



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2025년06월10일

(11) 등록번호 10-2818627

(24) 등록일자 2025년06월04일

(51) 국제특허분류(Int. Cl.)

C07K 14/315 (2006.01)

(52) CPC특허분류

C07K 14/315 (2013.01)

C07K 2319/90 (2013.01)

(21) 출원번호 10-2018-7000488

(22) 출원일자(국제) 2016년06월03일

심사청구일자 2021년05월25일

(85) 번역문제출일자 2018년01월05일

(65) 공개번호 10-2018-0050640

(43) 공개일자 2018년05월15일

(86) 국제출원번호 PCT/GB2016/051640

(87) 국제공개번호 WO 2016/193746

국제공개일자 2016년12월08일

(30) 우선권주장

1509782.7 2015년06월05일 영국(GB)

(56) 선행기술조사문헌

WO2011098772 A1

WO2012127060 A1

WO2012142113 A2

(73) 특허권자

옥스포드 유니버시티 이노베이션 리미티드

영국 오엑스2 0제이비 옥스포드 웨스트 웨이 3 벅스틴 코트

(72) 발명자

호왈쓰, 마크

영국 오엑스1 3큐유 옥스포드샤이어, 옥스포드 사우스 파크스 로드 디파트먼트 오브 바이오케미스트리

(74) 대리인

특허법인지평

전체 청구항 수 : 총 29 항

심사관 : 한정선

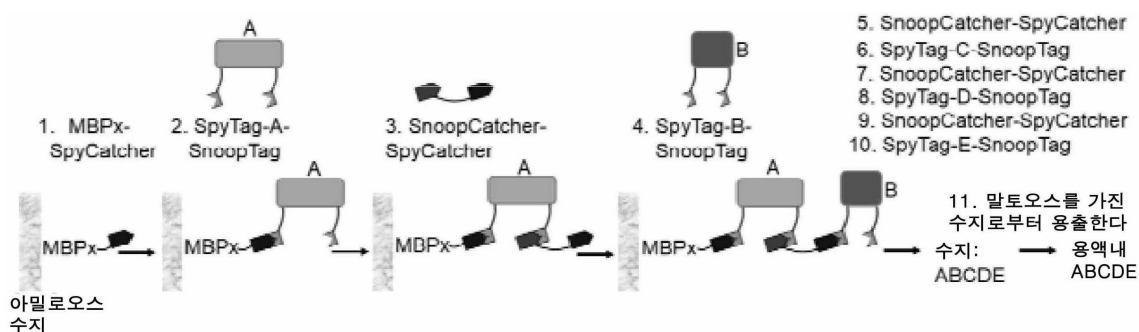
(54) 발명의 명칭 융합 단백질 합성용 방법 및 생성물

(57) 요약

본 발명은 융합 단백질의 생산 방법을 제공하고, 상기 방법은 하기 단계를 포함한다: a) 상기 단백질 사이 이소펩타이드 결합의 형성을 가능하게 하는 조건하에 제2 단백질과 제1 단백질을 접촉시키는 단계로서, 상기 제1 단백질 및 상기 제2 단백질 각각은 펩타이드 링커를 포함하고, 상기 펩타이드 링커는 연결 단백질을 형성하기 위해

(뒷면에 계속)

대표도



상기 제2 단백질에 상기 제1 단백질을 연결하는 이소펩타이드 결합을 형성하도록 반응하는 펩타이드 링커의 쌍인, 단계; 및 b) 상기 제3 단백질과 상기 연결 단백질 사이 이소펩타이드 결합의 형성을 가능하게 하는 조건하에 제3 단백질과 (a)로부터의 연결 단백질을 접촉시키는 단계로서, 상기 제3 단백질은 (a)로부터 연결 단백질에 추가 펩타이드 링커와 반응하는 펩타이드 링커를 포함하고, 상기 펩타이드 링커는 융합 단백질을 형성하기 위해 상기 연결 단백질에 상기 제3 단백질을 연결하는 이소펩타이드 결합을 형성하도록 반응하는 펩타이드 링커의 쌍이고, (a)에서 사용된 펩타이드 링커의 상기 쌍은 (b)에서 사용된 펩타이드 링커의 쌍과 직교이다. 융합 단백질의 합성에서 펩타이드 링커 및 상기 링커의 직교 쌍의 용도가 또한 제공된다. 상기 링커를 포함하는 제조된 단백질, 상기 단백질 및 링커를 인코딩하는 핵산 분자, 상기 핵산 분자를 포함하는 벡터 및 상기 벡터 및 핵산 분자를 포함하는 숙주 세포가 또한 고려된다.

명세서

청구범위

청구항 1

융합 단백질의 생산 방법으로서,

a) 제1 단백질과 제2 단백질을 상기 단백질들 사이에 이소펩타이드 결합의 형성을 가능하게 하는 조건하에서 접촉시키는 단계로서, 상기 제1 단백질 및 상기 제2 단백질 각각은 펩타이드 링커를 포함하고, 상기 펩타이드 링커는 연결 단백질을 형성하기 위해 상기 제2 단백질에 상기 제1 단백질을 연결하는 이소펩타이드 결합을 형성하도록 반응하는 펩타이드 링커의 쌍인, 단계; 및

b) (a)로부터의 상기 연결 단백질과 제3 단백질을 상기 제3 단백질과 상기 연결 단백질 사이에 이소펩타이드 결합의 형성을 가능하게 하는 조건하에서 접촉시키는 단계로서, 상기 제3 단백질은 (a)로부터의 상기 연결 단백질에서 추가 펩타이드 링커와 반응하는 펩타이드 링커를 포함하고, 상기 펩타이드 링커는 융합 단백질을 형성하기 위해 상기 연결 단백질에 상기 제3 단백질을 연결하는 이소펩타이드 결합을 형성하도록 반응하는 펩타이드 링커의 쌍인, 단계;

를 포함하되,

(a)에서 사용된 상기 펩타이드 링커의 쌍이 (b)에서 사용된 상기 펩타이드 링커의 쌍과 직교이고,

상기 펩타이드 링커의 직교 쌍들은:

(1) (i) 서열 식별 번호: 1의 아미노산 서열 또는 서열 식별 번호: 1의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 9에서 리신 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 2의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음; 및

(ii) 서열 식별 번호: 2의 아미노산 서열 또는 서열 식별 번호: 2의 아미노산 서열과 적어도 90% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 55에서 글루타메이트 또는 아스파르테이트 잔기, 위치 94에서 트레오닌 잔기, 위치 100에서 글리신 잔기 및 위치 106에서 아스파라긴 또는 아스파르테이트 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 1의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

(2) (i) 서열 식별 번호: 5의 아미노산 서열 또는 서열 식별 번호: 5의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 8에서 아스파르테이트 또는 아스파라긴 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 6의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음; 및

(ii) 서열 식별 번호: 6의 아미노산 서열 또는 서열 식별 번호: 6의 아미노산 서열과 적어도 90% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 8에서 리신 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 5의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

(3) (i) 서열 식별 번호: 9의 아미노산 서열 또는 서열 식별 번호: 9의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 17에서 아스파라긴 또는 아스파르테이트 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 10의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음; 및

(ii) 서열 식별 번호: 10의 아미노산 서열 또는 서열 식별 번호: 10의 아미노산 서열과 적어도 90% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 9에서 리신 잔기 및 위치 70에서 글루타메이트 또는 아스파르테이트 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 9 또는 서열 식별 번호: 109의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

(4) (i) 서열 식별 번호: 109의 아미노산 서열 또는 서열 식별 번호: 109의 아미노산 서열과 적어도 80% 서열

동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 17에서 아스파라긴 또는 아스파르테이트 잔기, 위치 11에서 글리신 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 10의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음; 및

(ii) 서열 식별 번호: 10의 아미노산 서열 또는 서열 식별 번호: 10의 아미노산 서열과 적어도 90% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 9에서 리신 잔기 및 위치 70에서 글루타메이트 또는 아스파르테이트 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 9 또는 서열 식별 번호: 109의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음; 및

(5) (i) 서열 식별 번호: 13의 아미노산 서열 또는 서열 식별 번호: 13의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 7에서 아스파르테이트 또는 아스파라긴 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 14의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음; 및

(ii) 서열 식별 번호: 14의 아미노산 서열 또는 서열 식별 번호: 14의 아미노산 서열과 적어도 90% 서열 동일성을 가진 서열을 포함하는 펩타이드 링커, 상기 아미노산 서열은 위치 56에서 글루타메이트 또는 아스파르테이트 잔기, 및 위치 10에서 리신 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 13의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

중에서 선택되며,

상기 펩타이드 링커의 직교 쌍들은 (1) 및 (4), (1) 및 (5), (1) 및 (3), (1) 및 (2), (2) 및 (5), (3) 및 (5), 또는 (4) 및 (5)를 포함하는, 방법.

청구항 2

청구항 1에 있어서, 하기 단계를 포함하는, 융합 단백질의 생산 방법으로서:

- a) 제1 펩타이드 링커를 포함하는 제1 단백질을 제공하는 단계;
- b) 제2 단백질과 상기 제1 단백질을 접촉시키는 단계로서, 상기 제2 단백질은 제2 펩타이드 링커 및 제3 펩타이드 링커를 포함하고, 상기 제1 펩타이드 링커 및 상기 제2 펩타이드 링커가 이소펩타이드 결합 형성을 가능하게 하는 조건하에서 접촉되며, 이로써 상기 제1 및 제2 단백질을 연결시키는, 단계; 및
- c) 제3 단백질과 상기 연결된 제1 및 제2 단백질을 접촉시키는 단계로서, 상기 제3 단백질은 제4 펩타이드 링커를 포함하고, 상기 제3 펩타이드 링커 및 상기 제4 펩타이드 링커가 이소펩타이드 결합 형성을 가능하게 하는 조건하에서 접촉되며, 이로써 상기 제2 및 제3 단백질을 연결시켜 융합 단백질을 생산하는 단계,

상기 제1 및 제2 펩타이드 링커가 상기 제3 및 제4 펩타이드 링커로 구성되는 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍인, 방법.

청구항 3

청구항 1 또는 2에 있어서, 상기 방법이 추가로 상기 융합 단백질 확장의 단계를 포함하며, 여기서 상기 융합 단백질에 연결되는 신규한 단백질은 상기 융합 단백질내 이전의 이소펩타이드 결합을 형성하는데 사용된 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍의 일부를 형성하는 펩타이드 링커를 포함하고, 상기 신규한 단백질내 상기 펩타이드 링커는 상기 융합 단백질의 단백질에서 펩타이드 링커와 이소펩타이드 결합을 형성할 수 있으며, 상기 방법은 상기 신규한 단백질이 상기 융합 단백질내 펩타이드 링커와 이소펩타이드 결합 형성을 가능하게 하는 조건하에서 상기 융합 단백질과 상기 신규한 단백질을 접촉시키는 단계를 포함하는, 방법.

청구항 4

청구항 1 또는 2에 있어서, 상기 융합 단백질이 분지형, 선형 또는 원형 구조를 갖는, 방법.

청구항 5

청구항 4에 있어서, 상기 융합 단백질이 원형 융합 단백질을 형성할 수 있는, 방법.

청구항 6

청구항 1 또는 2에 있어서, 상기 방법이 고형 상에서 수행되는, 방법.

청구항 7

청구항 6에 있어서, 상기 고형 상으로부터 용합 단백질 용출의 단계를 추가로 포함하는, 방법.

청구항 8

청구항 1에 있어서, 서열 식별 번호: 109의 아미노산 서열과 적어도 80% 서열 동일성을 가진 아미노산 서열을 포함하는 상기 펩타이드 링커는 위치 20에서 이소류신 잔기, 위치 21 및 22에서 프롤린 잔기 및 위치 23에서 리신 잔기를 더 포함하는, 방법.

청구항 9

청구항 1 또는 2에 있어서, 바이러스성-유사 입자(viral-like particles)를 꾸미기(decorating) 위해 사슬 내부에 단백질 항원을 연결하기 위한, 방법.

청구항 10

펩타이드 링커로서,

(i) 서열 식별 번호: 1의 아미노산 서열 또는 서열 식별 번호: 1의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열, 상기 아미노산 서열은 위치 9에서 리신 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 2의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

(ii) 서열 식별 번호: 2의 아미노산 서열 또는 서열 식별 번호: 2의 아미노산 서열과 적어도 90% 서열 동일성을 가진 서열, 상기 아미노산 서열은 위치 55에서 글루탐레이트 또는 아스파르테이트 잔기, 위치 94에서 트레오닌 잔기, 위치 100에서 글리신 잔기 및 위치 106에서 아스파라긴 또는 아스파르테이트 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 1의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

(iii) 서열 식별 번호: 5의 아미노산 서열 또는 서열 식별 번호: 5의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열, 상기 아미노산 서열은 위치 8에서 아스파르테이트 또는 아스파라긴 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 6의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

(iv) 서열 식별 번호: 6의 아미노산 서열 또는 서열 식별 번호: 6의 아미노산 서열과 적어도 90% 서열 동일성을 가진 서열, 상기 아미노산 서열은 위치 8에서 리신 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 5의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

(v) 서열 식별 번호: 9의 아미노산 서열 또는 서열 식별 번호: 9의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열, 상기 아미노산 서열은 위치 17에서 아스파라긴 또는 아스파르테이트 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 10의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

(vi) 서열 식별 번호: 10의 아미노산 서열 또는 서열 식별 번호: 10의 아미노산 서열과 적어도 90% 서열 동일성을 가진 서열, 상기 아미노산 서열은 위치 9에서 리신 잔기 및 위치 70에서 글루탐레이트 또는 아스파르테이트 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 9 또는 109의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음; 또는

(vii) 서열 식별 번호: 109의 아미노산 서열 또는 서열 식별 번호: 109의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열, 상기 아미노산 서열은 위치 17에서 아스파라긴 또는 아스파르테이트 잔기, 위치 11에서 글리신 잔기를 포함하고, 상기 펩타이드 링커는 서열 식별 번호: 10의 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있음;

을 포함하는, 펩타이드 링커.

청구항 11

청구항 10에 있어서, 서열 식별 번호: 109의 아미노산 서열과 적어도 80% 서열 동일성을 가진 서열을 포함하는

상기 펩타이드 링커는 위치 20에서 이소류신 잔기, 위치 21 및 22에서 프롤린 잔기 및 위치 23에서 리신 잔기를 더 포함하는, 펩타이드 링커.

청구항 12

청구항 10에 있어서, 하기인, 펩타이드 링커:

- (a) (i)에서 상기 펩타이드 링커가 서열 식별 번호: 38의 아미노산 서열을 포함함; 및/또는
- (b) (ii)에서 상기 펩타이드 링커가 서열 식별 번호: 39의 아미노산 서열을 포함함 ; 및/또는
- (c) (iii)에서 상기 펩타이드 링커가 서열 식별 번호: 42의 아미노산 서열을 포함함; 및/또는
- (d) (iv)에서 상기 펩타이드 링커가 서열 식별 번호: 43의 아미노산 서열을 포함함; 및/또는
- (e) (v) 또는 (vii)에서 상기 펩타이드 링커가 서열 식별 번호: 46의 아미노산 서열을 포함함; 및/또는
- (f) (vi)에서 상기 펩타이드 링커가 서열 식별 번호: 47의 아미노산 서열을 포함함.

청구항 13

하기를 포함하는, 청구항 1 또는 2의 방법에서 사용하기 위한 펩타이드 링커의 쌍:

- (1) 청구항 10의 부분 (i)에 정의된 펩타이드 링커 및 청구항 10의 부분 (ii)에 정의된 펩타이드 링커;
- (2) 청구항 10의 부분 (iii)에 정의된 펩타이드 링커 및 청구항 10의 부분 (iv)에 정의된 펩타이드 링커; 또는
- (3) 청구항 10의 (v) 또는 청구항 10의 부분 (vii)에 정의된 펩타이드 링커 및 청구항 10의 부분 (vi)에 정의된 펩타이드 링커.

청구항 14

재조합 또는 합성 폴리펩타이드로서,

- (a) 폴리펩타이드 및 (b) 청구항 10에 정의된 펩타이드 링커를 포함하는, 재조합 또는 합성 폴리펩타이드.

청구항 15

청구항 14에 있어서, 상기 (a)의 폴리펩타이드는 원핵 세포, 진핵 세포, 바이러스, 박테리오파지 또는 마이코플라스마 유래인, 재조합 또는 합성 폴리펩타이드.

청구항 16

청구항 14에 있어서, 상기 (a)의 폴리펩타이드는 바이러스 유래인, 재조합 또는 합성 폴리펩타이드.

청구항 17

청구항 14에 있어서, 상기 (a)의 폴리펩타이드는 효소, 구조적 단백질, 항체, 항원, 프리온, 수용체, 리간드, 사이토카인, 케모카인 또는 호르몬인, 재조합 또는 합성 폴리펩타이드.

청구항 18

핵산 분자로서,

- (i) 청구항 10에 정의된 펩타이드 링커; 또는
- (ii) (a) 폴리펩타이드 및 (b) 청구항 10에 정의된 펩타이드 링커를 포함하는 재조합 또는 합성 폴리펩타이드; 를 인코딩하는, 핵산 분자.

청구항 19

청구항 18에 있어서, 상기 핵산 분자가 서열 식별 번호: 3, 4, 7, 8, 11, 12, 40, 41, 44, 45, 48 또는 49의 뉴클레오타이드 서열 또는 서열 식별 번호: 3, 4, 7, 8, 11, 12, 40, 41, 44, 45, 48 또는 49의 서열과 적어도 90% 서열 동일성을 가진 뉴클레오타이드 서열을 포함하는, 핵산 분자.

청구항 20

청구항 18에 있어서, 상기 (a)의 폴리펩타이드는 원핵 세포, 진핵 세포, 바이러스, 박테리오파지 또는 마이코플라스마 유래인, 핵산 분자.

청구항 21

청구항 18에 있어서, 상기 (a)의 폴리펩타이드는 바이러스 유래인, 핵산 분자.

청구항 22

청구항 18에 있어서, 상기 (a)의 폴리펩타이드는 효소, 구조적 단백질, 항체, 항원, 프리온, 수용체, 리간드, 사이토카인, 케모카인 또는 호르몬인, 핵산 분자.

청구항 23

청구항 18 또는 19의 핵산 분자를 포함하는 벡터.

청구항 24

인체 내의 세포를 제외한 세포로서,

청구항 18 또는 19의 핵산 분자, 또는 청구항 18 또는 19의 핵산 분자를 포함하는 벡터를 함유하는, 세포.

청구항 25

청구항 1 또는 2의 방법으로부터 수득된 또는 수득가능한, 융합 단백질.

청구항 26

청구항 25에 있어서, 상기 융합 단백질은 예방접종(vaccination)을 위한 것인, 융합 단백질.

청구항 27

청구항 1 또는 2의 방법으로부터 수득된 또는 수득가능한 적어도 하나의 융합 단백질을 포함하는, 고형 기질.

청구항 28

청구항 27에 있어서, 상기 기질이 어레이인, 고형 기질.

청구항 29

청구항 1 또는 2의 방법으로부터 수득된 또는 수득가능한 적어도 2개의 융합 단백질을 포함하는, 융합 단백질의 라이브러리.

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

발명의 설명

기술 분야

[0001]

본 발명은 융합 단백질 (즉, 아래 한정된 바와 같이, 2개 이상의 공유적으로 연결 단백질을 포함하는 폴리머)의 합성 (즉 생산, 생성 또는 어셈블리) 및 특히 이소펩타이드 결합을 형성하도록 반응하는 펩타이드 링커의 직교 쌍을 이용하는 융합 단백질의 모듈러 (예를 들면 단계식) 합성에 관한 것이다. 본 발명은 융합 단백질 합성, 특히 고품 상(phase) 합성을 위한 신규한 방법의 제공에 관한 것이다. 상기 방법은 유익하게는 융합 단백질, 예를 들면 융합 단백질 어레이를 포함하는 다양한 생성물의 생산에서 사용될 수 있다. 본 발명은 또한 융합 단백질의 합성에서 펩타이드 링커 및 상기 링커의 직교 쌍의 용도를 제공한다. 상기 링커를 포함하는 재조합 단백질, 상기 단백질 및 링커를 인코딩하는 핵산 분자, 상기 핵산 분자를 포함하는 벡터 및 상기 벡터 및 핵산 분자를 포함하는 숙주 세포가 또한 제공된다. 상기 재조합 폴리펩타이드 및/또는 핵산 분자/벡터를 포함하는 키트가 또한 제공된다. 본 발명의 방법에 의해 수득된 융합 단백질 및, 상기 융합 단백질을 포함하는, 생성물, 예를 들면 어레이 및 라이브러리가 또한 고려된다.

배경 기술

[0002]

생물학적 사건은 일반적으로 다중 단백질의 협력 활성화에 의존하고 복합체에서 단백질의 정확한 배열은 그의 기능에 영향을 미치고 기능을 결정한다. 따라서, 제어된 방식으로 복합체에서 개별적인 단백질을 배열하는 능력은 단백질 기능 특성규명에서 유용한 도구를 나타낸다. 또한, 소위 "융합 단백질"을 형성하기 위한 다중 단백질의 콘주게이션은 유용한 특징을 가진 분자를 초래할 수 있다. 예를 들어, 단백질의 단일 종류 클러스터링은 종종

크게 생물학적 신호, 예를 들면 백신에서 반복 항원 구조를 향상시킨다. 상이한 활성을 가진 단백질 클러스터링은 또한 개선된 활성, 예를 들면 효소에 의한 기질 채널링을 가진 복합체를 초래할 수 있다.

[0003] 그러나, 정확한 인공 "융합 단백질" 내부에 단백질의 상이한 종류 클러스터링은 수많은 문제를 마주치고 있다. 예를 들어, 개별적인 단백질 또는 단백질 도메인은 하나의 긴 열린 해독틀 내부에 유전자 상으로 연결될 수 있지만, 단백질 합성 및 미스폴딩에서 오차는 곧 제한적이다. 대안적인 방법은 개별적으로 단백질 또는 단백질 도메인 발현 및 그 다음 함께 이들 "모듈" 또는 "유닛" 연결에 집중하고 있다. 예를 들어, 방법은 양호하게 특성 규명된 상호작용 파트너, 예컨대 바이오틴/아비딘을 함유하기 위한 단백질 변형, 이로써 단백질이 비-공유 상호작용을 통해 복합체를 형성하게 하는 것에 집중하고 있다. 다른 방법은, 공유 결합, 즉 디설파이드 브릿지를 통해 단백질을 연결하기 위해, 단백질 내의 반응성 기, 특히 시스테인 잔기에 의존하고 있다. 그러나, 최상의 비-공유결합 또는 가역적 공유결합조차 융합 단백질의 재배열을 허용한다. 따라서, 존재하는 방법은, 이들이 분리하게 어려운 융합 단백질 및/또는 다양한 환경을 거쳐, 예를 들면 환원 조건에서 안정적이지 않은 융합 단백질의 미한정된 혼합물을 통상적으로 초래하는 한에 있어서는, 제한된다.

[0004] 융합 단백질 합성용 시스템의 중요한 특징은 상기 융합 단백질 내의 개별적인 단백질 (즉 모듈, 도메인 또는 유닛) 사이 분자적으로-한정된 연결, 임의의 템플레이트로부터 독립, 및 각각의 단백질 (즉 모듈, 도메인 또는 유닛)의 단순 발현을 포함한다. 혼합물 내에 불완전한 사슬의 혼한 결과인, 단지 몇개의 단계 이후 불균일 생성물의 우연한 합성을 최소화하기 위해 각각의 반응에 대하여 거의 정량적 수율을 갖는 것이 또한 아주 바람직하다. 융합 단백질 내에 각각의 모듈의 기능에 최소 파괴를 위하여, 큰 단백질 융합 도메인보다는 비교적 작은 펩타이드 태그로 개별적인 모듈이 변형되는 것이 또한 바람직하다. 그러나, 존재하는 융합 단백질 합성 방법은 이들 기준을 충족시킬 수 없었다.

[0005] 따라서, 융합 단백질의 개선된 합성 방법에 대한 필요성 및 요구가 있고 비가역적 공유결합을 생성하기 위해 이소펩타이드 결합을 형성하는 펩타이드 링커가 융합 단백질 합성을 위한 모듈러 (예를 들면 단계식), 및 고수율화, 방법에서 사용될 수 있다는 것이 이제 밝혀지고 있다.

[0006] 이소펩타이드 결합은 카복실/카복사미드와 아미노기 사이 형성된 아미드 결합이고, 여기서 카복실 또는 아미노기의 적어도 하나는 단백질 주요-사슬 (단백질의 골격) 밖이다. 상기 결합은 생물학적 조건하에서 화학적으로 비가역적이고 이들은 대부분의 프로테아제에 저항성이다. 사실상, 단백질 사이 이소펩타이드 결합은 가장 강력한 측정된 단백질 상호작용이도록 결정되고 있다.

[0007] 이소펩타이드 결합 형성은, 예를 들어 트랜스글루타미나제 효소에 의해 효소 촉매화될 수 있다. 이소펩타이드 결합은 단백질 복합체의 강도 및/또는 안정성, 예를 들면 세포외 매트릭스 구조의 안정화 또는 응혈의 보강을 개선하기 위해 천연 환경에서 통상적으로 발견된다.

[0008] HK97 박테리오파아지 캡시드 형성 및 그램-양성 박테리아 섬모에서 확인됨에 따라 이소펩타이드 결합은 또한 자발적으로 형성할 수 있다. 자발적 이소펩타이드 결합 형성은, 근처의 글루타메이트 또는 아스파르테이트에 의해 촉진된, 아스파라긴 또는 아스파르테이트의 C_g 기에서 리신으로부터 e-아미노기의 친핵성 공격을 통해, 단백질 폴딩 이후 발생하는 것으로 제안되고 있다.

[0009] 자발적 이소펩타이드 결합 형성을 할 수 있는 단백질은 유익하게는 서로 공유적으로 결합하는 및 따라서 비가역적 상호작용을 제공하는 펩타이드 태그/결합 파트너 쌍을 개발하는데 사용되고 있다 (참고 예를 들면 W02011/098772 본원에서 참고로 편입됨). 이 점에 있어서, 자발적 이소펩타이드 결합 형성을 할 수 있는 단백질은, 펩타이드 태그 및 펩타이드 태그용 결합 파트너를 제공하기 위해, 별개의 단편으로서 표현될 수 있고, 여기서 2개 단편은 이소펩타이드 결합 형성에 의해 공유적으로 재구성을 할 수 있다. 펩타이드 태그 및 결합 파트너 쌍에 의해 형성된 이소펩타이드 결합은 비-공유 상호작용이 빠르게 해리할 조건하에, 예를 들면 오랜 기간 (예를 들면 주) 동안, 고온에서 (적어도 95 °C까지에서), 강한 힘으로, 또는 가혹한 화학 처리 (예를 들면 pH 2-11, 유기 용매, 세제 또는 변성제)를 수반하여 안정적이다.

[0010] 간단히 말해서, 펩타이드 태그/결합 파트너 쌍은 이소펩타이드 결합을 자발적으로 형성할 수 있는 임의의 단백질 (이소펩타이드 단백질)로부터 유래될 수 있고, 여기에서 단백질의 도메인은 이소펩타이드 결합에서 관여된 잔기 중 하나 (예를 들면 리신)를 포함하는 펩타이드 태그 및 이소펩타이드 결합에서 관여된 다른 잔기 (예를 들면 아스파라긴 또는 아스파르테이트)를 포함하는 펩타이드 결합 파트너를 생산하기 위해 개별적으로 발현된다. 일부 사례에서, 펩타이드 태그 또는 결합 파트너 중 하나는 이소펩타이드 결합을 형성하기 위해 요구된 하나 이상의 다른 잔기 (예를 들면 글루타메이트)를 포함한다. 그러나, 개별적으로, 즉 별개의 3개 펩타이드

(도메인, 모듈 또는 유닛)으로서 이소펩타이드 결합 형성에 관여된 잔기를 포함하는 도메인을 발현하는 것이 가능하다는 것이 밝혀졌다. 이 점에 있어서, 상기 펩타이드 태그는 이소펩타이드 결합에서 관여된 잔기 중 하나 (예를 들면 리신)를 포함하고, 펩타이드 결합 파트너는 이소펩타이드 결합에서 관여된 다른 잔기 (예를 들면 아스파라긴 또는 아스파르테이트)를 포함하고 제3 펩타이드는 이소펩타이드 결합 형성에 관여된 하나 이상의 다른 잔기 (예를 들면 글루타메이트)를 포함한다. 모든 3개 펩타이드의 혼합은 이소펩타이드 결합을 형성하도록 반응하는 잔기를 포함하는 2개 펩타이드, 즉 펩타이드 태그 및 결합 파트너 사이 이소펩타이드 결합의 형성을 초래한다. 따라서, 제3 펩타이드는 펩타이드 태그 및 결합 파트너의 콘주게이션을 매개하지만 일부 결과적인 구조를 형성하지 않는다, 즉 제3 펩타이드는 펩타이드 태그 또는 결합 파트너에 공유결합되지 않는다. 이와 같이, 제3 펩타이드는 단백질 리가아제 또는 펩타이드 리가아제로서 보여질 수 있다. 이는 관심 단백질에 융합될 필요가 있는 펩타이드 태그 및 결합 파트너의 크기를 최소화함에 따라 특히 유용하여, 이로써 펩타이드 태그 또는 결합 파트너의 첨가에 의해 야기된 원치않는 상호작용, 예를 들면 미스폴딩의 가능성을 감소시킨다.

[0011] 아래에 더 상세히 논의된 바와 같이, 하나 이상의 이소펩타이드 결합을 자발적으로 형성할 수 있는 다양한 단백질 (소위 "이소펩타이드 단백질")은 확인되고 있고, 상기 논의된 바와 같이, 펩타이드 태그/결합 파트너 쌍 및 선택적으로 펩타이드 리가아제를 생산하기 위해 변형될 수 있다. 하나 이상의 이소펩타이드 결합을 자발적으로 형성할 수 있는 추가 단백질은 하나 이상의 이소펩타이드 결합을 자발적으로 형성하도록 공지되는 단백질의 것과 그의 구조를 비교함으로써 확인될 수 있다. 특히, 이소펩타이드 결합을 자발적으로 형성할 수 있는 다른 단백질은 공지된 이소펩타이드 단백질 예를 들면 주요 필린 단백질 Spy0128로부터의 것과 그의 결정 구조 비교, 및 특히 이소펩타이드 단백질의 형성에 종종 관여된 Lys-Asn/Asp-Glu/Asp 잔기 비교에 의해 확인될 수 있다. 추가로, 다른 이소펩타이드 단백질은 표준 데이터베이스 검색 툴을 이용하는 단백질 데이터 뱅크를 이용하여 공지된 이소펩타이드 단백질의 구조적 동족체에 대한 선별에 의해 확인될 수 있다. SPASM 서버 (<http://eds.bmc.uu.se/eds/spana.php?spasm>)는 이소펩타이드 결합의 Lys-Asn/Asp-Glu/Asp의 3D 구조적 템플레이트를 표적하는데 사용될 수 있거나 이소펩타이드 단백질은 또한 서열 상동성 단독에 의해 확인될 수 있다.

[0012] 현저히, 이소펩타이드 결합을 형성하는 단백질은 W02011/098772 (본원에서 참고로 편입됨)에 기재된 바와 같이 신규 설계될 수 있다. Rosetta는 이소펩타이드 단백질을 신규 설계하는데 사용될 수 있고 상기 소프트웨어는 http://depts.washington.edu/ventures/UW_Technology/Express_Licenses/rosetta.php에서 찾아질 수 있다. (참고 또한 Macromolecular modeling with rosetta, Das.R, Baker.D, Annu Rev Biochem, 2008, 77, 363-82). 추가적으로, RASMOT-3D PRO 서버는 <http://biodev.extra.cea.fr/rasmot3d/>에서 잔기의 적절한 배향에 대하여 단백질 데이터베이스를 검색하는데 사용될 수 있다.

발명의 내용

과제의 해결 수단

[0013] 본 발명자들은 상기 펩타이드 태그/결합 파트너 쌍이 다중 단백질을 공유적으로 결합하기 위해, 즉 융합 단백질을 생산하기 위해 펩타이드 링커로서 사용될 수 있다는 것을 유익하게는 결정하였다. 특히, 본 발명자들은 펩타이드 태그/결합 파트너 펩타이드의 직교 (즉 상호간에 무반응성 또는 비-동족) 쌍이 2개 이상의 단백질의 융합 (예를 들면 콘주게이션, 연결)에서 유용성, 즉 융합 단백질의 생산 (합성, 작제, 어셈블리)를 찾는 것을 실증하였다. 아래 실시예에서 상세히 실증된 바와 같이, 본 발명의 방법 및 용도는 순차적인 이소펩타이드 결합 형성에 기반된 사슬 내부에 단백질을 연결하는 모듈리 (예를 들면 단계식) 및 고수율화 접근법을 제공한다. 특히, 본원에서 기재된 방법 및 용도는 통계적인 혼합물의 생성 없이 단백질 사슬의 제어된 (즉 특이적, 표적화된) 확장을 가능하게 한다. 각각의 단백질 유닛 (모듈, 도메인)이 비가역적 연결, 즉 이소펩타이드 결합에 의해 연결되는 융합 단백질을 초래하기 때문에 이전의 방법보다 특히 유리하다. 따라서, 연결이 시스테인 잔기의 반응에 의지함에 따라, 유리 시스테인 잔기 및/또는 디설파이드 결합을 함유하는 단백질에 적용가능하다. 더욱이, 사슬에 첨가되는 각각의 단백질 유닛은, 단백질 내의 다양한 위치에서, 즉 단백질상의 N-말단, C-말단 또는 내부 부위에서 편입될 수 있는, 2개의 작은 펩타이드 태그로만 변형될 필요가 있다. 따라서, 융합 단백질의 각각의 단백질 유닛은 완전히 유전자 상으로 인코딩될 수 있다, 즉 본 방법은 비천연 (즉 비-표준) 아미노산의 용도 또는 아미노산 잔기의 번역후 변형에 의존하지 않는다. 따라서, 본 발명은 고도로 특이적이고 중간체의 정제를 요구하지 않는 융합 단백질의 단순 및 확장가능한 합성 방법을 제공한다.

[0014] 본 발명의 방법의 대표적인 예는, 본 발명의 고형 상 구현예를 보여주는, 도 1에서 도시된다. 그러나, 이는 본 발명의 범위에 제한되도록 의도된 것은 결코 아니고 다양한 다른 순열은 아래 설명으로부터 숙련가에 분명해질

것이고 첨부된 청구항에서 한정된 바와 같이 본 발명에 의해 포함되도록 의도된다.

- [0015] 도 1은 펩타이드 링커의 2개 쌍, 일명 SpyTag/SpyCatcher 및 SnoopTag/Snoop Catcher를 보여주고, 여기에서 각각의 쌍, 즉 각각의 "Tag" 및 "Catcher"는 구체적으로 및 자발적으로 반응하여 이소펩타이드 결합을 형성하여 이로써 "Tag" 펩타이드를 "Catcher" 펩타이드에 연결한다. 이 점에 있어서, 상기 쌍은 직교이고, 이들이 상호간에 무반응성임, 즉 SpyTag 및 SpyCatcher가 이소펩타이드 결합을 형성하기 위해 SnoopCatcher 또는 SnoopTag와 반응할 수 없다는 것을 의미한다. 더 상세히 이하에 논의된 바와 같이, 일부 구현예에서, "Tags"는 펩타이드 태그로서 보여질 수 있고 "Catcher" 펩타이드는 결합 파트너 단백질로서 보여질 수 있다.
- [0016] 따라서, 단계 1에서, 제1 단백질, MBPx (말토오스-결합 단백질의 변형된 버전, 아래 논의됨)가 제공되고, 여기에서 상기 단백질은 펩타이드 링커, 예를 들면 MBPx 폴리펩타이드를 인코딩하는 핵산 분자의 재조합 발현을 통한, SpyCatcher (즉 펩타이드 링커의 제1 쌍의 제1 일부) 및 단일 열린 핵독톨에서 SpyCatcher 펩타이드 링커를 편입시키도록 변형되고 있다. 상기 대표적인 예에서, MBPx 단백질은 확장하는 융합 단백질을 고품 상 (아미로오스 수지)에서 고정되도록 하는 정제 또는 고정화 태그로서 사용된다. 그러나, 이것이 본 발명의 필수적인 특징이 아니라는 아래 논의로부터 분명해질 것이다. 예를 들어, 본 방법은 불균일 (즉 고품 상) 또는 균일 (즉 용액 내)일 수 있고, 불균일이면, 임의의 적합한 정제/고정화 태그가 사용될 수 있다, 즉 태그가 단백질 또는 펩타이드 태그인 것이 필수적이지 않다.
- [0017] 단계 2에서, 제1 단백질 (MBPx-SpyCatcher)은 2개 펩타이드 링커를 편입시키기 위해 변형되고 있는 제2 단백질 (A)와 접촉된다. 하나의 펩타이드 링커는 링커 (SpyTag)의 제1 쌍의 제2 일부이고, 제1 일부는 제1 단백질 (SpyCatcher)의 도메인을 형성한다. 다른 펩타이드 링커는 펩타이드 링커 (SnoopTag)의 제2 쌍의 제1 일부이고; 상기 논의된 바와 같이, 링커의 제2 쌍은 링커의 제1 쌍과 반응하지 않는다. 따라서, 함께 제1 및 제2 단백질 접촉시, 링커의 제1 쌍은 (예를 들면 자발적으로) 반응하여 SpyCatcher와 SpyTag 펩타이드 링커 사이 특이적 이소펩타이드 결합을 형성하여, 이로써 함께 제1 단백질 (MBPx-SpyCatcher) 및 제2 단백질 (SpyTag-A-SnoopTag)을 연결하여 융합 단백질을 형성한다.
- [0018] 단계 3에서, 융합 단백질 (MBPx-SpyCatcher-SpyTag-A-SnoopTag)은 2개 펩타이드 링커, SnoopCatcher 및 SpyCatcher를 포함하는 추가 단백질과 접촉된다. 따라서, 하나의 펩타이드 링커 (SnoopCatcher)는 펩타이드 링커의 제2 쌍의 제2 일부이고 다른 펩타이드 링커 (SpyCatcher)는 펩타이드 링커의 제1 쌍 출신이다. 이들 펩타이드 링커는 스페이서 예를 들면 펩타이드 스페이서, 또는 최종 융합 단백질 내부에 편입되는 단백질을 통해 연결될 수 있다. 추가 단백질 (SnoopCatcher-SpyCatcher)과 융합 단백질 (MBPx-SpyCatcher-SpyTag-A-SnoopTag) 접촉시, 링커의 제2 쌍은 (예를 들면 자발적으로) 반응하여 이소펩타이드 결합을 형성하여, 이로써 융합 단백질을 확장한다. 대안적으로 보면, SnoopCatcher-SpyCatcher 단백질의 첨가는 추가 확장을 위한 융합 단백질의 기능화 또는 활성화로서, 즉 융합 단백질에 반응성 기 (반응성 펩타이드 링커)의 첨가에 의해 보여질 수 있다.
- [0019] 단계 4에서, 단계 3 (MBPx-SpyCatcher-SpyTag-A-SnoopTag-SnoopCatcher-SpyCatcher)으로부터 확장된 융합 단백질은, A 단백질과 비슷한 2개 펩타이드 링커를 편입시키는, 추가 단백질 (B)와 접촉된다 (SpyTag-B-SnoopTag). 재차, 이소펩타이드 결합은, 융합 단백질을 추가로 확장하기 위해, 함께 반응할 수 있는 펩타이드 링커, 즉 제1 쌍, SpyCatcher 및 SpyTag 사이 형성된다.
- [0020] 요망된 융합 단백질의 모든 단백질 유닛이 함께 연결되고 있을 때까지 상기 방법이 반복될 수 있다는 것이 분명할 것이다. 융합 단백질은 고품 상으로부터, 예를 들면 말토오스로 간단히 용출될 수 있고, 추가 정제 없이 사용될 수 있다. 융합 단백질의 말단 단백질이, 단백질에서 자유 펩타이드 링커, 예를 들면 끝에서 두번째 단백질 유닛과 반응할 수 있는, 융합 단백질의 단일 펩타이드 링커만을 편입시키도록 변형될 필요가 있다는 것을 유의해야 한다. 실시예에서 논의된 바와 같이, 본 발명자는, 겔 전기영동 및 질량 분광분석법에 의해 입증되고 있는, 10 단백질 유닛을 함유하는 융합 단백질의 합성을 실증하였다.
- [0021] 이론에 의한 구속됨 없이, 펩타이드 링커 쌍, 예를 들면 SnoopTag/SnoopCatcher, SpyTag/SpyCatcher 등에서 아미노산 잔기의 정확한 배향이 펩타이드 링커 사이 비가역적 이소펩타이드 결합의 형성 및 친핵성 공격을 촉진시키는 것으로 생각된다. 상기에서 언급된 바와 같이, 리신은 각각의 이들 쌍에서 아스파르테이트 또는 아스파라긴과 반응한다. SpyTag 펩타이드는 반응성 아스파르테이트를 갖고 그래서 SnoopCatcher의 반응성 아스파라긴과 반응할 수 없다. SnoopTag 펩타이드는 반응성 리신을 갖고 그래서 SpyCatcher의 반응성 리신과 반응할 수 없다. 따라서 이들 2개 펩타이드 링커 쌍은 직교이고 펩타이드 링커의 임의의 직교 쌍이 융합 단백질을 생성하기 위해 본 발명의 방법에서 사용될 수 있다는 것이 분명할 것이다. 이 점에 있어서, 강력한 및 프로그래밍가능한 융합 단백질의 생성을 가능하게 하는 펩타이드 링커 쌍의 직교, 상호간에 무반응성, 특성이다. 특히, 성장하는 융합

단백질 사슬이 고품 상에 부착되면, 반응 모듈 (즉 융합 단백질에 연결되는 다음 단백질)은 큰 과잉으로 첨가될 수 있어서, 이로써 반응을 완료하게 한다. 이는 미반응된 빌딩 블록이 세정 제거될 수 있고, 그래서 분리 (즉 미반응된 성분으로부터 성장하는 융합 단백질의 분리)는 각각의 단계에서 불필요하다는 것을 의미한다. 따라서, 한번에 한 단계 연장은 직교 연결의 작은 수를 이용하는 사슬 성장을 허용한다. 그러므로, 본 발명자에 의해 개발된 본 방법은 이전에 기재된 단백질-커플링 방법보다, 특히 융합 단백질 생성물의 안정성 및 개별적인 반응 단계의 단순함에 관하여 우월하다.

[0022] 따라서, 그의 가장 넓게, 본 발명은 융합 단백질의 생산용 펩타이드 링커의 적어도 2개의 직교 쌍의 용도로서 보여질 수 있고, 여기에서 펩타이드 링커의 각각의 쌍은 반응하여 이소펩타이드 결합을 형성한다.

[0023] 특히, 각각의 펩타이드 링커 쌍의 펩타이드 링커는 서로 반응하여 이소펩타이드 결합을 형성한다. 상기에서 언급된 바와 같이, 각각의 펩타이드 링커는 융합 단백질의 유닛 (예를 들면 도메인 또는 모듈)을 형성할 단백질의 일부 (예를 들면 도메인)를 형성한다. 다시 말해서, 함께 연결되는 단백질은 적어도 하나의 펩타이드 링커 (예를 들면 2, 3, 4 펩타이드 링커 등)를 편입시키도록 변형될 수 있고, 여기에서 융합 단백질의 생산에서 사용된 펩타이드 링커의 각각의 쌍은 상기 융합 단백질의 생산에서 사용된 펩타이드 링커의 적어도 하나의 다른 쌍과 직교이다.

[0024] 따라서, 일부 구현예에서, 펩타이드 링커의 직교 쌍은 적어도 2개의 단백질 유닛 (예를 들면 도메인 또는 모듈)을 함유하는 융합 단백질의 생산에서 사용된다. 예를 들어, 도 1에서 보여준 대표적인 구현예에서, 단백질 A를 단백질 B와 접합하는데 사용된 단백질은 링커 유닛으로서 보여질 수 있다, 즉 링커 유닛 (링커 단백질)은 단백질 A를 단백질 B와 단지 접합하는 기능을 한다. 따라서, 융합 단백질은 적어도 2개의 기능성 단백질, 즉 링커로서 이외의 기능을 가진 단백질을 함유 또는 포함하는 것으로 보여질 수 있다. 다른 구현예에서, 융합 단백질은 (즉 그의 기능과 무관한) 적어도 3개의 단백질을 함유 또는 포함하는 것으로 보여질 수 있다.

[0025] 추가 구현예에서, 융합 단백질은 적어도 3개의 기능성 단백질을 함유 또는 포함하는 것으로 보여질 수 있다. 예를 들어, 도 1에서 보여준 대표적인 구현예와 관련하여, 단백질 A를 단백질 B와 단지 접합하는데 사용된 링커 단백질이 펩타이드 링커에 더하여 단백질 (예를 들면 기능성 단백질)을 함유하면, 융합 단백질의 단백질 유닛 (도메인 또는 모듈)로서 보여질 수 있다. 따라서, 융합 단백질은 적어도 3개의 기능성 단백질, 즉 링커 이외에, 또는 링커에 더하여, 기능을 가진 단백질을 함유 또는 포함하는 것으로 보여질 수 있다.

[0026] 대안적으로 보면, 본 발명은 융합 단백질의 생산 방법 (예를 들면 생성, 합성, 어셈블링 등)을 제공하고, 상기 방법은 하기 단계를 포함한다:

[0027] a) 상기 단백질 사이 이소펩타이드 결합의 형성을 가능하게 하는 조건하에 제2 단백질과 제1 단백질을 접촉시키는 단계로서, 상기 제1 단백질 및 상기 제2 단백질 각각은 펩타이드 링커를 포함하고, 상기 펩타이드 링커는 연결 단백질을 형성하기 위해 상기 제2 단백질에 상기 제1 단백질을 연결하는 이소펩타이드 결합을 형성하도록 (서로) 반응하는 펩타이드 링커의 쌍인, 단계; 및

[0028] b) 상기 제3 단백질과 상기 연결 단백질 사이 이소펩타이드 결합의 형성을 가능하게 하는 조건하에 제3 단백질과 (a)로부터의 연결 단백질을 접촉시키는 단계로서, 상기 제3 단백질은 (a)로부터 연결 단백질에서 추가 펩타이드 링커와 반응하는 펩타이드 링커를 포함하고, 상기 펩타이드 링커는 융합 단백질을 형성하기 위해 상기 연결 단백질에 상기 제3 단백질을 연결하는 이소펩타이드 결합을 형성하도록 (서로) 반응하는 펩타이드 링커의 쌍인, 단계,

[0029] 여기에서 (a)로부터/에서 사용된 펩타이드 링커의 상기 쌍은 (b)로부터/에서 사용된 펩타이드 링커의 상기 쌍과 직교임.

