



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0139149
(43) 공개일자 2015년12월11일

(51) 국제특허분류(Int. Cl.)

C12N 15/70 (2006.01) C07D 417/02 (2006.01)
C07D 417/06 (2006.01) C12N 15/09 (2006.01)
C12P 1/04 (2006.01) C12P 17/14 (2006.01)

(21) 출원번호 10-2014-0067238

(22) 출원일자 2014년06월02일

심사청구일자 2014년06월02일

(71) 출원인

이화여자대학교 산학협력단

서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)

(72) 발명자

윤여준

서울 강남구 압구정로 151, 126동 202호 (압구정동, 현대아파트)

박한오

대전광역시 유성구 엑스포로448 엑스포아파트 208동 601호

(뒷면에 계속)

(74) 대리인

손민

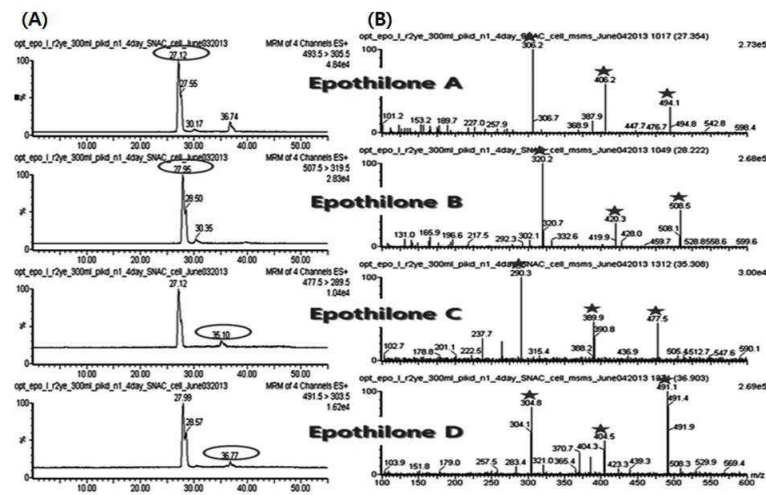
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 에포틸론 발현 카세트 및 이의 용도

(57) 요약

본 발명은 스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자와 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트, 이를 포함하는 에포틸론 발현용 벡터, 및 이를 이용한 에포틸론의 제조방법에 관한 것으로, 본 발명에 따른 에포틸론 발현 카세트 및 발현용 벡터는 생장 속도가 빠르고, 유전자 조작이 용이한 이중숙주 스트렙토마이시스 베네주엘라에 도입할 수 있어, 단시간 동안 높은 수율로 에포틸론을 생산할 수 있으며, 본 발명에 따른 에포틸론의 생산방법에 의하여 생산된 에포틸론은 항암 물질로 유용하게 사용할 수 있다.

대표도 - 도6



(72) 발명자

정소선

대전광역시 대덕구 덕암로 265번길 81 덕암아파트
716호

송재경

충청남도 아산시 신창면 온천대로 725

김응수

서울특별시 강남구 언주로30길 26 삼성타워팰리스
G동 5006호

김은지

서울 서대문구 신촌로 215-2, 803호 (북아현동, 전
진빌딩)

김은지

서울 송파구 송파대로 567, 508동 305호 (잠실동,
잠실5단지아파트)

나지혜

서울특별시 강서구 방화대로 49길 6-16 중앙하이빌
501호

이 발명을 지원한 국가연구개발사업

과제고유번호 1711002649

부처명 미래창조과학부

연구관리전문기관 한국연구재단

연구사업명 원천기술개발사업-글로벌프론티어사업

연구과제명 (지능형바이오시스템)거대 인공 이차대사산물 생합성 유전자집단의 발현시스템 개발

기 여 율 1/2

주관기관 이화여자대학교 산학협력단

연구기간 2011.09.29 ~ 2020.08.31이 발명을 지원한 국가연구개발사업

과제고유번호 1711000101

부처명 미래창조과학부

연구관리전문기관 한국연구재단

연구사업명 기초연구사업(학술진흥)(2013)-중견연구자지원사업-도약연구지원사업(도전)

연구과제명 미생물 유래 천연물의 생합성경로 규명 및 재설계

기 여 율 1/2

주관기관 이화여자대학교 산학협력단

연구기간 2013.06.01 ~ 2016.05.31

명세서

청구범위

청구항 1

스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 소란지움 셀룰로섬(*Sorangium cellulosum*)에서 유래한 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트.

청구항 2

제1항에 있어서,

상기 유전자간 영역은 *epoA* 유전자의 업스트림에 위치하는 영역(서열번호 11), *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 12), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 13)을 포함하는 군에서 이루어진 어느 하나인, 에포틸론 발현 카세트.

청구항 3

스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자인 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로섬(*Sorangium cellulosum*)에서 유래한 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트.

청구항 4

제3항에 있어서,

상기 유전자간 영역은 *epoD* 유전자의 업스트림에 위치하는 영역(서열번호 14), *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 15), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 16), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 17), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 18), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 19), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 20)을 포함하는 군에서 이루어진 어느 하나인, 에포틸론 발현 카세트.

청구항 5

스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 스트렙토마이시스 베네주엘라의 피크로마이신(pikromycin) 또는 데소사민(desosamin) 생합성 유전자 집단에서 유래한 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트.

청구항 6

제5항에 있어서,

상기 유전자간 영역은 *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 21), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 22)을 포함하는 군에서 이루어진 어느 하나인, 에포틸론 발현 카세트.

청구항 7

스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자인 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신(pikromycin) 또는 데소사민(desosamin) 생합성 유전자 집단에서 유래한 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트.

청구항 8

제7항에 있어서,

상기 유전자간 영역은 *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 23), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 24), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 25), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 26), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 27), 및 *orf14* 유전자의 다운스트림에 위치하는 영역(서열번호 28)을 포함하는 군에서 이루어진 어느 하나인, 에포틸론 발현 카세트.

청구항 9

스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), *epoC*(서열번호 4), *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신(pikromycin) 또는 데소사민(desosamin) 생합성 유전자 집단에서 유래한 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트.

청구항 10

제9항에 있어서,

상기 유전자간 영역은 *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 21), *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 22), *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 23), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 24), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 25), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 26), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 27), 및 *orf14* 유전자의 다운스트림에 위치하는 영역(서열번호 28)을 포함하는 군에서 이루어진 어느 하나인, 에포틸론 발현 카세트.

청구항 11

제1항 내지 제10항 중 어느 한 항에 따른 에포틸론 발현 카세트를 포함하는, 에포틸론 발현용 벡터.

청구항 12

제11항에 있어서,

상기 벡터는,

(a) pDHS702 벡터에 제1항에 따른 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 소란지움 셀룰로섬에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pDHS702-optSoC_APBC_Sv;

(b) pDHS618 벡터에 제3항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8),

orf3(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pDHS618-optSoC_DEF_Sv;

(c) pDHS702 벡터에 제5항에 따른 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pDHS702-optPD_APB_C_Sv;

(d) pDHS618 벡터에 제7항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pDHS618-optPD_DEF_Sv;

(e) pSBAC 벡터에 제1항에 따른 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSBAC-optSoC_APB_C_Sv;

(f) pSBAC 벡터에 제3항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSBAC-optSoC_DEF_Sv;

(g) pSBAC 벡터에 제5항 및 제7항에 따른 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), *epoC*(서열번호 4), *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSBAC-optPD_APB_CDEF_Sv;

(h) pSET152 벡터에 제3항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSET152-optSoC_DEF_Sv; 및

(i) pSET152 벡터에 제7항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSET152-optPD_DEF_Sv를 포함하는 군에서 선택되는 어느 하나인 것인, 에포틸론 발현용 벡터.

청구항 13

제11항에 따른 에포틸론 발현용 벡터가 하나 이상 도입된 스트렙토마이시스 베네주엘라를 배양하는 단계를 포함하는, 에포틸론 A, B, C, 및 D를 포함하는 군에서 선택되는 어느 하나 이상의 에포틸론의 제조방법.

청구항 14

제13항에 있어서,

상기 에포틸론 발현용 벡터는 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)를 포함하는 제1항, 제2항, 제5항, 및 제6항 중 어느 한 항에 따른 제1 발현 카세트, 및 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)를 포함하는 제3항, 제4항, 및 제7항 내지 제10항 중 어느 한 항에 따른 제2 발현 카세트를 포함하는 군에서 선택되는 어느 하나 이상의 발현 카세트를 포함하는 것인, 에포틸론의 제조방법.

청구항 15

제14항에 있어서,

상기 발현 카세트는 pDHS702, pDHS618, pSBAC, 및 pSET152로 이루어지는 군에서 선택되는 어느 하나의 벡터에 합성된 것인, 에포틸론의 제조방법.

청구항 16

제13항에 있어서,

상기 벡터는,

(a) pDHS702 벡터에 제1항에 따른 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pDHS702-optSoC_APBC_Sv;

(b) pDHS618 벡터에 제3항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pDHS618-optSoC_DEF_Sv;

(c) pDHS702 벡터에 제5항에 따른 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pDHS702-optPD_APBC_Sv;

(d) pDHS618 벡터에 제7항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pDHS618-optPD_DEF_Sv;

(e) pSBAC 벡터에 제1항에 따른 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSBAC-optSoC_APBC_Sv;

(f) pSBAC 벡터에 제3항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSBAC-optSoC_DEF_Sv;

(g) pSBAC 벡터에 제5항 및 제7항에 따른 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), *epoC*(서열번호 4), *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSBAC-optPD_APBCDEF_Sv;

(h) pSET152 벡터에 제3항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로솜에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSET152-optSoC_DEF_Sv; 및

(i) pSET152 벡터에 제7항에 따른 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 포함하는 pSET152-optPD_DEF_Sv를 포함하는 군에서 선택되는 어느 하나인, 에포틸론의 제조방법.

청구항 17

제16항에 있어서,

상기 벡터 pDHS702-optSoC_APBC_Sv 및 pDHS618-optSoC_DEF_Sv가 도입된 스트렙토마이시스 베네주엘라를 배양하는 단계를 포함하는, 에포틸론의 제조방법.

청구항 18

제16항에 있어서,

상기 벡터 pDHS702-optPD_APB_C_Sv 및 pDHS618-optPD_DEF_Sv가 도입된 스트렙토마이시스 베네주엘라를 배양하는 단계를 포함하는, 에포틸론의 제조방법.

청구항 19

제16항에 있어서,

상기 벡터 pSBAC-optSoC_APB_C_Sv 및 pSBAC-optSoC_DEF_Sv가 도입된 스트렙토마이시스 베네주엘라를 배양하는 단계를 포함하는, 에포틸론의 제조방법.

청구항 20

제16항에 있어서,

상기 벡터 pSBAC-optSoC_APB_CDEF_Sv가 도입된 스트렙토마이시스 베네주엘라를 배양하는 단계를 포함하는, 에포틸론의 제조방법.

발명의 설명

기술 분야

[0001]

본 발명은 스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자와 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트, 이를 포함하는 에포틸론 발현용 벡터, 및 이를 이용한 에포틸론의 제조방법에 관한 것이다.

배경 기술

[0002]

식물 및 미생물 등에서 생산되는 폴리케타이드(polyketide) 계열의 화합물은 복잡한 구조의 천연 화합물(natural product)로서, 현재까지 약 10,000 가지 이상의 화합물이 알려져 있다. 그 중 에포틸론은 16환계 폴리케타이드 매크로락톤(16-membered polyketide macrolactone)의 일종으로 항암 화학요법제인 탁솔(Taxol)과 유사하게 튜불린(tubulin)에 결합하여 미세소관(microtubule)의 분해를 억제하는 강력한 항암제이다.

[0003]

에포틸론 B로부터 반합성 되는 익사베필론(성분명: Ixabepilone, 상품명: Ixampra, Bristol-Myers Squibb)은 2007년 하반기에 유방암의 치료제로 FDA의 승인을 얻었다. 특히 안트라사이클린계(anthracycline)와 탁산계(taxane) 치료에 내성이 생겼거나, 또는 탁산계에 내성이 있고 더 이상의 안트라사이클린계 치료가 금기인 전이성 유방암 또는 국소 진행성 유방암의 치료에 사용이 가능하다. 또한 익사베필론은 탁솔과 다르게 부작용을 유발하는 가용화제(solubilizing agent) 없이 제형화할 수 있으며, 통상적인 화학 치료법(chemotherapy)에 내성이 있는 암세포에 강한 항암 작용을 보이는 장점이 있어 차세대 항암제로 대두되고 있다.

[0004]

에포틸론의 생산은 므소박테리움(myxobacterium)인 소랑지움 셀룰로섬(*Sorangium cellulosum*)에 의해 생산되는 이차대사산물로써 분리되었으나, 16시간에 이르는 긴 생장시간(doubling time)과 유전자 조작의 어려움으로 인하여 그 산업적 생산을 위한 유용성이 떨어지는 문제점이 있다.

[0005]

한편, 경제적인 산업적 가치가 매우 높으나 목적하는 이차대사산물의 대량생산에 기술적 어려움이 있는 경우, 현재의 화학합성을 대체할 수 있는 효율적 방안 중 하나로 대량 생산하고자 하는 이차대사산물의 생합성 유전자집단을 자연생산 생물체로부터 확보하고, 이를 최적화된 이종숙주에서 발현하는 이종숙주 생산시스템(heterologous expression system)이 연구되고 있다. 미국 및 독일의 연구팀에서 이종숙주 생산시스템을 이용하여 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*), 므소코쿠스 잔서스(*Myxococcus xanthus*), 에스케리치아 콜라이(*Escherichia coli*), 그리고 스트렙토마이시스 베네주엘라에서 에포틸론의 생합성을 시도하여 왔다. 독일 Rolf Miller 연구팀에서는 에포틸론 생합성 유전자를 최적화하여 합성한 후 므소코쿠스 잔서스에서 발현하는 기술(Oßwald C et al., ACS Synth Biol., 2012. 10)을 보고한 바 있다.

[0006]

또한, 본 발명자들은 방선균 중 스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에서의 에포틸론 생산방

법에 대하여 출원한 바 있다(한국 특허등록 제 10-0949313호). 그러나 상기 발명들은 에포틸론 B의 경우 약 0.1 $\mu\text{g/L}$, 에포틸론 D의 경우 약 0.4 $\mu\text{g/L}$ 의 수율을 나타내어, 에포틸론을 산업 수요에 맞추어 생산할 만큼 수율이 높지 않다는 문제가 있었는데, 에포틸론의 생산 효율을 높이기 위한 기술이 여전히 요구되고 있다.

[0007] 이러한 배경 하에, 본 발명자들은 유용한 항암물질인 에포틸론의 수율을 높이는 방법을 개발하고자 예의 연구 노력한 결과, 에포틸론 생합성 유전자집단을 이중숙주 스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 적합하도록 최적화하여 합성하고, 상기 합성된 유전자 사이에 소란지움 셀룰로섬 유래, 또는 스트렙토마이시스 베네주엘라의 피크로마이신(pikromycin) 또는 데소사민(desosamin) 생합성 유전자 집단 유래의 유전자간 영역(intergenic region)을 삽입한 발현 카세트를 포함하는 벡터를 스트렙토마이시스 베네주엘라에 도입할 경우 에포틸론의 수율을 현저히 향상시킴을 확인하고 본 발명을 완성하였다.

발명의 내용

해결하려는 과제

[0008] 본 발명의 하나의 목적은 스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 소란지움 셀룰로섬(*Sorangium cellulosum*)에서 유래한 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트를 제공하는 것이다.

[0009] 본 발명의 하나의 목적은 스트렙토마이시스 베네주엘라에 코돈 최적화된 에포틸론 생합성 유전자인 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로섬에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 제공하는 것이다.

[0010] 본 발명의 하나의 목적은 스트렙토마이시스 베네주엘라에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*, *epoP*, *epoB*, 및 *epoC*와, 스트렙토마이시스 베네주엘라의 피크로마이신(pikromycin) 또는 데소사민(desosamin) 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 제공하는 것이다.

[0011] 본 발명의 하나의 목적은 스트렙토마이시스 베네주엘라에 코돈 최적화된 에포틸론 생합성 유전자인 *epoD*, *epoE*, *epoF*, *orf6*, *orf3*, 및 *orf14*와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 데소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 제공하는 것이다.

[0012] 본 발명의 하나의 목적은 스트렙토마이시스 베네주엘라에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), *epoC*(서열번호 4), *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 데소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 에포틸론 발현 카세트를 제공하는 것이다.

[0013] 본 발명의 다른 목적은 상기 에포틸론 발현 카세트를 포함하는, 에포틸론 발현용 벡터를 제공하는 것이다.

[0014] 본 발명의 또 다른 목적은 상기 에포틸론 발현용 벡터가 하나 이상 도입된 스트렙토마이시스 베네주엘라를 배양하는 단계를 포함하는, 에포틸론 A, B, C, 및 D을 포함하는 군에서 선택되는 어느 하나 이상의 에포틸론의 제조 방법을 제공하는 것이다.

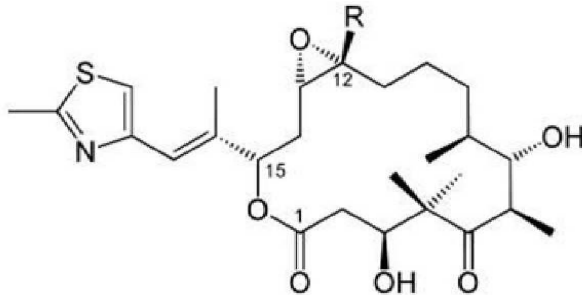
과제의 해결 수단

[0015] 상기의 목적을 달성하기 위한 하나의 양태로서, 본 발명은 스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*)에 코돈 최적화된 에포틸론 생합성 유전자와, 하나 이상의 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트를 제공한다.

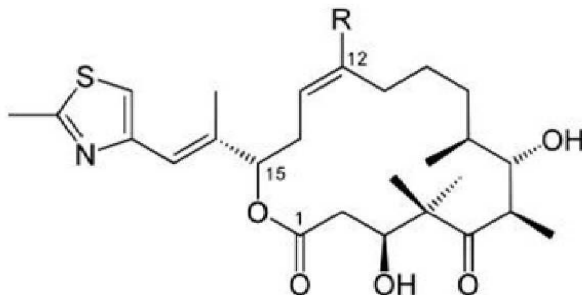
[0016] 본 발명에서 용어, "에포틸론"은 메틸티아졸(methylthiazole)기가 올레핀 결합(olefinic bond)에 의하여 결합된 16환계 폴리케타이드 매크로락톤(polyketide macrolactone)의 일종으로, 저성장(low-growing) 미소박테리움

(myxobacterium)인 소란지움 셀룰로섬에 의해 생산되는 이차 대사산물로써 최초 분리되었다. 상기 에포틸론은 하기 화학식 1로 표시되며 R이 H인 에포틸론 A, 하기 화학식 1로 표시되며 R이 CH₃인 에포틸론 B, 하기 화학식 2로 표시되며 R이 H인 에포틸론 C, 하기 화학식 2로 표시되며 R이 CH₃인 에포틸론 D를 포함할 수 있다. 에포틸론 C와 D는 에포틸론 A와 B의 생합성 초기단계의 중간물질이다.

화학식 1



화학식 2



본 발명의 에포틸론은, 에포틸론의 유사체, 유도체, 역사베필론을 포함하는 의미로 사용될 수 있다.

상기 에포틸론은 튜블린 중합 에세이에서 항종양 활성을 가지는 것으로 보고되었으며(Bollag, D. et al., Cancer Res., 55:2325, 1995), 탁솔이나 탁소텔 등의 항암제처럼 미세소관(microtubule)을 안정화시키고 세포 분열을 억제함으로써 암 세포의 증식을 막는 것으로 알려져 있는 유용한 의약품의 원료이다.

본 발명에서 용어, "발현 카세트"란, 에포틸론 생합성 유전자를 포함하고 있어서 에포틸론을 발현시킬 수 있는 단위 카세트를 의미한다. 또한, 본 발명에서 발현 카세트는 발현구조체와 혼용될 수 있다. 본 발명에 따른 발현 카세트는 본 발명의 목적 상, 스트렙토마이시스 베네주엘라에 도입되어 에포틸론을 생산할 수 있다.

본 발명에서 용어, "코돈 최적화(codon optimization)"는 코돈의 숙주에서의 에포틸론 생합성 유전자의 발현을 보다 효율적으로 하기 위하여 유전자의 코돈을 상기 숙주 유전자에서 높은 빈도로 사용되는 코돈으로 치환하는 것을 의미한다.

상기 숙주는 본 발명의 벡터가 도입되어 에포틸론을 제조할 수 있는 세포는 제한 없이 포함되나, 바람직하게는 방선균, 스트렙토마이시스 속이며, 더욱 바람직하게는 스트렙토마이시스 베네주엘라(*Streptomyces venezuelae*) 또는 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*)이다. 스트렙토마이시스 베네주엘라와 스트렙토마이

시스 실리칼라는 방선균의 일종으로, 스트렙토마이시스 속에 속하는 토양 미생물이며 그람 양성균이다.

[0024] 상기 에포틸론 생합성 유전자는 바람직하게는 소란지움 셀룰로섬에서 유래한 유전자일 수 있다.

[0025] 소란지움 셀룰로섬은 기존에 에포틸론을 생산하는 점액 세균으로 이용되어 왔다. 그러나 소란지움 셀룰로섬은 에포틸론의 수율이 낮을 뿐만 아니라 생장시간(더블링 타임)이 길며 유전자 조작의 곤란으로 인하여 유용성이 떨어진다는 단점이 있었다. 또한 에포틸론의 유기합성방법이 보고된바 있지만, 합성방법이 복잡하다는 단점이 있다. 최근, 에포틸론 생합성 유전자 클러스터의 규명으로 인하여 대장균(*E. coli*)에서의 에포틸론의 이종생산이 연구되고 있으나, 생산되는 에포틸론의 양이 채 1 μ g/L도 되지 않으며, 낮은 온도에서(15 $^{\circ}$ C) 장기간 배양(6일)이 요구되는 단점이 있었다.

[0026] 이에 본 발명자들은 에포틸론의 생산수율을 향상시키기 위하여 이종숙주로 스트렙토마이시스 베네주엘레를 사용하고 이종숙주에 코돈 최적화된 에포틸론 생합성 유전자를 포함하는 벡터를 완성하였다.

[0027] 본 발명의 에포틸론 생합성 유전자는 에포틸론의 생합성에 필요한 *epoA*, *epoP*, *epoB*, *epoC*, *epoD*, *epoE*, *epoF*, *orf6*, *orf3*, 및 *orf14*으로 이루어진 군에서 선택되는 어느 하나 이상의 유전자를 의미한다. 상기 유전자는 본 발명의 목적 상, 스트렙토마이시스 베네주엘레에 코돈 최적화된 서열로서 서열번호 1로 표시되는 염기서열을 갖는 *epoA*, 서열번호 2로 표시되는 염기서열을 갖는 *epoP*, 서열번호 3으로 표시되는 염기서열을 갖는 *epoB*, 서열번호 4로 표시되는 염기서열을 갖는 *epoC*, 서열번호 5로 표시되는 염기서열을 갖는 *epoD*, 서열번호 6로 표시되는 염기서열을 갖는 *epoE*, 서열번호 7으로 표시되는 염기서열을 갖는 *epoF*, 서열번호 8로 표시되는 염기서열을 갖는 *orf6*, 서열번호 9으로 표시되는 염기서열을 갖는 *orf3*, 및 서열번호 10로 표시되는 염기서열을 갖는 *orf14*이다.

[0028] 본 발명에서 용어, "유전자간 영역(intergenic region)"은 유전자와 유전자 사이에 위치하는 DNA서열로, 비해독성(noncoding) 부분을 의미한다(Tropp, Burton E. (2008). Molecular Biology: Genes to Proteins). 이는 인트라제닉 영역(intragenic region)인 인트론(intron)과는 차이가 있다. 유전자간 영역은 기능이 없는 것으로 알려져 있을 뿐, 발현 유전자의 특이적인 발현 증대를 위한 유전자간 영역이 어느 부분인지는 전혀 보고된 바 없다.

[0029] 본 발명자들은 상기 유전자간 영역으로 소란지움 셀룰로섬의 유전자간 영역, 또는 이종숙주로 사용되는 스트렙토마이시스 베네주엘레의 피크로마이신 또는 데소사민 유전자 집단의 유전자간 영역을 사용하여 벡터를 제조한 결과, 에포틸론 유전자의 발현이 매우 향상되어 에포틸론의 수율이 향상됨을 확인하였다.

