



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0139151

(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 10-2014-0067240

(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

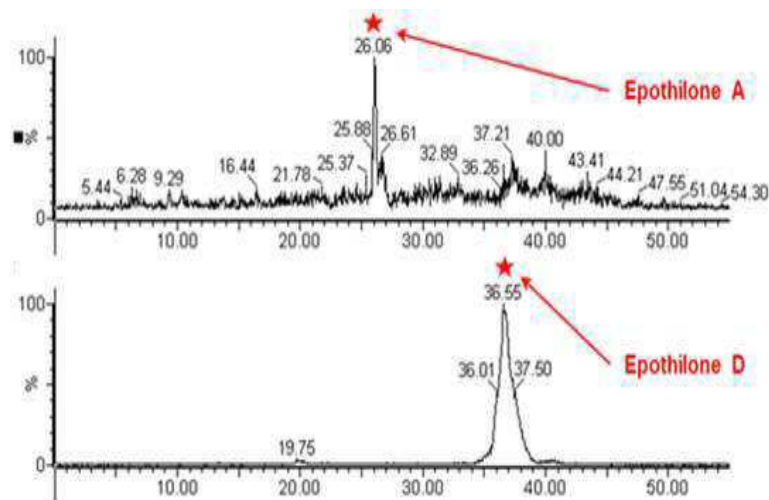
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 에포틸론 발현용 벡터 및 이의 용도

(57) 요약

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

대표도 - 도4



(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

act 프로모터, 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*)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 발현용 벡터.

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 발현용 벡터.

청구항 3

act 프로모터, 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*)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 발현용 벡터.

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 발현용 벡터.

청구항 5

제1항 내지 제4항 중 어느 한 항에 따른 에포틸론 발현용 벡터가 하나 이상 도입된 스트렙토마이시스 실리칼라를 배양하는 단계를 포함하는, 에포틸론의 제조방법.

청구항 6

제5항에 있어서,

상기 에포틸론은 에포틸론 A, B, C, 및 D를 포함하는 군에서 선택되는 어느 하나 이상인 것인, 에포틸론의 제조방법.

청구항 7

제5항에 있어서,

상기 에포틸론 발현용 벡터는 pDHS702, pDHS618, pSBAC, 및 pSET152로 이루어지는 군에서 선택되는 어느 하나의 벡터에 합성된 것인, 에포틸론의 제조방법.

청구항 8

제5항에 있어서,

상기 에포틸론 발현용 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 pDHS702-optSoC_APBSc; 또는

(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 셀룰로섬에서 유래한 유전자간 영역을 포함하는 pDHS618-optSoC_DEF_Sc인, 에포틸론의 제조방법.

청구항 9

제8항에 있어서,

상기 벡터 pDHS702-optSoC_APBSc 및 pDHS618-optSoC_DEF_Sc가 도입된 스트렙토마이시스 실리칼라를 배양하는 단계를 포함하는, 에포틸론의 제조방법.

발명의 설명

기술 분야

[0001] 본 발명은 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*)에 코돈 최적화된 에포틸론 생합성 유전자와 유전자간 영역(intergenic region)을 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 미소박테리움(myxobacterium)인 소란지움 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 스트렙토마

이시스 베네주엘라(*Streptomyces venezuelae*), 믹소코쿠스 잔서스(*Myxococcus xanthus*), 에스케리치아 콜라이(*Escherichia coli*), 그리고 스트렙토마이시스 실리칼라에서 에포틸론의 생합성을 시도하여 왔다. 독일 Rolf Miller 연구팀에서는 에포틸론 생합성 유전자를 최적화하여 합성한 후 믹소코쿠스 잔서스에서 발현하는 기술(Oßwald C et al., ACS Synth Biol., 2012. 10)을 보고한 바 있다.

[0006] 또한, 본 발명자들은 방선균 중 스트렙토마이시스 베네주엘라에서의 에포틸론 생산방법에 대하여 출원한 바 있다(한국 특허등록 제 10-0949313호). 그러나 상기 발명들은 에포틸론 B의 경우 약 0.1 μ g/L, 에포틸론 D의 경우 약 0.4 μ g/L의 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 향암물질인 에포틸론의 수율을 높이는 방법을 개발하고자 예의 연구 노력한 결과, 에포틸론 생합성 유전자집단을 이중숙주 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*)에 적합하도록 최적화하여 합성하고, 상기 합성된 유전자 사이에 소란지움 셀룰로섬 유래의 유전자간 영역(intergenic region)을 삽입한 발현 카세트를 포함하는 벡터를 스트렙토마이시스 실리칼라에 도입할 경우 에포틸론의 수율을 현저히 향상시킴을 확인하고 본 발명을 완성하였다.

발명의 내용

해결하려는 과제

[0008] 본 발명의 하나의 목적은 *act* 프로모터, 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*)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 발현용 벡터를 제공하는 것이다.

[0009] 본 발명의 하나의 목적은 *act* 프로모터, 스트렙토마이시스 실리칼라에 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 발현용 벡터를 제공하는 것이다.

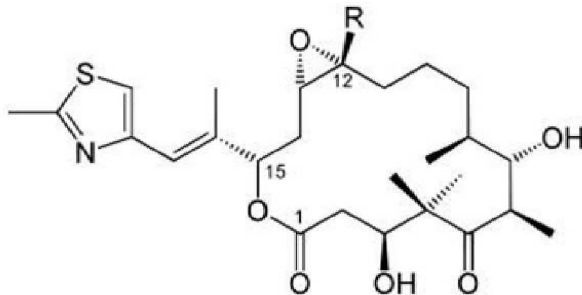
[0010] 본 발명의 또 다른 목적은 상기 에포틸론 발현용 벡터가 하나 이상 도입된 스트렙토마이시스 실리칼라를 배양하는 단계를 포함하는, 에포틸론의 제조방법을 제공하는 것이다.

과제의 해결 수단

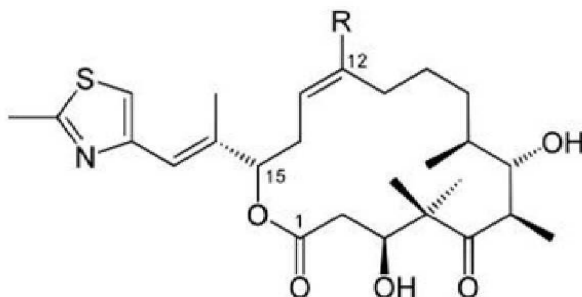
[0011] 상기의 목적을 달성하기 위한 하나의 양태로서, 본 발명은 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*)에 코돈 최적화된 에포틸론 생합성 유전자와, 하나 이상의 유전자간 영역(intergenic region)을 포함하는 에포틸론 발현 카세트를 제공한다.

[0012] 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) 믹소박테리움(myxobacterium)인 소란지움 셀룰로섬에 의해 생산되는 이차 대사산물로 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



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

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 실리칼라에 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 실리칼라이다. 스트렙토마이시스 실리칼라(*Streptomyces coelicolor*)는 방선균의 일종으로, 스트렙토마이시스 속에 속하는 토양 미생물이며 그람 양성균이다.

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

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 이종생산

이 연구되고 있으나, 생산되는 에포틸론의 양이 채 1 μ g/L도 되지 않으며, 낮은 온도에서(15℃) 장기간 배양(6일)이 요구되는 단점이 있었다.

[0022] 이에 본 발명자들은 에포틸론의 생산수율을 향상시키기 위하여 이중숙주로 스트렙토마이시스 실리칼라를 사용하고 이중숙주에 코돈 최적화된 에포틸론 생합성 유전자를 포함하는 벡터를 완성하였다.

[0023] 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 실리칼라에 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.

[0024] 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.

[0025] 본 발명자들은 상기 유전자간 영역으로 소란지움 셀룰로섬의 유전자간 영역을 사용하여 벡터를 제조한 결과, 에포틸론의 생산성이 매우 향상되었음을 확인하였다.

[0026] 하나의 구체예로서, 본 발명의 에포틸론 발현 카세트는 스트렙토마이시스 실리칼라에 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.

[0027] 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.

[0028] 상기 유전자간 영역을 포함하는 본 발명의 에포틸론 발현 카세트가 pDHS702 벡터에 합성된 벡터 pDHS702-optSoC_APB_C-Sc의 개열지도를 도 2a에 도시하였다.

[0029] 또 하나의 구체예로서, 본 발명의 에포틸론 발현 카세트는 스트렙토마이시스 실리칼라에 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.

[0030] 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.

[0031] 상기 유전자간 영역을 포함하는 본 발명의 에포틸론 발현 카세트가, pDHS618 벡터에 합성된 벡터 pDHS618-optSoC_DEF-Sc의 개열지도는 도 2b에 도시하였다.

[0032] 또 하나의 양태로서, 본 발명은 상기 에포틸론 발현 카세트를 포함하는, 에포틸론 발현용 벡터를 제공한다.

[0033] 본 발명에서 용어, 벡터는 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, 막 표적화 또는 분비를 위한 서열 등을 적절히 선택하고 목적에 따라 다양하게 조합할 수 있다.

- [0034] 상기 에포틸론 발현 카세트가 벡터에 포함되는 방법은 특별히 제한되지 아니하며 당해 기술 분야에서 통상적으로 사용되는 방법이라면 어느 것이든 사용할 수 있다.
- [0035] 상기 벡터는 pDHS702, pDHS618, pSBAC, 및 pSET152로 이루어지는 군에서 선택되는 어느 하나의 벡터에 합성된 것일 수 있다.
- [0036] 상기 벡터는 *act* 프로모터를 사용하는 것일 수 있다.
- [0037] 상기 pDHS702 벡터에 합성되는 벡터는 pDHS702-optSoC_APB_C_Sc 또는 pDHS702-optSoC_DEF_Sc, 또는 pDHS702-optPD_APB_CDEF_Sc일 수 있다. 또 다른 구체예로서, 상기 pDHS618 벡터에 합성되는 벡터는 pDHS618-optSoC_APB_C_Sc, pDHS618-optPD_DEF_Sc, 또는 pDHS618-optPD_APB_CDEF_Sc일 수 있다. 바람직하게는 pDHS702-optSoC_APB_C_Sc 또는 pDHS618-optSoC_DEF_Sc일 수 있다.
- [0038] 하나의 구체예로서, 본 발명의 벡터는 pDHS702-optSoC_APB_C_Sc일 수 있다. 상기 벡터 pDHS702-optSoC_APB_C_Sc는 *act* 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_APB_C_Sc의 개열지도도 도 2a에 도시하였다. 본 발명자들은 상기 벡터 pDHS702-optSoC_APB_C_Sc를 스트랩토마이시스 실리칼라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.
- [0039] 다른 구체예로서, 본 발명의 벡터는 pDHS618-optSoC_DEF_Sc일 수 있다. 상기 벡터 pDHS618-optSoC_DEF_Sc는 *act* 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_Sc의 개열지도도 도 2b에 도시하였다. 본 발명자들은 상기 벡터 pDHS618-optSoC_DEF_Sc를 스트랩토마이시스 실리칼라에 도입하여 배양한 결과, 에포틸론 A, B, C, D의 발현이 향상됨을 확인하였다.
- [0040] 또 하나의 양태로서, 본 발명은 상기 에포틸론 발현용 벡터가 도입된 스트랩토마이시스 실리칼라를 배양하는 단계를 포함하는, 에포틸론의 제조방법을 제공한다.
- [0041] 상기 에포틸론은 에포틸론 A, B, C, 및 D를 포함하는 군에서 선택되는 어느 하나 이상일 수 있다.
- [0042] 상기 벡터가 스트랩토마이시스 실리칼라에 도입되고, 배양하는 방법은 특별히 제한되지 아니하며 당해 기술 분야에서 스트랩토마이시스 실리칼라에 대하여 통상적으로 사용되는 방법이라면 어느 것이든 사용할 수 있다.
- [0043] 하나의 구체예로서, 본 발명의 에포틸론의 제조방법은 *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, 제7항, 및 제8항 중 어느 한 항에 따른 제2 발현 카세트를 포함하는 군에서 선택되는 어느 하나 이상을 포함할 수 있다.
- [0044] 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 제

2 발현 카세트가 합성된 것일 수 있다.

[0045] 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 실리칼라에 도입될 수 있다. 도입되는 방법은 특별히 제한되지 아니하며 당해 기술 분야에서 스트렙토마이시스 실리칼라에 대하여 통상적으로 사용되는 방법이라면 어느 것이든 사용할 수 있다.

[0046] 다른 구체예로서, 본 발명의 에포틸론의 제조방법에 사용되는 발현 카세트는 pDHS702, pDHS618, pSBAC, 및 pSET152로 이루어지는 군에서 선택되는 어느 하나의 벡터에 합성된 것일 수 있다.

[0047] 또 다른 구체예로서, 본 발명의 에포틸론의 제조방법에 사용되는 벡터는 상기한 pDHS702-optSoC_APB_C_Sc, pDHS702-optSoC_DEF_Sc, 또는 pDHS702-optPD_APB_CDEF_Sc, pDHS618-optSoC_APB_C_Sc, pDHS618-optPD_DEF_Sc, 또는 pDHS618-optPD_APB_CDEF_Sc일 수 있다. 바람직하게는 pDHS702-optSoC_APB_C_Sc 또는 pDHS618-optSoC_DEF_Sc일 수 있으나 이에 제한되지 않는다. 상기 각 벡터에 대한 설명은 상기한 것과 같으므로 생략한다. 상기 발현 카세트는 먼저 제작한 후 벡터로 합성하여 스트렙토마이시스 실리칼라에 도입할 수도 있고, 한 번에 벡터를 합성하여 도입할 수도 있다.

[0048] 또 다른 구체예로서, 본 발명의 에포틸론의 제조방법은 상기 2개의 벡터 pDHS702-optSoC_APB_C_Sc 및 pDHS618-optSoC_DEF_Sc가 도입된 스트렙토마이시스 실리칼라를 배양하여 에포틸론 A, B, C, D를 합성할 수 있다. 본 발명자들은 pDHS702-optSoC_APB_C_Sc 및 pDHS618-optSoC_DEF_Sc를 스트렙토마이시스 실리칼라에 도입하여 배양한 결과, 에포틸론 A의 수율은 약 1 내지 10 $\mu\text{g/L}$, 에포틸론 D의 수율은 7 내지 10 $\mu\text{g/L}$ 으로 나타나, 에포틸론 생산 수율이 현저히 증가함을 확인하였다.