[0030] 더욱 또 다른 측면으로부터 보면, 본 발명은 융합 단백질의 생산 (예를 들면 생성, 합성, 어셈블링 등) 방법을 제공하고, 상기 방법은 하기 단계를 포함한다:

[0031] a) 제1 펩타이드 링커를 포함하는 제1 단백질을 제공하는 단계;

[0032] b) 제2 단백질과 상기 제1 단백질을 접촉시키는 단계로서, 상기 제2 단백질은, 상기 제1 펩타이드 링커 및 상기 제2 펩타이드 링커가 이소펩타이드 결합 형성을 가능하게 하는 조건하에, 제2 펩타이드 링커 및 제3 펩타이드 링커를 포함하고, 이로써 상기 제1 및 제2 단백질을 연결시키는, 단계; 및

[0033] c) 제3 단백질과 상기 연결된 제1 및 제2 단백질을 접촉시키는 단계로서, 상기 제3 단백질은, 상기 제3 펩타이드 링커 및 상기 제4 펩타이드 링커가 이소펩타이드 결합 형성을 가능하게 하는 조건하에, 제4 펩타이드 링커를

포함하고, 이로써 상기 제2 및 제3 단백질을 연결시켜 융합 단백질을 생산하는 단계,

- [0034] 여기에서 상기 제1 및 제2 펩타이드 링커가 상기 제3 및 제4 펩타이드 링커로 구성되는 펩타이드 링커의 상기 쌍과 직교인 펩타이드 링커의 쌍임.
- [0035] 상기 언급된 바와 같이, 일부 구현예에서, 제2 단백질은 제1 및 제3 단백질 사이 링커로서 기능할 수 있다. 따라서, 융합 단백질은 2개 "기능성" 단백질, 즉 2개 단백질 유닛 (모듈, 도메인 등)을 함께 연결하는 것 이외의 기능을 갖는 단백질을 포함하는 것으로서 보여질 수 있다. 따라서, 일부 구현예에서, 제2 단백질은 링커 단백질, 즉 펩타이드 링커 및 선택적으로 스페이서 도메인, 예를 들면 펩타이드 스페이서의 상이한 직교 쌍으로부터 각각 적어도 2개의 펩타이드 링커를 함유하는 단백질로서 보여질 수 있다.
- [0036] 따라서, 일부 구현예에서, 제2 단백질은 상기 제1 단백질이 상기 제3 단백질에 연결되도록 (상기에 접합되도록) 하기 위해 제1 단백질을 기능화 또는 활성화하는 링커 단백질로서 보여질 수 있다. 유사하게, 추가 단백질이 융합 단백질에 첨가되는 경우 (즉 융합 단백질이 확장되는 경우) 링커 단백질은 상기 하나 이상의 단백질이 상기 추가 단백질과 연결되도록 하기 위해 융합 단백질에서 하나 이상의 단백질을 기능화 또는 활성화하는데 사용될 수 있다.
- [0037] 상기 논의된 바와 같이, 펩타이드 링커의 직교 쌍의 용도는 다수의 단백질 유닛을 함유하는 융합 단백질의 생산을 용이하게 한다. 따라서, 추가 단백질은 융합 단백질의 한 단백질에서 펩타이드 링커와 이소펩타이드 결합을 형성할 수 있는 적어도 하나의 펩타이드 링커를 포함하는 추가 단백질과 융합 단백질의 접촉에 의해 융합 단백질에 첨가될 수 있다 (즉 융합 단백질은 확장 (예를 들면 연장, 신장)될 수 있다). 이 점에 있어서, 신규한 단백질에서 펩타이드 링커는 융합 단백질에서 이전의 이소펩타이드 결합을 형성하는데 사용된 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍의 일부를 형성한다.
- [0038] 따라서, 일부 구현예에서, 본 방법은 추가로 융합 단백질의 확장의 단계를 포함하고, 여기에서 융합 단백질에 연결되는 신규한 단백질 (즉 추가의 또는 추가 단백질)은 융합 단백질에서 이전의 이소펩타이드 결합을 형성하는데 사용된 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍의 일부를 형성하는 펩타이드 링커를 포함하고, 여기에서 신규한 단백질에서 펩타이드 링커는 융합 단백질의 단백질에서 펩타이드 링커와 이소펩타이드 결합을 형성할 수 있고, 상기 방법은 상기 신규한 단백질 (특히 상기 신규한 단백질내 펩타이드 링커)가 융합 단백질에서 펩타이드 링커와 이소펩타이드 결합 형성을 가능하게 하는 조건하에 상기 융합 단백질과 상기 신규한 단백질의 접촉을 포함한다.
- [0039] 따라서, 일부 구현예에서, 제3 단백질은, 융합 단백질에 첨가되는, "추가" 단백질, 예를 들면 추가의 또는 신규한 단백질로서 보여질 수 있다. 그러므로, 융합 단백질 확장은 상기 방법에서 단계 (c) 반복으로서 보여질 수 있고, 여기에서 추가 단백질에 있어서 펩타이드 링커는 융합 단백질에 첨가된 이전의 단백질을 결합하는데 사용된 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍이다.
- [0040] 일부 구현예에서, 융합 단백질에 첨가되는 신규한 단백질 (즉 추가 또는 추가의 단백질)은 (예를 들면 융합 단백질 사슬의 추가 확장을 허용하기 위해) 적어도 제2 펩타이드 링커를 포함한다. 따라서, 제2 펩타이드 링커 (및 신규한 단백질에서 임의의 추가 펩타이드 링커)는 융합 단백질 및 신규한 단백질을 연결 (접합)하는데 사용된 펩타이드 링커의 쌍과 직교이다.
- [0041] 따라서, 더욱 추가 구현예에서, 상기 융합 단백질의 생산 방법은 상기 융합 단백질의 확장의 단계를 포함할 수 있고, 여기에서 상기 제3 단백질은 제5 펩타이드 링커를 포함하고 상기 방법은 제4 단백질과 상기 융합 단백질의 접촉의 단계를 포함하고, 여기에서 상기 제4 단백질은, 상기 제5 펩타이드 링커 및 상기 제6 펩타이드 링커가 이소펩타이드 결합 형성을 가능하게 하는 조건하에, 제6 펩타이드 링커를 포함하여, 이로써 상기 제3 및 제4 단백질을 연결시켜 상기 융합 단백질을 확장시키고, 여기에서 상기 제5 및 제6 펩타이드 링커는 상기 제3 및 제4 펩타이드 링커로 구성되는 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍을 형성한다.
- [0042] 도 1에서 나타난 바와 같이, 펩타이드 링커의 2개 직교 쌍을 이용하여 다중 단백질 유닛 (예를 들면 3 초과 단백질 유닛, 예를 들면 4, 5, 6, 7, 8, 9, 10 이상 단백질 유닛, 예컨대 12, 15, 20 이상 단백질 유닛)을 포함하는 융합 단백질을 생성하는 것이 가능하다. 따라서, 일부 구현예에서 제5 및 제6 펩타이드 링커로 구성되는 펩타이드 링커의 쌍은 제1 및 제2 펩타이드 링커로 구성되는 펩타이드 링커의 쌍과 동일하다.
- [0043] 따라서, 더욱 추가 구현예에서, 융합 단백질은 추가로 확장될 수 있고, 여기에서 상기 제4 단백질은 제7 펩타이드 링커를 포함하고 상기 방법은 제5 단백질과 상기 융합 단백질의 접촉의 단계를 포함하고, 여기에서 상기 제5 단백질은, 상기 제7 펩타이드 링커 및 상기 제8 펩타이드 링커가 이소펩타이드 결합 형성을 가능하게 하는 조건

하에, 제8 펩타이드 링커를 포함하고, 이로써 상기 제4 및 제5 단백질을 연결시켜 상기 융합 단백질을 확장시키고, 여기에서 상기 제7 및 제8 펩타이드 링커는 상기 제5 및 제6 펩타이드 링커로 구성되는 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍을 형성한다.

[0044] 일부 구현예에서 제7 및 제8 펩타이드 링커로 구성되는 펩타이드 링커의 쌍은 제3 및 제4 펩타이드 링커로 구성되는 펩타이드 링커의 쌍과 동일하다.

[0045] 융합 단백질 사슬이 상기 기재된 단계의 반복에 의해 확장될 수 있다는 것이 분명할 것이고, 예를 들면 여기에서 제5 단백질은 제10 펩타이드 링커를 포함하는 제6 단백질 및 제9 펩타이드 링커를 포함하고 여기에서 상기 제9 및 제10 펩타이드 링커는 상기 제7 및 제8 펩타이드 링커로 구성되는 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍을 형성한다. 일부 구현예에서, 제9 및 제10 펩타이드 링커로 구성되는 펩타이드 링커의 쌍은 제1 및 제3 펩타이드 링커 및/또는 상기 제5 및 제6 펩타이드 링커로 구성되는 펩타이드 링커의 쌍과 동일하다.

[0046] 따라서, 일부 구현예에서, 펩타이드 링커의 적어도 2개의 직교 쌍은 융합 단백질을 형성하기 위해 단백질을 연결 (접합)하는데 대안적으로 사용될 수 있다. 대안적으로 보면, 융합 단백질에 첨가되는 신규한 또는 추가 단백질은 융합 단백질에서 이전에 첨가된 단백질을 연결하는데 사용된 펩타이드 링커의 쌍과 직교인 펩타이드 링커의 쌍의 일부를 형성하는 적어도 하나의 펩타이드 링커를 포함한다.

[0047] 본 발명이 펩타이드 링커의 2개 직교 쌍을 이용하여 성공적으로 작업될 수 있는 동안, 펩타이드 링커의 2개 초과 직교 쌍이 본 발명의 방법 및 용도에서 이용될 수 있다는 것이 분명할 것이다. 따라서, 상기 제공된 대표적인 예의 문맥에 있어서, 일부 구현예에서 제5 및 제6 펩타이드 링커로 구성되는 펩타이드 링커의 쌍은 제1 및 제2 펩타이드 링커로 구성되는 펩타이드 링커의 쌍과 상이하고, 바람직하게는 직교이다. 이하에 논의된 바와 같이, 직교 펩타이드 링커의 2개 초과 쌍의 용도는 복합체 융합 단백질 구조, 예를 들면 분지형 구조의 생산을 가능하게 할 것이다. 따라서, 아래 상세히 논의된 바와 같이, 본 발명자는, 본 발명의 추가 구현예를 형성하는, 펩타이드 링커의 몇 개의 직교 쌍을 개발하고 있다.

[0048] 예를 들어, 3개 단백질, 1, 2 및 3을 포함하는 융합 단백질은 상기 기재된 방법에 따라 생산될 수 있고, 여기에서 단백질 1은 펩타이드 링커 A를 포함하고, 단백질 2는 펩타이드 링커 A' 및 B를 포함하고 단백질 3은 펩타이드 링커 B'를 포함한다. 이 점에 있어서, 펩타이드 링커 A 및 A' (펩타이드 링커의 쌍)은 반응하여 이소펩타이드 결합을 형성하고 펩타이드 링커 B 및 B' (펩타이드 링커의 쌍)은 반응하여 이소펩타이드 결합을 형성하고, 여기에서 펩타이드 링커 쌍 A/A' 및 B/B'는 직교이다 (즉 이소펩타이드 결합을 형성하기 위해 다른 쌍과 반응하지 않는다). 펩타이드 링커의 제3 직교 쌍의 용도는 분지형 구조의 생산을 가능하게 할 것이다. 예를 들어, 단백질 2는 제3 펩타이드 링커 C를 포함할 수 있고 제4 단백질, 4는 펩타이드 링커 C'를 포함할 수 있고, 여기에서 C 및 C' (펩타이드 링커의 쌍)은 반응하여 이소펩타이드 결합을 형성하고 여기에서 펩타이드 링커 A/A', B/B' 및 C/C'는 직교이다. 융합 단백질 1-2-3이 C 및 C'가 이소펩타이드 결합 형성을 가능하게 하는 조건하에 단백질 4와 접촉되는 경우, 수득한 융합 단백질은 분지형, 즉 1-2(-4)-3일 것이다 (참고 도 13A). 대안적으로, 융합 단백질 1-2-4는 B 및 B'가 분지형 융합 단백질, 1-2(-4)-3을 생산하기 위해 이소펩타이드 결합 형성을 가능하게 하는 조건하에 단백질 3과 접촉될 수 있다. 숙련가는 복합체 분지화 구조가 직교 펩타이드 링커의 3개 쌍을 이용하여 생성될 수 있고 분지화 구조의 복잡성이 펩타이드 링커의 추가의 직교 쌍의 용도에 의해 추가로 증가될 수 있다는 것을 이해할 것이다. 특히, 펩타이드 링커의 2개 초과 직교 쌍의 용도는 유익하게는 비대칭 분지화 구조를 생성하는데 사용될 수 있다.

[0049] 따라서, 일부 구현예에서, 본 발명의 방법 및 용도는 펩타이드 링커의 2개 초과 직교 쌍, 예를 들면 펩타이드 링커의 3, 4, 5, 6, 7, 8, 9 또는 10 이상 직교 쌍을 이용한다.

[0050] 분지화는 또한 펩타이드 링커의 2개 직교 쌍을 이용하여 달성될 수 있다. 예를 들어, 5개 단백질, 1-5를 포함하는 분지형 융합 단백질은 단백질 중 하나에서 추가의 펩타이드 링커 포함에 의해 생산될 수 있고, 예를 들면 단백질 2는 펩타이드 링커의 2개 직교 쌍으로부터 4 펩타이드 링커를 포함할 수 있다. 상기 대표적인 구현예에서, 단백질 1은 펩타이드 링커 A를 포함하고, 단백질 2는 펩타이드 링커 A' 및 3 펩타이드 링커 B를 포함한다. 단백질 3, 4 및 5 각각은 펩타이드 링커 B'를 포함하고, 여기에서 펩타이드 링커 A 및 A' (펩타이드 링커의 쌍)은 반응하여 이소펩타이드 결합을 형성하고 펩타이드 링커 B 및 B' (펩타이드 링커의 쌍)은 반응하여 이소펩타이드 결합을 형성하고, 여기에서 펩타이드 링커 쌍 A/A' 및 B/B'는 직교이다. 따라서, 단백질 3-5와 융합 단백질 1-2의 접촉은 단백질 3-5 모두가 서로 독립적으로 단백질 2에 연결되는 분지형 융합 단백질을 초래할 것이다 (참고 도 13B). 단백질 3-5가 동일한 또는 상이한 단백질일 수 있다는 것이 분명할 것이다. 또한, 하나 이상의 단백질 3-5는 융합 단백질의 각각의 분지의 확장 (예를 들면 별도의, 독립적인 확장)을 용이하게 하기 위해 펩타이드

링커의 직교 쌍으로부터 추가의 펩타이드 링커를 포함할 수 있다.

- [0051] 따라서, 일부 구현예에서 융합 단백질은 분지형일 수 있다. 다른 구현예에서, 융합 단백질은 선형일 수 있다. 일부 구현예에서, 예를 들면 펩타이드 링커의 2개 초과 직교 쌍이 사용되는 경우, 융합 단백질은 비대칭 분지로 구성될 수 있다, 즉 융합 단백질은 비대칭 구조를 가질 수 있다. 따라서, 일부 구현예에서, 본 발명은 분지형 융합 단백질의 생산 방법을 제공한다. 일부 구현예에서, 본 발명은 선형 융합 단백질의 생산 방법을 제공한다.
- [0052] 용어 "분지형"은 2개 이상의 단백질 유닛이, 서로 독립적으로, 즉 독립적으로 (개별적으로) 형성된 이소펩타이드 결합을 통해, 융합 단백질의 동일한 내부 단백질 유닛 (비-말단 단백질 유닛)에 연결되는 (결합되는, 접합되는) 융합 단백질을 지칭한다. 내부 단백질 유닛 또는 비-말단 단백질 유닛은 융합 단백질에서 적어도 2개의 다른 단백질 유닛에 이소펩타이드 결합에 의해 연결되는 (결합되는, 접합되는) 단백질로서 한정될 수 있다. 말단 단백질 유닛은 융합 단백질에서 하나의 다른 단백질 유닛에만 이소펩타이드 결합을 통해 연결되는 (결합되는, 접합되는) 단백질로서 한정될 수 있다. 따라서, 상기 논의된 및 도 13에서 보여진 대표적인 예에서, 단백질 1 및 3에 이소펩타이드 결합을 통해 연결되기 때문에 단백질 2는 내부 단백질 유닛 또는 비-말단 단백질 유닛이고, 여기에서 단백질 4 및 5는 융합 단백질의 "분지"로서 보여질 수 있다. 단백질 1, 3, 4 및 5는 말단 단백질 유닛으로서 보여질 수 있다. 따라서, 분지형 융합 단백질은 2개 초과 말단 단백질 유닛을 포함한다.
- [0053] 용어 "선형"은 모든 내부 단백질 유닛이 융합 단백질에서 2개의 다른 단백질 유닛에만 연결되어, 이로써 단백질 유닛의 선형 사슬을 생성하는 융합 단백질을 지칭한다. 따라서, 선형 융합 단백질은 2개의 말단 단백질 유닛만을 포함한다.
- [0054] 더욱 다른 구현예에서, 융합 단백질은 원형일 수 있다. 예를 들어, 상기 융합 단백질, 1-2-3을 고려하여, 단백질 1이 또한 펩타이드 링커 C를 함유하고 단백질 3이 또한 펩타이드 링커 C'를 함유하면, 단백질 1 및 3은 이소펩타이드 결합에 의해 연결될 수 있고, 이로써 원형 단백질을 형성한다. 따라서, 일부 구현예에서, 선형 단백질은 원형가능한, 즉 원형 융합 단백질을 형성할 수 있는 것으로 보여질 수 있다. 이 점에 있어서, 이하에 논의된 바와 같이, 하나 이상의 펩타이드 링커는 그의 반응을 예방 또는 지연하기 위해 차단 또는 보호될 수 있다. 따라서, 상기 예를 이용하여, 펩타이드 링커 C 및/또는 C'가 차단되면, 융합 단백질은 펩타이드 링커가 반응하여 이소펩타이드 결합을 형성하게 하기 위해 C 및/또는 C'의 미차단에 의해 원형화될 수 있고 원형가능한 선형 융합 단백질일 것이다.
- [0055] 따라서, 일부 구현예에서, 본 발명은 원형 또는 원형가능한 융합 단백질의 생산 방법을 제공한다.
- [0056] 따라서, 용어 "원형"은 임의의 말단 단백질 유닛을 함유하지 않는 융합 단백질을 일반적으로 지칭한다. 그러나, 하나 이상의 내부 단백질 유닛이 융합 단백질에서 적어도 3개의 다른 단백질 유닛에 이소펩타이드 결합에 의해 연결되는 원형 융합 단백질을 포함하는, "분지형 원형" 융합 단백질을 생산하는 것이 가능함이 분명할 것이다.
- [0057] 용어 "직교"는 본원에서 사용된 바와 같이 상호간에 무반응성인 분자, 예를 들면 서로 반응할 수 없거나 서로 반응할 수 있는 대응하는 분자와 비교된 경우 감소된 효율로 반응하는 분자를 지칭한다. 본 발명의 펩타이드 링커, 특히 펩타이드 링커의 쌍의 문맥에서, 용어 직교는 이소펩타이드 결합을 형성하기 위해 펩타이드 링커의 다른 쌍과 반응할 수 없는 또는 대응하는 분자, 예를 들면 이소펩타이드 결합을 자발적으로 형성할 수 있는 내인성 단백질과 비교된 경우 감소된 효율로 반응하는 펩타이드 링커의 쌍 또는 이소펩타이드 결합을 형성하기 위해 서로 효율적으로 반응할 수 있는 펩타이드 링커의 쌍을 지칭한다. 반응에 대한 무능은 이소펩타이드 결합을 형성하기 위해 반응하는 샘플에서 펩타이드 링커의 5% 이하, 예를 들면 4%, 3%, 2% 또는 1% 이하로서 보여질 수 있다. 감소된 효율은 이소펩타이드 결합을 형성하기 위한 펩타이드 링커의 각각의 쌍의 능력과 비교된 이소펩타이드 결합을 형성하기 위해 반응하는 직교 펩타이드 링커의 쌍의 5% 미만 효율, 예를 들면 4%, 3%, 2% 또는 1% 미만 효율로서 보여질 수 있다. 반대로, 이소펩타이드 결합을 형성하기 위해 효율적으로 반응하는 펩타이드 링커의 쌍은 적어도 95% 효율, 예를 들면 적어도 96%, 97%, 98%, 99% 또는 100% 효율로 반응할 수 있다, 즉 샘플에서 펩타이드 링커의 쌍의 펩타이드 링커의 적어도 95%는 반응하여 이소펩타이드 결합의 형성을 가능하게 하는 조건하에 이소펩타이드 결합을 형성한다. 예를 들어, 펩타이드 링커의 2개 쌍, A/A' 및 B/B'는 A 및 A'가 이소펩타이드 결합을 형성하기 위해 B 및/또는 B'와 반응할 수 없는 경우 또는 A 및 A'가 B 및/또는 B'와 반응하여 A 및 A' 및/또는 B 및 B' 사이 이소펩타이드 결합 형성과 비교된 경우 5% 미만 효율로 이소펩타이드 결합을 형성하는 경우 직교로서 보여질 수 있다.
- [0058] 대안적으로 보면, 이소펩타이드 결합 형성을 가능 또는 용이하게 하는 조건하에 이소펩타이드 결합을 형성하기 위해 효율적으로 함께 반응하는 2개 펩타이드 링커는 펩타이드 링커의 동족 쌍으로서 한정될 수 있고, 여기에서

용어 "동족"은 함께 기능하는, 즉 함께 반응하여 이소펩타이드 결합을 형성하는 성분을 지칭한다. 따라서, 이소펩타이드 결합 형성을 가능 또는 용이하게 하는 조건하에 이소펩타이드 결합을 형성하기 위해 효율적으로 함께 반응하는 2개 펩타이드 링커는 또한 펩타이드 링커의 "상보적" 쌍으로서 지칭될 수 있다. 이와 같이, 펩타이드 링커의 직교 쌍은 비-동족 쌍 또는 비-상보적 쌍으로서 보여질 수 있다. 예를 들어, 상기 기재된 대표적인 예에 기반하여, 펩타이드 링커 쌍 A/A'는 펩타이드 링커의 동족 또는 상보적 쌍으로서 보여질 수 있고, 반면에 A/A' 및 B/B'는 A 및 A'가 이소펩타이드 결합 형성을 가능 또는 용이하게 하는 조건하에 이소펩타이드 결합을 형성하기 위해 B 및/또는 B'와 효율적으로 반응할 수 없는 한에 있어서는 비-동족 또는 비-상보적 쌍이다.

[0059] 본 발명의 방법 및 용도에서 사용하기 위한 펩타이드 링커는 이소펩타이드 결합을 자발적으로 형성할 수 있는 단백질로부터 유래될 수 있다. 특히, "이소펩타이드 결합을 자발적으로 형성할 수 있는 단백질" (또한 본원에서 "이소펩타이드 단백질"로서 지칭됨)은 효소 또는 다른 서브스틴스의 부재하에 및/또는 화학 변형 없이, 그의 단백질 사슬 내에, 즉 분자내로 이소펩타이드 결합을 형성할 수 있는 것이다. 이소펩타이드 결합을 형성하기 위한 2개의 반응성 잔기는 따라서 단일 단백질 사슬 내에 포함된다. 따라서, 분자간으로, 즉 다른 펩타이드 또는 단백질 사슬 또는 유닛과, 단지 이소펩타이드 결합을 형성하는 단백질은 본 발명에서 사용된 바와 같이 이소펩타이드 단백질인 것으로 고려되지 않는다. 특히, 분자간 이소펩타이드 결합을 갖는 HK97 캡시드 서브유닛은 제외된다.

[0060] 용어 "이소펩타이드 결합"은 본원에서 사용된 바와 같이 카복실 또는 카복사미드기와 아미노기 사이 아미드 결합을 지칭하고 이의 적어도 하나는 단백질 주요 사슬로부터 유래되지 않거나 대안적으로 보면 단백질 골격의 일부가 아니다. 이소펩타이드 결합은 단일 단백질 내에 형성할 수 있거나 2개 펩타이드 또는 펩타이드 및 단백질 사이 발생할 수 있다. 따라서, 이소펩타이드 결합은 단일 단백질 내에 분자내로 또는 분자간으로 즉 2개 펩타이드/단백질 분자 사이, 예를 들면 2개 펩타이드 링커 사이 형성할 수 있다. 전형적으로, 이소펩타이드 결합은 단백질 또는 펩타이드 사슬의 리신 잔기와 아스파라긴, 아스파르트산, 글루타민, 또는 글루탐산 잔기 또는 말단 카복실기 사이에서 발생할 수 있거나 단백질 또는 펩타이드 사슬의 알파-아미노 말단과 아스파라긴, 아스파르트산, 글루타민 또는 글루탐산 사이 발생할 수 있다. 이소펩타이드 결합에서 관여된 쌍의 각각의 잔기는 본원에서 반응성 잔기로서 지칭된다. 본 발명의 바람직한 구현예에서, 이소펩타이드 결합은 리신 잔기와 아스파라긴 잔기 사이 또는 리신 잔기와 아스파르트산 잔기 사이 형성할 수 있다. 특히, 이소펩타이드 결합은 리신의 측사슬 아민과 아스파라긴의 카복사미드기 또는 아스파르테이트의 카복실기 사이 발생할 수 있다.

[0061] 이소펩타이드 결합에서 관여된 잔기 사이 거리는 잔기 내에 특정한 C 원자로부터 측정된다. 따라서, 리신이 이소펩타이드 결합에서 관여되는 경우, 거리는 리신의 C-엡실론 원자로부터 측정되고; 아스파르트산이 이소펩타이드 결합에서 관여되는 경우, 거리는 아스파르트산의 C-감마 원자로부터 측정되고; 아스파라긴이 이소펩타이드 결합에서 관여되는 경우, 거리는 아스파라긴의 C-감마 원자로부터 측정되고 글루탐산이 이소펩타이드 결합에서 관여되는 경우, 거리는 글루탐산의 C-델타 원자로부터 측정된다. 이소펩타이드 결합에서 관여된 반응성 잔기의 (거리가 계산되는) 이들 원자는 본원에서 "관련된 원자"로서 지칭된다.

[0062] 전형적으로, 이소펩타이드 결합이 형성하기 위해, 반응성 잔기 예를 들면 반응성 리신 및 아스파라긴/아스파르테이트 잔기 (및 특히 이의 관련된 원자; 리신에 대하여 C-엡실론 원자 및 아스파라긴/아스파르테이트에 대하여 C-감마 원자)는 공간, 예를 들면 이들이 유래되는 이소펩타이드 단백질에서 서로 매우 근접해서 배치되어야 한다. 따라서, 특히, 반응성 잔기 예를 들면 리신 및 아스파라긴/아스파르테이트 (및 특히 이의 관련된 원자)는 (이들이 유래되는) 폴딩된 단백질에서 서로의 4 옹스트롬 이내이고 서로의 3.8, 3.6, 3.4, 3.2, 3.0, 2.8, 2.6, 2.4, 2.2, 2.0, 1.8 또는 1.6 옹스트롬 이내일 수 있다. 특히, 반응성 잔기 (및 더욱 특히 그의 관련된 원자)는 이들이 유래되는 이소펩타이드 단백질에서 서로의 1.81, 2.63 또는 2.60 옹스트롬 이내일 수 있다.

[0063] 일반적으로 본 발명의 펩타이드 링커가 유래될 수 있는 이소펩타이드 단백질은, 이소펩타이드 결합의 형성에서 관여되는, 2개의 다른 반응성 아미노산 잔기 예를 들면 리신 및 아스파라긴/아스파르테이트에 매우 근접해서 글루탐산 또는 아스파르트산 잔기를 포함할 수 있다. 특히, 글루탐산의 C-델타 원자 또는 아스파르트산 잔기의 C-감마 원자는, 폴딩된 단백질 구조에서, 이소펩타이드 결합에 관여된, 반응성 아스파라긴/아스파르테이트 잔기로부터, 예를 들면 반응성 아스파라긴/아스파르테이트 잔기의 C-감마 원자로부터 5.5 옹스트롬 이내일 수 있다. 예를 들어, 글루탐산 (예를 들면 이의 C-델타 원자)는 이소펩타이드 결합에서 반응성 아스파라긴/아스파르테이트 잔기 예를 들면 이의 C-감마 원자로부터 5.4, 5.2, 5.0, 4.8, 4.6, 4.4, 4.2, 4.0, 3.8, 3.6, 3.4, 3.2 또는 3.0 옹스트롬 이내일 수 있다. 특히, 글루탐산 잔기, 예를 들면 이의 C-델타 원자는 아스파라긴/아스파르테이트 잔기 예를 들면 이의 C-감마 원자로부터 4.99, 3.84 또는 3.73 옹스트롬일 수 있다.

- [0064] 추가로, 글루탐산 잔기, 예를 들면 이의 C-델타 원자는, 이소펩타이드 결합에서 관여된, 반응성 리신 잔기, 예를 들면 이의 C-엡실론 원자의 6.5 옹스트롬 이내, 예를 들어 6.3, 6.1, 5.9, 5.7, 5.5., 5.3, 5.1, 4.9, 4.7, 4.5, 4.3 또는 4.1 옹스트롬 이내일 수 있다. 특히, 글루탐산 잔기, 예를 들면 이의 C-델타 원자는 반응성 리신, 예를 들면 이의 C-엡실론 원자로부터 6.07, 4.80 또는 4.42 옹스트롬일 수 있다.
- [0065] 글루탐산 잔기 (또는 아스파르트산 잔기)는 이전에 논의된 바와 같이 이소펩타이드 결합의 형성 유도를 도울 수 있다.
- [0066] 상기 논의된 바와 같이, 본 발명의 방법 및 용도에서 사용하기 위한 펩타이드 링커는 이소펩타이드 단백질의 반응성 도메인을 2개 또는 3개 도메인으로 분할함으로써 획득될 수 있다. 따라서, 펩타이드 링커의 각각의 쌍은 리신 잔기를 포함하는 펩타이드 및 아스파르트레이트 또는 아스파라긴 잔기를 포함하는 펩타이드로 구성되고, 여기에서 상기 잔기 (즉 리신 및 아스파르트레이트 또는 리신 및 아스파라긴)은 이소펩타이드 결합의 형성에서 관여되고 (즉 반응하여 이소펩타이드 결합을 형성하고), 이로써 상기 펩타이드 링커를 연결 (접합)한다.
- [0067] 일부 바람직한 구현예에서, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성은 자발적이다. 따라서, 펩타이드 링커 중 하나는 펩타이드 링커에서 리신 및 아스파라긴 또는 아스파르트레이트 잔기 사이 이소펩타이드 결합의 형성을 용이하게 하는, 예를 들면 유도하는 또는 촉매화하는 글루탐산 또는 아스파르트산 잔기를 포함한다. 일부 구현예에서, 글루탐산 또는 아스파르트산 잔기는 상기 설명된 근접 기준의 하나 이상을 충족시킨다.
- [0068] 따라서, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성이 자발적인 구현예에서, 펩타이드 링커 중 하나는 펩타이드 태그로서 보여질 수 있고 다른 펩타이드 링커 (즉 이소펩타이드 결합의 형성을 용이하게 하는, 예를 들면 유도하는 또는 촉매화하는 글루탐산 또는 아스파르트산 잔기를 포함하는 링커)는 펩타이드 결합 파트너, 즉 아래 추가로 한정된 바와 같이 펩타이드 태그에 대한 결합 파트너로서 보여질 수 있다.
- [0069] 용어 "자발적"은 본원에서 사용된 바와 같이 존재하는 임의의 다른 제제 (예를 들면 효소 촉매) 없이 및/또는 단백질 또는 펩타이드의 화학 변형 없이, 예를 들면 1-에틸-3-(3-디메틸아미노프로필) 카보디이미드 (EDC)를 이용하는 원상태 화학 결합 또는 화학 커플링 없이 결합, 예를 들면 단백질에서 또는 펩타이드 또는 단백질 사이 (예를 들면 2개 펩타이드 또는 펩타이드 및 단백질, 즉 본 발명의 펩타이드 링커 사이) 형성할 수 있는 이소펩타이드 또는 공유 결합을 지칭한다. 따라서, C-말단 티오에스테르를 갖기 위해 펩타이드 또는 단백질을 변형시키는 원상태 화학 결합은 수행되지 않는다.
- [0070] 따라서, 자발적 이소펩타이드 결합은 단백질이 스스로 단리되는 경우 형성할 수 있거나 공유 또는 이소펩타이드 결합은 단리된 경우 또는 화학 변형 없이 2개 펩타이드 또는 펩타이드 및 단백질 (즉 본 발명의 펩타이드 링커) 사이 형성할 수 있다. 자발적 이소펩타이드 또는 공유 결합은 따라서 효소 또는 다른 외인성 서브스텐스의 부재 하에 또는 화학 변형 없이 저절로 형성할 수 있다. 특히 그러나, 자발적 이소펩타이드 또는 공유 결합은 근접-유도된 방식으로 결합의 형성을 허용하기 위해 결합에서 관여된 펩타이드/단백질 중 하나에서 (즉 펩타이드 링커 중 하나에서) 또는 단백질에서 글루탐산 또는 아스파르트산 잔기의 존재를 요구할 수 있다.
- [0071] 자발적 이소펩타이드 또는 공유 결합은, 예를 들면 1, 2, 3, 4, 5, 10, 15, 20, 25 또는 30 분 이내, 또는 1, 2, 4, 8, 12, 16, 20 또는 24 시간 이내, 본 발명의 펩타이드 링커를 포함하는 2개 이상의 단백질, 예를 들면 펩타이드 태그 및 결합 파트너 사이 접촉 이후 또는 단백질의 생산 거의 직후 형성할 수 있다. 결합은 일정 범위의 조건하에, 예컨대 포스페이트-완충된 염수 (PBS) 또는 트리스-완충된 염수 (TBS)에서 pH 4.0-9.0, 예를 들면 5.0, 5.5, 6.5, 7.0, 7.5, 8.0 또는 8.5에, 그리고 0-40°C, 예를 들면 1, 2, 3, 4, 5, 10, 12, 15, 18, 20, 22 또는 25°C에 형성할 수 있다. 숙련가는 다른 적당한 조건을 쉽게 결정할 수 있을 것이다.
- [0072] 따라서, 일부 구현예에서, "이소펩타이드 결합의 형성을 가능하게 하는 조건하에" 본원에서 한정된 바와 같이 펩타이드 링커를 포함하는 접촉 단백질은 완충된 조건에서, 예를 들면 완충액, 예컨대 PBS 또는 TBS로 평형화되고 있는 고형 상 (예를 들면 칼럼)에서 또는 완충된 용액에서 상기 단백질 접촉을 포함한다. 접촉의 단계는 임의의 적합한 pH, 예컨대 pH 4.0-9.0, 예를 들면 4.5-8.5, 5.0-8.0, 5.5-7.5, 예컨대 약 pH 6.2, 6.4, 6.6, 6.8, 7.0, 7.2, 7.4, 7.6, 7.8 또는 8.0에서일 수 있다. 추가적으로 또는 대안적으로, 접촉의 단계는 임의의 적합한 온도, 예컨대 약 0-40°C, 예를 들면 약 1-39, 2-38, 3-37, 4-36, 5-35, 6-34, 7-33, 8-32, 9-31 또는 10-30°C, 예를 들면 약 10, 12, 15, 18, 20, 22 또는 25°C에서일 수 있다. 숙련가는 조건이 본 발명의 방법에서 사용된 펩타이드 링커의 특징에 의존하여 적응될 필요가 있을 수 있고 조건이 적합한지를 쉽게 결정할 수 있을 것임을 이해할 것이다.
- [0073] 일부 구현예에서, "이소펩타이드 결합의 형성을 가능하게 하는 조건하에" 본원에서 한정된 바와 같이 펩타이드

링커를 포함하는 접촉 단백질은 화학 샤프론, 예를 들면 펩타이드 링커의 반응성을 향상 또는 개선하는 분자의 존재하에 상기 단백질의 접촉을 포함한다. 일부 구현예에서, 화학 샤프론은 TMAO (트리메틸아민 N-옥사이드)이다. 일부 구현예에서, 화학 샤프론, 예를 들면 TMAO는 반응에서 적어도 약 0.2M, 예를 들면 적어도 0.3, 0.4, 0.5, 1.0, 1.5, 2.0 또는 2.5M, 예를 들면 약 0.2-3.0M, 0.5-2.0M, 1.0-1.5M의 농도로 존재한다.

[0074] 일부 구현예에서, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성은 자발적이지 않다, 즉 이소펩타이드 결합의 형성은 반응에 첨가되는 성분에 의해 유도 또는 촉매화된다. 이소펩타이드 결합의 형성을 유도 또는 촉매화하는 성분은 펩타이드, 예를 들면 폴리펩타이드 예컨대 효소, 예컨대 트랜스글루타미나제일 수 있다. 바람직한 구현예에서, 이소펩타이드 결합의 형성을 유도 또는 촉매화하는 성분은 이소펩타이드 단백질로부터 유래된 펩타이드, 즉 펩타이드 링커에서 리신과 아스파라긴 또는 아스파르테이트 잔기 사이 이소펩타이드 결합의 형성을 용이하게 하는, 예를 들면 유도 또는 촉매화하는 글루탐산 또는 아스파르트산 잔기를 포함하는 이소펩타이드 단백질의 도메인 또는 단편일 수 있다. 펩타이드 링커에서 리신과 아스파라긴 또는 아스파르테이트 잔기 사이 이소펩타이드 결합의 형성을 용이하게 하는, 예를 들면 유도 또는 촉매화하는 펩타이드는, 구체적으로, 2개 펩타이드 링커 사이 이소펩타이드 결합의 형성을 있는 그대로 유도할 수 있는 한에 있어서는, 단백질 리가아제 또는 펩타이드 리가아제로서 보여질 수 있다.

[0075] 따라서, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성이 자발적이지 않은 구현예에서, 즉 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도하는 성분 (예를 들면 펩타이드, 예를 들면 펩타이드 리가아제)가 개별적으로 제공되는 구현예에서, 펩타이드 링커 모두는 아래 한정된 바와 같이 펩타이드 태그로서 보여질 수 있다. 따라서, 펩타이드 링커 (펩타이드 태그) 사이 이소펩타이드 결합의 형성을 유도하는 펩타이드는 펩타이드 리가아제 또는 펩타이드 링커 쌍 결합 파트너로서 보여질 수 있다.

[0076] 따라서, 일부 구현예에서, 본 발명은 추가로 상기 단백질 사이 이소펩타이드 결합의 형성을 가능하게 하는 조건하에 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도할 수 있는 성분 (예를 들면 펩타이드)와 연결되도록 단백질의 접촉의 단계를 포함한다. 일부 구현예에서, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도할 수 있는 성분은 상기 단백질내 펩타이드 링커에서 리신 및 아스파라긴 또는 아스파르테이트 잔기 사이 이소펩타이드 결합의 형성을 유도하는 글루탐산 또는 아스파르트산 잔기를 포함하는 펩타이드이다.

[0077] 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도할 수 있는 성분 (예를 들면 펩타이드)는 함께 연결된 단백질이 서로 접촉되는 경우 이전, 이후 또는 동시발생적으로 반응에 첨가될 수 있다. 일부 구현예에서, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도할 수 있는 성분 (예를 들면 펩타이드)는 함께 연결된 단백질이 서로 접촉된 이후 반응에 첨가될 수 있다.

[0078] 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도할 수 있는 성분 (예를 들면 펩타이드)의 용도는 큰 개입 펩타이드 도메인의 존재 없이 융합 단백질의 단백질 유닛이 연결되도록 (접합되도록) 하기 때문에 특히 유리하다. 대안적으로 보면, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도할 수 있는 성분 (예를 들면 펩타이드)의 용도는 작은 펩타이드 링커 (예를 들면 펩타이드 태그), 즉 상기 펩타이드 링커 사이 이소펩타이드 결합을 형성할 수 있는 동족 펩타이드 링커 쌍에서 각각의 펩타이드 링커의 최소 펩타이드 서열의 용도를 용이하게 한다.

[0079] 일부 구현예에서, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도할 수 있는 펩타이드 및 동족 펩타이드 링커의 쌍은 동일한 이소펩타이드 단백질로부터 유래된다.

[0080] 이소펩타이드 결합을 자발적으로 형성할 수 있는 단백질은 적어도 하나의 상기 결합을 형성할 수 있고 1 초과, 예를 들어 2, 3, 4, 5, 6, 7, 8, 9, 10 이상 이소펩타이드 결합을 포함할 수 있다. 특히 1 초과 자발적으로 형성된 이소펩타이드 결합이 단백질 내에 존재하면, 이소펩타이드 단백질로부터 몇 개의 상이한 펩타이드 링커 쌍을 개발하는 것이 가능할 수 있다. 일부 구현예에서, 동일한 이소펩타이드 단백질로부터 유래된 상이한 펩타이드 링커 쌍은 직교일 수 있다. 단일 또는 단지 2개의 이소펩타이드 결합을 포함하는 이소펩타이드 단백질로부터 펩타이드 링커의 각각의 쌍을 개발하는 것이 본 발명에서 바람직하다.

[0081] 하나 이상의 이소펩타이드 결합을 자발적으로 형성할 수 있는 공지된 단백질의 예는 하기를 포함한다: *연쇄상구균 파이오제네스*로부터 Spy0128 (Kang et al, Science, 2007, 318(5856), 1625-8), Spy0125 (Pointon et al, J. Biol. Chem., 2010, 285(44), 33858-66) 및 FbaB (Oke et al, J. Struct Funct Genomics, 2010, 11(2), 167-80), *스타필로코쿠스 아우레우스*의 Cna (Kang et al, Science, 2007, 318 (5856), 1625-8), *엔테로코쿠스 파에칼리스*의 ACE19 단백질 (Kang et al, Science, 2007, 318(5856), 1625-8), *바실러스 세레우스*로부터 BcpA

필린 (Budzik et al, PNAS USA, 2007, 106(47), 19992-7), 연쇄상구균 아갈락티아에로부터 마이너 필린 GBS52 (Kang et al, Science, 2007, 318(5856), 1625-8), 코라이네박테리움 디프테리아에로부터 SpaA (Kang et al, PNAS USA, 2009, 106(40), 16967-71), 연쇄상구균 류탄스로부터 SpaP (Nylander et al, Acta Crystallogr Sect F Struct Biol Cryst Commun., 2011, 67(Pt1), 23-6), RrgA (Izore et al, Structure, 2010, 18(1), 106-15), 연쇄상구균 뉴모니아에로부터 RrgB (El Mortaji et al, J. Biol. Chem., 2010, 285(16), 12405-15) 및 RrgC (El Mortaji et al, J. Biol. Chem., 2010, 285(16), 12405-15), 연쇄상구균 코르도니아로부터 SspB (Forsgren et al, J Mol Biol, 2010, 397(3), 740-51). 상기 논의된 바와 같이, 임의의 이들 단백질은 본 발명의 방법 및 용도에서 사용하기 위해 펩타이드 링커 (특히 동족 펩타이드 링커 쌍)을 생성하는데 사용될 수 있다.

[0082] 융합 단백질을 형성하기 위해 연결된 단백질에서 펩타이드 링커의 배열 또는 순서는 특히 중요하지 않다. 예를 들어, 요망된 융합 단백질의 제1 단백질은 펩타이드 태그 (A)를 포함할 수 있고 제2 단백질은 제1 단백질 (A')에서 펩타이드 태그에 대한 동족인 펩타이드 결합 파트너 및 제3 단백질 (B')에서 펩타이드 태그에 대한 동족인 펩타이드 결합 파트너를 포함할 수 있다. 대안적으로, 요망된 융합 단백질의 제1 단백질은 펩타이드 결합 파트너 (A')를 포함할 수 있고 제2 단백질은 제1 단백질 (A)에서 펩타이드 결합 파트너에 대한 동족인 펩타이드 태그 및 제3 단백질 (B)에서 펩타이드 결합 파트너에 대한 동족인 펩타이드 태그를 포함할 수 있다. 이 점에 있어서, 2개 단백질 (예를 들면 제1 단백질 및 제2 단백질, 추가 단백질을 가진 융합 단백질 등.)을 연결하는데 사용된 펩타이드 링커의 쌍이 융합 단백질을 확장하는데 사용된 펩타이드 링커의 쌍과 직교인 것이 충분하다. 이하에 논의된 바와 같이, 직교 펩타이드 링커는 다양한 방식으로 달성될 수 있다.

[0083] 따라서, 일부 바람직한 구현예에서, 펩타이드 링커 (A/A')의 제1 쌍은 반응성 리신 잔기를 가진 하나의 펩타이드 링커, A, (예를 들면 펩타이드 태그) 및 반응성 아스파르테이트 또는 아스파라긴 잔기를 가진 하나의 펩타이드 링커, A', (예를 들면 펩타이드 결합 파트너)를 포함하고 펩타이드 링커 (B/B')의 제2 쌍은 반응성 아스파르테이트 또는 아스파라긴 잔기를 가진 하나의 펩타이드 링커, B, (예를 들면 펩타이드 태그) 및 반응성 리신 잔기를 가진 하나의 펩타이드 링커, B', (예를 들면 펩타이드 결합 파트너)를 포함한다. 상기 제공된 예를 이용하여, B'와 반응하기 위해 A에 적합한 경로는 없고 A'와 반응하기 위해 B에 적합한 경로는 없다. 따라서, 펩타이드 링커 쌍은 서로 직교이다.

[0084] 추가 구현예에서, 펩타이드 링커 (A/A')의 제1 쌍은 반응성 리신 잔기를 가진 하나의 펩타이드 링커, A, (예를 들면 펩타이드 태그) 및 반응성 아스파르테이트 또는 아스파라긴 잔기를 가진 하나의 펩타이드 링커, A', (예를 들면 펩타이드 결합 파트너)를 포함하고 펩타이드 링커 (B/B')의 제2 쌍은 반응성 리신 잔기를 가진 하나의 펩타이드 링커, B, (예를 들면 펩타이드 태그) 및 반응성 아스파르테이트 또는 아스파라긴 잔기를 가진 하나의 펩타이드 링커, B', (예를 들면 펩타이드 결합 파트너)를 포함한다. 대안적으로, 펩타이드 링커 (A/A')의 제1 쌍은 반응성 아스파르테이트 또는 아스파라긴 잔기를 가진 하나의 펩타이드 링커, A, (예를 들면 펩타이드 태그) 및 반응성 리신 잔기를 가진 하나의 펩타이드 링커, A', (예를 들면 펩타이드 결합 파트너)를 포함하고 펩타이드 링커 (B/B')의 제2 쌍은 반응성 아스파르테이트 또는 아스파라긴 잔기를 가진 하나의 펩타이드 링커, B, (예를 들면 펩타이드 태그) 및 반응성 리신 잔기를 가진 하나의 펩타이드 링커, B', (예를 들면 펩타이드 결합 파트너)를 포함한다. 이들 구현예에서, 펩타이드 링커 (펩타이드 태그) A 및 B는 이들이 적어도 하나의 (예를 들면 2, 3) "앵커" 잔기에서 크기의 실질적인 차이를 갖도록 선택될 수 있어서, A 및 B' 및 B 및 A'의 비-공유 도킹 (즉 A 및 B' 및 B 및 A' 사이 상호작용)이 비효율적이고, 이로써 최소 교차-반응이 있다는 것을 확인한다.

[0085] 용어 "앵커 잔기"는 펩타이드 링커의 소수성 코어를 향하는 및 동족 펩타이드 링커 쌍 (예를 들면 펩타이드 태그)의 다른 펩타이드 링커로부터 반응성 잔기를 수용하는 동족 펩타이드 링커 쌍 (예를 들면 펩타이드 결합 파트너)에서 펩타이드 링커 중 하나의 β -가닥에서 아미노산 잔기를 지칭한다. β -가닥은 용매를 향하는 잔기와 소수성 단백질 코어를 향하는 잔기 사이 교대하고 잔기 배향은 펩타이드 링커가 유래되는 이소펩타이드 단백질에서 자발적 이소펩타이드 결합을 형성하는 도메인의 구조로부터 한정된다. 이는 당해 기술에서 공지된 임의의 적합한 방법, 예를 들면 X-선 결정학, 핵자기 공명 또는 동결-전자 현미경검사에 의해 결정될 수 있다.

[0086] 소형 앵커 잔기는 알라닌 및 발린을 포함한다. 중간 크기 앵커 잔기는 류신, 이소류신 및 메티오닌을 포함한다. 대형 앵커 잔기는 페닐알라닌 및 트립토판을 포함한다. 따라서, 일부 구현예에서 적어도 하나의 소형 앵커 잔기는 중간 크기 또는 대형 앵커 잔기로 대체될 수 있다. 일부 구현예에서 적어도 하나의 중간 크기 앵커 잔기는 소형 또는 대형 앵커 잔기로 대체될 수 있다. 더욱 추가 구현예에서, 적어도 하나의 대형 앵커 잔기는 중간 크기 또는 소형 앵커 잔기로 대체될 수 있다.

[0087] 일부 구현예에서, 펩타이드 링커의 직교 쌍은 상이한 이소펩타이드 단백질 또는 동일한 이소펩타이드 단백질의

상이한 도메인으로부터 유래될 수 있다. 일부 구현예에서, 펩타이드 링커의 직교 쌍은 신규 생산된다.

[0088]

신규 생산되는 펩타이드 링커의 쌍은, 바람직하게는 글루탐산 또는 아스파르트산 잔기와 함께, 이소펩타이드 결합의 자발적 형성을 위하여 2개의 요구된 반응성 아미노산 잔기를 보유해야 한다. 따라서, 상기 기재된 바와 같이, 하나의 펩타이드 링커는 반응성 리신 잔기를 포함하고 다른 펩타이드 링커는 반응성 아스파라긴 또는 아스파르테이트 잔기를 포함한다. 바람직한 구현예에서, 펩타이드 링커 중 하나는 또한 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도 또는 용이하게 하는 글루탐산 또는 아스파르트산 잔기를 포함한다. 그러나, 상기 언급된 바와 같이, 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도 또는 용이하게 하는 글루탐산 또는 아스파르트산 잔기를 포함하는 성분 (예를 들면 펩타이드, 예를 들면 펩타이드 리가아제)는 개별적으로 제공될 수 있다.

[0089]

동족 펩타이드 링커 쌍에서 어떤 펩타이드 링커도 이소펩타이드 결합의 형성에서 관여된 반응성 잔기 모두를 포함하지 않는 것이 분명할 것이다, 즉 펩타이드 링커의 동족 쌍에서 각각의 펩타이드 링커는 하나의 반응성 잔기, 즉 리신 잔기 또는 아스파르테이트/아스파라긴 잔기를 포함한다.

[0090]

펩타이드 링커 중 하나가 상기 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도 또는 용이하게 하는 글루탐산 또는 아스파르트산 잔기를 포함하는 구현예에서, 전형적으로 상기 글루탐산 또는 아스파르트산 잔기는 이소펩타이드 결합에서 관여된 링커에서 잔기의 6.5 옹스트롬 이내, 예를 들면 6.0, 5.5, 5.0, 4.5, 4.0, 3.5 또는 3.0 옹스트롬 이내이다. 이들 거리는 특히 각각의 잔기 내에 관련된 원자, 즉 이소펩타이드 결합 형성에 관여된 원자 사이 거리를 지칭한다. 2개 펩타이드 링커가 서로 근접하는 경우, 예를 들면 제1 및 제2 단백질이 함께 접촉되는 경우, 결합에서 관여된 2개 반응성 잔기 (및 더욱 특히, 그의 관련된 원자)는 공간에서 서로 4 옹스트롬, 바람직하게는 3.8, 3.6, 3.4, 3.2, 3.0, 2.8, 2.6, 2.4, 2.2, 2.0, 1.8 또는 1.6 옹스트롬 이내이어야 한다.

[0091]

숙련가는 이소펩타이드 결합 형성에서 관여된 잔기의 pKa가 또한 신규 이소펩타이드 단백질 설계 경우 고려되어야 한다는 것을 즉시 인지할 것이다. 예를 들어, 바람직하게는, 반응성 리신 잔기는, 중성 pH에서 소수성 코어에서 리신이 매장되는 것을 요구할 수 있는, 반응 이전 탈양성자화되어야 한다.

[0092]

바람직하게는, 펩타이드 링커의 직교 쌍이 상이한 이소펩타이드 단백질 또는 동일한 이소펩타이드 단백질의 상이한 도메인으로부터 유래될 수 있는 동안, 동일한 이소펩타이드 단백질로부터, 특히 이소펩타이드 단백질의 동일한 도메인으로부터 펩타이드 링커의 직교 쌍을 생산하는 것이 가능하다. 예를 들어, 펩타이드 링커의 동족 쌍으로부터 하나의 펩타이드 링커는 쌍에서 다른 펩타이드 링커와 반응하지 않는 (또는 효율적으로 반응하지 않는) 정도로 변형될 수 있다. 변형은, 펩타이드 링커 사이 반응을 예방하는 변형의 반전 또는 제거가 이소펩타이드 결합을 형성하기 위해 효율적으로 반응하도록 펩타이드 링커 쌍의 수용력을 재구성하도록, 가역적일 수 있다. 따라서, 예로써, 동족 펩타이드 링커 쌍 A/A'의 펩타이드 링커 중 하나는 변형될 수 있고, 예를 들면 A는 블록기의 첨가에 의해 변형되어, 펩타이드 링커 B를 생산하고, 여기에서 B는 이소펩타이드 결합을 형성하기 위해 A' 또는 A와 효율적으로 반응할 수 없다. B로부터 블록기의 제거는, 이소펩타이드 결합을 형성하기 위해 A'와 반응할 수 있는, 펩타이드 링커 B'를 초래한다.

