



등록특허 10-2315246



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2021년10월21일  
(11) 등록번호 10-2315246  
(24) 등록일자 2021년10월14일

(51) 국제특허분류(Int. Cl.)  
*C12N 15/70* (2006.01) *C07K 1/18* (2006.01)  
*C07K 14/34* (2006.01) *C12N 1/08* (2006.01)  
*C12N 1/20* (2006.01) *C12P 21/00* (2006.01)

(52) CPC특허분류  
*C12N 15/70* (2013.01)  
*C07K 1/18* (2013.01)

(21) 출원번호 10-2016-7026716

(22) 출원일자(국제) 2015년03월02일

심사청구일자 2020년03월02일

(85) 번역문제출일자 2016년09월27일

(65) 공개번호 10-2016-0128362

(43) 공개일자 2016년11월07일

(86) 국제출원번호 PCT/US2015/018338

(87) 국제공개번호 WO 2015/134402

국제공개일자 2015년09월11일

(30) 우선권주장

61/947,234 2014년03월03일 미국(US)

(56) 선행기술조사문현

JP2011512796 A

(뒷면에 계속)

전체 청구항 수 : 총 23 항

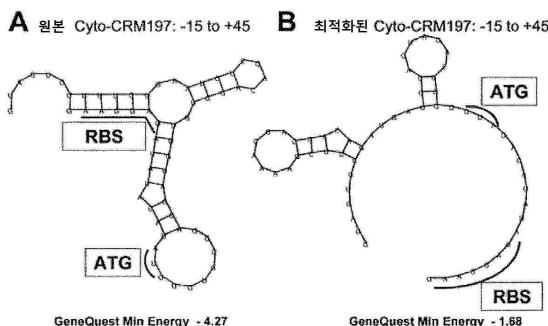
심사관 : 유성전

(54) 발명의 명칭 대장균에서 재조합 CRM197의 향상된 생산

**(57) 요 약**

재조합 CRM197 단백질을 인코딩하는 발현 벡터를 포함하는 감소된 계놈 또는 네이티브 K12 균주 대장균(*E. coli*) 박테리아 및 CRM197의 생산에서 그들의 사용이 제공된다. CRM197 단백질은 대장균 숙주의 주변 세포질에 대해 발현된 CRM197 단백질을 통제하는 신호 서열에 융합될 수 있다.

**대 표 도 - 도1**



(52) CPC특허분류

*C07K 14/34* (2013.01)  
*C12N 1/08* (2013.01)  
*C12N 1/20* (2021.05)  
*C12P 21/00* (2013.01)  
*C12N 2523/00* (2013.01)

(72) 발명자

**노비, 로버트, 이.**

미국, 위스콘신 53593, 베로나, 스프링 로즈 로드  
1717

**핸커, 테伦스, 엠.**

미국, 위스콘신 53711, 매디슨, 아파트 셀2더블유,  
먼로 스트리트 2901

**스테펜, 에릭, 에이.**

미국, 위스콘신 53572, 마운트 호롭, 대니 드라이  
브 1812

**블래트너, 프레드릭, 알.**

미국, 위스콘신 53715, 매디슨, 더블유. 올린 에비  
뉴 710

**최, 현식**

미국, 위스콘신 53719, 매디슨, 이스트 패스 7526

**포스파이, 기요르기**

헝가리, 에이치-6727 세게드, 마유스 1 유. 74

**랜드리, 카를로스, 에프.**

미국, 위스콘신 53711, 피치버그, 셀 106, 트와별  
톤 파이크 드라이브 3001

(56) 선행기술조사문헌

KR1020120099672 A  
WO2003070880 A2  
WO2007024756 A2  
JP2013507113 A

## 명세서

### 청구범위

#### 청구항 1

주변 세포질에 CRM197 단백질의 전이를 통제하는 OmpA, OmpF 또는 YtfQ 신호 서열에 융합된 CRM197 단백질을 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터를 포함하는 감소된 게놈 대장균을 성장시키는 것을 포함하며, 상기 뉴클레오티드 서열은 대장균에서 발현을 위해 최적화된 코돈이며 발현 제어 서열에 작동 가능하게 연쇄되며, 재조합 CRM197 단백질의 발현을 유도하는 것을 포함하는 감소된 게놈 대장균 숙주에서 재조합 CRM197을 생산하는 방법이며,

상기 감소된 게놈 대장균 숙주의 네이티브 모 대장균(*E. coli*) 균주는 *ilvG* 유전자의 네이티브-2 프레임시프트 변이를 포함하는 K-12 균주이고,

상기 감소된 게놈 대장균 숙주는 적어도 다음의 DNA 세그먼트: 대장균 K-12 균주 MG1655의 b0245-b0301, b0303-b0310, b1336-b1411, b4426-b4427, b2441-b2450, b2622-b2654, b2657-b2660, b4462, b1994-b2008, b4435, b3322-b3338, b2349-b2363, b1539-b1579, b4269-b4320, b2968-b2972, b2975-b2977, b2979-b2987, b4466-4468, b1137-b1172, b0537-b0565, b0016-b0022, b4412-b4413, b0577-b0582, b4415, b2389-b2390, b2392-b2395, b0358-b0368, b0370-b0380, b2856-b2863, b3042-b3048, b0656, b1325-b1333, b2030-b2062, b2190-b2192, b3215-b3219, b3504-b3505, b1070-b1083, b1878-b1894, b1917-b1950, b4324-b4342, b4345-b4358, b4486, b0497-b0502, b0700-b0706, b1456-b1462, b3481-b3484, b3592-b3596, b0981-b0988, b1021-b1029, b2080-b2096, b4438, b3440-b3445, b4451, b3556-b3558, b4455, b1786, b0150-b0153 및 b2945 또는 다른 대장균 K-12 균주의 상응하는 DNA 세그먼트가 결여됨으로써, 가용성 CRM197 리터당 적어도 1g의 수율이 수득되는, 방법.

#### 청구항 2

제 1항에 있어서,

상기 네이티브 모 대장균 균주는 *ilvG* 유전자의 네이티브-2 프레임시프트 변이를 포함하는 K-12 균주이며, 상기 감소된 게놈 대장균 숙주는 다음의 변형: (i) 오로테이트 포스포 리보실 전이효소 활성을 향상하는 *rph* 유전자의 제거 (ii) 활성 아세토하이드록시 산 신타아제 II 생산을 부활하기 위해 *ilvG* 유전자의 네이티브-2 프레임시프트 변이의 보정 및 (iii) *iclR* 및 *arpA* 유전자의 모두 또는 일부의 제거를 포함하는, 방법.

#### 청구항 3

제 1항 또는 제 2항에 있어서,

상기 네이티브 모 대장균 균주는 K-12 균주 MG1655인, 방법.

#### 청구항 4

제 1항 또는 제 2항에 있어서,

가용성 CRM197 리터당 적어도 2g의 수율이 수득되는, 방법.

#### 청구항 5

제 1항에 있어서,

상기 신호 서열은 OmpF 또는 YtfQ인, 방법.

#### 청구항 6

제 5항에 있어서,

상기 신호 서열은 YtfQ인, 방법.

### 청구항 7

제 1항 또는 제 2항에 있어서,

상기 감소된 게놈 대장균은 적어도 다음의 DNA 세그먼트: 대장균 K-12 균주 MG1655의 b0315-b0331, b0333-b0341, b0346-b0354, b2481-b2492, b2219-b2230, b4500, b3707-b3723, b0644-b0650, b4079-4090, b4487, b4092-b4106, b0730-b0732, b3572-b3587, b1653, b2735-b2740, b2405-b2407, b3896-b3900, b1202, b4263-b4268, b0611, b2364-b2366, b0839, b0488-b0500 및 b0502가 추가적으로 결여된, 방법.

### 청구항 8

제 1항 또는 제 2항에 있어서,

상기 감소된 게놈 대장균은 균주 MDS42metab인, 방법.

### 청구항 9

제 7항에 있어서,

상기 감소된 게놈 대장균은 균주 MDS69metab인, 방법.

### 청구항 10

제 1항 또는 제 2항에 있어서,

상기 감소된 게놈 대장균은 기능성 *recA*(b2699) 유전자를 포함하는, 방법.

### 청구항 11

제 1항 또는 제 2항에 있어서,

상기 감소된 게놈 대장균은 *relA* 유전자의 코딩 서열의 위치 547 또는 548에서 적어도 원 포인트 변이를 갖는 *relA* 유전자를 포함하며, 상기 변이는 위치 547에서 G → A 변이, 위치 547에서 G → T 변이, 위치 548에서 C → G 변이, 또는 위치 548에서 C → T 변이 중 하나 이상으로부터 선택되는, 방법.

### 청구항 12

제 1항 또는 제 2항에 있어서,

상기 감소된 게놈 대장균은 기능성 *polB*(b0060), 기능성 *dinB*(b0231) 및 기능성 *umuD*C(b1183-b1184) 유전자가 결여된, 방법.

### 청구항 13

제 1항 또는 제 2항에 있어서,

상기 감소된 게놈 대장균은 삽입 서열을 포함하지 않는, 방법.

### 청구항 14

제 1항 또는 제 2항에 있어서,

CRM197 뉴클레오티드 서열은 대장균 숙주 세포에서 발현을 위해 최적화된 코돈인, 방법.

### 청구항 15

삭제

### 청구항 16

제 1항 또는 제 2항에 있어서,

상기 방법은 다음:

(1) 감소된 게놈 대장균을 성장시키는 스텝 (a)는 상기 감소된 게놈 대장균을 19시간까지 동안 37°C에서 성장시

킨 후 CRM197의 발현을 유도하는 스텝 (b) 이전 및 다음에 20-30°C에서 성장이 뒤따르며;

(2) 스텝 (a) 및 (b)는 혈청, 효모 추출물 또는 다른 동물의 부산물을 포함하지 않는 증식 배지에서 수행되며;

(3) 스텝 (a)에서 pH가 6.5와 7.5 사이의 범위이며;

(4) pH는 인산 완충액, Tris 완충액 또는 히스티딘 완충액을 사용하여 유지되며;

(5) 발현은 IPTG의 적당량의 첨가에 의해 스텝 (b)에서 유도되며; 및/또는

(6) 세제의 부재 하에서 배양된 감소된 계놈 대장균 세포를 기계적으로 파쇄하고, 그 결과로 생성된 세포 용해물을 가용성 분획을 수득하기 위해 원심 분리하는 스텝 (c)를 더 포함하는 것 중 하나 이상을 포함하는, 방법.

### 청구항 17

제 16항에 있어서,

상기 방법은 발효기에서 이행되는, 방법.

### 청구항 18

제 16항에 있어서,

상기 완충액은 인산 완충액인, 방법.

### 청구항 19

제 16항에 있어서,

세제의 부재 하에서 배양된 감소된 계놈 대장균 세포를 기계적으로 파쇄하고, 그 결과로 생성된 세포 용해물을 가용성 분획을 수득하기 위해 원심 분리하는 추가의 스텝 (c)를 포함하며, CRM197은 소수성 상호 작용 크로마토그래피 및/또는 음이온 교환 크로마토그래피를 포함하는 하나 이상의 정제 스텝에 의해 상기 가용성 분획으로부터 정제되는, 방법.

### 청구항 20

제 1항 또는 제 2항에 있어서,

상기 감소된 계놈 대장균은 기능성 *recA*(b2699) 유전자가 결여된, 방법.

### 청구항 21

주변 세포질에 CRM197 단백질의 전이를 통제하는 *OmpA*, *OmpF* 또는 *YtfQ* 신호 서열에 융합된 CRM197 단백질을 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터를 포함하며, 상기 뉴클레오티드 서열은 발현 제어 서열에 작동 가능하게 연쇄되는, 발현 벡터를 포함하는 감소된 계놈 대장균 숙주이며,

상기 감소된 계놈 대장균 숙주의 네이티브 모 대장균(*E. coli*) 균주는 K-12 균주이고,

상기 감소된 계놈 대장균 숙주는 적어도 다음의 DNA 세그먼트: 대장균 K-12 균주 MG1655의 b0245-b0301, b0303-b0310, b1336-b1411, b4426-b4427, b2441-b2450, b2622-b2654, b2657-b2660, b4462, b1994-b2008, b4435, b3322-b3338, b2349-b2363, b1539-b1579, b4269-b4320, b2968-b2972, b2975-b2977, b2979-b2987, b4466-4468, b1137-b1172, b0537-b0565, b0016-b0022, b4412-b4413, b0577-b0582, b4415, b2389-b2390, b2392-b2395, b0358-b0368, b0370-b0380, b2856-b2863, b3042-b3048, b0656, b1325-b1333, b2030-b2062, b2190-b2192, b3215-b3219, b3504-b3505, b1070-b1083, b1878-b1894, b1917-b1950, b4324-b4342, b4345-b4358, b4486, b0497-b0502, b0700-b0706, b1456-b1462, b3481-b3484, b3592-b3596, b0981-b0988, b1021-b1029, b2080-b2096, b4438, b3440-b3445, b4451, b3556-b3558, b4455, b1786, b0150-b0153 및 b2945 또는 다른 대장균 K-12 균주의 상응하는 DNA 세그먼트의 제거를 포함하는, 발현 벡터를 포함하는 감소된 계놈 대장균 숙주.

### 청구항 22

제 21항에 있어서,

상기 네이티브 모 대장균 균주는 *iIVG* 유전자의 네이티브-2 프레임시프트 변이를 포함하는 K-12 균주이며, 상기 감소된 게놈 대장균 숙주는 다음의 변형: (i) 오로테이트 포스포 리보실 전이효소 활성을 향상하는 *rph* 유전자의 제거 (ii) 활성 아세토하이드록시 산 신타아제 II 생산을 부활하기 위해 *iIVG* 유전자의 네이티브-2 프레임시프트 변이의 보정 및 (iii) *iclR* 및 *arpA* 유전자의 모두 또는 일부의 제거를 포함하는, 발현 벡터를 포함하는 감소된 게놈 대장균 숙주.

#### 청구항 23

제 2항에 있어서,

상기 *iIVG* 유전자의 네이티브-2 프레임시프트 변이의 보정은 위치 982와 983 사이에서 두 개의 뉴클레오티드의 삽입인, 방법.

#### 청구항 24

제 22항에 있어서,

상기 *iIVG* 유전자의 네이티브-2 프레임시프트 변이의 보정은 위치 982와 983 사이에서 두 개의 뉴클레오티드의 삽입인, 발현 벡터를 포함하는 감소된 게놈 대장균 숙주.

#### 청구항 25

삭제

#### 청구항 26

삭제

#### 청구항 27

삭제

#### 청구항 28

삭제

#### 청구항 29

삭제

#### 청구항 30

삭제

#### 청구항 31

삭제

#### 청구항 32

삭제

#### 청구항 33

삭제

#### 청구항 34

삭제

#### 청구항 35

삭제

청구항 36

삭제

### 발명의 설명

#### 기술 분야

##### 관련출원에 대한 상호 참조

[0001] 본 출원은 2014년 3월 3일자로 출원된 미국 출원 번호 61/947,234의 이익을 주장하며, 그 내용을 본 명세서에 참고로 포함한다.

##### 발명의 분야

[0004] 본 발명은 대장균(*E. coli*), 바람직하게는 감소된 계놈 대장균 K12 균주에서 재조합 CRM197의 생산에 관한 것이다.

### 배경 기술

##### 발명의 배경

[0006] 디프테리아 독소(DTx)는 A(활성) 도메인과 이황화물 교상결합에 의해 서로 연쇄된 B(결합) 도메인을 함유하는 535 아미노산의 단일 폴리펩티드 사슬로 합성된 코리네박테리움 디프테리아의 이성분(二成分) 외(外)독소이다. 독소는 세포 수용체(HB-EGF 수용체)에 결합되어 A 도메인이 단백질 분해 분열에 의하여 B 영역으로부터 방출되는 세포내 섭취에 의해 세포로 들어간다. 그 후 A 도메인은 B 도메인에 의해 이루어진 기공을 통해 엔도솜을 빠져나와 단백질 합성을 억제하여 궁극적으로 세포 사멸을 초래하는 세포질로 들어간다.

[0007] CRM197은 효소같이 비활성 및 비-독성 단백질을 제공하는 글리신(G52E)에 대한 글루탐산의 단일 아미노산 치환을 포함하는 Dtx의 변이 형태이다. CRM197은 피포성의(encapsulated) 박테리아에 비해 결합 백신에 대해 이상적인 탑체(carrier)인 것으로 밝혀졌다. 결합 백신은, 열약한 면역원성 및 T 세포 독립적인 협막다당에 공유 결합으로 연쇄된 CRM197을 포함하여, 높은 면역원성이며 항원에 대해 오래 지속하는 면역력을 초래하는 결합 항원을 생성한다.

[0008] 운반 단백질로 CRM197을 포함하는 백신은 성공적으로 수백만의 아이들을 면역하는데 사용했으며, Menveo® , 수막염균의 혈청균 A-C-W135-Y에 대비하여 4가의 결합 백신, Menjugate® 및 Meningitec®(수막염균의 혈청형 C에 대비하여), Vaxem-Hib® 및 HibTITER®(헤모필루스 인플루엔자 B형, Hib에 대비하여), 및 다 가의 폐렴 구균성 복합제 Prevnar™ 를 포함한다.

[0009] 파상풍 및 디프테리아 독소와 달리, CRM197은 화학 해독을 필요로 하지 않기 때문에, 균질성으로 정제될 수 있으며 접합을 위해 직접적으로 사용될 수 있다. CRM197은  $\beta$  파지의 다수의 리소젠 또는 슈도모나스 형광성 (*pseudomonas fluorescens*)의 플라스미드 시스템으로부터 발현된 코리네박테리움 디프테리아 C7의 발효에 의해 현재 제조된다. CRM197(C. 디프테리아 발효하는 동안 배양 배지로 방출되는)의 수율은 수십 mg/l ~ 200mg/l 이르기까지 낮은 범위이고 생물학적 안정성 레벨 2 시설을 요구하기 때문에, CRM197 밀리그램당 약 미국 \$500의 소매가격을 초래한다. 백신의 단일 투여량은 일반적으로 약 10 및 60 $\mu$ g의 CRM197을 포함하며 1억 5천만 이상 용량이 매년 사용된다. 결합 CRM197 백신에 대한 현재 수요가 공급을 능가하여, 어린이 수백만의 건강이 위험에 처하고 있는 개발도상국의 예방 접종 프로그램을 차수해 지연 가져왔다.

[0010] 또한, 난소암과 같은 암 치료에 CRM197에 대한 가능한 치료 용도는, 일부 암에서 높게 발현되는 혜파린-결합 표피 성장 인자(pro-HB-EGF)의 가용성 형태를 결합하는 CRM197의 능력을 기반으로 최근 보고되었다. 이 치료 가능성에 대한 연구 및 개발은 현재의 생산 방법에서 더욱 많은 균주에 중점을 두고 있다.

[0011] CRM197의 높은 가격과 공급 부족에 기여하는 하나의 큰 요인은 주력 대장균의 생산에 있어 다양한 CRM197을 생성하는 역사적 능력이 없다는 것이다. CRM197의 불용성 형태가 대장균에서 상대적으로 적당한 수율로 발효될 수 있지만, 불용성 생성물의 일부분만 가용성 형태로 전환될 수 있다(Stefan 외., 2011). 대장균에서 다양한의 가용성 CRM197을 생산하는 것은 더욱 어려운 도전이었다. 확실하고 저렴하게 치료 용도에 대한 많은 양의 CRM197을 생산하는 방법은 당 업계에서 상당한 진보라 여겨질 것이다.

## 발명의 내용

### 해결하려는 과제

### 과제의 해결 수단

#### 발명의 요약

본 발명은, 대장균 숙주 세포에서 재조합 CRM197 단백질의 생산 방법에 관한 것이다. 몇몇 실시예에서, 상기 방법은, 재조합 CRM197 단백질의 발현에 적합한 조건하에서 발현 제어 서열에 작동가능하게 연쇄된 CRM197 단백질을 인코딩하는 핵산 서열을 포함하는 발현 벡터를 포함하는 감소된 계놈 대장균을 배양하는 것을 포함한다. CRM197의 수율의 상당한 증가는 BL21과 같은 야생형 대장균 균주에서 생산에 비해 본 발명에 따른 감소된 계놈 대장균 숙주 세포에서 달성된다. CRM197 단백질을 인코딩하는 핵산 서열은 바람직하게는 대장균 숙주 세포에서 발현을 위해 최적화된 코돈이다. 바람직한 실시예에서, 네이티브 모 대장균 균주는 K12 균주이다. 다른 실시예에서, 상기 방법은 재조합 CRM197 단백질의 발현에 적합한 조건하에서 발현 제어 서열에 작동 가능하게 연쇄된 CRM197 단백질을 인코딩하는 핵산 서열을 포함하는 발현 벡터를 포함하는 네이티브 K12 균주 대장균을 배양하는 것을 포함한다.

하나의 양태에서, CRM197 단백질을 인코딩하는 핵산 서열은 대장균 숙주 세포(바람직하게는 감소된 계놈 대장균 숙주 세포)의 주변 세포질에 CRM197 단백질의 전이를 통제하는 신호 서열을 인코딩하는 핵산 서열과 융합하여, 가용성 CRM197 리터당 약 1g 내지 리터당 10g의 수율이 달성된다. 본 발명의 이러한 양태에 따르면, 대장균 숙주(바람직하게는 감소된 계놈 대장균 숙주)는 대장균 주변 세포질에 CRM197의 전이를 통제할 수 있는 아미노산 서열을 갖는 폴리펩티드를 인코딩하는 5' 신호 서열 부분 및 그의 네이티브 신호 서열을 결여한 CRM197 단백질을 인코딩하는 3' CRM197 부분을 포함하는 핵산 서열을 포함하는 발현 벡터를 포함한다. 바람직하게는, CRM197의 발현이 유도 가능하여 상기 방법은 (a) 대장균(바람직하게는 감소된 계놈 대장균)을 성장시키고 (b) CRM197의 발현을 유도하는 스텝을 포함한다. 바람직하게는, 상기 방법은 발효기에서 수행된다.

관련 양태에서, (예를 들어, 감소된 계놈)대장균 숙주 세포는 단백질 코딩 서열에 작동 가능하게 연쇄된 유도성 프로모터(예를 들어, 락크 파생의 프로모터)를 포함하는 발현 벡터에 의해 형질 전환되며, CRM197의 발현은 적당량의 인듀서(예를 들어, 이소프로필  $\beta$ -D-1-티오갈락토파라노시드(IPTG))를 추가함으로써 유도된다. 바람직하게는, 진탕 플라스크 조건하에서, 유도는 약 0.1 내지 약 1.5(더욱 바람직하게는 약 0.2 내지 약 0.9, 더욱더 바람직하게는 약 0.3 내지 약 0.6)의 600nm(파장 1 OD 유닛은 약  $0.8 \times 10^9$  세포/ml에 대응하는 것으로)에 광학 밀도(OD)로 발생한다. 발효 조건하에서 유도는 바람직하게는 약 100 내지 400, 더 바람직하게는 약 150 내지 300, 가장 바람직하게는 200 내지 275 사이(예를 들어, 230 및 250)의 OD600로 발생한다. 다른 관련 양태에서, 성장 및/또는 유도시 배양의 pH는 약 6.5 내지 약 7.5이고, 성장 및/또는 유도 온도는 약 20°C 내지 약 30°C(바람직하게는 약 25°C)이며 성장 배지는 혈청, 효소 추출물 및 동물성 부산물이 없다. 특히 바람직한 실시예에서, (예를 들어, 감소된 계놈)대장균의 성장은 37°C(예를 들어, 1 내지 3시간)에서 상대적으로 짧은 초기 배양을 하고 나서 유도 이전과 이후에 20°C 내지 30°C(바람직하게는 25°C에서)에서 성장이 뒤따르는 것을 포함하거나 유도 이전과 이후에 20°C 내지 30°C(바람직하게는 25°C에서)에서 지속적인 성장을 포함한다.

관련 실시예에서, 수득된 가용성 CRM197의 수율은 적어도 약 0.5g/L, 적어도 약 0.7g/L, 적어도 약 1.0g/L, 적어도 약 1.3g/L, 적어도 약 1.5g/L, 적어도 약 1.7g/L, 적어도 약 2.0g/L, 적어도 약 2.3g/L, 적어도 약 2.5g/L, 적어도 약 2.7g/L, 적어도 약 3.0g/L, 적어도 약 3.5g/L, 적어도 약 3.7g/L, 적어도 약 4.0g/L, 적어도 약 4.5g/L, 적어도 약 5g/L, 적어도 약 5.5g/L, 적어도 약 6.0g/L, 적어도 약 7.0g/L, 적어도 약 8.0g/L, 적어도 약 9.0g/L 또는 적어도 약 10.0g/L이다. 다른 실시예에서, 수득된 가용성 CRM197의 수율은 약 1.0g/L 내지 약 10.0g/L, 약 1.0g/L 내지 약 9.0g/L, 약 1.0g/L 내지 약 8.0g/L, 약 1.0g/L 내지 약 7.0g/L, 약 1.0g/L 내지 약 6.0g/L, 약 1.0g/L 내지 약 5.0g/L, 약 1.0g/L 내지 약 4.0g/L, 약 1.0g/L 내지 약 3.0g/L 또는 약 1.0g/L 내지 약 2.0g/L이다. 다른 실시예에서, 수득된 가용성 CRM197의 수율은 약 2.0g/L 내지 약 10.0g/L, 약 2.0g/L 내지 약 9.0g/L, 약 2.0g/L 내지 약 8.0g/L, 약 2.0g/L 내지 약 7.0g/L, 약 2.0g/L 내지 약 6.0g/L, 약 2.0g/L 내지 약 5.0g/L, 약 2.0g/L 내지 약 4.0g/L, 약 2.0g/L 내지 약 4.0g/L 또는 약 2.0g/L 내지 약 3.0g/L이다. 다른 실시예에서, 수득된 가용성 CRM197의 수율은 약 3.0g/L 내지 약 10.0g/L, 약 3.0g/L 내지 약 9.0g/L, 약 3.0g/L 내지 약 8.0g/L, 약 3.0g/L 내지 약 7.0g/L, 약 3.0g/L 내지 약 6.0g/L, 약 3.0g/L 내지 약 5.0g/L, 약 3.0g/L 내지 약 4.0g/L, 약 3.0g/L 내지 약 4.0g/L 또는 약 3.0g/L 내지 약 3.0g/L이다.

5.0g/L 또는 약 3.0g/L 내지 약 4.0g/L이다.

[0017] 관련 양태에서, 5' 신호 서열 부분은 DsbA, TolB 및 TorT 분비 신호와 같은 신호 인식 입자(SRP) 의존성 신호 서열, OmpF, OmpT, OmpA, PhoA, MalE, LamB, LivK 및 PelB 분비 신호와 같은 Sec-의존성 신호 서열, 또는 TorA 및 Suf1 분비 신호와 같은 트윈 아르기닌 친위(TAT) 신호 서열을 인코딩한다. 일부 실시예에서, 5' 신호 서열 부분은 Sec-의존성 신호 서열, 바람직하게는 OmpA 또는 OmpF 분비 신호를 인코드한다. 특히 바람직한 실시예에서, 5' 신호 서열 부분은 ompF 분비 신호를 인코드한다.

[0018] 다른 바람직한 실시예에서, 5' 신호 서열 부분은 MgIB, MalE, OppA, RbsB, Agp, FkpA, YtfQ, HdeA, HdeB, OmpC 및 GlnH 분비 신호로부터 선택되는 신호 서열을 인코드한다. 특히 바람직한 실시예에서, 5' 신호 서열 부분은 YtfQ 분비 신호를 인코드한다.

[0019] 또 다른 관련 양태에서, 대장균 숙주 세포는 발현 제어 서열에 작동 가능하게 연쇄된 주변 세포질에서 CRM197의 전위 및/또는 폴딩을 보조하기 위한 하나 이상의 단백질을 인코딩하는 서열을 포함하는 하나 이상의 핵산을 추가로 포함한다. 주변 세포질에서 CRM197의 전위 및/또는 폴딩을 보조하기 위한 하나 이상의 단백질을 인코딩하는 서열을 포함하는 핵산은 CRM197을 인코딩하는 뉴클레오티드 서열을 포함하는 동일한 발현 벡터의 일부분일 수 있거나 다른 발현 벡터에 위치될 수 있다. CRM197의 전위 및/또는 폴딩을 보조하기 위한 단백질은 Skp, DnaK, DnaJ, Caf1M 및 Caf1A와 같은 샤프론; DsbA, DsbB, DsbC 및 DsbD과 같은 이황화 결합 형성 단백질; PpiA, PpiD, FkpA 및 SurA와 같은 펩티딜-프로릴 시스-트랜스 이성화 효소; MBP, GST 및 티오래독신과 같은 가용성 파트너 단백질; YebF, MalE, HlyA, 히루딘, OmpF 및 Spy와 같은 분비 경로 단백질; YccA와 같은 단백질 분해 효소 억제제; 및 PspA와 같은 수출 포화(export saturation)를 완화하는 단백질을 포함하나 이에 제한하지 않는다.

[0020] 다른 실시예에서, CRM197 단백질을 인코딩하는 뉴클레오티드 서열은 신호 서열에 융합되지 않으며, 이로 인해 리터당 약 2g 내지 리터당 약 20g의 불용성 CRM197의 수율이 달성된다. 본 발명의 이런 양태에 의하면, (예를 들어, 감소된 게놈)대장균 숙주는 그의 네이티브 신호 서열을 결여한 CRM197 단백질을 인코딩하는 핵산 서열을 포함하는 발현 벡터를 포함하며, 이로 인해 CRM197 단백질은 대장균 숙주의 세포질에서 발현된다.