발명의 효과

[0049] 본 발명에 따른 에포틸론 발현 카세트 및 발현용 벡터는 성장속도가 빠르고, 유전자 조작이 용이한 이중속주 스트렙토마이시스 실리칼라에 도입할 수 있어, 단시간 동안 높은 수율로 에포틸론을 생산할 수 있다.

[0050] 또한 본 발명에 따른 에포틸론의 제조방법에 의하여 생산된 에포틸론은 항암물질으로 유용하게 사용할 수 있다.

도면의 간단한 설명

[0051] 도 1은 본 발명에 따른 벡터에 포함되는 에포틸론 생합성 유전자 및 그의 제한효소를 나타낸 것이다.

도 2a는 본 발명에 따른 에포틸론 발현용 벡터 pDHS702-optSoC_APB_C_Sc의 개열지도를, 도 2b는 pDHS618-optSoC_DEF_Sc의 개열지도를 도시한 것이다.

도 3은 본 발명에 따른 에포틸론 발현용 벡터 pDHS702-optSoC_APB_C_Sc 및 pDHS618-optSoC_DEF_Sc를 스트렙토마이시스 실리칼라에 도입한 후 RT-PCR을 통하여 전사여부를 확인한 결과를 나타낸 것이다.

도 4는 본 발명에 따른 벡터 pDHS702-optSoC_APB_C_Sc와 pDHS618-optSoC_DEF_Sc를 도입한 스트렙토마이시스 실리칼라 균주에서의 ESI-MS/MS 분석 결과를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0052] 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.

[0053] 균주, 플라스미드, 배양조건

[0054] 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 널

리 이용되는 플라스미드이다.

[0055] 에포틸론을 스트렙토마이세스 실리칼라에에서 발현시키기 위한 대장균-스트렙토마이세스 셔틀(shuttle) 벡터로 pDHS702, pDHS618 또는 pSBAC를 사용하였다. 상기 벡터들은 *act* 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.

[0056] 조작된 돌연변이 균주인 스트렙토마이세스 실리칼라 DHS2001은 한국 등록특허 제10-0949313호에 언급된 균주를 사용하였다.

[0057] 대장균은 플라스미드의 선별을 위하여 암피실린 (ampicillin, 50 μ g/mL) 또는 아프라마이신 (apramycin, 25 μ g/mL)이 첨가된 LB배지에서 배양하였다.

[0058] 실시예 1: 에포틸론 생합성 유전자 합성 및 발현벡터 제조

[0059] 에포틸론 생합성 유전자는 소란지움 셀룰로섬의 유전자 정보를 바탕으로 이중숙주 스트렙토마이세스 실리칼라의 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.

[0060] 코돈 최적화된 에포틸론 생합성 유전자 단편과 각각의 제한효소를 도 1에 나타내었다.

[0061] 상기 합성된 유전자 단편들은 각각의 제한효소를 이용하여 서브클로닝 단계를 거쳐 결합시켜 pDHS702 및 pDHS618에 도입하였다. 서브클로닝 및 도입 방법은 한국 등록특허 제10-0949313호에 기재된 방법을 일부 변형하여 사용하였다.

[0062] 이를 통하여 하기 표 1과 같이 벡터 pDHS702-optSoC_APBC_Sc, pDHS618-optSoC_DEF_Sc를 완성하였다. 각 벡터의 개열지도는 도 2a 및 도 2b에 도시하였다.

표 1

[0063] 실시예 1에 따른 벡터의 특징

벡터	프로모터	벡터	에포틸론 생합성 유전자 및 유전자간 영역(IGR)
pDHS702-optSoC_APBC_Sc	act	pDHS702	IGR- <i>epoA</i> - <i>epoP</i> -IGR- <i>epoB</i> - <i>epoC</i> -IGR
pDHS618-optSoC_DEF_Sc	act	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

[0064] 실시예 2: 이중숙주에서 에포틸론의 발현

[0065] (1) 스트렙토마이세스 실리칼라에의 형질전환

[0066] 실시예 1에서 제조한 pDHS702-optSoC_APBC_Sc와 pDHS618-optSoC_DEF_Sc를 스트렙토마이세스 실리칼라 M1154에 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.

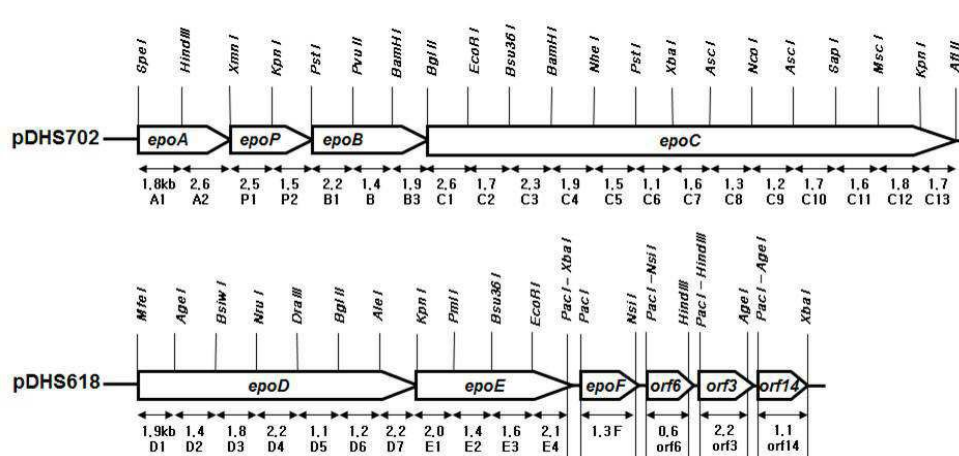
[0067] 상기 R2YE 배지의 조성은 103g 수크로오스 (sigma), 0.25g K₂SO₄ (sigma), 10.12g MgCl₂·6H₂O (sigma), 10g 글루코스 (sigma), 0.1g 카사미노 애시드(casamino acids, Difco), 5g 이스트 추출물(yeast extract, Difco), 5.73g TES 버퍼 (sigma), 2ml 미량원소용액(trace element solution), 22g agar (1L 기준)이며, 고온 멸균 (autoclave) 후에 10ml 0.5% KH₂PO₄ (sigma), 4ml 5M CaCl₂·2H₂O (sigma), 15ml L-프롤린(L-proline, sigma), 7mL 1N NaOH(sigma) (1L 기준)을 첨가하여 제조하였다.

[0068] (2) 에포틸론의 생산여부 확인

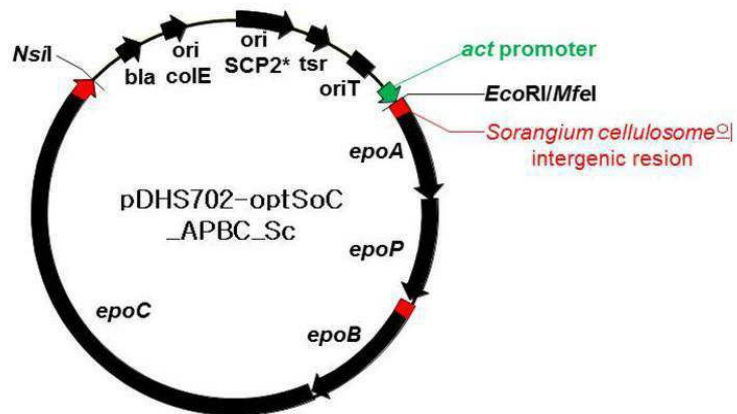
- [0069] 상기 배양 용액을 동량의 에틸아세테이트 (ethyl acetate)와 혼합하고, 분별깔때기로 층 분리를 2회 반복하고, 유기층을 진공회전농축기(rotary evaporator)로 농축하였다. 얻어진 용출물을 100 μ l의 메탄올(methanol)로 다시 녹이고 상기 용액을 Phenomenex Synergi Polar-RP 컬럼(1504.6mm, 4 μ m)을 장착한 LC-ESI-MS/MS (엑체크로마토그래피/전자스프레이 이온화-질량분석기; Waters/Micromass QuattroMicro/MS)로 분석하였다.
- [0070] 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.
- [0071] ESI-MS/MS로 분석한 결과, 상기 형질전환된 스트렙토마이시스 실리칼라 균주에서 에포틸론 A, B, C, D를 모두 생산하는 것을 확인하였다.
- [0072] 특히 26.06분의 용출물은 에포틸론 A이며, 36.55분의 용출물은 에포틸론 D임을 확인하였다. 에포틸론 표준물질로 동일하게 실험한 결과 용출시간 및 분획 형태가 동일하였다. 에포틸론 A의 수율은 약 1~10 μ g/L 및 에포틸론 D의 수율은 약 7~10 μ g/L 생산되는 것을 확인하였다(도 4).
- [0073] 이는 본 발명에 따른 발현 카세트 및 이를 포함하는 에포틸론 발현용 벡터에 의하여 에포틸론 생산 수율이 현저히 증가함을 확인한 것이다.
- [0074] 이와 같이, 본 발명에 따른 에포틸론 발현용 벡터는 생장속도가 빠르고, 유전자 조작이 용이한 이중숙주 스트렙토마이시스 실리칼라에 도입할 수 있어, 단시간 동안 높은 수율로 에포틸론을 생산할 수 있으며, 본 발명에 따른 에포틸론의 생산방법에 의하여 생산된 에포틸론은 항암물질로 유용하게 사용할 수 있다.
- [0075] 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

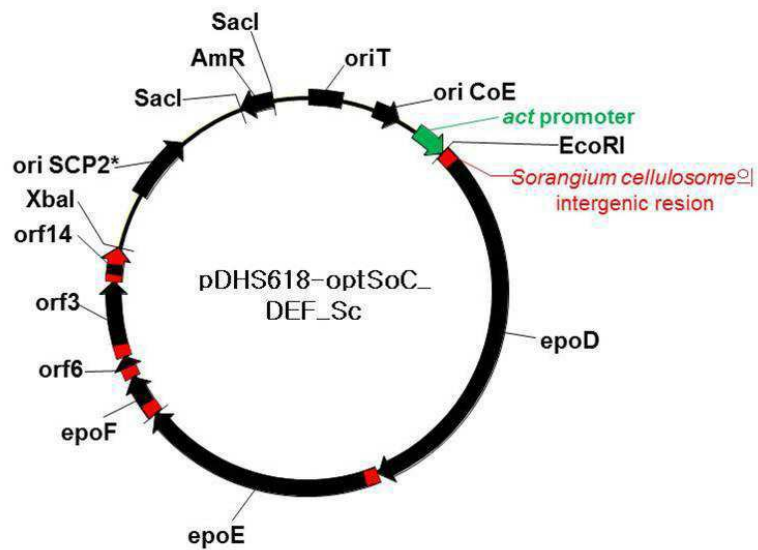
도면1



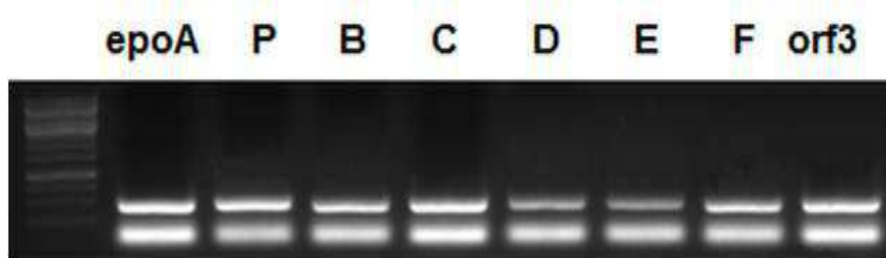
도면2a



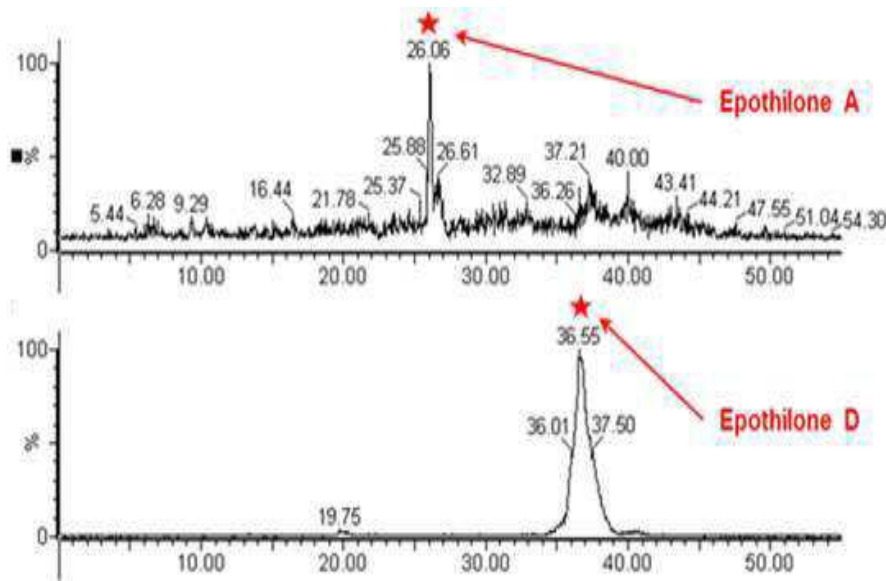
도면2b



도면3



도면4



서열목록

<110> EWhA UNIVERSITY - INDUSTRY COLLABORATION FOUNDATION

<120> EXPRESSION VECTOR AND ITS USE

<130> KPA140533-KR

<160> 20

<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 gccgatcga ggcgcccgc gaggaccga tcgcatcgt cggagcctcc	60
tgccgcctcc cgggcggcgt catcgacctc tccggcttct ggaccctcct cgaaggttcc	120
cgcgacaccg tcggccgcgt cccggccgaa cgctgggacg ccgctgcctg gttcgaccgc	180
gaccccgacg ccccgggcaa gacccggtc acccgcgct ctttctcag tgacgtcgcc	240
tgttcgacg catccttctt cggtatctcc ccgcgcgagg cgctccgcat ggaccggcc	300
caccgtctcc tctcgaggt ctgctgggag gccctcgaga acgcgcgat cgccccgtcc	360
gccctcgtcg gcacggagac cggcgtcttc atcggaatcg gcccgagcga gtacgaggcc	420