[0030] 하나의 구체예로서, 본 발명의 에포틸론 발현 카세트는 스트렙토마이시스 베네주엘레에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 소란지움 셀룰로섬에서 유래한 유전자간 영역을 포함하는 것일 수 있다.

[0031] 상기 유전자간 영역은 *epoA* 유전자의 업스트림에 위치하는 영역(서열번호 11), *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 12), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 13)을 포함하는 군에서 이루어진 어느 하나일 수 있다. 바람직하게는 상기 3개의 모든 영역을 포함할 수 있다.

[0032] 상기 유전자간 영역을 포함하는 본 발명의 에포틸론 발현 카세트가 pDHS702 벡터에 합성된 벡터 pDHS702-optSoC_APB_C_Sv의 개열지도를 도 1a에 도시하였으며, 본 발명의 에포틸론 발현 카세트가 pSBAC벡터에 합성된 벡터 pSBAC-optSoC_APB_C_Sv의 개열지도를 도 3a에 도시하였다.

[0033] 또 하나의 구체예로서, 본 발명의 에포틸론 발현 카세트는 스트렙토마이시스 베네주엘레에 코돈 최적화된 에포틸론 생합성 유전자인 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 소란지움 셀룰로섬에서 유래한 유전자간 영역을 포함하는 것일 수 있다.

[0034] 상기 유전자간 영역은 *epoD* 유전자의 업스트림에 위치하는 영역(서열번호 14), *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 15), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 16), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 17), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 18), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 19), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 20)을 포함하는 군에서 이루어진 어느 하나일 수 있다. 바람직하게는 상기 7개의 모든 영역을 포함할 수 있다.

- [0035] 상기 유전자간 영역을 포함하는 본 발명의 에포틸론 발현 카세트가, pDHS618 벡터에 합성된 벡터 pDHS618-optSoC_DEF_Sv의 개열지도는 도 1b에 도시하였으며, pSBAC 벡터에 합성된 벡터 pSBAC-optSoC_DEF_Sv의 개열지도는 도 3b에 도시하였고, pSET152 벡터에 합성된 벡터 pSET152-optSoC_DEF_Sv의 개열지도는 도 4a에 도시하였다.
- [0036] 또 하나의 구체예로서, 본 발명의 에포틸론 발현 카세트는 스트렙토마이시스 베네주엘라에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 것일 수 있다. 스트렙토마이시스 베네주엘라는 PKS 유전자 클러스터로부터 항생제로 사용되는 피크로마이신을 생산하며 PKS는 6개의 모듈로 구성되어 있다. 본 발명자들은 유전자간 영역을 이중숙주인 스트렙토마이시스 베네주엘라가 본래 보유하고 있는 것으로서, 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 벡터에 도입한 결과, 에포틸론의 생합성 수율이 향상됨을 확인하였다. 바람직하게 상기 유전자간 영역은 *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 21), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 22)을 포함하는 군에서 이루어진 어느 하나일 수 있다. 바람직하게는 상기 2개의 모든 영역을 포함할 수 있다.
- [0037] 상기 유전자간 영역을 포함하는 본 발명의 에포틸론 발현 카세트가, pDHS702 벡터에 합성된 벡터 pDHS702-optPD_APB_C_Sv의 개열지도는 도 2a에 도시하였다.
- [0038] 또 하나의 구체예로서, 본 발명의 에포틸론 발현 카세트는 스트렙토마이시스 베네주엘라에 코돈 최적화된 에포틸론 생합성 유전자인 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 것일 수 있다.
- [0039] 상기 유전자간 영역은 *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 23), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 24), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 25), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 26), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 27), 및 *orf14* 유전자의 다운스트림에 위치하는 영역(서열번호 28)을 포함하는 군에서 이루어진 어느 하나일 수 있다. 바람직하게는 상기 6개의 모든 영역을 포함할 수 있다.
- [0040] 상기 유전자간 영역을 포함하는 본 발명의 에포틸론 발현 카세트가, pDHS618 벡터에 합성된 벡터 pDHS618-optPD_DEF_Sv의 개열지도는 도 2b에 도시하였고, 본 발명의 에포틸론 발현 카세트가, pSET152 벡터에 합성된 벡터 pSET152-optPD_DEF_Sv의 개열지도는 도 4b에 도시하였다.
- [0041] 또 하나의 구체예로서, 본 발명의 에포틸론 발현 카세트는 스트렙토마이시스 베네주엘라에 코돈 최적화된 에포틸론 생합성 유전자인 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), *epoC*(서열번호 4), *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)와, 스트렙토마이시스 베네주엘라의 피크로마이신 또는 테소사민 생합성 유전자 집단에서 유래한 유전자간 영역을 포함하는 것일 수 있다.
- [0042] 상기 유전자간 영역은 *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 21), *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 22), *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 23), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 24), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 25), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 26), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 27), 및 *orf14* 유전자의 다운스트림에 위치하는 영역(서열번호 28)을 포함할 수 있다.
- [0043] 상기 유전자간 영역을 포함하는 본 발명의 에포틸론 발현 카세트가, pSBAC 벡터에 합성된 벡터 pSBAC-optPD_APB_CDEF_Sv의 개열지도는 도 3c에 도시하였다.
- [0044] 또 하나의 양태로서, 본 발명은 상기 에포틸론 발현 카세트를 포함하는, 에포틸론 발현용 벡터를 제공한다.

- [0045] 본 발명에서 용어, "벡터"는 숙주세포에 DNA를 도입하여 에포틸론을 미생물에서 효율적으로 발현시키기 위한 수단으로서, 상기 벡터의 제작 시에는 에포틸론 생합성 유전자뿐만 아니라, 에포틸론을 생산하고자 하는 숙주세포의 종류에 따라 프로모터(promoter), 종결자(terminator), 인핸서(inhancer) 등과 같은 발현 조절 서열, 목표적화 또는 분비를 위한 서열 등을 적절히 선택하고 목적에 따라 다양하게 조합할 수 있다.
- [0046] 상기 에포틸론 발현 카세트가 벡터에 포함되는 방법은 특별히 제한되지 아니하며 당해 기술 분야에서 통상적으로 사용되는 방법이라면 어느 것이든 사용할 수 있다.
- [0047] 상기 벡터는 pDHS702, pDHS618, pSBAC, 및 pSET152로 이루어지는 군에서 선택되는 어느 하나의 벡터에 합성된 것일 수 있다.
- [0048] 상기 pDHS702 벡터에 합성되는 벡터는 pDHS702-optSoC_APBC_Sv 또는 pDHS702-optPD_APBC_Sv일 수 있다. 또 다른 구체예로서, 상기 pDHS618 벡터에 합성되는 벡터는 pDHS618-optSoC_DEF_Sv 또는 pDHS618-optPD_DEF_Sv일 수 있다. 상기 벡터 pSBAC에 합성되는 벡터는 pSBAC-optSoC_APBC_Sv, pSBAC-optSoC_DEF_Sv, 또는 pSBAC-optPD_APBCDEF_Sv일 수 있다. 상기 pSET152 벡터에 합성되는 벡터는 pSET152-optSoC_DEF_Sv 또는 pSET152-optPD_DEF_Sv 일 수 있다.
- [0049] 하나의 구체예로서, 본 발명의 벡터는 pDHS702-optSoC_APBC_Sv일 수 있다. 상기 벡터 pDHS702-optSoC_APBC_Sv는 *pikA* I 프로모터를 사용하는 pDHS702 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유전자를 포함하며, 상기 유전자는 바람직하게는 서열번호 1로 표시되는 염기서열을 갖는 *epoA*, 서열번호 2로 표시되는 염기서열을 갖는 *epoP*, 서열번호 3으로 표시되는 염기서열을 갖는 *epoB*, 및 서열번호 4로 표시되는 염기서열을 갖는 *epoC*를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoA* 유전자의 업스트림, 즉 프로모터와 *epoA* 유전자 사이에 위치하는 영역(서열번호 11), *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 12), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 13) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 3개의 영역을 모두 포함할 수 있다. 상기 벡터 pDHS702-optSoC_APBC_Sv의 개열지도를 도 1a에 도시하였다. 본 발명자들은 상기 벡터 pDHS702-optSoC_APBC_Sv를 스트랩토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.
- [0050] 다른 구체예로서, 본 발명의 벡터는 pDHS618-optSoC_DEF_Sv일 수 있다. 상기 벡터 pDHS618-optSoC_DEF_Sv는 *pikA* I 프로모터를 사용하는 pDHS618 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유전자를 포함하며, 상기 유전자는 바람직하게는 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoD* 유전자의 업스트림에 위치하는 영역(서열번호 14), *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 15), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 16), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 17), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 18), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 19), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 20) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 7개의 영역을 모두 포함할 수 있다. 상기 벡터 pDHS618-optSoC_DEF_Sv의 개열지도를 도 1b에 도시하였다. 본 발명자들은 상기 벡터 pDHS618-optSoC_DEF_Sv를 스트랩토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.
- [0051] 또 다른 구체예로서, 본 발명의 벡터는 pDHS702-optPD_APBC_Sv일 수 있다. 상기 벡터 pDHS702-optPD_APBC_Sv는 *pikA* I 프로모터를 사용하는 pDHS702 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유전자를 포함하며, 상기 유전자는 바람직하게는 서열번호 1로 표시되는 염기서열을 갖는 *epoA*, 서열번호 2로 표시되는 염기서열을 갖는 *epoP*, 서열번호 3으로 표시되는 염기서열을 갖는 *epoB*, 및 서열번호 4로 표시되는 염기서열을 갖는 *epoC*를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 21), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 22) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 3개의 영역을 모두 포함할 수 있다. 상기 벡터 pDHS702-optPD_APBC_Sv의 개열지도를 도 2a에 도시하였다. 본 발명자들은 상기 벡터 pDHS702-optPD_APBC_Sv를 스트랩토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.
- [0052] 또 다른 구체예로서, 본 발명의 벡터는 pDHS618-optPD_DEF_Sv일 수 있다. 상기 벡터 pDHS618-optPD_DEF_Sv는 *pikA* I 프로모터를 사용하는 pDHS618 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유

전자를 포함하며, 상기 유전자는 바람직하게는 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 23), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 24), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 25), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 26), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 27), 및 *orf14* 유전자의 다운스트림에 위치하는 영역(서열번호 28) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 3개의 영역을 모두 포함할 수 있다. 상기 벡터 pDHS618-optPD_DEF_Sv의 개열지도를 도 2b에 도시하였다. 본 발명자들은 상기 벡터 pDHS618-optPD_DEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.

[0053]

또 다른 구체예로서, 본 발명의 벡터는 pSBAC-optSoC_APBc_Sv일 수 있다. 상기 벡터 pSBAC-optSoC_APBc_Sv는 *pika* I 프로모터를 사용하는 pSBAC 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유전자를 포함하며, 상기 유전자는 바람직하게는 서열번호 1로 표시되는 염기서열을 갖는 *epoA*, 서열번호 2로 표시되는 염기서열을 갖는 *epoP*, 서열번호 3으로 표시되는 염기서열을 갖는 *epoB*, 및 서열번호 4로 표시되는 염기서열을 갖는 *epoC*를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoA* 유전자의 업스트림, 즉 프로모터와 *epoA* 유전자 사이에 위치하는 영역(서열번호 11), *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 12), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 13) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 3개의 영역을 모두 포함할 수 있다. 상기 벡터 pSBAC-optSoC_APBc_Sv의 개열지도를 도 3a에 도시하였다. 본 발명자들은 상기 벡터 pSBAC-optSoC_APBc_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.

[0054]

또 다른 구체예로서, 본 발명의 벡터는 pSBAC-optSoC_DEF_Sv일 수 있다. 상기 벡터 pSBAC-optSoC_DEF_Sv는 *pika* I 프로모터를 사용하는 pSBAC 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유전자를 포함하며, 상기 유전자는 바람직하게는 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoD* 유전자의 업스트림에 위치하는 영역(서열번호 14), *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 15), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 16), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 17), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 18), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 19), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 20) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 3개의 영역을 모두 포함할 수 있다. 상기 벡터 pSBAC-optSoC_DEF_Sv의 개열지도를 도 3b에 도시하였다. 본 발명자들은 상기 벡터 pSBAC-optSoC_DEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.

[0055]

또 다른 구체예로서, 본 발명의 벡터는 pSBAC-optPD_APBcDEF_Sv일 수 있다. 상기 벡터 pSBAC-optPD_APBcDEF_Sv는 *pika* I 프로모터를 사용하는 pSBAC 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유전자를 포함하며, 상기 유전자는 바람직하게는 서열번호 1로 표시되는 염기서열을 갖는 *epoA*, 서열번호 2로 표시되는 염기서열을 갖는 *epoP*, 서열번호 3으로 표시되는 염기서열을 갖는 *epoB*, 서열번호 4로 표시되는 염기서열을 갖는 *epoC*, *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoP*와 *epoB* 유전자 사이에 위치하는 영역(서열번호 21), *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 22), *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 23), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 24), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 25), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 26), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 27), 및 *orf14* 유전자의 다운스트림에 위치하는 영역(서열번호 28) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 3개의 영역을 모두 포함할 수 있다. 상기 벡터 pSBAC-optPD_APBcDEF_Sv의 개열지도를 도 3c에 도시하였다. 본 발명자들은 상기 벡터 pSBAC-optPD_APBcDEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.

[0056]

또 다른 구체예로서, 본 발명의 벡터는 pSET152-optSoC_DEF_Sv일 수 있다. 상기 벡터 pSET152-optSoC_DEF_Sv는 *ermE** 프로모터를 사용하는 pSET152 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유전자를 포함하며, 상기 유전자는 바람직하게는 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoD* 유전자의 업스트림에 위치하는 영역(서열번호 14), *epoD*와 *epoE* 유전자 사

이에 위치하는 영역(서열번호 15), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 16), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 17), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 18), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 19), 및 *epoC* 유전자의 다운스트림에 위치하는 영역(서열번호 20) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 3개의 영역을 모두 포함할 수 있다. 상기 벡터 pSET152-optSoC_DEF_Sv의 개열지도도 도 4a에 도시하였다. 본 발명자들은 상기 벡터 pSET152-optSoC_DEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.

[0057]

또 다른 구체예로서, 본 발명의 벡터는 pSET152-optPD_DEF_Sv일 수 있다. 상기 벡터 pSET152-optPD_DEF_Sv는 *ermE** 프로모터를 사용하는 pSET152 벡터에 합성할 수 있다. 상기 벡터는 코돈 최적화된 에포틸론 생합성 유전자를 포함하며, 상기 유전자는 바람직하게는 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)를 포함한다. 상기 벡터는 유전자간 영역을 하나 이상 포함하며, 상기 유전자간 영역은 *epoD*와 *epoE* 유전자 사이에 위치하는 영역(서열번호 23), *epoE*와 *epoF* 유전자 사이에 위치하는 영역(서열번호 24), *epoF*와 *orf6* 유전자 사이에 위치하는 영역(서열번호 25), *orf6*과 *orf3* 유전자 사이에 위치하는 영역(서열번호 26), *orf3*과 *orf14* 유전자 사이에 위치하는 영역(서열번호 27), 및 *orf14* 유전자의 다운스트림에 위치하는 영역(서열번호 28) 중 어느 하나 이상을 포함할 수 있다. 바람직하게는 상기 3개의 영역을 모두 포함할 수 있다. 상기 벡터 pSET152-optPD_DEF_Sv의 개열지도도 도 4b에 도시하였다. 본 발명자들은 상기 벡터 pSET152-optPD_DEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.

[0058]

또 하나의 양태로서, 본 발명은 상기 에포틸론 발현용 벡터가 도입된 스트렙토마이시스 베네주엘라를 배양하는 단계를 포함하는, 에포틸론의 제조방법을 제공한다.

[0059]

상기 에포틸론은 에포틸론 A, B, C, 및 D를 포함하는 군에서 선택되는 어느 하나 이상일 수 있다.

[0060]

상기 벡터가 스트렙토마이시스 베네주엘라에 도입되고, 배양하는 방법은 특별히 제한되지 아니하며 당해 기술 분야에서 스트렙토마이시스 베네주엘라에 대하여 통상적으로 사용되는 방법이라면 어느 것이든 사용할 수 있다.

[0061]

하나의 구체예로서, 본 발명의 에포틸론의 제조방법은 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), 및 *epoC*(서열번호 4)를 포함하는 제1 발현 카세트, 및 *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)를 포함하는 제2 발현 카세트를 포함하는 군에서 선택되는 어느 하나 이상을 포함할 수 있다.

[0062]

상기 제1 발현 카세트 및 제2 발현 카세트는 1개의 벡터에 합성되어 1개의 벡터로 합성될 수 있다. 상기 발현 카세트를 이용하여 벡터를 합성하는 방법은 특별히 제한되지 아니하며 당해 기술 분야에서 통상적으로 사용되는 방법이라면 어느 것이든 사용할 수 있다. 합성된 벡터는 바람직하게는 상기 제1 발현 카세트의 다운스트림에 제2 발현 카세트가 합성된 것일 수 있다.

[0063]

상기 제1 발현 카세트 및 제2 발현 카세트는 각각 별개의 벡터에 합성될 수 있다. 별개의 벡터에 합성될 경우에는, 제1 발현 카세트를 포함하는 제1 벡터, 및 제2 발현 카세트를 포함하는 제2 벡터가 스트렙토마이시스 베네주엘라에 도입될 수 있다. 도입되는 방법은 특별히 제한되지 아니하며 당해 기술 분야에서 스트렙토마이시스 베네주엘라에 대하여 통상적으로 사용되는 방법이라면 어느 것이든 사용할 수 있다.

[0064]

다른 구체예로서, 본 발명의 에포틸론의 제조방법에 사용되는 발현 카세트는 pDHS702, pDHS618, pSBAC, 및 pSET152로 이루어지는 군에서 선택되는 어느 하나의 벡터에 합성된 것일 수 있다.

[0065]

또 다른 구체예로서, 본 발명의 에포틸론의 제조방법에 사용되는 벡터는 상기한 pDHS702-optSoC_APB_C_Sv, pDHS618-optSoC_DEF_Sv, pDHS702-optPD_APB_C_Sv, pDHS618-optPD_DEF_Sv, pSBAC-optSoC_APB_C_Sv, pSBAC-optSoC_DEF_Sv, pSBAC-optPD_APB_CDEF_Sv, pSET152-optSoC_DEF_Sv, 또는 pSET152-optPD_DEF_Sv일 수 있으나 이에 제한되지 않는다. 상기 각 벡터에 대한 설명은 상기한 것과 같으므로 생략한다. 상기 발현 카세트는 먼저 제작한 후 벡터로 합성하여 스트렙토마이시스 베네주엘라에 도입할 수도 있고, 한 번에 벡터를 합성하여 도입할 수도 있다.

[0066]

또 다른 구체예로서, 본 발명의 에포틸론의 제조방법은 상기 2개의 벡터 pDHS702-optSoC_APB_C_Sv 및 pDHS618-optSoC_DEF_Sv가 도입된 스트렙토마이시스 베네주엘라를 배양하여 에포틸론 A, B, C, D를 합성할 수 있다. 본 발명자들은 pDHS702-optSoC_APB_C_Sv 및 pDHS618-optSoC_DEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배

양한 결과, 에포틸론 A의 수율은 약 10 μ g/L, 에포틸론 B의 수율은 약 5 μ g/l으로 나타났으며, 에포틸론 C와 D는 각각 에포틸론 A와 B의 수율의 1/10 정도의 수율로 나타나, 에포틸론 생산 수율이 현저히 증가함을 확인하였다(도 6).

[0067] 또 다른 구체예로서, 본 발명의 에포틸론의 제조방법은 상기 2개의 벡터 pDHS618-optPD_APB_C_Sv 및 pDHS702-optPD_DEF_Sv가 도입된 스트렙토마이시스 베네주엘라를 배양하여 에포틸론 A, B, C, D를 합성할 수 있다. 본 발명자들은 pDHS618-optPD_APB_C_Sv 및 pDHS702-optPD_DEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 생산 수율이 현저히 증가함을 확인하였다.

[0068] 또 다른 구체예로서, 본 발명의 에포틸론의 제조방법은 상기 2개의 벡터 pSBAC-optSoC_APB_C_Sv 및 pSBAC-optSoC_DEF_Sv가 도입된 스트렙토마이시스 베네주엘라를 배양하여 에포틸론 A, B, C, D를 합성할 수 있다. 본 발명자들은 pSBAC-optSoC_APB_C_Sv 및 pSBAC-optSoC_DEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 생산 수율이 현저히 증가함을 확인하였다.

[0069] 또 다른 구체예로서, 본 발명의 에포틸론의 제조방법은 1개의 벡터 pSBAC-optSoC_APB_CDEF_Sv가 도입된 스트렙토마이시스 베네주엘라를 배양하여 에포틸론 A, B, C, D를 합성할 수 있다. 본 발명자들은 pSBAC-optSoC_APB_CDEF_Sv를 스트렙토마이시스 베네주엘라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 생산 수율이 현저히 증가함을 확인하였다.

[0070] 또 다른 구체예로서, 상기 벡터 pDHS702-optSoC_APB_C_Sv, pDHS702-optPD_APB_C_Sv, 및 pSBAC-optSoC_APB_C_Sv를 포함하는 군에서 선택되는 어느 하나의 제1 발현 벡터, 및 pDHS618-optSoC_DEF_Sv, pDHS618-optPD_DEF_Sv, pSBAC-optSoC_DEF_Sv, pSET152-optSoC_DEF_Sv, 및 pSET152-optPD_DEF_Sv를 포함하는 군에서 선택되는 어느 하나의 제2 발현 벡터가 도입된 스트렙토마이시스 베네주엘라를 배양하여 높은 수율로 에포틸론 A, B, C, D를 합성할 수 있다.

[0071] 바람직하게는 pDHS702-optSoC_APB_C_Sv, pDHS702-optPD_APB_C_Sv, 또는 pSBAC-optSoC_APB_C_Sv 중 어느 하나의 벡터와 벡터 pDHS618-optSoC_DEF_Sv를 도입하는 방법, pDHS702-optSoC_APB_C_Sv, pDHS702-optPD_APB_C_Sv, 또는 pSBAC-optSoC_APB_C_Sv 중 어느 하나의 벡터와 벡터 pDHS618-optPD_DEF_Sv를 도입하는 방법, pDHS702-optSoC_APB_C_Sv, pDHS702-optPD_APB_C_Sv, 또는 pSBAC-optSoC_APB_C_Sv 중 어느 하나의 벡터와 벡터 pSBAC-optSoC_DEF_Sv를 도입하는 방법, pDHS702-optSoC_APB_C_Sv, pDHS702-optPD_APB_C_Sv, 또는 pSBAC-optSoC_APB_C_Sv 중 어느 하나의 벡터와 벡터 pSET152-optSoC_DEF_Sv를 도입하는 방법, 및 pDHS702-optSoC_APB_C_Sv, pDHS702-optPD_APB_C_Sv, 또는 pSBAC-optSoC_APB_C_Sv 중 어느 하나의 벡터와 벡터 pSET152-optPD_DEF_Sv를 도입하는 방법으로 에포틸론을 생산할 수 있다.

발명의 효과

[0072] 본 발명에 따른 에포틸론 발현 카세트 및 발현용 벡터는 성장속도가 빠르고, 유전자 조작이 용이한 이중속주 스트렙토마이시스 베네주엘라에 도입할 수 있어, 단시간 동안 높은 수율로 에포틸론을 생산할 수 있다.

[0073] 또한 본 발명에 따른 에포틸론의 제조방법에 의하여 생산된 에포틸론은 항암물질으로 유용하게 사용할 수 있다.

도면의 간단한 설명

[0074] 도 1a는 본 발명에 따른 에포틸론 발현용 벡터 pDHS702-optSoC_APB_C_Sv의 개열지도, 도 1b는 pDHS618-optSoC_DEF_Sv의 개열지도 도시한 것이다.

도 2a는 본 발명에 따른 에포틸론 발현용 벡터 pDHS702-optPD_APB_C_Sv의 개열지도, 도 2b는 pDHS618-optPD_DEF_Sv의 개열지도 도시한 것이다.