[0021] 몇몇 양태에서, 본 발명은 (예를 들어, 감소된 게놈)대장균 숙주 세포에서 재조합 CRM197 단백질을 생산하는 방법에 관한 것으로, 상기 방법은, 주변 세포질에 CRM197 단백질의 전이를 통제하는 신호 서열에 융합된 CRM197 단백질을 인코딩하는 뉴클레오티드 서열을 발현 벡터에 결찰; 발현 벡터에 의해 대장균 숙주 세포를 형질 전환; 및 재조합 CRM197 단백질의 발현에 적합한 배양 배지에서 형질 전환된 대장균 숙주 세포를 배양;을 포함하며, 가용성 CRM197의 수율은 약 1 내지 10g/L, 바람직하게는 약 2 내지 10g/L이며, 또한 배양으로부터 대장균 세포를 수확하고, 세제의 부재 하에 기계적 방법으로 수확된 세포를 용해시키는 것을 추가로 포함한다. 선택적으로, 상기 방법은 그 결과로 발생된 용해물(예를 들어, 가용성 및 불용성 분획을 단리하기 위해 원심 분리에 의해)의 가용성 분획을 수득하고 그 가용성 분획(대장균 숙주에 의해 생산된 가용성 CRM197을 포함하는)이 하나 이상의 정제 스텝을 거치도록 하는 것을 추가로 포함한다. 일 실시예에서, 가용성 CRM197은 소수성 상호 작용 크로마토그래피 및/또는 음이온 교환 크로마토그래피를 행하였다. 바람직한 실시예에서, 대장균 숙주 세포는 감소된 게놈 대장균 숙주 세포이다.

[0022] 다른 양태에서, 본 발명은 발현 벡터를 포함하는 (예를 들어, 감소된 게놈)대장균 숙주 세포에 관한 것이며, 상기 발현 벡터는 대장균 주변 세포질에 CRM197의 전이를 통제할 수 있는 아미노산 서열을 갖는 폴리펩티드를 인코딩하는 5' 신호 서열 부분 및 발현 제어 서열에 작동 가능하게 연쇄된 그의 네이티브 신호 서열을 결여한 CRM197 단백질을 인코딩하는 3' CRM197 부분을 포함하는 핵산 서열을 포함한다. 바람직한 실시예에서, 대장균 숙주 세포는 감소된 게놈 대장균 균주 MDS42에서 제거된 유전자를 적어도 결여한 또는 감소된 게놈 대장균 균주 MDS69에서 제거된 유전자를 적어도 결여한 감소된 게놈 대장균 숙주 세포이다.

[0023] 본 발명의 이들 및 다른 실시예들은 이하의 본 명세서에서 더 상세히 기술된다.

## 도면의 간단한 설명

### 도면의 간단한 설명

도 1은 CRM197의 불용성 형태를 검사 대상으로 실험에 사용된 CRM197 서열에서 헤어핀 구조의 방출을 초래하는 DNA 서열 변화를 도시한다. 최적화된 서열(B)은 높은 최소한의 에너지(원본 서열에 대한 -4.27에 비해 최적화된 서열에 대한 -1.68)를 생성하고 원본 서열에 비해 시작 부위(ATG)와 리보솜 결합 부위(RBS) 모두의 인식을 향상

시키는 이차 구조를 인도(relax) 한다.

도 2는 감소된 게놈 대장균 균주 MDS42 $recA$ ( $recA$  제거를 갖는 MDS42 균주)에서 CRM197의 세포질 발현을 도시한다. 진탕 플라스크 배양은 0.5의 광학 밀도(OD)에 최소 배양 배지에서 배양한 다음 표시된 바와 같이 0 또는 250  $\mu$ M IPTG로 유도 하였다. 4-12% 구배 아크릴아미드 Bis-Tris 겔로 전기 영동한 후 GelCode 염색 시약을 사용하여 단백질을 염색했다. 레인당 0.04 OD의 물질을 적재했다. M=마커 레인; T=전체 세포 단백질; S=가용 분획; I=불용성 분획. 다량의 세포 기질(cytosolic) CRM197은 불용성 CRM197의 생산에서 구조 및 균주의 장인한 특성을 검증하는 불용성 분획(화살표)에 존재했다.

도 3은 감소된 게놈 균주 MDS42 $recA$ 에서 CRM197의 주변 세포질 전달에 관하여 검사된 신호 서열을 도시한다. 상단 패널(A)은 CRM197 서열을 최적화한 코돈의 5' 말단에 융합된 이종 신호 서열을 예시한다. 하부 왼쪽 패널(B)은 대표적인 세 개의 대장균 분비 경로의 신호 서열의 예시이다. 하부 오른쪽 패널(C)은 주변 세포질에서 CRM197의 전위 및/또는 폴딩을 보조하는 능력을 시험하기 위해 CRM197와 공동 발현된 단백질을 예시한다.

도 4는 OmpA 및 OmpF 신호 서열을 사용하는 MDS42 $recA$ 에서 CRM197 발현의 단백질 겔 분석을 도시한다. 샤프론으로  $\Delta$ aa1-5Ycca(OmpA+CRM197+/-Ycca) 또는  $\Delta$ aa1-5Ycca(OmpF+CRM197+/-Ycca)을 사용하여 25°C에서 조기(패널 A와 C; 0.01의 OD<sub>600</sub>으로 추가된 인듀서) 및 후기(패널 B와 D; 약 0.4의 OD<sub>600</sub>으로 추가된 인듀서) 유도. 화살표는 각 인듀서 시리즈에서 CRM197의 최고 수준을 나타낸다. CRM 197 바로 아래에 이동한 내인성 대장균 단백질을 주목하라. 겔 방법은 도 2에 대해 기술된 바와 같다. 주변 세포질 샘플은 원형질막 완충액(Periplastin Buffer)(Epicentre, Madison, WI)의 도움으로 조제하였다. 2 OD 샘플은 1.5ml 애펜 도르프 튜브에서 10 분 동안 7,500xg으로 원심 분리에 의해 수화하였다. 상청액을 제거하고, 세포 펠릿을 50  $\mu$ l 원형질막 완충액(200mM Tris-HCl[pH 7.5, 20% 자당, 1mM EDTA, 및 30U/ $\mu$ l Ready-Lyse 리소자임)에서 조심스럽게 재 혼탁하였다. 실온에서 5분 후, 얼음 냉수 50  $\mu$ l를 재 혼탁된 펠릿에 빠르게 첨가하였다. 혼합물을 15분 동안 4000xg으로 원심 분리하여 스페로플라스트로부터 주변 세포질 분획을 분획하기 전에 5분 동안 얼음 위에서 배양하였다. 주변 세포질 분획을 나타내는 상청액을 SDS-PAGE 분석을 위해 제조하였다. 0.12 OD 단위에 상당하는 양을 적재하였다.

도 5는 MDS42 $recA$  숙주 세포에서 CRM197 발현에 탄소원으로서 포도당 또는 글리세롤을 사용하는 효과를 비교한다. 패널 A는 전체 세포 단백질(레인당 0.04 OD 적재)을 도시한다. 패널 A는 단리된 주변 세포질 단백질(레인당 0.12 OD 적재)을 도시한다. CRM197의 높은 수준은 포도당-상보 배양 배지에서 생성됨이 주목된다. “42 만(only)”으로 표시된 레인은 발현 벡터가 없는 MDS42 $recA$ (즉, CRM197을 발현시키지 않는)를 갖는 제어 레인이다.

도 6은 MDS42 플랫폼에 내장된 네 개의 균주 및 MDS69 플랫폼에 내장된 네 개의 균주를 포함하는 ompA 신호 서열에 융합된 CRM197에 대해 코딩하는 발현 벡터를 지닌 여덟 개의 감소된 게놈 대장균 균주에서 CRM197 발현의 단백질 겔 분석을 도시한다. 시험된 MDS42 균주는 (i) MDS42, (ii) MDS42 $recA$ , (iii) MDS42 metab(*rph* 및 *ilvG* 프레임시프트 변이의 보정과 *iclR* 및 *arpA* 유전자의 제거를 포함하는 MDS42 균주) 및 (iv) MDS42 Blon metab(IS 삽입이 원형 대장균 *lon* 프로모터의 -10 영역으로부터 -35 영역을 단리시키는, B 균주 대장균의 *lon* 프로모터 영역의 서열을 모방하는 론 단백질 분해효소(b0439) 프로모터 영역의 변형을 추가로 포함하는 MDS42metb)이다. 시험된 MDS69 균주는 (i) MDS69metab(*rph* 및 *ilvG* 프레임시프트 변이의 보정 및 *iclR* 및 *arpA* 유전자의 제거를 포함하는 MDS69 균주) (ii) MDS69 Blon metab(상술한 론 단백질 분해효소 프로모터 변형을 추가로 포함하는 MDS69metab) (iii) MDS69 lpp metab(lpp 유전자(MG1655의 뉴클레오티드 1755260-1755687)의 제거를 추가로 포함하는 MDS69metab 및 (iv) MDS69 Blon, lpp metab이다. 패널 A 및 C는 표시된 IPTG 농도로 유도 후 단리된 전체 세포 단백질을 도시한다. 패널 B 및 D는 병렬로 수행된 주변 세포질 단리를 나타낸다. 겔 방법은 도 2에 기술된 바와 같다.

도 7은 MDS42 및 MDS69 단백질 분해효소 균주에서 CRM197 발현의 단백질 겔 분석을 도시한다. 패널 A 및 C는 표시된 IPTG 농도로 유도 후 단리된 전체 세포 단백질을 도시한다. 패널 B 및 D는 병렬로 수행된 주변 세포질 분리를 도시한다. 겔 방법은 도 2에 기술된 바와 같다.

도 8은 10리터 규모로 정의된 최소 배양 배지에서 OmpA 신호 서열에 융합된 CRM197에 대해 코딩하는 발현 벡터를 지닌 감소된 게놈 대장균 균주 MDS42metab의 유가(fed-batch) 발효가 뒤따르는 단백질 겔 및 웨스턴 블로트(Western blot)의 발효 샘플을 도시한다. 패널 A: 전체 세포 단백질(TCP) 및 주변 세포질(Peri) 제제는 표시된 발효 ODs에서 수집했다. IPTG는 OD=200으로 첨가됐다. 패널 B: 항-디프테리아 독소 항체로서 웨스턴 블로팅은 결정적으로 CRM197을 식별하는 데 사용했다. CRM197은 유도 전에 발현되지 않았음이 주목된다. SFC= 진탕 플라스크 제어. TCP 및 Peri 샘플 각각에 대해 레인당 0.04 및 0.12 OD를 각각 적재했다. 겔 방법은 도 2에서 기술

된 바와 같다.

도 9는 (i)종래 세제-기반 완충액(페널 A와 페널 B의 왼쪽 세 번째 레인)을 사용하여 단리된 전체 세포 단백질(TCP) 및 (ii) MDS42recA 균주에서 OmpA-CRM197의 발현이 뒤따르는, 주변 세포질 단백질 제제(페널 B의 오른쪽 세 번째 레인)의 단백질 겔을 도시한다. 페널 A: 전체 세포 단백질의 샘플(왼쪽)은 고속(21K g)으로 10분 동안 원심 분리한 후 재분석(오른쪽). CRM197은 가용성 분획에서 확인되지 않았다. 페널 B: CRM197이 세제에 노출된 세포질 분획(왼쪽)에서 불용성이며 세제에 노출되지 않은 주변 세포질 분획(오른쪽)에서 가용성임을 나타내는 전체 세포 단백질(TCP) 및 주변 세포질 분획의 고속 원심 분리. 겔 방법은 도 2에 기술된 바와 같다. 화살표는 가용성 분획에서 나타나지 않는 TCP에서 CRM197을 나타낸다.

도 10은 세제 및 기계적 용해의 단백질 겔 및 CRM197 용해성을 도시한다. CRM197에 융합된 ompA를 인코딩하는 발현 벡터를 지닌 MDS42recA 세포를 도 8에 대해 기술된 바와 같이 유가 발효를 실시했다. 세포는 (A) 세제(Bugbuster®(Novagen) 세포막을 분해하는 비-이온성 세제의 전매 혼합물) 또는 (B) 가용화제의 존재하에 기계적(초음파) 용해를 사용하여 용해시켰다. 세제가 없을 때의 용해가 높은 수준의 가용성 CRM197의 결과를 가져옴을 주목하라. GSH:GSSG=글루타티온의 산화 비율에 따라 감소됨; M=마커; Sol=가용성 분획; TCP=전체 세포 단백질. 인듀서(IPTG)=35 μM.

도 11은 OmpA-CRM197을 인코딩하는 발현 벡터를 지닌 MDS69metab 숙주 세포의 발효를 도시하는 그래프이다. CRM197은 1.95g/L의 최대 수율을 달성하는 30시간 시점까지의 유가 발효에서 증가하는 것을 확인하였다.

도 12는 발효 샘플에서 많은 양의 가용성 CRM197을 나타내는 도 11의 발효로 부터 단백질 겔을 도시한다. 균주 MDS69metab를 탄소원으로서 포도당을 갖는 최소 배양 배지에서 유가 발효를 행했다. 인듀서(22 시간) 첨가 이전 또는 28 또는 29시간에 수집된 샘플은 초음파 및 단리된 다른 세포 분획에 의해 균질화 했다. CRM197은 가용성(Sol) 분획에서 독점적으로 발견되었다. TCP=전체 세포 단백질 분획; Insol=불용성 분획.

도 13은 CRM197의 2 칼럼 정제의 결과를 도시한다. 페널 A: 도 12에 도시된 28시간 발효 샘플의 50 OD는 마이크로 유동기(MF)로 균질화를 행했다. 가용성 및 불용성(IS) 분획을 단리하고 가용성 물질(25 OD당량)을 페널 세파로스 칼럼 상에 적재했다. “가용성”으로 레이블된 세 개의 레인은 각각 적재 전 가용성 샘플의 0.1, 0.07 및 0.04 OD이다. CRM197-함유 물질은 10mM 염화나트륨, 10mM 인산 완충액, pH7.5로 용출시켰다. 5개의 연속 2.5ml의 샘플을 수집(왼쪽에서 오른쪽)하고 원으로 표시된 분획은 풀링(pool)했다. 주요 용출된 샘플과 떨어져 정제되고 단지 증류수를 사용하여 뒤따르는 용출이 확인되는 적은 양의 미처리된 CRM197(화살표)를 주목하라. 페널 B: 페널 A에서 풀링된 샘플을 DEAE 세파로스 칼럼 상에 적재하고 다른 염 농도(0.1M 염화나트륨, 뒤이어 1M 염화나트륨 및 최종 1.5M 염화나트륨)를 사용하여 용출하였다. 높은 순도의 CRM197은 1M 염화나트륨에서 용출되었다. 항-디프테리아 독소 항체(1:1000 희석) 웨스턴 블로트는 병렬로 실행된 단백질-염색된 겔과 함께 도시된다. F, 흐름 통과; W, 칼럼 세척; MF=마이크로 유동화가 뒤따르는 전체 균질화; IS=마이크로 유동화 균질화의 원심 분리가 뒤따르는 불용성 분획; S=마이크로 유동화 균질화의 원심 분리가 뒤따르는 가용성 분획.

도 14는 야생형 B균주 BLR(DE3)에 비해 MDS42recA에서 단백질 겔 분석을 도시한다. 세포는 주변 세포질에 단백질을 통제하는 OmpA-CRM197 융합을 인코딩하는 발현 벡터로 형질 전환하였고 진탕 플라스크에서 19시간 동안 25°C에서 성장하였다. CRM197의 발현은 표시된 농도에서 IPTG의 첨가에 의해 OD=0.3(후기)으로 유도하였다. 주변 세포질 단백질은 단리하고 분석하였다. CRM197 바로 아래로 전위한 내성 대장균 단백질을 주목된다.

도 15는 MDS69metab 및 MDS69metab low 변이 속도 숙주에서 pSX2-OmpA CRM197 및 pSX2-OmpF CRM197 발현에서 발생되는 주변 세포질 분획의 단백질 겔 분석을 도시한다. OmpF-CRM197 클론은 동일 균주에서 동일 유도 조건을 사용하는 OmpA-CRM197 클론(레인 8-9) 보다 MDS69metab low mut 숙주(레인 5-6)에서 가용성 주변 세포질 CRM197을 더 많이 생산한다. 레인 11-15는 세포 수확 후 배양 배지 상청액의 샘플을 나타내고 이러한 유도 및 성장 조건하에서 배양 배지에 방출되는 CRM197은 거의 조금도 없음을 보여줍니다.

도 16a-16c는 CRM197 DNA와 아미노산 서열을 도시한다. 도 16a는 주변 세포질 공간에서 성숙 CRM197 단백질의 발현 및 처리를 통제하는 신호 서열 갖는 발현에 대해 최적화된 CRM197 ORF의 DNA 서열을 도시한다. 도 16b는 세포질에서 발현에 대해 최적화된 CRM197 ORF의 DNA 서열을 도시한다. 도 16c는 신호 서열 처리(도 16a) 또는 N-말단 메티오닌 제거(도 16b) 후 생산된 성숙 CRM197 단백질의 아미노산 서열을 도시한다.

도 17은 감소된 계놈 균주 MDS69metab에서 CRM197의 주변 세포질 전달에 관하여 검사되고, 대장균 K 및/또는 B 균주에서 그들의 상대적 존재비(abundance)에 따라 분류된 신호 서열을 도시한다.

도 18은 다음의 신호 서열: OmpF, MalE, HdeA, OppA, HdeB 및 GlnH(유도 A) 및 감소된 계놈 대장균 균주

MDS69metab에서 OD~0.3으로 첨가된 25  $\mu$ M 인듀서(IPTG)로 OmpF, Mg1B, Agp, OmpC, RbsB, FkpA, 및 YtfQ(유도 B)을 갖는 주변 세포질 CRM197(캘리퍼스 분석)의  $\mu$ g/OD(상단 패널) 및  $\mu$ g/ml을 도시한다.

도 19는 다음의 신호 서열: OmpF, MalE, HdeA, OppA, HdeB 및 GlnH(유도 A) 및 감소된 계놈 대장균 균주 MDS69metab에 OD~0.3으로 첨가된 35  $\mu$ M 인듀서(IPTG)로 OmpF, Mg1B, Agp, OmpC, RbsB, FkpA, 및 YtfQ(유도 B)을 갖는 주변 세포질 CRM197(캘리퍼스 분석)의  $\mu$ g/OD(상단 패널) 및  $\mu$ g/ml을 도시한다.

도 20은 다음의 신호 서열: OmpF, MalE, HdeA, OppA, HdeB 및 GlnH(유도 A) 및 감소된 계놈 대장균 균주 MDS69metab에 OD~0.3으로 첨가된 50  $\mu$ M 인듀서(IPTG)로 OmpF, Mg1B, Agp, OmpC, RbsB, FkpA, 및 YtfQ(유도 B)을 갖는 주변 세포질 CRM197(캘리퍼스 분석)의  $\mu$ g/OD(상단 패널) 및  $\mu$ g/ml을 도시한다.

도 21은 0, 25, 35 및 50  $\mu$ M 인듀서로 OmpF 신호 서열과 감소된 계놈 대장균 MDS69metab에서 OD ~ 0.3으로 첨가된 0, 25, 35, 50, 75, 100, 150 및 250  $\mu$ M 인듀서로 YtfQ 신호 서열로 수득된 주변 세포질 CRM197(캘리퍼스 분석)의  $\mu$ g/OD(상단 패널) 및  $\mu$ g/ml을 도시한다. 결과는 두 세트 실험의 평균이다.

도 22는 50  $\mu$ M(OmpF) 또는 OD=0.3(후기)으로 첨가된 50, 75, 100, 150, 250  $\mu$ M(YtfQ)의 인듀서 농도로 OmpF 또는 YtfQ 신호 서열로 MDS69 metab의 배양 배지(M) 및 주변 세포질(P)에서 CRM197 수율을 비교하는 단백질 젤 분석이다. 샘플은 분석을 위해 지정된 OD에서 수집했다.

도 23은 OmpF 또는 YtfQ 신호 서열로 MDSmetab에서 그리고 표시된 인듀서 농도로 ompF 신호 서열로 BL21(DE3) 세포에서 CRM197의 주변 세포질 발현을 비교했다. 인듀서는 OD ~ 2(매우 늦은 유도)로 첨가하고, 샘플은 캘리퍼스로 분석했다. (가용성)주변 세포질 CRM197(OD600=250으로 추정)의 g/L은 인듀서 각각의 농도로 표시했다.

## 발명을 실시하기 위한 구체적인 내용

### 발명의 상세한 설명

[0025] 본 발명은 다양한 형태로 구현될 수 있지만, 여러 실시예들의 이하의 설명은 본 발명의 예시로서 고려되어야 한다는 이해 하에 이루어지며, 본 발명을 예시된 특정 실시예로 본 발명을 제한하려는 것은 아니다. 표제는 단지 편의상 제공된 것으로, 어떠한 방식으로든 본 발명을 제한하는 것으로 해석되어서는 안 된다. 임의의 표제 아래에 도시된 실시예는 임의의 다른 표제 아래에 도시된 실시예와 결합될 수 있다.

[0027] 본 출원에서 지정된 다양한 범위의 수치의 사용은, 명확히 달리 표시하지 않는 한, 정해진 범위 내의 최소 및 최대값 모두 단어 “약.” 앞에 있더라도 근사치로 언급된다. 이와 같이, 언급된 범위 위 및 아래에 약간의 변형이 그 범위 내의 값과 동일한 결과를 실질적으로 달성하기 위해 사용될 수 있다. 본 출원에 사용된, 수치를 참조할 때 용어 “약” 및 “대략”은 쟁점의 해당 당업자에게 그들의 보통 및 통상적인 의미를 갖는다. 또한, 범위의 개시는 인용된 최소값과 최대값 사이의 모든 값을 포함하는 연속적인 범위뿐만 아니라 이러한 값들에 의해 형성될 수 있는 임의의 범위로 의도된다. 이는 상부 및/또는 하부 유한의 경계를 포함 또는 포함하지 않는 것으로 형성될 수 있는 범위를 포함한다. 또한 이것은 주어진 공개된 수치를 다른 공개된 수치로 나눔으로써 유도될 수 있는 비율을 포함한다. 따라서, 당업자는 대다수 이러한 비율, 범위, 비율 범위는 본 출원에 제시된 데이터와 수치 및 본 발명의 일례인 다양한 실시예로부터 명확하게 유도될 수 있음을 이해할 것이다.

[0028] 본 명세서에 사용된 “감소된 계놈” 박테리아들은 약 1% 내지 약 75%가 제거된 계놈(단백질 코딩 유전자), 예를 들어, 약 5%, 약 10%, 약 20%, 약 30% 약 40%, 약 50% 또는 약 60%의 제거된 계놈을 갖는 박테리아를 의미한다. 일 실시예에서, 본 발명의 실행에서 사용되어진 감소된 계놈 박테리아는 바람직하게는 네이티브 모 균주의 계놈보다 적어도 이 퍼센트(2%) 및 이십 퍼센트(20%)(이들 사이의 임의의 수를 포함)까지 작게 되도록 유전자 공학에 의해 생성된 계놈을 갖는다. 바람직하게는, 계놈은 네이티브 모 균주의 계놈보다 적어도 오 퍼센트(5%) 및 삼십 퍼센트(30%)까지 작다. 더욱 바람직하게는, 계놈은 팔 퍼센트(8%)까지 십사 퍼센트(14%)까지 이십 퍼센트(20%)까지(이들 사이의 임의의 수를 포함) 또는 네이티브 모 균주보다 더 작다. 대안적으로, 계놈은 네이티브 모 균주보다 20% 미만, 30% 미만, 40% 미만 또는 50% 미만으로 작게 되도록 유전자 조작에 의해 조작될 수 있다. 용어 “네이티브 모 균주”는 일반적으로 과학계에 의해 근본 균주 계보(strain line)의 표본이며 그의 계놈에서 일련의 제거는 더 작은 계놈을 갖는 박테리아 균주의 생성이 이뤄질 수 있다고 이해되는 자연 또는 네이티브 환경에서 확인된 박테리아 균주를 의미한다. 또한, 네이티브 모 균주는 유전자 조작에 의해 생성된 균주에 대해 비교되는 균주를 말하며, 유전자 조작에 의해 생성된 균주는 네이티브 모 균주의 완전 상보체(complement)보다 작다. 계놈이 일련의 제거 후 작아지는 백분율은 “전체 제거 후 제거된 염기쌍의 총 개수”를 “전체 제거 전 계놈에서 염기쌍의 총 개수”에 의해 나누고 나서 100을 곱함으로써 계산한다.

유사하게, 게놈이 네이티브 모 균주보다 작아지는 백분율은 작은 게놈을 갖는 균주에서 뉴클레오티드의 총 개수(그것이 생산되는 과정에 관계없이)를 네이티브 모 균주에서 뉴클리오티드의 총 개수로 나누고 나서 100을 곱함으로써 계산한다.

[0029] 일 실시예에서, “감소된 게놈” 박테리아는 상기 양의 게놈의 제거가 최소 배양 배지에서 성장하는 미생물의 능력에 영향을 미치지 않는 박테리아를 의미한다. 두 개 또는 그 이상의 유전자 제거가 “허용 불가능한 영향” 이든 아니든 본 문맥 내에서 최소 배양 배지에서 성장하는 미생물의 능력은 특정 애플리케이션에 의존한다. 예를 들어, 증식률에서 30% 감소는 하나의 애플리케이션에서 허용될 수 있지만 다른 애플리케이션에서는 허용될 수 없다. 또한, 게놈으로부터 DNA 서열의 제거의 역효과는 배양 조건을 변경하는 것과 같은 조치에 의해 감소될 수 있다. 이러한 조치는 다르게 허용 불가능한 역효과를 허용 가능한 것으로 돌릴 수 있다. 일 실시예에서, 증식률은 대략 모 균주와 동일하다. 그러나, 모 균주에 비해 약 5%, 10%, 15%, 20%, 30%, 40%로부터 약 50%까지 작은 범위인 증식률은 본 발명의 범위 내에 있다. 특히, 본 발명의 박테리아의 두 배 시간은 약 15분에서 약 3시간 범위일 수 있다. 적합한 감소된 게놈 박테리아의 예뿐만 아니라, 대장균과 같은 박테리아로부터 DNA를 제거하는 방법의 비-제한적 예는, 미국 특허 번호 6,989,265, 7,303,906, 8,119,365, 8,039,243 및 8,178,339에 개시되어 있고, 이들 각각을 본 명세서에 참조로 포함한다.

[0030] 본 명세서에서 사용된 용어 “b 번호”는 Blattner 외., *Science* 277:1453-1474(1997)에 기술된 바와 같이 K-12 MG1655 균주의 유전자 각각에 할당된 고유 ID를 말한다.

[0031] 본 명세서에서 사용된 용어 “CRM197”은 CRM197에 글루탐산을 갖는 글리신(야생형 독소에서 위치 52)의 치환을 선도하는 단일 G → A 전이를 갖는 디프테리아 독소 변이체인, 가교 반응(cross-reacting) 물질 197(CRM197)을 말한다. 이 미스센스 변이는 ADP-리보실 전이효소 활성의 손실에 대해 책임이 있다. 예를 들어, Giannini 외., *Nucleic Acids Res.* 12(10):4063-4069 (1984)를 참조하라.

[0032] 여러 실시예에서, 감소된 게놈 대장균 숙주 세포에서 재조합 CRM197 단백질을 생산하는 방법이 제공된다. 예를 들어, 야생형 대장균 숙주 균주에 비해 감소된 게놈 대장균 숙주 균주를 사용하여 불용성 또는 가용성 형태로 놀라울 정도로 높은 수율의 재조합 CRM197을 생산할 수 있음이 확인되었다. 한 양태에서, 이 방법은 숙주 세포의 세포질에서 불용성 CRM197의 생산을 증가하도록 이끈다. 다른 양태에서, 이 방법은 숙주 세포의 주변 세포질에서 가용성 CRM197의 생산을 증가하도록 이끈다. 바람직한 실시예에서, 감소된 게놈 대장균 숙주 세포를 생성하는 데 사용되는 네이티브 모 대장균 균주는 K-12 균주 MG1655 등의 K-12 균주이다.

[0033] 일부 실시예에서, K-12 MG1655 등의 네이티브 K-12 균주는 본 명세서에 기술된 방법에 따라 재조합 CRM197을 생산하기 위해 사용된다.

[0034] 본 발명에 따른 용도를 위해 CRM197의 뉴클레오티드 서열은 재조합 DNA 기술을 사용하여 제조될 수 있다. 예를 들어, CRM197은 화학적으로 합성될 수 있거나 코리네박테리아 과지  $\beta$  (Greenfield 외., *Pro Nat Acad Sci*, 80. : 6953-6957 (1993))에 의해 이행된 디프테리아 독소에 대해 야생형 구조 유전자의 공지된 뉴클레오티드 서열을 기반으로 부위-특이적 변이 유도에 의해 제조될 수 있다. 바람직하게는, CRM197의 뉴클레오티드 서열은 대장균에서 발현에 대해 최적화된다.

