



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0119026
(43) 공개일자 2022년08월26일

- (51) 국제특허분류(Int. Cl.)
 - C12N 9/00 (2006.01) C12N 15/70 (2006.01)
 - C12N 15/77 (2006.01) C12N 15/78 (2006.01)
 - C12N 9/10 (2006.01) C12P 13/10 (2006.01)
 - C12P 7/40 (2006.01)
- (52) CPC특허분류
 - C12N 9/93 (2013.01)
 - C12N 15/70 (2013.01)
- (21) 출원번호 10-2022-7020423
- (22) 출원일자(국제) 2020년12월14일
 심사청구일자 없음
- (85) 번역문제출일자 2022년06월15일
- (86) 국제출원번호 PCT/EP2020/085882
- (87) 국제공개번호 WO 2021/122400
 국제공개일자 2021년06월24일
- (30) 우선권주장
 19218204.6 2019년12월19일
 유럽특허청(EPO)(EP)
- (71) 출원인
 에보닉 오퍼레이션스 게엠베하
 독일 45128 에센 렐링하우저 슈트라쎬 1-11
- (72) 발명자
 슈나이더 프랑크
 독일 33790 할레 키스커슈트라쎬 14
 안코비치 프랑크
 독일 48336 자쎬베르크 케텔슈트라쎬 10아
- (74) 대리인
 특허법인코리아나

전체 청구항 수 : 총 26 항

(54) 발명의 명칭 구아니디노아세트산의 발효 생산 방법

(57) 요약

본 발명은 구아니디노아세트산 (GAA) 을 생산할 수 있도록 형질전환된 미생물 및 상기 미생물을 이용하는 GAA 의 발효 생산 방법에 관한 것이다. 본 발명은 또한 크레아틴의 발효 생산 방법에 관한 것이다.

(52) CPC특허분류

C12N 15/77 (2013.01)

C12N 15/78 (2013.01)

C12N 9/1003 (2013.01)

C12P 13/10 (2013.01)

C12P 7/40 (2022.01)

C12Y 201/03003 (2013.01)

C12Y 201/04001 (2013.01)

C12Y 403/02001 (2013.01)

C12Y 603/05005 (2013.01)

명세서

청구범위

청구항 1

야생형 미생물에서 각각의 효소 활성화에 비해 카르바모일포스페이트 신타아제의 기능을 갖는 효소의 증가된 활성을 포함하고, L-아르기닌:글리신 아미디노트랜스퍼라제의 활성을 갖는 단백질을 코딩하는 적어도 하나의 이중 유전자를 포함하는 미생물.

청구항 2

제 1 항에 있어서, 카르바모일포스페이트 신타아제의 기능을 갖는 효소의 증가된 활성이 카르바모일포스페이트 신타아제의 기능을 갖는 효소를 코딩하는 유전자의 돌연변이 및/또는 과발현에 의해 달성되는 미생물.

청구항 3

제 1 항 또는 제 2 항에 있어서, L-아르기닌:글리신 아미디노트랜스퍼라제의 활성이 L-아르기닌:글리신 아미디노트랜스퍼라제를 코딩하는 유전자의 돌연변이 및/또는 과발현에 의해 증가되는 미생물.

청구항 4

제 1 항 내지 제 3 항 중 어느 한 항에 있어서, 야생형 미생물의 능력에 비해 L-아르기닌 생산 능력이 개선된 미생물.

청구항 5

제 4 항에 있어서, 아르기니노숙시네이트 리아제의 기능을 갖는 효소의 활성이 야생형 미생물에서의 각각의 효소 활성화에 비해 증가되는 미생물.

청구항 6

제 4 항 또는 제 5 항에 있어서, 오르니틴 카르바모일트랜스퍼라제의 기능을 갖는 효소의 활성이 야생형 미생물에서의 각각의 효소 활성화에 비해 증가되는 미생물.

청구항 7

제 4 항 내지 제 6 항 중 어느 한 항에 있어서, 아르기니노숙시네이트 신테타제의 기능을 갖는 효소의 활성이 야생형 미생물에서의 각각의 효소 활성화에 비해 증가되는 미생물.

청구항 8

제 4 항 내지 제 7 항 중 어느 한 항에 있어서, 효소의 증가된 활성이 각각의 효소를 코딩하는 유전자의 돌연변이 및/또는 과발현에 의해 달성되는 미생물.

청구항 9

제 4 항 내지 제 8 항 중 어느 한 항에 있어서, 아르기닌 오페론 (*argCJBDFR*) 이 과발현되는 미생물.

청구항 10

제 4 항 내지 제 8 항 중 어느 한 항에 있어서, 아르기닌 반응성 리프레서 단백질 ArgR 을 코딩하는 *argR* 유전자가 약독화되거나 결실되는 미생물.

청구항 11

제 4 항 내지 제 8 항 또는 제 10 항 중 어느 한 항에 있어서, 글루타메이트 디히드로게나제, 오르니틴 아세틸트랜스퍼라제, 아세틸글루타메이트 키나제, 아세틸글루타미드포스페이트 리덕타제 및 아세틸오르니틴 아미노트랜스퍼라제를 각각 코딩하는 *gdh*, *argJ*, *argB*, *argC* 및/또는 *argD* 를 포함하는, L-오르니틴 및 L-아르기닌의 생합성 경로의 효소를 코딩하는 유전자 중 적어도 하나 이상이 과발현되는 미생물.

청구항 12

제 1 항 내지 제 11 항 중 어느 한 항에 있어서, L-아르기닌:글리신 아미디노트랜스퍼라제의 기능을 갖는 단백질이 SEQ ID NO: 2 에 따른 아미노산 서열과 적어도 70% 동일한 아미노산 서열을 포함하는 미생물.

청구항 13

제 1 항 내지 제 11 항 중 어느 한 항에 있어서, L-아르기닌:글리신 아미디노트랜스퍼라제의 기능을 갖는 단백질이 SEQ ID NO: 16 에 따른 아미노산 서열과 적어도 70% 동일한 아미노산 서열을 포함하는 미생물.

청구항 14

제 1 항 내지 제 11 항 중 어느 한 항에 있어서, L-아르기닌:글리신 아미디노트랜스퍼라제의 기능을 갖는 단백질이 SEQ ID NO: 24 에 따른 아미노산 서열과 적어도 70% 동일한 아미노산 서열을 포함하는 미생물.

청구항 15

제 1 항 내지 제 11 항 중 어느 한 항에 있어서, L-아르기닌:글리신 아미디노트랜스퍼라제의 기능을 갖는 단백질이 SEQ ID NO: 20 에 따른 아미노산 서열과 적어도 70% 동일한 아미노산 서열을 포함하는 미생물.

청구항 16

제 1 항 내지 제 15 항 중 어느 한 항에 있어서, 미생물이 코리네박테리움 (*Corynebacterium*) 속, 엔테로박테리아세아에 (*Enterobacteriaceae*) 속 또는 슈도모나스 (*Pseudomonas*) 속에 속하는 미생물.

청구항 17

제 16 항에 있어서, 미생물이 코리네박테리움 글루타미쿰 (*Corynebacterium glutamicum*) 인 미생물.

청구항 18

제 16 항에 있어서, 미생물이 에스케리치아 콜라이 (*Escherichia coli*) 인 미생물.

청구항 19

제 16 항에 있어서, 미생물이 슈도모나스 푸티다 (*Pseudomonas putida*) 인 미생물.

청구항 20

a) 발효 배지에서 제 1 항 내지 제 19 항 중 어느 한 항에서 정의된 미생물을 배양하는 단계, 및 b) 배지에 구아니디노 아세트산 (GAA) 을 축적하여 GAA 함유 발효 브로쓰를 형성하는 단계를 포함하는, 구아니디노 아세트산 (GAA) 의 발효 생산 방법.

청구항 21

제 20 항에 있어서, GAA 함유 발효 브로쓰로부터 GAA 를 분리하는 단계를 추가로 포함하는 구아니디노 아세트산 (GAA) 의 발효 생산 방법.

청구항 22

제 20 항에 있어서, GAA 함유 발효 브로쓰를 건조 및/또는 과립화하는 단계를 추가로 포함하는 구아니디노 아세트산 (GAA) 의 발효 생산 방법.