도 3a는 본 발명에 따른 에포틸론 발현용 벡터 pSBAC-optSoC_APB_C_Sv의 개열지도, 도 3b는 pSBAC-optSoC_DEF_Sv의 개열지도, 도 3c는 pSBAC-optPD_APB_CDEF_Sv의 개열지도 도시한 것이다.

도 4a는 본 발명에 따른 에포틸론 발현용 벡터 pSET152-optSoC_DEF_Sv의 개열지도, 도 4b는 pSET152-optPD_DEF_Sv의 개열지도 도시한 것이다.

도 5는 본 발명에 따른 벡터에 포함되는 에포틸론 생합성 유전자 및 그의 제한효소를 나타낸 것이다.

도 6은 본 발명에 따른 벡터 pDHS702-optSoC_APB_C_Sv와 pDHS618-optSoC_DEF_Sv를 도입한 스트렙토마이시스 베네주엘라 균주에서의 LC-ESI-MS/MS 분석 결과를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

이하 본 발명을 하기 예에 의해 상세히 설명한다. 다만, 하기 예는 본 발명을 예시하기 위한 것일 뿐, 하기 예에 의해 본 발명의 범위가 제한되는 것은 아니다.

균주, 플라스미드, 배양조건

본 발명에서 서브-클로닝(subcloning)을 위하여 대장균(*E. coli*) DH5(Invitrogen), 플라스미드 pLitmus28(New England Biolabs)을 사용하였다. 본 발명에 사용된 대장균 DH5는 서브클로닝에 널리 이용되는 가장 잘 알려진 균주 중에 하나이다. 이 균주는 redAI 과 endAI 유전자의 변이로 인하여 클로닝된 내부 유전자(insert)의 안정성과 플라스미드의 획득 수율을 높여주는 장점이 있다. 한편, 플라스미드 pLitmus28은 3kb 이하의 크기를 가지며, 암피실린(ampicillin)에 저항성을 가지고 다양한 개열지도(cleavage map)를 가지고 있어서 서브클로닝에 널리 이용되는 플라스미드이다.

에포틸론을 스트렙토마이세스 베네주엘라에서 발현시키기 위한 대장균-스트렙토마이시스 셔틀(shuttle) 벡터로 pDHS702, pDHS618 또는 pSBAC를 사용하였다. 상기 벡터들은 *pikA* I 프로모터와 각각 티오스트렙톤(thiostrepton) 및 아프라마이신(apramycin) 내성 유전자를 포함하고 있다. 상기 셔틀벡터 pDHS702와 pDHS618은 미국 미시간 대학교 미생물학과 데이비드 셔먼(David H. Sherman) 교수로부터 확보하였다.

조작된 돌연변이 균주인 스트렙토마이시스 베네주엘라 DHS2001은 한국 등록특허 제10-0949313호에 언급된 균주를 사용하였다.

대장균은 플라스미드의 선별을 위하여 암피실린(ampicillin, 50 µg/mL) 또는 아프라마이신(apramycin, 25 µg/mL)이 첨가된 LB배지에서 배양하였다.

실시예 1: 에포틸론 생합성 유전자 합성 및 발현벡터 제조

에포틸론 생합성 유전자는 소란지움 셀룰로소의 유전자 정보를 바탕으로 이중숙주 스트렙토마이시스 베네주엘라의 코돈 선호도에 맞추어 바이오니아에서 ~2kb의 사이즈로 합성하였다. 합성된 에포틸론 생합성 유전자는 *epoA*(서열번호 1), *epoP*(서열번호 2), *epoB*(서열번호 3), *epoC*(서열번호 4), *epoD*(서열번호 5), *epoE*(서열번호 6), *epoF*(서열번호 7), *orf6*(서열번호 8), *orf3*(서열번호 9), 및 *orf14*(서열번호 10)이다.

상기 합성된 유전자 단편들은 각각의 제한효소를 이용하여 서브클로닝 단계를 거쳐 결합시켜 pDHS702, pDHS618 및 pSBAC에 도입하였다. 서브클로닝 및 도입 방법은 한국 등록특허 제10-0949313호에 기재된 방법을 일부 변형하여 사용하였다.

이를 통하여 하기 표 1과 같이 벡터 pDHS702-optSoC_APB_C_Sv, pDHS618-optSoC_DEF_Sv, pDHS702-optPD_APB_C_Sv, pDHS618-optPD_DEF_Sv, pSBAC-optSoC_APB_C_Sv, pSBAC-optSoC_DEF_Sv, pSBAC-optPD_APB_CDEF_Sv, pSET152-optSoC_DEF_Sv, pSET152-optPD_DEF_Sv를 완성하였다. 각 벡터의 개열지도는 도 1a 내지 도 4b에 도시하였다. 코돈 최적화된 에포틸론 생합성 유전자 단편과 각각의 제한효소를 도 5에 나타내었다.

표 1

실시예 1에 따른 벡터의 특징

벡터	프로모터	벡터	에포틸론 생합성 유전자 및 유전자간 영역(IGR)
pDHS702-optSoC_APB_C_Sv	<i>pikA</i> I	pDHS702	IGR- <i>epoA</i> - <i>epoP</i> -IGR- <i>epoB</i> - <i>epoC</i> -IGR

pDHS618- optSoC_DEF_Sv	<i>pikA</i> I	pDHS618	IGR- <i>epoD</i> -IGR- <i>epoE</i> -IGR- <i>epoF</i> -IGR- <i>orf6</i> -IGR- <i>orf3</i> -IGR- <i>orf14</i> -IGR
pDHS702- optPD_APB_C_Sv	<i>pikA</i> I	pDHS702	<i>epoA</i> - <i>epoP</i> -IGR- <i>epoB</i> - <i>epoC</i> -IGR
pDHS618- optPD_DEF_Sv	<i>pikA</i> I	pDHS618	<i>epoD</i> -IGR- <i>epoE</i> -IGR- <i>epoF</i> -IGR- <i>orf6</i> -IGR- <i>orf3</i> -IGR- <i>orf14</i> -IGR
pSBAC- optSoC_APB_C_Sv	<i>pikA</i> I	pSBAC	IGR- <i>epoA</i> - <i>epoP</i> -IGR- <i>epoB</i> - <i>epoC</i> -IGR
pSBAC- optSoC_DEF_Sv	<i>pikA</i> I	pSBAC	IGR- <i>epoD</i> -IGR- <i>epoE</i> -IGR- <i>epoF</i> -IGR- <i>orf6</i> -IGR- <i>orf3</i> -IGR- <i>orf14</i> -IGR
pSBAC- optSoC_APB_CDEF_Sv	<i>pikA</i> I	pSBAC	<i>epoA</i> - <i>epoP</i> -IGR- <i>epoB</i> - <i>epoC</i> -IGR- <i>epoD</i> -IGR- <i>epoE</i> -IGR- <i>epoF</i> -IGR- <i>orf6</i> -IGR- <i>orf3</i> -IGR- <i>orf14</i> -IGR
pSET152- optSoC_DEF_Sv	<i>ermE</i> *	pSET152	IGR- <i>epoD</i> -IGR- <i>epoE</i> -IGR- <i>epoF</i> -IGR- <i>orf6</i> -IGR- <i>orf3</i> -IGR- <i>orf14</i> -IGR
pSET152- optPD_DEF_Sv	<i>ermE</i> *	pSET152	<i>epoD</i> -IGR- <i>epoE</i> -IGR- <i>epoF</i> -IGR- <i>orf6</i> -IGR- <i>orf3</i> -IGR- <i>orf14</i> -IGR

실시예 2: 이중숙주에서 에포틸론의 발현

(1) 스트렙토마이시스 베네주엘레에의 형질전환

실시예 1에서 제조한 pDHS702-optSoC_APB_C_Sv와 pDHS618-optSoC_DEF_Sv를 스트렙토마이시스 베네주엘레 DHS2001로 도입하고, 상기 균주를 아프라마이신 (25 μ g/mL), 티오스트렙톤 (50 μ g/mL) 및 카나마이신 (50 μ g/mL)를 함유한 R2YE 배지를 이용하여 30℃에서 4일간 배양하였다. R2YE 조성은 103g 수크로오스 (sigma), 0.25g K₂SO₄ (sigma), 10.12g MgCl₂·6H₂O (sigma), 10g 글루코스 (sigma), 0.1g 카사미노 애시드(casamino acids, Difco), 5g 이스트 추출물(yeast extract, Difco), 5.73g TES 버퍼 (sigma), 2ml 미량원소용액(trace element solution), 22g agar (1L 기준)이며, 고온 멸균 autoclave 후에 10ml 0.5% KH₂PO₄ (sigma), 4ml 5M CaCl₂·2H₂O (sigma), 15ml L-프롤린(L-proline, sigma), 7ml 1N NaOH(sigma) (1L 기준)을 첨가하여 제조하였다.

또한 pDHS702-optPD_APB_C_Sv 및 pDHS618-optPD_DEF_Sv의 두 벡터를 스트렙토마이시스 베네주엘레 DHS2001 균주로 도입하고, pSBAC-optSoC_APB_C_Sv 및 pSBAC-optSoC_DEF_Sv의 두 벡터를 스트렙토마이시스 베네주엘레 DHS2001 균주로 도입하고, pSBAC-optPD_APB_CDEF_Sv 벡터를 스트렙토마이시스 베네주엘레 DHS2001 균주로 도입하여 형질전환하고, 상기과 동일한 방법으로 배양하였다.

(2) 에포틸론의 생산여부 확인

상기 배양 용액을 동량의 에틸아세테이트 (ethyl acetate)와 혼합하고, 분별깔때기로 층 분리를 2회 반복하고, 유기층을 진공회전농축기(rotary evaporator)로 농축하였다. 얻어진 용출물을 100 μ l의 메탄올(methanol)로 다시 녹이고 상기 용액을 Phenomenex Synergi Polar-RP 컬럼(1504.6mm, 4 μ m)을 장착한 LC-ESI-MS/MS (액체크로마토그래피/전자스프레이 이온화-질량분석기; Waters/Micromass Quattro micro/MS)로 분석하였다.

분석조건으로는 250 μ l/min의 유속으로 용출시켰고, 0.1%(v/v) 포름산 (formic acid) 이 첨가된 증류수 용액(A)과 0.1%(v/v)의 포름산이 첨가된 80%(v/v) 아세토니트릴 (acetonitrile) 용액(B)를 이용하여 수행하였다. 상기 용액 B를 30%에서 50%로 5분간, 50%에서 90%로 3분간 적용하고, 90%에서 4분간 유지시킨 후 30%로 15분간의 구배로 컬럼의 재평형을 위하여 용출시켰다.

분석 결과, 먼저, pDHS702-optSoC_APB_C_Sv와 pDHS618-optSoC_DEF_Sv를 도입한 스트렙토마이시스 베네주엘레 균주에서 도 6 좌측에 도시된 것과 같이 에포틸론 A, B, C, D를 모두 생산하는 것을 확인하였다. 다른 스트렙토마이시스 베네주엘레 균주에서도 에포틸론 A, B, C, D를 모두 생산하는 것을 확인하였다.

(3) 에포틸론 생산량 정량분석

다중 반응 모니터링 모드(multiple reactions monitoring mode)로 ESI-MS/MS (전자스프레이 이온화-질량분석기/

질량분석기)로 정량분석하였다. 이는 양성자화된 분자 이온(parents)이 에포틸론 A, B, C, D에 특이적인 단편으로 전이되는 것을 탐지하기 위해 셋팅된 2개의 질량 이온(mass ion)을 선별하는 방식으로 수행되었다. 에포틸론 표준물질 A, B는 시그마 (Sigma)에서 구입하였고, 에포틸론 C, D는 소란지움 셀룰로섬에서 분리하였다. 에포틸론의 생산량은 배양액 세 군데에서 채취하여 추출한 추출물의 평균으로 계산하였다.

[0096] 도 6 우측 그래프는 pDHS702-optSoC_APBC_Sv와 pDHS618-optSoC_DEF_Sv를 도입한 스트렙토마이시스 베네주엘레균주를 ESI-MS/MS로 분석한 결과를 나타낸 것이다. 도시된 것과 같이 27.12분의 용출물은 전형적인 단편 부분인 m/z 306.2, 406.2 및 494.1을 보여 에포틸론 A임을 확인하였으며, 27.95분의 용출물은 320.2, 420.3 및 508.5를 보여 에포틸론 B임을 확인하였다. 또한, 35.10분의 용출물은 290.3, 289.9 및 477.5를 보여 에포틸론 C임을 확인하였고, 36.77분의 용출물은 304.8, 404.5 및 491.1을 보여 에포틸론 D임을 확인하였다. 에포틸론 표준물질로 동일하게 실험한 결과 용출시간 및 분획 형태가 동일하였다.

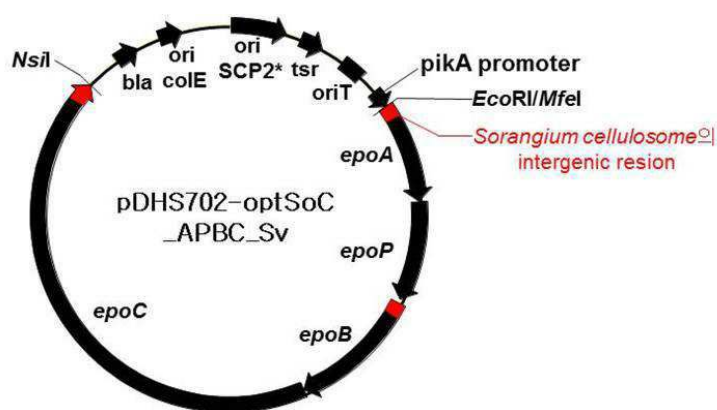
[0097] 에포틸론 A의 수율은 약 10 μ g/L, 에포틸론 B의 수율은 약 5 μ g/l으로 나타났으며, 에포틸론 C와 D는 각각 에포틸론 A와 B의 수율의 1/10 정도, 즉 약 1 μ g/l의 수율로 생산하였다. 이는 본 발명에 따른 발현 카세트 및 이를 포함하는 에포틸론 발현용 벡터에 의하여 에포틸론 생산 수율이 현저히 증가함을 확인한 것이다.

[0098] 이와 같이, 본 발명에 따른 에포틸론 발현용 벡터는 성장속도가 빠르고, 유전자 조작이 용이한 이중숙주 스트렙토마이시스 베네주엘레에 도입할 수 있어, 단시간 동안 높은 수율로 에포틸론 A, B, C, D를 모두 생산할 수 있으며, 본 발명에 따른 에포틸론의 생산방법에 의하여 생산된 에포틸론은 항암물질으로 유용하게 사용할 수 있다.

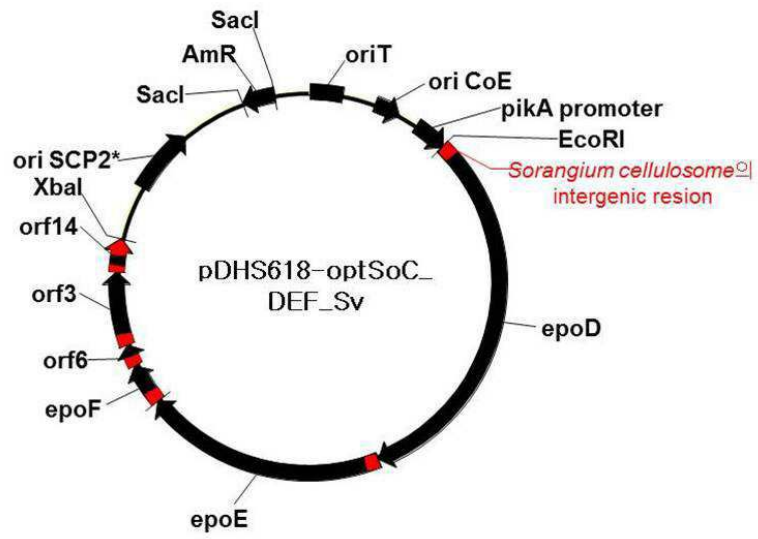
[0099] 이상의 설명으로부터, 본 발명이 속하는 기술분야의 당업자는 본 발명이 그 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 이와 관련하여, 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로서 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허 청구범위의 의미 및 범위 그리고 그 등가 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면