[0035] 이종성 핵산의 다양한 서열의 특성은, 번역 개시 영역의 변형(modification), mRNA 구성 요소의 교번(alternation) 및 다른 코돈 성향의 사용을 포함하나 이들에 제한됨 없이 최적화될 수 있다. 대장균 숙주 세포에서 발현을 개선시키기 위해 핵산 서열을 최적화하는 방법은 당해 분야에 공지되어 있으며, 예를 들어, 미국 특허 번호 제7,561,972호에 기술되어있으며, 그 내용은 본 명세서에 참고로 포함된다. 바람직하게는, 대장균에서 발현을 위한 CRM197의 뉴클레오티드 서열의 최적화는 적어도 코돈 최적화를 포함한다. 대장균에 거의 사용되지 않은 이종성 핵산 서열에서 코돈의 존재는 인코드된 단백질의 번역을 지연하여 대장균 숙주 세포에서 감소된 발현을 초래할 수 있다. 따라서, 일 양태에서, 대장균에서 일반 코돈 사용은 대장균에서 CRM197의 발현을 최적화하는 데 사용된다. 대장균에서 발현을 위한 CRM197의 최적화는 바람직하게는 방해 이차 구조(Interfering secondary structure)의 최소화를 포함한다. 방해 이차 구조는 전사 및 번역을 지연함으로써 대장균에서 이종성 단백질의 감소된 발현을 초래할 수 있다. 예를 들면, 개시 부위에서 mRNA의 이차 구조는 번역 효율에 대해 역 상관관계를 나타낸다. 신호 서열을 인코딩하는 상류 영역에 부착될 경우 대장균의 주변 세포질에서 발현을 위해 최적화된, 모범적인 CRM197 뉴클레오티드 서열은 SEQ ID NO: 1(도 16a)로서 제공된다. 상류 ATG 개시 코돈에 부착될 경우 대장균의 세포질에서 발현을 위해 최적화된, 모범적인 CRM197 뉴클레오티드 서열은 SEQ ID NO: 3(도 16b)으로서 제공된다. 본 발명의 방법은 SEQ ID NO: 1로 명시된 CRM197 뉴클레오티드 서열로 한정되지 않는다는 것이 이해되어야 한다. 대장균에서 발현을 위한 이종성 뉴클레오티드 서열의 최적화를 위한 추가 전략은 당

해 분야에 공지되어 있어 상기 기술된 전략들에 추가 또는 대안으로 사용될 수 있다.

[0036]

발현 시스템을 포함하는 대장균과 같은 유전자 조작으로 생성된 박테리아 숙주 세포로부터 재조합 이종성 단백질을 생산하는 공정은 당업자에게 잘 알려져 있다. 재조합 CRM197은 이러한 방법들 중 임의의 방법을 통해(예를 들어, 감소된 게놈) 대장균 숙주 세포에서 발현될 수 있다. 한 양태에서, 본 발명의 방법은 발현 시스템을 포함하는 감소된 게놈 대장균 숙주 세포에 관한 것이며, 상기 발현 시스템은 유도 가능 프로모터에 동작 가능하게 연쇄된 CRM197을 인코딩하는 뉴클레오티드 서열을 포함하여, 프로모터가 유도될 경우 숙주 세포에서 발현된다. 바람직한 양태에서, 상기 프로모터는 적당량의 IPTG를 첨가함으로써 유도된다. 감소된 게놈 대장균 숙주 세포 내로 폴리 뉴클레오티드의 도입은, Davis 외., *Basic Methods in Molecular Biology*(1986) 및 Sambrook 외., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.(1989)에 기술된 것과 같은, 인산칼슘 형질 주입, 미세 주입, 전기 천공, 접합, 감염 등을 포함하나 이들로 제한하지 않는 여러 표준 분자 생물학 기술 중 하나에 의해 달성될 수 있다. 유사하게, 폴리 뉴클레오티드를 유지, 전달 또는 발현 및/또는 숙주에서 폴리펩티드를 발현하는데 적합한 임의의 시스템 또는 벡터는 본 발명을 실시하는데 사용될 수 있다. 예를 들어, 적절한 DNA 서열은 표준 기술에 의해 플라스미드와 같은 벡터에 삽입될 수 있다.

[0037]

본 발명의 하나의 양태는 (예를 들어, 감소된 게놈) 대장균 숙주 세포에서 CRM197의 주변 세포질 발현에 관한 것이다. 주변 세포질에서의 단백질 발현은 산업용으로 사용되었으며, Hanahan, J. Mol. Biol., 166:557-580(1983); Hockney, Trends Biotechnol., 12:456-632(1994); 및 Hannig 외., Trends Biotechnol., 16:54-60(1998)에서 검토되었으며, 이들 각각은 본 명세서에 참고로 포함되었다. 따라서, 여러 실시예에서, 재조합 CRM197 단백질의 발현에 적합한 조건하에서 발현 제어 서열에 작동 가능하게 연쇄된 신호 서열에 융합된 CRM197 단백질을 인코딩하는 핵산 서열을 포함하는 발현 벡터를 포함하는 (예를 들어, 감소된 게놈) 대장균을 성장시키는 방법이 제공되며, 상기 신호 서열은 대장균 숙주의 주변 세포질에 CRM197 단백질의 전이를 통제한다. 이러한 방법에 의하면, 놀란 만큼 높은 수율의 완전한 가용성 CRM197이 생산되고, 실질적으로 모든 가용성 CRM197을 회수할 수 있었다.

[0038]

단백질 상에 신호 서열의 존재는 주변 세포질 공간 내 대장균의 내막에 걸쳐 새롭게 번역된 단백질의 전이를 용이하게 한다. 신호 서열은 이어서 절단되며; 따라서 네이티브 C. 디프테리아 신호 서열을 대장균의 주변 세포질에 CRM197의 전이를 통제하는 신호 서열로의 대체는 궁극적으로 동일한 아미노산 서열을 갖는 성숙 단백질을 초래한다.

[0039]

대장균 주변 세포질에 대해 이종성 단백질을 통제할 수 있는 신호 서열의 대표적인 예는 다음에 열거했다. 본 발명의 방법에 유용한 신호 서열은 아래에 열거된 것에 한정되지 않는다는 것은 당연하다. 바람직하게는, 신호 서열은 대장균에서 발현 시 주변 세포질에 대해 적어도 70, 80, 90 또는 100% 폴리펩티드의 통제를 초래한다.

신호 서열	아미노산 서열
PelB (펩테이트 리아제 B)	MKYLLPTAAAGLLLLAAQPAMA
OmpA (외막 단백질 A)	MKKTAIAIAVALAGFATVAQA
StII (내열성 엔테로톡신 2)	MKKNIAFLASMVFVSIATNAYA
Endoxylanase	MFKFKKKFLVGLTAAFMSISMFSATASA
PhoA (알칼리 포스파타아제)	MKQSTIALALLPLLFTPVTKA
OmpF (외막 단백질 F)	MMKRNILAVIVPALLVAGTANA
PhoE (외막 기공 단백질 E)	MKKSTLALVVMGIVASASVQA
MalE (말토스-결합 단백질)	MKIKTGARILALSALTMMFSASALA
OmpC (외막 단백질 C)	MKVVLSSLLVPALLVAGAANA
Lpp (무레인 지질단백질)	MKATKVLGAVILGSTLLAG
LamB ( $\lambda$ 수용체 단백질)	MMITLRKLPLAVAVAAGVMSAQAMA
OmpT (프로테아제 VII)	MRAKLLGIVLTTPIAIISSFA
LTB (이열성 엔테로톡신 서브유닛 B)	MNKVKCYVLFTALLSSLYAHG
MglB (메틸 갈락토스 트랜스포터)	MNKKVLTLSAVMASMLFGAAAHA
OppA (올리고펩티드 트랜스포터)	MTNITKRSLSVAAGVLAALMAGNVALA
RbsB (서브유닛 리보오스 트랜스포터)	MNMKKLATLVSALVSATVSANAMA
Agp (글루코스-1 포스파타아제, 3-펩타아제)	MNKTLIAAAVAGIVLLASNAQA
FkpA (펩티딜-프롤릴 시스-트랜스 이소머라아제)	MKSLFKVTLLATTMAVALHAPITFA
YtfQ (갈락토퓨라노스 결합 단백질, 서브유닛 ABC 트랜스포터)	MWKRLLIVSAVSAAMSSMALA
HdeA (산성조건에 의해 유도된 스트레스 반응)	MKKVLGVILGGLLLLPVVSNA
HdeB (산성조건에 의해 유도된 스트레스 반응)	MNISSLRKAFIFMGAVAALSLVNAQSALA
GlnH (글루타민 ABC 트랜스포터의 서브유닛)	MKSVLKVSLAALTAFAVSSHA

[0040]

[0041] 본 발명에 따른 사용을 위한 추가적인 신호 서열은, CpdB(3'-뉴클레오티다아제/2', 3'-환상 뉴클레오티드 2'-포스포디에스터라아제), YdeN(추정 술퍼타아제), OsmY(고삼투압 스트레스에 의해 유도된), ArtI(서브 유닛 아르기닌 ABC 트랜스포터), GltL(글루타메이트 ABC 트랜스포터), 및 CybC(시토크롬 b562)을 포함하나 이에 제한되지 않는다.

[0042]

바람직한 실시예에서, 신호 서열은 ytfQ, OmpA 및 OmpF 신호 서열로부터 선택된다. 특히 바람직한 실시예에서, 신호 서열은 OmpF 신호 서열이다. 다른 특히 바람직한 실시예에서, 신호 서열은 YtfQ 신호 서열이다.

[0043]

본 명세서에 기술된 방법에 따른 임의의 감소된 계놈 대장균 균주는 숙주 세포로서 사용될 수 있다. 일 양태에서, 감소된 계놈 대장균은 그의 네이티브 모 균주의 계놈보다 5%와 30% 사이 또는 5% 와 20% 사이와 같이 (이들 사이의 임의의 수를 포함하는) 적어도 이 퍼센트(2%) 및 사십 퍼센트(40%)까지 작게 되도록 유전자 조작에 의해 생성된 계놈을 갖는다. 계놈이 일련의 제거 후 작아지는 백분율은 “전체 제거 후 제거된 염기쌍의 총 개수”를 “전체 제거 전 모 균주의 계놈에서 염기쌍의 총 개수”에 의해 나누고 나서 100을 곱함으로써 계산한다. 또 다른 양태에서, 감소된 계놈 박테리아는 4.41Mb와 3.71Mb 사이, 4.41Mb와 3.25Mb 사이 또는 4.41Mb와 2.78Mb 사이인 계놈을 보유하고 있다. 본 명세서에 기술된 방법에 따른 사용하기 위한 감소된 계놈 대장균 균주는 국제 특허 공개 공보 번호 WO2003/070880에 기술된 방법에 의해 모 대장균 균주의 누적 계놈 제거에 의해 생산될 수 있다.

[0044]

모 대장균 균주는 임의의 대장균 균주일 수 있지만 바람직하게는 K-12 균주(예를 들어, MG1655(ATCC 번호 47076) 또는 W3110(ATCC 번호 27325)) 또는 B 균주이다. 특히 바람직한 모 대장균 균주는 4,639,674 염기쌍을 가지고 있는 계놈을 갖는 K-12 균주 MG1655(주석 베전 m56, NCBI 수탁 번호 U000961)이다.

[0045]

하나의 양태에서, 모 대장균 균주는, 본 명세서에 참고로 포함된 미국 특히 번호 8,178,339의 테이블 2-20에 열거된 유전자 중 하나 이상을 결여한 K-12 균주이다. 바람직한 실시예에서, 감소된 계놈 대장균 K-12 균주는 최소한 다음의 유전자(Blattner 외., Science, 277:1453-74 및 GenBank 수탁 번호 400096에 제시된 지정을 기반으로 “b” 번호에 의해 식별되는): b0245-b0301, b0303-b0310, b1336-b1411, b4426-b4427, b2441-b2450,

b2622-b2654, b2657-b2660, b4462, b1994-b2008, b4435, b3322-b3338, b2349-b2363, b1539-b1579, b4269-b4320, b2968-b2972, b2975-b2977, b2979-b2987, b4466-4468, b1137-b1172, b0537-b0565, b0016-b0022, b4412-b4413, b0577-b0582, b4415, b2389-b2390, b2392-b2395, b0358-b0368, b0370-b0380, b2856-b2863, b3042-b3048, b0656, b1325-b1333, b2030-b2062, b2190-b2192, b3215-b3219, b3504-b3505, b1070-b1083, b1878-b1894, b1917-b1950, b4324-b4342, b4345-b4358, b4486, b0497-b0502, b0700-b0706, b1456-b1462, b3481-b3484, b3592-b3596, b0981-b0988, b1021-b1 029, b2080-b2096, b4438, b3440-b3445, b4451, b3556-b3558, 및 b4455가 결여되어 있으며, 이를 유전자들은 감소된 게놈(또는 다중 제거) 균주 MDS39를 생성하기 위해 대장균 K-12 MG1655에서 제거된 유전자들이다. 또 다른 바람직한 실시예에서, 감소된 게놈 대장균 K-12 균주는 다음의 유전자: b1756이 추가로 결여되어 있으며, 이 유전자는 감소된 게놈 균주 MDS40을 생산하기 위해 MDS39로부터 제거된 유전자이다. 또 다른 바람직한 실시예에서, 감소된 게놈 대장균 K-12 균주는 다음의 유전자: 감소된 게놈 균주 MDS42를 생성하기 위해 MDS41로부터 제거된 유전자인 b2945(*endA*)가 추가로 결여되어 있다. 또 다른 실시예에서, 감소된 게놈 대장균 K-12 균주는 다음의 유전자: b0315-b0331, b0333-b0341 및 b0346-b0354 중 임의의 것이 추가로 결여되어 있으며, 이 유전자는 감소된 게놈 균주 MDS43을 생산하기 위해 MDS42로부터 제거된 유전자이다. 또 다른 실시예에서, 감소된 게놈 대장균 K-12 균주는 다음의 유전자: b2481-b2492, b2219-b2230, b4500, b3707-b3723, b0644-b0650, b4079-4090, b4487, b4092-b4106, b0730-b0732, b3572-b3587, b1653, b2735-b2740, b2405-b2407, b3896-b3900, b1202, b4263-b4268, b0611, b2364-b2366, b0839, b0488-b0500, b0502 중 임의의 것이 추가로 결여되어 있으며, 이 유전자는 MDS60을 생산하기 위해 MDS43으로부터 제거된 유전자이다. 또 다른 실시예에서, 감소된 게놈 대장균 K-12 균주는 다음의 유전자: b0566-b0575, b2209, b0160-b0161, b1431-b1444, b3643, b1037-b1043, b0383, b0226-b0234, b2115-b2132 중 임의의 것이 추가로 결여되어 있으며, 이 유전자는 MDS69을 생산하기 위해 MDS60으로부터 제거된 유전자이다. 특정 실시예에서, 본 명세서에 기술된 방법에 사용하기 위한 감소된 게놈 대장균 K-12 균주는 MDS41, MDS42, MDS60 또는 MDS69이다.

[0046]

본 발명에서 사용을 위한 대장균 숙주 세포는, 비록 기능성 *recA* 유전자(b2699)를 결여한 대장균도 CRM197을 생산하기 위한 숙주 세포로서 사용될 수 있지만, 바람직하게는 기능성 *recA* 유전자 (b2699)를 포함한다. 예를 들어, 감소된 게놈 대장균, 예를 들어, 균주 MDS40, MDS41, MDS42 또는 MDS69와 같은 감소된 게놈 대장균은 변형된 대장균 K-12 균주로부터 유전자 전체 또는 부분 제거에 의해 b2699의 불활성화에 의해 변형될 수 있다. 일 실시예에서, *OmpA* 신호 서열에 융합된 CRM197은 기능적 *recA* 유전자를 결여한 감소된 게놈 대장균 숙주에서 발현된다.

[0047]

다른 양태에서, 감소된 게놈 대장균은, 본 명세서에 그 내용이 참고로 포함된, WIPO 공개 공보 번호 2013/059595에 기술된 바와 같은, *Pol II*, *Pol IV* 및 *Pol V*를 인코딩하는 유전자로 이루어진 군으로부터 선택된 하나 이상의 비-기능성 유전자를 포함한다. 일 실시예에서, 감소된 게놈 대장균은 비-기능성 *PolB*(b0060에 의해 인코드된, 대장균 K12 MG1655 게놈에서 좌표 63429-65780) 및 *DinB*(b0231에 의해 인코드된, MG1655의 게놈에서 좌표 250898-251953) 유전자를 갖는다. 다른 실시예에서, 감소된 게놈 대장균은 비-기능성 *PolB*, *DinB* 및 *UmuDC*(b1183-b1184에 의해 인코드된, MG1655의 게놈에서 좌표 1229990-1231667) 유전자를 갖는다. 바람직하게는, 유전자(들)은 전체 또는 부분 제거에 의해 불활성화된다. 예를 들어, *polB*, *dinB* 및 *umuDC* 유전자는 균주 MDS40, MDS41, MDS42 또는 MDS69와 같은 감소된 게놈 대장균에서 비-기능성으로 될 수 있다.

[0048]

또 다른 양태에서, 감소된 게놈 대장균(예를 들어, 균주 MDS40, MDS41, MDS42 또는 MDS69)은 (a) 오로테이트 포스포 리보실 전이효소 활성을 향상 (b) 활성 아세토하이드록시 산 신타아제 II 생산 및 (c) *iclR* 및 *arpA* 유전자 산물의 발현을 감소시키기 위해 유전자가 조작되었다.

[0049]

피리미딘 뉴클레오티드의 합성을 촉매하는 효소인 대장균 오로테이트 포스포 리보실 전이효소는 b-번호 b3642인 *pryE* 유전자에 의해 인코드된다. *pryE* 유전자는 상류 *rph* 유전자(b3643)를 갖는 오페론에 존재한다. *pryE* 유전자는 *rph* 유전자의 코딩 영역에서 a-1 프레임시프트 변이로 인해 MG1655 및 W3310과 같은 대장균 K-12 균주에서 차선의 수준으로 발현된다. 오로테이트 포스포 리보실 전이효소 활성은 *pryE*의 변역 개시 부위에 더 가깝게 *rph-pryE* 오페론의 프로모터를 가져오기 위해 *rph* 코딩 서열을 완전히 제거하는 제거에 의해 향상될 수 있다. 대안적으로, 그 내용이 참고로서 포함된 미국 특허 번호 8,293,505에 기술된 방법 중 임의의 방법 오로테이트 포스포 리보실 전이효소 활성을 향상시키기 위해 사용될 수 있다.

[0050]

대장균 아세토하이드록시 산 신타아제 II는 일반적으로 *ilvG* 유전자에 의해 인코드된, 큰 서브 유닛 및 *ilvM* 유전자(b3769)에 의해 인코드된, 작은 서브 유닛으로 구성된다. 대장균 K-12 균주 MG1655의 *ilvG* 서열은 변질되어

실질적으로 GenBank 수탁 번호 AAC77488.1에서 제시된 바와 같은, 유사 유전자(pseudo gene)(b-번호 b4488)이다. *i1vG* 유사 유전자는 두 개의 개별적인 코딩 서열, *i1vG\_1*(b3767)과 *i1vG\_2*(b3768)로 구성되어 있다. MG1655 와 같은 K-12 균주에서 *i1vG* 유사 유전자 서열은 다른 대장균 균주(예를 들어, B 균주, 0 균주, 등)에서 확인된 완전한 *i1vG* 유전자에 비해 위치 983 및 984에서 뉴클레오티드 GT의 제거를 포함한다. 이러한 뉴클레오티드의 제거은 프레임시프트 변이를 초래하고 K-12 *i1vG* 유사 유전자 서열의 위치 982-984에서 뉴클레오티드 TGA는 *i1vG\_1*에 대응하는 *i1vG*의 잘린 형태로 인한 조기 종결 코돈 역할을 한다. 따라서, *i1vG*의 정상적인 유전자 산물은 발현되지 않고 아세토하이드록시 산 신타아제 II는 대장균 K-12 균주에 존재하지 않는다. 감소된 게놈 대장균은 *i1vG* 유전자에서 네이티브-2 프레임시프트 변이를 상보하는 변이를 도입함으로써 활성 아세토하이드록시 산 신타아제 II를 생산하도록 변형될 수 있다. 대안으로, 감소된 게놈 대장균은 본 명세서에 참고로 그 전체 내용이 포함된, 미국 특허 번호 7,300,776의 방법 중 임의의 방법에 의해 활성 아세토하이드록시 산 신타아제 II를 생산하도록 변형될 수 있다.

[0051]

대장균 K 균주의 *icIR* 및 *arpA* 유전자는 각각 글리옥실레이트 선트 효소 및 아세틸-CoA 합성 효소의 발현을 조절하는 조절 단백질을 인코딩하는 유전자에 인접한다. *icIR*(이소시트르산 분해효소 레귤레이터) 유전자, b-번호 b4018은 NCBI Entrez GeneID 번호 948524에 기술되어 있다. *arpA*(안카린-유사 조절 단백질) 유전자, b-번호 b4017은 NCBI Entrez GeneID 번호 944933에 기술되어 있다. *arpA* 유전자는 BL21DE3 및 K-12 균주 서열과 관련된 REL606과 같은 B 균주의 게놈 서열에서 부분적으로 제거되는 것으로 확인됐다. *icIR* 및 *arpA* 유전자는, 예를 들어 미국 특허 번호 6,989,265의 라인 41, 칼럼 14 내지 라인 45, 칼럼 8에 기술된 “무 상흔” 제거 방법에 의해, *icIR* 및 *arpA* 유전자 서열의 전부 또는 일부의 제거에 의해 감소된 게놈 대장균에서 비활성화(예를 들어, 비-기능성 상태가 되도록 만들어진)될 수 있다.

[0052]

다른 실시예에서, 감소된 게놈 대장균은 본 명세서에 참고로 그 내용이 포함된, 미국 특허 번호 8,367,380에 기술된 임의의 변이를 함유하는 *relA* 유전자를 포함한다. 예를 들면, 균주 MDS40, MDS41, MDS42 또는 MDS69와 같은 감소된 게놈 대장균 균주는 이들 변이 중 임의의 변이를 포함하도록 변형될 수 있다.

[0053]

본 발명에 따라 사용하기 위한 감소된 게놈 대장균은 상술한 변형의 임의의 조합을 포함할 수 있다. 일부 바람직한 실시예에서, 감소된 게놈 대장균은 적어도 MDS42의 제거를 포함하거나 또는 적어도 MDS69의 제거를 포함하며, (a) 오로테이트 포스포 리보실 전이효소 활성을 향상 (b) 활성 아세토하이드록시 산 신타아제 II 생산 및 (c) CRM197의 주변 세포질 생산을 위해 균주로서 사용되는 *icIR* 및 *arpA* 유전자 산물의 발현을 감소시키기 위해 유전자가 조작되었다. 바람직하게는 감소된 게놈 대장균은 기능성 *recA* 유전자를 포함한다.

[0054]

여러 가지 단백질 코딩 유전자는 감소된 게놈 대장균을 형성하기 위해 제거될 수 있다. 대장균 및 다른 박테리아에서, 제거될 수 있는 DNA 서열의 유형은 일반적으로 유기체 또는 그 유기체의 유전자 산물의 안정성에 부정적인 역효과를 미치게 하는 것들을 포함한다. 불안정성을 야기하는 이러한 요소는 전이 인자, 삽입 서열, 및 게놈 불안정성의 역할을 하는 다른 “이기 DNA” 요소를 포함하나 이에 제한되지 않는다. 예를 들면, 삽입 서열 (IS) 요소 및 이와 연관된 전이 요소는 종종 박테리아 게놈에서 확인되고, 따라서 제거 대상이다. IS 서열은 대장균에서 일반적이며, 이들 모두는 제거될 수 있다. 이 문서의 명확성을 위해, 우리는, 완전하든 또는 결손되었든, 게놈에서 한 지점에서 다른 지점으로 이동할 수 있는 DNA 요소를 일반적으로 참조하기 위해 용어 IS 요소 및 전이 요소를 사용한다. 과학 기술에서 IS 요소의 유해한 효과의 한 예는 서열화를 위해 전과 중에 숙주 대장균의 게놈에서 플라스미드로 도약(hop)할 수 있다는 사실이다. 이러한 인공산물(artifact)은 모든 IS 요소의 숙주 세포로부터의 제거에 의해 예방할 수 있다. 특정 애플리케이션에 대해, 활성 및 비활성 프로파지와 같은 게놈 불안정성과 관련된 다른 특정 유전자도 제거될 수 있다. 특히 바람직한 실시예에서, 본 발명에 따른 감소된 게놈 대장균 숙주는 모든 삽입 서열(즉, 삽입 서열을 포함하지 않는)로부터 제거되었다. 관련 양태에서, 감소된 게놈 대장균 숙주는 IS1, IS2, IS3, IS5, IS150 및 IS186 삽입 서열 모두 결여되어 있다.

[0055]

또한, 본 발명의 감소된 게놈 박테리아는, 예를 들어 박테리아의 성장과 대사를 위해 불필요한 특정 유전자, 유사 유전자, 프로파지, 바람직하지 않은 내인성 제한-변형 유전자, 병원성 유전자, 독소 유전자, 섬모 유전자, 주변 세포질 단백질 유전자, 침습성 유전자, 리포다당류 유전자, 클래스 III 분비 시스템, 파지 독성 결정 인자, 파지 수용체, 병원성 유전자균, RHS 요소, 알 수 없는 기능의 서열 및 같은 네이티브 모 종의 박테리아의 두 균주 사이에서 공통점을 찾을 수 없는 서열이, 그러나 이것에 제한되지 않고, 결여되게 설계될 수 있다. 또한, 세포 생존에 요구되지 않는 다른 DNA 서열은 제거되거나 생략될 수 있다.

[0056]

특히 바람직한 실시예에서, 감소된 게놈 대장균은 네이티브 모 대장균 K 균주의 게놈 보다 오 퍼센트(5%) 및 삼십 퍼센트(30%) 사이로 작고 모든 삽입 서열(IS) 요소가 결여된 게놈을 갖도록 제공된다. 대장균 MG1655의 게놈

지도상에 IS 요소의 위치는 본 명세서에 그 내용이 참고로 포함된, 미국 특허 번호 8,178,339의 도 1 및 표 2에 도시됐다. 일반적으로 대장균에서 발생하고 제거할 수 있는 삽입 서열 요소는 IS1, IS2, IS3, IS4, IS5, IS30, IS150, IS186, IS600, IS911 및 IS10을 포함하나 이에 제한되지 않는다. 바람직하게는, 네이티브 모 대장균 균주는 대장균 K-12 균주 MG1655이다.

[0057] 또 다른 특허 바람직한 실시예에서, 감소된 게놈 대장균은 다음의 유전자: b0018, b0150, b0152-b0153, b0161, b0227, b0250, b0291-b0293, b0297, b0316, b0329, b0365, b0371, b0376, b0383-b0384, b0494, b0497-b0498, b0545, b0553, b0559, b0562, b0565, b0567, b0569, b0572-b0574, b0611, b0700, b0704, b0839, b0983-b0986, b1023-b1024, b1072, b1079-b1080, b1083, b1038-b1039, b1041-b1043, b1329, b1357, b1369, b1377, b1383, b1386, b1435-b1436, b1440, b1562, b1878, b1889, b1920, b1995, b2000, b2062, b2123, b2126, b2131-b2132, b2190, b2209, b2487, b2637, b2647, b2945, b3043, b3046-b3048, b3215-b3216, b3219, b3325, b3329, b3338, b3482, b3579, b3584, b3586, b3593, b3596, b4080, b4088, b4105, b4280, b4290-b4292, b4309-b4311, b4314, b4316-b4320, b4412, b4415, b4455 및 b4487, 단독으로 또는 임의의 조합을 포함하는 하나 이상의 주변 세포질 단백질 유전자의 제거(들)을 포함하나 이것에 제한되지 않는다.

[0058] 본 발명의 또 다른 양태에서, K-12 MG1655와 같은 네이티브 K-12 균주는 본 명세서에 기술된 방법에 따른 재조합 CRM197을 생산하기 위해 사용된다.

[0059] 재조합 단백질은 재조합 단백질의 적당한 풀딩을 제공할 수 있는, Skp, DnaK, DnaJ, Caf1M, Caf1A, DsbA, DsbB, DsbC, DsbD, PpiA, PpiD, FkpA, SurA, MBP, GST, YebF, MalE, HlyA, 히루딘, OmpF, Spy, YccA; 및 PspA를 포함하나 이에 제한되지 않는 아프론/이황화 결합 형성 효소와 공동 발현될 수 있다. 재조합 단백질의 주변 세포질 발현에 유용한 이러한 단백질의 핵산 서열은, 본 명세서에 참고로 그들 각각을 포함한, 미국 특허 번호 5,747,662; 5,578,464 및 6,022,952에 기술된 것을, 제한 없이, 포함한다.

[0060] CRM197을 인코딩하는 발현 벡터로 형질 전환된 대장균 숙주 세포(감소된 게놈 또는 네이티브 K12 균주)는 임의의 발효 형식으로 배양할 수 있다. 예를 들어, 진탕 플라스크 배양, 배지, 유가 배양, 반-연속 및 연속 발효 모드는 본 명세서에서 사용할 수 있다. 본 명세서에서 사용된 “발효”는 리터럴 발효를 사용하는 실시예 및 다른 비-발효 배양 모드를 사용하는 실시예 모두를 포함한다. 또한, 1리터 규모 및 더 큰 발효 용적을 포함하는 임의의 규모의 발효가 사용될 수 있다. 일 실시예에서, 발효 용적은 1리터이거나 적어도 1리터이다. 다른 실시예에서, 발효 용적은 적어도 5리터, 10리터, 15리터, 20리터, 25리터, 50리터, 75리터, 100리터, 200리터, 500리터, 1,000리터, 5,000리터, 10,000리터, 50,000리터, 또는 그 이상이다.

[0061] 다양한 실시예에서, 발효 배양 배지는 리치 배양 배지, 최소 배양 배지 및 미네랄 염 배양 배지 중에서 선택될 수 있다. 바람직한 실시예에서, 최소 배양 배지 또는 미네랄 염 배양 배지가 선택된다. 바람직하게는 이 배양 배지는 혈청 및 동물에서 유도된 산물이 없거나 또는 실질적으로 없다. 미네랄 염 배양 배지는 일반적으로 미네랄 염 및 탄소원(예를 들어, 포도당, 자당, 또는 글리세롤)으로 구성된다. 미네랄 염 배양 배지를 만드는 데 사용된 미네랄 염은 예를 들어, 인산칼륨, 황산 또는 염화암모늄, 황산 또는 염화마그네슘, 및 염화칼슘과 같은 트레이스 미네랄, 봉산, 및 철, 구리, 망간 및 아연의 황산염들 중에서 선택된 것들을 포함한다. 펩톤, 트립톤, 아미노산, 또는 효모 추출물과 같은 유기 질소원은 미네랄 염 배양 배지에 전혀 포함되지 않는다. 대신에, 무기 질소원이 사용되고, 이것은 예를 들면, 암모늄염, 암모니아 수용액, 및 암모니아 가스 중에서 선택할 수 있다. 바람직한 미네랄 염 배양 배지는 탄소원으로서 포도당을 포함할 것이다. 또한, 미네랄 염 배지에 비해, 최소 배양 배지는 미네랄 염 및 탄소원을 포함할 수 있지만, 예를 들어, 아미노산, 비타민, 펩톤, 또는 다른 성분의 낮은 수준으로, 비록 이들이 매우 최소한의 수준으로 첨가될지라도, 상보될 수 있다.