[0093]

가역적 또는 제거적 블록기의 용도는 당해 분야에 공지되어 있다. 따라서, 펩타이드 링커의 직교 쌍을 생산하기 위해 펩타이드 링커의 동족 쌍으로부터 하나의 펩타이드 링커에 블록기의 첨가는 펩타이드 링커의 보호기 첨가 또는 펩타이드 링커 케이지로서 보여질 수 있다. 블록 (예를 들면 보호, 마스킹 또는 케이지) 기는 이소펩타이드 결합을 형성하기 위해 펩타이드 링커 쌍의 다른 펩타이드 링커와 효율적으로 반응하도록 펩타이드 링커의 수용력을 재구성하는 당해 기술에서 공지된 임의의 적당한 수단에 의해 제거될 수 있다. 블록기의 제거 (예를 들면 탈보호, 마스킹해제, 케이지해제)는, 블록기의 성질에 의존하여, 화학적, 효소적 또는 광 반응을 통해 달성될 수 있다. 블록기의 적합한 예는 큰 부피의 모이어티, 예컨대 반응을 입체적으로 방해할 수 있는 및 효소, 예컨대 하기의 사용에 의해 제거될 수 있는 단백질을 포함한다: 담배 Etch 바이러스 프로테아제 (하기에서 검토된 바와 같음: Bioorg Med Chem. 2012 Jan 15;20(2):571-82. doi: 10.1016/j.bmc.2011.07.048. Epub 2011 Jul 30. Cleavable linkers in chemical biology. Leriche G, Chisholm L, Wagner A. 테트라진과 반응에 의해 화학적으로 탈케이지되는, 트랜스-사이클로옥텐-케이지된 리신, (N-(((E)-사이클로옥트-2-엔-1-일)-옥시)카보닐-L-리신 (Nat Chem Biol. 2014 Dec;10(12):1003-5. doi: 10.1038/nchembio.1656. Epub 2014 Nov 2. 살아 있는 세포에서 탈케이지하는 딜스-알더 반응-유발된 생물직교 단백질. Li J, Jia S, Chen PR) 또는, 당해 기술에서 잘 알려진 바와 같이, 적절한 파장의 광에 의해 탈케이지되는 o-니트로벤질 또는 쿠마린 기로 케이지된 리신 (참고 예를 들면 Chem Rev. 2013 Jan 9;113(1):119-91. doi: 10.1021/cr300177k. Epub 2012 Dec 21. 화학 및

생물학에서 광제거가능 보호기: 반응 기전 및 효능. Klan, Solomek T, Bochet CG, Blanc A, Givens R, Rubina M, Popik V, Kostikov A, Wirz J.)

- [0094] 블록기의 용도는 펩타이드 링커의 추가의 직교 쌍의 생산에 제한될 필요는 없다. 예를 들어, 블록기는, 예를 들면 멀티플렉스 반응에서, 융합 단백질의 확장을 제어하는데 특히 유용할 수 있다. 예로써, 다중 융합 단백질은 단일 고품 상 기질에서, 예를 들면 다양한 상이한 융합 단백질을 포함하는 어레이를 생산하기 위해 합성될 수 있다. 고품 상에서 각각의 융합 단백질의 물리적 분리는, 예를 들면 핵산 어레이의 생성과 유사한 광-반응성 블록기를 이용하여 기질에서 펩타이드 링커의 선택적 차단해제를 용이하게 할 것이다. 펩타이드 링커의 선택적 차단해제는 하나의 확장 반응에서 (예를 들면 고품 상에서 특이적 위치에 있어서) 단일 융합 단백질, 또는 융합 단백질의 세트의 확장 및 차후의 반응에서 상이한 융합 단백질, 또는 융합 단백질의 세트의 확장을 가능하게 할 것이다.
- [0095] 따라서, 일부 구현예에서, 하나 이상의 펩타이드 링커는 블록기, 즉 가역적 블록기를 포함할 수 있다. 일부 구현예에서, 블록기는 블록기를 제거하는 광, 예를 들면 UV 광, 화학제 또는 효소와 융합 단백질을 접촉시킴으로써 제거될 수 있다.
- [0096] 따라서, 일부 구현예에서, 본 발명의 방법은 융합 단백질에서 펩타이드 링커로부터 블록기의 차단해제 또는 제거의 단계를 포함할 수 있다.
- [0097] 대표적인 구현예에서, 본 발명은 융합 단백질의 생산 (예를 들면 생성, 합성, 어셈블링 등.) 방법을 제공하고, 상기 방법은 하기 단계를 포함한다:
- [0098] a) 상기 단백질 사이 이소펩타이드 결합의 형성을 가능하게 하는 조건하에 제2 단백질과 제1 단백질을 접촉시키는 단계로서, 상기 제1 단백질 및 상기 제2 단백질 각각은 펩타이드 링커를 포함하고, 상기 펩타이드 링커는 연결 단백질을 형성하기 위해 상기 제2 단백질에 상기 제1 단백질을 연결하는 이소펩타이드 결합을 형성하도록 (서로) 반응하는 펩타이드 링커의 쌍인, 단계; 및
- [0099] b) 상기 제3 단백질과 상기 연결 단백질 사이 이소펩타이드 결합의 형성을 가능하게 하는 조건하에 제3 단백질과 (a)로부터의 연결 단백질을 접촉시키는 단계로서, 상기 제3 단백질은 (a)로부터 연결 단백질에서 추가 펩타이드 링커와 반응하는 펩타이드 링커를 포함하고, 상기 펩타이드 링커는 융합 단백질을 형성하기 위해 상기 연결 단백질에 상기 제3 단백질을 연결하는 이소펩타이드 결합을 형성하도록 (서로) 반응하는 펩타이드 링커의 쌍인, 단계,
- [0100] 여기에서 (a)로부터 펩타이드 링커의 상기 쌍은 (b)로부터 펩타이드 링커의 상기 쌍과 직교이고
- [0101] 연결 단백질에서 추가 펩타이드 링커는 블록기를 포함하고 상기 제3 단백질과 상기 연결 단백질 사이 이소펩타이드 결합의 형성을 가능하게 하는 조건은 블록기 제거에 연결 단백질 처리를 포함함.
- [0102] 일부 구현예에서 블록기는 상기 제3 단백질과 연결 단백질의 접촉의 단계 이전 제거될 수 있다 (펩타이드 링커는 차단해제될 수 있다). 일부 구현예에서, 블록기는 상기 제3 단백질과 연결 단백질의 접촉의 단계 이후 또는 동시발생적으로 제거될 수 있다 (펩타이드 링커는 차단해제될 수 있다).
- [0103] 용어 "펩타이드 링커"는 본원에서 사용된 바와 같이 일반적으로 이소펩타이드 단백질로부터 직접적으로 설계 또는 유래될 수 있는 펩타이드, 올리고펩타이드 또는 폴리펩타이드를 지칭하고, 예를 들면 펩타이드 링커는 이소펩타이드 단백질의 단편 또는 이의 변형일 수 있다. 펩타이드, 올리고펩타이드 및 폴리펩타이드에 의해 의미되는 크기 경계에 관한 표준 정의는 없지만, 전형적으로 펩타이드는 2-20 개의 아미노산을 포함하는 것, 및 21-39 개의 아미노산인 올리고펩타이드로서 보여질 수 있고 폴리펩타이드는 적어도 40 개의 아미노산을 포함하는 것으로 보여질 수 있다. 따라서, 본원에서 한정된 바와 같은 펩타이드 링커는 적어도 6 개의 아미노산, 예를 들면 6-300 개의 아미노산을 포함하는 것으로 보여질 수 있다.
- [0104] 일부 구현예에서, 펩타이드 링커는 펩타이드 태그로서 지칭될 수 있고 6-50 개의 아미노산 길이, 예를 들면 7-45, 8-40, 9-35, 10-30, 11-25 개의 아미노산 길이일 수 있고, 예를 들면 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20 개의 아미노산을 포함할 수 있거나 상기로 구성될 수 있다. 펩타이드 링커 또는 태그는 특이적으로 이소펩타이드 결합을 통해 제2 펩타이드 링커에 공유적으로 결합하고, 여기에서 또 다른 펩타이드 링커는, 아래 한정된 바와 같이, 펩타이드 태그 또는 펩타이드 결합 파트너로서 보여질 수 있다. 이소펩타이드 결합을 형성하기 위해 (예를 들면 구체적으로 및 효율적으로) 서로 반응하는 2개 펩타이드 링커 (예를 들면 펩타이드 태그 및 펩타이드 태그 또는 펩타이드 태그 및 펩타이드 결합 파트너)는 펩타이드 링커의 쌍, 특히

펩타이드 링커의 동족 쌍으로서 한정될 수 있다.

- [0105] 따라서, 상기에서 언급된 바와 같이, 펩타이드 링커는, 이소펩타이드 결합의 형성에서 관여되는, 적어도 하나의 아미노산 잔기, 예를 들면 리신 또는 아스파라긴/아스파르테이트를 포함해야 한다. 따라서, 펩타이드 링커의 쌍에서 각각의 펩타이드 링커는 이소펩타이드 결합의 형성에서 관여되는 상이한, 즉 상보적, 반응성 아미노산 잔기를 포함해야 한다, 즉 하나의 펩타이드 링커는 리신 잔기를 포함하고 다른 펩타이드 링커는 아스파라긴 또는 아스파르테이트 잔기를 포함한다.
- [0106] 일부 구현예에서, 펩타이드 링커의 쌍은 2개 펩타이드 태그를 포함한다. 전형적으로, 2개 펩타이드 태그는 이소펩타이드 결합을 형성하기 위해 자발적으로 반응하지 않는다, 즉 이들은 상기 한정된 바와 같이 상기 펩타이드 태그/링커 사이 이소펩타이드 결합의 형성을 유도 또는 촉매화하는 성분 (예를 들면 펩타이드, 예를 들면 펩타이드 리가아제)의 첨가를 요구한다.
- [0107] 일부 구현예에서, 펩타이드 링커 (즉 펩타이드 링커의 동족 쌍에서 펩타이드 링커 중 하나)는, 이소펩타이드 단백질로부터 유래 또는 설계되는 및 이소펩타이드 결합을 통해 (바람직하게는 자발적 반응을 통해) 펩타이드 태그에 공유적으로 결합할 수 있는 펩타이드 (특히 올리고펩타이드 또는 폴리펩타이드)로서 한정될 수 있는, 펩타이드 결합 파트너로서 지칭될 수 있다. 일부 구현예에서, 펩타이드 결합 파트너는 공유적으로 결합하는 펩타이드 태그, 즉 그의 대응하는 펩타이드 태그 또는 링커로서 동일한 이소펩타이드 단백질로부터 설계 또는 유래될 수 있다.
- [0108] 일반적으로, 펩타이드 결합 파트너는 그의 대응하는 펩타이드 태그보다 더 크고 펩타이드 태그와 비교된 이소펩타이드 단백질의 더 큰 단편 또는 부분을 포함하거나 상기로 구성된다. 특히, 이소펩타이드 결합의 형성에서 관여되는 잔기 (즉 리신 또는 아스파라긴/아스파르테이트) 포함에 더하여 펩타이드 결합 파트너는 펩타이드 링커, 예를 들면 펩타이드 태그 및 펩타이드 결합 파트너 사이 이소펩타이드 결합의 형성을 유도 또는 용이하게 하는 글루탐산 또는 아스파르트산 잔기를 포함한다.
- [0109] 따라서, 펩타이드 결합 파트너는 펩타이드 태그를 구성하도록 설계된 단편과 중첩하는 이소펩타이드 단백질의 단편을 포함할 수 있거나 펩타이드 태그의 것과 비교된 이소펩타이드 단백질의 별도의 및 별개의 단편을 포함할 수 있다. 따라서, 펩타이드 결합 파트너의 서열은 설계된 펩타이드 태그의 것과 중첩할 수 있거나 펩타이드 태그 및 펩타이드 결합 파트너는 이소펩타이드 단백질의 2개 별도의 단편을 포함할 수 있거나 상기로 구성될 수 있다. 일부 구현예에서, 펩타이드 태그는 이소펩타이드 단백질의 서열에 기반되지 않을 수 있고, 예를 들면 펩타이드 태그 (펩타이드 링커)는 신규 설계될 수 있다.
- [0110] 펩타이드 결합 파트너의 크기에 관해 특정한 제한이 없는 동안, 실제로 본 발명의 방법 및 용도에서 사용하기 위한 펩타이드 링커의 크기를 최소화하는 것이 바람직하다.
- [0111] 따라서, 일부 구현예에서, 펩타이드 링커 (예를 들면 펩타이드 결합 파트너)는 50-300 아미노산 길이, 예를 들면 60-250, 70-225, 80-200 개의 아미노산 길이일 수 있고, 예를 들면 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 또는 200 개의 아미노산을 포함할 수 있거나 상기로 구성될 수 있다.
- [0112] 따라서, 일부 구현예에서, 펩타이드 링커의 쌍은 펩타이드 태그 및 펩타이드 결합 파트너를 포함하고, 여기에서 상기 펩타이드 링커는 자발적으로 반응하여 이소펩타이드 결합을 형성한다.
- [0113] 펩타이드 링커 사이 이소펩타이드 결합 형성이 자발적이지 않은 경우 (예를 들면 펩타이드 링커 둘 모두가 펩타이드 태그인 경우) 상기 펩타이드 태그/링커 사이 이소펩타이드 결합의 형성을 유도 또는 촉매화하는 펩타이드는 상기 한정된 바와 같이 이소펩타이드 단백질 또는 펩타이드 결합 파트너로부터 유래된 펩타이드 (예를 들면 펩타이드 리가아제)로서 보여질 수 있다. 특히, 펩타이드는 펩타이드 링커 사이 이소펩타이드 결합의 형성을 유도 또는 용이하게 하는 글루탐산 또는 아스파르트산 잔기를 포함하지만, 중요하게는 리가아제는 펩타이드 링커 쌍에서 펩타이드 링커 중 어느 하나와 이소펩타이드 결합을 형성하기 위해 반응하는 아미노산 잔기를 함유하지 않는다. 일부 구현예에서, 펩타이드 리가아제는 50-300 개의 아미노산 길이, 예를 들면 60-250, 70-225, 80-200 개의 아미노산 길이일 수 있고, 예를 들면 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 또는 200 개의 아미노산을 포함할 수 있거나 상기로 구성될 수 있다.
- [0114] 따라서, 펩타이드 링커 (예를 들면 펩타이드 태그 및/또는 펩타이드 결합 파트너)는 그러므로 이소펩타이드 단백질의 전체 단백질 서열로 구성되지 않고 길이가 더 짧다. 예를 들어, 펩타이드 링커는 이소펩타이드 단백질에

서 존재하는 아미노산 잔기의 수의 5, 10, 20, 30, 40 또는 50% 미만을 포함할 수 있다.

- [0115] 펩타이드 링커 또는 펩타이드 링커의 쌍이 이소펩타이드 단백질 (특히 하나 이상의 이의 단편)의 서열에 기반될 수 있는 동안, 펩타이드 링커의 서열이 유래되는 이소펩타이드 단백질의 부분의 서열과 상이할 수 있다는 것이 숙련가에 의해 쉽게 이해될 것이다. 따라서, 일부 구현예에서 펩타이드 링커 또는 펩타이드 링커의 쌍은 유래되는 이소펩타이드 단백질의 서열과 비교된 경우 돌연변이 또는 변경을 포함할 수 있다. 이하에 논의된 바와 같이, 일부 돌연변이는 펩타이드 링커의 안정성 및/또는 기능을 개선하기 위해, 예를 들면 펩타이드 링커 사이 자발적 이소펩타이드 결합 형성의 반응 속도를 개선하기 위해 펩타이드 링커 서열에 도입될 수 있다.
- [0116] 따라서, 일부 구현예에서, 펩타이드 링커는 이소펩타이드 단백질의 단편을 포함할 수 있거나 상기로 구성될 수 있고, 여기에서 상기 단편은 상기 제시된 크기 기준을 충족시키고 유래된 이소펩타이드 단백질의 비교할만한 영역과 적어도 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100% 서열 동일성을 포함한다.
- [0117] 또한, 상기 언급된 바와 같이, 이소펩타이드 단백질은 공지된 이소펩타이드 단백질, 즉 공지된 이소펩타이드 단백질과 서열 유사성 또는 동일성을 가진 단백질의 구조적 동족체 검색에 의해 확인될 수 있다. 그와 같은 동족체는 기능적으로 동등한 단백질로서 보여질 수 있고 본 발명의 펩타이드 링커의 생산에서 유용성을 찾을 수 있다.
- [0118] 일부 구현예에서, 본 발명의 방법 및 용도에서 사용하기 위한 펩타이드 링커의 쌍은 임의의 적합한 이소펩타이드 단백질로부터 유래될 수 있다. 상기에서 언급된 바와 같이, 다양한 이소펩타이드 단백질은 당해 기술에 공지되어 있다. 예를 들어, 펩타이드 링커는, 서열 식별 번호. 23에서 설명된 바와 같은 아미노산 서열을 갖는, 주요 필린 단백질 Spy0128로부터 유래될 수 있고 서열 식별 번호. 24에서 설명된 바와 같은 뉴클레오타이드 서열에 의해 인코딩된다. 2개 이소펩타이드 결합은 단백질에서 형성된다. 1개 이소펩타이드 결합은 서열 식별 번호. 23내 위치 179에서 리신과 서열 식별 번호. 23내 위치 303에서 아스파라긴 (반응성 잔기) 사이 형성된다. 자발적 이소펩타이드 결합을 유도하는 글루탐산 잔기는 서열 식별 번호. 23내 위치 258에서 발견된다. 따라서, 서열 식별 번호: 23에서 제시된 이소펩타이드 단백질로부터 발생된 펩타이드 링커의 쌍은 바람직하게는 위치 303에서 반응성 아스파라긴을 포함하는 단백질의 단편을 포함하는 펩타이드 링커 및 위치 179에서 반응성 리신을 포함하는 단백질의 단편을 포함하는 펩타이드 링커를 포함할 것이다. 일부 구현예에서, 펩타이드 링커 중 하나는 위치 258에서 글루탐산 잔기를 또한 함유하는 단편을 포함할 것이다. 일부 구현예에서, 위치 258에서 글루탐산 잔기를 포함하는 단백질의 단편은 개별적으로, 즉 상기 한정된 바와 같은 펩타이드 리가아제로서 제공될 수 있다.
- [0119] 주요 필린 단백질 Spy0128에서 또 다른 이소펩타이드 결합은 서열 식별 번호. 23의 위치 36에서 리신 잔기와 서열 식별 번호. 23의 위치 168에서 아스파라긴 잔기 사이 발생한다. 이소펩타이드 형성을 유도하는 글루탐산 잔기는 서열 식별 번호. 23내 위치 117에서 발견된다. 따라서, 서열 식별 번호: 23에서 제시된 이소펩타이드 단백질로부터 발생된 펩타이드 링커의 쌍은 바람직하게는 위치 36에서 반응성 리신 잔기를 포함하는 단백질의 단편을 포함하는 펩타이드 링커 및 위치 168에서 반응성 아스파라긴을 포함하는 단백질의 단편을 포함하는 펩타이드 링커를 포함할 것이다. 일부 구현예에서, 펩타이드 링커 중 하나는 위치 117에서 글루탐산 잔기를 또한 함유하는 단편을 포함할 것이다. 일부 구현예에서, 위치 117에서 글루탐산 잔기를 포함하는 단백질의 단편은 개별적으로, 즉 상기 한정된 바와 같은 펩타이드 리가아제로서 제공될 수 있다.
- [0120] *E. 파에칼리스*로부터 부착소 단백질의 도메인인, ACE19는 또한 자발적으로 이소펩타이드 결합을 형성한다. ACE19는 서열 식별 번호. 27에서 제시된 바와 같은 아미노산 서열을 갖고 서열 식별 번호. 28에서 제시된 바와 같은 뉴클레오타이드 서열에 의해 인코딩된다.
- [0121] 이소펩타이드 결합은 서열 식별 번호. 27의 위치 181에서 리신 잔기와 서열 식별 번호. 27의 위치 294에서 아스파라긴 잔기 사이 발생한다. 결합은 서열 식별 번호. 27내 위치 213에서 아스파르트산 잔기에 의해 유도된다. 따라서, 서열 식별 번호: 27에서 제시된 이소펩타이드 단백질로부터 발생된 펩타이드 링커의 쌍은 바람직하게는 위치 294에서 반응성 아스파라긴 잔기를 포함하는 단백질의 단편을 포함하는 펩타이드 링커 및 위치 181에서 반응성 리신 잔기를 포함하는 단백질의 단편을 포함하는 펩타이드 링커를 포함할 것이다. 일부 구현예에서, 펩타이드 링커 중 하나는 위치 213에서 아스파르트산 잔기를 또한 함유하는 단편을 포함할 것이다. 일부 구현예에서, 위치 213에서 아스파르트산 잔기를 포함하는 단백질의 단편은 개별적으로, 즉 상기 한정된 바와 같은 펩타이드 리가아제로서 제공될 수 있다.
- [0122] 서열 식별 번호. 29에서 설명된 아미노산 서열을 갖는 *S. 아우레우스*로부터 콜라겐 결합 도메인은 하나의 자발적으로 형성된 이소펩타이드 결합을 포함한다. 이소펩타이드 결합은 서열 식별 번호. 29의 위치 176에서 리신과

서열 식별 번호. 29의 위치 308에서 아스파라긴 사이 발생한다. 이소펩타이드 결합을 유도하는 아스파르트산 잔기는 서열 식별 번호. 29의 위치 209이다. 따라서, 서열 식별 번호: 29에서 제시된 이소펩타이드 단백질로부터 발생된 펩타이드 링커의 쌍은 바람직하게는 위치 176에서 반응성 리신을 포함하는 단백질의 단편을 포함하는 펩타이드 링커 및 위치 308에서 반응성 아스파라긴을 포함하는 단백질의 단편을 포함하는 펩타이드 링커를 포함할 것이다. 일부 구현예에서, 펩타이드 링커 중 하나는 위치 209에서 아스파르트산 잔기를 또한 함유하는 단편을 포함할 것이다. 일부 구현예에서, 위치 209에서 아스파르트산 잔기를 포함하는 단백질의 단편은 개별적으로, 즉 상기 한정된 바와 같은 펩타이드 리가아제로서 제공될 수 있다.

[0123] *연쇄상구균 파이오제네스*로부터 FbaB는, 서열 식별 번호. 25에서 설명된 아미노산 서열을 갖고, 서열 식별 번호. 26에서 설명된 뉴클레오타이드 서열에 의해 인코딩되고 하나의 자발적으로 형성된 이소펩타이드 결합을 포함하는, 도메인, CnaB2를 포함한다. CnaB2 도메인에서 이소펩타이드 결합은 서열 식별 번호. 25의 위치 15에서 리신과 서열 식별 번호. 25의 위치 101에서 아스파르트산 잔기 사이 형성된다. 이소펩타이드 결합을 유도하는 글루탐산 잔기는 서열 식별 번호. 25의 위치 61이다. 따라서, 서열 식별 번호: 25에서 제시된 이소펩타이드 단백질로부터 발생된 펩타이드 링커의 쌍은 바람직하게는 위치 15에서 반응성 리신을 포함하는 단백질의 단편을 포함하는 펩타이드 링커 및 위치 101에서 반응성 아스파르트산을 포함하는 단백질의 단편을 포함하는 펩타이드 링커를 포함할 것이다. 일부 구현예에서, 펩타이드 링커 중 하나는 위치 61에서 글루탐산 잔기를 또한 함유하는 단편을 포함할 것이다. 일부 구현예에서, 위치 61에서 글루탐산 잔기를 포함하는 단백질의 단편은 개별적으로, 즉 상기 한정된 바와 같은 펩타이드 리가아제로서 제공될 수 있다 (예를 들면 서열 식별 번호: 34).

[0124] RrgA 단백질은, 서열 식별 번호. 21에서 설명된 바와 같은 아미노산 서열을 갖고 서열 식별 번호. 22에서 설명된 바와 같은 뉴클레오타이드 서열에 의해 인코딩되는, *연쇄상구균 뉴모니아*로부터 접착 단백질이다. 이소펩타이드 결합은 서열 식별 번호. 21내 위치 742에서 리신과 서열 식별 번호. 21내 위치 854에서 아스파라긴 사이 형성된다. 결합은 서열 식별 번호. 21내 위치 803에서 글루탐산 잔기에 의해 유도된다. 따라서, 서열 식별 번호: 21에서 제시된 이소펩타이드 단백질로부터 발생된 펩타이드 링커의 쌍은 바람직하게는 위치 854에서 반응성 아스파라긴을 포함하는 단백질의 단편을 포함하는 펩타이드 링커 및 위치 742에서 반응성 리신을 포함하는 단백질의 단편을 포함하는 펩타이드 링커를 포함할 것이다. 일부 구현예에서, 펩타이드 링커 중 하나는 위치 803에서 글루탐산 잔기를 또한 함유하는 단편을 포함할 것이다. 일부 구현예에서, 위치 803에서 글루탐산 잔기를 포함하는 단백질의 단편은 개별적으로, 즉 상기 한정된 바와 같은 펩타이드 리가아제로서 제공될 수 있다.

[0125] PsCs 단백질은, 서열 식별 번호. 31에서 설명된 바와 같은 아미노산 서열을 갖고 서열 식별 번호. 32에서 설명된 바와 같은 뉴클레오타이드 서열에 의해 인코딩되는, *연쇄상구균 인터메디우스*로부터 포르(por) 분비 시스템 C-말단 분류 도메인 단백질의 단편이다. 이소펩타이드 결합은 서열 식별 번호. 31내 위치 405에서 리신과 서열 식별 번호. 31내 위치 496에서 아스파르트레이트 사이 형성된다. 따라서, 서열 식별 번호: 31에서 제시된 이소펩타이드 단백질로부터 발생된 펩타이드 링커의 쌍은 바람직하게는 위치 496에서 반응성 아스파르트레이트를 포함하는 단백질의 단편을 포함하는 펩타이드 링커 및 위치 405에서 반응성 리신을 포함하는 단백질의 단편을 포함하는 펩타이드 링커를 포함할 것이다.

[0126] 따라서, 일부 구현예에서, 본 발명의 방법에서 사용하기 위한 펩타이드 링커의 쌍은 서열 식별 번호: 21, 23, 25, 27, 29 또는 31의 어느 하나에서 제시된 바와 같은 아미노산 서열을 포함하는 이소펩타이드 단백질 또는 서열 식별 번호: 21, 23, 25, 27, 29 또는 31의 어느 하나에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 단백질로부터 유래될 수 있다.

[0127] 일부 구현예에서, 상기 이소펩타이드 단백질 서열은 비교되는 서열 (서열 식별 번호: 21, 23, 25, 27, 29 또는 31)과 적어도 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100% 동일성이다.

[0128] 바람직하게는, 상기 한정된 이소펩타이드 단백질로부터 유래된 펩타이드 링커는 상기 기재된 크기 및 서열 동일성 기준을 충족시킨다.

[0129] 서열 동일성은 당해 기술에서 공지된 임의의 적당한 수단에 의해, 예를 들면 가변성 팜팩터(pamfactor), 및 12.0에서 갭 창출 패널티 세트 및 4.0에서 갭 연장 패널티 세트, 및 2개 아미노산의 윈도우를 가진 FASTA pep-cmp를 이용하는 SWISS-PROT 단백질 서열 데이터뱅크를 이용하여 결정될 수 있다. 아미노산 서열 동일성을 결정하기 위한 다른 프로그램은 위스콘신 대학교로부터 유전학 컴퓨터 그룹 (Genetics Computer Group; GCG) 버전 10 소프트웨어 패키지의 베스트핏(BestFit) 프로그램을 포함한다. 상기 프로그램은 하기 디폴트 값을 가진 스미스 앤드 워터맨(Smith and Waterman)의 국부 상동성 알고리즘을 이용한다: 갭 창출 패널티 - 8, 갭 연장 패널티 = 2, 평균 매치 = 2.912, 평균 미스매치 = -2.003.

- [0130] 바람직하게는 상기 비교는 서열의 전체 길이에 대해 실시되지만, 비교의 더 작은 윈도우, 예를 들면 200, 100 또는 50 개 미만 인접 아미노산에 대해 실시될 수 있다.
- [0131] 바람직하게는 상기 서열 동일성-관련된 단백질은 인용된 서열 식별 번호에서 제시되는 폴리펩타이드와 기능적으로 동등하다. 본원에서 지칭되는 바와 같이, "기능성 동등물"은 모 분자 (즉 서열 상동성을 보여주는 분자)에 비해 이소펩타이드 결합의 자발적으로 형성함에 있어서 일부 감소된 효능을 보여줄 수 있는 상기 논의된 이소펩타이드 단백질의 동족체를 지칭하지만, 바람직하게는 효율적이거나 더욱 효율적이다.
- [0132] 일부 구현예에서, 펩타이드 링커의 직교 쌍은 상기 한정된 이소펩타이드 단백질의 임의의 2개 이상으로부터 유래될 수 있다. 바람직한 구현예에서, 펩타이드 링커의 제1 쌍은 서열 식별 번호: 21에서 제시된 바와 같은 아미노산 서열을 갖는 이소펩타이드 단백질로부터 유래되고 펩타이드 링커의 제2, 직교, 쌍은 서열 식별 번호: 25에서 제시된 바와 같은 아미노산 서열을 갖는 이소펩타이드 단백질로부터 유래된다. 이하에 논의된 바와 같이, 일부 구현예에서, 펩타이드 링커의 2개 직교 쌍은 동일한 이소펩타이드 단백질, 예를 들면 서열 식별 번호: 21로부터 유래될 수 있다. 펩타이드 링커의 다른 직교 쌍은 서열 식별 번호: 21 및 23, 21 및 27, 21 및 29, 21 및 31, 25 및 27, 25 및 29 또는 25 및 31에서 제시된 바와 같은 아미노산 서열을 갖는 이소펩타이드 단백질로부터 유래될 수 있다. 숙련가는 펩타이드 링커의 임의의 2개 쌍이 본원에서, 특히 실시예에서 개시된 방법에 기반된 직교인지를 쉽게 결정할 수 있을 것이다. 예를 들어, 펩타이드 링커의 상이한 쌍으로부터 펩타이드 링커의 다양한 조합은, 예를 들면 용액에서, 적합한 기간, 예를 들면 1-24 시간 동안, pH 4-9, 예를 들면 pH 7에서, 1-40°C, 예를 들면 25°C에서 예를 들면 PBS에서 이소펩타이드 결합 형성을 용이하게 하는 조건하에 접촉될 수 있다. 샘플은, 임의의 펩타이드 링커가 반응되는지를 결정하기 위해, 예를 들면 겔 전기영동 (예를 들면 SDS-PAGE)에 의해, 즉 접합된 펩타이드 조사에 의해 분석될 수 있다, 참고 예를 들면 도 7. 따라서, 본 발명의 방법에서 사용하기 위한 펩타이드 링커의 직교 쌍은 이소펩타이드 단백질의 임의의 적합한 조합으로부터 유래될 수 있다.
- [0133] 본 발명자는 유익하게는 본 발명의 방법 및 용도에서 특정한 유용성을 갖는 펩타이드 링커의 쌍을 개발하였다. 이 점에 있어서, 본 발명자는 펩타이드 링커 쌍이 상기 한정된 바와 같이 RrgA 단백질로부터 유래될 수 있다는 것을 확정하였다. 그러나, 아래 실시예에서 상세히 기재된 바와 같이, 본 발명자는 펩타이드 링커의 반응성을 개선하기 위해 원상태 RrgA 서열에 비해 펩타이드 링커 내부에 돌연변이를 도입하였다. 구체적으로, 글리신 잔기는 β -가닥을 안정화시키기 위해 트레오닌 잔기로 대체되었고 아스파르트레이트 잔기는 반응 부위에 근접한 헤어핀 턴을 안정화시키기 위해 글리신 잔기로 대체되었다.
- [0134] 따라서, 본 발명은 하기를 포함하는 펩타이드 링커를 제공한다:
- [0135] (i) 서열 식별 번호: 1에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 1에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 9에서 리신 잔기를 갖는, 서열; 또는
- [0136] (ii) 서열 식별 번호: 2에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 2에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 55에서 글루타메이트 또는 아스파르트레이트 잔기, 위치 94에서 트레오닌 잔기, 위치 100에서 글리신 잔기 및 위치 106에서 아스파라긴 또는 아스파르트레이트 잔기를 포함하는, 서열.
- [0137] 일부 구현예에서, (i)에서 펩타이드 링커는 서열 식별 번호: 38에서 제시된 바와 같은 아미노산 서열을 포함하고/하거나 (ii)에서 펩타이드 링커는 서열 식별 번호: 39에서 제시된 바와 같은 아미노산 서열을 포함한다.
- [0138] 추가 구현예에서, 본 발명은 하기를 포함하는 펩타이드 링커를 제공한다:
- [0139] (i) 서열 식별 번호: 5에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 5에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 8에서 아스파르트레이트 또는 아스파라긴 잔기를 포함하는, 서열; 또는
- [0140] (ii) 서열 식별 번호: 6에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 6에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 8에서 리신 잔기를 포함하는, 서열.
- [0141] 일부 구현예에서, (i)에서 펩타이드 링커는 서열 식별 번호: 42에서 제시된 바와 같은 아미노산 서열을 포함하고/하거나 (ii)에서 펩타이드 링커는 서열 식별 번호: 43에서 제시된 바와 같은 아미노산 서열을 포함한다.

- [0142] 더욱 추가 구현예에서, 본 발명은 하기를 포함하는 펩타이드 링커를 제공한다:
- [0143] (i) 서열 식별 번호: 9에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 9에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 17에서 아스파라긴 또는 아스파르테이트 잔기를 포함하는, 서열; 또는
- [0144] (ii) 서열 식별 번호: 10에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 10에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 9에서 리신 잔기 및 위치 70에서 글루타메이트 또는 아스파르테이트 잔기를 포함하는, 서열.
- [0145] 일부 구현예에서, (i)에서 펩타이드 링커는 서열 식별 번호: 109에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 109에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열을 포함하고, 여기에서 상기 아미노산 서열은 위치 17에서 아스파라긴 또는 아스파르테이트 잔기, 위치 11에서 글리신 잔기 및 바람직하게는 위치 20에서 이소류신 잔기, 위치 21 및 22에서 프롤린 잔기 및 위치 23에서 리신 잔기를 포함한다.
- [0146] 일부 구현예에서, (i)에서 펩타이드 링커는 서열 식별 번호: 46에서 제시된 바와 같은 아미노산 서열을 포함하고/하거나 (ii)에서 펩타이드 링커는 서열 식별 번호: 47에서 제시된 바와 같은 아미노산 서열을 포함한다.
- [0147] 일부 구현예에서, 상기 펩타이드 링커 서열은 비교되는 서열 (서열 식별 번호: 1, 2, 5, 6, 9, 10 또는 109)와 적어도 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100% 동일성이다.
- [0148] 바람직한 구현예에서, 상기 각각의 일부 (i)에서 한정된 펩타이드 링커는 상기 각각의 일부 (ii)에서 한정된 바와 같은 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있다. 예를 들어, 서열 식별 번호: 1 또는 그의 변이체에서 제시된 바와 같은 아미노산 서열을 포함하는 펩타이드 링커는 서열 식별 번호: 2 또는 그의 변이체에서 제시된 바와 같은 아미노산 서열을 포함하는 펩타이드 링커와 이소펩타이드 결합을 자발적으로 형성할 수 있다. 유사하게, 서열 식별 번호: 5 및 6 또는 그의 변이체를 포함하는 펩타이드는 서로 이소펩타이드 결합을 자발적으로 형성할 수 있고, 서열 식별 번호: 9 및 10 또는 그들의 변이체 (예를 들면 서열 식별 번호: 109)를 포함하는 펩타이드는 서로 (예를 들면 서열 식별 번호: 109 및 10) 이소펩타이드 결합을 자발적으로 형성할 수 있다.
- [0149] 따라서, 본 발명은 하기를 포함하는 본 발명의 방법 및 용도에서 사용될 수 있는 펩타이드 링커의 쌍을 제공한다:
- [0150] (1) 상기 한정된 바와 같이 서열 식별 번호: 1 및 2 또는 그들의 변이체, 예를 들면 서열 식별 번호: 38 및 39를 포함하는 펩타이드 링커;
- [0151] (2) 상기 한정된 바와 같이 서열 식별 번호: 5 및 6 또는 그들의 변이체, 예를 들면 서열 식별 번호: 42 및 43을 포함하는 펩타이드 링커;
- [0152] (3) 상기 한정된 바와 같이 서열 식별 번호: 9 및 10 또는 그들의 변이체, 예를 들면 서열 식별 번호: 46 및 47을 포함하는 펩타이드 링커; 또는
- [0153] (4) 상기 한정된 바와 같이 서열 식별 번호: 109 및 10 또는 그들의 변이체를 포함하는 펩타이드 링커.
- [0154] 따라서, 상기 한정된 펩타이드 링커의 각각의 쌍은 동족 펩타이드 링커 쌍으로서 한정될 수 있다.
- [0155] 일부 구현예에서, 상기 한정된 각각의 펩타이드 링커 쌍 (즉 각각의 동족 펩타이드 링커 쌍)은 다른 펩타이드 링커 쌍과 직교 (즉 비-동족)로서 보여질 수 있고, 예를 들면 쌍 (1)은 쌍 (2), (3) 및/또는 쌍 (4)와 직교이고, 쌍 (2)는 쌍 (1), (3) 및/또는 쌍 (4)와 직교이고, 쌍 (3)은 쌍 (1) 및/또는 쌍 (2)와 직교이고 쌍 (4)는 쌍 (1) 및/또는 (2)와 직교이다. 일부 구현예에서, 이들 직교 쌍은 본 발명의 방법 및 용도에서 사용하기 위한 펩타이드 (동족) 링커의 바람직한 직교 (비-동족) 쌍을 나타낸다. 펩타이드 링커의 추가 바람직한 직교 쌍은 아래 한정된다.
- [0156] 상기 논의된 바와 같이, 본 발명의 펩타이드 링커는 융합 단백질의 합성에서 특정한 유용성을 찾고, 여기에서 펩타이드 링커는 융합 단백질을 형성하기 위해 또 다른 단백질 유닛에 연결되는 (접합되는) 단백질 유닛에 편입된다 (예를 들면 단백질 유닛의 도메인을 형성한다, 또는 단백질 유닛에 연결된다). 따라서, 추가 구현예에서, 본 발명은 상기 한정된 바와 같은 폴리펩타이드 및 펩타이드 링커를 포함하는 재조합 또는 합성 폴리펩타이드를

제공한다.

- [0157] 본 발명의 펩타이드 링커가 예를 들면 (본원에서 참고로 편집된) W02011/098772에서 기재된 바와 같이 펩타이드 태그로서 다른 방법 및 용도에서 유용성을 찾을 수 있다는 것이 분명할 것이다.
- [0158] 본 발명의 방법 및 용도에서 사용될 수 있는 다른 펩타이드 링커는 하기를 포함한다:
- [0159] (i) 서열 식별 번호: 13에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 13에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 7에서 아스파르테이트 또는 아스파라긴 잔기를 포함하는, 서열; 또는
- [0160] (ii) 서열 식별 번호: 14에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 14에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 56에서 글루타메이트 또는 아스파르테이트 잔기, 및 위치 10에서 리신 잔기를 포함하는, 서열; 또는
- [0161] (iii) 서열 식별 번호: 33에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 33에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 8에서 리신 잔기를 포함하는, 서열; 또는
- [0162] (iv) 서열 식별 번호: 17에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 17에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 11에서 아스파르테이트 또는 아스파라긴 잔기를 포함하는, 서열; 또는
- [0163] (v) 서열 식별 번호: 18에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 18에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열로서, 상기 아미노산 서열이 위치 241에서 글루타메이트 또는 아스파르테이트 잔기 및 위치 162에서 리신 잔기를 포함하는, 서열.
- [0164] 일부 구현예에서, 상기 펩타이드 링커 서열은 비교되는 서열 (서열 식별 번호: 13, 14, 17, 18, 또는 33)과 적어도 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100% 동일성이다.
- [0165] 본 발명의 방법 및 용도에서 사용될 수 있는 다른 펩타이드 링커 쌍은 하기를 포함한다:
- [0166] (5) 상기 한정된 바와 같이 서열 식별 번호: 13 및 14 또는 그들의 변이체를 포함하는 펩타이드 링커;
- [0167] (6) 상기 한정된 바와 같이 서열 식별 번호: 13 및 33 또는 그들의 변이체를 포함하는 펩타이드 링커; 또는
- [0168] (7) 상기 한정된 바와 같이 서열 식별 번호: 17 및 18 또는 그들의 변이체를 포함하는 펩타이드 링커.
- [0169] 일부 구현예에서, 동족 펩타이드 링커 쌍이 상기 (6)에서 한정된 쌍을 포함하는 경우, 반응은 또한 이소펩타이드 결합의 형성을 유도 또는 촉매화하는 성분을 포함한다. 예를 들어, 반응은 펩타이드 리가아제를 포함하고, 바람직하게는 여기에서 상기 펩타이드 리가아제는 서열 식별 번호: 34에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 34에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열을 포함한다.
- [0170] 일부 구현예에서, 상기 펩타이드 리가아제 서열은 비교되는 서열 (서열 식별 번호: 34)와 적어도 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100% 동일성이다.
- [0171] 상기 (1)-(7)로부터 선택된 펩타이드 링커 쌍의 임의의 직교 쌍이 본 발명의 방법 및 용도에서 사용될 수 있는 동안, 펩타이드 링커의 특히 바람직한 직교 쌍은, 상기 한정되는, 임의의 하나의 하기 쌍을 포함한다: (1) 및 (4), (1) 및 (5), (1) 및 (6), (1) 및 (3), (1) 및 (2), (2) 및 (4), (2) 및 (5), (2) 및 (6), (3) 및 (5), (3) 및 (6), (4) 및 (5) 그리고 (4) 및 (6).
- [0172] 융합 단백질을 형성하기 위해 또 다른 단백질에 연결되는 단백질 내에 펩타이드 링커의 위치는 특히 중요하지 않다. 따라서, 일부 구현예에서 펩타이드 링커는 융합 단백질에서 연결되는 단백질 또는 재조합 또는 합성 폴리펩타이드의 N-말단 또는 C-말단에서 위치할 수 있다. 일부 구현예에서, 펩타이드 링커는 융합 단백질에서 연결되는 단백질 또는 재조합 또는 합성 폴리펩타이드 내에 내부적으로 위치할 수 있다. 따라서, 일부 구현예에서 펩타이드 링커는 융합 단백질에서 연결되는 단백질 또는 재조합 또는 합성 폴리펩타이드의 N-말단, C-말단 또는 내부 도메인으로서 보여질 수 있다.
- [0173] 일부 구현예에서, 융합 단백질 안팎에 연결되는 단백질과 펩타이드 링커 사이, 하나 이상의 스페이서, 예를 들면 펩타이드 스페이서를 포함하는 것이 유용할 수 있다. 따라서, 단백질 및 펩타이드 링커는 서로 직접적으로

연결될 수 있거나 이들은 하나 이상의 스페이스 서열에 의해 간접적으로 연결될 수 있다. 따라서, 스페이스 서열은 융합 단백질에서 연결되는 단백질 또는 재조합 또는 합성 폴리펩타이드의 2개 이상의 개별적인 부분을 분리시키거나 사이에 공간을 둘 수 있다. 일부 구현예에서, 스페이스는 펩타이드 링커의 N-말단 또는 C-말단일 수 있다. 일부 구현예에서, 스페이스는 펩타이드 링커의 양측일 수 있다.

[0174] 스페이스 서열의 정확한 성질은 임계적이지 않고 가변성 길이 및/또는 서열일 수 있고, 예를 들어 1-40, 더욱 특히 2-20, 1-15, 1-12, 1-10, 1-8, 또는 1-6 잔기, 예를 들면 6, 7, 8, 9, 10 이상 잔기를 가질 수 있다. 대표적인 예의 방식으로 스페이스 서열은, 존재한다면, 1-15, 1-12, 1-10, 1-8 또는 1-6 잔기 등을 가질 수 있다. 잔기의 성질은 임계적이지 않고 이들은 예를 들어 임의의 아미노산, 예를 들면 중성 아미노산, 또는 지방족 아미노산일 수 있거나, 대안적으로 이들은 소수성, 또는 극성 또는 충전 또는 구조-형성 예를 들면 프롤린일 수 있다. 일부 바람직한 구현예에서, 링커는 세린 및/또는 글리신-풍부 서열이다.

[0175] 예시적인 스페이스 서열은 따라서 하나 이상의 상기 잔기로 구성된 임의의 단일 아미노산 잔기, 예를 들면 S, G, L, V, P, R, H, M, A 또는 E 또는 디-, 트리- 테트라- 펜타- 또는 헥사-펩타이드를 포함한다. 대표적인 및 바람직한 스페이스 서열은 서열 식별 번호: 36 또는 37에서 제시된 바와 같은 아미노산 서열을 포함한다.

[0176] 본 발명의 재조합 또는 합성 폴리펩타이드는 또한 (예를 들면 본 발명의 방법 및 용도에서 사용에 앞서 및/또는 이하에 논의된 바와 같이 융합 단백질의 확장 동안) 그의 정제를 용이하게 하기 위해 정제 모이어티 또는 태그를 포함할 수 있다. 임의의 적합한 정제 모이어티 또는 태그는 폴리펩타이드 내부에 편입될 수 있고 상기 모이어티는 당해 기술에 공지되어 있다. 예를 들어, 일부 구현예에서, 재조합 또는 합성 펩타이드는 펩타이드 정제 태그 또는 모이어티, 예를 들면 His-태그 서열을 포함할 수 있다. 그와 같은 정제 모이어티 또는 태그는 폴리펩타이드 내에 임의의 위치에서 편입될 수 있다. 일부 바람직한 구현예에서, 정제 모이어티는 폴리펩타이드의 N- 또는 C-말단에서 또는 상기를 향하여 (즉 상기의 5, 10, 15, 20 아미노산 내에) 위치한다.

[0177] 본 발명의 대표적인 재조합 또는 합성 폴리펩타이드는 서열 식별 번호: 50-59의 어느 하나에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 50-59의 어느 하나에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 서열을 가진 폴리펩타이드를 포함하고, 여기에서 상기 폴리펩타이드는 상기 한정된 바와 같은 펩타이드 링커를 포함한다.

[0178] 바람직하게는 재조합 또는 합성 폴리펩타이드는 상기 한정된 서열 동일성 요건을 충족시키고, 예를 들면 비교되는 서열과 적어도 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100% 동일성이다.

[0179] 상기 언급된 바와 같이, 본 발명의 이점은 융합 단백질을 형성하기 위해 함께 연결되는 단백질에서 편입된 펩타이드 링커 (예를 들면 본 발명의 재조합 또는 합성 폴리펩타이드)가 완전히 유전자 상으로 인코딩될 수 있다는 사실에서 발생한다. 따라서, 추가 측면에서, 본 발명은 상기 한정된 바와 같은 펩타이드 링커 또는 폴리펩타이드를 인코딩하는 핵산 분자를 제공한다.

[0180] 일부 구현예에서, 상기 한정된 펩타이드 링커를 인코딩하는 핵산 분자는 서열 식별 번호: 3, 4, 7, 8, 11, 12, 40, 41, 44, 45, 48, 49 또는 110의 어느 하나에서 제시된 바와 같은 뉴클레오타이드 서열 또는 서열 식별 번호: 3, 4, 7, 8, 11, 12, 40, 41, 44, 45, 48, 49 또는 110의 어느 하나에서 제시된 바와 같은 서열과 적어도 70% 서열 동일성을 가진 뉴클레오타이드 서열을 포함한다.

[0181] 일부 구현예에서, 상기 한정된 재조합 또는 합성 폴리펩타이드를 인코딩하는 핵산 분자는 서열 식별 번호: 60-69의 어느 하나에서 제시된 바와 같은 뉴클레오타이드 서열 또는 서열 식별 번호: 60-69의 어느 하나에서 제시된 바와 같은 서열과 적어도 70% 서열 동일성을 가진 뉴클레오타이드 서열을 포함한다.

[0182] 바람직하게는, 상기 핵산 분자는 비교되는 서열과 적어도 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100% 동일성이다.

[0183] 핵산 서열 동일성은, 예를 들면 디폴트 값 및 가변성 팜팩터, 및 6 뉴클레오타이드의 윈도우와 12.0에서 갭 창출 패널티 세트 및 4.0에서 갭 연장 패널티 세트를 수반한, GCG 패키지를 이용하는 FASTA 조사에 의해 결정될 수 있다. 바람직하게는 상기 비교는 서열의 전체 길이에 대해 실시되지만, 비교의 더 작은 윈도우, 예를 들면 600, 500, 400, 300, 200, 100 또는 50 미만 인접 뉴클레오타이드에 대해 실시될 수 있다.

[0184] 본 발명의 핵산 분자는 왓슨-크릭 유형 또는 유사한 염기 쌍 상호작용에 참여할 수 있는 리보뉴클레오타이드 및/또는 데옥시리보뉴클레오타이드 뿐만 아니라 합성 뉴클레오타이드 잔기로 구성될 수 있다. 바람직하게는, 핵산 분자는 DNA 또는 RNA이다.