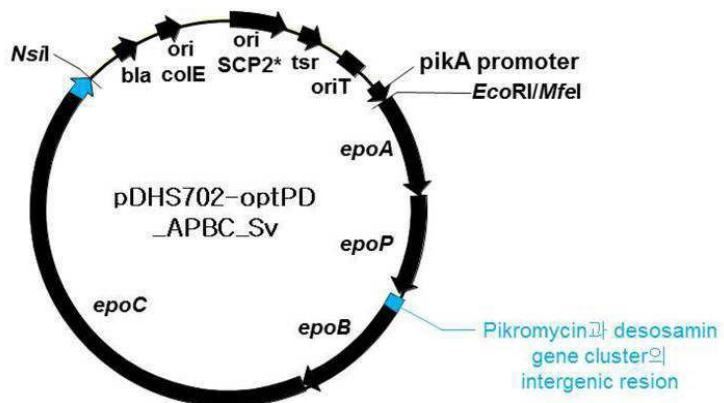
도면1a



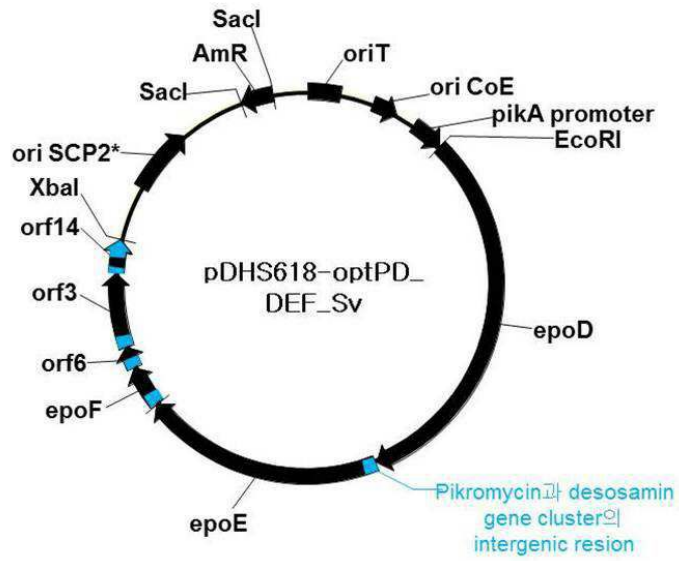
도면1b



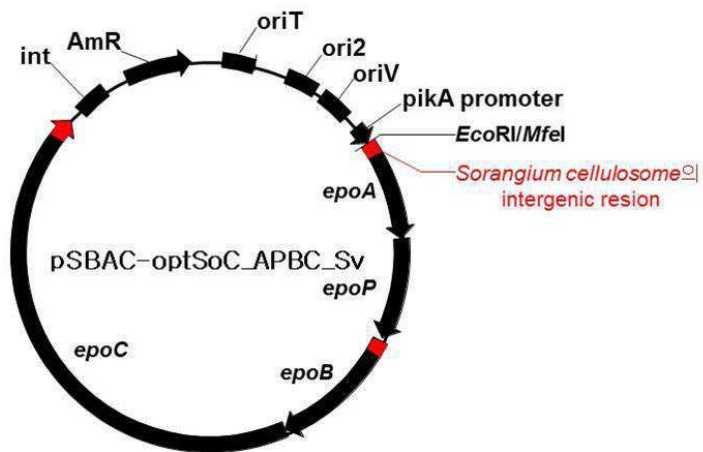
도면2a



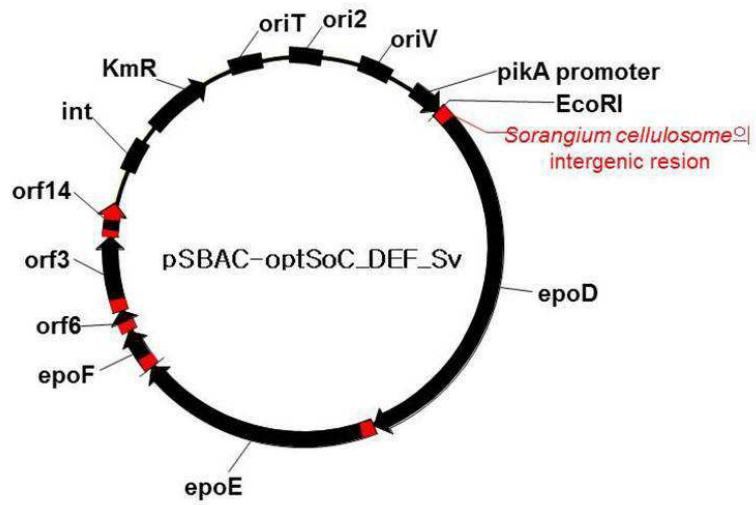
도면2b



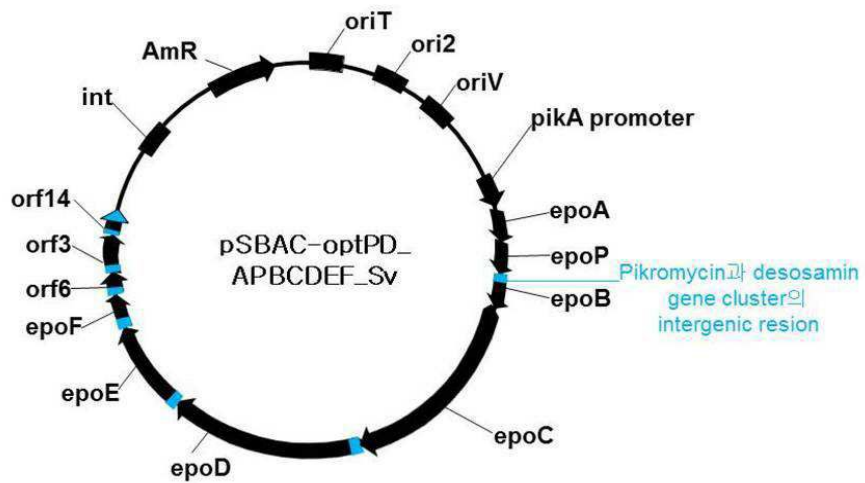
도면3a



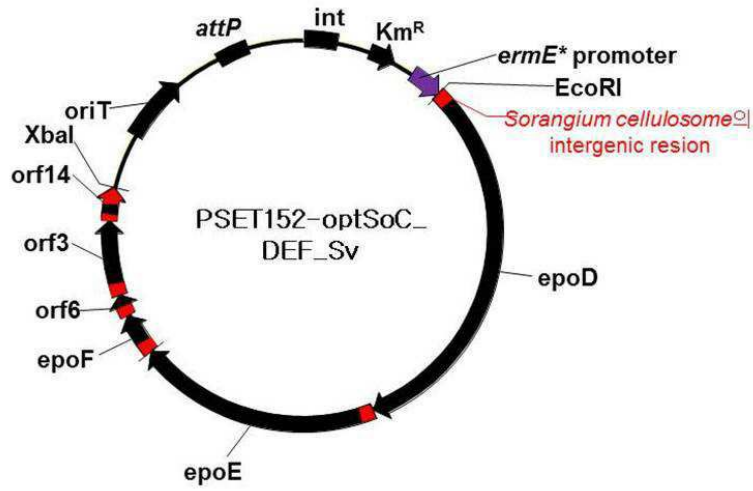
도면3b



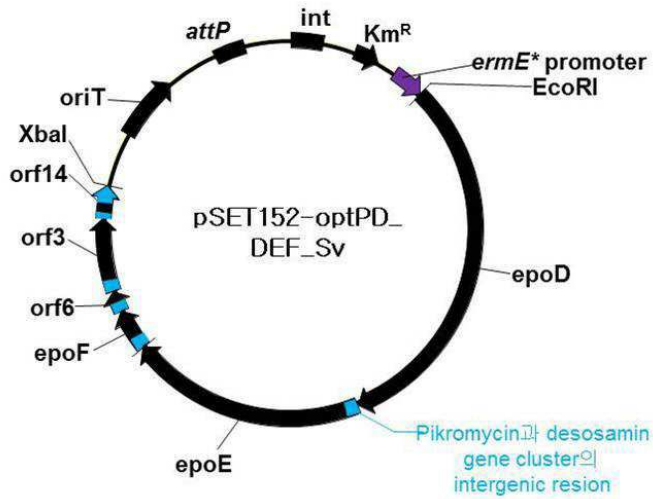
도면3c



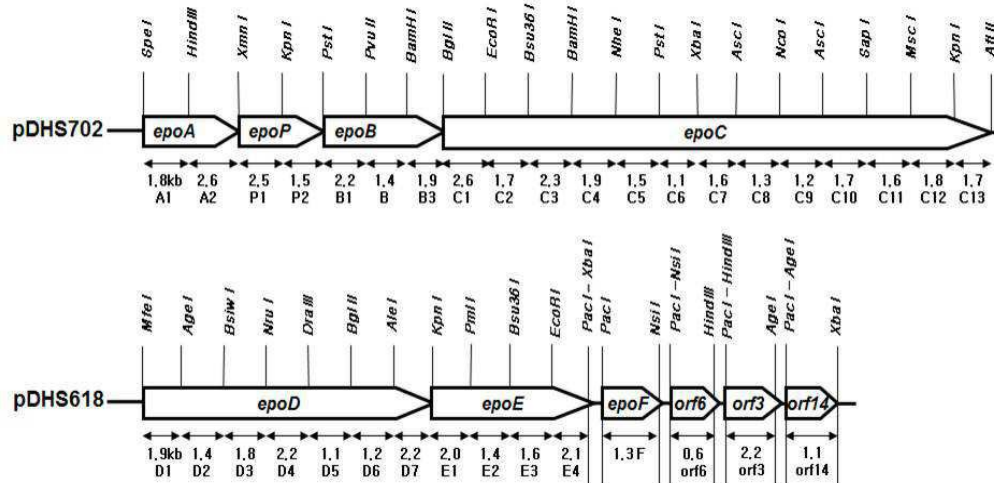
도면4a



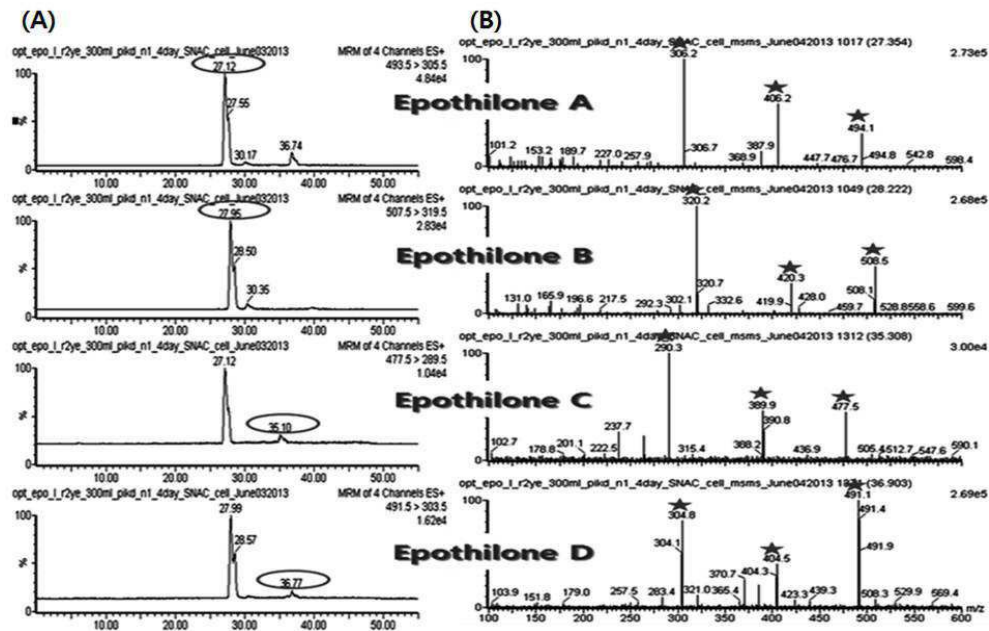
도면4b



도면5



도면6



서열목록

<110> EWEH UNIVERSITY - INDUSTRY COLLABORATION FOUNDATION

<120> EXPRESSION CASSETTES AND THEIR USE

<130> KPA140462-KR

<160> 28

<170> KopatentIn 2.0

<210> 1

<211> 4266

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in Streptomyces venezuelae

<400> 1

atggccgacc gccgatcgca gcgcgccgcc gaggaccgca tcgcatcgct cggagcttcc 60

tgccgcttcc cgggcggcgt catcgacctc tccggtttct ggaccttctc cgaaggttcc 120

cgcgacaccg tcggccgcgt cccggccgaa cgctgggacg ccgctgcctg gttcgaccgg 180

gaccccgacg ccccgggcaa gacccgggtc acccgcgctt ctttctcag tgacgtcgcc 240

tgcttcgacg catctttctt cggtatctcc ccgcgcgagg cgctccgcat ggaccgggcc 300

caccgtctcc tcctcagagt ctgctgggag gccctcgaga acgcccccat cgcctccgtcc	360
gccctcgtcg gcacggagac cggcgtcttc atcggaatcg gcccagagcga gtacgaggcc	420
gccctcccgc agggccaccgc ctccgccgag atcgacgccc acggcgggtct cggcaccatg	480
ccctccgtcg gcgcgggccc catcagctac gccctcggcc tccgcggccc gtgcgtcgcc	540
gtcgataccg cctactcttc ctccctcgtc gccgtgcacc tcgcctgcca gtcctccgt	600
tccggcgagt gctccaccgc cctcgccggc ggctgtctcc tcatgtcttc ccgtccacc	660
ctcgtctggc tctccaagac ccgcgccctc gcccgcgacg gccgtgcaa ggccttctcc	720
gccgaagccg acggtctcgg tcgcggtgag ggctgcgccc tcgtcgtcct gaagcgcttc	780
tccggcggcc gggccgacgg cgaccgcac ctcgccgtca tccgcggctc cgccatcaac	840
cacgacggag ccagctccgg gctcaccgtc ccgaacggct cctcccagga gatcgtcttc	900
aagcgcgccc tcgccgacgc cggctgcgcc gcctcctcgg tcggctacgt cgaggcgcac	960
ggcaccggga cgacctcgg cgacccgac gagatccagg ccctcaacgc cgtctacggc	1020
ctcggctcgc acgtcgccac ccgctcttc atcggtccg tcaagaccaa cctcgccac	1080
ccggaatacg cctccggcat caccggcctc ctcaaggctc tactctccct ccagcacggc	1140
cagatccgg cccacctcca cgcccaggcc ctcaaccccc gcatctcttg gggcgacctc	1200
cgcctacccg tcacccgcgc ccgcaccccg tggccggact ggaacacgcc gcgcgcgcc	1260
gggtgtctct ccttcggcat gtcgggaac aacgcccacg tcgtcctcga ggaggcccg	1320
gccgccacct gcacccgcc ggccttgag cgtccggccg agtcctcgt cctctccgcg	1380
cggaccgcct ccgccctcga tgcacggcc gccgcctcc gcgaccact cgagacctac	1440
ccgtcccagt gcctcggcga cgtcgtttc tcctcgcca ctacgcgtc cgccatggag	1500
caccgcctcg ccgtcgccgc cacatccgt gagggcctcc gcgtgcctt cgacgtgcc	1560
gccagggtc agacctccc gggcgccgtc aggtccatcg ccgactctc caggggcaag	1620
cttgcccttc tcttcaccgg ccagggtgcc cagacgtcg gcatgggccc cggcctctac	1680
gacgtctggt ccgccttcg cgaggccttc gacctctgcg tccgcctctt caaccaggag	1740
ctcgaccgcc cgtccgcga ggtcatgtgg gccgagccgg cctccgtcga cgccgccctc	1800
ctcgaccaga ccgccttcac ccagccggcc ctgttcacct tcgagtacgc gctcgggcc	1860
ctctggcgct cctggggggt cgagccggag ctctcgccg ggcactccat cggcgagctc	1920
gtcggccct gcgtcgccgg cgtcttctc ctcgaggacg ccgtcttctt cgtcgccgcc	1980
cgcggtcgtc tcatgcaggc cctcccggcc ggccggcgca tggcttccat cgaggccccg	2040
gaggccgacg tcgccgcgc cgtcgccgg caccgcgct ccgtctccat cgccgcgtc	2100
aacgccccg accaggtcgt catcgccgg gccggccagc cggtcacgc aatcgccgc	2160

gccatggcgc cccgcggggc cgcaccaag gccctccacg tcagccacgc cttccactcc	2220
ccgtcatg cccgatgct cgaggccttc ggccgcgtcg ccgagtcctg ctctaccgc	2280
cgcgcagca tegtctctg ctccaacctc tccgggaagg cctgtaccga cgaggtctcc	2340
tccccgggat actgggtcgc ccacgccgc gaggtctac gcttcgccga tggagtcaag	2400
gccctccacg ccgcaggcgc cggcaccttc gtcgaggtcg gccgaagtc caccctctc	2460
ggctctctcc cggcctgcat gcccgacgcc cgtccggccc tctcgcctc ctcccgccc	2520
ggccgcgacg aaccggccac cgtcctcgag gcccttgccg gcctctgggc cgtcggcggc	2580
ctcgtctctt gggccggtct cttcccgctc ggccggccc gcgtcccgct ccgacctac	2640
ccgtggcagc gcgagcgcta ctggatcgac accaaggcgg acgacgccgc ccgcggcgac	2700
cgcgcgccc cgggcgcggc ccacgacgag gtcgaggagg gcggcgccgt ccgtggcggc	2760
gaccgccct cgcgccct cgaccaccgc ccgccggaat ccggccgcgc cgagaaggtc	2820
gaggccgcgc gcgacaggcc gttccgactc gagatcgacg agccgggctg cctcgaccac	2880
ctcgtctctc gcgtaccga acgcgcgcc ccgggcctcg gcgaggttga gatcgccgc	2940
gacccgcgg gcctctctt caacgacgtg cagctcgccc tcggcatggt ccggaacgac	3000
ctccgggca agccgaacct gccgtctc ctcggcgggg agtgcgcgg ccgcatctc	3060
gccgtcggcg agggcgta cggcctctc gtcggccagc cggtcacgc actctccgc	3120
ggcgcgttcg ccaccacgt caccaccagc gccgactcg tctcccgcg accgcaggcc	3180
ctctccgca tcgaggccgc cgcatgccc gtcgcctacc tcaccgcgtg gtacgccctc	3240
gaccgcatc cccgctcca gccggcgaa cgcgtctca tccacgccgc caccggcggc	3300
gtcggcctc ccgcctgca gtgggcccag cagtcggcg ccgaggtaca cgccaccgc	3360
ggcacgccg agaagcgagc ctacctgag tccctcggcg tgcgtactg ctccgactc	3420
cgtccgacc gcttcgtggc cgacgtccgc gcctggaccg gcggggaggg cgtcgacgtc	3480
gtctcaact cctctccgg cgagctcatc gacaagtct tcaacctct ccgtccac	3540
ggccgcttcg tcgagctcgc caagcgcgac tgctacgccg acaaccagct cggcctccgc	3600
ccgttctcc gcaacctctc cttctcctc gtcgacctc gcggcatgat gctcgagcg	3660
ccggcccgtg tccgcgcct cctcgaggag ctctcggcc tcacgccgc cggtgtctc	3720
acccgcgcg cgatgccac gctcccatc gcccgctgg ccgacgctt ccgtccatg	3780
gcccaggccc agcacctcgc caagctctc ttgacctcg gagaccgga ggtccagatc	3840
cgcacccga cccatgccg gcgcggccc tccaccggcg accgcacct cctcgaccg	3900

ctcgectcgg ccgccccgcg cgcgcgcgcc gccgccctcg aggccttcct ccgcacccag 3960
gtctcccagg tcttcgcac cccggaaatc aagtcggag ccgagggcct cttcaccgc 4020
ctcggcatgg actccctcat ggccgtcgag ctccgcaacc gcatcgaagc ctccctcaag 4080
ctcaagctct ccaccacatt cctctccacc tccccgaaca tcgccctcct cgcaccagaac 4140
ctcctcgacg ccttggccac cgcctctcc ctccgagcgc tggctgccga gaacctccgc 4200
gccggcgctc agaacgactt cgtctctcc ggagccgacc aggactggga gatcatcgcg 4260
ctgtga 4266

<210> 2

<211> 4233

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in
Streptomyces venezuelae

<400> 2

gtgacatca accagctcct caacgagctc gagcaccagg gtatcaagct cgcggccgac 60
ggcgagcgcc tccagatcca ggccccgaag aacgccctca acccgaaact cctcgccgc 120
atctccgagc acaagagcac catctcacg atgtccgcc agcgctccc ggcgaggtcc 180
atcgtcccgg cccggccga gaggcacgcc ccgtcccgc tcaccgacat ccaggagtcc 240

tactggctcg gccgcaccgg ggcttcacg gtccctccg gcatccacgc ctaccgcgag 300
tacgactgca cggacctcga cgtcccgccg ctctcccgcg ccttcgcaa ggtcgtcgcc 360
cgccacgaca tgcctccgcg ccacacctc ccggacatga tgcaggtcat cgagccgaag 420
gtcgacgccg acatcgagat catcgacctg cggggcctcg accgctccac ccgcgaggcc 480
cgcctcgtct cgctccgca cgcgatgtcc caccgcatct acgacaccga gcgcccgcg 540
ctctaccacg tagtcgccg cgcctcgac gagcgccaga cccgctcgt cctcagcatc 600
gacctcatca acgtcgacct cggcagcctc agcatcatct tcaaggactg gctctcgttc 660

tacgaggacc cggagacgtc cctcccggtc ctccagctgt cctaccgga ctacgtcctc 720
gccctcgaat cccgaagaa gtccgaggcg caccagaggt ccatggacta ctggaagcgc 780
cgcatcgccg agtccccgcc gccccgacc ctcccgatga aggcggacc gtccacctc 840
aaggagatcc gcttccgcca caccgagcag tggctccgt cggactcctg gggccggctc 900
aagcgccgcg ttggcgagcg cggccttacc ccgacggcg tcctcctcgc cgccttctcc 960
gaggtcatcg gccgtggtc cgcgtccccg cgttcaccc tcaacatcac cctcttcaac 1020

cgctccccg tccacccccg cgtcaacgac ataaccggcg acttcacctc catggtcctg	1080
ctggacatcg acacaaccg cgacaagtcc ttcgagcagc gcgccaagcg catccaggag	1140
cagctctggg aggccatgga ccactgcgac gtctccggca tcgaggtgca gcgcgaagcc	1200
gcccgcgtcc tggcatcca gcgcggcgcc ctcttcccgg tegtctcac ctccgccctc	1260
aaccagcagg tcgtggcggt caccagctc cagcgctcg gcaccccggt ctacacctcc	1320
accagacgc cgcagctcct cctcgaccac cagctctacg aacacgacgg cgacctctg	1380
ctcgctggg acatcgtcga cggcgtcttc ccgccggacc tcctcgacga catgctcgag	1440
gcctacgtcg tcttctccg ccgcctcacc gaggagccct ggggggagca ggtccgctgc	1500
tccttcccgc ccgccagct cgaggccgc gcctccgcca acgccacgaa cgccctctc	1560
tccgagcaca cgctccacgg cctcttcgcc gcccgcgtcg agcagctccc gatgcagctc	1620
gcggtcgtct ccgccgcaa gacctcacc tacgaggagc tcagccgccc ctcccgcgg	1680
ctcggggccc gcctccgca gcaggcgcc cgcccgaaca ccctcgtcgc ggtcgtcatg	1740
gagaagggtc gggagcaggt cgtcgccgtc ctccgggtcc tcgagtcgg cgccgcctac	1800
gtcccgatcg acgcggacct gccggccgag cgcattccact acctctcga ccacggcgag	1860
gtcaagctcg tctcaccga gccgtggctc gacggcaagc tctctggcc gccgggcac	1920
cagcgctcc tegtctcga ggccggcgtc gaggcgacg gcgaccagcc gcccatgatg	1980
ccgatccaga ccccgctcga cctcgctac gtcatctata cctcgggtc caccggcctc	2040
ccgaaggcgc tcatgatcga ccaccgggc gccgtcaaca ccatcctcga catcaacgag	2100
cgttctgaga tcggccccgg cgaccgcgtc ctccgctct cctccctcag cttcgacctc	2160
tcctgttacg acgtcttcgg gatctcgcc gccggcgga ccatcgtcgt cccggatgcc	2220
agcaagctcc gcgacccggc ccattgggcc gagtcatcg agcgcgagaa ggtaaccgtc	2280
tggaaactcc tgccggccct catcgcatg ctgctcgagc atttcgaggg ccgcccgac	2340
tccttcgccc gctccctcgc gctctccctc ctctccggcg actggatccc ggtgggcctc	2400
ccggcgagc tccaggccat ccgccgggc gtctccgtca tctccctcgg cggcgccacc	2460
gaggcctcca tctggtccat cgggtaccg gtccgcaacg tcgacctcag ctgggcctcc	2520
atcccgtagc ggcgcccgt ccgcaaccag acgttccacg tcctcgacga agccctcgaa	2580
ccgcgcccgg tctgggtccc gggccagctc tacatcgcg gcgtggggct cgccctcggc	2640
tactggcgcg acgaggagaa gacgcgcaag tccttctctg tgcacccgga gaccggcgag	2700
cgctctaca agaccggcga cctcggccgc tacctccgg acggcaacat cgaattcatg	2760

ggccgcgagg acaaccagat caagctccgc gggtaccgcg tcgagctcgg cgagattgag 2820

gagaccctca agtcccaccc gaacgtccgc gacgccgtga tcgtcccggc cggcaacgac 2880

gctgcgaaca agtcctctct cgcttacgtc gtgccggagg ggaccgcg ccgcgcgcgc 2940

gagcaggatg cctccctcaa gaccgagcgc atcgacgccc gcgccacgc cgccgaagcc 3000

gacggcctct ccgacggcga gcgcgtccag ttcaagctgg cccggcacgg cctccgcgc 3060

gacctcgacg gcaagccggc cgtcgatctc accggccagg acccgcgca agccggcctc 3120

gacgtctacg cccgccgcgc ctccgtccgc acgttcttgg aggcccgat cccgttcgtc 3180

gagttcgggc gcttctcttc ctgcctctcc tccgtggaac cggacggcgc caccctcccc 3240

aagttccgct acccgctccgc gggtccacc taccgggtcc agacctacgc ctacgtgaag 3300

tccggccgca tcgagggcgt cgacgaaggc ttctactact accaccggtt cgagcacctg 3360

ctgtgaagc tctccgacca cggcatcgag cgcggcgccc acgtgcgcca aaacttcgac 3420

gtcttcgacg agggcggtt caacctctc ttctcgcc gcctcgacgc catcgagtcc 3480

ttgtacggct ccagctcccg cgagttctgc ctgctcgagg ccggtacat ggcccagctg 3540

ctcatggagc agggcccttc ctgcaacatc gggtctgcc cggtcggcca gttcaacttc 3600

gagcaggctc gcccggtctc cgacctccga cactccgacg tctacgtcca cggcatgtc 3660

ggcgcccgcg tcgaccgcgc gcagttccag gtctgcaccc tcggccagga ctctccccg 3720

cgccgcgcca ccaccgcgg cgccccgcgc ggccgcgagc agcacttcgc cgacatgtc 3780

cgcgacttcc tccgaccaa gctccggag tacatggtcc cgaccgtctt cgtcgagtc 3840

gacgccctcc cgtcacgtc caacggcaag gtcgaccgca aggcctccg cgagcgcaag 3900

gacacctctc ccccgcgcca ctccggccac accgccccgc gcgacgcgt cgaggagatc 3960

ctcgtcgccg tcgtccgca ggtactcggc ctcgaggctg tcgggctgca gcagtccttc 4020

gtcgacctcg gcgcgacctc catcacatc gtccgcatgc gctccctctc ccagaaacgc 4080

ctcgaccgcg agatcgccat caccgagctc ttccagtacc cgaacctggg ctcgctcgcc 4140

tccggcctcc gccgcgactc gcgcgatctc gaccagcgc cgaacatgca ggaccgcgtc 4200

gaggtccgcc gcaagggcgc ccgccgtcc tga 4233

<210> 3

<211> 5499

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in
Streptomyces venezuelae

<400> 3

atggaagagc aggagtcctc cgccatcgcc gtcacgaggga tgtccggccg cttcccgggc	60
gcccgcgacc tcgacgagtt ctggcgcaac ctccgcgacg gcaccgaggc cgtccagcgc	120
ttctccgagc aggagctcgc cgctccgggt gtcgaccggg ccctcgtcct cgaccctcc	180
tacgtccgcg ccggcagcgt ccttgaggac gtcgacgct tcgacgccgc cttcttcggc	240
atctccccgc gcgaggccga gcttatggac ccgcagcacc gcatcttcat ggagtgcgcc	300
tgggaggccc tcgagaacgc cggatacgac ccgaccgcct acgagggtc catcggcgtc	360
tacgccggcg ccaacatgic ctctacctg accagcaacc tccacgagca cccggccatg	420
atgcgctggc cgggctggtt ccagaccctc atcggcaacg acaaggacta cctcgccacc	480
cacgtctcct accgcctgaa cctccgcggc cgtccatct cgtccagac ggctgttcc	540
acctccctcg tcgccgtcca cctcgcatgc atgtccctcc tcgaccgtga gtgcgacatg	600
gctctcgccg gcggcatcac cgtccgcac cgcaccgcg ccggctacgt ctacccgag	660
ggcggcatct tctccccgga cggccactgc cgcgccttcg acgccaaggc caacggcacc	720
atcatgggca acggatgcgg cgtcgtgctc ctcaagccgc tcgaccgcgc cctctccgac	780
ggtgaccggg tccgcgccgt catctcggc tccgccacca acaacgacgg cgcccgaag	840
atcggcttca ccgccccgtc tgaggtcggc caggcccagg ccatcatgga ggccctcgcc	900
ctcgcgggcg tcgaagcccg ctccatccag tacatcgaga cccacggcac cggcacctc	960
ctcggcgacg ccatcgaaac cgccgccctc cgccgcgtct tcggcaggga cgcctccgcc	1020
cgcaggtcct gcgccatcgg ctccgtcaag accggaatcg gccacctga gtccgccgcc	1080
ggcatcgccg gcctcatcaa gaccgtctc gccctcgaa accgccagct cccccgtcc	1140
ctcaacttcg agtccccgaa cccgtccatc gacttcgct cctccccgtt ctacgtcaac	1200
acctccctca aggactggaa caccggctcc accccgcgcc gcgccggcgt ctctctctc	1260
ggcatcgccg gcaccaacgc ccacgtggtc ctcgaggagg cccccccgc caagctcccg	1320
gccgccgccc ccgccgctc cgccgagctc ttcgtcgtct ccgccaagtc cgccgccgcc	1380
ctcgacgtcg ccgccgccg cctccgcgac cactgcagg cccaccagg gatctccctc	1440
ggcgacgtcg cttctccct cgcgaccacc cgtccccga tggagcaccg cctcgccatg	1500
gccccccga gcccgaggc cctccgcgaa ggctcgacg ccgccgccg cggccagacc	1560
ccgccgggtg ccgtccggc ccgtgtctc ccgggcaacg tcccgaagg cgtcttcgtc	1620
ttccgggggc agggttccca gtgggtcggc atgggtcgcc agtcctcgc cgaggagccg	1680