청구항 23

제 1 항 내지 제 19 항 중 어느 한 항에 있어서, 구아니디노아세테이트 N-메틸트랜스퍼라제의 활성을 갖는 효소를 코딩하는 유전자를 추가로 포함하는 미생물.

청구항 24

제 23 항에 있어서, 구아니디노아세테이트 N-메틸트랜스퍼라제의 활성을 갖는 효소를 코딩하는 유전자가 과발현

되는 미생물.

청구항 25

a) 발효 배지에서 제 22 항 또는 제 24 항에서 정의된 미생물을 배양하는 단계, 및 b) 배지에 크레아틴을 축적하여 크레아틴 함유 발효 브로쓰를 형성하는 단계를 포함하는, 크레아틴의 발효 생산 방법.

청구항 26

제 25 항에 있어서, 크레아틴 함유 발효 브로쓰로부터 크레아틴을 단리하는 단계를 추가로 포함하는 크레아틴의 발효 생산 방법.

발명의 설명

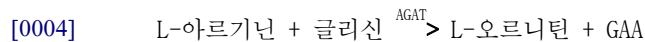
기술 분야

[0001] 본 발명은 구아니디노아세트산 (GAA) 을 생산할 수 있도록 형질전환된 미생물 및 상기 미생물을 이용하는 GAA 의 발효 생산 방법에 관한 것이다. 본 발명은 또한 크레아틴의 발효 생산 방법에 관한 것이다.

배경 기술

[0002] GAA 는 동물 사료 첨가제로서 사용되는 유기 화합물이다 (WO2005120246 A1/ US2011257075 A1). GAA 는 크레아틴의 천연 전구체이다 (예를 들어, Humm et al., Biochem. J. (1997) 322, 771-776). 따라서, GAA 의 보충은 유기체 내의 크레아틴의 최적의 공급을 허용한다.

[0003] 본 발명은 출발 물질로서 산업용 공급 원료 (예를 들어, 암모니아, 암모늄 염 및 글루코스 또는 당 함유 기질) 를 사용하는 발효 공정에 의해 GAA 를 제조하는 방법에 관한 것이다. 생물학적 시스템에서, GAA 및 오르니틴은 L-아르기닌:글리신-아미도트랜스퍼라제 (AGAT; EC 2.1.4.1) 의 촉매 작용에 의해 출발 물질로서 아르기닌 및 글리신으로부터 형성되며, 이는 크레아틴 생합성의 제 1 단계이다 (US 20060200870 A1):



[0005] Guthmiller et al. (J Biol Chem. 1994 Jul 1;269(26):17556-60) 은 E. 콜라이 (*E. coli*) 에서 효소를 클로닝하고 이중적으로 발현시킴으로써 래트 신장 AGAT 를 특징화하였다. Muenchhoff et al. (FEBS Journal 277 (2010) 3844-3860) 은 또한 E. 콜라이에서 효소를 클로닝하고 이중적으로 발현시킴으로써 원핵생물로부터 AGAT 의 제 1 특징화를 보고한다. Sosio et al. (Cell Chemical Biology 25, 540-549, May 17, 2018) 은 스트렙토마이세스 중에서 슈도우리디마이신에 대한 생합성 경로를 설명하였다. 이들은 PumN, L-아르기닌:글리신-아미도트랜스퍼라제 (AGAT) 에 의해 촉매되는 글리신과 L-아르기닌의 반응에 의한 GAA 및 L-오르니틴의 형성을 중간 반응으로서 기술한다. Humm et al. 은 인간 AGAT 를 인코딩하는 제조할 유전자를 대장균에서 발현시키고, AGAT 의 활성 부위 잔기로서 시스템인-407 을 확인하였다 (Biochem. J. (1997) 322, 771-776).

[0006] Mijts et al. (WO 2018079687 A1) 은 미생물 중의 목적 물질, 예를 들어 바닐린 및 바닐산의 생산의 맥락에서 크레아틴이 L-아르기닌 및 글리신으로부터 생산될 수 있음을 개시하고 있다. 저자는 이것이 L-아르기닌 생합성 효소, 글리신 생합성 효소, 및 L-아르기닌 및 글리신의 크레아틴으로의 전환을 촉매하는 효소를 사용함으로써 달성될 수 있음을 추가로 제안한다. L-아르기닌 및 글리신은 AGAT (EC 2.1.4.1) 의 작용에 의해 구아니디노아세테이트 (GAA) 및 오르니틴을 생성하기 위해 조합될 수 있고, GAA 는 메틸 공여체로서 S-아데노실 메티오닌 (SAM) 을 사용하여 구아니디노아세테이트 N-메틸트랜스퍼라제 (GAMT, EC 2.1.1.2) 의 작용에 의해 크레아틴을 생성하기 위해 메틸화될 수 있다. 저자는 L-아르기닌 생합성 효소의 예인 폴리아민의 생산의 맥락에서 또한 공지된 L-오르니틴 생합성 효소뿐만 아니라 소위 L-오르니틴 사이클로부터 잘 공지된 효소, 즉 카르바모일 포스페이트 신타아제 (*carAB*), 오르니틴 카르바모일 트랜스퍼라제 (*argF*, *argI*), 아르기니노숙시네이트 신타아제 (*argG*), 아르기니노숙시네이트 리아제 (*argH*) 를 포함할 수 있다 (참조: Marc et al., Eur. J. Biochem. 267, 5217-5226, 200).

[0007] 그러나, 미생물, 특히 박테리아에서, GAA 합성에서 출발 물질 중 하나, 즉 L-아르기닌의 생산을 증가시키기 위한 몇 가지 더 구체적인 접근법이 또한 문헌으로부터 공지되어 있다. L-아르기닌 생산을 위한 코리네박테리움 글루타미쿰 (*C. glutamicum*) 의 대사 공학에 대한 개요는 ark et al. (NATURE COMMUNICATIONS | DOI: 10.1038/ncomms5618) 에 의해 제공된다. 이들은 이미 L-아르기닌 생산 *C. 글루타미쿰* 균주의 L-아르기닌 생

산자, 예를 들어 ATCC 21831 (Nakayama and Yoshida 1974, US3849250 A) 에 대한 랜덤 돌연변이유발 및 스크리닝을 제안하며, 대사의 시스템-와이드 분석에 기초한 단계적 합리적 대사 공학은 균주 공학 단계 전반에 걸쳐 L-아르기닌 생산의 점진적 증가를 초래한다. Yim et al. (J Ind Microbiol Biotechnol (2011) 38:1911-1920) 은 *C. 글루타미쿰*에서 염색체 argR 유전자를 파괴함으로써 L-아르기닌 생합성 경로를 제어하는 중심 리프레이 단백질 ArgR 을 코딩하는 유전자인 argR 의 불활성화가 아르기닌-생산 균주를 개선시킨다는 것을 보여줄 수 있었다. Ginesy et al. (Microbial Cell Factories (2015) 14:29) 은 아르기닌 생산 증대를 위한 *E. 콜라이*의 성공적인 공학을 보고한다. 그 중에서도, 그들은 argR 리프레이 유전자의 결실을 제안하였다.

[0008] Kurahashi et al. (EP1057893 A1) 은 재조합 DNA 기술을 이용하여 L-아르기닌 생합성 효소를 증진시킴으로써, 예를 들어 에스케리키아 속에 속하는 미생물로부터 유래된 아세틸오르니틴 데아세틸라제, N-아세틸글루탐산- γ -세미알데히드 데히드로게나제, N-아세틸 글루타모키나제 및 아르기니노숙신아제에 대한 유전자를 함유하는 DNA 단편 및 벡터 DNA 를 포함하는 재조합 DNA 를 갖도록 제조된 *코리네박테리움* 또는 *브레비박테리움* 속에 속하는 미생물을 이용함으로써 미생물의 L-아르기닌 생산 능력을 증가시키는 방법에 대한 보고한다. 개선된 L-아르기닌 생산을 위해, 상기 저자는 세포내 글루타메이트 디히드로게나제 (GDH) 의 활성이 향상되고, L-아르기닌 생산 능력이 있는 미생물을 추가로 제안하였다.