gccctccgc aggccaccgc ctccgccgag atcgacgccc acggcggctc cggcaccatg	480
ccctccgtcg gcgccggcgc catcagctac gccctcggcc tccgcggccc gtgcgtcgcc	540
gtcgataccg cctactctc ctccctcgtc gccgtgcacc tcgcctgcca gtcctccgt	600
tccggcgagt gctccaccgc cctcgccggc ggctgtctcc tcatgtctc cccgtccacc	660
ctcgtctggc tctccaagac ccgcgccctc gcccgcgacg gccgtgcaa ggccttctcc	720
gccgaagcgc acggcttcgg tcgcgggtgag ggctgcgcgc tcgtcgtcct gaagcgcctc	780
tccggcggcc gggccgacgg cgaccgcatc ctgcgcgtca tccgcggctc cgccatcaac	840
cacgacggag ccagctccgg gctcaccgtc ccgaacggct cctcccagga gatcgtctc	900
aagcgcggcc tcgccgacgc cggtcgcgc gcctcctcgg tcggctacgt cgaggcgcac	960
ggcaccggga cgaccctcgg cgaccgcatc gagatccagg cctcaacgc cgtctacggc	1020
ctcggctcgc acgtcgccac cccgtctctc atcggctcgc tcaagacaa cctcgccac	1080
ccggaatacg cctccggcat caccggcctc ctcaaggctc tactctcct ccagcacggc	1140
cagatcccg cccacctcca cgcccaggcc ctcaaccccc gcatctctg gggcgacctc	1200
cgcctcaccg tcaccgcgc ccgcaccccg tggccggact ggaacacgc gcgccgcgc	1260
ggtgtctcct ccttcggcat gtcgggaacg aacgcccacg tcgtcctcga ggaggcccg	1320
gccgccact gcaccccgcc ggcccctgag cgtccggcgc agtcctcgt cctctccgcg	1380
cggaccgcct ccgcctcga tgcccaggcc gcccgccctc gcgaccact cgagacctac	1440
ccgtcccagt gcctcggcga cgtcgcttcc tccctcgcca ctacgcgctc cgccatggag	1500
caccgcctcg ccgtcgccgc cacatccgt gagggcctcc gcgtgcctc cgacgtgcc	1560
gcccagggtc agacctcccc gggcgccgtc aggtccatcg ccgactctc caggggcaag	1620
cttgccctcc tcttcaccgg ccagggtgcc cagacgtcg gcatgggccc cggcctctac	1680
gacgtctggt ccgccttcgg cgaggccttc gacctctgcg tccgcctctt caaccaggag	1740
ctcgaccgcc cgtccgcga ggtcatgtgg gccgagccgg cctccgtcga cgccgccctc	1800
ctcgaccaga ccgccttcac ccagccggcc ctgttcacct tcgagtacgc gctcgcggcc	1860
ctctggcgct cctggggggt cgagccggag ctctcgcgc ggcaactcat cggcgagctc	1920
gtcgccgct gcgtcgccgg cgtcttctcc ctcgaggacg ccgtcttct cgtcgccgcc	1980
cgcggtcgtc tcatgcaggc cctccggcc ggccggcgca tggcttccat cgaggcccg	2040
gaggccgacg tcgcgcgcgc cgtcgccgc cagcccggt ccgtctccat cgccgcgctc	2100
aacgccccgg accaggctgt catcgccggc gccggccagc cggctccacg aatcgccgcc	2160

gccatggcgc cccgcggggc cgcaccaag gccctccacg tcagccacgc cttccactcc	2220
ccgtcatg cccgatgct cgaggccttc ggccgcgtcg ccgagtcctg ctctaccgc	2280
cgcgcagca tegtctctg ctccaacctc tccgggaagg cctgtaccga cgaggctctc	2340
tccccgggat actgggtcgc ccacgccgc gaggtctac gcttcgccga tggagtcaag	2400
gccctccacg ccgcaggcgc cggcaccttc gtcgaggtcg gccgaagtc caccctctc	2460
ggtctctg ccgcctgcat gcccgacgc cgtccggccc tctcgcctc ctccgcgcgc	2520
ggccgcgacg aaccggccac cgtcctcgag gcccttgccg gcctctgggc cgtcggcggc	2580
ctcgtctctt gggccggtct cttcccgctc ggccggccgc gcgtccgct cccgacctac	2640
ccgtggcagc gcgagccta ctggatcgac accaaggcgc acgacgccgc ccgcggcgac	2700
cgcgcgcgc cgggcgcgcgc ccacgacgag gtcgaggagg gcggcgccgt ccgtggcggc	2760
gaccgccct cgcgccct cgaccaccgc ccgcggaat ccggccgcgc cgagaaggtc	2820
gaggccgcgc gcgacaggcc gttccgactc gagatcgacg agccggcgct cctcgaccac	2880
ctcgtctctc gcgtaccga acgcgcgcgc ccgggcctcg gcgaggttga gatcgccgc	2940
gacccgcgc gcctctctt caacgacgtg cagctcgccc tcggcatggt cccggacgac	3000
ctccgggca agccgaacct gccgtctctc ctccggggg agtgcgcgcgc cgcatcgtc	3060
gccgtcggc agggcgta cggcctctc gtcggccagc cggtcacgc actctccgc	3120
ggcgcttcg ccaccacgt caccaccgc gccgactcg tctcccgcg accgcaggcc	3180
ctctccgca tcgaggcgc cgcatgccc gtcgcctacc tcaccgcgtg gtacgcctc	3240
gaccgcatc cccgctcca gccgggcgaa cgcgtctca tcacgccgc caccggcggc	3300
gtcgccctc ccgcctgca gtgggcccag cagtcggcg ccgaggtaca cgccaccgc	3360
ggcacgccg agaagcgagc ctacctgag tccctcgcg tgcgtactg ctccgactc	3420
cgtccgacc gcttcgtggc cgacgtccgc gcctggaccg gcggggaggg cgtcgacgtc	3480
gtctcaact cctctccgc cgagctcatc gacaagtcct tcaacctct ccgtccac	3540
ggcgcttcg tcgagctcg caagcgcgac tgctacgcgc acaaccagct cggcctccgc	3600
ccgtctctc gcaacctctc cttctctctc gtcgacctc gcggcatgat gctcgagcg	3660
ccggcccgtg tccgcgcct cctcgaggag ctctcgcc tcacgccgc cgggtgtctc	3720
acccgcgcgc cgatgccac gctcccatc gcccgctgg ccgacgctt ccgtccatg	3780
gcccaggccc agcacctcgc caagctctc ttgacctcg gagaccgga ggtccagatc	3840
cgcacccga cccatgccgc cgcgggccc tccaccggcg accgcacct cctcgaccgc	3900

ctcgctccg ccgccccgc cgcgcgcgc gccgcctcg aggccttcct cgcacccag 3960
gtctcccagg tctccgcac cccggaaatc aagtcggag ccgagccct cttacccgc 4020
ctcggcatgg actccctcat ggccgtcgag ctccgcaacc gcatcgaagc ctccctcaag 4080
ctcaagctct ccaccacatt cctctccacc tccccgaaca tcgccttcct 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 ggccgagctc 180
atcgtcccgg ccccgccga gagcacgcc ccgtcccgc tcaccgacat ccaggagctc 240

tactggctcg gccgaccgg ggcttcacg gtccctccg gcatccacgc ctaccgcgag 300
tacgactgca cggacctcga cgtcccgcc ctctcccgcg ccttcgcaa ggtcgtcgcc 360
cgccacgaca tgctccgcgc ccacacctc ccgacatga tgcaggtcat cgagccgaag 420
gtcgacgccg acatcgagat catcgacctg cggggcctcg accgctccac ccgagaggcc 480
cgcctcgtct cgctccgca cgcgatgtcc caccgatct acgacaccga gcgcccgcg 540
ctctaccacg tagtcgccg ccgcctcgac gagcgccaga cccgcctcgt cctcagcatc 600
gacctcatca acgtcgacct cggcagcctc agcatcatct tcaaggactg gctctcgttc 660

tacgaggacc cggagacgtc cctcccggtc ctccagctgt cctaccgga ctacgtcctc 720
gccctcgaat cccgaagaa gtccgaggcg caccagaggt ccatggacta ctggaagcgc 780
cgcatcgccg agctcccgc gccccgacc ctcccgatga aggcggacc gtccacctc 840
aaggagatcc gcttccgcca caccgagcag tggctccgt cggactcctg gggccggctc 900
aagcgccgcg ttggcgagcg cggccttacc ccgacggcg tcctcctcgc cgccttctcc 960
gaggtcatcg gccgtggtc cgcgtcccc 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
tccctccgc ccgccagct cgaggccgc gcctccgcca acgccacgaa cgccctctc	1560
tccgagcaca cgtccacgg cctcttcgcc gcccgcgtcg agcagctccc gatgcagctc	1620
gcggtcgtct ccgccgcaa gacctcacc tacgaggagc tcagccgccc ctcccgcgg	1680
ctcggggccc gcctccgca gcaggcgcc cgcccgaaca ccctcgtcg ggtcgtcatg	1740
gagaagggtc gggagcaggt cgtcgccgtc ctcgcggtcc tcgagtcgg cgccgcctac	1800
gtcccgatcg acgcggaact gccggccgag cgcattccact acctctcga ccacggcgag	1860
gtcaagctcg tctcaccga gccgtggctc gacggcaagc tctctggcc gccgggcac	1920
cagcgctcc tegtctcga ggccggcgtc gaggcgagc gcgaccagcc gccatgatg	1980
ccgatccaga ccccgctcga cctcgctac gtcatctata cctcgggctc caccggcctc	2040
ccgaaggcgc tcatgatcga ccaccgggc gccgtcaaca ccatcctcga catcaacgag	2100
cgtctcgaga tcggccccgg cgaccgcgtc ctgcctctt cctccctcag cttcgacctc	2160
tccgtgtacg acgtcttcgg gatctcgcc gccggcgga ccatcgtcgt cccggatgcc	2220
agcaagctcc gcgacccggc ccattgggcc gagctcatg agcgcgagaa ggtaaccgtc	2280
tggaaactcg tgccggccct catcgcatg ctgctgagc atttcgaggg ccgcccgac	2340
tccctcgccc gctccctgcg 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 gacgcgaag tccttctctg tgcacccgga gaccggcgag	2700
cgcctctaca agaccggcga cctcgccgc tacctccgg acggcaacat cgaattcatg	2760

ggccgcgagg acaaccagat caagctccgc gggtaccgcg tcgagctcgg cgagattgag 2820

gagaccctca agtcccaccc gaacgtccgc gacgccgtga tcgtcccggc cggcaacgac 2880

gctgcgaaca agtctctcct cgcctacgtc gtgccggagg ggacccgccg ccgcgccgcc 2940

gagcaggatg cctccctcaa gaccgagcgc atcgacgccc gcgccacgc cgccgaagcc 3000

gacggcctct ccgacggcga gcgcgtccag ttcaagctgg cccggcacgg cctccgccgc 3060

gacctcgacg gcaagccggc cgtcgatctc accggccagg acccgcgca agccggcctc 3120

gacgtctacg cccgccgccg ctccgtccgc acgttcttgg agggcccgat cccgttcgtc 3180

gagttcgggc gcttctcttc ctgcctctcc tccgtggaac cggacggcgc caccctcccc 3240

aagttccgct acccgctccg gggtccacc taccgggtcc agacctacgc ctacgtgaag 3300

tccggccgca tcgagggcgt cgacgaaggc ttctactact accaccggtt cgagcacctg 3360

ctgtgaagc tctccgacca cggcatcgag cgcggcgccc acgtgcgcca aaacttcgac 3420

gtcttcgacg agggggcggt caacctctc ttctcgcc gcctcgacgc catcgagtcc 3480

ttgtacggct ccagctcccg cgagttctgc ctgctcgagg ccggtacat ggcccagctg 3540

ctcatggagc agggcccttc ctgcaacatc gggtctgcc cggtcggcca gttcaacttc 3600

gagcagggtc gcccggtcct cgacctccga cactccgacg tctacgtcca cggcatgtc 3660

ggcgcccgcg tcgaccccg gcagttccag gtctgcaccc tcggccagga ctctccccg 3720

cgcccgcca ccaccccgcg cgcgcccg ggcccgagc agcacttcgc cgacatgtc 3780

cgcgacttc tccgaccaa gctccggag tacatggtcc cgaccgtctt cgtcgagtc 3840

gacgccctcc cgtcacgtc caacggcaag gtcgaccgca aggcctccg cgagcgcaag 3900

gacacctcct ccccgcgcca ctccggccac accgccccgc gcgacgcgt cgaggagatc 3960

ctcgtcgccg tcgtccgca ggtactcggc ctcgaggtcg tcgggctgca gcagtccttc 4020

gtcgacctcg gcgcgacctc catcacatc gtccgcatgc gctccctcct ccagaaacgc 4080

ctcgaccgcg agatcgccat caccgagctc ttccagtacc cgaacctggg ctcgctcgcc 4140

tccggcctcc gccgcgactc gcgcgatctc gaccagcgc cgaacatgca ggaccgcgtc 4200

gaggtccgcc gcaagggcgg 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
gccccgcgacc 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 ctacgccgag	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 tcgaagcccc 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 gaattcgct cctccccgtt ctacgtcaac	1200
acctccctca aggactggaa caccggctcc accccgcgcc gcgccggcgt ctctctctc	1260
ggcatcgccg gcaccaacgc ccacgtggtc ctcgaggagg cccccgccgc caagctcccg	1320
gccgccgccc ccgccgctc cgccgagctc ttctgtgtct ccgccaagtc cgccgccgcc	1380
ctcgacgtcg ccgccgccg cctccgcgac cactgcagg cccaccaggg gatctccctc	1440
ggcgacgtcg ccttctccct cgcgaccacc cgtccccga tggagcaccg cctcgccatg	1500
gccccccga gcccgaggc cctccgcgaa ggctcgacg ccgccgccg cggccagacc	1560
ccgccgggtg ccgtccggc ccgtgtctcc ccgggcaacg tcccgaaggc cgtcttcgtc	1620
ttccgggggc aggggttcca gtgggtcggc atgggtcgcc agtcctcgc cgaggagccg	1680