gtgttccacg ccgcccctctc cgcttgcgac cgcgccatcc aggccgaggc cggctgtgtcc	1740
ctctctcgccg aactcgccgc cgacgaaggc tctcccagc tcgagcgcat cgacgtcgtc	1800
cagccgggtcc tcttcgcctc cgccgtcgcc ttccgccc tctggcgag ctggggcgtc	1860
gccccggacg tcgtcattgg aactccatg ggcgaggctc ccgcccga cgtcgccggc	1920
gcccctctct tggaggacgc agtcgccatc atctgccgcc gctcccgcct cctccgcgc	1980
atctccggcc agggcgagat ggccgtcacc gagctgtccc tggccgaggc cgaggccgcc	2040
ctccgcggt acgaggaccg cgtctccgtc gccgtcagca actccccgcg ctccaccgtc	2100
ctctccggcg agccggccgc catcggcgag gtctctctct ccctcaacgc caagggggtc	2160
ttctgccgcc gcgtaaaggt cgacgtcgcc tcccactccc cgcaggtcga cccgtccgc	2220
gaggacctcc tcgcccctc cgccggcctc cccccggggg ccgcccgcgt cccgatgcgc	2280
tccaccgtca ccggcgctat ggtcgccggc ccggaactcg gcgccaacta ctggatgaac	2340
aacctccgcc agccgggtccg cttcgccgag gtctgtccagg cccaactcca gggcgccac	2400
ggcctcttcg tcgagatgtc gccgcacccc atctcacca cctccgtcga ggagatgcgc	2460
cgcgccgcc agcgagccgg tgcgcagta ggctccctcc gccggggcca ggacgaacgc	2520
ccggccatgc tcgaggccct cgccaccctc tgggcccagg gttaccgggt cccgtggggc	2580
cgcctcttcc cgccggcgcc cgccgcgta ccgtcccga cctaccgtg gcagcgcgag	2640
cgtactgga tcgaggcgcc ggccaagtcc gcccgggcg accgccgtgg cgtccgcgcc	2700
ggcgccacc cgtctctcgg cgagatgcag acccttcca cccagacctc caccgcctc	2760
tgggaaacca ccctgcacct caagcctc ccgtggctcg gagatcacg cgtccagggc	2820
gccgtcgtct tcccgggcgc cgctacctc gagatggcca tctctcggg cgccgaggcc	2880
ctcggggacg gcccgctcca gatcaccgac gtctctctcg ccgaggccct cgccttcgcc	2940
ggtgacgccg ccgtcctcgt ccaggctgtc actaccgagc agccgtcagg ccgctccag	3000
ttccagatcg ccagccgggc cccgggcgcc ggccatgcct ccttcgcgt ccacgccgc	3060
ggagctctcc tcagggtcga gcgcaccgag gtcccggccg gtctaccct ctccgccgtc	3120
cgcgccgcc tccaggctc catccggcc gccgccact acgccgagct caccgagatg	3180
ggcctccagt acggcccggc cttccagggt atcgccgagc tctggcgcg cgagggagag	3240
gcccctggcc gcgtccgcct cccggacgcc gccggatccg ccgccgagta ccgctccac	3300
cccgcctcc tcgacgcctg cttccagatc gtcggctccc tcttcgccg ctccggcgag	3360
gccacgccgt gggctccgggt cgagctcggc tccctccgcc tgetccagcg cccttccggc	3420
gaactctggt gccacgccg cgtcgtcaac cagggccacc agaccccgga ccgccagggc	3480
gccgacttct gggctcgtcga ctctccggc gccgtcgtcg ccgaggctcg cggcctcgtg	3540

gcccagcgcc tccccggcgg cgtccgccgc cgcgaggagg acgactggtt cctcgagctc	3600
gagtgggagc cggccgccgt cggcaccgcc aaggtcaacg ccgacgctg gtcctcctc	3660
ggcgggtgggg gggcctcgg agccgccctc cgcgccatgc tcgaagccgg cggccacgca	3720
gtcgtccacg ccgccgagaa caacacctcc gccgccggcg tccgcgact cctcgccaag	3780
gccttcgacg ggcaggcccc gaccgccgtg gtccacctcg gtcctcctga cggcggcggc	3840
gagctcgacc cgggactcgg cggccaggga gccctcgacg ccccgcgctc cggcgcgctc	3900
tccccggacg cctcgaccc ggccctcgtc cgcggctgcg actcggctct ctggaccgtc	3960
caggcgctcg ccggcatggg cttccgcgac gccccgcgcc tctggctctt gaccgcggc	4020
gcccaggccg tcggtgccgg cgacgtctcc gtcacccagg ccccgctcct cggcctcggc	4080
cgcgtcatcg ccatggagca tgccgacctc cgttgcgccc gcgtcgacct cgaccggcc	4140
cggccggaag gcgagctcg cgcctcctc gcagagctcc tcgctgacga cggcaggcc	4200
gaggtcgccc tcagggcgcg cgagcgctgc gtcgcgcgca tcgtccgccg ccagccggag	4260
accgccccgc gaggtcgcat cgagtcctgc gtcgccaccg acgtcaccat ccgcgccgac	4320
tccacctacc tcgtaccgg cggcctcggc ggccctcggc tctccgtcgc aggtcggctc	4380
gccgagcgcg gcgccgtca cctcgtctc gtcggccgat ccggcgccgc ttcgtcgag	4440
cagcgcgccg ccgtcgccgc cctcgaggcc cgcggcgccc gcgtcacct cgccaaggcc	4500
gacgtcgccg accgagccca gctcgagcgc atcctccgcg aggtcaccac cagcggcatg	4560
ccgtccgcg gcgtcgtcca cgcggccggc atcctcgacg acggcctcct catgcagcag	4620
accccgcccc gcttcgcaa ggtcatggcc ccgaaggctc agggcgccct ccacctccac	4680
ggcctcacc cgcaggcccc gctctcttc ttcgtctct acgcctccgg cgtgggcctc	4740
ctcggctccc cgggacaggg caactacgcc gccccaaca cattcctga cgcctcggc	4800
caccaccgcc gcgccaggg tctccggcc ctctccgtcg actggggcct cttcgccgag	4860
gtcggcatgg ccgccgccca ggaggaccgt ggcgtcggc tcgtctccg cggtatgcgc	4920
tcctcacc cggacagagg cctctccgc ctcggccgc tccttgagtc cggccgcgcc	4980
caggtcggcg tcatgccgt caacccccgc ctctgggtcg aactctacc ggccgccgcc	5040
agtccccga tgcctcccc cctcgtcacc gccaccgcg cctcgccgg cggcccgcc	5100
ggcgacggcg acctcctccg ccgcctcggc gccgccgagc cgtccgccg ctcgcctc	5160
ctcgagccgc tctccgcgc ccagatctcc caggtcctcc gcctcccgga gggcaagatc	5220
gaggtggacg ccccgctcac ctccctcggc atgaactccc tcattgggtct cgagctccgc	5280
aaccgcatcg aggcatgct cggcatcacc gtcccgcca cctcctctg gacctaccg	5340
accgtcgccg cctctccgg ccacctcgc cgcgaggcct gcgaggccgc cccggtcgag	5400

tccccgcaca ccaccgccga ttccggccgta gagatcgagg aaatgtccca ggacgacctc 5460

accagctca tcgccgcca gttcaaggcc ctcacgtga 5499

<210> 4

<211> 21774

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in
Streptomyces venezuelae

<400> 4

gtgaccaccc gtggcccgcac cgcccagcag aaccgcgtca agcaggccgc catcatcatc 60

cagcgcctcg aggagcgctt cgccggcctc gcccaggccg agtcgagcg caccgagccg 120

atgccatcg tcggcatcgg ctgccgttc cggggcggcg ccgacgccc ggaggccttc 180

tgggagctcc tcgacgccga gcgcgacgcc gtccagccgc tcgacatcg ctgggcccctc 240

gtcggcgctg ccccggtcga ggccgtccc cactgggccc gcctcctcac cgagccgac 300

gactgcttcg acgcccgtt cttcggcctc tccccgcgc agggccgctc cctcgaccg 360

cagcaccgcc tctcctcga agtcgcctgg gaggcctcg aggacgccc catccgccc 420

cgctccatcg acggctccc caccggcgtc ttctcggcg ccttcaccgc cgactacgc 480

cgcaccgtcg ccgcctccc gcgcgaggag cgcgacgct actccgccac cggcaaatg 540

ctctccatcg ccgcggcgcc cctcagctac accctcgcc tccagggcc gtgcctcacc 600

gtcgacaccg cctgtctctc ctccctctc gccatccacc tcgctgccc ctccctccgc 660

gccggtgagt ccgacctgc cctcgccgc ggctctcctg ccctcctctc cccggacatg 720

atggaggccg ccgcccgcac ccaggccctc tccccggacg gccgtgccc caccttcgac 780

gcctccgcca acggcttctg ccgcggcgag ggctgcggcc tcgtcgtcct caagcgctc 840

tccgatccc agcgcgacgg cgaccggatc tgggccttga tccgcggctc cgccatcaac 900

cacgacggtc gctccaccg cctcaccgcc ccgaacgtc tcgccagga gaccgtctc 960

cgcgaggccc tccgtccgc ccacgtcgag gccggcgccc tcgactacgt cgagaccac 1020

ggcaccggca cctccctcgg cgacctatc gaggtcgagg ccctccgcgc caccgtcggc 1080

ccggcccctg ccgacggcac ccgtgcgtc ctccggcgcg tcaagaccaa catcgccac 1140

ctcaggccg ccgcggcgt cgccggcctc atcaaggccg ccctctcct caccacgag 1200

cgcacccgc gcaacctca cttccgact ctcaaccgc gcatccgct cgagggtcc 1260

gccctcgccc tcgccaccga gccggtccc tggccgcgca ccgaccgcc gcgttcgcc 1320

ggcgtctcct ccttcggcat gtccggcacc aacgcccacg tcgtcctcga ggaggccccc	1380
gccgtcgaac tctggccggc cgcgccggag cgctccgccg agctcctcgt cctctccgga	1440
aagtccgagg gcgccctcga cgcgccaggcc gcccgccctcc gcgagcacct cgacatgcac	1500
ccggaactcg gcctcggcga cgtcgccttc tcccttgcca ccaccgcctc cgccatgtcc	1560
caccgcctcg ccgtcgcctg cacctcccgc gagggcctcc tcgccgcctc ctcgccctc	1620
gcccagggcc agaccccgcc cggcgccgcc cgtgcatcg cctcctcctc ccgcggaag	1680
ctcgccttc tcttcaccgg ccaggcgcc cagaccccg gcatggcgcc cggcctgtgc	1740
gccgcctggc cggccttcgg caggccttc gaccgctcgc tcgccctctt cgaccgcgag	1800
ctcgaccgcc cgctccgca ggtcatgtgg gccgaggccg gctccgcca gtcctcctc	1860
ctcgaccaga ccgccttcac ccaaccggcc ctcttcgccg tcgagtacgc cctcaccgcc	1920
ctctggcgct cctggggggt cgagccggag ctctcgtcg gccattccat cggcgagctc	1980
gtcgtgcct gcgtcgcgg cgtcttctc cttgaggacg gcgtccgct cgtcgccgcc	2040
cgcggccgcc tcatgcagg cctcagtcc ggccggcgcca tggatccct cggcgccccg	2100
gaggcggaag tcgccgcgc cgtcgccccg cagccgcct ccgtttctat cgcgcgcgc	2160
aacggcccg agcaggtcgt catcgccggc gtcgagcagg ccgtccaggc catcgccga	2220
ggcttcgccg ccgcggcgcc ccgcaccaag cgcctccacg tctccacgc ctccactcc	2280
ccgtcatgg agccgatgct cgaggagtcc ggccgcgtcg ccgcctcgt cacctatgc	2340
cgcctcgtcgt tctcctcgt ctccaacctc tccgcaagg tcgtaccga cgagctctcc	2400
gccccgggct actgggtccg ccacgtccgc gagccgtcc gcttcgccga cggcgtcaag	2460
gccctccacg aggcggcgcc cggcaccttc gttgaggtcg gccgaagcc gacctcctc	2520
ggcctcctcc ccgcctgct cccggaggcc gagccgacct tctcgcctc cctccgcgc	2580
ggccgcgagg aggcgcggc cgtcctcag gccctcgcc gcctctgggc cgcggcgcc	2640
tccgtctcct ggccggcgct ctcccgacc gccggccgcc gcgtcccgct ccgacctac	2700
ccgtggcagc gccagcgcta ctggatcgag gccccggccg agggcctcgg cgccaccgcc	2760
gccgacgccc tcgccagtg gttctaccgc gtcgactggc cggagatgcc gcgtcctcc	2820
gtcgactccc gccgcggcgc ctcggcgccg tggctcgttc tcgccgaccg cggcgccgc	2880
ggtgaggccg ccgcgcggc cctctcctcc cagggtgct cctcgccgt cctccagcc	2940
ccggccgagg cctccgctg cgcggagcag gtacccagg ccctcgccg ccgcaacgac	3000
tggcaggcg tctctacct ctggggcctc gacccgtcg tcgaggcagg cgcctccgcc	3060

gaggaggtcg ccaaggtcac ccacctcgcc gccgccccgg tctctgcct catccaggcc	3120
ctcggcaccg gcccgcgtc cccgcgcctc tggatcgta cccgcggcgc ctgcaccgtc	3180
ggcggcgagc cggacgccgc cccgtgccag gccgcccctc ggggcatggg cgcgtcgcc	3240
gccctcgagc acccgggctc ctggggcggc ctggtcgacc tcgacccgga ggagtcctcg	3300
accgaggtcg aggcctctgt cgccgagctc ctctccccgg acgccgagga ccagctcgcc	3360
ttccgccagg gcccgcccg cgccgccgc ctctcgccg ccccgccgga gggcaacgcc	3420
gccccggtct cctctccgc cgagggtcc tacctcgta ccggcggcct cggcgccctc	3480
ggcctctcg tcgccgtg gctcgtcgag cgcggcgccg gccacctgt cctcatctc	3540
cgccacggcc tcccgaccg cgaggagtgg ggccgcgacc agccgccgga ggtccgcgc	3600
cgcctcgccg ccatcgaggc cctcgaggcc caggcgccc gcgtcacgt cgcgcgcgtc	3660
gacgtcgccg acgccgagg catggccgc ctctcgccg ccgtcgagcc gccgtccga	3720
ggcgtcgcc acgcccgcc cctctcgac gacggcctc tcgccacca ggacgccggc	3780
agactcgccc gggctctcg cccgaagtc gagggcgct gggctctcca caccctacc	3840
cgcgagcagc cgtcgacct ctctgctt ttctctccg cctccggcgt ctccggctc	3900
atcgccagg gctctacgc cgccggcaac gccttctcg acgcctcgc cgacctagg	3960
cgacccagg gcctcgccgc cctctccatc gcctggggcc tctgggccga gggcgcatg	4020
ggctcccagg cccagcgag ggagcacgag gcctccgca tctgggcat gccgacctc	4080
cgcgccctgg ccgcatgga gtggctctc ggaccccg ccacccagcg cgtcgatc	4140
cagatggact gggccacgc cggcgccgc ccgcgcgacg cctcccgcg ccgttctgg	4200
gaccgctcg tcaccgca caaggaggcc tctctctcg ccgtccggc cgtcgagcg	4260
tggcgcaac cctccgtgt cgagaccgc tccgccctc acgagctgt ccgcggcgtc	4320
gtcgccggcg tcatgggtt caccgaccg ggcaactct acgtccgccc cggttctgc	4380
gagcagggcc tcgactcct catggcgtc gagatccga aggcctcca gggcgagctc	4440
ggcatgccgc tctccgca cctcgcttc gaccaccga ccgtcgagcg cctcgtcgag	4500
tacctctct cccaggcct cgagctccg gaccgaccg acgtccgtc cgtccgctc	4560
ccggccaccg aagaccgat cgccatgct ggccgcccct gccgttccc gggcgggcgtc	4620
gaggacctc agtctactg gcagctctc accgaggcg tcgtcgtct caccgaggtc	4680
ccggccgacc gctggaacg cgccgacgg cggtcccg gctccggcg ggcacagcg	4740
cagacctacg tcccgcgcg cggttctc cgcgaggtc agaccttca cgccgcttc	4800

ttccacatct ccccgcgca ggccatgtcc ctcgaccgc agcagcgct cctcctcgag	4860
gtctcctggg aggccatcga gcgcgccggc caggaccgt ccgcctccg cgagtcgccg	4920
accggcgctt tcgtcggcgc cggcccgaac gagtacgcc agcgcgtcca ggagctcgcc	4980
gacgaggccg caggcctcta ctccggcacc ggcaacatgc tctccgtcgc cgccggccgc	5040
ctctcttct tctcggcct ccatggccc accctcgccg tcgataccgc ctgctcctcc	5100
tccctcgtcg cctccacct cggctgtcag tccctccgc gcggcgagt cgaccaggcc	5160
ctcgtcggcg gcgtcaacat gctcctctcc ccgaagacct tcgcctcct ctcgccatg	5220
cacgccctct cccggggcgg ccgtcgaag accttctcc cgcagccga cggtacgcc	5280
cgcccgagg gctcgccgt cgtcgtctc aagcgctct ccgagccca gcgcgaccgc	5340
gaccgatcc tcgcgtcat ccgcggcacc gccatcaacc acgacggccc gtctcggc	5400
ctcaccgtcc cgtcgggccc ggcccaggag gccctcctcc gccaggcct cgcccacgcc	5460
ggcgtcgtcc cgcccgacgt cgacttcgtc gagtccacg gcaccggcac cgccctcgcc	5520
gaccgatcg aggtccgcgc cctctccgac gtctacggc agggccgccc ggccgaccgc	5580
ccgtcatcc tcggcgccgc caaagccaac ctcgccaca tggagccgc cgccggcctc	5640
gcccgcctcc tcaaggcgt cctcgccctc ggccaggagc agatccggc ccagccggag	5700
ctcggcgagc tcaaccgct cctcccggtg gaggccctcc cggtcgctt cgcccgccc	5760
gccgtccgt ggcccgcac cgaccgccc cgcttcgccc gcgtctctc cttcggcatg	5820
tccggacca acgcccagc cgtctcgag gaggccccag ccgtcgagct ctggccggcc	5880
gccccggagc gatccgcca gctcctcgt ctctccgca agtccgagg cgccctcgac	5940
gcccaggccg ccgcctccg cgagcacctc gacatgcacc cggagctcgg cctcggcgac	6000
gtcgcttct cctcgccac cacacgtcc gccatgaacc accgcctgc cgtcgccgtc	6060
acctcccgcg agggcctcct cgccgcctc tccgccgtc cccaggcca gacccgccc	6120
ggcgccgccc gctgcatgc ctctcctcc cgcggaagc tcgcttctt cttaccggc	6180
caggcgccc agacccggg catgggccc ggctctgct ccgcctggc cgcttccgc	6240
gaggccttc accgtgct cgccctctc gatcgaggt tggaccgcc gctccgag	6300
gtcatgtggg ccgagccgg ctctgccgag tccctcctc tcgaccagac cgcttcacc	6360
cagccggccc tcttcacgt cgagtacgc ctaccgccc tctggcgctc ctggggcgtc	6420
gagccggagc tcgtcgccg ccattccgc ggcaactcg tcgcccctg cgtcgccggc	6480
gtcttctccc tcgaggacgg cgttcgctc gtcgccccc gcggccgact catgcaggc	6540
ctctccgccc gcggcgccat ggtctcctc ggcccccgg aggcgaggt cgcccgccc	6600
gtcgccccgc acgcccctc cgtctccat gccgccgta acggcccga gcaggtcgtc	6660

atcgccggcg tgcagcaggc cgtccaggcc atcgccggcg gcttcgccgc ccgcggcgcc	6720
cgcaccaaac gcctccacgt ctcccacgcc tcccactccc cgctcatgga gccgatgtc	6780
gaggagttag gccgcgtcgc cgctccgtc acctaccgcc gcccgtcgt ctccctcgtc	6840
tccaacctct ccggcaaggc cgtcgccgac gagctctccg ccccgggcta ctgggtccgc	6900
cacgtccgcg aggcggtccg cttcgccgac ggcgtaagg ccctccacga ggccggcgcc	6960
ggcaccttag tgcaggtcgg ccgaagccg accctcctcg gcctcctccc ggctgtcctc	7020
ccggaggccg aaccgacct cctcgctcc ctccgcgcc gccgcgagga ggccgccggc	7080
gtctcagagg ccctggcgcc cctctgggcc gccggcggt ccgtctctg gcctggcgtc	7140
ttcccgaccg ccggccggcg cgtcccgtc ccgacctacc cgtggcagcg tcagcgctac	7200
tggccggaca tcgagccgga ttcccgccgc caccgcccgc ccgacccgac ccagggtgg	7260
ttctaccgcg tcgactggcc ggagatccg cgcagtctcc agaagtccga ggaggcctcc	7320
cgcggctcct ggctcgtcct cgtgacaag ggccggcgtc gcgaggcgt cgcgcgcc	7380
ctctccacc gcggttgcc gtgcgtcgtc ctccacgcc cgcagagac ctccgccacc	7440
gccgagctcg tcaccgaggc cgcggcgcc cgtccgact ggaggtcgt cctctacctc	7500
tggggcctcg acgccgtcgt ccggcccgag gcctccatcg acgagatcgg cgacgccacc	7560
cgcgcgccca ccgccccgt cctcggcctc gcccgcttcc tctccacgt ctctgtctcc	7620
ccgcgcctct gggtcgtcac ccgcggcgcc tgcctcgtcg gcgacgagcc ggcaatcgcc	7680
ccctgccagg ccgcctctg gggcatgggc cgcgtcgccg ccctcgagca cccggcgca	7740
tggggcgcc tcgtcacct cgacccgcgc gcctccccgc cgcaggcctc cccgattgac	7800
ggcgagatgc tcgtaccga gctcctctcc caggagaccg aggaccagt cgccttcgc	7860
cacggccgcc gccacgcgc ccgcctcgtc gccccccgc cgcagggcca ggccgccccg	7920
gtctccctct ccgcgaggc ctctacctc gtcaccggcg gcctcgcggg cctcggcctc	7980
atcgtcgccc agtggctcgt cgagctcggc gcccgccacc tcgtcctcac ctcccgccgc	8040
ggcctcccgg accgccaggc ctggtgcgag cagcagccgc cggagatccg cgcgcgcatc	8100
gcccgctcg aggcctcga ggcccgggc gcccgctca ccgtcgccgc cgtcgacgtc	8160
gccgagctcg agccgatgac cgccctcgtc tctccgtgg agccgccgt ccgcggcgtc	8220
gtccacgccg ccggcgtctc cgtcatgcgc ccgtcgccg agaccgacga gacctctc	8280
gagtcgtcc tcccccga ggtcgccggc tctggctcc tccaccgct cctccacggc	8340
cgcgcgtcg acctctcgt cctcttctcc tccggcgccg ccgtctgggg ctccactcc	8400
cagggcgct acgcccgc caacgcctc ctgacggcc tcgccacct ccgcccgtcc	8460
cagtcctcc cgccctctc cgtcgctgg gccctctggg ccgaggcgcg catggccgac	8520

gccgaggccc acgcccgtt gtccgacatc ggctctctcc cgatgtccac ctccgcgcc	8580
ctctccgccc tccagcgctt cgtcgagacc ggctctgccc agcgaccgt caccgcgatg	8640
gactgggccc gtttcgcccc ggtctacacc gcccgcgccc gcccaacct cctctccgcc	8700
ctctgctcgc gcccgacat catcgccccg tccccgcgg ccgcccac ccgcaactgg	8760
cgcggcctct ccgtcgcca ggcccgcgtc gccctccacg agatcgtcca cggtgccgtc	8820
gcccgctcc tcggttctt cgaccgtcc gccctcgacc cgggcatggg cttcaacgag	8880
cagggcctcg actccttgat ggccgtcgag atccgaacc tcctccaggc cgagctcgat	8940
gtccgcctct ccaccacct cgccttcgac caccgaccg tccagcgctt cgtcgagcac	9000
ctctctgtcg acgtctcaa gctcgaggac cgctccgaca cccagcacgt ccgtccctc	9060
gcctccgacg agccgatcg catcgctggc gccgcctgcc gcttcccggg cggcgctcgag	9120
gacctcgagt cctactggca gctcctgcc gagggcgtcg tcgtatccgc cgaggtcccg	9180
gccgaccgtt gggacgccgc cgactggtac gaccggacc cggagatccc gggccgcacc	9240
tacgtacca agggcgctt cctccgcgac ctccagccc tcgacccac cttcttcgc	9300
atctccccgc gcgagggcat gtccctcgac ccgcagcagc gcctctctt cgaggtctcc	9360
tgggaggccc tcgagtcgc cggcatgcc ccggacacac tccgcgactc cccgaccggc	9420
gtcttctgtc gcgcccggcc gaacgagtac tacaccagc gcctccgtgg cttcaccgac	9480
ggcgccgccc gcctctacgg cggcaccggc aacatgctt ccgtaccgc cggacgcctc	9540
tccttcttcc tcggcctcca cggcccggc ctcccatgg acaccgctg ctctctctcc	9600
ctctgtcgac tccacctgc ctgccagagt ctccgctcg gcgagtgcga ccaggccctc	9660
gtcggcgcg tcaacgtct cctcgcccc gagaccttcg tctctcttc ccgcatgcgc	9720
gccctctccc cggacggcgc ctgcaagacc ttctccgccc acgccgacgg ctacggcccgc	9780
ggcgagggct gcgctgtct cgtctcaag cgcctccgc acgccagcg cgcggcgac	9840
tccatctctg cctcatccg cggctccgcc gtcaaccacg acggcccgtc ctccggtctc	9900
accgtcccga acggcccggc ccagcaggcc ctctccgcc aggcctctc ccaggccggc	9960
gtctccccgg tcgacgtcga cttcgtcgag tgccacggca ccggcaccgc cctcggcgac	10020
ccgatcgagg tccaggcct ctccgaggtc tacggcccgg gccgtccgg cgaccggccg	10080
ctctctctg gcgcccga ggccaacgt gccacctcg aggcgcctc cggcctcgcc	10140
tcctctctca aggcgtctt cgcctccgc cagagcaga tcccgccca gccggagctc	10200
ggcgagctca acccgacct cccgtggaac accctccgg tcgctgtccc gcgcaaggcc	10260