- [0185] 상기 기재된 핵산 분자는 발현 대조군 서열, 또는 그와 같은 재조합 DNA 분자를 함유하는 재조합 DNA 클로닝 벡터 또는 벡터에 작동가능하게 연결될 수 있다. 이는 본 발명의 방법 및 용도에서 사용하기 위한 단백질의 세포내 발현, 예를 들면 본 발명의 폴리펩타이드의 발현을 허용하고, 유전자 생성물로서, 이의 발현은 관심 세포 내부에 도입된 유전자(들)에 의해 유도된다. 유전자 발현은 관심 세포에서 프로모터 활성에서 유도되고 계층에서 편입을 위하여 또는 독립적인 복제 또는 일시적 형질감염/발현을 위하여 선형 또는 원형 핵산 (예를 들면 DNA) 벡터의 임의의 형태로 삽입될 수 있다. 적합한 형질전환 또는 형질감염 기술은 문헌에서 양호하게 기재된다. 대안적으로, 네이키드 핵산 (예를 들면 DNA) 분자는 본 발명의, 및 본 발명에서 사용하기 위한 단백질 및 폴리펩타이드의 생산을 위하여 세포 내부에 직접적으로 도입될 수 있다. 대안적으로 핵산은 시험관내 전사에 의해 mRNA로 전환될 수 있고 관련된 단백질은 시험관내 번역에 의해 생성될 수 있다.
- [0186] 적절한 발현 벡터는 본 발명의 핵산 분자와 해독를 매칭에서 연결된 적절한 대조군 서열 예컨대 예를 들어 번역 (예를 들면 개시 및 정지 코돈, 리보솜 결합 부위) 및 전사 조절 요소 (예를 들면 프로모터-오퍼레이터 영역, 종료 정지 서열)을 포함한다. 적절한 벡터는 (박테리오파아지 및 진핵 바이러스 둘 모두를 포함하는) 플라스미드 및 바이러스를 포함할 수 있다. 적합한 바이러스성 벡터는 배큘로바이러스 및 또한 아테노바이러스, 아테노-관련된 바이러스, 헤르페스 및 백시니아/폭스 바이러스를 포함한다. 많은 다른 바이러스성 벡터는 당해 기술에서 기재된다. 바람직한 벡터는 박테리아 및 포유동물 발현 벡터 pGEX-KG, pEF-neo 및 pEF-HA를 포함한다.
- [0187] 상기 언급된 바와 같이, 본 발명의 폴리펩타이드는 추가의 서열 (예를 들면 폴리펩타이드의 정제를 용이하게 하기 위한 펩타이드/단백질 태그)를 포함할 수 있고 따라서 핵산 분자는, 발현시 융합 단백질을 생산하기 위해, 추가의 펩타이드 또는 폴리펩타이드, 예를 들면 His-태그, 말토오스-결합 단백질을 인코딩하는 DNA와 편리하게 융합될 수 있다.
- [0188] 따라서 추가 측면에서 보면, 본 발명은, 상기 한정된 바와 같은 핵산 분자를 포함하는, 벡터, 바람직하게는 발현 벡터를 제공한다.
- [0189] 본 발명의 다른 측면은, 벡터 핵산 내부에 본 발명의 펩타이드 링커 및/또는 폴리펩타이드를 인코딩하는 본 발명의 핵산 분자 삽입을 포함하는, 본 발명에 따른 재조합 핵산 분자의 제조 방법을 포함한다.
- [0190] 바람직하게는 벡터에 함유된, 본 발명의 핵산 분자는 임의의 적절한 수단에 의해 세포 내부에 도입될 수 있다. 적합한 형질전환 또는 형질감염 기술은 문헌에서 양호하게 기재된다. 다양한 기술은 공지되고 발현을 위하여 원핵 또는 진핵 세포 내부에 상기 벡터를 도입하는데 사용될 수 있다. 이러한 목적을 위해 바람직한 숙주 세포는 곤충 세포주, 효모, 포유동물 세포주 또는 *E. 콜리*, 예컨대 균주 BL21/DE3을 포함한다. 본 발명은 또한 상기 한정된 바와 같은 핵산 분자, 특히 벡터를 함유하는 형질전환된 또는 형질감염된 원핵 또는 진핵 숙주 세포까지 확장한다.
- [0191] 따라서, 또 다른 측면에서, 상기 기재된 바와 같은 핵산 분자 및/또는 벡터를 함유하는 재조합 숙주 세포가 제공된다.
- [0192] "재조합"은 핵산 분자 및/또는 벡터가 숙주 세포 내부에 도입되고 있는 것을 의미한다. 숙주 세포는 핵산 분자의 내인성 카피를 자연적으로 함유할 수 있거나 그렇지 않을 수 있지만, 핵산 분자 및/또는 벡터의 외인성 또는 추가 내인성 카피가 도입되고 있다는 점에서 재조합이다.
- [0193] 본 발명의 추가 측면은, 상기 한정된 바와 같은 본 발명의 펩타이드 링커 및/또는 폴리펩타이드의 제조 방법을 제공하며, 상기 방법은 상기 한정된 바와 같은 핵산 분자를 함유하는 숙주 세포를, 상기 펩타이드 링커 및/또는 폴리펩타이드를 인코딩하는 상기 핵산 분자가 발현되는 조건하에서 배양하고 상기 분자 (펩타이드 링커 및/또는 폴리펩타이드)를 회수함으로써 생산되는 것을 포함한다. 상기 발현된 펩타이드 링커 및/또는 폴리펩타이드는 본 발명의 추가 측면을 형성한다.
- [0194] 일부 구현예에서, 본 발명의, 또는 본 발명의 방법 및 용도에서 사용하기 위한 펩타이드 링커 및/또는 폴리펩타이드는, 예를 들면 아미노산 또는 더 작은 합성으로 생성된 펩타이드의 결합에 의해, 또는 더욱 편리하게 위에서 기재된 바와 같이 상기 폴리펩타이드를 인코딩하는 핵산 분자의 재조합 발현에 의해 합성으로 생성될 수 있다.
- [0195] 본 발명의 핵산 분자는 당해 기술에서 공지된 임의의 적당한 수단에 의해 합성으로 생성될 수 있다.
- [0196] 따라서, 본 발명의 펩타이드 링커 및/또는 폴리펩타이드는 단리된, 정제된, 재조합 또는 합성된 펩타이드 링커 또는 폴리펩타이드일 수 있다. 상기 언급된 바와 같이, 용어 "폴리펩타이드"는 용어 "단백질"과 상호교환적으로

본원에서 사용된다. 상기 언급된 바와 같이, 용어 폴리펩타이드 또는 단백질은 전형적으로 적어도 40 개 연속 아미노산 잔기, 예를 들면 적어도 50, 60, 70, 80, 90, 100, 150 개의 아미노산, 예컨대 40-1000, 50-900, 60-800, 70-700, 80-600, 90-500, 100-400 개의 아미노산을 포함하는 임의의 아미노산 서열을 포함한다.

[0197] 유사하게, 본 발명의 핵산 분자는 단리된, 정제된, 재조합 또는 합성된 핵산 분자일 수 있다.

[0198] 따라서, 대안적으로 보면, 본 발명의 펩타이드 링커, 폴리펩타이드 및 핵산 분자는 바람직하게는 비-원상태, 즉 비-천연 발생, 분자이다.

[0199] 표준 아미노산 명명법은 본원에서 사용된다. 따라서, 아미노산 잔기의 완전한 명칭은 1개의 문자 암호 또는 3개의 글자 약어와 상호교환적으로 사용될 수 있다. 예를 들어, 리신은 K 또는 Lys로 치환될 수 있고, 이소류신은 I 또는 Ile로 치환될 수 있고, 등등이다. 또한, 용어 아스파르테이트 및 아스파르트산, 및 글루타메이트 및 글루탐산은 본원에서 상호교환적으로 사용되고 asp 또는 D, 또는 glu 또는 E, 각각으로 대체될 수 있다.

[0200] 본 발명의, 및 본 발명에서 사용하기 위한 펩타이드 링커 및 폴리펩타이드가 재조합으로 생산될 수 있다는 것이 구상중이고, 이는 본 발명의 바람직한 구현예인 동안, 본 발명의 펩타이드 링커가 다른 수단에 의해 융합 단백질에서 연결되는 단백질에 접합될 수 있다는 것이 분명할 것이다. 다시 말해서, 펩타이드 링커 및 단백질은 임의의 적당한 수단에 의해, 예를 들면 재조합으로 개별적으로 생산될 수 있고, 그 뒤에 접합(연결)되어 본 발명의 방법에서 사용될 수 있는 펩타이드 링커-단백질 콘주게이트를 형성할 수 있다. 예를 들어, 본 발명의 펩타이드 링커는, 상기 기재된 바와 같이, 합성으로 또는 재조합으로 생산될 수 있고, 비-펩타이드 링커 또는 스페이서, 예를 들면 화학 링커 또는 스페이서를 통해 (본 발명의 방법에 따라 융합 단백질에서 연결되는) 단백질에 접합될 수 있다.

[0201] 따라서, 일부 구현예에서, 융합 내부에 편입되는 펩타이드 링커 및 단백질은 직접적으로 결합을 통해 또는 간접적으로 연결 기를 통해 함께 연결될 수 있다. 연결 기가 이용되는 경우, 상기 기는 연결 기를 통해 펩타이드 링커 및 단백질 성분의 공유결합에 대하여 제공하기 위해 선택될 수 있다. 관심 연결 기는 단백질 성분의 성질에 의존하여 널리 다양할 수 있다. 연결 기는, 존재하는 경우, 많은 구현예에서 생물학적으로 불활성이다.

[0202] 다양한 연결 기는 당해 분야의 숙련가에 공지되고 본 발명에서 용도를 찾는다. 대표적인 구현예에서, 연결 기는 일반적으로 적어도 약 50 달톤, 일반적으로 적어도 약 100 달톤이고, 연결 기가 스페이서를 함유하면 1000 달톤 이상, 예를 들어 최대 1000000 달톤만큼 클 수 있지만, 일반적으로 약 500 달톤을 초과하지 않을 것이고 일반적으로 약 300 달톤을 초과하지 않을 것이다. 일반적으로, 상기 링커는 펩타이드 링커 및 단백질 성분에 공유결합할 수 있는 반응성 작용기와 어느 하나의 말단에서 종료된 스페이서 기를 포함할 것이다. 관심 스페이서 기는 지방족 및 불포화된 탄화수소 사슬, 헤테로원자 예컨대 산소 (에테르 예컨대 폴리에틸렌 글리콜) 또는 질소 (폴리아민)를 함유하는 스페이서, 펩타이드, 탄수화물, 가능하게는 헤테로원자를 함유할 수 있는 환형 또는 비환형 시스템을 포함할 수 있다. 스페이서 기는 또한 금속 이온의 존재가 복합체를 형성하기 위해 2개 이상의 리간드를 배위하도록 금속에 결합하는 리간드로 구성될 수 있다. 특이적 스페이서 요소는 하기를 포함한다: 1,4-디아미노헥산, 크실릴렌디아민, 테레프탈산, 3,6-디옥사옥탄이산, 에틸렌디아민-N,N-디아세트산, 1,1'-에틸렌비스(5-옥소-3-피롤리딘카복실산), 4,4'-에틸렌디피페리딘. 잠재적인 반응성 작용기는 하기를 포함한다: 친핵성 작용기 (아민, 알코올, 티올, 하이드라자이드), 친전자성 작용기 (알데하이드, 에스테르, 비닐 케톤, 에폭사이드, 이소시아네이트, 말레이미드), 고리화부가 반응, 디설파이드 결합 형성, 또는 금속에 결합할 수 있는 작용기. 특이적 예는 하기를 포함한다: 1차 및 2차 아민, 하이드록삼산, N-하이드록시석신이미드 에스테르, N-하이드록시석신이미드 카보네이트, 옥시카보닐이미다졸, 니트로페닐에스테르, 트리플루오로에틸 에스테르, 글리시딜 에테르, 비닐설펜, 및 말레이미드. 상기 대상체 블록킹 시약에서 용도를 찾을 수 있는 특이적 링커 기는 하기를 포함한다: 헤테로작용성 화합물, 예컨대 아지도벤조일 하이드라자이드, N-[4-(p-아지도살리실아미노)부틸]-3'-[2'-피리딘디티오]프로피온아미드), 비스-설포석신이미드 수베레이트, 디메틸아디프이미데이트, 디석신이미드타르트레이트, N-말레이미도부틸옥시석신이미드 에스테르, N-하이드록시 설포석신이미드-4-아지도벤조에이트, N-석신이미드 [4-아지도페닐]-1,3'-디티오프로피오네이트, N-석신이미드 [4-아이오도아세틸]아미노벤조에이트, 글루타르알데하이드, 및 석신이미드-4-[N-말레이미도메틸]사이클로헥산-1-카복실레이트, 3-(2-피리딘디티오)프로피온산 N-하이드록시석신이미드 에스테르 (SPDP), 4-(N-말레이미도메틸)-사이클로헥산-1-카복실산 N-하이드록시석신이미드 에스테르 (SMCC), 등등.

[0203] 일부 구현예에서, 이들 분자의 콘주게이션을 용이하게 하기 위해 및/또는 펩타이드 링커 및/또는 단백질의 안정성을 개선하기 위해 펩타이드 링커 및/또는 단백질에서 하나 이상의 잔기를 변형시키는 것이 유용할 수 있다. 따라서, 일부 구현예에서, 본 발명의, 또는 본 발명에서 사용하기 위한 펩타이드 링커, 폴리펩타이드 또는 단백질

질은 비천연 또는 비-표준 아미노산을 포함할 수 있다.

[0204] 일부 구현예에서, 본 발명의, 또는 본 발명에서 사용하기 위한 펩타이드 링커, 폴리펩타이드 또는 단백질은 하나 이상의, 예를 들면 적어도 1, 2, 3, 4, 5 개의 비-종래의 아미노산, 예컨대 10, 15, 20 개 이상 비-종래의 아미노산, 즉 표준 유전자 암호에 의해 코딩되지 않는 측사슬을 보유하는 아미노산, 본원에서 일명 "비-코딩된 아미노산"을 포함할 수 있다 (참고 예를 들면 표 1). 이들은 대사 과정을 통해 형성되는 아미노산 예컨대 오르니틴 또는 타우린, 및/또는 인공적으로 변형된 아미노산 예컨대 9H-플루오렌-9-일메톡시카보닐 (Fmoc), (tert)-(부)틸 (옥)시 (카)보닐 (Boc), 2,2,5,7,8-펜타메틸크로만-6-설포닐 (Pmc) 보호된 아미노산, 또는 벤질옥시-카보닐 (Z)기를 갖는 아미노산으로부터 선택될 수 있다.

[0205] 본 발명의, 또는 본 발명에서 사용하기 위한 펩타이드 링커 또는 폴리펩타이드에서 사용될 수 있는 비-표준 또는 구조적 유사체 아미노산의 예는 하기이다: D 아미노산, 아미드 등배체 (예컨대 N-메틸 아미드, 레트로-인버스 아미드, 티오아미드, 티오에스테르, 포스포네이트, 케토메틸렌, 하이드록시메틸렌, 플루오로비닐, (E)-비닐, 메틸렌아미노, 메틸렌티오 또는 알칸), L-N 메틸아미노산, D-a 메틸아미노산, D-N-메틸아미노산. 비-종래의, 즉 비-코딩된, 아미노산의 예는 표 1에서 열거된다.

표 1

비-종래의 아미노산	코드	비-종래의 아미노산	코드
α-아미노부티르산	Abu	L-N-메틸알라닌	Nmala
α-아미노-α-메틸부티레이트	Mgab	L-N-메틸아르기닌	Nmarg
아미노시클로프로판-카복실레이트	Cpro	L-N-메틸아스파라긴	Nmasn
아미노이소부티르산	Aib	L-N-메틸아스파르트산	Nmasp
아미노노르보르닐-카복실레이트	Norb	L-N-메틸시스테인	Nmcys
시클로헥실알라닌	Chexa	L-N-메틸글루타민	Nmgln
시클로펜틸알라닌	Cpen	L-N-메틸글루탐산	Nmglu
D-알라닌	Dal	L-N-메틸히스티딘	Nmhis
D-아르기닌	Darg	L-N-메틸이소류신	Nmile
D-아스파르트산	Das	L-N-메틸류신	Nmleu
D-시스테인	Dcys	L-N-메틸리신	Nmlys
D-글루타민	Dgln	L-N-메틸메티오닌	Nmmet
D-글루탐산	Dglu	L-N-메틸노르류신	Nmnle
D-히스티딘	Dhis	L-N-메틸노르발린	Nmnva
D-이소류신	Dile	L-N-메틸오르니틴	Nmorn
D-류신	Dleu	L-N-메틸페닐알라닌	Nmphe
D-리신	Dlys	L-N-메틸프롤린	Nmpro
D-메티오닌	Dmet	L-N-메틸세린	Nmser
D-오르니틴	Dorn	L-N-메틸트레오닌	Nmthr
D-페닐알라닌	Dphe	L-N-메틸트립토판	Nmtrp
D-프롤린	Dpro	L-N-메틸티로신	Nmtyr
D-세린	Dser	L-N-메틸발린	Nmval
D-트레오닌	Dthr	L-N-메틸에틸글리신	Nmetg
D-트립토판	Dtrp	L-N-메틸-t-부틸글리신	Nmtbug
D-티로신	Dtyr	L-노르류신	Nle
D-발린	Dval	L-노르발린	Nva
D-α-메틸알라닌	Dmala	α-메틸-아미노이소부티레이트	Maib
D-α-메틸아르기닌	Dmarg	α-메틸-γ-아미노부티레이트	Mgab
D-α-메틸아스파라긴	Dmasn	α-메틸시클로헥실알라닌	Mchexa
D-α-메틸아스파르테이트	Dmasp	α-메틸시클로펜틸알라닌	Mcpen
		α-메틸-α-나프틸알라닌	Manap
		α-메틸페닐알라닌	Mpen

[0206]

D- α -메틸히스티딘	Dmhis	N-(3-아미노프로필)글리신	Norn
D- α -메틸이소류신	Dmile	N-아미노- α -메틸부티레이트	Nmaabu
D- α -메틸류신	Dmleu	α -나프틸알라닌	Anap
D- α -메틸리신	Dmlys	N-벤질글리신	Nphe
D- α -메틸메티오닌	Dmmet	N-(2-카바밀에틸)글리신	Ngln
D- α -메틸오르니틴	Dmorn	N-(카바밀메틸)글리신	Nasn
D- α -메틸페닐알라닌	Dmphe	N-(2-카복시에틸)글리신	Nglu
D- α -메틸프롤린	Dmpro	N-(카복시메틸)글리신	Nasp
D- α -메틸세린	Dmser	N-시클로부틸글리신	Ncbut
D- α -메틸트레오닌	Dmthr	N-시클로헥실글리신	Nchep
D- α -메틸트립토판	Dmtrp	N-시클로헥실글리신	Nchex
D- α -메틸티로신	Dmtyr	N-시클로데실글리신	Ncdec
D- α -메틸발린	Dmval	N-실코도데실글리신	Ncdod
D-N-메틸알라닌	Dnmala	N-시클로옥틸글리신	Ncoct
D-N-메틸아르기닌	Dnmarg	N-시클로프로필글리신	Ncpro
D-N-메틸아스파라긴	Dnmasn	N-시클로운데실글리신	Ncund
D-N-메틸아스파르테이트	Dnmasp	N-(2,2-디페닐에틸)글리신	Nbhm
D-N-메틸시스테인	Dnmcys	N-(3,3-디페닐프로필)글리신	Nbhe
D-N-메틸글루타민	Dnmglu	N-(3-구아니디노프로필)글리신	Narg
D-N-메틸글루탐에이트	Dnmglu	N-(1-하이드록시에틸)글리신	Nthr
D-N-메틸히스티딘	Dnmhis	N-(하이드록시에틸)글리신	Nser
D-N-메틸이소류신	Dnmile	N-(이미다졸릴에틸)글리신	Nhis
D-N-메틸류신	Dnmleu	N-(3-인돌릴리에틸)글리신	Nhtrp
D-N-메틸리신	Dnmlys	N-메틸- γ -아미노부티레이트	Nmgabu
N-메틸시클로헥실알라닌	Nmchexa	D-N-메틸메티오닌	Dnmmet
D-N-메틸오르니틴	Dnmorn	N-메틸시클로펜틸알라닌	Nmcpen
N-메틸글리신	Nala	D-N-메틸페닐알라닌	Dnmphe
N-메틸아미노이소부티레이트	Nmaib	D-N-메틸프롤린	Dnmpro
N-(1-메틸프로필)글리신	Nile	D-N-메틸세린	Dnmser
N-(2-메틸프로필)글리신	Nleu	D-N-메틸트레오닌	Dnmthr
D-N-메틸트립토판	Dnmtrp	N-(1-메틸에틸)글리신	Nval
D-N-메틸티로신	Dnmtyr	N-메틸라-나프틸알라닌	Nmanap
D-N-메틸발린	Dnmval	N-메틸페니실라민	Nmpen
γ -아미노부티르산	Gabu	N-(p-하이드록시페닐)글리신	Nhtyr
L-t-부틸글리신	Tbug	N-(티오메틸)글리신	Ncys
L-에틸글리신	Etg	페니실라민	Pen
L-homo 페닐알라닌	Hphe	L- α -메틸알라닌	Mala
L- α -메틸아르기닌	Marg	L- α -메틸아스파라긴	Masn
L- α -메틸아스파르테이트	Masp	L- α -메틸-t-부틸글리신	Mtbug

[0207]

L- α -메틸시스테인	Mcys	L-메틸에틸글리신	Metg
L- α -메틸글루타민	Mglu	L- α -메틸글루타메이트	Mglu
L- α -메틸히스티딘	Mhis	L- α -메틸호모페닐알라닌	Mhphe
L- α -메틸이소류신	Mile	N-(2-메틸티오에틸)글리신	Nmet
L- α -메틸류신	Mleu	L- α -메틸리신	Mlys
L- α -메틸메티오닌	Mmet	L- α -메틸노르류신	Mnle
L- α -메틸노르발린	Mnva	L- α -메틸오르니틴	Morn
L- α -메틸페닐알라닌	Mphe	L- α -메틸프롤린	Mpro
L- α -메틸세린	Mser	L- α -메틸트레오닌	Mthr
L- α -메틸트립토판	Mtrp	L- α -메틸티로신	Mtyr
L- α -메틸발린	Mval	L-N-메틸호모페닐알라닌	Nmhphe
N-(N-(2,2-디페닐에틸))	Nnbhm	N-(N-(3,3-디페닐프로필))	Nnbhe
카바밀메틸)글리신		카바밀메틸)글리신	
1-카복시-1-(2,2-디페닐- Nmbc		L-O-메틸 세린	Omser
에틸아미노)시클로프로판		L-O-메틸 호모세린	Omhse

[0208]

[0209]

일부 구현예에서 본 발명의 방법은, 고품 상을 이용하여, (상기 기재된 바와 같이) 불균일하게 수행될 수 있고, 예를 들어, 여기에서 성장하는 융합 단백질, 바람직하게는 융합 단백질 사슬에서 제1 또는 제2 단백질은 고품 상에서 고정될 수 있고, 세정 단계의 이용을 허용한다. 따라서, 일부 구현예에서, 방법은 고품 상 방법 (즉 불균일 방법)이다. 대안적으로 보면, 방법은 고품 상 또는 고품 기질에서 수행된다. 고품 상 검정의 용도는 이점을 제공한다. 예를 들어, 세정 단계는 차후의 라운드의 반응 (즉 융합 단백질에 추가 단백질의 첨가)를 방해할 수 있는 과잉의, 미반응된 단백질 및/또는 성분, 예를 들면 펩타이드 리가아제, 펩타이드 링커 등의 차단해제 (케이징해제, 마스킹해제, 탈보호)에 관여된 성분의 제거에 일조할 수 있다.

[0210]

고품 상에서 융합 단백질의 고정화는 다양한 방식으로 달성될 수 있다. 융합 단백질은 임의의 편리한 방식으로 지지체에 고정, 즉 결합될 수 있다. 일부 구현예에서, 융합 단백질의 제1 또는 제2 단백질은 고품 지지체에서 고정된다. 따라서, 일부 구현예에서, 방법은 고품 지지체에서 제1 단백질의 고정화의 단계를 포함할 수 있다. 일부 구현예에서, 방법은 고품 지지체에서 제1 및 제2 단백질을 포함하는 연결 단백질의 고정화의 단계를 포함할 수 있다.

[0211]

따라서 고정화의 방식 또는 수단 및 고품 지지체는, 당해 기술에서 널리 공지된 및 문헌에서 기재된 바와 같이 임의의 수의 고정화 수단 및 고품 지지체로부터, 결정에 따라, 선택될 수 있다. 따라서, 융합 단백질은 (예를 들면 화학적으로 가교결합된) 융합 단백질에서 예를 들어 적어도 하나의 단백질의 도메인 또는 모이어티를 통해 지지체에 직접적으로 결합될 수 있다. 일부 구현예에서, 융합 단백질은 링커 기에 의해, 또는 중개자 결합 기 (들)에 의해 (예를 들면 바이오틴-스트렙타비딘 상호작용에 의해) 간접적으로 결합될 수 있다. 따라서, 융합 단백질은 고품 지지체에 공유 또는 비-공유결합될 수 있다. 연결은 가역적 (예를 들면 절단가능) 또는 비가역적 연결일 수 있다. 따라서, 일부 구현예에서, 연결은 효소적으로, 화학적으로 또는 광으로 절단될 수 있고, 예를 들면 연결은 광-감수성 연결일 수 있다.

[0212]

따라서, 일부 구현예에서, 융합 단백질에서 포함되는 단백질은 지지체에서 제공된 고정화용 수단 (예를 들면, 그의 결합 파트너, 즉 동족 결합 파트너, 예를 들면 스트렙타비딘 또는 항체에 결합할 수 있는, 친화도 결합 파트너, 예를 들면 바이오틴 또는 합텐)으로 제공될 수 있다. 일부 구현예에서, 지지체에서 고정되는 단백질은 결합 단백질, 예를 들면 말토오스 결합 단백질, 항체 등일 수 있다. 융합 단백질과 고품 지지체 사이 상호작용은 세정 단계를 허용할만큼 충분히 강력해야 한다, 즉 융합 단백질과 고품 지지체 사이 상호작용은 세정 단계에 의해 파괴 (상당히 파괴)되지 않는다. 예를 들어, 바람직하게는, 각각의 세정 단계로, 융합 단백질의 5% 미만, 바람직하게는 4, 3, 2, 1, 0.5 또는 0.1% 미만은 고품 상으로부터 제거 또는 용출된다. 이 점에 있어서, 본 발명자는 말토오스에 대하여 개선된 결합 친화도를 갖는 변형된 말토오스 결합 단백질을 개발하였고 따라서 본 발명의 방법에서 특정한 유용성을 찾는다.

[0213]

따라서, 본 발명의 추가 측면은 서열 식별 번호: 70에서 제시된 바와 같은 아미노산 서열 또는 서열 식별 번호: 70에서 제시된 바와 같은 아미노산 서열과 적어도 70% 동일성을 가진 서열을 포함하는 말토오스 결합 단백질을 제공한다.

[0214]

일부 구현예에서, 상기 말토오스 결합 단백질은 비교되는 서열과 적어도 75, 80, 85, 90, 95, 96, 97, 98, 99

또는 100% 동일성이다.

- [0215] 바람직하게는, 서열 식별 번호: 70에서 제시된 바와 같은 아미노산 서열과 적어도 70% 동일성을 가진 말토오스 결합 단백질은 서열 식별 번호: 70에서 제시된 바와 같은 아미노산 서열로 구성되는 단백질과 기능적으로 동등하다, 즉 서열 식별 번호: 70에서 제시된 바와 같은 아미노산 서열로 구성되는 단백질과 동일한 친화도, 또는 상기보다 더 큰 친화도로 말토오스를 결합할 수 있다. 예를 들어, 본 발명의 말토오스 결합 단백질은 0.2 μ M 미만, 예를 들면 0.1, 0.08, 0.05, 0.03, 또는 0.01 μ M 이하의 말토오스에 대한 결합 친화도를 갖는다. 바람직한 구현예에서, 서열 식별 번호: 70에서 제시된 바와 같은 아미노산 서열과 적어도 70% 동일성을 가진 말토오스 결합 단백질은 위치 312 및 317에서 발린을 포함한다.
- [0216] 본 발명은 또한 상기 한정된 말토오스 결합 단백질을 인코딩하는 핵산 분자를 제공한다. 일부 구현예에서, 핵산 분자는 서열 식별 번호: 71에서 제시된 바와 같은 뉴클레오타이드 서열 또는 서열 식별 번호: 71에서 제시된 바와 같은 뉴클레오타이드 서열과 적어도 70% 서열 동일성을 가진 서열을 포함한다.
- [0217] 일부 구현예에서, 말토오스 결합 단백질은 본원에서 한정된 바와 같은 펩타이드 링커를 포함한다 (예를 들면 상기에 접합된다). 더욱 추가 구현예에서, 말토오스 결합 단백질은 상기 한정된 바와 같은 1 개 초과 (예를 들면 2 또는 3 개) 아미노산 서열을 포함한다, 즉 반복된 서열을 포함한다.
- [0218] 융합 단백질, 예를 들면 융합 단백질에서 편입되는 제1 단백질은 융합 단백질 내부에 편입되는 추가 단백질 (예를 들면 제2 단백질)과 접촉되기 이전 또는 이후 고정될 수 있다. 추가로, 그와 같은 "고정화가능한" 융합 단백질은 지지체와 함께 추가 단백질과 접촉될 수 있다.
- [0219] 고정 지지체는 고정화, 분리 등에 현재 널리 사용 또는 제안되는 임의의 공지된 지지체 또는 매트릭스일 수 있다. 이들은 입자 (예를 들면 자기, 파라-자기 또는 비-자기일 수 있는 비드), 시트, 겔, 필터, 막, 섬유, 모세관, 슬라이드, 어레이 또는 미세적정 스트립, 튜브, 플레이트 또는 웰 등의 형태를 취할 수 있다.
- [0220] 지지체는 유리, 실리카, 라텍스 또는 폴리머 재료로 만들어질 수 있다. 융합 단백질의 결합을 위하여 높은 표면적을 나타내는 물질이 적합하다. 상기 지지체는 불규칙한 표면을 가질 수 있고 예를 들어 다공성 또는 미립자 예를 들면 입자, 섬유, 웹, 소결물 또는 체일 수 있다. 미립자 물질, 예를 들면 비드, 특히 폴리머 비드는 그의 더 큰 결합능 때문에 유용하다.
- [0221] 편리하게, 본 발명에 따라 사용된 미립자 고정 지지체는 구형 비드를 포함할 것이다. 비드의 크기는 임계적이지 않지만, 이들은 예를 들어 적어도 1 및 바람직하게는 적어도 2 μ m의 직경의 정도일 수 있고, 바람직하게는 10 이하, 및 예를 들면 6 μ m 이하의 최대 직경을 가질 수 있다.
- [0222] 크기 (예를 들면 5% 미만의 직경 표준 편차를 갖는 크기)가 실질적으로 균일한 것인, 단분산 입자는 이들이 반응의 매우 균일한 재현성을 제공한다는 이점을 갖는다. 대표적인 단분산 폴리머 입자는 US-A-4336173에서 기재된 기술에 의해 생산될 수 있다.
- [0223] 그러나, 조작 및 분리를 돕기 위해, 자기 비드가 유리하다. 용어 "자기"는 본원에서 사용된 바와 같이 지지체가 자기장에서 배치된 경우 여기에 부여된 자기 모멘트를 가질 수 있고, 따라서 그 장의 작용하에 대체가능하다는 것을 의미한다. 다시 말해서, 자기 입자를 포함하는 지지체는, 이소젯타이드 결합 형성 단계 이후 입자 분리의 신속한, 단순한 및 효율적인 방식을 제공하는, 자기 응집에 의해 쉽게 제거될 수 있다.
- [0224] 일부 구현예에서, 고정 지지체는 아밀로오스 수지이다.
- [0225] 융합 단백질에서 끝에서 두번째와 최종 단백질 사이 이소젯타이드 결합의 형성시, 고정 지지체로부터 단백질을 제거 또는 용출하는 것이 바람직할 수 있다. 따라서, 일부 구현예에서, 방법은 고정 지지체로부터 융합 단백질의 용출 또는 제거의 단계를 포함한다.
- [0226] 상기에서 나타난 바와 같이, 특정 프로토콜에서 본 발명의 방법은 동일한 고정 지지체, 예를 들면 어레이에서 2 개 이상의 융합 단백질의 동시 생산을 허용할 수 있다. 따라서, 일부 구현예에서, 본 발명의 방법은 멀티플렉스 및/또는 고 처리 포맷으로서 보여질 수 있다.
- [0227] 추가 구현예에서, 본 발명은 본 발명의 방법으로부터 수득된 또는 수득가능한 융합 단백질을 제공한다. 일부 구현예에서, 융합 단백질은 고정 기질에서 고정된다. 따라서, 더욱 추가 구현예에서, 본 발명은 본 발명의 방법으로부터 수득된 또는 수득가능한, 적어도 하나의 융합 단백질을 포함하는 고정 기질을 제공한다. 일부 구현예에서, 고정 기질은 본 발명의 방법으로부터 수득된 또는 수득가능한 2개 이상의 융합 단백질 (상기한 서열을 가진

융합 단백질)을 포함하는 어레이 (즉 단백질 어레이, 특히 융합 단백질 어레이)의 형태일 수 있다. 일부 구현예에서, 어레이는 (상이한 구조 또는 서열을 가진) 적어도 10, 20, 50, 100, 200, 300, 400, 500, 1000, 1500, 2000, 5000 또는 10000 융합 단백질 즉 상이한 융합 단백질을 포함한다.

[0228] 일부 구현예에서, 본 발명의 방법으로부터 수득된 또는 수득가능한 2개 이상의 융합 단백질은 융합 단백질의 라이브러리를 형성하기 위해 함께 혼합될 수 있다. 따라서, 추가 구현예에서, 본 발명은 본 발명의 방법으로부터 수득된 또는 수득가능한, 적어도 2개의 융합 단백질을 포함하는 융합 단백질의 라이브러리를 제공한다. 일부 구현예에서, 라이브러리는 (상이한 구조 또는 서열을 가진) 적어도 10, 20, 50, 100, 200, 300, 400, 500, 1000, 1500, 2000, 5000 또는 10000 융합 단백질, 즉 상이한 융합 단백질을 포함한다. 일부 구현예에서, 라이브러리는 고정 기질, 예를 들면 비드 또는 입자에서 고정된 융합 단백질을 포함할 수 있다. 예를 들어, 각각의 고정 기질, 예를 들면 비드 또는 입자는 상이한 융합 단백질을 포함할 수 있다.

[0229] 본 발명의 방법이 불균일 구현예를 이용하여 예시되고 있는 동안, 방법이 (즉 용액내) 균일하게 이용될 수 있다는 것이 본원에서 개시내용으로부터 쉽게 분명할 것이다. 그러나, 융합 단백질의 혼합물의 생산을 방지하기 위해, 일부 구현예에서 각각의 라운드의 확장 이후 반응에서 다른 성분으로부터 융합 단백질을 분리시키는 것이 필요할 수 있다. 분리 또는 정제는 임의의 적당한 수단에 의해 달성될 수 있다. 예를 들어, 융합 단백질 사슬에서 단백질 중 하나는 정제 태그를 포함할 수 있거나 반응, 예를 들면 친화성 크로마토그래피에서 다른 성분으로부터 융합 단백질의 분리를 용이하게 할 결합 단백질 (예를 들면 말토오스 결합 단백질)일 수 있다. 추가적으로 또는 대안적으로, 다른 정제/분리 방법, 예를 들면 이온교환 크로마토그래피, 크기-제외 크로마토그래피, 초원심분리, 스핀-여과, 투석, 투석-여과 등은 이용될 수 있다.

[0230] 따라서, 일부 구현예에서, 본 발명의 방법은 이소펩타이드 결합 형성의 단계 이후 융합 단백질의 분리 또는 정제의 단계를 포함할 수 있다.

[0231] 추가 구현예에서, 본 발명은 키트, 특히 본 발명의 방법 및 용도에서, 즉 융합 단백질의 생산 또는 합성에서 사용하기 위한 키트를 제공하고, 여기에서 상기 키트는 하기를 포함한다:

[0232] (a) 상기 한정된 바와 같은 펩타이드 링커를 포함하는 재조합 또는 합성 폴리펩타이드; 및

[0233] (b) (a)의 폴리펩타이드에서 펩타이드 링커와 이소펩타이드 결합을 형성할 수 있는 한정된 바와 같은 펩타이드 링커를 포함하는 재조합 또는 합성 폴리펩타이드; 및/또는

[0234] (c) 상기 한정된 바와 같은 펩타이드 링커를 인코딩하는, 핵산 분자, 특히 벡터; 및/또는

[0235] (d) (b)의 핵산 분자에 의해 인코딩된 펩타이드 링커와 이소펩타이드 결합을 형성할 수 있는 펩타이드 링커를 인코딩하는, 핵산 분자, 특히 벡터,

[0236] 선택적으로 여기에서 (a) 및/또는 (b)의 재조합 또는 합성 폴리펩타이드는 (a) 및 (b)의 폴리펩타이드에서 펩타이드 링커와 직교인 펩타이드 링커의 쌍의 일부인 추가 펩타이드 링커를 포함함.

[0237] 본 발명의 방법 및 용도는 *시험관내* 방법 및 용도, 즉 융합 단백질의 합성용 *시험관내* 방법으로서 한정될 수 있다.

[0238] 본 발명의 방법이 융합 단백질을 형성하기 위해 함께 임의의 특이적 단백질 연결에 제한되지 않는다는 것이 분명할 것이다. 따라서, 방법은 본원에서 한정된 바와 같은 임의의 단백질 또는 폴리펩타이드, 즉 임의의 요망된 단백질 또는 폴리펩타이드를 이용할 수 있다. 다시 말해서, 본 발명은 융합 단백질 내부에 포함 또는 편입되도록 요망되는 임의의 단백질 또는 폴리펩타이드를 이용할 수 있다. 더욱이, 본 발명의 재조합 또는 합성 폴리펩타이드는 본 발명의 펩타이드 링커에 연결된 임의의 단백질을 포함할 수 있다. 단백질은 임의의 적합한 공급원으로부터 유래 또는 수득될 수 있다. 예를 들어, 단백질은 생물학적 및 임상 샘플, 예를 들면 유기체 (진핵, 원핵)의 임의의 세포 또는 조직 샘플, 또는 임의의 체액 또는 그로부터 유래된 조제물, 뿐만 아니라 샘플 예컨대 세포 배양물, 세포 제제, 세포 용해물 등으로부터 *시험관내* 번역 또는 정제될 수 있다. 단백질은 환경 샘플로부터 유래 또는 수득, 예를 들면 정제될 수 있고, 예를 들면 토양 및 물 샘플 또는 음식 샘플이 또한 포함된다. 샘플은 새롭게 제조될 수 있거나 이들은 예를 들면 저장을 위하여 임의의 편리한 방식으로 사전-처리될 수 있다.

[0239] 상기 언급된 바와 같이, 바람직한 구현예에서, 융합 단백질에서 편입되는 단백질은 재조합으로 생산될 수 있고 따라서 상기 단백질을 인코딩하는 핵산 분자는, 모든 원핵 또는 진핵 세포, 바이러스, 박테리오파아지, 마이코플라스마, 원형질 및 세포소기관을 포함하는, 임의의 적합한 공급원, 예를 들면 임의의 바이러스성 또는 세포

물질로부터 유래 또는 수득될 수 있다. 상기 생물학적 물질은 따라서 포유동물 및 비-포유동물 동물 세포, 식물 세포, 청조류-녹조류를 포함하는 조류, 진균, 박테리아, 원생동물 등의 모든 유형을 포함할 수 있다. 일부 구현예에서, 융합 단백질에서 함께 연결되는 단백질은 합성 단백질일 수 있다.

[0240] 대표적인 예로서, 본 발명에 따라 융합 단백질에서 연결되는 단백질은 효소, 구조적 단백질, 항체, 항원, 프리온, 수용체, 리간드, 사이토카인, 케모카인, 호르몬 등등 또는 이들의 임의의 조합일 수 있다.

[0241] 일부 구현예에서, 본 발명의 및 방법에서 사용하기 위한 재조합 또는 합성 폴리펩타이드는 펩타이드 링커가 유래되는 이소펩타이드 단백질에 대한 상이한 이소펩타이드 단백질 또는 이소펩타이드 단백질이 아니다.

[0242] 일부 구현예에서, 융합 단백질은 반복된 구조를 포함하고, 예를 들면 동일한 단백질은 함께 연결될 수 있다. 대안적으로 보면, 융합 단백질은 동일한 서열의 2개 이상의 단백질 유닛을 함유할 수 있다. 융합 단백질이 동일한 서열의 2개 이상의 단백질 유닛을 함유하면, 이들 단백질 유닛은 연속적, 예를 들면 단백질 유닛을 함께 연결하는 펩타이드 링커에 의해서만 분리될 수 있거나, 이들은 비-연속적 또는 비-순차적일 (예를 들면 상이한 서열을 가진 하나 이상의 단백질에 의해 분리될) 수 있다. 일부 바람직한 구현예에서, 융합 단백질은 상이한 서열을 가진 적어도 2개의 단백질, 예를 들면 상이한 서열을 가진 적어도 2, 3, 4, 5, 6 단백질을 포함한다. 상이한 서열을 가진 단백질은, 융합 단백질의 목적에 의존하여, 임의의 적합한 순서로 배열될 수 있다.

[0243] 더욱 추가 구현예에서, 단백질은 본원에서 한정된 바와 같이 2개 이상의 펩타이드 링커 및 선택적으로, 상기 펩타이드 링커를 연결하는, 하나 이상의 스페이서, 예를 들면 펩타이드 스페이서로 구성될 수 있다. 이 점에 있어서, 단백질은, 상기 기재된 바와 같이, 비-기능성 단백질로서 또는 링커 단백질/펩타이드로서 보여질 수 있다. 이들 구현예에서, 융합 단백질에서 다른 단백질은 상이한 단백질 또는 기능성 단백질이다, 즉 펩타이드 링커 및 스페이서 이외의 서열을 포함한다. 따라서, 일부 구현예에서 융합 단백질은 서열 식별 번호: 56-59의 어느 하나에서 제시된 바와 같은 아미노산 서열을 포함하는 또는 서열 식별 번호: 56-59의 어느 하나에서 제시된 바와 같은 아미노산 서열과 적어도 70% 서열 동일성을 가진 하나 이상의 단백질을 포함하고, 여기에서 상기 단백질은 상기 한정된 바와 같은 적어도 2개의 펩타이드 링커를 포함한다. 예를 들어, 비-기능성 단백질은, 즉 제1 및 제3 단백질을 연결하는, 융합 단백질에서 제2 단백질, 또는, 즉 제3 및 제5 단백질을 연결하는, 융합 단백질에서 제4 단백질 등등으로서 사용될 수 있다. 상기 대표적인 예에서, 제2 및 제4 단백질은 동일한 단백질 또는 상이한 단백질일 수 있다. 따라서, 일부 구현예에서, 융합 단백질에서 단백질 유닛은 링커 단백질 대안적으로, 예를 들면 기능 단백질-링커 단백질-기능성 단백질, 또는 링커 단백질-기능성 단백질-링커 단백질 등을 포함할 수 있다.

[0244] 일부 구현예에서, 상기 단백질은 비교되는 서열과 적어도 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100% 동일성이다.

[0245] "융합 단백질"은, 본원에서 한정된 바와 같이 공유 결합, 바람직하게는 이소펩타이드 결합에 의해 함께 연결된, 적어도 2개의 단백질 유닛, 예를 들면 2, 3, 4, 5, 6, 7, 8, 9 또는 10 이상 단백질 유닛, 예컨대 15, 20, 25 또는 50 단백질 유닛을 포함하는 폴리머로서 한정될 수 있다. 단백질 유닛은 적어도 40 연속 아미노산을 포함하는 분자로서 한정될 수 있고, 바람직하게는 여기에서 단백질은 생체내 기능을 갖고, 예를 들면 여기에서 단백질은 구체적으로 하나 이상의 생물학적 성분과 상호작용할 수 있고, 예를 들면 여기에서 단백질은 생체내 활성이다. 따라서, 융합 단백질은 본원에서 한정된 바와 같이 공유 결합, 바람직하게는 이소펩타이드 결합에 의해 함께 연결된, 적어도 2개의 단백질 유닛, 예를 들면 2, 3, 4, 5, 6, 7, 8, 9 또는 10 이상 단백질 유닛, 예컨대 15, 20, 25 또는 50 단백질 유닛을 포함하는 초구조체, 거대분자, 초분자 또는 다단백질로서 보여질 수 있다.

[0246] 융합 단백질에서 2개 이상의 단백질에 대하여 본 발명의 문맥에서 용어 "연결하다", "연결된" 또는 "연결하는"은 상기 단백질에서 편입되는 펩타이드 링커 (예를 들면 상기 단백질의 도메인을 형성하는 펩타이드 링커) 사이 형성하는 공유 결합, 특히 이소펩타이드 결합을 통해 상기 단백질의 결합 또는 접합을 지칭하였다.

[0247] 본 발명이 이소펩타이드 결합을 형성하기 위해 함께 반응하는 쌍 링커에 관하여 상기 기재되는 동안, 링커의 각각의 (동족) 쌍은 대안적으로 이소펩타이드 결합을 형성하기 위해 (상기 단백질을 연결/접합하기 위해) 반응하는 2개의 별개의 또는 분리가능한 부분 (태그, 태그 및 결합 파트너)로 형성되는 단일 펩타이드 링커로서 보여질 수 있다. 따라서, 상기 관점으로부터 보여지면, 본 발명은 융합 단백질의 생산을 위하여 2개 직교 펩타이드 링커의 용도로서 보여질 수 있고, 여기에서 각각의 펩타이드 링커는 이소펩타이드 결합을 형성하기 위해 반응하는 2개의 분리가능한 부분을 포함하거나 상기로 이루어지고 여기에서 링커의 각각의 부분은 함께 연결 (접합)되는 단백질에서 편입된다 (단백질의 도메인을 형성한다).

- [0248] 본원에서 기재된 방법 및 용도 그리고 본원에서 기재된 방법으로부터 수득된 또는 수득가능한 융합 단백질이 광범위한 유용성을 갖는다는 것이 분명할 것이다. 대안적으로 보면, 본원에서 기재된 방법에 의해 생산된 융합 단백질은 다양한 산업에서 이용될 수 있다. 예를 들어, 본 발명의 방법은 예방접종을 위한 융합 단백질의 생산에 유용할 수 있다. 이 점에 있어서, 항원 다량체화가 크게 향상된 면역 반응을 제공하기 때문에, 방법은, 직접적으로 주사된 또는 바이러스성-유사 입자 (VLPs)를 꾸미는데 사용된, 사슬 내부에 단백질 항원의 연결에 유용할 수 있다.
- [0249] 본 발명의 방법은 향상된 효소 특성, 예를 들면 기질 채널링을 가진 융합 단백질의 생산에 유용할 수 있다. 이 점에 있어서, 효소는 종종 세포 내부 경로에서 기능하기 위해 합치고 종래에 세포 외부 (시험관내)에서 함께 다중 효소를 연결하는 것이 어렵게 되고 있다. 따라서, 본 발명의 방법은, 산업 전환의 범위에서 및 진단에 유용할 수 있는, 다중-단계 효소 경로의 활성을 향상시키는데 사용될 수 있다.
- [0250] 본 발명의 융합 단백질은 또한 그의 안정성에 대하여 개선된 특성을 가질 수 있다, 즉 융합 단백질에서 단백질 유닛의 안정성은 독립적인 단백질로서 그의 안정성에 비하여 향상될 수 있다. 특히, 융합 단백질은 단백질 유닛의 열안정성을 개선할 수 있다. 이 점에 있어서, 효소는 많은 공정에서 귀중한 도구이지만 불안정하고 회수하기 어렵다. 효소 폴리머는 온도, pH 및 유기 용매에 대해 더 큰 안정성을 갖고 산업 공정에서 효소 폴리머를 사용하는 증가된 요구가 있다. 그러나, 본 발명에 앞서, 효소 폴리머 생성은 글루타르알데하이드 비-특이적 반응을 통상적으로 사용하였고 이는 많은 잠재적으로 유용한 효소를 손상 또는 변성시킬 것이다 (즉 효소의 활성을 감소시킬 것이다). 본 발명에 따라 이소펩타이드 결합을 통해 사슬 (폴리머) 내부에 단백질의 부위-특이적 연결은, 예컨대 동물 사료에 첨가된 진단 또는 효소에서, 효소 회복력을 향상시킨다고 예상된다. 특히 바람직한 구현예에서, 효소는, 상기 논의된 바와 같이 원형화에 의해 안정화될 수 있다.
- [0251] 본 발명의 방법은 또한 항체 폴리머의 생산에서 유용성을 찾을 것이다. 이 점에 있어서, 항체는 의약품의 가장 중요한 부류 중 하나이고 표면에 부착되어 종종 사용된다. 그러나, 샘플에서 항원 혼잡, 및 따라서 상기 샘플에서 상기 항원의 포착은 표면 가까이에서 비효율적이다. 항체의 사슬 확장에 의해, 포착 효율이 개선될 것이 기대된다. 이는 순환하는 중앙 세포 단리에서 특히 귀중할 것이고, 현재는 초기 암 진단을 가능하게 하는 가장 유망한 방식 중 하나이다. 또한 상이한 특이성의 항체는 임의의 요망된 순서로 조합될 수 있다.
- [0252] 더욱 추가 구현예에서, 본 발명의 방법은 세포 신호전달 활성화를 위한 약물의 생산에서 유용성을 찾을 수 있다. 이 점에 있어서, 세포 기능을 활성화하기 위한 가장 효과적인 방식의 다수는 단백질 리간드를 통해서이다. 그러나, 사실상 단백질 리간드는 일반적으로 단독으로 작동하지 않지만 다른 신호전달 분자의 특이적 조합으로 작동할 것이다. 따라서, 본 발명의 방법은, 세포 신호전달의 최적의 활성화를 제공할 수 있는, 맞춤형 융합 단백질 (즉 단백질 팀)의 생성을 허용한다. 이들 융합 단백질 (단백질 팀)은 세포 생존, 분할, 또는 분화 제어에 적용될 수 있다.
- [0253] 더욱 추가 구현예에서, 본 발명의 펩타이드 링커, 특히 본 발명의 링커의 쌍은 줄기세포의 성장을 위한 하이드로겔의 생성, 생체적합물질의 제조, 염료 또는 효소로 항체 작용화 및 고리화에 의한 효소 안정화에서 유용성을 찾을 수 있다.
- [0254] 본 발명은 이제 하기 도면과 관련하여 하기 비-제한적인 예에서 더 상세히 기재될 것이다:

도면의 간단한 설명

- [0255] **도 1**은 펩타이드 링커의 2개 직교 쌍, SnoopTag/Snoop Catcher 및 SpyTag/SpyCatcher를 이용하는 융합 단백질의 고형 상 합성의 대표적인 예의 도식을 보여준다.
- 도 2**는 SnoopTag 및 SnoopCatcher 펩타이드 링커 쌍이 유래되는 RrgA 단백질에서 이소펩타이드 결합 형성의 도식을 보여준다 (단백질 데이터뱅크 ID 2WW8에 기반된 넘버링).
- 도 3**은 SnoopTag의 반응성 Lys (KA) 또는 SnoopCatcher의 반응성 Asn (NA)의 알라닌 돌연변이를 가진 대조군과 함께 SnoopCatcher으로 SnoopTag-MBP 반응을 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.
- 도 4**는 (A) SnoopCatcher 대 SnoopTag-MBP의 1:1 또는 2:1 비로 SnoopTag 반응의 시간경과를 도시하는 그래프; (B) SnoopCatcher 대 SnoopTag-MBP의 2:1 비에서 SnoopCatcher로 SnoopTag-MBP 반응을 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진; (C) SnoopCatcher 대 SnoopTag-MBP의 1:1, 2:1 또는 4:1 비로 SnoopTag 반응의 시간경과를 도시하는 그래프; 및 (D) SnoopCatcher 대 SnoopTag-MBP의 4:1 비에서 SnoopCatcher로

SnoopTag-MBP 반응을 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.