gtgttccacg ccgcccctctc cgcttgcgac cgcgccatcc aggccgaggc cggctgtgtcc	1740
ctctctgccg aactcgccgc cgacgaaggc tctcccagc tcgagcgcat cgacgtcgtc	1800
cagccgggtcc tcttcgccct cgccgtcgcc ttccgccc tctggcgag ctggggcgctc	1860
gccccggacg tcgtcattgg acactccatg ggcgagggtcg ccgccgccca cgtcgccggc	1920
gcccctctct tggaggacgc agtcgccatc atctgccgcc gctcccgcct cctccgccgc	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 tcgccgcct cgccggcctc cggccggggg ccgccgcgt 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 ggccaagtc gcccgggcg accgccgtgg cgtccgcgcc	2700
ggcgccacc cgtctctcgg cgagatgcag acccttcca cccagacct caccgcctc	2760
tgggaaacca ccctcgacct caagccctc ccgtggctcg gagatcacg cgtccagggc	2820
gccgtcgtct tcccgggcgc cgctacctc gagatggcca tctctcggg cgccgaggcc	2880
ctcggggacg gcccgctcca gatcaccgac gtctctctcg ccgaggccct cgccttcgcc	2940
ggtgacgccg ccgtcctcgt ccaggtcgtc actaccgagc agccgtcagg ccgcctccag	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 ccgcctccac	3300
cccgcctcc tcgacgcctg cttccagatc gtcggctccc tcttcgccg ctccggcgag	3360
gccacgccgt gggctccgggt cgagctcggc tccctccgcc tgetccagcg cccttccggc	3420
gaactctggt gccacgccg cgtcgtcaac cacggccacc agaccccgga ccgccagggc	3480
gccgacttct gggctcgtcga ctctccggc gccgtcgtcg ccgagggtctg 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 cggcggaggc	4200
gaggtcgccc tcaggggcgg cgagcgctgc gtcgcgcgca tcgtccgccg ccagccggag	4260
accgccccgc gaggtcgcat cgagtcctgc gtcccgaccg acgtcaccat ccgcgccgac	4320
tccacctacc tcgtaccgg cggcctcggc ggctcggtc tctccgtcgc aggttggtc	4380
gccgagcgcg gcgccgtca cctcgtctc gtcggccgat ccggcgccgc ttcgtcgag	4440
cagcgcgccg ccgtcgccgc cctcgaggcc cgcggcgccc gcgtcacgt 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
agttcccga tgetctccg cctcgtcacc gccaccgcg cctcgccgg cggcccgcc	5100
ggcgacggcg acctcctcg ccgcctcgc gccgccgagc cgtccgccg ctcgcctc	5160
ctcgagccgc tcttcgcgc 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 ttcggccgta 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 cttcggcatc tccccgcgc agggccgctc cctcgaccg 360

cagcaccgcc tctcctcga agtcgcctgg gaggcctcg aggacgccc catccgccc 420

cgctccatcg acggctccc caccggcgtc ttctcggcg ccttcaccgc cgactacgc 480

cgcaccgtcg ccgcctccc gcgcgaggag cgcgacgct actccggcac cggcaaatg 540

ctctccatcg ccgcggcgcc cctcagctac accctcgcc tccagggcc gtgcctcacc 600

gtcgacaccg cctgtctctc ctccctctc gccatccacc tcgctgccc ctccctccgc 660

gccggtgagt ccgacctgc cctcgccgc ggctctccg 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

ccggcccgt ccgacggcac ccgtgctgc ctccggccg tcaagaccaa catcgccac 1140

ctcgaggccg ccgcggcgt cgccggcctc atcaaggccg ccctctcct caccacgag 1200

cgcacccgc gcaacctcaa cttccgact ctaaccgcg 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 cacctccgcg gagggcctcc tcgccgcctc ctcgccctc	1620
gcccagggcc agaccccgcc cggcgccgcc cgtgcatcg cctcctcctc ccgcggcaag	1680
ctcgccttcc tcttcaccgg ccaggcgccc cagaccccg gcatgggcgg cggcctgtgc	1740
gccgcctggc cggccttcgg cgaggcttcc gaccgctcgg tcgccctctt cgaccgcgag	1800
ctcgaccgcc cgctccgca ggtcatgtgg gccgaggccg gctccgccga gtcctcctc	1860
ctcgaccaga ccgccttcac ccaaccggcc ctcttcgccg tcgagtacgc cctcaccgcc	1920
ctctggcgct cctggggggt cgagccggag ctctcgtcgg gccattccat cggcgagctc	1980
gtcgtgcct gcgtcgcgg cgtcttctcc cttgaggacg gcgtccgct cgtcgccgcc	2040
cgcggccgcc tcatgcagg cctcagtcc ggccggcgcca tggatccct cggcgccccg	2100
gaggcggaag tcgccccgc cgtcgccccg caccgcct ccgtttctat cgcgcgcgtc	2160
aacggcccgg agcaggtcgt catcgccggc gtcgagcagg ccgtccaggc catcgccga	2220
ggcttcgccg ccccgggcgc ccgcaccaag cgcctccag tctccacgc ctccactcc	2280
ccgtcatgg agccgatgct cgaggagtcc ggccgcgtcg ccgcctcgt cacctatcg	2340
cgcctcgtcgt tctcctcgt ctccaacctc tccggcaagg tcgtaccga cgagctctcc	2400
gccccgggct actgggtccg ccacgtccgc gaggcctcc gcttcgccga cggcgtcaag	2460
gccctccacg aggcgggcgc cggcaccttc gttgaggtcg gccgaagcc gacctcctc	2520
ggcctcctcc ccgcctgct cccggaggcc gagccgacct tctcgcctc cctccgcgc	2580
ggccgcgagg aggcgcgcgg cgtcctcag gccctcggcc gcctctgggc cgcggcggc	2640
tccgtctcct ggccgggctt ctcccgacc gccggccgcc gcgtccgct cccgacctac	2700
ccgtggcagc gccagcgcta ctggatcgag gccccggccg agggcctcgg cgccaccgcc	2760
gccgacgccc tcgccagtg gttctaccgc gtcgactggc cggagatgcc gcgtcctcc	2820
gtcgactccc gccgcgccg ctcggcggcg tggctcgttc tcgccgaccg cggcggcgtc	2880
ggtgaggccg ccgcgcgcg cctctcctcc cagggtgct cctcgccgt cctccagcc	2940
ccggccgagg cctccgctg cgccgagcag gtacccagg ccctcggcgg ccgcaacgac	3000
tggcaggcgc 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 tcgaccgga ggagtcctcg	3300
accgaggtcg aggcctctgt cgccgagctc ctctccccgg acgccgagga ccagctcgcc	3360
ttccgccagg gcccgccgc cgccgccgc ctctcgccg ccccgccgga gggcaacgcc	3420
gccccggtct cctctccgc cgagggtcc tacctcgta ccggcggcct cggcgccctc	3480
ggcctctcg tcgccgtg gctcgtcgag cgcggcgcc gccacctgt cctcatctc	3540
cgccacggcc tcccgaccg cgaggagtgg ggccgcgacc agccgccgga ggtccgcgc	3600
cgcctcgccg ccatcgaggc cctcgaggcc caggcgccc gcgtcacgt cgcgcgctc	3660
gacgtcgccg acgccgagg catggccgc ctctcgccg ccgtcgagcc gccgtccga	3720
ggcgtcgtc acgccgccg cctctcgac gacggcctc tcgccacca ggacgccgc	3780
agactcgccc gggctctcg cccgaagtc gagggcgct gggctctcca caccctacc	3840
cgcgagcagc cgtcgacct ctctgtctt ttctctccg cctccggcgt ctctggctc	3900
atcgccagg gctcctacg cgccggcaac gccttctcg acgcctcgc cgacctagg	3960
cgcacccagg gcctcgccgc cctctccatc gcctggggcc tctgggccga gggcgcatg	4020
ggctcccagg cccagcgag ggagcacgag gcctccgca tctgggcat gccgacctc	4080
cgcgcccctg cgcctatgga gtggctctc ggaccccg ccacccagc cgtcgtcatc	4140
cagatggact gggcccacg cggcgccgc ccgcgcgac cctcccgcg ccgtttctg	4200
gaccgctcg tcaccgcc caaggaggcc tctctctcg ccgtccggc cgtcgagcg	4260
tggcgcaac cctccgtgt cgagaccgc tccgccctc acgagctgt ccgcggcgtc	4320
gtcgccggcg tcatgggtt caccgaccg ggcactctg acgtccgcg cggttcgcc	4380
gagcagggcc tcgactcct catggccgtc gagatccga aggcctcca gggcgagctc	4440
ggcatgccg tctccgcc cctcgcttc gaccaccga ccgtcgagc cctcgtcgag	4500
tacctctct cccaggcct cgagctccg gaccgaccg acgtccgtc cgtccgctc	4560
ccggccaccg aagaccgat cgccatgtc ggcgccgct gccgttccc gggcggcgtc	4620
gaggacctg agtctactg gcagctctc accgaggcg tcgtcgtct caccgaggtc	4680
ccggccgacc gctggaacg cgccgacgg cggtcccg gctccggcg ggcaccg	4740
cagacctag tcccgcgcg cggttctc cgcgaggtc agaccttga cgccgcttc	4800

ttccacatct ccccgcgca ggccatgtcc ctcgaccgc agcagcgct ctcctcgag	4860
gtctcctggg aggccatcga gcgcgccgc caggaccgt ccgcctccg cgagtcgcc	4920
accggcgtct tcgtgggcgc cggcccgaac gagtacgcc agcgcgtcca ggagctcgc	4980
gacgaggccg caggcctcta ctccggcacc ggcaacatgc tctccgtcgc cgccggccgc	5040
ctctcttct tctcggcct ccatggccc accctcgcc tcgataccgc ctgctctcc	5100
tccctcgtcg cctccacct cggctgtcag tccctccgc gcggcgagt cgaccaggcc	5160
ctcgtcggcg gcgtcaacat gctcctctcc ccgaagacct tcgcctcct ctcccgcag	5220
cacgccctct cccggggcgg ccgtcgaag accttctcc ccgacgccga cggtacgcc	5280
cgcccgagg gctcgccgt cgtcgtctc aagcgctct ccgacgcca gcgcgaccgc	5340
gaccgatcc tcgcgtcat ccgcggcacc gccatcaacc acgacggccc gtctcggc	5400
ctcaccgtcc cgtcgggccc ggcccaggag gccctctcc gccaggcct cgcccacgcc	5460
ggcgtcgtcc cgcccgacgt cgacttcgtc gagtccacg gcaccggcac cgccctcggc	5520
gaccgatcg aggtccgcgc cctctccgac gtctacggcc agggccgccc ggccgaccgc	5580
ccgtcatcc tcggcgccgc caaagccaac ctcgccaca tggagccgc cgccggcctc	5640
gcccgcctcc tcaaggcgt cctcgccctc ggccaggagc agatccggc ccagccggag	5700
ctcggcgagc tcaaccgct cctccgtgg gaggccctcc cggtcgctt cgcccgccc	5760
gccgtccgt ggcccgcac cgaccgccc cgcttcgccc gcgtctctc cttcggcatg	5820
tccggacca acgcccagc cgtctcgag gaggccccag ccgtcgagct ctggccggcc	5880
gccccggagc gatccgcca gctcctcgt ctctccgca agtcagagg cgccctcgac	5940
gcccaggccg ccgcctccg cgagcacctc gacatgcacc cggagctcgg cctcggcgac	6000
gtcgcttct cctcgccac cacacgtcc gccatgaacc accgcctcg cgtcgccgtc	6060
acctcccgcg agggcctct cgccgcctc tccgctcg cccaggcca gacccgccc	6120
ggcgccgccc gctgcatgc ctctctctc cgccgcaagc tcgcttct cttaccggc	6180
caggcgccc agacccggg catgggcgc ggctctcgc ccgctggcc cgcttccgc	6240
gaggccttcg accgtgct cgccctctc gatcgaggt tggaccgccc gctcccgag	6300
gtcatgtggg ccgagccgg ctctgccgag tccctctcc 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 gcggcccat ggtctcctc ggcccccgg aggcgaggt cgccgccc	6600
gtcgccccgc acgcccctc cgtctccat gccgccgta acggcccga gcaggtcgtc	6660