[0009] 아르기닌-생합성 오페론인 argR 의 발현을 저해하는 유전자가 불활성화된 유전자 재조합 균주의 이용 방법이 Suga et al. (US7160705 B2) 에 의해 보고되었다. 특히, 아르기닌 오페론을 조절하는 argR 내 결실은 아르기닌 생산의 중요한 인자로서 여겨져 왔다.

[0010] *코리네박테리움* 미생물에서, 아르기닌 생합성에 관여하는 argCJBDFR 유전자는 오페론 형태로 구성되어 세포내 아르기닌에 의한 피드백 저해를 거침으로써 (Sakanyan et al., Microbiology, 142:9-108, 1996), 이의 높은 L-아르기닌 생산 수율에 대한 제한을 가하였다.

[0011] 그러나, Bae et al. (EP3153573 A1) 은 *C. 글루타미쿰*에서 L-아르기닌의 생산 수율을 증가시키려는 시도로, 아르기닌 리프레이 (argR) 에서 어떠한 결실도 없이, 아르기닌 오페론 및 오르니틴 카르바모일트랜스퍼라제 (ArgF, ArgF2) 의 활성을 향상시킴으로써, 모체 L-아르기닌-생산 균주에 비해 L-아르기닌을 더 높은 수율로 생산할 수 있음을 발견하였다.

[0012] 아르기닌 오페론은 L-아르기닌 생합성 기전에 관여하는 효소를 인코딩하는 유전자로 구성된 오페론으로, 특히 아르기닌 오페론은 L-아르기닌 생합성의 순환형 단계를 구성하는 효소를 인코딩하는 유전자로 구성된다. 구체적으로, 아르기닌 오페론은 N-아세틸글루타미드 포스페이트 리덕타제 (ArgC), 글루타메이트 N-아세틸트랜스퍼라제 (ArgJ), N-아세틸글루타메이트 키나제 (ArgB), 아세틸오르니틴 아미노트랜스퍼라제 (ArgD), 오르니틴 카르바모일트랜스퍼라제 (ArgF) 및 아르기닌 리프레이 (ArgR) 로 구성된다. 이들 효소는 L-아르기닌 생합성의 연속 효소 반응에 관여하며, argR 에 의해 인코딩된 아르기닌 리프레이에 의해 조절된다 (WO 2006/057450 A1).

[0013] Fan Wenchao 은 *코리네박테리움 글루타미쿰*과 같은 비병원성 미생물의 발효에 의한 크레아틴의 생산 방법을 개시한다 (CN106065411 A). 미생물은 다음의 바이오전환 기능을 갖는다: 글루코스의 L-글루탐산으로의 전환; L-글루탐산의 N-아세틸-L-글루탐산으로의 전환; N-아세틸-L-글루탐산의 N-아세틸-L-글루탐산 세미알데히드로의 전환; N-아세틸-L-글루탐산 세미알데히드의 N-아세틸-L-오르니틴으로의 전환; N-아세틸-L-오르니틴의 L-오르니틴으로의 전환; L-오르니틴의 L-시트룰린으로의 전환; L-시트룰린의 아르기니노-숙신산으로의 전환; 아르기니노-숙신산의 L-아르기닌으로의 전환; L-아르기닌의 구아니디노아세트산으로의 전환; 및 마지막으로 구아니디노아세트산의 크레아틴으로의 전환. Fan Wenchao 는 미생물이 N-아세틸글루타메이트-신타아제, N-아세틸오르니틴- δ -아미노트랜스퍼라제, N-아세틸오르니티나제, 오르니틴-카르바모일 트랜스퍼라제, 아르기니노숙시네이트 신테타제, 글리신 아미디노-트랜스퍼라제 (EC: 2.1. 4.1), 및 구아니디노아세테이트 N-메틸트랜스퍼라제 (EC: 2.1.1.2) 로 이루어진 균으로부터 선택된 하나 이상의 효소를 과발현한다고 제안한다. 미생물은 바람직하게는 글리신 아미노트랜스퍼라제 (L-아르기닌:글리신 아미디노트랜스퍼라제) 및 구아니디노아세테이트 N-메틸트랜스퍼라제를 과발현한다.

발명의 내용

해결하려는 과제

[0014] 지금까지, 야생형 형태에 비해 증가된 GAA 생산에 적합한 미생물 및 이러한 미생물을 이용한 GAA 생산에 대한 각각의 방법은 보고되지 않았다.

[0015] 따라서, 본 발명의 근본적인 과제는 구아니디노아세트산 (GAA) 을 생산할 수 있도록 형질전환된 미생물 및 상기 미생물을 이용하는 GAA 의 발효 생산 방법을 제공하는 것이다.

과제의 해결 수단

[0016] 이 과제는 야생형 미생물에서 각각의 효소 활성화에 비해 카르바모일포스페이트 신타아제 (EC 6.3.4.16) 의 기능을 갖는 효소의 증가된 활성을 포함하고, L-아르기닌:글리신 아미디노트랜스퍼라제 (AGAT) 의 기능을 갖는 단백질을 코딩하는 적어도 하나의 이중 유전자를 포함하는 미생물에 의해 해결된다.

[0017] 이중 유전자는 유전자가 자연적으로 이 유전자를 갖지 않는 숙주 유기체 내에 삽입되었음을 의미한다. 숙주 내 이중 유전자의 삽입은 재조합 DNA 기술에 의해 수행된다. 재조합 DNA 기술을 거친 미생물은 형질전환, 유전자 변형 또는 재조합체로 불린다. 따라서, 본 발명에 따른 미생물은 재조합체이다.

[0018] 카르바모일포스페이트 신타아제의 기능을 갖는 효소의 증가된 활성은 카르바모일포스페이트 신타아제의 기능을 갖는 효소를 코딩하는 유전자의 돌연변이 및/또는 과발현에 의해 달성될 수 있다.

[0019] L-아르기닌:글리신 아미디노트랜스퍼라제의 활성화는 또한 L-아르기닌:글리신 아미디노트랜스퍼라제를 코딩하는 유전자의 돌연변이 및/또는 과발현에 의해 증가될 수 있다.

[0020] L-아르기닌:글리신 아미디노트랜스퍼라제 (AGAT) 의 기능을 갖는 단백질은 아미디노트랜스퍼라제 패밀리에 속한다. 아미디노트랜스퍼라제 패밀리는 각각 크레아틴 및 스트렙토마이신 생합성에 관여하는 효소인 글리신 (EC:2.1.4.1) 및 이노사민 (EC:2.1.4.2) 을 포함한다. 이 패밀리는 또한 아르기닌 테이미나제, EC:3.5.3.6 을 포함한다. 이들 효소는 반응을 촉매한다: 아르기닌 + H₂O <=> 시트룰린 + NH₃. 또한 이 패밀리에서 발견되는 것은 스트렙토코쿠스 항종양 당단백질이다. L-아르기닌:글리신-아미디노트랜스퍼라제 (AGAT) 활성을 갖는 효소 또는 단백질은 또한, 하기 간행물: Pissowotzki K et al., Mol Gen Genet 1991;231:113-123 (PUBMED:1661369 EPMC:1661369); D'Hooghe I et al., J Bacteriol 1997;179:7403-7409 (PUBMED:9393705 EPMC:9393705); Kanaoka M et al., Jpn J Cancer Res 1987;78:1409-1414 (PUBMED:3123442 EPMC:3123442) 에서 설명되는 바와 같은, PFAM 패밀리: Amidinotransf (PF02274) (: Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", Nucleic Acids Res. 45(D1):D200-D203.) 에 속하는 보존된 도메인을 갖는 것으로 기술된다. AGAT 의 특정 예는 무레아 프로듀센스 (*Moorea producens*), 호모 사피엔스 (*Homo sapiens*), 라투스 노르베지쿠스 (*Rattus norvegicus*), 갈레옵테루스 바리에가투스 (*Galeopterus variegatus*), 및 실린드로스페르모프시스 라시보르스키 (*Cylindrospermopsis raciborskii*) 의 것이다.

[0021] 본 발명에 따른 미생물은 야생형 미생물의 능력에 비해 L-아르기닌 생산 능력이 이상적으로 개선된다. 이러한 특성은 천연 L-아르기닌 생산자인 미생물의 선별에 의해 달성될 수 있거나 돌연변이에 의해 L-아르기닌을 생산하는 능력을 취득할 수 있다.