[0062] 실시예에서, 타겟 배양 세포 밀도는 인듀서, 바람직하게는 IPTG가 단백질 생산을 개시하기 위해 첨가되는 시점에 도달한다. 유도에서 세포 밀도, 인듀서의 농도, pH 및 온도는 발현의 최적 조건을 결정하기 위해 변화될 수 있다고 생각되고 있다.

[0063] 바람직한 실시예에서, 배양의 pH는 약 6.5 내지 7.5이다.

[0064] 형질 전환된 감소된 게놈 대장균의 성장, 배양 및/또는 발효는 생존을 허용하는 온도 범위 내에서 수행하지만, 바람직하게는 약 20°C 내지 약 30°C에서, 보다 바람직하게는 약 25°C이다. 또 다른 바람직한 실시예에서, 배양은 37°C(예를 들어, 1 내지 3시간)에서 상대적으로 짧은 초기 배양을 하고 나서 유도 이전 및 이후에 약 20°C 내지 약 30°C에서, 바람직하게는 약 25°C에서 성장이 뒤따르는 것을 포함한다. 다른 실시예에서, 배양은 유도 이전 및 이후에 약 25°C에서 성장을 포함한다.

- [0065] 실시예에서, 진탕 플라스크 조건하에서, 인듀서는 20 ~ 30°C, 바람직하게는 25°C의 배양 온도에서 약 0.1 내지 약 1.5, 더 바람직하게는 약 0.2 내지 약 0.9, 더욱 더 바람직하게는 약 0.3 내지 약 0.6)의 600nm에 광학 밀도 (OD)로 첨가된다. 600nm에, 1 OD 단위는 약  $0.8 \times 10^9$  세포/ml에 상응한다. 다른 실시예에서, 발효 조건하에서, 인듀서는 약 100 내지 400, 더 바람직하게는 약 150 내지 300, 가장 바람직하게는 230과 250 사이의 OD600으로 첨가된다.
- [0066] 본 발명의 방법은 야생형 대장균 B 균주에서와 같은 종래의 발현 시스템에 비하여 적절히 처리된 CRM197의 수준의 증가를 위해 제공한다. 특정 실시예에서, 방법은 가용성 CRM197의 증가를 제공한다. 이러한 맥락에서, 용어 “가용성”은 생리학적 조건하에서 완충액에 10-30분 동안 빠른 속도로 회전시키고자 했을 경우 단백질이 약 5,000 내지 20,000 x 중력 사이의 원심 분리로 침전되지 않는 것을 의미한다. 반대로, “불용성”은 생리학적 조건하에서 완충액을 10-30분 빠른 속도로 회전시키고자 했을 경우 단백질이 5000 내지 20,000 x 중력 사이의 원심 분리에 의해 침전될 수 있다는 것을 의미한다.
- [0067] 본 발명의 방법 (예를 들어, 감소된 게놈) 대장균 숙주 세포에서 생산된 재조합 CRM197의 회수를 포함할 수 있다. 가용성 단백질로서 주변 세포질에서 생산했을 경우, 가용성 형태의 재조합 CRM197의 회수는 바람직하게는 세제 및 가용화제의 부재 하에서 대장균 숙주 세포를 기계적으로 용해시킴으로써 수행된다. 기계적 분쇄는 일반적으로 초음파(Neppiras 및 Huges, Biotechnology and Bioengineering, 6:247-270(1964)), 마이크로 유동화 (Sauer 외., Biotechnology and Bioengineering, 33:1330-1342(1989)), 또는 비드 밀링(Limon-Lason 외., Biotechnology and Bioengineering, 21(5):745-774 (1979))을 포함한다. 당 업계에 공지된 다른 기계적 방법들이 사용될 수 있다.
- [0068] 재조합 CRM197은 황산 암모늄 또는 에탄올 침전, 산 추출, 음이온 또는 양이온 교환 크로마토그래피, 포스포 셀룰로오스 크로마토그래피, 소수성 상호 작용 크로마토그래피, 친화성 크로마토그래피, 히드록시 아파타이트 크로마토그래피, 면역 정제(immunopurification) 방법 등을 포함하는, 그러나 이들에 제한 없이, 당 업계에 공지된 표준 기술에 의해 정제될 수 있다. 바람직한 실시예에서, 재조합 CRM197의 정제는 소수성 상호 작용 크로마토그래피 및/또는 음이온 교환 크로마토그래피를 포함한다.
- [0069] CRM197의 수율은 모세관 갤 전기영동 및 웨스턴 블로트 분석과 같은 당업자에게 공지된 방법에 의해 측정할 수 있다. 또한, 활성 분석은 단백질 수율에 관한 정보를 제공할 수 있다. 단백질 수율의 유용한 측정은 배양 용적당 재조합 단백질 양(예를 들어, 단백질 그램/배양 리터), 활성 단백질의 퍼센트 또는 분획(예를 들어, 활성 단백질/분석에 사용된 단백질 양), 전체 세포 단백질의 퍼센트 또는 분획, 단백질 양/세포 및 건조 바이어매스의 퍼센트 또는 비율을 포함한다.
- [0070] CRM197을 평가하기 위한 활성 분석은 당 업계에 공지되어 있으며, 문헌에 기술되어 있으며, 면역 분석, 예를 들어, 웨스턴 블로트 분석 및 ELISA뿐만 아니라 수용체 결합 분석, 예를 들어, 디프테리아 독소 수용체(proHB-EGF) 결합을 포함할 수 있다. 일 실시예에서, 활성은 분석된 전체량(즉, 분석에서 사용된 CRM197의 전체량에 대하여 분석에 의해 활성이 있는 것으로 측정되는 CRM197의 양을 기반으로)과 비교하여 추출 상청액에서 % 활성 재조합 CRM197 단백질로 표시하였다. 다른 실시예에서, 활성은 표준, 예를 들어, 네이티브 단백질(즉, 각각의 샘플로부터 동일 양의 단백질이 분석에 사용되는 표준 샘플에서 활성 단백질의 양에 대하여 상청액 추출 샘플에서 활성 CRM197 단백질의 양을 기반으로)과 비교하여 추출 상청액에서 % 활성 재조합 CRM197 단백질로 표시하였다. 실시예에서, 재조합 CRM197 단백질의 약 60% 내지 약 100%, 약 70% 내지 약 100%, 약 80% 내지 약 100%, 약 90% 내지 약 100%, 약 95% 내지 약 100%, 또는 99% 내지 약 100%는 활성이 있는 것으로 측정되었다.
- [0071] 또한, CRM197의 정체(identity)를 확인하는 수단은 당해 분야에 공지되어 있으며, 예를 들어 단백질은 MALDI-TOF 질량 분석법, N-말단 서열 분석 또는 웹티드 맵핑을 사용하는 웹티드 질량 평거프린트에 의해 분석할 수 있다.
- [0072] 다음은 본 발명의 바람직한 실시예들이다.
- [0073] 감소된 게놈 대장균 K12 균주 숙주에서 재조합 CRM197을 생산하는 방법은 재조합 CRM197 단백질의 발현에 적합한 조건하에서, 발현 제어 서열에 작동 가능하게 연쇄된 감소된 게놈 대장균의 주변 세포질에 CRM197 단백질의 전이를 통제하는 OmpF 또는 YtfQ 신호 서열을 인코딩하는 뉴클레오티드 서열에 융합된 CRM197 단백질을 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터를 포함하는 감소된 게놈 대장균을 배양하는 것을 포함함으로써, 가용성 CRM197 리터당 적어도 1g, 바람직하게는 2g의 수율이 수득되고, 배양 조건은 동물의 혈청 또는 다른 동물의 부산물이 없는 최소 배양 배지에서 대장균 숙주 세포를 배양하는 것을 포함한다.

[0074]

감소된 게놈 대장균 K12 균주 숙주에서 재조합 CRM197을 생산하는 방법은 재조합 CRM197 단백질의 발현에 적합한 조건하에서, 발현 체어 서열에 작동 가능하게 연쇄된 감소된 게놈 대장균의 주변 세포질에 CRM197 단백질의 전이를 통제하는 OmpF 또는 YtfQ 신호 서열을 인코딩하는 뉴클레오티드 서열에 융합된 CRM197 단백질을 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터를 포함하는 감소된 게놈 대장균 K12 균주를 배양하는 것을 포함함으로써, 가용성 CRM197 리터당 적어도 1g, 바람직하게는 2g의 수율이 수득되고, 상기 감소된 대장균 K12 균주는 적어도 다음의 DNA 세그먼트: 대장균 K-12 균주 MG1655의 b0245-b0301, b0303-b0310, b1336-b1411, b4426-b4427, b2441-b2450, b2622-b2654, b2657-b2660, b4462, b1994-b2008, b4435, b3322-b3338, b2349-b2363, b1539-b1579, b4269-b4320, b2968-b2972, b2975-b2977, b2979-b2987, b4466-4468, b1137-b1172, b0537-b0565, b0016-b0022, b4412-b4413, b0577-b0582, b4415, b2389-b2390, b2392-b2395, b0358-b0368, b0370-b0380, b2856-b2863, b3042-b3048, b0656, b1325-b1333, b2030-b2062, b2190-b2192, b3215-b3219, b3504-b3505, b1070-b1083, b1878-b1894, b1917-b1950, b4324-b4342, b4345-b4358, b4486, b0497-b0502, b0700-b0706, b1456-b1462, b3481-b3484, b3592-b3596, b0981-b0988, b1021-b1029, b2080-b2096, b4438, b3440-b3445, b4451, b3556-b3558, b4455, b1786, b0150-b0153 및 b2945가 제거되며, 선택적으로 다음의 추가적인 변형: (i) b4017, b4018 및 b3643의 제거 및 (ii) b4488의 위치 982에서 AT 디뉴클레오티드의 삽입을 가지며, 배양 조건은 동물의 혈청 또는 다른 동물의 부산물이 없는 최소 배양 배지에서 대장균 숙주 세포를 배양하는 것을 포함한다.

[0075]

감소된 게놈 대장균 K12 균주 숙주에서 재조합 CRM197을 생산하는 방법은 재조합 CRM197 단백질의 발현에 적합한 조건하에서, 발현 체어 서열에 작동 가능하게 연쇄된 감소된 게놈 대장균의 주변 세포질에 CRM197 단백질의 전이를 통제하는 OmpF 또는 YtfQ 신호 서열을 인코딩하는 뉴클레오티드 서열에 융합된 CRM197 단백질을 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터를 포함하는 감소된 게놈 대장균 K12 균주를 배양하는 것을 포함함으로써, 가용성 CRM197 리터당 적어도 1g, 바람직하게는 2g의 수율이 수득되고, 상기 감소된 대장균 K12 균주는 적어도 다음의 DNA 세그먼트: 대장균 K-12 균주 MG1655의 b0245-b0301, b0303-b0310, b1336-b1411, b4426-b4427, b2441-b2450, b2622-b2654, b2657-b2660, b4462, b1994-b2008, b4435, b3322-b3338, b2349-b2363, b1539-b1579, b4269-b4320, b2968-b2972, b2975-b2977, b2979-b2987, b4466-4468, b1137-b1172, b0537-b0565, b0016-b0022, b4412-b4413, b0577-b0582, b4415, b2389-b2390, b2392-b2395, b0358-b0368, b0370-b0380, b2856-b2863, b3042-b3048, b0656, b1325-b1333, b2030-b2062, b2190-b2192, b3215-b3219, b3504-b3505, b1070-b1083, b1878-b1894, b1917-b1950, b4324-b4342, b4345-b4358, b4486, b0497-b0502, b0700-b0706, b1456-b1462, b3481-b3484, b3592-b3596, b0981-b0988, b1021-b1029, b2080-b2096, b4438, b3440-b3445, b4451, b3556-b3558, b4455, b1786, b0150-b0153, b2945, b0315-b0331, b0333-b0341, b0346-b0354, b2481-b2492, b2219-b2230, b4500, b3707-b3723, b0644-b0650, b4079-4090, b4487, b4092-b4106, b0730-b0732, b3572-b3587, b1653, b2735-b2740, b2405-b2407, b3896-b3900, b1202, b4263-b4268, b0611, b2364-b2366, b0839, b0488-b0500 및 b0502가 제거되며, 선택적으로 다음의 추가적인 변형: (i) b4017, b4018 및 b3643의 제거 및 (ii) b4488의 위치 982에서 AT 디뉴클레오티드의 삽입을 가지며, 배양 조건은 동물의 혈청 또는 다른 동물의 부산물이 없는 최소 배양 배지에서 대장균 숙주 세포를 배양하는 것을 포함한다.

[0076]

### 실시예 1

[0077]

### 감소된 게놈 대장균 숙주에서 불용성 CRM197의 세포질 발현

[0078]

CRM197은  $\beta$  과자의 다수의 리소젠으로부터, 또는 슈도모나스 형광성(*pseudomonas flurorescens*)에서 재조합 플라스미드 시스템으로부터 발현된 코리네박테리움 디프테리아 C7의 발효에 의해 현재 제조된다. C. 디프테리아에서 CRM197의 수율은 낮고(많아야 ~ 200mg/ℓ) 생물학적 안정성 레벨 2(BSL2) 시설을 요구한다. *P.* 형광성에서 생산은 높은 수율(약 2g/L)의 결과를 가져온다; 그러나, 두 숙주는 병원성 기능을 가지는 다수의 모바일 요소, 암호를 지닌 프로파지 및 유전자 잔재를 유지한다. 박테리아 발효에서, 삽입 서열(IS) 요소의 이동성은 관심의 유전자를 불활성화시키는 삽입을 야기할 수 있다. 그 결과, 발효 실패 또는 절단된 산물의 불필요한 발현이 있을 수 있으며, 이 모두 경제적 문제 및 잠재적 위험이 있다. 또한, 그의 독성 모(parent)에 CRM197의 반전은 최악의 결과를 초래할 수 있다. CRM197의 반전은 조직 배양 세포에서 검출된 독성 활동에 기여할 수 있다(Qiao 외., 2008). 따라서, 감소된 변이율을 갖는 발현 시스템은 반전에 대한 가장 낮은 수준의 위험성을 결부하여 생각되는 가장 높은 신뢰성 및 생산성을 제공할 수 있다.

[0079]

CRM197의 높은 가격과 공급 부족에 기여하는 하나의 큰 요인은 주력 대장균의 생산에서 다량의 CRM197을 생성하는 역사적 능력이 없다는 것이다. CRM197은 박테리아의 세포질에서 발현될 경우 불용성이며 표준 상업용 대장균 균주에서 만들 경우 사용하기 전에 채-풀딩을 요구한다. 비교적 적은 양의 CRM197이 종래의 균주에서 생산되기

때문에, 감소된 게놈 대장균 균주, MDS42는 진탕 플라스크 배양에서 불용성 CRM197을 위한 숙주 생산으로 시험되었다.

[0080] 감소된 게놈 균주 MDS42는, 본 명세서에 참고로 포함된, 국제 특허 공개 공보 WO 2003/070880에 기술된 방법에 의해 생산하였다. 간략하게, 감소된 게놈 균주의 시리즈(MDS01-MDS39)는 모 균주 대장균 MG1655로부터 핵산 서열의 39 누적 제거(대략 게놈의 약 14.1%)의 시리즈를 만듦으로써 생산하였다. K-12 서열 및 IS 데이터베이스의 모든 서열을 포함하는 게놈 스캔 칩(NimbleGen 시스템, Madison, WI)에 대한 교배는 모든 IS 요소를 결여하도록 설계된 최초 균주인, MDS39는 그의 생산 중 새로운 위치로 전위한 IS 요소의 추가적인 사본을 예기치 않게 포함된 것으로 나타났다. 이러한 IS 요소는 MDS40을 생산하기 위해 제거하였다. *fhuACDB*(tonA 유전자 자리)는 MDS41을 생산을 위해 MDS40에서 제거하였다. MDS01-MDS41 생산하기 위해 만들어진 누적 제거 각각의 위치 및 기능은 본 명세서에 그 전체내용이 참조로 포함된, 미국 특허 번호 8,178,339의 표 2에서 확인할 수 있다. 그리고 나서 *endA* 유전자는 MDS42를 생산을 위해 MDS41에서 제거하였다. MDS69를 생성하기 위해 MDS42에서 스몰일곱 추가 핵산 제거를 하였다. MDS42 및 MDS42(MDS43, MDS44 ... MDS69 등)를 기반으로 하는 모든 균주는 삽입 서열이 없다.

[0081] 불용성 CRM197의 생산을 위해, CRM197 서열에서 헤어핀 구조의 방출을 야기하는 DNA 서열 변화를 포함하는 변형된 CRM197 서열은 사용하였다. 최적화된 CRM197 서열은 번역 개시를 억제하며 시작 부위(ATG)와 리보솜 결합 부위(RBS) 모두의 인식을 향상시키는 이차 구조를 제거한다. 도 1을 참조하라.

[0082] 네이티브 CRM197 신호 서열은 제거하고, 최적화된 CRM197 서열(cyto-CRM197, SEQ ID NO: 3)은 PCR로 증폭하고, 카나마이신 저항성 카세트를 포함하고 복제된 서열의 발현을 구동하는 락토오스 유도성 프로모터를 사용하는 발현 벡터 pSX2로 보조 복제(subclone)하였다. CRM197(그의 네이티브 신호 서열을 결여한)을 포함하는 플라스미드 pSX2는 감소된 게놈 대장균 균주 MDS42로 형질 전환하였고 진탕 플라스크 배양으로 검사하였다. 요약하면, 3 ml 배양은 0.2% 포도당 및 50 µg/ml 카나마이신으로 상보되고 초기 OD<sub>600</sub>=0.075로 20ml 배양을 접종하는데 사용되는 Korz 최소 배양 배지(Korz DJ 외., J. Biotechnol., 39(1):59-65(1995))에서 포화로 성장하였다. 그 후, 최적 수준의 불용성 세포질 CRM197을 생산하는 성장 온도와 인ду서(IPTG)의 농도는 진탕 플라스크를 사용하여(플라스미드가 선택할 수 있는 항생제 카나마이신으로 상보된 최소 배양 배지에서) 측정하였다. 최적 IPTG 농도는 250 µM인 것으로 측정되었다. 도 2는 전체 세포 단백질(T) 및 가용성 분획(S) 및 불용성 분획(I)을 분석하는 단백질 겔의 일례이며, IPTG의 첨가 이전에 0.5의 OD<sub>600</sub>(후기 유도)으로 성장한 세 개의 배양에서 전체 세포 단백질의 고속 원심 분리가 뒤따른다. 놀랍게도 많은 양의 cyto-CRM197이 불용성 분획(도 2, 화살표 참조)에 나타났다. 단백질 표준에 대해 정량화하면, 진탕 플라스크 결과는 OD<sub>600</sub>의 적당한 발효에서 10 내지 12g/L의 cyto-CRM197을 예측한다. 감소된 게놈 대장균 숙주 세포에서 불용성 CRM197 생산은 종래 대장균 균주보다 10배 높았다.

### 실시예 2

#### 감소된 게놈 대장균 숙주에서 가용성 CRM197의 주변 세포질 발현

[0083] 다음으로, 감소된 게놈 대장균 균주에서 가용성 CRM197의 생산은 주변 세포질 공간에 CRM197의 발현을 통제함으로써 시험하였다. CRM197은 대장균에서 가용성 형태로 생산하기 어렵다는 것을 주지로 입증했다. 올바른 단백질 폴딩 및 이황화물 교상결합의 형성을 위한 최적의 비-감소 환경을 제공함으로써 주변 세포질 공간에 대해 고도로 발현된 단백질의 보급(export)은 안정성을 지원한다. 이를 위해, 공동 발현된 사프론 단백질의 수와 결합하여, 여섯 신호 서열을 신호 서열 및 CRM197의 주변 세포질 전달의 높은 수준을 부여하는 사프론 단백질을 확인하기 위해 검사하였다. 도 3은 검사된 신호 서열과 공동 발현 사프론 단백질을 도시한다. 검사된 신호 서열은 세 개의 대장균 분비 경로 각각으로부터 대표적인 신호 서열을 포함하였다.

[0084] 대장균에 대해 최적화된 코돈(SEQ ID NO: 1)인 CRM197 오픈 리딩 프레임(ORF)은 DNA 2.0(Menlo Park, CA)에서 주문하였다. CRM197 ORF는 Pe1B 신호 서열을 인코딩하는 서열에 의해 선행되었다. *pe1B* 및 CRM197 ORF는 pSX2 발현 벡터에 클로닝을 촉진하기 위해 설계된 서열의 측면에 배치(f1ank)되었다. 5' 플랭킹 서열-Pe1B 신호 서열-CRM197 ORF(종결 코돈을 포함하는)-3' 플랭킹 서열은, 플랭킹 서열은 밀줄로, Pe1B 신호 서열을 인코딩하는 뉴클레오티드 서열은 굵게, 그리고 CRM197 ORF는 일반 텍스트로서 아래 표 1에 제공한다.

[0085] 표 1 : *pe1B*(굵게)-CRM197 뉴클레오티드 서열(일반 텍스트)+플랭킹 서열(밀줄)

[0086] CCTCTAGAAATAATTTGTTAACCTTTGAAGGAGATACATATGAAATACTTGCTGCCAACGCCGCCCTGCTGCTCGCAGCACAGCCGGC

TATGGCAGGCTATTCCGCTGGTCGGCGAAGTGGTGATATCGGCTTGCTGCCTACAACCTCGTTGAAAGCATCATTAAACCTGTTCAGGTGGTCCACAACA  
GCTATAATGCCCGCGTACAGCCGGGTACAAGACCCAACCGTCTCGCACGATGGCTATCGGGTGTCTGGAACCGTCGAAGATAGCATCATTGTA  
CCGGTTCCAGGGCGAGCGGGCATGACATCAAGATTACTGCAGAAAATACCCCGCTGCCATCGCAGGTGTCCTGCTGCTACGATTCCGGTAAGCTGG  
ACGTTAACAAAAGCAAACCCACATTCTGTGAACGGTCGTAAGATTGCATGCCTGCGATTGACGGCAGCTCACCTCTGCCGTCCGAAGAGGCC  
CGGTCTACGTTGGTAAAGGTGTCACGCGAACCTGCACGTGGCTTACCGCAGCAGCTCGAGAAAATCATAGCAATGAGATTCTAGGCACAGCATTG  
GCGTTCTGGTTACCAAAAGACGGTGACCACCAAAACTGAGCCTGTTGAGATCAAAGCTAACTCGAGCCCCAAGGGCAGACC  
CCCT

[0089] *Pe1B-CRM197* ORF를 인코딩하는 뉴클레오티드 서열은 pSX2 벡터에 클로닝하기 위해 필요한 측면 영역(flanking region)을 생성하기 위해 센스 프라이머(GGAGATATACATATGAAATACTGCTGCCAACC)와 안티-센스 프라이머 CTTGTTAGCAGCCGATTAGCTTTGATCTCAAAGAACAA를 갖는 DNA 2.0 클론으로부터 증폭된 PCR이다.

[0090] 대안적인 신호 서열은 2 또는 3스텝 PCR 처리를 사용하여 CRM197 ORF에 융합시켰다. 첫 번째 스텝에서, 신호 서열의 C-말단 코딩 영역 및 CRM197의 N-말단 코딩 영역 모두를 커버하는 센스 프라이머는 CRM197의 C-말단 코딩 영역을 커버하는 안티-센스 프라이머와 함께 사용하였다. 두 번째 스텝에서, 신호 서열의 ORF를 완료하는 프라이머는 CRM197의 C-말단 코딩 영역을 커버하는 동일한 프라이머로 사용하였다. *OmpA-CRM197* 구조의 경우에는, 신호 서열의 N-말단 영역을 커버하는 짧은 프라이머 및 CRM197의 C-말단 코딩 영역을 커버하는 동일한 프라이머를 포함하는 세 번째 스텝을 사용하였다. CRM197 ORF에 대장균 *ompA* 및 *OmpF* 신호 서열을 융합하는 데 사용되는 프라이머는 아래에 기술한다.

[0091] 다음의 프라이머는 CRM197 ORF에 대장균 *ompA* 인코드된 신호 서열을 융합하는데 사용하였다. 스텝 1에 대해, 센스 프라이머=5'-  
GCTACCGTAGCGCAGGCCGGTGGGATGATGTTGTGGA-3' 및 안티센스 프라이머=5'-  
CTTGTTAGCAGCCGATTAGCTTTGATCTCAAAGAACAA-3'. 스텝 2에 대해, 센스 프라이머 5'=-  
GGAGATATACATATGAAAAGACAGCTATCGCATTGCACTGGCAGTGGCTGGCTGGCTACCGTAGGCCAGGCC-3' 및 안티센스 프라이머  
5'=-CTTGTTAGCAGCCGATTAGCTTTGATCTCAAAGAACAA-3'. 스텝 3에 대해, 센스 프라이머 5'=-  
GGAGATATACATATGAAAAGACAGCTATCG-3' 및 안티센스 프라이머 5'=-CTTGTTAGCAGCCGATTAGCTTTGATCTCAAAGAACAA-3'.

[0092] 다음 프라이머는 CRM197 ORF에 *ompF* 인코드된 신호 서열을 융합하는데 사용하였다. 스텝 1에 대해, 센스 프라이머 5'=-GTTAGTAGCAGGTACTGCAAACGCTGGTGGGATGATGTTGTGGA-3' 및 안티센스 프라이머 5'=-  
CTTGTTAGCAGCCGATTAGCTTTGATCTCAAAGAACAA-3'. 스텝 2에 대해, 센스 프라이머 5'=-  
GGAGATATACATATGAAAGCGCAATATTCTGGCAGTGCATCGCCCTGCTCTGTTAGCAGGTACTGCAAACGCT-3' 및 안티센스 프라이머  
5'=-CTTGTTAGCAGCCGATTAGCTTTGATCTCAAAGAACAA-3'.

[0093] 완료된 신호 서열-CRM197 PCR 산물은 pSX2 발현 벡터에 복제하였다. 신호 서열-CRM197 PCR 산물의 말단은 pSX2 벡터의 서열을 중첩시키는 서열의 15bp를 소유하였다. pSX2 벡터는 복제 반응을 촉진하는 제한 효소 *Kpn* I 및 *Sac* I로 선형화 하였다. 복제 반응은 재조합 pSX2 발현 벡터를 생성하기 위해, *MDS42*, *MDS42recA* 또는 *IS609*의 추가적인 제거를 갖는 *MDS42recA*로 형질 전환하였다. 신호 서열-CRM197 영역과 플랫킹 벡터 서열은 서열 분석에 의해 확인하였다.

[0094] 도 3(B)에 예시된(그의 네이티브 신호 서열을 결여한)CRM197 서열과 신호 서열의 조합을 함유하는 플라스미드 pSX2는 감소된 게놈 대장균 균주 *MDS42* 또는 *MDS42recA*(*recA* 유전자(*recA1819* 대립 유전자)의 제거를 갖는 *MDS42* 균주)로 형질 전환하여 진탕 플라스크 배양으로 검사하였다. 신호 서열 및 샤프론 단백질에 추가하여, 검사된 배양 변수는 온도, 인듀서(IPTG) 농도 및 인듀서가 첨가(초기[0.01의 OD<sub>600nm</sub>] 또는 후기[0.4의 OD<sub>600nm</sub>])되는 시점을 포함한다. 다음의 조건은 CRM197의 주변 세포질 분비에 최적인 것으로 측정하였고 이러한 조건은 후속 실험에 사용하였다: (i) 짧은 37°C 배양(예를 들어, 2시간)에 의해 전 처리된 약 25°C의 성장 온도 (ii) 후기 유도(약 0.4의 OD<sub>600</sub>으로 IPTG 첨가) 및 (iii) 15와 35 μM(cyto-CRM197의 최적 발현을 위해 요구되는 약 1/10) 사이 인듀서(IPTG) 농도.

[0095] 요약하면, 3ml 배양은 0.2% 포도당 및 50 μg/ml 카나마이신으로 상보되고 초기 OD<sub>600</sub>=0.075로 20ml 배양을 접종하는데 사용되는 Korz 최소 배양 배지(Korz DJ 외., J. Biotechnol., 39(1):59-65(1995))에서 포화로 성장하였다. 20ml의 배양(차단된 125ml 삼각 플라스크에서)은 2시간 동안 37°C 진탕 배양기(250rpm)에 배치하였다. 그 후, 배양은 25°C 진탕 배양기로 변경하고 OD<sub>600</sub>이 0.3-0.4 사이 때까지 관찰하였다. 이때, IPTG는 지정된 농도로 첨가하였다. 유도된 배양은 25°C 진탕 배양기에서 밤새 배양하였다. 총 유도 시간은 18~22시간 사이였다. 유도

후, 배양의 OD<sub>600</sub>을 측정하였다. 2 OD 단위를 나타내는 배양 분취량은 주변 세포질 샘플들을 생성하기 위해 처리하였다. 주변 세포질 샘플은 원형질막 완충액(Epicentre, Madison, WI)의 도움으로 제조하였다. 2 OD 샘플을 1.5ml 에펜 도르프 투브에서 10분 동안 7500xg으로 원심 분리에 의해 수화하였다. 상청액을 제거하고, 세포 펠릿을 50 μl 원형질막 완충액(200mM Tris-HCl[pH 7.5, 20% 자당, 1mM EDTA, 및 30U/μl Ready-Lyse 리소자임]에서 조심스럽게 재 혼탁하였다. 실온에서 5분 후, 얼음냉수 50μl를 재 혼탁된 펠릿에 빠르게 첨가하였다. 혼합물을 15분 동안 4000xg으로 원심 분리하여 스페로플라스트로부터 주변 세포질 분획을 분획하기 전에 5분 동안 얼음 위에서 배양하였다. 주변 세포질 분획을 나타내는 상청액을 SDS-PAGE 분석을 위해 제조하였다. 0.12 OD 단위에 상당하는 양을 레인당 적재하였다.