atcgccggcg tgcagcaggc cgtccaggcc atcgccggcg gcttcgccgc ccgcggcgcc	6720
cgcaccaaac gcctccacgt ctcccacgcc tcccactccc cgctcatgga gccgatgtc	6780
gaggagttag gccgcgtcgc cgcctccgtc acctaccgcc gcccgtcgt ctcctcgtc	6840
tccaacctct ccggcaaggc cgtcgccgac gagctctccg ccccgggcta ctgggtccgc	6900
cacgtccgcg aggcctgtccg cttcgccgac ggcgtaagg ccctccacga ggccggcgcc	6960
ggcaccttag tgcaggtcgg ccgaagccg acctcctcg gcctcctccc ggctgtcctc	7020
ccggaggccg aaccgacct cctcgctcc ctccgcgcc gccgcgagga ggccgccggc	7080
gtcctcgagg 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 gcgaggccgt cgcgcgcc	7380
ctctccacc gccgttgcc gtgcgtcgtc ctccacgcc cgcgagagac ctccgccacc	7440
gccgagctcg tcaccgaggc cgcggcgccg cgtccgact ggaggtcgt cctctacctc	7500
tggggcctcg acgccgtcgt cggcgccgag gcctccatcg acgagatcgg cgcgccacc	7560
cgcgcgccca ccgccccgt cctcggcctc gcccgcttcc tctccacgt ctctgtctcc	7620
ccgcgcctct gggtcgtcac ccgcggcgcc tgcctcgtcg gcgacgagcc ggcaatgcc	7680
ccctgccagg ccgcctctg gggcatgggc cgcgtcgccg ccctcgagca cccggcgca	7740
tggggcgcc tcgtcacct cgaaccgcgc gcctccccgc cgcaggcctc cccgattgac	7800
ggcgagatgc tcgtaccga gctcctctcc caggagaccg aggaccagt cgccttcgc	7860
cacggccgcc gccacgcgc ccgcctcgtc gccccccgc cgcagggccca ggccgccccg	7920
gtctccctct ccgcgaggc ctctacctc gtaccggcg gcctcgccgg cctcggcctc	7980
atcgtcgccc agtggctcgt cgagctcggc gcccgccacc tcgtcctcac ctcccgccgc	8040
ggcctcccgg accgccaggc ctggtgcgag cagcagccgc cggagatccg cgcgcgcatc	8100
gcccgctcg aggcctcga ggcccgccgc gcccgctca ccgtcgccgc cgtcgacgtc	8160
gccgagctcg agccgatgac cgcctcgtc tctccgtgg agccgccgt ccgcgccgtc	8220
gtccacgccg ccggcgtctc cgtcatgcgc ccgtcgccg agaccgacga gacctctc	8280
gagtccttcc tcccccga ggtcgccggc tctggctcc tccaccgct cctccacggc	8340
cgcgcgtcg acctcttct cctcttctcc tccggcgccg ccgtctgggg ctcccactcc	8400
cagggcgct acgcccgcc caacgcctc ctgacggcc tcgccacct ccgcgcctcc	8460
cagtcctcc ccgccctct cgtcgctgg gccctctggg ccgaggcgcc catggccgac	8520

gccgaggccc acgcccgtt gtccgacatc ggctctctcc cgatgtccac ctccgcgcc	8580
ctctccgccc tccagcgctt cgtcgagacc ggctctgccc agcgaccgt caccgcatg	8640
gactgggccc gtttcgccc ggtctacacc gcccgcgccc gcccaacct cctctccgc	8700
ctctcgccg gcccgacat catcgccccg tccccgcgg ccgcccac ccgcaactgg	8760
cgcggcctct ccgtcgcca ggcccgcgtc gccctccac agatcgcca cggcgccgtc	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 catcgtcggc gccgcctgcc gcttcccggg cggcgtcgag	9120
gacctcgagt cctactggca gctcctgcc gaggcgctcg tcgtatccgc cgaggtcccg	9180
gccgaccgtt gggacgccgc cgactgttac 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 gcgcgggccc gaacgagtac tacaccagc gcctccgtgg cttcaccgac	9480
ggcgccgccc gcctctacgg cggcaccggc aacatgctt ccgtaccgc cggacgcctc	9540
tccttcttcc tcggcctcca cggcccgcac ctcccatgg acaccgctg ctctctctcc	9600
ctctctgac tccacctgc ctgccagagt ctccgctcg gcgagtgca ccaggccctc	9660
gtcggcgccg tcaactctt cctcgcccc gagaccttcg tctctcttc ccgcatgcgc	9720
gccctctccc cggacggccg ctgcaagacc ttctccgccc acgccgacgg ctacggccgc	9780
ggcgagggct gcgcgctgt cgtctcaag cgcctccgc acgccagcg cgcggcgac	9840
tccatctctg ccctcatccg cggctccgcc gtcaaccac acggcccgtc ctccggtctc	9900
accgtcccga acggcccgcc ccagcaggcc ctctccgcc aggcctctc ccaggccggc	9960
gtctccccgg tcgacgtcga cttctcgag tgccacggca ccggcaccgc cctcgcgac	10020
ccgatcgagg tccaggccct ctccgaggtc tacggcccgg gccgtccgg cgaccggccg	10080
ctctctctg gcgcggccaa ggccaacgtc gccacctcg aggcgcctc cggcctcgcc	10140
tcctctctca aggcgctct cgcctccgc cagagcaga tcccgccca gccggagctc	10200
ggcgagctca acccgacct cccgtggaac accctccgg tcgctctcc gcgcaaggcc	10260

gtcccgtggg gccgcggcgc ccgcccgcgc cgcgccgcgc tctccgcctt cggcctctcc	10320
ggcaccaacg tccacgtcgt cctcgaggag gccccggagg tcgagccggc cccggccgcc	10380
ccggccccgc cggtcgagct cgtcgtcttc tccccaagt ccgcccgcgc cctcgacgcc	10440
gccgccccgc gcctctccgc ccacctctcc gccaccgg agctctccct cggcgacgtc	10500
gccttctccc tcgccaccac ccgctccccg atggagcacc gcctcgccat cgccaccacc	10560
tcccgcgagg cctccgcgg cgcctcgac gccgcccc 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 gcccgctcg accagaccgc ctacgccag	10860
ccggccctct tcgcctcga gtacgccctc gccgccctct ggcgctcttg gggcgtcgag	10920
ccgcacgtcc tcctcgcca ttcaatcggc gagctcgtcg ccgctgtgt cgcggcgctc	10980
ttctccctcg aggacgccgt ccgcctcgtc gccgcccgcg gccgctcat gcaggccctc	11040
ccggccggcg gcgccatggt cgccatgcc gcctccgagg ccgaggtcgc cgcctcgtc	11100
gccccgcacg ccgccaccgt ctccatgcc gccgtcaacg gcccggaacg cgtcgtcatc	11160
gccggcgccg aggtccaggt cttggccctc ggcgccacct tcgccgcccg 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 ctctcggc tcctcccggc ctgcctcggc	11520
gaggccgacg ccgtcctcgt ccgctccctc cgcgccgacc gctccgagtg tgaagtcgtc	11580
ctcgccgccc tcggcgctg gtacgcctgg ggccggcgccc tcgactggaa gggcgctctc	11640
cctgacggcg cccgccgct cgcctcccg atgtaccctt ggacgcgca gcgccactgg	11700
atggacctca cccgcgctc cgcgccccg gccggcatcg ccggccgctg gccgtcgc	11760
ggagtccgcc tctgcatgcc gggcgccgtc ctccaccacg tccttccat cggccccgc	11820
caccagccgt tcctcggcga tcacctgtc ttccggcaagg tcgtcgtccc gggcgcttc	11880
cacgtcgcg tcactcttc catcgccgc gagcgtggc cggagcgcgc catcgagctc	11940
accggcgctc agttcctcaa ggccatagcc atggagccgg accaggaggt cgagctccac	12000
gccgtcctca ccccgaggc cgcggcgac ggtacctct tcgagctcgc caccctgcc	12060
gccccggaga ccgagcgccg ctggaccacc cacgcccgcg 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 tcctcgacaa cggtctcgcc	12360
gtctccctcc tctccaccg ctcgagccg gaggacgacg gcaccccgcc gctcccgttc	12420
gccgtcgagc gcgtccgctg gtggcgcgcc ccggtcggcc gcgtccgctg cggcggcgctc	12480
ccgcgtctcc aggccttcgg cgtctctctc ttctctctcg tcgacgagac cggcgaggtc	12540
gtcgcggagg tcgagggctt cgtctgccgc cgcgtccgc 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 ctccccggcc ggcgctcatct gcctctggga ggccggcgcc	12840
cacgaggagg ccccgccgc cgcccagcgc gtcgcaaccg agggcctctc cgtcgtccag	12900
gccctccgcg accgcgccgt ccgcctctgg tgggtcacca tgggcgccgt cgcgctcgag	12960
gccggcgagc gcgtccaggc cgccaccgcc cctgtctggg gcctcggccg caccgtcatg	13020
caggagcgcc cggagctctc ctgcacctc gtcgacctcg agccggaggc cgacgcccgc	13080
cgtccgccc acgtctctct ccgcgagctc ggccgcgccc acgacgaaac ccaggtagcc	13140
ttccgctccg gcaagcgccc cgtcgcccgc ctctcaagg ccaccactcc ggagggcctc	13200
ctcgtcccg acgccgagtc ctaccgctc gaggccggcc agaagggcac cctcgaccag	13260
ctccgctcg ccccgcccca gcgcgcgcc cccggcccgg gcgaggtcga gatcaaggtc	13320
accgctccg gcctcaactt ccgaaccgtc ctgccgtcc tcggcatgta cccggcgac	13380
gccggcccga tggcgccga ctgcgccgc gtcgccaccg ccgtcggcca gggcgtccgc	13440
catgttgccg tcggcgagc cgtcatgact ctggcacc cccaccgctt cgtcaccgtc	13500
gacccccgc tcgtcgtccg ccagccggcc ggcctcacc cggcccaggc cgccaccgtc	13560
ccggtcgcct tcctaccgc ctggctcgcc ctccacgacc tcggaaacct ccgcccggc	13620
gagcgctcc tcatccacgc cgccgccgc ggcgtcggca tggcagccgt ccagatcgcc	13680
cgctggatcg gcgccgaggt cttcgccacc gcttccccgt ccaagtgggc agcgtccag	13740
gccatggcg tccccgcac ccacatgcc tcctcccga ccttcgagtt cgccgagacc	13800
ttccgccaag tcaccggcgg ccgcggcgtc gacgtcgtcc tcaacgcct cgccggcgag	13860
ttcgtcgacg cctcctctc cctctctcc accggcgac gcttctcga gatgggcaag	13920
accgacatcc gcgaccgcgc cgccgtcgcc gcccccacc cgggcgtccg ctaccgctc	13980

ttcgacatcc tcgagctcgc cccggaccgc acccgcgaga tcctcgagcg cgtggtcgag	14040
ggcttcgccc ccggccatct ccgtgccctc ccggtccacg ccttcgcat caccaaggcc	14100
gaggccgcct tccgtttcat ggcccaggcc cgcaccagg gcaaggtcgt cctcctccc	14160
gccccgtccg ccgccccgct cgcgccgacc ggcaccgtcc tcctcacgg cggcctcggc	14220
gcccctggcc tccacgtcgc ccgtggctc gcccagcagg gcgtcccgca catggtcctc	14280
acaggccgcc gcggcctcga cactccgggc gccccaagg ccgtcgccga gatcgaggca	14340
ctcgggcgccc gcgtcacat cgcgcctcc gacgtcgccg accgcaacgc cctcgaggcc	14400
gtcctccagg ccatcccggc cgaatggccg ctccaggcg tcattcacgc cgcaggcgcc	14460
ctcgacgacg gcgtcctcga cgagcagacc accgaccgt tctcccgct cctcgcccc	14520
aaggtcaccg gcgcctggaa cctccacgag ctccaccgg gcaacgacct cgccttcttc	14580
gtactcttct cctccatgtc cggcctctc ggctccgccc gccagtcaa ctacgccc	14640
gcaaacacct tcctcgacgc cctcgccc caccgcccg ccgagggcct cgcgcacag	14700
tcctcgctt ggggccccgtg gtccgacggc ggcatggccg ccggcctctc tgcgcctc	14760
caggccccgc tcgcccgcga cggcatgggc gccctctccc cagcccagg caccgcctc	14820
cttggccagg cctcgcccc cccggagacc cagctcgccg ccatgtcct cgacgtccgc	14880
gccgcctccc aggcctccgg cgcgcctgc ccgccgtct ggcgccct cgtccgcgc	14940
gaggccccgc acaccgccgc cggcgcccag ggcgccctcg ccgccgcct cggcgccctc	15000
ccggaggccc gacgcgcga cgaggtccgc aaggctgtcc aggcgcgat cgcgcgctc	15060
ctctcctggt ccgccctc cgcgtcccgt gtcgaccgc cgtctcga tctcgcccta	15120
gactccctca ccgcgtcga gtcgcgaac gtctcgcc agcgcgtcgg cgcaccctc	15180
ccggccacc tcgccttcga ccaccgacc gtcgacgcc tcaccgctg gctcctcgac	15240
aaggctcctg ccgtcgccga gccgtccgtc tcctccgca agtctcccc gcaggtcgcc	15300
ctcgacgagc cgatcgcat catcgcatc ggctgtcgt tcccgggcgg cgtcgccgac	15360
ccggagtctt tctgggcct cctcgaggag ggctccgac ccgtcgtcga ggtccgcac	15420
gagcgtggg acatcgatgc cttctacgac ccggaccgg acgtccgccc caagatgacc	15480
accgcttcg gcgcttct ctcgacatc gaccgttcg acccgccctt cttcgccatc	15540
tccccgcgcg aggcaccac catggaccg cagcagcgc tcctcctcga gacctcctgg	15600
gaggccttcg agcgcgcgg cactctccc gagcgtctca tgggctcga caccggcgtc	15660
ttcgtcgac tcttctacca ggagtacgc gccctcgccg 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 tcaccccgcc caccctcgtc gaggctctcc gcctccgcgg cctcgccccg	15960
gacggccgct gcaagtcctt ctccgccgcc gccgacggcg tcggctggtc cgagggtgctc	16020
gccatgtctc tcctcaagcc gctccgcgac gcccagcgcg 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 cgcgccgggc cgcctcgccg accgccgctt cgtcataggc	16320
tctgtcaagt ccaacatcgg ccataccag gccgccgccc gcgtcgccgg cgtcatcaag	16380
gtagccctcg ccttcgagcg cggcctcatc ccgcgtccc tccacttga cgcgccgaac	16440
ccgcacattc cgtgggtccg gctcgccgtc caggctcgccg ccaagccggt cgagtggacc	16500
cgaacggcg tcccgcgccg cgcggcgctc tctccttcg gcgtctccgg caccaacgcc	16560
cacgtcttcc tcgaggaggc cccggccgcc gccttcgccc cggccgcgcg ccgtccgcc	16620
gagctcttcg tcctctccgc caagtcgcc gccgccctcg acgcacaggc ggcccgcctc	16680
tccgccatg tcgtcgccca cccggagctc ggcttcggcg acctcgctt ctccctcgcc	16740
accaccgct cccgatgac ctaccgctc gccgttgccg ccacctccg cgaggccctc	16800
tccgccgcc tcgacaccgc cgcgccggg caggccccgc cggccgcggc ccgcggccac	16860
gcctccaccg gctccgccc gaaggctgtc ttctgtctcc cgggccaggg ctcccagtgg	16920
ctcggcatgg gccagaagct cctctccgag gagccggtct tccgcgacgc cctctccgcc	16980
tgcgaccgcg ccatccaggc cgaggccggc tggctcttgc tcgccgagct cgcgcggac	17040
gagaccacat cccaactcgg ccgcatcgac gtcgtccagc cggccctctt cgccatcgag	17100
gtcgccctct ccgccctctg gcgtctctgg ggcttcgagc cggacgccgt cgtcgccac	17160
tccatggcg aggtcgccgc cgcacacgtc gccggcgccc tctcctcga agacgccgtc	17220
gccatcatct gccgccgtc cctcctctt cgcgcgatct ccggccaggg tgagatggcc	17280
gtcgtcgagc tctcctcgc cgaggctgaa gccgccctcc tcggctacga ggaccgcctc	17340
tccgtagccg tctccaactc cccgcgtcc accgtcctcg cggcgagcc ggccgcctc	17400
gccgaggtcc tcgccatcct cgcgccaaag ggctcttct gccgccgct caaggtcgac	17460
gtcgccctcc actccccga gatcgaccg ctccgcgacg agctcctgc cgcctcggc	17520
gagctcgaac cgcgccaggc caccgtctcc atgcgtcca ccgtcacctc caccatcatg	17580