[0022] 본 발명에 따르면 야생형 미생물의 능력에 비해 L-아르기닌 생산 능력이 향상된 미생물은 야생형 미생물에서 각각의 효소 활성화에 비해 아르기니노숙시네이트 리아제 (E.C. 4.3.2.1) 의 기능을 갖는 효소의 활성이 증가될 수 있다.

[0023] 또한, 본 발명에 따른 미생물에서, 오르니틴 카르바모일트랜스퍼라제 (EC 2.1.3.3) 의 기능을 갖는 효소의 활성을 야생형 미생물에서의 각각의 효소 활성화에 비해 증가시킬 수 있다.

[0024] 본 발명에 따른 미생물에서, 아르기니노숙시네이트 신테타제 (E.C. 6.3.4.5) 의 기능을 갖는 효소의 활성을 야생형 미생물에서의 각각의 효소 활성화에 비해 증가시킬 수 있다.

[0025] 미생물에서 증가된 효소 활성화는, 예를 들어, 상응하는 내인성 유전자의 돌연변이에 의해 달성될 수 있다. 효소 활성을 증가시키기 위한 추가 측정은 효소를 코딩하는 mRNA 를 안정화시키는 것일 수 있다.

[0026] 상기 언급된 효소의 활성 증가는 각 효소를 코딩하는 유전자를 과발현시킴으로써 달성될 수도 있다. 다시 말해, 이 과제는 바람직하게는 야생형 유기체의 능력에 비해 L-아르기닌을 생산하는 능력이 개선된 및/또는 카르바모일포스페이트 신타아제 (EC 6.3.4.16) 의 기능을 갖는 단백질을 코딩하는 적어도 하나 이상의 과발현된 유전자 (예를 들어, carA, carB) 를 갖고, L-아르기닌:글리신 아미디노트랜스퍼라제 (AGAT, 예를 들어, EC 2.1.4.1) 의 기능을 갖는 단백질을 코딩하는 유전자를 추가로 포함하는 미생물에 의해 해결된다.

[0027] 또한, 본 발명에 따른 미생물은 바람직하게는 오르니틴 카르바모일트랜스퍼라제의 기능 (EC 2.1.3.3) 을 갖는

단백질을 코딩하는 유전자 (예를 들어, *argF/argF2/argI*), 아르기니노숙시네이트 신테타제 (E.C. 6.3.4.5) 의 기능을 갖는 단백질을 코딩하는 유전자 (예를 들어, *argG*), 및 아르기니노숙시네이트 리아제 (E.C. 4.3.2.1) 의 기능을 갖는 단백질을 코딩하는 유전자 (예를 들어, *argH*) 로 이루어진 군으로부터 선택되는 적어도 하나 이상의 과발현된 유전자를 포함한다.

- [0028] 유전자의 과발현은 일반적으로 유전자의 카피수를 증가시킴으로써 및/또는 유전자를 강력한 프로모터와 기능적으로 연결시킴으로써 및/또는 리보솜 결합 부위를 향상시킴으로써 및/또는 개시 코돈 또는 전체 유전자의 코돈 사용 최적화 또는 상기 언급된 모든 방법의 선택을 포함하는 조합에 의해 달성된다.
- [0029] 본 발명의 문맥에서 L-아르기닌을 생산하는 능력이 향상된 미생물은 자신의 필요를 초과하여 L-아르기닌을 생산하는 미생물을 의미한다. 이러한 L-아르기닌 생산 미생물의 예는 예를 들어 *C. 글루타미쿰* ATCC 21831 또는 Park et al. (NATURE COMMUNICATIONS | DOI: 10.1038/ncomms5618) 또는 Ginesy et al. (Microbial Cell Factories (2015) 14:29) 에 의해 개시된 것들이다.
- [0030] 본 발명에 따른 미생물의 일 구현예에서, 아르기닌 오페론 (*argCJBDFR*) 은 과발현될 수 있다.
- [0031] 대안적으로는, 본 발명에 따른 미생물에서 아르기닌 반응성 리프레서 단백질 ArgR 을 코딩하는 *argR* 유전자는 약독화되거나 결실될 수 있다.
- [0032] 본 발명의 추가의 구현예에서, 및 임의로 상기 언급된 변형에 더하여, 글루타메이트 디히드로게나제, 오르니틴 아세틸트랜스퍼라제, 아세틸글루타메이트 키나제, 아세틸글루타미포스페이트 리덕타제 및 아세틸오르니틴 아미노트랜스퍼라제를 각각 코딩하는 *gdh*, *argJ*, *argB*, *argC* 및/또는 *argD* 를 포함하는, L-아르기닌의 생합성 경로의 효소를 코딩하는 적어도 하나 이상의 유전자가 본 발명에 따른 미생물에서 과발현된다.
- [0033] 표 1 은 상이한 종, 즉 *E. 콜라이*, *C. 글루타미쿰* 및 슈도모나스 푸티다 (*P. putida*) 에서 아르기닌 생합성에 관여하거나 이에 기여하는 효소의 상이한 명칭을 나타낸다.
- [0034] 본 발명의 미생물에 있어서, L-아르기닌:글리신 아미디노트랜스퍼라제 기능을 갖는 단백질을 코딩하는 유전자가 추가로 과발현될 수 있다. 유전자의 과발현은 일반적으로 유전자의 카피수를 증가시킴으로써 및/또는 유전자를 강력한 프로모터와 기능적으로 연결시킴으로써 및/또는 리보솜 결합 부위를 향상시킴으로써 및/또는 전체 유전자의 개시 코돈의 코돈 사용 최적화 또는 상기 언급된 모든 방법 또는 선택을 포함하는 조합에 의해 달성된다.

[0035]

표 1: 효소의 명칭

효소 명칭	Alias	EC 번호	E. 콜라이	C. 글루타미쿰	P. 푸티다
글루타메이트 데히드로게나제	NADP-SPECIFIC GLUTAMATE DEHYDROGENASE	1.4.1.4 (P. 푸티다 에서는 1.4.1.2)	gdhA	gdh	gdhA/gdhB
N-아세틸 글루타모키나제	ACETYLGLUTAMATE KINASE	2.7.2.8	argB	argB	argB
N-아세틸글루타밀 포스페이트 리덕타제	N-아세틸-감마-글루타밀포스페이트 리덕타제	1.2.1.38	argC	argC	argC
N-아세틸글루탐산- γ -세 미알데히드 데히드로게나제	ACETYLORNITHINE AMINOTRANSFERASE (e.c. 이작용성 아세틸로니틴 아미노트랜스퍼라제/ 숙시닐디아미노피멜레이트 아미노트랜스퍼라제(에서))	2.6.1.11 / 2.6.1.17	argD	argD	aruC
아세틸로니틴 데아세틸라제	이작용성 아세틸로니틴 데아세틸라제 / GLUTAMATE N-ACETYLTRANSFERASE	2.3.1.35 / 2.3.1.1		argJ	argJ

[0036]

효소 명칭	Alias	EC 번호	E. 콜라이	C. 글루타미쿰	P. 푸티다
아세틸로니틴		3.5.1.16	argE		argE
데아세틸라제					
	GLUTAMATE N-ACETYLTRANSFERASE	2.3.1.1	argA		argA
오르니틴 카르바모일트랜스퍼라제	오르니틴 카르바모일트랜스퍼라제 1	2.1.3.3	argI	argF/argF2	arcB/argF
제					
아르기니노숙시네이트 신태타제	ARGININOSUCCINATE SYNTHASE	6.3.4.5	argG	argG	argG
아르기니노숙시네이트 리아제	ARGININOSUCCINATE LYASE	4.3.2.1	argH	argH	argH
카르바모일-포스페이트 신태타제	카르바모일-포스페이트 신태타제	6.3.5.5	carAB	carAB	carAB
카르바메이트 키나제	카르바메이트 키나제	2.7.2.2	ybcF/yqeA		arcC

[0037]

[0038]

본 발명의 미생물에서 L-아르기닌:글리신 아미디노트랜스퍼라제 (AGAT) 의 기능을 갖는 단백질은 SEQ ID NO: 2 또는 SEQ ID NO: 4 에 따른 아미노산 서열 (무레아 프로듀센스 (*Moorea producens*) 의 "AGAT_Mp") 에 대해 적어도 70% 상동, 바람직하게는 적어도 80% 또는 적어도 90% 상동인 아미노산 서열을 포함할 수 있다. 본 발명의 추가의 구현예에서, L-아르기닌:글리신 아미디노트랜스퍼라제의 아미노산 서열은 SEQ ID NO: 2 또는 SEQ ID NO: 4 에 따른 아미노산 서열과 동일하다 (참조: Database UniProt, 15 February 2017, "Glycine amidinotransferase", XP055706853, EBI 수탁 번호 UNIPROT: A0A1D8TKD3). 무레아 프로듀센스 AGAT 를 코딩하는 야생형 DNA 의 서열은 SEQ ID NO:1 이고, C. 글루타미쿰에 대해 코돈 최적화된 상응하는 DNA 서열은 SEQ ID NO:3 이다. P. 푸티다에 대해 코돈 최적화된 무레아 프로듀센스 AGAT 유전자에 대한 상응하는 DNA 서열은 SEQ ID NO: 33 이다.

