcacagg gcttctgac gctcatgctt 2400

cgttcctggg ctaagctgaa ggttcaacca tccgaggagc agccatggg ccgcctgaca 2460

gccttgctgc tccagcgta cccagcctc acctcccaac tctttatcga catgtcaccg 2520

ctcatccct tctggctgt cctgacctc atgcgttcc caccgtccct ttggccaac 2580

gacagtgtcc tggctgcat caggatcac agctcaggaa tgaagcctga acagaaggag 2640

gccttgcaa aacgactgt ggccccag ctgtttggag aagtgcctga ttggccccag 2700

gagctgctgt gggcagccct gcctctgctt cccatctgc ctctggagag ctttctccag 2760

ctcagccctc accagatcca ggccctggag gatagctggc cagtagcaga tcttgggccg 2820

ggacacgccc gacatgtgct tcttagccta gtaaaccaga gcatggagga tggggaggag 2880
 caggtgctca ggcttgggtc cctcgctgtt ttcttgagtc ctgaggagct acagagtctg 2940
 gtgcccttga gtgatccaat ggggcctgta gaacagggtc tgctggaatg tgcggccaat 3000
 gggaccctca gcccagaagg acgggtggca tatgaacttc tgggagtgtt gcgttcatct 3060
 ggaggaactg tcttaagccc ccgagagctg agggctctggg caccctctctt tccccagctg 3120
 ggctccgct tctgcagga gctctcagag acccagctta gagccatgct tctgccccta 3180
 caggagacca gtgtcacacc tgcccaggct gtctgttgt ttggaaggct ccttcctaag 3240

 catgatctgt ccttggagga actctgctcc ctgcaccctc tctgccagg tctcagcccc 3300
 cagacactcc aggccatccc taagagagtt ctgggttggtg cttgttctg cctgggcccct 3360
 gaactgtcaa ggctttcagc ttgccagatt gcagctctgc tgcagacctt tgggtaaaa 3420
 gatggtgtta aaaaataggg tgcagcaggt gccggctcag ccgtgtgcat tcttgggcag 3480
 cccaccactt ggccagactg cctgcttccc ctgctcccat taaagctgct acagctggac 3540
 gctgcagctc ttctggcaaa ccgaagactc tatcggcagc tgccttggtc tgagcaacag 3600
 gcacagtctc tctggaagaa aatgcaagtg cctaccaacc tgagcctgag gaatctgcag 3660

 gctctgggca acttggcagg aggcattgacc tgcgagtctc tgcagcagat cagctcaatg 3720
 gttgactctc ttgatgtgtt acacatgctc taccagctgc ccactggtgt tgcagagagc 3780
 ctgcccgcct gtatctggac agagctacag cggaggatga caatgccaga gccagagctg 3840
 accaccctag ggccagaact gactgaactt gacacaaagc tactcctgga cttgccgac 3900
 cagctgatgg acagattgtc caatgattcc attatgttgg tgggtggagat ggtccaaggc 3960
 gctccagagc agctgctggc actgacccca ctccaccaga cagccttggc agagcgagca 4020
 cttaaaaacc tggtccaaa ggagacccca atctccaaag aagtgtgga gacactgggc 4080

 ccttgggtg gattcctggg aatagagagc acgcgacgga tccctttacc cattctactg 4140
 tctcatctca gtcagctgca gggcttctgc ctaggagaga catttgccac agagctggga 4200
 tggctgctgt tgcaggagcc tgttcttggg aaaccagaat tgtggagcca ggtgaaata 4260
 gagcaagctg gacgcctagt attcactctg tctgctgagg ctatttctc gatccccagg 4320
 gaggctttgg gccagagac actggagagg ctcttgggaa agcatcaaag ctgggagcag 4380
 agcagagtgg gccatctgtg tggggagtca cagcttggcc acaagaaaag agctctggtg 4440
 gctgggattg tgcatccagc tgctgagggt ctccaagagc ctgtacaaa ctgtgcagac 4500

 atacggggaa ccttcccagc ggcctggtct gcgacacaaa tctcagagat ggaactctca 4560
 gactttgaag actgcctgtc actatttctg ggagatccag gacttgggtc tgaggaaacta 4620
 cgggcagcca tgggcaaggc caagcagttg tggggtcccc ctgaggatt ccgtcctgag 4680

cagatcttgc agctgggccc tctcctgata ggcttaggag aacgggaact gcaggagctt 4740
accttgggtg actgggggtg gctgagcagc ctggggcaaa tagatggctg gagttccatg 4800
cagctccgag ccgtggtctc cagtttctta aggcagagtg gtcggcatgt gagccacctg 4860
gacttcattt atctgacagc actgggttac acagtctgtg gattgcgacc agaggagtta 4920

cagcacatca gcagttggga gtttagccaa gcagctctct tctgggtag ctgcatctc 4980
ccgtgctctg aggaacagct ggaagttctg gcctatctcc ttgtgttgcc tgggtggcttt 5040
ggcccagtca gtaactgggg gcctgagatc ttactgaaa ttggcacaat agcagctggc 5100
atccagacc tggtcttttc agcattactg cggggacaga tccaaggcct gactcctctt 5160
gccatttctg tcattctgc tccaagttt gcagtggctt tcaaccccat ccagttatct 5220
agtctacca ggggtcaggc cgtagctgtt actcctgaac agctggccta tctgagtcct 5280
gagcagcggc gagcagttgc atgggcccac cacgaaggga aggagatccc agagcagctg 5340

ggtcgaaact cagcctgggg tctctacgac tggttccaag cctcctgggc cctggcattg 5400
cccgtcagca tttttggcca cctattaacg cgtacgcggc cgctcgagca gaaactcatc 5460
tcagaagagg atctgtaata g 5481

<210> 39

<211> 1825

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 39

Met Ala Leu Ser Leu Gln Pro Gln Leu Leu Leu Leu Ser Leu Leu

1 5 10 15

Pro Gln Glu Val Thr Ser Ala Pro Thr Gly Pro Gln Ser Leu Asp Ala

20 25 30

Gly Leu Ser Leu Leu Lys Ser Phe Val Ala Thr Leu Asp Gln Ala Pro

35 40 45

Gln Arg Ser Leu Ser Gln Ser Arg Phe Ser Ala Phe Leu Ala Asn Ile

50 55 60

Ser Ser Ser Phe Gln Leu Gly Arg Met Gly Glu Gly Pro Val Gly Glu

65 70 75 80

Pro Pro Pro Leu Gln Pro Pro Ala Leu Arg Leu His Asp Phe Leu Val
85 90 95

Thr Leu Arg Gly Ser Pro Asp Trp Glu Pro Met Leu Gly Leu Leu Gly
100 105 110

Asp Val Leu Ala Leu Leu Gly Gln Glu Gln Thr Pro Arg Asp Phe Leu
115 120 125

Val His Gln Ala Gly Val Leu Gly Gly Leu Val Glu Ala Leu Leu Gly
130 135 140

Ala Leu Val Pro Gly Gly Pro Pro Ala Pro Thr Arg Pro Pro Cys Thr
145 150 155 160

Arg Asp Gly Pro Ser Asp Cys Val Leu Ala Ala Asp Trp Leu Pro Ser
165 170 175

Leu Met Leu Leu Leu Glu Gly Thr Arg Trp Gln Ala Leu Val Gln Leu
180 185 190

Gln Pro Ser Val Asp Pro Thr Asn Ala Thr Gly Leu Asp Gly Arg Glu
195 200 205

Pro Ala Pro His Phe Leu Gln Gly Leu Leu Gly Leu Leu Thr Pro Ala
210 215 220

Gly Glu Leu Gly Ser Glu Glu Ala Leu Trp Gly Gly Leu Leu Arg Thr
225 230 235 240

Val Gly Ala Pro Leu Tyr Ala Ala Phe Gln Glu Gly Leu Leu Arg Val
245 250 255

Thr His Ser Leu Gln Asp Glu Val Phe Ser Ile Met Gly Gln Pro Glu
260 265 270

Pro Asp Ala Ser Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln Leu Leu
275 280 285

Leu Trp Gly Met Arg Asn Asn Leu Ser Trp Asp Ala Arg Ala Leu Gly
290 295 300

Phe Leu Ser Gly Ser Pro Pro Pro Pro Ala Leu Leu His Cys Leu
305 310 315 320

Ser Arg Gly Val Pro Leu Pro Arg Ala Ser Gln Pro Ala Ala His Ile

325 330 335
 Ser Pro Arg Gln Arg Arg Ala Ile Ser Val Glu Ala Leu Cys Glu Asn
 340 345 350
 His Ser Gly Pro Glu Pro Pro Tyr Ser Ile Ser Asn Phe Ser Ile Tyr
 355 360 365
 Leu Leu Cys Gln His Ile Lys Pro Ala Thr Pro Arg Pro Pro Pro Thr
 370 375 380
 Thr Pro Arg Pro Pro Pro Thr Thr Pro Gln Pro Pro Pro Thr Thr Thr
 385 390 395 400