gccggcccg agctcgtcgc ctctactgg gccgacaacg tccgccagcc ggtccgcttc	17640
gccgaggccg tccagtcct catggaggac ggccacggcc tcttcgtcga gatgtcccca	17700
cacccgatcc tcaccacctc cgtcgaggag atccgccgcg ccaccaagcg cgaggcgctc	17760
gccgtcgggt cctccgcg cgccaagac gagcgctct ccatgctcga ggccctcggc	17820
gcccctctggg 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 gcctcccgtg gctcggcgac caccgctcc agggcgccgt cgtcttccc	18120
ggcgccgct accctcagat ggccctctcc tccggcgccg aagccctcg cgatggccc	18180
ctccaggct cgcacgtcgt cctcgccgag gccctcgct tcgccgacga cccccggcc	18240
gctgtccagg tcatggccac cgaggagcga ccgggccc tccagttcca cgtcgctcc	18300
agagtcgg gccacggcg cgccgcttc cgtccacg ccgcggcg cctccgccag	18360
atcgagcgc ccgaggtccc ggcccgctc gacctcgag ccctccgc ccgcctccag	18420
gcctccgcc cgcccgccgc cactacgcc gccctcgcc agatgggct cgagtacggc	18480
ccggccttc agggcctcgt cgagctctgg cgccgaggg gggaggcgct tggccgctg	18540
cgctcccg aagccgccc ctcgccgcc gcctgccgcc tccaccggc cctcctcgac	18600
gcctgcttc acgtctctc cgcttcgcc gaccgccc aggccaccc gtgggtccc	18660
gtcgagatcg gctccctcc ctggttcag ccccgctcc gcgagctctg gtgtcacgcc	18720
cgtccgtct cccacggcaa gccgacccc gaccgccc ccaccgact ctgggtcgtc	18780
gactccacc gcgccatcgt cgccgagatc tccggcctc tcgccagcg cctcgccggc	18840
ggcgtccgc gcccgagga ggacgactgg tcatggagc cgccctggga gccgaccgc	18900
gtccgggct ccgaggtcat ggccggcg tggctcctca tcggctcgg cgccggcctc	18960
ggcgccccc tccactccgc cctaccgag gccggccact ccgtagtcca cgcgaccgc	19020
cgcggcacca gcgcccggg cctccaggcc ctctcaccg cctcctcga cggccaggcc	19080
ccgacgtccg tcgtccacct cggtccctc gacgagcgc gcgtcctcga cgccgacgc	19140
ccgttcgatg ccgatccct cgaggagtcc ctgctccgc gctcgactc cgtcctctgg	19200
accgtccagg ccgtcgccgg cgccggcttc cgccaccgc cgccctctg gctcgtcacc	19260
cgcggcgccc aggccatcgg cgccggcgac gtctccgtc ccaggcccc cctcctcggc	19320
ctggccgcg tcatgcct cgagcacgcc gagtccgt gcgccgcat cgacctcgac	19380
ccggcccc gccgacggc ggtcgacgag ctctcgccg agtccctgc cgacgacgc	19440

gaggaggagg tcgccttcg cgccggcgag cgccgcgtcg cccgcctcgt ccgcgcctc	19500
ccagagaccg actgccgcga gaagatcgag ccggccgagg gccgcccggtt ccgcctcgag	19560
atcgacggct ccggcgctct cgacgacctc gtccctccgcg ccaccgagcg ccgcccgccg	19620
ggcccgccgag aggtcgagat tggcgctcgag gccggcgccg tcaacttcct cgacgtcatg	19680
cgcgccatgg gcatctaccc tggggccggc gacggcccg tgcctcctcg cgccgagtgc	19740
tccggccgca tcgtcgccat gggcgaggcg gtcgagtcct tccgcatcgg ccaggacgtc	19800
gtcgccgtcg ccccgcttc cttcggcacc cacgtcacca tcgatcccg catgctcgcc	19860
ccgcgcccg ccgcactcac cgccgccag gccgccgccc tcccggtcgc cttcatgacc	19920
gcctggtacg gcctcgtcca cctcgccgcg ctccgcgcg gcgagcggt cctcatccac	19980
tcagccaccg gcggcacccg cctcgccgcc gtccagatcg ccgccacct cggcgccgag	20040
atcttcgcca ccgccggtac ccgggagaag cgccctggc tccgcgagca gggcatcgcc	20100
cacgtcatgg actccgctc cctcgacttc gccgagcagg tcctcgccgc caccaaggcg	20160
gagggcgctg acgtcgtct caactccctc tccggcgccg ccacgacgc ctcctctcc	20220
acctcgtcc cggacggccg cttcatcgaa ctccgcaaga ccgacatcta cgccgatcgc	20280
tccttgggcc ttgcacactt ccgcaagtcc ctctctact ccgcgtcga cctcgccggc	20340
ctcgccgtcc gccgccgga gcgcgtgcc gccctcctcg ccgaggtcgt cgatctctc	20400
gcccgccgag ccctccagcc gctcccggtc gagatcttc cgctctccg cgccgccgac	20460
gccttcgca agatggccca ggcccagcac ctccgcaagc tcgtcctcgc cctcgaggac	20520
ccggacgtcc gcatccggt cccggcgag tccggcgctg ccacccgcgc cgacggcgcc	20580
tacctcgtca ccggcgccct cggcgccctc ggctctcag tcggcggtg gctcgccgag	20640
cagggcgccg gccacctcgt cctcgtcggc cgctccggcg ccgtctcgc cgagcagcag	20700
accgcccgtg ccgcctcga ggcccacggc gcccgctca ccgtcgccg cgccgacgtc	20760
gccgaccgag ccagatgga gcgcatectc cgcgaggta ccgcctcgg catgcccctc	20820
cgccggcgctg tccacgccg cggcatectc gacgacggac tcctcatgca gcagacccg	20880
gccgcttcc gcgcgtcat ggccccgaag gtccggcgcg ccctccacct ccacgcctc	20940
accgcgaag ccccgctct cttcttcgtc ctctacggt ccggcgccgg cctcctcggc	21000
tccccgggcc agggcaacta cgccgccgcc aacaccttc tcgagccct cgccaccac	21060
cgccgcgccc agggcctccc ggccctctcc atcgactggg gcctcttcgc cgacgtcggc	21120
ctcgccgccc gccagcagaa ccggggcgcc gcctcgtca ccgcggcac 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
gaggtcgctc gcgccaggt ctcccaggtc ctccgcctct ccgagggcaa gctcgacgtc	21480
gacggccgc tcacctcct cggcatggac tccctcatgg gcctcgagct ccgcaaccgc	21540
atcgaggcgc tctcggcat caccatgcc gccaccctcc tctggacctc cccgaccgtc	21600
gcccctct cgcgccct cgcctccac gtcgtctcca ccggcgacgc cgagtcgcc	21660
cgcggccgc acaccggctc cgtcgcccg accaccacg aggtcgctc cctcgacgag	21720
gacggcctct tcgcctcat cgacgagtc ctgcccgc cggcaagcg 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 gcgagggtca gctcctcgag cgtctccgc aggtcacct cgcctccgc	60
aagacctca acgagcgca caccctcgag ctcgagaaga ccgagccgat cgccatctg	120
ggcatcggtt gccgttccc gggcgggcc gccaccccg aggccttctg ggagctctc	180
gacgacggcc gcgacccat ccgccgctc gaagagcgt gggccctct aggcgtcgac	240
ccggcgacg acgtcccgc ctggcgccgc ctctcaccg aggcctatga tggcttcgac	300
gccgcttct tcggcatgc cccgcgcgag gccgctccc tcgaccgca gcaccgctc	360
ctctcgagg tcgctggga gggcttcgaa gacgccgca tcccgccgc ctccctcgtc	420
ggctcccga ccggcgtctt cgtcggcgtc tgcgccacc agtacctca cgcgcgctc	480
gcccaccagc cgcgcgagga acgcgacgcc tactccacca ccggcaacat gctttccatc	540
gccggcgac gcctctcta caccctcggc ctccagggcc cgtgccttac cgtcgatacc	600
gcctgctcct cctccctct cgcctccac ctgcctgcc gtccctccg cgcggcgaa	660
tccgacctcg ccctcgccg cgcgctaac atgtcctct ccccgacac catgcgcgc	720
ctcgccgca cccaggcct ctcccgaac ggccgctgcc agaccttga cgcctccgc	780
aacggcttcg tccggcgca gggtgcggc ctcatcgtc 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 tctcgagggc	1140
ggcccgcgcg tcgccggcct catcaaggcc accctctccc tccaccacga gcgtatccg	1200
cgcaacctca acttcgcac cctcaaccg cgcatccga tcgagggcac cgcctcgcc	1260
ctgccaccg agccggctcc gtggccgcgc accggccga ccgcttcgc cggcgtctcc	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 gcctccggcg gctccgccc gaaggtcgtc	1680
tctgtcttc cgggccaggg ctccagtg gtcggaatgg gccgcaagct catggccgag	1740
gaaccggtct tccgcgcgc cctcgaaggc tgcgaccgc ccatcgaggc agaggccgt	1800
tggtcctcc tcgcgagct ctccgcgc gaggcgcct ccagctcgg ccgcatcgac	1860
gtcgtccagc cggctctct cgccatggag gtcgcctct ccgcctctg gcgtcctgg	1920
ggcgtcgagc cggaggcgt cgtcgccac tccatggcg aggtagccgc cgcacgctc	1980
ggcgcgcg tctcctcga ggacgcgt gccatcatc gccgcgctc ccgctctc	2040
cgcgcgctc ccggccagg cgagatggc ctgctcgagc tctcctcga ggaggccgag	2100
ggccctca gaggtcacga gggcgctc tcggtcgcg tctccaactc cccgcgtcc	2160
actgtctcg ccggcgagcc ggccgcctc tcagaggtcc tcgcccct caccgcaag	2220
ggcgtctct ggccaggt caaggtcgac gtcgcatccc actcccgca ggtcgaccg	2280
ctccgcagg agctcatgc cgcctcggc gccatccgc cgcgcgcgc cgcgtccg	2340
atgcgtcca ctgtaccgg cggcgtcgc gccggccgg agctcggcg ctcctactgg	2400
ggcgacaacc tcccgagcc ggtccgttc gccgcccg cccaggcct cctcgaggc	2460
ggccggccc tctcatcga aatgtcccg 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
gccgcccgcc tctccaccg cggcctctcc tgcaccgtcc tccacgcctc cgtgacgcc	2940
tccaccgtcg ccgagcaggt ctccgaggcc gcctcccgcc gcaacgactg gcaggcgctc	3000
ctctacctct ggggcctcga cgcgtcgtc gaccccgcg cctccgccga cgaggctctcc	3060
gaggccaccc gccgcgccac cgcgccgtc ctcggcctcg tacgtttct ctcgccgcc	3120
ccgacccgc cgcgtttctg ggtcgtcacc cgcggcgctt 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 gcgacgtcgc cccgatctcc	3420
ctctccgccg agggctccta cctcgtcacc ggccggcctcg gcggcctcgg cctcctcgtc	3480
gcccgttggc tcgtcgagcg cggcgccccgc cacctcgtcc tcacctccg ccacggactc	3540
ccggagcgcc aggcctcgg cggcgagcag ccgccggagg cccgcgcccg catcgccgcc	3600
gtcgagggcc tcgaggccca gggcgccccgc gtcaccgtcg ccgccgtcga cgtcgcgag	3660
gccgatccga tgaccgcct cctcgccgcc atcgagccgc cgtccgagg cgtcgtccac	3720
gcccgccgcg tcttccccgt ccgccacctc gccgagaccg acgaggccct cctcgaatct	3780
gtctccgac cgaaggtcgc cggcagttgg ctctccacc gtctcctccg cgaccgaccg	3840
ctcgacctct tcgtcctctt ctctccggc gccccgtctt gggcgggcaa gggccagggc	3900
gcctacgccg ccgccaacgc ctctctgac ggcctcgccc accaccgccg cgccactcg	3960
ctcccgccc tctcctcgc ctggggcctc tgggccgagg gcggcatggt cgacccaag	4020
gcccacgcc gactctcga catcggtc ctcctgatgg ccaccggccc ggccctctcc	4080
gcctcgagc gcctcgtcaa cactccgcc gtccagcgt ccgtcaccg catggactgg	4140
gcccgttctg ccccggtcta cgcgcccg gcgcgccga acctcctctc cgcctcgtc	4200
gctgaggacg agcgcgccgc ctccccgcc gtcccgaccg ccaaccgat ctggcgcggc	4260
ctcagcgtcg ccgagtcgg ctcgccttg tacgagctcg tcccgggcgt cgcgccgtc	4320
ctcggttct ccgaccggg cgcactgac gtccggcgag gcttcgccga gcagggcctc	4380
gactccctca tggccctcga gatccgaac cgcctccagc gcgagctcgg cgagcgctc	4440