[0039]

본 발명의 미생물에서 L-아르기닌:글리신 아미디노트랜스퍼라제의 기능을 갖는 단백질은 SEQ ID NO: 16 또는 SEQ ID NO: 26 에 따른 아미노산 서열 (실린드로스페르모프시스 라시보르스키 (*Cylindrospermopsis raciborskii*) ATW205 의 "AGAT_cyrA") 에 대해 적어도 70% 상동, 바람직하게는 적어도 80% 또는 적어도 90% 상동인 아미노산 서열을 포함할 수 있다. 본 발명의 추가의 구현예에서, L-아르기닌:글리신 아미디노트랜스퍼라제의 아미노산 서열은 SEQ ID NO:16 또는 SEQ ID NO:26 에 따른 아미노산 서열과 동일하다. 실린드로스페

르모프시스 라시보르스키 (*Cylindrospermopsis raciborskii*) AGAT 를 코딩하는 야생형 DNA 의 서열은 SEQ ID NO:15 이고, *C. 글루타미쿰*에 대해 코돈 최적화된 상응하는 DNA 서열은 SEQ ID NO:25 이다.

[0040] 본 발명의 미생물에서 L-아르기닌:글리신 아미디노트랜스퍼라제의 기능을 갖는 단백질은 SEQ ID NO: 23 (갈레옵테루스 바리에가투스 (*Galeopterus variegatus*) 의 "AGAT_Gv"), 바람직하게는 SEQ ID NO: 24 또는 SEQ ID NO:32 에 따른 아미노산 서열에 대해 적어도 70% 상동, 바람직하게는 적어도 80% 또는 적어도 90% 상동인 아미노산 서열을 포함할 수 있다. 본 발명의 추가의 구현예에서, L-아르기닌:글리신 아미디노트랜스퍼라제의 아미노산 서열은 SEQ ID NO:24 또는 SEQ ID NO:32 에 따른 아미노산 서열과 동일하다. *C. 글루타미쿰*에 대해 코돈 최적화된 상응하는 갈레옵테루스 바리에가투스 AGAT DNA 의 서열은 SEQ ID NO: 31 이다.

[0041] 본 발명의 미생물에서 L-아르기닌:글리신 아미디노트랜스퍼라제의 기능을 갖는 단백질은 SEQ ID NO:18 (호모 사피엔스 (*Homo sapiens*) 의 "AGAT_Hs"), 바람직하게는 SEQ ID NO: 20 또는 SEQ ID NO:28 에 따른 아미노산 서열, 예를 들어 SEQ ID NO: 21 또는 SEQ ID NO: 22 또는 SEQ ID NO:30 (라투스 노르베지쿠스 (*Rattus norvegicus*) 의 각 "AGAT Rn"; *C. 글루타미쿰*에 대해 코돈 최적화된, 상응하는 DNA 는 SEQ ID NO: 29 임) 에 따른 아미노산 서열에 대해 적어도 70% 상동, 바람직하게는 적어도 80% 또는 적어도 90% 상동인 아미노산 서열을 포함할 수 있다. 본 발명의 추가의 구현예에서, L-아르기닌:글리신 아미디노트랜스퍼라제의 아미노산 서열은 SEQ ID NO:18 에 따른 아미노산 서열과 동일하다. 호모 사피엔스 AGAT 를 코딩하는 야생형 DNA 의 서열은 SEQ ID NO:17 이고, *C. 글루타미쿰*에 대해 코돈 최적화된 상응하는 DNA 서열은 SEQ ID NO:27 이다.

[0042] 본 발명의 미생물은 코리네박테리움 (*Corynebacterium*) 속, 바람직하게는 코리네박테리움 글루타미쿰 (*Corynebacterium glutamicum* (*C. glutamicum*)), 또는 엔테로박테리아세아에 (*Enterobacteriaceae*) 속, 바람직하게는 에스케리치아 콜라이 (*Escherichia coli* (*E. coli*)), 또는 슈도모나스 (*Pseudomonas*) 속, 바람직하게는 슈도모나스 푸티다 (*Pseudomonas putida* (*P. putida*)) 에 속할 수 있다.

[0043] 일반적으로, 미생물에서 증가된 효소 활성은, 예를 들어, 상응하는 내인성 유전자의 돌연변이에 의해 달성될 수 있다. 또한 상응하는 유전자의 과발현에 의해 효소 활성이 향상될 수 있다.

[0044] 일반적으로, 본 발명에 따른, 유전자의 과발현은 유전자의 카피수를 증가시킴으로써 및/또는 조절 인자의 향상에 의해, 예를 들어, 유전자를 강력한 프로모터와 기능적으로 연결시킴으로써 및/또는 리보솜 결합 부위를 향상시킴으로써 및/또는 개시 코돈 또는 전체 유전자의 코돈 사용 최적화에 의해 달성된다. 유전자 발현에 긍정적인 영향을 미치는 이러한 조절 인자의 향상은, 예를 들어, 프로모터의 유효성을 증가시키기 위해 구조 유전자의 업스트림의 프로모터 서열을 변형함으로써 또는 상기 프로모터를 보다 효과적이거나 소위 강한 프로모터로 완전히 대체함으로써 달성될 수 있다. 프로모터는 유전자의 업스트림에 위치한다. 프로모터는 약 40 내지 50 개의 염기쌍으로 이루어진 DNA 서열로서, RNA 중합효소 할로효소의 결합 부위 및 전사 시작 지점을 구성하며, 이에 의해 조절된 폴리뉴클레오티드 또는 유전자의 발현 강도가 영향을 받을 수 있다. 일반적으로, 강한 프로모터를 선택함으로써, 예를 들어, 본래의 프로모터를 강하고 천연적인 (본래 다른 유전자에 할당된) 프로모터로 대체함으로써, 또는 주어진 천연적인 프로모터의 특정 영역 (예를 들어, 소위 -10 및 -35 영역) 을, 예를 들어, *C. 글루타미쿰*의 경우 M. Patek et al. (*Microbial Biotechnology* 6 (2013), 103-117) 에 의해 교시된 바와 같이, 컨센서스 서열을 향해 변형시킴으로써, 박테리아에서 유전자의 과발현 또는 발현의 증가를 달성할 수 있다. "기능적 연결" 은 프로모터와 유전자의 순차적인 배열로, 이는 유전자의 전사를 초래하는 것을 의미하는 것으로 이해된다.

[0045] 유전자 코드는 퇴보되는데, 이는 특정 아미노산이 다수의 상이한 트리플릿에 의해 인코딩될 수 있다는 것을 의미한다. 용어 코돈 사용은 특정 유기체가 전형적으로 동일한 빈도를 갖는 특정 아미노산에 대해 모든 가능한 코돈을 사용하지 않을 것이라는 관찰을 의미한다. 대신에, 유기체는 전형적으로 특정 코돈에 대한 특정 선호도를 나타낼 것이며, 이는 이들 코돈이 유기체의 전사된 유전자의 코딩 서열에서 더 빈번하게 발견됨을 의미한다. 미래의 숙주에 외래인, 즉 상이한 종으로부터 유래된 특정 유전자가 미래의 숙주 유기체에서 발현되어야 한다면, 상기 유전자의 코딩 서열은 상기 미래의 숙주 유기체의 코돈 사용에 대해 조정되어야 한다 (즉, 코돈 사용 최적화).