 Gln Pro Ile Pro Asp Thr Thr Gln Pro Pro Pro Val Thr Pro Arg Pro
 405 410 415
 Pro Pro Thr Thr Pro Gln Pro Pro Pro Ser Thr Ala Val Ile Cys Gln
 420 425 430
 Thr Ala Val Trp Tyr Ala Val Ser Trp Ala Pro Gly Ala Arg Gly Trp
 435 440 445
 Leu Gln Ala Cys His Asp Gln Phe Pro Asp Gln Phe Leu Asp Met Ile
 450 455 460

 Cys Gly Asn Leu Ser Phe Ser Ala Leu Ser Gly Pro Ser Arg Pro Leu
 465 470 475 480
 Val Lys Gln Leu Cys Ala Gly Leu Leu Pro Pro Pro Thr Ser Cys Pro
 485 490 495
 Pro Gly Leu Ile Pro Val Pro Leu Thr Pro Glu Ile Phe Trp Gly Cys
 500 505 510
 Phe Leu Glu Asn Glu Thr Leu Trp Ala Glu Arg Leu Cys Val Glu Asp
 515 520 525

 Ser Leu Gln Ala Val Pro Pro Arg Asn Gln Ala Trp Val Gln His Val
 530 535 540
 Cys Arg Gly Pro Thr Leu Asp Ala Thr Asp Phe Pro Pro Cys Arg Val
 545 550 555 560
 Gly Pro Cys Gly Glu Arg Cys Pro Asp Gly Gly Ser Phe Leu Leu Met
 565 570 575

Val Cys Ala Asn Asp Thr Leu Tyr Glu Ala Leu Val Pro Phe Trp Ala
580 585 590

Trp Leu Ala Gly Gln Cys Arg Ile Ser Arg Gly Gly Asn Asp Thr Cys
595 600 605

Phe Leu Glu Gly Met Leu Gly Pro Leu Leu Pro Ser Leu Pro Pro Leu
610 615 620

Gly Pro Ser Pro Leu Cys Leu Ala Pro Gly Pro Phe Leu Leu Gly Met
625 630 635 640

Leu Ser Gln Leu Pro Arg Cys Gln Ser Ser Val Pro Ala Leu Ala His
645 650 655

Pro Thr Arg Leu His Tyr Leu Leu Arg Leu Leu Thr Phe Leu Leu Gly
660 665 670

Pro Gly Thr Gly Gly Ala Glu Thr Gln Gly Met Leu Gly Gln Ala Leu
675 680 685

Leu Leu Ser Ser Leu Pro Asp Asn Cys Ser Phe Trp Asp Ala Phe Arg
690 695 700

Pro Glu Gly Arg Arg Ser Val Leu Arg Thr Val Gly Glu Tyr Leu Gln
705 710 715 720

Arg Glu Glu Pro Thr Pro Pro Gly Leu Asp Ser Ser Leu Ser Leu Gly
725 730 735

Ser Gly Met Ser Lys Met Glu Leu Leu Ser Cys Phe Ser Pro Val Leu
740 745 750

Trp Asp Leu Leu Gln Arg Glu Lys Ser Val Trp Ala Leu Arg Thr Leu
755 760 765

Val Lys Ala Tyr Leu Arg Met Pro Pro Glu Asp Leu Gln Gln Leu Val
770 775 780

Leu Ser Ala Glu Met Glu Ala Ala Gln Gly Phe Leu Thr Leu Met Leu
785 790 795 800

Arg Ser Trp Ala Lys Leu Lys Val Gln Pro Ser Glu Glu Gln Ala Met
805 810 815

Gly Arg Leu Thr Ala Leu Leu Leu Gln Arg Tyr Pro Arg Leu Thr Ser

820	825	830
Gln Leu Phe Ile Asp Met Ser Pro Leu Ile Pro Phe Leu Ala Val Pro		
835	840	845
Asp Leu Met Arg Phe Pro Pro Ser Leu Leu Ala Asn Asp Ser Val Leu		
850	855	860
Ala Ala Ile Arg Asp His Ser Ser Gly Met Lys Pro Glu Gln Lys Glu		
865	870	875
Ala Leu Ala Lys Arg Leu Leu Ala Pro Glu Leu Phe Gly Glu Val Pro		
885	890	895
Asp Trp Pro Gln Glu Leu Leu Trp Ala Ala Leu Pro Leu Leu Pro His		
900	905	910
Leu Pro Leu Glu Ser Phe Leu Gln Leu Ser Pro His Gln Ile Gln Ala		
915	920	925
Leu Glu Asp Ser Trp Pro Val Ala Asp Leu Gly Pro Gly His Ala Arg		
930	935	940
His Val Leu Arg Ser Leu Val Asn Gln Ser Met Glu Asp Gly Glu Glu		
945	950	955
Gln Val Leu Arg Leu Gly Ser Leu Ala Cys Phe Leu Ser Pro Glu Glu		
965	970	975
Leu Gln Ser Leu Val Pro Leu Ser Asp Pro Met Gly Pro Val Glu Gln		
980	985	990
Gly Leu Leu Glu Cys Ala Ala Asn Gly Thr Leu Ser Pro Glu Gly Arg		
995	1000	1005
Val Ala Tyr Glu Leu Leu Gly Val Leu Arg Ser Ser Gly Gly Thr		
1010	1015	1020
Val Leu Ser Pro Arg Glu Leu Arg Val Trp Ala Pro Leu Phe Pro		
1025	1030	1035
Gln Leu Gly Leu Arg Phe Leu Gln Glu Leu Ser Glu Thr Gln Leu		
1040	1045	1050
Arg Ala Met Leu Pro Ala Leu Gln Gly Ala Ser Val Thr Pro Ala		
1055	1060	1065

Gln Ala Val Leu Leu Phe Gly Arg Leu Leu Pro Lys His Asp Leu		
1070	1075	1080
Ser Leu Glu Glu Leu Cys Ser Leu His Pro Leu Leu Pro Gly Leu		
1085	1090	1095
Ser Pro Gln Thr Leu Gln Ala Ile Pro Lys Arg Val Leu Val Gly		
1100	1105	1110
Ala Cys Ser Cys Leu Gly Pro Glu Leu Ser Arg Leu Ser Ala Cys		
1115	1120	1125
Gln Ile Ala Ala Leu Leu Gln Thr Phe Arg Val Lys Asp Gly Val		
1130	1135	1140
Lys Asn Met Gly Ala Ala Gly Ala Gly Ser Ala Val Cys Ile Pro		
1145	1150	1155
Gly Gln Pro Thr Thr Trp Pro Asp Cys Leu Leu Pro Leu Leu Pro		
1160	1165	1170
Leu Lys Leu Leu Gln Leu Asp Ala Ala Ala Leu Leu Ala Asn Arg		
1175	1180	1185
Arg Leu Tyr Arg Gln Leu Pro Trp Ser Glu Gln Gln Ala Gln Phe		
1190	1195	1200
Leu Trp Lys Lys Met Gln Val Pro Thr Asn Leu Ser Leu Arg Asn		
1205	1210	1215
Leu Gln Ala Leu Gly Asn Leu Ala Gly Gly Met Thr Cys Glu Phe		
1220	1225	1230
Leu Gln Gln Ile Ser Ser Met Val Asp Phe Leu Asp Val Val His		
1235	1240	1245
Met Leu Tyr Gln Leu Pro Thr Gly Val Arg Glu Ser Leu Arg Ala		
1250	1255	1260
Cys Ile Trp Thr Glu Leu Gln Arg Arg Met Thr Met Pro Glu Pro		
1265	1270	1275
Glu Leu Thr Thr Leu Gly Pro Glu Leu Ser Glu Leu Asp Thr Lys		
1280	1285	1290
Leu Leu Leu Asp Leu Pro Ile Gln Leu Met Asp Arg Leu Ser Asn		

