



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0105289  
(43) 공개일자 2013년09월25일

- (51) 국제특허분류(Int. Cl.)  
    *A61K 38/16* (2006.01)  *A61K 39/395* (2006.01)  
    *A61P 35/00* (2006.01)
- (21) 출원번호  10-2012-7030156
- (22) 출원일자(국제)  2011년05월20일  
    심사청구일자  없음
- (85) 번역문제출일자  2012년11월16일
- (86) 국제출원번호  PCT/EP2011/058307
- (87) 국제공개번호  WO 2011/144756  
    국제공개일자  2011년11월24일
- (30) 우선권주장  
    10163564.7  2010년05월21일  
    유럽특허청(EPO)(EP)  
    61/428,016  2010년12월29일  미국(US)
- (71) 출원인  
    **엑스엘-프로테인 게엠베하**  
    독일 85354 프라이징 리제 마이트너 스트라쎬 30
- (72) 발명자  
    **스케라 아르네**  
    독일 85354 프라이징 막스-레흐너-스트라쎬 19  
    **바인더 울리**  
    독일 85354 프라이징 앵거스트라쎬 7  
    **스클랍스키 마르틴**  
    독일 85354 프라이징 간젠플러스트라쎬 31에이
- (74) 대리인  
    **특허법인대평양**

전체 청구항 수 : 총 31 항

(54) 발명의 명칭 **생합성 프롤린 / 알라닌 랜덤 코일 폴리펩티드 및 이들의 용도**

(57) 요약

본 발명은 생합성 랜덤 코일 폴리펩티드 또는 생합성 랜덤 코일 폴리펩티드 절편 또는 생합성 콘주게이트에 관한 것으로, 상기 생합성 랜덤 코일 폴리펩티드, 상기 생합성 랜덤 코일 폴리펩티드 절편 또는 상기 생합성 콘주게이트는 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하며, 상기 아미노산 서열은 약 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다. 상기 약 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기는 이중성 폴리펩티드 또는 이중성 폴리펩티드 컨스트럭트의 성분(들)일 수 있다. 또한, 이러한 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 또는 상기 콘주게이트의 용도 및 사용 방법이 기재된다. 상기 용도에는 특히 의학 용도, 진단 용도 또는 식품 산업에서의 용도뿐만 아니라 종이 산업, 오일 회수 등과 같은 다른 산업용 용도가 포함될 수 있다. 특히 (i) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어지고, (ii) (a) 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 또는 폴리펩티드 (b) 소분자 약물로 이루어진 균으로부터 선택되는 약물을 포함하는 약물 콘주게이트가 제공된다. 또한, 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및/또는 생물학적 활성의 이중성 단백질을 인코딩하는 핵산 분자뿐만 아니라 상기 핵산 분자를 포함하는 벡터 및 세포가 개시된다. 또한, 본 발명은 본 발명의 화합물을 포함하는 조성물뿐만 아니라 본 발명의 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편, 생물학적 활성 단백질, 약물 콘주게이트 또는 핵산 분자, 벡터 및 세포의 특정 용도를 제공한다. 또한, 본 발명의 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편의 제조 방법 및/또는 수득 방법뿐만 아니라 본 발명의 생물학적 활성의 이중성 단백질, 및/또는 폴리펩티드 컨스트럭트 또는 약물 콘주게이트와 같은 콘주게이트의 제조 방법 및/또는 수득 방법이 제공된다.

대표도

```

gccGCTCCAGCTGCACCTGCTCCAGCAGCACCTGCTGCACCAGCTCCGGCTGCTCCTGCT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CGAGGTCGACGTGGACGAGGTCGTCGTGGACGACGTGGTCGAGGCCGACGAGGACGAcgg
AlaAlaProAlaAlaProAlaProAlaAlaProAlaAlaProAlaProAlaAlaProAla
    
```

## 특허청구의 범위

### 청구항 1

(i) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하며, 여기서 상기 아미노산 서열이 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어지는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편, 및

(ii) (a) 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 또는 폴리펩티드 및 (b) 소분자 약물로 이루어진 군으로부터 선택되는 약물을 포함하는 약물 컨주게이트.

### 청구항 2

청구항 1에 있어서,

상기 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편은 약 50개 내지 약 3000개 아미노산 잔기로 이루어진 아미노산 서열을 포함하는 약물 컨주게이트.

### 청구항 3

청구항 1 또는 청구항 2에 있어서,

상기 프롤린 잔기는 상기 아미노산 서열의 약 10% 초과 및 약 75% 미만으로 구성하는 약물 컨주게이트.

### 청구항 4

청구항 1 내지 청구항 3 중 어느 한 항에 있어서,

상기 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편은 복수의 아미노산 반복체를 포함하고, 상기 반복체는 프롤린 및 알라닌 잔기로 이루어지며 6개를 초과하는 연속 아미노산 잔기는 동일하지 않은 약물 컨주게이트.

### 청구항 5

청구항 1 내지 청구항 4 중 어느 한 항에 있어서,

상기 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편은 이들 서열의 전체 또는 일부로서 하기로 이루어진 군으로부터 선택되는 아미노산 서열 또는 이러한 서열의 환상 순열 버전 또는 다량체(들)를 포함하는 약물 컨주게이트:

AAPAAPAAPAAPAAPAAP(서열 번호 1);

AAPAAAPAAPAAPAAP(서열 번호 2);

AAAPAAAPAAAPAAAP(서열 번호 3);

AAPAAPAAPAAPAAPAAP(서열 번호 4);

APAAAPAPAAAPAPAAAP(서열 번호 5);

AAAPAAPAPPAAAPAAAPPA(서열 번호 6); 및

APAPAPAPAPAPAPAPAP(서열 번호 51).

### 청구항 6

청구항 1 내지 청구항 5 중 어느 한 항에 있어서,

생물학적 활성을 갖는 상기 폴리펩티드, 상기 생물학적 활성 단백질 또는 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 상기 폴리펩티드는 결합 단백질, 항체 단편, 사이토카인, 성장 인자, 호르몬 또는 효소로 이루어진 군으로부터 선택되는 약물 컨주게이트.

### 청구항 7

청구항 6에 있어서,

생물학적 활성을 갖는 상기 폴리펩티드가 결합 단백질이고, 상기 결합 분자는 항체, Fab 단편, F(ab')<sub>2</sub> 단편, CDR-유래 펩티드 모사체, 단일쇄 가변 단편(scFv), 도메인 항체, 렉틴, 면역글로불린 도메인, 피브로넥틴 도메인, 단백질 A 도메인, SH3 도메인, 안키린 반복 도메인, 및 리포칼린으로 이루어진 군으로부터 선택되는 약물 콘주게이트.

**청구항 8**

청구항 1 내지 청구항 7 중 어느 한 항에 있어서,

상기 생물학적 활성 단백질은 과립구 집락 자극 인자, 인간 성장 호르몬, 알파-인터페론, 베타-인터페론, 감마-인터페론, 종양 괴사 인자, 에리트로포이에틴, 응고 인자 VIII, gp120/gp160, 가용성 종양 괴사 인자 I 및 II 수용체, 레테플라아제, 엑센딘-4, 아나킨라, 인터루킨-2, 호중구 젤라티나아제 연관 리포칼린, 난포 자극 호르몬, 글루코세레브로시다아제, 티모신 알파 1, 글루카곤, 소마토스타틴, 아테노신 데아미나아제, 인터루킨 11, 응고 인자 VIIa, 응고 인자 IX, 헤마티드, 람다-인터페론, 렙틴, 인터루킨-22 수용체 서브유닛 알파(IL-22ra), 인터루킨-22, 히알루로니다아제, 섬유아세포 성장 인자 18, 섬유아세포 성장 인자 21, 글루카곤 유사 펩티드 1, 오스테오프로테게린, IL-18 결합 단백질, 성장 호르몬 방출 인자, 가용성 TACI 수용체, 트롬보스폰딘-1, 가용성 VEGF 수용체 Flt-1, 및 IL-4 뮤테인으로 이루어진 군으로부터 선택되는 약물 콘주게이트.

**청구항 9**

청구항 1 내지 청구항 8 중 어느 한 항에 있어서,

상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드는 상기 약물 콘주게이트의 증가된 생체 내 및/또는 시험관 내 안정성을 매개하는 약물 콘주게이트.

**청구항 10**

청구항 9에 있어서,

상기 증가된 생체 내 안정성은 상기 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편이 없는 대조 폴리펩티드 또는 대조 콘주게이트의 안정성과 비교할 때, 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하는 상기 약물 콘주게이트의 연장된 혈장 반감기이며, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진 약물 콘주게이트.

**청구항 11**

청구항 1 내지 청구항 10 중 어느 한 항에 있어서,

상기 소분자는 디곡시제닌, 플루오레신, 독소루비신, 칼리케아미신, 캄프토테신, 푸마길린, 텍사메타손, 겔다나마이신, 파클리탁셀, 도세탁셀, 이리노테칸, 시클로스포린, 부프레노르핀, 날트렉손, 날록손, 빈데신, 반코마이신, 리스페리돈, 아리피프라졸, 팔로노세트론, 그라니세트론, 시타라빈 NX1838, 류프롤리드, 고세렐린, 부세렐린, 옥트레오티드, 테두글루티드, 실렌지티드, 아바렐릭스, 엔푸비르티드, 그렐린, 알파 4 인테그린 저해제, 안티센스 핵산, 소간섭 RNA, 마이크로 RNA, 스테로이드, DNA 또는 RNA 앵타머 및 펩티드 및/또는 펩티드 모사체로 이루어진 군으로부터 선택되는 약물 콘주게이트.

**청구항 12**

청구항 1 내지 청구항 11 중 어느 한 항에 따른 약물 콘주게이트를 포함하는 조성물.

**청구항 13**

청구항 12에 있어서,

약학적으로 허용가능한 담체(들)를 선택적으로 추가로 포함하는 약학적 조성물, 또는 진단 조성물에 허용가능한 담체(들)를 선택적으로 추가로 포함하는 진단 조성물인 조성물.

**청구항 14**

청구항 1 내지 청구항 11 중 어느 한 항에서 정의된 약물 콘주게이트에 포함되는 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 분자 또는 청구항 6 내지 청구항 8 중 어느 한 항에서 정의된 생물학적 활성 단백질을 포함하고 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하는 단백질 콘주게이트를 인코딩하는 핵산 분자로서, 상기 아미노산 서열이 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진 핵산 분자.

**청구항 15**

청구항 1 내지 청구항 10 중 어느 한 항에서 정의된 약물 콘주게이트를 인코딩하는 핵산 분자로서,

- (i) 번역된 아미노산 및/또는 리더 서열을 인코딩하는 핵산 서열;
- (ii) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 서열로서, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어지는 핵산 서열;
- (iii) 생물학적 활성 단백질을 인코딩하거나 또는 생물학적 및/또는 치료적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 상기 폴리펩티드를 인코딩하는 핵산 서열; 및
- (iv) 번역 정지 코돈이거나 이를 나타내는 핵산 서열을 포함하는 핵산 분자.

**청구항 16**

청구항 15에 있어서,

- (ii) 및 (iii)에서 정의된 핵산 분자 부분 또는 절편은 상기 약물 콘주게이트를 인코딩하는 핵산 분자상에서 이들의 위치가 상호 교환되는 핵산 분자.

**청구항 17**

청구항 15 또는 청구항 16에 있어서,

- (i) 및 (ii)에서 정의된 부분 또는 절편간에 및/또는 (ii) 및 (iii)에서 정의된 부분 또는 절편간에, 프로테아제 및/또는 화학적 절단 부위 및/또는 인식 부위를 선택적으로 포함하는 핵산 분자.

**청구항 18**

프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산으로서, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어지며,

- (i) 번역된 아미노산 및/또는 리더 서열을 인코딩하는 핵산 서열;
- (ii) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 서열; 및
- (iii) 번역 정지 코돈이거나 이를 나타내는 핵산 서열을 포함하는 핵산 분자.

**청구항 19**

청구항 18에 있어서,

- (i) 및 (ii)간에 프로테아제 및/또는 화학적 절단 부위 및/또는 인식 부위를 선택적으로 포함하는 핵산 분자.

**청구항 20**

청구항 14 내지 청구항 19 중 어느 한 항에 따른 핵산을 포함하는 벡터.

**청구항 21**

청구항 14 내지 청구항 19 중 어느 한 항에 따른 핵산 또는 청구항 20에 따른 벡터를 포함하는 숙주 세포.

**청구항 22**

청구항 21에 있어서,

상기 숙주 세포는 진핵생물 숙주 세포인 숙주 세포.

**청구항 23**

청구항 22에 있어서,

상기 진핵생물 숙주 세포는 진균 또는 동물 세포인 숙주 세포.

**청구항 24**

청구항 1 내지 청구항 11 중 어느 한 항의 약물 콘주게이트에 포함되는 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편의 제조 방법, 상기 랜덤 코일 폴리펩티드 또는 상기 랜덤 코일 폴리펩티드 절편을 포함하는 생물학적 활성 단백질 또는 약물 또는 식품 콘주게이트의 제조 방법 및/또는 청구항 1 내지 청구항 8 중 어느 한 항에 정의된 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하며 상기 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 추가로 포함하는 폴리펩티드의 제조 방법으로서,

청구항 21 내지 청구항 23 중 어느 한 항에 따른 세포를 배양하는 단계, 배양물로부터 또는 상기 세포로부터 상기 랜덤 코일 폴리펩티드 또는 생물학적 활성 단백질 및/또는 상기 생물학적 활성 단백질 또는 상기 폴리펩티드를 분리하는 단계를 포함하는 방법.

**청구항 25**

호르몬 결핍-관련 장애, 자가-면역 질환, 증식성 장애, 암, 빈혈, 혈관신생 질환, 감염성/염증성 질환, 알러지성 장애, 혈전증, 심근경색, 재협착, 당뇨병, 불임, 고셔 질환, 만성 B형 간염, C형 간염, 저혈당, 말단비대, 아테노신 데아미나아제 결핍, 혈소판 감소증, 혈우병, 빈혈, 비만, 알츠하이머 질환, 지방이상증, 건선, 전이성 흑색종, 골관절염, 이상지질혈증, 류머티스성 관절염, 전신성 홍반성 루푸스, 다발성 경화증, 천식, 골다공증 및 재관류 손상 또는 다른 신장 질환의 치료를 위한, 청구항 1 내지 청구항 11 중 어느 한 항에 따른 약물 콘주게이트, 청구항 12 또는 청구항 13의 조성물, 청구항 14 내지 청구항 19 중 어느 한 항에 따른 핵산, 청구항 20에 따른 벡터, 청구항 21 내지 청구항 23 중 어느 한 항에 따른 세포, 또는 상기 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 포함하고 청구항 24의 방법에 의해 제조되는 생물학적 활성 단백질 또는 폴리펩티드.

**청구항 26**

상기 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편, 생물학적 활성 단백질, 또는 약물 콘주게이트의 증가된 생체 내 및/또는 시험관 내 안정성을 갖는 약제로서의 용도를 위한, 청구항 1 내지 청구항 11 중 어느 한 항에 따른 약물 콘주게이트, 청구항 12 또는 청구항 13의 조성물, 청구항 14 내지 청구항 19 중 어느 한 항에 따른 핵산, 청구항 20에 따른 벡터 또는 청구항 21 내지 청구항 24 중 어느 한 항에 따른 세포.

**청구항 27**

화장료, 화장 트리트먼트, 영양 또는 식품 기술 또는 종이 산업에서의, 청구항 1 내지 청구항 5 중 어느 한 항에서 정의되거나 청구항 14 내지 청구항 19 중 어느 한 항의 핵산 분자 또는 이들의 일부에 의해 인코딩되는 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편의 용도.

**청구항 28**

청구항 1 내지 청구항 11 중 어느 한 항에서 정의된 콘주게이트의 화장료, 화장 트리트먼트에 사용될 화합물, 식품 또는 음료, 또는 관심 산업의 콘주게이트의 제조 방법으로서,

청구항 21 내지 청구항 23 중 어느 한 항의 세포 배양, 및 상기 배양 또는 상기 세포로부터 상기 랜덤 코일 폴리펩티드, 상기 생물학적 활성 단백질 및/또는 생물학적 활성 또는 관심 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하고 상기 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 추가로 포함하는 상기 생물학적 활성 단백질 또는 상기 폴리펩티드를 분리하는 단계를 포함하는 방법.

**청구항 29**

청구항 28에 있어서,

상기 랜덤 코일 폴리펩티드, 상기 생물학적 활성 단백질 및/또는 생물학적 활성 또는 관심 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하고 상기 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 추가로 포함하는 상기 생물학적 활성 단백질 또는 상기 폴리펩티드는 성장 배지 또는 배양 배지, 세포 용해물, 세포막 분획, 상기 (숙주)세포의 원형질막으로부터 단리되는 방법.

**청구항 30**

청구항 28 또는 청구항 29에 있어서,

상기 랜덤 코일 폴리펩티드는 관심 분자에 연결되고/되거나 커플링되는 방법.

**청구항 31**

청구항 30에 있어서,

상기 연결 또는 상기 커플링은 화학적 연결 또는 커플링인 방법.

**명세서**

**기술분야**

[0001]

본 발명은 생합성 랜덤 코일 폴리펩티드 또는 생합성 랜덤 코일 폴리펩티드 절편 또는 콘주게이트에 관한 것으로, 상기 생합성 랜덤 코일 폴리펩티드 또는 생합성 랜덤 코일 폴리펩티드 절편 또는 콘주게이트는 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하며, 상기 아미노산 서열은 약 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다. 상기 약 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기는 이종성(heterologous) 폴리펩티드 또는 이종성 폴리펩티드 컨스트럭트의 성분(들)일 수 있다. 또한, 이러한 생합성 랜덤 코일 폴리펩티드, 상기 폴리펩티드 절편 또는 상기 콘주게이트의 용도 및 사용 방법이 기재된다. 상기 용도에는 특히 의학 용도, 진단 용도 또는 식품 산업에서의 용도뿐만 아니라 종이 산업, 오일 회수 등에서와 같은 다른 산업용 용도가 포함될 수 있다. 본 발명은 또한, 본원에서 제공되는 생합성 랜덤 코일 폴리펩티드 또는 생합성 랜덤 코일 폴리펩티드 절편 또는 콘주게이트의 특정 용도(들)에 관한 것으로, 상기 생합성 랜덤 코일 폴리펩티드 또는 생합성 랜덤 코일 폴리펩티드 절편 또는 콘주게이트는 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함한다. 본원에서 제공되는 생합성 랜덤 코일 폴리펩티드 또는 생합성 랜덤 코일 폴리펩티드 절편의 아미노산 서열은 약 50개 이상, 약 100개 이상, 약 150개 이상, 약 200개 이상, 약 250개 이상, 약 300개 이상, 약 350개 이상 또는 약 400개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다. 상기 약 50개 이상, 약 100개 이상, 약 150개 이상, 약 200개 이상, 약 250개 이상, 약 300개 이상, 약 350개 이상 또는 약 400개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기는 바람직하게는 (a) 이종성 폴리펩티드 또는 이종성 폴리펩티드 컨스트럭트의 성분이거나, 바람직하게는 (b) 콘주게이트, 예컨대 약물 콘주게이트, 식품 또는 화장품 성분 또는 첨가제와의 콘주게이트, 생물학적 활성 화합물과의 콘주게이트 또는 입체 활성 화합물과의 콘주게이트의 성분이다. 특히 이종성 단백질이 본원에서 제공되며, 여기서 이러한 단백질은 둘 이상의 도메인을 포함하고, 상기 둘 이상의 도메인의 제 1 도메인에는 생물학적 활성과 같은 활성을 갖고/갖거나 매개하는 아미노산 서열이 포함되며, 상기 둘 이상의 도메인의 제 2 도메인에는 본 발명의 생합성 랜덤 코일 폴리펩티드 또는 프롤린/알라닌 폴리펩티드 절편이 포함된다. 본 발명은 특히 (i) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편(상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다), 및 (ii) (a) 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 또는 폴리펩티드 및 (b) 소분자 약물로 이루어진 균으로부터 선택되는 약물을 포함하는 약물 콘주게이트에 관한 것이다. 본 발명의 추가 요지는 본원에서 제공되는 생합성 랜덤 코일 폴리펩티드 또는 프롤린/알라닌 폴리펩티드 절편 및 추가로, (a) 상기 생합성 랜덤 코일 폴리펩티드 또는 프롤린/알라닌 폴리펩티드 절편에 연결되고/되거나 커플링된 약학적으로 또는 의학적으로 유용한 분자(들), 예컨대 소분자, 펩티드 또는 생체거대분자(예컨대 단백질, 핵산, 탄수화물, 지질 소포) 등을 포함하는 약물 콘주게이트이다. 또한, 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및/또는 생물학적 활성의 이종성 단백질을 인코딩하는 핵산 분자뿐만 아니라 상기 핵산 분자를 포함하는 벡터 및 세포가 개시된다. 또한, 본원에 기재되는 본 발명의 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 해당 약물 또는 식품

콘주게이트, 즉 본원에서 정의된 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 식품 성분 또는 식품 첨가제를 포함하는 콘주게이트의 제조 방법이 개시된다. 또한, 화장품 성분 또는 첨가제 또는 생물학적 또는 입체 활성 화합물을 특히 포함하는 대응 콘주게이트(하나의 성분으로 본원에 개시되는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함함)가 개시된다. 또한, 본 발명은 본 발명의 화합물(즉 본원에 개시되는 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편 및 이를 인코딩하는 핵산 분자)을 포함하는 조성물뿐만 아니라 상기 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편, 상기 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 포함하는 생물학적 활성 단백질, 본 발명의 약물 콘주게이트, 식품 콘주게이트 또는 핵산 분자, 박터 및 세포의 특별 용도를 제공한다. 또한, 본 발명의 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편의 제조 방법 및/또는 수득 방법뿐만 아니라 본 발명의 생물학적 활성의 이중성 단백질, 및/또는 폴리펩티드 컨스트럭트 또는 약물 콘주게이트의 제조 방법 및/또는 수득 방법이 제공된다. 또한, 본원에서 정의되는 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 위한(또는 이를 포함하는 분자 및 콘주게이트를 위한) 의학적, 약학적뿐만 아니라 진단적 용도가 제공된다. 이러한 의학적 또는 약학적 용도에는 혈장 증량제 등으로서의 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편의 용도가 포함될 수 있다. 그러나, 본원에 제공되는 수단 및 방법은 약학적, 의학적 및 생물학적 용도에만 제한되는 것이 아니며 종이 산업, 오일 회수 등에서와 같은 다른 산업 분야에서도 채용될 수 있다.

### 배경 기술

- [0002] 신장 여과에 의한 혈액 순환으로부터의 신속한 제거는 소분자(작은 단백질 및 펩티드 포함)의 전형적인 특성이 다. 그러나 신장 사구체의 포어 크기를 초과하도록 겔보기 분자 크기를 확장함으로써 치료적 단백질의 혈장 반감기를 의학적으로 유용한 수일 범위로 확장시킬 수 있다. 이러한 효과를 달성하는 하나의 전략은 생물학적 체제와 합성 중합체 폴리에틸렌 글리콜(PEG)의 화학적 콘주게이션이다. 이는 몇몇 허가된 약물, 예를 들어 PEG-인터페론 알파2a(Pegasys®), PEG-G-CSF(Neulasta®) 및 최근의 PEG화 알파TNF-Fab 단편(Cimzia®)으로 이어졌다. 그럼에도 불구하고, 상기 "PEG화" 기술은 몇몇 단점을 갖는다: 임상 등급의 PEG 유도체는 고가이며, 이들의 재조합 단백질에 대한 공유 커플링에는 추가적 하단 처리 및 정제 단계가 필요하여 수율을 낮추고 비용을 상승시킨다. 또한, PEG는 생분해성이 아니어서 연속 처리시 신장 상피의 공포화(vacuolation)와 같은 부작용을 일으킬 수 있다; 예로 [Gaberer-Porekar (2008) Curr Opin Drug Discov Devel 11:242-50; Knop (2010) Angew Chem Int Ed Engl 49:6288-308 또는 Armstrong in: Veronese (Ed.), "PEGylated Protein Drugs: Basic Science and Clinical Applications"; Birkhaeuser Verlag, basel 2009] 참고.
- [0003] PEG 기술의 일부 단점을 극복하기 위해, 특정한 재조합 폴리펩티드 모사체가 당분야에 제공되었으며, 이들 중 일부는 천연 생성 아미노산 서열 또는 합성 아미노산 연장부(stretches)에 기반한다.
- [0004] 대부분의 천연 아미노산 서열은 폴딩된 입체형태(2차 구조)를 취하는 경향이 있거나 폴딩되지 않은 경우에는 통상 불용성이며 응집물을 형성하기 때문에, 생리학적 용액 중에서 이상적인 랜덤 사슬과 같이 거동하지 않는다. 사실상 폴리펩티드의 랜덤 사슬 거동을 조사하기 위한 대부분의 전통적 실험은 변성 조건, 즉 요소 또는 염화구아니디늄과 같은 화학적 변성제의 존재하에 수행되었다(예로 [Cantor (1980) Biophysical Chemistry. W.H. Freeman and Company, New York] 참고). 따라서 이러한 기술은 일반적으로 폴딩, 응집뿐만 아니라 비특이적 흡착에 저항하여, 폴딩된 치료적 단백질 도메인에 유전적으로 융합된 경우에도 생리학적 완충 조건 및 온도하에 안정한 랜덤 사슬을 제공하는 특정 아미노산 서열에 의존한다. 이러한 상황하에서, 상기 재조합 PEG 모사체는 이들의 분자 질량만을 기준으로 통상 예측되는 것보다 훨씬 더 큰 크기 증가를 부여하여 결국 신장 여과를 지연시키고 부착된 생물학적 체제의 혈장 반감기를 상당한 수준으로 효과적으로 연장할 수 있다.
- [0005] 그러나 많은 이들 기술은 추가적 독점권 및 단점을 갖는다.
- [0006] 예를 들어, 천연 생성 반복 아미노산 서열은 의학 과학 및 약학 접근법에서 이들의 유용성에 대해 평가되었다. 이들 접근법 중 하나는 트리파노소마 크루지(*Trypanosoma cruzi*)의 트랜스-시알리다아제에 관한 것이다. 이것은 680개 아미노산 잔기의 촉매 도메인에 이어 다양한 수의 12머 아미노산 반복체를 포함하는 "출혈 급성상 항원(SAPA)"으로 마무리되는 C-말단 반복 도메인을 함유한다. 천연 서열 DSSAHSTPSTPA를 가지는 13개의 친수성 및 (생리학적 pH에서) 음으로 하전된 대응 아미노산 반복체를 함유하는 트랜스-시알리다아제에 대한 마우스에서의 약동학(PK) 연구는 C-말단 반복 서열이 결실된 재조합 효소에 비해 5배 더 긴 혈장 반감기를 나타내었다 (Buscaglia (1999) Blood 93:2025-32). 동일한 트랜스-시알리다아제, 즉 그 76kDa 촉매 도메인을 트리파노소마 크루지 단백질 항원 13에서 발견되는 서열 EPKSA의 하전된 13개 아미노산 반복체에 융합 후 유사한 반감기



연장 효과가 관찰되었다. SAPA 및 항원 13 유래의 반복체는 모두 이러한 동종이량체성 효소의 C-말단으로 유전적 융합한 후 스킴스토소마 자포니쿰(*Schistosoma japonicum*)의 이종성 단백질 글루타치온 S-트랜스퍼라아제(GST)의 혈장 반감기를 7-8배 연장할 수 있었다(Buscaglia, 위 인용문 참고). 그러나 인간 병원체 유래의 이들 천연 생성 반복 아미노산 서열은 원리상 치료적 단백질의 약동학 최적화에 매력적으로 보일 수 있지만 면역원성이 높은 것으로 나타났다([Affranchino (1989), Mol Biochem Parasitol 34:221-8 또는 Buscaglia (1998), J Infect Dis 1998;177:431-6] 참고).

[0007] 또 다른 접근법은 젤라틴에 관한 것이다. 가수분해 및 변성된 동물 콜라겐인 젤라틴은 Gly-Xaa-Yaa 반복체의 긴 연장부를 함유하며, 여기서 Xaa 및 Yaa는 대부분 각각 프롤린 및 4-히드록시프롤린을 형성한다. 주로 자연적으로 배치된 라이신 측쇄의 ε-아미노기를 통한 젤라틴의 숙시닐화는 이 생체중합체의 친수성을 증가시키고 그 등전위점(pI)을 낮춘다. 개질 측쇄의 음으로 하전된 카르복실레이트기들간 분자 내 정전기적 반발이 분자를 다소 연장된 입체형태로 펼치는 것으로 여겨진다. 생성되는 부피 확장으로 숙시닐화 젤라틴은 인간에서 혈장 증량제로 사용하기 위한 거대분자가 되며, 특히 Volplex®(Beacon Pharmaceuticals Ltd) 또는 Gelofusine®(B. Braun Melsungen AG)으로 시판된다. 또한, 과립구-집락-자극 인자(G-CSF)의 인공적 젤라틴 유사 폴리펩티드로의 유전적 융합에 의해 반감기 연장 효과가 달성되었다(Huang (2010) Eur J Pharm Biopharm 74:435-41). 이를 위해, 천연 젤라틴의 모든 소수성 측쇄가 친수성 잔기로 교체되어 아미노산 G, P, E, Q, N, S, 및 K를 다양한 순서로 포함하는 116개 아미노산의 젤라틴 유사 단백질(GLK)이 얻어졌다. G-CSF는 그 N-말단이 상기 GLK 서열의 4개 사본과 함께 융합되어 피치아 파스토리스(*Pichia pastoris*)에서 분비되었다. 피치아 파스토리스는 GLK 융합 단백질을 위해 바람직한 제조 개체로 나타났다; 그러나 재조합 젤라틴 단편은 예를 들어 [Olsen (2003), Adv Drug Deliv Rev 55:1547-67]에 나타난 바와 같이 *E. coli*에서 낮은 수율로만 발견될 수 있다는 것이 알려져 있으므로, GLK가 다른 개체에서도 제조될 수 있는지가 결정되어야 한다.

[0008] 엘라스틴은 여러 조직에서 세포 외 기질의 성분이다. 친수성 Lys/Ala-풍부 도메인 및 소수성 엘라스토머 도메인의 반복 서열로 이루어지는 가용성 전구체 트로포엘라스틴으로부터 형성된다. 친수성 도메인 내 라이신 측쇄의 효소적 가교는 불용성 엘라스틴 형성으로 이어진다. 엘라스틴 유사 폴리펩티드(ELP)는 트로포엘라스틴의 소수성 도메인에서 유래되는 인공적으로 설계된 반복 아미노산 서열이다. ELP의 가장 일반적인 반복 서열 모티프는 V-P-G-X-G로, 여기서 "X"는 Pro를 제외한 임의의 아미노산일 수 있다(MacEwan (2010) Biopolymers 94:60-77; Kim (2010) Adv Drug Deliv Rev 62:1468-78). 적합한 ELP가 치료적 단백질에 융합되어 *E. coli*에서 제조될 수 있다. 결과적으로 주입 후 겔 유사 축적부위를 형성하는 ELP의 능력은 다른 비구조화 폴리펩티드와 다른 기전에 의하기는 하지만, 부착된 생물학적 체제의 생체 내 반감기를 크게 연장할 수 있다. 그러나 ELP 부착은 IL-1 유도 림프구 증식 생체분석에서 인터루킨-1 수용체 길항제에 대해 나타난 바와 같은 융합 파트너의 생체 활성을 방해할 수 있다(Shamji (2007) Arthritis Rheum. 11:3650-3661). 또한, ELP는 내인성 프로테아제, 예컨대 콜라게나아제에 의해 분해된다. 또한, 응집된 단백질은 일반적으로 면역원성이 되기 더 쉽다.

[0009] 추가 접근법은 폴리음이온 중합체의 사용에 관한 것이다. 예를 들어, 폴리글루타메이트(PG)가 암 치료를 위해 가용성이 낮은 세포독성 소분자 약물에 화학적으로 커플링 되었다. 대응 제품은 현재 임상 III상 연구 중인 파클리탁셀 약물 콘쥬게이트인 Opaxio™이다. 파클리탁셀 PG 콘쥬게이트의 반감기는 비개질 화합물에 비해 3 내지 14배 연장되었다(Singer (2005) J Control Release 109:120-6). 추가 융합 단백질, 예를 들어 그 N-말단에서 C-말단에 84개 잔기의 PG 꼬리를 수반하는 IFN-알파2 또는 175개의 연속 Glu 잔기의 연장부와 융합된 G-CSF는 *E. coli*의 세포질에서 가용성 상태로 제조되었다(WO2002/077036 참고). 효율적인 번역을 위해, N-말단 융합에는 리더 펩티드가 필요하며, 이것은 이후에 담배 에치 바이러스(Tabacco Etch Virus, TEV) 프로테아제 절단에 의해 제거되었다. G-CSF 및 INF α2의 폴리글루타메이트 융합은 세포 배양 분석에서 생활성을 나타내었다. 그러나 현재까지 이들 PG 융합물에 대한 약동학 데이터는 보고되지 않았다. 또한, PG 융합물의 높은 음전하는 인공적 정전기적 인력 또는 반발 효과로 인해 생체분자 상호작용(예로 표적 수용체 또는 가용성 인자의 결합)에 대해 일반적으로 불리하다.

[0010] WO 2006/081249에는 3 내지 6개 아미노산의 약 2 내지 500개 반복 단위를 갖는 폴리펩티드 서열이 기재되며, 여기서 G, N 또는 Q는 주요 성분을 나타내고, 소량 성분은 A, S, T, D 또는 E일 수 있다. 이러한 아미노산 조성은 진핵생물 발현계에서 Asn 측쇄의 N-연관 글리코실화를 위한 글리코실화 세쿠온 Asn-Xaa-Ser/Thr(여기서 Xaa는 Pro를 제외한 임의의 아미노산이다)의 통합을 허용한다. 부피가 큰 용매화 탄수화물 구조를 갖는 번역 후 개질을 포함하는 생성 융합 단백질의 증가된 거대분자 크기는 유전적으로 콘쥬게이션된 단백질의 약동학을 연장할 수 있다. 이러한 올리고당 부착("글리코엔지니어링")은 일반적으로 단백질 분해에 대한 감수성을 감소시키고 수력학적 부피를 증가시킬 수 있다(Sinclair (2005) J Pharm Sci 94:1626-35). 단점은 글리코실화된 생체거대분자



의 내재적 분자 이중성이 생물공학적 제조 및 품질 제어 동안 추가적 노력을 요한다는 것이다.

[0011] WO 2010/091122(및 WO 2007/103515) 및 [Schellenberger (2009) *Nat Biotechnol* 27:1186-90]에는 잔기 P, E, S, T, A 및 G를 둘러싸고 포함하는 비구조화 비반복 아미노산 중합체가 개시된다. 상기에서 추가 설명된 PSTAD 반복체와 마찬가지로 조성을 나타내는 상기 아미노산 세트는 응집을 일으킬 수 있고 HLA/MHC-II 매개 면역 반응을 유도할 수 있는 소수성 측쇄-특히 F, I, L, M, V 및 W-를 배제함으로써 바이오약제 개발에 적합한 큰 분자 크기를 갖는 용매화된 폴리펩티드를 얻기 위한 서열에 대해 체계적으로 스크리닝되었다. 또한, 가교 잠재성이 있는 Cys 잔기, 음으로 하전된 세포막과 상호작용할 수 있는 양이온성 아미노산 K, R 및 H, 그리고 잠재적으로 가수분해되기 쉬운 아미드 측쇄의 N 및 Q는 제외되었다(Schellenberger (2009) 위 인용문 참고). 녹색 형광 단백질(GFP)에 융합된 PESTAG 잔기 세트를 포함하는 비반복 서열을 인코딩하는 합성 유전자 라이브러리가 *E. coli*에서의 가용성 발현 수준에 대해 스크리닝되었으며, 생성 서브세트가 유전적 안정성, 단백질 용해도, 열안정성, 응집 경향 및 오염 프로필에 대해 추가 조사되었다. 마지막으로 216개 Ser 잔기(25.0몰%), 72개 Ala 잔기(8.3몰%) 및 각 Pro, Thr, Glu, 및 Gly의 144개 아미노산(16.7몰%)을 함유하는 864개 아미노산 서열을 GLP-1 수용체 작동제 엑센딘-4(E-XTEN) 및 몇몇 다른 생물학적 제제에 대한 융합에 대해 추가 평가하였다. 융합 단백질-전형적으로 이후 절단 제거되는 셀룰로오스 결합 도메인을 수반함-은 *E. coli*의 세포질에서 가용성 상태로 제조되고 단리되었다. 원형 광이색형(circular dichroism, CD) 분광측정에 의한 E-XTEN의 조사에서는 2차 구조의 부재가 드러난 반면, 크기 배제 크로마토그래피(SEC) 동안 융합 단백질은 84 kDa 단백질에 대해 예측되는 것보다 실질적으로 적은 보유를 나타내어 수력학적 부피가 증가하였음을 나타내었다(Schellenberg (2009) 위 인용문). PESTAG 폴리펩티드의 불규칙적 구조 및 연관된 수력학적 반경의 증가는 XTEN 서열에 걸쳐 분포된 높은 전체적 음전하를 수반하는 아미노산간에 정전기적 반발에 의해 선호될 수 있다(WO 2010/091122 참고). 그러나 추가 연구[Geething (2010) *PLoS One* 2010;5:e10175]에서는 XTEN이 그 치료적 융합 파트너의 역할을 감소시키는 것으로 나타났다. 세포 배양 분석에서, 글루카곤 XTEN 융합은 비개질 펩티드의 생활성의 단지 15%를 나타내었다. 인간 성장 호르몬(hGH)의 XTEN 융합에서는 수용체 친화도에 대해 더 큰 손실(17배 증가한 EC<sub>50</sub>)이 나타났다; WO 2010/144502 참고.

[0012] 또한, 가장 작으며 구조적으로 가장 단순한 아미노산인 글리신은 이론적 관점에 기준하여 입체형태적으로 가장 유연한 아미노산으로 간주되어 왔다; 예로 [Schulz GE, Schirmer RH. *Principles of Protein Structure*. Springer, New York 1979] 참고. 또한, 컴퓨터 시뮬레이션에서는 Gly 중합체에 2차 구조가 없으며, 용액 중에서 랜덤 코일을 형성할 수 있다는 것이 나타났다; [Shental-Bechor (2005) *Biophys J* 88:2391-402] 참고. 화학적 측면에서, 폴리글리신은 폴리에테르 PEG에 대해 일정한 유사성을 나타내는 선형 미분기 폴리아미드로, 둘 다 본질적으로 수소 결합 및 고도로 용매화된 극성기에 의해 규칙적으로 방해되는 짧은 반복 탄화수소 단위로 제조되며, 사슬을 따라 여러 회전 자유도를 갖는 일차원적인 거대분자이다. 따라서, 폴리글리신은 치료적 단백질의 혈장 반감기 연장에 대해 가장 단순한 유전적으로 인코딩할 수 있는 PEG 모사체를 형성할 수 있다. 이러한 개념이 각각 "단독 아미노산 중합체(HAP)" 또는 글리신 풍부 서열(GRS)의 형태로 채용되었다; [Schlappschy (2007) *Protein Eng Des Sel* 20:273-84]; WO 2007/103515 참고. 그러나, Gly의 화학적으로 합성된 순수 중합체는 낮은 수용성을 나타낸다는 것이 오랫동안 알려져 왔다; 특히 [Bamford CH 등, *Synthetic Polypeptides-Preparation, Structure, and Properties*. Academic Press, New York 1956] 참고. 따라서, 수소 결합 세린 알코올 측쇄(WO 2007/103515 뿐만 아니라 Schlappschy (2007) 위 인용문) 또는 추가로 음으로 하전된 글루타메이트 잔기(WO 2007/103515)를 도입함으로써 친수성을 증가시키려는 상이한 시도가 있었다. 융합 단백질의 도메인을 유연한 방식으로 연결하기 위한 (Gly<sub>4</sub>Ser)<sub>n</sub>의 조성을 갖는 펩티드 스페이서가 이미 당분야에 설명되었다는 것이 주지된다. 분석용 SEC에서 이러한 융합 단백질에 대해 상당히 증가된 수력학적 부피가 검출되었다. 200개 잔기의 HAP 버전의 경우에는 미융합 Fab 단편 대비 겔보기 크기 증가가 120%이었던 반면 실제 질량은 단지 29% 더 커서 폴리글리신 태그의 용매화된 랜덤 코일 구조로 인한 증강된 수력학적 부피의 효과를 드러냈다. 또한, CD 차 스펙트럼은 HAP 부분에 대해 불규칙한 2차 구조에 특징적인 것이었다. 마지막으로, 마우스에서 200개 잔기의 HAP을 수반하는 Fab 단편의 말단 혈장 반감기는 대략 3배 연장되었다. 이 효과는 중간 정도기는 해도 특정(특화된) 진단 용도, 예컨대 생체 내 조영에 적절할 수 있다; [Schlappschy (2007); 위 인용문] 참고. 불행하게도, 더 긴 (Gly<sub>4</sub>Ser)<sub>n</sub> 반복 서열을 갖는 융합 단백질의 제조는 응집물을 형성하는 경향 증가로 인해 실현 가능성이 더 적은 것으로 나타나, PEG 모사체와 같은 다소 순수한 글리신 중합체의 사용을 자연적으로 제한하였다.

[0013] WO 2008/155134에서는 Pro, Ala, 및 Ser(즉 PAS) 잔기의 적절한 혼합물을 갖는 서열이 이들의 특징적인 2차 구조 선호성을 상호 취소시켜 안정하게 불규칙화된 폴리펩티드가 생성됨이 개시된다. 그러나 WO 2008/155134에서는 또한 세린 및 알라닌(SA) 잔기만으로 이루어진 도메인, 즉 2가지 유형의 아미노산만을 포함하는 도메인을 갖

는 융합 단백질은 랜덤 코일을 형성하지 않고 그 대신에  $\beta$ -시트 구조를 형성한다고 언급한다.

[0014] 폴리펩티드의 화학적 합성은 널리 공지되어 있고 당분야에서 설명되어 왔다. Izuka는 프롤린을 함유하는 폴리펩티드의 화학적 합성을 개시한다(Izuka (1993), Bull. Chem. Soc. Jpn 66, 1269-1272 참고). 이러한 코폴리펩티드는 각각 글리신, L-알라닌, L- $\alpha$ -아미노부티르산(Abu), L-노르발린(Nva) 또는 L-류신, 및 프롤린의 랜덤 서열을 함유하며, 화학적 공중합에 의해 합성된다. Izuka는 이러한 코폴리펩티드가 주로 정의된 콜라겐 유사 입체형태를 갖는다고 개시한다. 나아가, 이 문헌에서는 프롤린 및 알라닌(또는 프롤린 및 L- $\alpha$ -아미노부티르산)의 코폴리펩티드가 수중에서 부분적으로 가용성인 반면 다른 코폴리펩티드는 완전히 불용성이었다고 기재한다. Izuka는 프롤린/알라닌 코폴리펩티드가 부분적인 불규칙화 입체형태를 가질 수 있다고 추정한다. Izuka는 랜덤 프롤린/알라닌 서열을 갖는 화학적으로 합성된 폴리펩티드가 주로 콜라겐 유사 입체형태, 즉 구조화된 입체형태로 생성된다고 강조한다.

**발명의 내용**

**해결하려는 과제**

[0015] 즉, 본 발명의 근원적인 기술적 문제는 실제 랜덤 코일 입체형태를 갖는 거대한 폴리펩티드를 제공하는 것이다. 상기 기술적 문제점은 특허청구범위에서 구체화되는 구현예의 제공에 의해 및 본원에서 제공되는 바에 따라 해결된다.

[0016] 따라서, 본 발명은 약 50개 이상의, 특히 약 100개 이상의, 특히 약 150개 이상의, 특히 약 200개 이상의, 특히 약 250개 이상의, 특히 약 300개 이상의, 특히 약 350개 이상의, 특히 약 400개 이상의 프롤린 및 알라닌 아미노산 잔기로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편의 제공 및 용도에 관한 것이다. 따라서 본 발명은 50개 이상의 아미노산 잔기의 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 제공하는 것에 관한 것으로, 상기 아미노산 서열은 프롤린 및 알라닌 아미노산 잔기만으로 이루어지며, 하나 이상의 프롤린 및 하나 이상의 알라닌을 포함한다. 본 발명은 또한 (i) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편(상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다), 및 (ii) (a) 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 또는 폴리펩티드 및 (b) 소분자 약물로 이루어진 군으로부터 선택되는 약물을 포함하는 약물 콘주게이트를 제공한다. 본원에서 제공되는 실제 랜덤 코일 입체형태를 갖는 폴리펩티드 및 실제 랜덤 코일 입체형태를 갖는 폴리펩티드 절편은 또한 화장료 용도의 맥락뿐만 아니라 식품 산업 및 음료 제조에서의 용도에서 유용하다. 실제 랜덤 코일 입체구조를 보이는 본원에서 제공되는 거대한 폴리펩티드는 단독으로 및 주로 프롤린(P, Pro) 및 알라닌(A, Ala) 잔기로 이루어지며, 50개 이상의 아미노산, 특히 약 100개 이상의, 특히 약 150개 이상의, 특히 약 200개 이상의, 특히 약 250개 이상의, 특히 약 300개 이상의, 특히 약 350개 이상의, 특히 약 400개 이상의 프롤린 및 알라닌 아미노산 잔기를 포함한다. 아미노산 P 및 A는 모두 본원에서 제공되는 실제 랜덤 코일 입체형태를 갖는 거대한 폴리펩티드 및 실제 랜덤 코일 입체형태를 갖는 폴리펩티드 절편에 존재해야 한다. 또한, 본원에 개시되는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 분자뿐만 아니라 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 생물학적 활성 단백질과 같이 (공유 연결된) 관심 단백질을 포함하는 약물 또는 식품 콘주게이트가 본원에 제공된다.

[0017] 본원에서 제공되는 약물 또는 식품 콘주게이트에 사용되기 위한, 그리고 약 50개 이상의, 약 100개 이상의, 약 150개 이상의, 약 200개 이상의, 약 250개 이상의, 약 300개 이상의, 약 350개 이상의, 약 400개 이상의 프롤린 (P) 및 알라닌(A) 아미노산 잔기로 이루어진 아미노산 서열을 포함하는 본원에 기재되는 생합성 랜덤 코일 폴리펩티드 또는 생합성 랜덤 코일 폴리펩티드 절편은 특히 이중성 맥락에서, 즉 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 약학적으로 또는 의학적으로 유용한 분자, 예컨대 소분자, 펩티드 또는 생체거대분자, 예컨대 단백질, 핵산, 탄수화물, 지질 소포 등을 포함하는 생물학적 활성의 이중성 단백질, 단백질 컨스트럭트 및/또는 약물 콘주게이트에서 사용된다. 첨부된 실시예에 예시되는 바와 같이, 본 발명자들은 본원에 정의되는 바와 같은 실제 랜덤 코일 폴리펩티드 및 생물학적 활성 단백질 또는 단백질 연장부로 이루어진 약물 콘주게이트뿐만 아니라 프롤린 및 알라닌 아미노산 잔기(즉 두 아미노산 P 및 A)만으로 이루어진 본원에 기재되는 랜덤 코일 폴리펩티드를 포함하고/하거나 연결된 소분자 또는 소분자 약물로 이루어진 약물 콘주게이트를 성공적으로 제공할 수 있었다.

[0018] 따라서, 본 발명은 특히 둘 이상의 도메인을 포함하는 생물학적 활성의 이중성 단백질을 제공하며, 여기서 (a)

상기 둘 이상의 도메인의 제 1 도메인은 상기 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열을 포함하며; (b) 상기 둘 이상의 도메인의 제 2 도메인은 약 50개 이상의, 약 100개 이상의, 약 150개 이상의, 약 200개 이상의, 약 250개 이상의, 약 300개 이상의, 약 350개 이상의, 약 400개 이상의 프롤린 및 알라닌 아미노산 잔기로 이루어진 아미노산 서열로 이루어진 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함한다. 본 발명에 따르면, 상기 "제 1 도메인" 및 상기 "제 2 도메인"은 천연 생성 코딩 핵산 서열, 예컨대 개방 해독들 등에서 추론되는 가설상의 단백질 또는 천연(즉 천연 생성) 단백질에 포함되지 않는다.

[0019] 또한, 본 발명은 약 50개 이상의, 약 100개 이상의, 약 150개 이상의, 약 200개 이상의, 약 250개 이상의, 약 300개 이상의, 약 350개 이상의, 약 400개 이상의 프롤린 및 알라닌 아미노산 잔기로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편에 콘주게이션된 약학적으로, 치료적으로 및/또는 의학적으로 유용한 분자(들), 예컨대 소분자(들), 펩티드(들) 또는 생체거대분자(들), 예컨대 단백질(들), 핵산(들), 탄수화물(들), 지질 소포(들) 등으로 이루어진 약물 콘주게이트를 제공한다. 또다시, 본원에 개시되는 콘주게이트의 맥락에서 "생물학적 활성"이라는 용어는 순수한 생물학적 분자에 제한되지 않고 의학적 활성, 치료적 활성, 약학적 활성 분자 등을 또한 포함한다는 것이 주지된다. 당분야 숙련자에게는 본원에서 제공되는 수단 및 방법이 약학적 및 의학적 용도에만 제한되는 것이 아니라 화장품, 식품, 음료 및 영양 기술, 오일 산업, 종이 산업 등을 비제한적으로 포함하는 광범위한 기술에도 채용될 수 있다는 것이 자명하다.

[0020] 화학적으로 합성된 코폴리펩티드(예컨대 Izuka, 위 인용문에 기재된 바와 같은)와는 대조적으로, 본원에서 제공되는 랜덤 코일 폴리펩티드는 생합성적으로 제조된다. 본원에서 사용되는 "생합성"이라는 용어는 생물공학적인 방법(화학적 합성에 대조적으로)의 수단에 의한 합성을 나타낸다. 이러한 생물공학적인 방법은 당분야에 널리 공지되어 있으며, 또한 본원에서 더 후술된다. 본 발명의 랜덤 코일 폴리펩티드의 생합성으로 정의된 서열의 프롤린 및 알라닌 잔기, 정의된 길이 및/또는 정의된 비의 프롤린 및 알라닌 잔기를 갖는 폴리펩티드를 제조할 수 있다. 나아가, 본 발명에 따라 제공되는 폴리펩티드는 실질적으로 순수하다, 즉 제조되는 폴리펩티드는 본질적으로 균일하며, 상기 특징(즉, 정의된 서열, 정의된 길이 및/또는 정의된 아미노산 비)을 공유한다. 약 50개 이상의, 특히 약 100개 이상의, 특히 약 150개 이상의, 특히 약 200개 이상의, 특히 약 250개 이상의, 특히 약 300개 이상의, 특히 약 350개 이상의, 특히 약 400개 이상의 프롤린 및 알라닌 아미노산 잔기로 이루어진 랜덤 코일 폴리펩티드는 본 발명에 따르면, 예를 들어 생물학적 활성의 이중성 폴리펩티드/폴리펩티드 컨스트럭트에 및/또는 약물 또는 식품 콘주게이트에 뿐만 아니라 비제한적으로 종이 산업, 오일 산업 등을 포함하는 다른 산업 분야에서 유용한 다른 콘주게이트에 포함된다.

**과제의 해결 수단**

[0021] 전체적으로, 본 발명의 폴리펩티드의 상기 특징은 폴리펩티드의 안정한 랜덤 코일이 형성되고 이러한 랜덤 코일 폴리펩티드가 놀랍고 유리한 특성을 갖도록 해준다. 예를 들어, 본 발명의 폴리펩티드는 수용액 중에서 완전 가용성이며, 증가된 수력학적 부피를 갖는다. 놀랍게도 본원에 정의되는 랜덤 코일 폴리펩티드는 또한 증가된 생체 내/시험관 내 안정성을 부여할 수 있다. 이는 특히 의학적 용도, 예를 들어 본 발명의 랜덤 코일 폴리펩티드를 포함하는 생물학적 활성 단백질 또는 약물 콘주게이트에 있어 중요하다. 그러나 본 발명의 랜덤 코일 폴리펩티드의 여러 유리한 특성은 이들의 의학적 분야뿐만 아니라 다른 분야, 예컨대 화장품/화장 트리트먼트에서 또는 영양 및 식품 기술 분야, 예컨대 유제품 산업 및 육류 가공에서의 사용을 가능케 한다. 식품 산업 등에서 유용한 콘주게이트의 예에는 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 본원에 개시되는 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 이러한 기술에 유용한 화합물, 예컨대 유화제로 사용되는 비이온성 계면활성제인 폴리옥시프로필렌 또는 폴리옥시에틸렌 중합체를 포함하는 콘주게이트가 있다. 또한, 생화학적 방법 및 기술적 공정, 예컨대 종이 제조, 오일 회수 등에서 본원에 정의된 바와 같은 생합성 랜덤 코일 폴리펩티드의 용도가 포함된다. 본원에서 제공되는 프롤린 및 알라닌 잔기만으로 이루어진 생합성 랜덤 코일 폴리펩티드의 (그리고 또한 상기 생합성, 실제 랜덤 코일 폴리펩티드를 포함하는 본원에 개시되는 콘주게이트 및 컨스트럭트, 예컨대 약물 또는 식품 콘주게이트/컨스트럭트의) 놀랍고 유리한 특징이 아래에 보다 자세하게 기재된다. 또한, 이러한 본 발명의 생합성 랜덤 코일 폴리펩티드를 채용하는 예시적 용도 및 수단과 방법이 아래에 제공된다. 또한, 상기 생합성 랜덤 코일 폴리펩티드뿐만 아니라 생물학적 활성의 이중성 폴리펩티드 또는 폴리펩티드 컨스트럭트 및 상기 랜덤 코일 폴리펩티드를 포함하는 본원에 개시된 콘주게이트 및 컨스트럭트, 예컨대 약물 컨스트럭트의 제조 방법 및 수단이 본원에 제공된다.

[0022] 본 발명의 맥락에서, 놀랍게도 균일한 조성의 프롤린-알라닌 중합체/폴리펩티드는 안정한 랜덤 코일 입체형태를



형성한다는 것이 밝혀졌다. 이는 생합성 프롤린/알라닌 (공)중합체/폴리펩티드의 랜덤 코일 구조가 원형 광이색형(CD) 분광측정에 의해 확인된 첨부된 실시예에서도 나타났다. 확립된 Chou-Fasman 방법(Chou and Fasman (1974), Biochemistry 13, 223-245)에서는 도 7에 나타난 바와 같이 프롤린 및 알라닌으로 구성된 중합체/폴리펩티드(또는 이들의 절편)의 100% α-나선형 2차 구조를 예측하였으므로, 이러한 생합성, 실제 랜덤 코일 폴리펩티드/중합체의 수득 및 사용은 놀라운 것이었다. 그러나 본원에서는 놀랍게도 균일한 조성을 갖는 프롤린-알라닌 중합체/폴리펩티드가 안정한 랜덤 코일 입체형태를 형성한다는 것을 발견하고 실험적으로 나타내었다. 이는 또한 프롤린/알라닌 (공)중합체/폴리펩티드의 랜덤 코일 구조가 원형 광이색형(CD) 분광측정 및 크기 배제 크로마토그래피(SEC)와 같은 실험적 기법에 의해 확인된 첨부된 실시예에서도 드러난다.

[0023] 본 발명의 폴리펩티드/중합체와는 대조적으로, 예를 들어 [Izuka (1993), 위 인용문]에 기재된 화학적으로 합성된 폴리펩티드는 임의의/정의되지 않은 확률적 서열 및 다양한 길이를 갖는다. 따라서, 화학적으로 합성된 폴리펩티드에는 다양한 프롤린/알라닌 비, 길이 등을 갖는 완전히 상이한 펩티드들의 혼합물이 포함된다. Izuka가 언급한 바와 같이, 이러한 혼합물의 화학적으로 합성된 폴리펩티드는 랜덤 코일을 형성하지 않으며(또는 부분적으로는 형성하지 않으며), 따라서 아래에 제공되고 본원에 기재되는 생합성 폴리펩티드의 어떠한 유리한 특성도 갖지 않는다. 따라서, 본 발명은 본원에 개시되는 본 발명의 생합성 랜덤 코일 폴리펩티드/중합체를 포함하는 조성물을 포함하며 이에 관한 것으로, 상기 생합성 랜덤 코일 폴리펩티드/중합체는 특히 프롤린 및 알라닌 잔기만을 포함하는 이들의 서열에 의해 정의된다. 하나의 특정 구현예에서, 본 발명은 하나의 성분으로 본원에 개시된 이러한 랜덤 코일 폴리펩티드/중합체를 포함하는 콘주게이트, 예컨대 약물 또는 식품 콘주게이트에 관한 것이다. 상기 조성물에 포함되는 이러한 본 발명의 생합성 랜덤 코일 폴리펩티드/중합체는 하나의 구현예에서 균일한 길이를 갖는다.

[0024] 앞에서 언급된 바와 같이, 프롤린 및 알라닌 잔기만으로 이루어진 본 발명의 생합성 랜덤 코일 폴리펩티드(또는 랜덤 코일 폴리펩티드 절편)는 놀랍게도 안정한 랜덤 코일 입체형태를 형성한다. 본원에서 사용되는 "랜덤 코일"이라는 용어는 일반적으로 아미노산 중합체/아미노산 서열/폴리펩티드를 포함하는 임의 입체형태의 중합체성 분자에 관한 것으로, 여기서 상기 중합체성 구조를 형성하는 개별 단량체성 성분은 본질적으로 인접 단량체성 성분을 향해 무작위로 배향되는 반면, 상기 인접 단량체성 성분에 여전히 화학적으로 결합되어 있다. 특히 "랜덤 코일 입체형태"를 취하고/갖고/형성하는 폴리펩티드, 아미노산 서열 또는 아미노산 중합체에는 실질적으로 정의된 2차 및 3차 구조가 없다. 본 발명의 폴리펩티드의 맥락에서, 중합체성 구조(즉 폴리펩티드/아미노산 서열)를 형성하는 단량체성 성분은 단일 아미노산, 예컨대 프롤린 및 알라닌 자체이거나 또는 펩티드 연장부, 예컨대 아래에 더 기재되고 정의되는 "아미노산 반복체"/"아미노산 카세트"/"카세트 반복체"/"구성 블록"/"모듈"(또는 이들의 단편)이다.

[0025] 폴리펩티드 랜덤 코일의 성질 및 이들의 실험적 확인 방법은 당분야 숙련자에게 공지되어 있으며, 과학 문헌 [Cantor (1980) Biophysical Chemistry, 2nd ed., W. H. Freeman and Company, New York; Creighton (1993) Proteins-Structures and Molecular Properties, 2nd ed., W. H. Freeman and Company, New York; Smith (1996) Fold Des 1:R95-R106]에 기재되어 있다. 본원에서 사용되는 "절편"이라는 용어는 본원에서 정의된 생합성 랜덤 코일 폴리펩티드의 부분을 나타내며, 여기서 상기 부분은 본원에 기재된 생합성 랜덤 코일 폴리펩티드의 내부 부분일 수 있다. 이러한 "절편"은, 예를 들어 본원에 정의된 바와 같은 하나(또는 그 이상)의 아미노산(들)이, 예컨대 본 발명의 폴리펩티드의 시작으로부터 및/또는 말단으로부터 결실된 생합성 랜덤 코일 폴리펩티드일 수 있다. 또한, 상기 "절편"은, 예를 들어 생물학적 활성 단백질을 갖는 융합 단백질의 더 거대한 단백질 또는 폴리펩티드로서 사용되거나 그 일부를 형성할 수 있다. 상기 "융합 단백질"은 또한 본 발명의 이중성, 생물학적 활성 폴리펩티드/단백질/폴리펩티드 컨스트럭트의 한 예일 수 있다. 본원에서 사용되는 "이중성"이라는 용어는 본원에서 아래에 정의된다.