[0046] 상기 언급된 과정은 또한, a) 상기 정의된 본 발명에 따른 미생물을 적합한 배지에서 적합한 조건 하에서 배양하는 단계, 및 b) 배지에 구아니디노 아세트산 (GAA) 을 축적하여 GAA 함유 발효 브로쓰를 형성하는 단계를 포함하는, 구아니디노 아세트산 (GAA) 의 발효 생산 방법에 의해 해결된다.

[0047] 본 발명에 따른 방법은 배지에 글리신을 첨가하는 단계 및/또는 L-아르기닌을 첨가하는 단계 및/또는 L-오르니

틴을 첨가하는 단계를 더 포함할 수 있다. 바람직하게는, 배지에 0.1 내지 300 g 글리신/1 배지, 바람직하게는 0.82 g 글리신/1 배지 범위의 농도의 글리신 및/또는 0.1 내지 200 g L-아르기닌/1 배지, 바람직하게는 1.9 g L-아르기닌/1 배지 범위의 농도를 얻도록 L-아르기닌을 보충한다.

[0048] 본 발명의 방법은 발효 브로쓰로부터 GAA 를 분리하는 단계를 더 포함할 수 있다.

[0049] 본 발명에 따른 방법은 GAA 함유 발효 브로쓰를 건조 및/또는 과립화하는 단계를 더 포함할 수 있다.

[0050] 본 발명은 또한 구아니디노아세테이트 N-메틸트랜스퍼라제 (EC: 2.1.1.2) 의 활성을 갖는 효소를 코딩하는 유전자를 추가로 포함하는, 상기 정의된 미생물에 관한 것이다. 바람직하게는, 구아니디노아세테이트 N-메틸트랜스퍼라제의 활성을 갖는 효소를 코딩하는 유전자가 과발현된다.

[0051] 또한, 본 발명은 a) 구아니디노아세테이트 N-메틸트랜스퍼라제의 활성을 갖는 효소를 코딩하는 유전자를 포함하는 본 발명에 따른 미생물을 적합한 배지에서 적합한 조건 하에 배양하는 단계, 및 b) 상기 배지 내에 크레아틴을 축적하여 크레아틴 함유 발효 브로쓰를 형성하는 단계를 포함하는, 크레아틴의 발효 생산 방법에 관한 것이다.

[0052] 바람직하게는, 상기 방법은 크레아틴 함유 발효 브로쓰로부터 크레아틴을 분리하는 단계를 추가로 포함한다. 크레아틴은 등전점법 및/또는 이온 교환법에 의해 발효 브로쓰로부터 추출될 수 있다. 대안적으로, 크레아틴은 물 중에서 재결정화하는 방법에 의해 추가로 정제될 수 있다.

발명을 실시하기 위한 구체적인 내용

[0053] 실험 부문

[0054] A) 재료 및 방법

[0055] 화학물질

[0056] 스트렙토마이세스 카나마이세티쿠스 (*Streptomyces kanamyceticus*) 로부터의 카나마이신 용액을 Sigma Aldrich (St. Louis, USA, Cat. no. K0254) 로부터 구입하였다. IPTG (이소프로필 β-D-1-티오갈락토피라노사이드) 를 Carl-Roth (Karlsruhe, Germany, Cat. no. 2316.4.) 로부터 구입하였다. 달리 언급되지 않는다면, 모든 다른 화학물질은 Merck (Darmstadt, Germany), Sigma Aldrich (St. Louis, USA) 또는 Carl-Roth (Karlsruhe, Germany) 로부터 분석적으로 순수하게 구입하였다.

[0057] 세포 증식을 위한 배양

[0058] 달리 언급되지 않는다면, 배양/인큐베이션 절차는 다음과 같이 수행하였다:

[0059] a. Merck (Darmstadt, Germany; Cat. no. 110285) 사의 LB 브로쓰 (MILLER) 를 액체 배지에서 *E. 콜라이* 균주를 배양하는데 사용하였다. 액체 배양물 (3 개의 배플이 있는 100 ml Erlenmeyer 플라스크 당 10 ml 액체 배지) 을 Infors GmbH (Bottmingen, Switzerland) 사의 Infors HT Multitron 표준 인큐베이터 진탕기에서 30°C 및 200 rpm 으로 인큐베이션하였다.

[0060] b. Merck (Darmstadt, Germany; Cat. no. 110283) 사의 LB 브로쓰 (MILLER) 를 아가 플레이트 상에서 *E. 콜라이* 균주를 배양하는데 사용하였다. 아가 플레이트를 VWR (Radnor, USA) 사의 INCU-Line® 미니 인큐베이터에서 30°C 에서 인큐베이션하였다.

[0061] c. Merck (Darmstadt, Germany; Cat. no. 110493) 사의 뇌 심장 인퓨전 브로쓰 (BHI) 를 액체 배지에서 *C. 글루타미쿰* 균주를 배양하는데 사용하였다. 액체 배양물 (3 개의 배플이 있는 100 ml Erlenmeyer 플라스크 당 10 ml 액체 배지) 을 Infors GmbH (Bottmingen, Switzerland) 사의 Infors HT Multitron 표준 인큐베이터 진탕기에서 30°C 및 200 rpm 으로 인큐베이션하였다.

[0062] d. Merck (Darmstadt, Germany; Cat. no. 113825) 사의 뇌 심장 아가 (BHI-아가) 를 아가 플레이트 상에서 *C. 글루타미쿰* 균주의 배양에 사용하였다. 아가 플레이트를 Kelvitron® 온도 제어기 (Hanau, Germany) 가 있는 Heraeus Instruments 사의 인큐베이터에서 30°C 에서 인큐베이션하였다.

[0063] e. 전기천공 후 *C. 글루타미쿰* 을 배양하기 위하여, BHI-agar (Merck, Darmstadt, Germany, Cat. no. 113825) 에 134 g/l 소르비톨 (Carl Roth GmbH + Co. KG, Karlsruhe, Germany), 2.5 g/l 효모 추출물 (Oxoid/ThermoFisher Scientific, Waltham, USA, Cat. no. LP0021) 및 25 mg/l 카나마이신을 보충하였다.

아가 플레이트를 Kelvitron® 온도 제어기 (Hanau, Germany) 가 있는 Heraeus Instruments 사의 인큐베이터에서 30℃ 에서 인큐베이션하였다.

[0064] **박테리아 현탁액의 광학 밀도 결정**

[0065] a. 진탕 플라스크 배양물 중의 박테리아 현탁액의 광학 밀도를 Eppendorf AG (Hamburg, Germany) 사의 BioPhotometer 를 사용하여 600 nm (OD600) 에서 측정하였다.

[0066] b. Tecan Group AG (Maennedorf, Switzerland) 사의 GENios™ 플레이트 판독기를 사용하여 Wouter Duetz (WDS) 마이크로 발효 시스템 (24-웰 플레이트) 에서 생산된 박테리아 현탁액의 광학 밀도를 660 nm (OD660) 에서 측정하였다.

[0067] **원심분리**

[0068] a. 최대 부피가 2 ml 인 박테리아 현탁액을 Eppendorf 5417 R 벤치탑 원심분리 (13.000 rpm 에서 5 분) 를 사용하여 1.5 ml 또는 2 ml 반응 튜브 (예를 들어, Eppendorf Tubes® 3810X) 에서 원심분리하였다.

[0069] b. 최대 부피가 50 ml 인 박테리아 현탁액을 4.000 rpm 에서 10 분 동안 Eppendorf 5810 R 벤치탑 원심분리를 사용하여 15 ml 또는 50 ml 반응 튜브 (예를 들어, Falcon™ 50 ml Conical Centrifuge Tubes) 에서 원심분리하였다.

[0070] **DNA 단리**

[0071] 플라스미드 DNA 를 제조사의 지시에 따라 Qiagen (Hilden, Germany, Cat. No. 27106) 사의 QIAprep Spin Miniprep Kit 를 사용하여 *E. 콜라이* 세포로부터 단리하였다.