도 5는 (A) SnoopTag-MBP와 SnoopCatcher 사이 이소펩타이드 결합 형성의 pH-의존을 도시하는 막대형 차트; 및 (B) SnoopTag-MBP와 SnoopCatcher 사이 이소펩타이드 결합 형성의 온도-의존을 도시하는 차트를 보여준다.

도 6은 (A) 염, 환원제 및 세제에서 SnoopTag-MBP와 SnoopCatcher 사이 이소펩타이드 결합 형성의 의존을 도시하는 막대형 차트; 및 (B) SnoopTag-MBP와 SnoopCatcher 사이 이소펩타이드 결합 형성의 TMAO-의존을 도시하는 그래프를 보여준다.

도 7은 SnoopTag/SnoopCatcher 및 SpyTag/SpyCatcher 직교 반응성을 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.

도 8은 (A) PsCsTag/PsCsCatcher, SnoopTag/SnoopCatcher 및 SpyTag/SpyCatcher 직교 반응성을 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진; 및 (B) RrgATag/RrgACatcher, SnoopTag/SnoopCatcher 및 SpyTag/SpyCatcher 직교 반응성을 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.

도 9는 (A) 고품 상 융합 단백질 합성을 분석하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다. 레인 1-3은 단리에서 MBPx-SpyCatcher, SnoopTag-Affi-SpyTag 및 SpyCatcher-SnoopCatcher를 보여준다. MBPx-SpyCatcher는 아밀로오스 수지에 결합되었고 SnoopTag-아피바디-SpyTag 및 SpyCatcher-SnoopCatcher와 단계식 반응은 수행되었다. 각각의 단계 이후, 하나의 분취량의 샘플은 말토오스를 가진 수지로부터 용출되었다 (레인 4-13). 샘플은 임의의 추가 정제 없이 분석되었고; (B) 고품 상 융합 단백질 합성을 분석하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다. 레인 1-3은 단리에서 바이오틴-SpyCatcher, SnoopTag-Affi-SpyTag 및 SpyCatcher-SnoopCatcher를 보여준다. 바이오틴-SpyCatcher는 스트렙타비딘 아가로오스에 결합되었고 SnoopTag-Affi-SpyTag 및 SpyCatcher-SnoopCatcher와 단계식 반응은 수행되었다. 각각의 단계 이후, 하나의 분취량의 샘플은 바이오틴을 가진 아가로오스로부터 용출되었다 (레인 4-13). 샘플은 임의의 추가 정제 없이 분석되었다.

도 10은 (A) 10량체 융합 단백질, 바이오틴-SpyCatcher:(SnoopTag-Affi-SpyTag:SpyCatcher-SnoopCatcher)₄:SnoopTag-Affi-SpyTag의 동일성을 시험하기 위한 전기분무 이온화 질량 분광분석법을 도시하는 그래프 및 (B) 10량체 융합 단백질, MBPx-SpyCatcher:(SnoopTag-Affi-SpyTag:SpyCatcher-SnoopCatcher)₄:SnoopTag-Affi-SpyTag의 크기-제외 크로마토그래피 분석을 도시하는 그래프를 보여준다. 삽입그림은 분자량 표준을 보여준다.

도 11은 (A) 10량체 융합 단백질, MBPx-SpyCatcher:(SnoopTag-Affi-SpyTag:SpyCatcher-SnoopCatcher)₄:SnoopTag-Affi-SpyTag의 열안정성을 분석하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진; 및 (B) 10량체 융합 단백질, 바이오틴-SpyCatcher:(SnoopTag-Affi-SpyTag:SpyCatcher-SnoopCatcher)₄:SnoopTag-Affi-SpyTag의 시간-의존적 안정성을 분석하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.

도 12는 고품 상 융합 단백질 합성을 분석하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다. 레인 1-3은 단리에서 MBPx-SpyCatcher, SnoopTag-mEGFP-SpyTag 및 SpyCatcher-SnoopCatcher를 보여준다. MBPx-SpyCatcher는 아밀로오스 수지에 결합되었고 SnoopTag-mEGFP-SpyTag 및 SpyCatcher-SnoopCatcher와 단계식 반응은 수행되었다. 각각의 단계 이후, 하나의 분취량의 샘플은 말토오스를 가진 수지로부터 용출되었다 (레인 4-9). 샘플은 임의의 추가 정제 없이 분석되었고; (B) 고품 상 융합 단백질 합성을 분석하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다. 레인 1-3은 단리에서 MBPx-SpyCatcher, SnoopTag-SpyTag-Affi3 및 SpyCatcher-SnoopCatcher를 보여준다. 단계식 반응은 수행되었고 (A)에서와 같이 분석되었다.

도 13은 본 발명의 방법을 이용하여 수득될 수 있는 2개의 단순 분지형 융합 단백질 구조의 만화를 보여준다.

도 14는 RrgACatcher와 반응된 MBP에 융합된 미돌연변이된 RrgATag (서열 식별 번호: 9)와 RrgACatcher와 반응된 MBP에 융합된 돌연변이된 RrgATag (RrgATag2.0, 서열 식별 번호: 111)의 활성을 비교하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.

도 15는 RrgACatcher와 반응된 (SUMO에 융합된) 다양한 RrgATag 펩타이드 링커 돌연변이체를 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.

도 16은 RrgATag2 대 RrgACatcher의 1:1, 2:1 또는 4:1 비로 RrgATag2 반응의 시간경과를 도시하는 그래프를

보여준다. 삽입 그래프는 반응의 첫 8 분 동안 반응을 보여준다.

도 17은 SnoopTag, SnoopCatcher, SpyTag, SpyCatcher 및 RrgATag2와 RrgACatcher 반응성을 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.

도 18은 RrgATag2/RrgACatcher, SnoopTag/SnoopCatcher 및 SpyTag/SpyCatcher 직교 반응성을 특성규명하는 쿠마씨 염색을 수반한 SDS-PAGE 겔의 사진을 보여준다.

발명을 실시하기 위한 구체적인 내용

실시예

실시예 1 - 자발적 이소펩타이드 결합을 형성하는 펩타이드 링커의 동족 쌍의 설계 및 합성

RrgA (서열 식별 번호: 21)은, 인간에 있어서 패혈증, 폐렴 및 수막염을 유발할 수 있는 그람-양성 박테리움인, *연쇄상구균 뉴모니아*로부터 부작소이다. 자발적 이소펩타이드 결합은 잔기 Lys742와 Asn854 사이 RrgA의 D4 면역글로불린-유사 도메인에서 형성한다 (도 2). 본 발명자는 SnoopTag (잔기 734-748, 서열 식별 번호: 1)로 지칭되는 펩타이드 링커의 쌍 및 SnoopCatcher (잔기 749-860, 서열 식별 번호: 2)로 명명한 단백질로 D4 도메인을 분할한다.

그러나, 본 발명자들은 본 발명에서 사용하기 위한 안정적인 펩타이드 링커의 쌍을 형성하기 위해 SnoopCatcher 펩타이드 링커 내부에 2개 돌연변이를 도입하는 것이 필요하다는 것을 알아내었다. 이 점에 있어서, 본 발명자는 β -가닥을 안정화시키기 위해 SnoopCatcher에 G842T 돌연변이 및 반응 부위에 근접한 헤어핀 턴을 안정화시키기 위해 D848G를 도입하였다.

SnoopTag 펩타이드는 말토오스 결합 단백질 (MBP) 및 His-Tag (서열 식별 번호: 50)에 융합된 재조합 폴리펩타이드로서 발현되었다. SnoopCatcher는 His-Tag (서열 식별 번호: 39)에 융합된 재조합 폴리펩타이드로서 발현되었다. SnoopTag-MBP 및 SnoopCatcher는 *에스캐리치아 콜라이*의 사이토솔에서 가용성 단백질로서 효율적으로 발현되었고 Ni-NTA 친화성 크로마토그래피에 의해 정제되었다. SnoopTag-MBP 및 SnoopCatcher는, 간단히 혼합시, SDS에서 비등에 안정적인 복합체를 형성하였다 (도 3). SnoopTag (SnoopTag KA-MBP)의 추정 반응성 Lys742 및 SnoopCatcher (SnoopCatcher NA)의 추정 반응성 Asn854에서 돌연변이는 반응을 폐지시켰다 (도 3). 전기분무 이온화 질량 분광분석법은 SnoopCatcher와 합성 SnoopTag 펩타이드 사이 이소펩타이드 결합 형성으로부터 NH_3 의 손실을 지지하였고; *E. coli* 과발현에 혼한 아세틸화된 및 글루코닐화된 부산물은 또한 관측되었다.

1:1로 SnoopCatcher 대 SnoopTag-MBP 반응은 ~80% 수율로 진행하였다. 그러나, 2-배 과잉의 SnoopCatcher로, SnoopTag-MBP는 정량적으로 반응하였다 (도 4A 및 B). 유사하게 과잉의 SnoopTag-MBP로, SnoopCatcher는 ~100% 소비되었다 (도 4C 및 D).

본 발명자들은 반응이 pH 6-9에서 효율적으로 진행하였지만, pH 5에서 느려졌다는 것을 추가로 확립하였다 (도 5A). 반응은 실온에서 가장 빨랐지만 또한 4°C 및 37°C에서 발생하였다 (도 5B). 시스테인은 SnoopTag 및 SnoopCatcher에서 부재이고 그래서, 예상대로, 반응은 디티오프레이톨 (DTT)에 비감수성이었다. 특이적 완충액 성분은, PBS에서 뿐만 아니라 세제 트리톤 X-100 및 Tween 20, 또는 고 염 (1 M NaCl)의 존재하에 반응으로, 요구되지 않았다 (도 6A). 화학 샤프론 트리메틸아민 N-옥사이드 (TMAO)은 보통의 향상을 제공하였다 (도 6B).

아미드 결합의 자발적 가수분해는 정상적으로 중성 조건하에 수년이 걸리지만 우리는 가수분해가 이러한 특정한 단백질 환경에서 가속화되었는지를 시험하였다. 우리는 과잉의 대안적인 SnoopTag-연결 단백질과 또는 암모니아와 경쟁함으로써 SnoopTag-MBP/SnoopCatcher 상호작용의 절단을 조사하였지만 가역성을 관측하지 못했다.

펩타이드 링커의 추가 쌍은 펩타이드 링커의 SnoopTag/SnoopCatcher 쌍에 상이한 방향으로 D4 면역글로불린-유사 도메인의 분할에 의해 RrgA 단백질로부터 개발되었다. 펩타이드 링커의 상기 쌍은 RrgATag (서열 식별 번호: 9) 및 RrgACatcher (서열 식별 번호: 10)으로 지칭되었다. 펩타이드 링커의 쌍은 또한 PsCsTag (서열 식별 번호: 5) 및 PsCsCatcher (서열 식별 번호: 6)으로 지칭되는 PsCs 단백질 (서열 식별 번호: 31)에 기반하여 개발되었다.

펩타이드 링커의 각각의 쌍은 상기 논의된 SnoopTag/SnoopCatcher 쌍에 유사한 다양한 조건하에 이소펩타이드 결합을 자발적으로 형성할 수 있다.

실시예 2 - 펩타이드 링커의 쌍의 교차-반응성 조사

- [0267] 이소펩타이드 결합을 형성하기 위해 자발적으로 반응하는, 펩타이드 태그 및 결합 파트너, SpyTag 및 SpyCatcher (서열 식별 번호: 13 및 14)는 이전에 개발되었다 (W02011/098772).
- [0268] SnoopTag는 반응성 Lys를 갖고, 반면에 SpyTag는 반응성 Asp를 갖고, 그래서 본 발명자는 SnoopTag/SnoopCatcher 및 SpyTag/SpyCatcher 쌍이 완전히 직교일 것이다, 즉 교차-반응성을 보여주지 않을 것이라고 가정하였다. 다양한 조합으로 펩타이드 링커 혼합시, 펩타이드 링커의 각각의 동족 쌍이 효율적으로 반응하였지만, 쌍 사이 교차-반응의 흔적이 밤새 인큐베이션 이후조차 발견되지 않았음이 밝혀졌다 (도 7). 상기 결과는 SnoopTag/SnoopCatcher 쌍이 SpyTag/SpyCatcher에 직교인 것을 확인하였다.
- [0269] 본 발명자는 또한 SnoopTag/SnoopCatcher 및 SpyTag/SpyCatcher 쌍과 교차-반응성에 대하여 PsCsTag/PsCsCatcher 쌍 및 RrgATag/ RrgACatcher 쌍을 시험하였다. 도 8A 및 B에서 나타낸 바와 같이, 상당한 교차-반응성은 "PsCs" 쌍과 "Spy" 쌍 또는 "Snoop" 쌍 사이 발견되지 않았다. 유사하게, 상당한 교차-반응성은 "RrgA" 쌍과 "Spy" 쌍 또는 "Snoop" 쌍 사이 발견되지 않았다. 따라서, 펩타이드 링커의 각각의 쌍은 펩타이드 링커의 다른 쌍과 직교이다.
- [0270] 실시예 3 - 펩타이드 링커의 2개 직교 쌍을 이용한 융합 단백질의 합성
- [0271] 본 발명자들은 펩타이드 링커의 상기 직교 쌍이 융합 단백질을 합성하는데 성공적으로 사용될 수 있다는 것을 입증하기 위해 펩타이드 링커의 "Spy" 및 "Snoop" 쌍을 사용하였다.
- [0272] 아밀로오스 수지와 *E. 콜리* MBP의 상호작용은 친화도 정제에서 널리 사용된다: MBP-융합은 전형적으로 폴딩하고 양호하게 발현하고 낮은 비-특이적 수지 결합을 보여준다. MBP는, 프로테아제 제거를 위한 필요성을 피하는, 말토오스를 이용하여 선택적 온화한 용출을 보여준다. 말토오스용 야생형 MBP의 친화도는 1.2 μ M이고, 이는 단백질 정제에 실제적이지만 융합 단백질 합성에서 여러 라운드의 세정 및 사슬 확장에 불충분하다. 따라서, 본 발명자들은 그의 말토오스-결합 안정성을 개선하기 위해 돌연변이된 MBP를 개발하였다. 먼저, 본 발명자들은 돌연변이 A312V 및 I317V 도입 및 잔기 172, 173, 175 및 176) 결실에 의해 폴리펩타이드 서열을 변형시켰다. 두번째로, MBP 돌연변이체 (서열 식별 번호: 70)은 MBPx (His₆-MBPmt-링커-MBPmt)를 생성하기 위해 테넨-연결되었다.
- [0273] 초기 사슬 빌딩을 위하여, 본 발명자들은 *E. 콜리* 사이토솔에서 효율적으로 발현된 비-면역글로불린 스캐폴드인, 아피바디를 편입시켰다. HER2에 대한 아피바디는 그의 N-말단에서 SnoopTag와 그리고 그의 C-말단에서 SpyTag (SnoopTag-Affi-SpyTag, 서열 식별 번호: 72)와 연결되었다. 아피바디 유닛은 SnoopCatcher (또한 *E. 콜리*에서 가용성 단백질로서 효율적으로 발현된, SpyCatcher-SnoopCatcher (서열 식별 번호: 56))에 나선 스페이서를 통해 연결된 SpyCatcher를 이용하여 브릿징되었다 (도 1). 각각의 연결이 공유적이기 때문에, 사슬 합성은 수지로부터 용출하기 위한 말토오스의 첨가 및 그 다음 융합 단백질의 확장을 따르기 위해 쿠마씨 염색을 수반한 SDS-PAGE 이전 상청액 비등이 이어질 수 있다 (도 9A). (아밀로오스 수지에 결합된) MBPx-SpyCatcher는 SnoopTag-Affi-SpyTag와 정량적으로 반응하였다 (도 9A, 라인 5). 상기 작제물은 그 다음 SpyCatcher-SnoopCatcher와 정량적으로 반응하였다 (도 9A, 라인 6). SnoopTag-Affi-SpyTag 및 SpyCatcher-SnoopCatcher의 순차적 첨가는, 생성물 10 유닛으로 길게 확장하는, 효율적인 사슬 성장을 가능하게 하였다 (10량체, 도 9A, 라인 13).
- [0274] 고품 상 부착의 상이한 종류로 고품 상 확장을 입증하기 위해, 변형된 SpyCatcher 단백질, AviTag-SpyCatcher는 부위-특이적 N-말단 바이오티닐화를 허용하기 위해 생성되었다. 스트렙타비딘-코팅된 비드에 바이오티닐화된 SpyCatcher 연결 이후, 융합 단백질 사슬은 동일한 방식으로 10량체의 길이로 조립되었고 자유 바이오티닌으로 용출되었다 (도 9B).
- [0275] 조립된 10량체를 검증하기 위해, 관측된 및 예상된 질량 사이 양호한 관련성을 보여주는, 전기분무 이온화 질량 분광분석법은 수행되었다 (도 10A). 질량 분광분석법이 동일성의 양호한 징후를 제공하는 동안, 더 낮은 분자량 부산물이 더욱 효율적으로 이온화하기 때문에, SDS-PAGE는 순도 평가에 훨씬 더 양호하다. 아피바디는, 약간의 자가-회합을 보여주는, 일반적으로 모노머성이고, 따라서 10량체가 응집물을 형성하였는지를 분석하기 위해 크기-제외 크로마토그래피 (SEC)는 수행되었다. SEC는, 구상 단백질 표준으로 보정된 10량체의 예상된 모노머성 질량과 일치하는, 1개의 주요 피크를 제공하였고, 이들 조건하에 최소 10량체 자가-회합이 있었다는 것을 나타냈다 (도 10B).
- [0276] 열안정성을 평가하기 위해, 10량체는 일정 범위의 온도에서 간단히 가열되었고 심지어 70°C에서 크게 가용성이

었다 (도 11A). 저장에 대한 10량체 완전성은 또한 시험되었고 약간의 분해 및 약간의 용해도 손실은 1 또는 4 일 후 관측되었다 (도 11B).

[0277] 사슬 내부에 AffiHER2의 초기 편입으로부터 팽창하여, 다른 단백질 유닛이 직교 이소펩타이드 결합 형성을 이용하여 효율적으로 편입될 수 있다는 것이 보여졌다 (도 12). 이 점에 있어서, 형광 단백질 융합 단백질 사슬은 생성되었다 (도 12A). 병-브러쉬 융합 단백질 폴리머는 N-말단 (SnoopTag-SpyTag-Affi-Affi-Affi)에서 태그 모듈과 HER2에 대해 텐덤-연결된 아피바디의 연결에 의해 또한 생산되었다 (도 12B).

[0278] 요약하면, 본 발명자는, 펩타이드 링커 사이 자발적 이소펩타이드 결합 형성을 통해, 융합 단백질의 합성에 대한 모듈러 접근법을 개발하였다. 본 발명의 방법에 따라 생성된 융합 단백질은 비가역적 아미드 결합을 통해 연결되고, 그래서 (프로테아제로부터 보호되면) 경시적으로 안정적이고 SDS-PAGE에 의한 용이한 분석을 허용한다. 개시, 확장 및 방출 단계는, 산화환원 상태와 독립적으로, 온화한 조건을 이용하고, 그래서 광범위한 단백질에 적용가능해질 수 있다. 사슬이 성장하기 위한 유일한 단일 방식으로, 생성물은 분자적으로 한정되고, 재현성을 선호하고 기능을 정확하게 튜닝한다. 또한, 서브유닛은, 상기 기재된 병-브러쉬 중합체 구조로 보여진 바와 같이, N- 내지 C-배향으로 연결될 필요가 없다. 모듈의 화학 변형은 요구되지 않고, 시간-소비 및 제어하기 어려운 바이오콘주게이션 단계를 피하고, 그래서 방법은 재조합 단백질을 발현할 수 있는 임의의 실험실에 접근가능하다. 자발적 이소펩타이드 결합 형성은 - 카복실산 또는 카복사미드와 아민 - 낮은 고유 반응성을 갖는 2개 작용기 사이 단순 반응 경로의 이점을 갖고 그래서 부반응이 거의 없다.

[0279] 본 실시예가 펩타이드 링커의 "Spy" 및 "Snoop" 쌍을 이용하여 융합 단백질 합성을 이용하는 동안, 본 발명에 따라 펩타이드 링커의 임의의 직교 쌍이 본 발명의 방법에서 이용될 수 있고, 상기 논의된 바와 같이, 펩타이드 링커의 2개 초과 직교 쌍 이용이 복합체 구조, 예를 들면 분지형 또는 원형 구조를 가진 융합 단백질 합성에 특히 유리할 수 있다는 것이 분명할 것이다.

[0280] 실시예 4 - RrgA 단백질에 기반된 펩타이드 링커의 개선된 동족 쌍의 설계 및 합성

[0281] 실시예 1에서 기재된 RrgATag는 RrgATag/RrgACatcher 펩타이드 링커 쌍에 비해 개선된 활성을 가진 펩타이드 링커의 쌍의 생산 목적으로 다양한 변형에 적용되었다.

[0282] 본 발명자는 RrgATag2.0으로 언급되는 - 아스파르트산을 글리신 (D11G)으로 - 위치 11에서 치환을 포함하는 돌연변이체 RrgATag 펩타이드 링커를 합성하였다 (참고 아래 표 2). RrgATag2.0 (서열 식별 번호: 111) 및 RrgATag (서열 식별 번호: 9)는 말토오스 결합 단백질 (MBP)에 연결된 융합 단백질로서 발현되었고 RrgACatcher와 그의 반응성은 비교되었다. 반응은 pH 7.4 및 실온에서 포스페이트 완충된 염수 (PBS)에 6시간 동안 수행되었다. 10 μ M의 각각의 단백질은 각각의 반응에서 사용되었다.

[0283] 도 14는 RrgATag2.0이 RrgATag와 비교된 경우 RrgACatcher으로 반응성을 크게 증가시켰다는 것을 보여준다.

[0284] 본 발명자는, 확장, 절단, 치환 및 이들의 조합을 포함하는, RrgATag (서열 식별 번호: 9)에 비해 다양한 돌연변이를 포함하는 추가 8개 펩타이드 링커를 합성하였다. 표 2는 돌연변이체 RrgATag 펩타이드 링커의 서열을 보여주고, 여기에서 치환 및 확장은 밑줄 친다.

표 2

펩타이드 명칭 (서열 식별 번호:)	서열	RrgATag (서열 식별 번호: 9)에 대한 변형
RrgATag (서열 식별 번호: 9)	DIPATYEFTNDKHYITNEP	-
RrgATag2 (서열 식별 번호: 109)	DIPATYEFTNGKHYITNEP <u>IPPK</u>	D11G (치환) 4 아미노산 C-말단 확장
RrgATag2.0 (서열 식별 번호: 111)	DIPATYEFTNGKHYITNEP	D11G (치환)
RrgATag2.1 (서열 식별 번호: 113)	DIPATYEFTNGKHYITNE	D11G (치환) 1 아미노산 C-말단 결실
RrgATag2.2 (서열 식별 번호: 115)	DIPATYEFTNGKHYITN	D11G (치환) 2 아미노산 C-말단 결실
RrgATag2.3 (서열 식별 번호: 117)	ATYEFTNGKHYITNEP	D11G (치환) 3 아미노산 N-말단 결실
RrgATag2.4 (서열 식별 번호: 119)	KHYITNEP	11 아미노산 N-말단 결실
RrgATag2.5 (서열 식별 번호: 121)	GKHYITNEP	D11G (치환) 10 아미노산 N-말단 결실
RrgATag2.6 (서열 식별 번호: 123)	NGKHYITNEP	D11G (치환) 9 아미노산 N-말단 결실
RrgATag2.7 (서열 식별 번호: 125)	<u>IVPQ</u> DIPATYEFTNGKHYITNEP	D11G (치환) 4 아미노산 N-말단 확장

[0285]

[0286]

돌연변이된 RrgATag 펩타이드 링커는 SUMO (작은 유비퀴틴 조절제) 단백질에 연결된 융합 단백질로서 발현되었고 융합 단백질은 RrgACatcher (서열 식별 번호: 10)으로 반응성에 대하여 시험되었다. 반응은 pH 7.4 및 실온에서 포스페이트 완충된 염수 (PBS)에 30 분 동안 수행되었다. 10 μ M의 각각의 단백질은 각각의 반응에서 사용되었다. 도 15는 변형된 RrgATag 펩타이드 링커의 단지 4개가 RrgACatcher: RrgATag 2.0, RrgATag2.3, RrgATag2 및 RrgATag2.7로 관측가능한 반응성을 보여주었다는 것을 나타낸다. 그러나, RrgATag2는, 상기 논의된 바와 같이, RrgATag에 비해 증가된 활성을 갖는, RrgATag2.0에 비해 활성에서 유의미한 증가를 보여준다. 따라서 RrgATag2는 RrgATag와 비교에서 RrgACatcher로 상당히 개선된 반응성을 갖는다.

[0287]

(SUMO를 가진 융합 단백질의 형태로) RrgATag2와 RrgACatcher 사이 반응의 속도는 도 16에서 보여지고 과잉의 RrgATag2가 반응의 속도를 증가시킨다는 것을 나타낸다. 그러나, 반응은 완료, 즉 RrgATag2의 모든 농도에서, RrgACatcher의 100% 소비에 근접하였다.

[0288]

이론에 의한 구속됨 없이, RrgATag2의 상당히 개선된 활성이 RrgATag에 비해 변형/돌연변이의 조합의 결과라는 것이 가정된다. 이 점에 있어서, 원상태 RrgA 서열에 기반되는, 상기 C-말단 확장은 RrgACatcher 펩타이드 링커와 양호한 상호작용을 형성한다고 생각된다. 더욱이, RrgATag2 펩타이드 링커의 중간에서 D 내지 G 돌연변이 (즉 측사슬의 크기에서 감소)가 (전체 길이 도메인에서 존재하는 결정 구조에서 나타난 바와 같이) 펩타이드에서 헤어핀 턴을 안정화시킨다는 것이 가정된다.

[0289]

실시예 5 - 개선된 RrgATag2 펩타이드 링커의 교차-반응성 조사

[0290]

RrgATag2/RrgACatcher 펩타이드 링커 쌍은, 상기 실시예 3에서 기재된 바와 같이, SnoopTag/SnoopCatcher 및 SpyTag/SpyCatcher 펩타이드 링커 쌍에 대한 교차-반응성에 대하여 시험되었다. RrgATag2 펩타이드 링커는, 실시예 4에서 기재된 바와 같이, SUMO에 연결된 융합 단백질로서 발현되었다.

[0291]

도 17은 상당한 교차-반응성이 RrgACatcher 펩타이드 링커와 SpyTag 또는 SnoopTag 펩타이드 링커 사이 발견되지 않았다는 것을 보여준다. 도 18은 상당한 교차-반응성이 RrgATag2 펩타이드 링커와 SpyCatcher 또는 SnoopCatcher 펩타이드 링커 사이 발견되지 않았다는 것을 보여준다. 따라서, 펩타이드 링커의 각각의 쌍은 펩타이드 링커의 다른 쌍과 직교이다.

- [0292] **물질 및 방법**
- [0293] **클로닝**
- [0294] KOD Hot Start DNA 폴리머라제 (Roche)는 모든 PCR 및 부위 지향적 돌연변이유발을 수행하는데 사용되었다. Gibson Assembly® Master Mix (NEB)는 제조자의 지침에 따라 사용되었다. 작제물은 화학적으로 능숙한 *E. coli* DH5 α (Life Technologies) 내부에 초기에 클로닝되었다.
- [0295] pET28a SpyTag-MBP (Addgene 플라스미드 ID 35050), 글루타티온-S-전달효소-BirA 및 pDEST14-SpyCatcher (GenBank JQ478411, Addgene 플라스미드 ID 35044)는 문헌 [B. Zakeri et al., 2012 (Proc Natl Acad Sci U S A 109, E690-697)]에 기재되어 있다.
- [0296] pET28a SnoopCatcher는 *연쇄상구균 뉴모니아*에 부착소 RrgA의 잔기 749-860 (단백질 데이터뱅크 ID 2WW8에 기재된 넘버링)으로부터 DNAWorks 프라이머-매개된 어셈블리에 의해 생성되었고, HindIII로 소화되었고 pET28a 내부에 아클론화되었다. SnoopTag로 반응을 최적화하기 위해, G842T 돌연변이는 상기 작제물에서 5'-GTGCCGAGGATATTCGGCTACATATGAATTACCAACG (서열 식별 번호: 73)으로 퀵체인지(QuickChange), 및 5'-GCTACATATGAATTACCAACGGTAAACATTATATCACCAATGAACC (서열 식별 번호: 74) 및 그의 역방향 보체로 D848G 돌연변이에 의해 실시되었다. SnoopCatcher는 (fMet 절단을 추정하는) 132 잔기 길이이고 N-말단 트롬빈 절단 부위 및 His₆ 태그를 갖는다. pET28a SnoopCatcher NA는 정방향 프라이머 5'-ACATTATATCACCGTGAACCGATACCGCCG (서열 식별 번호: 75) 및 그의 역방향 보체를 이용하여 A로 N854의 퀵체인지에 의해 pET28a SnoopCatcher로부터 생산되었다.
- [0297] pET28a SnoopTag-MBP는 2 단계로 생성되었다. 먼저 RrgA의 D4 도메인의 N-말단 β -가닥에 기반된 반응성 펩타이드는 하기를 이용하는 부위 지향적, 리가아제-독립적인 돌연변이유발 (SLIM) PCR (Chiu et al., 2004)에 의해 pET28a-SpyTag-MBP 내부에 클로닝되었다 (잔기 734-748): 5'-GGTAGTGGTGAAAGTGGTAAAATCGAAGAAG (서열 식별 번호: 76), 5'-AAACTGGGCGATATTGAATTTATTAAGTGAACAAAACGATAAAGGTAGTGGTGAAGTGGTAAAATCGAAGAAG (서열 식별 번호: 77), 5'-TCCCATATGGCTGCCGCGCG (서열 식별 번호: 78) 및 5'-TTTATCGTTTTGTTCACCTTAATAAATTCATATCGCCGAGTTTCCCATATGGCTGCCGCGCG (서열 식별 번호: 79). 펩타이드의 3 C-말단 잔기는 5'-GAATTTATTAAGTGAACAAAGGTAGTGGTGAAGTGGTAAAATCG (서열 식별 번호: 80) 및 그의 역방향 보체로 퀵체인지를 이용하여 제거되었다. SnoopTag의 무반응성 버전인, pET28a SnoopTag KA-MBP는 5'-GGGCGATATTGAATTTATTGCAGTGAACAAAGGTAGTGG (서열 식별 번호: 81) 및 그의 역방향 보체를 이용하여 pET28a SnoopTag-MBP에서 A로 K742의 퀵체인지에 의해 생성되었다.
- [0298] pET28a MBP-SpyCatcher는, 중첩 확장 PCR을 통해, MBP의 C-말단에서 Gly/Ser 스페이스와 SpyCatcher 융합에 의해 생성되었다. SpyCatcher는 정방향 프라이머 5'-GTTCCGGGCGGTAGTGGTGCATGCTTATACCTTATCAGTTTATCAAGTGAGCAAG (서열 식별 번호: 82) 및 역방향 프라이머 5'-TACTAAGCTTCTATTAATATAGCGTCACCTTTAGTTGCTTTGCCATTTACAG (서열 식별 번호: 83)을 이용하여 pDEST14-SpyCatcher로부터 증폭되었다. 정방향 프라이머 5'-ATCTCATATGGGCAGCAGCCATCATCATCATCAC (서열 식별 번호: 84) 및 역방향 프라이머 5'-GTATCAACCATGGCACCCTACCGCCGAAACCGAGCTCGAATTAGTCTGCG (서열 식별 번호: 85)는 pET28a SpyTag-MBP로부터 MBP를 증폭시키는데 사용되었다. 2개의 수득한 PCR 생성물은 혼합되었고 SpyCatcher 정방향 프라이머 및 MBP 역방향 프라이머를 이용하여 증폭되었고, NdeI 및 HindIII로 소화되었고, pET21 내부에 아클론화되었다. 아밀로오스에 대한 MBP-SpyCatcher의 친화도를 증가시키기 위해 우리는 먼저 정방향 프라이머 5'-GTCTTACGAGGAAGAGTTGGTGAAAGATCCACGTGTGGCCGCACTATGGAACGC (서열 식별 번호: 86) 및 그의 역방향 보체를 이용하여 퀵체인지에 의해 MBP에서 A312V 및 I317V 돌연변이를 실시하였다. 잔기 172, 173, 175 및 176은 5'-GGGTTATGCGTTCAAGTATGGCGACATTAAGACGTGGGCG (서열 식별 번호: 87) 및 그의 역방향 보체로 퀵체인지를 이용하여 MBP로부터 결실되었다. 우리는 그 다음 5'-CACCATCACCATCAGATTACGATAGTGCTACCCATATTAATTCTC (서열 식별 번호: 88) 및 그의 역방향 보체를 이용하는 퀵체인지에 의해 SpyCatcher의 N-말단을 단축시켰다. 심지어 추가로 아밀로오스 수치로부터 해리를 감소시키기 위해, 상기 돌연변이체 MBP의 탠덤 연결은 MBPx-SpyCatcher (N-말단 His₆ 태그-MBPmt-스페이스-MBPmt-스페이스-SpyCatcher)를 제공하기 위해 생성되었다. MBPx는 증폭되었고 정방향 프라이머 5'-GGCGGATCCGGAGGTGGATCCGAAAGATAGAGGAGGGTAACTGGTAATCTGG (서열 식별 번호: 89), 역방향 프라이머 5'-CCTATAGTGAGTCGTATTAATTTTCG (서열 식별 번호: 90), 정방향 프라이머 5'-CGAAATTAATACGACTCACTATAGG (서열 식별 번호: 91) 및 역방향 프라이머 5'-TCCGATCCACCTCCGGATCCGCGGAAGTGAATTCGTCTGCGCGTCTTCAGG (서열 식별 번호: 92)를 이용하여 Gibson 어셈블리를 통해 MBPx-SpyCatcher에 융합되었다.

- [0299] pET28a SpyCatcher-SnoopCatcher는 단계에서 생성되었다. 초기에 SpyCatcher는 SnoopCatcher의 N-말단에서 Gly/Ser 스페이서로 융합되었고, 그 다음 Gly/Ser 스페이서는 α -나선 스페이서 (서열 PANLKALEAQKQKEQRQAAEELANAKLKEQLEK, 서열 식별 번호: 93)으로 대체되었다. 정방향 프라미어 5'-CTTTAAGAAGGAGATATACATATGTCGTACTACCATCACCATC (서열 식별 번호: 94) 및 역방향 프라이머 5'-CCGCTGCTTCCGGATCCAATATGAGCGTCACCTTTAGTTG (서열 식별 번호: 95)는 pDEST14-SpyCatcher로부터 SpyCatcher 부분을 증폭시키는데 사용되었다. SnoopCatcher 일부는 pET28a SnoopCatcher로부터 정방향 프라미어 5'-CATATTGGATCCGGAAGCAGCGCCTGGTGCCGCGCGGATCCCATGAAGCCGCTGC (서열 식별 번호: 96) 및 역방향 프라이머 5'-GTGGTGGTGGTGGTCTCGAGTTATTATTTTCGGCGGTATCGGTTC (서열 식별 번호: 97)을 이용하여 클로닝되었다. SpyCatcher 및 SnoopCatcher 융합 이후, Gly/Ser 스페이서는 정방향 프라미어 5'-CTAAAGGTGACGCTCATATTGGATCCCCGCCAACCTGAAGGCCCTGGAGGCCAGAGAGAGACAGGCCGCCGAGGAGC (서열 식별 번호: 98) 및 역방향 프라이머 5'-CACGGCACCACGACGCGCTTCATATGGGATCCCTTCTCCAGCTGCTCCTTCAGTTCTTGGCGTTGCCAGCTCCTCGGCGCCTGTC (서열 식별 번호: 99)를 이용하여 안정적인 α -나선 링커로 대체되었다. 35 잔기는 정방향 프라미어 5'-CACCATCACCATCAGATTACGATAGTGCTACCCATATTAATTCTC (서열 식별 번호: 100) 및 그의 역방향 보체를 이용하는 퀵 체인지를 통해 SpyCatcher의 N-말단으로부터 결실되었다.
- [0300] pET28a SnoopTag-Affi-SpyTag (HER2-스페이서-SpyTag에 대한 N-말단 His₆ 태그-SnoopTag-스페이서-아피바디)는 pET28a-KTag-AffiHer2-SpyTag로부터 정방향 프라미어 5'-GTGAACAAAGGCAGTGGTGAGTCGGGATCCGGAGCTAGCATGACTGGTGG (서열 식별 번호: 101) 및 역 5'-CATCACGATGTGGGCACCGAACCTTCCCCGGATCCCTCGAGGCCTTTCGG (서열 식별 번호: 102)를 이용하는 Gibson 어셈블리에 의해 생성되었다.
- [0301] Taq DNA 폴리머라제에 대한 아피바디인, pET28a SnoopTag-AffiTaq-SpyTag는 5'-CTACCAACCTAAACGGGTACAAGTAAAGGCTTTCATAGACTCGTAAAGGATGACCAAGCCAAAGCGC (서열 식별 번호: 103) 및 5'-GTTGAATATCTCCAAGTAGCCACCTAGCTCCTGTTGAACTGTTGTCTACTTCTTGTGAATTTGTTGTCCACGCC (서열 식별 번호: 104)를 이용하는 pET28a SnoopTag-AffiHer2-SpyTag로부터 역 PCR에 의해 생성되었다.
- [0302] pET28a SnoopTag-mEGFP-Spy태그는 pET28a SnoopTag-Affi-SpyTag내 BamHI 부위에서 mEGFP 치환에 의해 그리고 스페이서를 확장하기 위한 PCR에 의해 클로닝되었다. pET28a SnoopTag-SpyTag-Affi3는 Gly/Ser 스페이서에 의해 연결된 AffiHER2의 탠덤 카피의 PCR 어셈블리에 의해 생성되었다.
- [0303] N-말단에서 부위-특이적 바이오티닐화를 위하여 펩타이드 태그를 함유하는, AviTag-SpyCatcher는 5'-GATTACGACATCCCAACGACCGAAAACCTG (서열 식별 번호: 105), 5'-GCCTGAACGATATTTTGAAGCGCAGAAAATTGAATGGCATGAAGGCGATTACGACATCCCAACGACCGAAAACCTG (서열 식별 번호: 106), 5'-GTGATGTGATGGTGTAGTACGACATATG (서열 식별 번호: 107) 및 5'-TGCCATTCAATTTTCTGCGCTTCAAAAATATCGTTCAGGCCGCTGCCGTGATGGTGTAGTACGACATATG (서열 식별 번호: 108)을 이용하는 pDEST14-SpyCatcher로부터 SLIM PCR에 의해 클로닝되었다.
- [0304] 모든 돌연변이 및 작제물은 서열분석에 의해 확인되었다.
- [0305] 단백질 발현 및 정제
- [0306] 단백질은 *E. 콜리* BL21 DE3 RIPL (Agilent)에서 발현되었다. 콜로니는 pET28a 벡터용 0.5 mg/mL 카나마이신 및 pET21용 0.1 mg/mL 암피실린을 함유하는 LB내 37°C에서 밤새 성장되었다. 밤새 배양물은 적절한 항생제와 0.8% 글루코오스를 함유하는 LB내 1:100 희석되었고 37°C, 200 rpm에서 OD₆₀₀ 0.5-0.6으로 성장되었고 30°C, 200 rpm에서 4시간 동안 0.4 mM IPTG로 유도되었다. 단백질은 Ni-NTA (Qiagen)에서 표준 방법으로 정제되었고 TBS (50 mM Tris HCl pH 8.0 및 50 mM NaCl)로 3회 투석되었다.
- [0307] MBPx-SpyCatcher의 정제를 위하여, Ni-NTA로부터 용출 이후, 완충액은 4°C에서 20 mM Tris HCl pH 8.0에 투석에 의해 교환되었고, 4차 고성능 (Q-HP) 수지 (GE Healthcare)에 장입되었고 1 mL/min의 유량으로 0-0.15 M NaCl의 10 칼럼 용적 (즉 10 mL) 선형 구배에 의해 용출되었다. 추가의 용출 단계는 1.5 mL/min의 유량에서 0.15-0.35 M NaCl의 선형 구배 및 0.5 mL 분획 수집으로 수행되었다. 수집된 분획은 TBS에 투석되었고, Vivaspın 원심 농축기 5 kDa 컷오프 (GE Healthcare)를 이용하여 농축되었고 -80°C에서 저장되었다.
- [0308] SnoopTag-Affi-SpyTag는 4 °C에서 20 mM 2-(N-모폴리노)에탄설폰산 (MES) pH 5.8에 투석되었고 설포프로필 고

성능 (SP-HP) 수지 (GE Healthcare)에 장입되었다. 단백질은 0.2-0.5 M NaCl의 선형 구배 적용 및 1 mL 분획 수집에 의해 용출되었다. 용출된 분획은 Vivaspin 원심 농축기 5 kDa 컷오프 (GE Healthcare)를 이용하여 1-2 mg/mL로 농축되었고, TBS pH 8.0에 투석되었고 -80℃에서 저장되었다.

[0309] SpyCatcher-SnoopCatcher의 정제를 위하여, Ni-NTA로부터 용출 이후, 완충액은 4℃에서 20 mM Tris HCl pH 8.0에 투석에 의해 교환되었고, 4차 고성능 (Q-HP) 수지에 장입되었고 0.2-0.5 M NaCl의 선형 구배로 용출되었다. 수집된 분획은 TBS에 투석되었고, Vivaspin 원심 농축기 5 kDa 컷오프를 이용하여 농축되었고 -80℃에서 저장되었다.

[0310] 정제된 AviTag-SpyCatcher는 25℃에서 1시간 동안 5 mM MgCl₂, 1 mM ATP, 380 μM D-바이오틴 및 7 μM GST-BirA를 함유하는 PBS pH 7.4에서 바이오틴화되었다. 1 시간 인큐베이션 이후, 추가 GST-BirA는 첨가되어 14 μM의 최종 농도를 제공하였고 반응은 25℃에서 추가 1시간 동안 인큐베이션되었다. GST-BirA는, 30 분 동안 엔드-오버-엔드 회전하면서, 25℃에서 50 μL의 슬러리 Hi-Cap 글루타티온 매트릭스 (Qiagen)와 반응 혼합물의 인큐베이션에 의해 제거되었다. 수지는 1 분 동안 4,000 g에서 스핀 다운되었다. 상청액은 수집되었고 밤새 4℃에서 PBS에 투석되었다. 완전한 바이오틴화를 확인하기 위해, 스트렙타비딘 겔-이동 검정은 기재된 바와 같이 수행되었다.

[0311] **SDS-PAGE**

[0312] SDS-PAGE는 200 V에서 XCell SureLock 겔 컨테이너 (Life Technologies)를 이용하는 표시된 백분을 폴리아크릴아미드 겔에서 수행되었다. 겔은 Instant Blue Coomassie 얼룩 (Triple Red Ltd.)으로 염색되었고 밴드는 Gel Doc XR 이미지 및 이미지 Lab 3.0 소프트웨어 (Bio-Rad)를 이용하여 밀도측정으로 분석되었다. 모든 작동 완충액은, 높은 M_w 생성물의 해상도를 개선하기 위해 트리스-아세테이트인 도 9A를 제외하고, 트리스-글리신이었다.

[0313] **이소펩타이드 결합 재구성**

[0314] SnoopTag와 SnoopCatcher 사이 공유 결합의 형성을 평가하기 위해, 단백질은 1.5 M 트리메틸아민 N-옥사이드 (TMAO; Sigma-Aldrich)를 함유하는 TBS pH 8.0에 10 μM 최종 농도에서 각각 혼합되었다. TMAO는 화학 샤프론으로서 작용한다. 반응은 6× SDS 장입 완충액 (0.23 M 트리스-HCl, pH 6.8, 24% v/v 글리세롤, 120 μM 브로모페놀 블루, 0.23 M SDS) 첨가에 의해 중단되었다. 샘플은, 16% 폴리아크릴아미드 겔에서 SDS-PAGE 이전, 95℃에서 5 분 동안 Bio-Rad C1000 열 순환기를 이용하여 그 뒤에 가열되었다.

[0315] 직교성을 시험하기 위해, 10 μM SnoopTag-MBP 및 10 μM SnoopCatcher 또는 SpyCatcher는 SDS-PAGE 이전 1시간 동안 25℃에서 TBS pH 8.0에 인큐베이션되었다. 유사하게 10 μM SpyTag-MBP 및 10 μM SnoopCatcher 또는 SpyCatcher는 상기와 같이 인큐베이션되었다.

[0316] 다른 펩타이드 링커 쌍을 위하여, 10 μM RrgATag-MBP 또는 10 μM PsCsTag-MBP 및 10 μM SnoopCatcher, SpyCatcher, SnoopTag-MBP 또는 SpyTag-MBP는 SDS-PAGE 이전 24시간 동안 25℃에서 PBS pH 7.4에 인큐베이션되었다.

[0317] pH-의존성을 평가하기 위해, 각각의 단백질은 10 μM으로 석시네이트-포스페이트-글리신 완충액 (12.5 mM 석신산, 43.75 mM NaH₂PO₄, 43.75 mM 글리신)에서 혼합되었고, pH 4.0 내지 pH 9.0 범위의, 넓은 pH 범위에 걸쳐 적절한 완충을 가능하게 하도록 선택되었고 25℃에서 15 분 동안 인큐베이션되었다.

[0318] 온도의 효과를 결정하기 위해, 10 μM SnoopTag-MBP 및 10 μM SnoopCatcher는 15분 동안 표시된 온도에서 1.5 M TMAO를 함유하는 포스페이트 완충된 염수 (PBS, 10 mM Na₂HPO₄ 137 mM NaCl, 27 mM KCl, 1.8 mM KH₂PO₄) pH 8.0에 혼합되었다. PBS는 Tris 완충액의 pH가 온도에 따라 실질적으로 변화하기 때문에 TBS 대신에 사용되었다.

[0319] 완충액 조성물에 대한 감수성을 조사하기 위해, 단백질은 25℃에서 15분 동안 1 M NaCl과 1% 트리톤 X-100 (w/v), 1% Tween 20 (v/v), 10 mM 에틸렌 디아민 테트라아세테이트 (EDTA), 10 mM MgCl₂, 10 mM DTT 또는 50 mM Tris pH 8.0을 함유하는 PBS pH 8.0, TBS pH 8.0 또는 TBS pH 8.0에 인큐베이션되었다.

[0320] 반응 속도는 1.5 M TMAO를 함유하는 TBS pH 8.0에 표시된 농도에서 SnoopTag-MBP 및 SnoopCatcher 반응 및 25℃에서 다양한 시간 동안 인큐베이션에 의해 결정되었다. 반응은 SDS-PAGE에 앞서 상기 기재된 바와 같이 SDS-장입 완충액에서 중단되었다. % 재구성은, SnoopTag-MBP, SnoopCatcher 및 공유 부가물의 밴드 강도의 합계에 의해 분할된, 공유 부가물의 밴드 강도×100으로서 계산되었다.