tccgccaccc tcgccttcga ccacccgacc gtcgagcgcc tcgtcgccca cctgctcacc	4500
gacgtcttga agctcgagga ccgctccgac acccgccaca tccgctccgt cgccgcccac	4560
gacgacatcg ccacgtcggt cgccgcatgc cgcttcccgg gcggagacga aggcctcgag	4620
acctactggc gccacctcgc cgagggcgat gtcgtctcca ccgaggtccc ggccgaccgc	4680
tgggcgcccg ccgactggta cgacccggac ccggagggtcc cgggccgcac ctacgtcgcc	4740
aagggcgccct tcctccgca cgtccgctcc ctgcagcccg cttcttcgc catctccccg	4800
cgcgaggcca tgtccctcga cccgcagcag cgctctctcc tcgaagtctc ctgggaggcc	4860
atcgagcgcg ccggccagga ccgatggcc cttcgcgagt ccgccaccgg cgtcttcgtc	4920
ggcatgatcg gctccgagca cgccgaacgc gtccagggcc tcgacgacga cgccgccctc	4980
ctctacggca ccacggcaa cctctctca gtcgcccgg gccgcctctc cttcttcctc	5040
ggcctccacg gcccgacat gaccgtgat accgcctgtt cctctcct cgtcgccctc	5100
catctgcct gtcagtcct ccgcctcgga gagtgcgacc aggcctcgc cggcggtcc	5160
tccgtctcc tctccccg ctccttcgtc gccgatccc gcatgcgct cctctccccg	5220
gacggccgt gcaagacct ctcgccgcc gccgacgat tcgcccgcgc cgagggtgc	5280
gccgtcgtc tctcaagcg tctccgac gccacgcgc acccgaccc gatcctcgcc	5340
gtcgtccgt ccacgccat caaccacgac ggcccgtct ccggcctcac cgtcccgctc	5400
ggcccggccc agcaggccct cctccgcag gccctcgccc aggcggcggt cgcccggcc	5460
gaggtcgact tcgtcgagt ccacggcacc ggacccgcc tcggcgaccc gatcgaggtc	5520
caggccctcg gcgccgtcta cggccgggc cgcccggccg agcggccgt ctggctcggt	5580
gctgtaaagg ccaacctcgg ccacctcgag gccgccgcc gccctgccgg cgtctcaag	5640
gtctctctcg cctcgagca cgagcagatc ccggctcaac cggagctcga cgagctcaac	5700
ccgcacatcc cgtgggcca gctcccgtc gccgtcgtc gccgcgcgt gccgtggccg	5760
cgccggcccc gcccgcccg cgccggcgtc tccgccttcg gcctcagtgg caccaacgcc	5820
cacgtcgtc tcgaggagg cccggccgtc gagccggtc ccgccgtcc ggagcgcgcc	5880
gccgagctct tcgtctctc cgccaagtcc gccgccgcc tcgacgcca ggccgccgc	5940
ctccgcgacc acctcgagaa gcacgtcgag ctccgctcg gcacgtcgc cttctccctc	6000
gccaccaccc gctccgcat ggagcacga ctcccgctc cgccctctc ccgcgaggcc	6060
ttgcgcggcg cctctccgc cgccgccag ggccacaccc cgccggcgcc cgtccgcggc	6120
cgccctccg gcggctccgc cccgaaggtc gtcttcgtct tcccgggcca gggctcccag	6180
tgggtcgga tgggcccga gctcatggc gaggagccgg tcttccgcgc cgccctcgag	6240
ggctgcgacc gcgccatcga ggccgaagcg ggctgggtcc tcctcggcga gctctccgc	6300

gacgaggccg cctcccagct cggacgcac gacgtcgtcc agccggctcct cttcgccatg	6360
gaagtcgccc tctccgccct ctggcgctcc tggggcgctcg aaccagaggc cgtcgtcggc	6420
cactccatgg gcgaggtcgc cgcgcgccac gtcgccggcg cctctcctct cgaggacgcc	6480
gtcgccatca tctgccgccc ctcccgcttc ctcgccgcca tctccggaca gggagagatg	6540
gccctcgtcg agctctcctt cgaggaggcc gagggcgctc tccgcggcca cgagggccgc	6600
ctctccgtcg ccgtctccaa ctccccgcgc tccaccgtcc tcgccggcga gccggccgcc	6660
ctctccgagg tctcgcgcgc cctcaccgcc aaggcgctct tctggcgcca ggtcaaggtc	6720
gacgtcgctt cccactcccc gcaggtcgac ccgtcccgcg aggagctcat cgcgccttt	6780
ggcgccatcc gcccgcgcg cgcgcgcgtc ccgatgcgtt ccaccgtcac cggcggcgtc	6840
atcgctggcc cggagctcgg cgcctcctac tgggcccaca acctccgcca gccggtccgc	6900
ttcgccgccc ccgcccaggc cctcctcgag ggcggcccgg cctctttcat cgagatgtcc	6960
ccgcaccga tctcgtccc gccgctcgac gagatccaga cggccgcga gcaggcgga	7020
gcccgctcg gctcctccg gcgcggccag gacgaacgc ccacctctt cgaagccctc	7080
ggcaccctgt gggcctccgg ctaccgggtc tctggggccc gcctcttccc ggccggcggc	7140
cgcgcgtcc cgctccccac ctaccgtgg cagcacgagc gctactggat cgaggactcc	7200
gtccacggct ctaagccgtc cctccgcttc cgcagctcc gcaacggcg caccgaccac	7260
ccgtcctcg gcgccccgt cctcgtctcc gcccgcccgg gcgcccact ctgggaacag	7320
gccctcagcg acgagctct ctcctacctc tccgagcacc gcgtccacgg cgaggccgtc	7380
ctcccgctcg ccgctacgt cgagatggcc ctcgccgccc gcgtcgacct ctacggcacc	7440
gccacctcg tgtcagca gctcgccctc gagcgcgctt tggccgtccc gtccgagggc	7500
ggtcgcatcg tccaggtcgc cctctccgag gagggcccgg gcccgccctc cttccaggtc	7560
tctccccgcg aagaggcagg ccgtcctgg gttcgccacg ccaccggcca cgtctgctcc	7620
ggccagtcct ccgcgctcgg cgcctcaag gagccccgt gggagatcca gcgcgctgc	7680
ccgtccgtcc tctctccga ggccctctac ccgtcctca acgagcacgc cctcgactac	7740
ggcccgctgt tccaggcgct cgaacaggtc tggctcgga ccggcgaggt cctcgccgc	7800
gtccggctcc caggcgacat ggctcctcc tccggcgct accgatcca cccggccctc	7860
ctcgagcctt gcttcagggt cctcaccgcc ctcctacca cccccgaatc catcgagatc	7920
cgcgcgcccc tcaccgaact ccacgagcgg gacctccgc gctcccgcgc cccggtcaac	7980
caggccgtct ccgacacctg gctctgggac gccgccctgg acggcggccg ccgccaatcc	8040

gcctccgtcc cggtcgacct cgtcctcggc tccttccacg ccaagtggga ggttatggag	8100
cgctcgcgcc aggcctacat catcggcacc ctccgcatct ggaacgtctt ctgcgcgcc	8160
ggcgagcgcc acaccatcga cgagctcttc gtcgggtcc agatctccgt cgtctaccgc	8220
aaggtcatca agcgtggat ggagcacctc gtcgcatatg gcatcctcgt cggcgacggc	8280
gagcacttcg tctcctccca gccgctcccg gagccggacc tcgcccgt cctcgaggag	8340
gccggccgcg tcttcgccga cctcccggtc ctcttcgagt ggtgcaagtt cgcggcgag	8400
cgctcgcgcg acgtcctcac cggcaagacc ctgcccctcg agatcctctt cccgggcggc	8460
tccttcgata tggccgagcg catctaccgc gattccccga tcgccgcta ctccaacggc	8520
atcgtacgcg gcgtcgtcga gtcgcagcc cgcgtcgtcg ccccgagcgg catgttctcc	8580
atcctcgaga tcggcgccgg caccggcgcc accaccgccc cgtcctccc ggtcctcttc	8640
ccggaccgca ccgagtacca cttaccgac gtctccccgc tcttctcgc cgcgcagag	8700
cagcgcttcc gcgactacc gttctcaaa tacggcatcc tcgacgtcga ccaggaaccg	8760
gccggacagg gctacgcaca ccagcgttc gacgtcatcg tagccgcaa cgtcatccac	8820
gccacccgcg acatccgcgc caccgccaag cgcctgctct 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 cagtcatcc tctcccgcgc cccgggcatc	9120
gccggcgccg cctgcgactc ctccggcgag tccgccaccg agagcccggc cgcgcgcgc	9180
gtccgacagg agtgggcca cggtccgcc gacgtcgtgc accgcatggc cctcgagcgc	9240
atgtacttcc atcgccgtcc gggccgccag gtctgggtcc acggccgct ccgcaccggt	9300
ggcggcgcct tcaccaaggc cctcgccggc gacctcctcc tcttcagga caccggacag	9360
gtcgtcgcgc aggtccaggg cctccgcctc ccgagctcg aggcctcgc cttcgccccg	9420
cgcgaccgcg gcgaggagt gctctacga ctcgagtggc agcgcaagga cccgatcccg	9480
gaggccccgg ccgcccctc gtctctctcc gccggcgctt ggctcgtcct catggaccag	9540
ggcggcaccg gcgcccctt cgtatcctc ctcgagggcc gcggcgaggc ctgctccgc	9600
gtcatcgccg gcaccgcta cgctgcctc gccccgggc tctaccaggt cgaccggcc	9660
cagccggacg gcttccacac cctctcaga gacgccttcg gcgaagaccg catctgccgc	9720
gccgtcgtcc acatgtggtc cctcgacgcc accccgccc gcgagcgcgc caccgccgag	9780
tcctccagg ccgaccagct cctcggtcc ctctccgcc tctcctcgt ccaggccctc	9840
gtccgccgcc gctggcgcaa catccgcgc ctctggctcc tcacccgcgc 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 cgcctcgtg atggtcatgc agggcgcccc ccacgtctc	10260
ctcgtcgacc gcggcgccgc ctccgagcc tcccgcgacg ccctccgag tatggccgag	10320
gccggagccg aagtccagat agtcgaagcc gacgtcgccc gcccgacga cgtggccccg	10380
ctctctcca agatcgaacc gtccatgccg ccgtccgcg gcatcgtcta cgtcgacggc	10440
acctccagg gcgactctc catgctcgag ctgacgccc gccgcttcaa ggagtggatg	10500
tatccgaagg tctcggcg ctggaacctc cagccctca cccgcgaccg ttcctcgac	10560
ttctctgtec tctactctc cggcacctcc ctctcggcc tcccgggcca gggctccgc	10620
gcccgccgac acgttttct cgacgccatc gccaccacc gctcaaggt cggcctcacc	10680
gccaatgtcca tcaactgggg cctctctcc gaggcctcct ccccgccac cccgaacgac	10740
ggcgccgccc gcctcgagta ccgcgccatg gagggcctca ccctcgagca gggcgccgcc	10800
gccctcggac gccctctgc ccgcccgcg gcacaggtcg gcgtcatgcg cctcaatttg	10860
cgtcagtgga tcgagttcta cccgaacgt gcccgccctg ccctctgggc cgagctctc	10920
aaggaaacgc accgcgccga ccgcgccgcc tccaacgct ccaacctcg cgaggccctc	10980
cagtcgccc gcccgagga ccgccaactc atctcgaga agcacctctc cgagctctc	11040
ggcgccggcc tcgcctccc gccggagcgc atcgagcgtc acgtccggt ctccaacctc	11100
ggcatggact ccctcatcgg cctcgagctc cgcaaccgca tcgaggccgc tctcgccatc	11160
accgtcccg ccacctctc ttggacctac cctaactcg ccgcctctc cggtctctc	11220
ctcgacatcc tcttccgaa cgcggcgcc acccacgccc cggccaccga gcgcgagaag	11280
tccttcgaaa acgacgccgc cgacctgaa gccctccgc gcatgaccga cgagcagaag	11340
gacgccctcc tcgccgagaa gctcgcccag 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 gagccgtgg acggattcga cgccgccttc	300
ttcggcacct ccccgccga gggccgtcc ctgaccgcgc 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 gcgaggcct ccagtccgcc	960
cgcgtcgacg ccggcgccat cggctacgtc gagaccacg gcactggcac ttcgctcggc	1020
gaccgatcgc aggtcgaggc cctccgcgcc gtctcggcc cgcccgcgc cgatggctcc	1080
cgtcgtctcc tcggcgccgt gaagaccaac ctgggccacc tcgagggcgc cgccggcgtc	1140
gcgggcctca tcaaggccgc cctcgccctc caccacgaac tcatcccgcg caacctccac	1200
tccacaccc tcaaccgcg catccgcatc gagggcacgc cctcgcct cgccaccgag	1260
ccggtcccggt ggccgcgcgc cggccgccgc cgcttcgcgc gctctcgc attcggcctc	1320
tccggcacca acgtccacgt cgtcctcgag gagggcccg ccaccgtcct cgccccggcc	1380
acccggggcc gctccgccga gctcctcgtc ctctccgcca agtcgccgc cgccctcgac	1440
gcccaggccg cagcctctc cgacacatc gccgcctacc cggagcaggg cctcggcgac	1500
gtcgccttct ccttggcttc caccgctcc ccgatggaac atcgctcgc cgtcgccgcc	1560
accagccgcg aggcctccg ctccgccctc gaggtcgccg cccaggcca gacccggcc	1620
ggtgccgccc gcggccgcgc cgctcctcc ccgggcaagc tcgccttct cttcgccgga	1680
cagggggccc aagtccggg catgggcgc ggctctggg aggcctggcc ggccttcgc	1740
gaaaccttcg accgtgcgt caccctctc gaccgcgagc tccaccagcc gctctcgag	1800