[0072] **중합효소 연쇄 반응 (PCR)**

[0073] 프루프 리딩 (높은 충실도) 폴리머라제를 사용한 PCR 을 사용하여 생어 (Sanger) 서열분석 또는 DNA 어셈블리를 위해 원하는 DNA 분절을 증폭시켰다. 비-프루프-리딩 폴리머라제 키트를 *E. 콜라이* 또는 *C. 글루타미쿰* 콜로니로부터 직접 원하는 DNA 단편의 존재 또는 부재를 결정하기 위해 사용하였다.

[0074] a. New England BioLabs Inc. (Ipswich, USA, Cat. No. M0530) 사의 Phusion® High-Fidelity DNA Polymerase Kit (Phusion Kit) 를 제조사의 지시에 따라 선택된 DNA 영역의 주형-정확한 증폭에 사용하였다 (표 2 참조).

[0075] 표 2: New England BioLabs Inc. 사의 Phusion® High-Fidelity DNA Polymerase Kit 를 이용한 PCR 을 위한 열순환 조건.

PCR 프로그램			
단계	시간 [min.:sec.]	T [°C]	설명
1	00:30	98	초기 변성 단계
2	00:05	98	변성 단계
3	00:30	60	어닐링 단계
4	00:xx	72	신장 단계 DNA kb 당 1 min
			단계 2 내지 4 반복: 35 x
5	05:00	72	최종 신장 단계
6	유지	4	냉각 단계

[0076]

[0077] b. Qiagen (Hilden, Germany, Cat. No.201203) 사의 Taq PCR Core Kit (Taq Kit) 를 DNA 의 원하는 분절을 그의 존재를 확인하기 위해 증폭하는데 사용하였다. 키트는 제조사의 지시에 따라 사용하였다 (표 3 참조).

[0078] 표 3: Qiagen 사의 Taq PCR Core Kit 를 이용한 PCR 을 위한 열순환 조건.

PCR 프로그램			
단계	시간 [min.:sec.]	T [°C]	설명
1	05:00	94	초기 변성 단계
2	00:30	94	변성 단계
3	00:30	52	어닐링 단계
4	01:20	72	신장 단계 DNA kb 당 1 min
			단계 2 내지 4 반복: 35 x
5	04:00	72	최종 신장 단계
6	유지	4	냉각 단계

[0079]

[0080] c. Takara Bio Inc (Takara Bio Europe S.A.S., Saint-Germain-en-Laye, France, Cat. No. RR350A/B) 사의 SapphireAmp® Fast PCR Master Mix (Sapphire Mix) 를 제조자의 지침에 따라 *E. 콜라이* 또는 *C. 글루타미쿰* 으로부터 취한 세포에서 원하는 DNA 분절의 존재를 확인하기 위한 대안으로서 사용하였다 (표 4 참조).

[0081] 표 4: Takara Bio Inc. 사의 SapphireAmp® Fast PCR Master Mix (Sapphire Mix) 를 사용한 PCR 을 위한 열순환 조건.

PCR 프로그램			
단계	시간 [min.:sec.]	T [°C]	설명
1	01:00	94	초기 변성 단계
2	00:05	98	변성 단계
3	00:05	55	어닐링 단계
4	00:05	72	신장 단계
			단계 2 내지 4 반복: 30 x
5	04:00	72	최종 신장 단계
6	유지	4	냉각 단계

[0082]

[0083] d. 모든 올리고뉴클레오타이드 프라이머는 McBride and Caruthers (1983) 에 의해 기술된 포스포라미다이트 방법을 사용하여 Eurofins Genomics GmbH (Ebersberg, Germany) 에 의해 합성되었다.

[0084] e. PCR 주형으로서, 단리된 플라스미드 DNA 또는 액체 배양물로부터 단리된 총 DNA 의 적합하게 희석된 용액 또는 박테리아 콜로니에 함유된 총 DNA (콜로니 PCR) 를 사용하였다. 상기 콜로니 PCR 을 위해, 아가 플레이트 상의 콜로니로부터 이쥘시개로 세포 물질을 취하고, 세포 물질을 PCR 반응 튜브 내로 직접 위치시킴으로써 주형을 제조하였다. 세포 물질을 SEVERIN Elektrogeraete GmbH (Sundern, Germany) 사의 전자레인지 오븐 유형 Mikrowave & Grill 에서 800 W 에서 10 초 동안 가열한 다음, PCR 반응 튜브 내의 주형에 PCR 시약을 첨가하였다.

- [0085] f. 모든 PCR 반응은 Eppendorf AG (Hamburg, Germany) 사의 PCR 사이클러 타입 Mastercycler 또는 Mastercycler nexus 구매에서 수행하였다.
- [0086] **DNA 의 제한효소 분해**
- [0087] 제한 효소 분해에 대해서는 New England BioLabs Inc. (Ipswich, USA) 사의 "FastDigest restriction endonucleases (FD)" (ThermoFisher Scientific, Waltham, USA) 또는 제한 엔도뉴클레아제를 사용하였다. 반응은 제조사의 매뉴얼의 지시에 따라 수행되었다.
- [0088] **DNA 단편의 크기 결정**
- [0089] a. 작은 DNA 단편의 크기 (<1000 bps) 는 일반적으로 Qiagen (Hilden, Germany) 사의 QIAxcel 을 사용하는 자동 모세관 전기영동에 의해 결정되었다.
- [0090] b. DNA 단편이 단리되어야 하는 경우 또는 DNA 단편이 1000 bps 초과인 경우, DNA 를 TAE 아가로스 겔 전기영동에 의해 분리하고, GelRed® Nucleic Acid Gel Stain (Biotium, Inc., Fremont, Canada) 으로 염색하였다. 염색된 DNA 는 302 nm 에서 눈에 보였다.
- [0091] **PCR 증폭물 및 제한 단편의 정제**
- [0092] PCR 증폭물 및 제한 단편을 제조사의 지침에 따라, Qiagen (Hilden, Germany; Cat. No. 28106) 사의 QIAquick PCR Purification Kit 를 사용하여 세정하였다. DNA 를 30 μ l 의 10 mM Tris*HCl (pH 8.5) 로 용리시켰다.
- [0093] **DNA 농도 결정**
- [0094] DNA 농도는 PEQLAB Biotechnologie GmbH, 2015 년부터 VWR brand (Erlangen, Germany) 사의 NanoDrop Spectrophotometer ND-1000 을 사용하여 측정하였다.
- [0095] **어셈블리 클로닝**
- [0096] 플라스미드 벡터는 New England BioLabs Inc. (Ipswich, USA, Cat. No. E5520) 로부터 구입한 "NEBuilder HiFi DNA Assembly Cloning Kit"를 사용하여 어셈블리하였다. 선형 벡터 및 적어도 하나의 DNA 삽입물을 함유하는 반응 혼합물을 50°C 에서 60 분 동안 인큐베이션하였다. 어셈블리 혼합물 0.5 μ l 를 각 형질전환 실험에 사용하였다.
- [0097] **E. 콜라이의 화학적 형질전환**
- [0098] 플라스미드 클로닝을 위해, 화학적으로 적격인 "NEB® Stable Competent *E. coli* (High Efficiency)" (New England BioLabs Inc., Ipswich, USA, Cat. No. C3040) 을 제조사의 프로토콜에 따라 형질전환시켰다. 성공적으로 형질전환된 세포를 25 mg/1 카나마이신이 보충된 LB 아가 상에서 선별하였다.
- [0099] **C. 글루타미쿰의 형질전환**
- [0100] 플라스미드-DNA 로의 *C. 글루타미쿰*의 형질전환은 Ruan *et al.* (2015) 에 의해 기술된 바와 같이 "Gene Pulser Xcell" (Bio-Rad Laboratories GmbH, Feldkirchen, Germany) 을 사용하는 전기천공을 통해 수행하였다. 전기천공은 1 mm 전기천공 큐벳 (Bio-Rad Laboratories GmbH, Feldkirchen, Germany) 에서 1.8 kV 및 5 ms 로 설정된 고정된 시간 상수로 수행하였다. 형질전환된 세포를 134 g/l 소르비톨, 2.5 g/l 효모 추출물 및 25 mg/l 카나마이신을 함유하는 BHI-아가 상에서 선별하였다.
- [0101] **뉴클레오티드 서열 결정**
- [0102] DNA 분자의 뉴클레오티드 서열은 Applied Biosystems® (Carlsbad, CA, USA) 3730xl DNA Analyzers 상에서, Sanger *et al.* (Proceedings of the National Academy of Sciences USA 74, 5463 - 5467, 1977) 의 디데옥시 사슬 종결 방법을 사용하여, 사이클 서열분석에 의해 Eurofins Genomics GmbH (Ebersberg, Germany) 에 의해 결정하였다. Scientific & Educational Software (Denver, USA) 사의 Clonemanager Professional 9 소프트웨어를 사용하여 서열을 시각화하고 평가하였다.
- [0103] **E. 콜라이 및 C. 글루타미쿰 균주의 글리세롤 스톱**
- [0104] *E. 콜라이* 및 *C. 글루타미쿰* 균주의 장기간 저장을 위해 글리세롤 스톱을 제조하였다. 선별된 *E. 콜라이* 클론을 2 g/l 글루코스로 보충된 10 ml LB 배지에서 배양하였다. 선별된 *C. 글루타미쿰* 클론을 2 g/l 글루코스로 보충된 10 ml 2 배 농축 BHI 배지에서 배양하였다. 플라스미드 함유 *E. 콜라이* 및 *C. 글루타미쿰* 균주의