[0026] 본 발명에서 제공되고 본 발명의 맥락에서 사용될 랜덤 코일 폴리펩티드(또는 이들의 랜덤 코일 절편)는, 예를 들어 수용액 중에서 또는 생리학적 조건에서 랜덤 코일 입체형태를 취한다/형성한다. "생리학적 조건"이라는 용어는 당분야에 공지되어 있으며, 단백질이 보통 이들의 천연 폴딩된 입체형태를 취하는 조건에 관한 것이다. 보다 구체적으로 "생리학적 조건"이라는 용어는 전형적으로 더 고등 형태의 생명, 특히 포유류, 가장 바람직하게는 인간에서 유효한 생물학적 파라미터에 관한 것이다. "생리학적 조건"이라는 용어는 이들이 포유류, 특히 인간의 체내에서(특히 체액에서) 보통 발견되는 바와 같은 생화학적 및 생물학적 파라미터에 관한 것일 수 있다. 상기 "생리학적 조건"은 건강한 신체에서 발견되는 해당 파라미터뿐만 아니라 질환 상태 또는 인간 환자에서 발견되는 파라미터에 관한 것일 수 있다. 예를 들어, 아픈 포유류 또는 인간 환자는 상기 포유류 또는 상기 인간이 열을 겪을 때 더 높지만 여전히 "생리학적" 온도 조건을 가질 수 있다. 단백질이 이들의 천연 입

체형태/상태를 취하는 "생리학적 조건"에 있어서, 가장 중요한 파라미터는 온도(인체의 경우 37°C), pH(인간 혈액의 경우 7.35~7.45), 삼투압 농도(280~300 mmol/kg H<sub>2</sub>O), 그리고 필요한 경우, 단백질 함량(66~85 g/혈청 l)이다. 그러나 당분야의 숙련자는 생리학적 조건에서 이러한 파라미터가 변할 수 있다는 것, 예컨대 온도, pH, 삼투압 농도, 및 단백질 함량이 주어진 신체 또는 조직액, 예컨대 혈액, 뇌척수액, 복강액 및 림프에서 다를 수 있다는 것을 인지한다[Klinke (2005) Physiologie, 5th ed., Georg Thieme Verlag, Stuttgart]. 예를 들어, 뇌척수액에서 삼투압 농도는 약 290 mmol/kg H<sub>2</sub>O일 수 있고 단백질 농도는 0.15 g/l 내지 0.45 g/l 일 수 있는 반면, 림프에서 pH는 약 7.4일 수 있고, 단백질 함량은 3 g/l 내지 5 g/l 일 수 있다. 폴리펩티드(또는 이들의 절편)/아미노산 서열이 하기 본원에서 기재되는 방법을 이용하는 실험적 조건 하에 랜덤 코일 입체형태를 형성하는지/취하는지를 결정할 때, 생물리학적 파라미터, 예컨대 온도, pH, 삼투압 농도 및 단백질 함량은 생체 내에서 보통 발견되는 생리학적 조건과 다를 수 있다. 1°C 내지 42°C, 또는 바람직하게는 4°C 내지 25°C의 온도는 시험관 내 생리학적 조건 하에서 단백질의 생물리학적 특성 및 생물학적 활성을 평가하고/하거나 확인하는데 유용한 것으로 간주될 수 있다.

[0027] 특히 실험적 설정(예를 들어 단백질 구조의 결정, 특히 CD 측정 및 당분야의 숙련자가 단백질/아미노산 연장부의 구조적 특성을 결정할 수 있도록 하는 다른 방법) 또는 약학적 조성물을 위한 완충액, 용매 및/또는 부형제에서의 일부 완충액은 시험관 내 "생리학적 용액"/"생리학적 조건"을 나타내는 것으로 간주된다. 상기 완충액의 예로는, 예컨대 인산염 완충 식염수(PBS: 115 mM NaCl, 4 mM H<sub>2</sub>PO<sub>4</sub>, 16 mM Na<sub>2</sub>HPO<sub>4</sub> pH 7.4), 트리스 완충액, 아세테이트 완충액, 시트레이트 완충액 또는 유사한 완충액, 예컨대 첨부된 실시예에서 사용되는 것들이 있다. 일반적으로, "생리학적 용액 조건"을 나타내는 완충액의 pH는 6.5 내지 8.5 범위, 바람직하게는 7.0 내지 8.0 범위, 가장 바람직하게는 7.2 내지 7.7 범위여야 하고, 삼투압 농도는 10 내지 1000 mmol/kg H<sub>2</sub>O 범위, 보다 바람직하게는 50 내지 500 mmol/kg H<sub>2</sub>O 범위 및 가장 바람직하게는 200 내지 350 mmol/kg H<sub>2</sub>O 범위여야 한다. 선택적으로, 생리학적 용액 조건을 나타내는 완충액의 단백질 함량은 0 내지 100 g/l 범위일 수 있고, 스스로 생물학적 활성을 갖는 단백질을 무시하고 전형적인 안정화 단백질, 예를 들어 인간 또는 소 혈청 알부민을 사용할 수 있다.

[0028] 본원에서는 폴리펩티드(또는 이들의 절편)가 생리학적 조건 하에서뿐만 아니라 보다 일반적으로 수용액 중에서 랜덤 코일 입체형태를 형성하는 것으로 나타났다. "수용액"이라는 용어는 당분야에 널리 공지되어 있다. "수용액"은 물(H<sub>2</sub>O) 함량이 약 20% 이상, 약 30% 이상, 약 40% 이상, 약 50% 이상, 약 60% 이상, 약 70% 이상, 약 80% 이상 또는 약 90% 이상 H<sub>2</sub>O(중량/중량)인 용액일 수 있다. 따라서, 본 발명의 폴리펩티드(또는 이들의 절편)는 더 넓은 범위의 온도, pH 값, 삼투압 농도 또는 단백질 함량을 갖는, 가능하게는 다른 혼화성 용매를 함유하는 수용액 중에서 또는 수성 분산액 중에서 랜덤 코일 입체형태를 형성할 수 있다. 이는 특히 의학적 치료 또는 생체 내 진단을 벗어난 랜덤 코일 폴리펩티드(또는 이들의 절편)의 용도, 예를 들어 화장품, 영양 또는 식품 기술에서 상관성이 있다.

[0029] 따라서, 본 발명의 맥락에서 본 발명의 프롤린/알라닌 생합성 폴리펩티드(또는 이들의 절편)의 랜덤 코일 입체형태가 약학적 조성물, 예컨대 액체 약학 제제/생물학적 제제 또는 동결건조된 약학적 조성물 중에서 유지되고/유지되거나 그 맥락에서 사용된다는 것이 또한 포함된다. 이는 특히, 본 발명의 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)를 포함하는 본원에서 제공되는 생물학적 활성의 이중성 단백질 또는 약물 콘주게이트의 맥락에서 중요하다. 바람직하게는, "생리학적 조건"은 해당 완충액 시스템, 용매 및/또는 부형제에서 사용된다. 그러나, 예를 들어 동결건조된 또는 건조된 조성물(예컨대 약학적 조성물/생물학적 제제)에서, 본원에서 제공되는 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)의 랜덤 코일 입체형태가 일시적으로 존재하지 않고/않거나 검출될 수 없다는 것이 포함된다. 그러나, 상기 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)는 해당 완충액/용액/부형제/용매에서 재구성 후 또는 신체에 투여 후 그 랜덤 코일을 다시 취할/형성할 것이다. 폴리펩티드(또는 이들의 절편)가 랜덤 코일 입체형태를 형성하는지/취하는지를 결정하기 위한 방법은 당분야에 공지되어 있다(Cantor (1980) 위 인용문; Creighton (1993) 위 인용문; Smith (1996) 위 인용문). 상기 방법에는 본원에서 아래에 예시되는 원형 광이색형(CD) 분광측정이 포함된다. CD 분광측정은 물질에 의한 우- 및 좌-원편광의 흡광도 차이를 측정하는 흡광도 분광측정 방법을 나타낸다. 단백질의 2차 구조는 대략 190 내지 250 nm의 파장으로 원자외선 스펙트럼을 이용하여 CD 분광측정으로 결정할 수 있다. 이러한 파장에서, α-나선, 평행 및 반-평행 β-시트, 및 랜덤 코일 입체형태 각각이 특징적 형태 및 크기의 CD 스펙트럼을 나타내므로, 폴리펩티드에서 통상 발견되는 상이한 2차 구조를 분석할 수 있다. 따라서, CD 분광측정을 이용하여 당업자는 쉽게 폴리펩티드(또는 이들의 절편)가 생리학적 조건에서 또는 수용액 중에서 랜덤 코일 입체형태를 형성하는지/취하는지를 결

정할 수 있다. 다른 확립된 생물리학적 방법에는 핵 자기 공명(NMR) 분광측정, 흡광 분광측정, 원적외선 및 라만 분광측정, 크기 배제 크로마토그래피, 분석용 원심분리 또는 동적/정적 광 산란을 통한 수력학적 부피의 측정뿐만 아니라 마찰 계수 또는 내재 점도의 측정이 포함된다(Cantor (1980) 위 인용문; Creighton (1993) 위 인용문; Smith (1996) 위 인용문).

[0030] 상기 실험적 방법에 부가하여, 단백질 2차 구조의 예측을 위한 이론적 방법이 설명되었다. 이러한 이론적 방법의 하나의 예는, 예를 들어 X-선 결정학으로 해결된 공지 단백질 구조에 근거한  $\alpha$ -나선,  $\beta$ -시트 및 회전에서 각 아미노산의 상대빈도 분석에 근거하는 Chou-Fasman 방법이다(Chou and Fasman, 위 인용문). 그러나 단백질 2차 구조의 이론적 예측은 신뢰할 수 없는 것으로 알려져 있다. 아래에서 본원에 예시되는 바와 같이, Chou-Fasman 방법에 따라  $\alpha$ -나선 2차 구조를 취할 것으로 예측되는 아미노산 서열이 실험적으로 랜덤 코일을 형성하는 것으로 나타났다. 따라서, 이론적 방법, 예컨대 Chou-Fasman 알고리즘은 첨부된 실시예 및 도면에서 예시되는 바와 같이 주어진 폴리펩티드가 랜덤 코일 입체형태를 취하는지 여부에 대해 제한된 예측값만을 가질 수 있다. 그럼에도 불구하고, 상술된 이론적 예측은 종종 주어진 폴리펩티드/아미노산 서열의 추정 2차 구조의 평가에서 종종 첫 번째 접근법이다. 랜덤 코일 구조의 이론적 예측은 주어진 폴리펩티드/아미노산 서열이 실제로 랜덤 코일 입체형태를 갖는지 여부를 상기 실험적 수단으로 확인하는 것이 중요할 수 있다는 것을 나타낸다.

[0031] 대부분의 아미노산, 특히 소수성 아미노산의 단독-중합체는 통상 수용액 중에서 불용성이다(Bamford (1956) *Synthetic Polypeptides-Preparation, Structure, and Properties*, 2nd ed., Academic Press, New York). 몇몇 친수성 아미노산의 단독-중합체는 2차 구조를 형성하는 것으로 알려져 있고, 예를 들어 Ala의 경우는  $\alpha$ -나선(Shental-Bechor (2005) *Biophys J* 88:2391-2402) 및 Ser의 경우는  $\beta$ -시트(Quadrifoglio (1968) *J Am Chem Soc* 90:2760-2765)인 반면, 가장 강직한 단독올리고펩티드인 폴리프롤린(Schimmel (1967) *Proc Natl Acad Sci USA* 58:52-59)은 수용액 중에서 II형 트랜스 나선을 형성한다(Cowan (1955) *Nature* 176:501-503).

[0032] 중합체 생물리학의 이론적 논리를 이용하면, 200개 아미노산 잔기 사슬의 랜덤 코일 직경은, 예를 들어 폴리-글

$$\sqrt{\langle r^2 \rangle_0} = l \cdot \sqrt{n \cdot C_\infty}$$

리신의 경우 약 75Å일 것이다-의 말단에서 말단까지 거리의 제곱의 평균 제곱근으로 계산되며, 폴리(Gly)에 대한  $n = 200$ 이고, 각각의  $C\alpha-C\alpha$  거리에 대한 회전가능 결합의 길이  $l = 3.8\text{\AA}$ 이고, '특징적 비'인  $C_\infty \sim 2.0$ 이다(Brant (1967) *J Mol Biol* 23:47-65; Creighton (1993) 위 인용문). 이러한 관계는 당분야의 숙련자가 랜덤 사슬 아미노산 중합체의 수력학적 부피가 (a) 더 긴 사슬 길이  $l$ 을 사용해서 또는 (b) 더 큰 특징적 비  $C_\infty$ 를 나타내는 아미노산을 사용해서 연장될 수 있다고 예측할 것임을 보여준다.  $C_\infty$ 은 분자적 랜덤 사슬의 내재적 강성도에 대한 척도이며, 대부분의 아미노산에 대해 일반 값 9를 갖는다(Brant (1967) 위 인용문). 측쇄가 없는 Gly과 또한 아미노산 Pro만이 상당히 더 작은 값을 나타낸다. 따라서 Gly 및 Pro(변성 조건하에서) 랜덤 코일 단백질의 크기를 감소시키는데 기여할 것으로 예측된다(Miller (1968) *Biochemistry* 7:3925-3935). 따라서 프롤린 잔기를 포함하는 아미노산 서열은 상대적으로 컴팩트한 수력학적 부피를 가질 것으로 예측된다. 그러나 이러한 교시와는 대조적으로, 프롤린 및 알라닌 잔기의 혼합물을 포함하는 본 발명의 아미노산 중합체/폴리펩티드의 수력학적 부피는 예측되는 수력학적 부피에 비해 분석용 겔 투과/크기 배제 크로마토그래피에 의해 결정되는 현저히 증가된 수력학적 부피를 갖는 것으로 본원에 나타난다. 사실상, 각각 단독으로는 정의된 2차 구조를 갖는 단독올리고펩티드를 형성하는 경향을 갖는 상기 두 아미노산(프롤린 및 알라닌)의 혼합물을 포함하는 폴리펩티드가 생리학적 조건하에서 랜덤 코일 입체형태를 취한다는 것은 놀랍다. 이러한 본 발명의 프롤린/알라닌 폴리펩티드는, 예를 들어 동일한 수의 Gly 잔기를 포함하는 단독-중합체에 비해 더 큰 수력학적 반경을 가지며, 생물학적 활성 단백질 또는 컨스트럭트, 즉 본 발명에 따른 생물학적 활성의 이종성 단백질 또는 약물 콘주게이트에 대해 더 우수한 용해도를 부여한다.

[0033] 상기에서 언급된 바와 같이, 본 발명의 생합성 랜덤 코일 프롤린/알라닌 폴리펩티드는 이들이 쉬운 수단 및 방법에 의해 정의된, 균일한 길이를 취할 수 있다는 점에서 화학적으로 합성된 폴리펩티드와 상이하다. 선행 기술은 펩티드의 길이의 관점에서 수많은 편차를 갖는 폴리펩티드의 혼합물/조성물을 제공하는 반면, 본 발명은 정의된 길이를 갖는 생합성 랜덤 코일 폴리펩티드의 혼합물/조성물을 제공할 수 있다. 바람직하게는, 본질적으로 상기 혼합물/조성물에 포함된 본 발명의 모든 폴리펩티드는 정의된 동일 길이를 가져서, 동일한 생화학적 특징을 공유한다. 이러한 균일한 조성물은 생합성 랜덤 코일 폴리펩티드를 채용할 수 있는 다양한 의학적, 화장품, 영양 용도에서 더 유리하다. 또한, 특히 의학적 또는 약학적 맥락에서, 50개 이상의, 특히 약 100개 이상의, 특히 약 150개 이상의, 특히 약 200개 이상의, 특히 약 250개 이상의, 특히 약 300개 이상의, 특히 약 350개 이상의, 특히 약 400개 이상의 프롤린 및 알라닌 아미노산 잔기로 이루어진 아미노산 서열을 포함하는 본원



에서 정의된 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편은 또한, 예를 들어 부상, 화상, 수술 등의 이후에 손상된 혈액 혈장 상황에 연관되고/되거나 관련된 장애의 예방, 완화 및/또는 치료에 사용될 수 있다. 따라서 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편의 하나의 의학적 용도는 혈장 증량제로서의 용도이다. 그러나 본 발명에 따르면 본원에 기재되는 약물 콘주게이트 및 이중성 폴리펩티드 또는 이중성 폴리펩티드 컨스트럭트가 손상된 혈액 혈장량 또는 혈액 혈장 함량에 관련된 장애 또는 손상된 혈액 부피에 관련된 장애의 의학적 또는 약학적 개입의 맥락에서 사용될 수 있음이 또한 주지된다.

[0034] 따라서, 본 발명은 하나의 구현예에서 특히 이중성 단백질/폴리펩티드/폴리펩티드 컨스트럭트 또는 약물 콘주게이트에 포함되는 경우, 약 50개 이상의 프롤린 및 알라닌 아미노산 잔기, 약 100개 이상의 프롤린 및 알라닌 아미노산 잔기, 약 150개 이상의 프롤린 및 알라닌 아미노산 잔기 또는 약 200개 이상의 프롤린 및 알라닌 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 폴리펩티드(또는 이들의 절편)에 관한 것이다. 본 발명은 또한 약 200개 이상의 프롤린 및 알라닌 아미노산 잔기, 보다 바람직하게는 약 300개 이상의 프롤린 및 알라닌 아미노산 잔기, 특히 바람직하게는 약 400개 이상의 프롤린 및 알라닌 아미노산 잔기, 보다 특히 바람직하게는 약 500개 이상의 프롤린 및 알라닌 아미노산 잔기 및 가장 바람직하게는 약 600개 이상의 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드에 관한 것이다. 랜덤 코일 입체형태를 형성하는 아미노산 서열은 최대 약 3000개의 프롤린 및 알라닌 아미노산 잔기, 최대 약 2000개의 프롤린 및 알라닌 아미노산 잔기, 최대 약 1500개의 프롤린 및 알라닌 아미노산 잔기, 최대 약 1200개의 프롤린 및 알라닌 아미노산 잔기, 최대 약 800개의 프롤린 및 알라닌 아미노산 잔기로 이루어질 수 있다. 따라서, 프롤린/알라닌 아미노산 서열 연장부는 약 50개, 약 100개, 약 150개, 약 200개, 약 250개, 약 300개, 약 350개, 약 400개, 약 500개, 약 600개, 약 700개, 약 800개, 약 900개 내지 약 3000개의 프롤린 및 알라닌 아미노산 잔기로 이루어질 수 있다. 특정 구현예에 있어서, 본 발명의 생합성 아미노산 서열은 약 200 내지 약 3000개의 프롤린 및 알라닌 잔기, 약 200 내지 약 2500개의 프롤린 및 알라닌 잔기, 약 200 내지 약 2000개의 프롤린 및 알라닌 잔기, 약 200 내지 약 1500개의 프롤린 및 알라닌 잔기, 약 200 내지 약 1000개의 프롤린 및 알라닌 잔기, 약 300 내지 약 3000개의 프롤린 및 알라닌 잔기, 약 300 내지 약 2500개의 프롤린 및 알라닌 잔기, 약 300 내지 약 2000개의 프롤린 및 알라닌 잔기, 약 300 내지 약 1500개의 프롤린 및 알라닌 잔기, 약 300 내지 약 1000개의 프롤린 및 알라닌 잔기, 약 400 내지 약 3000개의 프롤린 및 알라닌 잔기, 약 400 내지 약 2500개의 프롤린 및 알라닌 잔기, 약 400 내지 약 2000개의 프롤린 및 알라닌 잔기, 약 400 내지 약 1500개의 프롤린 및 알라닌 잔기, 약 400 내지 약 1000개의 프롤린 및 알라닌 잔기, 약 500 내지 약 3000개의 프롤린 및 알라닌 잔기, 약 500 내지 약 2500개의 프롤린 및 알라닌 잔기, 약 500 내지 약 2000개의 프롤린 및 알라닌 잔기, 약 500 내지 약 1500개의 프롤린 및 알라닌 잔기, 약 500 내지 약 1000개의 프롤린 및 알라닌 잔기, 약 600 내지 약 3000개의 프롤린 및 알라닌 잔기, 약 600 내지 약 2500개의 프롤린 및 알라닌 잔기, 약 600 내지 약 2000개의 프롤린 및 알라닌 잔기, 약 600 내지 약 1500개의 프롤린 및 알라닌 잔기, 약 600 내지 약 1000개의 프롤린 및 알라닌 잔기, 약 700 내지 약 3000개의 프롤린 및 알라닌 잔기, 약 700 내지 약 2500개의 프롤린 및 알라닌 잔기, 약 700 내지 약 2000개의 프롤린 및 알라닌 잔기, 약 700 내지 약 1500개의 프롤린 및 알라닌 잔기, 약 700 내지 약 1000개의 프롤린 및 알라닌 잔기, 약 800 내지 약 3000개의 프롤린 및 알라닌 잔기, 약 800 내지 약 2500개의 프롤린 및 알라닌 잔기, 약 800 내지 약 2000개의 프롤린 및 알라닌 잔기, 약 800 내지 약 1500개의 프롤린 및 알라닌 잔기, 약 800 내지 약 1000개의 프롤린 및 알라닌 잔기를 포함한다. 본 발명의 내용에서 자명하듯이, 더 거대한 생합성 아미노산 서열(본질적으로 프롤린 및 알라닌으로 이루어진 서열)도 본 발명의 범위 내에 속하며, 둘 이상의 도메인의 하나의 도메인으로 상기 생물학적 활성을 가지며/가지거나 매개하는 아미노산 서열을, 그리고 둘 이상의 도메인의 또 다른 도메인으로 약 50개 이상의 프롤린 및 알라닌 아미노산 잔기, 약 100개 이상의 프롤린 및 알라닌 아미노산 잔기, 약 150개 이상의 프롤린 및 알라닌 아미노산 잔기, 약 200개 이상의, 약 250개 이상의, 약 300개 이상의, 약 350개 이상의, 약 400개 이상의 프롤린 및 알라닌 아미노산 잔기로 이루어진 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하는 본원에 정의된 생물학적 활성 단백질 또는 단백질 컨스트럭트에 쉽게 채용될 수 있다. 이러한 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편은 이중성 단백질/단백질 컨스트럭트의 생합성 랜덤 코일 부분에 해당한다. 이러한 생합성 프롤린/알라닌 연장부는 최대 약 3000개의 프롤린 및 알라닌 아미노산 잔기로 이루어진다. 이러한 아미노산 서열(프롤린/알라닌 연장부)은 아래에서 더 설명되는 바와 같이 주요 또는 고유 잔기로서 프롤린 및 알라닌을 포함한다.

[0035] 본원에서 정의되는 프롤린(P) 및 알라닌(A) 아미노산 잔기만으로 이루어지며 랜덤 코일 입체형태를 형성하고/취하고/갖는 본원에서 정의된 생합성 아미노산 서열이 포함된다. 가장 단순한 경우에서, 생합성 폴리펩티드 또는 폴리펩티드 절편은 본원에서 정의된 바와 같은 랜덤 코일 입체형태를 갖는 아미노산 서열로 이루어진다. 그러

나 생합성 폴리펩티드(또는 이들의 절편)는 본원에 기재되는 랜덤 코일 입체형태를 형성하는/취하는/가지는 아미노산 서열에 부가하여, 랜덤 코일 입체형태의 형성에 기여하지 않거나 이들 자체에 랜덤 코일 입체형태를 형성할/취할/가질 수 없는 추가 아미노산 서열/아미노산 잔기를 포함할 수 있다. 본 발명의 요지에서 벗어나지 않고, 상기 생합성 폴리펩티드(또는 이들의 절편)는 또한 생합성 "랜덤 코일" 폴리펩티드 또는 폴리펩티드 절편이다. 추가 아미노산 서열/아미노산 잔기는, 예를 들어 링커로서 유용할 수 있다. 특히 생합성 랜덤 코일 폴리펩티드의 이량체, 삼량체, 즉 일반적 다량체가 본 발명의 맥락에서 또한 열거되며, 이러한 다량체는 랜덤 코일 입체형태를 형성하지 않는 아미노산 서열/잔기에 의해 연결될 수 있다. 이러한 랜덤 코일 폴리펩티드를 포함할 수 있는 단백질의 예로는 본원에 정의된 프롤린 및 알라닌 아미노산 잔기로 이루어진 랜덤 코일 폴리펩티드에 부가하여 생물학적 활성을 가지는/매개하는 또 다른 폴리펩티드를 포함할 수 있는 본원에 제공되는 생물학적 활성 단백질이 있다. 마찬가지로 상기 컨스트럭트는 본원에서 기재되는 이종성, 생물학적 활성 단백질 또는 폴리펩티드 컨스트럭트일 수 있다.

[0036] "약 50/100/150/200/300/400/500/600/700/800개 등 이상의 아미노산 잔기"라는 용어는 정확한 아미노산 잔기의 수에 제한되지 않고, 약 1~20%, 예컨대 10% 내지 20% 추가 잔기 또는 약 1~20%, 예컨대 약 10% 내지 20% 적은 잔기를 포함하는 아미노산 연장부를 포함한다. 예를 들어 "약 100개 이상의 아미노산 잔기"에는 또한 본 발명의 요지에서 벗어나지 않고 약 80 내지 100개 및 약 100 내지 120개 아미노산 잔기가 포함될 수 있다. 예를 들어 "약 200개 이상의 아미노산 잔기"에는 또한, 본 발명의 요지에서 벗어나지 않고 약 160 내지 200개 및 약 200 내지 240개 아미노산 잔기가 포함될 수 있다. 상기에서 본원에 주어진 정의 및 설명은 또한 필요한 변경을 가하여 "최대 약 3000/2000/1500/1200/800개 아미노산 잔기" 등이라는 용어에도 적용된다. 따라서 "약"이라는 용어는 더 긴 아미노산 서열(예컨대 최대 3000개 아미노산 잔기를 포함하거나 이들로 이루어진 아미노산 서열)의 맥락에서 정확한 수의 아미노산 잔기에 한정 또는 제한되지 않는다. 따라서 "최대 약 3000/2000/1500/1200/800개 아미노산 잔기"라는 용어는 또한 본 발명에서 벗어나지 않고 10% 내지 20% 추가 또는 10% 내지 20% 더 적은 잔기를 포함하는 아미노산 연장부를 포함할 수 있다.

[0037] 또한, 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편)는 아미노산 잔기, 특히 주성분인 프롤린 및 알라닌의 정의된 함량 또는 비를 특징으로 한다. 상기에서 언급된 바와 같이, 본 발명은 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편에 관한 것으로, 특히 이종성 생물학적 활성 단백질/단백질 컨스트럭트/폴리펩티드 또는 약물 콘주게이트에 포함되는 경우 상기 아미노산 서열은 약 50개 이상의, 약 100개 이상의, 약 150개 이상의, 약 200개 이상의 약 250개 이상의, 약 300개 이상의, 약 350개 이상의, 약 400개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다. 본원에서 사용되는 "단독으로"라는 용어는 바람직하게는 약 90% 이상 또는 약 95% 이상의 아미노산이 프롤린 및 알라닌이며, 프롤린 및 알라닌이 대다수를 구성하지만 유일한 아미노산 잔기는 아닐 수 있고, 즉 이러한 본 발명의 아미노산 서열이 반드시 100% 프롤린 및 알라닌 아미노산 연장부는 아니라는 것을 의미한다. 따라서 본 발명의 생합성 폴리펩티드/아미노산 서열은 또한 아미노산 서열이 랜덤 코일 입체형태를 형성하고/취하고/갖는 한 프롤린 및 알라닌 이외의 다른 아미노산을 소량 성분으로 포함할 수 있다. 이러한 랜덤 코일 입체형태는 본원에서 제공되는 수단 및 방법에 의해 쉽게 결정될 수 있다. 따라서, 또한 "단독으로"라는 용어의 맥락에서, 소량(약 10% 미만 또는 약 5% 미만)의 다른 아미노산 잔기가 포함될 수 있다. 상기 "다른" 소량 아미노산 잔기는 본원에서 아래에 정의된다.

[0038] 따라서, 본 발명은 하나의 구현예에서 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편)에 관한 것으로, 여기서 아미노산 서열은 주로 프롤린 및 알라닌으로 이루어지며, 프롤린 잔기는 아미노산 서열의 약 10% 초과 및 75% 미만을 차지한다. 알라닌 잔기는 상기 아미노산 서열(또는 아미노산 서열로 이루어진 경우 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편) 나머지의 25% 이상 내지 90%를 차지한다.

[0039] 바람직하게는, 아미노산 서열은 약 10% 초과, 바람직하게는 약 12% 초과, 더욱 바람직하게는 약 14% 초과, 특히 바람직하게는 약 18% 초과, 보다 특히 바람직하게는 약 20% 초과, 더욱 특히 바람직하게는 약 22%, 23% 또는 24% 초과 및 가장 바람직하게는 약 25% 초과인 프롤린 잔기를 포함한다. 아미노산 서열은 바람직하게는 약 75% 미만, 보다 바람직하게는 70%, 65%, 60%, 55% 또는 50% 미만의 프롤린 잔기를 포함하며, 더 낮은 값이 바람직하다. 더욱 바람직하게는, 아미노산 서열은 약 48%, 46%, 44%, 42% 미만의 프롤린 잔기를 포함한다. 약 41%, 40%, 39%, 38%, 37% 또는 36% 미만의 프롤린 잔기를 포함하는 아미노산 서열이 특히 바람직하며, 더 낮은 값이 바람직하다. 가장 바람직하게는, 아미노산 서열은 약 35% 미만의 프롤린 잔기를 포함한다; 또한 본원에서 하기에 제공되는 PA 컨스트럭트를 참고.

[0040] 반대로, 아미노산 서열은 바람직하게는 약 90% 미만, 보다 바람직하게는 88%, 86%, 84%, 82% 또는 80% 미만의

알라닌 잔기를 포함하며, 더 낮은 값이 바람직하다. 보다 바람직하게는, 아미노산 서열에는 약 79%, 78%, 77%, 76% 미만의 알라닌 잔기가 포함되며, 더 낮은 값이 바람직하다. 가장 바람직하게는, 아미노산 서열에는 약 75% 미만의 알라닌 잔기가 포함된다.

[0041] 약 25% 초과, 바람직하게는 약 30% 초과, 더욱 바람직하게는 약 35% 초과, 특히 바람직하게는 약 40% 초과, 보다 특히 바람직하게는 약 45% 또는 50% 초과, 더욱 특히 바람직하게는 약 52%, 54%, 56%, 58% 또는 59% 초과, 알라닌 잔기를 포함하는 아미노산 서열이 본원에서 또한 바람직하며, 더 높은 값이 바람직하다. 더욱 바람직하게는, 아미노산 서열은 약 60%, 61%, 62%, 63% 또는 64% 초과, 알라닌 잔기, 가장 바람직하게는 약 65% 초과, 알라닌 잔기를 포함한다.

[0042] 따라서, 랜덤 코일 폴리펩티드(또는 이들의 절편)는 약 25% 프롤린 잔기, 및 약 75% 알라닌 잔기로 이루어진 아미노산 서열을 포함할 수 있다. 대안적으로, 랜덤 코일 폴리펩티드(또는 이들의 절편)는 약 35% 프롤린 잔기 및 약 65% 알라닌 잔기로 이루어진 아미노산 서열을 포함할 수 있다. 상기 본원에서 사용되는 "약 X%"라는 용어는 정확한 수의 백분율에 제한되는 것이 아니라 10% 내지 20% 추가 또는 10% 내지 20% 더 적은 잔기의 값을 또한 포함한다. 예를 들어 10%라는 용어는 각각 11% 또는 12% 및 9% 및 8%에 관한 것일 수도 있다.

[0043] 그러나 상기에서 그리고 아래에서 본원에 더 상세히 언급되는 바와 같이, 상기 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편), 및 특히 아미노산 서열은 소량 성분으로 프롤린 및 알라닌과 상이한 추가적 아미노산을 또한 포함할 수 있다. 상기 본원에서 이미 논의된 바와 같이, 상기 소량 성분(들), 즉 프롤린 또는 알라닌 이외의 또 다른 아미노산(들)은 본 발명의 생합성 랜덤 코일 폴리펩티드/중합체의 약 10% 미만, 약 9% 미만, 약 8% 미만, 약 7% 미만, 약 6% 미만, 약 5% 미만, 약 4% 미만, 약 3% 미만 또는 약 2% 미만을 차지할 수 있다.

[0044] 당분야의 숙련자는 프롤린 및 알라닌 이외의 다른 잔기가 상기 아미노산 서열/폴리펩티드(폴리펩티드 절편)에서 소량 성분으로 포함되는 경우 아미노산 서열/폴리펩티드(또는 이들의 절편)가 또한 랜덤 코일 입체형태를 형성할 수 있다는 것을 인지한다. 본원에서 사용되는 "소량 성분"이라는 용어는 최대 5% 또는 최대 10% 아미노산 잔기가 본 발명의 생합성 랜덤 코일 폴리펩티드/본 발명의 중합체에서 프롤린 또는 알라닌과 상이하다는 것을 의미한다.-이는 랜덤 코일 폴리펩티드(또는 이들의 절편)에 포함된 100개 아미노산 중 최대 10개가 프롤린 및 알라닌과 상이할 수 있다, 바람직하게는 최대 8%, 즉 100개 아미노산 중 최대 8개가 프롤린 및 알라닌과 상이할 수 있고, 보다 바람직하게는 최대 6%, 즉 100개 아미노산 중 최대 6개가 프롤린 및 알라닌과 상이할 수 있다, 더욱 바람직하게는 최대 5%, 즉 100개 아미노산 중 최대 5개가 프롤린 및 알라닌과 상이할 수 있고, 특히 바람직하게는 최대 4%, 즉 100개 아미노산 중 최대 4개가 프롤린 및 알라닌과 상이할 수 있다, 보다 특히 바람직하게는 최대 3%, 즉 100개 아미노산 중 최대 3개가 프롤린 및 알라닌과 상이할 수 있고, 더욱 특히 바람직하게는 최대 2%, 즉 100개 아미노산 중 최대 2개가 프롤린 및 알라닌과 상이할 수 있으며, 가장 바람직하게는 최대 1%, 즉 100개 아미노산 중 최대 1개가 프롤린 및 알라닌과 상이할 수 있다는 것을 의미한다. 프롤린 및 알라닌과 상이한 상기 아미노산은 번역 후 개질된 아미노산 또는 비천연 아미노산을 포함하여, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Tyr, 및 Val으로 이루어진 군으로부터 선택될 수 있다(예로 [Budisa (2004) Angew Chem Int Ed Engl 43:6426-6463 또는 Young (2010) J Biol Chem 285:11039-11044] 참고). 생합성 랜덤 코일 폴리펩티드/컨스트럭트/중합체(또는 이들의 단편)의 "소량 성분"(즉 프롤린 및 알라닌 이외의 아미노산)이 "다른 아미노산"/"상이한 아미노산"으로 Ser(들)을 포함하는 경우, 상기 Ser 아미노산/Ser 아미노산들은 바람직하게는 이러한 (소량) 아미노산 잔기의 50% 미만, 보다 바람직하게는 40% 미만, 30% 미만, 20% 미만 또는 10% 미만을 형성한다. 가장 바람직한 구현예에서, 본원에서 기재되는 생합성 랜덤 코일 폴리펩티드/컨스트럭트/중합체 또는 (예로)본원에서 기재되는 융합 단백질의 랜덤 코일 폴리펩티드 부분은 세린 잔기(들)를 포함하지 않는다. 본원에서는 이러한 "소량" 아미노산(프롤린 및 알라닌 이외의 것)이 본원에서 제공되는 생합성 랜덤 코일 폴리펩티드/컨스트럭트/본원에 기재되는 중합체 또는 (예로)융합 단백질의 랜덤 코일 폴리펩티드 부분에 존재하지 않는 것이 일반적으로 바람직하다. 본 발명에 따르면, 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편)/아미노산 서열은 특히 프롤린 및 알라닌 아미노산 잔기만으로 이루어질 수 있다(즉 다른 아미노산 잔기가 랜덤 코일 폴리펩티드 또는 아미노산 서열에 존재하지 않는다).

[0045] 상기는 랜덤 코일 폴리펩티드(또는 이들의 절편)에 포함되는 아미노산 서열의 전체적 길이 및 프롤린/알라닌 함량에 관한 것이지만, 하기는 구체적인 예시적인 아미노산 서열(또는 이들의 단편)의 보다 상세한 내용에 관한 것이다.

[0046] 하나의 구현예에서, 예를 들어 수용액 중에서 또는 생리학적 조건 하에서 랜덤 코일 입체형태를 취하는 아미노산 서열/폴리펩티드(본원에서 정의되는 랜덤 코일 폴리펩티드 또는 이들의 절편)는 복수의 "아미노산 반복체"



/"아미노산 카세트"/"카세트 반복체"를 포함할 수 있으며, 상기 "아미노산 반복체"/"아미노산 카세트"/"카세트 반복체"/"구성 블록"/"모듈"(이러한 용어는 본원에서 상호교환적으로 사용된다)은 주로 또는 배타적으로 프롤린(Pro, P) 및 알라닌(Ala, A) 아미노산 잔기(본원에서 "PA" 또는 "AP"로 나타냄)로 이루어지며, 6개를 초과하는 연속 아미노산 잔기는 동일하지 않다. 예시적인 "구성 블록"은, 예컨대 "AP"이며, 이는 또한, 본 발명의 기능적 생합성 랜덤 코일 도메인의 첨부된 예시적 예에 제공되었다. 이러한 예시적 예는 APAPAPAPAPAPAPAPAP(서열 번호 51), 즉 "폴리 PA" "아미노산 반복체"/"아미노산 카세트"/"카세트 반복체"의 형태로 또한 제공되는 서열 "P1A1"이다. 바람직한 구현예에서, 상기 정의된 "아미노산 반복체"/"아미노산 카세트"/"카세트 반복체" 등을 포함하는 아미노산 서열/폴리펩티드에서 포함하는 5개를 초과하는 연속 아미노산 잔기는 동일하지 않다. 다른 대안적 구현예가 예시되는 개별 구성 블록의 맥락에서 하기에 본원에서 제공된다.

[0047] 본 발명에 따른 랜덤 코일 폴리펩티드(또는 이들의 절편)에서, 아미노산 반복체는 동일하거나 동일하지 않을 수 있다. 프롤린 및 알라닌 잔기로 이루어진 "아미노산 반복체", "구성 블록", "모듈", "반복체", "아미노산 카세트" 등의 비제한적 예가 하기에 본원에서 제공된다; 예로 서열 번호 1, 서열 번호 2, 서열 번호 3, 서열 번호 4, 서열 번호 5, 서열 번호 6 및 서열 번호 51 참고(동봉된 서열 목록에는 또한 상기 "반복체"/"모듈" 등을 인코딩하는 예시적 핵산 서열이 포함된다. 본원과 함께 출원된 상기 서열 목록에 첨부된 서열은 본 명세서 및 설명의 일부를 구성한다). 또한 이러한 서열의 (동일하고/하거나 동일하지 않은) 단편의 용도가 본원에 열거되며, 여기서 "단편"은 2개 이상의 아미노산을 포함하고 하나 이상의 프롤린 및/또는 알라닌, 바람직하게는 하나 이상의 프롤린 및 하나 이상의 알라닌을 포함한다. 랜덤 코일 폴리펩티드(또는 이들의 절편)의 생성을 위해 본 발명에 따라 채용될 이러한 서열의 "단편"은 상기 서열 번호 1, 2, 3, 4, 5, 6 및 51로 이루어진 군으로부터 선택되는 아미노산 서열의 3개 이상, 바람직하게는 4개 이상, 보다 바람직하게는 5개 이상, 더욱 바람직하게는 6개 이상, 더욱 더 바람직하게는 8개 이상, 특히 바람직하게는 10개 이상, 보다 특히 바람직하게는 12개 이상, 더욱 특히 바람직하게는 14개 이상, 더욱 더 특히 바람직하게는 16개 이상, 가장 바람직하게는 18개 이상의 연속 아미노산으로 이루어질 수 있다(여기에서 서열 번호 51은 예시적 "AP" 또는 "PA" 반복으로 이루어진다는 것이 주지된다).

[0048] 본원에서 주어진 교시에 근거하여, 당분야의 숙련자는, 예를 들어 수성 또는 생리학적 조건하에서 랜덤 코일 입체형태를 형성하며 주로 본원에서 정의되는 프롤린 및 알라닌으로 구성된 추가 아미노산 서열/폴리펩티드를 용이하게 생성할 수 있는 입장에 있다. 본원에서 정의된 랜덤 코일 폴리펩티드(또는 이들의 절편)의 구성 블록 또는 모듈로서 사용될 랜덤 코일 입체형태 형성 아미노산 서열/폴리펩티드의 추가 예는 특히 상기 나타낸 특정 "구성 블록", "중합체 카세트", 또는 "중합체 반복체"의 조합 및/또는 단편 또는 환상 순열 버전을 포함할 수 있다. 따라서, 랜덤 코일 폴리펩티드/아미노산 서열의 예시된 모듈/서열 단위/중합체 반복체/중합체 카세트는 또한 본 발명에 따른 추가 모듈/서열 단위/중합체 반복체/중합체 카세트를 형성하기 위해 새로 조합될 수 있는 개별 단편을 제공할 수 있다.

[0049] "모듈(들)", "서열 단위(들)", "중합체 반복체(들)", "중합체 카세트(들)" 및 "구성 블록(들)"이라는 용어는 본원에서 동의어로 사용되며, 본원에서 정의된 랜덤 코일 폴리펩티드(또는 이들의 절편)/아미노산 서열을 형성하는데 사용될 수 있는 개별 아미노산 연장부에 관한 것이다.

[0050] 아미노산 반복체(본 발명의 생합성 랜덤 코일 폴리펩티드의 "구성 블록" 등으로 사용됨)는 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30개 또는 그 이상의 아미노산 잔기로 이루어질 수 있으며, 각각의 반복체는 프롤린 및 알라닌 잔기(들)를 포함한다. 그러나 첨부된 서열 번호 51에 나타낸 바와 같이, 상기 "구성 블록"은 또한 본원에서 제공되는 2 아미노산 잔기 P 및 A, 즉 "PA" 또는 "AP"의 형태로만 이루어질 수도 있다. 하나의 구현예에서, 본 발명에 따른 아미노산 반복체는 50개 초과 아미노산 잔기를 포함하지 않는다. 그러나 당분야의 숙련자에게는, 예를 들어 상기 본 발명의 생합성 랜덤 코일 폴리펩티드/중합체가 예컨대 약 100개 초과 아미노산, 약 150개 초과 아미노산, 약 200개 초과 아미노산 등을 포함하는 경우, 상기 "반복체"가 50개 초과 아미노산 잔기도 포함할 수 있다는 것이 자명하다. 따라서, 이러한 "반복체"에 포함되는 아미노산 잔기의 최대량은 본원에서 제공되는 생합성 폴리펩티드(또는 이들의 절편)/중합체의 전체적 길이에 의해 조건 지워진다.

[0051] 그러나 상기 반복체 등을 포함하는 생합성 랜덤 코일 폴리펩티드/아미노산 서열은 바람직하게는 상기에서 본원에 정의되고 설명되는 바와 같은 전체적 길이 및/또는 프롤린/알라닌 함량을 가져야 하며, 즉 약 50개, 약 100개, 약 150개, 약 200개, 약 250개, 약 300개, 약 350개, 약 400개 내지 약 3000개의 아미노산으로 이루어지고/이루어지거나 약 10% 초과 및 약 75% 미만의 프롤린 잔기를 포함해야 한다. 이 맥락에서 상기 본원에서 주어

진 모든 정의가 필요한 변경을 가하여 여기에 적용된다.

[0052] 본원에서 상세히 논의되고 상기 본원에서 제공되는 바와 같이, 본 발명은 약학적, 의학적 및/또는 의약적 설정에 특히 유용한 생물학적 활성의 이중성 단백질(들) 또는 단백질 컨스트럭트(들)을 제공한다. 이러한 생물학적 활성의 이중성 단백질/단백질 컨스트럭트는 상기 둘 이상의 도메인의 하나 이상의 도메인으로서 단독으로 프롤린 및 알라닌 잔기로 이루어진 아미노산 서열을 포함하는 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하며, 상기 아미노산 서열은 약 50개, 약 100개, 약 150개, 약 200개, 약 250개, 약 300개, 약 350개, 약 400개 내지 약 3000개의 프롤린(Pro) 및 알라닌(Ala) 잔기로 이루어진다.

[0053] 본원에 개시되는 생물학적 활성의 이중성 단백질, 폴리펩티드 또는 단백질 컨스트럭트의 맥락에서, "이중성"이라는 용어는 상기 단백질, 폴리펩티드 또는 단백질 컨스트럭트 내의 둘 이상의 도메인에 관한 것이며, 상기 둘 이상의 도메인의 제 1 도메인은 정의된 생물학적 활성을 부여, 소유 및/또는 매개하고, 상기 둘 이상의 도메인의 제 2 도메인은 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 생합성 랜덤 코일 폴리펩티드를 포함하며, 상기 둘 이상의 도메인은 자연계에서 서로 선택적으로 연결되어 발견되지 않거나 자연계에 존재하는 하나의 코딩 핵산 서열(예컨대 개방 해독틀)에 의해 인코딩되지 않는다. 본원에서 제공되며 본 발명의 생물학적 활성의 이중성 단백질/단백질 컨스트럭트에서 채용되는 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 생합성 랜덤 코일 폴리펩티드/폴리펩티드 절편은 바람직하게는 (화학적으로) 추가 개질되지 않고, 예를 들어 이들은 바람직하게는 글리코실화되거나 히드록실화되지 않는다.

[0054] 특정한 천연 생성 단백질 또는 자연계에서 발견되는 서열 분석된 핵산 연장부에서 추정되는 가설상의 단백질은 상대적으로 높은(즉 평균 초과) 함량의 프롤린 및 알라닌을 포함하는 것으로 설명된다는 것이 주지된다. 예를 들어, 리슈마니아 주요 균주 Friedlin에 대한 가설상의 동종성 단백질이 설명되었다(Ivens (2005) Science 309, 436-442). 1514개의 코돈 삼중자를 포함하는 공개된 해독틀에는 240개 Ala, 132개 Pro, 34개 Lys 및 4개 Val 코돈으로 구성된 412개 삼중자의 연장부가 포함된다. 생리학적 완충 조건하에서 양으로 하전된 Lys 잔기는 이 서열에 걸쳐 거의 균일하게 분포되어 가용화 효과를 시사한다. 그러나 본원의 개시에서 자명하듯이, 이러한 천연 생성 핵산 분자 또는 개방 해독틀에서 추정되며 평균을 초과하는 높은 프롤린 및 알라닌 함량을 포함하는 가설상의 동종성 단백질은 본 발명의 일부가 아니다. 본 발명은 의학적/약학적 맥락에서 특히 유용하고, 단리되는 방식으로 천연 생성되지 않으며, 단독으로 프롤린 및 알라닌 잔기로 이루어진 아미노산 서열을 포함하고, 상기 아미노산 서열이 약 50개, 약 100개, 약 150개, 약 200개, 약 250개, 약 300개, 약 350개, 약 400개 내지 약 3000개의 프롤린(Pro) 및 알라닌(Ala) 잔기로 이루어진 다소 거대한 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편이 제공된다는 사실에 근거한다. 단리되는 방식으로 천연 생성되지 않는 본원에 기재되는 단리된 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편은 또한 약학적, 의학적 및/또는 의약적 설정에서 특히 유용한 본원에 개시되는 생물학적 활성의 이중성 단백질(들) 또는 단백질 컨스트럭트(들)에 또한 포함된다. 이러한 생물학적 활성의 이중성 단백질/단백질 컨스트럭트는 상기 둘 이상의 도메인의 하나 이상의 도메인으로서 단독으로 프롤린 및 알라닌 잔기로 이루어진 아미노산 서열을 포함하는 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하며, 상기 아미노산 서열은 약 50개, 약 100개, 약 150개, 약 200개, 약 250개, 약 300개, 약 350개, 약 400개 내지 약 3000개의 프롤린(Pro) 및 알라닌(Ala) 잔기로 이루어진다.

[0055] 또한, 아라비노갈락탄 단백질(AGP), Pro-풍부 단백질, 및 익스텐신은 히드록시프롤린(Hyp)-풍부 당단백질(HRG P)로 알려져 있으며 식물계에 걸쳐 발견되는 큰 당단백질 그룹에 속한다. Ala-Pro 반복체 (AP)<sub>51</sub>을 포함하는 이러한 한 AGP 모티브는 유전자삽입 애기장대(*Arabidopsis thaliana*)에서 N-말단 신호 서열 및 C-말단 녹색 형광 단백질을 갖는 합성 당모듈 펩티드로 발견되었으며, 프롤린 히드록실라아제 및 히드록시프롤린 잔기의 후속 O-글리코실화의 기질로서 연구되었다(Estevez (2006) Plant Physiol. 142, 458-470). 또 다시, 물 분자와 수소 결합을 형성할 수 있는 개시된 히드록실화되고/되거나 글리코실화된 Pro 측쇄는 가용화 효과를 갖는 것으로 나타난다.

[0056] 본원에 기재되는 "(적어도) 하나의 도메인으로서 단독으로 프롤린 및 알라닌 잔기로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 펩티드 절편을 포함하는 생물학적 활성 단백질 또는 단백질 컨스트럭트"는 보통 천연 생성되지 않고 따라서 "이중성"인 단백질 또는 단백질 컨스트럭트에 관한 것임이 주지된다. 또한, 그리고 식물계에서 설명되는 프롤린-풍부 서열과는 대조적으로, 본원에 기재되는 생합성 랜덤 코일 폴리펩티드/폴리펩티드 절편은 바람직하게는 화학적으로 개질되지 않는다, 즉 이들은 바람직하게는 글리코실화되거나 히드록실화되지 않는다.

[0057] 특히 본 발명의 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편의 장점은 이들의 내재적으로 친수성이지만

전하를 띄지 않는 특징에 있다. 따라서 본원에 기재되는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 연장부에서 "소량" 아미노산(프롤린 및 알라닌 이외의 것)으로서, 소수성 측쇄를 갖지 않는 아미노산, 예컨대 Val, Ile, Leu, Met, Phe, Tyr 또는 Trp 및/또는 하전된 측쇄를 갖지 않는 아미노산, 예컨대 Lys, Arg, Asp 또는 Glu이 바람직하다. 본 발명에 따르면, (상기 개별 아미노산이 그럼에도 불구하고 본 발명의 생합성 랜덤 코일 폴리펩티드/폴리펩티드 절편에 포함되는 경우) 본원에서 정의된 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편) 내에서 소수성 측쇄를 갖는 각각의 개별 아미노산, 예컨대 Val, Ile, Leu, Met, Phe, Tyr 또는 Trp, 및/또는 하전된 측쇄를 갖는 각각의 개별 아미노산, 예컨대 Lys, Arg, Asp, 또는 Glu의 전체적 함량은 8%, 7%, 6%, 5%, 4%, 3%, 2% 또는 1%를 초과하지 않는다는 것이 포함된다.

[0058] 본 발명의 생합성 랜덤 코일 폴리펩티드/아미노산 서열은 서열  $(Pro)_x-(Ala)_y$ 인 조합된 프롤린/알라닌 연장부를 포함하는 개별 블록의 동일계능 연속집합체를 포함할 수 있으며, 여기서 x는 1 내지 바람직하게는 15, 보다 바람직하게는 1 내지 10, 더욱 바람직하게는 1 내지 5의 정수 값을 가질 수 있고, y는 1 내지 바람직하게는 15, 보다 바람직하게는 1 내지 10, 더욱 바람직하게는 1 내지 5의 정수 값을 가질 수 있고, x 및 y는 후속 블록간에 다를 수 있다. 상기 x 및 y는 또한 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15인 정수일 수 있다.

[0059] 수용액 또는 생리학적 조건하에서 랜덤 코일 입체형태를 형성하는 아미노산 서열/폴리펩티드는 하기 화학식 I을 가질 수 있다:

[0060]  $[Pro_xAla_y]_n$  [화학식 I]

[0061] 식 중, x는 독립적으로 1 내지 5의 정수에서 선택된다. 또한, 각각의 n에 대해, y는 독립적으로 1 내지 5의 정수에서 선택된다. 마지막으로 n은 랜덤 코일 폴리펩티드(또는 이들의 절편)/아미노산 서열이 바람직하게는 약 50개 이상의, 약 100개 이상의, 약 150개 이상의, 약 200개 이상의, 약 250개 이상의, 약 300개 이상의, 약 350개 이상의, 약 400개 이상의 아미노산 잔기 및 최대 약 3000 아미노산 잔기로 이루어지는 한, 임의의 정수이다. 또한, 이러한 맥락에서, 상기 동일계능 연속집합체를 포함하거나 상기 화학식 I을 갖는 폴리펩티드/아미노산 서열은 바람직하게는 상기 정의되고 본원에서 설명되는 전체적 길이 및/또는 프롤린/알라닌 함량을 갖고, 즉 약 50개, 약 100개, 약 150개, 약 200개, 약 250개, 약 300개, 약 350개, 약 400개 내지 약 3000개의 아미노산으로 이루어지고/이루어지거나 약 10% 초과 및 약 75% 미만의 프롤린 잔기를 포함한다는 것이 주지된다. 또다시 이러한 맥락에서 상기 본원에서 주어진 모든 정의가 필요한 변경을 가하여 여기에 적용된다.

[0062] 본 발명은 또한 AAPAAPAPAAPAAPAPAAP(서열 번호 1); AAPAAPAPAAPAAPAPAAP(서열 번호 2); AAPAAPAPAAPAAPAPAAP(서열 번호 3,  $[Pro_1Ala_3]_5$ 에 대한 예임); AAPAAPAPAAPAAPAPAAP(서열 번호 4); AAPAAPAPAAPAAPAPAAP(서열 번호 5); AAPAAPAPAAPAAPAPAAP(서열 번호 6) 및 APAPAPAPAPAPAPAPAP(서열 번호 51,  $[Pro_1Ala_1]_{10}$ 에 대한 예임) 또는 이러한 서열의 전체 또는 일부로서 이러한 서열의 환상 순열 버전 또는 다량체(들)로 이루어진 군으로부터 선택되는 아미노산 연장부를 포함하는 랜덤 코일 폴리펩티드(폴리펩티드 절편(들))/아미노산 서열에 관한 것이다. 따라서 랜덤 코일 폴리펩티드(이들의 폴리펩티드 절편(들))/아미노산 서열은 생성 생합성 랜덤 코일 폴리펩티드가 단독으로 프롤린 및 알라닌 아미노산 잔기로 이루어지는 한, 아미노산 연장부 AAPAAPAPAAPAAPAPAAP(서열 번호 1), AAPAAPAPAAPAAPAPAAP(서열 번호 1); AAPAAPAPAAPAAPAPAAP(서열 번호 2); AAPAAPAPAAPAAPAPAAP(서열 번호 3); AAPAAPAPAAPAAPAPAAP(서열 번호 4); AAPAAPAPAAPAAPAPAAP(서열 번호 5); AAPAAPAPAAPAAPAPAAP(서열 번호 6) 및 APAPAPAPAPAPAPAPAP(서열 번호 51)뿐만 아니라 이러한 모티브의 조합 또는 상기 모티브의 단편 및 부분의 조합을 포함할 수 있으며, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다.

[0063] 또한, 상기 아미노산 서열의 환상 순열 버전을 본 발명에 따라 사용할 수 있다. 예컨대 AAPAAPAPAAPAAPAPAAP(서열 번호 1)의 예시적인 환상 순열 버전을, 예를 들어 첫 번째 알라닌을 제거하고 상기 서열의 말단에 또 다른 알라닌을 첨가함으로써 쉽게 생성할 수 있다. 이어서 서열 번호 1의 이러한 환상 순열 버전은 AAPAAPAPAAPAAPAPAAP(서열 번호 7)일 것이다. 나아가, 서열 번호 1의 환상 순열 버전의 비제한적 예는 다음과 같다:

[0064] PAAPAPAPAPAPAPAPAAA(서열 번호 8),

[0065] AAPAPAPAPAPAPAPAAA(서열 번호 9),



- [0066] APAPAAPAPAPAPAAAAPA(서열 번호 10),
- [0067] PAPAAPAPAPAPAAAAPAA(서열 번호 11),
- [0068] APAAPAPAPAPAAAAPAAP(서열 번호 12),
- [0069] PAAPAAPAPAPAAAAPAAPA(서열 번호 13),
- [0070] AAPAAPAPAPAAAAPAAPAP(서열 번호 14),
- [0071] APAAPAPAPAAAAPAAPAPA(서열 번호 15),
- [0072] PAAPAPAPAAAAPAAPAPAA(서열 번호 16) 등.
- [0073] 본 발명의 교시에 근거하여, 당업자는 서열 번호 2, 서열 번호 3, 서열 번호 4, 서열 번호 5, 서열 번호 6 및 서열 번호 51(상기 서열 번호 51은 전적으로 "AP" 반복체에 근거하며, 환상 순열 버전은 전적으로 "PA" 또는 "AP" 반복체/구성 블록에 근거할 수 있다)에 나타낸 아미노산 연장부의 해당 환상 순열 버전을 용이하게 생성할 수 있는 위치에 있다.
- [0074] 상기 환상 순열 버전은 또한 본원에 따라 사용될 수 있는 본원에서 제공되는 폴리펩티드/아미노산 서열의 추가 "모듈"/"구성 블록" 등의 예로서 간주될 수 있다.
- [0075] 본원에서 제공되는 아미노산 연장부의 "모듈" 및 (더 짧은) 단편 또는 환상 순열 버전이 또한 본원에서 정의된 랜덤 코일 폴리펩티드(또는 이들의 절편)/아미노산 서열에 대한 "모듈", "반복체" 및/또는 구성 블록으로서 사용될 수 있다는 것이 당분야의 숙련자에게 자명하다.
- [0076] 상기에 따라, 랜덤 코일 입체형태를 형성하는 랜덤 코일 폴리펩티드/아미노산 서열은 바람직하게는 서열 번호 1, 서열 번호 2, 서열 번호 3, 서열 번호 4, 서열 번호 5, 서열 번호 6 및 서열 번호 51에 나타낸, 임의의 상기 아미노산 연장부(또는 환상 순열 버전 또는 이들의 단편)의 다량체를 포함할 수 있다. 이러한 서열이 본 발명의 맥락을 제한하는 것은 전혀 아님이 주지된다.
- [0077] 또다시, 상기 아미노산 연장부(또는 이들의 단편), 환상 순열 버전(또는 이들의 단편)을 포함하는 폴리펩티드/아미노산 서열은 바람직하게는 상기 정의되고 본원에서 설명되는 전체적 길이 및/또는 프롤린/알라닌 함량을 가져야 한다, 즉 약 50개, 약 100개, 약 150개, 약 200개, 약 250개, 약 300개, 약 350개, 약 400개 내지 약 3000개 아미노산으로 이루어지고/이루어지거나 약 10% 초과 및 약 75% 미만의 프롤린 잔기를 포함해야 한다. 이러한 맥락에서 본원에서 상기 주어진 모든 정의가 필요한 변경을 가하여 여기에 적용된다. 또한, "단편"이라는 용어는 상기에서 정의되었다.
- [0078] 상기에서 언급된 바와 같이, 본 발명의 맥락에서 놀랍게도 본원에서 제공되는 생합성 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)/중합체가 상대적으로 큰 수력학적 부피를 특징으로 한다는 것이 발견되었다. 겔보기 크기로도 불리는 이러한 수력학적 부피는 분석용 겔 여과(크기 배제 크로마토그래피, SEC으로도 공지되어 있음)로 쉽게 결정될 수 있다. 바람직하게는 랜덤 코일 폴리펩티드(또는 이들의 절편)는 10 kDa 이상, 바람직하게는 25 kDa 이상, 보다 바람직하게는 50 kDa 이상, 더욱 바람직하게는 100 kDa 이상, 특히 바람직하게는 200 kDa 이상, 가장 바람직하게는 400 kDa 이상의 겔보기 크기를 갖는다. 당분야의 숙련자는 특정 단백질의 수력학적 부피를 쉽게 결정할 수 있다. 상기 방법에는 하기 본원에서 예시된 바와 같은 동적/정적 광 산란, 분석용 원심분리 또는 분석용 겔 여과가 포함될 수 있다. 분석용 겔 여과는 거대분자의 수력학적 부피를 측정하기 위한 당분야에 공지된 방법을 나타낸다. 대안적으로 구형 폴리펩티드의 수력학적 부피는 그 분자량으로 추정될 수 있다 (Creighton (1993) 위 인용문). 바람직하게는 약 50개 이상의, 약 100개 이상의, 약 150개 이상의, 약 200개 이상의, 약 250개 이상의, 약 300개 이상의, 약 350개 이상의, 약 400개 이상 내지 약 3000개의 프롤린 및 알라닌 아미노산 잔기로 이루어지며 랜덤 코일 입체형태를 갖는 본 발명의 폴리펩티드의 본원에서 기재되는 수력학적 부피는 분자량에 근거하여 해당하는 풀딩된, 구형 단백질에 대해 추정될 수력학적 부피에 관해 예상치 못하게 높은 값을 나타낸다. 하기는 특히 본 발명의 바람직한 구현예를 나타내는 상기 본원에서 기재되고 정의된 바와 같은 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하는 생물학적 활성의 이중성 단백질 또는 단백질 컨스트럭트에 관한 것이다. 이론에 구애되지 않고, 본 발명의 맥락에서 놀랍게도 본원에서 제공되고 단독으로 프롤린 및 알라닌으로 이루어진 생합성 랜덤 코일 폴리펩티드 연장부가 동일한 전체 아미노산 잔기수를 갖지만 단독으로 프롤린, 알라닌 및 세린으로 이루어진 대응 생합성 랜덤 코일 연장부에 비해서도 더 높은 수력학적 부피를 제공할 수 있다는 것이 발견되었다(WO 2008/155134에 제공됨).