1295	1300	1305
Asp Ser Ile Met Leu Val Val	Glu Met Val Gln Gly	Ala Pro Glu
1310	1315	1320
Gln Leu Leu Ala Leu Thr Pro	Leu His Gln Thr Ala	Leu Ala Glu
1325	1330	1335
Arg Ala Leu Lys Asn Leu Ala	Pro Lys Glu Thr Pro	Ile Ser Lys
1340	1345	1350
Glu Val Leu Glu Thr Leu Gly	Pro Leu Val Gly Phe	Leu Gly Ile
1355	1360	1365
Glu Ser Thr Arg Arg Ile Pro	Leu Pro Ile Leu Leu	Ser His Leu
1370	1375	1380
Ser Gln Leu Gln Gly Phe Cys	Leu Gly Glu Thr Phe	Ala Thr Glu
1385	1390	1395
Leu Gly Trp Leu Leu Leu Gln	Glu Pro Val Leu Gly	Lys Pro Glu
1400	1405	1410
Leu Trp Ser Gln Asp Glu Ile	Glu Gln Ala Gly Arg	Leu Val Phe
1415	1420	1425
Thr Leu Ser Ala Glu Ala Ile	Ser Ser Ile Pro Arg	Glu Ala Leu
1430	1435	1440
Gly Pro Glu Thr Leu Glu Arg	Leu Leu Gly Lys His	Gln Ser Trp
1445	1450	1455
Glu Gln Ser Arg Val Gly His	Leu Cys Gly Glu Ser	Gln Leu Ala
1460	1465	1470
His Lys Lys Ala Ala Leu Val	Ala Gly Ile Val His	Pro Ala Ala
1475	1480	1485
Glu Gly Leu Gln Glu Pro Val	Pro Asn Cys Ala Asp	Ile Arg Gly
1490	1495	1500
Thr Phe Pro Ala Ala Trp Ser	Ala Thr Gln Ile Ser	Glu Met Glu
1505	1510	1515
Leu Ser Asp Phe Glu Asp Cys	Leu Ser Leu Phe Ala	Gly Asp Pro
1520	1525	1530

Gly Leu	Gly Pro Glu Glu Leu	Arg Ala Ala Met Gly	Lys Ala Lys
1535	1540	1545	
Gln Leu	Trp Gly Pro Pro Arg	Gly Phe Arg Pro Glu	Gln Ile Leu
1550	1555	1560	
Gln Leu	Gly Arg Leu Leu Ile	Gly Leu Gly Glu Arg	Glu Leu Gln
1565	1570	1575	
Glu Leu	Thr Leu Val Asp Trp	Gly Val Leu Ser Ser	Leu Gly Gln
1580	1585	1590	
Ile Asp	Gly Trp Ser Ser Met	Gln Leu Arg Ala Val	Val Ser Ser
1595	1600	1605	
Phe Leu	Arg Gln Ser Gly Arg	His Val Ser His Leu	Asp Phe Ile
1610	1615	1620	
Tyr Leu	Thr Ala Leu Gly Tyr	Thr Val Cys Gly Leu	Arg Pro Glu
1625	1630	1635	
Glu Leu	Gln His Ile Ser Ser	Trp Glu Phe Ser Gln	Ala Ala Leu
1640	1645	1650	
Phe Leu	Gly Ser Leu His Leu	Pro Cys Ser Glu Glu	Gln Leu Glu
1655	1660	1665	
Val Leu	Ala Tyr Leu Leu Val	Leu Pro Gly Gly Phe	Gly Pro Val
1670	1675	1680	
Ser Asn	Trp Gly Pro Glu Ile	Phe Thr Glu Ile Gly	Thr Ile Ala
1685	1690	1695	
Ala Gly	Ile Pro Asp Leu Ala	Leu Ser Ala Leu Leu	Arg Gly Gln
1700	1705	1710	
Ile Gln	Gly Leu Thr Pro Leu	Ala Ile Ser Val Ile	Pro Ala Pro
1715	1720	1725	
Lys Phe	Ala Val Val Phe Asn	Pro Ile Gln Leu Ser	Ser Leu Thr
1730	1735	1740	
Arg Gly	Gln Ala Val Ala Val	Thr Pro Glu Gln Leu	Ala Tyr Leu
1745	1750	1755	
Ser Pro	Glu Gln Arg Arg Ala	Val Ala Trp Ala Gln	His Glu Gly

1760	1765	1770	
Lys Glu Ile Pro Glu Gln Leu	Gly Arg Asn Ser Ala	Trp Gly Leu	
1775	1780	1785	
Tyr Asp Trp Phe Gln Ala Ser	Trp Ala Leu Ala Leu	Pro Val Ser	
1790	1795	1800	
Ile Phe Gly His Leu Leu Thr	Arg Thr Arg Pro Leu	Glu Gln Lys	
1805	1810	1815	
Leu Ile Ser Glu Glu Asp Leu			
1820	1825		
<210> 40			
<211> 60			
<212> DNA			
<213> Artificial Sequence			
<220><223> Description of Artificial Sequence: Synthetic			
oligonucleotide			
<400> 40			
gctctcagcc tctggccct gctgctgctg ctgctgctgc tgcctgtgct gtcctttgca	60		
<210> 41			
<211> 63			
<212> DNA			
<213> Artificial Sequence			
<220><223> Description of Artificial Sequence: Synthetic			
oligonucleotide			
<400> 41			
gctctgagcc tccagcccca gctgctcctt ctctgtgcgc tctgcccga ggaagtgact	60		
tca		63	
<210> 42			
<211> 312			
<212> DNA			
<213> Artificial Sequence			
<220><223> Description of Artificial Sequence: Synthetic			
polynucleotide			
<400> 42			

tgcctgtcat acgaaaccga gatactgaca gtagaatatg gccttctgcc aatcggaag 60
attgtggaga aacggataga atgcacagtt tactctgtcg ataacaatgg taacatttat 120
actcagccag ttgccagtgc gcacgaccgg ggagagcagg aagtattcga atactgtctg 180
gaggatggaa gtctcattag ggccactaag gaccacaaat ttatgacagt cgatggccag 240

atgtgccta tagacgaaat ctttgagcga gagttggacc tcatgagagt tgacaacctt 300
cctaattaat ag 312

<210> 43

<211> 20

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 43

Ala Leu Ser Leu Trp Pro Leu Leu Leu Leu Leu Leu Leu Leu Leu

1 5 10 15

Leu Ser Phe Ala

20

<210> 44

<211> 21

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 44

Ala Leu Ser Leu Gln Pro Gln Leu Leu Leu Leu Ser Leu Leu Pro

1 5 10 15

Gln Glu Val Thr Ser

20

<210> 45

<211> 102

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 45

Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu

1 5 10 15

Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser

20 25 30

Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His

35 40 45

Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser

50 55 60

Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln

65 70 75 80

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

85 90 95

Val Asp Asn Leu Pro Asn

100

<210> 46

<211> 105

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 46

atcaagatag ctacaaggaa gtatcttggc aaacaaaacg tttatgatat tggagtcgaa 60

agagatcaca actttgctct gaagaacgga ttcatagctt ctaat 105

<210> 47

<211> 18

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 47

acgcgtacgc ggccgctc 18

<210> 48

<211> 36

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 48

gagcagaaac tcatctcaga agaggatctg taatag 36

<210> 49

<211> 35

<212> PRT

<213> Artificial Sequence

<220><223>

> Description of Artificial Sequence: Synthetic

polypeptide

<400> 49

Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn Val Tyr Asp

1 5 10 15

Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn Gly Phe Ile

20 25 30

Ala Ser Asn

35

<210> 50

<211> 6

<212> PRT

<213> Artificial Sequence

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

peptide

<

400> 50

Thr Arg Thr Arg Pro Leu

1 5

<210> 51

<211> 3219

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 51

atggctctga gctccagcc ccagctgctc cttctctgt cgctcctgcc gcaggaagt	60
acttcagccc ctactgggcc tcagctcttg gatgctggtc tctcccttct gaagtcattc	120
gtagccactc tggaccaagc tctcagcgt tcctcagcc agtcacggtt ctctgcgttc	180
ctggccaaca tttcttcac cttccagctt gggaggatgg gggagggacc ggtgggagag	240
ccccacctc tccagcccc tgcacttga cttcatgatt tcctcgtgac actgagaggt	300
agcccagact gggagccaat gctagggctt ctgggagatg tgctggcact cctgggacag	360
gaacagactc cccgggactt tttggtgcac caggcaggtg tactgggtgg acttgtagag	420
gcattgttgg gagcgttagt tcctggaggc cccctgccc cactcgacc cccatgcacc	480
cgtgatggcc cttctgactg tgctctggct gctgattggt tgccttctct gatgttgta	540
ttagagggtg cacgctggca ggccctgggt cagttgcagc ccagtgtgga cccaaccaat	600
gccacaggtc ttgatgtag agagccagct cctcactttt tacagggtct gctgggcttg	660
cttaccag caggagagtt gggctctgag gaggtctttt ggggtggtct gctgcgcaca	720
gtggggggccc cctctatgc tgccttcag gaggggctac tgcgagtcac tcattctctg	780
caagatgagg tcttttctat tatgggacag ccagagcctg atgccagtgg gcagtgccag	840
ggaggcaacc ttcaacagct gcttttatgg ggcattgcga acaaccttc tgggacgcc	900
cgagcactgg gttttctatc tggatcacca cctccacccc ctgctctcct gactgcctg	960
agcagagggtg tgcctctgcc cagggttcc cagcctgcgg ctcacatcag cctcgacag	1020
cggcgagcca tctctgtgga ggccctctgc gagaaccact caggcccaga gccacctac	1080
agcatctcca acttctccat ctacttctc tgccagcaca tcaagcctgc cccccgcgg	1140
ccccctcta ccacccacg gcctctctc accacccac agccccctcc taccactaca	1200
cagccattc ctgacactac acagccccct cctgtcacc caaggcctcc tctaccacc	1260
ccacaacccc ctctagcac agctgtcctc tgccagacag ctgtatggtg cgcagtctcg	1320
tgggcaccag gtgcccaggg ttggtccaa gcctgccatg atcagtttcc tgatcaattt	1380
ctggatatga tctgcggcaa cctctcattt tcagccctgt ctggccccag tcgtcctttg	1440
gtaaagcagc tctgtgctgg cttgtctcca cccccacta gctgtccacc aggctgatc	1500
cctgtgcccc tcacccaga aatattctgg ggctgtttcc tggagaatga gacactgtgg	1560