[0096] 주변 세포질로 CRM197의 최고 분비를 야기시키는 가장 성공적인 신호 서열 및 유도 특성은 도 4에 도시하였다. 주변 세포질 신호 OmpA 및 OmpF는 주변 세포질로 CRM197의 가장 큰 이동을 촉진하게 하는 것으로 확인하였다. 세 가지 공동 발현 샤프론 단백질의 어떤 한 것도 주변 세포질 전달에 별다른 영향이 없었다(발현이 도 4에서 YccA가 있고 없고 비교한 예와 같이). OmpA 및 OmpF는 주변 세포질 CRM197의 대략 유사한 양을 야기시키는 것으로 나타났기 때문에, OmpA-CRM197은 후속 실험에 사용하였다.

[0097] *ompA* 및 *ompF*를 포함하는 *sec*-의존성 경로의 구성 요소의 발현이 이화 작용 산물 억제를 행할 수 있기 때문에, *ompA*-CRM197의 생산을 위한 탄소원으로서 글리세롤의 영향은 상기에 기술된 조건하에서 전탕 플라스크 배양에서 감소된 계놈 대장균 균주 MDS42에서 포도당과 비교하였다. 도 5에서 예시된 바와 같이, 글리세롤로 상보된 최소 배양 배지는 포도당에 비해 크게 낮은 수준의 CRM197 발현을 야기했다. 따라서, 포도당은 모든 후속 실험에서 탄소원으로 사용하였다.

[0098] 다음으로, 여러 가지 다른 감소된 계놈 대장균 숙주 세포에서 주변 세포질 CRM197의 생산을 비교하였다. 따라서, MDS42 또는 MDS69 균주 배경(background) 내 일련의 제거는 (i) 세포 대사를 최적화하는 제거 또는 (ii) CRM197 발현에 부정적인 영향을 미칠 수 프로테아제(예를 들어, Blon)의 수준을 제거 또는 감소하는 제거를 함유한 MDS42에 대해 상기 기술된 최적의 조건을 기반으로 전탕 플라스크 배양에서 주변 세포질(가용성) CRM197의 생산에 미치는 그들의 영향을 검사하였다. MDS42를 기반으로 다음의 감소된 계놈 대장균 균주를 시험하였다: MDS42recA, MDS42metab, 및 MDS42Blon/metab. MDS42metab는 (i) *icIR*(NCBI Entrez GeneID 번호 948524에 기술된, b-번호 b4018) 및 *arpA* 유전자(NCBI Entrez GeneID 번호 944933에 기술된, b-번호 b4017)를 제거 (ii) *rph* 유전자(b3643)를 제거(따라서, 하류 *pyrE* 유전자의 전사 수준을 증가시키는), 및 (iii) 위치 982(활성 아세토하이드록시 산 신타아제 II의 발현을 야기시키는)에서 AT 디뉴클레오티드를 삽입에 의해 *ilvG*의 프레임시프트 변이를 보정함으로써 생산하였다. MDS42Blon/metab는 MDS42metab 대해 기술된 변형뿐만 아니라 원형 대장균 *lon* 프로모터의 -10 영역에서 -35 영역을 단리하는 B 균주 대장균의 *lon* 프로모터 영역의 서열을 모방하는 *lon* 프로테아제(b0439) 프로모터의 변형도 함유한다. MDS69를 기반으로 다음의 감소된 계놈 대장균 균주를 시험하였다: MDS69metab(MDS42metab에 대해 상술한 바와 같이 변형된 MDS69 균주), MDS69Blon/metab(Blon 프로테아제 변형을 포함하기 위해 추가로 교번된 MDS69metab, MDS69lpp/metab(지질 단백질 lpp(b1677)를 제거하기 위해 추가로 변형된 MDS69metab, 및 MDS69Blon/lpp/metab(Blon 프로테아제 변형 및 지질 단백질 lpp 유전자 제거 모두를 포함하도록 추가로 변형된 MDS69metab).

[0099] 도 6은 이들 균주에서 OmpA-CRM197 발현을 비교한다. 검사된 여덟 균주 중, 최고 수준의 CRM197 발현은 대사 활성(대사 균주) 향상을 목표로 제거를 함유하는 이들 균주(MDS42 또는 MDS69 배경에서)에서 분명하였다. 그러나, 시험된 모든 균주는 전체 세포 단백질 제제(도 6 패널 A와 C)에서 단백질 겔 상에서 또한 분명한 놀랍게도 다양한 주변 세포질 CRM197(도 6 패널 B와 D)을 함유하였다. 중요한 것은, 이들 결과는 프로테아제 제거 Blon은 대사 균주에서 CRM197 수준에 영향을 미치지 않았다는 것을 나타낸다. 또한, *sec*-의존성 주변 세포질 수송 시스템을 “해소(free up)”하기 위해, 그 부재로, 생각되는 매우 풍부한 지질 단백질 Lpp의 제거는 주변 단백질 CRM197 수준에 영향을 미친다는 것을 확인하지 못하였다. 이러한 실험에서 배양 배지는 후-유도를 단리하였고 CRM197 방출에 대해 검사하였다. CRM197은 검사된 어떤 균주로부터의 배양 배지에서 확인되지 않았다. 표 2는 가장 높은 주변 세포질 CRM197 발현 수준을 생성한 MDS 균주에 대한 수율 결과의 요약이다. 이러한 결과는 동일한 겔에서 실행된 단백질 표준과 지적된 네 가지 균주로부터 CRM197의 균주 강도를 비교함으로써 수득하였다. 전탕 플라스크 값은 100 또는 200 OD<sub>600</sub>에 도달하는 발효에서 CRM197의 양을 예측하기 위해 추정하였다. 표 2의 나타낸 네 가지 균주는 이들 균주가 종래 균주에서 현재 가능한 것보다 훨씬 더 많은 CRM197을 생성하는 능력을 가지고 있음을 시사하는 유가 발효에서 일반적으로 300의 ODs에 도달한다.

표 2

pSX2-ompA CRM197를 갖는 균주 발현	주변 세포질 샘플	적재된 #ODs	캘리브레이트 된 용량 ngs/레인	주변세포질 에서 타겟 단백질%	100 ODs 에서 평균 g/L	200 ODs 에서 평균 g/L
MDS42 대사	Peri, Late 25 $\mu$ M IPTG, 25°C o/n	0.06 0.03	1102 499	48% 53%	1.75	3.5
MDS69 대사	Peri, Late 25 $\mu$ M IPTG, 25°C o/n	0.06 0.03	1047 449	43% 46%	1.62	3.24
MDS42 △프로테아제	Peri, Late 35 $\mu$ M IPTG, 25°C o/n	0.06 0.03	825 330	47% 45%	1.24	2.48
MDS69 △프로테아제	Peri, Late 25 $\mu$ M IPTG, 25°C o/n	0.06 0.03	1028 506	28% 30%	1.7	3.4

[0101] CRM197은 높은 품질의 CRM197 도전(Bishai 외., J. Bacteriol., 169:5140-51(1987); 운반 단백질의 재조합 생산, GEN News, Dec. 1, 2012)의 생산을 제공한 단백질 분해 분열에 매우 민감하다. 실험의 개별 세트에서, 주변 세포질 CRM197의 생산은 감소된 게놈 대장균 균주에서 프로테아제 유전자의 표적된 제거가 주변 세포질에서 CRM197의 증가를 초래하는지 여부를 측정하기 위해 프로테아제 제거 균주의 시리즈에서 검사하였다. 따라서, 유전자를 인코딩 다음의 프로테아제는 개별적으로 조합하여 제거하였다: *degP*(b0161), *prc*(b1830), *hptX*(b1829)뿐만 아니라 *lon* 프로모터 영역의 부분. 프로테아제 유전자의 제거는, 개별적으로 또는 조합하여, CRM197 발현 수준에 영향을 미치지 않았다. 프로테아제 유전자의 특정 조합을 제거하도록 변형된, 감소된 게놈 대장균 균주 MDS42는 CRM197의 주변 세포질 발현에 아무런 영향을 미치지 않는다는 것을 예시하는 도 7을 참조하라. 이 데이터는 MDS42 또는 MDS69을 기반으로 감소 게놈 대장균에서 생산 시 CRM197의 단백질 분해 분열은 이들 균주에서 프로테아제 활성의 낮은 수준으로 인해 아마도 발생하지 않는 것을 나타낸다.

### 실시예 3

#### 유가 발효에서 CRM197 생산

[0102] 다음으로, 감소된 게놈 대장균 균주에서 CRM197의 상업적 규모-상승을 검사하였다. 따라서, MDS42 대사 균주에서 OmpA-CRM197은 10리터 규모로 정의된 최소 배양 배지에서 유가 발효를 행하였다. 발효 조건은 0.18 OD로 접종되는 37°C에서 배치(batch) 단계를 포함하며 배치 배양 배지에서 1% 포도당이 소비될 때 까지(~ 7.5 시간) 성장하도록 허용했다. 유가 배치 단계는 배치 배양 배지가 포도당이 고갈될 때 발생하는 DO 스파이크에 의해 촉발되었다. 공급(feed)은 중량 측정(~ 12.5시간) 적으로 제어된 0.3Mu(1/h)의 성장 속도를 생산하는 지수 공급 속도(exponential feed rate)로 시작하였다. 유도 시점은 유효한 인산이 거의 고갈된 시점인 것으로 측정하였다. 유도 시점 이전 약 2시간 시점에서, 온도는 25°C로 변경시키고, 공급 속도는 0.2Mu(1/hr)의 성장 속도를 생산하는 속도로 저하시켰다. 인듀서(100uM)를 첨가하면, 공급은 약 7시간 동안 시간당 80g의 포도당이 첨가되도록 일정 속도로 변경하였다. 발효 OD<sub>600</sub>은 300에 접근하였고 도 8에 예시된 바와 같이 타겟된 CRM197 주변 세포질의 매우 높은 수준을 발생하였다. 최적 조건에서 두 번째 발효는 발효 시험에서 높은 수준의 일관성을 나타내는 약 2g/L의 주변 세포질 CRM197 수준을 야기시켰다.

[0103] 상기 기술된 결과는 진탕 플라스크 및 10L 유가 발효 모두에서 MDS42 및 MDS69와 같은 감소된 게놈 대장균 생산 숙주에서 수득된 가용성 CRM197의 놀라운 수율을 보여준다.

[0104] 예비 발효 분석 시 관찰된 하나의 문제는 전체 세포 단백질 단리에서 CRM197의 가용성 형태의 감소였다. 주변 세포질 단리 방법은 대규모로 적용 가능하지 않기 때문에, 가용성 CRM197 단리의 일반적인 방법이 개발되었다. 초기 실험은 가용성 형태로 단리될 수 있는 불용성인 종래의 전체 세포 단백질(TCP) 단리의 결과 CRM197이 관찰되는지 여부를 측정하기 위해 수행하였다. 따라서, MDS42 *recA* 균주에서 OmpA-CRM197은 상술한 바와 같이(인듀서의 첨가 이전에 25°C에서 단기간 배양이 뒤따르는 37°C에서 배양을 포함하는) 10리터 규모로 정의된 최소 배양 배지에서 유가 발효시켰다. 주변 세포질 CRM197의 많은 양을 함유하는, 세포는 전체 세포 단백질(TCP)을 단리하기 위해 상업적으로 유효한 비-이온 세제-기반 완충액으로서 표준 세제 침지를 하였다. 전체 세포 단백질 샘플은 고속(21k g)으로 10분 동안 원심 분리하여 가용성 분획을 단리하였다. TCP의 샘플과 가용성 분획의 샘플을 분석하였다. 도 9에 도시된 바와 같이, 패널 A는, CRM197의 가용성 주변 세포질 형태는 세제 균질화에 의해 완전히 불용성을 제공하였다. 반대로, 주변 세포질 제제(상술한 바와 같이)가 고속 원심 분리되었을 경우, 주변 세포질 CRM197은 예상대로 가용성 형태로 유지하였다. 도 9, 패널 B를 참조하라.

[0105] 불용성인 CRM197의 분획을 회수하기 위한 시도로, 세제-기반 박테리아 세포 용해는 세제 용해에 비해 CRM197을

생성하기 위한 생산-수준 플랫폼에 더 전도성이 있고 상업적 규모-상승 과정에서 주변 세포질을 단리시킬 필요를 제거할 수 있는 기계적 방법의 세포 용해와 비교하였다. 또한, 하기 표 3에 기술된 바와 같이 용해는 단백질의 가용화를 향상시키는 것으로 알려진 화학제의 존재하에서 수행하였다.

[0109] 표 3: 용해 방법 및 가용화제의 목록.

가용화를 향상하는 제제	초음파 분해 후 가용성 CRM197 %
이미다졸, 250 mM	107%
트레할로오스, 50 or 250 mM	64%, 79%
산화된 상태에 따라 5:1로 감소된 글루타티온	104%
글리세롤, 10%	68%
자당, 10%	75%
제제(agent) 없음	88%

[0110]

[0111] 초음파 분해 및 마이크로 유동화는 50mM TrisHCl 완충액(pH 8)에서 수행하고, 모든 용해 방법은 리소자임과 벤조나제(Novagen, Darmstadt, 독일)의 상업적 혼합물인, Lysonase™의 존재하에 이행하였다. 그 후 표 3에 열거된 각각의 제제는 별도의 제제에서 시험하였다. 도 10은 세제 또는 기계적 용해에 의해 수행된 일련의 단리의 일례이다. 가용성 CRM197은 세제 용해를 사용하여 수득할 수 없었고 가용성 CRM197에서 단지 작은 증가는 가용화제를 포함하는 세제 용해를 사용하여 명백했다. 가용성 CRM197의 양을 적당히 향상시키는 글리세롤 및 자당은 단독 세제(도 10, 패널 A)와 비교할 경우 가용성 분획에서 확인하였다. 그러나 기계적 용해는 가용성 분획에서 CRM197의 수준을 인상적으로 증가시켰다. 사실, CRM197 수준의 인상적인 증가는 초음파 분해(도 10, 패널 B) 또는 마이크로 유동화를 사용하든, 기계적 용해를 받은 모든 샘플에서 가용성 분획에서 명백했다. 또한, 기계적 용해의 결과 수득된 가용성 CRM197의 양은 가용화제(표 3에 모든 다른 제제와 “제제 없음”과 비교)에 의한 것과 현저하게 차이가 없었다. 기계적 용해 방법으로부터 생성된 결과의 집계는, MDS42(25-35  $\mu$ M IPTG로 25°C 및 후기 유도에서 성장이 뒤따르는 깊은 37°C 배양을 포함하는 배양 조건을 사용하는)에서 CRM197이 전체 세포 단백질의 7.2-8.3% 및 6.3과 7.7% 사이의 가용성 단백질을 포함하는 것을 암시한다. 이를 결과는 기계적 용해가 대규모 상업적 발효에 사용되는 세포 파쇄의 표준 방법이고 다량의 가용성 CRM197을 생성하는 능력을 내포하기 때문에 흥미를 자아내고 있다.

[0112] 상기 언급한 데이터를 기반으로, 가용성 CRM197을 생성하기에 적합한 상업적 프로토콜은 25°C에서 세포질 신호서열(예를 들어, *ompA* 또는 *ompF*에 의해 인코드된)에 융합된 CRM197 코딩 서열을 인코딩하는 발현 벡터를 지닌 감소된 게놈 대장균 숙주의 발효를 포함하며, 적당한 완충액(예를 들어, pH ~8에서 50mM Tris-HCl 완충액)에서 기계적 수단(예를 들어, 초음파 분해 또는 마이크로 유동화)으로 용해된, 세포가 저속 원심 분리에 의해 수집된다. 이물질을 제거하는 원심 분리가 뒤따르고, 이어서, 가용성 CRM197은 상청액으로부터 단리하였다. 진탕 플라스크에서 배양은 25°C 그리고 25~35mM IPTG에서 배양하여 CRM197의 95 내지 100%는 가용성 형태로 단리하였다.

[0113] *ompA-CRM197* 융합을 함유하는 발현 벡터를 지닌 감소된 게놈 대장균 균주 MDS69metab(전술한 바와 같이)를 사용하는 발효의 결과의 요약은 하기 표 4에 나타냈다. 이러한 발효는 정의된 최소 배양 배지 및 대수 증식에서 늦게 인ду서 IPTG의 첨가를 사용하는 유가 배양 조건하에서 일어났다. 발효 규모는 10리터였다. 인ду서 농도를 변경함으로써, 주변 세포질 CRM197의 양은 0.5에서 약 2g/L로 증가하였다.

표 4: 포도당 공급을 사용하여 균주 T69 대사의 유가 발효		최대 수율@ 29 시간	
발효	유도 수준	OD	가용성 CRM <sub>197</sub> 수율
Ferm 157	25 $\mu$ M	256	0.74 g/L
Ferm 158	50 $\mu$ M	262	1.62 g/L
Ferm 159	100 $\mu$ M	291	1.96 g/L

[0114]

[0115] 도 11은 100  $\mu$ M 인듀서 농도를 사용한 발효의 특성을 도시한다. 가용성(Sol) 및 불용성(Insol) 분획으로 단리한 전체 세포 단백질(TCP)을 비교하는 이 발효에서 겔은 발효 중(도 12를 참조) 가용성 CRM197의 강력한 발현을 명확하게 나타냈다.

[0116]

감소된 계놈 대장균 숙주 균주의 유가 발효에서 가용성 CRM197의 생산을 위한 최적의 조건은 다음과 같았다. 온도에 관해서는, 인듀서(이 경우에 IPTG)의 첨가 이전에 20과 25°C 사이의 온도 변동이 뒤따르는 37°C에서 배양 함으로써 배치 단계에서 성장의 개시는 최적이었다. 최적의 pH 범위는 6.5와 7.5 사이(예를 들어, 6.5, 7.0 또는 7.5)이다. 최적의 인듀서 농도는 100과 250  $\mu$ M IPTG(성장의 늦은 로그 단계 중 추가) 사이이다. 배양 배지 조건에 관해서는, 최소 배양 배지 조건은 충분하며 저렴한 비용 및 동물 유도 산물이 없는 정의된 조건의 이점을 갖도록 측정하였다. 중요한 것은, 종래의 대장균 균주는 최소한의 배양 배지에서 견고하게 성장하지 않는다는 것이다. 이러한 최적의 조건을 사용하여, 가용성 CRM197의 적어도 4g/L의 목표 수율은 감소된 계놈 대장균 숙주 균주(예를 들어, MDS42 또는 MDS69)를 사용하여 10L 규모 발효에서 확실하게 생산할 수 있는 것으로 추정된다.

[0117]

#### 실시예 4

[0118]

#### CRM197의 다운 스트림 처리

[0119]

감소된 계놈 대장균에서 CRM197의 생산 및 기계적 용해에 뒤따라, CRM197은 정제될 수 있다. 발효 조건하에서 MDS69metab로부터 생산된 CRM197이 정제 의무가 있는지 여부를 측정하기 위해, 작은 규모 정제는 소수성 상호 작용 크로마토그래피(페닐 세파로스)와 음이온 교환 크로마토그래피(DEAE-셀룰로오스)의 조합을 사용하여 수행하였다. 도 11에 나타낸 28시간 발효 샘플의 50 OD 유닛은 10mM 인산나트륨 완충액(pH 7.5) 용액에서 마이크로 유동화기(MF)를 사용하여 균질화하였다. 그 결과로 발생된 균질물을 10분 동안 21,000g로 원심 분리하고, 가용성(IS) 및 불용성 분획을 단리하였다. 웨스턴 블로팅 및 디프테리아 독소(DPX)에 대비하여 다른 항체를 사용함으로써, CRM197은 가용성 분획에서 매우 풍부한 것으로 확인되었다(도 13, 패널 A는 마이크로 유동화기(MF) 및 재 현탁된 불용성(IS) 분획을 세 가지 농도(0.1, 0.07 및 0.04 OD)에서 프리-컬럼 가용성 분획과 비교한다). 가용성 분획(25 OD 상당)은 여과하여(0.45  $\mu$ m), 13%(wt/vol) 황산암모늄을 가져와 10mM의 염화나트륨, 10mM의 인산나트륨 완충액, pH 7.5에서 미리 평형화한 패널 세파로스 칼럼(패널 세파로스 HP HiTrap, 제너럴 일렉트릭) 상에 적재하였다. 칼럼은 0.6M 황산암모늄, 6mM 인산나트륨 완충액, pH 7.5를 사용하여 세척하고, CRM197은 저염 조건(10mM의 염화나트륨, 10mM의 인산나트륨 완충액, pH 7.5) 하에서 용출하였다. 그 후, 다섯 개의 2.5mL 용출된 분획은 반-DPX 웨스턴 블로트 및 단백질 염색(도 13, 패널 A, 10mM NaCl로 레이블된 레인)으로 분석하였다. 소량의 미처리 CRM197(도 13, 패널 A, 화살표)은 중류수로 최종 세척한 주요 용출된 샘플로부터 떨어져 정제하였다. 그 후, 도 13 패널 A에 동그라미 된 분획은 풀링하고, 중류수로 1:2 희석하여 10mM 염화나트륨, 10mM 인산나트륨, pH 7.5에서 평형화한 DEAE 세파로스 패스트 플로우(파마시아) 1mL를 함유하는 칼럼 상 적재하였다. 샘플을 적재하고 흐름을 통해 수집한 후, 칼럼은 3 용량의 50mM 염화나트륨, 0.5mM의 인산나트륨 완충액(pH 7.5)으로 세척하였다. CRM197은 염화나트륨 농도 증가: 100mM의 NaCl(2회 3mL), 1M NaCl(3mL) 및 1.5M의 NaCl(3mL)을 사용하여 용출하였다. SDS-PAGE 분석은, 비록 여전히 상당량이 칼럼에 결합되어 남아 있지만, 가장 고순도 가용성 CRM197은 1M의 NaCl을 사용하여 용출한다고 밝혀졌다.

[0120]

이러한 결과는 감소된 계놈 대장균 숙주 균주에서 생산된 CRM197은 매우 가용성이며, 기존의 정제 방법을 사용하여 고순도로 단리될 수 있음을 나타낸다.

[0121]

#### 실시예 5

[0122] 야생형 균주에 비해 감소된 게놈 대장균 숙주에서 CRM197 생산

[0123] 유사한 조건하에서 감소된 게놈 대장균에서 CRM197의 주변 세포질 생산과 야생형 대장균 균주에서 CRM197의 생산을 비교하였다. 따라서, CRM197 대장균 BLR(DE3) 균주는 *OmpA-CRM197* 융합을 지닌 pSX2 벡터로 형질 전환하였고 주변 세포질 생산을 평가하였고 감소된 게놈 대장균 균주 MDS42*recA*에서 CRM197의 주변 세포질 생산과 비교하였다. 발효 조건은 상술한 바와 같다. 37°C에서 짧은 성장 개시 단계에 뒤따라, 세포는 19시간 동안 25°C에서 0.2% 포도당(및 BLR(DE3) 배양을 위해 31μg/ml 이소류신)으로 상보된 Korz 배양 배지에서 성장시켰다. CRM197의 발현은 15 또는 25mM IPTG에 의해 OD=0.3로 유도하였다.

[0124] 도 14에 예시된 바와 같이, 주변 세포질 CRM197의 생산에서 적어도 ~ 5배 증가가 야생형 B 균주에 비해 감소된 게놈 대장균 숙주에서 관찰되었다.

[0125] 추가 실험은 *OmpF-CRM197* 융합은 *OmpA-CRM197* 융합에 비해 감소된 게놈 대장균 숙주에서 실제로 더 많은 양의 가용성 주변 세포질 CRM197을 초래하는 것을 나타낸다. 감소된 게놈 대장균 숙주 균주 MDS69*metab* 및 MDS69*lowmut*(*polB*(b0060), *dinB*(b0231) 및 *umuD*(b1183-b1184)의 제거를 추가로 포함하는 MDS69 균주)는 *OmpF-CRM197* 융합을 인코딩하는 발현 벡터로 형질 전환하였고 CRM197의 주변 세포질 발현은 동일한 조건하에서 *OmpA-CRM197* 융합을 인코딩하는 발현 벡터를 지닌 MDS69*lowmut* 숙주에서의 것과 비교하였다. 37°C에서 짧은 성장 개시 단계에 뒤따라, 세포는 23시간 동안 25°C에서 0.2% 포도당으로 상보된 Korz 배양 배지에서 성장시켰다. CRM197의 발현은 25 또는 35mM IPTG에 의해 OD=0.3 내지 0.34로 유도하였다. 주변 세포질 단백질을 단리하였고 각각의 균주에서 가용성 CRM197의 발현은 분석하였다. 도 15에 예시된 바와 같이, CRM197의 더 높은 수율은 *OmpA-CRM197* 구조체에 비해 *OmpF-CRM197* 구성체에 의해 수득했다.

[0126] 실시예 6

[0127] 감소된 게놈 대장균 균주에서 다양한 신호 서열을 갖는 CRM197 생산을 시험

[0128] 신호 서열은 주변 세포질 분획(Han, Mee-Jung 외., Journal of Bioscience and Bioengineering, 117(4):437-442(2014))의 2D 겔 분석에 의해 측정된 바와 같이 대장균 B와 K 균주의 주변 세포질에서 존재비를 기반으로 선택하였다. 표 5는 선택된 신호 서열 및 B와 K 균주의 주변 세포질에서 상대적 존재비를 열거하였다.:

[0129] 표 5

단백질	주변 세포질 분획(B 및/또는 K)에서 존재비	유전자/단백질 기능
MglB	K	메틸 갈락토스 트랜스포터
MalE	B + K	말토스 트랜스포터
OppA	B + K	올리고펩티드 트랜스포터
RbsB	B + K	서브유닛 리보오스 트랜스포터
Agp	B > K	글루코스-1 포스파타아제, 3-피타아제
FkpA	B > K	펩티딜-프롤릴 시스-트랜스 이소머라아제; 단백질 폴딩에서
YtfQ	B ≫ K	갈락토퓨라노오스 결합 단백질, 서브유닛 ABC 트랜스포터
HdeA	K	산성조건에 의해 유도된 스트레스 반응
HdeB	K	산성조건에 의해 유도된 스트레스 반응
GlnH	B > K	글루타민 ABC 트랜스포터의 서브유닛

[0130] [0131] 신호 서열과 표 5에 예시된 CRM197 서열(그의 네이티브 신호 순서를 결여한) 및 도 17(MglB, MalE, OppA, RbsB, Agp, FkpA, YtfQ, HdeA, HdeB 또는 GlnH; *OmpF* 및 *OmpC*는 당연히 시험했다)의 조합을 함유하는 플라스미드 pSX2는 감소된 게놈 대장균 균주 MDS69*metab*(도 18 ~ 21에서 T69 Meta)로 형질 전환시키고, 진탕 플라스크 배양에서 검사하였다. 전술한 바와 같이, MDS69*metab*는 MDS69 배경(background)에서 다음의 변형을 포함한다. (i) *icIR*(NCBI Entrez GeneID 번호 948524에 기술된, b-번호 b4018) 및 *arpA* 유전자(NCBI Entrez GeneID 번호 944933에 기술된, b-번호 b4017)의 제거 (ii) *rph* 유전자(b3643)의 제거, 및 (iii) 위치 982에서 AT 디뉴클레오티드의 삽입에 의해 *iIVG*의 프레임시프트 변이의 보정. 요약하면, MOPS 최소 배지-카나마이신(MMM/Kan)-포도당 면선 접종 플레이트에서 형질 전환된 박테리아의 콜로니 형성 유닛은 0.2% 포도당 및 50 μg/ml 카나마이신으로 상보된 3ml Korz 최소 배지에서 재 혼탁하여 스타터 배양을 생성하기 위해 37°C에서 밤새 배양하였다. 스타터 배양은 OD600=0.05로 125mL 삼각-플라스크에서 20ml Korz/0.2% 포도당/Kan을 접종하기 위해 사용하

였고 1.5시간 동안 37°C에서 성장시키고 그 후 25°C로 변경하고 OD<sub>600</sub> ~ 0.3까지 성장시켰다. 그 시점에서, 인듀서(IPTG)는 25 μM, 35 μM, 50 μM 농도(후기 유도)로 첨가하였다. 그 후, 후기 유도는 20시간 동안 25°C에서 성장시키고, 배양의 2 ODs를 수확하였다. 전체 세포 단백질은 BugBuster+ 리소나아제를 사용하여 제조하였고 주변 세포질 및 스페로플라스트 분획은 중심지 원형질막 방법(Epicentre Periplasting Method)을 사용하여 제조하였다.

[0132] 도\_18-20은 표시된 신호 서열(유도 A - OmpF, MalE, HdeA, OppA, HdeB, GlnH; 유도 B - OmpF, MglB, Agp, OmpC, RbsB, FkpA, YtfQ)을 사용하여 각각 25 μM, 35 μM 또는 50 μM 인듀서 농도에서 (가용성)CRM197의 주변 세포질 수율을 도시했다. 양호한 수율은 25 μM 인듀서 농도에서 모든 신호 서열로서 수득하였다(도 18). 흥미롭게도, 35 μM 인듀서 농도에서, YtfQ 신호 서열을 갖는 CRM197의 수율이 다른 시험된 신호 서열로서 인듀서 농도에서 수득된 CRM197의 수율에 비해 현저하게 증가했다(도 19). 이러한 효과는 50 μM 인듀서 농도에서 더 크게 되었고, 다른 시험된 신호 서열을 갖는 CRM197의 수율은 이 인듀서 농도에서 현저하게 감소된 반면, 나머지 CRM197의 수율은 높았다(도 20). 따라서, CRM197와 YtfQ 신호 서열의 조합은 시험된 다른 신호 서열을 갖는 CRM197의 조합 보다 현저하게 넓은 유도를 갖는다는 것을 측정하였다.