- [0321] 경쟁 태그로 가역성을 시험하기 위해, 10 μ M SnoopCatcher는 15 μ M SnoopTag-MBP로 6시간 동안 인큐베이션되었고 그 다음 130 μ M의 최종 농도에서 SnoopTag-Affi-SpyTag는 모두 25℃에서 16시간 동안 첨가되었다. 암모니아로 가역성을 시험하기 위해, 10 μ M SnoopCatcher는 10 μ M SnoopTag-MBP로 2시간 동안 1.5 M TMAO를 함유하는 TBS pH 8.0에 인큐베이션되었고 그 다음 (1 M의 최종 농도로) TBS pH 8.0 또는 NH_4Cl pH 9.0은 모두 25℃에서 16시간 동안 첨가되었다.
- [0322] **질량 분광분석법**
- [0323] 20 μ M SnoopTag-MBP 및 20 μ M SnoopCatcher는 25℃에서 2시간 동안 PBS pH 7.4에 인큐베이션되었다. 질량 분광분석법 분석은 Micromass LCT 비과시간 전기분무 이온화 MS (Micromass)를 이용하여 수행되었고 m/z 스펙트럼은 최대 엔트로피 알고리즘 및 V4.00.00 소프트웨어 (Waters)를 이용하여 분자량 프로파일로 전환되었다. ExPASy ProtParam은, N-말단 fMet가 절단되고 이소헵타이드 결합 형성을 위하여 17.0 Da 차감하면서, 단백질의 아미노산 서열에 기반된 분자 질량을 예측하는데 사용되었다. 비-효소 글루코닐화는 종종 *E. 콜리*에서 His-태그된 단백질의 발현으로부터 관측되고 178 Da 첨가한다. 유사하게, *E. 콜리*-발현된 단백질은 또한 어느 정도의 아세틸화를 경험할 수 있다.
- [0324] 10량체는 15 μ M으로 농축되었고 10 kDa 컷오프를 가진 Amicon Ultra 0.5 mL 원심 필터 (Millipore)를 이용하여 200 mM 암모늄 아세테이트에 완충액-교환되었다. 측정은, 250 mM 암모늄 아세테이트에서 10 mg/mL 세슘 아이오다이드를 이용하여 보정된, 1세대 Synapt 고해상도 질량 분광분석법 사중극자 비과시간 질량 분광분석기 (Waters)에서 수행되었다. 2.5 μ L 분취액의 샘플은, 가정에서 제조된, 금-코팅된 모세관을 통해 나노-전기분무 이온화에 의해 전달되었다. 기기적 파라미터는 아래와 같았다: 공급원 압력 6.0 mbar, 모세관 전압 1.20 kV, 콘 전압 150 V, 트랩 에너지 30 V, 수송 에너지 10 V, 바이어스 전압 5 V, 및 트랩 압력 0.0163 mbar. 질량 스펙트럼은 연해졌고 피크-중심화되었고 질량은 MassLynx v4.1 (Waters)를 이용하여 배정되었다.
- [0325] **융합 단백질의 고형 상 합성**
- [0326] 40 μ L의 슬러리 아밀로오스 수지 (NEB)는 1 mL poly-prep 칼럼 (Bio-Rad)에 인가되었고, 1 mL MilliQ 물로 린스되었고 1 mL TBS pH 8.0으로 평형화되었다. 80 μ L의 최종 용적으로 TBS pH 8.0내 320 pmol 탠덤 MBP-SpyCatcher는 수지에 첨가되었고 ThermoMixer 컴포트 (Eppendorf)에서 700 rpm 진탕하면서 1시간 동안 25℃에서 인큐베이션되었다. 미반응된 단백질은 중력 유동에 의해 칼럼으로부터 제거되었고 수지는 1 mL 세정 완충액 (500 mM NaCl을 가진 50 mM Tris HCl pH 8.0)으로 세정되었다. 80 μ L의 최종 용적으로 TBS pH 8.0내 3 nmol SnoopTag-Affi-SpyTag는 수지에 첨가되었고 700 rpm 진탕하면서 1시간 동안 25℃에서 인큐베이션되었다. 미반응된 SnoopTag-Affi-SpyTag는 중력 유동에 의해 칼럼으로부터 제거되었고 수지는 1 mL 세정 완충액으로 세정되었다. 1.5 M TMAO를 가진 TBS pH 8.0내 4 nmol SpyCatcher-SnoopCatcher는 수지에 첨가되었고 700 rpm 진탕하면서 2시간 동안 25℃에서 인큐베이션되었다. 미반응된 SpyCatcher-SnoopCatcher는 중력 유동에 의해 칼럼으로부터 제거되었고 수지는 1 mL 세정 완충액으로 세정되었다. 사슬은 상기 기재된 조건에 따라 SnoopTag-Affi-SpyTag 및 SpyCatcher-SnoopCatcher의 순차적 첨가에 의해 생산되었다. 사슬은, 수지 세정 이후, 50 mM D-말토오스 (Sigma)를 함유하는 40 μ L TBS pH 8.0 첨가에 의해 용출되었고 700 rpm 진탕하면서 10분 동안 25℃에서 인큐베이션되었다. 사슬은 17,000 g에서 10 초 동안 1.5 mL 마이크로원심관에 칼럼 원심분리에 의해 수집되었다. SnoopTag-mEGFP-SpyTag 및 SnoopTag-SpyTag-Affi3를 함유하는 사슬은 정확하게 동일한 방식으로 합성되었다.
- [0327] 각각의 단계 이후 SDS-PAGE 시험을 위하여, 샘플은 이전에 기재된 바와 같이 용출되었고, 6 \times SDS 장입 완충액으로 혼합되었고 SDS-PAGE 이전 5분 동안 95℃에서 가열되었다.
- [0328] 바이오티닐화된-SpyCatcher-기반 어셈블리를 위하여, 40 μ L의 슬러리 모노머성 아비딘 수지 (Thermo Scientific)는 1 mL poly-prep 칼럼에 인가되었고, 상기과 같이 린스되었고 평형화되었다. 80 μ L의 최종 용적으로 TBS pH 8.0내 4 μ M 바이오티닐화된-SpyCatcher는 수지에 첨가되었고 700 rpm 진탕하면서 1시간 동안 25℃에서 인큐베이션되었다. 미반응된 바이오티닐화된-SpyCatcher는 중력 유동에 의해 칼럼으로부터 제거되었고, 수지는 1 mL 세정 완충액으로 세정되었고, SnoopTag-Affi-SpyTag 및 SpyCatcher-SnoopCatcher의 순차적 첨가는 상기 기재된 바와 같이 수행되었다. 수지 세정 이후, 사슬은 TBS pH 8.0내 칼럼 40 μ L 1 mM D-바이오틴에 인가 및 700 rpm 진탕하면서 4시간 동안 25℃에서 인큐베이션에 의해 용출되었다. 사슬은 미리 나타난 바와 같이 수집되었고 16 및 8% 트리스-글리신 겔에서 SDS-PAGE에 의해 분석되었다.
- [0329] **겔 여과 크로마토그래피**

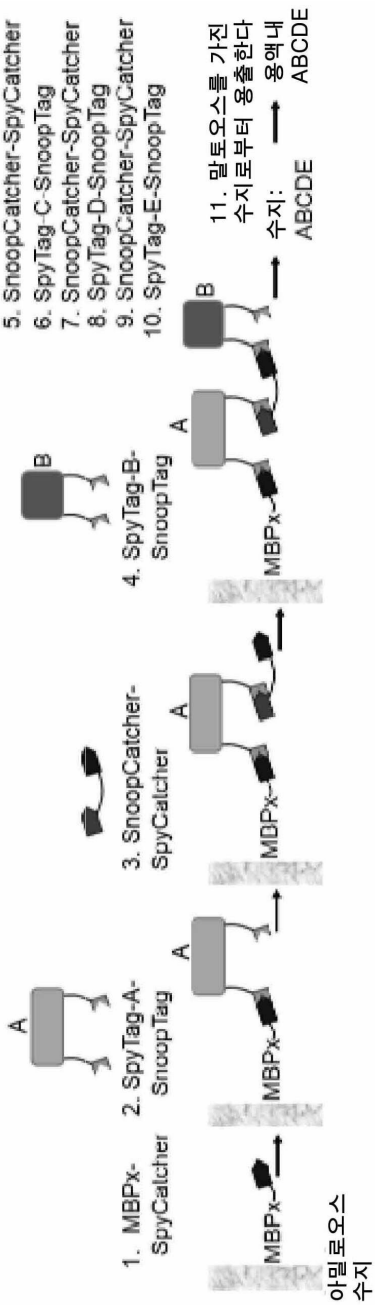
[0330] 10량체 사슬은 Superdex 200 GL 10/300 칼럼 (24 mL 총 용적) (GE Healthcare)에서 겔 여과 크로마토그래피에 의해 분석되었다. 칼럼은 겔 여과 표준 (티로글로불린 670 kDa, IgG 158 kDa, 난백알부민 44 kDa, 미오글로빈 17 kDa, 및 비타민 B12 1.35 kDa) (Bio-Rad) 이용에 의해 보정되었다. 샘플은, AKTA 정제장치 10 (GE Healthcare)에서 280 nm에 측정된 흡광도 프로파일로, 500 mM NaCl을 가진 50 mM Tris HCl pH 8.0내 0.4 mL/min에서 용출되었다.

[0331] **사슬의 안정성 시험**

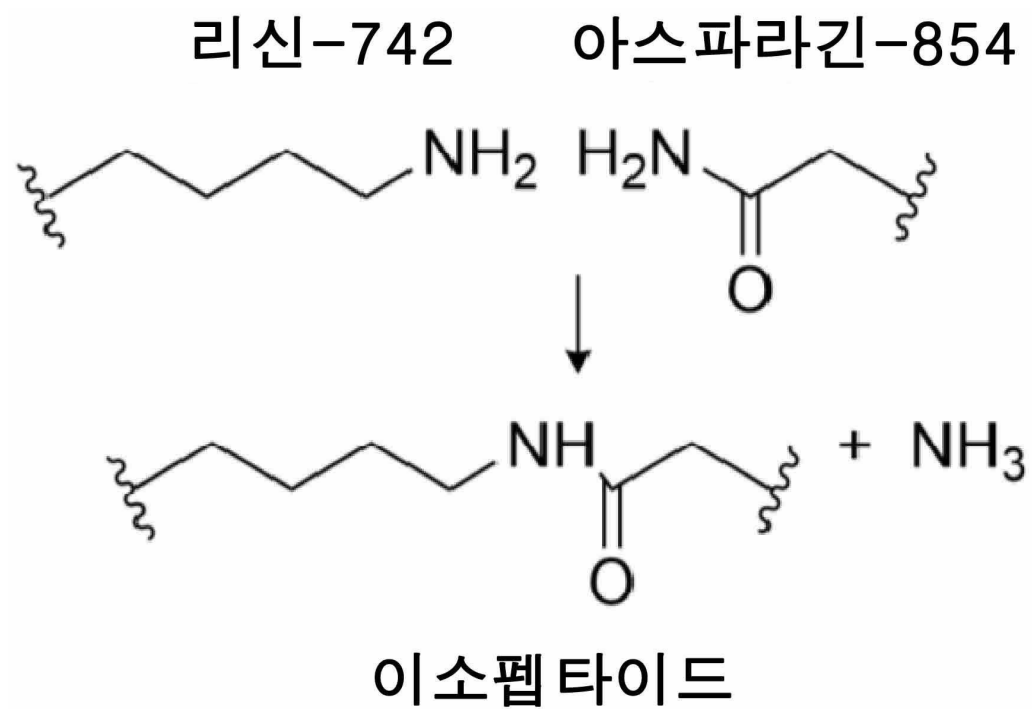
[0332] 온도-안정성 시험을 위하여, 30 μ L의 최종 용적으로 3 μ M에서 150 mM 암모늄 아세테이트 pH 8.0내 10량체 사슬은 3 분 동안 25, 37, 50, 60 또는 70 $^{\circ}$ C에서 인큐베이션되었고 Bio-Rad C1000 열 순환기에서 3 $^{\circ}$ C/s로 10 $^{\circ}$ C까지 냉각되었다. 샘플은 그 다음 17,000 g로 4 $^{\circ}$ C에서 30분 동안 스핀되어 응집물을 제거하였고 상청액은 8% 트리스-글리신 겔에서 쿠마씨 염색을 수반한 SDS-PAGE에 의해 분석되었다. 시간-의존적 안정성 시험을 위하여, 30 μ L의 최종 용적으로 0.1% 아지드화나트륨, 1 mM 페닐메틸설포닐 플루오라이드 (PMSF), 1 mM EDTA, 및 EDTA-무 혼합된 프로테아제 억제제 (Roche)를 함유하는 150 mM 암모늄 아세테이트 pH 8.0내 3 μ M에서 10량체 사슬은 25 $^{\circ}$ C에서 1 또는 4 일 동안 인큐베이션되었다. 각각의 시점에서 샘플은 17, 000 g로 4 $^{\circ}$ C에서 30분 동안 스핀되었고 상청액은 8% 트리스-글리신 겔에서 쿠마씨 염색을 수반한 SDS-PAGE에 의해 분석되었다.