gctgaacggt tgtgtgtgga ggacagtctg caggctgtgc ccccaggaa ccaggcttgg 1620
gttcagcatg tgtgtcgggg cccacacttg gacgccactg attttccacc gtgccgcgtt 1680
ggacctgtg gggaacgctg cccagatggg ggcagcttcc tgctcatggt ctgtgccaat 1740
gacactctgt atgaagcctt gggtcccttc tgggcttggc tagcaggcca atgcagaatt 1800
agtcgtggag gaaatgatac ttgctttcta gaaggcatgc tgggccccctt gttgccctct 1860
ctgccccctc tgggaccatc cccactctgt ctggtcctg gtctttttct gcttggcatg 1920

ttatccaggt tgccacgctg tcagtccctc gtgccagccc tcgcccacc cagcgccta 1980
cattacctcc tgcgcctact gaccttctt ctgggtccag ggactggggg tgccgagacg 2040
caggggatgt taggtcaagc cctgtctctc tctagtctcc cagacaactg ttcattctgg 2100
gatgccttcc gcccagaggg ccggagaagt gtactgagga cagtcggaga gtacttgacg 2160
cgggaagagc caaccccc accgcttagac tctccctca gcctcggtc tggatatgac 2220
aagatggagc ttctgtcctg cttcagtcct gtactgtggg atctactcca gagagagaag 2280
agcgtttggg cctgaggac cctggtgaag gcctacctgc gcatgcctcc agaagacctt 2340

cagcagcttg tgctttcagc agagatggag gctgcacagg gcttccctgac gctcatgctt 2400
cgttctctgg ctaagctgaa ggttcaacca tccgaggagc aggccatggg ccgcctgaca 2460
gccttctctg tccagcggtc cccacgctc acctcccaac tctttatcga catgtcaccg 2520
ctcatccctc tctggctgt cctgacctc atgcgcttcc caccgtccct ttgggccaac 2580
gacagtgtcc tggctgccat cagggatcac agctcaggaa tgaagcctga acagaaggag 2640
gccctggcaa aacgactgct ggccccctgag ctgtttggag aagtgcctga ttggccccag 2700
gagctgctgt gggcagccct gcctctgctt ccccatctgc ctctggagag ctttctccag 2760

ctcagccctc accagatcca ggccctggag gatagctggc cagtagcaga tcttgggccg 2820
ggacacgccc gacatgtgct tcgtagccta gtaaaccaga gcatggagga tggggaggag 2880
caggtgctca ggcttgggtc cctcgctctc ctgtcatacg aaaccgagat actgacagta 2940
gaatatggcc ttctgccaat cgggaagatt gtggagaaac ggatagaatg cacagtttac 3000
tctgtcgata acaatggtaa catttatact cagccagtig cccagtggca cgaccgggga 3060
gagcaggaag tattcgaata ctgtctggag gatggaagtc tcattagggc cactaaggac 3120
cacaaattta tgacagtca tggccagatg ctgcctatag acgaaatctt tgagcgagag 3180

ttggacctca tgcgagttga caaccttcct aattaatag 3219

<210> 52

<211> 1071

<212> PRT

<213> Artificial Sequence

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

<400> 52

Met Ala Leu Ser Leu Gln Pro Gln Leu Leu Leu Leu Ser Leu Leu

1 5 10 15

Pro Gln Glu Val Thr Ser Ala Pro Thr Gly Pro Gln Ser Leu Asp Ala

20 25 30

Gly Leu Ser Leu Leu Lys Ser Phe Val Ala Thr Leu Asp Gln Ala Pro

35 40 45

Gln Arg Ser Leu Ser Gln Ser Arg Phe Ser Ala Phe Leu Ala Asn Ile

50 55 60

Ser Ser Ser Phe Gln Leu Gly Arg Met Gly Glu Gly Pro Val Gly Glu

65 70 75 80

Pro Pro Pro Leu Gln Pro Pro Ala Leu Arg Leu His Asp Phe Leu Val

85 90 95

Thr Leu Arg Gly Ser Pro Asp Trp Glu Pro Met Leu Gly Leu Leu Gly

100 105 110

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

115 120 125

Val His Gln Ala Gly Val Leu Gly Gly Leu Val Glu Ala Leu Leu Gly

130 135 140

Ala Leu Val Pro Gly Gly Pro Pro Ala Pro Thr Arg Pro Pro Cys Thr

145 150 155 160

Arg Asp Gly Pro Ser Asp Cys Val Leu Ala Ala Asp Trp Leu Pro Ser

165 170 175

Leu Met Leu Leu Leu Glu Gly Thr Arg Trp Gln Ala Leu Val Gln Leu

180 185 190

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

195 200 205

Pro Ala Pro His Phe Leu Gln Gly Leu Leu Gly Leu Leu Thr Pro Ala

210 215 220

Gly Glu Leu Gly Ser Glu Glu Ala Leu Trp Gly Gly Leu Leu Arg Thr

225 230 235 240

Val Gly Ala Pro Leu Tyr Ala Ala Phe Gln Glu Gly Leu Leu Arg Val

245 250 255

Thr His Ser Leu Gln Asp Glu Val Phe Ser Ile Met Gly Gln Pro Glu

260 265 270

Pro Asp Ala Ser Gly Gln Cys Gln Gly Gly Asn Leu Gln Gln Leu Leu

275 280 285

Leu Trp Gly Met Arg Asn Asn Leu Ser Trp Asp Ala Arg Ala Leu Gly

290 295 300

Phe Leu Ser Gly Ser Pro Pro Pro Pro Pro Ala Leu Leu His Cys Leu

305 310 315 320

Ser Arg Gly Val Pro Leu Pro Arg Ala Ser Gln Pro Ala Ala His Ile

325 330 335

Ser Pro Arg Gln Arg Arg Ala Ile Ser Val Glu Ala Leu Cys Glu Asn

340 345 350

His Ser Gly Pro Glu Pro Pro Tyr Ser Ile Ser Asn Phe Ser Ile Tyr

355 360 365

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

370 375 380

Thr Pro Arg Pro Pro Pro Thr Thr Pro Gln Pro Pro Pro Thr Thr Thr

385 390 395 400

Gln Pro Ile Pro Asp Thr Thr Gln Pro Pro Pro Val Thr Pro Arg Pro

405 410 415

Pro Pro Thr Thr Pro Gln Pro Pro Pro Ser Thr Ala Val Ile Cys Gln

420 425 430

Thr Ala Val Trp Tyr Ala Val Ser Trp Ala Pro Gly Ala Arg Gly Trp

435 440 445

Leu Gln Ala Cys His Asp Gln Phe Pro Asp Gln Phe Leu Asp Met Ile

450 455 460

Cys Gly Asn Leu Ser Phe Ser Ala Leu Ser Gly Pro Ser Arg Pro Leu

465 470 475 480
 Val Lys Gln Leu Cys Ala Gly Leu Leu Pro Pro Pro Thr Ser Cys Pro

 485 490 495
 Pro Gly Leu Ile Pro Val Pro Leu Thr Pro Glu Ile Phe Trp Gly Cys
 500 505 510
 Phe Leu Glu Asn Glu Thr Leu Trp Ala Glu Arg Leu Cys Val Glu Asp
 515 520 525
 Ser Leu Gln Ala Val Pro Pro Arg Asn Gln Ala Trp Val Gln His Val
 530 535 540
 Cys Arg Gly Pro Thr Leu Asp Ala Thr Asp Phe Pro Pro Cys Arg Val