gtcccgtggg gccgcggcgc ccgcccgcgc cgcgccggcg tctccgcctt cggcctctcc	10320
ggcaccaacg tccacgtcgt cctcgaggag gccccggagg tcgagccggc cccggccgcc	10380
ccggcccgcc cggtcgagct cgtcgtcttc tccccaagt ccgcccgcgc cctcgacgcc	10440
gccgcccgcc gcctctccgc ccacctctcc gccaccgg agctctccct cggcgacgtc	10500
gccttctccc tcgccaccac ccgctccccg atggagcacc gcctcgccat cgccaccacc	10560
tcccgcgagg ccctccggcg cgcctcgac gcccccgcgc agcaaaagac cccgcagggc	10620
gccgtccgcg gcaaggccgt ctctccgcgc ggcaagctcg ccttctctt caccggccag	10680
ggcgcacaga tgccgggcat gggccgcggc ctctacgaga cctggccggc cttccgcgag	10740
gccttcgacc gctcgtcgc cctcttcgac cgcgaaatcg accagccgct ccgcgaggtc	10800
atgtgggccc caccgggctt cgcgcaggct gcccgccctg accagaccgc ctacgccag	10860
ccggccctct tcgcctcga gtacgccctc gcccgccctt ggcgctcttg gggcgctcag	10920
ccgcacgtcc tcctcgcca ttcaatcggc gagctcgtcg ccgctgtgt cgcggcgctc	10980
ttctccctcg aggacgccgt ccgctcgtc gcccccgcgc gccgctcat gcaggccctc	11040
ccggccggcg gcgccatggt cgccatgcc gcctccgagg ccgaggtcgc cgcctcgtc	11100
gccccgcacg ccgccaccgt ctccatgcc gccgtcaacg gcccggaacg cgtcgtcatc	11160
gccggcgccg aggtccagggt cttggccctc ggcgccacct tcgcgcgccg cggcatccgc	11220
accaagagac ttgcgcttc ccacgccttc cactccccgc tcatggacc gatgctcgag	11280
gacttcacg cgtcgcgc caccatgcc taccgcgcc cggaccgcc ggtcgtctcc	11340
aacgtaccg gccacgtcgc tggtcggag atcgccacc cggagtactg ggtccgccac	11400
gtccgctccg ccgtccgtt cggcgacggc gccaaggccc tcacgccgc cggcgccgcc	11460
acattcgtcg aggtcggccc gaagccggtc ctctcggcc tcctcccggc ctgcctcggc	11520
gaggccgacg ccgtcctcgt ccgctccctc cgcgccgacc gctccgagtg tgaagtcgtc	11580
ctcgcgcccc tcggcgctg gtacgcctgg ggccggcgccc tcgactggaa gggcgctctc	11640
cctgacggcg cccgccgct cgcctcccg atgtaccctt ggacgcgca gcgccactgg	11700
atggacctca cccgcgctc cgcgccccc gccggcatcg ccggccgctg gccgtcgc	11760
ggagtccgcc tctgcatgcc gggcgccgtc ctccaccacg tccttccat cggcccgcgc	11820
caccagccgt tcctcggcga tcacctgtc ttccggcaagg tcgtcgtccc gggcgccctc	11880
cacgtcgcg tcatctctc catcgccgc gagcgtggc cggagcgcgc catcgagctc	11940
accggcgctc agttcctcaa ggccatagcc atggagccgg accaggaggt cgagctccac	12000
gccgtcctca ccccgaggc cgcggcgac ggtacctct tcgagctcgc caccctgcc	12060
gccccggaga ccgagcgccg ctggaccacc cacgccgcgc gccgcgtcca gccaccgac	12120

ggcgccccgg gcgcctcccc gcgcctcgag gtctctgagg accgcgcat ccagccgctt	12180
gacttcgccg gcttctctga ccgcctctcc gccgtccgca tcggctgggg cccgctctgg	12240
cgctggctcc aggacggccg cgtcggcgac gaggcctccc tcgccacctt agtcccgacc	12300
taccggaacg cccacgacgt cgccccgctc caccgatcc tcctcgaaa cggtctcgcc	12360
gtctccctcc tctccaccg ctcgagccg gaggacgacg gcaccccgcc gctcccgttc	12420
gccgtcgagc gcgtccgctg gtggcgcgcc ccggtcggcc gcgtccgctg cggcggcgctc	12480
ccgcgctccc aggccttcgg cgtctctctc ttctctctcg tcgacgagac cggcggaggtc	12540
gtcggcgagg tcgagggctt cgtctgccgc cgcgctccgc gcgaggtctt cctccgccag	12600
gagtccggcg cctccaccgc cgcctcttat cgcctcgact ggccggaggc cccgctcccg	12660
gacccccgg ccgagcgcat cgaggagtcc tgggtcgtcg tcgccgcccc gggtctcgag	12720
atggccgccc ccttcgccac ccgcctcaac cgtgcgtcc tcgccgagcc gaagggcctc	12780
gaggccgccc tcgccggcgt ctcgccgccc ggctctcatc gcctctggga ggccggcgcc	12840
cacgaggagg ccccgccgc cgcacgagc gcgcaaccg agggcctctc cgtcgtccag	12900
gccctccgcg accgcgccgt ccgcctctgg tgggtcacca tggcgccgt cgcgctcgag	12960
gccggcgagc gcgtccaggt cgcaccgcc cctgtctggg gcctcggccg caccgtcatg	13020
caggagcgcc cggagctctc ctgcacctc gtgcacctc agccggaggc cgacgcccgc	13080
cgtccgccc acgtctctct ccgcgagctc ggccgcccgc acgacgaaac ccaggtagcc	13140
ttccgctccg gcaagcgccc cgtcgcccgc ctctcaagg ccaccactc ggagggcctc	13200
ctcgtcccg acgccgagtc ctaccgctc gagcccgccc agaaggcac cctcgaccag	13260
ctccgctcg ccccgcccca gcgcgcgccc cccggcccgg gcgaggtcga gatcaaggtc	13320
accgctccg gcctcaactt ccgaaccgct ctcgccgtcc tcggcatgta cccggcgac	13380
gccggcccga tggcgccgca ctgcgccggt gtgccaccg ccgtcggcca gggcgtccgc	13440
catgttgccg tcggcgagc cgtcatgact ctcggcacc tccaccgctt cgtcaccgtc	13500
gacccccgc tcgtcgtccg ccagccggcc ggcctcacc cggcccaggc cgccaccgtc	13560
ccggtcgcct tcctcaccgc ctggctcgcc ctccacgacc tcggaaacct ccgcccggc	13620
gagcgctcc tcatccacgc cgccgccggc ggctcggca tggcagccgt ccagatcgcc	13680
cgctggatcg gcgccgaggt cttcgccacc gcttccccgt ccaagtgggc agcgtccag	13740
gccatggcg tcccgcgac ccacatcgcc tcctcccga cctcagatt cgccgagacc	13800
ttccgccaag tcaccggcgg ccgcggcgct gacgtcgtcc tcaacgcct cgccggcgag	13860
ttcgtcgag cctcctctc cctcctctcc accggcgac gcttctcga gatgggcaag	13920
accgacatcc gcgaccgccc cgccgtcgcc gcccccacc cgggcgtccg ctaccgctc	13980

ttcgacatcc tcgagctcgc cccggaccgc acccgcgaga tcctcgagcg cgtggtcgag	14040
ggcttcgcgc cggccatct cgtgccctc cgggtccacg ccttcgcat caccaaggcc	14100
gaggccgcct tccgttcat ggcccaggcc cgccaccagg gcaaggtcgt cctcctcccg	14160
gccccgtcgc cggccccgct cgccccgacc ggcaccgtcc tcctcacggc cggcctcggc	14220
gcccctcgcc tccacgtcgc cgcctggctc gcccagcagg gcgtcccgca catggtcctc	14280
acaggccgcc gggcctcga cactccgggc gccccaagg cgtcgccga gatcgaggca	14340
ctcgccgccc gcgtcacat cgccgcctcc gacgtcgccg accgcaacgc cctcgaggcc	14400
gtcctccagg ccatcccgc cgaatggcgc ctccaggcgc tcattcacgc cgcaggcgcc	14460
ctcgacgacg gcgtcctcga cgagcagacc accgaccgt tctcccgct cctcgccccg	14520
aaggtcacgc ggcctggaa cctccacgag ctccaccgc gcaacgacct cgccttcttc	14580
gtactcttct cctccatgct cggcctctc ggctccgcgc gccagtcaa ctacgcgcc	14640
gcaaacacct tcctcgacgc cctcgccgc caccgccgcg ccgagggcct cgccgcacag	14700
tcctcgcct ggggccccgtg gtccgacggc ggcatggcgc cggcctctc tgcgcctc	14760
caggccccgc tcgcccgcga cggcatgggc gccctctccc cagcccaggc caccgcctc	14820
cttggccagg cctcgcccc cccggagacc cagctcggcg ccatgtcct cgacgtcgc	14880
gccgcctccc aggctccgc cgccgccgtc ccgccgtct ggcgccct cgtccgcgc	14940
gaggccccgc acaccgcgc cggcgcccag ggcgccctcg ccgccgcct cggcgccctc	15000
ccggaggccc gacgcgcga cgaggtcgc aaggtcgtcc aggcgcgat cgccgcgtc	15060
ctctcctggt ccgccctc cgccgtccg gtcgaccgc cgtctcga tctcgcccta	15120
gactccctca ccgcgtcga gtcgcgaac gtctcgcc agcgcgtcgg cgccaccctc	15180
ccggccacc tcgccttga ccaccgacc gtcgacgcc tcaccgctg gctcctgac	15240
aaggtcctcg ccgtgcgcga gccgtccgtc tcctccgca agtctcccc gcaggtcgcc	15300
ctcgacgagc cgatcgcat catcgcatc ggctgtcgt tcccgggcgg cgtcgccgac	15360
ccggagtct tctgggcct cctcgaggag ggctccgac cgtcgtcga ggtccgcac	15420
gagcgtggg acatcgatgc cttctacgac ccggaccgg acgtccgcgg caagatgacc	15480
accgcttcg gcgcttct ctcgacatc gaccgttcg acccgccctt ctteggcatc	15540
tccccgcgc aggccaccac catggaccg cagcagcgc tcctcctga gactcctgg	15600
gaggccttc agcgcgcgg cactctccg gagcgtctca tgggctcga caccggcgtc	15660
ttcgtcgac tcttctacca ggagtacgc gccctgcgc gcggcatcga ggccttcgac	15720

ggctacctcg gcaccggcac caccgcctcc gtcgcctccg gccgcattctc ctacgtcctc	15780
ggcctcaagg gcccgtcctt caccgtcgac accgcctgct cctcctccct cgtcgccgtc	15840
cacctcgctt gccaggccct cgcgcgcggc gagtgtctccg tcgccctcgc cggcggcgta	15900
gccctcatgc tcaccccggc caccttcgtc gaggttctccc gcctccgcgg cctcgccccg	15960
gacggccgct gcaagtcttt ctccgccgcc gccgacggcg tcggctgggtc cgagggtgtc	16020
gccatgtctc tcctcaagcc gtcgcgcgac gccagcgcg acggcgaccc gatcctcgcc	16080
gtcatccgcg gcaccgccgt caaccaggac ggccgctcca atggcctcac cgcgccgaac	16140
ggctcctccc agcaggaggt catccgccgc gccctcgagc aggcgggctt cgcgccggcc	16200
gagctctcct acgtcgagtg ccacggcacc ggcaccaccc tcggcgaccc gatcgaggtc	16260
caggccctcg gcgcgtctt cgcaccgggc cgcctgtccg accgccgctt cgtcataggc	16320
tctgtcaagt ccaacatcgg ccataccag gccgccgccc gcgtcgccgg cgtcatcaag	16380
gtagccctcg ccttcgagcg cggcctcatc ccgcgtccc tccacttga cgcgccgaac	16440
ccgcacattc cgtgggtccg gtcgcgcgtc caggtcgcgg ccaagccggt cgagtggacc	16500
cgaacggcg tccgcgcgg cgcggcggtc tctccttcg gcgtctccgg caccaacgcc	16560
cacgtcttcc tcgaggaggc cccggccgcc gccttcgccc cggccgcgc ccgtccgcc	16620
gagctcttcg tcctctccgc caagtccgcc gccgccctcg acgcacaggc ggcccgcctc	16680
tccgccatg tcgtcgccca cccggagctc ggcttcggcg acctcgctt ctccctcgcc	16740
accaccgct cccgatgac ctaccgctc gccgttgccg ccacctccg cgaggccctc	16800
tccgccccc tcgacaccgc cgcaccggg caggccccgc cggccgcggc ccgcggccac	16860
gcctccaccg gtcgcgcgc gaaggctgtc ttctgtcttc cgggccaggc ctcccagtgg	16920
ctcggcatgg gccagaagct cctctccgag gagccggtct tccgcgacgc cctctccgcc	16980
tgcgaccgcg ccatccaggc cgaggccggc tggctcttgc tcgccgagct cgcgcgcgac	17040
gagaccacat cccaactcgg ccgcatcgac gtcgtccagc cggccctctt cgccatcgag	17100
gtcgccctct ccgccctctg gcgtctctgg ggcttcgagc cggacgccgt cgtcgccac	17160
tccatggcg aggtcgccgc cgcacgctc gccggcgccc tctcctcga agacgcccgc	17220
gccatcatct gccgcgctc cctcctctt cgcgcgcatc ccggccaggc tgagatggcc	17280
gtcgtcgagc tctcctcgc cgaggctgaa gccgccctc tcggctacga ggaccgcctc	17340
tccgtagccg tctccaactc cccgcgtcc accgtcctcg cggcgagcc ggccgcctc	17400
gccgaggtcc tcgccatcct cgcgccaaag ggctcttct gccgccgct caaggtcgac	17460
gtcgccctcc actccccga gatcgaccg ctccgcgac agctcctcgc cgcctcggc	17520
gagctcgaac cgcgccaggc caccgtctcc atgcgtcca ccgtcacctc caccatcatg	17580

gccggcccg agctcgtcgc ctctactgg gccgacaacg tccgccagcc ggtccgcttc	17640
gccgaggccg tccagtcctt catggaggac ggccacggcc tcttcgtcga gatgtcccca	17700
cacccgatcc tcaccacctc cgtcgaggag atccgccgcg ccaccaagcg cgagggcgtc	17760
gccgtcgggt cctccgcg cgccaagac gagcgctct ccatgctcga ggccctcggc	17820
gccctctggg tccacggcca ggccgtcggc tgggagcgcc tcttctcgc cggcggcgcc	17880
ggcctccgcc gcgtcccgct ccgacctac ccgtggcagc gcgagcgcta ctgggtcgac	17940
gccccgaccg gcggcgccgc cgccggctcc cgttcgccc acgccggctc ccaccgctc	18000
ctggcgaaa tgcagacct ctccaccag cgtccaccc gcgtctggga gaccacctc	18060
gacctcaagc gcctcccggt gctcggcgac caccgctcc agggcgccgt cgtcttcccg	18120
ggcgccgctt acctcgagat ggccctctcc tccggcgccg aagccctcgg cgatggcccg	18180
ctccaggctt ccgacgtcgt cctcgccgag gccctcgct tcgccgacga cccccggcc	18240
gctgtccagg tcatggccac cgaggagcga ccgggccgcc tccagttcca cgtcgctcc	18300
agagtcggg gccacggcgg cgccgcttc cgtccacag ccgcggcggt cctccgccag	18360
atcgagcgcg ccgaggtccc ggcccgctc gacctcgag ccctccgcgc ccgcctccag	18420
gcctccgccc cgccgcgcgc cactacgcc gccctcgccg agatgggctt cgagtacggc	18480
ccggccttcc agggcctcgt cgagctctgg cgccgagagg gggaggcgct tggccgctg	18540
cgctcccg aagccgcgg ctcgccgcc gcctgccgcc tccaccggc cctcctcgac	18600
gcctgcttc acgtctctc cgcttcgcc gaccgcggcg aggccacccc gtgggtcccg	18660
gtcgagatcg gctccctccg ctggttcag cccccgtccg gcgagctctg gtgtcacgcc	18720
cgtccgtct cccacggcaa gccgacccg gaccgccgt ccaccgactt ctgggtcgtc	18780
gactccaccg gcgcatcgt cgccgagatc tccggcctcg tcgccagcg cctcgccggc	18840
ggcgtccgcc gcccgagga ggacgactgg ttcatggagc cggcctggga gccgaccgcc	18900
gtcccggtt ccgaggtcat ggccggcgcg tggctcctca tcggctcgg cgccggcctc	18960
ggcgccccc tccactcgc cctaccgag gccggccact ccgtagtcca cgcgaccggc	19020
cgcggcacca gcgcccgcg cctccaggcc ctctcaccg cctccttca cgccaggcc	19080
ccgacgtccg tcgtccacct cggtccctc gacgagcgcg gcgtcctcga cgccgacgcc	19140
ccgttcgatg ccgatgcct cgaggagtcc ctgctccgc gctcgactc cgtcctctgg	19200
accgtccagg ccgtcgccgg cgccggcttc cgcgaccgc cgccctctg gctcgtcacc	19260
cgcggcgccc aggccatcgg cgccggcgac gtctccgtcg ccaggcccc cctcctcggc	19320
ctcgcccgcg tcatcgctt cgagcacgcc gagtccgct gcgccgcat cgacctcgac	19380
ccggccccgc gcgacggcga ggtcgacgag ctctcgccg agtcctcgc cgacgacgcc	19440

gaggaggagg tgccttccg cggcggcgag cgccgcgtcg cccgcctcgt ccgcgcctc	19500
ccagagaccg actgccgca gaagatcgag ccggccgagg gcccccgtt ccgcctcgag	19560
atcgacggct cggcgctcct cgacgacctc gtctctcccg ccaccgagcg ccgcccgccg	19620
ggccccggcg aggtcgagat tgcgctcgag gccggcgcc tcaacttcct cgacgtcatg	19680
cgcgcatgg gcatctaccc tgggccccggc gacggccccg tcgccctcgg cgccgagtgc	19740
tccggccgca tcgtcgccat gggcgagggc gtcgagtcct tccgcatcgg ccaggacgtc	19800
gtcgccgtcg ccccgttctc cttcggcacc cacgtcacca tcgatcccg catgctcgcc	19860
ccgcgccccg ccgcactcac cgccgccag gccgcgccc tcccggtcgc cttcatgacc	19920
gcctggtacg gcctcgtcca cctcgccgc ctcgcgcgg gcgagcggt cctcatccac	19980
tcagccaccg gcggcacccg cctcgccgcc gtccagatcg ccgccacct cggcgccgag	20040
atcttcgcca ccgccggtac cccggagaag cgccctggc tccgcgagca gggcatcgcc	20100
cacgtcatgg actccgctc cctcgacttc gccgagcagg tcctcgccgc caccaagggc	20160
gagggcgctg acgtcgtcct caactccctc tccggcgccg ccacgacgc ctcctctcc	20220
acctcgtcc cggacggccg cttcatcgaa ctcggcaaga ccgacatcta cgccgatcgc	20280
tccttgggcc ttgcacactt ccgcaagtcc ctctctact ccgcgctga cctcgccggc	20340
ctcgccgtcc gccccccgga gcgcgtcgcc gccctcctcg ccgaggtcgt cgatctctc	20400
gcccgcgcg ccctccagcc gctcccgtc gagatcttc cgctctcccg cgccgccgac	20460
gccttcgca agatggccca ggcccagcac ctcggcaagc tcgtcctcgc cctcgaggac	20520
ccggacgtcc gcatccggt cccgggagag tccggcgctg ccacccgcgc cgacggcgcc	20580
tacctcgtca ccggcgccct cggcgccctc ggctctcag tcggcggtg gctcgccgag	20640
cagggcgccg gccacctcgt cctcgtcgcc cgctccggcg ccgtctccgc cgagcagcag	20700
accgcccgtg ccgcctcga ggcccacggc gccgcgtca ccgtcgcccg cgccgacgtc	20760
gccgaccgag cccagatgga gcgcatectc cgcgaggta ccgcctccgg catgccccctc	20820
cgcggcgctg tccacgccgc cggcatectc gacgacggac tcctcatgca gcagaccccg	20880
gcccgttcc gcgcgtcat ggccccgaag gtccggcgcg ccctccacct ccacgcctc	20940
accgcgaag ccccgctctc cttcttcgtc ctctacggt ccggcgccgg cctcctcggc	21000
tccccgggcc agggcaacta cgccgccgcc aacaccttc tcgacgcct cgcccaccac	21060
cgccgcgccc agggcctccc ggccctctcc atcgactggg gcctcttcgc cgacgtcggc	21120
ctcgccggcg gccagcagaa ccggggcgcc gcctcgtca cccgcggcac ccgtccctc	21180

accccgagacg agggcctctg ggccctcgag cgcctcctcg acggcgacgc caccaggcc	21240
ggcgtcatgc cgttcgacgt ccgccaatgg gtcgagttct acccgccgc cgcctcctcc	21300
cgcgcctct cccgcctcat gaccgcgcgc cgcgtcgct cggccgcct cgcggcgac	21360
cgcgacttgc tcgagcgct cgcaccgcc gagccggcg cccgagccgg catgctccag	21420
gaggtcgtcc gcgcccaggt ctcccaggtc ctccgcctct ccgagggcaa gctcgacgtc	21480
gacgccccgc tcacctcct cggcatggac tccctcatgg gcctcgagct ccgcaaccgc	21540
atcgaggcgc tctcggcat caccatgcc gccaccctcc tctggacctc cccgaccgtc	21600
gcccctct cgcgccct cgcctccac gtcgtctcca ccggcgacgc cgagtcgcc	21660
cgcggccgc acaccggctc cgtcgcccc accaccacg aggtcgctc cctcgacgag	21720
gacggcctct tcgcctcat cgacgagtc ctgccccgc ccgcaagcg ctga	21774
<210> 5	
<211> 11394	
<212> DNA	
<213> Artificial Sequence	
<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in	
Streptomyces venezuelae	
<400> 5	
atgaccgacc gcgagggta gtcctcgag cgtctccgc aggtcacct cgcctccgc	60
aagacctca acgagcgga caccctcgag ctcgagaaga ccgagccgat cgccatctg	120
ggcatcggt gcccttccc gggcgggcc ggcaaccgc aggccttctg ggagctctc	180
gacgacggc gcgacccat ccgcccgtc gaagagcgt gggcctctg aggcgtcgac	240
ccggcgacg acgtcccgc ctggcgccgc ctctcaccg aggcctatga tggcttcgac	300
gccgcttct tcggcatgc cccgcgcgag gccgctccc tcgaccgca gcaccgctc	360
ctctcgagg tcgctggga gggcttcga gacccgga tcccgcgcg ctccctcgtc	420
ggctcccga ccggcgtct cgtcggcgtc tgcgccacc agtacctca cgcgcgctc	480
gcccaccagc cgcgcgagga acgcgacgc tactccacca ccgcaacat gctttccatc	540
gccgcccgc gcctctcta caccctcggc ctccaggcc cgtgccttac cgtcgatacc	600
gcctgctct cctccctct cgcctccac ctgcctgcc gtccctccg cgcggcgaa	660
tccgacctc ccctcgccg cgcgctaac atgtctctt ccccgacac catgcgcgc	720
ctgcccga cccaggcct ctcccgaac ggccgtgcc agaccttga cgcctccgc	780
aacggcttc tccggcgga gggtgcggc ctcatcttc tcaagcct ctccgacgc	840