도면

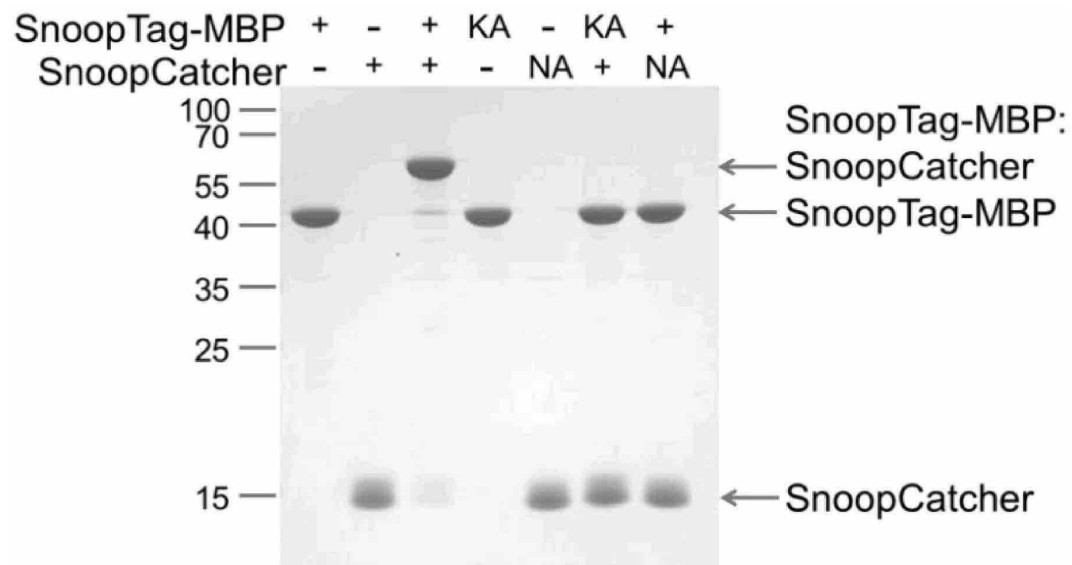
도면1



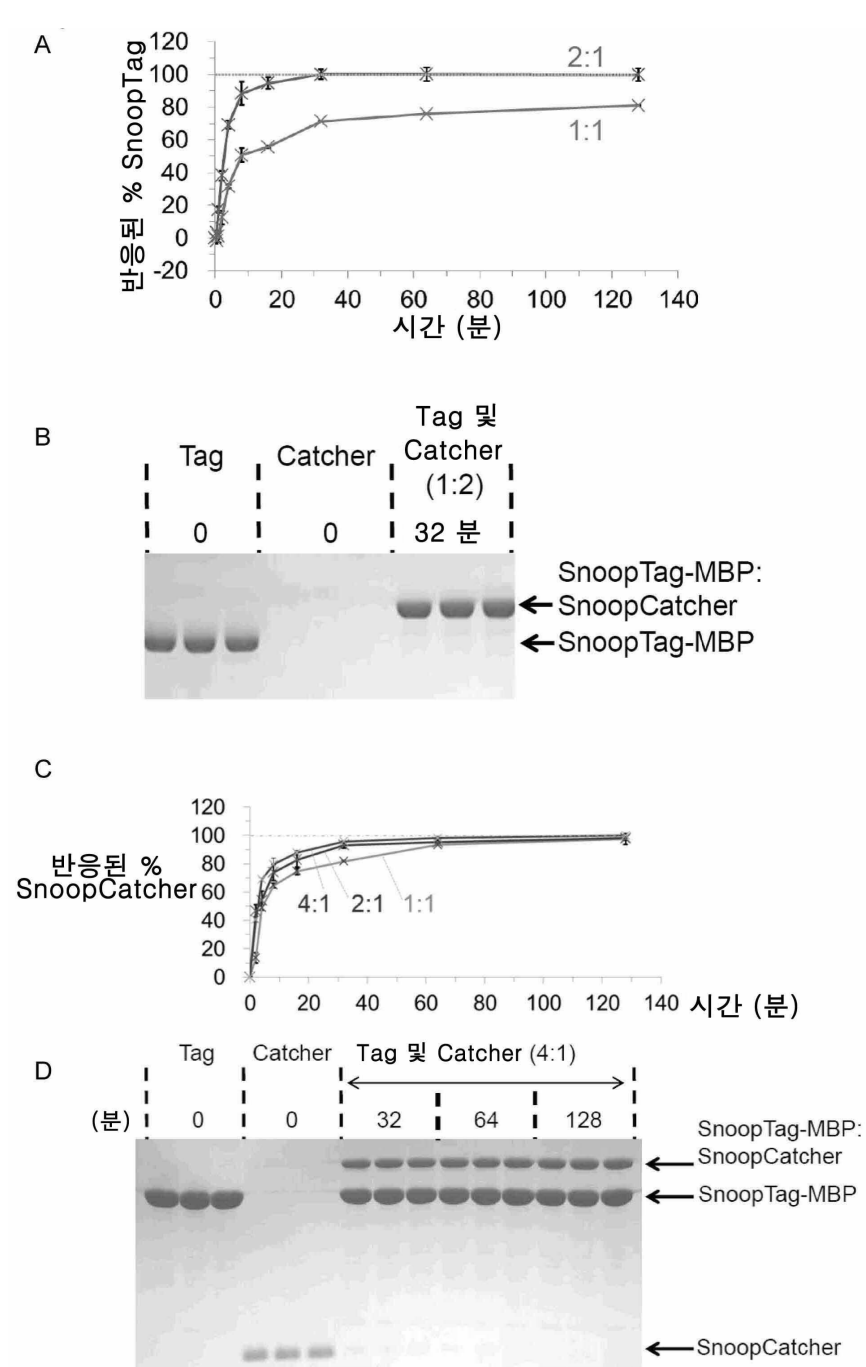
도면2



도면3

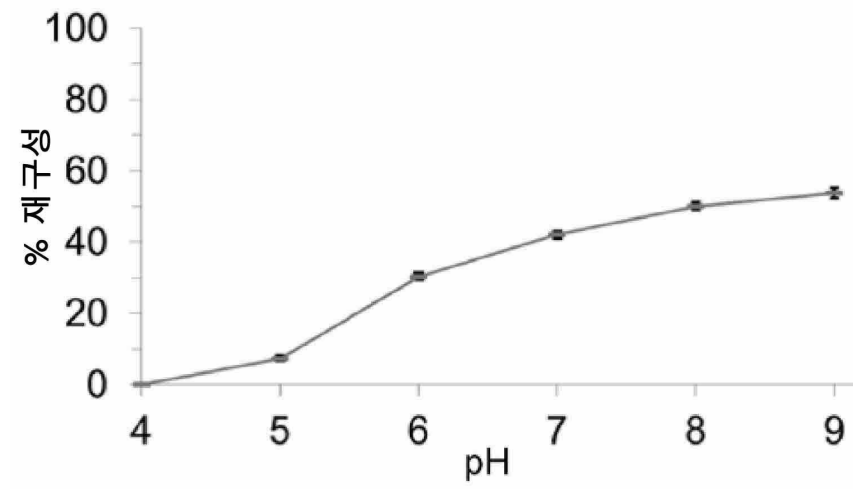


도면4

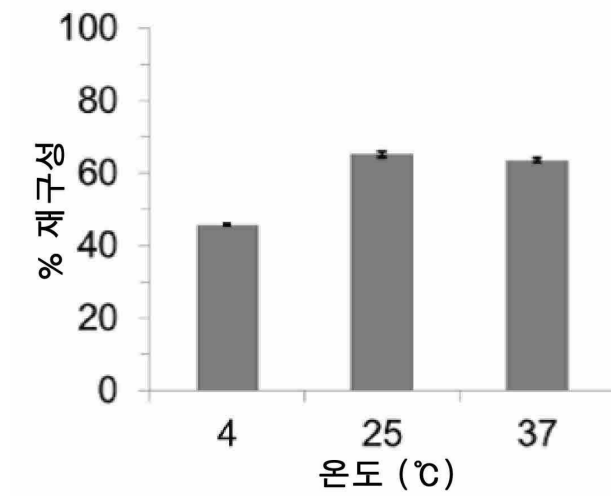


도면5

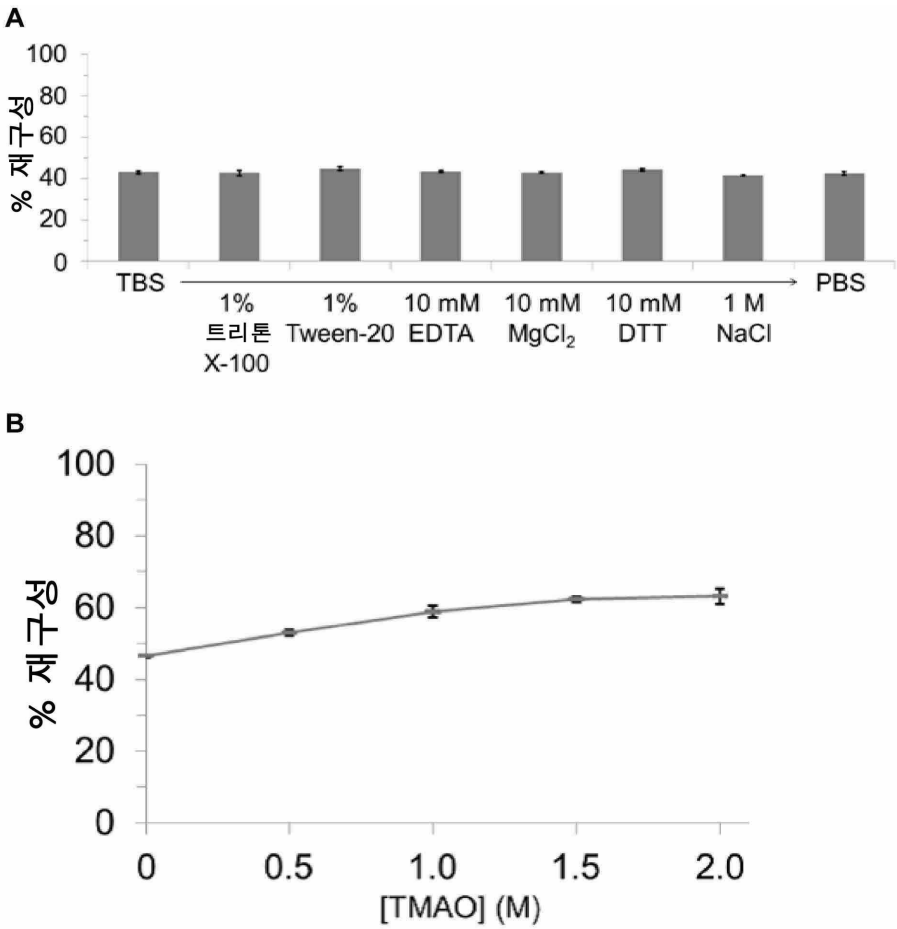
A



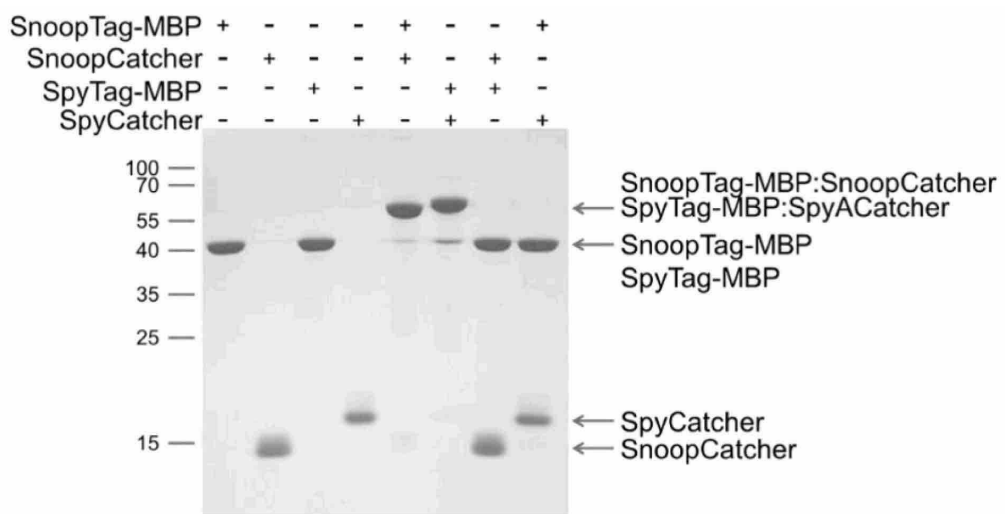
B



도면6

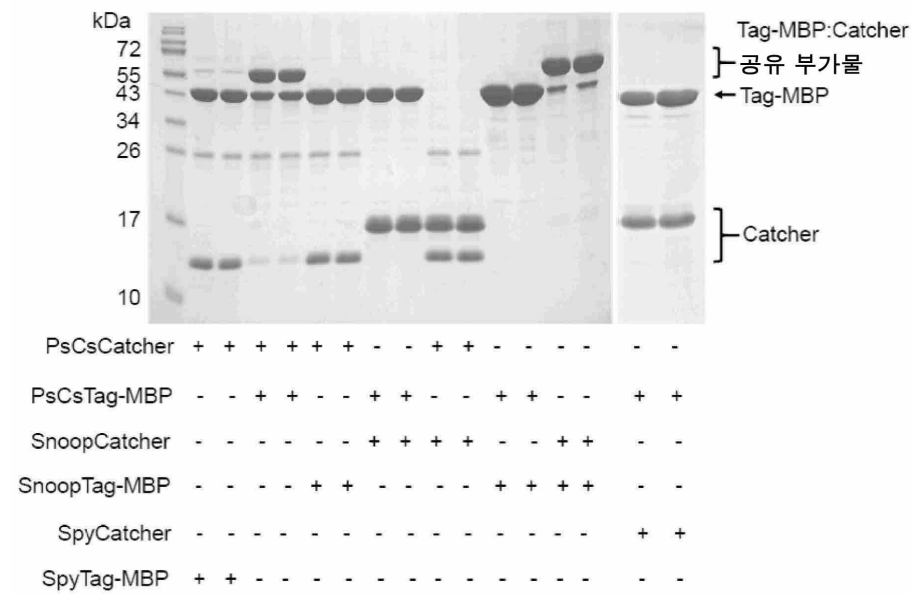


도면7

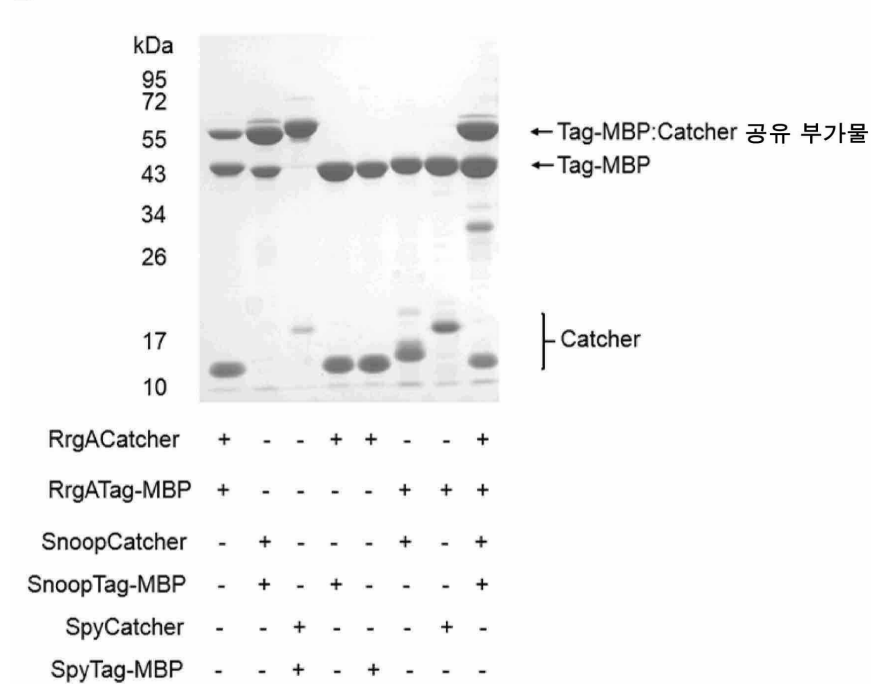


도면8

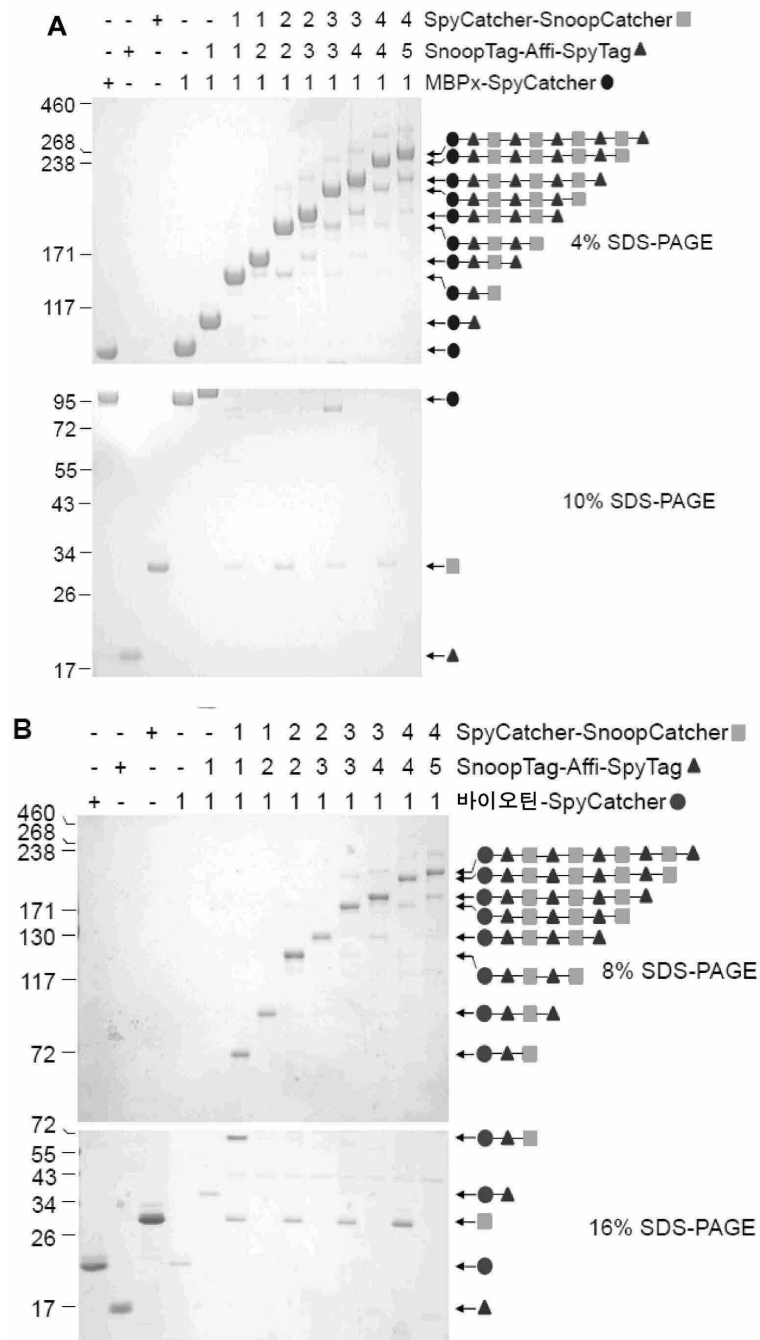
A



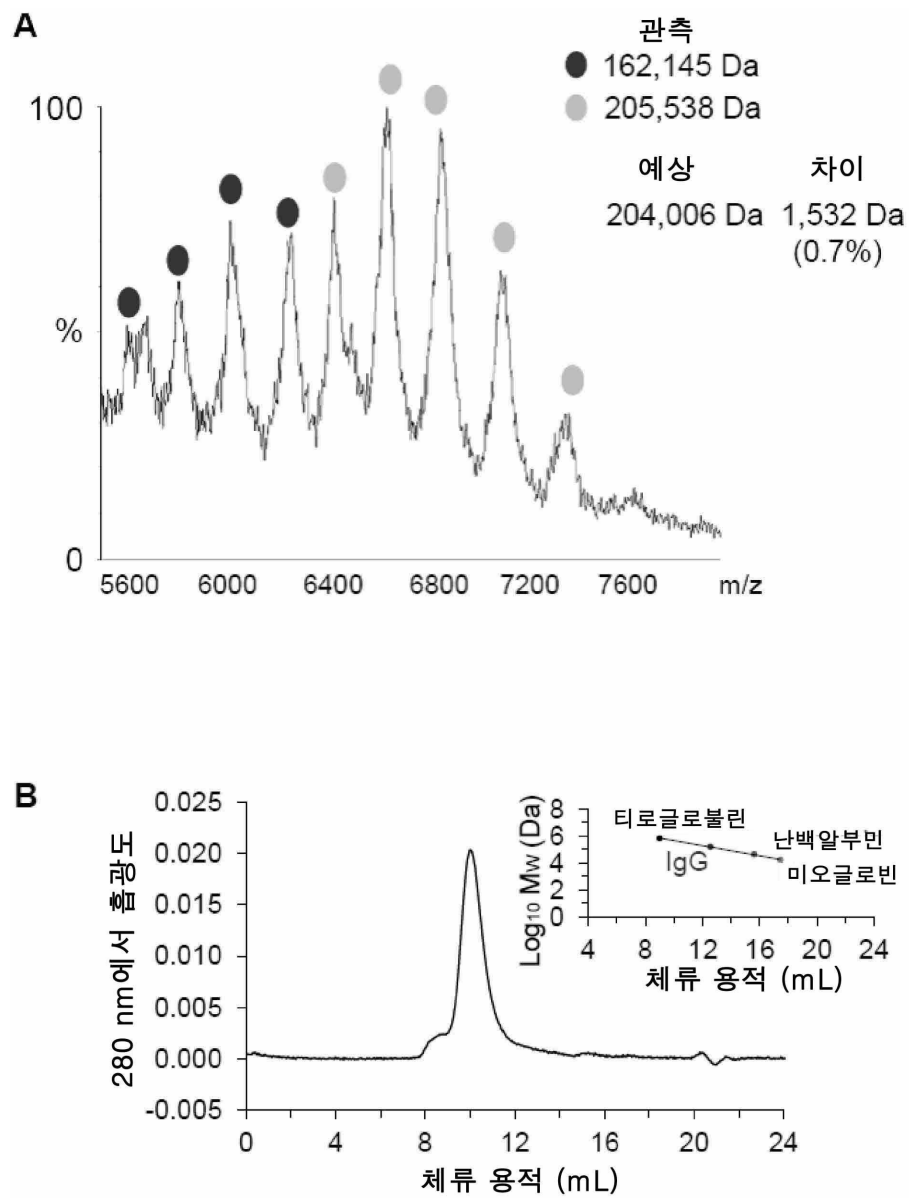
B



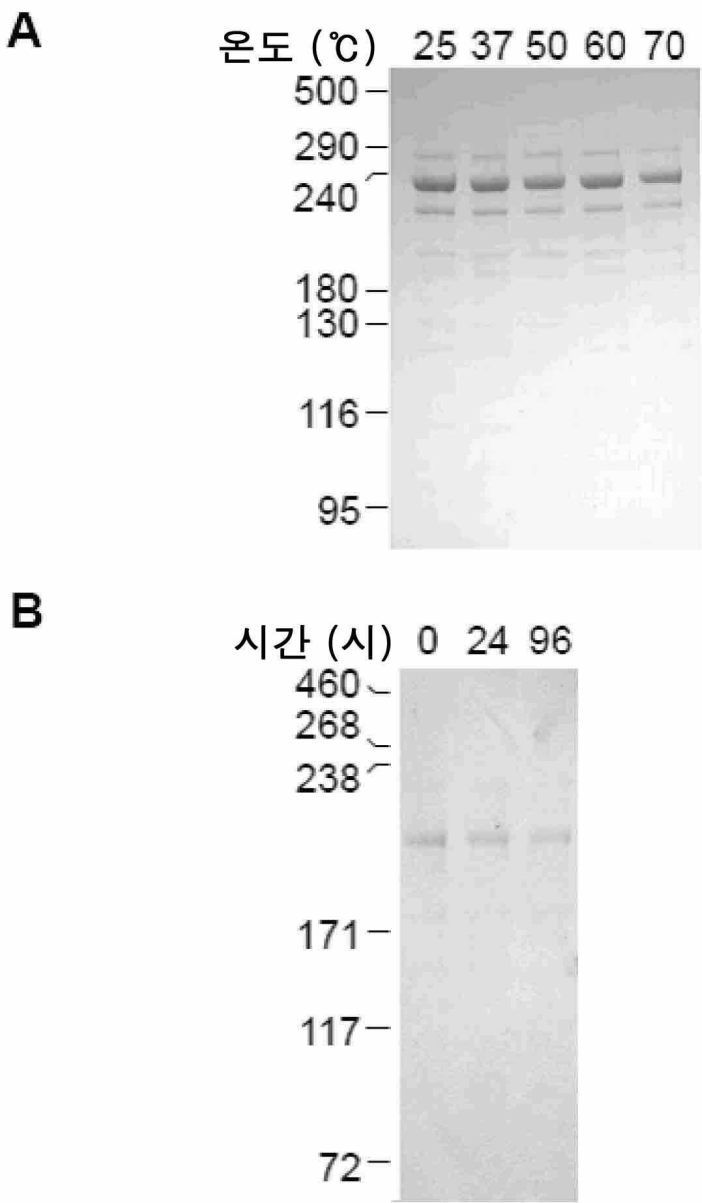
도면9



도면10

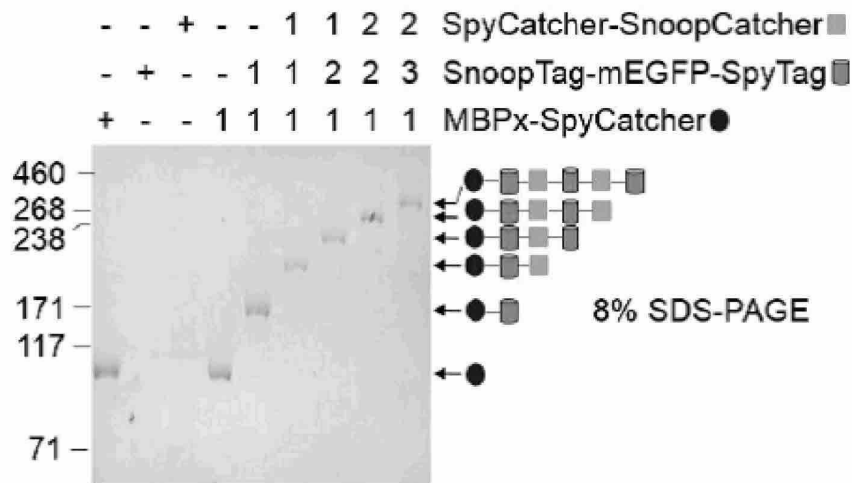


도면11

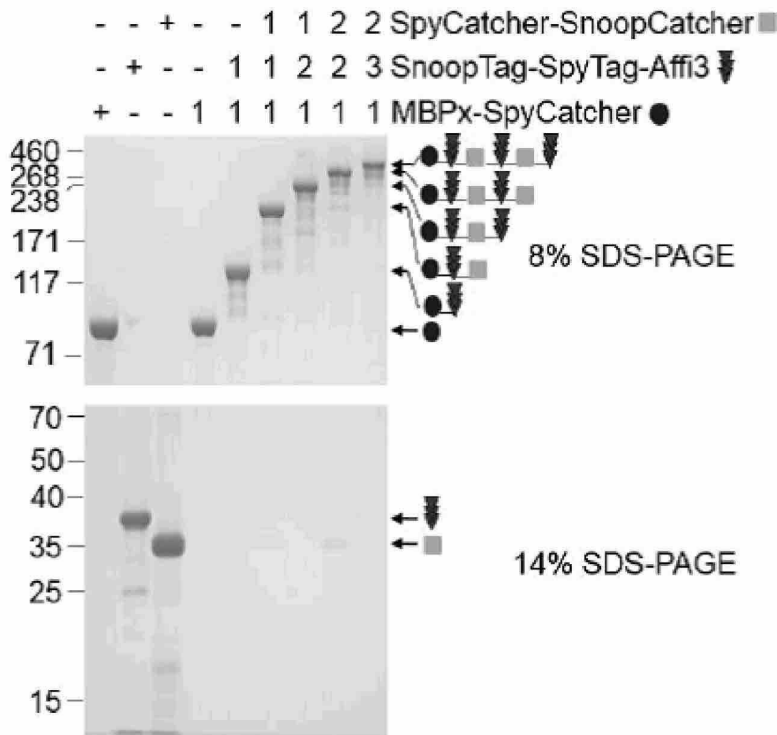


도면12

A

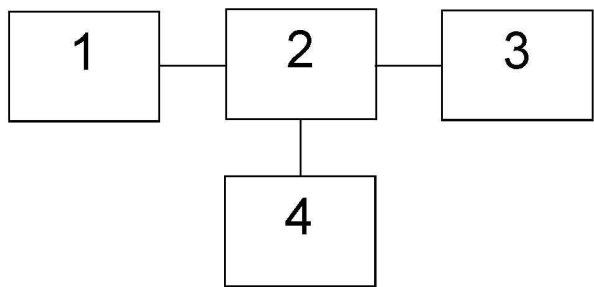


B

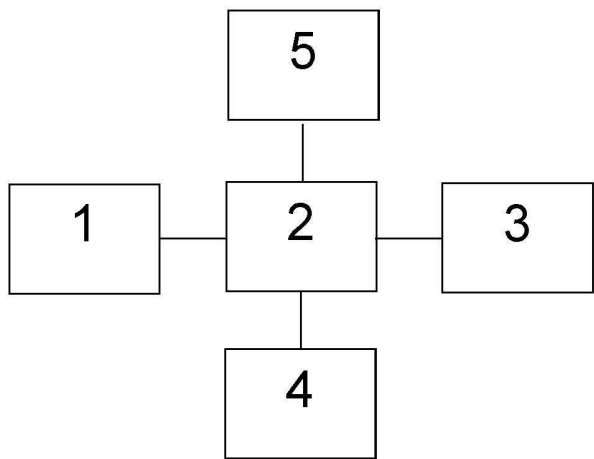


도면13

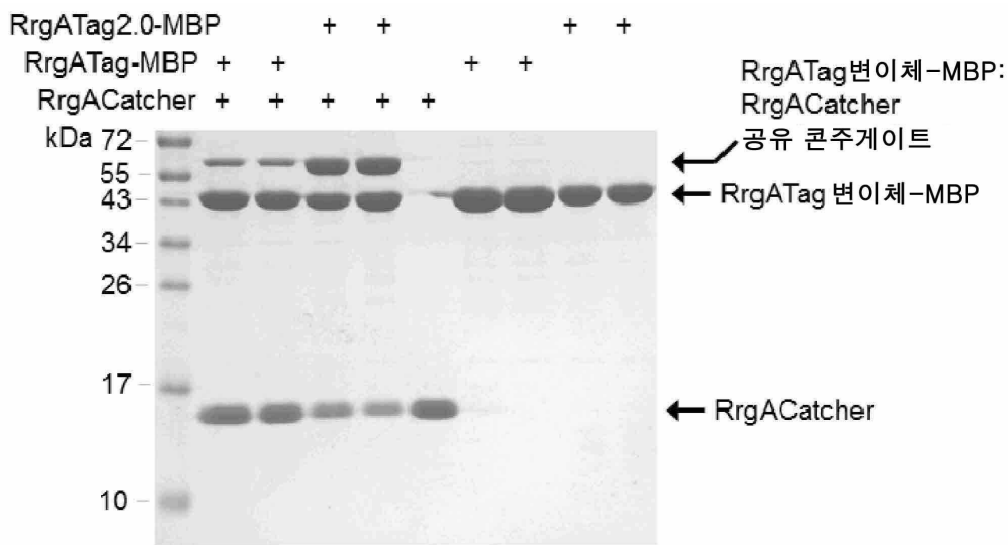
A



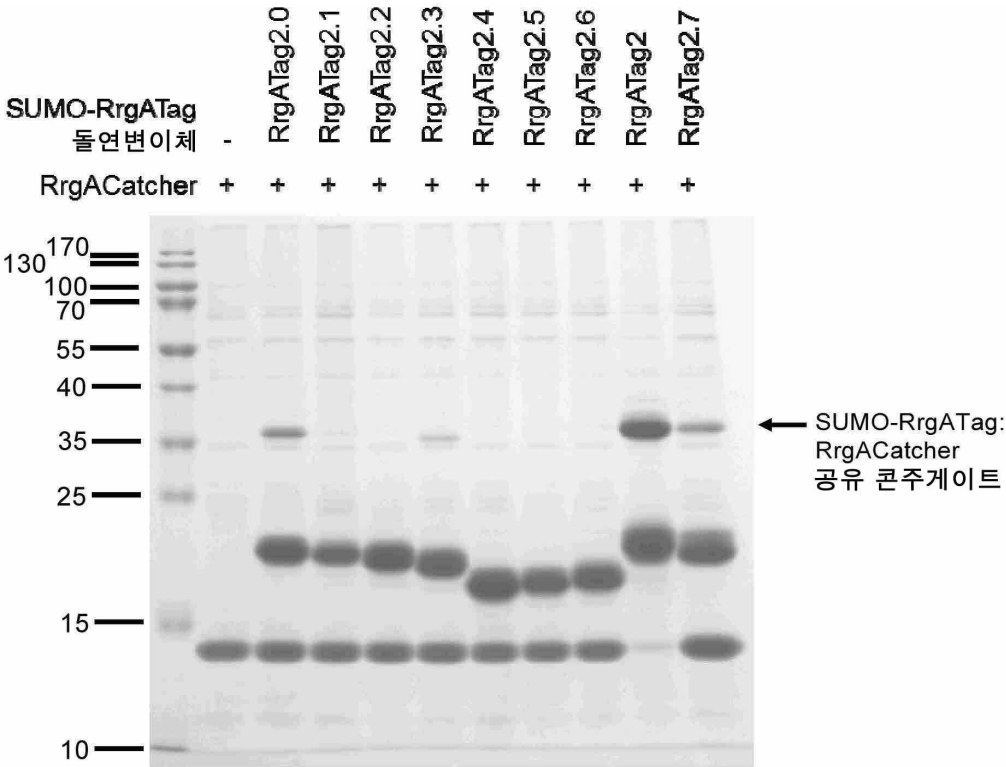
B



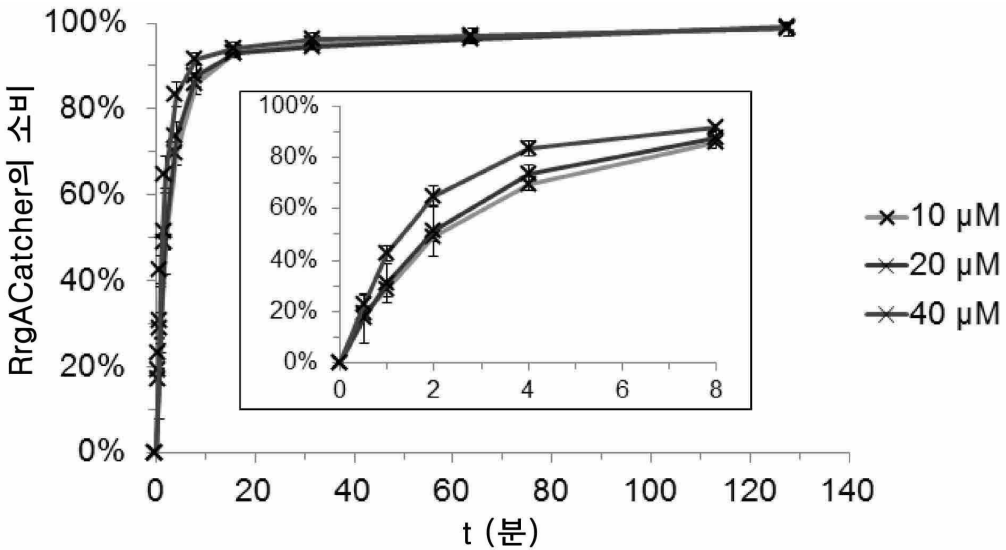
도면14



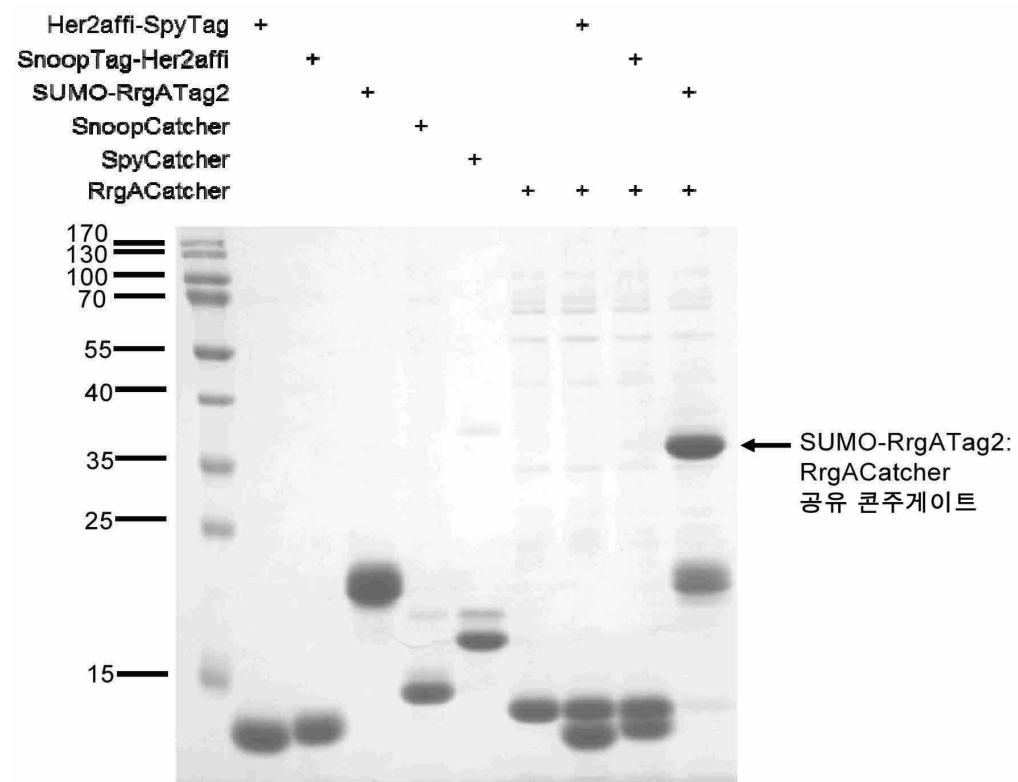
도면15



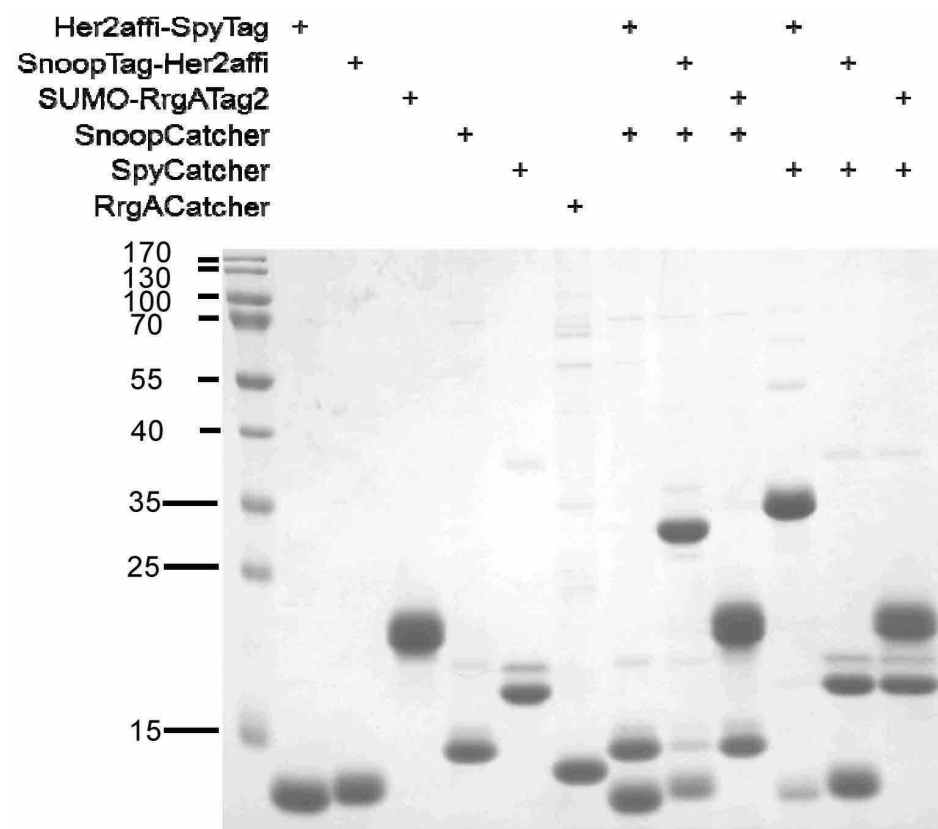
도면16



도면17



도면18



서열 목록

SEQUENCE LISTING

<110> Isis Innovation Limited

<120> Methods and products for fusion protein synthesis

<130> 20.123967/01

<150> GB1509782.7

<151> 2015-06-05

<160> 126

<170> PatentIn version 3.5

<210> 1

<211> 12

<212> PRT

<213> Artificial sequence

<220><223> SnoopTag

<400> 1

Lys Leu Gly Asp Ile Glu Phe Ile Lys Val Asn Lys

1 5 10

<210> 2

<211> 112

<212> PRT

<213> Artificial sequence

<220><223> SnoopCatcher

<400> 2

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

1 5 10 15

Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asn

20 25 30

Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Lys Asn Leu Ser Asp

35 40 45

Gly Lys Tyr Arg Leu Phe Glu Asn Ser Glu Pro Ala Gly Tyr Lys Pro

50 55 60

Val Gln Asn Lys Pro Ile Val Ala Phe Gln Ile Val Asn Gly Glu Val

65 70 75 80

Arg Asp Val Thr Ser Ile Val Pro Gln Asp Ile Pro Ala Thr Tyr Glu
85 90 95

Phe Thr Asn Gly Lys His Tyr Ile Thr Asn Glu Pro Ile Pro Pro Lys
100 105 110

<210> 3

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> SnoopTag

<400> 3

aaactgggcg atattgaatt tattaagtg aacaaa 36

<210> 4

<211> 336

<212> DNA

<213> Artificial sequence

<220><223> SnoopCatcher

<400> 4

aagccgctgc gtggtgccgt gtttagcctg cagaaacagc atcccgacta tcccgatatac 60

tatggcgcca ttgatcagaa tgggacctat caaaatgtgc gtaccggcga agatggtaaa 120

ctgaccttta agaattctgag cgatggcaaa tatcgctgt ttgaaaatag cgaacccgct 180

ggctataaac cggatgcagaa taagccgatt gtggcggttc agattgtgaa tggcgaagtg 240

cgtgatgtga ccagcattgt gccgcaggat attccggcta catatgaatt taccaacggt 300

aaacattata tcaccaatga accgataccg ccgaaa 336

<210> 5

<211> 13

<212> PRT

<213> Artificial sequence

<220><223> PsCsTag

<400> 5

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

1 5 10

<210> 6

<211> 91

<212> PRT

<213> Artificial sequence

<220><223> PsCsCatcher

<400> 6

Glu Gln Asp Val Val Phe Ser Lys Val Asn Val Ala Gly Glu Glu Ile

1 5 10 15

Ala Gly Ala Lys Ile Gln Leu Lys Asp Ala Gln Gly Gln Val Val His

20 25 30

Ser Trp Thr Ser Lys Ala Gly Gln Ser Glu Thr Val Lys Leu Lys Ala

35 40 45

Gly Thr Tyr Thr Phe His Glu Ala Ser Ala Pro Thr Gly Tyr Leu Ala

50 55 60

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

65 70 75 80

Lys Asp Ala Asn Gly Asn Gly Val Lys Ala Asp

85 90

<210> 7

<211> 39

<212> DNA

<213> Artificial sequence

<220><223> PsCsTag

<400> 7

ggcaacaaac tgaccgtgac cgatcaggcg gcgccgagc 39

<210> 8

<211> 273

<212> DNA

<213> Artificial sequence

<220><223> PsCsCatcher

<400> 8

gaacaggatg tgggtgttag caaagtgaat gtggctggcg aggaaattgc gggagcgaaa 60

attcagttga aagacgcgca gggccagggtg gtgcatagct ggaccagcaa agcgggccaa 120

agcgaaaccg tgaagctgaa agccggcacc tatacctttc atgaggcgag cgcaccgacc 180

ggctatctgg cggtagaccga tattaccttt gaagtggatg tgcagggcaa agttacagtg 240

aaagatgcga atggcaatgg tgtgaaagcg gat 273

<210> 9

<211> 19

<212> PRT

<213> Artificial sequence

<220><223> RrgATag

<400> 9

Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Asp Lys His Tyr Ile Thr

1 5 10 15

Asn Glu Pro

<210> 10

<211> 104

<212> PRT

<213> Artificial sequence

<220><223> RrgACatcher

<400> 10

Lys Leu Gly Asp Ile Glu Phe Ile Lys Val Asn Lys Asn Asp Lys Lys

1 5 10 15

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

20 25 30

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

35 40 45

Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Lys Asn Leu Ser Asp Gly

50 55 60

Lys Tyr Arg Leu Phe Glu Asn Ser Glu Pro Ala Gly Tyr Lys Pro Val

65 70 75 80

Gln Asn Lys Pro Ile Val Ala Phe Gln Ile Val Asn Gly Glu Val Arg

85 90 95

Asp Val Thr Ser Ile Val Pro Gln

100

<210> 11
 <211> 57
 <212> DNA
 <213> Artificial sequence
 <220><223> RrgATag
 <400> 11
 gatattccgg ctacatatga atttaccac gataaacatt atatcaccaa tgaaccg 57
 <210> 12
 <211> 312
 <212> DNA
 <213> Artificial sequence
 <220><223> RrgACatcher
 <400> 12
 aaactgggcg atattgaatt tattaaagtg aacaaaaacg ataaaaagcc gctgcgtggt 60
 gccgtgttta gcctgcagaa acagcatccc gactatcccg atatctatgg cgcgattgat 120
 cagaatggga cctatcaaaa tgtgcgtacc ggccaagatg gtaaaactgac ctttaagaat 180
 ctgagcgatg gcaaatatcg cctgtttgaa aatagcgaac ccgctggcta taaaccggtg 240
 cagaataagc cgattgtggc gtttcagatt gtgaatggcg aagtgcgtga tgtgaccagc 300
 attgtgccgc ag 312
 <210> 13
 <211> 13
 <212> PRT
 <213> Artificial sequence
 <220><223> SpyTag
 <400> 13
 Ala His Ile Val Met Val Asp Ala Tyr Lys Pro Thr Lys
 1 5 10
 <210> 14
 <211> 83
 <212> PRT
 <213> Artificial sequence
 <220><223> SpyCatcher
 <400> 14

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

Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala Pro Asp Gly Tyr Glu
 50 55 60
 Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu Gln Gly Gln Val Thr
 65 70 75 80
 Val Asn Gly

<210> 15

<211> 39

<212> DNA

<213> Artificial sequence

<220><223> SpyTag

<400> 15

gcccacatcg tgatggtgga cgcctacaag cgcacgaag 39

<210> 16

<211> 249

<212> DNA

<213>

> Artificial sequence

<220><223> SpyCatcher

<400> 16

gatagtgcta cccatattaa attctcaaaa cgtgatgagg acggcaaaga gttagctggt 60

gcaactatgg agttgctgta ttcactctggt aaaactatta gtacatggat ttcagatgga 120

caagtgaag atttctacct gtatccagga aaatatacat ttgtcgaaac cgcagcacca 180

gacggttatg aggtagcaac tgctattacc ttacagtta atgagcaagg tcaggttact 240

gtaaatggc 249

<210> 17

<211> 16

<212> PRT

<213> Artificial sequence

<220><223> Isopeptag

<400> 17

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

<210> 18

<211> 282

<212> PRT

<213> Artificial sequence

<220><223> Pilin-C

<400> 18

Ala Thr Thr Val His Gly Glu Thr Val Val Asn Gly Ala Lys Leu Thr
1 5 10 15
Val Thr Lys Asn Leu Asp Leu Val Asn Ser Asn Ala Leu Ile Pro Asn
20 25 30

Thr Asp Phe Thr Phe Lys Ile Glu Pro Asp Thr Thr Val Asn Glu Asp
35 40 45
Gly Asn Lys Phe Lys Gly Val Ala Leu Asn Thr Pro Met Thr Lys Val
50 55 60
Thr Tyr Thr Asn Ser Asp Lys Gly Gly Ser Asn Thr Lys Thr Ala Glu
65 70 75 80
Phe Asp Phe Ser Glu Val Thr Phe Glu Lys Pro Gly Val Tyr Tyr Tyr
85 90 95

Lys Val Thr Glu Glu Lys Ile Asp Lys Val Pro Gly Val Ser Tyr Asp
100 105 110
Thr Thr Ser Tyr Thr Val Gln Val His Val Leu Trp Asn Glu Glu Gln
115 120 125
Gln Lys Pro Val Ala Thr Tyr Ile Val Gly Tyr Lys Glu Gly Ser Lys
130 135 140
Val Pro Ile Gln Phe Lys Asn Ser Leu Asp Ser Thr Thr Leu Thr Val
145 150 155 160

Lys Lys Lys Val Ser Gly Thr Gly Gly Asp Arg Ser Lys Asp Phe Asn
165 170 175
Phe Gly Leu Thr Leu Lys Ala Asn Gln Tyr Tyr Lys Ala Ser Glu Lys
180 185 190
Val Met Ile Glu Lys Thr Thr Lys Gly Gly Gln Ala Pro Val Gln Thr
195 200 205
Glu Ala Ser Ile Asp Gln Leu Tyr His Phe Thr Leu Lys Asp Gly Glu
210 215 220

Ser Ile Lys Val Thr Asn Leu Pro Val Gly Val Asp Tyr Val Val Thr
225 230 235 240
Glu Asp Asp Tyr Lys Ser Glu Lys Tyr Thr Thr Asn Val Glu Val Ser
245 250 255
Pro Gln Asp Gly Ala Val Lys Asn Ile Ala Gly Asn Ser Thr Glu Gln
260 265 270
Glu Thr Ser Thr Asp Lys Asp Met Thr Ile
275 280

<210> 19
<211> 48
<212> DNA

<213> Artificial sequence

<220><223> Isopeptag

<400> 19

accgataaag atatgacat tacctttacc aacaaaaaag atgcggaa 48

<210> 20

<211> 846

<212> DNA

<213> Artificial sequence

<220><223> Pilin-C

<400> 20

gctacaacag ttcacgggga gactgttgta aacggagcca aactaacagt tacaaaaaac 60

cttgatttag ttaatagcaa tgcattaatt ccaaatacag attttacatt taaaatcgaa 120

cctgatacta ctgtcaacga agacggaaat aagtttaaag gtgtagcttt gaacacaccg 180

atgactaaag tcacttacac caattcagat aaaggtggat caaatacgaa aactgcagaa 240

tttgattttt cagaagttac ttttgaaaaa ccaggtgttt attattacaa agtaactgag 300

gagaagatag ataaagttcc tgggtgtttct tatgatacaa catcttacac tgttcaagtt 360

catgtcttgt ggaatgaaga gcaacaaaaa ccagtagcta cttatatgtt tggttataaa 420

gaaggtagta aggtgccaat tcagttcaaa aatagcttag attctactac attaacggtg 480

aagaaaaaag tttcaggtac cgggtggagat cgtcttaaag attttaattt tgggtctgact 540

ttaaaagcaa atcagtatta taaggcgtca gaaaaagtca tgattgagaa gacaactaaa 600

ggtggtcaag ctctgttca aacagaggct agtatagatc aactctatca ttttaccttg 660

aaagatggtg aatcaatcaa agtcacaaat cttccagtag gtgtggatta tgttgcact 720

gaagacgatt acaaatcaga aaaatataca accaacgtgg aagttagtcc tcaagatgga 780

gctgtaaaaa atatcgagg taattcaact gaacaagaga catctactga taaagatatg 840

accatt 846

<210> 21

<211> 893

<212> PRT

<213> Streptococcus pneumoniae

<400> 21

Met Leu Asn Arg Glu Thr His Met Lys Lys Val Arg Lys Ile Phe Gln

1 5 10 15

Lys Ala Val Ala Gly Leu Cys Cys Ile Ser Gln Leu Thr Ala Phe Ser

20 25 30

Ser Ile Val Ala Leu Ala Glu Thr Pro Glu Thr Ser Pro Ala Ile Gly

35 40 45

Lys Val Val Ile Lys Glu Thr Gly Glu Gly Gly Ala Leu Leu Gly Asp

50 55 60

Ala Val Phe Glu Leu Lys Asn Asn Thr Asp Gly Thr Thr Val Ser Gln

65 70 75 80

Arg Thr Glu Ala Gln Thr Gly Glu Ala Ile Phe Ser Asn Ile Lys Pro

85 90 95

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

100 105 110

Ser Thr Lys Gln Trp Thr Val Glu Val Glu Lys Asn Gly Arg Thr Thr
 115 120 125
 Val Gln Gly Glu Gln Val Glu Asn Arg Glu Glu Ala Leu Ser Asp Gln
 130 135 140

 Tyr Pro Gln Thr Gly Thr Tyr Pro Asp Val Gln Thr Pro Tyr Gln Ile
 145 150 155 160
 Ile Lys Val Asp Gly Ser Glu Lys Asn Gly Gln His Lys Ala Leu Asn
 165 170 175
 Pro Asn Pro Tyr Glu Arg Val Ile Pro Glu Gly Thr Leu Ser Lys Arg
 180 185 190
 Ile Tyr Gln Val Asn Asn Leu Asp Asp Asn Gln Tyr Gly Ile Glu Leu
 195 200 205

 Thr Val Ser Gly Lys Thr Val Tyr Glu Gln Lys Asp Lys Ser Val Pro
 210 215 220
 Leu Asp Val Val Ile Leu Leu Asp Asn Ser Asn Ser Met Ser Asn Ile
 225 230 235 240
 Arg Asn Lys Asn Ala Arg Arg Ala Glu Arg Ala Gly Glu Ala Thr Arg
 245 250 255
 Ser Leu Ile Asp Lys Ile Thr Ser Asp Ser Glu Asn Arg Val Ala Leu
 260 265 270

 Val Thr Tyr Ala Ser Thr Ile Phe Asp Gly Thr Glu Phe Thr Val Glu
 275 280 285
 Lys Gly Val Ala Asp Lys Asn Gly Lys Arg Leu Asn Asp Ser Leu Phe
 290 295 300
 Trp Asn Tyr Asp Gln Thr Ser Phe Thr Thr Asn Thr Lys Asp Tyr Ser
 305 310 315 320
 Tyr Leu Lys Leu Thr Asn Asp Lys Asn Asp Ile Val Glu Leu Lys Asn
 325 330 335

 Lys Val Pro Thr Glu Ala Glu Asp His Asp Gly Asn Arg Leu Met Tyr
 340 345 350
 Gln Phe Gly Ala Thr Phe Thr Gln Lys Ala Leu Met Lys Ala Asp Glu

355	360	365	
Ile Leu Thr Gln Gln Ala Arg Gln Asn Ser Gln Lys Val Ile Phe His			
370	375	380	
Ile Thr Asp Gly Val Pro Thr Met Ser Tyr Pro Ile Asn Phe Asn His			
385	390	395	400
Ala Thr Phe Ala Pro Ser Tyr Gln Asn Gln Leu Asn Ala Phe Phe Ser			
405	410	415	
Lys Ser Pro Asn Lys Asp Gly Ile Leu Leu Ser Asp Phe Ile Thr Gln			
420	425	430	
Ala Thr Ser Gly Glu His Thr Ile Val Arg Gly Asp Gly Gln Ser Tyr			
435	440	445	
Gln Met Phe Thr Asp Lys Thr Val Tyr Glu Lys Gly Ala Pro Ala Ala			
450	455	460	
Phe Pro Val Lys Pro Glu Lys Tyr Ser Glu Met Lys Ala Ala Gly Tyr			
465	470	475	480
Ala Val Ile Gly Asp Pro Ile Asn Gly Gly Tyr Ile Trp Leu Asn Trp			
485	490	495	
Arg Glu Ser Ile Leu Ala Tyr Pro Phe Asn Ser Asn Thr Ala Lys Ile			
500	505	510	
Thr Asn His Gly Asp Pro Thr Arg Trp Tyr Tyr Asn Gly Asn Ile Ala			
515	520	525	
Pro Asp Gly Tyr Asp Val Phe Thr Val Gly Ile Gly Ile Asn Gly Asp			
530	535	540	
Pro Gly Thr Asp Glu Ala Thr Ala Thr Ser Phe Met Gln Ser Ile Ser			
545	550	555	560
Ser Lys Pro Glu Asn Tyr Thr Asn Val Thr Asp Thr Thr Lys Ile Leu			
565	570	575	
Glu Gln Leu Asn Arg Tyr Phe His Thr Ile Val Thr Glu Lys Lys Ser			
580	585	590	
Ile Glu Asn Gly Thr Ile Thr Asp Pro Met Gly Glu Leu Ile Asp Leu			
595	600	605	

Gln Leu Gly Thr Asp Gly Arg Phe Asp Pro Ala Asp Tyr Thr Leu Thr
610 615 620

Ala Asn Asp Gly Ser Arg Leu Glu Asn Gly Gln Ala Val Gly Gly Pro
625 630 635 640

Gln Asn Asp Gly Gly Leu Leu Lys Asn Ala Lys Val Leu Tyr Asp Thr
645 650 655

Thr Glu Lys Arg Ile Arg Val Thr Gly Leu Tyr Leu Gly Thr Asp Glu
660 665 670

Lys Val Thr Leu Thr Tyr Asn Val Arg Leu Asn Asp Glu Phe Val Ser
675 680 685

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

Val Glu Gln Asn Thr Val Arg Asp Phe Pro Ile Pro Lys Ile Arg Asp
705 710 715 720

Val Arg Lys Tyr Pro Glu Ile Thr Ile Ser Lys Glu Lys Lys Leu Gly
725 730 735

Asp Ile Glu Phe Ile Lys Val Asn Lys Asn Asp Lys Lys Pro Leu Arg
740 745 750

Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr Pro Asp Ile
755 760 765

Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asn Val Arg Thr Gly
770 775 780

Glu Asp Gly Lys Leu Thr Phe Lys Asn Leu Ser Asp Gly Lys Tyr Arg
785 790 795 800

Leu Phe Glu Asn Ser Glu Pro Ala Gly Tyr Lys Pro Val Gln Asn Lys
805 810 815

Pro Ile Val Ala Phe Gln Ile Val Asn Gly Glu Val Arg Asp Val Thr
820 825 830

Ser Ile Val Pro Gln Asp Ile Pro Ala Gly Tyr Glu Phe Thr Asn Asp
835 840 845

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

850	855	860	
Arg Thr Gly Gly Ile Gly Met Leu Pro Phe Tyr Leu Ile Gly Cys Met			
865	870	875	880
Met Met Gly Gly Val Leu Leu Tyr Thr Arg Lys His Pro			
	885	890	

<210> 22

<211> 2682

<212> DNA

<213> Streptococcus pneumoniae

<400> 22

atgctgaacc gcgaaaccca tatgaaaaaa gtaagaaaga tatttcagaa ggcagttgca	60
ggactgtgct gtatatctca gttgacagct ttttcttcga tagttgcttt agcagaaacg	120
cctgaaacca gtccagcgat aggaaaagta gtgattaagg agacaggcga aggaggagcg	180
cttctaggag atgccgtctt tgagttgaaa aacaatacgg atggcacaac tgtttcgcaa	240
aggacagagg cgcaaacagg agaagcgata ttttcaaaca taaaacctgg gacatacacc	300
ttgacagaag cccaacctcc agttggttat aaacctcta ctaaacaatg gactgttgaa	360
gttgagaaga atggtcggac gactgtccaa ggtgaacagg tagaaaatcg agaagaggct	420
ctatctgacc agtatccaca aacagggact tatccagatg ttcaaacacc ttatcagatt	480
attaaggtag atggttcgga aaaaaacgga cagcacaagg cgttgaatcc gaatccatat	540
gaacgtgtga ttccagaagg tacactttca aagagaattt atcaagtga taatttggat	600
gataaccaat atggaatcga attgacggtt agtgggaaaa cagtgtatga acaaaaagat	660
aagtctgtgc cgtcggatgt cgttatcttg ctcgataact caaatagtat gagtaacatt	720
cgaacaaga atgctcgacg tgcggaaaga gctggtgagg cgacacgttc tcttattgat	780
aaaattacat ctgattcaga aaatagggtg gcgcttgtga cttatgcttc cactatcttt	840
gatgggaccg agtttacagt agaaaaaggg gtagcagata aaaacggaaa gcgattgaat	900
gattctcttt ttiggaatta tgatcagacg agttttacaa ccaataccaa agattatagt	960
tatttaaagc tgactaatga taagaatgac attgtagaat taaaaataa ggtacctacc	1020
gaggcagaag accatgatgg aaatagattg atgtaccaat tcggtgccac ttttactcag	1080
aaagctttga tgaaggcaga tgagattttg acacaacaag cgagacaaaa tagtcaaaaa	1140
gtcatcttcc atattacgga tgggtgccca actatgtcgt atccgattaa ttttaatcat	1200
gctacgtttg ctccatcata tcaaaatcaa ctaaatgcat ttttagtaa atctcctaat	1260

aaagatggaa tactattaag tgattttatt acgcaagcaa ctagtggaga acatacaatt 1320

gtacgcggag atgggcaaag ttaccagatg tttacagata agacagttta tgaaaaaggt 1380

gctcctgcag ctttccaggt taaacctgaa aaatattctg aaatgaaggc ggctggttat 1440

gcagttatag gcgatccaat taatggtgga tatatttggc ttaattggag agagagtatt 1500

ctggcttate cgtttaattc taatactgct aaaattacca atcatggtga ccctacaaga 1560

tggtactata acgggaatat tgctcctgat gggtatgatg tctttacggt aggtattggt 1620

attaacggag atcctggtag ggatgaagca acggctacta gttttatgca aagtatttct 1680

agtaaacctg aaaactatac caatgttact gacacgacaa aaatattgga acagttegaat 1740

cgttattttcc acaccatcgt aactgaaaag aaatcaattg agaatggtac gattacagat 1800

ccgatgggtg agttaattga tttgcaattg ggcacagatg gaagatttga tccagcagat 1860

tacactttaa ctgcaaacga tggtagtcgc ttggagaatg gacaagctgt aggtggtcca 1920

caaatgatg gtggtttgtt aaaaaatgca aaagtgtctt atgatacgac tgagaaaagg 1980

attcgtgtaa caggtctgta ccttggaaac gatgaaaaag ttacgttgac ctacaatgtt 2040

cgtttgaatg atgagtttgt aagcaataaa ttttatgata ccaatggtcg aacaacctta 2100

catcctaagg aagtagaaca gaacacagtg cgcgacttcc cgattcctaa gattcgtgat 2160

gtgcggaagt atccagaaat cacaatttca aaagagaaaa aacttggtga cattgagttt 2220

attaaggtca ataaaaatga taaaaaacca ctgagagggtg cggctcttag tcttcaaaaa 2280

caacatccgg attatccaga tatttatgga gctattgatc aaaatggcac ttatcaaaat 2340

gtgagaacag gtgaagatgg taagttgacc tttaaaaatc tgtcagatgg gaaatatcga 2400

ttatttgaaa attctgaacc agctgggttat aaaccgttc aaaataagcc tatcgttgcc 2460

ttccaaatag taaatggaga agtcagagat gtgacttcaa tcgttcaca agatatacca 2520

gcgggttacg agtttacgaa tgataagcac tatattacca atgaacctat tcctccaaag 2580

agagaatate ctgaaactgg tggatatcga atgttgccat tctatctgat aggttgcatg 2640

atgatgggag gagttctatt atacacacgg aaacatccgt aa 2682

<210> 23

<211> 340

<212> PRT

<213> Streptococcus pyogenes

<400> 23

Met Lys Leu Arg His Leu Leu Leu Thr Gly Ala Ala Leu Thr Ser Phe

1 5 10 15

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

260	265	270	
Ser Pro Gln Asp Gly Ala Val Lys Asn Ile Ala Gly Asn Ser Thr Glu			
275	280	285	
Gln Glu Thr Ser Thr Asp Lys Asp Met Thr Ile Thr Phe Thr Asn Lys			
290	295	300	
Lys Asp Phe Glu Val Pro Thr Gly Val Ala Met Thr Val Ala Pro Tyr			
305	310	315	320
Ile Ala Leu Gly Ile Val Ala Val Gly Gly Ala Leu Tyr Phe Val Lys			
325	330	335	
Lys Lys Asn Ala			
340			
<210>	24		
<211>	1023		
<212>	DNA		
<213>	Streptococcus pyogenes		
<400>	24		
atgaaattac gtcacttact attaacggga gcagccctaa ctagttttgc tgctacaaca		60	
gttcacgggg agactgttgt aaacggagcc aaactaacag ttacaaaaa ccttgattta		120	
gttaatagca atgcattaat tccaaataga gatattacat ttaaaatcga acctgatact		180	
actgtcaacg aagacggaaa taagtittaaa ggtgtagctt tgaacacacc gatgactaaa		240	
gtcacttaca ccaattcaga taaaggtgga tcaaataga aaactgcaga atttgatttt		300	
tcagaagtta cttttgaaaa accaggtgtt tattattaca aagtaactga ggagaagata		360	
gataaagttc ctggtgtttc ttatgataca acatcttaca ctgttcaagt tcatgtcttg		420	
tggaatgaag agcaacaaaa accagtagct acttatattg ttggttataa agaaggtagt		480	
aaggtgccaa ttcagttcaa aaatagctta gattctacta cattaacggt gaagaaaaaa		540	
gtttcaggta ccggtggaga tcgctctaaa gattttaatt ttggtctgac tttaaaagca		600	
aatcagtatt ataaggcgtc agaaaaagtc atgattgaga agacaactaa aggtggtcaa		660	
gtctctgttc aaacagaggc tagtatagat caactctatc attttacctt gaaagatggt		720	
gaatcaatca aagtcacaaa tcttcagta ggtgtggatt atgtgtgcac tgaagacgat		780	
tacaaatcag aaaaatatat aaccaactg gaagttagtc ctcaagatgg agctgtaaaa		840	
aatatcgag gtaattcaac tgaacaagag acatctactg ataaagatat gaccattact		900	

tttacaata aaaaagactt tgaagtgccacacaggagtag caatgactgt ggcacatat 960
attgcttttag gaattgtagc agttgggtgga gctctttact ttgttaaaaa gaaaaatgct 1020
taa 1023

<210> 25

<211> 102

<212> PRT

<213> Streptococcus pyogenes

<400> 25

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

1 5 10 15

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

20 25 30

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly Gln Val Lys

35 40 45

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

50 55 60

Pro Asp Gly Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu

65 70 75 80

Gln Gly Gln Val Thr Val Asn Gly Lys Ala Thr Lys Gly Asp Ala His

85 90 95

Ile Val Met Val Asp Ala

100

<210> 26

<211> 309

<212> DNA

<213> Streptococcus pyogenes

<400> 26

atgacaattg aagaagatag tgctacccat attaaattct caaaactga tattgacggc 60

aaagagtttag ctggtgcaac tatggagttg cgtgattcat ctggtaaaac tattagtaca 120

tggatttcag atggacaagt gaaagatttc tacctgatgc caggaaaata tacatttgtc 180

gaaaccgcag caccagacgg ttatgaggta gcaactgcta ttacctttac agttaatgag 240

caaggtcagg ttactgtaaa tggcaaagca actaaagtg acgctcatat tgtcatggtt 300

gatgcttga

309

<210> 27

<211> 674

<212> PRT

<213> Enterococcus faecalis

<400> 27

Met Thr Lys Ser Val Lys Phe Leu Val Leu Leu Leu Val Met Ile Leu

1 5 10 15

Pro Ile Ala Gly Ala Leu Leu Ile Gly Pro Ile Ser Phe Gly Ala Glu

20 25 30

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

35 40 45

Leu Tyr Gln Gly Glu Met Thr Ser Ile Lys Val Ser Phe Ser Asp Lys

50 55 60

Glu Asn Gln Lys Ile Lys Pro Gly Asp Thr Ile Thr Leu Thr Leu Pro

65 70 75 80

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

85 90 95

Asn Leu Asn Gly Leu Gly Glu Val Phe Ile Tyr Lys Asp His Val Val

100 105 110

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

115 120 125

Phe Ser Phe Gly Ile Lys Thr Leu Ile Thr Asn Ser Ser Gln Pro Asn

130 135 140

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

145 150 155 160

Ile Glu Gly Val Thr Asn Thr Glu Thr Gly Gln Ile Glu Arg Asp Tyr

165 170 175

Pro Phe Phe Tyr Lys Val Gly Asp Leu Ala Gly Glu Ser Asn Gln Val

180 185 190

Arg Trp Phe Leu Asn Val Asn Leu Asn Lys Ser Asp Val Thr Glu Asp

195 200 205

Ile Ser Ile Ala Asp Arg Gln Gly Ser Gly Gln Gln Leu Asn Lys Glu
210 215 220

Ser Phe Thr Phe Asp Ile Val Asn Asp Lys Glu Thr Lys Tyr Ile Ser
225 230 235 240

Leu Ala Glu Phe Glu Gln Gln Gly Tyr Gly Lys Ile Asp Phe Val Thr
245 250 255

Asp Asn Asp Phe Asn Leu Arg Phe Tyr Arg Asp Lys Ala Arg Phe Thr
260 265 270

Ser Phe Ile Val Arg Tyr Thr Ser Thr Ile Thr Glu Ala Gly Gln His
275 280 285

Gln Ala Thr Phe Glu Asn Ser Tyr Asp Ile Asn Tyr Gln Leu Asn Asn
290 295 300

Gln Asp Ala Thr Asn Glu Lys Asn Thr Ser Gln Val Lys Asn Val Phe
305 310 315 320

Val Glu Gly Glu Ala Ser Gly Asn Gln Asn Val Glu Met Pro Thr Glu
325 330 335

Glu Ser Leu Asp Ile Pro Leu Glu Thr Ile Asp Glu Trp Glu Pro Lys
340 345 350

Thr Pro Thr Ser Glu Gln Ala Thr Glu Thr Ser Glu Lys Thr Asp Thr
355 360 365

Thr Glu Thr Ala Glu Ser Ser Gln Pro Glu Val His Val Ser Pro Thr
370 375 380

Glu Glu Glu Asn Pro Asp Glu Gly Glu Thr Leu Gly Thr Ile Glu Pro
385 390 395 400

Ile Ile Pro Glu Lys Pro Ser Val Thr Thr Glu Glu Asn Gly Thr Thr
405 410 415

Glu Thr Ala Glu Ser Ser Gln Pro Glu Val His Val Ser Pro Thr Glu
420 425 430

Glu Glu Asn Pro Asp Glu Ser Glu Thr Leu Gly Thr Ile Glu Pro Ile
435 440 445

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

450 455 460
 Thr Ala Glu Ser Ser Gln Pro Glu Val His Val Ser Pro Ala Glu Glu
 465 470 475 480
 Glu Asn Pro Asp Glu Ser Glu Thr Leu Gly Thr Ile Leu Pro Ile Leu
 485 490 495

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

 Glu Ser Ser Gln Pro Glu Val His Val Ser Pro Thr Lys Glu Ile Thr
 565 570 575
 Thr Thr Glu Lys Lys Gln Pro Ser Thr Glu Thr Thr Val Glu Lys Asn
 580 585 590
 Lys Asn Val Thr Ser Lys Asn Gln Pro Gln Ile Leu Asn Ala Pro Leu
 595 600 605
 Asn Thr Leu Lys Asn Glu Gly Ser Pro Gln Leu Ala Pro Gln Leu Leu
 610 615 620

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

<210> 28

<211> 2025

<212> DNA

<213> Enterococcus faecalis

<400> 28

atgacaaaaa gtgtaaaatt tttagtgtta ctgttggtaa tgattctacc aattgcgggg	60
gcgttattga ttggtccaat ttcgtttggc gccgaattga gcaaaagtgc aatcgttgac	120
aaagtagaat tagatcacac tactttatat caaggagaga tgacctcaat taaagtatct	180
tttagtgaca aagaaaatca gaaaataaaa cctggagata ctattacttt aactttacca	240
gacgcactag ttggaatgac cgagaacgat agttcaccac gaaaaatcaa tttaatggt	300
ttagggggaag tttttatcta taaagatcat gttgtagcaa catttaacga aaaagttgaa	360
tctttacata atgtgaatgg gcatttttct ttcgggatta aaacgcttat caccaatagt	420
tctcaaccga atgtgataga aacggatttc ggaacagcaa cggcgactca acgtttgacg	480
attgaaggag tgactaacac agagactggc caaattgagc gagactatcc gtttttttat	540
aaagtaggcg atttggctgg agagtcaaat caagtacgtt ggttttttaa tgtgaacctc	600
aataaatccg atgtcacaga agatatttca attgcggatc gacaaggaag tggccaacaa	660
ttaaataaag agagttttac atttgatatt gtgaatgaca aagaaactaa atatatttca	720
cttgccgagt ttgagcaaca aggttatggc aaaattgact tcgtaacaga taatgacttt	780
aacttacgtt tttatcgga taaagcacgc tttacttctt ttatcgtcgc ttacacttcg	840
acaatcacag aagcaggcca acatcaagca acatttgaaa atagtatatga catcaattat	900
caactaaaca atcaagacgc aacgaatgaa aaaaatacat cacaggttaa aaatgttttt	960
gtagaaggcg aggcaagcgg caatcaaaat gtggaaatgc caacagaaga aagtctagac	1020
attcctttag agacaataga tgaatgggaa ccaagacac ctacttcgga acaggcaaca	1080
gaaacaagtg aaaaagacaga cacaacagaa accgcagaaa gcagccaacc agaagttcat	1140
gtctcaccaa cagaagaaga aaatccagat gaaggtgaaa cactaggcac gattgagcca	1200
atcatacctg aaaaaccaag tgtgacaact gaagagaatg gcacgacaga aactgcagaa	1260
agcagccaac cagaagtta tgtctacca acagaagaag aaaatccaga tgaaagtgaa	1320
acactaggca cgattgagcc aatcatacct gaaaaaccaa gtgtgacaac tgaagagaac	1380
ggcacaacag aaaccgcaga aagcagccaa ccagaagttc atgtctcacc agcggagaa	1440
gaaaaaccag atgaaagtga aacgttaggt acaattttac caatcctacc tgaaaaacca	1500
agtgtgacaa ctgaagagaa tggcacaacg gaaactgcag aaagcagtca accagaagtc	1560
catgtgtcgc caacggaaga agaaaatcca gatgaaagtg aaacactagg cacgattgca	1620
ccaatcatac ctgaaaaacc aagcgtaaca actgaagaga atggtataac ggaaacggca	1680

gaaagcagcc agccagaagt tcatgtctca ccaacaaaag aaattactac aactgagaaa 1740

aaacagccat ccacagaaac aactgtggag aaaaataaaa atgttacatc aaaaaatcaa 1800

ccacaaatac taaacgtcc attaaataca ttgaaaaatg aaggaagccc acagttggt 1860

ccccaactgc ttagtgaacc aattcaaaaa ttaaatgaag caaacgggca acgagaactt 1920

cccaaacag gcacaacaaa aacaccgttt atgctaatac caggaatact ggcaagtaca 1980

tttgccgttt taggtgtaag ttatctacaa atcagaaaaga attaa 2025

<210> 29

<211> 331

<212> PRT

<213> Staphylococcus aureus

<400> 29

Met Asn Lys Asn Val Leu Lys Phe Met Val Phe Ile Met Leu Leu Asn

1 5 10 15

Ile Ile Thr Pro Leu Phe Asn Lys Asn Glu Ala Phe Ala Ala Arg Asp

20 25 30

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

35 40 45

Glu Asp Gly Gly Lys Thr Thr Val Lys Met Thr Phe Asp Asp Lys Asn

50 55 60

Gly Lys Ile Gln Asn Gly Asp Met Ile Lys Val Ala Trp Pro Thr Ser

65 70 75 80

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

85 90 95

Lys Gly Glu Gln Val Gly Gln Ala Val Ile Thr Pro Asp Gly Ala Thr

100 105 110

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

115 120 125

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

130 135 140

Asp Asp Lys Val Ala Thr Ile Thr Ser Gly Asn Lys Ser Thr Asn Val

145 150 155 160

Thr Val His Lys Ser Glu Ala Gly Thr Ser Ser Val Phe Tyr Tyr Lys
165 170 175
Thr Gly Asp Met Leu Pro Glu Asp Thr Thr His Val Arg Trp Phe Leu
180 185 190
Asn Ile Asn Asn Glu Lys Ser Tyr Val Ser Lys Asp Ile Thr Ile Lys
195 200 205
Asp Gln Ile Gln Gly Gly Gln Gln Leu Asp Leu Ser Thr Leu Asn Ile
210 215 220
Asn Val Thr Gly Thr His Ser Asn Tyr Tyr Ser Gly Gln Ser Ala Ile
225 230 235 240
Thr Asp Phe Glu Lys Ala Phe Pro Gly Ser Lys Ile Thr Val Asp Asn
245 250 255
Thr Lys Asn Thr Ile Asp Val Thr Ile Pro Gln Gly Tyr Gly Ser Tyr
260 265 270
Asn Ser Phe Ser Ile Asn Tyr Lys Thr Lys Ile Thr Asn Glu Gln Gln
275 280 285
Lys Glu Phe Val Asn Asn Ser Gln Ala Trp Tyr Gln Glu His Gly Lys
290 295 300
Glu Glu Val Asn Gly Lys Ser Phe Asn His Thr Val His Asn Ile Asn
305 310 315 320
Ala Asn Ala Gly Ile Glu Gly Thr Val Lys Gly
325 330

<210> 30

<211> 993

<212> DNA

<213> Staphylococcus aureus

<400> 30

atgaacaaaa atgtgttgaa gtttatggtc ttataatgt tattaaatat catcacacct	60
ttatttaata aaaatgaagc atttcagca cgagatattt catcaacgaa tgttacagat	120
ttaactgtat caccgtctaa gatagaagat ggtggtaaaa cgacagtaaa aatgacgttc	180
gacgataaaa atggaaaaat acaaaatggt gacatgatta aagtggcatg gccgacaagc	240
ggtacagtaa agatagaggg ttatagtaaa acagtacat taactgttaa aggtgaacag	300

gtgggtcaag cagttattac accagacggt gcaacaatta cattcaatga taaagtagaa 360

aaattaagtg atgtttcggg atttcagaa tttgaagtac aaggaagaaa tttaacgcaa 420

acaaatactt cagatgacaa agtagctacg ataacatctg ggaataaatc aacgaatgtt 480

acggttcata aaagtgaagc gggaacaagt agtgttttct attataaaac gggagatatg 540

ctaccagaag atacgacaca tgtacgatgg tttttaaata ttaacaatga aaaaagttat 600

gtatcgaaag atattactat aaaggatcag attcaagggt gacagcagtt agatttaagc 660

acattaaaca ttaatgtgac aggtacacat agcaattatt atagtgaca aagtgcatt 720

actgattttg aaaaagcctt tccaggttct aaaataactg ttgataatac gaagaacaca 780

attgatgtaa caattccaca aggctatggg tcatataata gtttttcaat taactacaaa 840

acaaaaatta cgaatgaaca gcaaaaagag ttgttaata attcacaagc ttggtatcaa 900

gagcatggta aggaagaagt gaacgggaaa tcatttaatc atactgtgca caatattaat 960

gctaatagccg gtattgaagg tactgtaaaa ggt 993

<210> 31

<211> 738

<212> PRT

<213> Streptococcus intermedius

<400> 31

Met Lys Lys Arg Arg Gly Gln Phe Phe Lys Ser Ala Ile Ser Phe Leu

1 5 10 15

Val Val Phe Leu Met Val Met Val Ser Ile Ile Tyr Pro Ser Ser Lys

20 25 30

Ile Lys Ala Asp Gly Phe Pro Asn Asp Ala Thr Gly Val Ser Pro Asn

35 40 45

Gly Lys Tyr Tyr Ser Ala Gly Arg Glu Asn Arg Leu Gly Met Val Thr

50 55 60

Ser Asp Glu Leu His Thr Ala Thr Glu Leu Phe Gly Phe Cys Met Ala

65 70 75 80

Asn Ser Lys Lys Tyr Pro Gly Tyr Asp Ser Lys Lys Asp Glu Tyr Phe

85 90 95

Gly Val Tyr Glu Gln Ile Leu Asn Leu Asn Lys Glu Ser Phe Asn Lys

100 105 110

Leu Val Arg Asp Asn His Thr Tyr Gly Asn Ile Pro Thr Ser Pro Glu
 115 120 125
 Glu Leu Trp Asp Lys Val Ser Lys Leu Ile Tyr Ile Tyr Leu Lys Asp
 130 135 140

 Pro Thr Asn Val Ile Gly Gln Ala Gly Trp Thr Asn Pro Gln Asp Ala
 145 150 155 160
 Met Tyr Glu Phe Tyr Thr Val Val Gln Gln Glu Ile Trp Arg Tyr Thr
 165 170 175
 Asp Gly Gln Lys Val Asp Lys Asp Thr Asn Ser Tyr Leu Tyr Tyr Lys
 180 185 190
 Tyr Ser Lys Gln Gly Gln Lys Ala Val Tyr Leu Leu Arg Asp Ala Val
 195 200 205

 Asn Ser Ile Ser Ile Pro Ser Asn Phe Glu Leu Arg Gly Tyr Lys Pro
 210 215 220
 Glu Trp Val Gln Gly Gln Lys Gly Tyr Gln Ala Ile Val Thr Gly Arg
 225 230 235 240
 Leu Lys Val Asp Gln Pro Val Gly Glu Ile Lys Thr Thr Val Thr Ala
 245 250 255
 Gly Gly Lys Thr Ser Ser Glu Asn Asp Ile Ala Thr Leu Lys Ala Gln
 260 265 270

 Asp Val Ile Gly Gly Val Glu Val Ser Asp Lys Ile Thr Tyr Ser Gly
 275 280 285
 Leu Tyr Pro Asn Thr Glu Tyr Asp Val Ile Gly Glu Ile Tyr Glu Val
 290 295 300
 Lys Asp Gly Glu Leu Val Asn Pro Gly Arg Pro Val Ser Val Val Asn
 305 310 315 320
 Ser Gly Asp Asp Leu Lys Thr Asp Ala Thr Gly Lys Gly Lys Trp Thr
 325 330 335

 Leu Asn Phe Gly Lys Leu Asp Leu Glu Ala Gly Lys Ser Tyr Val Val
 340 345 350
 Phe Glu Lys Val Val Ser Leu Lys Asn Val Ile Asp Thr Asp Gly Asp

355	360	365
Gly Lys Pro Asp Lys Lys Gln Glu Leu Ser His Asn Asp Pro Lys Asp		
370	375	380
Lys Ser Gln Thr Phe Thr Ile Leu Pro Lys Glu Ile Val Glu Gln Asp		
385	390	395
		400
Val Val Phe Ser Lys Val Asn Val Ala Gly Glu Glu Ile Ala Gly Ala		
405	410	415
Lys Ile Gln Leu Lys Asp Ala Gln Gly Gln Val Val His Ser Trp Thr		
420	425	430
Ser Lys Ala Gly Gln Ser Glu Thr Val Lys Leu Lys Ala Gly Thr Tyr		
435	440	445
Thr Phe His Glu Ala Ser Ala Pro Thr Gly Tyr Leu Ala Val Thr Asp		
450	455	460
Ile Thr Phe Glu Val Asp Val Gln Gly Lys Val Thr Val Lys Asp Ala		
465	470	475
		480
Asn Gly Asn Gly Val Lys Ala Asp Gly Asn Lys Leu Thr Val Thr Asp		
485	490	495
Gln Ala Ala Pro Ser Val Pro Asn Glu Gln Asp Val Val Phe Ser Lys		
500	505	510
Val Asn Val Ala Gly Glu Glu Ile Ala Gly Ala Lys Ile Gln Leu Lys		
515	520	525
Asp Ala Gln Gly Gln Val Val His Ser Trp Thr Ser Lys Ala Gly Gln		
530	535	540
Ser Glu Thr Val Lys Leu Lys Ala Gly Thr Tyr Thr Phe His Glu Ala		
545	550	555
		560
Ser Ala Pro Thr Gly Tyr Leu Ala Val Thr Asp Ile Thr Phe Glu Val		
565	570	575
Asp Val Gln Gly Lys Val Thr Val Lys Asp Ala Asn Gly Asn Gly Val		
580	585	590
Lys Ala Asp Gly Asn Lys Leu Thr Val Thr Asp Gln Ala Ala Pro Ser		
595	600	605

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

Leu Lys Ala Gly Thr Tyr Thr Phe His Glu Ala Ser Ala Pro Thr Gly
660 665 670
Tyr Leu Ala Val Thr Asp Ile Thr Phe Glu Val Asp Val Gln Gly Lys
675 680 685
Val Thr Val Lys Asp Ala Asn Gly Asn Gly Val Lys Ala Asp Gly Asn
690 695 700
Lys Leu Thr Val Thr Asp Gln Ala Ala Pro Ser Val Pro Asn Glu Gln
705 710 715 720

Asp Val Val Phe Ser Lys Val Asn Val Ala Gly Glu Glu Ile Ala Gly
725 730 735

Ala Lys

<210> 32

<211> 2215

<212> DNA

<213> Streptococcus intermedius

<400> 32

atgaaaaaga gaagaggaca atttttcaaa agtgcaattt cgtttttggt tgtatttttg	60
atggtaatgg taagtatcat ttacccatct tcaaaaatta aagcagatgg atttcctaat	120
gatgctacgg gagtatcgcc aaatggtaaa tattactcgg caggagaga aaaccgttta	180
ggaatggtta catcagatga attgcataca gctacagaat tattcggttt ttgtatggca	240
aatagcaaga aatatccagg atatgattca aaaaaggatg agtattttgg ggtgtatgaa	300
caaatcttaa accttaataa agaaagcttt aataagcttg ttagagataa tcatacgtat	360
ggtaacattc ctacaagtcc agaggaactt tgggataaag tatctaaact gatttatatt	420
tatttgaaag acctacaaa tgttattgga caagctgggt ggacgaatcc acaggatgca	480
atgtatgaat ttatactgt tgtacaacag gaaatatggc gttatacaga tggacaaaag	540

gtggataaag acaccaattc atatttgtat tataaatatt caaaacaagg tcaaaaagca 600
 gtgtacttac tgcgtgacgc tgtgaatagc atcagtatac ctagtaattt tgaacttcgt 660

ggctataaac ctgaatgggt tcaaggtaa aaaggatacc aagctattgt aactggtaga 720
 ttgaaagtag atcaacctgt cggggaaata aagactacag taacagcagg tggaaaaacc 780
 tcaagtgaac acgacattgc tacattgaag gcgcaagacg ttataggtgg ggttgaagtc 840
 tctgataaga taacatatag tggctctttat ccaaatacag aatatgatgt tataggtgaa 900
 atttacgaag taaaagatgg agaacttgtt aatccaggac gaccgggttc tgtagtcaat 960
 agtgggtgacg atttaaaaac agatgcaaca ggaaaaggga aatggacatt aaactttgga 1020
 aagcttgatt tagaagcagg aaaatcctat gtggctcttg aaaaagttgt ttcattaaaa 1080

aacgtgatag atacagatgg agatggaaaa ccggataaaa aacaagaact atcgcataat 1140
 gatccaaaag ataatcgca aacatttaca attttaccta aggaaatagt tgaacaagac 1200
 gttgtcttca gtaaggtaga tgtggctggg gaagaaatcg ctggtgcaa gatccaactg 1260
 aaggatgcgc aaggtaagt tgttcattcc tggacttcta aagcgggtca aagtgaacg 1320
 gtcaaatga aagctggcac ctatactttc catgaagcat ccgctccgac tggttacttg 1380
 gccgtaacgg atatacatt cgaagtagat gttcaaggaa aagtacgggt taaggatgcc 1440
 aacggcaatg gtgttaaggc ggatggtaat aagttaacgg tgaccgatca agctgctcct 1500

agcgtaccga atgaacaaga cgttgtcttc agtaaggtaga atgtggctgg tgaagaaatc 1560
 gctggtgcga agatccaact gaaggatgcg caaggtaag ttgttcattc ctggacttct 1620
 aaagcgggtc aaagtgaac ggtcaaatg aaagctggca cctatacttt ccatgaagca 1680
 tccgctccga ctggttactt ggccgtaacg gatatacat tcgaagtaga tgttcaagga 1740
 aaagtacgg ttaaggatgc caacggcaat ggtgttaagg cggatggtaa taagttaacg 1800
 gtgaccgatc aagctgctcc tagcgtaccg aatgaacaag acgttgtctt cagtaagggtg 1860
 aatgtggctg gtgaagaaat cgctggtgcg aagatccaac tgaaggatgc gcaaggtaa 1920

gttgttcatt cctggacttc taaagcgggt caaagtgaac cggtaaatg gaaagctggc 1980
 acctatactt tccatgaagc atccgctccg actggttact tggccgtaac ggatacaca 2040
 ttcgaagtag atgttcaagg aaaagtacg gtttaaggatg ccaacggcaa tgggtttaag 2100
 gcggatggta ataagttaac ggtgaccgat caagctgctc ctacgctacc gaatgaacaa 2160
 gacgttgtct tcagtaagggt gaatgtggct ggtgaagaaa tcgctggtgc gaaga 2215