- [0079] 일반적인 인간 혈장 단백질, 예컨대 혈청 알부민(HSA) 및 인간화된 항체를 포함하는 면역글로불린(Ig)은 전형적으로 2 내지 3주의 긴 반감기를 나타내며, 이는 엔도솜 재생을 일으키는 이들의 신생아 Fc 수용체(FcRn)와의 특이적 상호작용에 기인한다(Ghetie (2002) Immunol Res, 25:97-113). 대조적으로, 관심 약학 제제인 대부분의 다른 단백질, 특히 재조합 항체 단편, 호르몬, 인터페론 등은 빠른 (혈액) 제거를 겪는다. 이는 특히 그 크기가 신장 여과를 위한 역치 값인 약 70 kDa 미만인 단백질에 있어서 실제적인 것이다(Caliceti (2003) Adv Drug Deliv Rev 55:1261-1277). 이러한 경우, 미개질된 약학적 단백질의 혈장 반감기는 1시간보다 상당히 더 적어서 이를 대부분의 치료적 용도에 대해 본질적으로 쓸모없게 만들 수 있다. 지속되는 약리학적 작용 및 개선된 환자 순응도를 또한 얻기 위해-필요한 투여 간격을 수일 또는 심지어 수주로 연장하면서-생약학적 약물 개발 목적을 위해 몇몇 전략이 이미 확립되었다.
- [0080] 먼저, 천연 혈장 단백질의 재생 기전이 Ig의 Fc 부분을 갖는 융합 단백질, 예를 들어 TNF  $\alpha$  수용체의 세포외 도메인 및 인간 IgG1 간의 하이브리드인 Enbrel®(Goldenberg (1999) Clin Ther 21:75-87), 또는 혈청 알부민을 갖는 융합 단백질, 예를 들어 IFN $\alpha$ 와 HSA의 대응 융합물인 Albuferon®(알브인터페론 알파-2b, ZALBIN™, JOULFERON®)(Osborn (2002) J Pharmacol Exp Ther 303:540-548)의 제조에 채용되었다. 예를 들어 스트랩토코코스 단백질 G의 박테리아 알부민 결합 도메인(ABD)(Makrides (1996) J Pharmacol Exp Ther 277:534-542) 또는 파지 디스플레이 라이브러리로부터 HSA에 대해 선택된 펩티드(Dennis (2002) J Biol Chem, 277:35035-35043; Nguyen (2006) Protein Eng Des Sel 19:291-297)와의 융합을 통해 알부민 결합 기능이 부가된 생약학적 체제를 위한 캐리어 담체로 기능하는 600  $\mu$ M의 높은 혈장 농도를 갖는 알부민을 또한 간접적 방식으로 사용하였다.
- [0081] 두 번째로, 생약학적 체제의 혈장 반감기를 증가시키기 위한 기본적으로 상이한 방법론은 고도 용매화되고 생리학적으로 불활성인 화학적 중합체와의 콘주게이션으로, 치료적 단백질의 수력학적 반경을 대략 3~5 nm의 사구체 포어 크기를 넘어서서 효과적으로 확장시키는 것이다(Caliceti (2003) 위 인용문). Lys 측쇄를 통해 무작위적이거나(Clark (1996) J Biol Chem 271:21969-21977) 또는 특이적으로 도입된 Cys 잔기에 의한(Rosendahl (2005) BioProcess International:52-60) 생화학적으로 온화한 조건하 폴리에틸렌 글리콜(PEG)의 활성화된 유도체와의 공유 커플링은 중등 정도로 성공적이며, 현재 몇몇 허가된 약물에 적용되고 있다. 해당 장점은 특히 특이적인 약리학적 활성을 보유한 작은 단백질과 함께, 예를 들어 화학적으로 PEG화된 재조합 IFN $\alpha$ -2a인 Pegasys®에서 달성되었다(Harris (2003) Nat Rev Drug Discov, 2:214-221; Walsh (2003) Nat Biotechnol 21:865-870).
- [0082] 그러나 생물학적 활성 단백질과 합성 중합체의 화학적 커플링은 생약학적 체제의 개발 및 제조에 대해 단점을 갖는다. 적합한 PEG 유도체는 특히 고순도가 필요하기 때문에 고가이며, 이들의 재조합 단백질과의 콘주게이션에는 추가적 시험관 내 처리 및 정제 단계가 필요하여, 수율을 낮추고 제조 비용을 높인다. 사실상, PEG는 종종 알데히드 및 퍼옥시드로 오염되며(Ray (1985) Anal Biochem 146:307-312) 내재적으로 산소의 존재하에 저장 시 화학적으로 분해되기 쉽다. 또한, 치료적 단백질의 약학적 기능은 그 생화학적 활성 부위 근처의 아미노산 측쇄가 PEG화 공정에 의해 개질되는 경우 방해될 수 있다. 또한, 합성 중합체와의 화학적 커플링은 통상 생체 내 활성의 실질적 편차를 나타낼 수 있는 이중성 분자 혼합물을 생성한다.
- [0083] 세 번째로, 혈장 반감기를 연장하기 위해 새로운 N-연관 글리코실화 공통 서열이 도입되는 생물학적 활성 단백질의 글리코실화 유사체의 사용이 제안되었다; [WO 02/02597; Perlman (2003) J Clin Endocrinol Metab 88:2327-2335; 또는 Elliott (2003) Nat Biotechnol 21:414-420] 참고. 그러나 기재되는 당으로 엔지니어링된 단백질은 변형된 생체 내 활성을 나타내어, 새로운 탄수화물 측쇄가 엔지니어링된 단백질의 생물학적 활성에 영향을 미친다는 것을 나타낸다. 또한, 추가적 탄수화물 측쇄는 생성 생물학적 활성 분자의 항원성을 증가시킬 수 있어서 실질적인 안전성 우려사항이 제기된다. 또한, 트리파노소마 크루지 유래 인공적 반복 서열 PSTAD를 포함하는 융합 단백질은 트랜스-시알리다아제의 현장 반감기 연장을 유도하는 것으로 보고되었다(Alvarez (2004) JBC 279:3375-3381). 그러나 이러한 트리파노소마 크루지 유래 반복체는 체액성 면역 반응을 유도하는 것으로 보고되었다(Alvarez (2004) 위 인용문). 따라서, 생물학적 활성 단백질의 작용을 연장하기 위한 대안적 전략이 필요하다.
- [0084] 본원에서 개시되고 본 발명에 따라 단독으로 프롤린 및 알라닌으로 이루어진 생합성 아미노산 서열/폴리펩티드는 놀랍게도 특히 생리학 조건하에서 랜덤 코일 입체형태를 취하는 것으로 밝혀졌다. 따라서 이들은 본원에서 하기에 정의된 생물학적 활성 단백질(들)/폴리펩티드(들)의 "제 2 도메인", 즉 생리학 조건하에서 랜덤 코일 입체형태를 형성하는 폴리펩티드 연장부를 포함하며 생물학적 활성("기능적") 단백질(들) 또는 폴리펩티드(들)를 제공하여 증가된 생체 내 및/또는 시험관 내 안정성, 특히 증가된 혈장 반감기를 매개하기에 유리한 분자들이다. 상기 랜덤 코일 도메인에 융합된 기능적 단백질의 수력학적 부피는 본원에서 언급되는 표준 방법을

이용하여 추정할 수 있는 바와 같이 현저하게 증가된다. 랜덤 코일 도메인은 생물학적 활성 단백질의 제 1 도메인의 생물학적 활성을 방해하지 않는 것으로 여겨지므로, 이것이 융합된 기능적 관심 단백질에 의해 매개되는 생물학적 활성이 본질적으로 보존된다. 또한, 본원에 개시된 바와 같은 랜덤 코일 도메인을 형성하는 아미노산 중합체/폴리펩티드는 특히 혈액 혈장 중에서의 단백질해, 면역원성, 등전위점/정전기적 거동, 세포 표면 수용체에 대한 결합뿐만 아니라 내재화에 있어서 생물학적으로 상당히 불활성인 것으로 여겨지지만 여전히 생분해성이어서, 합성 중합체, 예컨대 PEG에 비해 뚜렷한 장점을 제공한다.

- [0085] 상기에 따르면, 본 발명은 본원에 기재되는 생합성 랜덤 코일 폴리펩티드를 포함하는 생물학적 활성 단백질에 관한 것이다. 본원에서 기재되는 생합성 랜덤 코일 폴리펩티드를 포함하는 상기 생물학적 활성 단백질/단백질 컨스트럭트는 이중성 생물학적 활성 단백질/단백질 컨스트럭트이다. 특히, 둘 이상의 도메인을 포함하거나 이로써 이루어지는 생물학적 활성의 이중성 단백질(들)이 또한 본원에 개시되며, 여기서
- [0086] (a) 상기 둘 이상의 도메인의 제 1 도메인은 상기 생물학적 활성을 가지고/가지거나 매개하는 아미노산 서열을 포함하거나 이로써 이루어지며;
- [0087] (b) 상기 둘 이상의 도메인의 제 2 도메인은 본원에 기재되고 정의되는 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하거나 이로써 이루어진다.
- [0088] 본 발명에 따르면 "제 1 도메인" 및 상기 "제 2 도메인"은 동일한 단백질 내에서 천연 생성되지 않거나 자연계에서 발견되는 코딩 핵산 서열(예컨대 개방 해독틀)에 의해 인코딩되는 동일한 가설상 단백질의 부분으로 예측되지 않는 단백질 연장부에 관한 것이 주지된다.
- [0089] 랜덤 코일 폴리펩티드 또는 이들의 폴리펩티드 절편의 맥락에서 상기 본원에서 주어진 정의 및 설명이 상기 랜덤 코일 폴리펩티드(또는 이들의 폴리펩티드 절편(들))를 포함하는 생물학적 활성 단백질의 맥락에서 필요한 변경을 가하여 적용된다.
- [0090] 바람직하게는 상기 랜덤 코일 입체형태는 상기 생물학적 활성 단백질의 증가된 생체 내 및/또는 시험관 내 안정성, 예컨대 생물학적 표본 또는 생리학적 환경에서의 생체 내 및/또는 시험관 내 안정성을 매개한다.
- [0091] 예를 들어 수용액 또는 생리학적 조건하에서 랜덤 코일 입체형태를 취하는 본원에서 정의된, 추가적 "제 2 도메인"을 포함하는 단백질(예를 들어 약 200개 또는 약 400개 또는 약 600개 아미노산 잔기로 이루어지며, "구성 블록"으로 PA#1/서열 번호 1, PA#2/서열 번호 2, PA#3/서열 번호 3, PA#4/서열 번호 4, PA#5/서열 번호 5, PA#6/서열 번호 6 및/또는 PIA1/서열 번호 51을 포함하는 중합체)은 상기 랜덤 코일 입체형태가 없는 대조군에 비해 심지어 생체 내에서(특히 정맥내 투여되는 경우) 유리한 혈청 안정성 또는 혈장 반감기를 갖는 것이 본원에서 포함된다.
- [0092] WO 2008/155134(상기 본원에서 논의된)에서, 랜덤 코일 입체형태를 취하는 아미노산 서열을 갖는 도메인을 포함하는 생물학적 활성 단백질은 증가된 생체 내 및/또는 시험관 내 안정성을 갖는 것으로 나타났다. WO 2008/155134에 개시된 랜덤 코일 도메인은 특히 프롤린, 알라닌, 및 세린(PAS) 잔기로 이루어진다. 이러한 3 잔기의 존재는 수용액 중 안정한 가용성 랜덤 코일의 형성을 위한 필수 요건으로 상기 선행 기술 문서에 기재된다.
- [0093] 상기 본원에서 도입에 논의된 바와 같이, WO 2007/103515에는 주성분으로 다양한 아미노산, 특히 글리신, 아스파르테이트, 알라닌, 세린, 트레오닌, 글루타메이트 및 프롤린을 포함하는 비구조화된 재조합 중합체가 기재된다. 그러나 "비구조화된 재조합 중합체"라는 용어는 "생합성" 및 "랜덤 코일"이라는 용어와 대조적으로, 인식되는 뚜렷한 의미를 갖지 않는다.
- [0094] 또한, WO 2006/081249가 상기 본원에서 언급되었다. 이 문서는 주성분으로 Gly, Asn, 및 Gln 및 소량 성분으로 Ser, Thr, Asp, Gln, Glu, His, 및 Asn을 갖는 2 내지 500개 단위의 아미노산 반복체를 포함하는 폴리펩티드에 커플링된 생물학적 활성 단백질을 포함하는 단백질 콘주게이트를 기재한다. 상기 단백질 콘주게이트는 비콘주게이션된 생물학적 활성 단백질에 비해 증가되거나 감소된 혈장 반감기를 갖는 것으로 기재된다. 그러나 WO 2006/081249는 특정 아미노산 반복체가 콘주게이트의 혈장 반감기를 감소시키거나 증강시키는지 여부를 예측하는 어떠한 교시도 제공하지 않는다. 또한, WO 2006/081249는 콘주게이션된 단백질이 본 발명에서 나타난 바와 같은 랜덤 코일 입체형태를 형성하는 아미노산 반복체를 포함하는 경우 단백질의 혈장 반감기가 증가될 수 있다는 것을 교시하거나 제시하지 않는다. 또한, WO 2006/081249에 개시된 아미노산 반복체에는 Gly, Asn, 및 Gln에서 선택되는 둘 이상의 잔기를 포함하며, 이는 단독으로 프롤린 및 알라닌 아미노산 잔기로 이루어진 아미노

산 서열을 포함하는 본 발명의 생합성 랜덤 코일 폴리펩티드와는 명백히 대조되는 것이다.

- [0095] 놀랍게도, 선행 기술과는 대조적으로 단독으로 프롤린 및 알라닌 잔기를 포함하는(즉 바람직하게는 실질적인 양의 임의의 다른 아미노산을 포함하지 않고, 또한 실질적인 양의 세린 또는 어떠한 양의 세린도 포함하지 않는) 본원에서 제공되는 생합성 랜덤 코일 아미노산 서열이 또한 유용한 랜덤 코일 구조를 형성하는 것으로 나타났다. 이는 특히 세린 및 알라닌(SA) 잔기만으로 구성된, 즉 프롤린 잔기가 생략된 도메인을 가지는 융합 단백질에 대해, 두 가지 유형의 아미노산만을 포함하는 상기 도메인이 랜덤 코일이 아닌  $\beta$ -시트 구조를 형성한다는 것을 나타내는 WO 2008/155134의 주어진 개시에서는 예측되지 않는다. 이러한 세린-알라닌 도메인은 또한 본원에서 제공되는 "PAS" 또는 특히 "P/A" 서열에서 관찰되는 바와 같은 증가된 수력학적 부피를 나타내지 않았다.
- [0096] 본원에서 사용되는 "생물학적 활성"이라는 용어는 생체 상 물질의 생물학적 효과를 설명한다. 따라서 본원에서 사용되는 "생물학적 활성 단백질"이라는 용어는 상기 단백질 또는 폴리펩티드에 노출된 살아있는 세포/유기체에서 생물학적 효과를 유도할 수 있는 단백질에 관한 것이다. 그러나 본 발명의 맥락에서, "생물학적 활성 단백질"이라는 용어는 상기 생물학적 활성을 가지고/가지거나 매개하는 아미노산 서열(상기 제 1 도메인) 및 랜덤 코일 입체형태를 취하고/형성하고 단독으로 프롤린 및 알라닌으로 이루어진 본 발명의 아미노산 서열(상기 제 2 도메인)을 모두 포함하는 본 발명의 전체 단백질에 관한 것이다.
- [0097] 따라서 본 발명의 생물학적 활성 단백질의 상기 정의된 "제 1 도메인"에 대해 본원에서 사용되는 "생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열" 또는 "생물학적 활성을 갖는 아미노산 서열"이라는 용어는 상기 정의된 "생물학적 활성"을 매개하거나 갖거나 또는 매개하거나 가질 수 있는 것이다. 또한, "생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열" 또는 "생물학적 활성을 갖는 아미노산 서열"이라는 용어에는 생체 내 또는 시험관 내의 그 반감기가 연장되어야 하는 임의의 관심 단백질(및 이들의 기능적 단편, 예컨대 항체 단편, 막 수용체의 세포외 또는 세포내 도메인(들)을 포함하는 단편, 성장 인자 또는 사이토카인의 절단된 형태 등)이 포함된다. 본 발명의 하나의 구현예에서, 본 발명에 따른 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열은 임의의 "관심 단백질", 즉 임의의 약학적 또는 생물학적 관심 단백질 또는 치료제/진단제로서 유용한 임의의 단백질로부터 추정될 수 있다.
- [0098] 따라서, 생물학적 활성 단백질은 천연 제조되는 폴리펩티드 또는 재조합 DNA 기술에 의해 제조되는 폴리펩티드로부터 유래되는 생물학적 활성 아미노산 서열을 포함하는 제 1 도메인을 포함할 수 있다. 바람직한 구현예에서, 관심 단백질은 결합 단백질/결합 분자, 면역글로불린, 항체 단편, 수송 단백질, 막 수용체, 신호전달 단백질/펩티드, 예컨대 사이토카인, 성장 인자, 호르몬 또는 효소 등으로 이루어진 군으로부터 선택될 수 있다.
- [0099] 상기 본원에서 설명되는 바와 같이, 생물학적 활성 단백질의 제 2 도메인에 포함되는 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)는 특히 생리학적 조건하에서 랜덤 코일 입체형태를 형성한다. 이는 특히 개체 또는 환자에게 투여될 약학적 조성물의 일부를 형성할 수 있는 생물학적 활성 단백질의 맥락에서 상관성이 있다.
- [0100] 자연적으로(즉 생리학적 조건하에서) 생물학적 활성 단백질의 본 발명의 생합성 랜덤 코일 도메인(상기 "제 2 도메인")은 특히 생체 내에서 그리고 의학적 개입을 필요로 하는 포유류 또는 인간 환자에게 투여되는 경우 랜덤 코일 입체형태를 취하고/형성하고/가지는 것이 주지된다. 대조적으로, 천연 입체형태로서 랜덤하지 않은 2차 및/또는 3차 구조를 갖는 단백질은 비생리학적 조건(즉 변성 조건) 하에서 랜덤 코일 입체형태를 취하는 경향이 있다는 것이 당분야에 공지되어 있다. 그러나 이러한 변성된 단백질은 본 발명의 랜덤 코일 폴리펩티드를 포함하는 생물학적 활성 단백질에 비해 완전히 상이한 특징을 갖는다. 따라서 본원에서 제공되는 융합 단백질/융합 컨스트럭트의 "생물학적 활성 단백질" 및 생물학적 활성 부분은 본 발명의 생합성 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)와 조합되고/되거나 연결되는 경우에도 이들의 생물학적 기능을 유지한다는 것이 본 발명의 요지이다.
- [0101] 또한, 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)는 생리학적 조건하에서 용해도를 보유한다. 따라서 본 발명의 단백질 컨스트럭트(상기 정의된 "제 1" 및 "제 2 도메인"을 포함하는 것)는, 예를 들어 특정 조성물, 예컨대 동결건조물 또는 건조된 조성물의 형태인 경우, 일시적으로 또는 임시로 랜덤 코일 입체형태가 아닌 "제 2"의 랜덤 코일을 형성하는/취하는 도메인을 포함할 수 있다는 것이 또한 포함된다. 그러나 본 발명의 단백질 컨스트럭트의 이러한 "제 2 도메인"은, 예컨대 해당 완충액(바람직하게는 "생리학적" 완충액/부형제 및/또는 용매) 중에서 재구성 후에는 본원에서 정의된 랜덤 코일 입체형태를 다시 취한다는 것이 중요하다. 상기 "제 2 도메인"은 (필요한 경우 해당 재구성 후) 본 발명의 생물학적 활성 단백질의 증가된 생체 내 및/또는 시험관 내 안정성을 매개할 수 있다. 본원에서 정의되는 "제 2 도메인"은 본 발명의 랜덤 코일 폴리펩티드(또는 폴리펩티드



드 절편)로 이루어지는 것이 본원에서 바람직하다.

[0102] 본원에서 사용되는 "도메인"이라는 용어는 자동으로 특정 구조 및/또는 기능을 취할 수 있는 아미노산 서열의 임의 영역/부분에 관한 것이다. 따라서 본 발명의 맥락에서, "도메인"은 기능적 도메인 또는 구조적 도메인을 나타낼 수 있다. 본원에서 기재되는 본 발명의 단백질은 생물학적 활성을 가지고/갖거나 매개하는 하나 이상의 도메인/부분 및 랜덤 코일 입체형태를 형성하는 하나 이상의 도메인/부분을 포함한다. 그러나 본 발명의 단백질은 또한 3개 이상의 도메인으로 이루어질 수 있으며, 예컨대 본원에서 정의된 두 도메인/부분 또는 또 다른 도메인/부분 등의 사이에 추가적 링커 또는 스페이서 구조, 예컨대 프로테아제 감수성 절단 부위, 친화도 태그, 예컨대 His<sub>6</sub>-태그 또는 *Strep*-태그, 신호 펩티드, 보유 펩티드, 표적화 펩티드, 예컨대 막 전위 펩티드 또는 추가적 효과기 도메인, 예컨대 항-종양 독소에 연관된 종양 표적화용 항체 단편 또는 전구약물-활성화용 효소 등을 포함할 수 있다.

[0103] 또 다른 구현예에서, 본 발명의 생물학적 활성 단백질은 분석용 겔 여과(또한 크기 배제 크로마토그래피, SEC로 공지되어 있음)에 의해 결정되는 50 kDa 이상, 바람직하게는 70 kDa 이상, 보다 바람직하게는 80 kDa 이상, 더욱 바람직하게는 100 kDa 이상, 특히 바람직하게는 125 kDa 이상 및 가장 바람직하게는 150 kDa 이상의 수력학적 부피를 갖는다. 당분야의 숙련자는 특정 단백질의 수력학적 부피를 쉽게 결정할 수 있다. 예시적인 방법이 랜덤 코일 폴리펩티드의 맥락에서 상기 본원에서 설명되었다. 당분야의 숙련자는 본 발명의 생물학적 활성 단백질의 맥락에서 또한 상기 방법을 쉽게 채용할 수 있는 입장에 있다. 하기 본원에서 기재되는 상기 정의된 제 2 도메인, 즉 본원에서 제공되는 랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하거나 이로써 이루어진 도메인을 포함하는 본 발명의 생물학적 활성 단백질의 수력학적 부피는 이들의 분자량 또는 아미노산 잔기의 수/조성에 근거하여 대응하는 폴딩된 구형 단백질에 대해 추정되는 수력학적 부피에 관해 예상치 못하게 큰 수력학적 부피를 갖는 것으로 나타난다.

[0104] "생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열"을 포함하는 제 1 도메인도 또 다른 폴리펩티드 또는 아미노산 서열의 맥락에서 또는 이와는 연합 후에 그 생물학적 활성을 취할 수 있다는 것이 주지되어야 한다. 예를 들어 항체의 Fab 단편, 예컨대 항종양 항체 중 하나인 허셉틴(Herceptin)(Eigenbrot (1993) J. Mol. Biol. 229:969-995)은 두 상이한 폴리펩티드 사슬인 면역글로불린 경쇄 및 면역글로불린 중쇄의 단편으로 이루어지며, 이는 또한 사슬내 디설피드 결합(들)을 통해 연결될 수 있다. 본 발명에 따르면, 다른 사슬과의 연합에 의해 전체 생물학적 활성 단백질을 재구성하면서 이들 사슬 중 하나를 (예로 유전자 융합에 의해) 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)로 연결하기 충분할 수 있다. 이러한 재구성은, 예를 들어 첨부된 실시예에 기재된 대로, 또는 시험관 내 재구성에 의해, 예를 들어 재폴딩 프로토콜의 일부로서 동일한 숙주 세포내에서 상이한 폴리펩티드(한편에서는 하나의 사슬 및 랜덤 코일 폴리펩티드의 융합 단백질, 또 한편에서는 다른 사슬)의 공발현에 의해 달성될 수 있다.

[0105] 따라서 상기 단백질(두 별개의 폴리펩티드 사슬을 포함)도 본 발명에 따른 생물학적 활성 단백질로 간주된다. 이러한 경우, 본원에서 정의되는 제 1 도메인은 비공유적으로만 연결되는 두 별개의 폴리펩티드 사슬을 포함할 수 있다. 또한, 생물학적 활성 단백질/도메인의 독립적인 사슬은 각각 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)에 연결될 수 있다. 항체 단편에 더하여, 본 발명의 대상이 되는 몇몇 연합 폴리펩티드 사슬로 구성되는 여러 다른 동중- 또는 이중-올리고머 관심 단백질이 존재한다.

[0106] 본원에서 사용되는 "결합 단백질"이라는 용어는 잠재적인 결합 파트너(들)와 특이적으로 상호작용할 수 있어서, 잠재적인 결합 파트너(들)로서 상기 복수의 상이한 분자 풀로부터 상기 잠재적인 결합 파트너(들)만이 결합하거나 유의미하게 결합하는 정도로 상기 잠재적인 결합 파트너(들) 및 복수의 상이한 분자간을 구별할 수 있는 분자에 관한 것이다. 결합 단백질 및 잠재적인 결합 파트너 간 결합의 측정 방법은 당분야에 공지되어 있으며, 예컨대 ELISA, 등은 적정 열량측정, 평형 투석, 폴다운 분석, 표면 플라즈몬 공명 또는 Biacore 장치를 이용하여 일상적으로 수행될 수 있다. 본 발명의 맥락에서 유용한 예시적인 결합 단백질/결합 분자에는 항체, 항체 단편, 예컨대 Fab 단편, F(ab')<sub>2</sub> 단편, 단일쇄 가변 단편(scFv), 항체의 단리된 가변 영역(VL 및/또는 VH 영역), CDR, 단일 도메인 항체/면역글로불린, CDR-유래 펩티드 모사체, 렉틴, 면역글로불린 도메인, 피브로넥틴 도메인, 단백질 A 도메인, SH3 도메인, 안키린 반복 도메인, 리포칼린 또는 예를 들어 [Skerra (2000) J Mol Recognit 13:167-187, Gebauer (2009) Curr Opin Chem Biol 13:245-255, Binz (2005) Nat Biotechnol 23:1257-1268 또는 Nelson (2009) Nat Biotechnol 27:331-337]에 기재된 다양한 유형의 스캐폴드-유래 결합 단백질이 비제한적으로 포함된다.

[0107] 본 발명의 맥락에서 유용한 다른 예시적인 관심 생물학적 활성 단백질(특히 생물학적 활성 단백질의 제 1 도메

인에 포함되거나 제 1 도메인을 구성하는/이루는 단백질)에는 과립구 집락 자극 인자, 인간 성장 호르몬, 알파-인터페론, 베타-인터페론, 감마-인터페론, 람다-인터페론, 종양 괴사 인자, 에리트로포이에틴, 응고 인자, 예컨대 응고 인자 VIII, 응고 인자 VIIa, 응고 인자 IX, gp120/gp160, 가용성 종양 괴사 인자 I 및 II 수용체, 혈전용해제, 예컨대 레테플라아제, 대사성 효과를 갖는 펩티드, 예컨대 GLP-1 또는 엑센딘-4, 면역억제/면역조절 단백질, 예컨대 인터루킨-1 수용체 길항제 또는 아나킨라, 인터루킨-2 및 호중구 젤라티나아제 연관 리포칼린 또는 다른 천연 또는 조작된 리포칼린 또는 예를 들어 [Walsh (2003) Nat Biotechnol 21:865-870 또는 Walsh (2004) Eur J Pharm Biopharm 58:185-196]에 열거되거나 온라인 데이터베이스, 예컨대 <http://www.biopharma.com/approvals.html> 또는 <http://www.drugbank.ca>에 열거된 단백질 또는 화합물이 비제한적으로 포함된다. 본 발명의 맥락에서 채용될 수 있는 추가 생물학적 활성 단백질(특히 생물학적 활성 단백질의 제 1 도메인에 포함되거나 제 1 도메인을 구성하는/이루는 단백질)로는 특히 난포 자극 호르몬, 글루코세레브로시다아제, 타이모신 알파 1, 글루카곤, 소마토스타틴, 아데노신 데아미나아제, 인터루킨 11, 헤마티드, 랩틴, 인터루킨-20, 인터루킨-22 수용체 서브유닛 알파(IL-22r α), 인터루킨-22, 히알루로니다아제, 섬유아세포 성장 인자 18, 섬유아세포 성장 인자 21, 글루카곤 유사 펩티드 1, 오스테오프로테게린, IL-18 결합 단백질, 성장 호르몬 방출 인자, 가용성 TACI 수용체, 트롬보스폰딘-1, 가용성 VEGF 수용체 Flt-1, α-갈락토시다아제 A, 미오스타틴 길항제, 위 억제 폴리펩티드, 알파-1 안티트립신, IL-4 뮤테인 등이 있다. 본원의 개시에서 자명할 바와 같이, 본 발명은 또한 생합성 랜덤 코일 프롤린/알라닌 폴리펩티드 또는 프롤린/알라닌 폴리펩티드 절편 및 약학적으로 또는 의학적으로 유용한 분자, 예컨대 소분자, 펩티드 또는 생체거대분자, 예컨대 단백질, 핵산, 탄수화물, 지질 소포 등, 특히 약학적으로 또는 의학적으로 유용한 단백질, 예컨대 (비제한적으로) 결합 단백질/결합 분자, 면역글로블린, 항체 단편, 수송 단백질, 막 수용체, 신호전달 단백질/펩티드, 사이토카인, 성장 인자, 호르몬 또는 효소 등은 본원에서 정의된 약물 컨스트럭트에 포함될 수 있을 뿐만 아니라 상기 정의된 둘 이상의 도메인을 포함하거나 이로써 이루어진 본원에서 정의된 생물학적 활성의 이중성 단백질의 일부일 수 있다. 이러한 경우, 특히 상기 약학적으로 또는 의학적으로 유용한 단백질(또는 이들의 기능적 단편)은 상기 생물학적 활성을 가지고/가지거나 매개하는 아미노산 서열을 포함하거나 이로써 이루어진 상기 둘 이상의 도메인의 "제 1 도메인"일 수 있다. 이러한 맥락에서 기능적 단편은 생체 내 및/또는 시험관 내에서 의도하는 생물학적 또는 약학적 반응을 여전히 규명할 수 있고/있거나 의도하는 생물학적 활성을 여전히 갖거나 매개하는 상기 약학적으로 또는 의학적으로 유용한 단백질의 단편이다.

[0108] 상기 제 1 및 상기 제 2 도메인 사이에 삽입된 상기-언급된 폴리펩티드 링커/스페이서는 바람직하게는 두 도메인에 모두 공유적으로 연결된 복수의 친수성, 펩티드-결합된 아미노산을 포함한다. 또 다른 구현예에서, 상기 폴리펩티드 링커/스페이서는 생물학적 활성을 갖고/갖거나 매개하는 폴리펩티드를 포함하는 상기 제 1 도메인의 조절 방출을 허용하는 혈장 프로테아제 절단 부위를 포함한다. 상이한 유형 또는 길이의 링커는 특정 단백질의 최적 생물학적 활성을 획득하기 위해 부적절한 부담 없이 동정될 수 있다.

[0109] 바람직한 구현예에서, 본 발명의 생물학적 활성 단백질은 융합 단백질이다. 본원에서 기재되는 융합 단백질은 단일 멀티도메인 폴리펩티드에서 생물학적 활성을 매개할 수 있는 하나 이상의 도메인 및 본원에서 기재되는 생합성 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)를 포함하는 하나 이상의 다른 도메인을 포함할 수 있다. 또 다시, 본 발명은 하나의 도메인이 생물학적 활성을 매개하는 융합 단백질로 제한되지 않음이 주지된다. 또한, 하나의 부분/도메인은 본 발명의 프롤린/알라닌의 랜덤 코일 폴리펩티드/중합체이거나 이를 포함하며, 다른 부분/도메인은 또 다른 단백질 연장부/구조를 포함하는 다른 "융합 단백질"/"융합 컨스트럭트"가 본원에서 제공된다.

[0110] 특히 융합 단백질의 경우에, 본 발명에 따른 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)는 그 아미노 또는 카르복실 말단에 반드시 Pro 또는 Ala 잔기를 수반하는 것은 아니다. 대안적 구현예에서, 본 발명에 따른 생물학적 활성 단백질은 생물학적 활성을 갖고/갖거나 매개하는 관심 단백질 또는 폴리펩티드/폴리펩티드 연장부/펩티드/아미노산 서열이 랜덤 코일 입체형태, 특히 본원에서 제공되는 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)를 형성하고/취하고, 단독으로 프롤린 및 알라닌 잔기로 이루어진 아미노산 서열에 비펩티드 결합을 통해 콘주게이션되는 단백질 콘주게이트를 나타낼 수 있다. 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편에 있어서 단백질 가교를 위해 유용한 비펩티드 결합은 당분야에 공지되어 있으며, 여기서 상기 아미노산 서열은 본원에서 제공되는 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다. 이러한 비펩티드 결합에는 예컨대 Cys 측쇄 간 디설피드 결합, 화학적 가교제, 디숙신이미딜 수베레이트(DSS) 또는 설포숙신이미딜 4-[p-말레이미도페닐]부티레이트(Sulfo-SMPB), 금속 킬레이트기/착화기, 뿐만 아니라 비공유 단백질-단백질 상호작용에 의해 유도되는 티오에테



르 결합 또는 비펩티드 공유 결합이 포함될 수 있다.

- [0111] 본 발명의 "생물학적 활성 단백질"이 또한 둘 이상의 "생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열"을 포함할 수 있다는 것이 주지된다. 또한, 생물학적 활성 단백질은 또한 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편) 이상을 포함할 수 있다. 가장 단순한 경우에서, 생물학적 활성 단백질은 두 도메인, 즉 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열을 포함하는 제 1 도메인 및 생합성 폴리펩티드(또는 이들의 절편)를 포함하는 제 2 도메인으로 이루어진다. 본 발명은 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 본원에 개시되는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편에 연결된 "생물학적 또는 치료적 활성 단백질"에 제한되지 않음이 주지되며, 여기서 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다. 또한, 다른 산업, 예컨대 식품 또는 식품 또는 음료 산업, 화장품 산업 등에 관련되는 다른 관심 단백질 또는 분자가 본원에서 제공되는 수단 및 방법에 의해 제조될 수 있다.
- [0112] 당분야의 숙련자는 본 발명의 생물학적 활성 단백질에 포함되는 "생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열을 포함하는 도메인" 및 "랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하는 제 2 도메인"이 특정 순서로 조직될 수 있음을 인지한다.
- [0113] 따라서 본 발명의 맥락에서, 본 발명의 생물학적 활성 폴리펩티드의 본원에서 정의된 "제 1" 및 "제 2" 도메인의 순서는 상기 "제 1 도메인"(즉 관심 단백질; "상기 생물학적 활성을 가지고/가지거나 매개하는 아미노산 서열")이 아미노(N-) 말단에 위치하고 상기 "제 2 도메인"(즉 본원에서 제공되는 랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하는 도메인)이 생물학적 활성 단백질의 카르복시(C-) 말단에 위치하는 순서로 배열될 수 있다. 그러나 이 순서는 또한 역전될 수 있다, 예컨대 상기 "제 1 도메인"(즉 관심 단백질; "상기 생물학적 활성을 가지고/가지거나 매개하는 아미노산 서열")이 카르복시(C-) 말단에 위치하고 상기 "제 2 도메인"(즉 본원에서 제공되는 랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하는 도메인)이 생물학적 활성 단백질의 아미노(N-) 말단에 위치할 수 있다. 생물학적 활성 단백질이 하나의 제 1 도메인 및 하나의 제 2 도메인만으로 이루어진 경우, 도메인 순서는 다음과 같을 수 있다(N-말단에서 C-말단으로): 제 1 도메인(생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편)). 반대로 도메인 순서는 다음과 같을 수 있다(N-말단에서 C-말단으로): 제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 1 도메인(생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열).
- [0114] 상기 생물학적 활성을 가지고/가지거나 매개하는 아미노산 서열을 포함하거나 이로써 이루어진 둘 이상의 도메인을 본 발명의 단백질 컨스트럭트의 맥락에서 사용하는 것이 또한 포함된다. 예를 들어 생물학적 활성 단백질은 두 "제 1 도메인", 즉 생물학적 활성을 갖고/갖거나 매개하는 두 특정 아미노산 서열을 포함할 수 있고, 여기서 상기 생물학적 활성은 동일하거나 상이한 활성일 수 있다. 생물학적 활성 단백질이 이러한 두 "제 1 도메인", 즉 생물학적 활성을 갖고/갖거나 매개하는 두 특정 아미노산 서열, 및 하나의 "제 2 도메인"(생합성 랜덤 코일 폴리펩티드(또는 이들의 절편) 포함)으로 이루어진 경우, 도메인 순서는 다음과 같을 수 있다(N-말단에서 C-말단): 제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))- 제 1 도메인(특정한 (선택적으로 상이한) 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열).
- [0115] 생물학적 활성 단백질이 둘 이상의 "제 2 도메인"을 포함하는(즉 생물학적 활성 단백질이 둘 이상의 랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하는) 경우에 동일한 설명이 적용된다. 생물학적 활성 단백질이 이러한 두 "제 2 도메인", 즉 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하는 두 도메인, 및 하나의 "제 1 도메인"(생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열을 포함하는 도메인)으로 이루어진 경우, 도메인 순서는 다음과 같을 수 있다(N-말단에서 C-말단): 제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편)). 생물학적 활성 단백질이 둘 이상의 "제 2 도메인"을 포함하는 경우, 본원에서 이러한 "제 2 도메인"이 동일하거나 상이할 수 있다는 것이 포함된다.
- [0116] 상기에서 언급된 바와 같이, 생물학적 활성 단백질은 둘 이상의 "제 1 도메인", 즉 생물학적 활성을 갖고/갖거나 매개하는 둘 이상의 특정 아미노산 서열 및 둘 이상의 "제 2 도메인"(생합성 랜덤 코일 폴리펩티드(또는 이들의 절편))을 포함할 수 있으며, 여기서 이러한 "제 1 도메인"은 동일하거나 상이할 수 있고/있거나 상기 "제 2 도메인"은 동일하거나 상이할 수 있다. 이러한 경우, 하기 예시적인 도메인 순서를 고려할 수 있다(N-말단에서 C-말단):
- [0117] - 제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드

(또는 이들의 절편))-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편));

[0118] - 제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편));

[0119] - 제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열);

[0120] - 제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열);

[0121] - 제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열); 또는

[0122] - 제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 1 도메인(특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열)-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편))-제 2 도메인(랜덤 코일 폴리펩티드(또는 이들의 절편)).

[0123] 당분야 숙련자에 있어서, 추가 대응 도메인 순서(특히 3 이상의 "제 1 도메인" 또는 3 이상의 "제 2 도메인"이 생물학적 활성 단백질에 포함된 경우)를 쉽게 고려할 수 있다.

[0124] 본 발명의 폴리펩티드/생물학적 활성 단백질의 모든 구현예에서, 상기 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열을 포함하는 상기 도메인(들)은 또한, 의도하는 생물학적 기능을 갖는 주어진 단백질의 생물학적 활성 단편일 수 있다. 따라서 본원에서 정의된 "제 2 도메인"(바람직하게는 본원에서 제공되는 랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하는 도메인)은 또한 관심 단백질의 두 생물학적 활성 단편 사이 또는 두 관심 단백질의 생물학적 활성 단편 사이에 위치할 수 있다. 관심 "전장" 단백질/폴리펩티드의 맥락에서 (즉 아미노산 서열이 자체적으로 특정한 생물학적 활성을 갖는/매개하는 경우) 상기 본원에서 주어진 모든 설명 및 정의가 필요한 변경을 가하여 상기 단편의 맥락에서 적용된다.

[0125] 또다시, 상기 본 발명은 "생물학적 활성 기능"을 갖는 "도메인"을 포함하는 컨스트럭트에 제한되지 않는다. 본 발명의 컨스트럭트는 또한 다른 기능을 갖는 도메인을 포함할 수 있으며, 생물학적 활성에 제한되지 않는다. 이는 단지 본 발명의 구현예이며, 당분야의 숙련자에게는 본 발명의 요지를 벗어나지 않고 다른 컨스트럭트를 쉽게 제조하고 사용할 수 있음이 자명하다. 따라서, "특정 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열"의 맥락에서 상기 본원의 내용은 다른 컨스트럭트, 예를 들어 화장품, 식품 가공, 유제품, 종이 제조 등과 같은 다른 기술 분야에서 사용될 컨스트럭트에 필요한 변경을 가하여 적용된다. 상기 본원에서 언급된 바와 같이, 본 발명의 생합성 폴리펩티드/중합체는 또한 예컨대 소분자 등에 연결되어 사용될 수 있다.

[0126] 또다시, "제 1 생물학적 활성을 갖고/갖거나 매개하는 아미노산 서열"이라는 용어는 상기 생물학적 활성 또는 기능을 갖고/갖거나 매개하는 전장 폴리펩티드에 제한되지 않고 이들의 생물학적 및/또는 약리학적 활성 단편에도 적용됨이 지적되어야 한다. 특히 비제한적으로 본 발명의 "생물학적 활성 단백질"에 본원에서 정의되는 3 이상의 "제 1 도메인"이 포함되는 맥락에서, 이러한 "제 1 도메인"이 단백질 복합체의 상이한 부분 또는 단백질 복합체의 상기 부분의 단편이거나 이를 나타내는 것임이 또한 포함된다.

[0127] 본원에서 아래에 예시되는 바와 같이, 랜덤 코일 폴리펩티드를 포함하도록 개질된 본 발명의 생물학적 활성 단백질은 놀랍게도 상기 랜덤 코일 도메인이 없는 미개질된 생물학적 활성 단백질에 비해 증가된 생체 내 및/또는 시험관 내 안정성을 나타낸다. 본원에서 사용되는 "생체 내 안정성"이라는 용어는 생체로 투여되는 특정 물질의 생물학적 이용가능성 및 생물학적 활성의 유지 능력에 관한 것이다. 생체 내에서 물질은 배출, 신장 여과, 간 흡수, 응집, 분해 및/또는 다른 대사 공정으로 인해 제거되고/되거나 불활성화될 수 있다. 따라서 본 발명의 맥락에서 증가된 생체 내 안정성을 갖는 생물학적 활성 단백질은 신장을 통해(오줌으로) 또는 변을 통해 덜 빠르게 배출되고/되거나 단백질 분해, 특히 생물학적 유체, 예컨대 혈액, 뇌척수액, 복강액, 및 림프에서의 생체 내 단백질 분해에 대해 더 안정할 수 있다. 하나의 구현예에서, 생물학적 활성 단백질의 증가된 생체 내 안정성은 상기 생물학적 활성 단백질의 연장된 혈장 반감기로 발현된다. 특히 생물학적 활성 단백질의 증가된 생체 내

안정성으로 제 2 도메인이 없는 생물학적 활성 단백질에 비해 상기 제 2 도메인을 포함하는 상기 생물학적 활성 단백질의 혈장 반감기가 연장된다.

- [0128] 생물학적 활성 단백질의 생체 내 안정성 측정 방법은 당분야에 공지되어 있다. 하기 본원에서 예시된 바와 같이, 생물학적 활성 단백질은 웨스턴 블로팅 기법 또는 효소 연관 면역흡착 분석(ELISA)을 이용하여 혈액 혈장 중에서 특이적으로 검출될 수 있다. 그러나 당분야의 숙련자는 관심 단백질의 혈장 반감기를 특이적으로 측정하기 위해 다른 방법을 채용할 수 있다는 것을 인지한다. 상기 방법에는 방사활성으로 표지된 관심 단백질의 물리적 검출이 비제한적으로 포함된다. 단백질의 방사활성 표지 방법, 예컨대 방사성 요오드화는 당분야에 공지되어 있다.
- [0129] 본원에서 사용되는 "증가된 시험관 내 안정성"이라는 용어는 생물학적 활성 단백질의 분해 및/또는 응집에 저항하고 시험관 내 환경에서 그 원래의 생물학적 활성을 유지하는 능력에 관한 것이다. 생물학적 활성 단백질의 생물학적 활성 측정 방법은 당분야에 널리 공지되어 있다.
- [0130] 또한, 본원에 기재되고 정의된 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 상기 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편에 콘주게이션된 소분자 약물을 포함하는 약물 콘주게이트가 제공된다. 소분자의 비제한적 예로는 디곡시제닌, 플루오레신 독소루비신, 칼리케아미신, 캄프토테신, 푸마길린, 텍사메타손, 겔다나마이신, 파클리탁셀, 도세탁셀, 이리노테칸, 시클로스포린, 부프레노르핀, 날트렉손, 날록손, 빈데신, 반코마이신, 리스페리돈, 아리피프라졸, 팔로노세트론, 그라니세트론, 시타라빈, NX1838, 류프롤리드, 고세렐린, 부세렐린, 옥트레오티드, 테두글루티드, 실렌기티드, 아바렐릭스, 엔푸비르티드, 그랜린 및 유도체, 투블라이신, 플라틴 유도체, 알파 4 인테그린 저해제, 안티센스 핵산, 소 간섭 RNA, 마이크로 RNA, 스테로이드, DNA 또는 RNA 앵타머, 펩티드, 펩티드 모사체가 있다. 일반적으로 본 발명은 또한 본원에서 정의된 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 특히 약학적으로 또는 의학적으로 유용한 분자, 예컨대 소분자, 펩티드 또는 생체거대분자, 예컨대 단백질, 핵산, 탄수화물, 지질 소포 등을 포함하는 약물 컨스트럭트에 관한 것이다. 첨부된 예시적 실험 부분(예로 실시예 22 참고)에서, 본 발명의 컨스트럭트/콘주게이트, 또한 "화학적 소분자"가 본원에 개시되는 랜덤 코일 폴리펩티드에 콘주게이션된 컨스트럭트의 성공적 생성이 제공된다. 따라서 본 발명의 도면 및 해당 도면 설명에서의 실험적 정보가 예시적 예를 제공하며, 본원에 개시되는 약물 콘주게이트에는 (i) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편(상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다), 및 (ii) 예로서 디곡시제닌 및 플루오레신으로부터 선택되는 소분자가 포함된다. 이것이 단지 학술적 예만이 아님이 주지된다. 플루오레신 또는 플루오레신 유도체는 통상 진단제로 사용되며, 의학적 플루오레신 용액은 상표명 Fluorescein®, AK-FLUOR® 또는 Fluorescein®하에 시판된다. 이러한 화합물은 확실히 본원에서 제공되는 수단 및 방법의 이득을 얻을 수 있다. 디곡시제닌은 3가지 디지톡소오스 당을 또한 함유하는 심장활성 기능을 갖는 널리 공지된 이차적 식물 대사체인 디곡신의 스테로이드 부분을 형성한다. 디곡신, 및 좀 적은 정도로 밀접하게 연관된 화합물인 디지톡신은 심실 부정맥 및 울혈성 심부전의 치료를 위해 널리 사용된다(Hauptman (1999) Circulation 99:1265-1270). 모든 심장활성 스테로이드는 세포성 형질막에 위치하는  $Na^+/K^+$ -ATPase의 강력하고 매우 특이적인 저해제이므로 교감신경차단 또는 양의 수축촉진 효과를 발휘한다.
- [0131] 랜덤 코일 폴리펩티드 또는 이들의 폴리펩티드 절편의 맥락에서 상기 본원에서 주어진 정의 및 설명은 랜덤 코일 폴리펩티드(또는 폴리펩티드 이들의 절편)를 포함하며, (a) 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 또는 폴리펩티드 및 (b) 소분자 약물로 이루어진 균으로부터 선택되는 약물을 포함하는 약물 콘주게이트의 맥락에서 필요한 변경을 가하여 적용된다.
- [0132] 본원에서 정의되고 제공되는 랜덤 코일 입체형태를 형성하는 아미노산 중합체/랜덤 코일 폴리펩티드(또는 이들의 절편)는 소분자/소분자 약물에 콘주게이션될 수 있다. 이로써, 소분자/소분자 약물의 혈장 반감기 및/또는 용해도가 증가될 수 있고, 비특이적 독성이 감소될 수 있고, 신체내 표적 세포 또는 구조에 대한 활성 약물의 연장된 노출로 증강된 약력학을 일으킬 수 있다.
- [0133] 랜덤 코일 폴리펩티드 N-말단과 활성화된 약물 유도체, 예컨대 N-히드록시숙신이미드(NHS) 에스테르 유도체의 부위-특이적 콘주게이션(Hermanson (1996) Bioconjugate Techniques, Academic Press, San Diego, CA)이 가능하다. 일반적으로 N-말단 아미노기는 다양한 작용기, 예컨대 알데히드 및 케톤(예를 들어 수소화붕소나트륨 또는 수소화시안나트륨을 이용하여 아민으로 환원될 수 있는 시프(Schiff) 염기 형성) 또는 활성화된 카르복산 유도체(무수물, 클로라이드, 에스테르 등으로 아미드 형성) 또는 다른 반응성 화학제, 예컨대 이소시아네이트, 이소티오시아네이트, 설포닐 클로라이드 등에 화학적으로 커플링될 수 있다. 또한, 아미노산 중합체/폴리펩티드



의 N-말단은 먼저 적합한 보호기, 예를 들어 아세틸기, BOC기 또는 FMOC기로 개질될 수 있다(Jakubke (1996) Peptide. Spektrum Akademischer Verlag, Heidelberg, Germany). 또한, 아미노 말단은 피로글루타미드에 의해 보호될 수 있고, 이는 Pro/Ala 폴리펩티드 또는 폴리펩티드 절편에 선행하는 인코딩된 Gln 아미노산 잔기로부터 형성될 수 있다. 예컨대 일반 시약 EDC(N-(3-디메틸아미노프로필)-N'-에틸카르보디이미드) 및 NHS를 이용한 C-말단 카르복실레이트기의 활성화 후, 약물이, 예를 들어 자유 아미노기를 수반하는 경우 보호된 랜덤 코일 폴리펩티드의 C-말단으로의 부위-특이적인 커플링을 수행할 수 있다.

[0134] 대안적으로, 랜덤 코일 입체형태/랜덤 코일 폴리펩티드를 형성하는 아미노산 중합체의 N-말단 또는 C-말단은 말레이미드기를 제공하여 약물 분자의 일부로서 티올기로의 화학적 커플링을 허용하는 시판 링커 시약으로 개질될 수 있다. 이러한 방식으로, 균일한 약물 콘주게이트를 용이하게 획득할 수 있다. 당분야에 널리 공지되어 있는 유사한 기법(Hermanson (1996) 위 인용문)을 이용하여 랜덤 코일 폴리펩티드를 펩티드 또는 단백질 약물에 커플링할 수 있다. 이러한 펩티드 또는 단백질은 NHS 에스테르 또는 말레이미드 활성기를 통해 이들의 랜덤 코일 입체형태를 형성하는 아미노산 중합체로의 시험관 내 커플링을 허용하는 Lys 또는 Cys 측쇄를 수반하여 쉽게 제조될 수 있다. 일반적으로 랜덤 코일 폴리펩티드(또는 이들의 절편)를 포함하는 융합 단백질로 유사한 약물 콘주게이트를 제조할 수 있다. 그러나 첨부된 실시예 및 도에 예시된 바와 같이, 본 발명은 또한 본 발명의 혁신적 콘주게이트에 포함되는 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편의 제조를 제공한다.

[0135] 단일 부위-특이적인 콘주게이션에 대한 대안으로서, 랜덤 코일 폴리펩티드에는 N- 또는 C-말단에서 또는 내부적으로, 예를 들어 NHS 에스테르 또는 말레이미드 활성기를 이용하여 여러 소분자의 콘주게이션을 허용하는, 화학적 개질을 위해 적합한 추가 측쇄, 예컨대 이들의 ε-아미노기를 갖는 라이신 잔기, 이들의 티올기를 갖는 시스테인 잔기, 또는 비천연 아미노산이 제공될 수 있다.

[0136] 안정한 콘주게이션과 별도로, 전구약물은 일시적으로 랜덤 코일 폴리펩티드에 연결될 수 있다. 상기 연결은, 예를 들어 약하게 가용성인 항암제 캄프토테신이 PEG 중합체에 콘주게이션되어 증가된 생체분포, 감소된 독성, 증강된 효능 및 종양 축적을 달성하는 바와 같이(Conover (1998) Cancer Chemother Pharmacol, 42:407-414), 생리학적 pH에서 유사하게 개시되는 느린 가수분해 또는 효소적 기전을 통해 예측 가능한 방식으로 생체 내에서 절단되도록 설계될 수 있다. 추가 전구약물의 예로는 화학치료제, 예컨대 도세탁셀(Liu (2008) J Pharm Sci. 97:3274-3290), 독소루비신(Veronese (2005) Bioconjugate Chem. 16:775-784) 또는 파클리탁셀(Greenwald (2001) J Control Release 74:159-171)이 있다.

[0137] 또한, 소분자는 표적화 도메인, 예컨대 항체 단편에 유전적으로 융합된 아미노산 중합체/폴리펩티드를 포함하는 융합 단백질에 커플링되어 소분자 약물의 특이적 전달을 일으킬 수 있다. 후자의 경우, 표적화 도메인이, 예를 들어 내재화를 겪는 세포-표면 수용체에 대해 지정되는 경우, 세포독성 소분자와의 콘주게이션에 의해 면역독소를 쉽게 생성할 수 있다.

[0138] 상기에 따라, 본 발명은 또한 선택된 다른 화합물과의 또 다른 추가적 커플링을 위한 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 본원에 개시되는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편의 제공에 관한 것이다. 상기 또 다른 및/또는 추가적 커플링은 또 다른 화합물과의 또는 이에 대한 상기 생합성 랜덤 코일 폴리펩티드 또는 생합성 랜덤 코일 폴리펩티드 절편의 첫 번째 커플링이거나 이를 포함할 수 있다.

[0139] 또 다른 구현예에서, 본 발명은 본원에서 기재되는 랜덤 코일 폴리펩티드(또는 이들의 절편) 또는 생물학적 활성 단백질을 인코딩하는 핵산 분자에 관한 것이다. 따라서 상기 핵산 분자는 생물학적 활성을 갖는 폴리펩티드를 인코딩하는 핵산 서열 및 랜덤 코일 폴리펩티드(또는 이들의 절편)를 인코딩하는 핵산 서열...을 포함할 수 있다. 본원에서 사용되는 "핵산 분자"라는 용어는 핵산 분자, 예컨대 DNA 분자 및 RNA 분자를 포함하려는 것이다. 상기 핵산 분자는 단일쇄 또는 이중쇄일 수 있으나, 바람직하게는 이중쇄 DNA이다. 바람직하게는 상기 핵산 분자는 벡터에 포함될 수 있다.

[0140] 따라서, 본 발명은 또한 본원에서 제공되는 콘주게이트, 예컨대 본원에서 정의되는 약물 콘주게이트에 포함되는 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 분자, 또는 상기 정의된 생물학적 활성 단백질을 포함하고 추가적으로 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하는 단백질 콘주게이트를 인코딩하는 핵산 분자에 관한 것으로, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다.

[0141] 하나의 구현예에서, 콘주게이트, 예컨대 상기 정의된 바와 같은 약물 콘주게이트 또는 식품 콘주게이트를 인코

당하는 핵산 분자가 제공되며, 상기 핵산 분자는 하기를 포함한다:

- [0142] (i) 번역된 아미노산 및/또는 리더 서열을 인코딩하는 핵산 서열;
- [0143] (ii) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 서열(상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다);
- [0144] (iii) 생물학적 활성을 갖거나 이를 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 또는 상기 폴리펩티드 또는 관심 단백질, 예컨대 식품 산업과 같은 다른 산업 분야에서 채용될 단백질을 인코딩하는 핵산 서열; 및
- [0145] (iv) 번역 정지 코돈이거나 이를 나타내는 핵산 서열.
- [0146] 상기 언급된 (i)에서의 "번역된 아미노산 및/또는 리더 서열"은, 예를 들어 해당 출발 코돈에서 유래하는 출발 "M", 즉 메티오닌일 수 있고, 또한 예를 들어 리보솜 결합 부위를 포함하는 출발 코돈까지의 5' 서열과 같은 mRNA의 비번역된 서열을 포함할 수 있다. 그러나 이러한 서열은 또한, 예를 들어 원형질막 내로 또는 배양 배지 중 발현된 단백질의 분비를 위한 전통적 리더 및/또는 신호 서열을 포함할 수 있다. 원핵생물 신호 펩티드는, 예를 들어 OmpA, MalE, PhoA, DsbA, pelB, Afa, npr, STII이다. 진핵생물 신호 펩티드는, 예를 들어 꿀벌 멜리틴 신호 서열, 산성 당단백질 gp67 신호 서열, 마우스 IgM 신호 서열, hGH 신호 서열이다.
- [0147] 생물학적 활성을 갖거나 이를 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 또는 폴리펩티드뿐만 아니라 다른 관심 단백질, 예컨대 다른 산업 분야에서 채용될 단백질이 상기 본원에서 제공되었다. 상기 구현에는 상기 본원에서 예시되는 바와 같이 핵산 분자(부분/절편 (iii))에 대해 필요한 변경을 가하여 적용된다.
- [0148] 본원에서 제공되는 핵산 분자에서 채용될 번역 정지 코돈은 당분야에 널리 공지되어 있으며, 예컨대 코돈 UAA, UAG 또는 UGA이다.
- [0149] 상기 본원에서 제공되는 핵산 분자의 하나의 구현예에서, 상기 핵산 분자 부분/절편 (ii) 및 (iii)은 콘주게이트, 예컨대 약물 콘주게이트 또는 식품 콘주게이트를 인코딩하는 상기 핵산 분자 상 이들의 위치에서 상호 교환된다. 이러한 핵산 분자는 하기 순서의 부분/절편을 포함할 것이다:
- [0150] (i) 번역된 아미노산 및/또는 리더 서열을 인코딩하는 핵산 서열;
- [0151] (ii) 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 또는 상기 폴리펩티드 또는 관심 단백질, 예컨대 식품 산업과 같은 다른 산업 분야에서 채용될 단백질을 인코딩하는 핵산 서열;
- [0152] (iii) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 서열(상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다); 및
- [0153] (iv) 번역 정지 코돈이거나 이를 나타내는 핵산 서열.
- [0154] 상기 본원에서 제공되는 핵산 분자는 또한 선택적으로 부분/절편 (i) 및 (ii) 사이에 및/또는 부분/절편 (ii) 및 (iii) 사이에, 프로테아제 및/또는 화학적 절단 부위 및/또는 인식 부위를 포함할 수 있다. 이러한 화학적 절단 부위는 당분야에 널리 공지되어 있으며, 예를 들어 특정 개별 아미노산 서열을 포함한다(예로 Lottspeich and Engels (Hrsg.) (2006) Bioanalytik, 2. Auflage. Spektrum Akademischer Verlag, Elsevier, Muenchen, Germany 참고). 예를 들어 시아노젠 브로마이드 또는 시아노젠 클로라이드는 Met 잔기 뒤의 펩티드 결합을 절단한다; 히드록실아민은 아스파라기닐-글리실 결합을 절단한다; 포름산은 Asp-Pro를 절단한다; 2-(2'-니트로페닐설페닐)-3-메틸-3-브로모이놀레닌, 2-요오도소벤조산 또는 N-클로로숙신이미드는 Trp 뒤를; 2-니트로-5-티오시아나토벤조산은 Cys 뒤를 절단한다. 또한, Pro/Ala 폴리펩티드 또는 폴리펩티드 절편 앞의 잔기는 부위-지정 돌연변이화에 의해 Met으로 치환될 수 있고, 이어서 생성 융합 단백질은 BrCN으로 절단될 수 있음이 포함된다. 유사하게, 절단 부위를 포함하는 다른 아미노산 서열이 부위-지정 돌연변이화에 의해 재조합 융합 단백질 또는 그 인코딩 핵산 내로 도입될 수 있다.
- [0155] 또한, 유용한 프로테아제 인식/절단 부위가 당분야에 공지되어 있다. 여기에는 비제한적으로 하기가 포함된다: 트립신, 키모트립신, 엔테로키나아제, 담배 예치 바이러스(TEV) 프로테아제, PreScission 프로테아제, HRV 3C

프로테아제, SUMO 프로테아제, 소르타아제 A, 그랜자임 B, 퓨린, 트롬빈, 인자 Xa 또는 자가 절단성 인테인. 인자 Xa는 N-말단 융합 파트너 및 Pro/Ala 폴리펩티드 또는 폴리펩티드 절편 사이에 삽입될 수 있는 아미노산 서열 IleGluGlyArg의 C-말단에서 펩티드 결합을 가수분해한다. 단백질 분해 절단을 달성하는 특히 단순한 방법은 Pro/Ala 폴리펩티드 또는 폴리펩티드 절편의 N-말단 Lys 또는 Arg 측쇄의 삽입 또는 치환 후 트립신으로의 소화에 의한 것이며, 이는 내부 Lys 또는 Arg 측쇄가 배제되는 한 Pro/Ala 폴리펩티드 또는 폴리펩티드 절편 내를 절단하지 않는다. 예시적 인식 부위는 비제한적으로 D-D-D-D-K(엔테로키나아제), ENLYFQ(G/S)(TEV 프로테아제), I-(E/D)-G-R(인자 Xa), L-E-V-L-F-Q-G-P(HRV 3C), R-X-(K/R)-R(퓨린), LPXTG(소르타아제 A), L-V-P-R-G(트롬빈) 또는 I-E-X-D-X-G(그랜자임 B)이다.