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

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

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

Arg Glu Glu Pro Thr Pro Pro Gly Leu Asp Ser Ser Leu Ser Leu Gly
725 730 735

Ser Gly Met Ser Lys Met Glu Leu Leu Ser Cys Phe Ser Pro Val Leu
740 745 750

Trp Asp Leu Leu Gln Arg Glu Lys Ser Val Trp Ala Leu Arg Thr Leu
755 760 765

Val Lys Ala Tyr Leu Arg Met Pro Pro Glu Asp Leu Gln Gln Leu Val
770 775 780

Leu Ser Ala Glu Met Glu Ala Ala Gln Gly Phe Leu Thr Leu Met Leu
785 790 795 800

Arg Ser Trp Ala Lys Leu Lys Val Gln Pro Ser Glu Glu Gln Ala Met
805 810 815

Gly Arg Leu Thr Ala Leu Leu Leu Gln Arg Tyr Pro Arg Leu Thr Ser
820 825 830

Gln Leu Phe Ile Asp Met Ser Pro Leu Ile Pro Phe Leu Ala Val Pro
835 840 845

Asp Leu Met Arg Phe Pro Pro Ser Leu Leu Ala Asn Asp Ser Val Leu
850 855 860

Ala Ala Ile Arg Asp His Ser Ser Gly Met Lys Pro Glu Gln Lys Glu
865 870 875 880

Ala Leu Ala Lys Arg Leu Leu Ala Pro Glu Leu Phe Gly Glu Val Pro
885 890 895

Asp Trp Pro Gln Glu Leu Leu Trp Ala Ala Leu Pro Leu Leu Pro His
900 905 910

Leu Pro Leu Glu Ser Phe Leu Gln Leu Ser Pro His Gln Ile Gln Ala
915 920 925

Leu Glu Asp Ser Trp Pro Val Ala Asp Leu Gly Pro Gly His Ala Arg
930 935 940

His Val Leu Arg Ser Leu Val Asn Gln Ser Met Glu Asp Gly Glu Glu
945 950 955 960

Gln Val Leu Arg Leu Gly Ser Leu Ala Cys Leu Ser Tyr Glu Thr Glu

965 970 975
 Ile Leu Thr Val Glu Tyr Gly Leu Leu Pro Ile Gly Lys Ile Val Glu
 980 985 990
 Lys Arg Ile Glu Cys Thr Val Tyr Ser Val Asp Asn Asn Gly Asn Ile

995 1000 1005
 Tyr Thr Gln Pro Val Ala Gln Trp His Asp Arg Gly Glu Gln Glu
 1010 1015 1020
 Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser Leu Ile Arg Ala Thr
 1025 1030 1035
 Lys Asp His Lys Phe Met Thr Val Asp Gly Gln Met Leu Pro Ile
 1040 1045 1050
 Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu Met Arg Val Asp Asn

1055 1060 1065
 Leu Pro Asn
 1070
 <210> 53
 <211> 2745
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 polynucleotide

<400> 53
 atggctctga gcctccagcc ccagctgctc cttctcctgt cgctcctgcc gcaggaagtg 60
 acttcaatca agatagctac aaggaagtat cttggcaaac aaaacgttta tgatattgga 120
 gtcgaaagag atcacaactt tgctctgaag aacggattca tagcttctaa ttgtttcctg 180
 agtcctgagg agctacagag tctggtgccc ttgagtgatc caatggggcc tgtagaacag 240

 ggtctgctgg aatgtgcggc caatgggacc ctacagcccag aaggacgggt ggcataatgaa 300
 cttctgggag tgttgcttc atctggagga actgtcttaa gccccgaga gctgagggtc 360
 tgggcacctc tctttccca gctgggcctc cgcttcctgc aggagctctc agagaccag 420
 cttagagcca tgcttctgc cctacaggga gccagtgtca cacctgccca ggctgttctg 480
 ttgtttggaa ggctccttcc taagcatgat ctgtccctgg aggaactctg ctccctgcac 540
 cctctcctgc caggtctcag cccccagaca ctccaggcca tcctaagag agttctggtt 600

ggtgcttgtt cctgcctggg ccctgaactg tcaaggcttt cagcttgcca gattgcagct 660

 ctgctgcaga cctttcgggt aaaagatggt gttaaaaata tgggtgcagc aggtgccggc 720
 tcagccgtgt gcattcctgg gcagcccacc acttggccag actgcctgct tcccctgctc 780
 ccattaaagc tgctacagct ggacgctgca gctcttctgg caaacgaag actctatcgg 840
 cagctgcctt ggtctgagca acaggcacag tttctctgga agaaaatgca agtgcctacc 900
 aacctgagcc tgaggaatct gcaggctctg ggcaacttgg caggaggcat gacctgcgag 960
 tttctgcagc agatcagctc aatggttgac tttcttgatg tggtagacat gctctaccag 1020
 ctgcccactg gtgttcgaga gagcctgcgg gcctgtatct ggacagagct acagcggagg 1080

 atgacaatgc cagagccaga gctgaccacc ctagggccag aactgagtga acttgacaca 1140
 aagctactcc tggacttgcc gatccagctg atggacagat tgtccaatga ttccattatg 1200
 ttggtggtgg agatggtcca aggcgctcca gagcagctgc tggcactgac cccactccac 1260
 cagacagcct tggcagagcg agcacttaaa aacctggctc caaaggagac cccaatctcc 1320
 aaagaagtgc tggagacact gggcccttgg gttggattcc tgggaataga gagcacgcga 1380
 cggatccctt taccattctt actgtctcat ctcagtcagc tgcagggtct ctgcctagga 1440
 gagacatttg ccacagagct gggatggctg ctgttgcagg agcctgttct tggaaaacca 1500

 gaattgtgga gccaggatga aatagagcaa gctggacgcc tagtattcac tctgtctgct 1560
 gaggtctattt cctcgatccc cagggaggct ttgggccag agacactgga gaggttctg 1620
 ggaaagcatc aaagctggga gcagagcaga gtgggccatc tgtgtgggga gtcacagctt 1680
 gcccacaaga aagcagctct ggtagctggg attgtgcatc cagctgctga gggctctcaa 1740
 gagcctgtac caactgtgc agacatacgg ggaaccttcc cagcggcctg gtctgcgaca 1800
 caaatctcag agatggaact ctcaagcttt gaagactgcc tgtcactatt tgctggagat 1860
 ccaggacttg gtctgagga actacgggca gccatgggca aggccaagca gttgtggggt 1920

 cccctcgag gattccgtcc tgagcagatc ttgcagctgg gccgtctcct gataggtcta 1980
 ggagaacggg aactgcagga gcttaccttg gtggactggg gtgtgctgag cagcctgggg 2040
 caaatagatg gctggagttc catgcagctc cgagccgtgg tctccagttt cctaaggcag 2100
 agtggctggc atgtgagcca cctggacttc atttatctga cagcactggg ttacacagtc 2160
 tgtggattgc gaccagagga gttacagcac atcagcagtt gggagttag ccaagcagct 2220
 ctcttcctgg gtagcttgca tctccctgct tctgaggaac agctggaagt tctggcctat 2280
 ctcttctgtg tgcttggtgg ctttggccca gtcagtaact gggggcctga gatcttact 2340

gaaattggca caatagcagc tggcatccca gacctggctc tttcagcatt actgcgggga 2400
cagatccaag gcctgactcc tcttgccatt tctgtcattc ctgctcccaa gtttgagtg 2460
gtcttcaacc ccatccagtt atctagtctc accaggggctc aggccgtagc tgttactcct 2520
gaacagctgg cctatctgag tcctgagcag cggcgagcag ttgcatgggc ccaacacgaa 2580
gggaaggaga tcccagagca gctgggtcga aactcagcct ggggtctcta cgactggttc 2640
caagcctcct ggccctggc attgcccgtc agcatttttg gccacctatt aacgcgtacg 2700
cggccgctcg agcagaaact catctcagaa gaggatctgt aatag 2745

<210> 54

<211> 913

<212> PRT

<213> Artificial Sequence

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

polypeptide

<400> 54

Met Ala Leu Ser Leu Gln Pro Gln Leu Leu Leu Leu Ser Leu Leu

1 5 10 15

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

20 25 30

Lys Gln Asn Val Tyr Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala

35 40 45

Leu Lys Asn Gly Phe Ile Ala Ser Asn Cys Phe Leu Ser Pro Glu Glu

50 55 60

Leu Gln Ser Leu Val Pro Leu Ser Asp Pro Met Gly Pro Val Glu Gln

65 70 75 80

Gly Leu Leu Glu Cys Ala Ala Asn Gly Thr Leu Ser Pro Glu Gly Arg

85 90 95

Val Ala Tyr Glu Leu Leu Gly Val Leu Arg Ser Ser Gly Gly Thr Val

100 105 110

Leu Ser Pro Arg Glu Leu Arg Val Trp Ala Pro Leu Phe Pro Gln Leu

115 120 125

Gly Leu Arg Phe Leu Gln Glu Leu Ser Glu Thr Gln Leu Arg Ala Met

130 135 140

Leu Pro Ala Leu Gln Gly Ala Ser Val Thr Pro Ala Gln Ala Val Leu
 145 150 155 160
 Leu Phe Gly Arg Leu Leu Pro Lys His Asp Leu Ser Leu Glu Glu Leu
 165 170 175