cgccgcgacg ggcaccgcat ctgggccctc atccgcggct ccgccatcaa ccaggatggc	900
cgctccaccg gcctcaccgc accgaacgtc ctgcccagg gcgcctcct ccgcgaggcc	960
ctccgcaacg ccggcgctga ggccgaggcc atcggtaca tcgagacga cggcgcccg	1020
acctccctcg ggcacccgat cgagatcgag gccctcccg ccgtcgtcgg cccggcccg	1080
ggcgacggcg ccgctgctg cctcggcgct gtcaagacca acctcgcca tctcgaggc	1140
ggcccgcgcg tcgccggcct catcaaggcc accctctccc tccaccacga gcgtatccg	1200
cgcaacctca acttcgcac cctcaaccg cgcatccga tcgaggcac cgcctcgcc	1260
ctcgccaccg agccggctcc gtggccgcgc accggccga ccgcttcgc cggcgctcc	1320
tcttcggca tgtccggcac caacgccac gtcgtctcg aggaagccc ggccgtcga	1380
ccggaggcgg ccgcaccgga gcgcgccgc gagctcttcg tctctccgc caagtccgc	1440
ggccctcccg acgtcaggc cgcgcctc cgcgaccacc tcgagaagca cgtcgagtc	1500
ggcctcggcg acgtcgtct ctcctcgcc accaccgct ccgccatgga gcaccgcctc	1560
ggcgtcgccg cctctcccg cgaggccctc cgcggcgccc tctccgcgc cgcggaggc	1620
cacacccgc cggcgccgt ccggggcggc gcctcggcg gctccgccc gaaggtcgtc	1680
tctgtcttc cgggccaggg ctccagtg gtcggaatgg gccgcaagct catggccgag	1740
gaaccggtct tccgcgcgc cctcgaaggc tgcgaccgc ccatcgaggc agaggccgt	1800
tggtcctcc tcgcgagct ctccgccgac gaggccgct ccagctcgg ccgcatcgac	1860
gtcgtccagc cggctctct cgccatggag gtcgcccct ccgcctctg gcgtcctgg	1920
ggcgtcgagc cggaggccgt cgtcgccac tccatggcg aggtagccgc cgcacgctc	1980
ggcgccgcgc tctcctcga ggacgcgtc gccatcatc gccgcgctc ccgctctc	2040
cgcgcgcatc ccggccaggg cgagatggc ctgctcgagc tctcctcga ggaggccgag	2100
ggccctca gaggcacga gggccgctc tcggtcgcg tctccaactc cccgcgtcc	2160
actgtctcg ccggcgagcc ggccgcctc tcagaggctc tcgcccct caccgcaag	2220
ggcgtctct ggccaggt caaggtcgac gtcgcatccc actcccgca ggtcgaccg	2280
ctccgcgagg agctcatgc cgcctcggc gccatccgc cgcgcgcgc cgcgtccg	2340
atgcgtcca ctgtaccgg cggcgtcgc gccggcccg agtcggcg ctcctactgg	2400
ggcgacaacc tcccgagcc ggtccgttc gccgcccg cccaggcct cctcgaggc	2460
ggcccgccc tctcatcga aatgtccc caccgatcc tcgtccgccc gctcgacgag	2520
attcagaccg ccggcgagca gggaggcgcc gcagtcggct ccctccgccc cggccaggac	2580

gaacgcgccca cctctctcga ggccctcggc accctctggg cctccggcta cccggtctcc	2640
tgggccccgcc tcttccccgc cggcgggtcgc cgcgtcccg ccccgacgta cccgtggcag	2700
cacgagcgct gctggatcga ggtcgagccg gacgcccgcc gcctcgccgc cgcgacccg	2760
accaaggact ggttctaccg caccgactgg ccggagggtcc cgcgcgccgc ccctaagtcc	2820
gagaccgccc atggctctctg gctcctctc gccgaccgcg cggcggtcgg agaggccgtc	2880
gccgcccgc tctccaccg cggcctctcc tgcaccgtcc tccacgcctc cgtgacgcc	2940
tccaccgtcg ccgagcaggt ctccgaggcc gcctcccgcc gcaacgactg gcaggcgctc	3000
ctctacctct ggggcctcga cgcgtcgtc gaccccgcg cctccgccga cgaggctctcc	3060
gaggccaccc gcccgccac cgcgccgtc ctcggcctcg tacgtttct ctcgccgcc	3120
ccgacccgc cgcgtttctg ggtcgtcacc cgcggcgct gcaccgtcgg cggcgagccg	3180
gaggcctccc tctgccaggc cgcctcttgg ggcctcgccc gagtccgcgc tctcgagcac	3240
ccggccgcct gggcgggcct cgtcgacctc gacctcaga agtccccgac cgagatcgaa	3300
ccgtcgtcg ccgagctcct cagccctgac gccgaggacc agtcgcctt ccgtccggc	3360
cgcgccacg ccgccgcct cgtcgccgcc ccgccggagg gcgacgtcg cccgatctcc	3420
ctctccgccg agggctccta cctcgtcacc ggccggcctcg gcggcctcgg cctcctcgtc	3480
gcccgtggc tgcgcgagcg cggcgcccgc cactcgtcc tcacctccg ccacggactc	3540
ccggagcgcc aggcctcgg cggcgagcag ccgccggagg cccgcgccg catcgccgcc	3600
gtcgagggcc tcgaggccca gggcgcccgc gtcaccgtcg ccgccgtcga cgtcgcgag	3660
gccgatccga tgaccgcct cctcgccgcc atcgagccgc cgtccgagg cgtcgtccac	3720
gcccgggcg tcttccccgt ccgccacctc gccgagaccg acgaggccct cctcgaatct	3780
gtctccgac cgaaggtcgc cggcagttgg ctctccacc gtctcctccg cgaccgaccg	3840
ctcgacctct tgcctcttt ctctccggc gccccgtct gggcgggcaa gggccagggc	3900
gcctacgcc ccgccaacgc ctctctgac ggcctcgccc accaccgcc cgcacctcg	3960
ctcccgccc tctcctcgc ctggggcctc tgggccgagg gcggcatggt cgacccaag	4020
gcccacgcc gactctcga catcggtc ctcctgatgg ccaccggccc ggccctctcc	4080
gcctcagag gcctcgtcaa cactccgcc gtccagcgt ccgtcaccg catggactgg	4140
gcccgttctg ccccggtcta cgcgccgc ggcgccgca acctcctct cgcctcgtc	4200
gctgaggacg agcgcgccgc ctccccgcc gtccgaccg ccaaccgat ctggcgcggc	4260
ctcagcgtcg ccgagtcgg ctcgccttg tacgagctcg tccgcggcgt cgcgccgtc	4320
ctcggttct ccgaccggg cgcactgac gtccggcag gcttcgccga gcagggcctc	4380
gactccctca tggccctcga gatccgaac cgcctccagc gcgagctcgg cgagcgcctc	4440

tccgccaccc tcgccttcga ccacccgacc gtcgagcgcc tcgtcgccca cctgctcacc	4500
gacgtcttga agctcgagga ccgctccgac acccgccaca tccgctccgt cgccgccgac	4560
gacgacatcg ccacgtcggt cgccgcatgc cgcttcccgg gcggagacga aggcctcgag	4620
acctactggc gccacctcgc cgagggcatg gtcgtctcca ccgaggtccc ggccgaccgc	4680
tggcgcgccg ccgactggta cgacccggac ccggagggtcc cgggccgcac ctacgtcgcc	4740
aagggcgctt tcctccgca cgtccgctcc ctgacgccc cttcttctgc catctccccg	4800
cgcgaggcca tgtccctcga cccgcagcag cgcttctctc tcgaagtctc ctgggaggcc	4860
atcgagcgcg ccggccagga ccgatggcc cttcgcgagt ccgccaccgg cgtcttcgtc	4920
ggcatgatcg gctccgagca cgccgaacgc gtcaggggcc tcgacgacga cgccgccctc	4980
ctctacggca ccacggcaa cctctctca gtcgcccgg gccgcctctc cttcttctc	5040
ggcctccacg gcccgacat gaccgtcgat accgcctgtt cctctcct cgtcgccctc	5100
catctcgctt gtcagtcctt ccgctcggga gagtgcgacc aggcctcgc cgccggctcc	5160
tccgtctctc tctccccg ctccttcgtc gccgcatccc gcatgcgctt cctctccccg	5220
gacggccgct gcaagacctt ctccgccgc gccgacggat tcgcccgcgc cgagggtgc	5280
gccgtcgtcg tctcaagcg tctccgac gccacgccc acccgaccc gatcctcgcc	5340
gtcgtccgct ccacgccat caaccacgac ggcccgtctt ccggcctcac cgtcccgctc	5400
ggcccggccc agcaggccct cctccgccag gccctcgccc aggcggcggt cgcccggcc	5460
gaggtcgact tcgtcgagt ccacggcacc ggacccgcc tcggcgaccc gatcgaggtc	5520
caggccctcg gcgccgtcta cggccgggc cgcccggccg agcggccgt ctggctcggt	5580
gctgtaaagg ccaacctcgg ccacctcgag gccgccgccc gcctcgccgg cgtctcaag	5640
gtctctctcg cctcgagca cgagcagatc ccggctcaac cggagctcga cgagctcaac	5700
ccgcacatcc cgtgggcca gctcccgtc gccgtcgtc gccgcgcgt gccgtggccg	5760
cgccggcgccc gcccgccg cgccggcgtc tccgccttcg gcctcagtgg caccaacgcc	5820
cacgtcgtcc tcgaggaggc cccggccgtc gagccggtcg ccgccgtcc ggagcgcgcc	5880
gccgagctct tcgtctctc cgccaagtcc gccgccgcc tcgacgccc ggccggccgc	5940
ctccgcgacc acctcgagaa gcacgtcgag ctgggctcg gcgacgtcgc cttctccctc	6000
gccaccaccc gctccgcat ggagcaccga ctgccgtcg ccgctctctc ccgcgaggcc	6060
ttgcgcggcg cctctctcgc cgccgccag ggccacaccc cgccggcggc cgtccgccc	6120
cgccctccg gcggctcgc cccgaaggtc gtcttcgtct tcccgggcca gggctcccag	6180
tgggtcgga tgggcccga gctcatggcc gaggagccgg tcttccgccc cgccctcgag	6240
ggctgcgacc gcgccatcga ggccgaagcg ggctgggtccc tcctcgccga gctctccgc	6300

gacgaggccg cctcccagct cggacgcac gacgtcgtcc agccggctcct cttcgccatg	6360
gaagtcgccc tctccgcct ctggcgctcc tggggcgctc aaccagaggc cgtcgtcggc	6420
cactccatgg gcgaggtcgc cgcgcgccac gtcgccggcg cctctcctt cgaggacgcc	6480
gtcgccatca tctgcccgcg ctcccgctc ctcgccgca tctccggaca gggagagatg	6540
gccctcgtcg agctctcct cgaggaggcc gagccgctc tcccgggcca cgagggccgc	6600
ctctccgtcg ccgtctccaa ctccccgcgc tccaccgtcc tcgccggcga gccggccgcc	6660
ctctccgagg tctcgcgc cctaccgcc aaggcgctt tctggcgcca ggtcaaggtc	6720
gacgtcgct cccactcccc gcaggtcgac ccgtcccgcg aggagctcat cgcgcctt	6780
ggcgccatcc gcccgcgcg cgcgcgcgtc ccgatgcgt ccaccgtcac cggcggcgtc	6840
atcgctggcc cggagctcgg cgcctctac tggggcgaca acctccgcca gccggtcgc	6900
ttcgcccgcg ccgcccaggc cctcctcgtc ggcggcccg cctcttcat cgagatgtcc	6960
ccgcaccga tctcgtccc gccgctcgac gagatccaga cggccgcga gcaggcgga	7020
gcccgctcg gctcctcgc gcgcggccag gacgaacgc ccacctctt cgaagccctc	7080
ggcaccctgt gggcctcgg ctaccggtc tctgggccc gcctcttccc ggccggcggc	7140
cgcgcgtcc cgctccccac ctaccgtgg cagcacgagc gctactggat cgaggactcc	7200
gtccacggct ctaagccgtc cctccgctc cgcagctcc gcaacggcg caccgaccac	7260
ccgtcctcg gcgccccgt cctcgtctcc gcccgcccgg gcgcccact ctgggaacag	7320
gccctcagcg acgagctct ctctacctc tccgagcacc gcgtccacgg cgaggccgtc	7380
ctcccgctcg ccgctacgt cgagatggc ctcgcccgcg gcgtcgacct ctacggcacc	7440
gccacctcg tgcctagca gctcgccctc gagcgcgct tggccgtccc gtccgagggc	7500
ggtcgcatcg tccaggtcgc cctctccgag gagggcccg gcccgccctc cttccaggtc	7560
tctccccgcg aagaggcagg ccgtcctgg gttcgccacg ccaccggcca cgtctgctcc	7620
ggccagtcct ccgcgctcgg cgcctcaag gaggccccgt gggagatcca gcgcgctgc	7680
ccgtccgtcc tctctccga ggccctctac ccgtcctca acgagcacgc cctcgactac	7740
ggcccgctgt tccaggcgct cgaacaggtc tggctcgga ccggcgaggt cctcgccgc	7800
gtccggctcc caggcgacat ggctctctc tccggcgct accgatcca cccggccctc	7860
ctcgagcct gcttcagggt cctaccgcc ctctcacca cccccgaatc catcgagatc	7920
cgcgcgcccc tcaaccacct ccacgagcgg gacctccgc gctcccgcg cccggtcaac	7980
caggccgtct ccgacacctg gctctgggac gccgccttg acggcgggcg ccgccaatcc	8040

gcctccgtcc cggctgacct cgtcctcggc tccttccacg ccaagtggga gggttatggag	8100
cgctcgtccc aggcctacat catcggcacc ctccgcatct ggaacgtctt ctgcgccgcc	8160
ggcgagcgcc acaccatcga cgagctcttc gtcgggtccc agatctccgt cgtctaccgc	8220
aaggtcatca agcgtggat ggagcacctc gtcgccattg gcatcctcgt cggcgacggc	8280
gagcacttcg tctcctccca gccgtcccg gagccggacc tcgccgccgt cctcgaggag	8340
gccggccgcg tcttcgccga cctcccggtc ctcttcgagt ggtgcaagtt cgcggcgag	8400
cgctcgcgcg acgtcctcac cggcaagacc ctgccctcg agatcctctt cccgggcggc	8460
tccttcgata tggccgagcg catctaccgc gattccccga tcgcccgcta ctccaacggc	8520
atcgtacgcg gcgtcgtcga gtcgcagacc cgcgtcgtcg ccccgagcgg catgttctcc	8580
atcctcgaga tcggcgccgg caccggcgcc accaccgccg ccgtcctccc ggctcctctc	8640
ccggaccgca ccgagtacca cttaccgac gtctccccgc tcttcctcgc ccgcgcagag	8700
cagcgcttcc gcgactacc gttctcaaa tacggcatcc tcgacgtcga ccaggaaccg	8760
gccggacagg gctacgcaca ccagcgttc gacgtcatcg tagccgcaa cgtcatccac	8820
gccacccgcg acatccgcgc caccgccaag cgctgctct cctcctcgc cccgggcggc	8880
ctcctcgtcc tcgtcgaggg caccggccac ccgatctggt tcgacatcac caccggcctc	8940
atcgagggct ggcagaagta cgaggacgac cttcgcatcg accaccgct cctcccgcc	9000
cgcacctggt gcgacgtcct ccgcccgtc ggcttcgccg acgccgtctc cctcccggc	9060
gacggctccc cggccggcat cctcgccag cagtcaccc tctcccgcg cccgggcac	9120
gccggcgccg cctgcgactc ctccggcgag tccgccaccg agagcccggc cgcgcgcgc	9180
gtccgacagg agtgggcca cggtccgcc gacgtcgtgc accgcatggc cctcgagcgc	9240
atgtacttcc atcgccgtcc gggccgccag gtctgggtcc acggccgct ccgcaccggt	9300
ggcggcgccct tcaccaaggc cctcgccggc gacctcctcc tcttcgagga caccggacag	9360
gtcgtcgccg aggtccaggg cctccgcctc ccgagctcg aggcctcgc cttcgccccg	9420
cgcgaccgcg gcgaggagt gctctacga ctcgagtggc agcgcaagga cccgatcccg	9480
gaggccccgg ccgccctc gtcctcctcc gccggcgcc ggctcgtcct catggaccag	9540
ggcggcaccc gcgccccct cgtatccctc ctcgagggcc gcggcgaggc ctgcgtccgc	9600
gtcatcgccg gcaccgcta cgctgcctc gccccgggcc tctaccaggt cgaccggcc	9660
cagccggacg gcttccacac cctcctcaga gacgccttcg gcgaagaccg catctgccgc	9720
gccgtcgtcc acatgtggtc cctcgacgcc accccgccg gcgagcgcg caccgccgag	9780
tcctccagg ccgaccagct cctcggtccc ctctccgcc tctcctcgt ccaggccctc	9840
gtccgccgcc gctggcgcaa catccgcgc ctctggctcc tcaccgcgc cgtccacgcc	9900

gtcggcgccg aagacgccgc cgcctccgtc gccaggccc cggctcgggg cctcgccgc	9960
acctcgccc tcgagcacc ggagctccgc tgcacactcg tcgacgtcaa cccggccccg	10020
tccccggagg acgccgccgc cctcgccgtc gagctcggcg cctccgaccg cgaagatcag	10080
gtcgccctcc gctccgacgg ccgtacgtc gccgcctcg tccgtctctc cttctccggc	10140
aagccggcta cagactgcgg catccgcgcc gacggctcct acgtcatcac cgacggcatg	10200
ggcgtgtcg gcctctcgt cgcctcagg atggtcatgc agggcgcccc ccacgtctc	10260
ctcgtcgacc gcggcgccgc ctccgagcc tcccgcgacg ccctccgag tatggccgag	10320
gccggagccg aagtcagat agtcgaagcc gacgtcgccc gcccgacga cgtggccccg	10380
ctctctcca agatcgaacc gtccatgccg ccgtccgcg gcatcgtcta cgtcgacggc	10440
acctccagg gcgactctc catgctcgag ctgacgccc gccgcttcaa ggagtggatg	10500
tatccgaagg tctcggcgct ctggaacctc cagccctca cccgcgaccg ttcctcgac	10560
ttctctgtec tctactctc cggcacctcc ctctcggcc tccggggcca gggctccgc	10620
gcccgccgag acgttttct cgacgccatc gccaccacc gctcaaggt cggcctcacc	10680
gccatgtcca tcaactgggg cctctctcc gaggcctcct ccccgccac cccgaacgac	10740
ggcgccgccc gcctcgagta ccgcggcag gagggcctca ccctcgagca gggcgccgcc	10800
gccctcggac gccctctcgc ccgcccgcc gcacaggctg gcgtcatgag cctcaatttg	10860
cgtcagtgga tcgagttcta cccgaacgt gcccgccctg ccctctgggc cgagctctc	10920
aaggaaacgc accgcgccga ccgcgccgcc tccaacgct ccaacctcg cgaggccctc	10980
cagtcgccc gcccgaggga ccgccaactc atctcgaga agcacctctc cgagctctc	11040
ggcgccggcc tccgctccc gccggagcgc atcgagcgtc acgtccggt ctccaacctc	11100
ggcatggact cctcatcgg cctcgagctc cgcaaccgca tcgaggccgc tctcggcac	11160
accgtccgg ccacctctc ttggacctac cctaactcg ccgcctctc cggtctctc	11220
ctcgacatcc tcttccgaa cgcggcgcc acccacgccc cggccaccga gcgcgagaag	11280
tccttcgaaa acgacgccgc cgacctgaa gccctccgc gcatgaccga cgagcagaag	11340
gacgccctcc tcgccgagaa gctcgccag ctgcccaga tcgtcgggga gtga	11394

<210> 6

<211> 7320

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in
Streptomyces venezuelae

<400> 6

atggccacca ccaacgccgg caagctcgag cacgccctcc tctcatgga caagctcgcc	60
aagaagaacg cgtccctcga gcaggaaagg accgagccga tcgcaatcgt cggcatcggc	120
tgccgcttcc cgggaggcgc cgacaccccg gaggccttct gggagctcct cgactccggc	180
cgcgacgccg tccagccgct cgaccgccgc tgggccctcg tcggcgcca cccgtccgag	240
gaggtaccgc gctgggccgg cctcctcacc gaggccgtgg acggattcga cgccgccttc	300
ttcggcacct ccccgccga gggccgtcc ctcgaccgc agcagcgct cctcctcgag	360
gtcacctggg agggcctcga ggacgccgc atcgccccgc agagtctcga cggatccgc	420
accggcgctt tctcggcgc ctgctcctcc gactactccc acaccgtcgc ccagcagcgc	480
cgcgaggagc aggcgccta cgacatcacc ggcaacaccc tctccgtcgc cgccggccgt	540
ctctcctaca cctcggctt ccaggggccg tgcctcaccg tcgacaccgc ctgctcctcc	600
tccctcgtcg ccatccact cgctgccgg tccctccgcg cccgcgagtc cgacctgcc	660
ctcgcgggag gctcaacat gctcctctcc tccaagacca tgatcatgct cggcaggatc	720
caggccctct ccccggaagg ccactgccga accttcgacg cctccgcaa cggttcgtc	780
cgcggcgagg gctcggcat ggtcgtctcg aagcgctct cgcagccca gcgccacggc	840
gaccgcatct gggccctcat ccgcggctcc gccatgaacc aggcaggccg ctccaccggc	900
ctcatggccc cgaacgtcct cgcccaggag gccctcctcc gcgaggccct ccagtccgcc	960
cgcgtcgacg ccggcgccat cggctacgtc gagaccacg gcactggcac ttcgctcggc	1020
gacccgatcg aggtcgaggc cctccgcgcc gtctcggcc cgcccgcgc cgatggctcc	1080
cgtcgtctcc tcggcgccgt gaagaccaac ctgggccacc tcgagggcgc cgccggcgtc	1140
gcgggcctca tcaaggccgc cctcgccctc caccacgaac tcatcccgcg caacctccac	1200
ttccacaccc tcaaccgcg catccgcatc gagggcaccg cctcgcct cgccaccgag	1260
ccggtcccggt ggccgcgcgc cggccgcccg cgcttcgccc gctctcgcg attcggcctc	1320
tccggcacca acgtccacgt cgtcctcgag gaggcccccg ccaccgtcct cgccccggcc	1380
acccggggcc gctccgccga gctcctcgtc ctctccgcca agtcgccgc cgccctcgac	1440
gcccaggccg cagcctctc cgacacatc gccgcctacc cggagcaggg cctcggcgac	1500
gtcgccttct ccttggcttc caccgctcc ccgatggaac atcgccctgc cgtcgccgcc	1560
accagccgcg aggcctccg ctccgccctc gaggtcgccg cccagggccca gaccccgcc	1620
ggtgccgccc gcggccgcgc cgctcctcc ccgggcaagc tcgccttctt cttcgccgga	1680
cagggggccc aagtccggg catgggcgc ggctctggg aggcctggcc ggccttcgc	1740
gaaaccttcg accgtgcgt caccctcttc gaccgcgagc tccaccagcc gctctcgag	1800