- [0156] 상기 본원의 개시에서 자명하듯이, 본 발명은 선택 분자, 예컨대 유용한 단백질, 약학적 활성 폴리펩티드 또는 소분자, 진단적으로 유용한 폴리펩티드 또는 소분자, 또는 특히 다른 산업 분야, 예컨대 식품 또는 종이 산업 또는 오일 회수에서의 다른 유용한 단백질 또는 소분자와 콘주게이션될 수 있는 재조합적으로 제조되는 생합성 랜덤 코일 폴리펩티드 및 폴리펩티드 절편을 제공한다. 따라서, 본 발명은 또한 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산을 제공하며, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어지고, 상기 핵산 분자는 하기를 포함한다:
- [0157] (i) 번역된 아미노산 및/또는 리더 서열을 인코딩하는 핵산 서열;
- [0158] (ii) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 서열; 및
- [0159] (iii) 번역 정지 코돈이거나 이를 나타내는 핵산 서열.
- [0160] 상기 핵산 분자는 선택적으로 (i) 및 (ii) 사이에 프로테아제 및/또는 화학적 절단 부위 및/또는 인식 부위를 포함할 수 있다.
- [0161] 또한, 상기 핵산 분자에 있어서, 기재된 첫 번째 두 핵산 분자(즉 프로테아제 및/또는 화학적 절단 부위 및/또는 인식 부위)의 맥락에서 상기 본원에서 제공되는 구현예가 여기에 필요한 변경을 가하여 적용된다.
- [0162] 본 발명의 맥락에서 채용될 유용한 예시적 신호 서열은 비제한적으로 다음과 같은 원핵생물 서열: OmpA, MalE, PhoA, DsbA, pelB, Afa, npr, STII 또는 다음과 같은 진핵생물 서열을 포함한다: 꿀벌 멜리틴 신호 서열, 산성 당단백질 gp67 신호 서열, 마우스 IgM 신호 서열, hGH 신호 서열.
- [0163] 특히 본 발명의 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 인코딩하는 핵산 분자는 하기 본원에서 또한 제공되고 첨부된 실시예 및 도에 예시되는 바와 같은 방법에서 유용하다. 이러한 발현되는 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편은, 예컨대 이러한 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 발현하는 숙주 세포로부터 단리될 수 있다. 이러한 숙주 세포는, 예를 들어 본원에서 제공되는 벡터로 형질감염된 세포일 수 있다.
- [0164] 따라서 본원에서 기재되는 핵산 분자 또는 벡터로 세포를 형질감염시키는 것이 포함된다. 추가 구현예에서, 본 발명은 발현시 본 발명의 랜덤 코일 폴리펩티드(또는 이들의 절편) 또는 생물학적 활성 단백질을 인코딩하는 핵산 분자에 관한 것이다. 그러나 추가 구현예에서, 본 발명은 발현시 수용액 중에서 또는 생리학적 조건하에서 전적으로 또는 부분적으로 랜덤 코일 입체형태를 형성하는/취하는 본원에 개시되는 폴리펩티드를 인코딩하는 핵산 분자에 관한 것이다. 상기 핵산 분자는 폴리펩티드의 적절한 전사 및 번역을 보장하기 위해 당분야에 공지된 적합한 발현 조절 서열뿐만 아니라 기관으로의 표적화 또는 세포 분비를 보장하기 위한 신호 서열에 융합될 수 있다. 이러한 벡터는 추가 유전자, 예컨대 적합한 숙주 세포에서 적합한 조건하에 상기 벡터의 분비를 허용하는 마커 유전자를 포함할 수 있다.
- [0165] 바람직하게는 본 발명의 핵산 분자는 본원에 기재되는 생물학적 활성 단백질을 인코딩하는 핵산 분자가 원핵생물 또는 진핵생물 세포에서 발현될 수 있도록 하는 발현 조절 서열로 작동적으로 연결된 재조합 벡터에 포함된다. 상기 핵산 분자의 발현은 핵산 분자의 번역가능한 mRNA로의 전사를 포함한다. 원핵생물 숙주 세포에서의 발현을 허용하는 조절 성분에는, 예컨대 *E. coli*의 람다 PL, lac, trp, tac, ara, phoA, tet 또는 T7 프로모터가 포함된다. 진핵생물 세포, 바람직하게는 포유류 세포 또는 효모에서의 발현을 보장하는 가능한 조절 성분은 당분야의 숙련자에게 널리 공지되어 있다. 이들은 통상 전사 개시를 보장하는 조절 서열 및 선택적으로 전사 종료 및 전사체의 안정화를 일으키는 폴리-A 신호를 포함한다. 추가적 조절 성분에는 전사 증강제뿐만 아니라 번역 증강제 및/또는 자연적으로 연관된 또는 이중성 프로모터 영역이 포함될 수 있다. 진핵생물 숙주 세포에



서의 발현을 허용하는 조절 성분의 예로는 효모의 AOX1 또는 GAL1 프로모터 또는 CMV, SV40, RSV(라우스 육종 바이러스) 프로모터, CMV 증강제, SV40 증강제 또는 포유류 및 다른 동물 세포에서의 글로빈 인트론이 있다. 전사 개시에 관여하는 성분과는 별도로, 상기 조절 성분은 코딩 영역의 하류에 전사 종료 신호, 예컨대 SV40-폴리-A 부위 또는 tk-폴리-A 부위를 또한 포함할 수 있다.

[0166] 당분야의 숙련자에게 널리 공지된 방법을 이용하여 재조합 벡터를 구축할 수 있다(예를 들어 [Sambrook (1989), Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory NY and Ausubel (1989), Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, NY]에 기재된 기법 참고). 이러한 맥락에서, 적합한 발현 벡터, 예컨대 Okayama-Berg cDNA 발현 벡터 pcDV1(Pharmacia), pCDM8, pRc/CMV, pcDNA1, pcDNA3, pPICZ알파 A (Invitrogen), 또는 pSPORT1(GIBCO BRL)이 당분야에 공지되어 있다. 또한, 사용되는 발현 시스템에 따라, 폴리펩티드를 세포성 구획으로 보내거나 이를 배양 배지로 분비할 수 있는 리더 서열을 본 발명의 핵산 분자의 코딩 서열에 첨가할 수 있다.

[0167] 본 발명은 또한 본 발명의 랜덤 코일 폴리펩티드(또는 이들의 절편) 또는 생물학적 활성 단백질을 인코딩하는 핵산 분자를 포함하는 유전적 조작에 통상 채용되는 벡터, 특히 플라스미드, 코스미드, 바이러스, 및 박테리오파지에 관한 것이다. 바람직하게는 상기 벡터는 발현 벡터 및/또는 유전자 전달 또는 표적화 벡터이다. 바이러스, 예컨대 레트로바이러스, 박시니아 바이러스, 아데노 연관 바이러스, 헤르페스 바이러스 또는 소 유두종 바이러스 유래 발현 벡터를 표적화된 세포군 내로 본 발명의 폴리뉴클레오티드 또는 벡터를 전달하는데 사용할 수 있다.

[0168] 본 발명의 핵산 분자를 함유하는 벡터는 세포 유형에 따라 변하는, 널리 공지된 방법에 의해 숙주 세포내로 형질감염될 수 있다. 따라서, 본 발명은 또한 상기 핵산 분자 또는 상기 벡터를 포함하는 세포에 관한 것이다. 상기 방법에는, 예를 들어 [Sambrook (1989), 위 인용문] 및 [Ausubel (1989), 위 인용문]에 기재된 기법이 포함된다. 따라서 염화칼슘 형질감염 또는 전기천공이 보통 원핵생물 세포에 대해 이용되는 반면, 인산칼슘 치료 또는 전기천공이 다른 세포 숙주에 대해 이용될 수 있다(Sambrook (1989), 위 인용문). 추가 대안으로서, 본 발명의 핵산 분자 및 벡터는 표적 세포로 전달하기 위한 리포솜 내로 재구성될 수 있다. 숙주 세포에 존재하는 본 발명의 핵산 분자 또는 벡터는 숙주 세포의 게놈 내로 통합되거나 염색체 밖에서 유지될 수 있다. 따라서 본 발명은 또한 본 발명의 핵산 분자 및/또는 벡터를 포함하는 숙주 세포에 관한 것이다. 폴리펩티드 발현을 위한 숙주 세포는 당분야에 널리 공지되어 있으며, 원핵생물 세포뿐만 아니라 진핵생물 세포, 예컨대 *E. coli* 세포, 효모 세포, 무척추동물 세포, CHO 세포, CHO-K1 세포, HEK293 세포, HeLa 세포, COS-1 원숭이 세포, 흑색종 세포, 예컨대 Bowes 세포, 마우스 L-929 세포, Swiss 유래 3T3 세포주, Balb-c 또는 NIH 마우스, BHK 또는 HaK 햄스터 세포주 등을 포함한다.

[0169] 추가적인 측면에서, 본 발명에는 본 발명의 콘주게이트뿐만 아니라 본 발명의 (숙주)세포를 배양하고 상기 랜덤 코일 폴리펩티드(또는 이들의 절편) 또는 콘주게이트 또는 생물학적 활성 단백질을 본원에서 기재되는 배양에서 단리하는 것을 포함하는 본원에서 제공되는 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편) 또는 생물학적 활성 단백질의 제조 방법이 포함된다. 일반적으로 본 발명의 랜덤 코일 폴리펩티드(또는 이들의 절편), 랜덤 코일 도메인을 포함하는 콘주게이트 또는 생물학적 활성 단백질은 재조합 DNA 기술에 의해, 예컨대 본 발명의 생물학적 활성 단백질 또는 랜덤 코일 폴리펩티드(또는 이들의 절편)를 인코딩하는 기재된 핵산 분자 또는 벡터를 포함하는 세포를 배양하고, 상기 단백질/폴리펩티드를 배양으로부터 단리함으로써 제조될 수 있다. 본 발명의 생물학적 활성 단백질 또는 랜덤 코일 폴리펩티드(또는 이들의 절편)는 원핵생물 세포, 예컨대 *E. coli* BL21, KS272 또는 JM83, 또는 진핵생물 세포, 예컨대 피치아 파스토리스, 효모 균주 X-33 또는 CHO 세포를 포함하는 임의의 적합한 세포 배양 시스템에서 제조될 수 있다. 나아가 당분야에 공지된 적합한 세포주는 ATCC(American Type Culture Collection)와 같은 세포주 기탁처에서 수득할 수 있다.

[0170] "원핵생물"이라는 용어는 박테리아 세포를 포함하려는 것인 반면, "진핵생물"이라는 용어는 효모, 고등 식물, 곤충 및 포유류 세포를 포함하려는 것이다. 형질전환된 숙주는 최적 세포 성장을 달성하기 위해 당분야에 공지된 기법에 따라 발효기에서 성장 및 배양될 수 있다. 추가 구현예에서, 본 발명은 생물학적 활성 단백질 또는 랜덤 코일 폴리펩티드(또는 이들의 절편)의 발현에 적합한 조건하에 본 발명의 세포를 배양하고 세포 또는 배양 배지로부터 상기 단백질/폴리펩티드를 단리하는 것을 포함하는 상기 기재된 랜덤 코일 폴리펩티드(또는 이들의 절편) 또는 생물학적 활성 단백질의 제조 공정에 관한 것이다.

[0171] 본 발명의 랜덤 코일 폴리펩티드(또는 이들의 절편) 자체는 하나의 N-말단의 일차적인(또는 프롤린의 경우, 이차적인) 아미노기 및 중합체 C-말단에서 하나의 카르복실레이트기를 제외하고는 바람직하게는 임의 화학적 반응

기를 포함하지 않는다. 그러나 당분야의 숙련자에게는, 예를 들어 상기 랜덤 코일 폴리펩티드/중합체가 "융합 단백질"/"융합 컨스트럭트"의 일부인 경우 본원에서 제공되는 생합성 랜덤 코일 폴리펩티드/중합체가 화학적 반응기를 포함할 수 있다는 것이 자명하다. 또한, 상기에 기재된 바와 같이, 생합성 랜덤 코일 폴리펩티드(또는 이들의 절편)는, 예를 들어 (i) N-말단 Met 잔기/시작 코돈의 도움에 의한 세포질에서의 직접 발현; (ii) N-말단 신호 펩티드, 예를 들어 OmpA, PhoA(Monteilhet (1993) Gene. 1993 125:223-228), 펠리틴(Tessier (1991) Gene 98:177-183), 인터루킨 2(Zhang (2005) J Gene Med 7:354-365), hGH(Pecceu (1991) Gene 97(2):253-258) 등을 통한 분비에 뒤따르는 세포내 절단으로 야기되는 성숙 N-말단, 예컨대 Ala 또는 Pro; (iii) 또 다른 가용성 단백질, 예컨대 N-말단의 말토오스-결합 단백질과 배치된 프로테아제 절단 부위를 갖는 융합 단백질로서의 발현(Kapust and Waugh (2000) Protein Expr. Purif. 19:312-318)에 뒤따르는 시험관 내 또는 생체 내 특이적 프로테아제 절단에 의한 성숙 N-말단, 예컨대 Ala 또는 Pro를 갖는 아미노산 중합체/폴리펩티드의 방출과 같은 당분야의 숙련자에게 널리 공지된 방법에 따라 몇몇 방식으로 형질전환된 세포에서의 재조합 발현에 의해 제조될 수 있다. 또 다른 적합한 융합 파트너는 실시예 20 및 21에 기재된 바와 같은 SUMO 프로테아제에 의해 절단될 수 있는 SUMO 단백질이다. 나아가 융합 파트너에는 비제한적으로 글루타티온-S-트랜스퍼라아제, 티오레독신, 셀룰로오스-결합 도메인, 알부민-결합 도메인, 형광 단백질(예컨대 GFP), 단백질 A, 단백질 G, 인테인 등이 포함된다(Malhotra (2009) Methods Enzymol. 463:239-258).

[0172] 상기 설명된 바와 같이, 기재된 랜덤 코일 폴리펩티드(또는 폴리펩티드 절편)/중합체는 주로 알라닌 및 프롤린 잔기로 이루어지는 반면, O- 또는 N-글리코실화를 위해 필요한 세린, 트레오닌 또는 아스파라긴은 바람직하게는 존재하지 않는다. 따라서, 폴리펩티드 자체 또는 랜덤 코일 폴리펩티드(또는 폴리펩티드 이들의 절편)를 포함하는 생물학적 활성 단백질, 또는 일반적으로 랜덤 코일 폴리펩티드(또는 이들의 폴리펩티드 절편)를 포함하는 융합 단백질의 제조는 놀랍게도 Pro-Ala 서열내에서 바람직하게는 번역 후 개질이 없는 단분산성 산물을 생성할 수 있다. 이는 종종 복잡한 단백질의 생합성을 위해 선택되는 진핵생물 세포, 예컨대 중국 햄스터 난소 세포(CHO) 또는 효모에서의 재조합 단백질 제조를 위한 장점이다. 예를 들어 효모는 허가된 치료적 단백질, 예컨대 인슐린, 과립구-대식구 집락 자극 인자, 혈소판-유래 성장 인자 또는 허루딘(Gerngross (2004) Nat. Biotechnol. 22:1409-1414)의 제조를 위해 사용되었다. CHO 세포는 치료적 단백질, 예컨대 응고 인자 IX, 인터페론 β-1a, 테넥토플라아제(Chu (2001) Curr. Opin. Biotechnol. 12:180-187) 또는 고나도트로핀의 제조를 위해 사용되었으며, 당성분이 기능적 활성, 폴딩, 이량체화, 분비뿐만 아니라 수용체 상호작용, 신호 전달, 및 대사성 제거와 같은 여러 측면에 긍정적인 영향을 미칠 수 있다(Walsh (2006) Nat. Biotechnol. 24:-1241-1252). 따라서, 진핵생물 발현 시스템에서의 본 발명의 컨스트럭트, 랜덤 코일 폴리펩티드 및 콘주게이트의 제조가 또한 본 발명의 맥락에서 개시된다.

[0173] 본원에서 제공되는 수단 및 방법으로, 이제 (i) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편 및 (ii) 나아가 관심 분자, 예컨대 유용한 단백질, 단백질 절편 또는 소분자를 포함하는 본원에 개시되는 콘주게이트 및 분자를 제조하고 제공할 수 있다. 따라서 본 발명은 또한 상기 콘주게이트뿐만 아니라 생합성 랜덤 코일 폴리펩티드 및/또는 이를 포함하는 분자 또는 콘주게이트의 제조 또는 제작 방법을 제공한다. 따라서 본 발명은 또한 콘주게이트, 예컨대 약물 콘주게이트, 식품 콘주게이트, 진단 콘주게이트 등에 포함되는 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편의 제조 및/또는 제작 방법을 제공한다. 또한, 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 포함하는 생물학적 활성 단백질 또는 콘주게이트의 제조 및/또는 제작 방법을 제공한다. 또한, 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하고 추가적으로 상기 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 포함하는 폴리펩티드의 제조 및/또는 제작 방법이 제공된다. 이러한 방법에는 특히 (하나의 단계로서) 상기 본원에서 제공되는 (숙주) 세포의 배양 및 (추가 단계로서) 배양으로부터 또는 상기 세포로부터 상기 랜덤 코일 폴리펩티드 또는 생물학적 활성 단백질 및/또는 상기 생물학적 활성 단백질 및/또는 상기 폴리펩티드 콘주게이트의 단리가 포함된다. 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 상기 단리된 랜덤 코일, 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편뿐만 아니라 단리된 콘주게이트는 추가 처리될 수 있다. 예를 들어, 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편은 첨부된 실시예에 또한 나타낸 바와 같이 관심 분자에 화학적으로 연결되거나 커플링될 수 있다. 또한, 대안으로서, 관심 분자는 트랜스글루타미나아제(Besheer (2009) J Pharm Sci. 98:4420-8) 또는 다른 효소(Subul (2009) Org. Biomol. Chem. 7:3361-3371)를 통해 프롤린 및 알라닌 아미노산 잔기로 이루어진 아미노산 서열을 포함하는 상기 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편으로 효소적으로 콘주게이션될 수 있다.

[0174] 랜덤 코일 폴리펩티드(또는 이들의 절편) 및/또는 랜덤 코일 폴리펩티드(또는 이들의 절편) 및 관심 단백질, 예

컨대 생물학적 또는 치료적 활성 단백질 또는 예컨대 진단 방법에서 사용될 단백질을 포함하는 단백질 콘주게이트는 (특히) 성장 배지, 세포성 용해물, 원형질막 또는 세포막 분획으로부터 단리될 수 있다(또 다시, 본 발명은 의학적 또는 약학적 설정에서 유용한 (단백질) 콘주게이트로만 제한되지 않는다. 본원에서 제공되는 수단 및 방법은 또한 비제한적으로 식품 및 음료 산업, 영양 산업, 종이 산업, 바이오시약 산업, 연구 도구 및 시약 산업, 효소가 사용되는 산업, 화장료 산업, 오일 가공 및 오일 회수 등과 같은 다른 산업 분야에서의 용도를 갖는다). 본 발명의 발현된 폴리펩티드의 단리 및 정제는 황산암모늄 침전, 친화도 정제, 칼럼 크로마토그래피, 겔 전기영동 등을 포함하는 임의의 통상적 수단으로 수행될 수 있으며(Scopes (1982) "Protein Purification", Springer, New York, NY) 예컨대 본 발명의 생물학적 활성 단백질에 융합된 태그에 대해 지정된 모노클로날 또는 폴리클로날 항체의 사용이 관여될 수 있다. 예를 들어 단백질은 침부된 실시예에 기재된 스트렙타비딘 친화도 크로마토그래피를 이용하여 *Strep-tag II*를 통해 정제될 수 있다(Skerra (2000) *Methods Enzymol.* 326:271-304). 특히 약학적 사용/용도를 위해, 적어도 약 90 내지 95%의 동질성(단백질 수준에서)의 실질적으로 순수한 폴리펩티드가 바람직하며, 98 내지 99% 이상의 동질성이 가장 바람직하다. 제조 절차에 채용되는 숙주 세포/유기체에 따라, 본 발명의 폴리펩티드는 글리코실화되거나 비글리코실화될 수 있다.

[0175] 추가로 본 발명은 약제의 제조를 위한, 생물학적 활성 단백질, 랜덤 코일 폴리펩티드(또는 이들의 절편) 또는 콘주게이트, 예컨대 본 발명의 약물 콘주게이트, 본 발명의 핵산 분자, 본 발명의 벡터 또는 본 발명의 (숙주) 세포의 용도에 관한 것으로, 상기 생물학적 활성 단백질 또는 약물(또는 임의 다른 관심 소분자 또는 단백질)은 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하지 않거나 연결되지 않은 대조 분자에 비해 증가된 생체 내 및/또는 시험관 내 안정성을 가지며, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다.

[0176] 또 다른 구현예에서, 본 발명은 본원에서 기재되는 생물학적 활성 단백질 또는 약물 콘주게이트를 이러한 치료를 필요로 하는 포유류에게 투여하는 단계를 포함하는, 상기 생물학적 활성 단백질 또는 약물의 개선된 안정성의 이득을 얻는 질환 및/또는 장애의 치료 방법에 관한 것이다. 본 발명의 단백질 또는 약물 콘주게이트의 생물학적 활성에 따라, 당업자는 어느 질환/장애가 본 발명의 특정 생물학적 활성 단백질 또는 약물 콘주게이트로 치료될 것인지를 쉽게 결정할 수 있다. 일부 비제한적 예를 하기 표에 열거한다:

표 1

생물학적 활성 단백질 (또는 생물학적 활성 성분/ 이들의 단편) 또는 약물	치료할 장애/질환
과립구 집락 자극 인자	암 및/또는 화학치료 관련 호중구 감소증
인간 성장 호르몬	성장 호르몬 결핍 관련 저혈당 및/또는 성장 부진
인터페론 α	암, 바이러스 감염, C형 간염
인터페론 β	자가면역질환, 다발성 경화증
인터페론 γ	바이러스 감염
종양 괴사 인자	암
인터루킨-20	건선
α-갈락토시다아제 A	파브리 질환
미오스타틴 길항제	근육감소증
위 억제 폴리펩티드	2형 당뇨병
알파-1 안티트립신	효소 대체 치료, 낭성 섬유증, 만성 폐색성 폐 질환, 급성 호흡 증후군, 중증 천식
에리트로포이에틴	빈혈
응고 인자 VIII	혈우병
gp120/gp160	HIV
가용성 종양 괴사 인자 I 및 II 수용체	염증 질환
레테플라아제	혈전용해, 심근경색
엑센딘-4	당뇨병
인터루킨-1 수용체 길항제(IL-1ra; 아나킨라)	자가 면역 질환, 류머티스성 관절염
인터루킨-2	암
인슐린	당뇨병
아사프라기나아제	급성 림프모구 백혈병, 비호지킨성 임파종
온코나아제	악성 중피종 및 다른 유형의 암
스트렙토키나아제	혈전성 장애

호중구 젤라티나아제 연관 리포칼린	미생물 감염, 신장 재관류 손상
단일 도메인 항체, 단일쇄 및 CDR 모사 펩티드 및 CDR을 포함하는 다른 조작된 단편을 포함하는 항체 및 이들의 단편	면역학적, 종양학적, 혈관신생 및 감염성 질환 등
과립구-대식구 집락 자극 인자	화학치료 관련 호중구 감소증
난포 자극 호르몬	불임
글루코세레브로시다아제	고셔 질환
티모신 알파 1	만성 B형 간염, 압
글루카곤	저혈당
소마토스타틴	말단비대
아데노신 데아미나아제	아데노신 데아미나아제 결핍
인터루킨-11	혈소판 감소증
응고 인자 VIIa	혈우병
응고 인자 IX	혈우병
헤마티드	빈혈
인터페론 λ	C형 간염
렙틴	지방이상증, 비만, 알츠하이머 질환, I형 당뇨병
인터루킨-22 수용체 서브유닛 알파(IL-22ra)	건선
인터루킨 22	전이성 흑색종
히알루로니다아제	고혈압
섬유아세포 성장 인자 18	골관절염
섬유아세포 성장 인자 21	II형 당뇨병, 비만, 이상지질혈증, 대사성 장애
글루카곤 유사 펩티드 1	당뇨병
오스테오프로테게린	압, 골관절염, 류머티스성 관절염
IL-18 결합 단백질	류머티스성 관절염
성장 호르몬 방출 인자	HIV-연관 지방이상증
가용성 TACI 수용체	전신성 홍반성 루푸스, 다발성 경화증, 류머티스성 관절염
트로보스폰딘-1	압
가용성 VEGF 수용체 Flt-1	압
IL-4 뮤테인(IL-4 수용체 길항제)	천식
시클로스포린	기관 거부
푸마길린	압
날트렉손	알코올 의존성
옥트레오티드	말단비대, 유암종 종양
테두글루티드	짧은 창자 증후군, 크론씨 질환
고세렐린	진행성 전립선암, 유방암
캄프토테신	압
반코마이신	그람 양성 폐렴

[0178] 상기에 따라, 생물학적 활성 단백질, 랜덤 코일 폴리펩티드(또는 이들의 절편), 약물 콘주게이트, 핵산, 박터 또는 세포는 바람직하게는 증가된 생체 내 및/또는 시험관 내 안정성을 갖거나 부여하는 약제, 특히 생물학적 활성 단백질 및/또는 약물 성분의 제조를 위해, 예를 들어 호르몬 결핍 또는 관련 장애, 자가-면역 질환, 압, 빈혈, 혈관신생 질환, 감염성/염증성 질환, 혈전증, 심근경색, 당뇨병, 불임, 고셔 질환, 간염, 저혈당, 말단비대, 아데노신 데아미나아제 결핍, 혈소판 감소증, 혈우병, 빈혈, 비만, 알츠하이머 질환, 지방이상증, 건선, 전이성 흑색종, 골관절염, 이상지질혈증, 류머티스성 관절염, 전신성 홍반성 루푸스, 다발성 경화증, 천식, 골다공증, 및 재관류 손상 또는 다른 신장 질환의 치료를 위해 사용될 수 있다. 하나의 구현예에서, 생물학적 활성 단백질, 약물 콘주게이트, 핵산, 박터 또는 세포는 상기 생물학적 활성 단백질/약물 콘주게이트의 증가된 생체 내 및/또는 시험관 내 안정성을 갖는 약제로서의 용도를 위한 것이다. 유사하게 생물학적 활성 단백질, 랜덤 코일 폴리펩티드(또는 이들의 절편), 약물 콘주게이트, 핵산, 박터 또는 세포는, 예를 들어 호르몬 결핍 또는 관련 장애, 자가-면역 질환, 증식성 장애, 예컨대 압, 빈혈, 혈관신생 질환, 감염성 및/또는 염증성 질환, 혈전증, 심근경색 뇌졸중, 당뇨병, 불임, 음경 기능부전, 고셔 질환, 파브리 질환, 근육감소증, 남성 섬유증, 폐색성 폐 질환, 급성 호흡기 증후군, 간염, 저혈당, 말단비대, 아데노신 데아미나아제 결핍, 혈소판 감소증, 혈우병, 빈혈, 비만, 알츠하이머 질환, 지방이상증, 건선, 전이성 흑색종, 골관절염, 이상지질혈증, 류머티스성 관절염, 전신성 홍반성 루푸스, 다발성 경화증, 천식, 골다공증, 및 재관류 손상 또는 다른 신장 질환의 치료를 위한 치료에서의 용도를 위한 것이다.



- [0179] 본 발명은 또한 핵산 분자, 벡터뿐만 아니라 본원에서 제공되고 본 발명의 핵산 분자 또는 벡터를 포함하는 형질감염된 세포의 의학적 접근법, 예컨대 세포 기반 유전자 치료 접근법 또는 핵산 기반 유전자 치료 접근법에서의 용도에 관한 것이다.
- [0180] 추가 구현예에서, 본원에서 제공되는 랜덤 코일 폴리펩티드(또는 폴리펩티드 이들의 절편), 생물학적 활성의 이중성 단백질/단백질 컨스트럭트 또는 약물 또는 식품 콘주게이트 또는 생합성 랜덤 코일 폴리펩티드(또는 폴리펩티드 이들의 절편) 및/또는 본 발명의 핵산 분자 또는 벡터 또는 숙주 세포를 포함하는 다른 콘주게이트)는 조성물의 일부이다. 상기 조성물은 하나 이상의 본 발명의 랜덤 코일 폴리펩티드(또는 폴리펩티드 이들의 절편), 생물학적 활성 단백질, 식품 콘주게이트, 관심 콘주게이트, 약물 콘주게이트 또는 핵산 분자, 이를 인코딩하고/하거나 발현하는 벡터 및/또는 숙주 세포를 포함할 수 있다. 상기 조성물은 선택적으로 약학적으로 허용가능한 담체 및/또는 희석제를 추가로 포함하는 약학적 조성물일 수 있다. 추가 구현예에서, 본 발명은 이러한 약학적 조성물의 흡수를 필요로 하는 질환의 예방, 치료 또는 완화를 위한 약학적 조성물의 제조를 위한 본원에 기재되는 생물학적 활성 단백질, 랜덤 코일 폴리펩티드(또는 이들의 절편) 또는 약물 콘주게이트의 용도에 관한 것이다.
- [0181] 상기 본원에서 언급되는 바와 같이, 본원에 개시되는 콘주게이트, 예컨대 약물 콘주게이트 또는 진단 콘주게이트, 및/또는 생물학적 활성의 이중성 단백질/단백질 컨스트럭트(본 발명의 랜덤 코일 폴리펩티드 또는 이들의 폴리펩티드 절편을 포함하는 것)이 특히 의학적 또는 약학적 용도만을 위해 개시되는 것은 아니다. 또한, 상기 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편은 자체가 이러한 의학적 맥락에서, 예를 들어 "혈장 증량제" 또는 혈액 대용품으로서, 손상된 혈액 혈장 양 또는 혈액 혈장 함량에 관련된 장애의 완화, 예방 및/또는 치료에 또는 손상된 혈액 부피 관련 장애의 완화, 예방 및/또는 치료에 채용될 수 있다. 혈장 증량제로 치료되는 장애로는 비제한적으로 혈액 손실, 예컨대 부상, 수술, 화상, 외상 또는 급성 복증, 감염, 탈수 등에 관련된 장애가 있다. 그러나 이러한 의학적 용도는 본 발명의 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편에만 제한되지 않으며, 본원에서 개시된 바와 같은 특정 약물 콘주게이트 또는 특정 생물학적 활성의 이중성 단백질/단백질 컨스트럭트에도 연장될 수 있다.
- [0182] 하나의 구현예에서, 본원에서 기재되는 조성물은 선택적으로 검출을 위한 적합한 수단을 더 포함하는 진단 조성물, 예를 들어 조영 시약일 수 있으며, 상기 진단 조성물은 증가된 생체 내 및/또는 시험관 내 안정성을 갖는다.
- [0183] 본 발명의 조성물은 고체 또는 액체 형태일 수 있으며, 특히 분말(들), 정제(들), 용액(들) 또는 에어로졸(들)의 형태일 수 있다. 또한, 본 발명의 약제가 약학적 조성물의 의도하는 용도에 따라 추가적 생물학적 활성제를 포함할 수 있다는 것이 포함된다.
- [0184] 적합한 (약학적) 조성물의 투여는 상이한 방식, 예컨대 비경구, 피하, 정맥내, 동맥내, 복강내, 국소, 기관지내, 폐내 및 비강내 투여, 그리고 국소 치료를 의도하는 경우 병소내 투여에 의해 수행될 수 있다. 비경구 투여에는 복강내, 근육내, 피내, 피하, 정맥내 또는 동맥내 투여가 포함된다. 본 발명의 조성물은 또한 표적 부위로 직접, 예컨대 기관으로 특이적으로 수행되는 외부 또는 내부 표적 부위로의 바이오리스틱 전달에 의해 투여될 수 있다.
- [0185] 적합한 약학적 담체, 부형제 및/또는 희석제의 예는 당분야에 널리 공지되어 있으며, 인산염 완충 식염수 용액 또는 다른 완충액 용액, 물, 에멀션, 예컨대 오일/물 에멀션, 다양한 유형의 습윤제, 멸균 용액 등이 포함된다.
- [0186] 상기 담체를 포함하는 조성물은 널리 공지된 통상적 방법에 의해 제형화될 수 있다. 적합한 담체에는 본 발명의 생물학적 활성 단백질/약물 콘주게이트와 조합되는 경우 그 생물학적 및/또는 약학적 활성을 보유하는 임의의 재료가 포함될 수 있다(Remington's Pharmaceutical Sciences (1980) 16th edition, Osol, A. Ed, Mack Publishing Company, Easton, PA 참고). 비경구 투여용 제제에는 멸균 수성 또는 비수성 용액, 현탁액, 및 에멀션이 포함될 수 있다. 약학적 조성물의 맥락에서 채용되는 완충액, 용매 및/또는 부형제는 바람직하게는 상기 본원에서 정의되는 "생리학적"인 것이다. 비수성 용매의 예로는 프로필렌 글리콜, 폴리에틸렌 글리콜, 식물성 오일, 예컨대 올리브 오일, 및 주사가능한 유기 에스테르, 예컨대 에틸 올레레이트가 있다. 수성 담체에는 식염수 및 완충 배지를 포함하여 물, 알코올성 용액/수용액, 에멀션 또는 현탁액이 포함된다. 비경구 운반체에는 염화나트륨 용액, 링거 텍스트로오스, 텍스트로오스 및 염화나트륨, 락테이트화 링거액, 또는 고정 오일이 포함될 수 있다. 정맥내 운반체에는 유체 및 영양 보충물, 전해질 보충물(예컨대 링거 텍스트로오스에 기반한 보충물) 등이 포함될 수 있다. 보존제 및 항균제, 항산화제, 킬레이트제 및/또는 불활성 기체 등을 포함하는 다른 첨가제가 또한 존재할 수 있다. 또한, 본 발명의 약학적 조성물은 단백질성 담체, 예컨대 바람직하게는 인

간 기원의 혈청 알부민 또는 면역글로불린을 포함할 수 있다.

[0187] 이러한 약학적 조성물은 적합한 용량으로 개체에 투여될 수 있다. 투여 요법은 담당 의사 및 임상적 요인에 의해 결정될 것이다. 의학 분야에 널리 공지된 바와 같이, 임의의 한 환자를 위한 투여량은 환자의 체격, 신체 표면적, 연령, 투여될 특정 화합물, 성별, 시간 및 투여 경로, 일반 건강, 및 현재 투여되는 다른 약물을 포함하는 여러 요인에 의존한다. 약학적 활성 물질은 용량별로 1 µg 내지 20 mg/체중 kg, 예컨대 0.1 mg 내지 10 mg/체중 kg, 예컨대 0.5 mg 내지 5 mg/체중 kg의 양으로 존재할 수 있다. 요법이 연속 주입인 경우, 분당 1 µg 내지 10 mg/체중 kg의 범위여야 한다. 그러나 나타낸 예시적 범위 미만이거나 이를 초과하는 용량도 특히 상기 언급된 요인을 고려하여 포함된다.

[0188] 또한, 본 발명의 약학적 조성물은 약학적 조성물의 의도하는 용도에 따라 추가 생물학적 또는 약학적 활성제를 포함할 수 있다는 것이 포함된다. 이러한 추가 생물학적 또는 약학적 활성제는, 예컨대 항체, 항체 단편, 호르몬, 성장 인자, 효소, 결합 분자, 사이토카인, 케모카인, 핵산 분자 및 약물일 수 있다.

[0189] 본 발명이 약학적 조성물에 제한되지 않는다는 것이 주지된다. 또한, 연구에서 또는 진단제(들)로서 사용될 조성물이 포함된다. 예를 들어 본원에서 정의되는 랜덤 코일 도메인 또는 성분을 포함하는 생물학적 활성 단백질 또는 약물 콘주게이트가 진단 설정에서 사용된다는 것이 포함된다. 상기 목적을 위해, 본 발명의 생물학적 활성 단백질 또는 본 발명의 약물 콘주게이트는 검출을 허용하기 위해 표지될 수 있다. 상기 표지에는 비제한적으로 방사활성 표지(예컨대 [<sup>3</sup>H]수소 [<sup>125</sup>I]요오다이드 또는 [<sup>123</sup>I]요오다이드), 형광 표지(녹색 형광 단백질 (GFP)과 같은 형광 단백질 또는 플루오레신 이소티오시아네이트(FITC)와 같은 형광단 포함) 또는 NMR 표지(예컨대 가돌리늄 킬레이트)가 포함된다. 여기에서 정의된 표지 또는 마커는 전혀 제한적인 것이 아니며, 단순히 예시적 예를 나타낸다. 본 발명의 진단 조성물은 추적 또는 조영 실험 또는 진단 의학 설정에서 특히 유용하다. 첨부된 실시예 및 도에서, (i) 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편(상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기로 이루어진다), 및 (ii) 플루오레신 또는 디곡시제닌을 포함하는 콘주게이트를 포함하는 해당 컨스트럭트의 제조가 제공된다; 첨부된 도 13 및 14 및 해당 도 설명뿐만 아니라 예시적 실시예 22를 참고.

[0190] 본원에서 제공되는 수단 및 방법의 약학적 또는 진단적 용도만이 본 발명의 요지 내에 속하는 것은 아니다. 본원에서 제공되는 화합물/콘주게이트는 또한 특정한 다른 산업 분야, 예컨대 식품 산업, 음료 산업, 화장품 산업, 오일 산업, 종이 산업 등에서도 유용하다. 따라서 본 발명은 이러한 산업 분야에서 본원에서 제공되는 생합성 랜덤 코일 폴리펩티드의 용도를 또한 제공한다. 따라서 본 발명의 일부는 또한, 화장품, 화장 트리트먼트에서 사용될 화합물, 식품 또는 음료의 제조 방법이며, 상기 방법에는 본원에서 정의되는 랜덤 코일 폴리펩티드를 인코딩하거나 생물학적 활성 단백질을 인코딩하는 핵산 분자(또는 벡터) 및/또는 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 생물학적 활성 단백질 및/또는 폴리펩티드를 포함하는 세포의 배양이 포함된다. 이러한 방법에는 또한 상기 랜덤 코일 폴리펩티드, 상기 생물학적 활성 단백질 및/또는 활성, 예컨대 생물학적 활성을 갖거나 매개하는 아미노산 서열이거나 이를 포함하고 배양 또는 상기 세포로부터 상기 랜덤 코일 폴리펩티드 또는 랜덤 코일 폴리펩티드 절편을 추가적으로 포함하는 상기 생물학적 활성 단백질 또는 상기 폴리펩티드의 단리가 포함된다. 동일한 맥락에서, 다른 관심 콘주게이트, 예를 들어 상이한 산업 영역, 예컨대 오일 또는 종이 산업에서 유용한 콘주게이트를 제조할 수 있다. 당분야의 숙련자는 해당 분자적/제조법 컨스트럭트의 생성뿐만 아니라 프롤린 및 알라닌 아미노산 잔기만으로 이루어진 아미노산 서열을 포함하는 생합성 랜덤 코일 폴리펩티드 또는 폴리펩티드 절편을 포함하는 콘주게이트의 생성을 위해 본원에서 제공되는 수단 및 방법을 쉽게 채용할 수 있는 입장에 있으며, 상기 아미노산 서열은 50개 이상의 프롤린(Pro) 및 알라닌(Ala) 아미노산 잔기, 및 관심 소분자 또는 폴리펩티드로 이루어진다.

[0191] 또 다른 구현예에서, 본 발명은 랜덤 코일 폴리펩티드(또는 폴리펩티드 이들의 절편), 생물학적 활성 단백질, 약물 콘주게이트, 상기 생물학적 활성 단백질을 인코딩하고/하거나 활성(예를 들어 생물학적 활성)을 갖거나 매개하는 아미노산 서열이거나 이를 포함하는 상기 생물학적 활성 단백질을 인코딩하고/인코딩하거나 상기 폴리펩티드를 인코딩하는 핵산 분자, 상기 핵산 분자를 포함하는 벡터 또는 본원에서 기재되는 상기 핵산 또는 상기 벡터를 포함하는 세포를 포함하는 키트를 제공한다. 본 발명의 키트는 추가로 완충액(들), 저장 용액 및/또는 의학적, 과학적 또는 진단적 분석 및 목적의 수행을 위해 필요한 추가적 시약 또는 재료를 포함할 수 있다. 또한, 본 발명의 키트의 일부는 바이알 또는 병에 개별적으로 또는 용기 또는 다중용기 단위에 조합으로 포장될 수 있다.

[0192] 본 발명의 키트는 특히 본 발명의 방법을 수행하기 위해 유리할 수 있으며, 본원에서 언급되는 다양한 용도, 예

컨대 진단 키트, 연구 도구 또는 의학적 도구로서 채용될 수 있다. 추가적으로 본 발명의 키트에는 과학적, 의학적 및/또는 진단적 목적을 위해 적합한 검출 수단이 함유될 수 있다. 키트의 제작은 바람직하게는 당분야의 숙련자에게 공지된 표준 절차를 따른다.

**도면의 간단한 설명**

[0193] 본 발명을 하기 비제한적 도 및 실시예로 추가로 예시한다.

**도 1. PA#1 Pro/Ala 중합체/폴리펩티드 서열을 위한 유전자 설계.**

PA#1(서열 번호 1)에 대한 구성 블록의 뉴클레오티드 및 인코딩된 아미노산 서열이 두 상보적 올리고데옥시뉴클레오티드(상부/코딩쇄 올리고데옥시뉴클레오티드 서열 번호 17, 하부/비코딩쇄 올리고데옥시뉴클레오티드 서열 번호 18)의 혼성화에 의해 획득되었다. 생성 핵산은 각각 Ala 코돈 및 안티-코돈에 해당하는 두 점착성 말단(아래 첨자로 나타냄)을 가지며, 상호 호환성을 갖는다. 이러한 구성 블록의 반복 라이게이션(ligation)시, 다양한 길이의 Pro-Ala 폴리펩티드를 인코딩하는 동일계능 연속집합체는, 예를 들어 *SapI* 제한 부위(들)를 통해 획득되고 이어서 클로닝될 수 있다.

**도 2. Fab 단편 또는 인간 IFNa2b에 대한 융합으로서의 Pro/Ala 중합체/폴리펩티드 서열에 대한 클로닝 전략.**

(A) pASK75의 유도체인 pASK88-Fab-2x*SapI*(서열 번호 22) 상에 인코딩된 항체 Fab 단편의 면역글로불린 경쇄의 C-말단 주위의 뉴클레오티드 및 인코딩된 아미노산 서열 연장부(상부/코딩쇄 서열 번호 19, 하부/비코딩쇄 서열 번호 20, 인코딩된 아미노산 서열 서열 번호 21)을 Pro/Ala 중합체/폴리펩티드 서열의 서브클로닝 및 해당 생물학적 활성 단백질의 발현을 위해 사용하였다. 상기 뉴클레오티드 서열은 상호 역전된 배향의 *SapI* 인식 부위를 수반하여 절단시 도 1에 나타난 합성 유전자 카세트와 호환성을 갖는 돌출 DNA 말단이 생성된다. 경쇄의 인식 서열 및 C-말단 아미노산을 밀출쳐서 나타낸다.

(B) 도 1에 나타난 바와 같은 단일 카세트의 pASK88-Fab-2x*SapI* 플라스미드 내로의 삽입 후 20개 잔기를 갖는 PA#1 중합체/폴리펩티드의 뉴클레오티드 서열 및 인코딩된 아미노산 서열(상부/코딩쇄 서열 번호 23, 하부/비코딩쇄 서열 번호 24, 인코딩된 아미노산 서열 서열 번호 25). 이러한 10개의 반복 카세트의 유사한 라이게이션/삽입으로 200개 잔기를 갖는 중합체/폴리펩티드(서열 번호 26 및 27)를 코딩하는 플라스미드 벡터 pFab-PA#1(200)(서열 번호 28)이 생성되었다. Pro/Ala 중합체 인코딩 서열에 인접한 *SapI* 제한 부위가 표지된다(인식 서열을 밀출쳐서 나타낸다).

(C) pFab-PA#1(200)(서열 번호 28)의 플라스미드 지도. Fab-PA#1(200)의 중쇄(HC) 및 경쇄(LC)의 구조 유전자는 테트라사이클린 프로모터/작동유전자(*tet<sup>p/o</sup>*)의 전사 조절하에 있으며, 오페론은 지단백질 종료자(*t<sub>1pp</sub>*)로 끝난다. HC에는 박테리아 *OmpA* 신호 펩티드, 가변(VH) 및 제 1 인간 IgG1 중쇄 불변 C 도메인(CH)뿐만 아니라 His<sub>6</sub>-태그가 포함된다. LC에는 박테리아 *PhoA* 신호 펩티드, 가변(VL) 및 인간 경쇄 불변(CL) 도메인, 200개 잔기를 갖는 PA#1 중합체/폴리펩티드가 포함된다. *XbaI* 및 *HindIII* 제한 부위에 인접한 발현 카세트 밖 pFab-PA#1(200)의 플라스미드 골격은 일반 클로닝 및 발현 벡터 pASK75(Skerra (1994) Gene 151:131-135)의 골격과 동일하다. 단일 제한 부위를 나타낸다.

(D) pASK-IFNa2b(서열 번호 32) 상에 클로닝된 인간 IFNa2b의 N-말단 주위 뉴클레오티드 및 아미노산 서열 연장부(상부/코딩쇄 서열 번호 29, 하부/비코딩쇄 서열 번호 30, 인코딩된 아미노산 서열 서열 번호 31). Pro/Ala 중합체-인코딩 서열의 삽입을 위해 이용될 수 있는 단일 제한 부위 *SapI*을 표지한다(인식 서열을 밀출쳐서 나타낸다). *Strep*-태그 II의 두 C-말단 아미노산을 밀출쳐서 나타낸다. 성숙 IFNa2b의 첫 번째 아미노산이 +1로 표지된다.

(E) 도 1에 나타난 하나의 PA#1 중합체 서열 카세트의 삽입 후 IFNa2b의 N-말단의 뉴클레오티드 및 인코딩된 아미노산 서열 연장부(상부/코딩쇄 서열 번호 33, 하부/비코딩쇄 서열 번호 34, 인코딩된 아미노산 서열 서열 번호 35). Pro/Ala 중합체-인코딩 서열의 삽입 후 유지되는 단일 제한 부위 *SapI*이 표지된다(인식 서열을 밀출쳐서 나타낸다). 융합 단백질의 일부로서의 IFNa2b의 첫 번째 아미노산이 표지되며(1), *Strep*-태그 II의 두 C-말단 아미노산을 밀출쳐서 나타낸다. 10개의 반복된 PA#1 중합체 서열 카세트의 유사한 라이게이션/삽입으로 200개 잔기를 갖는 중합체/폴리펩티드를 코딩하는 플라스미드 벡터 pPA#1(200)-IFNa2b(서열 번호 36)가 생성되었다.

(F) pPA#1(200)-IFNa2b(서열 번호 37)의 플라스미드 지도. 생물학적 활성 단백질 PA#1(200)-IFNa2b에 대한 구

조 유전자(박테리아 OmpA 신호 펩티드, *Strep*-태그 II, 200개 잔기를 갖는 PA#1 중합체/폴리펩티드 절편, 및 인간 IFNa2b 포함)는 테트라사이클린 프로모터/작동유전자( $tet^{p/o}$ )의 전사 조절하에 있으며 지단백질 종료자( $t_{1pp}$ )로 끝난다. *Xba*I 및 *Hind*III 제한 부위가 인접한 발현 카세트 밖 플라스미드 골격은 일반 클로닝 및 발현 벡터 pASK75(Skerra (1994) 위 인용문)의 골격과 동일하다. 단일 제한 부위를 나타낸다.

**도 3. 정제된 재조합 Fab 단편 및 정제된 재조합 IFNa2b 뿐만 아니라 이들의 Pro/Ala 폴리펩티드/중합체 용합물의 SDS-PAGE에 의한 분석.**

재조합 단백질을 *E. coli* KS272(Strauch (1988) Proc. Natl. Acad. Sci. USA 85:1576-80)에서 원형질막 분비를 통해 제조하고 각각 고정화된 금속 또는 스트렙타비딘 친화도 크로마토그래피를 이용하여 His<sub>6</sub>-태그(Fab) 또는 *Strep*-태그 II(IFNa2b)의 수단에 의해 정제하였다.

(A) 정제된 재조합 Fab 및 200개 잔기를 갖는 그 PA#1 용합물의 12% SDS-PAGE에 의한 분석. 겔은 각각 Fab 및 Fab-PA#1(200)의 2  $\mu$ g 단백질 표본을 나타낸다. 좌측의 표본은 2-메르캅토에탄올을 사용하여 환원된 반면, 우측에 반복되는 표본은 비환원된 채로 남겨 두었다. 환원 조건하에 처리한 단백질 마커의 크기를 좌측 가장자리에 나타낸다. 사슬내 디설피드 가교의 환원 시, Fab 단편 및 그 200개 잔기의 PA#1 용합물은 두 균질한 밴드로 나타낸다. 환원된 Fab 단편의 경우, 각각 약 24 및 26 kDa의 분자 크기를 갖는 두 밴드는 분리된 LC 및 HC에 해당한다. 환원된 Fab-PA#1(200) 용합 단백질의 경우, 24 kDa의 밴드는 HC에 해당하는 반면, 약 90 kDa의 밴드는 PA#1(200) 폴리펩티드 절편과 용합된 LC에 해당한다. 비환원 조건하에서, Fab 단편 및 그 PA#1(200) 용합물은 각각 약 45 kDa 및 100 kDa의 겔보기 분자 크기를 갖는 하나의 균질한 밴드로 나타낸다. Fab-PA#1(200) 용합 단백질에 대한 두 겔보기 크기 값은 비환원된 Fab-PA#1(200)에 대한 64.3 kDa 및 단리된 LC-PA#1(200)에 대한 39.1 kDa의 계산 질량보다 상당히 더 크다. 계산 질량이 48.0 kDa인 Fab 단편 자체, 또는 그 용합되지 않은 경쇄는 본질적으로 정상적인 전기영동 이동성을 나타내기 때문에, 이러한 효과는 Pro/Ala 중합체/폴리펩티드 절편의 첨가에 기인한다.

(B) 정제된 재조합 IFNa2b 및 200개 잔기를 갖는 그 PA#1 용합 단백질의 12% SDS-PAGE에 의한 분석. 겔은 IFNa2b 및 PA#1(200)-IFNa2b 각각의 2  $\mu$ g 단백질 표본을 나타낸다. 좌측의 표본은 2-메르캅토에탄올로 환원된 반면 우측의 대응 표본은 미환원된 채로 남아 있다. 환원 조건하에 처리된 단백질 마커의 크기를 좌측 가장자리에 나타낸다. 두 단백질은 환원된 상태에서 겔보기 분자 크기가 약 20 kDa 및 약 80 kDa인 하나의 균질한 밴드를 나타낸다. 후자의 값은 PA#1(200)-IFNa2b에 대한 계산 질량 37.0 kDa 보다 상당히 더 크다. 계산 질량이 20.9 kDa인 IFNa2b 자체는 본질적으로 정상적인 전기영동 이동성을 나타내기 때문에, 이러한 효과는 Pro/Ala 중합체/폴리펩티드 절편의 첨가에 기인한다. 비환원된 상태의 IFNa2b는 약간 더 높은 전기영동 이동성을 가져서 그 두 분자내 디설피드 가교의 결과로 인한 보다 컴팩트한 형태를 시사한다.

**도 4. 정제된 재조합 Fab 및 IFNa2b 뿐만 아니라 이들의 PA#1(200) 용합물의 수력학적 부피에 대한 정량적 분석.**

(A) Fab 및 Fab-PA#1(200)의 분석용 크기 배제 크로마토그래피(SEC). 0.25 mg/ml 농도의 정제된 단백질 250  $\mu$ l를 PBS 완충액으로 평형화된 Superdex S200 10/300 GL 칼럼에 처리하였다. 280 nm에서의 흡광도를 모니터링하고 각각의 크로마토그래피 주행 피크를 값 1로 정규화하였다. 화살표는 칼럼의 공극 부피(7.8 ml)를 나타낸다.

(B) Superdex S200 10/300 GL 칼럼을 이용한 크로마토그램(A)의 적정 곡선. 마커 단백질(사이토크롬 c, 12.4 kDa; 카르본산 안하이드라아제, 29.0 kDa; 오발부민, 43.0 kDa; 소 혈청 알부민, 66.3 kDa; 알코올 탈수소효소, 150 kDa,  $\beta$ -아밀라아제, 200 kDa, 아포-페리틴, 440 kDa)의 분자량(MW)의 로그를 이들의 용출 부피에 대비하여 도시하고(검은색 원) 직선으로 조정하였다. Fab 단편 및 그 PA#1 용합 단백질의 관찰된 용출 부피(검은색 사각형)로부터, 이들의 겔보기 분자 크기를 다음과 같이 결정하였다. Fab: 31 kDa(실제 질량: 48.0 kDa); Fab-PA#1(200): 237 kDa(실제 질량: 64.3 kDa). 이러한 데이터는 PA#1 폴리펩티드와의 용합이 훨씬 더 증가된 수력학적 부피를 부여함을 보여준다.

(C) IFNa2b 및 PA#1(200)-IFNa2b의 분석용 크기 배제 크로마토그래피. 농도 0.25 mg/ml의 각각의 정제된 단백질 250  $\mu$ l를 인산염 완충 식염수 PBS로 평형화된 Superdex S200 10/300 GL 칼럼에 처리하였다. 280 nm에서의 흡광도를 모니터링하고 각각의 크로마토그래피 수행 피크를 값 1로 정규화하였다. 화살표는 칼럼의 공극 부피(7.8 ml)를 나타낸다.



(D) Superdex S200 10/300 GL 칼럼을 이용한 크로마토그램(C) 적정 곡선. 마커 단백질의 분자량(MW) 로그(B참고)를 이들의 용출 부피에 대비하여 도식화하고(검은색 원) 직선으로 조정하였다. IFNa2b 및 그 PA#1 융합 단백질의 관찰 용출 부피(검은색 사각형)로부터 이들의 겔보기 분자 크기를 다음과 같이 결정하였다. IFNa2b: 22.5 kDa(실제 질량: 20.9 kDa); PA#1(200)-IFNa2b: 229.0 kDa(실제 질량: 37.0 kDa). 이러한 데이터는 PA#1 폴리펩티드와의 융합이 훨씬 더 증가된 수력학적 부피를 제공함을 나타낸다.

**도 5. 제조합 단백질 및 이들의 PA#1 중합체/폴리펩티드 융합물의 원형 광이색형(CD) 분광측정에 의한 실험적 2차 구조 분석.**

실온에서 50 mM K<sub>2</sub>SO<sub>4</sub>, 20 mM K-포스페이트(pH 7.5) 중에서 스펙트럼을 기록하고 각각의 단백질에 대한 몰 타원율을  $\Theta_M$ 으로 정규화하였다.

(A) 정제된 제조합 Fab 및 Fab-PA#1(200)의 CD 스펙트럼. Fab 단편에 대한 CD 스펙트럼은 박테리아에 의해 생산되는 Fab 단편의 정확한 폴딩을 나타내는 216 nm 주위의 넓은 네거티브 최대 부위를 갖는 주로  $\beta$ -시트인 단백질의 전형적 특징을 나타낸다(Sreerama in: Circular Dichroism-Principles and Applications (2000) Berova, Nakanishi 및 Woody (Eds.) Wiley, New York, NY, pp. 601-620). Pro/Ala 중합체/폴리펩티드를 갖는 그 융합 단백질의 스펙트럼은 200 nm 이하에서 우세한 네거티브 밴드를 보여 랜덤 코일 입체형태를 시사한다. 또한, 220 nm 근처에 솔더(shoulder) 부위가 존재하는데, 이는 Fab 단편의  $\beta$ -시트 기여로 인한 것이며 융합 단백질의 일부도 정확히 폴딩되었음을 시사한다.

(B) Fab 단편에 대한 스펙트럼의 감산으로부터 수득되는 Fab-PA#1(200)에 대한 몰 차이 CD 스펙트럼. 차이 CD 스펙트럼은 200개 잔기의 PA#1 중합체/폴리펩티드 절편의 이차적 구조를 나타내며, 200 nm 주위의 강력한 최소 부위를 드러내어 완충 수용액 중 랜덤 코일 입체형태를 명확히 나타낸다(Greenfield (1969) Biochemistry 8:4108-4116; Sreerama (2000) 위 인용문; Fandrich (2002) EMBO J. 21:5682-5690).

(C) 정제된 제조합 IFNa2b 및 PA#1(200)-IFNa2b의 CD 스펙트럼. IFNa2b에 대한 CD 스펙트럼은 208 nm 및 220 nm 주위에 두 네거티브 밴드를 갖는 우세한  $\alpha$ -나선 단백질의 전형적 특징을 나타내며(Sreerama (2000) 위 인용문.), 이는 박테리아에 의해 생산되는 인간 IFNa2b의 정확한 폴딩을 시사한다. Pro/Ala 중합체/폴리펩티드를 갖는 그 융합 단백질의 스펙트럼은 200 nm 근처에 우세한 최소 부위를 갖는 특징적 일탈을 나타내어 랜덤 코일 입체형태를 시사한다. 또한, 220 nm 근처에 존재하는 솔더 부위는 IFNa2b의  $\alpha$ -나선 기여에 기인하는 것으로 IFNa2b가 융합 단백질의 일부로 정확히 폴딩되었음을 시사한다.

(D) IFNa2b에 대한 스펙트럼의 감산으로부터 수득되는 PA#1(200)-IFNa2b에 대한 몰 차이 CD 스펙트럼. 차이 CD 스펙트럼은 200개 잔기의 PA#1 중합체/폴리펩티드 절편의 이차적 구조를 나타내며, 200 nm 주위의 강력한 최소 부위를 드러내어 (B)에 나타난 것과 본질적으로 동일하다. 이는 또다시 본 발명에 따른 Pro 및 Ala 잔기를 포함하는 생물학적 중합체의 완충 수용액 중에서의 랜덤 코일 입체형태를 명확히 나타낸다.

**도 6. CHO 세포에서 인간 성장 호르몬(hGH) 및 유전적으로 인코딩된 PA#1 중합체 간 융합 단백질의 분비적 제조.**

(A) pASK75-His6-hGH(서열 번호 41) 상에 클로닝된 hGH의 N-말단 주위 뉴클레오티드 및 아미노산 서열 연장부(상부/코딩쇄 서열 번호 38, 하부/비코딩쇄 서열 번호 39, 인코딩된 아미노산 서열 서열 번호 40). 서브클로닝을 위해 *Hind*III(나타내지 않음)와 함께 사용될 수 있는 단일 제한 부위 *Nhe*I, 및 Pro/Ala 중합체-인코딩 서열의 삽입을 위해 사용될 수 있는 *Sap*I이 표지된다(인식 서열을 밑줄쳐서 나타낸다). His<sub>6</sub>-태그의 6개 아미노산을 밑줄쳐서 나타낸다. hGH의 첫 번째 아미노산이 +1로 표지된다.

(B) 도 1에 나타난 하나의 PA#1 중합체 서열 카세트의 삽입 후 hGH의 N-말단의 뉴클레오티드 및 인코딩된 아미노산 서열(상부/코딩쇄 서열 번호 42, 하부/비코딩쇄 서열 번호 43, 인코딩된 아미노산 서열 서열 번호 44). 서브클로닝을 위해 사용될 수 있는 단일 제한 부위 *Nhe*I, 및 Pro/Ala 중합체-인코딩 서열의 삽입 후 유지되는 *Sap*I이 표지된다(인식 서열을 밑줄쳐서 나타낸다). 융합 단백질의 일부인 hGH의 첫 번째 아미노산이 표지되며 (1) His<sub>6</sub>-태그의 아미노산을 밑줄쳐서 나타낸다. 10회 반복된 PA#1 중합체 서열 카세트의 유사한 라이게이션/삽입으로 성숙 융합 단백질을 코딩하는 플라스미드 벡터 pASK75-His6-PA#1(200)-hGH(서열 번호 45)가 생성된다.

(C) pASK75-His6-PA#1(200)-hGH(서열 번호 46)의 플라스미드 지도. 생물학적 활성 단백질 His6-PA#1(200)-hGH(박테리아 *OmpA* 신호 펩티드, His<sub>6</sub>-태그, 200개 잔기를 갖는 PA#1 중합체/폴리펩티드 절편, 및 인간 GH

포함)의 구조 유전자는 테트라사이클린 프로모터/작동유전자( $tet^{p/o}$ )의 전사적 조절하에 있으며, 지단백질 종료자( $t_{1pp}$ )로 끝난다. *Xba*I 및 *Hind*III 제한 부위에 인접한 발현 카세트 외부의 플라스미드 골격은 일반 클로닝 및 발현 벡터 pASK75(Skerra (1994) 위 인용문)의 골격과 동일하다. 단일 제한 부위를 나타낸다.