gtcatgtggg ccgaaccggg ctctcccg ctcctccctc tcgaccagac cgccttcacc	1860
cagccggccc tcttcgcct cgagtacgc ctcgccccc tcttcgctc ctggggcgtc	1920
gagccggagc tcgtcgccgg ccactccctt ggagaactcg tcgccgctg cgtcgccggc	1980
gtttttctccc tcgaggatgc cgtccgctc gtctcgccc gcggccgctt catgcaggcc	2040
ctcccgccg gcggcgccat ggtctccata gcggcgccgg aagctgacgt agcccgccg	2100
gtcggccgc acgcccctt ggtctccatc gccgccgtca acggcccgga gcaggtcgtc	2160
atcgccggcg ccgagaagt cgtccagcag atcgccggc ctttcgccc cgcggcgcc	2220
cgcaccaagc cgtccacgt gtccacgcc ttccactccc cgtcatgga cccgatgtc	2280
gaggccttca gacggtcac cgaatccgta acctaccgc gaccgtcat cgcgtcgtc	2340
tccaacctct ccgcaagcc gtgcaccgat gaggtctccg cccccggcta ctgggtccgc	2400
cacggcccg aggcggtccg cttcgccgac ggctcaagg ccctccacgc cgcggcgcc	2460
ggcctcttcg tcgaggtcgg ccgaagccg acctcctcg gcctcgtccc ggctgcctc	2520
ccggacgccc gcccggtcct cctccggcc tcccgcccg gccgcgacga ggccgctcc	2580
gccctcgaag ccctcgccgg attctgggtc gtccggcgct cggtcacctg gagcgagtc	2640
tccccctcg gcggccggc cgtccgctc ccgacctacc cgtggcaacg cgagcgctac	2700
tggatcgagg ccccggtcga ccgcgaggcc gacggcaccg gcccgcccg cgcggcgcc	2760
caccgctcc tcggcgaggt cttctccgtc tccaccacg ccggcctccg cctctgggag	2820
acgacctcg accgaagcg cctccgtgg ctcggcgagc accgcgcca gggtaggtt	2880
gttttcccg gcgccggcta cctcgagatg gccctcagct ccggcgccga gatcctcggc	2940
gacggccga tccaggtcac cgacgtcgtc ctcatcgaga ccctcacctt cgcggcgac	3000
acagccgtcc cggtcaggt cgtcaccacc gaggagcgcc cgggccgact ccgttccag	3060
gtcgctccc gcgagccggg cgagcgccgc gcccgttcc gcatccacgc ccgcgcgctc	3120
ctccgccga tcggccgctg tgagaccgg gcccgtcca acctcgccgc cctccgcgc	3180
cgcctccacg ccgccgtccc ggccgccgc atctacggcg ccctcgccga gatgggcctc	3240
cagtaggcc cggcccttcg cggcctcgcc gagctctggc gcggcgaggg cgaggccctc	3300
ggccgctcc gtctcccga ggcagccggc tccgccaccg cctaccagct ccaccggtc	3360
ctctcgacg cctcggtcca gatgatcgtc ggcccttcg ccgaccgca cgaggccacc	3420
ccttgggccc cggtagggt cggtccgtc cgcctcttcc agcgagccc gggcgagctc	3480
tggtgccacg cccgctcgt ctcgacggc cagcaggcct cctccaggtg gtccgccgac	3540

ttcgagctca tggacggaac cggcgccgtc gtcgccgaga tctccgcct cgtcgtcgag	3600
cgcctcgct caggagtcg cgcgcgcgac gccgacgact ggttcctcga actcgactgg	3660
gagccggccg cctcggcgg cccgaagatc accgccggtc gctggctcct cctcggcgag	3720
ggcggcgccg tcggcgctc cctctgctcc gccctcaagg ccgccggcca tgtcgtcgtc	3780
cacgccgcg gcgacgacac ctccaccgcc ggcatgcgcg ccctcctcgc caacgccttc	3840
gacggacagg ccccgaccgc agtcgtccac ctctcctccc tcgacggcgg cggacagctc	3900
ggccccggcc tcggcgccca gggcgccctc gacccccga ggtccccgga cgtcgacgcc	3960
gacgtctctg agtcgcctt catgaggggc tgcgactccg tctctctct cgtccaggcg	4020
ctcgtcgga tggacctcg caacgccccg cgcctctggc tctcaccgc cggcgcccag	4080
gccgccgcg ccggcgacgt ctccgtcgtc caggccccgc tctcggcct cggccgcacg	4140
atcgccctcg agcacgccga actcagggtg atctccgtcg acctcgacc gcccgagccg	4200
gagggcgag cgcacgcgt cctcgccgag ctctcgcgc acgacgtga agaggaggtc	4260
gccctcgcg gcggcgaccg cctcgtgcc cgcctcgtc accgcctccc ggacgcccag	4320
cgcgcgaga aggtcgagcc ggccggcgac cgcctgttc gcctcgagat cgacgagccg	4380
ggcgccctcg accagctcgt cctcgcgcc acggggcgcc gagccccggg cccgggcgag	4440
gtcgagatct ccgtcgaggc cgcggcctc gactccatcg acatccagct cgcctcggc	4500
gtcgccccga acgacctccc gggcgaggag atcgagccgc tcgtcttggg ctccgagtgc	4560
gccggccgca tcgtcgccgt cggcgagggc gtcaacggcc tcgtcgtcgg ccagcctgtc	4620
atcgccctcg ccgccggcgt cttcgtacc cagtcacca cctccgccac cctcgtctc	4680
ccgcgccgc tcggcctctc cgccaggag gccgccgcca tgcgctcgc ctacctacc	4740
gcctggtacg cctcgacaa ggtcgccac ctccaggccg gcgagcgcgt cctcatcac	4800
gctgaggccg gcggcgtcgg cctctgcgcc gtccgctggg cccagcgcgt cggcgccgag	4860
gtctaccca ccgccgacac cccgaaaac cgcgcctacc tcgagtcct cggcgtccgc	4920
tacgtctccg actccgctc cgggcgcttc gtaccgacg tccacgcctg gaccgacggc	4980
gagggcgtcg acgtcgtcct cgaactcctc tccggtgagc gcatcgacaa gtcctcatg	5040
gtcctccgcg cctcgggcgg cctcgtcaag ctggcgccc gcgacgactg cgccgacacc	5100
cagccgggac tcccgcgct cctccgcaac ttctctttt cccaggtcga cctccgaggc	5160
atgatgctcg accagccggc ccgcatccgc gccctcctcg acgagctctt cggectcgtc	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 gccctcgggt cgccgggtgg ctgcccgagc gcggcgccgg ccagctcgtc	5520
ctcgtgggcc gctccggcgc cgctccgcc gagcagcgcg ccgccgtcgc cgccctcgag	5580
gcccacggcg ccccggtcac cgtcgccaag gccgacgtcg ccgaccgtc ccagatcgag	5640
cgcgtcctcc gcgagggtcac cgctccggc atgccgtcc gcggcgtcgt ccacgccgcc	5700
ggcctcgtcg acgacggcct cctcatgcag cagaccccgg cccgttccg caccgtaatg	5760
ggccccaagg tccagggcgc cctccacctc cacaccctca cccgcgaggc cccgtctcc	5820
ttcttcgtcc tctacgcctc cgccgtggc ctttccggtt ccccgggcca gggaaactac	5880
gccgccgcca acgccttctc cgacgcctc agtcaccacc gccgcgcca gggcctcccg	5940
gccctctcca tcgactgggg catgttcacg gaggtcgga tggccgtcgc ccaggaaaac	6000
cgcgggtccc gccagatctc ccgcggcatg cgcggcatca cgcggatga aggcctctcc	6060
gcccctcgcc gcctcctcga gggcgaccgc gtccagaccg gcgtaatccc gatcaccccg	6120
cgccagtggg tcgagttcta cccggccacc gccgccagcc gccgcctctc ccgcctcgtc	6180
acgacgcagc gcgcggtcgc cgaccggacc gccggcgacc gcgacctct cgagcagctc	6240
gcttcgccc 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
tccgccggt ccttcgtcca cgtctcgcg ttccgtcccg tcgtcaagcc gcgcgcccgg	6600
ctcttctgct tccatggctc cggcggctcc ccggagggt tccgtctctg gtccgagaag	6660
tccgagtgg cgcacctga gatcgtgcc atgtggcacg accgtcctc cgctccgag	6720
gacgccccg gcaagaagta cgtccaggaa gccgcctccc tcattcagca ctacccgac	6780
gccccgttcg cctcgtcgg cttctccctc ggcttcgct tcgtcatggg caccgccgtc	6840
gagctcgct cccgcagcgg cgccccggc ccgtcgcgcg tcttcacct cggcggttcg	6900
ctcatctct cctccgagat caccgggag atggaaaccg atatcatgc caagctcttc	6960
ttccgcaacg ccgcgggttt cgtccgatcc acccagcagg tccaggccga cgcgccgcc	7020
gacaaggtca tcaccgacac catggtgcc ccggccccgg gcgactcaa ggaaccgccg	7080
gtcaagatcg ccgtcccaat cgtcgccatc gccggctccg 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 acgccagga ccccttccc gccatcgagc ggctccgca ggccacccg 120
atcttctact gggacgaggg ccggagctgg gtgtcaccc gctaccacga cgtctccgc 180
gtgttcaggg acgagcgctt cgccgtctcc cgggaggagt gggaatctc ggccgagtac 240

agctccgcca tcccgagct ctccgacatg aagaagtacg ggctcttcgg cctgcccccc 300
gaggaccacg cccgcgtccg caagctcgtg aaccgctct tcactcgcg 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 cgtcccgag gtcgacgagg agaccaagac cctcgtcgcc 600
tcggtcacg agggcctcgc cctcctgcac gacgtcctgg acgaacgcc cgcaaccg 660

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

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

cccgccttcc ggaacatcga gtcctgaac 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 ggtcctctc tcgctcgtgg ccctggccct cgtgctgctc 180
accgcccgtg cgttcggcga actcgcccgg cgcctccgcc agcccagggt cctcggcgag 240
ctcttcggcg gcgtcgtcct ggggtcccage gtggtcgggg cgctcgcgcc gggcttccac 300
cgggtcctct tccaggacce cgcggtcggc gtcgtcctct ccggcatctc ctggatcggg 360
gccctcgtgc tgcctcatat ggcgggcatc gaggtggacg tctccatcct ccgcaaggag 420

gcccggccgg gcgcctctc 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 tcatcgagcg cgagtcacatg	480
cgccgctcct acgcccaggt gaccctcgcc gcgggctcgt tctccgaggt cgccgctgg	540
gtctctcgtc ccatgacctc cagctcctac ggggcctccc cggccctcgc cgtcgccgc	600
tccgcccctc tcgctccggg ctctctctc ttcatgggcc tctcgggtcg ccgctcacg	660
cacctggcca tgcgtgggt cgccgacgcc acccgctct ccaaggcca ggtctccctc	720
gtctctgtcc tcacgttct cgccgccgcc ctacccagc gcctcgggtt ccaccgctg	780
ctcggggcct tcgcccctcg ggtctctc aactccgac cgcgaccaa ccgcccgtc	840
ctcgacggcg tccagacct cgtggccggt ctcttcgccc ccgtttctt cgtgctcgcc	900
ggaatgaggg tcgacgtcag ccagctccgc acgccggccg cctggggcac cgtcgccctc	960
ctctcgcca ccgccaccgc cgccaaggtc gtcccgccg ccctcggcgc ccgtctcggc	1020
ggcctccgcg gctcggaggc cgccctcgt gccgtcggc tcaacatgaa gggcggcacc	1080
gacctcatc tcgcatcgt cggagtcgag ctccgcccct tctccaacga ggcctacac	1140
atgtacgcc tcgtcgccct cgtcacgtc accgctccc ccgcccctt catctggctc	1200
gagaagcgcg ccccgccac ccaggaagag tccgcccgcc tcgagcgga ggaggccgc	1260
cgcgcgctt acatcccggg cgtcgagcgc atctctgtcc ccatcgtcgc ccacgccctc	1320
ccggggttcg ccacggacat cgtcgagtc atcgtcgct ccaagcgga gctcggggag	1380
accgtcgaca tcaccgagct ctccgtcgag cagcaggccc cgggcccgtc ccgcgccgc	1440
ggcgaggcct ccccgccct cgccaggctc ggcgcccggc tccgctcgg catctggcg	1500
cagcgccgag agctccgtg ctccatccag gccatctcc gtgctcccg cgaccagac	1560
ctctctgtca tcggcgcccg ctccccggc cgccccgtg gcatgtctt cggccgctg	1620
caggacgcca tcgtccagcg cgccgaaagc aacgtctcg tcgtcgtcgg cgaccgccc	1680
gccgcccagc gcgcgtccgc ccgcccgc atctgtccga tcacggcct cgagtacagc	1740
ttcgccgccc ccgacctgc cgcacacgt gccctcgct gggacgcgga gctcgtctc	1800
ctctctccg cccagacgga cccgggcgcc gtcgtctggc gcgaccgga gccgtcgcg	1860
gtccgcccgc tggcgcttc cgtcgtcgac gagccgtct tcccggtcg ccgctcggc	1920
gtccgctct ccagccgagt ccacgtcggc gccaccggt ccgacgagat caccgagag	1980
ctgccccg cgccctacga ctgctcgtc ctgggtgct acgaccagc cccctcggc	2040
cgctctacc tcggtccac cgtcgaatcc gtagtcgtc 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 cgagggcgtc 120

cactccctcg tggacaccgc cgacggcctc ctcctcctcc tcggcaagca ccgcagcgcg 180

cgcccgccgg acgccgagca ccccttcggc caccgcaagg agctctactt ctggaccctc 240

atcgtggcca tcatgatctt cgccggccgc ggccggcgtt 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 cgccgctcac 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 gtcctgcccc acgcgccgca ccccgcgcat cccgcgcacc 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	

aagcggcggg gcccgccacc cgggaggggtg gcgggcaccg ctttttcggc ctgctgggtc
taccggacca cctga

60

75