gtcatgtggg ccgaaccggg ctctcccg ctcctccctc tcgaccagac cgccttcacc	1860
cagccggccc tcttcgcct cgagtacgcc ctgccgccc tcttcgctc ctggggcgtc	1920
gagccggagc tcgtcgccgg ccactccctt ggagaactcg tcgccgctg cgtcgccggc	1980
gttttctccc tcgaggatgc cgtccgctc gtctcgccc gcggccgct catgcaggcc	2040
ctcccgccg gcggcgccat ggtctccata gcggcgccgg aagctgacgt agccggcgcc	2100
gtcggccgc acgcccctt ggtctccatc gccccgctca acggcccgga gcaggtcgtc	2160
atcgccggcg ccgagaagt cgtccagcag atcgccgccc ctttcgccc cgcggcgccc	2220
cgcaccaagc cgtccacgt gtccacgcc ttccactccc cgtcatgga cccgatgtc	2280
gaggccttca gacggtcac cgaatccgta acctaccgcc gaccgtccat cgcgtcgtc	2340
tccaacctct ccgcaagcc gtgcaccgat gaggtctccg cccccggcta ctgggtccgc	2400
cacggcccg aggcgctccg cttcgccgac ggctcaagg ccctccacgc cgcggcgccc	2460
ggcctcttcg tcgaggtcgg ccgaagccg acctcctcg gcctcgtccc ggctgcctc	2520
ccggacgccc gcccggtcct cctcccgccc tcccgcccg gccgcgacga ggccgctcc	2580
gccctcgaag ccctcgccgg attctgggtc gtccggcgct cggtcacctg gagcggagtc	2640
tccccctccg gcggccgccc cgtcccgtc ccgacctacc cgtggcaacg cgagcgctac	2700
tggatcgagg ccccggtcga ccgcgaggcc gacggcaccc gccgcgccc cgccggcggc	2760
caccgctcc tcggcgaggt cttctccgtc tccaccacg ccggcctccg cctctgggag	2820
acgacctcg accgaagcg cctccgtgg ctccggcagc accgcgcca gggtaggtt	2880
gttttcccg gcgccgcta cctcgagatg gccctcagct ccggcgccga gatcctcggc	2940
gacggccga tccaggtcac cgacgtcgtc ctcatcgaga ccctcacctt cgccggcgac	3000
acagccgtcc cggtcagggt cgtcaccacc gaggagcgcc cgggcccact ccgttccag	3060
gtcgctccc gcgagccggg cgagcgccgc gcccgttcc gcatccacgc ccgcgcgctc	3120
ctccgccga tcggccgctg tgagaccg ccgctcca acctcgccgc cctccgccc	3180
cgcctccacg ccgccgtccc ggccgccccc atctacggcg ccctcgccga gatgggcctc	3240
cagtaggcc cggcccttcg cgccctgcc gagctctggc gcggcgaggg cgaggccctc	3300
ggccgctcc gtctcccga ggcagccggc tccgccaccg cctaccagct ccacccggtc	3360
ctctcgacg cctcggtcca gatgatcgtc ggcccttcg ccgaccgca cgaggccacc	3420
ccttgggccc cggtagggt cggtccgtc cgcctcttcc agcgagccc gggcgagctc	3480
tggtgccacg cccgctcgt ctcgacggc cagcaggcct cctccaggtg gtccgcccac	3540

ttcgagctca tggacggaac cggcgccgtc gtcgccgaga tctccgcct cgtcgtcgag	3600
cgcttcgcct caggagtcg cgcgcgcgac gccgacgact ggttcctcga actcgactgg	3660
gagccggcgg cctcggcgg cccgaagatc accgccggtc gctggctcct cctcggcgag	3720
ggcggcgcc tcggcgctc cctctgtcc gccctcaagg ccgccggcca tgtcgtcgtc	3780
cacgccgccc gcgacgacac ctccaccgcc ggcatgcgcg ccctcctcgc caacgccttc	3840
gacggacagg ccccgaccgc agtcgtccac ctctcctccc tcgacggcgg cggacagctc	3900
ggccccggcc tcggcgccca gggcgccctc gacccccga ggtccccga cgtcgacgcc	3960
gacgtctcgc agtcgcctt catgaggggc tgcgactccg tctctctct cgtccaggcg	4020
ctcgtcggca tggacctcgc caacgccccg cgctctggc tctcaccgc cggcgcccag	4080
gccgccgccc cggcgacgt ctccgtcgtc caggccccgc tctcggcct cggccgcacg	4140
atcgccctcg agcacgccga actcaggtgt atctccgtcg acctcgacc gcccgagccg	4200
gagggcgaag ccgacgcgt cctcgccgag ctctcgcgcg acgacgtga agaggaggtc	4260
gccctcgcgc ggcgcgacc cctcgtgcc cgctcgtcc accgcctccc ggacgccag	4320
cgccgcgaga aggtcgagcc ggccggcgac cgcctgtcc gcctcgagat cgacgagccg	4380
ggcgccctcg accagctcgt cctccgcgcc acggggccgc gagccccggg cccgggcgag	4440
gtcgagatct ccgtcgaggc cgccggcctc gactccatcg acatccagct cgcctcggc	4500
gtgccccga acgacctccc gggcgaggag atcgagccgc tcgtcttggg ctccgagtgc	4560
gccggccgca tcgtcgcgt cggcgagggc gtaacggcc tcgtcgtcgg ccagcctgtc	4620
atcgccctcg ccgccggcgt cttcgtacc cagtcacca cctccgccac cctcgtctc	4680
ccgcgccgc tcggcctctc cgccaggag gccgccgcca tgccgtcgc ctacctacc	4740
gcctggtacg cctcgacaa ggtcgccac ctccaggccg gcgagcgcgt cctcatcac	4800
gctgaggccg gcggcgtcgg cctctgcgcc gtccgtggg cccagcgcgt cggcgccgag	4860
gtctaccca ccgccgacac cccgaaaac cgcgcctacc tcgagtcct cggcgtccgc	4920
tacgtctccg actccgctc cgggcgttc gtcaccgacg tccacgcctg gaccgacggc	4980
gagggcgtcg acgtcgtcct cgactcctc tccggtgagc gcatcgacaa gtcctcatg	5040
gtcctccgcg cctcggcgc cctcgtcaag ctccggccgc gcgacgactg cgccgacacc	5100
cagccgggac tcccgcgct cctccgcaac ttctctttt cccaggtcga cctccgcggc	5160
atgatgctcg accagccggc ccgcatccgc gccctcctcg acgagctctt cggcctcgtc	5220
gccgccggcg ccatctccc gctcggctcc ggctccgcg tcgggggctc cctcaccccg	5280

ccgccggtcg agaccttccc gatttcccgc gccgcagagg ccttcgcgcg catggcccag	5340
ggccagcacc tcggcaagct cgtcctcacc ctgcacgacc cggagggtccg catccgcgcc	5400
ccggccgagt cctcgggtcgc cgtccgcgcc gacggcacgt acctcgtcac cggcgggctc	5460
ggcggcctcg gcttccgggt cgccgggtgg ctgcccgagc gcggcgccgg ccagctcgtc	5520
ctcgtgggcc gctccggcgc cgctccgcc gagcagcgcg ccgccgtcgc cgccctcgag	5580
gcccacggcg cccgcgtcac cgtcgccaag gccgacgtcg ccgaccgtc ccagatcgag	5640
cgcgtcctcc gcgaggtcac cgctccggc atgccgtcc gcggcgtcgt ccacgccgcc	5700
ggcctcgtcg acgacggcct cctcatgcag cagaccccg cccgttccg caccgtaatg	5760
ggccccaagg tccagggcgc cctccacctc cacaccctca cccgcgaggc cccgtctcc	5820
ttcttcgtcc tctacgcctc cgccgtggc ctttccggtt ccccgggcca gggaaactac	5880
gccgccgcca acgccttctt cgacgcctc agtcaccacc gccgcgcca gggcctcccg	5940
gccctctcca tcgactgggg catgttcacg gaggtcgga tggccgtcgc ccaggaaaac	6000
cgcgggtccc gccagatctc ccgcggcatg cgcggcatca cgcgggatga aggcctctcc	6060
gcccctgccc gcctcctcga gggcgaccgc gtccagaccg gcgtaatccc gatcaccccg	6120
cgccagtggg tcgagttcta cccggccacc gccgccagcc gccgcctctc ccgcctcgtc	6180
acgacgcagc gcgcggtcgc cgaccggacc gccggcgacc gcgacctctt cgagcagctc	6240
gcttccgccg agccgtcgc ccgcgccggc ctctccagg acgtcgtccg cgtccaggtc	6300
tcccacgtcc tcgcctccc ggaggacaag atcgaagtgc acgccccgt ctctccatg	6360
ggcatggact ccttatgtc cctcgagctc cgcaaccgca tcgaggccgc cctcggcgtc	6420
gccgcaccgg cagccctcgg ctggacctac ccgaccgtcg ccgccatcac ccgtggctc	6480
ctgcagcagc cctcgtcgt ccgtctcggc ggccggctccg acaccgacga gtccaccgcc	6540
tccgccggct ccttcgtcca cgtctcgcg ttccgtcccg tcgtcaagcc gcgcgcccg	6600
ctcttctgct tccatggctc cggcgggtcc ccggagggt tccgtctctg gtccgagaag	6660
tccgagtggc ccgacctcga gatcgtgcc atgtggcacg accgtccct cgcctccgag	6720
gacgccccgg gcaagaagta cgtccaggaa gccgcctccc tcattccagca ctacgccgac	6780
gccccgttcg cctcgtcgg cttctccctc ggcttcgct tcgtcatggg caccgccgtc	6840
gagctcgct cccgcagcgg cgccccggcc ccgtcgcgg tcttcacctt cggcgggttcg	6900
ctcatctct cctccgagat caccgggag atggaaaccg atatcatcgc caagctcttc	6960
ttccgcaacg ccgcgggttt cgtccgatcc acccagcagg tccaggccga cgcgccgcc	7020
gacaaggtca tcaccgacac catggtcgcc ccggccccgg gcgactcaa ggaaccgccg	7080
gtcaagatcg ccgtcccaat cgtcgccatc gccgggtccg acgacgtcat cgtccgcgcg	7140

tccgacgtcc aggacctcca gtcccgacc accgagcgct tctacatgca cctcctcccg 7200
ggcgaccacg agttcctcgt cgaccgtggc cgcgagatca tgcacatcgt cgactccac 7260
ctcaaccgcg tctcgcgcg cgcaccacc tctccggcc cgctttcga ggccaagtga 7320
7320

<210> 7

<211> 1260

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in
Streptomyces venezuelae

<400> 7

atgaccagg agcaggcgaa ccagtcgag acgaagccgg cttcgactt caagcccttc 60
gccccgggt acgccgagga ccccttccc gccatcgagc ggctccgga ggccacccg 120
atcttctact gggacgaggg ccggagctgg gtgtcaccc gctaccacga cgtctccgc 180
gtgttcaggg acgagcgctt cgccgtctcc cgggaggagt gggaatcctc ggccgagtac 240

agctccgcca tcccgagct ctccgacatg aagaagtacg ggctcttcgg cctgcccccc 300
gaggaccacg cccgcgtccg caagctcgtg aaccgcctt tcacctcgcg ggccatcgac 360
ctcctccgcg cggaaatcca ggcaccgtc gaccagctcc tcgacgccg gtcggccag 420
gaggagttcg acgtcgtccg cgactacgcg gaaggcatcc cgatgcgggc gatctccgcg 480
ctctcaagg tccggccga gtgcgacgag aagtccgcc gcttcggctc cgccacggcc 540
cgcgactcg gcgtcggcct cgtcccgcag gtcgacgagg agaccaagac cctcgtcgcc 600
tcggtcacgg agggcctcgc cctcctgcac gacgtcctgg acgaacgccg ccgcaaccg 660

ctggagaacg acgtcctcac gatgctctc caggccgagg cggacggctc gcgcctctcc 720
accaaggagc tggtcgtctt cgtcggcgcc atcatcgccg ccggcaccga caccaccatc 780
tacctcatcg ccttcgccgt cctcaacctc ctccgctcgc ccgaggccct cgagctcgtc 840
aaggcggagc ccggcctcat gcggaacgcc ctcgacgagg tgcctcgctt cgacaacatc 900
ctccgcatcg gcacgggtcg gttcgccgc caggatctgg agtactcgcg cgctcgatc 960
aagaaggcg agatggctt cctcctcatc ccgagcgccc tccgcgacgg aacctcttc 1020
tccgccccg acgtcttcga cgtgcgtcga gacacggggg cgtccctcgc gtacggccgg 1080

ggcccgacg tctgccccgg cgtctccctc gcccgcctcg aggccgagat cgccgtcggc 1140
accatcttcc gccgttccc ggagatgaag ctgaaggaaa cgccggtgtt cgggtaccac 1200

cccgccttcc ggaacatcga gtccttgaac gtcctcctga agccgtccaa ggcgggctga 1260
1260

<210> 8

<211> 555

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in
Streptomyces venezuelae

<400> 8

atgacgtccg aggaggtccc cggcgccgcc ctgggcgccc agtcgtccct cgtccgggcc 60
cagcacgccg cgcgccacgt ccgcccgtgc acccgcgagg aggagccgcc cgcgctcatg 120
cacggactca ccgagcgcca ggtcctcttc tcgtcgttgg ccctggccct cgtgctgctc 180
accgcccgtg cgttcggcga actcgcccgg cgcctccgcc agcccagggt cctcggcgag 240
ctcttcggcg gcgtcgtcct ggggtcccage gtggtcgggg cgctcgcgcc gggcttccac 300
cgggtcctct tccaggacct cgcggtcggc gtcgtcctct ccggcatctc ctggatcggg 360
gccctcgtgc tgcctcatat ggcgggcatc gaggtggacg tctccatcct ccgcaaggag 420

gcccggccgg gcgcctcttc ggcctcggc gccatcgccc cgcgctccg caccggggc 480
cccctcgtcc agcgcatgca gggggccttc acgtgggacc tggacgtcag cccgcgccgc 540
agcgcgcagg cctga 555

<210> 9

<211> 2142

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in
Streptomyces venezuelae

<400> 9

atgcacgggc tgacggagcg ccaggctctc ctctcgtcgc tcacgtcgc cctcatctc 60

gtcaccgccc gcgcgtccgg cgagctcgcc cgcgcctcc gccagccgga agtcctcggc 120
gagctcttcg gcggcgtcgt cctcggcccg agcgtcgtcg gagccctcgc cccgggcttc 180
caccgcgccc tcttcaggga accgctgtc ggcgtcgtcc tctcggcat ctctggatc 240
ggcgccctcc tctcctcct catggccggc atcgaggtcg acgtcggcat cctccgaag 300
gaggcgccgc cgggggcctt ctcgccttc ggcgccatcg ccccccgct cgcgcggggc 360

gcagcgttct cggccctcgt cctcgaccgc cccctcccgt cggcctctt cctcgggac	420
gtactctccg tcaccgccgt ctccgtcatc gccaaagtcc tcctcgagcg cgagtcctatg	480
cgccgctcct acgcccaggt gacctcgcg gcgggcgtcg tctccgaggt cgccgctgg	540
gtctctgtcg ccatgacctc cagctcctac ggggcctccc cggccctcgc cgtcgccgc	600
tccgcccctc tcgctccggg cttctctctc ttcatgggcc tctcgggtcg ccgctcacg	660
cacctggcca tgcgtgggt cgccgacgcc acccgcgtct ccaaggcca ggtctccctc	720
gtctctgtcc tcacgttctt cgccgccgcc ctacccagc gcctcgggtt ccacccgtg	780
ctcggggcct tcgcccctcg ggtctctctc aactccgac cgcgaccaa ccgcccgtc	840
ctcgacggcg tccagacct cgtggccggg ctcttcgccc ccgttttctt cgtgctcgcc	900
ggaatgaggg tcgacgtcag ccagctccgc acgccggccg cctggggcac cgtcgccctc	960
ctctcgcca ccgccaccgc cgccaaggtc gtcccgccg ccctcggcgc ccgtctcggc	1020
ggcctccgcg gctcggaggc cgccctcgtg gccgtcggtc tcaacatgaa gggcggcacc	1080
gacctcatcg tcgcatcgt cggagtcgag ctccgcccct tctccaacga ggctacacc	1140
atgtacgccg tcgtcgccct cgtcacgtc accgcctccc ccgcccctt catctggctc	1200
gagaagcgcg ccccgccac ccaggaagag tccgcccgcc tcgagcgcca ggaggccgc	1260
cggcgcgctt acatcccggg cgtcgagcgc atctctgtcc ccatcgtcgc ccacgcctc	1320
ccgggggttc ccacggacat cgtcgagtc atctctgcct ccaagcgga gctcggggag	1380
accgtcgaca tcaccgagct ctccgtcgag cagcaggccc cgggcccgtc ccgcgccgc	1440
ggcgaggcct ccccgggcct cgccaggctc ggccccggc tccgctcgg catctggcg	1500
cagcgccgcg agctccgtgg ctccatccag gccatctcc gtgcgtccg cgaccagac	1560
ctctctgtca tcggcgcccg ctccccggc cgccccgtg gcatgtctt cggccgctg	1620
caggacgcca tcgtccagcg cgccgaaagc aacgtcctcg tcgtcgtcgg cgacccgccc	1680
gccgccgagc gcgcgtccgc ccgcccgcct ctctgtccga tcctcggcct cgagtacagc	1740
ttcgccgcg ccgacctgc cgcacacgtc gccctgcct gggacgcgga gctcgtctc	1800
ctctctccg cccagacgga cccgggcgcc gtctctggc gcgaccgga gccgtcgcg	1860
gtccgcgcg tggcgcggtt cgtcgtcgac gagccgtct tcccggtcg ccgctcggc	1920
gtccgctct ccagccgagt ccacgtcggc gccaccctt ccgacgagat caccgcgag	1980
ctgccccgcg cgccctacga cttgctcgtc ctccgctgct acgaccagc cccctcggc	2040
cgcctctacc tcggatccac cgtcgaatcc gtagtcgtcc gctcccgcgt cccggtggcc	2100

ctcctcgtgg cccacggtgg caccgcgag caggtccgct ga

2142

<210> 10

<211> 918

<212> DNA

<213> Artificial Sequence

<220><223> Codon Optimized DNA of Sorangium cellulosum for Expression in
Streptomyces venezuelae

<400> 10

atggagaagg agtcgcgcat cgccatctac ggcccatcg ccgccaacgt cgccatcgcc 60

gccgtcaagt tcatcgccgc ggccgtcacg ggctcgtcgg cgatgctcag cgagggcgctc 120

cactccctcg tggacaccgc cgacggcctc ctcctcctcc tcggcaagca ccgcagcgcg 180

cgcccgccgg acgccgagca ccccttcggc caccgcaagg agctctactt ctggaccctc 240

atcgtggcca tcatgatctt cgccggccgc ggccggcgtc ccatctacga ggccatcctc 300

cacctcctcc acccccggca gatcgaggac ccgacctgga actacgtcgt cctcggggcc 360

gcagcgggtg tgaaggcac ctcctgatc atctccatcc acgagttcaa gaagaaggac 420

gggcagggct acctcgccgc catgcgctcc agcaaggacc ccaccacctt caccatcgtc 480

ctcaggagca gcgcgcgct ggccggcctc acgatcgctt tcctcggcgt ctggctcgga 540

caccgcctcg gcaaccgta cctggacggc gccgcgtcca tcggcatcgg gtcgtcctc 600

gccgcgggtg ccgtcttctt cgctcccag tcccgcggcc tgctcgtcgg cgagtccgcc 660

gaccgggagc tctcgcgcg catccgcgcc ctggcctccg cggaccgggg cgtgtccgcc 720

gtggggcgcc cgctcacat gcacttcggc ccgcacgagg tctcgtcgt gctccggatc 780

gagttcgacg cggcgctcac gcctccggc gtggccgagg ccatcgagcg tatcgagacc 840

cgcatccggt ccgagcgccc cgacgtcaag cacatctacg tcgaggcgcg ctcctccac 900

cagcgcgccc gcgcgtga 918

<210> 11

<211> 207

<212> DNA

<213> Sorangium cellulosum

<400> 11

actttcaatt tttccgaggg gggttggtc tctggttctt caggaagcct gatcgggacg 60

agctaattcc catccatttt tttaggctc tgctcaaagg gattagatcg agtgagacag 120

ttcttttgca gtgcgcgaag aacctgggcc tcgaccggag gacgatcgac gtccgcgagc	180
gggtcagccg ctgaggatgt gcccgtc	207
<210> 12	
<211> 146	
<212> DNA	
<213> Sorangium cellulosum	
<400> 12	
gagcgccgaa caaaaccagg ccgagcgggc cgatgagccg caagcccgcc tgcgtcaccc	60
tgggactcat ctgatctgat cgcgggtacg cgtcgcgggt gtgcgcgttg agccgtgttg	120
ttcgaacgct gaggaacggt gagctc	146
<210> 13	
<211> 4	
<212> DNA	
<213> Sorangium cellulosum	
<400> 13	
ttgc	4
<210> 14	
<211> 183	
<212> DNA	
<213> Sorangium cellulosum	
<400> 14	
cggtaggcagc gctgagtgcg catctggctt ctcatgtcgt ctctacgggg gatggggaat	60
ccgcgcgccc gccggataca gggagcgtgg ctccaacgac ccacgaagtc gcttcgctcg	120
acgaagacgg gttgttcgcg ttgattgatg agtcactcgc gcgcgcggga aagaggtgat	180
tgc	183
<210> 15	
<211> 83	
<212> DNA	
<213> Sorangium cellulosum	
<400> 15	
cctgcccgac accaccggca ccaccggcac caccagcccc cctcacacac ggaacacgga	60
acggacaggc gagaacggga gcc	83
<210> 16	

<211> 102
 <212> DNA
 <213> Sorangium cellulosum
 <400> 16
 cccgaccgcg gtctgtcccc acgcgccgca ccccgcgcat cccgcgacc acccgccccc 60
 acacgcccac aaccccatcc acgagcggaa gaccacaccc ag 102
 <210> 17
 <211>
 > 68
 <212> DNA
 <213> Sorangium cellulosum
 <400> 17
 gcgccggccg tcgccgaggc cgtcctctcc tggctcgacg ccatcgaggg catcgagggg 60
 gcggggcaa 68
 <210> 18
 <211> 147
 <212> DNA
 <213> Sorangium cellulosum
 <400> 18
 gacaccgggg ccccggtccg cccggccccc cttcggacgg accggacggc tcggaccacg 60
 gggacgggtc agaccgtccc gtgtgtcccc gtccgggtcc cgtccgcccc atcccgcccc 120
 tccaccggca aggaaggaca cgacgcc 147
 <210> 19
 <211> 66
 <212> DNA
 <213> Sorangium cellulosum
 <400> 19
 gcaccgccca agaccccccg gggcgggacg tcccgggtgc accaagcaaa gagagagaaa 60
 cgaacc 66
 <210> 20
 <211> 75
 <212> DNA
 <213> Sorangium cellulosum
 <400> 20

aagcggcgggt gcccgccacc cgggaggggtg gcgggcaccg ctttttcggc ctgctgggtc 60
taccggacca cctga 75
<210> 21
<211>
104
<212> DNA
<213> Streptomyces venezuelae
<400> 21
tccttgccgc acggtgcct cccgccccgg acccgtccc gggcacctcg actcgaatca 60
cttcatgcgc gcctcgggcg cctccaggaa ctcaaggga cagc 104
<210> 22
<211> 108
<212> DNA
<213> Streptomyces venezuelae
<400> 22
cgtgacgccg tgaaagcggc ggtgcccccc acccgggagg gtggcgggca ccgctttttc 60
ggcctgctgg gtctaccgga ccacctgacc tgcaggaaac ttaagaaa 108
<210> 23
<211> 83
<212>
> DNA
<213> Streptomyces venezuelae
<400> 23
cctgccccgac accaccggca ccaccggcac caccagcccc cctcacacac ggaacacgga 60
acggacaggc gagaacggga gcc 83
<210> 24
<211> 102
<212> DNA
<213> Streptomyces venezuelae
<400> 24
cccgaccgcg gtcctgcccc acgcgccgca ccccgcgcat cccgcgcacc acccgcccc 60
acacgccac aaccccatcc acgagcggaa gaccacacc ag 102
<210> 25
<211> 68
<212> DNA

<

213> Streptomyces venezuelae

<400> 25

gcgccggccg tcgccgaggc cgtcctctcc tggctcgacg ccatcgaggg catcgagggg 60

gcgggcaa 68

<210> 26

<211> 147

<212> DNA

<213> Streptomyces venezuelae

<400> 26

gacaccgggg ccccggtccg cccggccccc cttcggacgg accggacggc tcggaccacg 60

gggacggctc agaccgtccc gtgtgtcccc gtccggctcc cgtccgcccc atcccgcccc 120

tccaccggca aggaaggaca cgacgcc 147

<210> 27

<211> 66

<212> DNA

<213> Streptomyces venezuelae

<400> 27

gcaccgccca agaccccccg gggcgggacg tcccgggtgc accaagcaaa gagagagaaa 60

cgaacc 66

<210> 28

<211> 101

<212> DNA

<213> Streptomyces venezuelae

<400> 28

cgccgtgaaa gcggcggtgc ccgccaccg ggagggtggc gggcaccgt ttttcggcct 60

gctgggtcta ccggaccacc tgagctagca aatctagaaa a 101