[0133] 상기 YtfQ 신호 서열을 갖는 CRM197에 대한 유도 범위를 추가로 평가하기 위해, 각각 8 IPTG(인듀서) 수준의 두 개의 배양은 상술한 방법에 따라 MDS69metab(0, 25, 35, 50, 75, 100, 150 및 250 μM)에서 시험하였다. 대조군으로서, 또한 CRM197 및 OmpF 신호 서열을 갖는 MDS69metab에 대한 4 IPTG 수준의 2개의 배양은 각각 (0, 25, 35, 50 μM) 시험하였다. 2 OD 샘플은 캘리퍼스로 전체 세포 단백질(TCP) 및 주변 세포질 분석을 위해 수집하였다.

[0134] 각각의 인듀서 수준에 대해 시험된 두 개의 배양의 평균화 결과를 도 21에 예시하였다. YtfQ 신호 서열(YtfQ-CRM197)과 조합한 CRM197의 수율은 모든 인듀서 수준이 100 μM에 걸쳐 여전히 높은 채로였다. 그러나 OmpF과 조합한 CRM197의 수율은 단지 25 및 35 μM 인듀서 농도에서 높았다. 도 22는 50 μM IPTG(OmpF) 및 50, 75, 100, 150 및 250 μM의 IPTG(YtfQ)에서 주변 세포질(P) 및 배양 배지(M)의 CRM197 수율 상에 OmpF 및 YtfQ 신호 서열의 효과를 비교한 단백질 겸이다. 주변 세포질 CRM197의 훨씬 적은 양은 OmpF 서열에 대해 50 μM 인듀서 농도에서 존재하는 반면 주변 세포질 CRM197의 상당한 다량은 YtfQ 신호 서열에 대해 50, 75 및 100 μM 인듀서 농도에서 명백했다.

[0135] “요약하면, MOPS 최소 배양 배지-카나마이신(MMM/Kan)-포도당 면선 접종 플레이트에서 형질 전환된 박테리아의 콜로니 형성 유닛은 0.2% 포도당 및 50 μg/ml 카나마이신으로 상보된 3ml Korz 최소 배양 배지에서 재 혼탁하여 스타터 배양을 생성하기 위해 37°C에서 밤새 배양하였다. 스타터 배양은 OD<sub>600</sub>=0.05로 125mL 삼각-플라스크에서 20ml Korz/0.2% 포도당/Kan을 접종하기 위해 사용하였고 1.5시간 동안 37°C에서 성장시키고 그 후 25°C로 변경하고 OD<sub>600</sub> ~ 0.3까지 성장시켰다. 그 시점에서, 인듀서(IPTG)는 25 μM, 35 μM, 50 μM 농도(후기 유도)로 첨가하였다. 그 후, 후기 유도는 20시간 동안 25°C에서 성장시키고, 배양의 2 ODs를 수확하였다. 전체 세포 단백질은 BugBuster+ 리소나아제를 사용하여 제조하였고 주변 세포질 및 스페로플라스트 분획은 중심지 원형질막 방법(Epicentre Periplasting Method)을 사용하여 제조하였다.”

[0136] 다음으로, MDS69metab에서 OmpF 또는 YtfQ 신호 서열 중 하나와 조합 및 대장균 B 균주(BL21DE3)에서 OmpF와 조합한 CRM197 수율에서 최후기 유도(OD<sub>600</sub> ~ 2)의 효과를 평가하였다. 요약하면, 0.2% 포도당 및 50 μg/ml 카나마이신으로 상보된 3ml Korz 최소 배양 배지는 형질 전환된 MDS69metab 또는 BL21DE3의 콜로니 형성 유닛으로 접종하였고, 37°C에서 밤새 배양했고, 25°C 스타터 배양을 생성하기 위해 25°C에서 밤새 성장한 125ml 삼각 플라스크에서 0.2% 포도당 및 50 μg/ml 카나마이신으로 상보된 15ml Korz 최소 배양 배지를 접종하는데 사용하였다. 스타터 배양은 500ml 삼각 플라스크에서 90ml Korz/0.2% 포도당/카나마이신을 OD<sub>600</sub>=0.1로 접종하는데 사용하였고, 25°C에서 OD<sub>600</sub>>2(포화된 또는 거의 포화된)까지 성장이 뒤따르고 나서, 다양한 IPTG 농도(최후기 유도)에서 유도를 위해 125ml의 삼각 플라스크에서 4×20 분취량으로 분할하였다. 유도는 20시간 동안 25°C에서 성장되고, 배양의 2 ODs를 분석을 위해 수확하였다. 전체 세포 단백질(TCP)을 BugBuster+ 리소나아제를 사용하여 제조하였다. 주변 세포질과 스페로플라스트 분획은 중심지 원형질막 방법을 사용하여 제조하였다. 도 23에 나타낸 바와 같이, 100 μM IPTG까지 좋은 세포질 발현은 아마도 불용성으로 인해 200mM 인듀서 농도에서 감소된 MDS69metab에서 CRM197과 OmpF 신호 서열의 조합에 대해 관찰하였다. 좋은 주변 세포질 발현은 어떠한 관찰된 불용성 CRM197을 가지지 않는 400 μM IPTG(최고의 농도에서 시험된)까지 MDS69metab에서 CRM197 및 YtfQ 신호 서열의 조합에 대해 관찰하였다. 약한 발현은 시험된 모든 인듀서 농도(25-200 μM)에서 CRM197과 BL21(DE3) 균

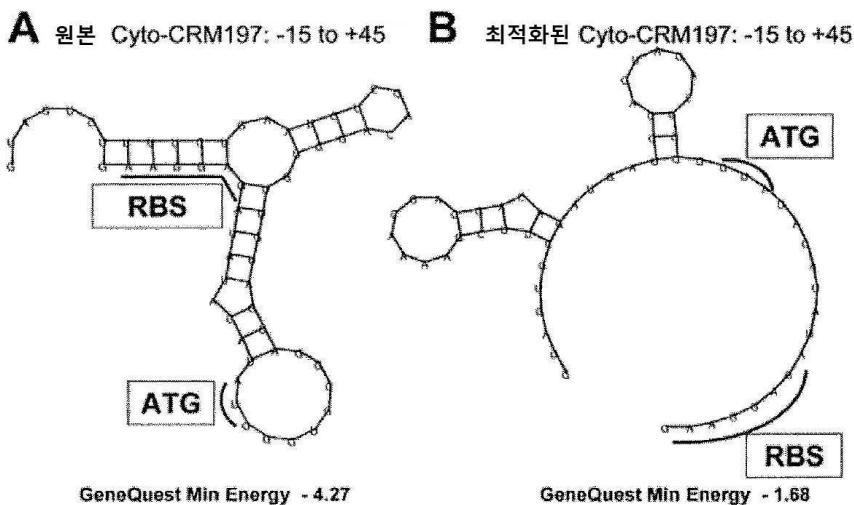
주에서 OmpF의 조합에 대해 관찰하였다. CRM197은 스페로플라스트에서 관찰되지 않았다.

[0137]

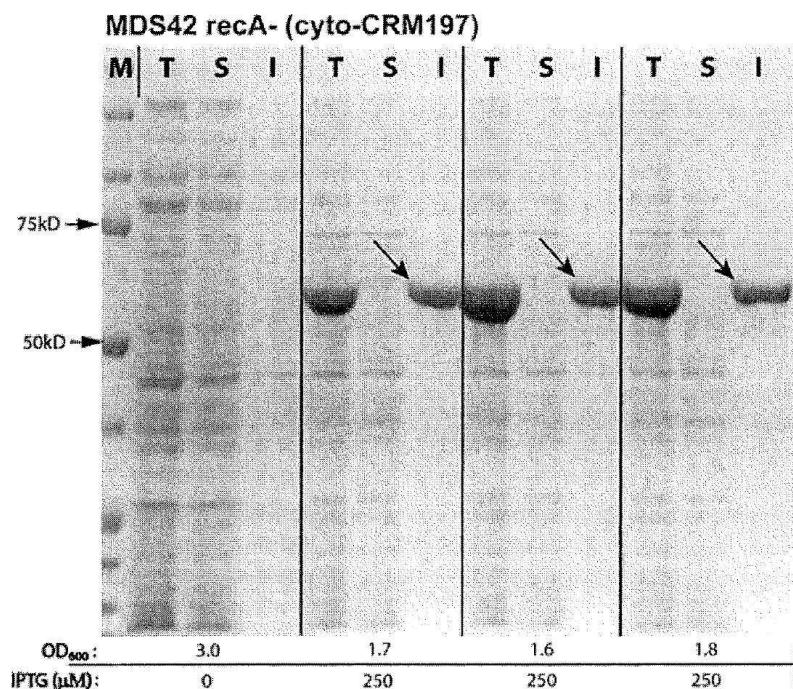
요약 - 제시된 데이터는 감소된 게놈 대장균 숙주에서 가용성 CRM197의 생산은 통상적인 방법으로 수득한 것에 10배인 수율을 가져온다는 것을 입증한다. 감소된 게놈 대장균 숙주에서 생산은 효율성을 높이고 제조비용을 감소시킬 것으로 예상된다. 또한, 감소된 게놈 대장균 숙주에서 생산은 비-감소된 게놈을 갖는 종래의 박테리아에서 생산된 것보다 더 깨끗하고 더 안전할 것이다. 이러한 개선은 약학적 단백질 산물의 생산에 광범위한 영향을 미칠 것이며 궁극적으로 그들을 필요로 하는 위험 집단을 위해 백신에 대한 접촉 기회가 넓어질 것이다. 또한, CRM197의 높은 수율은 신호 서열의 넓은 범위의 조합으로 관찰하였다. YtfQ는 K 균주 대장균 비해 B 균주 대장균에서 훨씬 더 많은 양으로 확인되었고 예시된 감소된 게놈 균주는 K 균주에 기반을 두기 때문에, CRM197과 조합한 YtfQ 신호 서열에 대해 관찰된 넓은 유도 범위는 놀라웠다. 인도서의 농도는 단백질의 생산 중 변할 수 있고, 따라서 이러한 숙주에서 CRM197과 조합한 YtfQ 신호 서열의 사용은 CRM197의 수율에서 더 증가를 초래하기 때문에, 감소된 게놈 대장균에서 YtfQ와 조합한 CRM197의 넓은 유도 범위는 상당한 장점이 있다.