 Cys Ser Leu His Pro Leu Leu Pro Gly Leu Ser Pro Gln Thr Leu Gln
 180 185 190
 Ala Ile Pro Lys Arg Val Leu Val Gly Ala Cys Ser Cys Leu Gly Pro
 195 200 205
 Glu Leu Ser Arg Leu Ser Ala Cys Gln Ile Ala Ala Leu Leu Gln Thr
 210 215 220
 Phe Arg Val Lys Asp Gly Val Lys Asn Met Gly Ala Ala Gly Ala Gly
 225 230 235 240

 Ser Ala Val Cys Ile Pro Gly Gln Pro Thr Thr Trp Pro Asp Cys Leu
 245 250 255
 Leu Pro Leu Leu Pro Leu Lys Leu Leu Gln Leu Asp Ala Ala Ala Leu
 260 265 270
 Leu Ala Asn Arg Arg Leu Tyr Arg Gln Leu Pro Trp Ser Glu Gln Gln
 275 280 285
 Ala Gln Phe Leu Trp Lys Lys Met Gln Val Pro Thr Asn Leu Ser Leu
 290 295 300

 Arg Asn Leu Gln Ala Leu Gly Asn Leu Ala Gly Gly Met Thr Cys Glu
 305 310 315 320
 Phe Leu Gln Gln Ile Ser Ser Met Val Asp Phe Leu Asp Val Val His
 325 330 335
 Met Leu Tyr Gln Leu Pro Thr Gly Val Arg Glu Ser Leu Arg Ala Cys
 340 345 350
 Ile Trp Thr Glu Leu Gln Arg Arg Met Thr Met Pro Glu Pro Glu Leu
 355 360 365

 Thr Thr Leu Gly Pro Glu Leu Ser Glu Leu Asp Thr Lys Leu Leu Leu
 370 375 380
 Asp Leu Pro Ile Gln Leu Met Asp Arg Leu Ser Asn Asp Ser Ile Met

385 390 395 400
 Leu Val Val Glu Met Val Gln Gly Ala Pro Glu Gln Leu Leu Ala Leu
 405 410 415
 Thr Pro Leu His Gln Thr Ala Leu Ala Glu Arg Ala Leu Lys Asn Leu
 420 425 430

 Ala Pro Lys Glu Thr Pro Ile Ser Lys Glu Val Leu Glu Thr Leu Gly
 435 440 445
 Pro Leu Val Gly Phe Leu Gly Ile Glu Ser Thr Arg Arg Ile Pro Leu
 450 455 460
 Pro Ile Leu Leu Ser His Leu Ser Gln Leu Gln Gly Phe Cys Leu Gly
 465 470 475 480
 Glu Thr Phe Ala Thr Glu Leu Gly Trp Leu Leu Leu Gln Glu Pro Val
 485 490 495

 Leu Gly Lys Pro Glu Leu Trp Ser Gln Asp Glu Ile Glu Gln Ala Gly
 500 505 510
 Arg Leu Val Phe Thr Leu Ser Ala Glu Ala Ile Ser Ser Ile Pro Arg
 515 520 525
 Glu Ala Leu Gly Pro Glu Thr Leu Glu Arg Leu Leu Gly Lys His Gln
 530 535 540
 Ser Trp Glu Gln Ser Arg Val Gly His Leu Cys Gly Glu Ser Gln Leu
 545 550 555 560

 Ala His Lys Lys Ala Ala Leu Val Ala Gly Ile Val His Pro Ala Ala
 565 570 575
 Glu Gly Leu Gln Glu Pro Val Pro Asn Cys Ala Asp Ile Arg Gly Thr
 580 585 590
 Phe Pro Ala Ala Trp Ser Ala Thr Gln Ile Ser Glu Met Glu Leu Ser
 595 600 605
 Asp Phe Glu Asp Cys Leu Ser Leu Phe Ala Gly Asp Pro Gly Leu Gly
 610 615 620

 Pro Glu Glu Leu Arg Ala Ala Met Gly Lys Ala Lys Gln Leu Trp Gly
 625 630 635 640

Pro Pro Arg Gly Phe Arg Pro Glu Gln Ile Leu Gln Leu Gly Arg Leu
645 650 655
Leu Ile Gly Leu Gly Glu Arg Glu Leu Gln Glu Leu Thr Leu Val Asp
660 665 670
Trp Gly Val Leu Ser Ser Leu Gly Gln Ile Asp Gly Trp Ser Ser Met
675 680 685

Gln Leu Arg Ala Val Val Ser Ser Phe Leu Arg Gln Ser Gly Arg His
690 695 700
Val Ser His Leu Asp Phe Ile Tyr Leu Thr Ala Leu Gly Tyr Thr Val
705 710 715 720
Cys Gly Leu Arg Pro Glu Glu Leu Gln His Ile Ser Ser Trp Glu Phe
725 730 735
Ser Gln Ala Ala Leu Phe Leu Gly Ser Leu His Leu Pro Cys Ser Glu
740 745 750

Glu Gln Leu Glu Val Leu Ala Tyr Leu Leu Val Leu Pro Gly Gly Phe
755 760 765
Gly Pro Val Ser Asn Trp Gly Pro Glu Ile Phe Thr Glu Ile Gly Thr
770 775 780
Ile Ala Ala Gly Ile Pro Asp Leu Ala Leu Ser Ala Leu Leu Arg Gly
785 790 795 800
Gln Ile Gln Gly Leu Thr Pro Leu Ala Ile Ser Val Ile Pro Ala Pro
805 810 815

Lys Phe Ala Val Val Phe Asn Pro Ile Gln Leu Ser Ser Leu Thr Arg
820 825 830
Gly Gln Ala Val Ala Val Thr Pro Glu Gln Leu Ala Tyr Leu Ser Pro
835 840 845
Glu Gln Arg Arg Ala Val Ala Trp Ala Gln His Glu Gly Lys Glu Ile
850 855 860
Pro Glu Gln Leu Gly Arg Asn Ser Ala Trp Gly Leu Tyr Asp Trp Phe
865 870 875 880

Gln Ala Ser Trp Ala Leu Ala Leu Pro Val Ser Ile Phe Gly His Leu

885	890	895	
Leu Thr Arg Thr Arg Pro Leu Glu Gln Lys Leu Ile Ser Glu Glu Asp			
900	905	910	
Leu			
<210> 55			
<211> 2124			
<212> DNA			
<213> Artificial Sequence			
<220><223> Description of Artificial Sequence: Synthetic			
polynucleotide			
<400> 55			
atggctctca gcctctggcc cctgctgctg ctgctgctgc tgctgctgct gctgtccttt	60		
gcagtgactc tggcccttac tgggcctcat tccttgacc ctggtctctc ctctctgaag	120		
tcattgctct cactctgga ccaggctccc cagggtccc tgagccgctc acggttcttt	180		
acattcttg ccaacatttc ttcttcttt gagcctggga gaatgggga aggaccagta	240		
ggagagcccc cactctcca gccgcctgct ctgcggtcc atgattttct agtgacactg	300		
agaggtagcc ccgactggga gccaatgcta gggctgctag gggatatgct ggcactgctg	360		
ggacaggagc agactcccc agatttctg gtgcaccagg caggggtgct ggggtgactt	420		
gtggaggtgc tgctgggagc cttagtctt gggggccccc ctacccaac tcggcccca	480		
tgcaccgtg atggccgctc tgactgtgtc ctggctgctg actggttgcc ttctctgctg	540		
ctgttgtag agggcacacg ctggcaagct ctggtgcagg tgcagcccag tgtggacccc	600		
accaatgcca caggcctcga tgggaggag gcagctctc actttttgca gggctctgtg	660		
ggtttgctta cccaacagg ggagctaggc tccaaggagg ctctttggg cggctctgta	720		
cgcacagtgg gggccccct ctatgctgcc ttccaggagg ggctgctccg tgtcactcac	780		
tcctgcagg atgaggtctt ctccattttg gggcagccag agcctgatac caatgggcag	840		
tgccaggag gtaacctta acagctgctc ttatggggcg tccggcaca ctttctctgg	900		
gatgtccagg cgtgggctt tctgtctgga tcaccacccc caccctgc ctccttcac	960		
tgcttagca cgggcgtgcc tctgccaga gcttctcagc cgtcagcca catcagcca	1020		
cgccaacggc gagccatcac tgtggaggcc ctctgtgaga accacttagg ccagcacca	1080		
ccctacagca ttccaactt ctccatccac ttgctctgcc agcacaccaa gcctgccact	1140		
ccacagccc atcccagcac cactgccatc tgccagacag ctgtgtgta tgcagtgtcc	1200		

tgggcaccag gtgccaaggt ctggctacag gctgcccag accagtttcc tgatgagttt	1260
ttggatgcga tctgcagtaa cctctccttt tcagccctgt ctggctccaa ccgccgcctg	1320
gtgaagcggc tctgtgctgg cctgctccca cccctacca gctgccctga aggctgccc	1380
cctgttcccc tcacccaga catcttttgg ggctgcttct tggagaatga gactctgtgg	1440
gctgagcgac tgtgtgggga ggcaagtcta caggctgtgc cccccagcaa ccaggcttgg	1500
gtccagcatg tgtgccaggg cccacccca gatgtcactg cctccccacc atgccacatt	1560
ggacctgtg gggaaacgtg cccggatggg ggagcttcc tggatgaggt ctgtgccaat	1620
gacacatgt atgaggtcct ggtgcccttc tggccttggc tagcaggcca atgcaggata	1680
agtcgtgggg gcaatgacac ttgcttccca gaagggtgc tgggccccct tctgccctct	1740
ctgccaccac tgggaccatc cccactctgt ctgacctg gcccttctt ccttgcatg	1800
ctatccagt tgccacgtg tcagtcctct gtcccagctc ttgtcaccc cacacgcta	1860
cactatctcc tccgctgct gacctcttc ttgggtccag gggctggggg cgctgaggcc	1920
caggggatgc tgggtcgggc cctactgtc tccagtctcc cagacaactg ctcttcttg	1980
gatgccttcc gccagaggg ccggcgaggt gtgctacgga cgattgggga atacctggaa	2040
caagatgagg agcagccaac cccatcaggc ttgaacca ctgtcaacc cagctctggt	2100
ataagcaaga tggagctgct ggcc	2124