(D) His<sub>6</sub>-PA#1(200)-hGH 융합 단백질(서열 번호 47)을 인코딩하는 pCHO-PA#1(200)-hGH의 플라스미드 지도. 인간 성장 호르몬 신호 펩티드(Sp), His<sub>6</sub>-태그, 200개 잔기를 갖는 PA#1 중합체/폴리펩티드 서열(PA#1(200)), 인간 성장 호르몬(hGH), 및 소 성장 호르몬 폴리아테닐화 신호(bGH pA)를 포함하는 구조 유전자는 사이토메갈리바이러스 프로모터(CMV<sup>P</sup>)의 전사적 조절하에 있다. 단일 제한 부위 *Nhe*I 및 *Hind*III를 나타낸다. 네오마이신 포스포트랜스퍼라아제에 대한 내성 유전자(neo)는 SV40 프로모터(SV40<sup>P</sup>)의 조절하에 있으며, SV40 폴리아테닐화 신호(SV40 pA)가 뒤따른다. 또한, 플라스미드는 증식 및 *E. coli*에서의 선별을 허용하는 박테리아 ColE1 복제 기점(ColE1-ori), 박테리오파지 f1 복제 기점(f1-ori), 및 β-락타마아제 유전자(bla)를 함유한다.

(E) 재조합 hGH 대비 CHO 세포에서 생산되는 200개 잔기의 유전적으로 인코딩된 PA#1 중합체 및 hGH간 융합 단백질의 웨스턴 블롯 분석. CHO-K1 세포를 pCHO-PA#1(200)-hGH(서열 번호 48) 또는 PA#1(200) 서열이 없는(그러나 His<sub>6</sub>-태그를 수반하는) hGH를 인코딩하는 유사한 플라스미드인 pCHO-hGH(서열 번호 49)로 형질감염시켰다. 형질감염 2일 후, 세포 배양 상청액 표본을 홀스 래디쉬 페록시다아제가 콘주게이션된 항-hGH 항체로 SDS-PAGE 및 웨스턴 블로팅하였다. 두 단백질은 화살표로 나타내는 겔보기 분자 크기 약 23 kDa(His<sub>6</sub>-hGH) 및 약 90 kDa(His<sub>6</sub>-PA#1-hGH)의 단일 밴드를 나타낸다. 또한, 배양 배지 중 혈청 단백질에 기인하는 약 60 kDa의 약한 밴드가 있다. His<sub>6</sub>-태그된 hGH는 계산 질량 23.5 kDa에 나타나지만, His<sub>6</sub>-PA#1-hGH의 겔보기 분자 크기는 그 계산 질량 39.5 kDa보다 훨씬 더 크다. 이 효과는 Pro-Ala 중합체의 친수성 랜덤 코일의 성질 때문이다.

**도 7. PA#1 Pro/Ala 폴리펩티드/중합체 서열에 대한 2차 구조의 이론적 예측.**

본 예시는 버지니아 대학의 서열 비교 및 2차 구조 예측 서버(URL: [http://fasta.bioch.virginia.edu/fasta\\_www2](http://fasta.bioch.virginia.edu/fasta_www2)) 상에서 구현되는 Chou-Fasman 방법(Chou and Fasman (1974) *Biochemistry* 13:222-245)에 따른 CHOFAS 컴퓨터 알고리즘으로부터의 결과를 나타낸다. 아미노 및 카르복시 말단에서의 경계 효과를 배제하기 위해, 도 1에 따른 20머 아미노산 반복체를 3 연속 복제물에 붙이고(합성 유전자 카세트의 반복 라이게이션/삽입 후 인코딩되는 것과 유사한 동일계놈 연속중합체 생성), 중심 20머 서열 블록(상자로 나타냄)에 대한 결과만을 고려하였다. PA#1 폴리펩티드 서열/절편(서열 번호 1)의 경우, Chou-Fasman 알고리즘은 100% α-나선 2차 구조를 예측한다. 이는 융합 단백질의 일부인 PA#1 폴리펩티드/폴리펩티드 절편에 대해 실험적으로 관찰되는 우세한 랜덤 코일 입체형태와 대비된다(도 5B/D 참고).

**도 8: BALB/c 마우스에서 200개 및 600개 잔기를 갖는 정제된 재조합 Fab 단편 및 그 PA#1 중합체 융합물의 약동학의 정량적 분석.**

실시에 16의 혈장 표본을 샌드위치 ELISA를 이용하여 Fab, Fab-PA#1(200), 및 Fab-PA#1(600) 농도에 대해 분석하였다. Fab, Fab-PA#1(200), 및 Fab-PA#1(600)의 혈장 반감기를 추정하기 위해, 측정된 농도 값을 정맥내 주사 후 시간에 대해 도식화하여 이중 지수 붕괴를 가정하여 수치적으로 조정하였다. 미융합 Fab 단편은 1.3 ± 0.1h의 제거 반감기를 가지며 매우 빠른 제거를 나타내었다. 대조적으로 Fab-PA#1(200) 및 Fab-PA#1(600)에 대해 결정된 제거상은 상당히 더 느려서, 각각 4.1 ± 1.8h 및 38.8 ± 11.2h의 말단 반감기를 가짐으로써, 미융합 Fab 단편 대비 200개 또는 600개 잔기를 갖는 Pro/Ala 중합체 융합물로 인한 약 3배 및 약 30배 연장된 순환을 나타내었다.

**도 9: 200개 잔기를 갖는 P1A1 또는 P1A3 중합체의 융합으로서의 정제된 재조합 Fab 단편의 분석.**

재조합 단백질을 원형질막 분비를 통해 *E. coli* KS272에서 생산하고 고정화된 금속 친화도 크로마토그래피를 이용하여 His<sub>6</sub>-태그 수단으로 정제하였다. 정제된 단백질을 12% SDS-PAGE로 분석하였다. 겔은 각각 Fab-P1A1(200) 및 Fab-P1A3(200) 뿐만 아니라 대조를 위한 미융합 Fab 단편(cf. 도 3A)의 2 μg 단백질 표본을 나타낸다. 좌측의 표본은 2-메르캅토에탄올로 환원된 반면 우측의 유사 표본은 환원되지 않은 채 남겨두었다. 환원 조건하에 처리한 단백질 마커의 크기를 좌측 가장자리에 나타낸다. 사슬간 디설피드 가교의 환원 후, Fab 단편 및 그 200개 잔기의 Pro/Ala 융합물은 두 균질한 밴드로 나타난다. 환원된 Fab 단편의 경우, 각각 약 24 및 26 kDa의 분자 크기를 갖는 두 밴드는 분리된 경쇄(LC) 및 중쇄 단편(HC)에 해당한다. 환원된 Fab-

P1A1(200) 융합 단백질의 경우, 24 kDa의 밴드는 HC에 해당하는 반면 약 90 kDa의 밴드는 P1A1(200) 폴리펩티드와 융합된 LC에 해당한다. 환원된 Fab-P1A3(200) 융합 단백질의 경우, 24 kDa의 밴드는 HC에 해당하는 반면, 약 75 kDa의 밴드는 P1A5(200) 폴리펩티드와 융합된 LC에 해당한다. 비환원 조건하에서, Fab 단편, 그 P1A1(200) 및 그 P1A3(200) 융합물은 각각 약 45 kDa, 110 kDa 및 90 kDa의 겔보기 분자 크기를 갖는 뚜렷한 단일 밴드로 나타난다. Fab-P1A1(200) 및 Fab-P1A3(200) 융합 단백질의 겔보기 크기는 비환원된 Fab-P1A1(200)에 대한 65.3 kDa 및 비환원된 Fab-P1A3(200)에 대한 64.0 kDa의 계산 질량보다 훨씬 더 크다. 또한, 대응 환원된 경색에 대한 겔보기 크기는 P1A1(200) LC에 대한 40.7 kDa 및 P1A3(200) LC에 대한 39.4 kDa의 계산 질량보다 훨씬 더 크다. 48.0 kDa의 계산 질량을 갖는 Fab 단편 자체 또는 23.4 kDa의 계산 질량을 갖는 그 미융합 경색은 본질적으로 정상적인 전기영동 이동성을 나타내기 때문에, 이러한 효과는 Pro/Ala 중합체/폴리펩티드 절편의 첨가에 기인한다.

**도 10. 정제된 재조합 Fab-P1A1(200) 및 Fab-P1A3(200) 융합 단백질의 수력학적 부피의 정량적 분석.**

Fab-P1A1(200) 및 Fab-P1A3(200)의 분석용 크기 배제 크로마토그래피(SEC). 농도 0.2 5mg/ml의 정제된 단백질 250  $\mu$ l를 PBS로 평형화된 Superdex S200 10/300 GL 칼럼에 처리하였다. 280 nm에서의 흡광도를 모니터링하고, 각각의 크로마토그래피 수행의 피크 값을 1로 정규화하였다. 화살표는 칼럼의 공극 부피(7.8 ml)를 나타낸다. 융합 단백질의 관찰된 용출 부피로부터, 이들의 겔보기 분자 크기를 하기와 같이 도 4B에 나타낸 대로 유사한 적정 곡선을 이용하여 결정하였다. Fab-P1A1(200): 180.7 kDa(실제 질량: 65.3 kDa); Fab-P1A3(200): 160.2 kDa(실제 질량: 64.0 kDa). 이러한 데이터는 단백질과 P1A1 및/또는 P1A5 폴리펩티드의 융합이 훨씬 더 증가된 수력학적 부피를 부여함을 보여준다.

**도 11. Fab-P1A1(200) 및 Fab-P1A3(200) 융합물의 원형 광이색형(CD) 분광측정에 의한 실험적 2차 구조 분석.**

스펙트럼을 실온에서 50 mM K<sub>2</sub>SO<sub>4</sub>, 20 mM K-포스페이트(pH 7.5) 중에서 기록하고 각각의 단백질에 대해 몰 타원율을  $\Theta_M$ 으로 정규화하였다.

(A) 정제된 재조합 Fab-P1A1(200) 및 Fab-P1A3(200)의 CD 스펙트럼. 두 Pro/Ala 중합체/폴리펩티드와의 Fab 융합 단백질 각각의 CD 스펙트럼은 랜덤 코일 입체형태를 시사하는 200 nm 아래에서 우세한 네거티브 밴드를 나타낸다. 또한, Fab 단편의  $\beta$ -시트 기여에서 기인하는 220 nm 근처의 솔더 부위가 존재하여 융합 단백질의 일부로서의 그 정확한 폴딩을 시사한다.

(B) 미융합 Fab 단편에 대한 스펙트럼의 감산으로 수득되는 Fab-P1A1(200) 및 Fab-P1A3(200)의 몰 차이 CD 스펙트럼(도 5A 참고). 차이 CD 스펙트럼은 각각 200개 잔기의 P1A1(서열 번호 51) 및 P1A3(서열 번호 3) 중합체/폴리펩티드 절편의 2차 구조를 나타내며, 완충 수용액 중 이들의 랜덤 코일 입체형태를 뚜렷이 시사하는 200 nm 근처의 강력한 최소 부위를 드러낸다(Greenfield (1969) Biochemistry 8:4108-4116; Sreerama (2000) 위 인용문; Fandrich (2002) EMBO J. 21:5682-5690).

**도 12: 단리된 생합성 Pro/Ala 중합체/폴리펩티드의 제조.**

(A) pSUMO-PA#1(200)(서열 번호 60)의 플라스미드 지도. 시작 메티오닌 코돈에 이어 라이신 코돈, 6개 연속 His 잔기의 N-말단 친화도 태그, 절단 가능한 작은 유비퀴틴-유사 개질제(SUMO) 단백질 Smt3p(Panavas (2009) Methods Mol Biol. 497:303-17), 및 200개 잔기를 갖는 PA#1 중합체/폴리펩티드 절편(서열 번호 60)을 포함하는 융합 단백질 MK-His(6)-SUMO-PA#1(200)의 구조 유전자는 박테리오파지 T7의 유전자 10 프로모터의 전사 조절 하에 있으며 t $\phi$  종료자로 끝난다. 추가 플라스미드 성분에는 복제 기점(ori), 앰피실린 내성 유전자(bla), 및 f1 복제 기점이 포함된다. NdeI 및 HindIII 제한 부위가 인접한 발현 카세트 외부의 플라스미드 골격은 SapI 제한 부위가 침묵 돌연변이에 의해 제거된 것을 제외하고는 일반 클로닝 및 발현 벡터 pRSET5a(Schoepfer (1993) 124:83-85)의 골격과 동일하다.

서열 번호 60이 동봉된 서열 목록(또한 본 설명 및 명세서의 일부임)에 제공되며 본원에서 아래에 재현된다.

```

gcacttttctg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa atacattcaa 60
atatgtatcc gctcatgaga caataaacct gataaatgct tcaataatat tgaaaaagga 120
agagtatgag tattcaacat ttccgtgtcg cccttattcc cttttttgcg gcattttgoc 180
ttcctgtttt tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg 240
gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt gagagttttc 300
gccccgaaga acggtttcca atgatgagca cttttaaagt tctgctatgt ggcgcggtat 360
tatcccgtat tgacgcccgg caagagcaac tcggtcgccg catacactat tctcagaatg 420
acttggttga gtaactacca gtcacagaaa agcatcttac ggatggcatg acagtaagag 480
aattatgcag tgctgcatac acctagatg ataacactgc ggccaactta cttctgacaa 540
cgatcggagg accgaaggag ctaaccgctt ttttgacaaa catgggggat catgtaactc 600
gccttgatog ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca 660
cgatgctgtg agcaatggca acaacgctgc gcaaactatt aactggcgaa ctacttactc 720
tagcttcccg gcaacaatta atgactgga ttggagcggg taaagtgtca ggaccacttc 780
tgcgctcggc ccttccggct ggcgtgttta ttgctgataa atctggagcc ggtgagcgtg 840
ggtctcgcgg tatcattgca gcaactgggc cagatggtaa gccctcccgt atcgtagtta 900
tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag 960
gtgcctcact gattaagcat ttgtaactgt cagaccaagt ttaactcatat atactttaga 1020
ttgattttaa acttcatttt taatttaaaa ggatctaggt gaagatcctt tttgataatc 1080
tcatgaccaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac cccgtagaaa 1140
agatcaaagg atcttcttga gatccttttt ttctgcgctg aatctgctgc ttgcaaaaa 1200
aaaaaccacc gctaccagcg agtggtttgg tgccggatca agagctacca actctttttc 1260
cgaaggtaac tggcttcagc agagcgcaga taccaaaatac tgccttcta gtgtagccgt 1320
agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgtc ctgctaatacc 1380
tgttaccagt ggtgctgcc agtggcgata agtgcgtgtc taccgggttg gactcaagac 1440
gatagttacc ggataaggcg cagcggctcg gctgaacggg gggttcgtgc acacagccca 1500
gcttggagcg aacgacctac accgaactga gatacctaca gcctgagcta tgagaaaagcg 1560
ccacgcttcc cgaagggaga aagcgggaca ggtatccggt aagcggcagg gtcggaaacg 1620
gagagcgcac gagggagctt acgaggggaa acgcctggta tctttatagt cctgtcgggt 1680
ttcggccact ctgacttgag cgtcgatttt tgtgatgctc gtcagggggg cggagcctat 1740
ggaaaaacgc cagcaacgcg gcccttttac ggttccctgg cttttgctgg ccttttctc 1800
acatgttctt tctgctgcta tccctgatt ctgtggataa ccgtattacc gcctttgagt 1860
gagctgatac cgtcgcgcgc agcgaacga ccgagcgcag cgagtcagtg agcgaggaag 1920
cggagaagcg cccaatacgc aaaaccgctc tccccgcgcg ttggccgatt cattaatgca 1980
ggatctcgat cccgcgaaat taatacagat cactataggg agaccacaac ggtttccctc 2040
tagaataaat tttgtttaac tttaaagagg agatatacat atgaaaacatc accaccatca 2100
ccattcggac tcagaagtca atcaagaagc taagccagag gtcaagccag aagtcaagcc 2160
tgagactcac atcaatttaa agtgttccga tggatcttca gaaatcttct ttaagatcaa 2220
aaagaccact cctttaagaa ggtgatgga agcgttcgct aaaagacagg gtaaggaat 2280
ggactcotta agattcttgt acgacggtat tagaattcaa gctgatcaga cccctgaaga 2340
tttgacatg gaggataaac atattattga ggtcacaga gaacagattg gtggcgcgc 2400
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2460
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2520
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2580
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2640
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2700
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2760
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2820
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2880
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 2940
tccagctgca cctgctccag cagcaactgc tgcaccagct ccggtgctc ctgctgccc 3000
aagagcaagc ttgatccggc tgcatacaa gcccgaaagg aagctgagtt ggtgctgccc 3060
accgctgagc aataactagc ataaccctt ggggcctcta aacgggtctt gaggggtttt 3120
ttgctgaaag gaggaactat atccggatc ggcgtaatag cgaagaggcc cgcaccgatc 3180
gccttccca acagttcgc agcctgatg gcgaatggga cgcgcctgt agcggcgcac 3240
taagcgcggc ggggtgtgtg gttacgcgca gcgtgaccgc tacacttgc agcgcctag 3300
cgcccgctcc tttcgtctt ttccttctc ttctgcacc gttgcgcggc tttcccgtc 3360
aagctctaaa tcgggggctc cctttagggt tccgatttag tgccttacgc cactcgacc 3420
ccaaaaact tgattaggt gatggttcc gtagtgggac atcgcctga tagacggtt 3480
ttcgccttt gacgttggg tccagttct ttaatagtg actctgttc caaactggaa 3540
caacactcaa cctatctc gtcctattct ttgatttata agggattttg ccgatttcgg 3600
cctattggtt aaaaaatgag ctgatttaac aaaaatttaa cgcgaatttt aacaaaaat 3660
taaccgttac aatttaggtg
    
```

(B) 박테리아에 의해 생산되는 His(6)-SUMO-PA#1(200) 융합 단백질 및 그 절단의 12% SDS-PAGE에 의한 분석. 겔은 *E. coli*에서 추출되고 고정화된 금속 친화도 크로마토그래피(IMAC) 및 크기 배제 크로마토그래피(SEC)를 통해 정제된 SUMO-PAS#1(200) 융합 단백질의 실시예 21에 기재된 바와 같은 Ub1-특이적 프로테아제 1(SUMO 프로테아제)로의 단백질 분해 절단 이전(레인 1) 및 이후(레인 2)를 나타낸다. 모든 표본은 2-메르캅토에탄올로 환원되었다. 환원 조건하에 처리한 단백질 마커(M)의 크기를 좌측 가장자리에 나타낸다. His(6)-SUMO-PA#1(200) 융합 단백질은 약 100 kDa의 겔보기 분자 크기를 갖는 균질한 단일 밴드로 나타난다. 따라서 SDS-PAGE에서 관찰된 His(6)-SUMO-PA#1(200) 융합 단백질에 대한 겔보기 크기는 계산 질량 28.3 kDa 보다 훨씬 더 크며, 이는 Pro/Ala 중합체/폴리펩티드의 존재 때문이다. 절단 후, 친수성 PA#1(200) 폴리펩티드는 쿠마시 블루에 의해 검출 가능한 정도로 염색되지 않는다; 따라서 단지 작은 잔여 분획의 융합 단백질 및 절단된 His(6)-SUMO 단백질만을 SDS 폴리아크릴아미드 겔(레인 2) 상에서 볼 수 있다. His(6)-SUMO 단백질은 그 계산 분자 질량 12.2 kDa



과 잘 일치하는 약 16 kDa의 겉보기 분자 크기를 갖는 균질한 밴드(레인 2)를 보여준다.

**도 13: 생합성 Pro/Ala 중합체/폴리펩티드와 화학적 화합물 및/또는 약물의 콘주게이션.**

(A-D) 생합성 PA#1(200) 중합체/폴리펩티드를 갖는 플루오레신 콘주게이트(서열 번호 61)의 제조를 분석용 크기 배제 크로마토그래피(SEC)를 통해 모니터링하였다. 패널은(상부에서 하부로) 정제된 His(6)-SUMO-PA#1(200)의 SEC 수행(A), SUMO 프로테아제의 존재하 절단 반응 후의 His(6)-SUMO-PA#1(200)(B), 플루오레신 NHS 에스테르로의 화학적 커플링 후 절단된 His(6)-SUMO-PA#1(200) 배치(C), 및 IMAC 정제 후 플루오레신-PA#1(200) 콘주게이트(D)를 보여준다. 농도 약 0.5 mg/ml의 단백질/폴리펩티드 250 μl를 Akta 정제기 시스템 상에서 PBS로 평형화된 Superdex S200 10/300 GL 칼럼에 처리하였다. 225 nm, 280 nm, 및 494 nm에서의 흡광도를 UV-900 UV/VIS 검출기(GE Healthcare)를 이용하여 모니터링하고, 각 크로마토그램의 뚜렷한 피크를 값 1로 정규화하였다. 화살표는 칼럼의 공극 부피(7.3 ml)를 나타낸다.

(E-K) 유리 플루오레신, 생합성 PA#1(200) 중합체/폴리펩티드, 및 그 플루오레신 콘주게이트의 SEC 및 UV VIS 분광측정을 통한 특징분석. 3 크로마토그램은 (상부에서 하부로) 정제된 PA#1(200)(E), 화학적 화합물 플루오레신(F) 및 정제된 플루오레신-PA#1(200) 콘주게이트(G)를 나타낸다. 4 UV/VIS 스펙트럼은 정제된 His(6)-SUMO-PA#1(200) 용합 단백질(H), 정제된 PA#1(200) 중합체/폴리펩티드(I), 유리 플루오레신(J), 및 정제된 플루오레신-PA#1(200) 콘주게이트(K)(모두 PBS 중)를 나타낸다. 화살표는 SUMO(280 nm), PA#1(200)(225 nm), 및 플루오레신(494 nm)의 특징적 흡광도 밴드/숄더 부위를 나타낸다.

(L) Superdex S200 10/300 GL 칼럼을 이용한 (A-G)의 크로마토그램에 대한 적정 곡선. 마커 단백질(아프로티닌, 6.5 kDa; 사이토크롬 C, 12.4 kDa; 카르보산 안하이드라아제, 29.0 kDa; 소 혈청 알부민, 66.3 kDa; 알코올 탈수소효소, 150 kDa; β-아밀라아제, 200 kDa; 아포-페리틴, 440 kDa)의 분자량(MW) 로그를 이들의 용출 부피(x)에 대해 도식화하고 직선으로 조정하였다. His(6)-SUMO-PA#1(200)(10.81 ml), PA#1(200)(11.51 ml), 플루오레신-PA#1(200)(11.49 ml) 및 플루오레신(27.57 ml)의 관찰된 용출 부피로부터, 이들의 겉보기 분자 크기를 하기와 같이 결정하였다. His(6)-SUMO-PA#1(200): 215.6 kDa, PA#1(200): 154.1 kDa(실제 질량: 16.1 kDa), 플루오레신-PA#1(200): 155.6 kDa(실제 질량: 16.6 kDa); SUMO: 25.7 kDa(실제 질량: 12.2 kDa); 플루오레신: 0.09 kDa(실제 질량: 0.33 kDa). 이러한 데이터는 Pro/Ala 폴리펩티드/중합체와의 용합이 미개질된 화합물에 비해 콘주게이션된 약물에 훨씬 더 증가된 수력학적 부피를 부여함을 나타낸다.

(M) 생합성 PA#1(200) 폴리펩티드/중합체 및 스테로이드 화합물 디곡시제닌간 화학적 콘주게이트의 전자 분무 이온화 질량 분광측정(ESI-MS)을 통한 특징분석. 디곡시제닌-PA#1(200)의 디컨볼루션한 ESI-MS 스펙트럼은 16671.4 Da의 질량을 드러내며, 이는 본질적으로 디곡시제닌-PA#1(200) 콘주게이트에 대한 계산 질량(16670.6 Da)과 일치한다.

**도 14: 생합성 PA#1(200) 폴리펩티드/중합체 및 소분자 약물간 화학적 콘주게이트의 예시.**

- (A) 생합성 PA#1(200)의 N-말단에 커플링된 플루오레신.
- (B) 생합성 PA#1(200)의 N-말단에 커플링된 디곡시제닌.

**발명을 실시하기 위한 구체적인 내용**

[0194] 본 발명의 더 나은 이해와 그 여러 장점을 제공하는 하기 예시적 비제한적 실시예에 의해 본 발명이 추가로 설명된다.

[0195] 달리 명시되지 않는 한, 예를 들어 [Sambrook (2001) 위 인용문]에 기재된 바와 같은 재조합 유전자 기술의 확립된 방법이 이용되었다.

**[0196] 실시예 1: Pro/Ala 아미노산 중합체/폴리펩티드를 위한 유전자 합성.**

[0197] 상기 본원에서 기재되는 Pro 및 Ala 잔기로 이루어진 아미노산 반복체는 본원에서 Pro/Ala 또는 "PA"로 나타낸다. Pro 및 Ala를 포함하는 반복 중합체 서열을 인코딩하는 유전자 단편(서열 번호 1에 해당하는 PA#1)을 도 1에 나타낸 두 상보적 올리고데옥시뉴클레오티드(서열 번호 17 및 서열 번호 18)의 혼성화, 이어서 이들의 상호 호환성이지만 비회문성 점착성 말단의 DNA 라이게이션을 통한 지정 방식의 동일계능 연속중합체 형성에 의해 수득하였다. 올리고데옥시뉴클레오티드를 ThermoScientific(Ulm, Germany)에서 구매하고 제조용 요소 폴리야크릴 아미드 겔 전기영동에 의해 정제하였다. 올리고데옥시뉴클레오티드의 핵산 서열을 도 1에 나타낸다(해당 점착

성 말단의 라이게이션시 하기 PA#1 서열 반복체의 일부가 되는 알라닌을 위한 추가적 GCC 코돈을 포함하는 서열 번호 17 및 18). 200 pmol의 두 올리고데옥시뉴클레오티드를 100  $\mu$ l의 50 mM 트리스/HCl(pH 7.6), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 1 mM ATP와 혼합하고 10 u 폴리뉴클레오티드 키나아제(MBI Fermentas, St. Leon-Rot, Germany)의 존재하에 37°C에서 30분 동안 인큐베이션함으로써 효소적 포스포릴화를 수행하였다. 80°C에서 10분 동안 변성 후 혼합물을 하룻밤 동안 실온으로 냉각하여 혼성화하였다. 이어서 상기 용액 50  $\mu$ l를 1u T4 DNA 리가아제(MBI Fermentas) 및 10  $\mu$ l 100 mM 트리스/HCl(pH 7.4), 50 mM MgCl<sub>2</sub>, 20 mM DTT, 10 mM ATP, 및 일부 경우에는 각각 5 mM의 dATP, dCTP, dGTP, 및 dTTP를 총 부피 100  $\mu$ l로 첨가하고 얼음 상에서 55분 동안 인큐베이션함으로써 라이게이션하였다. 70°C에서 10분 열 불활성화 후 라이게이션 산물을 TAE 완충액(40 mM 트리스, 20 mM 아세트산, 1 mM EDTA)의 존재하에 1.5%(w/v) 아가로오스 겔 전기영동에 의해 분리하였다. 에티뮴 브로마이드로의 염색 후, 300bp 길이의 조립된 유전자 절편에 해당하는 밴드를 절단하고 단리하였다.

**[0198] 실시예 2: Fab-PA#1 융합 단백질에 대한 발현 벡터로서의 pFab-PA#1(200)의 구축.**

**[0199]** 실시예 1의 PA#1의 20개 아미노산 서열을 코딩하는 합성 유전자 단편의 10머 반복체의 클로닝을 위해, 플라스미드 벡터 pASK88-Fab-2xSapI(서열 번호 22), 경쇄의 3' 말단에서 역상보적 배향의 두 SapI 제한 부위를 갖는 뉴클레오티드 서열(도 2A)을 포함하는 Fab 단편용 발현 플라스미드(Schlapschy (2007) Protein Eng. Des. Sel. 20:273-284)를 사용하였다. pASK75(Skerra, A. (1994) Gene 151:131-135)의 유도체인 상기 벡터를 SapI로 절단하고 새우 알칼린 포스파타아제(USB, Cleveland, OH)로 탈인산화하고 실시예 1에서 수득한 합성 DNA 단편의 300bp 카세트에 라이게이션하였다. 생성 중간체 플라스미드 pFab-PA#1(100)을 SapI로 절단하고 새우 알칼린 포스파타아제로 탈인산화하고 실시예 1에서 수득한 합성 DNA 단편의 300bp 카세트에 라이게이션하였다(도 2B에 예시된 대로, 그러나 PA#1(20) 중합체/폴리펩티드 카세트만으로). 생성 플라스미드는 설계된 pFab-PA#1(200)(서열 번호 28)였다(도 2C). 상기 플라스미드 상의 200개 잔기 PA#1 서열 반복체에 대한 코딩 영역은 두 SapI 절단 부위에 인접하여 정확한 절단, 나아가 5'-GCC 뉴클레오티드 오버행을 수반하는 전체 서열 카세트의 서브클로닝을 가능케함을 주지해야 한다.

**[0200]** *E. coli* XL1-Blue(Bullock (1987) Biotechniques 5:376-378)의 형질전환 후 플라스미드를 제조하고 클로닝된 합성 핵산 삽입물의 서열을 양측 서열 분석을 가능케 하는 올리고데옥시뉴클레오티드 프라이머뿐만 아니라 BigDye™ 종료자 키트를 이용하여 제한 분석 및 이중쇄 DNA 서열분석(ABI-Prism™ 310 유전 분석기, Perkin-Elmer Applied Biosystems, Weiterstadt, Germany)으로 확인하였다.

**[0201] 실시예 3: PA#1(200)-IFNa2b 융합 단백질에 대한 발현 벡터로서의 pASK-PA#1(200)-IFNa2b의 구축.**

**[0202]** 200개 잔기의 PA#1 서열 반복체 PA#1(200)과의 융합물로서의 IFNa2b를 인코딩하는 발현 플라스미드의 구축을 위해, pASK-IFNa2b(서열 번호 32)(도 2D)를 SapI로 절단하고 새우 알칼린 포스파타아제로 탈인산화하고 SapI로 절단 소화하여 앞서 구축한 플라스미드 pFab-PA#1(200)(실시예 2)으로부터 절단된 200개 잔기의 PA#1 폴리펩티드를 인코딩하는 유전자 단편과 라이게이션하였다(도 2E에 예시된 대로, 그러나 PA#1(20) 중합체/폴리펩티드 카세트만으로). *E. coli* JM83(Yanisch-Perron. (1985) Gene 33:103-119)의 형질전환 후, 플라스미드를 제조하고 정확한 삽입물의 존재를 제한 분석으로 확인하였다. 생성 플라스미드는 설계한 pPA#1(200)-IFNa2b였다(서열 번호 37)(도 2F).

**[0203] 실시예 4: Fab 단편 및 유전적으로 인코딩된 PA#1 중합체/폴리펩티드 간 융합 단백질의 박테리아를 이용한 생산 및 정제.**

**[0204]** Fab 단편(계산 질량: 48.0 kDa) 및 Fab-PA#1(200) 융합물(계산 질량: 64.3 kDa)을 폴딩 헬퍼 플라스미드 pTUM4(Schlapschy (2006) Protein Eng. Des. Sel. 19:385-390)와 함께 실시예 3의 해당 발현 플라스미드를 포함하는 *E. coli* KS272에서 22°C에서 100 mg/l 앰피실린 및 30 mg/l 클로람페니콜을 함유하는 2L LB 배지로 진탕기 플라스크 배양을 이용하여 제조하였다. 재조합 유전자 발현의 유도는 OD<sub>550</sub> = 0.5에서 0.4 mg 무수 테트라사이클린의 첨가에 의해 하룻밤 동안 수행하였다(전형적으로 수확 시 약 1.0의 OD<sub>550</sub> 생성). 50  $\mu$ g/ml 라이소자임을 함유하는 500 mM 수크로스, 1mM EDTA, 100mM 트리스/HCl(pH 8.0)의 존재하 원형질막 추출을 다른 곳에 기재된 바와 같이 수행한 뒤(Breustedt (2005) Biochim. Biophys. Acta 1764:161-173) 500 mM 베타인, 50 mM 나포스페이트(pH 7.5) 중 0 내지 200 mM의 이미다졸 구배로 고정화된 금속 친화도 크로마토그래피를 이용한 His<sub>6</sub>-태그 수단에 의해(Skerra (1994) Gene 141:79-84) 정제하였다.

**[0205]** 두 재조합 Fab 단편(도 3A)에 대해 균질한 단백질 제조물을 미융합 Fab에 대해서는 0.2 mg L<sup>-1</sup> OD<sup>-1</sup> 및 Fab-

PA#1(200)에 대해서는  $0.1 \text{ mg L}^{-1} \text{ OD}^{-1}$ 의 수율로 수득하였다. 높은 몰 농도의 트리스 완충액 시스템(Fling (1986) Anal. Biochem. 155:83-88)을 사용하여 SDS-PAGE를 수행하였다. Pro/Ala 중합체는 그 방향족 아미노산의 부재로 인해 UV 흡광도에 기여하지 않으므로 미융합 Fab 및 그 PA#1 중합체 용합물 모두에 대해  $68290 \text{ M}^{-1} \text{ cm}^{-1}$ 의 계산 흡광도 계수(Gill (1989) Anal. Biochem. 182:319-326)를 이용하여 280 nm에서의 흡광도에 따라 단백질 농도를 결정하였다.

[0206] 실시예 5: IFNa2b 및 유전적으로 인코딩된 PA#1 중합체/폴리펩티드 간 융합 단백질의 박테리아를 이용한 생산 및 정제.

[0207] IFNa2b(계산 질량: 20.9 kDa) 및 PA#1(200)-IFNa2b(계산 질량: 37.0 kDa)를 폴딩 헬퍼 플라스미드 pTUM4(Schlapschy (2006) 위 인용문)와 함께 실시예 3의 해당 발현 플라스미드를 포함하는 *E. coli* KS272에서 22°C에서 100 mg/l 앰피실린 및 30 mg/l 클로람페니콜을 함유하는 2L LB 배지로 진탕기 플라스크 배양을 이용하여 제조하였다. 제조합 유전자 발현의 유도는  $\text{OD}_{550} = 0.5$ 에서 0.4 mg 무수 테트라사이클린의 첨가에 의해 하루 동안 수행하였다(전형적으로 수확 시 약 1.0의  $\text{OD}_{550}$  생성). 50  $\mu\text{g/ml}$  라이소자임을 함유하는 500 mM 수크로스, 1 mM EDTA, 100 mM 트리스/HCl(pH 8.0)의 존재하 원형질막 추출을 다른 곳에 기재된 바와 같이 수행한 뒤(Breustedt (2005) 위 인용문) 150 mM NaCl, 1 mM EDTA, 100 mM 트리스/HCl(pH 8.0)의 존재하에 스트렙타비딘 친화도 크로마토그래피를 이용한 *Strep*-태그 II(Schmidt (2007) Nat. Protoc. 2:1528-1535)를 통해 정제하였다.

[0208] 두 제조합 IFNa2b 단백질(도 3B)에 대해 균질한 단백질 제조물을 IFNa2b에 대해서는  $0.15 \text{ mg L}^{-1} \text{ OD}^{-1}$  및 PA#1(200)-IFNa2b에 대해서는  $0.1 \text{ mg L}^{-1} \text{ OD}^{-1}$ 의 수율로 수득하였다. 높은 몰 농도의 트리스 완충액 시스템(Fling (1986) 위 인용문)을 사용하여 SDS-PAGE를 수행하였다. 미융합 IFNa2b 및 그 PA#1 중합체 용합물 모두에 대해  $23590 \text{ M}^{-1} \text{ cm}^{-1}$ 의 계산 흡광도 계수(Gill (1989) 위 인용문)를 이용하여 280 nm에서의 흡광도에 따라 단백질 농도를 결정하였다.

[0209] 실시예 6: Fab 단편 및 유전적으로 인코딩된 200개 잔기의 PA#1 중합체 간 제조합 융합 단백질에 대한 분석용 겔 여과에 의한 수력학적 부피 측정.

[0210] 수행 완충액으로 PBS(115 mM NaCl, 4 mM  $\text{KH}_2\text{PO}_4$ , 16 mM  $\text{Na}_2\text{HPO}_4$ ; pH 7.4)로 Akta 정제기 10 시스템(GE Healthcare)을 이용하여 1 ml/분의 유속으로 Superdex S200 HR 10/300 GL 칼럼(GE Healthcare Europe, Freiburg, Germany) 상에서 크기 배제 크로마토그래피(SEC)를 수행하였다. 실시예 4에 기재된 금속 친화도 크로마토그래피에서 수득한 정제된 Fab 단편 및 그 200개 잔기의 PA#1 용합물의 250  $\mu\text{l}$  표본을 PBS 중 0.25 mg/ml의 농도로 개별 처리하였다. 두 단백질 모두 도 4A에 나타난 바와 같은 균질한 단일 피크로 용출되었다.

[0211] 칼럼 적정을 위해(도 4B), 하기 구형 단백질의 적절한 혼합물 250  $\mu\text{l}$ (Sigma, Deisenhofen, Germany)를 0.2 mg/ml 내지 0.5 mg/ml의 단백질 농도로 PBS 중에서 처리하였다: 사이토크롬 c, 12.4 kDa; 카르본산 안하이드라아제, 29.0 kDa; 오발부민, 43.0 kDa; 소 혈청 알부민, 66.3 kDa; 알코올 탈수소효소, 150 kDa;  $\beta$ -아밀라아제, 200 kDa; 아포-페리틴, 440 kDa.

[0212] 그 결과, 200개 잔기의 PA#1 중합체/폴리펩티드를 갖는 융합 단백질은 동일한 분자량을 갖는 대응 구형 단백질 보다 훨씬 더 큰 크기를 나타내었다. Fab-PA#1(200)에 대한 겔보기 크기 증가는 미융합 Fab 단편 대비 7.4배였으나, 실제 질량은 단지 1.3배 더 컸다. 이 관찰은 본 발명에 따른 Pro/Ala 폴리펩티드 절편에 의해 생물학적 활성 Fab 단편에 부여되는 훨씬 더 증가된 수력학적 부피를 시사한다.

[0213] 실시예 7: 유전적으로 인코딩된 IFNa2b 및 200개 잔기의 PA#1 중합체 간 제조합 융합 단백질에 대한 분석용 겔 여과에 의한 수력학적 부피 측정.

[0214] 실시예 6에 기재된 바와 유사하게 Akta 정제기 10 시스템(GE Healthcare)을 이용하여 1 ml/분의 유속으로 Superdex S200 HR 10/300 GL 칼럼(GE Healthcare Europe, Freiburg, Germany) 상에서 IFNa2b 및 PA#1(200)-IFNa2b로 크기 배제 크로마토그래피를 수행하였다. 두 단백질 모두 도 4C에 나타난 바와 같은 균질한 단일 피크로 용출되었다.

[0215] 그 결과, 200개 잔기의 PA#1 중합체/폴리펩티드를 갖는 융합 단백질은 동일한 분자량을 갖는 대응 구형 단백질 보다 훨씬 더 큰 크기를 나타내었다(도 4D). PA#1(200)-IFNa2b에 대한 겔보기 크기 증가는 미융합 IFNa2b 단편

대비 10.2배였으나, 실제 질량은 단지 1.8배 더 컸다. 이 관찰은 본 발명에 따른 Pro/Ala 폴리펩티드 절편에 의해 생물학적 활성 인터페론에 부여되는 훨씬 더 증가된 수력학적 부피를 시사한다.

[0216] 실시예 8: 원형 광이색형 분광측정을 통한 Fab 단편에 융합된 생합성 PA#1 중합체에 대한 랜덤 코일 입체형태의 검출.

[0217] 석영 큐벳 106-QS(0.1 mm 경로 길이; Hellma, Mullheim, Germany)이 장착된 J-810 분광편광계 (Jasco, Groß-Umstadt, Germany)를 이용하여 2차 구조를 분석하였다. 50 mM K<sub>2</sub>SO<sub>4</sub>, 20 mM K-포스페이트(pH 7.5) 중 실시예 4에서 수득한 3.12 내지 15.4 μM 단백질 용액을 사용하여 16회 실험(밴드폭 1 nm, 스캔 속도 100 nm/분, 반응 4s)을 축적하여 실온에서 190 내지 250 nm의 스펙트럼을 기록하였다. 용액 블랭크에 대한 보정 후, 기기 소프트웨어를 이용해서 스펙트럼을 평활화하고, 하기 공식에 따라 몰 타원율 Θ<sub>M</sub>을 계산하였다:

$$\Theta_M = \frac{\Theta_{obs}}{c \cdot d}$$

[0218]

[0219] 여기서 Θ<sub>obs</sub>는 측정된 타원율, c는 단백질 농도[mol/l], d는 석영 큐벳의 경로 길이[cm]를 나타낸다. Θ<sub>M</sub> 값은 Kaleidagraph(Synergy Software, Reading, PA)를 이용하여 파장에 대해 도식화하였다.

[0220] 재조합 Fab에 대한 측정된 원형 광이색형(CD) 스펙트럼은 β-시트 우세 면역글로불린 폴딩과 일치한 반면, Fab-PA#1(200) 융합 단백질에 대한 스펙트럼은 랜덤 코일 입체형태의 상당한 기여를 나타내었다(도 5A). Pro/Ala 폴리펩티드 절편에 의한 분광측정 기여를 더 자세히 분석하기 위해, 미융합 Fab 단편에 대한 몰 차이 CD 스펙트럼을 Fab-PA#1(200)에 대한 것으로부터 후자의 스펙트럼을 감산하여 계산하였다(도 5B). 그 결과, 랜덤 코일 입체형태의 특징인 200 nm 근처의 강력한 최소 부위가 관찰되었다. 따라서 재조합 융합 단백질의 일부인 Pro/Ala 서열은 생리학적 완충액 조건하에 랜덤 코일 중합체로 존재하는 것으로 나타난다.

[0221] 실시예 9: 원형 광이색형 분광측정을 통한 IFNa2b에 융합된 유전적으로 인코딩된 PA#1 중합체에 대한 랜덤 코일 입체형태의 검출.

[0222] 3.6 내지 38.7 μM 단백질 용액을 사용하여 실시예 8에 기재된 바와 같이 IFNa2b 및 PA#1(200)-IFNa2b(실시예 5에서 수득)에 대한 CD 측정에 의해 2차 구조를 분석하였다. PA#1(200)-IFNa2b의 스펙트럼은 인터페론의 공지된 α-나선 번들 폴딩뿐만 아니라 랜덤 코일 입체형태를 시사하는 α-나선 2차 구조의 상당한 기여를 나타내었다(도 5C). Pro/Ala 중합체 융합 파트너에 의한 분광측정 기여를 더 자세히 분석하기 위해, 미융합 IFNa2b에 대한 몰 차이 CD 스펙트럼을 두 개별 스펙트럼의 감산에 의해 계산하였다(도 5D). 그 결과, 랜덤 코일 입체형태의 특징인 200 nm 근처의 강력한 최소 부위가 관찰되었다. 따라서 재조합 융합 단백질의 일부인 Pro/Ala 폴리펩티드 절편은 수성 완충액 조건하에 랜덤 코일 중합체로 존재하는 것으로 나타난다.

[0223] 실시예 10: Fab 단편, IFNa2b 및 이들의 200개 잔기의 PA#1 중합체 융합물의 2차 구조의 정량적 분석.

[0224] Fab 단편, Fab-PA#1(200), IFNa2b, 및 PA#1(200)-IFNa2b의 2차 구조 함량을 복합 CD 스펙트럼의 디컨볼루션을 위한 한 세트의 33개 염기 스펙트럼으로 2차 구조 디컨볼루션 프로그램 CDNN ver. 2.1(Bohm (1992) Protein Eng. 5:191-195)을 이용하여 실시예 8 및 9에서 측정된 해당 CD 스펙트럼으로부터 개별적으로 정량하였다. 상기 분석 결과를 하기 표에 제공한다:



	Fab	Fab-PA#1(200)	차이 : PA#1(200)	IFNa2b	PA#1(100)-IFNa2b	차이 : PA#1(200)
α-나선	9.5 %	7.5 %	2.1 %	38.2 %	31.0 %	0.7 %
반대 평형 β-시트	40.4 %	3.1 %	0 %	1.8 %	0.2 %	4.6 %
평형 β-시트	6.9 %	1.3 %	0.3 %	8.4 %	0.7 %	0.6 %
β-회전	6.2 %	50.4 %	78.6 %	19.2 %	75.2 %	69.7 %
랜덤 코일	37.2 %	63.4 %	94.8 %	35.9 %	64.4 %	97.5 %
Σ 합	100.2 %	125.8 %	175.8 %	103.5 %	171.4 %	170.0 %
Σ β-회전 및 랜덤 코일	43.4 %	113.8 %	173.4 %	55.1 %	139.6 %	169.1 %

[0225]

[0226]

그 공지된 번역글로블린 폴딩(Eigenbrot (1993) J. Mol. Biol. 229:969-995 참고)과 일치하는 재조합 Fab 단편의 우세한 β-시트 2차 구조 함량에 비해, 비구조화된 입체형태 분획(랜덤 코일 및 β-회전 포함)은 PA#1 중합체가 Fab 단편에 융합되는 경우 뚜렷이 증가한다. Pro/Ala 폴리펩티드 절편에 대한 차이 CD 스펙트럼은 뚜렷한 랜덤 코일 입체형태를 나타낸다. 2차 구조의 분석은 전체 2차 구조의 거의 100%를 포함하는 고분획의 비구조화된 입체형태(랜덤 코일 및 β-회전 포함)의 존재를 보여준다. 유사하게, α-나선 번들 단백질로서의 그 공지된 3차원 구조(Radhakrishnan (1996) Structure 4:1453-1463)와 일치하는 재조합 IFNa2b의 우세한 α-나선 2차 구조 함량에 비해, 전체 단백질에 대한 비구조화된 입체형태 분획은 PA#1 중합체가 IFNa2b에 융합되는 경우 뚜렷이 증가한다. Pro/Ala 폴리펩티드 절편에 대한 차이 CD 스펙트럼은 뚜렷한 랜덤 코일 입체형태를 나타낸다. 2차 구조 분석은 전체 2차 구조의 거의 100%를 포함하는 높은 분획의 비구조화된 입체형태(랜덤 코일 및 β-회전 포함)의 존재를 보여준다.

[0227]

Chou-Fasman 알고리즘(Chou and Fasman (1974) Biochemistry 13:222-245)을 이용하여 PA#1 중합체 서열의 이론적 분석을 수행할 경우, 상이한 결과가 수득되었다. 상기 분석 결과를 도 7에 나타낸다. 이 알고리즘은 실험적 데이터와 뚜렷이 대비되는 100% α-나선 2차 구조를 예측한다. 따라서 상기 알고리즘은 본 발명에 따른 아미노산 중합체에 대한 비구조화된 입체형태를 확실히 예측하는데 유용하지 않다.

[0228]

**실시예 11: His6-PA#1(200)-hGH 융합 단백질에 대한 발현 벡터로서의 pASK75-His6-PA#1(200)-hGH의 구축.**

[0229]

200개 잔기의 PA#1 서열 반복체 PA#1(200)을 갖는 융합물로서 hGH를 인코딩하는 발현 플라스미드의 구축을 위해, pASK75-His6-hGH(서열 번호 41)(도 6A)를 *SapI*로 절단하고 새우 알칼린 포스파타아제로 탈인산화하고, *SapI*로 제한 소화하여(도 6B에 예시된 대로, PA#1(20) 중합체/폴리펩티드 카세트만으로) 앞서 구축한 플라스미드 pFab-PA#1(200)(실시예 2)으로부터 절단된 200개 잔기의 PA#1 폴리펩티드를 인코딩하는 유전자 단편과 라이게이션하였다. *E. coli* JM83(Yanisch-Perron. (1985) 위 인용문)의 형질전환 후, 플라스미드를 제조하고 정확한 삽입물의 존재를 제한 분석으로 확인하였다. 생성 플라스미드는 pASK75-His6-PA#1(200)-hGH(서열 번호 46)로 명명하였다(도 6C).

[0230]

**실시예 12: CHO 세포에서 200개 잔기의 PA#1 중합체/폴리펩티드와 융합된 인간 성장 호르몬의 분비 제조를 위한 발현 벡터의 구축.**

[0231]

hGH PA#1 융합 단백질의 원핵생물 생산을 허용하는 pASK75(Skerra (1994) 위 인용문)의 유도체인 벡터 pASK75-His6-PA#1(200)-hGH(서열 번호 46)를 *NheI* 및 *HindIII*로 절단하였다. 상기 단편을 아가로오스 겔 전기영동을 통해 정제하고 대응 절단 pCHO(서열 번호 50)와 라이게이션하였다. *E. coli* XL1-Blue(Bullock (1987) 위 인용문.)의 형질전환 후, 플라스미드를 제조하고 단편의 정확한 삽입을 제한 분석을 통해 확인하였다. His<sub>6</sub> 태그에 융합된 hGH 신호 펩티드, PA#1(200) 폴리펩티드 절편, 및 인간 성장 호르몬(hGH)을 코딩하는 생성 플라스미드를

pCHO-PA#1(200)-hGH(서열 번호 48)로 명명하였으며 도 6D에 나타낸다.

[0232] **실시예 13: CHO 세포에서 인간 성장 호르몬(hGH) 및 유전적으로 인코딩된 PA#1 중합체 간 융합 단백질의 분비 제조.**

[0233] CHO-K1 세포(ATCC No. CCL-61)를 50% 융합성에 도달할 때까지 100 mm 플라스틱 디쉬에서 Quantum 263 배지(PAA Laboratories, Colbe, Germany) 중에 배양하였다. 세포를 Nanofectin 키트(PAA Laboratories, Colbe, Germany)를 이용해 8  $\mu$ g의 pCHO-PA#1(200)-hGH(서열 번호 48) 또는 대조로서 PA#1(200) 서열이 없는 hGH를 인코딩하는 유사한 플라스미드인 pCHO-hGH(서열 번호 49)로 형질감염하였다. 6시간 후, 세포 배양 배지를 7 ml의 Opti-MEM®-I 환원된 혈청 배지(Invitrogen, Darmstadt, Germany)로 교체하고 세포를 37°C 가습 분위기하에서 5% CO<sub>2</sub>와 함께 인큐베이션하였다. 2일 후, 세포 배양 상청액 20  $\mu$ l를 취해  $\beta$ -메르캅토에탄올을 함유하는 5  $\mu$ l의 SDS-PAGE 로딩 완충액으로 희석하였다. 95°C에서 5분 가열 후, 15  $\mu$ l의 각 표본으로 12% SDS-PAGE를 수행하였다. 반건조 블로팅 장치 수단에 의해 니트로셀룰로오스 막(Schleicher & Schuell, Dassel, Germany) 상에 전기 전달 후, 10 ml PBST(0.1% v/v Tween 20 함유 PBS)로 15분 동안 3회 막을 세척하였다. 막을 홀스 래디쉬 페록시다아제(Abeam, Cambridge, UK)에 콘주게이션된 항-인간 성장 호르몬 항체 ab1956의 1:1000 희석액 10 ml과 함께 인큐베이션하였다. 1시간 인큐베이션 및 막을 20 ml PBST로 5분 동안 2회 및 PBS로 5분 동안 2회 세척한 후, 15 ml의 SIGMAFAST™ 3,3-디아미노벤지딘 용액(Sigma-Aldrich Chemie, Munich, Germany)의 존재하에 발색 반응을 수행하였다. 물로 세척하여 반응을 정지시키고 막을 대기 건조하였다. 블롯으로 두 재조합 단백질 표본(도 6E)에 대한 신호가 나타나 CHO 세포에서 PA#1 폴리펩티드를 갖는 hGH 융합 단백질의 분비 제조가 입증되었다.

[0234] **실시예 14: hGH 및 유전적으로 인코딩된 PA#1 중합체/폴리펩티드 간 융합 단백질의 박테리아를 이용한 생산 및 정제.**

[0235] 인간 성장 호르몬(hGH)(계산 질량: 23.4 kDa), PA#1(200)-hGH(계산 질량: 39.6 kDa), PA#1(400)-hGH(계산 질량: 55.8 kDa) 및 PA#1(600)-hGH(계산 질량: 72.0 kDa)를 실시예 11의 대응 발현 플라스미드 또는 이들의 2배(400개 잔기 인코딩)의 또는 3배(600개 잔기) PA#1 서열 카세트 유도체를 각각 포함하는 *E. coli* KS272에서 제조하였다. 2.5 g/l 글루코오스, 0.5g/l 프롤린 및 100mg/l 엠피실린을 함유하는 2 l LB 배지를 이용하여 진탕기 플라스크 배양으로 22°C에서 박테리아를 이용한 생산을 수행하였다. 재조합 유전자 발현의 유도는 OD<sub>550</sub> = 0.5에서 3시간 동안 0.4 mg 무수 테트라사이클린을 첨가하여 수행하였다. 50  $\mu$ g/ml 라이소자임을 함유하는 500 mM 수크로오스, 1 mM EDTA, 100 mM 트리스/HCl(pH 8.0)의 존재하 원형질막 추출을 다른 곳에 기재된 바와 같이 수행한 후(Breustedt (2005) 위 인용문.) 완충액으로 40 mM Na-포스페이트(pH 7.5), 0.5 M NaCl로 HisTrap 고성능 친화도 칼럼(GE Healthcare)을 이용하여 His<sub>6</sub>-태그를 통해 정제하였다. 단백질을 0 내지 150 mM의 이미다졸 농도 구배(수행 완충액 중에 용해되고 HCl로 pH 7.5로 조정됨)를 이용하여 용출하고 PBS(115 mM NaCl, 4 mM KH<sub>2</sub>PO<sub>4</sub>, 16 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4)로 평형화된 Superdex 200-HR10/30 칼럼(GE Healthcare)을 이용하여 크기 배제 크로마토그래피를 통해 추가 정제하였다.

[0236] 크기 배제 크로마토그래피 후 응집의 징후 없이 hGH에 대해 1 mg L<sup>-1</sup> OD<sup>-1</sup>, PA#1(200)-hGH에 대해 0.3 mg L<sup>-1</sup> OD<sup>-1</sup>, PA#1(400)-hGH에 대해 0.3 mg L<sup>-1</sup> OD<sup>-1</sup> 및 PA#1(600)-hGH에 대해 0.2 mg L<sup>-1</sup> OD<sup>-1</sup>의 수율로 모든 재조합 hGH 융합 단백질에 대한 균질한 단백질 제조물을 수득하였다. 높은 몰 농도의 트리스 완충액 시스템(Fling (1986) 위 인용문)을 이용하여 SDS-PAGE를 수행하였다. 미융합 hGH 및 모든 그 PA#1 폴리펩티드 융합물에 대해 16050 M<sup>-1</sup> cm<sup>-1</sup>의 계산 흡광도 계수(Gill (1989) 위 인용문)를 이용하여 280 nm에서의 흡광도에 따라 단백질 농도를 결정하였다.

[0237] **실시예 15: 표면 플라즈몬 공명을 이용한 인간 성장 호르몬 수용체의 세포의 도메인에 대한 인간 성장 호르몬 및 그 PA#1 중합체 융합물의 결합 친화도 측정.**

[0238] 인간 성장 호르몬 수용체 Fc 융합 단백질(hGHR-Fc; R&D Systems)에 대한 hGH 및 그 PA#1 폴리펩티드 융합물의 친화도를 Biacore 2000 시스템(GE Healthcare) 상에서 표면 플라즈몬 공명(SPR) 실시간 측정을 통해 결정하였다. 먼저, 10 mM Na-아세테이트(pH 5.0) 중 100  $\mu$ g/ml 농도의 마우스 항-인간 IgG-Fc 포획 항체(Jackson Immuno Research) 15  $\mu$ l를 아민 커플링 키트(GE Healthcare)를 이용하여 CMDP 칩(XanTec bioanalytics)의 두 흐름 채널 표면으로 고정하였다. 이는 약 2700 반응 단위(RU)를 유도하였다. 흐름 완충액으로 PBS/T(0.05%(v/v) Tween 20 함유 PBS)로 평형화 후, 칩의 하나의 채널을 약 300 RU의 추가 신호에 도달할

때까지 5  $\mu\text{l}$ /분의 유속으로 2  $\mu\text{g}/\text{ml}$  hGHR-Fc로 충전하였다. 이어서 PBS/T 중 hGH 또는 그 PA#1 폴리펩티드 용합물 75  $\mu\text{l}$ 를 다양한 농도로 주입하고, 20  $\mu\text{l}$ /분의 연속 완충액 흐름하에 결합상 및 해리상을 측정하였다. 재생을 위해, 10 mM 글리신/HCl(pH 2.7)의 세 6  $\mu\text{l}$  펄스를 처리하였다. 고정화된 수용체가 없는 채널에 대해 측정된 대응 신호의 이중 감산 및 몇몇 완충액 블랭크 주입에서 결정된 평균화된 기준선에 의해 센소그램을 보정하였다(Myszka (1999) Mol. Recognit. 12:279-284). BIAevaluation 소프트웨어 버전 3.1(GE Healthcare)을 이용하여 1:1 랭뮤어(Langmuir) 결합 모델에 따라 7개 이상의 상이한 표본 주입으로부터 트레이스의 통괄 적합화에 의해 역학 데이터를 평가하였다. 역학에 대한 SPR 측정으로부터 수득하고 hGH 또는 그 PA#1 용합물 및 인간 성장 호르몬 수용체 간 복합체의 평형 상수로부터 유래된 값을 하기 표에 요약한다:

hGH 변이체	$k_{on} [10^5 \text{M}^{-1} \text{s}^{-1}]$	$k_{off} [10^{-6} \text{s}^{-1}]$	$K_D [\text{pM}]$
hGH	10.2	10.6	10.4
PA#1(200)-hGH	4.75	9.18	19.3
PA#1(400)-hGH	3.26	14.0	42.9
PA#1(600)-hGH	3.29	12.5	38.0

[0239]

[0240] 이러한 데이터는 상이한 길이의 PA#1 폴리펩티드를 갖는 hGH의 용합물이 수용체 결합을 크게 방해하지 않는 것을 보여준다. 모든 hGH PA#1 폴리펩티드 용합물은 PA#1 폴리펩티드가 없는 재조합 hGH에 비해 5배 내에서 수용체 결합 활성을 보유한다.

[0241] 실시예 16: Fab 단편 및 유전적으로 인코딩된 PA#1 중합체 간 재조합 용합 단백질에 대한 생체 내 연장된 혈장 반감기의 검출.

[0242] 생체 BALB/c 마우스(SPF 스톡 교배; TU Muenchen, Freising, Germany)는 하기 표에 따라 정맥내 주사되었다:

표 2

[0243]

군	A	B	D
평가 아이템	Fab	Fab-PA#1(200)	Fab-PA#1(600)
투여 경로	정맥내		
용량[mg/kg b.w.]	5.0	5.0	5.0
농도[mg/ml]	1.0	1.0	1.0
처리 부피[mg/kg b.w.]	5.0		
동물수/군	9	9	9
채혈 시점의 횟수	12	12	12
동물수/채혈 시점	3	3	3
채혈 시점의 횟수/동물	4/1	4/1	4/1

[0244] 정맥내 투여되는 평가 아이템의 총 부피를 투여일에 기록된 개체 체중 (b.w.)에 따라 계산하였다(예로 20 g 체중의 동물은 1 mg/ml의 평가 아이템 100  $\mu\text{l}$ 을 투여받음). 채혈은 하기 표에 따라 수행하였다:

		주사 후 채혈 시점												
평가 아이템	하위군	10 분	30 분	1 시간	2 시간	3 시간	4 시간	6 시간	8 시간	12 시간	24 시간	36 시간	48 시간	
Fab Fab-PA#1(200) Fab-PA#1(600)	1	x			x			x			x			
		x			x			x			x			
		x			x			x			x			
	2		x			x			x				x	
			x			x			x				x	
			x			x			x				x	
	3			x				x			x			x
				x				x			x			x
				x				x			x			x

[0245]

[0246] 각각의 물질(평가 아이템)을 9마리 동물에 모두-각각 3마리 동물씩의 3개 하위군 1~3으로 나눔-주사하고, 각각 상이한 시점에 4개 표본을 제공하였다. 혈액 표본(대략 50 μl)을 꼬리 정맥에서 취하여 4℃에서 30분 동안 보관하였다. 10000 g 및 4℃에서 10분 동안 원심분리 후, 상청액(혈장)을 즉시 냉동하고 -20℃에서 보관하였다.

[0247] ELISA에서 Fab 용합 단백질의 정량적 검출을 위해, 96웰 마이크로타이터 플레이트(Maxisorb, NUNC, Denmark)의 웰을 50 mM NaHCO<sub>3</sub>(pH 9.6) 중 재조합 Her2/ErbB2 외도메인 항원의 10 μg/ml 용액 50 μl로 4℃에서 하룻밤 동안 코팅하였다. 이어서 웰을 PBS 중 3%(w/v) BSA 200 μl로 1시간 동안 차단하고 PBS/T(0.1%(v/v) Tween 20 함유 PBS)로 3회 세척하였다. 혈장 표본을 미처리 동물의 0.5%(v/v) 마우스 혈장을 함유하는 PBS/T 중에서 연속 희석하여 처리하고 1시간 동안 인큐베이션하였다. 이어서 웰을 PBS/T로 3회 세척하고 PBS/T 중의 항-인간 Cκ 항체 알칼린 포스파타아제 콘주게이트의 1:1000 희석 용액 50 μl과 함께 1시간 동안 인큐베이션하였다. PBS/T로 2회 및 PBS로 2회 세척 후 100 mM 트리스/HCl(pH 8.8), 100 mM NaCl 중 0.5 μg/ml p-니트로페닐 포스페이트 50 μl, 기질로 5 mM MgCl<sub>2</sub>를 첨가하여 발색 반응을 시작하고, 25℃에서 15분 후 405 nm에서의 흡광도를 측정하였다. 혈장 표본 중의 Fab, Fab-PA#1(200), 및 Fab-PA#1(600) 농도를 측정된 신호와 미처리 동물의 0.5%(v/v) 마우스 혈장을 함유하는 PBS/T 중 정의된 농도의 대응 정제 단백질에 대한 연속 희석에서 결정된 표준 곡선을 비교하여 정량하였다.