## 도면

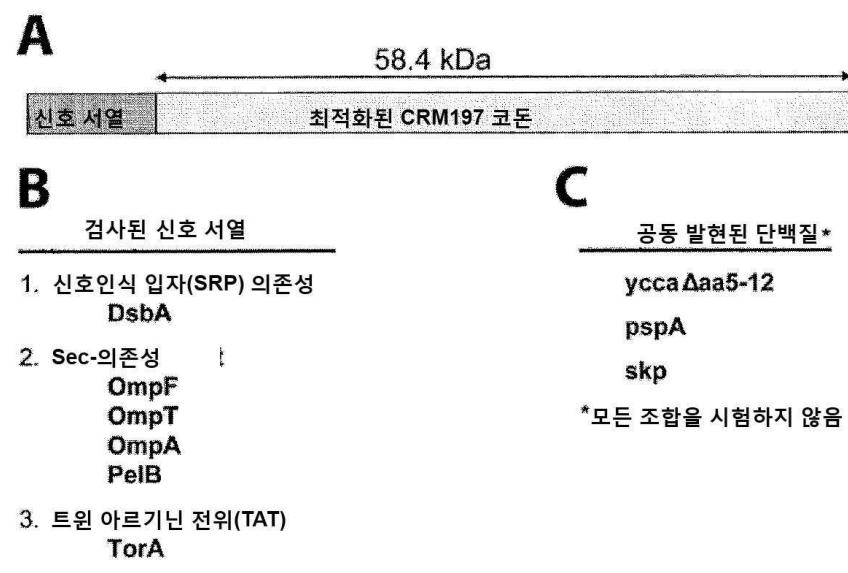
### 도면1



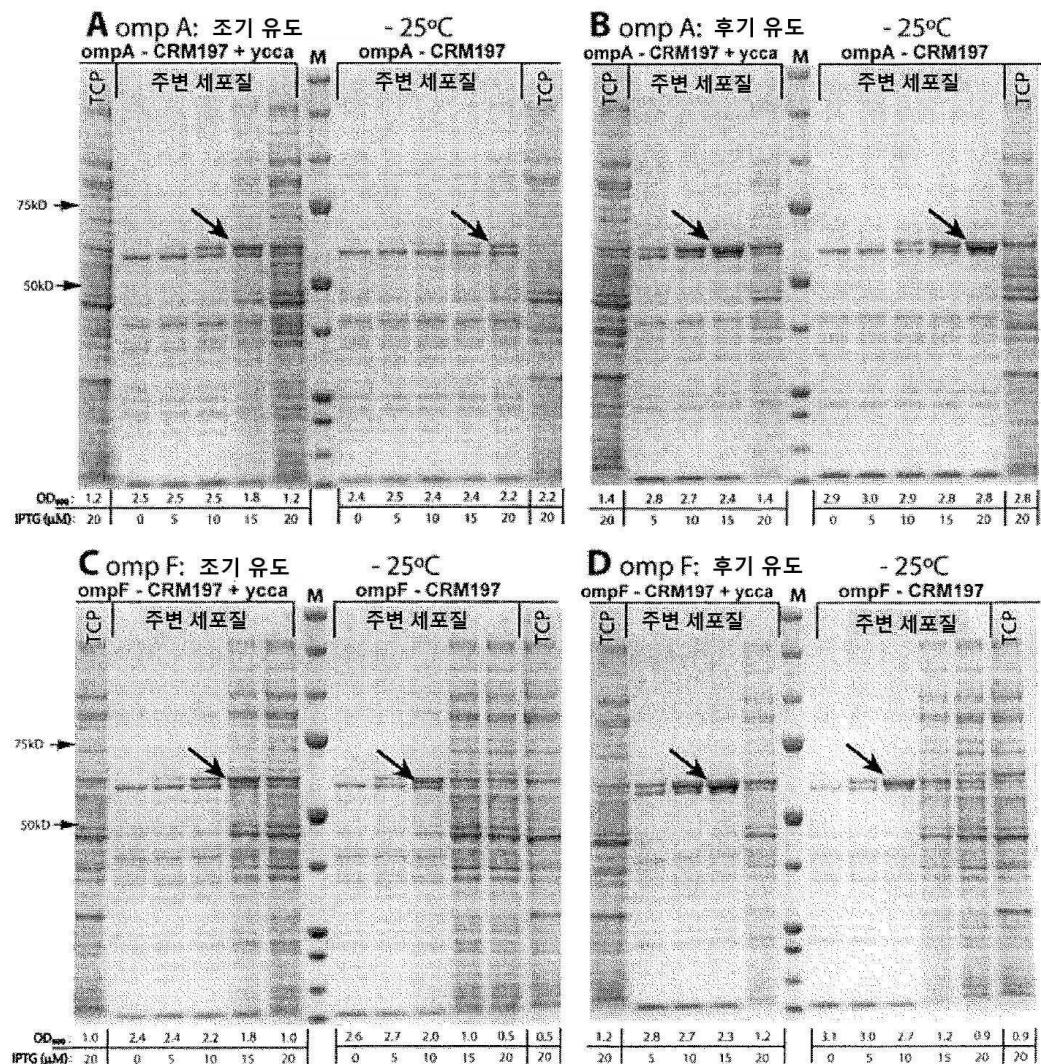
## 도면2



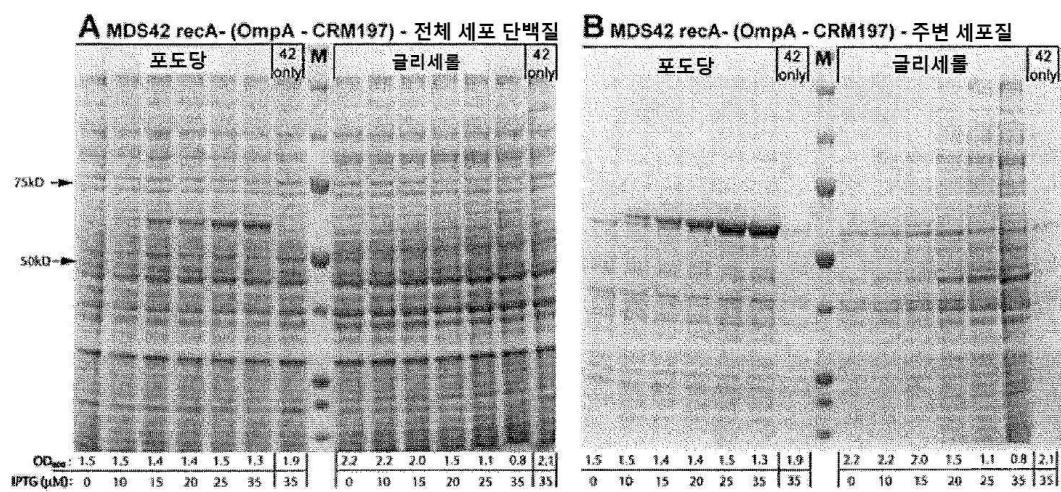
## 도면3



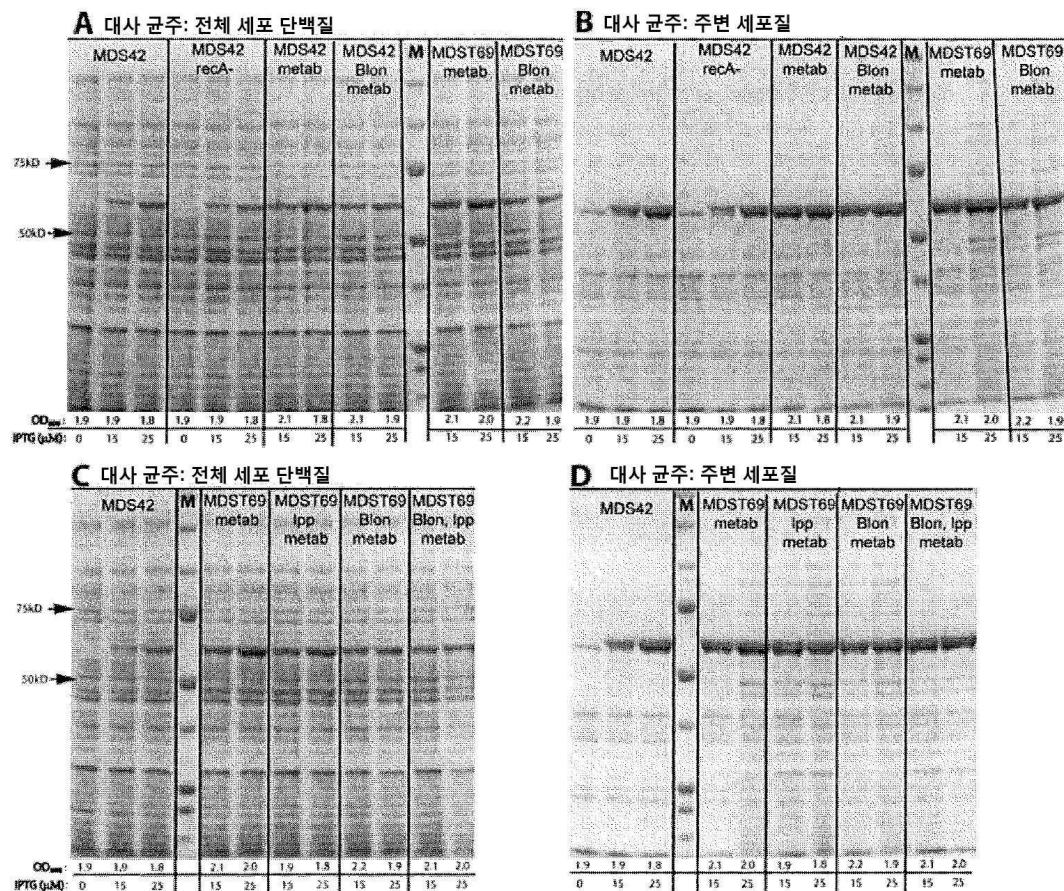
## 도면4



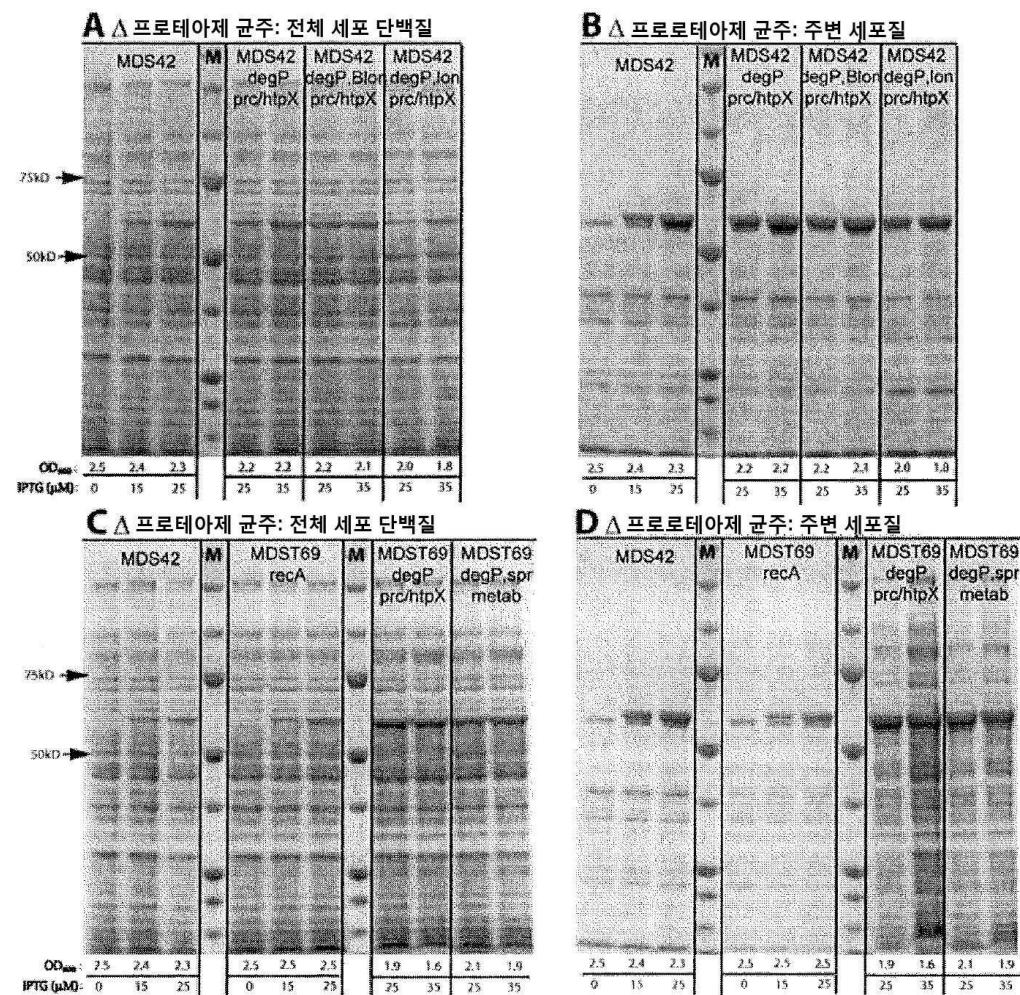
## 도면5



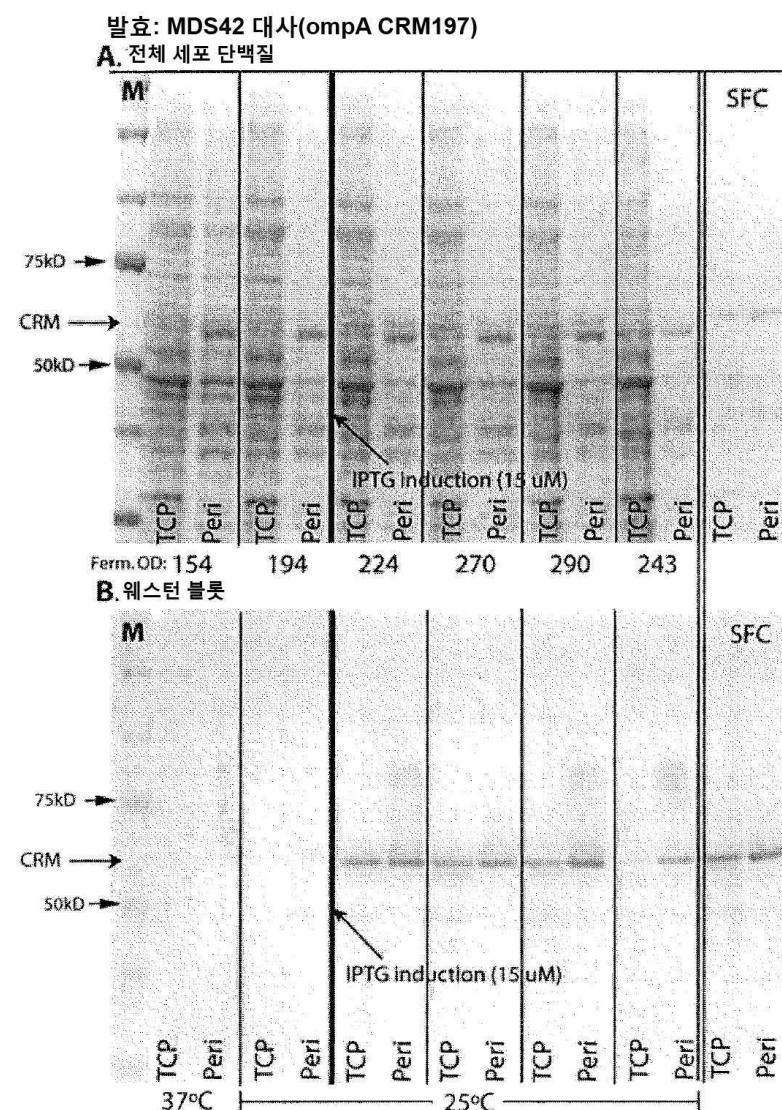
## 도면6



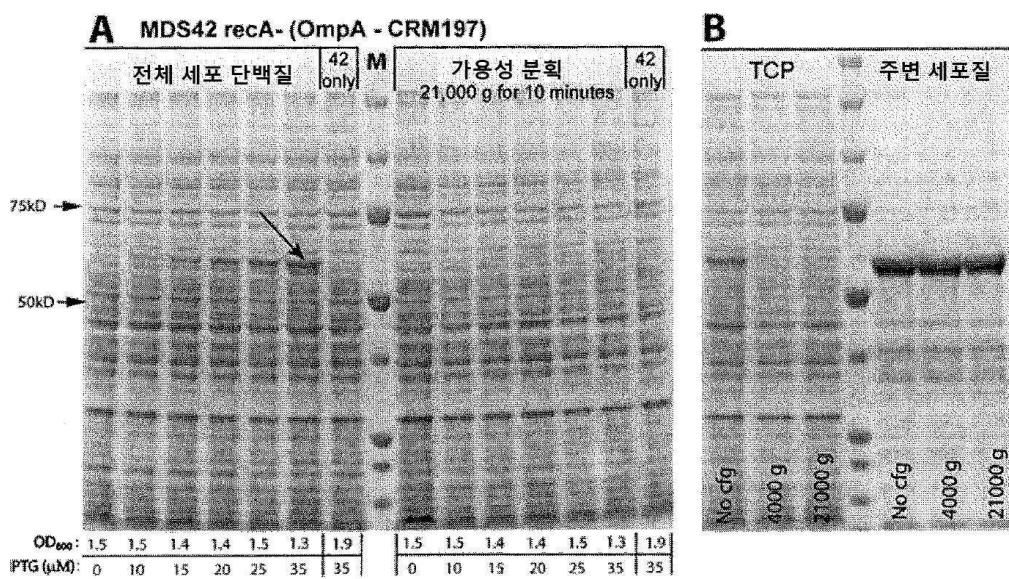
## 도면7



## 도면8



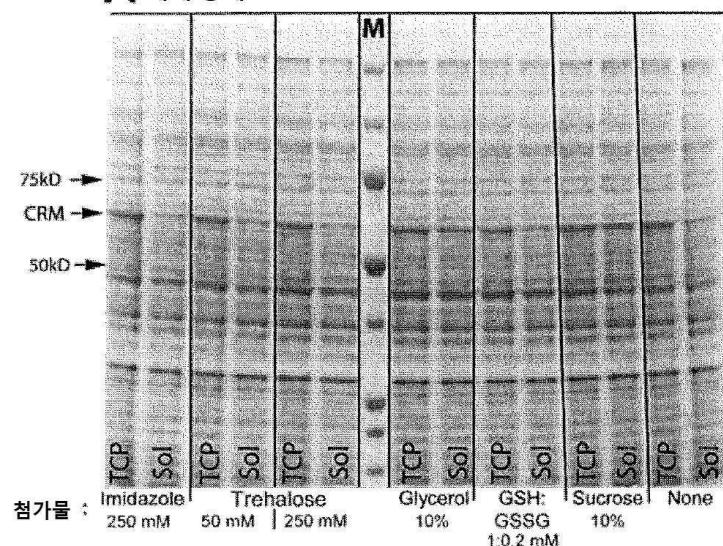
## 도면9



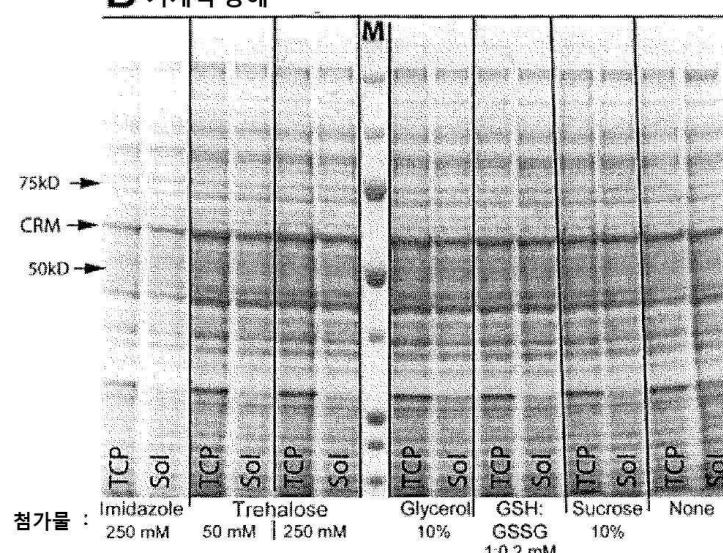
## 도면10

## MDS42 recA- (ompA - CRM197)

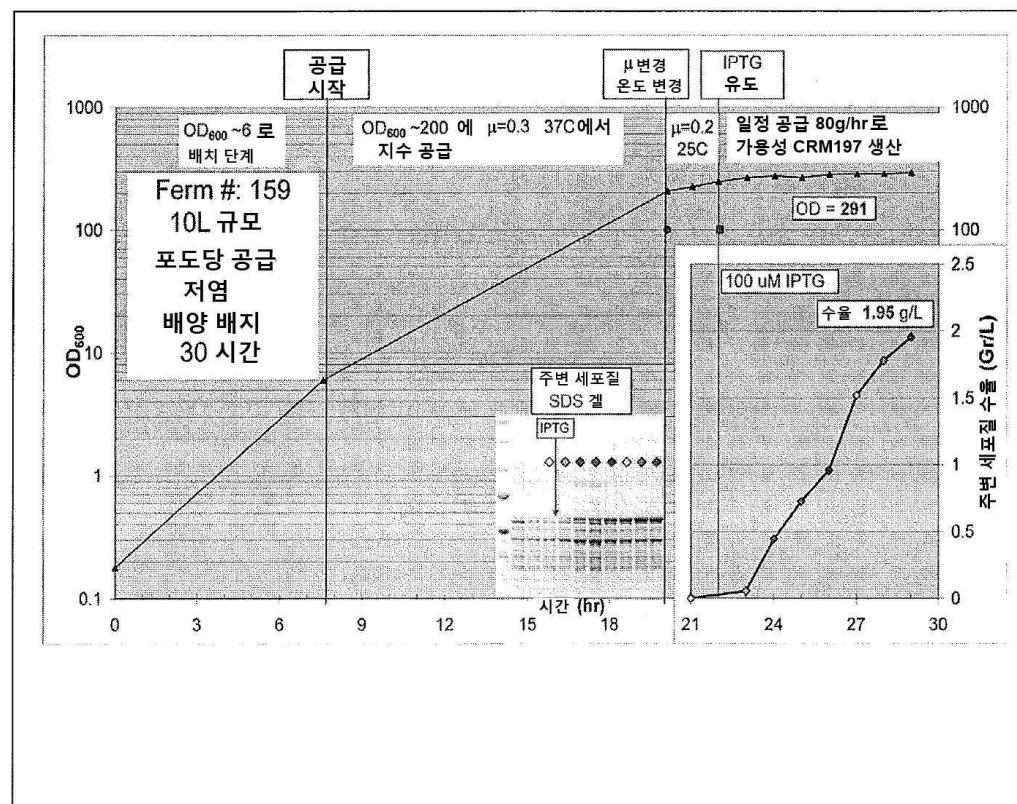
## A 세제 용해



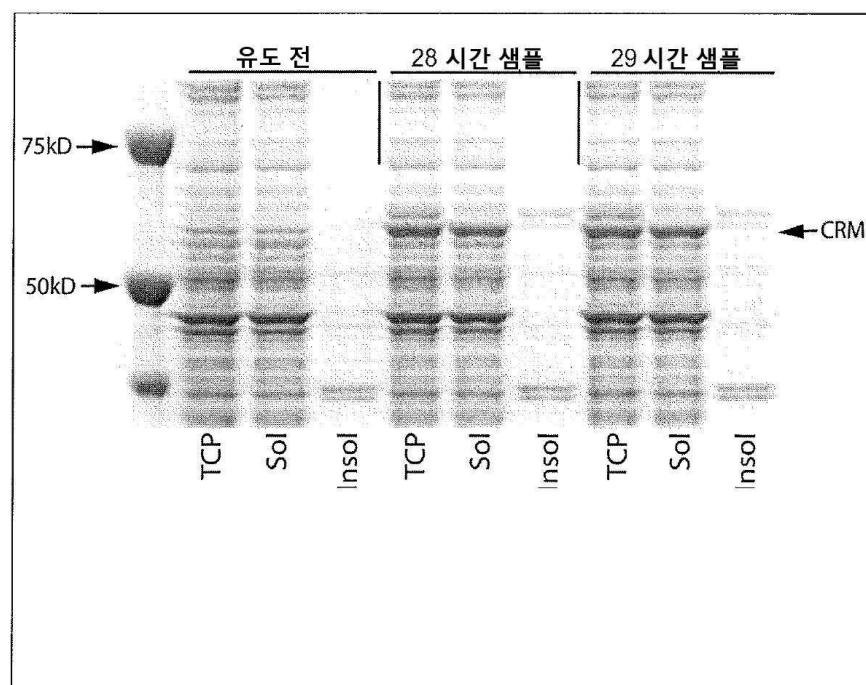
## B 기계적 용해



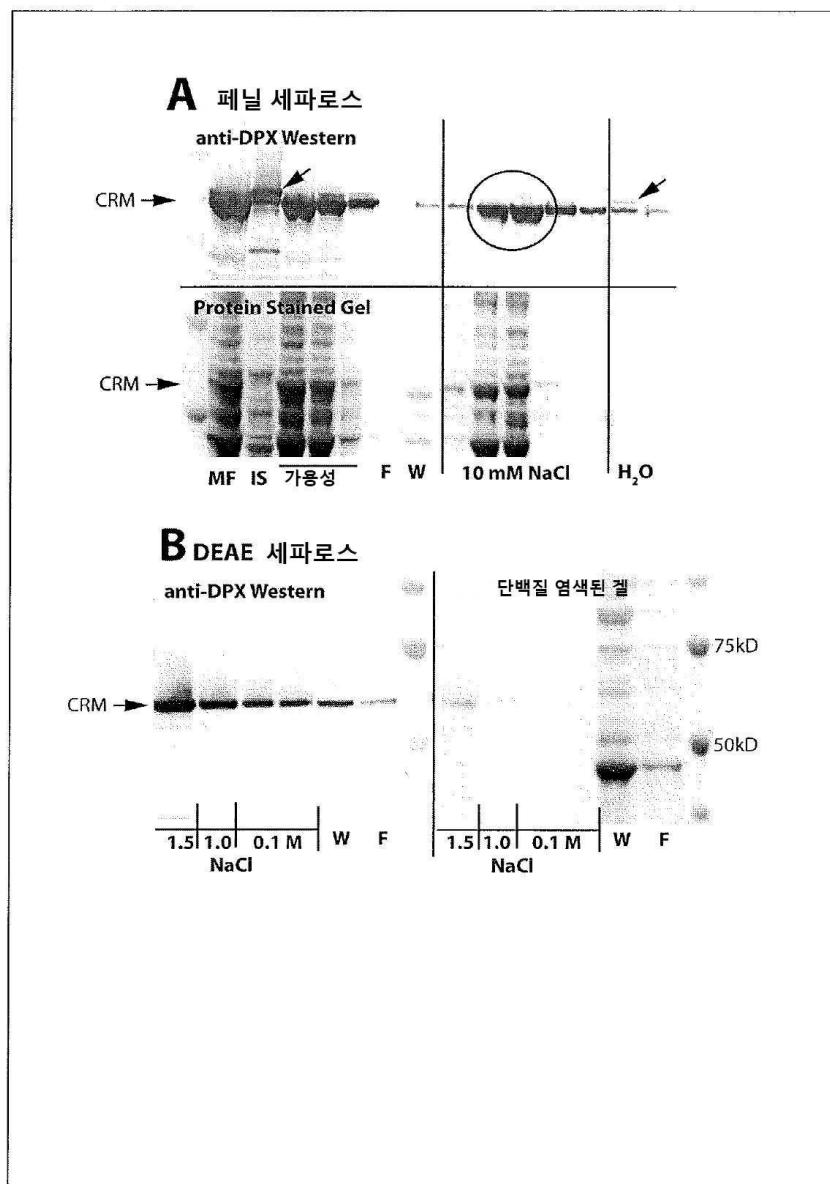
도면11



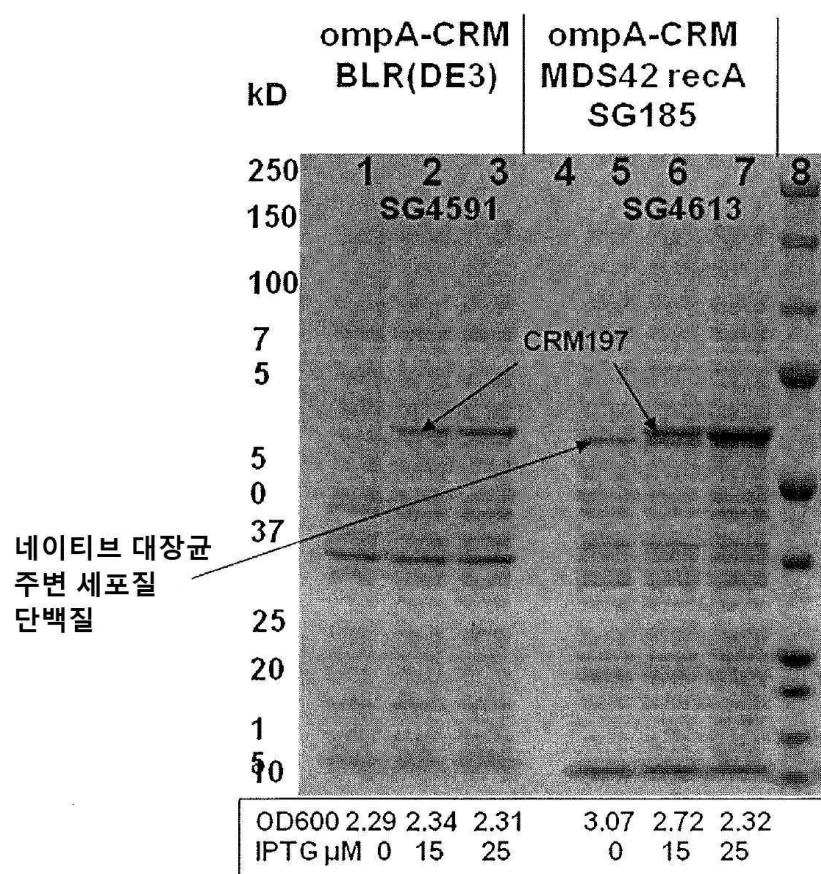
도면12



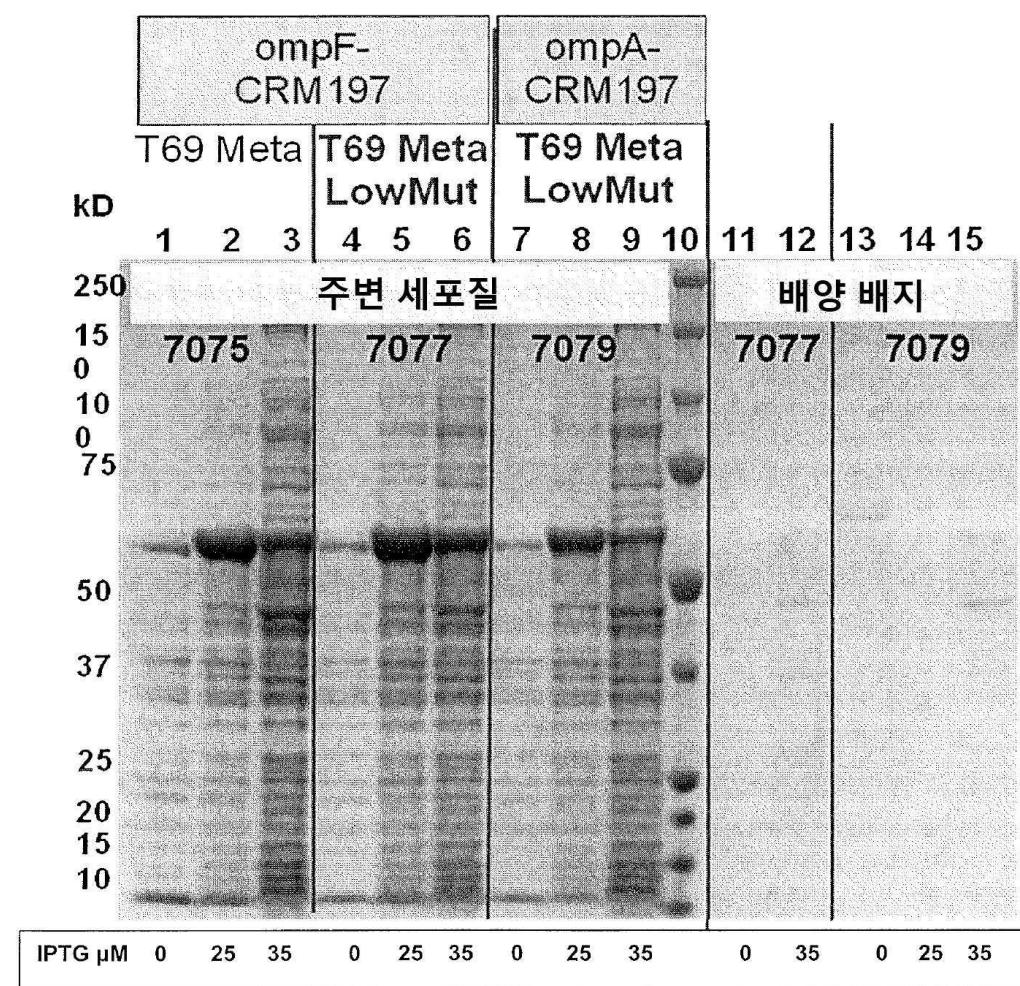
## 도면13



도면14



도면15



## 도면 16a

## A. 발현된 CRM197 합성 핵산 서열 (SEQ ID NO: 1)

GGTGGGATGATGTTGGACAGCTCTAAGTCTTGTGATGGAAAATTT  
 AGCTCGTACCAACGGTACGAAGCCAGGTTATGTCGACAGCATTCAAAAGG  
 TATCCAGAAACCGAAGTCCGGCACGCAGGGTAACACTACGACGACGATTGGA  
 AAGAGTTCTACAGCACCGACAACAAGTATGACCGCAGCGGGTTACAGCGTT  
 GACAATGAGAATCCGTTGAGCGGAAAGCGGGTGGTGTCAAAGTGAC  
 GTATCCGGGTCTGACCAAGGTCCTGGCGTTGAAAGITGATAACGCGGAAA  
 CCATTAAGAAAGAGCTGGGTCTGAGCTGACCGAGCCGTTGATGGAGCAA  
 GTTGGTACCGAAGAGTTATCAAACGTTCGGCATGGTGCAGGCCGCGT  
 TGCTGTCCTGCCTTCGCGGAGGGCAGCTCCAGCGTTGAGTATATCAA  
 TAACTGGGAGCAAGCAAAGCGCTGTCGAACTGAAATCAATTG  
 AACACGCGCGGTAACGTGGTCAAGATGCAATGTACGAGTATATGCCAG  
 GCCTGCGCTGGTAATCGTGTGCGAGCGTTGAGCAGCTTGTCTTGT  
 ATCAACCTGGATTGGGATGTGATCCGTGATAAGACCAAGACTAAGATCGA  
 GAGCCTGAAAGAACATGGCCGATTAAGAACAGATGTCGGAGAGCCCG  
 AATAAGACCGTGAGCGAAGAAAAGCCAAGCAGTATCTGGAAGAGTTCC  
 ACCAAACGGCTCTGGAGCATCCGGAGCTGAGCGAGCTGAAAACGGTTACG  
 GGCACCAACCCGGTGGTGCAGGTGCGAATTACGCGGCGTGGCAGTGAA  
 TGTGGCGCAGGTACTCGACTCCGAAACGGCGGACAATTGGAGAAAACCA  
 CCGCAGCGCTGAGCATTCTGCCGGCATCGCAGCGTTATGGGATTGCA  
 GATGGTGCTGTGCACCATAACACTGAAGAAATCGTAGCGCAAAGCATTGC  
 CCTGTCTAGCTTGATGGTGGCGCAGGCTATTCCGCTGGTGGCGAACTGGT  
 TGATATCGGCTTGCTGCCTACAACITCGITGAAAGCATTAACCTGTT  
 TCAGGTGGTCCACAACAGCTATAATGCCCAAGCGTACAGCCCCGGTCACA  
 AGACCCAACCGTCTGCACGATGGCTATCGGGTGTCTGGAACACGGTC  
 GAAGATAGCATTCGTACCGGTTCCAGGGCGAGAGCGGCCATGACAT  
 CAAGATTACTGCAGAAAATACCCCGCTGCCGATCGCAGGTGTCTGCTGC  
 CTACGATTCCGGGTAAGCTGGACGTTAACAAAGCAAACCCACATTCT  
 GTGAACGGTCGTAAGATTGCGATCGCTGCGATTGACGGCGACGT  
 CACCTTCTGCCGTCCGAAGAGGCCGGTACGTTGGTAATGGTGTGCACGC  
 GAACCTGCACTGGCGTTCCACCGCAGCAGCTGGAGAAAATCCATAGCA  
 ATGAGATTCTAGCGACAGCATTGGCGTTCTGGGTTACCAAAAGACGGTT  
 GACCATAACCAAGTCACATTCCAAACTGAGCCTGTTGAGATCAAAAG  
 CTAA

## 도면 16b

## B. 발현된 CRM197 합성 핵산 서열 (SEQ ID NO: 3)

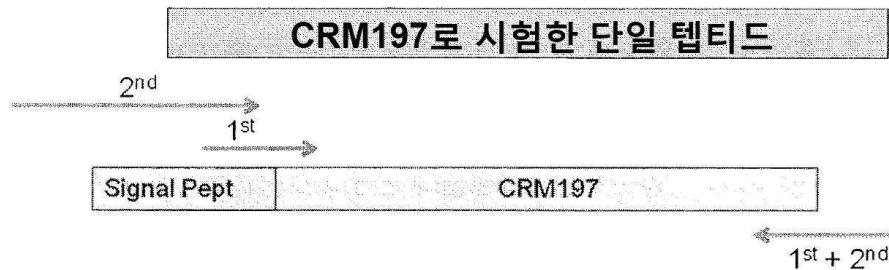
atggcgcagatgacgtagttagacacggcagcaaaagcttcgtatggaa  
 AACTTAGCTCGTACACCGTACGAAGCCAGGTTATGTCGACAGCATT  
 CAAAAAGGTATCCAGAAACCGAAGTCCGGCACGCAGGGTAACACGA  
 CGACGATTGGAAAGAGTTCTACAGCACCGACAACAAGTATGACGCAG  
 CGGGTTACAGCGTTGACAATGAGAATCCGTTGAGCGGCAAAGCGGGT  
 GGTGTTGTCAAAGTGACGTATCCGGGTCTGACCAAGGTCCTGGCGTT  
 GAAAGTTGATAACCGGGAAACCATTAAGAAAGAGCTGGGTCTGAGCC  
 TGACCGAGCCGTTGATGGAGCAAGTTGGTACCGAAGAGTTATCAAA  
 CGTTTGGCGATGGTGCAGGCCGTTGTCCTGCCTTCGC  
 GGAGGGCAGCTCCAGCGTTGAGTATATCAATAACTGGGAGCAAGCAA  
 AAGCGCTGCCGTCGAACTGAAATCAATTGAAACCGCGGTAAA  
 CGTGGTCAAGATGCAATGTACGAGTATATGCCAGGCCCTGCCTGG  
 TAATCGTGTTCGTCGCAGCGTTGGTAGCAGCTTGTCTGTATCACCT  
 GGATTGGGATGTGATCCGTGATAAGACCAAGACTAAGATCGAGAGCC  
 TGAAAGAACATGGCCCGATTAAGAACAGATGTCGGAGAGCCCGAAT  
 AAGACCGTGAGCGAAGAAAAGGCCAAGCAGTATCTGGAAGAGTTCCA  
 CCAAACGGCTCTGGAGCATCCGGAGCTGAGCGAGCTGAAAACGGTTA  
 CGGGCACCAACCCGGTGTGCAAGGTGCAATTACGGCGTGGCA  
 GTGAATGTGGCGCAGGTACTCGACTCCGAAACGGCGACAATTGGA  
 GAAAACCACCGCAGCGCTGAGCATTCTGCCGGCATCGGCAGCGTTA  
 TGGGCATTGCAGATGGTGTGCAACCATAACACTGAAGAACATCGTA  
 GCGCAAAGCATTGCCCTGCTAGCTTGATGGTGGCGCAGGCTATTCC  
 GCTGGTCGGCGAAGTGGTGTGATATCGGCTTGCTGCCTACAACCTCGT  
 TGAAAGCATCATTAACCTGTTCAAGGTGGTCCACAACAGCTATAATCG  
 CCCAGCGTACAGCCGGTCACAAGACCCAAACCGTTCTGCACGATG  
 GCTATGCGGTGTCCTGGAACACGGTCAAGATAGCATTCGTA  
 GCTTCCAGGGCGAGAGCGGCCATGACATCAAGATTACTGCAAGAAA  
 TACCCCGCTGCCGATCGCAGGTGTCCTGCTGCCTACGATTCCGGTA  
 AGCTGGACGTTAACAAAAGCAAAACCCACATTCTGTGAACGGTCGTA  
 AGATTGCGATGCGTTGTCGATTGACGGCGACGTCACCTCTGC  
 CGTCCGAAGAGCCGGTCTACGTTGGTAATGGTGTGCACGCGAACCT  
 GCACGTGGCCTTCACCGCAGCAGCTGGAGAAAATCCATAGCAATG  
 AGATTCTAGCGACAGCATTGGCGTCTGGGTTACCAAAAGACGGTT  
 GACCATACCAAAAGTCAATTCAAACGTGAGCCTGTTGAGATCAA  
 AGCtaa

## 도면 16c

## C. 발현되고, 처리된 CRM197 합성 아미노산 서열(SEQ ID NO: 2)

GADDVVDSSKSFMENFSSYHGTPGYVDSIQKGIQKPKSGTQGNYDDDWK  
 EFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETI  
 KKEGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYINNWEQ  
 AKALSVELEINFETRGKRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDW  
 DVIRDKTKTIESLKEHGPINKMSESPNKTVSEEKAKQYLEEFHQTALHPEL  
 SELKTVTGTNPVFAGANAAWAVNVAQVIDSETADNLEKTTAALSLPGIGSV  
 MGIADGAVIHNTTEEIVAQSIALLSMVAQAIPLVGEVDIGFAAYNFVESIINLF  
 QVVFHNSYNRPAYSPGHKTQPFLHDGYAVSWNTVEDSIIRTGFQGESGHDIKIT  
 AENTPLPIAGVLLPTIPGKLDVNKSKTTHISVNNGRKIRMRCRAIDGDVTFCRPKS  
 PVYVGNGVHANLHVAFHRSSEKIHSNEISSDSIGVLGYQKTVVDHTKVNSKLS  
 LFFEIKS