<210> 56

<211> 2238

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 56

atggctctga gctccagcc ccagctgtc ctctcctgt cgctcctgcc gcaggaagtg	60
acttcagccc ctactgggcc tcagcttttg gatgtgggc tctcccttct gaagtcattc	120
glagccactc tggaccaagc tctcagcgt tcctcagcc agtcacggtt ctctgcgttc	180
ctggccaaca tttcttcac ctccagctt gggaggatgg gggagggacc ggtgggagag	240
ccccacctc tccagcccc tgcaattcga ctcatgatt tctcgtgac actgagaggt	300
agcccagact gggagccaat gctagggtt ctgggagatg tgctggcact cctgggacag	360
gaacagactc cccgggactt ttgtgtgcac caggcaggtg tactgggtgg actttagag	420
gcattgttgg gacggttagt tctggaggc cccctgccc cactcgacc cccatgcacc	480

cgtgatggcc ctctgactg tgtcctggct gctgattggt tgccttctct gatgttgta 540
 ttagagggtg cacgctggca ggccctgggt cagttgcagc ccagltgga cccaaccaat 600
 gccacaggtc ttgatgtag agagccagct cctcactttt tacagggtct gctgggcttg 660
 cttaccccag caggagagtt gggctctgag gaggtctttt ggggtggtct gctgcgcaca 720

 gtggggggccc cctctatgc tgccttcag gaggggtac tgcgagtcac tcattctctg 780
 caagatgagg tcttttctat tatgggacag ccagagcctg atgccagtgg gcagtgccag 840
 ggaggcaacc ttcaacagct gcttttatgg ggcatgcgga acaacctttc ttgggacgcc 900
 cgagcactgg gttttctatc tggatcacca cctccacccc ctgctctcct gcaactgctg 960
 agcagagggtg tgcctctgcc cagggtctcc cagcctgcgg ctacatcag cctcgcagag 1020
 cggcgagcca tctctgtgga ggccctctgc gagaaccact caggcccaga gccaccctac 1080
 agcatctcca acttctccat ctacttctc tgcagcaca tcaagcctgc caccgcgg 1140

 cccctctcta ccacccacg gcctctcct accacccac agccctcc taccactaca 1200
 cagccattc ctgacactac acagccctt cctgtcacc caaggcctcc tctaccacc 1260
 ccacaacccc ctctagcac agctgtcatc tgcagacag ctgtatgga cgcagtctcg 1320
 tgggcaccag gtgcccagg ttggctcaa gcctgccatg atcagtttcc tgatcaattt 1380
 ctggatatga tctgcggcaa cctctcattt tcagccctgt ctggcccag tctcctttg 1440
 gtaaagcagc tctgtgtgg cttgtccca cccccacta gctgtccacc aggcctgac 1500
 cctgtgcccc tcacccaga aatattctgg ggctgtttcc tggagaatga gacactgtgg 1560

 gctgaacggt tgtgtgtgga ggacagtctg caggctgtgc cccgaggaa ccaggttgg 1620
 gtacagcatg tgtgtcgggg cccaccttg gacccactg atttccacc gtgccgcgtt 1680
 ggacctgtg gggaacgtg ccagatggg ggcagcttcc tgcctatggt ctgtgccaat 1740
 gacactctgt atgaagcctt ggttcccttc tgggcttggc tagcaggcca atgcagaatt 1800
 agtcgtggag gaaatgatac ttgctttcta gaaggcatgc tgggcccctt gttgccctct 1860
 ctgcccctc tgggaccatc cccactctgt ctggctctg gtccttttct gcttggcatg 1920
 ttatccaggt tgccacgtg tcagtcctcc gtccagccc tcgcccacc cagcgccta 1980

 cattacctc tgcgcctact gacctctt ctgggtccag ggactggggg tgccgagacg 2040
 caggggatgt taggtcaage cctgtctctc tctagtctcc cagacaactg ttcattcttg 2100
 gatgccttcc gccagaggg ccggagaagt gtactgagga cagtcggaga gtacttcag 2160
 cgggaagagc caacccacc aggttagac tctcctca gcctcggctc tggatatgac 2220
 aagatggagc ttctgtcc 2238

<210> 57

<211> 3201

<212> DNA

<213> Homo sapiens

<400> 57

tgctttagtc ctgtgctgtg ggatctgctc cagagggaaa agagtgtttg ggccctgcag	60
attctagtgc aggcgtagct gcatatgcc ccagaaaacc tccagcagct ggtgctttca	120
gcagagaggg aggctgcaca gggcttccctg acactcatgc tgcaggggaa gctgcagggg	180
aagctgcagg taccaccatc cgaggagcag gccctgggtc gcctgacagc cctgctgctc	240
cagcggtagc cagcctcac ctcccagctc ttcatgacc tgtcaccact catcccttc	300
ttggctgtct ctgacctgat gcgcttccca ccatccctgt tagccaacga cagtgtcctg	360
gctgccatcc gggattacag cccaggaatg aggctgaac agaaggaggc tctggcaaag	420
cgactgctgg cccctgaact gtttggggaa gtgcctgcct ggccccagga gctgctgtgg	480
gcagtgtgc cctgtctccc ccacctccct ctggagaact ttttgcagct cagccctcac	540
cagatccagg cctggagga tagctggcca gcagcaggtc tggggccagg gcatgccgc	600
catgtgtgc gcagcctggt aaaccagagt gtccaggatg gtgaggagca ggtacgcagg	660
cttgggcccc tcgctgttt cctgagccct gaggagctgc agagcctagt gcccctgagt	720
gatccaacgg ggccagtaga acgggggctg ctggaatgtg cagccaatgg gaccctcagc	780
ccagaaggac ggggtggcata tgaacttctg ggtgtgttgc gctcatctgg aggagcggtg	840
ctgagcccc gggagctgcg ggtctgggcc cctctcttct ctcagctggg cctccgcttc	900
cttcaggagc tgcagagcc ccagcttaga gccatgcttc ctgtcctgca gggaactagt	960
gttacacctg ctcaggctgt cctgtgtctt ggacggctcc ttcctaggca cgtatctacc	1020
ctggaggaac tctgtctctt gcaccttctg ctaccaggcc tcagccccca gacctccag	1080
gccatcccta ggcgagtcct ggtcggggct tgttctgccc tggccccga actgtcacgc	1140
ctctcagcct gccagaccgc agcactgtg cagacctttc gggttaaaga tgggtttaaa	1200
aatatgggta caacaggctg tggteccagct gtgtgtatcc ctggtcagcc tatteccacc	1260
acctggccag actgcctgct tcccctgctc ccattaaagc tgctacaact ggattccttg	1320
gtcttcttgg caaatcgaag acgctactgg gagctgccct ggtctgagca gcaggcacag	1380
tttctctgga agaagatgca agtaccacc aaccttacc tcaggaatct gcaggctctg	1440
ggcaccctgg caggaggcat gtcctgtgag ttctctgcagc agatcaactc catggtagac	1500
ttccttgaag tgggtgcacat gatctatcag ctgccacta gagttcgagg gaggctgagg	1560
gcctgtatct gggcagagct acagcggagg atggcaatgc cagaaccaga atggacaact	1620

gtagggccag aactgaacgg gctggatagc aagctactcc tggacttacc gatccagttg 1680
atggacagac tatccaatga atccattatg ttggtggtgg agctggtgca aagagctcca 1740