[0248] Fab, Fab-PA#1(200), 및 Fab-PA#1(600)의 혈장 반감기를 추정하기 위해, 농도 값 c(t)를 각각의 ELISA 측정 시점에 대해 결정하고 정맥내 주사 후 시간 t에 대해 도식화하였다. 이러한 데이터를 하기 공식에 따라 이중 지수 붕괴를 가정하여 KaleidaGraph 소프트웨어를 이용하여 수치적으로 조정하였다:

$$c(t) = c_{\alpha} e^{-\frac{\ln 2 \cdot t}{\tau_{\alpha/2}}} + (c_0 - c_{\alpha}) e^{-\frac{\ln 2 \cdot t}{\tau_{\beta/2}}}$$

[0249]

[0250] 식 중, τ<sup>α</sup><sub>1/2</sub> 및 τ<sup>β</sup><sub>1/2</sub>는 각각 분포상 α 및 제거상 β의 반감기 값이다. c<sub>0</sub>는 시점 0에서의 총 혈액 농도인 반면 c<sub>α</sub>는 분포상에 대한 농도 크기이다.

[0251] 도 8은 BALB/c 마우스에서 3 테스트 아이템에 대한 약동학을 나타낸다. 재조합 Fab는 단지 약 1.3시간의 제거 반감기를 가지며 신속한 혈액 제거를 나타낸 반면, Fab-PA#1(200) 및 Fab-PA#1(600) 용합 단백질은 각각 약 4.1 시간 및 38.8시간에 해당하는 값을 갖는 3배 및 29배 초과로 연장된 반감기를 갖는다. 이러한 데이터는 Fab 단편의 생체 내 혈장 반감기가 Pro/Ala 중합체/폴리펩티드와의 융합으로 인해 크게 연장되어 반감기가 아미노산 중합체의 길이 증가와 함께 더 길어짐을 입증한다.

[0252] 실시예 17: P1A1 및 P1A3 아미노산 중합체/폴리펩티드에 대한 유전자 합성 및 Fab-P1A1(200) 및 Fab-P1A3(200) 융합 단백질에 대한 발현 벡터로서의 pFab-P1A1(200) 및 pFab-P1A3(200)의 구축



[0253] Pro/Ala 폴리펩티드/중합체 P1A1(서열 번호 51) 및 또한 PA#3(서열 번호 3)으로 명명된 P1A3을 포함하는 반복 중합체 서열을 인코딩하는 유전자 단편을 각각 실시예 1에 기재된 바와 같이 P1A1에 대해 서열 번호 52 및 서열 번호 53 및 P1A3에 대해 서열 번호 54 및 서열 번호 55의 상보적 올리고데옥시뉴클레오티드쌍의 혼성화에 의해 수득하였다. 경쇄(LC)의 C-말단에서 200개 잔기의 대응 Pro/Ala 중합체/폴리펩티드 절편을 갖는 Fab 단편을 코딩하는 pFab-P1A1(200)(서열 번호 58) 및 pFab-P1A3(200)(서열 번호 59)(LC Fab-P1A1(200)의 아미노산 서열: 서열 번호 56; LC Fab-P1A3(200)의 아미노산 서열: 서열 번호 57)를 실시예 2에 기재된 pFab-PA#1(200)과 유사한 방식으로 구축하였다.

[0254] 하기 서열 번호 56, 57, 58 및 59도 또한 재현된다. 그러나 이러한 서열들은 또한 본 발명의 상기 개시 및 설명의 특정 부분인 첨부된 서열 목록에도 포함된다.

[0255] 서열 번호 56

```

Asp Ile Glu Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1           5                10                15
Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val Asn Thr Ala
                20                25                30
    
```

[0256]

Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile  
 35 40 45

Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg Phe Ser Gly  
 50 55 60

Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro  
 65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr Thr Thr Pro Pro  
 85 90 95

Thr Phe Gly Gln Gly Thr Lys Leu Glu Ile Lys Arg Thr Val Ala Ala  
 100 105 110

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

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala  
 130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln  
 145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser  
 165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr  
 180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser  
 195 200 205

Phe Asn Arg Gly Glu Cys Ser Ser Ala Pro Ala Pro Ala Pro Ala Pro  
 210 215 220

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 225 230 235 240

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 245 250 255

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 260 265 270

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 275 280 285

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 290 295 300

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 305 310 315 320

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 325 330 335

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 340 345 350

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 355 360 365

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 370 375 380

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 385 390 395 400

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro  
 405 410 415

[0257]

[0258]

Ala

[0259] 서열 번호 57

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

[0260]

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 225 230 235 240

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 245 250 255

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 260 265 270

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 275 280 285

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 290 295 300

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 305 310 315 320

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 325 330 335

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 340 345 350

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 355 360 365

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 370 375 380

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 385 390 395 400

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 405 410 415

Ala

[0261]

[0262] 서열 번호 58

```

accgacacc atcgaatggc cagatgatta attoctaatt tttgttgaca ctctatcatt 60
gatagagtta ttttacoact ccctatcagt gatagagaaa agtgaaatga atagttcgac 120
aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gaagtggcac 180
tggttggttt cgctaccgta gcgcaggccg aagttaaact gcaggaatcc ggtggtggtc 240
tggttcagcc aggtggttcc ctgcggctct cgtgtgtctc ttccggtttc aacatcaaaag 300
acacctacat coactggggt cgtcaggctc cgggtaaagg cctggaatgg gttgctcgta 360
tctaccgcac caacggttac accaggtatg ccgattcagt taaaggtcgt ttcaccatct 420
cggccgacac ttccaaaac accgcttacc tccagatgaa ctccctgcgt gctgaagaca 480
cagctgttta ttattgtctc cgttgggggt gtgacggttt ctaccgtatg gactactggg 540
gtcagggtac cctggtaacc gtctctcag cctccaccaa gggcccatcg gtcttcccc 600
tggcaccctc ctccaagagc acctctgggg gcacagcggc cctgggctgc ctggtcaagg 660
actacttccc cgaaccgggt acggtgtcgt ggaactcagg cgcctgacc agcggcgtgc 720
acaccttccc ggctgtccta cagtctcag gactctactc cctcagcagc gtggtgactg 780
tgccctccag cagcttgggc acccagacct acatctgcaa cgttaatcac aaaccagca 840
acaccaaggt cgacaagaaa gttgagccca aatcttgcca tcaccaccat caccattaat 900
aaccatggag aaaataaagt gaacaaaagc actattgcac tggcactctt accgttactg 960
tttacccctg tgacaaaagc cgacatcgag ctaccccaat cccctctctc cctgtccgct 1020
tcogttggcg accgtgttac catcacgtgt agggcctcgc aagacgtaaa caccgcgta 1080
gcgtggtatc agcagaaacc cgggaaagct ccgaaactgc tgatctatag cgttctcttc 1140
ctgtattccc gagtcccgag caggttcagt ggttcccggt ccggtaccga cttcaccctg 1200
    
```

[0263]



acgalatcct cccctcagcc ggaagacttc gctacctact actgtcaaca gcactacacc 1260  
 accccgcccga ccttcgggtca gggtagccaaa ctcgagatca aacggactgt ggtgcacca 1320  
 tctgtcttca tcttcccgcc atctgatgag cagttgaaat ctggaactgc ctctgttgtg 1380  
 tgcctgctga ataatctota tcccagagag gccaaagtac agtggagggl ggataacgcc 1440  
 ct.ccaatogy gtaactccca ggaagagtgc acagagcagg acagcaagga cagcacctac 1500  
 agcctcagca gccacctgac gctgagcaaaa gcagactacg agaaaaacaa agtctacgcc 1560  
 tgggaagtca cccatcaggg cctgagttcg cccgtcaca agagcttcaa ccgcgagag 1620  
 tgcctctctg cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 1680  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 1740  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 1800  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 1860  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 1920  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 1980  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 2040  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 2100  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 2160  
 cctgcaccag cccctgctcc tgcctccagca cctgcaccag cacctgctcc agcaccagct 2220  
 cctgcaccag cctgaagagc ttaagcttga cctgtgaagt gaaaaatggc gcacattgtg 2280  
 cgacattttt tttgtctgcc gtttaccgct actgcgtcac ggatctccac ggcctctgta 2340  
 ggggcgcaat aagcggggg ggtgtggtgg ttacggcag cgtgaccgct acacttgcca 2400  
 gggccctagc gcccgctcct ttgcctttct tccctcctt tctgcaccag ttcggcgct 2460  
 tcccctgca agctotaaat cgggggctcc ctttagggtt ccgatttagt gctttacggc 2520  
 acctcgacc caaaaaactt gatagggtg atggttcacg tagtgggcca tccgctgat 2580  
 agacggtttt tgcctcttg ccqgttgagt ccacgttctt taatagtgga ctcttgttcc 2640  
 aaactggaa cacaotcaac cctatctcgg tctattctt tgatttataa gggattttgc 2700  
 cgatttcggc ctattgggta aaaaatgagc tgatttaaca aaaatttaac gcgaatttta 2760  
 aaaaaatatt aacgtttaca atttcagggt gcactttctg gggaaatgtg cgcggaaacc 2820  
 ctatttggtt attttctaa atacattcaa atatgtatcc gctcatgaga caataaccct 2880  
 gataaatgct tcaataaat tgaaaaagga agagtatgag tattcaaat tcccggtctg 2940  
 ccttatttcc cttttttcgg gcattttgce ttctgtttt tgcctaccca gaaacgctg 3000  
 tgaagataaa agatgctgaa gatcagttgg gtcacagag gggttacatc gaactggatc 3060  
 tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca 3120  
 cttttaaagt tctgctatgt ggcgggtgat tatcccgat tgacggcggg caagagcaac 3180  
 tgggtgcgcc catacaatat tctcagaatg acttggttga gtaactaccca gtcacagaaa 3240  
 agcatcttac ggatggcatg acagtaagag aattatgcag tgcctgccata accatgagtg 3300  
 ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag ctaaccgctt 3360  
 ttttgcacaa catgggggat actgtaactc gccttgatcg ttgggaaacc gagctgaatg 3420  
 aagccatacc aaacgacggc cgtgacacca cgatgcctgt agcaatggca acaacgttgc 3480  
 qcaaacctat aactggcgaa ctacttactc tagcttcccg gcaacaattg atagactgga 3540  
 tggaggcggg taaagttgca ggaacccttc tgcgctcggc ccttccgctt ggtgtgttta 3600  
 ttgctgataa atctggagcc ggtgagcgtg gctctcggcg tatcattgca gcaactggggc 3660  
 cagatggtaa gccctccggt atcgtagtta totaacagac ggggagctag gcaactatgg 3720  
 atgaacgaaa tagacagatc gctgagatag gtgcctcact gatlaagcat tggtaggaat 3780  
 taatgatgct tgcgttagat aaaagtaaag tgattaacag cgcatttagag ctgcttaatg 3840  
 aggtcggaaat cgaaggttta acaaccgcta aactcgccta gaagctaggt gtagagcagc 3900  
 ctacattgta ttggcatgta aaaaaaagc gggctttgct cgaacgctta gccattgaga 3960  
 tgttagatag gcaccatact cacttttgcc ctttagaagg ggaaagctgg caagattttt 4020  
 tacgtaataa cgtaaaaagl tttagatgtg ctttactaag tcatcgcgat ggagcaaaa 4080  
 tacatttagg tacacgcctt acagaaaaac agratgaaac tctcgaanaa caattagcct 4140  
 ttttatgcca acaaggtttt tcaactagaga atgcattata tgcactcagc gcagtggggc 4200  
 attttacttt aggttgcgta ttggaaagatc aagagcatca agtogctaaa gaagaaagg 4260  
 aaacacctac lactgatagt atgccgcat tattaccaca agctatcgaa tlatttgatc 4320  
 accaaggtgc agagccagcc ttcttattcg gccctgaatt gatcaatgca ggattagaaa 4380  
 aacaacttaa atgtgaaagt gggctttaa agcagcataa cctttttccg tgatggtaac 4440  
 ttoactagt taaaaggatc taggtgaaga tcccttttga taatctcag accaaaaatcc 4500  
 cttaacgtga gttttcgttc cactgagcgt cagaccgcc agaaaagatc aaaggatctt 4560  
 cttgagatcc tttttttctg cgcgtaactc gctgcttga acaaaaaaaa ccaccgctac 4620  
 cagcgggtgt ttttttccg gatcaagagc taccactct ttttccgag gtaactggct 4680  
 tcagcagagc gcagatacca aatactgtcc tctagtgtg ccgctagtta ggcaccact 4740  
 tcaagaactc tgtagcaacc cctacatacc tgcctctgct aalctgttta ccagtggctg 4800

[0264]

ctgcnagtg ogataagtcg tgtcttaccg ggttggactc aagacgatag ttaccggata 4860  
 aggcgcagcg gtcgggctga acggggggtt cgtgcacaca gccagcttg gagcgaacga 4920  
 cctacaccga actgagatc ctacagcgtg agctatgaga aagcggccag cttcccgag 4980  
 ggagaaggcc ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgagg 5040  
 agcttccagg gggaaacgcc tggatcttt atagtctgt cgggtttcgc cacctctgac 5100  
 ttgagcgtcg atttttgtga tgcctcgtcag gggggcggag cctatggaaa aacgccagca 5160  
 acgggcctt tttacggttc ctggcctttt gctggccttt tgcctacatg 5210

[0265]

[0266] 서열 번호 59

accgcagacc	atcgaatggc	cagatgatta	attcctaatt	tttgttgaca	ctclatcatt	60
gatagagtta	ttttaccact	cctatocagt	gatagagaaa	agtgaaatga	atagttogac	120
aaaaatctag	ataacgaggg	caaaaaatga	aaaagacagc	tatcgcgatt	gcagtgccac	180
tggctgggtt	cgctaccgta	gcgcaggccg	aagttaaac	gcaggaatcc	gggtggtggtc	240
tggttcagcc	aggtgggttc	ctgcggctct	cggtgctgct	ttccggtttc	aacatcaaaag	300
acacctacat	ccaactgggtt	cgtaaggctc	cggttaaagg	cctggaaatgg	gttgctcgta	360
tctaccocgac	caacggttac	accaggatag	ccgattcagt	taaaggctcgt	ttcaccatct	420
cgcccgacac	ttccaaaaac	accgcttacc	tccagatgaa	ctccctgctg	gctgaagaca	480
cagctgttta	ttattgctcc	cgctgggggtg	gtgacggttt	ctacgctatg	gactactggg	540
gtcagggtag	cctggctacc	gtctctcag	cctccaccaa	gggcccctcg	gtcttcccc	600
tggcaacctc	ctccaagagc	acctctgggg	gcacagcggc	cctgggctgc	ctggctcaagg	660
actacttccc	cgaaaccggtg	acgggtgctg	ggaactcagg	cgccctgacc	agcggcgtgc	720
acaccttccc	ggctgtctcag	gactctactc	cctcagcagc	gtggtcagtg	780	
tgccctccag	cagcttgggc	accocagact	acatctgcaa	cgtaatcac	aaaccocagca	840
acaccaaggt	cgacaagaaa	ggtgagccca	aaccttgcca	tcaccaccat	caccattaat	900
aaccttggag	aaaataaagt	gaaacaaagc	actattgcac	tggcactctt	accgttactg	960
tttaccctcg	tgacaaaagc	cgacatcgag	ctcacccaat	ccccgtctc	cctgtccgct	1020
tccgttggcg	accgtgttac	ctcactggtg	agggcctgcg	aagacgtaaa	caccgocgta	1080
gcgtggtatc	agcaagaaac	cgggaaagct	ccgaaactgc	tgatctatag	cgcttctctc	1140
ctgtattccg	gagttccgag	caggttcagt	ggttcccgtt	ccggtaccga	cttcaccctg	1200
acgatatcct	ccctccagcc	ggaagaacttc	gctacctact	actgtcaaca	gcactacacc	1260
accccgcgca	ccttcgggtca	gggtaccaaa	ctcgagatca	aacggactgt	ggctgcacca	1320
tctgtcttca	tcttcccgcc	atctgatgag	cagttgaaat	ctggaaactgc	ctctgtttgtg	1380
tgcctgctga	ataacttcta	tcccagagag	gccaaagtac	agtggaaaggt	ggataacgcc	1440
ctccaatcgg	gtaactccca	ggagagtgtc	acagagcagg	acagcaagga	cagcaacctac	1500
agcctcagca	gcaccctgac	gctgagcaaa	gcagactacg	agaaacacaa	agtctacgcc	1560
tggaaagtca	cccacaggg	cctgagttcg	cccgtcacia	agagcttcaa	ccgcgagag	1620
tgtcttctcg	ccgctgcacc	tgtctgcagca	uctgtgcag	ctccagcagc	tgctcctgca	1680
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	1740
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	1800
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	1860
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	1920
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	1980
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	2040
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	2100
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	2160
gcagctccag	ccgctgcacc	tgctgcagca	cctgctgcag	ctccagcagc	tgctcctgca	2220
gcagctccag	cctgaagagc	ttaagcttga	cctgtgaaat	gaaaaatggc	gcacattgtg	2280
cgacattttt	tttgtctgac	gtttaccgct	actgctcac	ggatctccac	gcgccctgta	2340
gcggcgcatt	aagcggcggc	gggtgtggtg	ttacgcgcag	cgtgaccgct	acacttgcca	2400
gcgccttagc	gcccgcctct	ttcgtttct	tcccttctt	tctgcgccag	ttcgcggcct	2460
ttcccctgca	agctctaaat	cgggggctcc	ctttaggggt	ccgatttagt	gctttacggc	2520
acctcgaccc	caaaaaactt	gattaggggtg	atggttcacg	tagtgggcca	tgcacctgat	2580
agacggtttt	tgcacctttg	acgttggagt	ccacgttctt	taatagtggg	ctcttgttcc	2640
aaactggaac	aacactcaac	cctatctcgg	tctattcttt	tgatttataa	gggattttgc	2700
cgatttccgc	ctattgggta	aaaaatgagc	tgatttaaca	aaaatttaac	gcgaatttta	2760
acaaaatatt	aacgtttaca	atttcaggtg	gcacttttcc	gggaaatgtg	cgcggaaccc	2820
ctatttggtt	attttcttaa	atacattcaa	atatgtatcc	gctcatgaga	caataacccc	2880
gataaagct	tcaataatat	tgaaaaagga	agagtatgag	tattcaacat	ttccgtgctg	2940
cccttattcc	cttttttgcg	gcattttgcc	ttcctgtttt	tgctcaccca	gaaacgctgg	3000
tgaagagtaa	agatgctgaa	gatcagttgg	gtgcacaggt	gggttacatc	gaactggatc	3060

[0267]

```

tcaacagcgg taagatcctt gagagtttcc gcccogaaga acgttttcca atgatgagca 3120
cttttaaagt tctgctatgt ggcggcggtat tatcccgat tgaccgccgg caagagcaac 3180
tcggtcgccg catacaactat tctcagaatg acttggttga gtactcacca gtcacagaaa 3240
agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata accatgagtg 3300
ataacactgc ggccaactta cttctgacaa cgatcggag accggaaggag ctaaccgctt 3360
ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg 3420
aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc 3480
gcaactatt aactggcgaa ctacttactc tagcttcccg gcaacaattg atagactgga 3540
tgagagcggg taaagtgtgc ggaaccacttc tgcgctcggc ccttccggct ggctggtta 3600
ttgctgataa atcrggagcc ggtgagcgtg gctctcggg tatkattgca gcaactggggc 3660
cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg 3720
atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat 3780
taatgatgtc tegttagat aaaagfaaag tgattaacag cgattagag ctgcttaatg 3840
aggtcggaat cgaagggtta acaaccgcta aactcggcca gaagctaggt gttagcagc 3900
ctacattgta ttggcatgta aaaaataagc gggctttgct cgacgcctta gccattgaga 3960
tgtagatag gcaccatact cacttttgc ctttagaagg ggaagctgg caagatttt 4020
taogtaataa cgtataaagt tttagatgtg ctttactaag tcatcggat ggagcaaaag 4080
tacaattagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct 4140
ttttatgcca acaaggtttt tcaactagaga atgcattata tgcactcagc gcaqtggggc 4200
attttacttt aggttgcgta ttgggaagatc aagagcatca agtcgctaaa gaagaaagg 4260
aaacacttac tactgatagt atgcggcct tattacgaca agctatcgaa ttatttgatc 4320
accaagtgcc agagccagcc ttcttatcgg gccttgaatt gatcatatgc ggattagaaa 4380
aacaacttaa atgtgaaagt gggctttaa agcagcataa ccttttccg tgatggtaac 4440
ttcaactagt taaaaggatc laggtgaaaga tctttttga taatctcatg accaaaatcc 4500
cttaactgta gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt 4560
ctgagatcc ttttttctg cgcgtaactc gctgcttgca aacaaaaaaa ccaccgctac 4620
cagcgggtgt ttgtttgccc gatcaagagc taccactct ttttccgaag gtaactggct 4680
tcagcagagc gcagatacca aatactgtcc ttctagtga gccgtagtta ggccaccact 4740
tcaagaacte tgtaqcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg 4800
ctgcaagtgc cgataagtgc gtcttaccg ggttggaact aagacgatag ttaccggata 4860
aggcgcagcg gtcgggctga acggggggtt cgtgcacaca gccagcctg gagcgaacga 4920
cctacaccga actgagatc ctacagcgtg agctatgaga aagcggccag cttcccgaag 4980
ggagaaagcc ggacaggtat ccggtaaagc gcagggctcg aacaggagag cgcacgaggg 5040
agcttccagc gggaaacgct ttggtacttt atagtctgt cgggtttcgc caactctgac 5100
ttgagcgtcg atttttgtga tgcctctcag gggggcggag cctatggaaa aacgccagca 5160
acgcggcctt tttacggctc ctggcctttt gctggccttt tgctccatg 5210

```

[0268]

[0269] 실시예 18: 분석용 겔 여과에 의한 Fab 단편 및 유전적으로 인코딩된 P1A1 또는 P1A3 폴리펩티드/중합체간 재조합 융합 단백질에 대한 수력학적 부피의 측정.

[0270] 수행 완충액인 PBS로 Akta 정제기 10 시스템(GE Healthcare)을 이용하여 1 ml/분의 유속으로 Superdex S200 HR 10/300 GL 칼럼(GE Healthcare Europe, Freiburg, Germany) 상에서 SEC을 수행하였다. 실시예 4의 Fab-PA#1(200)에 대해 기재된 바와 유사하게 제조 및 정제된(도 9) Fab-P1A1(200) 및 Fab-P1A3(200) 융합 단백질의 250 μl 표본을 PBS 중에 0.25 mg/ml 농도로 개별 처리하였다. 두 단백질은 모두 도 10에 나타난 바와 같이 균질한 단일 피크로 용출되었다.

[0271] 그 결과, 200개 잔기의 P1A1 또는 P1A3 중합체/폴리펩티드를 갖는 융합 단백질은 대응 미융합 Fab 단편보다 훨씬 더 큰 크기를 나타내었다. Fab-P1A1(200) 및 Fab-P1A3(200)에 대한 겔보기 크기 증가는 Fab 단편(cf. 도 4B) 대비 각각 5.8배 및 5.2배였지만 실제 질량은 단지 1.4배 및 1.3배 더 컸다. 상기 관찰은 본 발명에 따른 생합성 P1A1 및 P1A3 폴리펩티드 절편에 의해 생물학적 활성 Fab 단편에 훨씬 증가된 수력학적 부피가 부여됨을 뚜렷이 나타낸다.

[0272] 실시예 19: 원형 광이색형(CD) 분광측정을 통한 Fab 단편에 융합된 생합성 P1A1 및 P1A3 중합체/폴리펩티드에 대한 랜덤 코일 입체형태의 검출.

[0273] Fab-P1A1(200) 및 Fab-P1A3(200)에 대한 CD 스펙트럼을 완충액으로 50 mM K<sub>2</sub>SO<sub>4</sub>, 20 mM K-포스페이트(pH 7.5)를 사용해서 실시예 4에 기재된 바와 유사하게 제조된 4.2 및 6.5 μM 단백질 용액을 각각 사용하여 실시예 8에 기재된 바와 같이 기록하였다.

[0274] Fab-P1A1(200) 및 Fab-P1A3(200) 융합 단백질에 대한 스펙트럼은 상당한 랜덤 코일 입체형태 분획을 나타내었다(도 11A). Pro/Ala 폴리펩티드 절편에 의한 분광학적 기여를 더욱 자세히 분석하기 위해, 미융합 Fab 단편에 대한 물 차이 CD 스펙트럼(실시예 8 참고)을 동일한 물 농도로 정규화한 후 Fab-P1A1(200) 및 Fab-P1A3(200) 각각에 대한 스펙트럼에서 후자의 스펙트럼을 감산하여 계산하였다(도 11B). 그 결과, 랜덤 코일 입체형태의 특징인 대략 200 nm 파장에서의 강력한 최소 부위가 관찰되었다. 따라서 재조합 융합 단백질의 일부로서의

P1A1 및 P1A3 서열은 생리학적 완충액 조건 하에 랜덤 코일 입체형태로 존재하는 것으로 나타난다.

- [0275] 실시예 20: His(6)-SUMO-PA#1(200) 융합 단백질에 대한 발현 벡터로서의 pSUMO-PA#1(200)의 구축.
- [0276] 200개 잔기의 PA#1 서열 반복체에 융합된 6개 잔기 His-태그 및 작은 유비퀴틴-유사 개질체(SUMO) 단백질 (Panavas (2009) Methods Mol. Biol. 497: 303-17)을 인코딩하는 발현 플라스미드의 구축을 위해, 사카로마이세스 세레비지에로부터의 SUMO 단백질[Smt3p; Uniprot: Q 12306으로도 알려져 있음]이 클로닝된 cDNA로부터 폴리머라아제 연쇄 반응(PCR)을 통해 증폭되었다. 5'-프라이머는 Met 시작 코돈(ATG) 및 추가적 Lys 코돈뿐만 아니라 His<sub>6</sub>-태그 인코딩 서열을 함유하는 *NdeI* 제한 부위를 도입한 반면, 3'-프라이머는 PCR 산물 내로 *HindIII* 및 *SapI* 제한 부위를 도입하였다. 생성 DNA 단편은 *NdeI* 및 *HindIII*로 소화되고, *SapI* 제한 부위가 침묵 돌연변이에 의해 제거된 플라스미드 pSA1(Schmidt (1994) J. Chromatogr. 676: 337-345)의 대응 절단된 유도체와 라이게이션 되었다. 생성 플라스미드를 *SapI*로 절단하고 새우 알칼린 포스포타아제에 탈인산화하고, *SapI*로 제한 절단된(도 2E에 예시된 것과 유사한 방식으로) 플라스미드 pFab-PA#1(200)(실시예 2에 기재됨)에서 절단된 200개 잔기의 PA#1 폴리펩티드 절편을 인코딩하는 유전자 단편과 라이게이션하였다. 생성 플라스미드는 pSUMO-PA#1(200)(서열 번호 60)로 명명되었으며, 도 12A에 나타난다.
- [0277] 실시예 21: 유전적으로 인코딩된 PA#1(200) 중합체/폴리펩티드의 박테리아를 이용한 발현 및 단리.
- [0278] PA#1(200)폴리펩티드(계산 질량: 16.1 kDa)를 먼저 발현 플라스미드 pSUMO-PA#1(200)(실시예 21에 기재됨)과 함께 T7 프로모터를 저해하는 플라스미드 pLysE(Studier (1991) J. Mol. Biol. 219: 37-44)를 포함하는 *E. coli* BLR(DE3)(NEB, Ipswich, MA, USA)의 세포질에서 작은 유비퀴틴-유사 개질체(SUMO) 단백질(계산 질량: 12.2 kDa)을 갖는 융합 단백질로 제조하였다. 2.5 g/l D-글루코오스, 0.5 g/l L-프롤린, 100 mg/l 엠포실린, 및 30 mg/l 클로람페니콜을 함유하는 2 l LB 배지를 갖는 진탕 플라스크 배양으로 30°C에서 박테리아를 이용한 생산을 수행하였다. 제조용 유전자 발현은 이소프로필-β-D-티오갈락토피라노시드(IPTG)를 최종 농도 0.5 mM로 첨가하여 유도하였다. 박테리아를 유도 3시간 후에 수확하고 100 mM NaCl, 40 mM Na-포스페이트(pH 7.5) 중에 재현탁하고 프렌치 프레스 셀(Thermo Scientific, Waltham, MA, USA)을 이용하여 용해시켰다. 용해물의 원심분리(15분, 15000 g) 후 용입체는 관찰되지 않았다.
- [0279] 가용성 융합 단백질을 함유하는 상청액을 70°C에서 15분 동안 인큐베이션하고 원심분리(15분, 15000 g)하여 열적으로 불안정한 숙주 세포 단백질을 제거하였다. His(6)-SUMO-PA#1(200) 융합 단백질을 Akta 정제기 시스템(GE Healthcare)에 연결된 12 ml Ni<sub>2</sub><sup>+</sup> 하전된 HisTrap 고성능 칼럼(GE Healthcare)을 이용하여 IMAC(Skerra (1994) Gene 141:79-84)을 통해 상청액으로부터 정제하고 500 mM NaCl, 40 mM Na-포스페이트(pH 7.5) 중 0 내지 150 mM의 이미다졸 구배로 용출하였다. 이어서 제조용 SEC 단계 후, His(6)-SUMO-PA#1(200) 융합 단백질(도 12B)의 균질한 제조물을 OD550 = 1인 1 l 박테리아 배양물 당 대략 5 mg의 수율로 수득하였다. His(6)-SUMO-PA#1(200) 폴리펩티드 용합물에 대해 1280 M<sup>-1</sup> cm<sup>-1</sup>의 계산 흡광도 계수(Gill (1989) 위 인용문)를 이용하여 280 nm에서의 흡광도에 따라 단백질 농도를 결정하였다. PA#1(200) 폴리펩티드 절편은 그 방향족 또는 방향족 함유 아미노산 측쇄의 부재로 인해 280 nm에서의 흡광도에 기여하지 않음을 주지하라.
- [0280] 생합성 PA#1(200) 폴리펩티드는 절단 완충액(0.2w/v% Igepal, 1 mM DTT, 150 mM NaCl, 50 mM 트리스-HCl, pH 8.0) 중 30°C에서 1시간 동안 사카로마이세스 세레비지에 유래의 Ubl-특이적 프로테아제 1(Invitrogen, Carlsbad, CA, USA) 2 U/mg으로 부위 특이적인 단백질 절단(Pro/Ala 폴리펩티드 절편에 선행하는 Gly-Gly 모티프 하류)에 의해 융합 단백질로부터 방출시켰다. 절단 공정은 높은 몰 농도의 트리스 완충액 시스템(Fling (1986) Anal. Biochem. 155:83-88)을 이용하여 SDS-PAGE(도 12B)로 확인하였다. 모두 His<sub>6</sub>-태그를 수반하는 절단된 His(6)-SUMO 단백질, 잔여 미절단 융합 단백질, 및 또한 SUMO 프로테아제를 제거하기 위해, 반응 혼합물을 수행 완충액으로 500 M NaCl, 20 mM 포스페이트(pH 7.5) 및 5 ml Ni<sub>2</sub><sup>+</sup> 하전 HisTrap 고성능 칼럼(GE Healthcare)을 이용해 또 다른 IMAC을 수행하였다. 이 시간 동안 용출 분획은 순수한 생합성 PA#1(200) 폴리펩티드를 함유하였다(도 13E). 이러한 방식으로 제조된 생합성 PA#1(200) 폴리펩티드/중합체(서열 번호 61)는 도 1에 나타낸 바와 같이 각각 20개 아미노산 잔기를 인코딩하는 10개의 라이게이션된 이중쇄 올리고데옥시뉴클레오티드 구성 블록의 인코딩된 조합 유전자 산물에서 기인하는 201개 아미노산 잔기와 함께 클로닝을 위해 사용된 *SapI* 제한 부위 하류의 삼중자 DNA 오버행에 의해 인코딩된 추가적 Ala 잔기를 포함한다.
- [0281] 실시예 22: PA#1(200)을 갖는 소분자/약물 콘쥬게이트의 제조 및 특징분석.



- [0282] 실시예 21의 His(6)-SUMO-PA#1(200) 융합 단백질의 미정제 단백분해 절단 반응 혼합물을 50 mM NaHCO<sub>3</sub>(pH 8.3)에 대해 4℃에서 2회 투석하고 무수 디메틸 포름아미드(DMF) 중 6-[플루오레신-5(6)-카르복사미도]헥산산 N-히드록시숙신이미드 에스테르(플루오레신-NHS 에스테르; Sigma-Aldrich)의 10배 몰 과량 용액과 혼합 후 실온에서 1시간 동안 인큐베이션하였다. 이를 위해, His(6)-SUMO-PA#1(200) 절단 혼합물의 2.5 mg/ml 용액 200 μl를 DMF 중에 용해된 플루오레신-NHS 에스테르 10 mM 용액 17.6 μl에 첨가하였다. 생성 혼합물을 실온에서 1시간 동안 인큐베이션하고 실시예 21에 기재된 바와 같이 IMAC에 처리하여 절단된 His(6)-SUMO 단백질, 잔여 미절단 융합 단백질, 및 SUMO 프로테아제를 제거하고, 0.5 ml/분의 유속으로 PBS로 평형화된 Superdex S200 10/300 GL 칼럼상에서 제조용 SEC으로 추가 정제하였다.
- [0283] 이어서 상이한 단계로부터의 표본을 0.5 ml/분의 유속으로 PBS로 평형화된 Superdex S200 10/300 GL 칼럼상에서 분석용 SEC을 통해 분석하였다. SUMO 단백질은 280 nm에서 그 방향족 측쇄를 통해 검출되고 Pro/Ala 폴리펩티드 또는 폴리펩티드 절편의 펩티드 결합을 포함하는 펩티드 결합은 225 nm에서 검출된 반면, 플루오레신은 494 nm에서 검출되었다(도 13A-G). 대조를 위해, 유리 플루오레신(Sigma-Aldrich) 용액 및 SEC에서 검출된 각각의 특징적 피크 분석의 UV/VIS 스펙트럼을 램다 9 기기(Perkin Elmer, Waltham, MA, USA)를 이용해 측정하였다(도 13H-K). 크로마토그래피 칼럼의 크기 적정을 위해(도 13L), 하기 구형 단백질들의 적절한 혼합물(Sigma-Aldrich) 250 μl를 0.2 내지 0.5 mg/ml 농도의 PBS 중에 처리하였다: 아프로티닌, 6.5 kDa; 사이토크롬 c, 12.4 kDa; 카르복산 안하이드라아제, 29.0 kDa; 소 혈청 알부민, 66.3 kDa; 알코올 탈수소효소, 150 kDa; β-아밀라아제, 200 kDa; 아포-페리틴, 440 kDa.
- [0284] 그 결과, 생합성 PA#1(200) 폴리펩티드/중합체와 플루오레신-NHS 에스테르의 커플링 후, 본질적으로 PA#1(200) 폴리펩티드/중합체의 크기 특성 및 소분자, 즉 플루오레신기의 분광학적 유의성을 나타내는 거대분자 콘주게이트가 IMAC 및 SEC을 통해 단리되었다. 이는 소분자가 생합성 Pro/Ala 폴리펩티드/중합체에 성공적으로 커플링되었으며, 본 발명에 따라 콘주게이션된 소분자 약물 또는 화합물의 수력학적 부피가 상당히 증가됨을 나타낸다.
- [0285] 생합성 Pro/Ala 폴리펩티드/중합체 및 식물 스테로이드 디곡시제닌간 유사한 콘주게이트를 제조하기 위해, 실시예 21의 정제된 PA#1(200) 폴리펩티드 0.1 mg을 상술된 바와 같이 50 mM NaHCO<sub>3</sub>(pH 8.3)에 대해 투석하였다. 정제된 PA#1(200) 폴리펩티드의 농도를 205 nm(Gill (1989) 위 인용문)에서 흡광도에 따라 결정하였다. PA#1(200) 폴리펩티드를 디곡시제닌-3-O-메틸카르보닐-ε-아미노카프로산 NHS 에스테르(DIG-NHS 에스테르; Roche Diagnostics, Mannheim, Germany)의 10배 몰 과량으로 커플링하였다. 상기 목적을 위해, 50 mM NaHCO<sub>3</sub>(pH 8.3) 중 정제된 PA#1(200) 폴리펩티드의 1 mg/ml 용액 100 μl를 무수 DMF 중에 용해된 DIG-NHS 에스테르의 30 mM 용액 2 μl에 첨가하고 반응 혼합물을 실온에서 1시간 인큐베이션하였다. 콘주게이트의 생성 용액을 7 kDa 컷오프의 Zeba™ 스피ن 탈염 칼럼(Thermo Scientific)을 이용하여 정제하고, 10 mM 암모늄 아세테이트 완충액(pH 6.8)에 대해 2회 투석하고 양이온 모드를 이용하여 Q-ToF Ultima 기기(Waters, Eschbronn, Germany) 상에서 ESI 질량 분광측정을 통해 분석하였다. 그 결과, 디곡시제닌-PA#1(200) 콘주게이트의 스펙트럼은 계산 질량 16670.6 Da과 본질적으로 일치하는 16671.4 Da의 질량을 나타내었다(도 13M). 이는 생합성 Pro/Ala 폴리펩티드/중합체, 특히 PA#1(200)가 소분자 약물과 효율적으로 콘주게이션될 수 있음을 뚜렷이 나타낸다.
- [0286] 본 발명은 하기 예시된 서열에 관한 것이며 이를 나타내고, 첨부된 서열 목록은 설명의 일부로 제공되며, 따라서 본 명세서의 일부이다.
- [0287] 서열 번호 1은 PA#1의 아미노산 서열을 나타낸다.
- [0288] 서열 번호 2는 PA#2의 아미노산 서열을 나타낸다.
- [0289] 서열 번호 3은 PA#3의 아미노산 서열을 나타낸다.
- [0290] 서열 번호 4는 PA#4의 아미노산 서열을 나타낸다.
- [0291] 서열 번호 5는 PA#5의 아미노산 서열을 나타낸다.
- [0292] 서열 번호 6은 PA#6의 아미노산 서열을 나타낸다.
- [0293] 서열 번호 7은 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0294] 서열 번호 8은 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.

- [0295] 서열 번호 9는 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0296] 서열 번호 10은 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0297] 서열 번호 11은 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0298] 서열 번호 12는 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0299] 서열 번호 13은 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0300] 서열 번호 14는 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0301] 서열 번호 15는 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0302] 서열 번호 16은 서열 번호 1의 환상 순열 버전의 아미노산 서열을 나타낸다.
- [0303] 서열 번호 17은 구성 블록 PA#1의 생성을 위해 사용되는 상부/코딩쇄 올리고데옥시뉴클레오티드의 핵산 서열을 나타낸다.
- [0304] 서열 번호 18은 구성 블록 PA#1의 생성을 위해 사용되는 하부/비코딩쇄 올리고데옥시뉴클레오티드의 핵산 서열을 나타낸다.
- [0305] 서열 번호 19는 pASK88-Fab-2xSapI 상에 인코딩된 항체 Fab 단편의 면역글로불린 경쇄 C-말단 근처의 핵산 서열 연장부(상부/코딩쇄)를 나타낸다.
- [0306] 서열 번호 20은 pASK88-Fab-2xSapI 상에 인코딩된 항체 Fab 단편의 면역글로불린 경쇄 C-말단 근처의 핵산 서열 연장부(하부/비코딩쇄)를 나타낸다.
- [0307] 서열 번호 21은 pASK88-Fab-2xSapI 상에 인코딩된 Fab 단편의 C-말단의 아미노산 서열을 나타낸다.
- [0308] 서열 번호 22는 pASK88-Fab-2xSapI의 핵산 서열을 나타낸다.
- [0309] 서열 번호 23은 하나의 PA#1(20) 중합체 삽입 후 Fab 경쇄 C-말단의 아미노산 서열을 인코딩하는 핵산 서열 연장부(상부/코딩쇄)를 나타낸다.
- [0310] 서열 번호 24는 하나의 PA#1(20) 중합체 삽입 후 Fab 경쇄 C-말단의 아미노산 연장부에 대한 핵산 서열(하부/비코딩쇄)을 나타낸다.
- [0311] 서열 번호 25는 하나의 PA#1(20) 중합체 삽입 후 Fab 경쇄 C-말단의 아미노산 서열 연장부를 나타낸다.
- [0312] 서열 번호 26은 pFab-PA#1(200) 상에 인코딩된 Fab 중쇄의 아미노산 서열을 나타낸다.
- [0313] 서열 번호 27은 pFab-PA#1(200) 상에 인코딩된 PA#1(200) 중합체에 융합된 Fab 경쇄의 아미노산 서열을 나타낸다.
- [0314] 서열 번호 28는 pFab-PA#1(200)의 핵산 서열을 나타낸다.
- [0315] 서열 번호 29는 *Strep*-태그 II(마지막 두 아미노산만) 및 INFa2b N-말단의 아미노산 서열을 인코딩하는 핵산 서열(상부/코딩쇄)을 나타낸다.
- [0316] 서열 번호 30은 *Strep*-태그 II(마지막 두 아미노산만) 및 INFa2b N-말단의 아미노산 서열을 인코딩하는 핵산 서열(하부/비코딩쇄)을 나타낸다.
- [0317] 서열 번호 31은 INFa2b N-말단 및 *Strep*-태그 II C-말단의 아미노산 서열을 나타낸다.
- [0318] 서열 번호 32는 pASK-IFNa2b의 핵산 서열을 나타낸다.
- [0319] 서열 번호 33은 하나의 PA#1 중합체 서열 카세트의 삽입 후 INFa2b N-말단 및 *Strep*-태그 II C-말단을 인코딩하는 핵산 서열 연장부(상부/코딩쇄)를 나타낸다.
- [0320] 서열 번호 34는 하나의 PA#1 중합체 서열 카세트의 삽입 후 INFa2b N-말단 및 *Strep*-태그 II C-말단의 핵산 서열 연장부(하부/비코딩쇄)를 나타낸다.
- [0321] 서열 번호 35는 하나의 PA#1 중합체 카세트와의 융합 후 INFa2b N-말단 및 *Strep*-태그 II C-말단의 아미노산 서열 연장부를 나타낸다.

- [0322] 서열 번호 36은 pPA#1(200)-IFNa2b 상에서 인코딩된 PA#1(200) 중합체와 융합된 IFNa2b 및 Strep-태그 II의 아미노산 서열을 나타낸다.
- [0323] 서열 번호 37은 pPA#1(200)-IFNa2b의 핵산 서열을 나타낸다.
- [0324] 서열 번호 38은 His<sub>6</sub>-hGH의 N-말단 근처 아미노산 서열을 인코딩하는 pASK75-His6-hGH 상의 핵산 서열 연장부(상부/코딩쇄)를 나타낸다.
- [0325] 서열 번호 39는 hGH의 N-말단 근처 아미노산 서열을 인코딩하는 pASK75-His6-hGH 상의 핵산 서열 연장부(하부/비코딩쇄)를 나타낸다.
- [0326] 서열 번호 40은 pASK75-His6-hGH 상에 인코딩된 His6-hGH N-말단의 아미노산 서열 연장부를 나타낸다.
- [0327] 서열 번호 41은 pASK75-His6-hGH의 핵산 서열을 나타낸다.
- [0328] 서열 번호 42는 PA#1(20) 중합체의 삽입 후 His6-hGH N-말단의 아미노산 서열을 인코딩하는 핵산 서열(상부/코딩쇄) 연장부를 나타낸다.
- [0329] 서열 번호 43은 하나의 PA#1 중합체 서열 카세트의 삽입 후 hGH N-말단을 인코딩하는 핵산 서열(하부/비코딩쇄)을 나타낸다.
- [0330] 서열 번호 44는 PA#1(20) 중합체의 삽입 후 His6-hGH N-말단의 아미노산 서열을 나타낸다.
- [0331] 서열 번호 45는 pASK75-His6-PA#1(200)-hGH 상에 인코딩된 성숙 His6-PA#1(200)-hGH의 아미노산 서열을 나타낸다.
- [0332] 서열 번호 46은 pASK75-His6-PA#1(200)-hGH의 핵산 서열을 나타낸다.
- [0333] 서열 번호 47은 pCHO-PA#1(200)-hGH 상에 인코딩된 His6-PA#1(200)-hGH의 아미노산 서열을 나타낸다.
- [0334] 서열 번호 48은 pCHO-PA#1(200)-hGH의 핵산 서열을 나타낸다.
- [0335] 서열 번호 49는 pCHO-hGH의 핵산 서열을 나타낸다.
- [0336] 서열 번호 50은 pCHO의 핵산 서열을 나타낸다.
- [0337] 서열 번호 51은 P1A1의 아미노산 서열을 나타낸다.
- [0338] 서열 번호 52는 P1A1에 대한 구성 블록의 생성을 위해 사용되는 상부/코딩쇄 올리고데옥시뉴클레오티드의 핵산 서열을 나타낸다.
- [0339] 서열 번호 53은 P1A1에 대한 구성 블록의 생성을 위해 사용되는 하부/비코딩쇄 올리고데옥시뉴클레오티드의 핵산 서열을 나타낸다.
- [0340] 서열 번호 54는 P1A3에 대한 구성 블록의 생성을 위해 사용되는 상부/코딩쇄 올리고데옥시뉴클레오티드의 핵산 서열을 나타낸다.
- [0341] 서열 번호 55는 P1A3에 대한 구성 블록의 생성을 위해 사용되는 하부/비코딩쇄 올리고데옥시뉴클레오티드의 핵산 서열을 나타낸다.
- [0342] 서열 번호 56은 pFab-P1A1(200) 상에 인코딩된 P1A1(200) 중합체와 융합된 Fab 경쇄의 아미노산 서열을 나타낸다.
- [0343] 서열 번호 57은 pFab-P1A3(200) 상에 인코딩된 P1A3(200) 중합체와 융합된 Fab 경쇄의 아미노산 서열을 나타낸다.
- [0344] 서열 번호 58은 pFab-P1A1(200)의 핵산 서열을 나타낸다.
- [0345] 서열 번호 59는 pFab-P1A3(200)의 핵산 서열을 나타낸다.
- [0346] 서열 번호 60은 pSUMO-PA#1(200)의 핵산 서열을 나타낸다.
- [0347] 서열 번호 61은 약물 콘주게이트의 제조에 사용되는 PA#1(200) 폴리펩티드/중합체를 나타낸다(하류 라이게이션 부위로부터 생성된 하나의 추가적 C-말단 Ala 잔기를 포함하는 10개의 20머 인코딩 유전자 카세트의 라이게이션에 의해 제조됨).

도면

도면1

```
gcccGCTCCAGCTGCACCTGCTCCAGCAGCACCCCTGCTGCACCAAGCTCCGGCTGCTCCTGCT
|||||
CGAGGTCGACGTGGACGAGGTCGCTGGACGACGTCGGACGACGTCGAGGCCGACGAGGACGACg
AlaAlaProAlaAlaProAlaProAlaAlaProAlaAlaProAlaProAlaAlaProAlaProAla
```

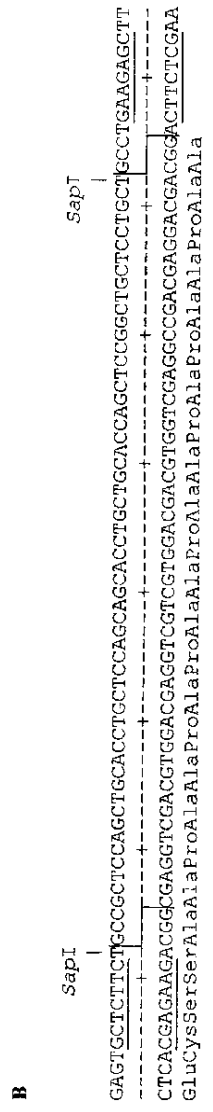


도면2a

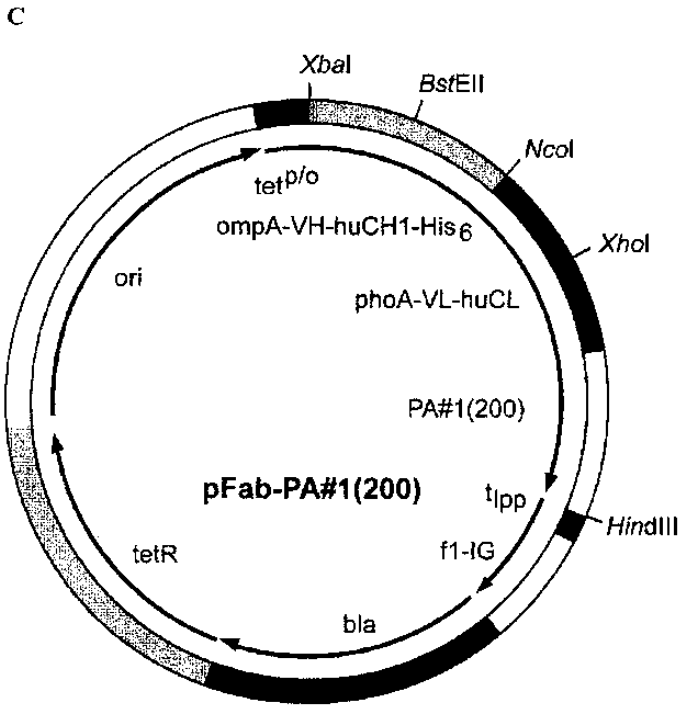
A

```
          SapI   SapI
          |     |
AAGAGCTTCAACCGGGAGAGTGCCTTCTGCCTGAAGAGCTTAAGCTT
+-----+-----+-----+-----+-----+-----+
TTCTCGAAGTTGGCCCTTCCACGAGAGACGGACTTTCGAATTCGAA
LysSerPheAsnArgGlyGluCysSerSerAlaEnd
```

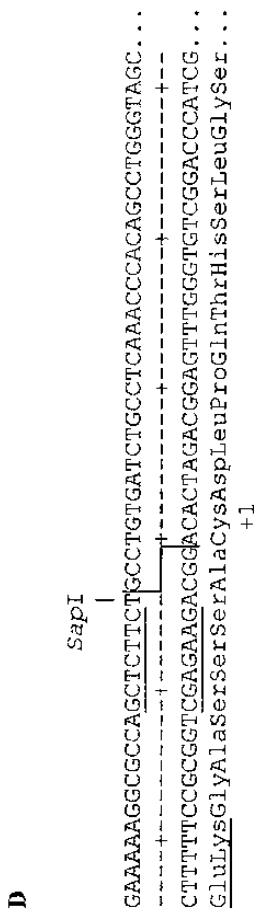
도면2b



도면2c



도면2d

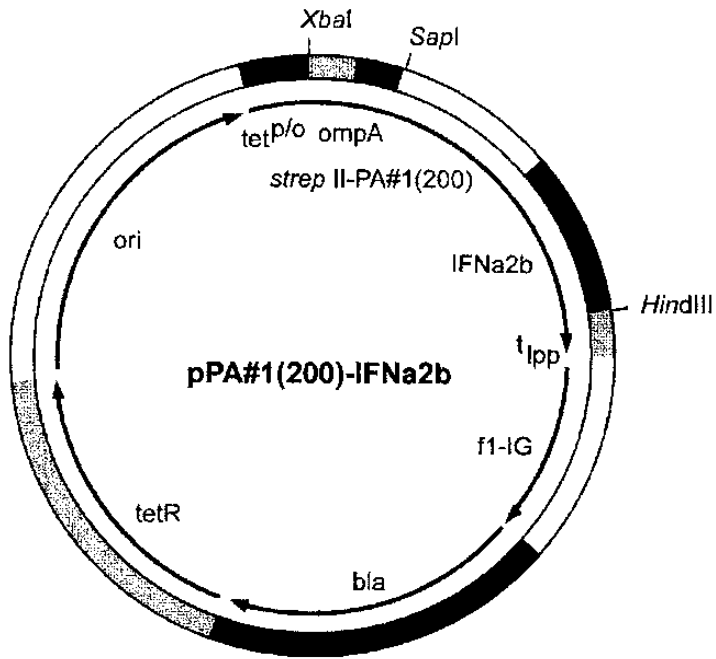






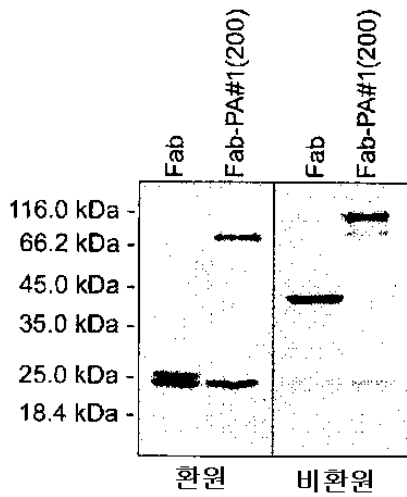
도면2f

F<sup>-</sup>



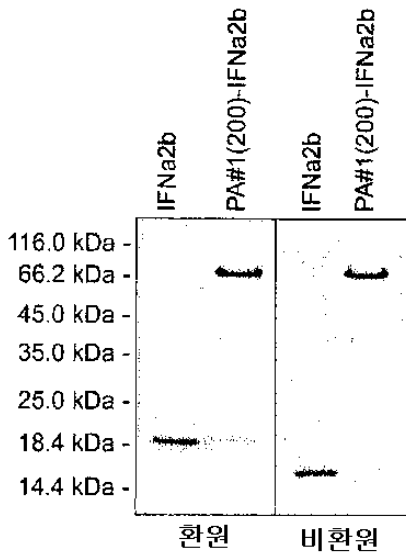
도면3a

A



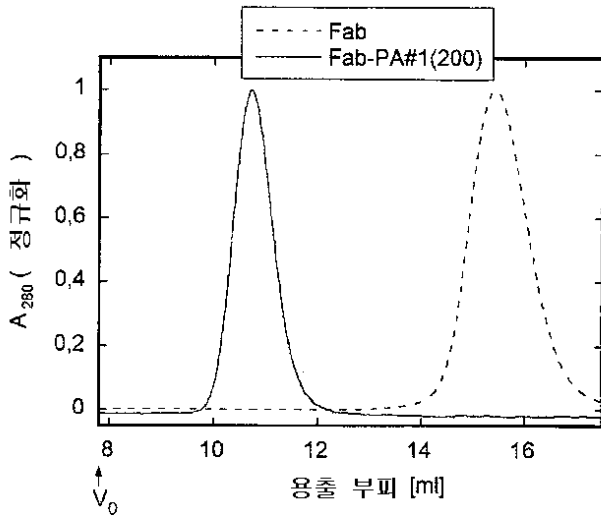
도면3b

B



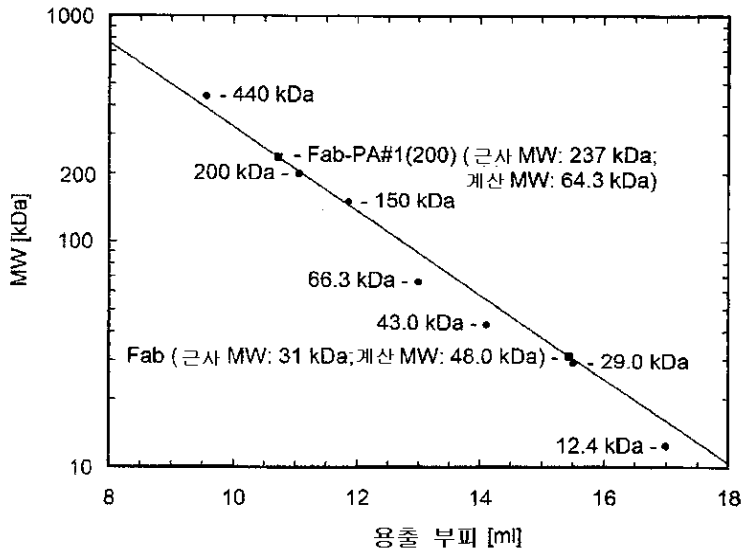
도면4a

A



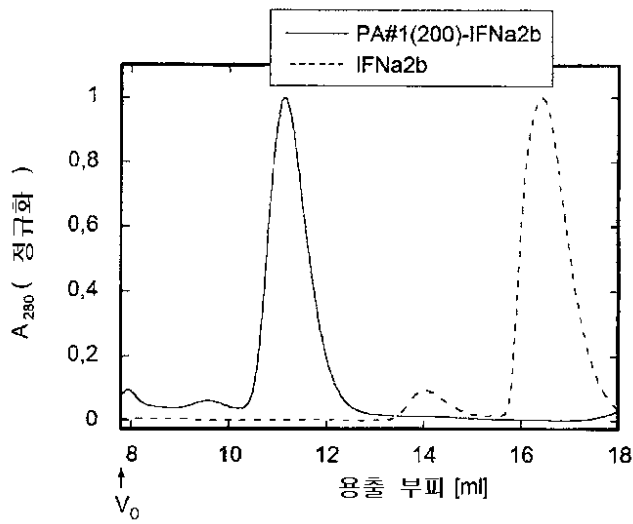
도면4b

B



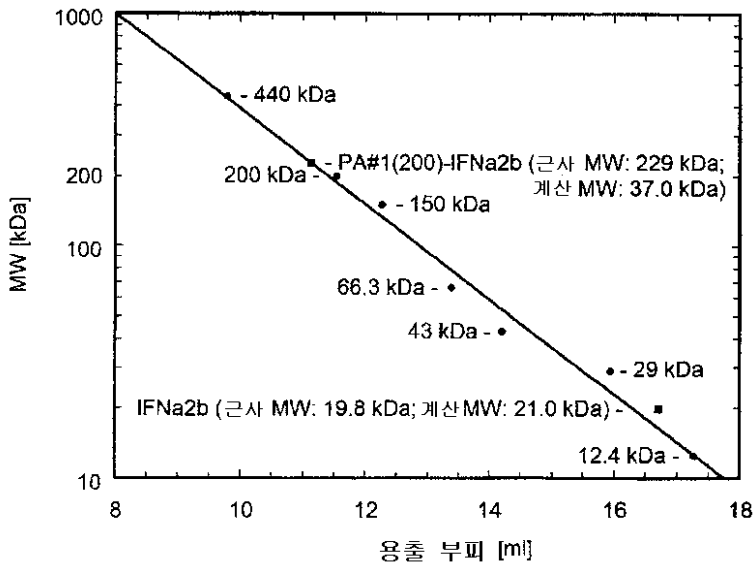
도면4c

C



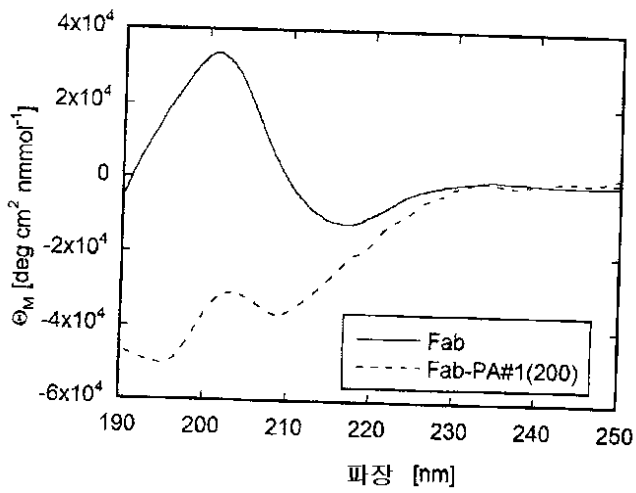
도면4d

D



도면5a

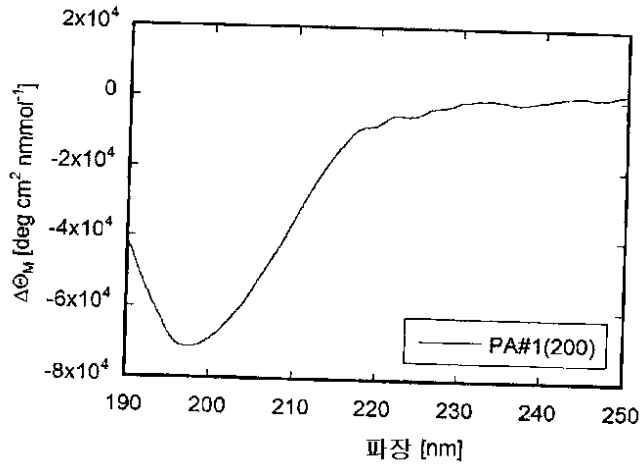
A





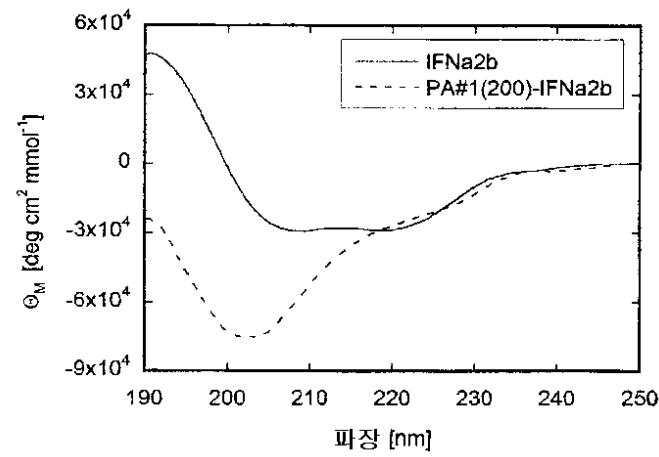
도면5b

B



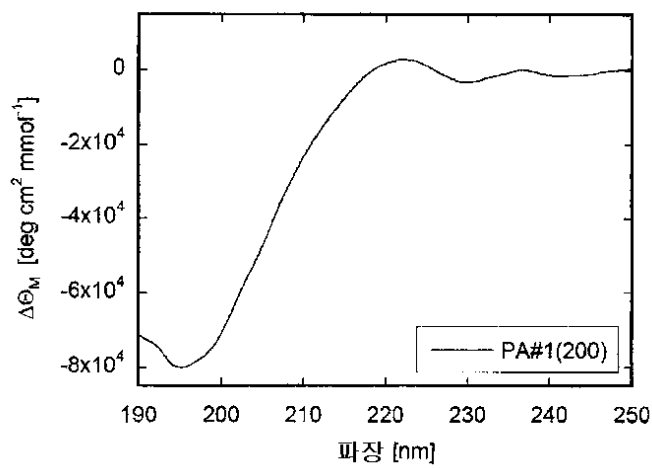
도면5c

C



도면5d

D



도면6a

A

```

NheI                               SspI
|                                   |
GCCGCTAGCCATCACCACCATCACCATGGCGCCAGCTCTTCTGCCTTCCCAACC...
-----+-----+-----+-----+-----+-----+-----+-----+
CGCGGATCGGTAGTGGTGGTGGTGGTACCGCGGTGAGAGACCGGAAGGGTTGG...
AlaAlaSerHisHisHisHisHisHisHisGlyAlaSerSerSerAlaPheProThr... -
                                                    +1
    
```