## 도면17



## 주요 주변 세포질 단백질의 신호 텝티드

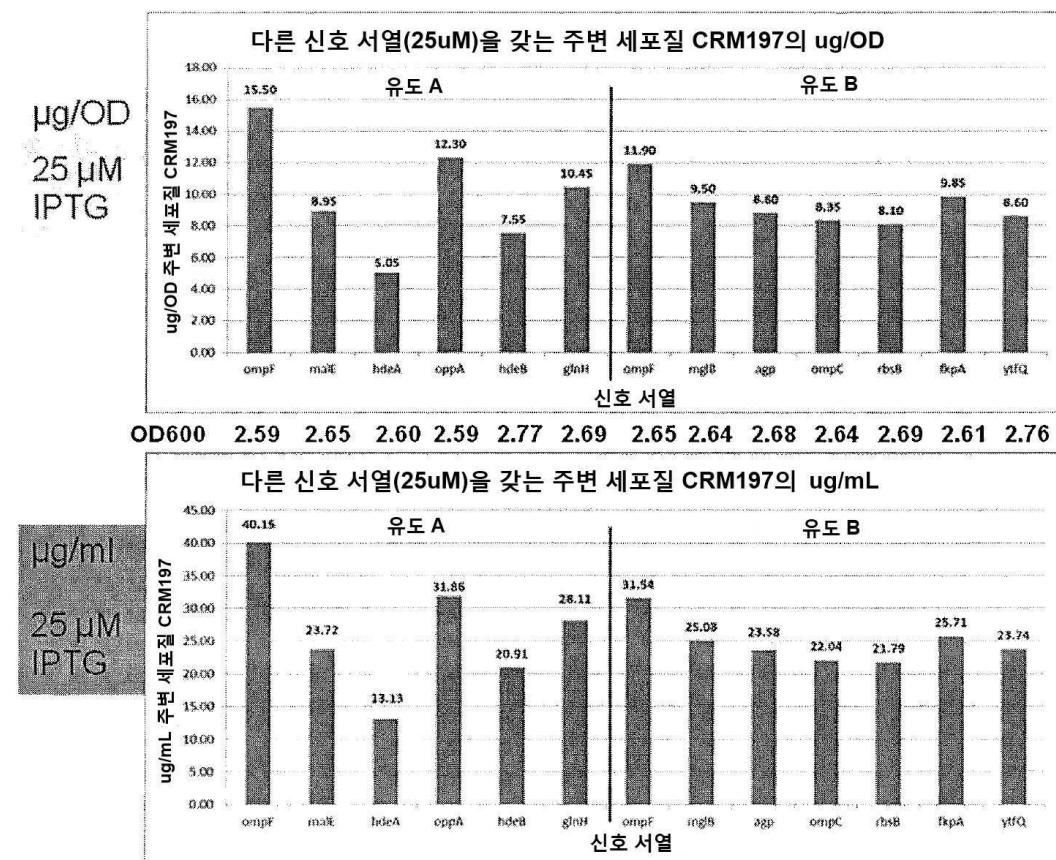
## 가장 풍부한 K-균주

malE  
hdeA  
oppA  
hdeB  
mglB  
agp  
rbsB

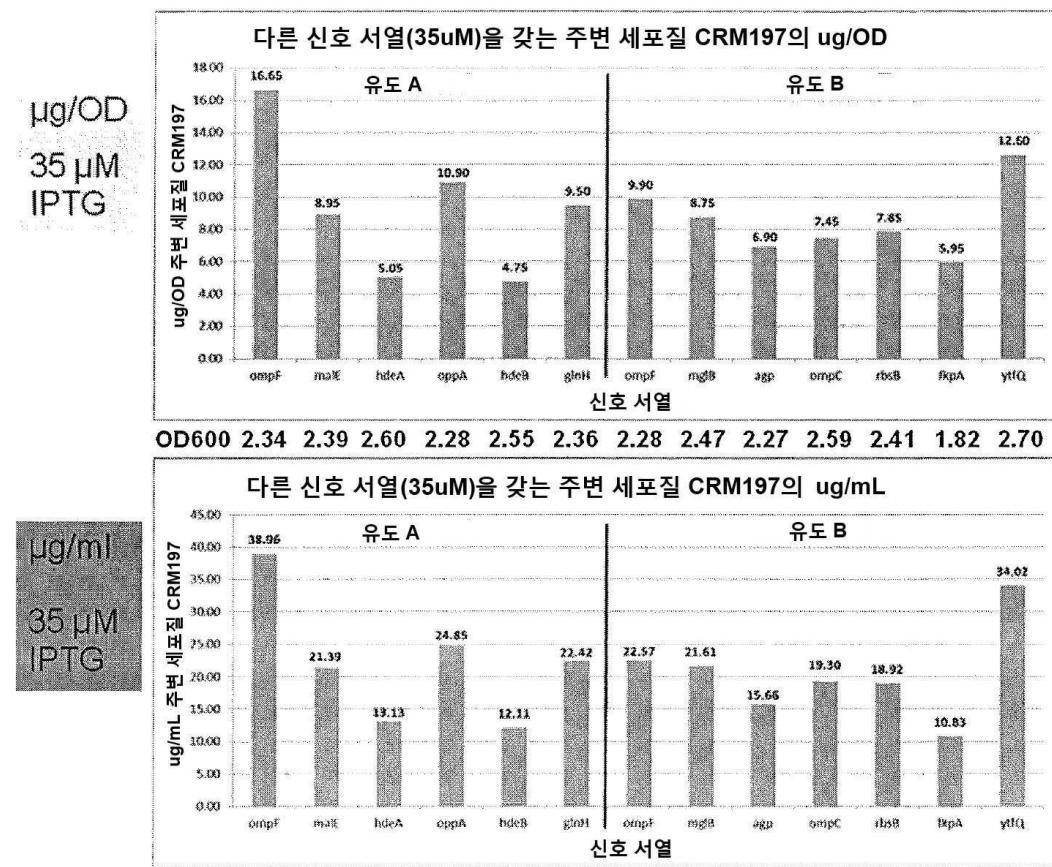
## 추가적인 후보 B-균주

fkpA  
glnH  
ytfQ

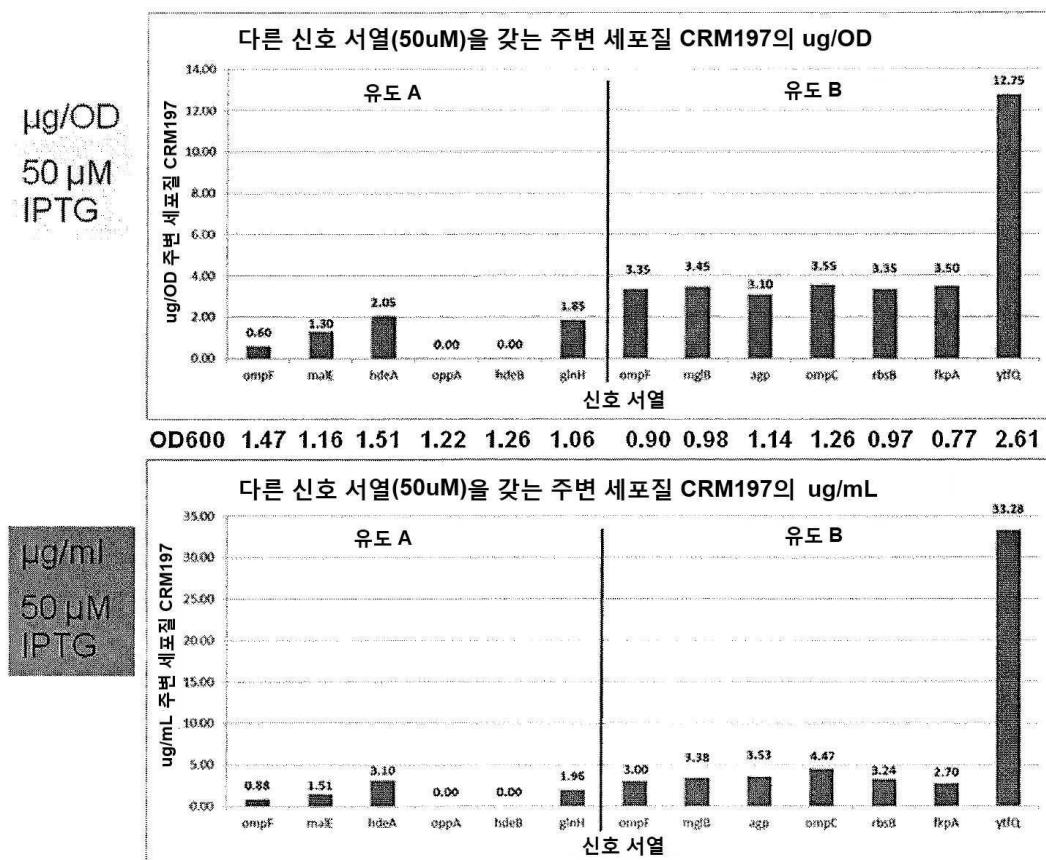
## 도면18



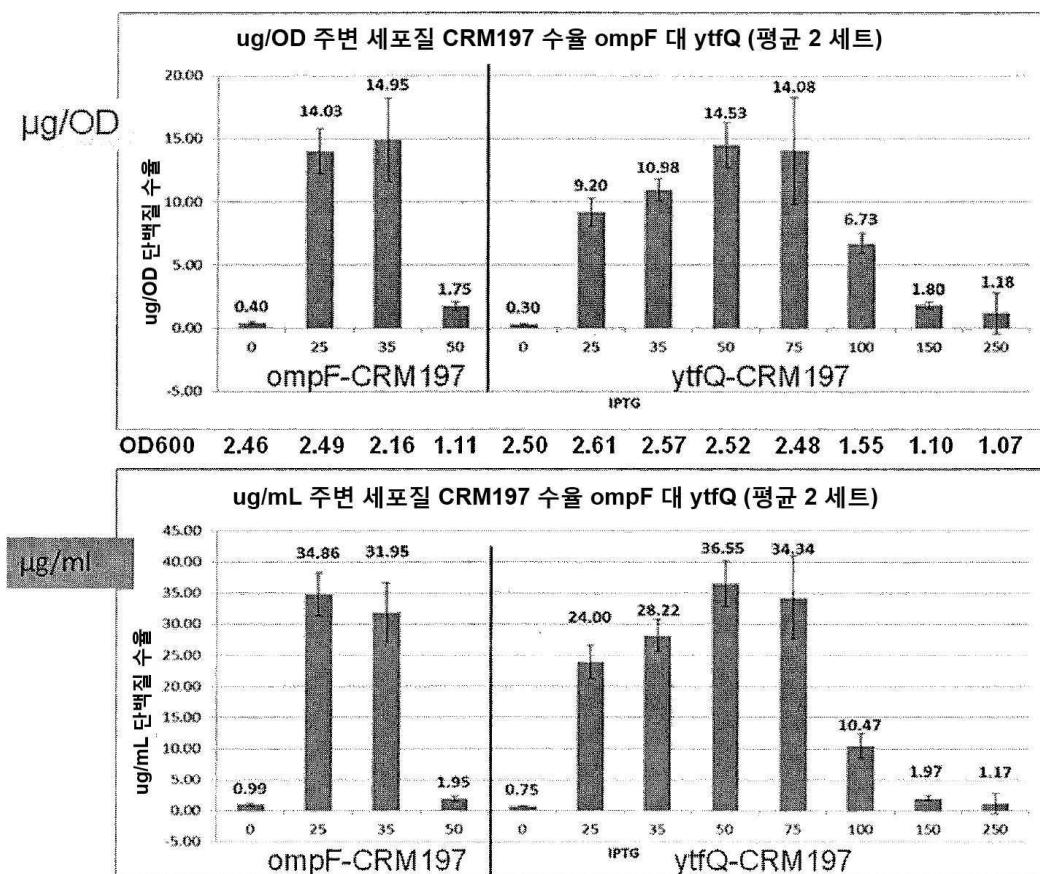
## 도면19



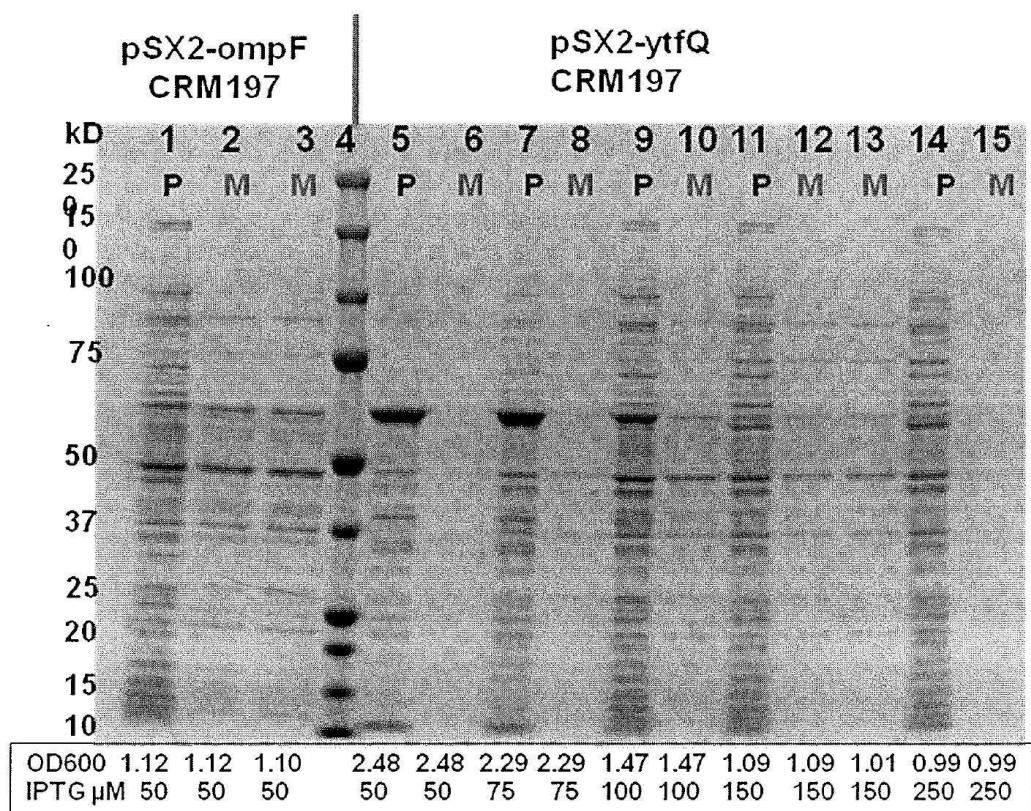
## 도면20



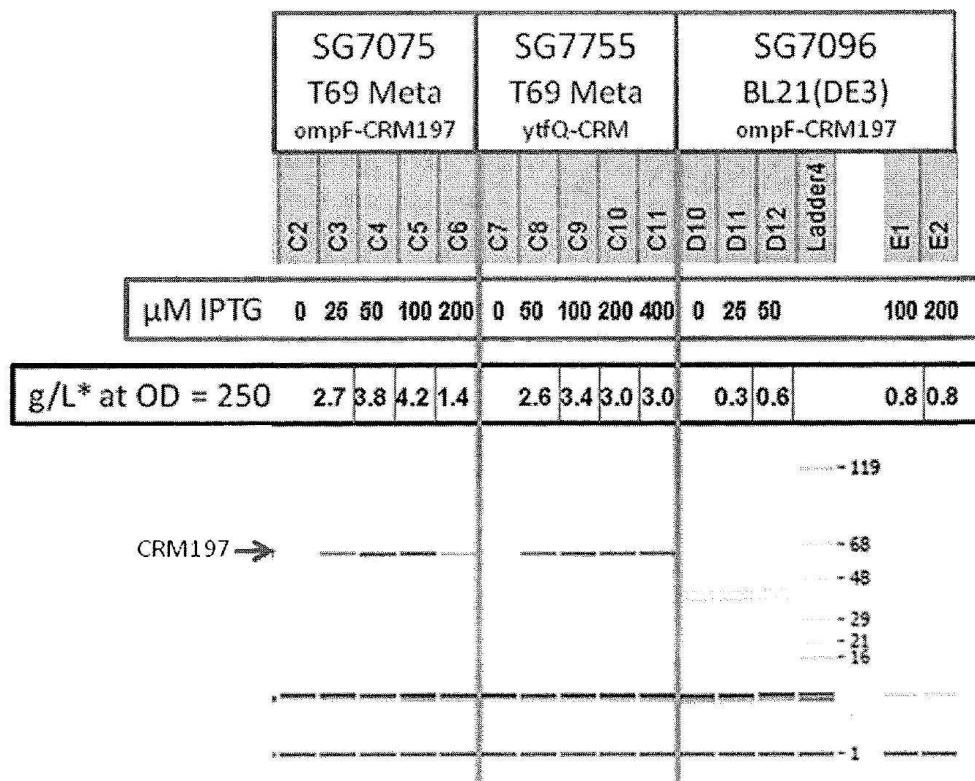
## 도면21



## 도면22



## 도면23



\* OD<sub>600</sub> = 250으로 추정

## 서열목록

SEQUENCE LISTING

<110> Scarab Genomics, LLC

<120> Enhanced Production of Recombinant CRM197 in E. Coli

<130> 010447-5037 WO

<140> PCT/US2015/18338

<141> 2015-03-02

<150> US 61/947,234

<151> 2014-03-03

<160> 38

<170> PatentIn version 3.5

<210> 1

<211> 1608

<212> DNA

## <213> Artificial Sequence

&lt;220&gt;&lt;223&gt; CRM197 sequence optimized for periplasmic expression

&lt;400&gt; 1

ggtgccgatg atttgtgga cagctctaag tctttgtga tggaaaactt tagctgtac	60
cacgtacga agccaggtt atgtcgacagc attcaaaaag gtatccagaa accgaagtc	120

ggcacgcagg gtaactacga cgacgattgg aaagagttct acagcacccga caacaagtat	180
gacgcagcgg gttacagcgt tgacaatgag aatccgttga gcggcaaaagc ggggtgggttt	240
gtcaaagtga cgtatccggg tctgaccaag gtcctggcgt taaaagttga taacgcggaa	300
accattaaga aagagctggg tctgagccgt accgagccgt tggatggagca agttggtacc	360
gaagagttt tcaaacgttt cggcgatggt gcggcccg tttgcctgtc cctgccttcc	420
gcggaggggca gctccagcgt tggatatac aataactggg agcaagcaaa agcgtgtcc	480
gtcgaactgg aatcaattt tggaaacgcgc ggttaaacgtt gtcaagatgc aatgtacgag	540

tatatggccc aggctgcgc tggtaatcgt gttcgctgcgc gcgttggtag cagcttgc	600
tgtatcaacc tggattggga tggatccgt gataagacca agactaagat cgagagccgt	660
aaagaacatg gcccattaa gaacaagatg tcggagagcc cgaataagac cgtgagcga	720
gaaaaggcca agcagtatct ggaagagttt caccacacgg ctctggagca tccggagctg	780
agcgagctga aaacggttac gggcaccaac ccgggtttcg caggtgcga ttacgcggcg	840
tggcagtga atgtggcgca ggtcatcgac tccgaaacgg cggacaattt ggagaaaacc	900
accgcagcgc tgacatttgc gcccggcatc ggcagcgtt tggcattgc agatgggtct	960

gtgcaccata acactgaaga aatcgtagcg caaagcatgt ccctgtctag cttgtatgg	1020
gcgcaggcta ttccgctggt cggcgaactg gttgatatcg gctttgtctgc ctacaacttc	1080
gttggaaagca tcattaacct gtttcaggtt gtcacacaaca gctataatcg cccagcgtac	1140
agcccggtc acaagaccca accgttcctg cacgtggct atgcgggttc ttggAACACG	1200
gtcgaagata gcatcattcg taccggttt cagggcgaga gcccgcattga catcaagatt	1260
actgcagaaa ataccccgct gccgatcgca ggtgtctgc tgcctacgt tccggtaag	1320
ctggacgtt acaaaagcaa aacccacatt tctgtgaacg gtcgtaaatgc tcgcattgc	1380

tgtcgatgcga ttgacggcga cgtcacccatc tgccgtccga agagcccggt ctacgttgg	1440
aatgggtgtc acgcgaacct gcacgtggcg tttcaccgcgc gcaatcgaa gaaaatccat	1500
agcaatgaga ttcttagcga cagcattggc gttctgggtt accaaaagac gttgaccat	1560
acccaaagtca attccaaact gagcgttgc tttgagatca aaagctaa	1608

&lt;210&gt; 2

&lt;211&gt; 535

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Expressed and processed CRM197 amino acid sequence

&lt;400&gt; 2

Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn

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

210	215	220	
Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu			
225	230	235	240
Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu			
245	250	255	
His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val			
260	265	270	
Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val			
275	280	285	
Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu			
290	295	300	
Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala			
305	310	315	320
Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser			
325	330	335	
Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp			
340	345	350	
Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe			
355	360	365	
Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His			
370	375	380	
Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp Asn Thr			
385	390	395	400
Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His			
405	410	415	
Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val			
420	425	430	
Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr			
435	440	445	
His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile			
450	455	460	

Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly

465 470 475 480

Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser

485 490 495

Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu

500 505 510

Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser

515 520 525

Leu Phe Phe Glu Ile Lys Ser

530 535

<210> 3

<211> 1611

<212> DNA

<213> Artificial Sequence

<220><223> CRM197 sequence optimized for expression in cytoplasm

<400> 3

atgggcgcag atgacgtagt agacagcagc aaaagcttcg tcatggaaaa cttagctcg	60
taccacggta cgaagccagg ttatgtcgac agcattcaaa aaggtatcca gaaaccgaag	120
tccggcacgc agggtaacta cgacgacgt tggaaagagt tctacagcac cgacaacaag	180
tatgacgcag cgggttacag cgttgacaat gagaatccgt tgagcggcaa agcgggtggt	240

gttgtcaaag tgacgtatcc gggcttgacc aaggctctgg ctttgaaagt tgataacgcg	300
gaaaccatta agaaagagct gggcttgagc ctgaccgagc ctttgatgga gcaagtttgt	360
accgaagagt ttatcaaacg ttccggcgat ggtgcgagcc gcgttgtcct gtccctgcct	420
ttcgcggagg gcagctccag ctttgagat atcaataact gggagcaagc aaaagcgctg	480
tccgtcaac tggaaatcaa ttttggaaacg cgccgtaaac gtggtaaga tgcaatgtac	540
gagtatatgg cccaggcctg cgctggtaat cgtgttcgtc gcagcgttgg tagcagcttg	600
tcttgatca acctggattt ggtgtgtatc cgtgataaga ccaagactaa gatcgagagc	660

ctgaaaagaac atggcccgat taagaacaag atgtcgaga gcccgaataa gaccgtgagc	720
gaagaaaagg ccaagcagta tctgaaagag ttccaccaaa cggctctgga gcatccggag	780
ctgagcgagc tggaaacggcgt tacggcacc aaccgggtt tcgcaggtgc gaattacgcg	840
gcgtggcag tgaatgtggc gcaggtcatc gactccgaaa cggcggacaa tttggagaaa	900
accaccgcag cgctgagcat tctgccggc atcggcagcg ttatggcat tgcaatgtat	960

gctgtgcacc ataacactga agaaatcgta gcgcaaagca ttgcctgtc tagcttgatg	1020
gtggcgcagg ctatccgct ggtcgccgaa ctgggtgata tcggcttgc tgcctacaac	1080

ttcggtgaaa gcatcattaa cctgttccag gtggtccaca acagctataa tcgcccagcg	1140
tacagcccg gtccacaagac ccaaccgttc ctgcacgatg gctatgcggt gtcttggAAC	1200
acggtcgaag atagcatcat tcgttccggt ttccaggcg agagcgccA tgacatcaag	1260
attactgcag aaaatcccc gctgccgatc gcaggtgtcc tgctgcctac gattccgggt	1320
aagctggacg ttaacaaaag caaaacccac atttctgtaa acggtcgtaa gattcgcatt	1380
cgttgtcgtg cgattgacgg cgacgtcacc ttctgccgtc cgaagagccc ggtctacgtt	1440
gttaatggtg tgcacgcgaa cctgcacgtg gcgtttcacc gcagcagctc ggagaaaatc	1500

catagcaatg agatttctag cgacagcatt ggcgttctgg gttacaaaaa gacggttgac	1560
cataccaaag tcaattccaa actgagcctg ttcttgaga tcaaaagcta a	1611

&lt;210&gt; 4

&lt;211&gt; 22

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; PelB signal sequence

&lt;400&gt; 4

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala

1 5 10 15

Ala Gln Pro Ala Met Ala

20

&lt;210&gt; 5

&lt;211&gt; 21

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; OmpA signal sequence

&lt;400&gt; 5

Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala

1 5 10 15

Thr Val Ala Gln Ala

20

&lt;210&gt; 6

&lt;211&gt; 23

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; StII signal sequence

&lt;400&gt; 6

Met Lys Lys Asn Ile Ala Phe Leu Leu Ala Ser Met Phe Val Phe Ser

1 5 10 15

Ile Ala Thr Asn Ala Tyr Ala

20

&lt;210&gt; 7

&lt;211&gt; 28

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Endoxylanase signal sequence

&lt;400&gt; 7

Met Phe Lys Phe Lys Lys Phe Leu Val Gly Leu Thr Ala Ala Phe

1 5 10 15

Met Ser Ile Ser Met Phe Ser Ala Thr Ala Ser Ala

20 25

&lt;210&gt; 8

&lt;211&gt; 21

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; PhoA signal sequence

&lt;400&gt; 8

Met Lys Gln Ser Thr Ile Ala Leu Ala Leu Pro Leu Leu Phe Thr

1 5 10 15

Pro Val Thr Lys Ala

20

&lt;210&gt; 9

&lt;211&gt; 22

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

<220><223> OmpF signal sequence

<400> 9

Met Met Lys Arg Asn Ile Leu Ala Val Ile Val Pro Ala Leu Leu Val

1 5 10 15

Ala Gly Thr Ala Asn Ala

20

<210> 10

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> PhoE signal sequence

<400

> 10

Met Lys Lys Ser Thr Leu Ala Leu Val Val Met Gly Ile Val Ala Ser

1 5 10 15

Ala Ser Val Gln Ala

20

<210> 11

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> MalE signal sequence

<400> 11

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr

1 5 10 15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala

20 25

<210> 12

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> OmpC signal sequence

<400> 12

Met Lys Val Lys Val Leu Ser Leu Leu Val Pro Ala Leu Val Ala

1 5 10 15

Gly Ala Ala Asn Ala

20

<210> 13

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Lpp signal sequence

<400> 13

Met Lys Ala Thr Lys Leu Val Leu Gly Ala Val Ile Leu Gly Ser Thr

1 5 10 15

Leu Leu Ala Gly

20

<210> 14

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> LamB signal sequence

<400> 14

Met Met Ile Thr Leu Arg Lys Leu Pro Leu Ala Val Ala Val Ala Ala

1 5 10 15

Gly Val Met Ser Ala Gln Ala Met Ala

20 25

<210> 15

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> OmpT signal sequence

<400> 15

Met Arg Ala Lys Leu Leu Gly Ile Val Leu Thr Thr Pro Ile Ala Ile

1 5 10 15

Ser Ser Phe Ala

20

&lt;210&gt; 16

&lt;211&gt; 21

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; LTB signal sequence

&lt;400&gt; 16

Met Asn Lys Val Lys Cys Tyr Val Leu Phe Thr Ala Leu Leu Ser Ser

1 5 10 15

Leu Tyr Ala His Gly

20

&lt;210&gt; 17

&lt;211&gt; 23

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; MglB signal sequence

&lt;400&gt; 17

Met Asn Lys Lys Val Leu Thr Leu Ser Ala Val Met Ala Ser Met Leu

1 5 10 15

Phe Gly Ala Ala Ala His Ala

20

&lt;210&gt; 18

&lt;211&gt; 26

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; OppA signal sequence

&lt;400&gt; 18

Met Thr Asn Ile Thr Lys Arg Ser Leu Val Ala Ala Gly Val Leu Ala

1 5 10 15

Ala Leu Met Ala Gly Asn Val Ala Leu Ala

20 25

&lt;210&gt; 19

&lt;211&gt; 25

&lt;212&gt; PRT

<213> Artificial Sequence

<220><223> RbsB signal sequence

<400> 19

Met Asn Met Lys Lys Leu Ala Thr Leu Val Ser Ala Val Ala Leu Ser

1 5 10 15

Ala Thr Val Ser Ala Asn Ala Met Ala

20 25

<210> 20

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Agp signal sequence

<400> 20

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

1 5 10 15

Ala Ser Asn Ala Gln Ala

20

<210> 21

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> FkpA signal sequence

<400> 21

Met Lys Ser Leu Phe Lys Val Thr Leu Leu Ala Thr Thr Met Ala Val

1 5 10 15

Ala Leu His Ala Pro Ile Thr Phe Ala

20 25

<210> 22

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> YtfQ signal sequence

<400> 22

Met Trp Lys Arg Leu Leu Ile Val Ser Ala Val Ser Ala Ala Met Ser

1 5 10 15

Ser Met Ala Leu Ala

20

<210> 23

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> HdeA signal sequence

<400> 23

Met Lys Lys Val Leu Gly Val Ile Leu Gly Gly Leu Leu Leu Pro

1 5 10 15

Val Val Ser Asn Ala

20

<210> 24

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> HdeB signal sequence

<400> 24

Met Asn Ile Ser Ser Leu Arg Lys Ala Phe Ile Phe Met Gly Ala Val

1 5 10 15

Ala Ala Leu Ser Leu Val Asn Ala Gln Ser Ala Leu Ala

20 25

<210> 25

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> GlnH signal sequence

<400> 25

Met Lys Ser Val Leu Lys Val Ser Leu Ala Ala Leu Thr Leu Ala Phe

1 5 10 15

Ala Val Ser Ser His Ala

20

&lt;210&gt; 26

&lt;211&gt; 1743

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; pelB-CRM197 fusion with 3' and 5' flanking sequence

&lt;400&gt; 26

cctctagaaa taatttgtt taacttttga aggagatata catatgaaat acttgctgcc	60
aaccggccgc gccggcctgc tgctgctcg agcacagccg gctatggcag gtgcggatga	120
tgttgtggac agctctaagt cttttgtat ggaaaacttt agctgttacc acggtaacgaa	180
gccaggttat gtcgacagca ttcaaaaagg tatccagaaa ccgaagtccg gcacgcaggg	240
taactacgac gacgatttga aagagttcta cagcacccgac aacaagtatg acgcagcggg	300

ttacagcggtt gacaatgaga atccgtttag cgccaaggcg ggtgggtttg tcaaagtgtac	360
gtatccgggt ctgaccaagg tcctggcggtt gaaagttgtat aacgcggaaa ccattaagaa	420
agagctgggt ctgagcctga ccgagccgtt gatggagccaa gttggtaccg aagagtttat	480
caaacgtttc ggcgatggtg cgagccgcgt tgcctgtcc ctgccttcg cggagggcag	540
ctccagcggtt gactatatca ataactggga gcaagcaaaa gcgcgttccg tcgaactggaa	600
aatcaatttt gaaacgcgcg gtaaacgtgg tcaagatgca atgtacgagt atatggccca	660
ggcctgcgtt ggtatcgtt ttcgtcgcag cggtggtagc agttgtctt gtatcaacct	720

ggatggat gtatccgtt ataagaccaa gactaagatc gagagcctga aagaacatgg	780
cccgattaag aacaagatgt cgagagccg gaataagacc gtgagcgaag aaaaggccaa	840
gcagtatctg gaagagttcc accaaacggc tctggagcat ccggagctga gcgcgttac	900
aacggttacg ggcaccaacc cgggttcgc aggtgcgtat tacgcggcgt gggcagtgaa	960
tgtggcgcag gtcacgtact ccgaaacggc ggacaatttg gagaacacca ccgcagcgct	1020
gagcattctg cgggcacatcg gcagcggtat gggcatttgc gatggtgcgt tgcaccataa	1080
cactgaagaa atcgtagcgc aaagcattgc cctgtctagc ttgtatggtgg cgccaggctat	1140

tccgctggtc ggcaactgg ttgatatcg ctttgcgttcc tacaacttcg ttgaaagcat	1200
cattaacctg ttccagggtgg tccacaacag ctataatcgc ccagcgatca gcccgggtca	1260
caagacccaa ccgttccgtc acgtggcta tgcgggtct tggAACACGG tcgaagatag	1320
catcattcgt accggtttcc agggcgagag cggccatgac atcaagatata ctgcagaaaa	1380

taccccgctg ccgatgcag gtgtcctgt gcctacgatt ccggtaa	1440
caaaagcaaa acccacattt ctgtgaacgg tcgtaaagg cgcatgcgtt gtcgtgcgt	1500
tgacggcgac gtacacccctt gccgtccgaa gagccggc tacgttgta atggtgtgca	1560
cgcgaacctg cacgtggcgt ttcaccgcag cagctggag aaaatccata gcaatgagat	1620
ttcttagcgac agcattggcg ttctgggta ccaaaagacg gttgaccata ccaaagtcaa	1680
ttccaaactg agcctgttct ttgagatcaa aagctaactc gagccccaa ggcgacaccc	1740
cct	1743
<210> 27	
<211> 33	
<212> DNA	
<213> Artificial Sequence	
<220><223> PCR sense primer (pelB-CRM197)	
<400> 27	
ggagatatac atatgaaata cttgctgcca acc	33
<210>	
28	
<211> 39	
<212> DNA	
<213> Artificial Sequence	
<220><223> PCR antisense primer (PelB-CRM197)	
<400> 28	
ctttgttagc agccgattag ctttgatct caaagaaca	39
<210> 29	
<211> 38	
<212> DNA	
<213> Artificial Sequence	
<220><223> PCR sense primer (OmpA-CRM197)	
<400> 29	
gctaccgtag cgccggccgg tgccgtatgtat gtttgttggaa	38
<210> 30	
<211> 39	
<212> DNA	
<213> Artificial Sequence	

&lt;220&gt;&lt;223&gt; PCR antisense primer (OmpA-CRM197)

&lt;400&gt; 30

ctttgttagc agccgattag ctttgatct caaagaaca 39

&lt;210&gt; 31

&lt;211&gt; 75

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; PCR sense primer 2 (OmpA-CRM197)

&lt;400&gt; 31

ggagatatac atatgaaaaa gacagctatc gcgattgcag tggcactggc tggttcgct 60

accgttagcgc aggcc 75

&lt;210&gt; 32

&lt;211&gt; 39

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; PCR antisense primer 2 (OmpA-CRM197)

&lt;400&gt; 32

ctttgttagc agccgattag ctttgatct caaagaaca 39

&lt;210&gt; 33

&lt;211&gt; 31

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; PCR sense primer 3 (OmpA-CRM197)

&lt;400&gt; 33

ggagatatac atatgaaaaa gacagctatc g 31

&lt;210&gt; 34

&lt;211&gt; 39

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; PCR antisense primer 3 (OmpA-CRM197)

&lt;400&gt; 34

ctttgttagc agccgattag ctttgatct caaagaaca 39

&lt;210&gt; 35

<211> 45  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> PCR sense primer (OmpF-CRM197)

<400> 35  
 gtttagtagca ggtactgcaa acgctggtgc ggatgatgtt gtgga 45

<210> 36  
 <211> 39  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> PCR antisense primer (OmpF-CRM197)

<400> 36  
 ctttgttagc agccgattag ctttgatct caaagaaca 39

<210> 37  
 <211> 78  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> PCR sense primer 2 (OmpF-CRM197)

<400> 37  
 ggagatatac atatgatgaa gcgcaatatt ctggcagtga tctccctgc tctgttagta 60  
 gcaggtactg caaacgct 78

<210> 38  
 <211> 39  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> PCR antisense primer 2 (OmpF-CRM197)  
 <400> 38  
 ctttgttagc agccgattag ctttgatct caaagaaca 39