gagcagctgc tggcactgac cccctccac caggcagccc tggcagagag ggcactacaa 1800
aacctggctc caaaggagac tccagtctca ggggaagtgc tggagacctt aggccctttg 1860
gttggattcc tggggacaga gagcacacga cagatcccc tacagatcct gctgtcccat 1920
ctcagtcagc tgcaaggctt ctgcctagga gagacatttg ccacagagct gggatggctg 1980
ctattgcagg agtctgttct tgggaaacca gatttgtgga gccaggatga agtagagcaa 2040
gctggacgcc tagtattcac tctgtctact gaggcaattt ccttgatccc cagggaggcc 2100
ttgggtccag agaccctgga gcggcttcta gaaaagcagc agagctggga gcagagcaga 2160

gttggacagc tgttaggga gccacagctt gctgccaaga aagcagccct gtagcaggg 2220
gtggtcgcac cagctgtgta ggatcttcca gaacctgtgc caaatgtgc agatgtacga 2280
gggacattcc cagcagcctg gtctgcaacc cagattgcag agatggagct ctcagacttt 2340
gaggactgcc tgacattatt tgcaggagac ccaggacttg ggcctgagga actgcgggca 2400
gccatgggca aagcaaaaca gttgtggggt ccccccggtt gatttcgtcc tgagcagatc 2460
ctgcagcttg gtaggctctt aataggtcta ggagatcggg aactacagga gctgatccta 2520
gtggactggg gagtgtgag caccctgggg cagatagatg gctggagcac cactcagctc 2580

cgcatgttg tctccagttt cctacggcag agtggtcggc atgtgagcca cctggacttc 2640
gttcatctga cagcgtcggg ttatactctc tgtggactgc ggccagagga gctccagcac 2700
atcagcagtt gggagttcag ccaagcagct ctcttctctg gcaccctgca tctccagtgc 2760
tctgaggaac aactggaggt tctggccac ctacttgtac tgcctggtgg gtttggccca 2820
atcagtaact gggggcctga gatcttact gaaattggca ccatagcagc tgggatccca 2880
gacctggctc ttccagcact gctgcgggga cagatccagg gcgttactcc tcttgccatt 2940
tctgtcatcc ctctcctaa atttgctgtg gtgtttagtc ccatccaact atctagtctc 3000

accagtgtc aggtgttggc tgctactcct gagcaaatgg cttttctgag tcttgagcag 3060
cgacgagcag ttgcatgggc ccaacatgag ggaaaggaga gccagaaca gcaaggtcga 3120
agtacagcct ggggcctcca ggactggtca cgaccttctt ggtccctggt attgactatc 3180
agcttcttg gccacctgct a 3201

<210> 58
<211> 3189
<212> DNA
<213> Mus musculus

<400> 58

tgcttcagtc ctgtactgtg ggatctactc cagagagaga agagcgtttg ggccctgagg	60
accctgggtga aggcctacct gcgcatgcct ccagaagacc ttcagcagct tgtgctttca	120
gcagagatgg aggcctgcaca gggcttccctg acgctcatgc ttcgttccctg ggctaagctg	180
aaggttcaac catccgagga gcagcccatg ggccgcctga cagccttgct gctccagcgg	240
taccacagcc tcacctccca actctttatc gacatgtcac cgctcatccc ctctctgget	300
gtccctgacc tcatgcgctt cccaccgtcc cttttggcca acgacagtgt cctggctgcc	360
atcagggatc acagctcagg aatgaagcct gaacagaagg aggccctggc aaaacgactg	420
ctggcccttg agctgtttgg agaagtgcct gattggcccc aggagctgct gtgggcagcc	480
ctgcctctgc ttccccatct gcctctggag agctttctcc agctcagccc tcaccagatc	540
caggccctgg aggatagctg gccagtagca gatcttgggc cgggacacgc ccgacatgtg	600
cttcgtagcc tagtaacca gagcatggag gatggggagg agcagggtgct caggcttggg	660
tccctcgcct gtttcttgag tcttgaggag ctacagagtc tggtgccctt gagtgatcca	720
atggggcctg tagaacaggg tctgctggaa tgtgcggcca atgggaccct cagcccagaa	780
ggacgggtgg catatgaact tctgggagtg ttgcgttcat ctggaggaaac tgtcttaagc	840
ccccgagagc tgagggtctg ggcacctctc ttccccagc tgggcctccg ctctctgcag	900
gagctctcag agaccagct tagagccatg ctctctgccc tacagggagc cagtgtcaca	960
ctgcccagg ctgttctgtt gtttgaagg ctcttctta agcatgatct gtccctggag	1020
gaactctgct cctgcaccc tctctgcca ggtctcagcc cccagacact ccaggccatc	1080
cctaagagag ttctggttgg tgcttgttcc tgctgggcc ctgaactgtc aaggctttca	1140
gcttgccaga ttgcagctct gctgcagacc ttctgggtaa aagatggtgt taaaaatatg	1200
ggtgcagcag gtgccgctc agccgtgtgc attctgggc agcccaccac ttggccagac	1260
tgcttcttc cctgtctcc attaaagctg ctacagctgg acgtgcagc tcttctggca	1320
aaccgaagac tctatcgga gctgccttgg tctgagcaac aggcacagtt tctctggaag	1380
aaaatgcaag tgcctaccaa cctgagcctg aggaatctgc aggctctggg caacttggca	1440
ggaggcatga cctgcgagtt tctgcagcag atcagctcaa tggttgactt tcttgatgtg	1500
gtacacatgc tctaccagct gccactggt gtctgagaga gcctgcgggc ctgtatctgg	1560
acagagctac agcggaggat gacaatgcca gagccagagc tgaccaccct agggccagaa	1620
ctgagtgaac ttgacacaaa gctactctg gacttgccga tccagctgat ggacagattg	1680
tccaatgatt ccattatgtt ggtggtggag atggtccaag gcgtccaga gcagctgctg	1740

gcactgaccc cactccacca gacagccttg gcagagcgag cacttaaaaa cctggctcca 1800

aaggagaccc caatctccaa agaagtgtg gagacactgg gcccttgggt tggattcctg 1860

ggaatagaga gcacgcgacg gatcccttta cccattctac tgtctcatct cagtcagctg 1920

cagggcttct gcctaggaga gacatttgcc acagagctgg gatggctgct gttgcaggag 1980

cctgttcttg gaaaaccaga attgtggagc caggatgaaa tagagcaagc tggacgccta 2040

gtattcactc tgtctgtga ggctatttcc tcgatcccca gggaggcttt gggcccagag 2100

acactggaga ggcttctggg aaagcatcaa agctgggagc agagcagagt gggccatctg 2160

tgtggggagt cacagcttgc ccacaagaaa gcagctctgg tagctgggat tgtgcatcca 2220

gctgctgagg gtctccaaga gcctgtacca aactgtgcag acatcgggg aaccttccca 2280

gcggcctggt ctgcgacaca aatctcagag atggaactct cagactttga agactgcctg 2340

tcactatttg ctggagatcc aggacttgg cctgaggaac tacgggcagc catgggcaag 2400

gccaaagcagt tgtgggggcc ccctcgagga ttccgtcctg agcagatctt gcagctgggc 2460

cgtctcctga taggtctagg agaacgggaa ctgcaggagc ttaccttggg ggactggggg 2520

gtgctgagca gcctggggca aatagatggc tggagtcca tgcagctccg agccgtggtc 2580

tccagtttcc taaggcagag tggtcggcat gtgagccacc tggacttcat ttatctgaca 2640

gcactgggtt acacagtctg tggattgcga ccagaggagt tacagcacat cagcagttgg 2700

gagtttagcc aagcagctct ctctctgggt agcttgcatc tcccgtgctc tgaggaaacag 2760

ctggaagtgc tggcctatct ccttgtgttg cctgggtggc ttggcccagt cagtaactgg 2820

gggcctgaga tcttcaactga aattggcaca atagcagctg gcatcccaga cctggctctt 2880

tcagcattac tgcggggaca gatccaaggc ctgactctc ttgccatttc tgtattcct 2940

gtcccaagt ttgcagtgtt cttcaacccc atccagttat ctagtctcac caggggtcag 3000

gccgtagctg ttactctga acagctggcc tatctgagtc ctgagcagcg gcgagcagtt 3060

gcatggggcc aacacgaagg gaaggagatc ccagagcagc tgggtcgaaa ctacgcctgg 3120

ggtctctacg actggttcca agcctcctgg gcctggcat tgcccgtcag cttttttggc 3180

cacctatta 3189

<210> 59

<211> 7

<212> PRT

<213> Artificial Sequence

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

<400> 59

Thr Leu Ala Val Pro Phe Lys

1 5

<210> 60

<211> 9

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 60

Glu Leu Leu Ser Cys Phe Ser Pro Val

1 5

<210> 61

<211> 9

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 61

Gly Ser Leu Ala Cys Phe Leu Ser Pro

1 5

<210> 62

<211> 9

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 62

Glu Leu Leu Ala Cys Phe Ser Pro Val

1 5

<210> 63

<211> 9

<212> PRT

<213> Artificial Sequence

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

<400> 63

Gly Pro Leu Ala Cys Phe Leu Ser Pro

1 5