도면6b

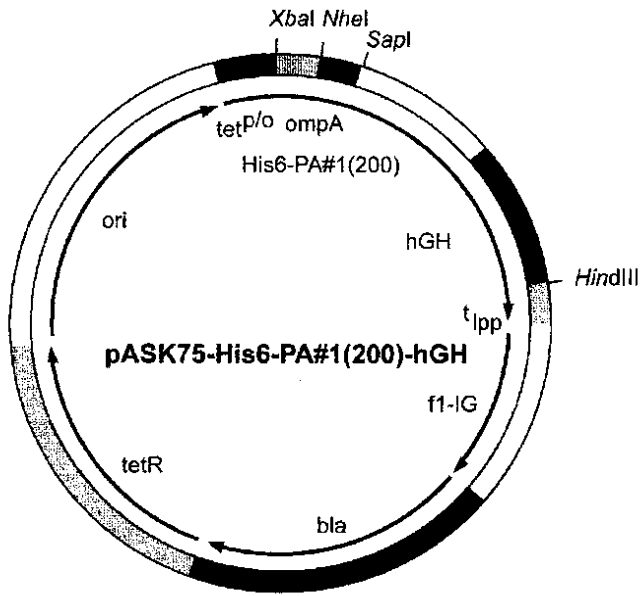
B

```

NheI                               SspI
|                                   |
GCCGTAGCCATCACCACCATCACCATGGCGCCAGCTCTTCTGCCTTCCCAACC...
-----+-----+-----+-----+-----+-----+-----+
CGCGGATCGGTAGTGGTGGTGGTGGTACCGCGGTGAGAGACCGGAAGGGTTGG...
AlaAlaSerHisHisHisHisHisHisHisGlyAlaSerSerSerAlaPheProThr...
GCTGCTCCTGCTGCCTTCCCAACC...
-----+-----+-----+-----+-----+-----+
CGAGGAGGACGACGAGGGTTGG...
AlaAlaProAlaAlaPheProThr...
(1)
    
```

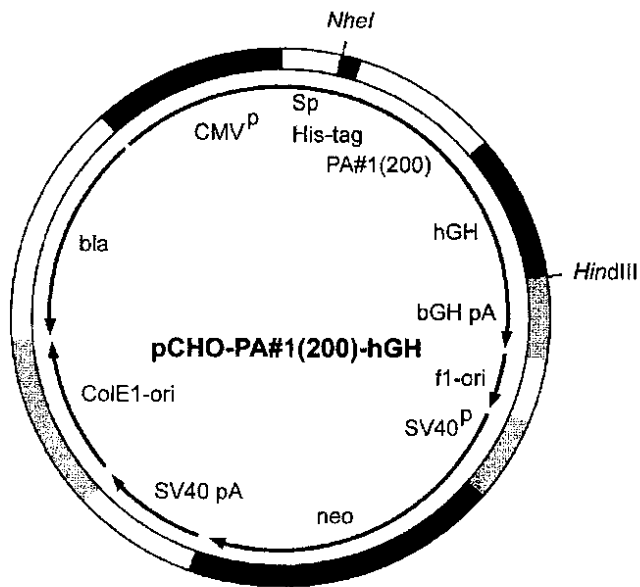
도면6c

C



도면6d

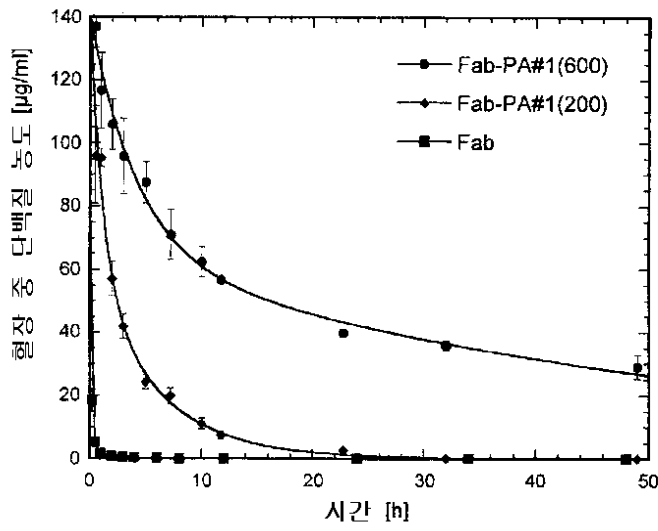
D



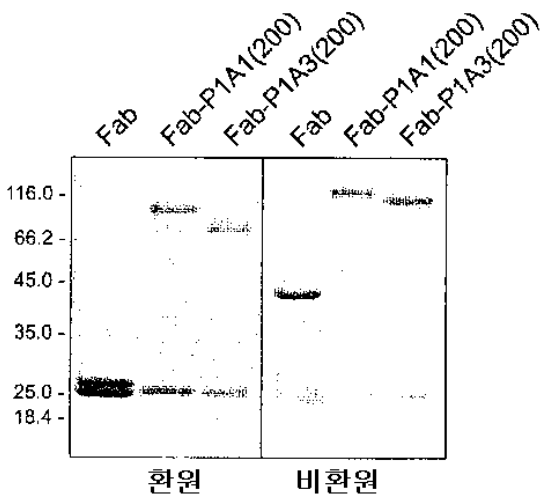




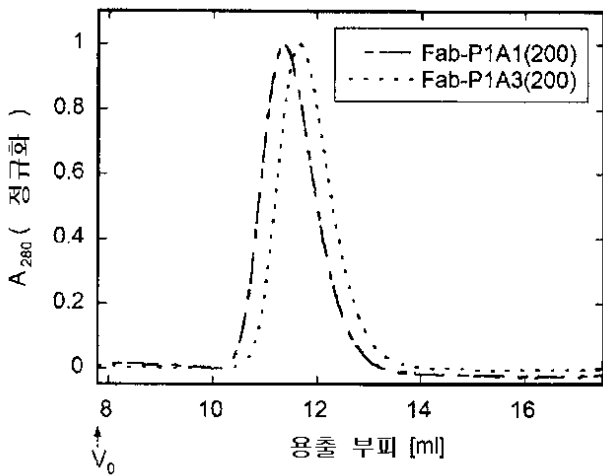
도면8



도면9

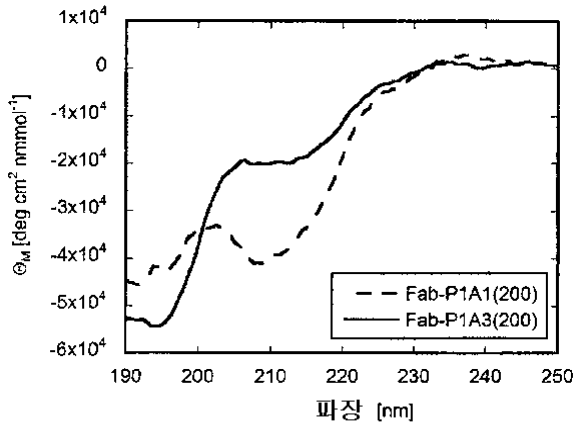


도면10



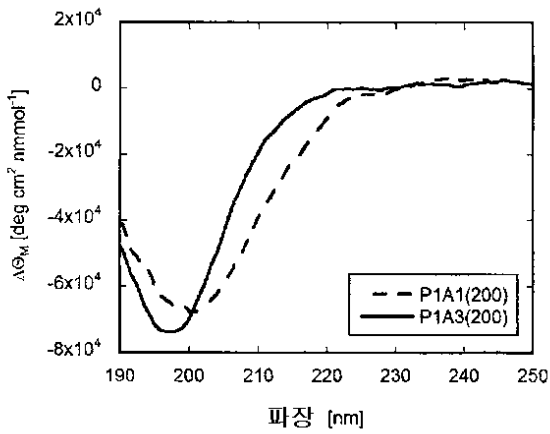
도면11a

A



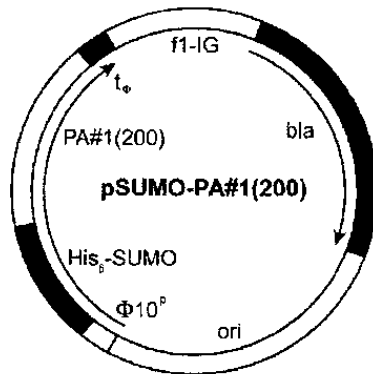
도면11b

B



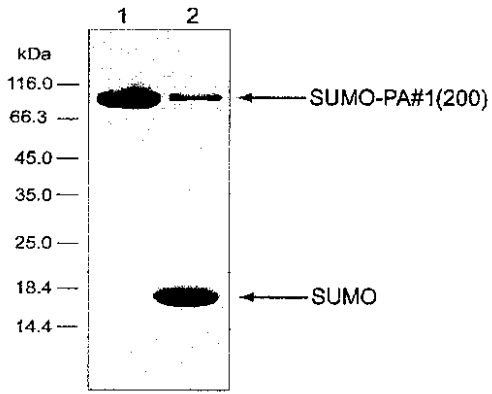
도면12a

A



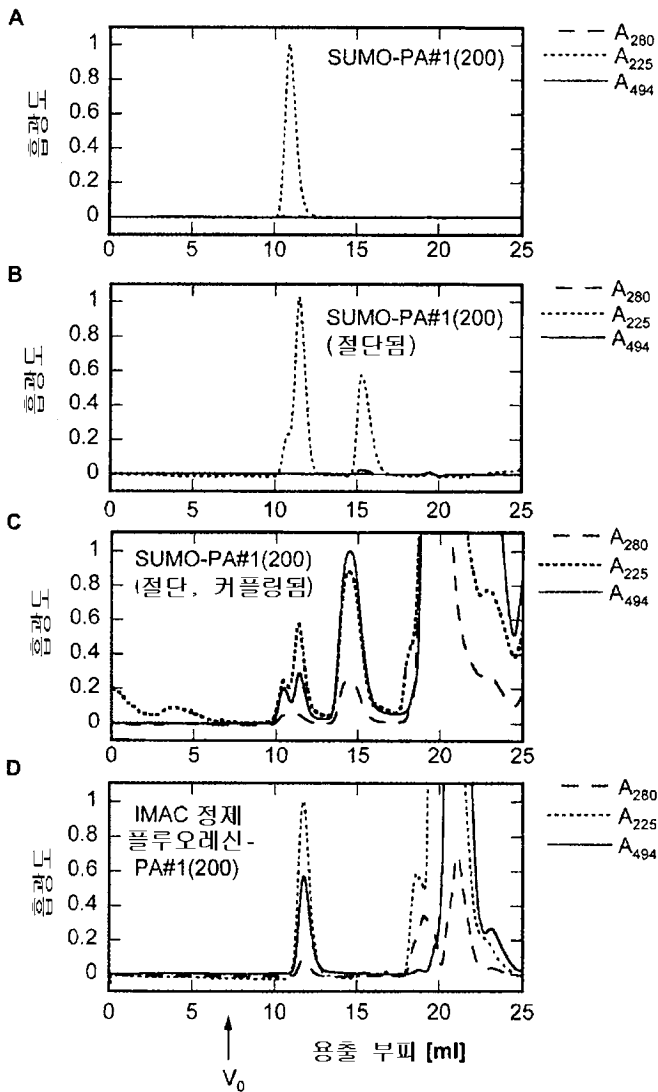
도면12b

B

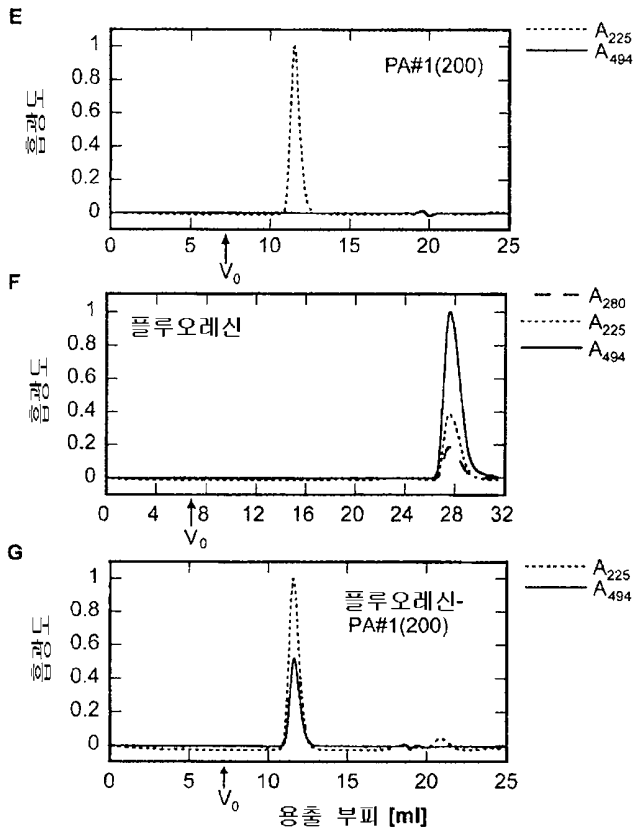


1: 절단 전 SUMO-PA#1(200)  
2: SUMO 프로테아제로 절단한 SUMO-PA#1(200)

도면13a

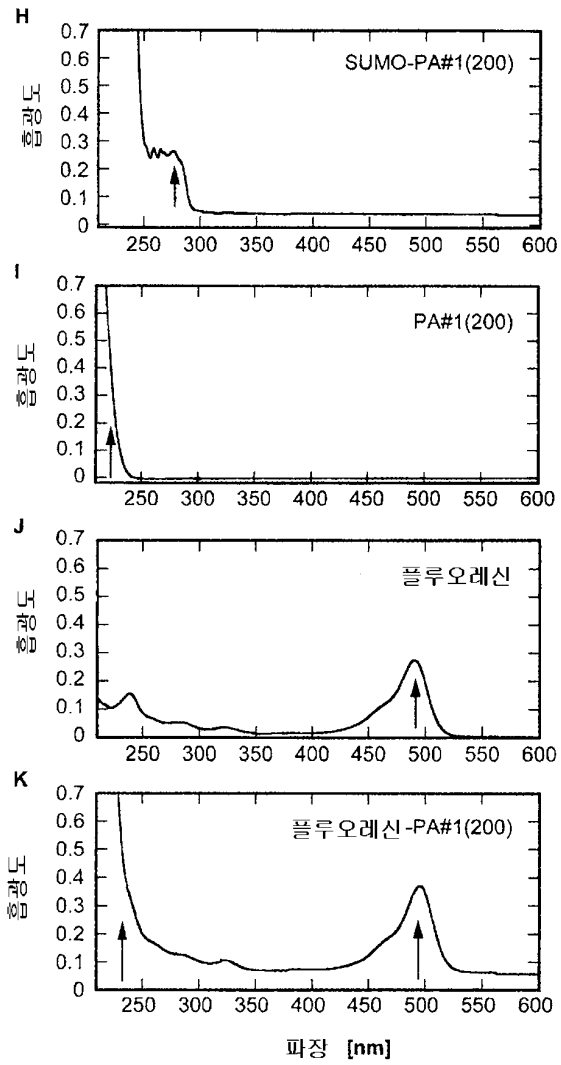


도면13e



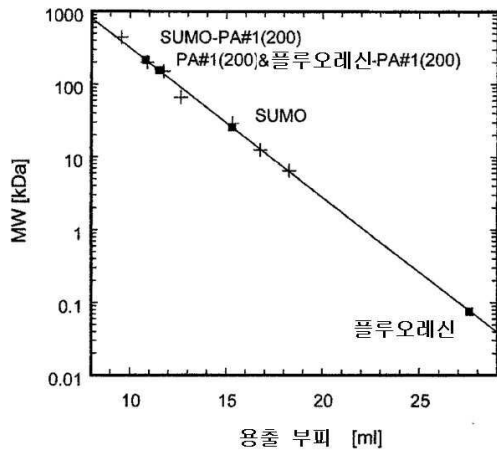


도면13h

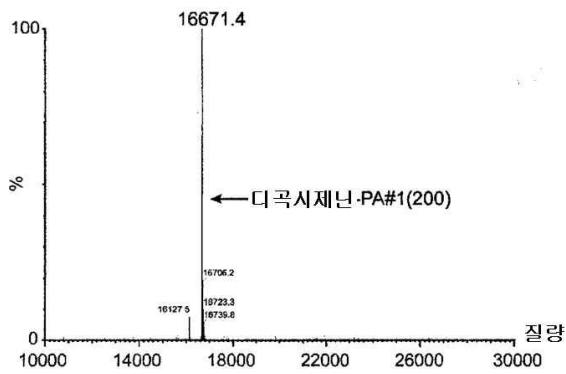


도면131

L

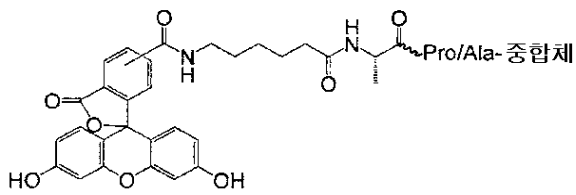


M



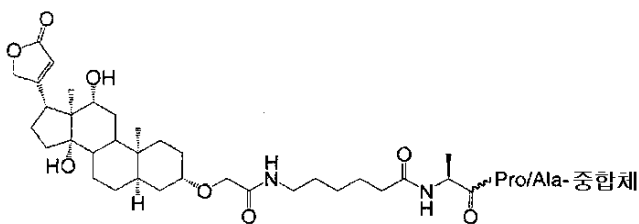
도면14a

A



도면14b

B



서열목록



<220><223> amino acid sequence of PA#3

<400>

> 3

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro

1 5 10 15

Ala Ala Ala Pro

20

<210> 4

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of PA#4

<400> 4

Ala Ala Pro Ala Ala Pro Ala Ala Pro Ala Ala Pro Ala Ala Pro Ala

1 5 10 15

Ala Pro Ala Ala Pro Ala Ala Pro

20

<210> 5

<

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of PA#5

<400> 5

Ala Pro Ala Ala Ala Pro Ala Pro Ala Ala Ala Pro Ala Pro Ala Ala

1 5 10 15

Ala Pro Ala Pro Ala Ala Ala Pro

20

<210> 6

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of PA#6

<400> 6

Ala Ala Ala Pro Ala Ala Pro Ala Ala Pro Pro Ala Ala Ala Ala Pro

1                    5                    10                    15

Ala Ala Pro Ala Ala Pro Pro Ala

20

<210> 7

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of a circular permuted version of SEQ ID

NO: 1

<400> 7

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala

1                    5                    10                    15

Ala Pro Ala Ala

20

<210> 8

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of a circular permuted version of SEQ ID

NO: 1

<400> 8

Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala

1                    5                    10                    15

Pro Ala Ala Ala

20

<210> 9

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of a circular permuted version of SEQ ID

NO: 1

<400> 9

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro



1 5 10 15

Ala Ala Ala Pro

20

<210> 10

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of a circular permuted version of SEQ ID

NO: 1

<400> 10

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala

1 5 10 15

Ala Ala Pro Ala

20

<210> 11

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of a circular permuted version of SEQ ID

NO: 1

<400> 11

Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala

1 5 10 15

Ala Pro Ala Ala

20

<210> 12

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of a circular permuted version of SEQ ID

NO: 1

<400> 12

Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala





ggcagcagga gcagccggag ctggtgcagc aggtgctgct ggagcagtg cagctggagc 60

60

<210> 19

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence stretch (upper/coding strand) around the C-terminus of the immunoglobulin light chain of an antibody Fab fragment as encoded on pASK88-Fab-2xSapI

<400> 19

aagagcttca accgcgaga gtgctcttct gcctgaagag cttaagctt 49

<210> 20

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence stretch (lower/non-coding strand) around the C-terminus of the immunoglobulin light chain of an antibody Fab fragment as encoded on pASK88-Fab-2xSapI

<400> 20

aagcttaagc tcttcaggca gaagagcact ctccgcggtt gaagctctt 49

<210> 21

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of the C-terminus of the light chain of the Fab fragment as encoded on pASK88-Fab-2xSapI

<400> 21

Lys Ser Phe Asn Arg Gly Glu Cys Ser Ser Ala

1 5 10

<210> 22

<211> 4610

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of pASK88-Fab-2xSapI

<400> 22

```

accgcacacc atcgaatggc cagatgatta attcctaatt ttgttgaca ctctatcatt      60
gatagagtta ttttaccact ccctatcagt gatagagaaa agtgaaatga atagttcgac      120
aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtggcac      180
tgcttggttt cgctaccgta gcgcaggccg aagttaaact gcaggaatcc ggtggtggtc      240

tggttcagcc aggtggttcc ctgcggctct cgtgtgctgc ttccggttc aacatcaaag      300
acacctacat cactgggtt cgtcagctc cgggtaaagg cctggaatgg gttgctcgta      360
tctaccggac caacggttac accaggtatg ccgattcagt taaaggctgt ttcaccatct      420
cggccgacac ttccaaaac accgcttacc tccagatgaa ctcctgcgt gctgaagaca      480
cagctgttta ttattgctcc cgttgggtg gtgacggtt ctacgctatg gactactggg      540
gtcaggttac cctggtcacc gtctctcag cctccacca gggcccatcg gtcttcccc      600
tggcacctc ctccaagagc acctctggg gcacagcggc cctgggctgc ctggtcaagg      660

actacttccc cgaaccggtg acggtgtcgt ggaactcagg cgcctgacc agcggcgtgc      720
acaccttccc ggctgtccta cagtctcag gactctact cctcagcagc gtggtgactg      780
tgccctccag cagcttgggc acccagacct acatctgcaa cgttaatcac aaaccagca      840
acaccaaggt cgacaagaaa gttgagccca aatcttgcca tcaccacat caccattaat      900
aaccatggag aaaataaagt gaaacaaagc actatgac tggcactctt accgttactg      960
tttaccctg tgacaaaagc cgacatcgag ctcaccaat ccccgctctc cctgtccget      1020
tccgttggcg accgtgttac catcacgtgt agggcctcgc aagacgtaaa caccgccgta      1080

gcgtggtatc agcagaaacc cgggaaagct ccgaaactgc tgatctatag cgcttcttc      1140
ctgtattccg gagtcccgag caggttcagt ggttccggt cgggtaccga cttaccctg      1200
acgatatcct cctccagcc ggaagacttc gctacctact actgtcaaca gcaactaccc      1260
accccgccga cttcggtea ggttaccaaa ctcgatca aacggactgt ggctgcacca      1320
tctgtcttca tcttccgccc atctgatgag cagttgaaat ctggaactgc ctctgttg      1380
tgcctgctga ataacttcta tcccagagag gccaaagtac agtggaaggt ggataacgcc      1440
ctccaatcgg gtaactccca ggagagtgtc acagagcagg acagcaagga cagcacctac      1500

agcctcagca gcacctgac gctgagcaaa gcagactacg agaaacaca agtctacgcc      1560
tgcgaagtca cccatcaggg cctgagttcg cccgtcaca agagcttcaa ccgcggagag      1620
tgctcttctg cctgaagagc ttaagcttga cctgtgaagt gaaaaatggc gcacattgtg      1680

```

cgacattttt ttgtctgcc gttaccgct actgcgtcac ggatctccac ggcacctgta 1740  
 gcggcgcat aagcgcggcg ggtgtgggg ttacgcgcag cgtgaccgct acacttgcca 1800  
 ggcacctagc gcccgtcct ttcgctttct tcccttctt tctcgccacg ttcgcccgtt 1860  
 ttccccgtca agctctaaat cgggggctcc ctttagggtt ccgatttagt gctttacggc 1920

acctcgacc caaaaaactt gattagggtg atggttcacg tagtgggcca tcgccctgat 1980  
 agacggtttt tcgccctttg acgttggagt ccacgttctt taatagtga ctcttgttcc 2040  
 aaactggaac aacactcaac cctatctcgg tctattcttt tgatttataa gggattttgc 2100  
 cgatttcggc ctattggtta aaaaatgagc tgatttaaca aaaatttaac gcgaatttta 2160  
 acaaaatatt aacgtttaca atttcagtg gcacttttcg gggaaatgtg cgcggaacc 2220  
 ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga caataaccct 2280  
 gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtcg 2340

ccctatttcc cttttttgcg gcattttgcc ttctgtttt tgetcaccca gaaacgtgg 2400  
 tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc 2460  
 tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca 2520  
 cttttaaagt tctgctatgt ggcgcggtat tateccgtat tgacgccggg caagagcaac 2580  
 tcggtcggcg catacactat tctcagaatg acttggttga gtactacca gtcacagaaa 2640  
 agcatcttac ggatggcatg acagtaagag aattatgcag tgetgccata accatgagtg 2700  
 ataacactgc ggccaactta ctctgacaa cgatcggagg accgaaggag ctaaccgctt 2760

ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg 2820  
 aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc 2880  
 gcaactatt aactggcgaa ctacttactc tagcttcccg gcaacaattg atagactgga 2940  
 tggaggcggg taaagtgca ggaccacttc tgcgctcggc ccttccggct ggctggttta 3000  
 ttgctgataa atctggagcc ggtgagcgtg gctctcggg tatcattgca gcactggggc 3060  
 cagatggtaa gcctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg 3120  
 atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat 3180

taatgatgic tcgtttagat aaaagtaaag tgattaacag cgcattagag ctgcttaatg 3240  
 aggtcggaat cgaaggttta acaaccgta aactcgcca gaagctaggt gtagagcagc 3300  
 ctacattgta ttggcatgta aaaaataagc gggctttgct cgacgctta gccattgaga 3360  
 tgttagatag gcaccatact cacttttgcc ctttagaagg ggaagctgg caagattttt 3420  
 tacgtaataa cgctaaaagt tttagatgtg ctttactaag tcatcgcgat ggagcaaaag 3480  
 tacatttagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct 3540



ttttatgcca acaaggtttt tcaactagaga atgcattata tgcactcagc gcagtggggc 3600

atfttacttt aggttgcgta ttggaagatc aagagcatca agtcgctaaa gaagaaaggg 3660

aaacacctac tactgatagt atgccgcatc tattacgaca agctatcgaa ttatttgatc 3720

accaaggtgc agagccagcc ttcttattcg gccttgaatt gatcatatgc ggattagaaa 3780

aacaacttaa atgtgaaagt gggctttaa agcagcataa cctttttccg tgatggtaac 3840

ttcaactagt taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc 3900

cttaactgta gttttcgttc cactgagcgt cagaccccg agaaaagatc aaaggatctt 3960

cttgagatcc tttttttctg cgcgtaatct gctgcttgca acaaaaaaaaa ccaccgctac 4020

cagcgggtgt ttgtttgccc gatcaagagc taccaactct ttttccgaag gtaactggct 4080

tcagcagagc gcagatacca aactactgtc ttctagtgtg gccgtagtta ggccaccact 4140

tcaagaactc ttagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg 4200

ctgccagtgg cgataagtcg tgtcttaccg ggttggactc aagacgatag ttaccggata 4260

aggcgcagcg gtcgggctga acgggggggt cgtgcacaca gccagcttg gagcgaacga 4320

cctacaccga actgagatcc ctacagcgtg agctatgaga aagcggccag cttcccgaag 4380

ggagaaaaggc ggacaggtat ccggttaagc gcagggtcgg aacaggagag cgcacgaggg 4440

agcttccagg gggaaacgcc tggatcttt atagtcctgt cgggtttcgc cacctctgac 4500

ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca 4560

acgcggcctt tttacggttc ctggccttt gctggccttt tgctcacatg 4610

<210> 23

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence stretch (upper/coding strand) encoding amino acid sequence of the C-terminus of the Fab light chain after insertion of one PA#1(20) polymer

<400

> 23

gagtgtcttt ctgccgtcc agctgcacct gctccagcag cacctgctgc accagctccg 60

gctgctcctg ctgcctgaag agctt 85

<210> 24

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence (lower/non-coding strand) for an amino acid stretch of the C-terminus of an Fab light chain after insertion of one PA#1(20) polymer

<400> 24

aagctcttca ggcagcagga gcagccggag ctggtgcagc aggtgctgct ggagcaggtg 60

cagctggagc ggcagaagag cactc 85

<210> 25

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence stretch of the C-terminus of an Fab light chain after insertion of one PA#1(20) polymer

<400> 25

Glu Cys Ser Ser Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala

1 5 10 15

Ala Pro Ala Pro Ala Ala Pro Ala Ala

20 25

<210> 26

<211> 229

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of a Fab heavy chain as encoded on pFab-PA#1(200)

<400> 26

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

1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Asn Ile Lys Asp Thr

20 25 30

Tyr Ile His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val

35 40 45

Ala Arg Ile Tyr Pro Thr Asn Gly Tyr Thr Arg Tyr Ala Asp Ser Val



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

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala

275 280 285

Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro

290 295 300

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala

305 310 315 320

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala

325 330 335

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro

340 345 350

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala

355 360 365

Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro

370 375 380

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala

385 390 395 400

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala

405 410 415

Ala

<210> 28

<211> 5210

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of pFab-PA#1(200)

<400> 28

accgcacacc atcgaatggc cagatgatta attcctaatt tttgttgaca ctctatcatt 60  
gatagagtta ttttaccact ccctatcagt gatagagaaa agtgaaatga atagttcgac 120  
aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtggcac 180  
tggctggttt cgctaccgta ggcgaggcgc aagttaaact gcaggaatcc ggtggtggtc 240  
tggttcagcc aggtggttcc ctgcggtctc cgtgtgctgc ttccggtttc aacatcaaag 300  
acacctacat cactgggtt cgtcaggctc cgggtaaagg cctggaatgg gttgctcgta 360

tctaccgcac caacggttac accaggtatg ccgattcagt taaaggtcgt ttcaccatct 420

cggccgcacac ttccaaaaac accgcttacc tccagatgaa ctccctgcgt gctgaagaca 480

cagctgttta ttattgctcc cgttgggggtg gtgacggttt ctacgctatg gactactggg 540

gtcaggggtac cctgggtcacc gtctcctcag cctecaccaa gggcccatcg gtcttccccc 600

tggcacccctc ctccaagagc acctctgggg gcacagcggc cctgggctgc ctggccaagg 660

actacttccc cgaaccggtg acggtgtcgt ggaactcagg cgcctgacc agcggcgtgc 720

acaccttccc ggctgtccta cagtctcag gactctactc cctcagcage gtggtgactg 780

tgccctccag cagcttgggc acccagacct acatctgcaa cgtaatcac aaaccagca 840

acaccaaggt cgacaagaaa gttgagccca aatcttgcca tcaccacat caccattaat 900

aaccatggag aaaataaagt gaaacaaagc actattgcac tggcactctt accgttactg 960

tttaccctg tgacaaaage cgacatcgag ctccaccaat ccccgctctc cctgtccgct 1020

tccgttggcg accgtgttac catcacgtgt agggcctcgc aagacgtaaa caccgccgta 1080

gcgtggatc agcagaaacc cgggaaagct ccgaaactgc tgatctatag cgcttcttc 1140

ctgtattccg gagtcccgag caggttcagt ggttccggt cgggtaccga cttaccctg 1200

acgatatcct cctccagec ggaagacttc gctacctact actgteaaca gcactacacc 1260

accccgccga ccttcggta ggttacaaa ctcgatca aacggactgt ggctgcacca 1320

tctgtcttca tcttccgcc atctgatgag cagttgaaat ctggaactgc ctctgttgtg 1380

tgcctgctga ataacttcta tcccagagag gccaaagtac agtggaaggt ggataacgcc 1440

ctccaatcgg gtaactccca ggagagtgtc acagagcagg acagcaagga cagcacctac 1500

agcctcagca gcacctgac gctgagcaaa gcagactacg agaaacacaa agtctacgcc 1560

tgcgaagtca cccatcaggg cctgagttcg cccgtcacia agagcttcaa ccgaggagag 1620

tgtcttctg ccgctccage tgcacctgct ccagcagcac ctgctgcacc agctccggct 1680

gctcctgctg ccgctccage tgcacctgct ccagcagcac ctgctgcacc agctccggct 1740

gctcctgctg ccgctccage tgcacctgct ccagcagcac ctgctgcacc agctccggct 1800

gctcctgctg ccgctccage tgcacctgct ccagcagcac ctgctgcacc agctccggct 1860

gctcctgctg ccgctccage tgcacctgct ccagcagcac ctgctgcacc agctccggct 1920

gctcctgctg ccgctccage tgcacctgct ccagcagcac ctgctgcacc agctccggct 1980

gctcctgctg ccgctccage tgcacctgct ccagcagcac ctgctgcacc agctccggct 2040

gctcctgctg ccgctccage tgcacctgct ccagcagcac ctgctgcacc agctccggct 2100



gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggt 2160  
gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggt 2220  
gctcctgctg cctgaagagc ttaagcttga cctgtgaagt gaaaaatggc gcacattgtg 2280  
cgacattttt ttgtctgcc gtttaccgt actgcgtcac ggatctccac gcgcctgta 2340  
gcggcgcatt aagcgcggcg ggtgtggagg ttacgcgcag cgtgaccgct acaactgcca 2400  
gcgcctagc gcccgtcct ttcgctttct tcccttctt tctcgccacg ttcgcccgt 2460  
ttcccgtca agctctaaat cgggggtccc ctttagggtt ccgatttagt gctttacggc 2520  
  
acctgaccc caaaaaactt gattagggtg atggttcacg tagtgggcca tcgccctgat 2580  
agacggtttt tcgccctttg acgttggagt ccacgttctt taatagtga ctcttgttcc 2640  
aaactggaac aacactcaac cctatctcgg tctattcttt tgatttataa gggattttgc 2700  
cgatttcggc ctattggtta aaaaatgagc tgatttaaca aaaatttaac gcgaattta 2760  
acaaaatatt aacgtttaca atttcaggtg gcacttttcg gggaaatgtg cgcggaacc 2820  
ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga caataaccct 2880  
gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttcctgtctg 2940  
  
cccttattcc cttttttggc gcattttgcc ttctgtttt tgetcaccca gaaacgtgg 3000  
tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc 3060  
tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca 3120  
cttttaaagt tctgctatgt ggcgcggtat tateccgtat tgacgccggg caagagcaac 3180  
tcggtcgccg catacactat tctcagaatg acttggttga gtactacca gtcacagaaa 3240  
agcatcttac ggatggcatg acagtaagag aattatgcag tctgcccata accatgagtg 3300  
ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag ctaaccgctt 3360  
  
ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg 3420  
aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc 3480  
gcaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaattg atagactgga 3540  
tggaggcgga taaagttgca ggaccacttc tgcgtcggc ccttccggtt ggctggttta 3600  
ttgtgataa atctggagcc ggtgagcgtg gctctcggg tatcattgca gcaactgggc 3660  
cagatggtaa gcctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg 3720  
atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat 3780  
  
taatgatgc tcgtttagat aaaagtaaag tgattaacag cgcattagag ctgcttaatg 3840  
aggtcggaat cgaaggttta acaaccgta aactcgcca gaagctaggt gtagagcagc 3900  
ctacattgta ttggcatgta aaaaataagc gggctttgct cgacgcctta gccattgaga 3960

tgttagatag gcaccatact cacttttgcc ctttagaagg ggaaagctgg caagatTTTT 4020  
 tacgtaataa cgctaaaagt ttagatgtg ctttactaag tcatcgcgat ggagcaaaag 4080  
 tacatttagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct 4140  
 ttttatgcca acaaggtttt tctactagaga atgcattata tgcactcagc gcagtggggc 4200  
  
 attttacttt aggttgcgta ttggaagatc aagagcatca agtcgctaaa gaagaaaggg 4260  
 aaacacctac tactgatagt atgccgcatc tattacgaca agctatcgaa ttatttgatc 4320  
 accaaggTgc agagccagcc ttcttattcg gccttgaatt gatcatatgc ggattagaaa 4380  
 aacaacttaa atgtgaaagt gggTcttaa agcagcataa ctttttccg tgatggtaac 4440  
 ttcactagtT taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc 4500  
 cttaacgtga gttttcgTtc cactgagcgt cagaccccgt agaaaagatc aaaggatctt 4560  
 cttgagatcc ttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac 4620  
  
 cagcggTggt ttgtttgceg gatcaagagc taccaactct ttttccgaag gtaactgget 4680  
 tcagcagagc gcagatacca aatactgtcc ttctagtgtg gccgtagtta ggccaccact 4740  
 tcaagaactc tgtagaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg 4800  
 ctgccagTgg cgataagteg tgtcttaccg ggttggactc aagacgatag ttaccggata 4860  
 aggcgcagcg gtcgggctga acggggggTt cgtgcacaca gccagcttg gagegaacga 4920  
 cctacaccga actgagatac ctacagcgtg agctatgaga aagcggccag cttcccgaag 4980  
 ggagaaagcg ggacaggtat ccggtaaagc gcagggtcgg aacaggagag cgcacgaggg 5040  
  
 agcttccagg gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc cacctctgac 5100  
 ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca 5160  
 acgcggcctt tttacggTtc ctggcctttt gctggccttt tgctcacatg 5210  
  
 <210> 29  
 <211> 57  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> nucleic acid sequence(upper/coding strand) encoding amino acid  
 sequence of the N-terminus of INFa2b and Strep-tag II (only the  
 last two amino acids)  
 <400> 29  
 gaaaaagcgc ccagctcttc tgctgtgat ctgcctcaaa cccacagcct gggtagc 57

<210> 30

<211> 57  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> nucleic acid sequence (lower/non-coding strand) encoding amino acid sequence of the N-terminus of INFa2b and Strep-tag II (only the last two amino acids)  
 <400> 30  
 gctaccagg ctgtgggttt gaggcagatc acaggcagaa gagctggcgc ctttttc 57  
 <210> 31  
 <211> 19  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> amino acid sequence of the C-terminus of Strep-tag II and the N-terminus of INFa2b  
 <400> 31  
 Glu Lys Gly Ala Ser Ser Ser Ala Cys Asp Leu Pro Gln Thr His Ser  
 1 5 10 15  
 Leu Gly Ser  
 <210> 32  
 <211> 3721  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> nucleic acid sequence of pASK-IFNa2b  
 <400> 32  
 ccacgaatg gccagatgat taattcctaa tttttgttga cactctatca ttgatagagt 60  
 tattttacca ctccctatca gtgatagaga aaagtgaaat gaatagttcg acaaaaatct 120  
 agataacgag ggcaaaaaat gaaaaagaca gctatcgca ttgcagtggc actggctggt 180  
 ttcgctaccg tagcgcagge cgetagctgg agccaccege agttcgaaaa aggcgccagc 240  
 tcttctgctt gtgatctgcc tcaaacccac agcctgggta gcaggaggac cttgatgctc 300  
 ctggcacaga tgaggagaat ctctcttttc tctgcttga aggacagaca tgactttgga 360  
 tttccccagg aggagtttgg caaccagttc caaaaggctg aaaccatccc tgtctccat 420  
 gagatgatcc agcagatctt caatctcttc agcacaaagg actcatctgc tgcttgggat 480

gagacctcc tagacaaatt ctacactgaa ctctaccage agctgaatga cctggaagcc 540  
 tgtgtgatac aggggtggg ggtgacagag actcccctga tgaaggagga ctccattctg 600

gctgtgagga aatacttcca aagaatcact ctctatctga aagagaagaa atacagccct 660  
 tgtgcctggg aggttctcag agcagaaatc atgagatctt tttctttgtc aacaaacttg 720  
 caagaaagt taagaagtaa ggaataagct tgacctgtga agtgaaaaat ggcgcacatt 780  
 gtgacacatt tttttgtct gccgtttacc gctactgctc cacggatctc cacgcgcct 840  
 gtagcggcgc attaacgcgg gcgggtgtgg tggttacgcg cagcgtgacc gctacacttg 900  
 ccagcgcct agcgcgccct ctttccgctt tcttccttc ctttctgcc acgttcgccg 960  
 gctttccccg tcaagetcta aatcgggggc tcccttagg gttccgattt agtgccttac 1020

ggcacctcga ccccaaaaaa ctigattagg gtgatggttc acgtagtggg ccatcgccct 1080  
 gatagacggt ttttcgccct ttgacttgg agtccactt ctttaatagt ggactcttgt 1140  
 tccaaactgg aacaacactc aacctatct cggctatctc ttttgattta taagggattt 1200  
 tgccgatttc ggctatttg ttaaaaaatg agctgattta acaaaaattt aacgcgaatt 1260  
 ttaacaaat attaacgctt acaatttcag gtggcacttt tcggggaaat gtgdcggaa 1320  
 cccctatttg tttattttc taaatacatt caaatatgta tccgctcatg agacaataac 1380  
 cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa catttccgtg 1440

tcgcccttat tccctttttt gcggcatttt gccttctgt ttttctcac ccagaaacgc 1500  
 tggigaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac atcgaactgg 1560  
 atctcaacag cggtaagatc ctgagagtt ttcgccccga agaactttt ccaatgatga 1620  
 gcacttttaa agttctgcta tgtggcggg tattatcccg tattgacgcc gggcaagagc 1680  
 aactcggctc ccgatacac tattctcaga atgacttggg tgagtactca ccagtcacag 1740  
 aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgtgcc ataaccatga 1800  
 gtgataacac tgcggccaac ttacttctga caacgatcgg aggaccgaag gagctaaccg 1860

cttttttga caacatgggg gatcatgtaa ctgccttga tcgttgggaa ccggagctga 1920  
 atgaagccat accaaacgac gacggtgaca ccacgatgcc tftagcaatg gcaacaact 1980  
 tgcgcaact attaactggc gaactactta ctctagcttc ccggcaacaa ttgatagact 2040  
 ggatggaggc ggataaagt gcaggaccac ttctgcctc gcccttccg gctggctggt 2100  
 ttattgctga taaatctgga gccggtgagc gtggtctctc cggtatcatt gcagcactgg 2160  
 ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt caggcaacta 2220

tggatgaacg aatagacag atcgctgaga taggtgcctc actgattaag cattggtagg 2280

aattaatgat gtctcgttta gataaaagta aagtgattaa cagcgatta gagctgctta 2340

atgaggtcgg aatcgaaggt ttaacaaccc gtaaacctgc ccagaagcta ggtgtagagc 2400

agcctacatt gtattggcat gtaaaaaata agcgggcttt gctcgacgcc ttagccattg 2460

agatgttaga taggcacat actcactttt gccctttaga aggggaaagc tggcaagatt 2520

ttttacgtaa taacgctaaa agttttagat gtgctttact aagtcacgc gatggagcaa 2580

aagtacattt aggtacacgg cctacagaaa aacagtatga aactctcgaa aatcaattag 2640

cctttttatg ccaacaaggt tttcactag agaatgcatt atatgcactc agcgcagtgg 2700

ggcattttac tttaggttgc gtattggaag atcaagagca tcaagtcgct aaagaagaaa 2760

gggaaacacc tactactgat agtatgccgc cattattacg acaagctatc gaattattg 2820

atcaccaagg tgcagagcca gccttcttat tcggccttga attgatcata tgcgattag 2880

aaaaacaact taaatgtgaa agtgggtctt aaaagcagca taacctttt ccgtgatggt 2940

aacttacta gtttaaaagg atctaggtga agatcctttt tgataatctc atgaccaaaa 3000

tccttaacg tgagttttc ttccactgag cgtcagacc cgtagaaaag atcaaaggat 3060

cttcttgaga tcctttttt ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc 3120

taccagcggg ggtttgtttg ccgatcaag agctaccaac tcttttccg aagtaactg 3180

gcttcagcag agcgcagata ccaaaactg tccttctagt gtagccgtag ttagccacc 3240

acttcaagaa ctctgtagca ccgctacat acctcgctct gctaatcctg ttaccagtgg 3300

ctgctgccag tggcgataag tcgtgtctta ccgggttga ctcaagacga tagttaccgg 3360

ataagcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc ttggagcgaa 3420

cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgc acgcttcccg 3480

aagggagaaa ggcggacagg tatccggtaa gcggcagggt cggaacagga gagcgcacga 3540

gggagcttcc agggggaac gcctggtatc tttatagtcc tgcgggttt cgccacctct 3600

gacttgagcg tcgatttttg tgatgctcgt cagggggcg gagcctatgg aaaaacgcca 3660

gcaacgcggc cttttaccg ttctggcct tttgctggcc tttgctcac atgaccgcac 3720

a 3721

<210> 33

<211> 90

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence stretch (upper/coding strand) encoding the

C-terminus of Strep-tag II and the N-terminus of IFNa2b after

insertion of one PA#1 polymer sequence cassette

<400> 33  
 gaaaaaggcg ccagctcttc tgccgctcca gctgcacctg ctcacagcage acctgctgca 60  
 ccagctccgg ctgctcctgc tgctgtgat 90  
 <210> 34  
 <211> 90  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> nucleic acid sequence stretch (lower/non-coding strand) of the  
 C-terminus of Strep-tag II and the N-terminus of IFNa2b after  
 insertion of one PA#1 polymer sequence cassette

<400> 34  
 atcacaggca gcaggagcag ccggagctgg tgcagcaggt gctgctggag caggtgcagc 60  
 tggagcggca gaagagctgg cgcctttttc 90  
 <210> 35  
 <211> 30  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> amino acid sequence stretch of the C-terminus of Strep-tag II and  
 the N-terminus of IFNa2b ater fusion with one PA#1 polymer  
 cassette

<400> 35  
 Glu Lys Gly Ala Ser Ser Ser Ala Ala Pro Ala Ala Pro Ala Pro Ala  
 1 5 10 15

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Cys Asp  
 20 25 30

<210> 36  
 <211> 381  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> amino acid sequence of IFNa2b and Strep-tag II fused with the



PA#1(200) polymer as encoded on pPA#1(200)-IFNa2b

<400> 36  
Ala Ser Trp Ser His Pro Gln Phe Glu Lys Gly Ala Ser Ser Ser Ala  
1 5 10 15  
Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala  
20 25 30  
Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala  
35 40 45  
Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala  
50 55 60  
Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala  
65 70 75 80  
Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala  
85 90 95  
Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala  
100 105 110  
Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala  
115 120 125  
Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala  
130 135 140  
Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala  
145 150 155 160  
Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala  
165 170 175  
Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala  
180 185 190  
Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala  
195 200 205  
Pro Ala Pro Ala Ala Pro Ala Ala Cys Asp Leu Pro Gln Thr His Ser  
210 215 220  
Leu Gly Ser Arg Arg Thr Leu Met Leu Leu Ala Gln Met Arg Arg Ile



cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 420  
cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 480  
cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 540  
cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 600  
cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 660  
cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 720

cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 780  
cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 840  
cctgctgcct gtgatctgcc tcaaacccac agcctgggta gcaggaggac cttgatgctc 900  
ctggcacaga ttaggagaat ctctcttttc tctctcttga aggacagaca tgactttgga 960  
tttccccagg aggagtttgg caaccagttc caaaaggctg aaacctccc tgtcctccat 1020  
gagatgatcc agcagatctt caatctcttc agcacaaagg actcatctgc tgcttgggat 1080  
gagaccctcc tagacaaatt ctacactgaa ctctaccagc agctgaatga cctggaagcc 1140

tgtgtgatac aggggggtggg ggtgacagag actcccctga tgaaggagga ctccattctg 1200  
gctgtgagga aatacttcca aagaatcact ctctatctga aagagaagaa atacagccct 1260  
tgtgcctggg aggttgtcag agcagaaatc atgagatctt tttctttgtc aacaaacttg 1320  
caagaaagt taagaagtaa ggaataagct tgacctgtga agtgaanaat ggcgcacatt 1380  
gtgcgacatt ttttttgtct gccgtttacc gctactgcgt cacggatctc cacgcgcct 1440  
gtagcggcgc attaagcgcg gcgggtgtgg tggttacgcg cagcgtgacc gctacacttg 1500  
ccagcgcctc agcgcgccct cctttcgtt tcttcccttc ctttctcgcc acgttcgccc 1560

gctttccccg tcaagctcta aatcgggggc tccctttagg gttccgattt agtgctttac 1620  
ggcacctcga cccccaaaaa cttgattagg gtgatggttc acgtagtggg ccatcgccct 1680  
gatagacggg ttttcgccct ttgacgttgg agtecacgtt ctttaatagt ggactcttgt 1740  
tccaaactgg aacaacactc aaccctatct cggctctatc ttttgattta taaggattt 1800  
tgccgatttc ggccatttgg ttaaaaaatg agctgattta acaaaaattt aacgcgaatt 1860  
ttaacaaaat attaacgctt acaatttcag gtggcacttt tcggggaaat gtgcgcgaa 1920  
cccctatttg tttatttttc taaatacatt caaatatgta tccgctcatg agacaataac 1980

cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa catttccgtg 2040  
tcgcccttat tccctttttt gcggcatttt gccttctgt ttttgctcac ccagaaacgc 2100  
tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac atcgaactgg 2160  
atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt ccaatgatga 2220

gcacttttaa agttctgcta tgtggcgcgg tattatcccg tattgacgcc gggcaagagc 2280  
aactcggctg ccgcatacac tattctcaga atgacttggg tgagtactca ccagtcacag 2340  
aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgtgcc ataaccatga 2400

gtgataacac tgcggccaac ttacttctga caacgatcgg aggaccgaag gagctaacgg 2460  
cttttttgca caacatgggg gatcatgtaa ctgccttga tcgttgggaa ccggagctga 2520  
atgaagccat accaaacgac gagcgtgaca ccacgatgcc tntagcaatg gcaacaacgt 2580  
tgcgcaaact attaactggc gaactactta ctctagcttc ccggcaacaa ttgatagact 2640  
ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttccg gctggctggt 2700  
ttattgctga taaatctgga gccggtgagc gtggctctcg cggtatcatt gcagcactgg 2760  
ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt caggcaacta 2820

tggatgaacg aatagacag atcgtgaga taggtcctc actgattaag cattggtagg 2880  
aattaatgat gtctcgttta gataaaagta aagtattaa cagcgatta gagctgctta 2940  
atgaggtcgg aatcgaaggt ttaacaacc gtaaacctgc ccagaagcta ggtgtagagc 3000  
agcctacatt gtattggcat gtaaaaaata agcgggcttt gctcgacgcc ttagccattg 3060  
agatgttaga taggcacat actcactttt gccctttaga aggggaaagc tggcaagatt 3120  
ttttacgtaa taacgctaaa agttttagat gtgctttact aagtcacgc gatggagcaa 3180  
aagtacattt aggtacacgg cctacagaaa aacagtatga aactctcgaa aatcaattag 3240

cctttttatg ccaacaaggt tttcactag agaatgcatt atatgcactc agcgcagtgg 3300  
ggcattttac tttaggttgc gtattggaag atcaagagca tcaagtcgct aaagaagaaa 3360  
gggaaacacc tactactgat agtatgccgc cattattacg acaagctatc gaattatttg 3420  
atcaccaagg tgcagagcca gccttcttat tcggccttga attgatcata tgcggattag 3480  
aaaaacaact taaatgtgaa agtgggtctt aaaagcagca taaccttttt ccgtgatggt 3540  
aacttacta gtttaaaagg atctaggtga agatcctttt tgataatctc atgaccaaaa 3600  
tcccttaacg tgagttttcg ttccactgag cgtcagacc cgtagaaaag atcaaaggat 3660

cttcttgaga tccttttttt ctgcgcgtaa tctgtctgtt gcaaacaaaa aaaccaccgc 3720  
taccagcggg ggtttgtttg ccggatcaag agctaccaac tctttttccg aaggtaaactg 3780  
gcttcagcag agcgcagata ccaatactg tccttctagt gtagccgtag ttagccacc 3840  
acttaagaa ctctgtagea ccgectacat acctcgctct gctaatctg ttaccagtgg 3900  
ctgctgccag tggcgataag tcgtgtctta ccgggttggg ctcaagacga tagttaccgg 3960  
ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc ttggagcgaa 4020

cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc acgcttcccg 4080

aagggagaaa ggcggacagg tatccggtaa gcggcagggc cggaacagga gagcgcacga 4140

gggagcttcc agggggaaac gcctggtatc tttatagtcc tgcgggttt cgccacctct 4200

gacttgagcg tcgatttttg tgatgctcgt cagggggcg gagcctatgg aaaaacgcca 4260

gcaacgcggc ctttttacgg ttcttgacct tttgctggcc tttgctcac atgacccgac 4320

a 4321

<210> 38

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence stretch (upper/ coding strand) on

pASK75-His6-hGH encoding the amino acid sequence around the  
N-terminus of His6-hGH

<400> 38

gccgctagcc atcaccacca tcacatggc gccagctctt ctgccttccc aacc 54

<210> 39

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence stretch (lower/non-coding strand) on

pASK75-His6-hGH encoding the amino acid sequence around the  
N-terminus of hGH

<400> 39

ggttgggaag gcagaagagc tggcgccatg gtgatggtgg tgatggctag cggc 54

<210> 40

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence stretch of the N-terminus of His6-hGH as  
encoded on pASK75-His6-hGH

<400> 40

Ala Ala Ser His His His His His His Gly Ala Ser Ser Ser Ala Phe

1                    5                    10                    15

Pro Thr

<210>    41

<211>    3793

<212>    DNA

<213>    Artificial Sequence

<220><223>    nucleic acid sequence of pASK75-His6-hGH

<400>    41

acccgacacc atcgaatggc cagatgatta attcctaatt ttgttgaca ctctatcatt                    60

gatagagtta ttttaccact ccctatcagt gatagagaaa agtgaaatga atagttcgac                    120

aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtggcac                    180

tggttggttt cgctaccgta ggcgaggccg ctagccatca ccaccatcac catggcgcca                    240

gctctttctc cttcccaacc attcccttat ccaggctttt tgacaacgct atgctccgcg                    300

cccatcgtct gcaccagctg gcctttgaca cctaccagga gtttgaagaa gcctatatcc                    360

caaaggaaca gaagtattca ttctgcaga acccccagac ctccctctgt ttctcagagt                    420

ctattccgac accctccaac agggaggaaa cacaacagaa atccaaccta gagctgctcc                    480

gcctcctcct gctgctcacc cagtcgtggc tggagcccgt gcagttcttc aggagtgtct                    540

tcgccaacag cctggtgtac ggcgcctctg acagcaacgt ctatgacctc ctaaaggacc                    600

tagaggaagg catccaaacg ctgatgggga ggctggaaga tggcagcccc cggactgggc                    660

agatcttcaa gcagacctac agcaagtctg acacaaactc acacaacgat gacgcactac                    720

tcaagaacta cgggctgctc tactgcttca ggaaggacat ggacaaggtc gagacattcc                    780

tgcgcatcgt gcagtgccg tctgtggagg gcagctgtgg cttctaagct tgacctgtga                    840

agtgaaaaat ggcgcacatt gtgcgacatt tttttgtct gccgtttacc gctactgcgt                    900

cacggatctc cacgcgccct gtagcggcgc attaaagcgc gcgggtgtgg tggttacgcg                    960

cagcgtgacc gctacacttg ccagcgccct agcgcgccgt ctttctgctt tcttcccttc                    1020

ctttctcgcc acgttcgccg gctttcccg tcaagctcta aatcgggggc tccttttagg                    1080

gttccgattt agtgccttac ggcacctcga cccccaaaaa cttgattagg gtgatggttc                    1140

acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg agtccacgtt                    1200

ctttaatagt ggactcttgt tccaaactgg aacaacactc aacctatct cggctctatc                    1260

ttttgattta taagggattt tgccgatttc gcctattgg ttaaaaaatg agctgattta                    1320

acaaaaatTT aacgcgaatt ttaacaaaat attaacgttt acaatttcag gtggcacttt 1380  
 tcggggaaat gtgcgcggaa cccctatttg tttatTTTT taaatacatt caaatatgta 1440  
 tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat 1500  
 gagtattcaa catttccgtg tcgcccttat tccctTTTT gcggcatttt gccttctctg 1560  
 ttttgctcac ccagaaacgc tggTgaaagt aaaagatgct gaagatcagt tgggtgcacg 1620  
 agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga 1680  
 agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgccg tattatcccg 1740  
  
 tattgacgcc gggcaagagc aactcggTcg ccgcatacac tattctcaga atgacttggT 1800  
 tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg 1860  
 cagtgtctcc ataaccatga gtgataaac tgcggccaac ttacttctga caacgatcgg 1920  
 aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa ctgccttga 1980  
 tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc 2040  
 tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta ctctagcttc 2100  
 ccggcaacaa ttgatagact ggaTggaggc ggataaagtt gcaggaccac ttctgcgctc 2160  
  
 ggcccttccg gctggetggT ttattgctga taaatctgga gccggTgagc gtggtctctg 2220  
 cggTatcatt gcagcactgg ggccagatgg taagccctcc cgtatcTtag ttatctacac 2280  
 gacggggagT caggcaacta tggatgaacg aaatagacag atcgtgaga taggtgcctc 2340  
 actgattaag cattgtagg aattaatgat gtctcgTTa gataaaagta aagtattaa 2400  
 cagcgatta gagctgctta atgaggtcgg aatcgaaggt ttaacaacc gTaaactcgc 2460  
 ccagaagcta ggtgtagagc agcctacatt gtattggcat gTaaaaaata agcgggcttt 2520  
 gctcgacgcc ttagccattg agatgttaga taggcacat actcactttt gccctttaga 2580  
  
 aggggaaagc tggcaagatt ttttacgtaa taacgctaaa agttttagat gtgctttact 2640  
 aagTatcgc gatggagcaa aagtacattt aggtacacgg cctacagaaa aacagtatga 2700  
 aactctgaa aatcaattag cttttttatg ccaacaaggt tttcactag agaatgatt 2760  
 atatgcactc agcgcagTgg ggcattttac tttaggtTgc gtattggaag atcaagagca 2820  
 tcaagtcgct aaagaagaaa gggaaacacc tactactgat agtatgccgc cattattacg 2880  
 acaagctatc gaattatttg atcaccaagg tgcagagcca gccttcttat tcggccttga 2940  
 attgatcata tgcggattag aaaaacaact taaatgtgaa agtgggtctt aaaagcagca 3000  
  
 taacctttt cgtgatggT aacttacta gtttaaaagg atctaggtga agatcctttt 3060  
 tgataatctc atgacaaaaa tccttaacg tgagttttcg ttccactgag cgTcagacc 3120  
 cgtagaaaag atcaaaggat cttcttgaga tcctttttt ctgcgcgtaa tctgctgctt 3180



gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg ccgatcaag agctaccaac 3240  
 tctttttccg aaggtaactg gcttcagcag agcgcagata ccaataactg tccttctagt 3300  
 gtagccgtag ttaggccacc acttcaagaa ctctgtagca ccgctacat acctcgtct 3360  
 gctaactctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta ccgggttgga 3420

ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac 3480  
 acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc gtgagctatg 3540  
 agaaagcgcc acgcttcccg aaggagagaa ggcggacagg tatccggtaa gcggcagggt 3600  
 cggaacagga gagcgcacga gggagcttc aggggggaaac gcctggatc tttatagtcc 3660  
 tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctctg caggggggcg 3720  
 gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg ttcttgccct tttgctggcc 3780  
 tttgctcac atg 3793

<210> 42

<211> 114

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence (upper/coding-strand) stretch encoding  
 amino acid sequence of the N-terminus of His6-hGH after insertion  
 of the PA#1(20) polymer

<400> 42

gccgctagcc atcaccacca tcaccatggc gccagctctt ctgccctcc agctgcacct 60  
 gctccagcag cacctgctgc accagctccg gctgctcctg ctgcttccc aacc 114

<210> 43

<211> 114

<212> DNA

<213> Artificial Sequence

<220

><223> nucleic acid sequence (lower/non-coding strand) encoding the  
 N-terminus of hGH after insertion of one PA#1 polymer sequence  
 cassette

<400> 43

ggttgggaag gcagcaggag cagccggagc tggtgcagca ggtgctgctg gagcaggctc 60  
 agctggagcg gcagaagagc tggcgccatg gtgatggtgg tgatggctag cggc 114

<210> 44

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of the N-terminus of His6-hGH after insertion  
of the PA#1(20) polymer

<400> 44

Ala Ala Ser His His His His His His Gly Ala Ser Ser Ser Ala Ala

1 5 10 15

Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala

20 25 30

Pro Ala Ala Phe Pro Thr

35

<210> 45

<211> 405

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of mature His6-PA#1(200)-hGH as encoded on  
pASK75-His6-PA#1(200)-hGH

<400

> 45

Ala Ser His His His His His His His Gly Ala Ser Ser Ser Ala Ala Pro

1 5 10 15

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro

20 25 30

Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala

35 40 45

Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro

50 55 60

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala

65 70 75 80

Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro

85 90 95

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro

100                      105                      110  
 Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala  
 115                      120                      125  
 Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro  
 130                      135                      140  
  
 Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala  
 145                      150                      155                      160  
 Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro  
 165                      170                      175  
 Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro  
 180                      185                      190  
 Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala  
 195                      200                      205  
 Pro Ala Ala Pro Ala Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe  
  
 210                      215                      220  
 Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp  
 225                      230                      235                      240  
 Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr  
 245                      250                      255  
 Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile  
 260                      265                      270  
 Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu  
 275                      280                      285  
  
 Leu Leu Arg Ile Ser Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val  
 290                      295                      300  
 Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser  
 305                      310                      315                      320  
 Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln  
 325                      330                      335  
 Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile  
 340                      345                      350

Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp

355 360 365

Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met

370 375 380

Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu

385 390 395 400

Gly Ser Cys Gly Phe

405

<210> 46

<211> 4393

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of pASK75-His6-PA#1(200)-hGH

<400> 46

gacccgacac catcgaatgg ccagatgatt aattcctaatt ttttgttgac actctatcat 60

tgatagagtt attttaccac tcctatcag tgatagagaa aagtgaatg aatagttcga 120

caaaaatcta gataacgagg gcaaaaaatg aaaaagacag ctatcgcat tgcagtggca 180

ctggctggtt tcgctaccgt agcgcaggcc gctagccatc accaccatca ccatggcgcc 240

agctcttctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 300

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 360

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 420

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 480

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 540

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 600

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 660

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 720

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 780

gctcctgctg ccgctccagc tgcacctgct ccagcagcac ctgctgcacc agctccggct 840

gctcctgctg ccttcccaac cattccctta tccaggcttt ttgacaacgc tatgctccgc 900

gcccatcgtc tgcaccagct ggcctttgac acctaccagg agtttgaaga agcctatc 960

ccaaaggaac agaagtattc attcctgcag aacccccaga cctccctctg tttctcagag 1020

tctattccga caccctccaa caggaggaggaa acacaacaga aatccaacct agagctgctc 1080  
 cgcatctccc tgctgctcat ccagtcgtgg ctggagcccg tgcagttcct caggagtgtc 1140  
 ttcgccaaca gcttgggtga cggcgcctct gacagcaacg tctatgacct cctaaaggac 1200  
 ctagagggaag gcatccaaac gctgatgggg aggctggaag atggcagccc cgggactggg 1260  
 cagatcttca agcagaccta cagcaagttc gacacaaact cacacaacga tgacgcacta 1320

ctcaagaact acgggctgct ctactgcttc aggaaggaca tggacaaggt cgagacattc 1380  
 ctgcgcatcg tgcagtgccg ctctgtggag ggcagctgtg gcttctaagc ttgacctgtg 1440  
 aagtgaaaaa tggcgcacat tgtgcgacat ttttttgtc tgccgtttac cgctactgcg 1500  
 tcacggatct ccacgcgccc tgtagcggcg cattaagcgc ggcgggtgtg gtggttacgc 1560  
 gcagcgtgac cgctacactt gccagcgcgc tagcgcgccg tcctttcgct ttcttcctt 1620  
 cctttctcgc cacgttcgcc ggctttcccc gtcaagctct aaatcggggg ctccttttag 1680  
 ggttccgatt tagtgcttta cggcacctcg accccaaaaa acttgattag ggtgatggtt 1740

cacgtagtgg gccatcgccc tgatagacgg tttttcgccc tttgacgttg gagtccacgt 1800  
 tctttaatag tggactcttg ttccaaactg gaacaacact caaccctatc tcggtctatt 1860  
 cttttgattt ataagggtt ttgccgattt cggcctattg gttaaaaaat gagctgattt 1920  
 aacaaaaatt taacgcgaat ttaacaaaa tattaacgtt tacaatttca ggtggcactt 1980  
 ttcggggaaa tgtgcgcgga acccctattt gtttatTTTT ctaatacat tcaaatatgt 2040  
 atccgctcat gagacaataa cctgataaa tgcttcaata atattgaaaa aggaagagta 2100  
 tgagtattca acatttccgt gtcgccctta ttccctttt tgcggcattt tgccttctg 2160

tttttctca ccagaaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac 2220  
 gagtgggtta catcgaactg gatctcaaca gcggtaagat ccttgagagt tttcccccg 2280  
 aagaacgttt tccaatgatg agcactttta aagtctgct atgtggcgcg gtattatccc 2340  
 gtattgacgc cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg 2400  
 ttgagtactc accagtcaaa gaaaagcatc ttacggatgg catgacagta agagaattat 2460  
 gcagtgctgc cataaccatg agtataaca ctgcggccaa cttacttctg acaacgatcg 2520  
 gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg 2580

atcgttggga accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc 2640  
 ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt 2700  
 cccggcaaca attgatagac tggatggagg cggataaagt tgcaggacca cttctgcgct 2760  
 cggcccttcc ggctggctgg tttattgctg ataaatctgg agccgggtgag cgtggtctc 2820  
 gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca 2880

cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct 2940  
cactgattaa gcattggtag gaattaatga tgtctcgttt agataaaagt aaagtgatta 3000

acagcgcatt agagctgctt aatgaggtcg gaatcgaagg ttaacaacc cgtaaacctg 3060  
cccagaagct aggtgtagag cagcctacat tgtattggca tgtaaaaaat aagcgggctt 3120  
tgctcgacgc cttagccatt gagatgtag ataggcacca tactcacttt tgccctttag 3180  
aaggggaaag ctggcaagat tttttacgta ataacgctaa aagttttaga tgtgctttac 3240  
taagtcacg cgatggagca aaagtacatt taggtacacg gcctacagaa aacagtatg 3300  
aaactctcga aatcaatta gcctttttat gccacaagg ttttctacta gagaatgcat 3360  
tatatgact cagcgcagtg gggcatttta ctttaggttg cgtattggaa gatcaagagc 3420

atcaagtcgc taaagaagaa agggaaacac ctactactga tagtatgccg ccattattac 3480  
gacaagctat cgaattatgt gatcaccaag gtgcagagcc agccttctta ttcggccttg 3540  
aattgatcat atgcggatta gaaaaacaac ttaaatgtga aagtgggtct taaaagcagc 3600  
ataacctttt tccgtgatgg taacttctact agtttaaaag gatctagggtg aagatccttt 3660  
ttgataatct catgacaaa atccctaac gtgagttttc gttccactga gcgtcagacc 3720  
ccgtagaaaa gatcaaagga tcttcttgag atcctttttt tctgcgcgta atctgctget 3780  
tgcaaaaaa aaaaccaccg ctaccagcgg tggtttgttt gccggatcaa gagctaccaa 3840

ctctttttcc gaaggtaact ggcttcagca gagcgcagat accaaatact gtccttctag 3900  
tgtagccgta gttagccac cacttcaaga actctgtagc accgcctaca tacctcgctc 3960  
tgctaactct gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg 4020  
actcaagacg atagttaccg gataagcgc agcggtcggg ctgaacgggg ggttcgtgca 4080  
cacagcccag cttggagcga acgacctaca ccgaactgag atacctacag cgtgagctat 4140  
gagaaagcgc cacgcttccc gaaggagaaa aggcggacag gtatccgta agcggcaggg 4200  
tcggaacagg agagcgcacg agggagcttc cagggggaaa cgcttggtat ctttatagtc 4260

ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt gtgatgctcg tcaggggggc 4320  
ggagcctatg gaaaaacgcc agcaacgcgg cttttttacg gttcctggcc ttttgctggc 4380  
cttttgctca cat 4393

<210> 47  
<211> 404  
<212> PRT  
<213> Artificial Sequence  
<220><223> amino acid sequence of His6-PA#1(200)-hGH as encoded on

pCHO-PA#1(200)-hGH

<400> 47  
 Ser His His His His His His Gly Ala Ser Ser Ser Ala Ala Pro Ala  
 1 5 10 15  
 Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala  
 20 25 30  
 Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro  
 35 40 45  
 Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala  
 50 55 60  
 Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro  
 65 70 75 80  
 Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala  
 85 90 95  
 Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala  
 100 105 110  
 Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro  
 115 120 125  
 Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala  
 130 135 140  
 Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro  
 145 150 155 160  
 Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala  
 165 170 175  
 Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala  
 180 185 190  
 Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro  
 195 200 205  
 Ala Ala Pro Ala Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp  
 210 215 220  
 Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr





attctattct ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata 240  
 gcaggcatgc tggggatgcg gtgggctcta tggcttctga ggcggaaaga accagctggg 300  
 gctctagggg gtatccccac gcgccctgta gcggcgcat t aagcgcggcg ggtgtggtgg 360  
 ttacgcgcag cgtgaccgct acacttgcca gcgccctagc gcccgtcct ttcgctttct 420  
 tcccttcctt tctcgccacg ttcgccggt tccccgtca agctctaaat cggggcatcc 480  
 ctttagggtt ccgatttagt gctttacggc acctcgacc caaaaaactt gattagggtg 540  
  
 atggttcacg tagtgggcca tcgccctgat agacggtttt tcgccctttg acgttggagt 600  
 ccacgttctt taatagtga ctcttgttcc aaactggaac aacactcaac cctatctcgg 660  
 tctattcttt tgatttataa gggattttgg ggatttcggc ctattggtta aaaaatgagc 720  
 tgatttaaca aaaatttaac gcgaattaat tctgtggaat gtgtgtcagt tagggtgtgg 780  
 aaagtcccca ggctccccag gcaggcagaa gtatgcaaag catgcatctc aattagtcag 840  
 caaccaggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcatc 900  
 tcaattagtc agcaaccata gtcccccccc taactccgcc catccccccc ctaactccgc 960  
  
 ccagttccgc ccattctccg ccccatggtc gactaatttt ttttatttat gcagaggccg 1020  
 aggccgcctc tgcctctgag ctattccaga agtagtgagg aggctttttt ggaggcctag 1080  
 gcttttgcaa aaagctcccg ggagcttgta tatccatttt cggatctgat caagagacag 1140  
 gatgaggatc gtttcgcat attgaacaag atggattgca cgcaggttct cgggccgctt 1200  
 ggggtggagag gctattcggc tatgactggg cacaacagac aatcggtctc tctgatgccg 1260  
 ccgtgttccg gctgtcagcg caggggcgcc cggttctttt tgtcaagacc gacctgtccg 1320  
 gtgccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc acgacgggcg 1380  
  
 ttccttgccg agctgtgctc gacgttgcca ctgaagcggg aagggactgg ctgctattgg 1440  
 gcgaagtgcc ggggcaggat ctctgtcat ctcaccttgc tctgcccag aaagtatcca 1500  
 tcatggctga tgcaatgccc cggtgcata cgcttgatcc ggetacctgc ccattcgacc 1560  
 accaagcga acatcgcatc gacgagcac gtactcggat ggaagccggt ctgttcgatc 1620  
 aggatgatct ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggctca 1680  
 aggcgcgcat gcccacggc gaggatctcg tcgtgacca tggcgalgcc tgcttgccga 1740  
 atatcatggt ggaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg 1800  
  
 cggaccgcta tcaggacata gcgttggcta cccgtgatat tgetgaagag cttggcggcg 1860  
 aatgggctga ccgcttctc gtgctttacg gtatcgccgc tcccattcg cagcgcacgc 1920  
 ccttctatcg ccttcttgac gagtcttct gagcgggact ctggggttcg aatgaccga 1980  
 ccaagcgag cccaactgc catcacgaga tticgattcc accgccgct tctatgaaag 2040

gttgggcttc ggaatcgttt tccgggacgc cggctggatg atcctccagc gcggggatct 2100  
catgctggag ttcttcgccc accccaactt gtttattgca gcttataatg gttacaaata 2160  
aagcaatagc atcacaaatt tcacaaataa agcatttttt tcaactgcatt ctagtgtggt 2220

tttgtccaaa ctcatcaatg tatcttatca tgtctgtata ccgtcgacct ctagnetagag 2280  
cttggcgtaa tcatggatcat agctgtttcc tgttgtaaat tgttatccgc tcacaattcc 2340  
acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaat gagtgagcta 2400  
actcacatta attgcgttgc gctcaactgcc cgctttccag tcgggaaacc tgtcgtgcca 2460  
gctgcattaa tgaatcggcc aacgcgcggg gagaggcggg ttgcgtattg ggcgctcttc 2520  
cgcttctctg ctcaactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc 2580  
tcaactcaaag gcggttaatac ggttatccac agaatcaggg gataacgcag gaaagaacat 2640

gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt 2700  
ccataggetc cgccccctg acgagcatca caaaaatcga cgetcaagtc agaggtggcg 2760  
aaacccgaca ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc 2820  
tcctgttccg acctgcgcg ttaccggata cctgtccgcc tttctccctt cgggaagcgt 2880  
ggcgctttct caatgetcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa 2940  
gctgggctgt gtgcacgaac cccccgttca gcccaccgc tgcgccttat ccgtaacta 3000  
tcgtcttgag tccaaccgg taagacacga cttatcgcca ctggcagcag ccaactgtaa 3060

caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa 3120  
ctacggctac actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacctt 3180  
cggaaaaaga gttggtagct ctgatccgg caaacaacc accgctggta gcggtggttt 3240  
ttttgtttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag atcctttgat 3300  
cttttctac gggctctgac ctcaagttaa cgaaaactca cgttaaggga ttttggteat 3360  
gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa gttttaaatc 3420  
aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc 3480

acctatctca ggcgatctgc tatttcgttc atccatagt gctgactcc ccgtcgtgta 3540  
gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga 3600  
cccacgctca ccgctccag atttatcagc aataaaccag ccagccggaa gggccgagcg 3660  
cagaagtggc cctgcaactt tatccgctc catccagtet attaatgtt gccgggaagc 3720  
tagagtaagt agttcggcag ttaatagttt gcgcaacgtt gttgccattg ctacaggcat 3780  
cgtgggtgca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc aacgatcaag 3840

gcgagttaca tgatcccca tgttgtgcaa aaaagcgggt agctccttcg gtcctccgat 3900

cgttgtcaga agtaagtgg ccgagtggt atcactcatg gttatggcag cactgcataa 3960

ttctcttact gtcacgcat ccgtaagatg cttttctgtg actggtgagt actcaaccaa 4020

gtcattctga gaatagtga tgcggcgacc gagttgctct tgcccggcgt caatacggga 4080

taataccgcg ccacatagca gaactttaaa agtgctcctc attggaaaac gttcttcggg 4140

gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccaactcgtgc 4200

accecaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg 4260

aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aatgttgaa tactcatact 4320

cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat 4380

atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt 4440

gccacctgac gtcgacggat cgggagatct cccgatcccc tatggtcgac tctcagtaca 4500

atctgctctg atgccgata gttaaaccag tatctgctcc ctgcttgtgt gttggaggtc 4560

gctgagtagt gcgagagcaa aatttaagct acaacaaggc aaggcttgac cgacaattgc 4620

atgaagaatc tgcttagggt taggcgtttt gcgctgcttc gcgatgtacg ggccagatat 4680

acgcgttgac attgattatt gactagttat taatagtaat caattacggg gtcattagtt 4740

catagcccat atatggagtt ccgcttaca taacttacgg taaatggccc gcctggctga 4800

ccgccaacg acccccggcc attgacgtca ataatgacgt atgttcccat agtaacgcca 4860

atagggactt tccattgacg tcaatgggtg gactatttac ggtaaactgc ccaactggca 4920

gtacatcaag tgtatcatal gccaaagtac cccctattg acgtcaatga cggtaaatgg 4980

cccgcctggc attatgcca gtacatgacc ttatgggact ttctacttg gcagtacatc 5040

tacgtattag tcatcgtat taccatggtg atgcggtttt ggcagtacat caatggcgt 5100

ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt 5160

ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg 5220

acgcaaatgg gcggtaggcg tgtacgggtg gaggtctata taagcagagc tctctggcta 5280

actagagaac ccaactgctta ctggcttacc gaaattaata cgactcacta tagggagacc 5340

caagctgcca ccatggctac aggtcccgg acgtccctgc tcttggtttt tggcctgctc 5400

tgctgcctt ggcttcaaga ggcagcgtc agccatcacc accatcacca tggcggcagc 5460

tctctgccc ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggtgct 5520

cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggtgct 5580

cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 5640  
 cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 5700  
 cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 5760  
 cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 5820  
 cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 5880  
 cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 5940  
 cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 6000

cctgctgccg ctccagctgc acctgctcca gcagcacctg ctgcaccagc tccggctgct 6060  
 cctgctgcct tcccaacct tccttatcc aggccttttg acaacgctat gctccgcgcc 6120  
 catcgtctgc accagctggc ctttgacacc taccaggagt ttgaagaagc ctatatccca 6180  
 aaggaacaga agtattcatt cctgcagaac cccagacct ccctctgttt ctcagagtct 6240  
 attccgacac cctccaacag ggaggaaaca caacagaaat ccaacctaga gctgctccgc 6300  
 atctccctgc tgctcatcca gtcgtggctg gagcccgtgc agttcctcag gagtgtcttc 6360  
 gccaacagcc tgggtgacgg cgctctgac agcaacgtct atgacctctt aaaggacctt 6420

gaggaaggca tccaaacgct gatggggagg ctggaagatg gcagcccccg gactgggcag 6480  
 atcttcaagc agacctacag caagttcgac acaactcac acaacgatga cgcactactc 6540  
 aagaactacg ggctgctcta ctgcttcagg aaggacatgg acaaggtcga gacattcttg 6600  
 cgcacgtgac agtgccgctc tgtggagggc agctgtggct tctaagcttg gcc 6653

<210> 49

<211> 6041

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of pCHO-hGH

<400> 49

ctattctata gtgtcaccta aatgctagag ctgctgatac agcctcgact gtgccttcta 60

gttgccagcc atctgttgtt tgccccctcc cegtgccttc cttgacctg gaagtgcca 120

ctcccactgt cctttcctaa taaatgagg aaattgcac gcattgtctg agtaggtgtc 180

attctattct ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata 240

gcaggcatgc tggggatgag gtgggctcta tggcttctga ggcggaaaga accagctggg 300

gctctagggg gtatccccac ggcctctgta gcggcgcat aagcgcggcg ggtgtggtgg 360

ttacgcgag cgtgaccgct aacttgcca ggcctctagc gcccgtcct ttcgctttct 420

tcctttcctt tctcgccacg ttccggcgtt ttccccgtca agctctaaat cggggcatcc 480  
  
 ctttaggggtt cggatttagt gctttacggc acctcgacce caaaaaactt gattagggtg 540  
 atggttcacg tagtgggcca tcgcccctgat agacggtttt tcgccctttg acgttggagt 600  
 ccacgttctt taatagtgga ctcttgttcc aaactggaac aacactcaac cctatctcgg 660  
 tctattcttt tgatttataa gggatthttg ggatttcggc ctattggtta aaaaatgagc 720  
 tgatttaaca aaaatttaac gcgaattaat tctgtggaat gtgtgtcagt tagggtgtgg 780  
 aaagtcccca ggctccccag gcaggcagaa gtatgcaaag catgcatctc aattagtcag 840  
 caaccagggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcatc 900  
  
 tcaattagtc agcaaccata gtcccccccc taactccgcc catccccccc ctaactccgc 960  
 ccagttccgc ccattctccg ccccatggct gactaatttt tttatttat gcagaggccg 1020  
 aggccgcctc tgccctctgag ctattccaga agtagtgagg aggccttttt ggaggcctag 1080  
 gcttttgcaa aaagctcccg ggagcttcta tatecatttt cggatctgat caagagacag 1140  
 gatgaggatc gtttcgatg attgaacaag atggattgca cgcaggttct ccggccgctt 1200  
 ggggtggagag gctattcggc tatgactggg cacaacagac aatcggtctc tctgatgccg 1260  
 ccgtgttccg gctgtcagcg caggggcgcc cggttctttt tgtcaagacc gacctgtccg 1320  
  
 gtgccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc acgacgggcg 1380  
 ttccctgcgc agctgtgctc gacgttgfca ctgaagcggg aagggactgg ctgctattgg 1440  
 gcgaagtgcc ggggcaggat ctctgtcat ctacacttgc tctgcccag aaagtatcca 1500  
 tcatggctga tgcaatgcgg cggctgcata cgcttgatcc ggctacctgc ccattcgacc 1560  
 accaagcgaa acatcgcac gacgagcac gtactcggat ggaagccggt cttgtcgatc 1620  
 aggatgatct ggacgaagag catcaggggc tcgcccagc cgaactgttc gccaggctca 1680  
 aggcgcgcat gcccacggc gaggatctcg tcgtgacca tggcgtatcc tgcttgccga 1740  
  
 atatcatggt ggaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg 1800  
 cggaccgcta tcaggacata gcgttgcta cccgtgatat tgctgaagag cttggcggcg 1860  
 aatgggctga ccgcttctc gtgctttacg gtatcgccgc tcccattcg cagcgcacg 1920  
 ccttctatcg ctttcttgac gagtcttct gacggggact ctggggttcg aaatgaccga 1980  
 ccaagcgacg cccaactgc catcacgaga ttctgattcc accgccgct tctatgaaag 2040  
 gttgggcttc ggaatcgtht tccgggacgc cggctggatg atctccagc gcggggatct 2100  
 catgctggag ttcttcgccc accccaactt gtttattgca gcttataatg gttacaaata 2160

aagcaatagc atcacaatt tcacaaataa agcatttttt tcaactgcatt ctagtgtg 2220  
tttgtccaaa ctcatcaatg tatcttatca tgtctgtata ccgtcgacct ctactagag 2280  
cttggcgtaa tcatggcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc 2340  
acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaatae gagtgagcta 2400  
actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgtcgtgcca 2460  
gctgcattaa tgaatcggcc aacgcgcggg gagaggcggg ttgcgtattg ggcgctcttc 2520  
cgcttctcgc ctcaactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc 2580  
  
tcaactcaaag gcggtaatac ggttatccac agaactcaggg gataacgcag gaaagaacat 2640  
gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt 2700  
ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg 2760  
aaacccgaca ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc 2820  
tctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt 2880  
ggcgttttct caatgctcac gctgtaggta tctcagttcg gtgtaggctg ttcgctccaa 2940  
gctgggctgt gtgcacgaac cccccgttca gcccagccgc tgcgcttat ccgtaacta 3000  
  
tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag ccaactggtaa 3060  
caggattagc agagcgagg atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa 3120  
ctacggctac actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacctt 3180  
cggaaaaaga gttggtagct ctgatccgg caaacaacc accgctggta gcggtggttt 3240  
ttttgtttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag atcctttgat 3300  
cttttctac gggctcgacg ctcactggaa cgaaaactca cgttaaggga ttttggctat 3360  
gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa gtttttaatc 3420  
  
aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgagc 3480  
acctatctca gcgatctgtc tatttcttc atccatagtt gctgactcc ccgtcgtgta 3540  
gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga 3600  
cccacgtca ccggtccag atttatcagc aataaaccag ccagccggaa gggccgagcg 3660  
cagaagtggg ctgcaactt tatccgctc catecagctt attaatgtt gccgggaagc 3720  
tagagtaagt agttcggag ttaatagtt gcgcaacgtt gttgccattg ctacaggcat 3780  
cgtgggtgca cgctcgtcgt ttggtatggc ttcaatcagc tccggttccc aacgatcaag 3840  
  
gcgagttaca tgatcccca tgttgtgcaa aaaagcgggt agtccttcg gtcctccgat 3900  
cgttgtcaga agtaagtgg ccgcagtgtt atcaactatg gttatggcag cactgcataa 3960  
ttctcttact gtcacccat ccgtaagatg cttttctgtg actgggtgagt actcaaccaa 4020

gtcattctga gaatagtgtg tgcggcgacc gagttgctct tgccccgcgt caatacggga 4080  
 taataccgcg ccacatagca gaactttaa agtgctcatc attggaaaac gttcttcggg 4140  
 gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccaactcgtgc 4200  
 acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg 4260

aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgtttaa tactcatact 4320  
 cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat 4380  
 atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt 4440  
 gccacctgac gtcgacggat cgggagatct cccgatcccc tatggtcgac tctcagtaca 4500  
 atctgctctg atgcccata gttaagccag tatctgctcc ctgcttgtgt gttggaggtc 4560  
 gctgagtagt gcgcgagcaa aatttaagct acaacaaggc aaggcttgac cgacaattgc 4620  
 atgaagaatc tgcttagggt taggcgtttt gcgctgcttc gcgatgtacg ggccagatat 4680

acgcgttgac attgattatt gactagtat taatagtaat caattacggg gtcattagtt 4740  
 catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga 4800  
 ccgccaacg acccccgcc attgacgtca ataatgacgt atgttcccat agtaacgcca 4860  
 atagggactt tccattgacg tcaatgggtg gactatttac ggtaaactgc ccaacttgca 4920  
 gtacatcaag tgtatcatal gcccaagtacg ccccctattg acgtcaatga cggtaaatgg 4980  
 cccgcctggc attatgccca gtacatgacc ttatgggact ttctacttg gcagtacatc 5040  
 tacgtattag tcatcgctat taccatggtg atgcggtttt ggcagtacat caatggcgt 5100

ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt 5160  
 ttgttttggc accaaaaata acgggacttt ccaaaaatgc gtaacaactc cgccccattg 5220  
 acgcaaatgg gcggtaggcg tgtacgggtg gaggtctata taagcagagc tctctggcta 5280  
 actagagaac ccaactgctta ctggcttacc gaaattaata cgactcacta tagggagacc 5340  
 caagctgccca ccatggctac aggcctcccg acgtccctgc tctggcttt tggcctgctc 5400  
 tgcctgcctt ggcttcaaga gggcagcgt agccatcacc accatcacca tggcgccttc 5460  
 ccaaccattc cttatccag gctttttgac aacgctatgc tccgcgcca tegtctgeac 5520

cagctggcct ttgacaccta ccaggagttt gaagaagcct atatccaaa ggaacagaag 5580  
 tattcattcc tgcagaacce ccagacctcc ctctgtttct cagagtctat tccgacacc 5640  
 tccaacaggg aggaaacaca acagaaatcc aacctagagc tgetccgat ctccctgctg 5700  
 ctcatccagt cgtggctgga gccctgacg ttctcagga gtgtcttcgc caacagcctg 5760  
 gtgtacggcg cctctgacag caacgtctat gacctcctaa aggacctaga ggaaggcatc 5820  
 caaacgctga tggggaggct ggaagatggc agccccgga ctgggcagat cttcaagcag 5880



acctacagca agttcgacac aaactcacac aacgatgacg cactactcaa gaactacggg 5940

ctgctctact gcttcaggaa ggacatggac aaggtcgaga cattcctgcg catcgtgcag 6000

tgccgctctg tggagggcag ctgtggcttc taagcttggc c 6041

<210> 50

<211> 5449

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of pCHO

<400> 50

ctattctata gtgtcaccta aatgctagag ctcgctgac agcctcgact gtgccttcta 60

gttgccagcc atctgttgtt tgcccctccc ccgtgccttc cttgaccctg gaaggtgcca 120

ctcccactgt cctttcctaa taaaatgagg aaattgcac gcattgtctg agtaggtgtc 180

attctattct ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata 240

gcaggcatgc tggggatgcg gtgggctcta tggcttctga ggcggaaaga accagctggg 300

gctctagggg gtaeccccac gcgccctgta gcggcgcat aagcgcggcg ggtgtggtgg 360

ttacgcgcag cgtgaccgct aactttgcca gcgccctagc gcccgctcct ttcgctttct 420

tccttctctt tctcgccacg ttcgccggtt tccccgtca agctctaaat cggggcatcc 480

ctttaggggt cggatttagt gctttacggc acctcgacc caaaaaactt gattaggggtg 540

atggttcacg tagtgggcca tcgccctgat agacggtttt tcgccctttg acgttggagt 600

ccacgttctt taatagtgga ctcttgttcc aaactggaac aacactcaac cctatctcgg 660

tctattcttt tgatttataa gggattttgg ggatttcggc ctattggtta aaaaatgagc 720

tgatttaaca aaaatttaac gcgaattaat tctgtggaat gtgtgtcagt tagggtgtgg 780

aaagtcccca ggctcccagc gcaggcagaa gtatgcaaag catgcatctc aattagtcag 840

caaccagggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcatc 900

tcaattagtc agcaaccata gtcccgcgcc taactccgcc catcccgcgcc ctaactccgc 960

ccagttccgc ccattctcgc cccatggct gactaatttt ttttatttat gcagaggcgc 1020

aggccgcctc tgccctgag ctattccaga agtagtgagg aggccttttt ggaggcctag 1080

gcttttgcaa aaagctcccg ggagcttgta tatccatttt cggatctgat caagagacag 1140

gatgaggatc gtttcgatg attgaacaag atggattgca cgcagggtct cgggccgctt 1200

gggtggagag gctattcggc tatgactggg cacaacagac aatcggtctc tctgatgccg 1260

ccgtgttccg getgtcagcg caggggcgcc cggttctttt tgtcaagacc gacctgtccg 1320

gtgccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc acgacgggcg 1380  
 ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg aagggactgg ctgctattgg 1440

gcgaagtgcc ggggcaggat ctctgtcat ctcaccttgc tcttgccgag aaagtatcca 1500  
 tcatggctga tgcaatgcgg cgctgcata cgcttgatcc ggetacctgc ccattcgacc 1560  
 accaagcgaa acatgcctc gagcgagcac gtactcggat ggaagccggt cttgtcgatc 1620  
 aggatgatct ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggctca 1680  
 aggcgcgcat gcccagcggc gaggatctcg tcgtgacca tggcgatgcc tgcttgccga 1740  
 atatcatggt ggaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg 1800  
 cggaccgcta tcaggacata gcgttgcta cccgtgatat tgctgaagag cttggcggcg 1860

aatgggctga cgcttcctc gtgctttacg gtatgcgcc tcccattcg cagcgcacgc 1920  
 ccttctatcg ccttcttgac gagtcttct gagcgggact ctggggttcg aaatgaccga 1980  
 ccaagcgag cccaacctgc catcacgaga tttegattcc accgcccct tctatgaaag 2040  
 gttgggcttc ggaatcgttt tccgggacgc cggctggatg atcctccagc gcggggatct 2100  
 catgctggag ttcttcgcc accccaactt gtttattgca gcttataatg gttacaaata 2160  
 aagcaatagc atcacaatt tcacaaataa agcatttttt tcaactgcatt ctagtgtgg 2220  
 tttgtccaaa ctcatcaatg tatcttata tgtctgata ccgtcgacct ctagctagag 2280

cttggcgtaa tcatggtcat agctgttcc tgtgtgaaat tgttatccgc tcacaattcc 2340  
 acacaacata cgagccggaa gcataaagt taaagcctgg ggtgcctaat gagtgagcta 2400  
 actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgctgtgcca 2460  
 gctgcattaa tgaatcgcc aacgcgcggg gagaggcggg ttgcgtattg ggcgctcttc 2520  
 cgcttctcg ctcaactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggatcagc 2580  
 tcaactcaag gcgtaatac ggttatccac agaatcaggg gataacgcag gaaagaacat 2640  
 gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt 2700

ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc agagtgggcg 2760  
 aaaccgaca ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc 2820  
 tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt cggaagcgt 2880  
 ggcgctttct caatgctcac gctgtagta tctcagttcg gtgtaggtcg ttcgctcaa 2940  
 gctgggctgt gtgcacgaac cccccgttca gcccagcgc tgcgcttat ccggtaaacta 3000  
 tcgtcttgag tccaaccgg taagacacga cttatcgcca ctggcagcag ccaactggtaa 3060

caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa 3120  
  
 ctacggctac actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacctt 3180  
 cggaaaaaga gttggtagct ctgatccgg caaacaacc accgctggta gcggtggttt 3240  
 ttttgtttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag atcctttgat 3300  
 cttttctacg gggctcgacg ctcagtggaa cgaaaactca cgtaaggga ttttggcat 3360  
 gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa gttttaaatc 3420  
 aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc 3480  
 acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtcgtgta 3540  
  
 gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga 3600  
 cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg 3660  
 cagaagtggc cctgcaactt tatccgctc catccagtct attaatggt gccgggaagc 3720  
 tagagtaagt agttcccgag ttaatagttt gcgcaacgtt gttgccattg ctacaggcat 3780  
 cgtgggtgca cgctcgtcgt ttggatggc ttcatcagc tccggttccc aacgatcaag 3840  
 gcgagttaca tgatcccca tgttgtgcaa aaaagcggtt agctccttcg gtcctccgat 3900  
 cgttgtcaga agtaagtgg ccgcagtgtt atcactcatg gttatggcag cactgcataa 3960  
  
 ttctcttact gtcatgcat ccgtaagatg cttttctgtg actggtgagt actcaaccaa 4020  
 gtcattctga gaatagtgtg tgcggcgacc gagttgctct tgccccgct caatacggga 4080  
 taataccgcg ccacatagca gaactttaa agtgctcacc attggaaaac gttcttcggg 4140  
 gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccaactcgtgc 4200  
 acccaactga tettcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg 4260  
 aaggcaaaat gccgcaaaaa aggaataag ggcgacacgg aaatgtttaa tactcatact 4320  
 cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat 4380  
  
 atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacattc cccgaaaagt 4440  
 gccacctgac gtcgacggat cgggagatct cccgatccc tatggtcgac tctcagtaca 4500  
 atctgctctg atgccgata gttaaagccag tatctgctcc ctgcttgtgt gttggaggtc 4560  
 gctgagtagt gcgcgagcaa aatttaagct acaacaaggc aaggcttgac cgacaattgc 4620  
 atgaagaatc tgcttagggt taggcgtttt gcgctgctc gcgatgtac ggccagatat 4680  
 acgcgttgac attgattatt gactagttat taatagtaat caattacggg gtcattagtt 4740  
 catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga 4800

ccgccaacg acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca 4860  
 atagggactt tccattgacg tcaatgggtg gactatttac ggtaaactgc ccacttggca 4920  
 gtacatcaag tgtatcatat gccaaagtacg ccccctattg acgtcaatga cggtaaatgg 4980  
 cccgcctggc attatgcccc gtacatgacc ttatgggact ttctacttg gcagtacatc 5040  
 tacgtattag tcatcgctat taccatgggtg atgcggtttt ggcagtacat caatgggcgt 5100  
 ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt 5160  
 ttgttttggc accaaaaatca acgggacttt ccaaaaatgtc gtaacaactc cgccccattg 5220

acgcaaatgg gcggtaggcg tgtacgggtg gaggtctata taagcagagc tctctggcta 5280  
 actagagaac ccactgctta ctggcttacc gaaattaata cgactcacta tagggagacc 5340  
 caagctgccca ccatggctac aggctcccgg acgtccctgc tctggcttt tggcctgctc 5400  
 tgcctgcctt ggcttcaaga gggcagcgt agctctagaa agcttggcc 5449

<210> 51

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of P1A1

<400> 51

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

1 5 10 15

Ala Pro Ala Pro

20

<210> 52

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of upper/coding strand

oligodesoxynucleotide used for the generation of the building

block for P1A1

<400> 52

gccctgctc ctgctccagc acctgcacca gcacctgctc cagcaccagc tctgcacca 60

60

<210> 53

<211> 60  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> nucleic acid sequence of lower/non-coding strand  
 oligodesoxynucleotide used for the generation of the building  
 block for P1A1  
 <400> 53  
 ggctggtgca ggagctgggtg ctggagcagg tgctgggtgca ggtgctggag caggagcagg 60  
 60  
 <210> 54  
 <211> 60  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> nucleic acid sequence of upper/coding strand  
 oligodesoxynucleotide used for the generation of the building  
 block for P1A3  
 <400> 54  
 gccgctgcac ctgctgcagc acctgctgca gctccagcag ctgctcctgc agcagctcca 60  
 60  
 <210> 55  
 <211> 60  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> nucleic acid sequence of lower/non-coding strand  
 oligodesoxynucleotide used for the generation of the building  
 block for P1A3  
 <400>  
 > 55  
 ggctggagct gctgcaggag cagctgctgg agctgcagca ggtgctgcag caggtgcagc 60  
 60  
 <210> 56  
 <211> 417  
 <212> PRT  
 <213> Artificial Sequence

<220><223> amino acid sequence of the Fab light chain fused with the

P1A1(200) polymer as encoded on pFab-P1A1(200)

<400> 56

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

1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val Asn Thr Ala

20 25 30

Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile

35 40 45

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

50 55 60

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

65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr Thr Thr Pro Pro

85 90 95

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

100 105 110

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

115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala

130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln

145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser

165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr

180 185 190

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

195 200 205

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

210 215 220

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

225                      230                      235                      240

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

245                      250                      255

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

260                      265                      270

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

275                      280                      285

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

290                      295                      300

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

305                      310                      315                      320

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

325                      330                      335

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

340                      345                      350

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

355                      360                      365

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

370                      375                      380

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

385                      390                      395                      400

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

405                      410                      415

Ala

<210> 57

<211> 417

<212> PRT

<213> Artificial Sequence

<220><223> amino acid sequence of the Fab light chain fused with the

P1A3(200) polymer as encoded on pFab-P1A3(200)

<400> 57

Asp Ile Glu Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly  
 1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val Asn Thr Ala  
 20 25 30

Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile  
 35 40 45

Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg Phe Ser Gly  
 50 55 60

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

65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr Thr Thr Pro Pro  
 85 90 95

Thr Phe Gly Gln Gly Thr Lys Leu Glu Ile Lys Arg Thr Val Ala Ala  
 100 105 110

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

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala  
 130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln  
 145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser  
 165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr  
 180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser  
 195 200 205

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

210 215 220

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro



225                    230                    235                    240  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          245                    250                    255  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          260                    265                    270  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          275                    280                    285

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          290                    295                    300  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 305                    310                    315                    320  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          325                    330                    335  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          340                    345                    350  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          355                    360                    365  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          370                    375                    380  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
 385                    390                    395                    400  
 Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro  
                          405                    410                    415

Ala

<210> 58

<211> 5210

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of pFab-P1A1(200)

<400> 58

acccgacacc atcgaatggc cagatgatta attcctaatt ttgttgaca ctctatcatt

60

gatagagtta ttttaccact ccctatcagt gatagagaaa agtgaaatga atagttcgac 120  
aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtgccac 180  
tggctggitt cgctaccgta ggcgagccg aagttaaact gcaggaatcc ggtggtggtc 240  
tggttcagcc aggtggttcc ctgcggtctc cgtgtgctgc ttccggtttc aacatcaaag 300  
acacctacat ccaactgggtt cgtcaggctc cgggtaaagg cctggaatgg gttgctcgt 360  
  
tctaccgac caacggttac accaggatg ccgattcagt taaaggtcgt ttaccatct 420  
cggccgacac ttccaaaaac accgcttacc tccagatgaa ctccctcgt gctgaagaca 480  
cagctgttta ttattgctcc cgttggggtg gtgacggttt ctacgctatg gactactggg 540  
gtcagggtag cctggtcacc gtctctcag cctccacaa gggcccatcg gtcttcccc 600  
tggcaccctc ctccaagagc acctctgggg gcacagcggc cctgggctgc ctggtcaagg 660  
actacttccc cgaaccggtg acggtgtcgt ggaactcagg cgcctgacc agcggcgtgc 720  
acaccttccc ggctgtccta cagtctcag gactctactc cctcagcagc gtggtgactg 780  
  
tgccctccag cagcttgggc acccagacct acatctgcaa cgttaatcac aaaccagca 840  
acaccaaggt cgacaagaaa gttgagccca aatcttgcca tcaccacat caccattaat 900  
aacatggag aaaataaagt gaaacaaagc actattgcac tggcactctt accgttactg 960  
tttaccctg tgacaaaagc cgacatcgag ctacaccaat ccccgctctc cctgtccgct 1020  
tccgttggcg accgtgttac catcacgtgt agggcctcgc aagacgtaaa caccgccgta 1080  
gcgtggtagc agcagaaacc cgggaaagct ccgaaactgc tgatctatag cgcttccttc 1140  
ctgtattccg gagtccgag caggttcagt ggttcccgtt ccggtaccga cttaccctg 1200  
  
acgatatcct cctccagcc ggaagacttc gctacctact actgteaaca gcactacacc 1260  
accccgcca ccttcggtca gggtagcaaa ctcgagatca aacggactgt ggctgcacca 1320  
tctgtcttca tcttcccgc atctgatgag cagttgaaat ctggaactgc ctctgttgtg 1380  
tgctgtctga ataacttcta tcccagagag gccaaagtac agtggaaagt ggataacgcc 1440  
ctccaatcgg gtaactcca ggagagtgtc acagagcagg acagcaagga cagcacctac 1500  
agcctcagca gcacctgac gctgagcaaa gcagactacg agaaacacaa agtctacgcc 1560  
tgcaagtca cccatcaggg cctgagttcg cccgtcacia agagcttcaa ccgaggagag 1620  
  
tgctctctg ccctgctcc tgetccagca cctgcaccag cacctgctcc agcaccagct 1680  
cctgcaccag ccctgctcc tgetccagca cctgcaccag cacctgctcc agcaccagct 1740  
cctgcaccag ccctgctcc tgetccagca cctgcaccag cacctgctcc agcaccagct 1800  
cctgcaccag ccctgctcc tgetccagca cctgcaccag cacctgctcc agcaccagct 1860  
cctgcaccag ccctgctcc tgetccagca cctgcaccag cacctgctcc agcaccagct 1920

cctgcaccag ccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct 1980  
 cctgcaccag ccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct 2040  
  
 cctgcaccag ccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct 2100  
 cctgcaccag ccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct 2160  
 cctgcaccag ccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct 2220  
 cctgcaccag cctgaagagc ttaagcttga cctgtgaagt gaaaaatggc gcacatttg 2280  
 cgacattttt ttgtctgcc gtttaccgct actgcgtcac ggatctccac gcgcctgta 2340  
 gcggcgcat aagcgcggcg ggtgtggtgg ttacgcgcag cgtgaccgct acaattgcca 2400  
 gcgccttagc gccctgctct ttcgctttct tcccttctt tctgccacg ttcgccggt 2460  
  
 ttccccgtca agctctaaat cgggggctcc ctttagggtt ccgatttagt gctttacggc 2520  
 acctgaccc caaaaaactt gattaggttg atggttcacg tagtgggcca tcgccctgat 2580  
 agacggtttt tcgcccttg acgttggagt ccacgttctt taatagtga ctcttgttcc 2640  
 aaactggaac aacactcaac cctatctcgg tctattctt tgatttataa gggattttgc 2700  
 cgatttcggc ctattggtta aaaaatgagc tgatttaaca aaaatttaac gcgaatttta 2760  
 acaaaatatt aacgtttaca atttcaggtg gcacttttcg gggaaatgtg cgcggaacc 2820  
 ctatttgttt attttctaa atacattcaa atatgatcc gctcatgaga caataaccct 2880  
  
 gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtcg 2940  
 ccctattcc ctttttgcg gcattttgcc ttccgtttt tgctcacca gaaacgctgg 3000  
 tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc 3060  
 tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca 3120  
 cttttaaagt tctgctatgt ggcgcggtat tatcccgat tgacgccggg caagagcaac 3180  
 tcggtcgccg catacactat tctcagaatg acttggttga gtactacca gtcacagaaa 3240  
 agcatcttac ggatggcatg acagtaagag aattatgcag tgetgccata accatgagtg 3300  
  
 ataacactgc ggccaactta ctctgacaa cgatcggagg accgaaggag ctaaccgctt 3360  
 ttttgcaaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg 3420  
 aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc 3480  
 gcaaaactatt aactggcgaa ctacttact tagcttccc gcaacaattg atagactgga 3540  
 tggaggcgga taaagtigca ggaccacttc tgcgctcggc ccttccggct ggctggttta 3600  
 ttgctgataa atctggagcc ggtgagcgtg gctctcgcg tatcattgca gcaactgggc 3660

cagatggtaa gcctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg 3720

atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat 3780

taatgatgic tcgtttagat aaaagtaaag tgattaacag cgcattagag ctgcttaatg 3840

aggtcggaat cgaaggttta acaacccgta aactcgccca gaagctaggt gtagagcagc 3900

ctacattgta ttggcatgta aaaaataagc gggccttgct cgacgcctta gccattgaga 3960

tgtagatag gcaccatact cacttttgcc ctttagaagg ggaaagctgg caagatTTTT 4020

tacgtaataa cgctaaaagt ttagatgtg ctttactaag tcacgcgat ggagcaaaag 4080

tacatttagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct 4140

TTTTatgcca acaaggtttt tcaactagaga atgcattata tgcactcagc gcagtggggc 4200

attttacttt aggttgcgta ttggaagatc aagagcatca agtcgctaaa gaagaaaggg 4260

aaacacctac tactgatagt atgccgccat tattacgaca agctatcgaa ttatttgatc 4320

accaaggtgc agagccagcc ttcttattcg gccttgaatt gatcatatgc ggattagaaa 4380

aacaacttaa atgtgaaagt gggctttaa agcagcataa ctttttccg tgatggtaac 4440

ttcaactagt taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc 4500

cttaactgta gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt 4560

cttgagatcc ttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac 4620

cagcgggtgg ttgtttgcc gatcaagagc taccaactct tttccgaag gtaactggct 4680

tcagcagagc gcagatacca aactactgcc ttctagtgtg gccgtagtta ggccaccact 4740

tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg 4800

ctgccagtgg cgataagtgc tgtcttaccg ggttggactc aagacgatag ttaccggata 4860

aggcgcagcg gtcgggctga acggggggtt cgtgcacaca gccagcttg gagcgaacga 4920

cctacaccga actgagatac ctacagcgtg agctatgaga aagcggccag cttcccgaag 4980

ggagaaaggc ggacaggtat ccgtaagcg gcagggtcgg aacaggagag cgcacgaggg 5040

agcttccagg gggaaacgcc tggatcttt atagtcctgt cgggtttcgc cacctctgac 5100

ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca 5160

acgcggcctt ttacgggttc ctggcctttt gctggccttt tgctcacatg 5210

<210> 59

<211> 5210

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of pFab-P1A3(200)

<400> 59

acccgacacc atcgaatggc cagatgatta attcctaatt ttgttgaca ctctatcatt 60  
  
 gatagagtta ttttaccact ccctatcagt gatagagaaa agtgaaatga atagttcgac 120  
 aaaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtggcac 180  
 tggctggttt cgctaccgta gcgcaggccg aagttaaact gcaggaatcc ggtggtggtc 240  
 tggttcagcc aggtggttcc ctgcggctct cgtgtgctgc ttccggttcc aacatcaaag 300  
 acacctacat ccaactgggtt cgtcaggctc cgggtaaagg cctggaatgg gttgctcgta 360  
 tctaccggac caacggttac accaggatg ccgattcagt taaaggctgt ttcaccatct 420  
 cgcccgacac ttccaaaaac accgcttacc tccagatgaa ctccctgcgt gctgaagaca 480  
  
 cagctgttta ttattgtccc cgttggggtg gtgacggttt ctacgctatg gactactggg 540  
 gtcagggtag cctggtcacc gtctctcag cctccacca gggcccatcg gtcttcccc 600  
 tggcaccctc ctccaagagc acctctgggg gcacagcggc cctgggctgc ctggtcaagg 660  
 actacttccc cgaaccggtg acggtgtcgt ggaactcagg cgcctgacc agcggcgtgc 720  
 acaccttccc ggctgtccta cagtctcag gactctact cctcagcagc gtggtgactg 780  
 tgcctccag cagcttgggc acccagacct acatctgcaa cgttaatcac aaaccagca 840  
 acaccaaggt cgacaagaaa gttgagccca aatcttgcca tcaccacat caccattaat 900  
  
 aaccatggag aaaataaagt gaaacaaagc actattgcac tggcactctt accgttactg 960  
 tttaccctg tgacaaaagc cgacatcgag ctacaccaat ccccgctctc cctgtccgct 1020  
 tccgttggcg accgtgttac catcacgtgt agggcctcgc aagacgtaaa caccgccgta 1080  
 gcgtggtatc agcagaaaacc cgggaaagct cggaaactgc tgatctatag cgcttcttc 1140  
 ctgtattccg gagtcccgag caggttcagt ggttcccggt cgggtaccga cttcacctg 1200  
 acgatatcct cctccagcc ggaagacttc gctacctact actgtcaaca gcactacacc 1260  
 accccgccga cttcggtea ggttacaaa ctcgagatca aacggactgt ggctgacca 1320  
  
 tctgtcttca tcttcccgcc atctgatgag cagttgaaat ctggaactgc ctctgtttgt 1380  
 tgctgctga ataacttcta tcccagagag gccaaagtac agtggaaggt ggataacgcc 1440  
 ctccaatcgg gtaactccca ggagagtgtc acagagcagg acagcaagga cagcacctac 1500  
 agcctcagca gcacctgac gctgagcaaa gcgactacg agaaacacaa agtctacgcc 1560  
 tgcaagtca cccatcaggg cctgagttcg cccgtcacia agagcttcaa ccgaggagag 1620  
 tgctcttctg ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 1680

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 1740

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 1800

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 1860

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 1920

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 1980

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 2040

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 2100

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 2160

gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca 2220

gcagctccag cctgaagagc ttaagcttga cctgtgaagt gaaaaatggc gcacattgtg 2280

cgacatTTTT tttgtctgcc gtttaccgct actgcgtcac ggatctccac gcgccctgta 2340

gcggcgcatt aagcgcggcg ggtgtggtag ttacgcgcag cgtgaccgct acacttgcca 2400

gcgccctagc gcccgctcct ttcgctttct tcccttcctt tctcgccacg ttcgccggct 2460

ttccccgta agctctaaat cgggggctcc ctttagggtt ccgatttagt gctttacggc 2520

acctcgacc caaaaaactt gattagggtg atggttcacg tagtgggcca tcgccctgat 2580

agacggTTTT tcgccctttg acgttggagt ccacgttctt taatagtga cttttgttcc 2640

aaactggaac aacctcaac cctatctcgg tctattcttt tgatttataa gggatTTTgc 2700

cgatttcggc ctattggtta aaaaatgagc tgatttaaca aaaattaac gcgaatttta 2760

acaaaatatt aacgtttaca atttcagtg gcacttttcg gggaaatgtg cgcggaacc 2820

ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga caataacct 2880

gataaatgct tcaataatat tgaaaaagga agagiatgag tattcaacat ttccgtgtcg 2940

cccttattcc cttttttgcg gcattttgcc ttctgtttt tgctcaccca gaaacgtgg 3000

tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc 3060

tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca 3120

cttttaaagt tctgctatgt ggcgcggfat tatcccgat tgacgccggg caagagcaac 3180

tcggtcgccg catacactat tctcagaatg acttggttga gtactcacca gtcacagaaa 3240

agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata accatgagtg 3300

ataaactgc ggccaactta cttctgacaa cgatcggagg accgaaggag ctaaccgctt 3360

ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg 3420

aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc 3480  
 gcaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaattg atagactgga 3540  
 tggaggcgga taaagttgca ggaccacttc tgcgctcggc ccttccggct ggctggttta 3600  
 ttgctgataa atctggagcc ggtgagcgtg gctctcggg tafcattgca gcaactggggc 3660  
 cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg 3720  
 atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat 3780  
 taatgatgic tcgtttagat aaaagtaaac tgattaacag cgcattagag ctgcttaatg 3840  
  
 aggtcggaat cgaaggttta acaaccgta aactcgcca gaagctaggt gtagagcagc 3900  
 ctacattgta ttggcatgta aaaaataagc gggctttgct cgacgcctta gccattgaga 3960  
 tgttagatag gcaccatact cacttttgcc ctttagaagg ggaaagctgg caagatTTTT 4020  
 tacgtaataa cgctaaaagt ttagatgtg ctttactaag tcatcgcgat ggagcaaaag 4080  
 tacatttagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct 4140  
 ttttatgcca acaaggtttt tctactagaga atgcattata tgcaactcagc gcagtggggc 4200  
 attttacttt aggttgcgta ttggaagatc aagagcatca agtcgctaaa gaagaaaggg 4260  
  
 aaacacctac tactgatagt atgccgcat tattacgaca agetatcgaa ttatttgatc 4320  
 accaaggtgc agagccagcc ttcttattcg gccttgaatt gatcatatgc ggattagaaa 4380  
 aacaacttaa atgtgaaagt gggctttaa agcagcataa ctttttccg tgatggtaac 4440  
 ttcactagtT taaaaggatc taggtgaaga tccttttga taatctcatg accaaaatcc 4500  
 cttaacgtga gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt 4560  
 cttgagatcc ttttttctg cgcgtaatct gctgcttgca acaaaaaaaaa ccaccgctac 4620  
 cagcgggtgt ttgtttgccg gatcaagagc taccaactct tttccgaag gtaactggct 4680  
  
 tcagcagagc gcagatacca aatactgtcc ttctagtgt gccgtagtta ggccaccact 4740  
 tcaagaactc ttagcaccgc cctacatacc tcgctctgct aatcctgtta ccagtggctg 4800  
 ctgccagtgg cgataagtgc tgtcttaccg ggttggactc aagacgatag ttaccggata 4860  
 aggcgcagcg gtcgggctga acggggggtt cgtgcacaca gccagcttg gagcgaacga 4920  
 cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag 4980  
 ggagaaaggc ggacaggtat ccgtaagcg gcagggtcgg aacaggagag cgcacgaggg 5040  
 agcttccagg gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc cacctctgac 5100  
  
 ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca 5160  
 acgcgccctt tttacggttc ctggcctttt gctggccttt tgctcacatg 5210

<210> 60

<211> 3680

<212> DNA

<213> Artificial Sequence

<220><223> nucleic acid sequence of pSUMO-PA#1(200)

<400> 60

```

gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa atacattcaa      60
atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat tgaaaaagga      120
agagtatgag tattcaacat ttccgtgtcg cccttattcc cttttttgcg gcattttgcc      180

ttcctgtttt tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg      240
gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt gagagttttc      300
gccccgaaga acgttttcca atgatgagca cttttaaagt tctgctatgt ggcgcggtat      360
tatcccgtat tgacgccggg caagagcaac tcggtcgccg catacactat tctcagaatg      420
acttggttga gtactcacca gtcacagaaa agcatcttac ggatggcatg acagtaagag      480
aattatgcag tgctgccata accatgagtg ataacactgc ggccaactta cttctgacaa      540
cgatcggagg accgaaggag ctaaccgctt ttttgcaaa catgggggat catgtaactc      600

gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca      660
cgatgcctgt agcaatggca acaacgttgc gcaaaactatt aactggcgaa ctacttactc      720
tagcttcccg gcaacaatta atagactgga tggaggcggg taaagttgca ggaccacttc      780
tgcgctcggc ccttccggtt ggctggttta ttgctgataa atctggagcc ggtgagcgtg      840
ggtctcggg tatcattgca gactggggc cagatggtaa gccctcccgt atcgtagtta      900
tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag      960
gtgcctcact gattaagcat tggttaactgt cagaccaagt ttactcatat atactttaga     1020

ttgatitaaa acttcatitt taatttataaa ggatctaggt gaagatcctt tttgataatc     1080
tcatgaccaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac cccgtagaaa     1140
agatcaaagg atcttcttga gatccttttt ttctgcgctt aatctgctgc ttgcaaacaa     1200
aaaaaccacc gtaccagcg gtggtttggt tgccggatca agagctacca actctttttc     1260
cgaaggtaac tggcttcagc agagcgcaga taccaatac tgtccttcta gtgtagccgt     1320
agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgtc ctgctaatcc     1380
tgttaccagt ggctgctgcc agtggcgata agtcgtgtct taccgggttg gactcaagac     1440

gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc acacagccca     1500
gcttggagcg aacgacctac accgaactga gatacctaca gcgtgagcta tgagaaagcg     1560

```



ccacgcttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg gtcggaacag 1620  
gagagcgcac gagggagctt ccagggggaa acgcctggta tctttatagt cctgtcgggt 1680  
ttcggcacct ctgactttag cgtcgatttt tgtgatgctc gtcagggggg cggagcctat 1740  
ggaaaaacgc cagcaacgcg gcctttttac ggttcctggc cttttgctgg ccttttgctc 1800  
acatgttctt tctgcgta tcccctgatt ctgtggataa ccgtattacc gcctttgagt 1860

gagctgatac cgctcggcg agccgaacga ccgagcgcag cgagtcagtg agcgaggaag 1920  
cggagaagcg cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt cattaatgca 1980  
ggatctcgat cccgcgaaat taatacgcact cactataggg agaccacaac ggtttccctc 2040  
tagaataat ttgttttaac ttaagaagg agatatacat atgaaacatc accaccatca 2100  
ccattcggac tcagaagtca atcaagaagc taagccagag gtcaagccag aagtcaagcc 2160  
tgagactcac atcaatttaa aggtgtccga tggatcttca gaaatcttct ttaagatcaa 2220  
aaagaccact cctttaagaa ggctgatgga agcgttcgct aaaagacagg gtaaggaaat 2280

ggactcctta agattcttgt acgacggtat tagaattcaa gctgatcaga cccctgaaga 2340  
tttggacatg gaggataacg atattattga ggctcacaga gaacagattg gtggcgccgc 2400  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2460  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2520  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2580  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2640  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2700

tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2760  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2820  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2880  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 2940  
tccagctgca cctgctccag cagcacctgc tgcaccagct cgggctgctc ctgctgccgc 3000  
aagagcaagc ttgatccggc tgctaacaaa gcccgaagg aagctgagtt ggctgctgcc 3060  
accgtgagc aataactagc ataaccctt ggggcctcta aacgggtctt gaggggtttt 3120

ttgctgaaag gaggaactat atccgatct ggcgtaatag cgaagaggcc cgcaccgatc 3180  
gcccctcca acagttgcgc agcctgaatg gcgaatggga cgcgccctgt agcggcgcac 3240  
taagcgcggc ggggtgtggtg gttacgcgca gcgtgaccgc tacacttgcc agcgccttag 3300  
cgcccctcc tttcgtttc ttccttctt ttctcgccac gttcggcggc tttccccgtc 3360  
aagctctaaa tcgggggctc cctttaggtt tccgatttag tgetttacgg cacctcgacc 3420

ccaaaaaact tgattaggt gatggtcac gtagtgggcc atcgccctga tagacggtt 3480  
 ttcgcccttt gacgttggag tccacgttct ttaatagtg actcttgttc caaactggaa 3540

caaacactcaa ccctatctcg gtctattctt ttgatttata agggattttg ccgatttcgg 3600  
 cctattggtt aaaaaatgag ctgatttaac aaaaatttaa cgcaatttt aacaaaatat 3660  
 taacgcttac aatttaggtg 3680

<210> 61

<211> 201

<212> PRT

<213> Artificial Sequence

<220><223> PA#1(200) polypeptide/polymer used for the preparation of drug  
 conjugates (made by ligation of 10 20mer encoding gene cassettes,  
 including one additional C-terminal Ala residue resulting from

the downstream ligation site)

<400> 61

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro  
 1 5 10 15

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala  
 20 25 30

Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro  
 35 40 45

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala  
 50 55 60

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala  
 65 70 75 80

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro  
 85 90 95

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala  
 100 105 110

Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro  
 115 120 125

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala

130

135

140

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala

145

150

155

160

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Pro

165

170

175

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala

180

185

190

Ala Pro Ala Pro Ala Ala Pro Ala Ala

195

200