<210> 33
 <211> 10
 <212> PRT

<213> Artificial sequence

<220><223> K-tag

<400> 33

Ala Thr His Ile Lys Phe Ser Lys Arg Asp

1 5 10

<210> 34

<211> 108

<212> PRT

<213> Artificial sequence

<220><223> SpyLigase

<400> 34

Met Ser Tyr Tyr His His His His His His Asp Tyr Asp Gly Gln Ser

1 5 10 15

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

20 25 30

Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly Gln Val Lys Asp

35 40 45

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

50 55 60

Asp Gly Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu Gln

65 70 75 80

Gly Gln Val Thr Val Asn Gly Lys Ala Thr Lys Gly Gly Ser Gly Gly

85 90 95

Ser Gly Gly Ser Gly Glu Asp Ser Ala Thr His Ile

100 105

<210> 35

<

211> 327

<212> DNA

<213> Artificial sequence

<220><223> SpyLigase

<400> 35

atgtcgtact accatcacca tcacatcac gattacgacg gtcagtccgg tgacggcaaa 60

gagttagctg gtgcaactat ggagttgcgt gattcatctg gtaaaactat tagtacatgg 120
 atttcagatg gacaagtgaag agatttctac ctgtatccag gaaaatatac atttgcgaa 180
 accgcagcac cagacgggta tgaggttagca actgctatta cctttacagt taatgagcaa 240
 ggtcagggta ctgtaaatgg caaagcaact aaaggtggga gtggtggcag cggaggtagt 300
 ggcgaggaca gcgctacca tatttaa 327

<210> 36

<211> 12

<212> PRT

<213> Artificial sequence

<220><223> Spacer

<400> 36

Ser Ser Gly Leu Val Pro Arg Gly Ser His Met Gly

1 5 10

<210> 37

<211> 6

<212> PRT

<213> Artificial sequence

<220><223> Spacer

<400> 37

Gly Ser Gly Glu Ser Gly

1 5

<210> 38

<211> 33

<212> PRT

<213> Artificial sequence

<220><223> SnoopTag-His

<400> 38

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Lys Leu Gly Asp Ile Glu Phe Ile Lys Val Asn

20 25 30

Lys

<210> 39

<211> 133

<212> PRT

<213> Artificial sequence

<220><223> SnoopCatcher-His

<400> 39

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Lys Pro Leu Arg Gly Ala Val Phe Ser Leu Gln

20 25 30

Lys Gln His Pro Asp Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn

35 40 45

Gly Thr Tyr Gln Asn Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe

50 55 60

Lys Asn Leu Ser Asp Gly Lys Tyr Arg Leu Phe Glu Asn Ser Glu Pro

65 70 75 80

Ala Gly Tyr Lys Pro Val Gln Asn Lys Pro Ile Val Ala Phe Gln Ile

85 90 95

Val Asn Gly Glu Val Arg Asp Val Thr Ser Ile Val Pro Gln Asp Ile

100 105 110

Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys His Tyr Ile Thr Asn Glu

115 120 125

Pro Ile Pro Pro Lys

130

<210> 40

<211> 99

<212> DNA

<213> Artificial sequence

<220><223> SnoopTag-His

<400> 40

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60

atgaaactgg gcgatattga atttattaaa gtgaacaaa 99

<210> 41

<211> 399

<212> DNA

<213> Artificial sequence

<220><223> SnoopCatcher-His

<400> 41

```
atgggcagca gccatcatca tcatcatcac agcagcggcc tggcgccgcg cggcagccat      60
atgaagccgc tgcgtgggtgc cgtgttttagc ctgcagaaac agcatcccga ctatcccgat    120
atctatggcg cgattgatca gaatgggacc tatcaaaatg tgcgtaccgg cgaagatggt      180
aaactgacct ttaagaatct gagcgtatggc aaatatcgcc tgtttgaaaa tagcgaaccc     240
gctggctata aaccggtgca gaataagccg attgtggcgt ttcagattgt gaatggcgaa      300
```

```
gtgcgtgatg tgaccagcat tgtgccgcag gatattccgg ctacatatga attaccaac      360
ggtaaacatt atatcaccaa tgaaccgata cgcgcgaaa                             399
```

<210> 42

<211> 35

<212> PRT

<213> Artificial sequence

<220><223> PsCsTag-His

<400> 42

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Gly Gly Asn Lys Leu Thr Val Thr Asp Gln Ala

20 25 30

Ala Pro Ser

35

<210> 43

<211> 112

<212> PRT

<213> Artificial sequence

<220><223> PsCsCatcher-His

<400> 43

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Glu Gln Asp Val Val Phe Ser Lys Val Asn Val
 20 25 30
 Ala Gly Glu Glu Ile Ala Gly Ala Lys Ile Gln Leu Lys Asp Ala Gln
 35 40 45

Gly Gln Val Val His Ser Trp Thr Ser Lys Ala Gly Gln Ser Glu Thr
 50 55 60
 Val Lys Leu Lys Ala Gly Thr Tyr Thr Phe His Glu Ala Ser Ala Pro
 65 70 75 80
 Thr Gly Tyr Leu Ala Val Thr Asp Ile Thr Phe Glu Val Asp Val Gln
 85 90 95
 Gly Lys Val Thr Val Lys Asp Ala Asn Gly Asn Gly Val Lys Ala Asp
 100 105 110

<210> 44

<211> 105

<212> DNA

<213> Artificial sequence

<220><223> PsCsTag-His

<400> 44

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60

atgggaggca acaaactgac cgtgaccgat caggcggcgc cgagc 105

<210> 45

<211> 336

<212> DNA

<213> Artificial sequence

<220><223> PsCsCatcher-His

<400> 45

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60

atggaacagg atgtggtgtt tagcaaagtg aatgtggctg gcgaggaaat tgcgggagcg 120

aaaattcagt tgaagacgc gcagggccag gtggtgcata gctggaccag caaagcgggc 180

caaagcgaaa ccgtgaagct gaaagccggc acctatacct ttcatgagcg gagcgcaccg 240

accggctatc tggcggtgac cgatattacc tttgaagtgg atgtgcaggg caaagttaca 300

gtgaaagatg cgaatggcaa tgggtgtgaaa gcggat 336

<210> 46

<211> 41

<212> PRT

<213> Artificial sequence

<220><223> RrgATag-His

<400> 46

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
1 5 10 15

Arg Gly Ser His Met Gly Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn
20 25 30

Asp Lys His Tyr Ile Thr Asn Glu Pro
35 40

<210> 47

<211> 125

<212> PRT

<213> Artificial sequence

<220><223> RrgACatcher-His

<400> 47

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
1 5 10 15

Arg Gly Ser His Met Lys Leu Gly Asp Ile Glu Phe Ile Lys Val Asn
20 25 30

Lys Asn Asp Lys Lys Pro Leu Arg Gly Ala Val Phe Ser Leu Gln Lys
35 40 45

Gln His Pro Asp Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly
50 55 60

Thr Tyr Gln Asn Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Lys
65 70 75 80

Asn Leu Ser Asp Gly Lys Tyr Arg Leu Phe Glu Asn Ser Glu Pro Ala
85 90 95

Gly Tyr Lys Pro Val Gln Asn Lys Pro Ile Val Ala Phe Gln Ile Val
100 105 110

Asn Gly Glu Val Arg Asp Val Thr Ser Ile Val Pro Gln

115 120 125

<210> 48

<211> 123

<212> DNA

<213> Artificial sequence

<220><223> RrgATag-His

<400> 48

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60

atgggagata ttccggctac atatgaattt accaacgata aacattatat caccaatgaa 120

ccg 123

<210> 49

<211> 375

<212> DNA

<213> Artificial sequence

<220><223> RrgACatcher-His

<400> 49

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60

atgaaactgg gcgatatga atttattaaa gtgaacaaaa acgataaaaa gccgctgcgt 120

ggtgccgtgt ttagcctgca gaaacagcat cccgactatc ccgatatcta tggcgcgatt 180

gatcagaatg ggacctatca aaatgtgcgt accggcgaag atggtaaact gacctttaag 240

aatctgagcg atggcaaata tcgcctgttt gaaaatagcg aaccgcgtgg ctataaaccg 300

gtgcagaata agccgattgt ggcgtttcag attgtgaatg gcgaagtgcg tgatgtgacc 360

agcattgtgc cgcag 375

<210> 50

<211> 409

<212> PRT

<213> Artificial sequence

<220><223> SnoopTag-MBP

<400> 50

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Lys Leu Gly Asp Ile Glu Phe Ile Lys Val Asn

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

Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val Leu Pro Thr
275 280 285

Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu Ser Ala Gly
290 295 300

Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu
305 310 315 320

Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn Lys Asp Lys
325 330 335

Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu Leu Ala Lys
340 345 350

Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys Gly Glu Ile
355 360 365

Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala Val Arg Thr
370 375 380

Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp Glu Ala Leu
385 390 395 400

Lys Asp Ala Gln Thr Asn Ser Ser Ser
405

<210> 51

<211> 509

<212> PRT

<213> Artificial sequence

<220><223> SnoopCatcher-MBP

<400> 51

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
1 5 10 15

Arg Gly Ser His Met Lys Pro Leu Arg Gly Ala Val Phe Ser Leu Gln
20 25 30

Lys Gln His Pro Asp Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn
35 40 45

Gly Thr Tyr Gln Asn Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe

50 55 60
 Lys Asn Leu Ser Asp Gly Lys Tyr Arg Leu Phe Glu Asn Ser Glu Pro
 65 70 75 80
 Ala Gly Tyr Lys Pro Val Gln Asn Lys Pro Ile Val Ala Phe Gln Ile
 85 90 95
 Val Asn Gly Glu Val Arg Asp Val Thr Ser Ile Val Pro Gln Asp Ile
 100 105 110
 Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys His Tyr Ile Thr Asn Glu

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

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

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

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

305 310 315 320

Gly Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp

325 330 335

Leu Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala

340 345 350

Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro

355 360 365

Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr

370 375 380

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

385 390 395 400

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

405 410 415

Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val

420 425 430

Asn Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu

435 440 445

Glu Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln

450 455 460

Lys Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr

465 470 475 480

Ala Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val

485 490 495

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

500 505

<210> 52

<211> 411

<212> PRT

<213> Artificial sequence

<220><223> PsCsTag-MBP

<400> 52

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
1 5 10 15

Arg Gly Ser His Met Gly Gly Asn Lys Leu Thr Val Thr Asp Gln Ala
20 25 30

Ala Pro Ser Gly Ser Gly Glu Ser Gly Lys Ile Glu Glu Gly Lys Leu
35 40 45

Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu Val
50 55 60

Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu His
65 70 75 80

Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly Asp
85 90 95

Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr Ala
100 105 110

Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln Asp
115 120 125

Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys Leu
130 135 140

Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn Lys
145 150 155 160

Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala Leu
165 170 175

Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn Leu
180 185 190

Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly Tyr
195 200 205

Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly Val
210 215 220

Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu Ile
225 230 235 240

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

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

Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala Val
 370 375 380
 Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp Glu
 385 390 395 400
 Ala Leu Lys Asp Ala Gln Thr Asn Ser Ser Ser
 405 410

<210> 53

<211> 488

<212> PRT

<213> Artificial sequence

<220><223> PsCsCatcherMBP

<400> 53

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15
 Arg Gly Ser His Met Glu Gln Asp Val Val Phe Ser Lys Val Asn Val
 20 25 30

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

275 280 285
Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala
290 295 300
Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys
305 310 315 320
His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn

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

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

450 455 460
Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys
465 470 475 480
Asp Ala Gln Thr Asn Ser Ser Ser

485

<210> 54

<211> 417

<212> PRT

<213> Artificial sequence

<220><223> RrgATag-MBP

<400> 54

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Gly Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn

20 25 30

Asp Lys His Tyr Ile Thr Asn Glu Pro Gly Ser Gly Glu Ser Gly Lys

35 40 45

Ile Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr

50 55 60

Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile

65 70 75 80

Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln

85 90 95

Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp

100 105 110

Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro

115 120 125

Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val

130 135 140

Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu

145 150 155 160

Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp

165 170 175

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

180 185 190

Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile

195 200 205

Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp

210 215 220

Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr

225 230 235 240

Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp
 245 250 255
 Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr
 260 265 270
 Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn
 275 280 285
 Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro
 290 295 300
 Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys
 305 310 315 320
 Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly
 325 330 335
 Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys
 340 345 350
 Ser Tyr Glu Glu Glu Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met
 355 360 365
 Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser
 370 375 380
 Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly
 385 390 395 400
 Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Asn Ser Ser
 405 410 415
 Ser

<210> 55

<211> 501

<212> PRT

<213> Artificial sequence

<220><223> RrgACatcher-MBP

<400> 55

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
 1 5 10 15

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

260 265 270
 Lys Gly Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr
 275 280 285
 Trp Pro Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn

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

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

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

 485 490 495
 Thr Asn Ser Ser Ser
 500

<210> 56

<211> 256

<212> PRT

<213> Artificial sequence

<220><223> SpyCatcher-SnoopCatcher

<400> 56

Met Ser Tyr Tyr His His His His His His Asp Tyr Asp Ser Ala Thr

1 5 10 15

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

20 25 30

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

35 40 45

Ile Ser Asp Gly Gln Val Lys Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr

50 55 60

Thr Phe Val Glu Thr Ala Ala Pro Asp Gly Tyr Glu Val Ala Thr Ala

65 70 75 80

Ile Thr Phe Thr Val Asn Glu Gln Gly Gln Val Thr Val Asn Gly Lys

85 90 95

Ala Thr Lys Gly Asp Ala His Ile Gly Ser Pro Ala Asn Leu Lys Ala

100 105 110

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

115 120 125

Ala Asn Ala Lys Lys Leu Lys Glu Gln Leu Glu Lys Gly Ser His Met

130 135 140

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

145 150 155 160

Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asn

165 170 175

Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Lys Asn Leu Ser Asp

180 185 190

Gly Lys Tyr Arg Leu Phe Glu Asn Ser Glu Pro Ala Gly Tyr Lys Pro

195 200 205

Val Gln Asn Lys Pro Ile Val Ala Phe Gln Ile Val Asn Gly Glu Val
210 215 220

Arg Asp Val Thr Ser Ile Val Pro Gln Asp Ile Pro Ala Thr Tyr Glu

225 230 235 240

Phe Thr Asn Gly Lys His Tyr Ile Thr Asn Glu Pro Ile Pro Pro Lys
245 250 255

<210> 57
<211> 256
<212> PRT
<213> Artificial sequence
<220><223> SnoopCatcher-SpyCatcher
<400> 57

Met Ser Tyr Tyr His His His His His His Lys Pro Leu Arg Gly Ala
1 5 10 15

Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr Pro Asp Ile Tyr Gly

20 25 30

Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asn Val Arg Thr Gly Glu Asp
35 40 45

Gly Lys Leu Thr Phe Lys Asn Leu Ser Asp Gly Lys Tyr Arg Leu Phe
50 55 60

Glu Asn Ser Glu Pro Ala Gly Tyr Lys Pro Val Gln Asn Lys Pro Ile
65 70 75 80

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

85 90 95

Val Pro Gln Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys His
100 105 110

Tyr Ile Thr Asn Glu Pro Ile Pro Pro Lys Gly Ser Pro Ala Asn Leu
115 120 125

Lys Ala Leu Glu Ala Gln Lys Gln Lys Glu Gln Arg Gln Ala Ala Glu
130 135 140

Glu Leu Ala Asn Ala Lys Lys Leu Lys Glu Gln Leu Glu Lys Gly Ser

145 150 155 160
 His Met Asp Tyr Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg Asp
 165 170 175
 Glu Asp Gly Lys Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp Ser
 180 185 190
 Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly Gln Val Lys Asp
 195 200 205
 Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala Pro

 210 215 220
 Asp Gly Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu Gln
 225 230 235 240
 Gly Gln Val Thr Val Asn Gly Lys Ala Thr Lys Gly Asp Ala His Ile
 245 250 255

<210> 58

<211> 86

<212> PRT

<213> Artificial sequence

<220><223> SpyTag-SnoopTag

<400> 58

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
 1 5 10 15

Arg Gly Ser His Met Ala His Ile Val Met Val Asp Ala Tyr Lys Pro
 20 25 30
 Thr Lys Gly Ser Pro Ala Asn Leu Lys Ala Leu Glu Ala Gln Lys Gln
 35 40 45
 Lys Glu Gln Arg Gln Ala Ala Glu Glu Leu Ala Asn Ala Lys Lys Leu
 50 55 60
 Lys Glu Gln Leu Glu Lys Gly Ser His Met Lys Leu Gly Asp Ile Glu
 65 70 75 80

Phe Ile Lys Val Asn Lys
 85

<210> 59

<211> 86

<212> PRT

<213> Artificial sequence

<220><223> SnoopTag-SpyTag

<400> 59

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Lys Leu Gly Asp Ile Glu Phe Ile Lys Val Asn

20 25 30

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

35 40 45

Glu Gln Arg Gln Ala Ala Glu Glu Leu Ala Asn Ala Lys Lys Leu Lys

50 55 60

Glu Gln Leu Glu Lys Gly Ser His Met Ala His Ile Val Met Val Asp

65 70 75 80

Ala Tyr Lys Pro Thr Lys

85

<210> 60

<211> 1230

<212> DNA

<213> Artificial sequence

<220><223> SnoopTag-MBP

<400> 60

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60

atgggaaaac tgggcgatat tgaatttatt aaagtgaaca aaggtagtgg tgaaagtgg 120

aaaatcgaag aaggtaaact ggtaatctgg attaacggcg ataaaggcta taacggtctc 180

gctgaagtcg gtaagaaatt cgagaaagat accggaatta aagtcaccgt tgagcatccg 240

gataaactgg aagagaaatt cccacagggt gcggcaactg gcgatggccc tgacattatc 300

ttctgggcac acgaccgctt tgggtggctac gctcaatctg gcctgttggc tgaaatcacc 360

ccggacaaa cgttccagga caagctgtat ccgtttacct gggatgccgt acgttacaac 420

ggcaagctga ttgcttacc gatcgtgtt gaagcgttat cgctgattta taacaaagat 480

ctgctgccga acccgccaaa aacctgggaa gagatcccg cgctggataa agaactgaaa 540

gcgaaaggta agagcgcgct gatgttcaac ctgcaagaac cgtacttcac ctggccgctg 600
attgtctgctg acggggggtta tgcgttcaag tatgaaaacg gcaagtacga cattaagac 660
gtggcgctgg ataacgctgg cgcgaaagcg ggtctgacct tcctggttga cctgattaaa 720
aacaacaca tgaatgcaga caccgattac tccatcgag aagctgcctt taataaaggc 780
gaaacagcga tgaccatcaa cgccccgtgg gcatggtcca acatcgacac cagcaaagtg 840
aattatggtg taacggtact gccgaccttc aagggtcaac catccaaacc gttcgttggc 900
gtgctgagcg caggtattaa cgccgccagt ccgaacaaag agctggcaaa agagttcctc 960

gaaaactatc tgcigactga tgaaggtctg gaagcggtta ataaagacaa accgctgggt 1020
gccgtagcgc tgaagtctta cgaggaagag ttggcgaaag atccacgtat tgccgccact 1080
atggaaaacg ccagaaaagg tgaaatcatg ccgaacatcc cgcagatgtc cgctttctgg 1140
tatgccgtgc gtactgcggt gatcaacgcc gccagcggtc gtcagactgt cgatgaagcc 1200
ctgaaagacg cgcagactaa ttcgagctcg 1230

<210> 61
<211> 1527
<212> DNA
<213> Artificial sequence
<220><223> SnoopCatcher-MBP
<400> 61

atgggcagca gccatcatca tcatcatcac agcagcggcc tggtgccgcg cggcagccat 60

atgaagccgc tgcgtggtgc cgtgtttagc ctgcagaaac agcatcccga ctatcccgat 120
atctatggcg cgattgatca gaatgggacc tatcaaatg tgcgtaccgg cgaagatggt 180
aaactgacct ttaagaatct gagcgatggc aaatatcgcc tgtttgaaaa tagcgaaccc 240
gttggtatata aaccggtgca gaataagccg attgtggcgt ttcagattgt gaatggcgaa 300
gtgcgtgatg tgaccagcat tgtgccgcag gatattccgg ctacatatga attaccaac 360
ggtaaacatt atatcaccaa tgaaccgata ccgccgaaag gtagtggtga aagtgtgtaa 420
atcgaagaag gtaaacctgg aatctggatt aacggcgata aaggctataa cggctctcgt 480

gaagtcggtg agaaattcga gaaagatacc ggaattaaag tcaccgttga gcatccggat 540
aaactggaag agaaattccc acaggttgcg gcaactggcg atggccctga cattatcttc 600
tgggcacacg accgctttgg tggctacgt caatctggcc tgttggctga aatcaccccg 660
gacaaagcgt tccaggacaa gctgtatccg ttacctggg atgccgtacg ttacaacggc 720
aagctgattg cttacccgat cgctgttgaa gcgttatcgc tgatttataa caaagatctg 780
ctgccgaacc cgccaaaaac ctgggaagag atcccggcgc tggataaaga actgaaagcg 840

aaagtaaga ggcgcgtgat gttcaacctg caagaaccgt acttcacctg gccgctgatt 900

gctgctgacg ggggttatgc gttcaagat gaaaacggca agtacgacat taaagacgtg 960

ggcgtggata acgctggcgc gaaagcgggt ctgaccttcc tggttgacct gattaaaaac 1020

aaacacatga atgcagacac cgattactcc atcgcagaag ctgcctttaa taaaggcgaa 1080

acagcgatga ccatcaacgg ccgctgggca tggccaaca tcgacaccag caaagtgaat 1140

tatgggtgaa cggtactgcc gaccttcaag ggtcaaccat ccaaaccgtt cgttggcgtg 1200

ctgagcgag gtattaacgc cgccagtcg aacaaagagc tggcaaaaga gttcctcgaa 1260

aactatctgc tgactgatga aggtctggaa gcggttaata aagacaaacc gctgggtgcc 1320

gtagcgctga agtcttacga ggaagagttg gcgaaagatc cacgtattgc cgccactatg 1380

gaaaacgccc agaaaggiga aatcatgccg aacatccgc agatgtccgc tttctggtat 1440

gccgtgcgta ctgcggtgat caacgccgcc agcggctcgc agactgtcga tgaagccctg 1500

aaagacgcgc agactaatc gagctcg 1527

<210> 62

<211> 1233

<212> DNA

<213> Artificial sequence

<220><223> PsCsTag-MBP

<400> 62

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60

atgggaggca acaaactgac cgtgaccgat caggcggcgc cgagcggtag tggtgaaagt 120

ggtaaaatcg aagaaggtaa actggtaatc tggattaacg gcgataaagg ctataacggt 180

ctcgtgaag tcggaagaa attcgagaaa gataccggaa ttaaagtcac cgttgagcat 240

ccgataaac tggaagagaa attccacag gttgcggcaa ctggcgatgg cctgacatt 300

atcttctggg cacacgaccg ctttgggtggc tacgtcaat ctggcctgtt ggctgaaatc 360

accccgaca aagcgttcca ggacaagctg tatccgttta cctgggatgc cgtacgttac 420

aacggcaagc tgattgctta ccgcatcgt gttgaagcgt tatcgtgat ttataacaaa 480

gatctgctgc cgaaccgcc aaaaacctgg gaagagatcc cggcgtgga taaagaactg 540

aaagcgaag gtaagagcgc gctgatgttc aacctgcaag aaccgtactt cacctggccg 600

ctgattgctg ctgacggggg ttatgcgttc aagtatgaaa acggcaagta cgacattaaa 660

gacgtggcgc tggataacgc tggcgcaaaa gcgggtctga ccttcctggt tgacctgatt 720

aaaaacaaac acatgaatgc agacaccgat tactccatcg cagaagctgc ctttaataaa 780

ggcgaaacag cgatgacat caacggcccg tgggcatggt ccaacatcga caccagcaaa 840
gtgaattatg gtgtaacggt actgccgacc ttcaagggtc aaccatcaa accgttcgtt 900
ggcgtgctga gcgcaggtat taacccgcc agtccgaaca aagagctggc aaaagagttc 960

ctcgaaaact atctgctgac tgatgaaggt ctggaagcgg ttaataaaga caaacgctg 1020
ggtgccgtag cgctgaagtc ttacaggaa gagttggcga aagatccacg tattgccgcc 1080
actatggaaa acgcccagaa aggtgaaatc atgccgaaca tcccgcagat gtccgctttc 1140
tggtatgccg tgcgtactgc ggtgatcaac gccgccagcg gtcgtcagac tgtcgatgaa 1200
gccctgaaag acgcccagac taattcgagc tcg 1233

<210> 63
<211> 1464
<212> DNA
<213> Artificial sequence
<220><223> PsCsCatcher-MBP
<400> 63

atgggcagca gccatcatca tcatcatcac agcagcggcc tggtgccgcg cggcagccat 60

atggaacagg atgtggtgtt tagcaaagtg aatgtggctg gcgaggaaat tgcgggagcg 120
aaaattcagt tgaagacgc gcagggccag gtggtgcata gctggaccag caaagcgggc 180
caaagcgaac ccgtgaagct gaaagccggc acctatacct ttcatgagcg gagcgcaccg 240
accggtatc tggcggtagc cgatattacc ttgaagtgg atgtgcaggc caaagttaca 300
gtgaaagatg cgaatggcaa tgggtgtgaaa gcggatggta gtggtgaaa tggtaaaatc 360
gaagaaggta aactggtaat ctggattaac ggcgataaag gctataacgg tctcgctgaa 420
gtcggtaaga aattcgagaa agataccgga attaaagtca ccgttgagca tccggataaa 480

ctggaagaga aattcccaca ggttgcggca actggcgatg gccctgacat tatcttctgg 540
gcacacgacc gctttggtgg ctacgtcaa tctggcctgt tggctgaaat caccgccgac 600
aaagcgttc aggacaagct gtatccgttt acctgggatg ccgtacgtta caacggcaag 660
ctgattgctt acccgatcgc tgttgaagcg ttatcgctga tttataacaa agatctgctg 720
ccgaaccgcg caaaaacctg ggaagagatc ccggcgtgg ataaagaact gaaagcgaac 780
ggtaagagcg cgtgatgtt caacctgcaa gaaccgtact tcacctggcc gctgattgct 840
gctgacgggg gttatgcgtt caagtatgaa aacggcaagt acgacattaa agacgtgggc 900

gtggataacg ctggcgcgaa agcgggtctg accttcttgg ttgacctgat taaaaacaaa 960
cacatgaatg cagacaccga ttactccatc gcagaagctg cctttaataa aggcgaacaa 1020

gcgatgacca tcaacggccc gtgggcatgg tccaacatcg acaccagcaa agtgaattat 1080
 ggtgtaacgg tactgccgac cttcaagggt caaccatcca aaccgttcgt tggcgtgctg 1140
 agcgcaggta ttaacgccgc cagtccgaac aaagagctgg caaaagagtt cctcgaaaac 1200
 tatctgctga ctgatgaagg tctggaagcg gtttaataaag acaaaccgct ggggtgccgta 1260
 gcgctgaagt cttacgagga agagtggcg aaagatccac gtattgccgc cactatggaa 1320

aacgcccaga aaggtgaaat catgccgaac atcccgcaga tgtccgcttt ctggtatgcc 1380
 gtgcgtactg cggatgatcaa cgccgccagc ggtcgtcaga ctgtcgatga agccctgaaa 1440
 gacgcgcaga ctaattcgag ctgc 1464

<210> 64
 <211> 1251
 <212> DNA
 <213> Artificial sequence
 <220><223> RrgATag-MBP
 <400> 64

atgggcagca gccatcatca tcatcatcac agcagcggcc tggcgccgcg cggcagccat 60
 atgggagata ttccggctac atatgaattt accaacgata aacattatat caccaatgaa 120
 ccgggtagtg gtgaaagtgg taaaatcgaa gaaggtaaac tggtaatctg gattaacggc 180

gataaaggct ataacggctc cgctgaagtc ggtaagaaat tcgagaaaga taccggaatt 240
 aaagtcaccg ttgagcatcc ggataaactg gaagagaaat tcccacaggt tgcggcaact 300
 ggcatggcc ctgacattat cttctgggca cagcaccgct ttggtggcta cgctcaatct 360
 ggcctgttgg ctgaaatcac cccggacaaa gcgttccagg acaagctgta tccgtttacc 420
 tgggatgccg tacgttacaa cggcaagctg attgcttacc cgatcgctgt tgaagcgta 480
 tcgctgattt ataacaaaga tctgctgccg aaccgcgcaa aaacctggga agagatcccg 540
 gcgctggata aagaactgaa agcgaaaggt aagagcgcgc tgatgttcaa cctgcaagaa 600

ccgtacttca cctggccgct gattgctgct gacggggggt atgcgttcaa gtatgaaaac 660
 ggcaagtacg acattaaaga cgtgggcgtg gataacgctg gcgcgaaagc gggctcgacc 720
 ttcttggttg acctgattaa aaacaaacac atgaatgcag acaccgatta ctccatcgca 780
 gaagctgcct ttaataaagg cgaaacagcg atgacatca acggcccgct ggcatggtcc 840
 aacatcgaca ccagcaaagt gaattatggt gtaacggtac tgccgacctt caagggtcaa 900
 ccatccaaac cgttcgttgg cgtgctgagc gcagggtatta acgccgccag tccgaacaaa 960
 gagctggcaa aagagttcct cgaaaactat ctgctgactg atgaaggtct ggaagcggtt 1020

aataaaagaca aaccgctggg tgccgtagcg ctgaagtctt acgaggaaga gttggcgaaa 1080
gatccacgta ttgccgccac tatggaaaac gccagaaaag gtgaaatcat gccgaacatc 1140
ccgcagatgt ccgctttctg gtatgccgtg cgtactgcgg tgatcaacgc cgccagcggc 1200
cgtcagactg tcgatgaagc cctgaaagac gcgcagacta attcgagctc g 1251
<210> 65
<211> 1503
<212> DNA
<213> Artificial sequence
<220><223> RrgACatcherMBP
<400> 65
atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60
atgaaactgg gcgatattga atttattaaa gtgaacaaaa acgataaaaa gccgctgcgt 120

ggtgccgtgt ttagcctgca gaaacagcat cccgactatc ccgatatcta tggcgcgatt 180
gatcagaatg ggacctatca aaatgtgcgt accggcgaag atggtaaaact gacctttaag 240
aatctgagcg atggcaata tcgcctgttt gaaaatagcg aaccgctgg ctataaaccg 300
gtgcagaata agccgattgt ggcgtttcag attgtgaatg gcgaagtgcg tgatgtgacc 360
agcattgtgc cgcagggtag tgggtaaaagt ggtaaaatcg aagaaggtaa actggtaatc 420
tggattaacg gcgataaagg ctataacggt ctgcgtgaag tcggtlaagaa attcgagaaa 480
gataccggaa ttaaagtcac cgttgagcat ccgataaac tggaagagaa attccacag 540

gttgcggcaa ctggcgatgg ccctgacatt atcttctggg cacacgaccg ctttgggtggc 600
tacgtcaat ctggcctgtt ggctgaaatc accccggaca aagcgttcca ggacaagctg 660
tatccgttta cctgggatgc cgtacgttac aacggcaagc tgattgctta cccgatcgt 720
gttgaagcgt tatcgtgat ttataacaaa gatctgctgc cgaacccgcc aaaaacctgg 780
gaagagatcc cggcgtgga taaagaactg aaagcgaaag gtaagagcgc gctgatgttc 840
aacctgcaag aaccgtactt cacctggccg ctgattgctg ctgacggggg ttatgcgttc 900
aagtatgaaa acggcaagta cgacattaaa gacgtgggcg tggataacgc tggcgcgaaa 960

gcgggtctga ctttctcgtt tgacctgatt aaaaacaaac acatgaatgc agacaccgat 1020
tactccatcg cagaagctgc ctttaataaa ggcgaaacag cgatgacat caacggcccc 1080
tgggcatggt ccaacatcga caccagcaaa gtgaattatg gtgtaacggt actgccgacc 1140
ttcaagggtc aaccatcaa accgttcgtt ggcgtgctga gcgcaggtat taacgccgcc 1200
agtccgaaca aagagctggc aaaagagttc ctcgaaaact atctgctgac tgatgaaggt 1260
ctggaagcgg ttaataaaga caaacgctg ggtgccgtag cgctgaagtc ttacaggaa 1320

gagttggcga aagatccacg tattgccgcc actatggaaa acgcccagaa aggtgaaatc 1380

atgccgaaca tcccgcatat gtccgctttc tggatatgccg tgcgtactgc ggtgatcaac 1440

gccgccagcg gtcgtcagac tgctgatgaa gccctgaaag acgcgagac taattcgagc 1500

tcg 1503

<210> 66

<211> 768

<212> DNA

<213> Artificial sequence

<220><223> SpyCatcher-SnoopCatcher

<400> 66

atgtcgtact accatcacca tcaccatcac gattacgata gtgctacca tattaattc 60

tcaaaacgtg atgaggacgg caaagagtta gctggtgcaa ctatggagtt gcgtgattca 120

tctggtaaaa ctattagtag atggatttca gatggacaag tgaaagattt ctacctgtat 180

ccaggaaaat atacatttgt cgaaaccgca gcaccagacg gttatgaggt agcaactgct 240

attaccttta cagttaatga gcaaggtagc gttactgtaa atggcaaagc aactaaaggt 300

gacgctcata ttggatcccc cgccaacctg aaggccctgg agggccagaa gcagaaggag 360

cagagacagg ccgccgagga gctggccaac gccaagaagc tgaaggagca gctggagaag 420

ggatcccata tgaagccgct gcgtggtagc gtgttttagc tgcagaaaca gcatcccgac 480

tatcccgata tctatggcgc gattgatcag aatgggacct atcaaatgt gcgtaccggc 540

gaagatggta aactgacctt taagaatctg agcgaaggca aatatcgctt gtttgaaaat 600

agcgaacccg ctggctataa accggtgcag aataagccga ttgtggcggt tcagattgtg 660

aatggcgaag tgcgtgatgt gaccagcatt gtgccgcagg atattccggc tacatatgaa 720

tttaccaacg gtaaacatta tatcaccaat gaaccgatac cgccgaaa 768

<210> 67

<211> 768

<212> DNA

<213> Artificial sequence

<220><223> SnoopCatcher-SpyCatcher

<400> 67

atgtcgtact accatcacca tcaccatcac aagccgctgc gtggtgccgt gtttagcctg 60

cagaaacagc atcccgaata tcccgaatac tatggcgaga ttgatcagaa tgggacctat 120

caaaatgtgc gtaccggcga agatggtaaa ctgaccttta agaattctgag cgatggcaaa 180

tatcgctgt ttgaaaatag cgaacccgct ggctataaac cggcgcagaa taagccgatt 240

gtggcgtttc agattgtgaa tggcgaagtg cgtgatgtga ccagcattgt gccgcaggat 300

attccggcta catatgaatt taccaacggt aacattata tcaccaatga accgataccg 360

ccgaaaggat ccccgccaa cctgaaggcc ctggaggccc agaagcagaa ggagcagaga 420

caggccgccc aggagctggc caacccaag aagctgaagg agcagctgga gaagggatcc 480

catatggatt acgatagtc taccatatt aaattctcaa aacgtgatga ggacggcaaa 540

gagttagctg gtgcaactat ggagttgcgt gattcatctg gtaaaactat tagtacatgg 600

atttcagatg gacaagtga agatttctac ctgtatccag gaaaatatac attgtcgaa 660

accgcagcac cagacggta tgaggtagca actgctatta cctttacagt taatgagcaa 720

ggtcaggta ctgtaatgg caaagcaact aaagtgacg ctcatatt 768

<210> 68

<211> 258

<212> DNA

<213> Artificial sequence

<220><223> SpyTag-SnoopTag

<400> 68

atgggcagca gccatcatca tcatcatcac agcagcggcc tggcgccgc cggcagccat 60

atggcccaca tcgtgatgtt ggacgcctac aagccgacga agggatcccc cgccaacctg 120

aaggccctgg agggccagaa gcagaaggag cagagacagg ccgccgagga gctggccaac 180

gccagaagc tgaaggagca gctggagaag ggatccata tgaaactggg cgatattgaa 240

tttattaaag tgaacaaa 258

<210> 69

<211> 258

<212> DNA

<213> Artificial sequence

<220><223> SnoopTag-SpyTag

<400> 69

atgggcagca gccatcatca tcatcatcac agcagcggcc tggcgccgc cggcagccat 60

atgaaactgg gcgatattga atttattaaa gtgaacaaag gatccccgc caacctgaag 120

gccctggagg ccgagaagca gaaggagcag agacaggccg ccgaggagct ggccaacgcc 180

aagaagctga aggagcagct ggagaaggga tcccatatgg cccacatcgt gatggtggac 240

gcctacaagc cgacgaag 258

<210> 70

<211> 366

<212> PRT

<213> Artificial sequence

<220><223> MBP_x

<400> 70

Lys Ile Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys Gly

1 5 10 15

Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr Gly

20 25 30

Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe Pro

35 40 45

Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala His

50 55 60

Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr

65 70 75 80

Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala

85 90 95

Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala

100 105 110

Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr

115 120 125

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

130 135 140

Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu

145 150 155 160

Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Gly Asp Ile Lys Asp

165 170 175

Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val

180 185 190
 Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile
 195 200 205
 Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly
 210 215 220
 Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val
 225 230 235 240

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

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

<210> 71

<211> 1098

<212> DNA

<213> Artificial sequence

<220><223> MBP_x

<400> 71

aaaatcgaag aaggtaaact ggtaatctgg attaacggcg ataaaggcta taacggtctc 60
 gctgaagtcg gtaagaaatt cgagaaagat accggaatta aagtcaccgt tgagcatccg 120
 gataaactgg aagagaaatt cccacaggtt gcggcaactg gcgatggccc tgacattatc 180

ttctgggcac acgaccgctt tggtaggctac gctcaatctg gcctgttggc tgaatcacc 240
ccggacaaaag cgttccagga caagctgtat ccgtttacct gggatgccgt acgttacaac 300
ggcaagctga ttgcttacc gatcgtgtt gaagcgttat cgctgattta taacaaagat 360

ctgctgccga acccgccaaa aacctgggaa gagatcccgg cgctggataa agaactgaaa 420
gcgaaaggta agagcgcgt gatgttcaac ctgcaagaac cgtacttcac ctggccgctg 480
attgctgctg acgggggtta tgcgttcaag tatggcgaca ttaaagacgt gggcgtggat 540
aacgctggcg cgaaagcggg tctgaccttc ctggttgacc tgattaaaaa caaacacatg 600
aatgcagaca ccgattactc catcgagaa gctgccttta ataaaggcga aacagcgatg 660
accatcaacg gcccgtaggc atggccaac atcgacacca gcaaagtga ttatggtgta 720
acggtactgc cgaccttcaa gggtaacca tccaaaccgt tcgttggcgt gctgagcgca 780

ggtattaacg ccgccagtc gaacaagag ctggcaaaag agttcctcga aaactatctg 840
ctgactgatg aaggtctgga agcggttaat aaagacaaac cgctgggtgc cgtagcgctg 900
aagtcttacg aggaagagtt ggtgaaagat ccacgtgtgg ccgccactat ggaaaacgcc 960
cagaaagggtg aaatcatgcc gaacatcccg cagatgtccg ctttctggta tgccgtgcgt 1020
actgcggtga tcaacgccgc cagcggtcgt cagactgtcg atgaagccct gaaagacgcg 1080
cagacgaatt ctagttcc 1098

<210> 72

<211> 137

<212> PRT

<213> Artificial sequence

<220><223> SnoopTag-AffiHER2-SpyTag

<400> 72

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro

1 5 10 15

Arg Gly Ser His Met Gly Lys Leu Gly Asp Ile Glu Phe Ile Lys Val

20 25 30

Asn Lys Gly Ser Gly Glu Ser Gly Ser Gly Ala Ser Met Thr Gly Gly

35 40 45

Gln Gln Met Gly Arg Asp Pro Gly Val Asp Asn Lys Phe Asn Lys Glu

50 55 60

Met Arg Asn Ala Tyr Trp Glu Ile Ala Leu Leu Pro Asn Leu Asn Asn

Gln Gln Lys Arg Ala Phe Ile Arg Ser Leu Tyr Asp Asp Pro Ser Gln
85 90 95
Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln Ala
100 105 110
Pro Lys Gly Leu Glu Gly Ser Gly Glu Gly Ser Gly Ala His Ile Val

Met Val Asp Ala Tyr Lys Pro Thr Lys

130 135

<210> 73

<211> 40

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 73

gtgccgcagg atattccggc tacatatgaa tttaccaacg 40

<210> 74

<211> 47

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 74

gctacatatg aatttaccaa cggtaaacaat tatatcacca atgaacc 47

<210> 75

 $\langle 211 \rangle$

31

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 75

acattatatac accgctgaac cgataccgcc g 31

<210> 76

<211> 31

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 76

ggtagtggtg aaagtggtaa aatcgaagaa g 31

<210> 77

<211> 76

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 77

aaactgggcg atattgaatt tattaaagtg aacaaaaacg ataaaggtag tggtgaaagt 60

ggtaaaatcg aagaag 76

<210> 78

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 78

tcccatatgg ctgccgcgcg 20

<210> 79

<211> 65

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 79

tttatcgttt ttgttcaatt taataaatc aatatcgccc agttttccca tatggctgcc 60

gcgcg 65

<210> 80

<211> 46

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 80

gaatttatta aagtgaacaa aggtagtggg gaaagtggta aaatcg 46

<210> 81

<211> 39

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 81

gggcgatatt gaatttattg cagtgaacaa aggtagtgg 39

<210> 82

<211> 57

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 82

gttcgggagg tagtgggtgcc atgggtgata cttatcagg tttatcaagt gagcaag 57

<210> 83

<211> 54

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 83

tactaagctt ctattaaata tgagcgacac ctttagttgc ttgcccattt acag 54

<210> 84

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 84

atctcatatg ggcagcagcc atcatcatca tcacac 37

<210> 85

<211> 52

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 85
glatcaacca tggcaccact accgcccga cccgagctcg aattagtctg cg 52
<210> 86
<211> 57
<212> DNA
<213> Artificial sequence

<220><223> Primer
<400> 86
gtcttacgag gaagagttgg tgaaagatcc acgtgtggcc gccactatgg aaaacgc 57
<210> 87
<211> 41
<212> DNA
<213> Artificial sequence

<220><223> Primer
<400> 87
gggttatgcg ttcaagtatg gcgacattaa agacgtgggc g 41
<210> 88
<211> 47
<212> DNA
<213> Artificial sequence

<220><223> Primer
<400> 88
caccatcacc atcagatta cgatagtgtc acccatatta aattctc 47
<210> 89
<211> 54
<212> DNA
<213> Artificial sequence

<220><223> Primer
<400> 89
ggcggatccg gaggtggatc cggaaagata gaggagggt aactggtaat ctgg 54
<210> 90
<211> 25
<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 90

cctatagtga gtcgtattaa ttctg 25

<210> 91

<211> 25

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 91

cgaaattaat acgactcact atagg 25

<210> 92

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 92

tccggatcca cctccggatc cgccggaact agaattcgtc tgcgcgtctt tcagg 55

<210> 93

<211> 34

<212> PRT

<213> Artificial sequence

<220><223> Helical linker

<400> 93

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

1 5 10 15

Gln Ala Ala Glu Glu Leu Ala Asn Ala Lys Lys Leu Lys Glu Gln Leu

20 25 30

Glu Lys

<210> 94

<211> 43

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 94

ctttaagaag gagatataca tatgtcgtac taccatcacc atc 43

<210> 95

<211> 40

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 95

ccgctgcttc cggatccaat atgagcgtca cctttagttg 40

<210> 96

<211> 58

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 96

catattggat ccggaagcag cggcctgggtg ccgcgcggat cccatatgaa gccgctgc 58

<210> 97

<211> 45

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 97

gtggtggtgg tggtgctcga gttattattt cggcgggtatc ggttc 45

<210> 98

<211> 90

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 98

ctaaaggatga cgctcatatt ggatcccccg ccaacctgaa ggccctggag gcccagaagc 60

agaaggagca gagacaggcc gccgaggagc 90

<210> 99

<211> 89

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 99

cacggcacca cgcageggct tcatatggga tcccttctcc agctgctcct tcagcttctt 60

ggcgttggcc agctcctcgg cggcctgtc 89

<210> 100

<211> 47

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 100

caccatcacc atcagatta cgatagtgtc acccatatta aattctc 47

<210> 101

<211> 50

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 101

gtgaacaaag gcagtgggtga gtcgggatcc ggagctagca tgactggtgg 50

<210> 102

<211> 51

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 102

catcacgatg tgggcaccgg aaccttcccc ggatccctcg aggcttttcg g 51

<210> 103

<211> 71

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 103
ctaccaacc taaacgggt acaagtaaag gctttcatag actcgctaag ggatgacca 60
agccaaagcg c 71

<210> 104
<211> 81
<212> DNA
<213> Artificial sequence
<220><223> Primer

<400> 104
gttgaatatt tccaagtag cccaccctag ctcttgttg aactgttgt ctactcttt 60
gttgaatttg ttgtccacgc c 81

<210> 105
<211> 30
<212> DNA
<213> Artificial sequence
<220><223> Primer

<400> 105
gattacgaca tccaacgac cgaaaacctg 30

<210> 106
<211> 77
<212> DNA
<213> Artificial sequence
<220><223> Primer

<400> 106
gcctgaacga tatTTTTgaa ggcagaaaa ttgaatggca tgaaggcgat tacgacatcc 60

caacgaccga aaacctg 77

<210> 107
<211> 33
<212> DNA
<213> Artificial sequence
<220><223> Primer

<400> 107
gtgatggtga tggatggtg agtacgacat atg 33

<210> 108
 <211> 80
 <212> DNA
 <213> Artificial sequence
 <220><223> Primer
 <400> 108
 tgccattcaa ttttctgcgc ttcaaaaata tcgttcaggc cgctgccgtg atggtgatgg 60
 tgatggtagt acgacatatg 80
 <210>
 > 109
 <211> 23
 <212> PRT
 <213> Artificial sequence
 <220><223> RrgATag2
 <400> 109
 Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys His Tyr Ile Thr
 1 5 10 15
 Asn Glu Pro Ile Pro Pro Lys
 20
 <210> 110
 <211> 69
 <212> DNA
 <213> Artificial Sequence
 <220><223> RrgATag2
 <400> 110
 gatattccgg ctacatatga atttaccacac ggtaaacatt atatcaccaa tgaaccgata 60
 ccgccgaaa 69
 <210> 111
 <211> 19
 <212> PRT
 <213> Artificial Sequence
 <220><223> RrgATag2.0
 <400> 111
 Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys His Tyr Ile Thr

1 5 10 15

Asn Glu Pro

<210> 112

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> RrgATag2.0

<400> 112

gatattccgg ctacatatga atttaccac ggtaaacatt atatcaccaa tgaaccg 57

<210> 113

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> RrgATag2.1

<400> 113

Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys His Tyr Ile Thr

1 5 10 15

Asn Glu

<210> 114

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> RrgATag2.1

<400> 114

gatattccgg ctacatatga atttaccac ggtaaacatt atatcaccaa tgaa 54

<210> 115

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> RrgATag2.2

<400> 115

Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys His Tyr Ile Thr

1 5 10 15

Asn

<210> 116

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> RrgATag2.2

<400> 116

gatattccgg ctacatatga atttaccac ggtaaactt atatcacaa t 51

<210> 117

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> RrgATag2.3

<400> 117

Ala Thr Tyr Glu Phe Thr Asn Gly Lys His Tyr Ile Thr Asn Glu Pro

1 5 10 15

<210> 118

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> RrgATag2.3

<400> 118

gtacatatg aatttaccaa cggtaaactt tatatcacca atgaaccg 48

<210> 119

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> RrgATag2.4

<400> 119

Lys His Tyr Ile Thr Asn Glu Pro

1 5

<210> 120

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> RrgATag2.4

<400> 120

aaacattata tcaccaatga accg 24

<210> 121

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> RrgATag2.5

<400> 121

Gly Lys His Tyr Ile Thr Asn Glu Pro

1 5

<210> 122

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> RrgATag2.5

<400> 122

ggtaaacatt atatcaccaa tgaaccg 27

<210> 123

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> RrgATag2.6

<400> 123

Asn Gly Lys His Tyr Ile Thr Asn Glu Pro

1 5 10

<210> 124

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> RrgATag2.6

<400> 124

aacggtaaac attatatcac caatgaaccg 30

<210> 125

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> RrgATag2.7

<400> 125

Ile Val Pro Gln Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys

1 5 10 15

His Tyr Ile Thr Asn Glu Pro

20

<210> 126

<211> 69

<212> DNA

<213> Artificial Sequence

<220><223> RrgATag2.7

<400> 126

attgtaccgc aggatattcc ggctacatat gaatttacca acggtaaaca ttatatcacc 60

aatgaaccg 69