배양물에 25 mg/l 카나마이신을 보충하였다. 3 개의 배플이 있는 100 ml Erlenmeyer 플라스크에 배지를 함유하였다. 이것을 콜로니로부터 취한 세포의 루프로 접종하였다. 배양물을 30°C 및 200 rpm 에서 18 h 동안 인큐베이션하였다. 상기 인큐베이션 기간 후, 1.2 ml 85% (v/v) 멸균 글리세롤을 배양물에 첨가하였다. 얻어진 글리세롤 함유 세포 현탁액을 2 ml 분량으로 분취하여 -80°C 에서 보관하였다.

[0105] 밀리리터-규모 배양에서의 GAA 생산

[0106] Duetz (2007) 에 따른 밀리리터-규모 배양 시스템을 사용하여 균주의 GAA-생산을 평가하였다. 이를 위해, 웰 당 2.5 ml 배지로 충전된 EnzyScreen BV (Heemstede, Netherlands, Cat. no. CR1424) 사의 24-딥웰 마이크로플레이트 (24 웰 WDS 플레이트) 를 사용하였다.

[0107] 균주의 예비배양은 10 ml 종자 배지 (SM) 에서 수행하였다. 3 개의 배플이 있는 100 ml Erlenmeyer 플라스크에 배지를 함유하였다. 이를 100 µl 의 글리세롤 스톱 배양물로 접종하고, 배양물을 30°C 및 200 rpm 에서 24 h 동안 인큐베이션하였다. 종자 배지 (SM) 의 조성은 표 5 에 제시한다.

[0108] 표 5: 종자 배지 (SM)

성분	농도 (g/l)
효모 추출물 FM902 (Angel Yeast Co.,LTD, Hubei, P.R.China)	10
우레아	1.5
KH ₂ PO ₄	0.5
K ₂ HPO ₄	0.5
MgSO ₄ * 7 H ₂ O	1
비오틴	0.0001
티아민 히드로클로라이드	0.0001
FeSO ₄ * 7 H ₂ O	0.01
MnSO ₄ * H ₂ O	0.01
글루코스	20
pH=7.0	

[0109]

[0110] 상기 인큐베이션 기간 후, 예비배양물의 광학 밀도 OD₆₀₀ 을 결정하였다. 0.1 의 OD₆₀₀ 에 2.5 ml 의 생산 배지 (PM) 를 접종하는데 필요한 부피를 예비배양물로부터 샘플링하고, 원심분리 (8000 g 에서 1 분) 하고, 상청액을 폐기하였다. 그 후, 세포를 100 µl 의 생산 배지에 재현탁시켰다.

[0111] 주 배양은 24 웰 WDS-플레이트의 웰의 2.4 ml 생산 배지 (PM) 함유 웰에 예비배양물로부터의 재현탁된 세포 각 100 µl 를 접종함으로써 시작하였다. 생산 배지 (PM) 의 조성은 표 6 에 제시한다.

[0112] 표 6: 생산 배지 (PM)

성분	농도 (g/l)
3-(N-모르폴리노)프로판술폰산 (MOPS)	40
효모 추출물 FM902 (Angel Yeast Co.,LTD, Hubei, P.R.China)	1.5
(NH ₄) ₂ SO ₄	10
NH ₄ Cl	15
트리스듦 시트레이트 * 2 H ₂ O	10
우레아	1
KH ₂ PO ₄	0.5
K ₂ HPO ₄	0.5
MgSO ₄ * 7 H ₂ O	1
비오틴	0.0001
티아민 히드로클로라이드	0.0001
FeSO ₄ * 7 H ₂ O	0.01
MnSO ₄ * H ₂ O	0.01
ZnSO ₄ * 7 H ₂ O	0.000015
CuSO ₄ * 5 H ₂ O	0.0004
소포제 XFO-1501 (Ivanhoe Industries Inc., Zion, USA)	0.5
글루코스	40
IPTG (이소프로필 β-D-1-티오갈락토피라노시드)	0.3 mM
pH=7.2	

[0113]

[0114] 주 배양물을 글루코스의 완전한 소비까지 Infors GmbH (Bottmingen, Switzerland) 사의 Infors HT Multitron 표준 인큐베이터 진탕기에서 30℃ 및 300 rpm 에서 72 h 동안 인큐베이션하였다. 현탁액 중의 글루코오스 농도를 LifeScan (Johnson & Johnson Medical GmbH, Neuss, Germany) 사의 혈당계 OneTouch Vita® 로 분석하였다.

[0115] 배양 후, 배양 현탁액을 딥 웰 마이크로플레이트로 옮겼다. 배양 현탁액의 일부를 적절히 희석하여 OD600 을 측정하였다. 배양물의 또다른 부분을 원심분리하고, 상층액 중의 GAA 의 농도를 하기 기재된 바와 같이 분석하였다.

[0116] **효모 펩톤 FM902 중의 L-아르기닌 및 글리신 함량의 측정**

[0117] 효모 추출물 FM902 (Angel Yeast Co.,LTD, Hubei, P.R.China) 는 다양한 펩티드와 아미노산을 함유하므로, L-아르기닌과 글리신의 함량을 다음과 같이 측정하였다.

[0118] 유리 아미노산을 측정하기 위해, 1 g 의 효모 추출물을 20 ml 의 물에 용해시켜 샘플을 제조하였다. 용액을 25 ml 의 총 부피까지 물로 채우고, 완전히 혼합하고, 0.2 μM 나일론 시린지 필터를 사용하여 여과하였다.

[0119] 총 아미노산 (유리 아미노산 + 펩티드에 결합된 아미노산) 을 측정하기 위해, 1 g 효모 추출물을 10 ml 6M HCl 에 용해시키고, 이들을 110℃ 에서 24 h 동안 인큐베이션함으로써 샘플을 제조하였다. 그 후, 총 부피 25 ml 까지 물을 첨가하였다. 용액을 완전히 혼합하고, 0.2 μM 나일론 시린지 필터를 사용하여 여과하였다.

[0120] 샘플 중의 L-아르기닌 및 글리신의 농도는 SYKAM Vertriebs GmbH (Füstenfeldbruck, Germany) 사의 SYKAM

S433 아미노산 분석기를 사용하여 이온 교환 크로마토그래피에 의해 결정하였다. 고체상으로서, SYKAM 사의 구형의, 폴리스티렌-기반 양이온 교환기 (Peek LCA N04/Na, 치수 150 x 4.6 mm) 를 갖는 컬럼을 사용하였다.

L-아미노산에 따라, 분리하는 용리를 위한 완충액 A 및 B 의 혼합물을 사용하는 등용매 실험에서 또는 상기 완충액을 사용하는 구배 용리에 의해 일어난다. 완충액 A 로서 20 l 에 263 g 트리소듐 시트레이트, 120 g 시트르산, 1100 ml 메탄올, 100 ml 37% HCl 및 2 ml 옥탄산 (최종 pH 3.5) 을 함유하는 수용액을 사용하였다.

완충액 B 로서 20 l 에 392 g 트리소듐 시트레이트, 100 g 붕산 및 2 ml 옥탄산 (최종 pH 10.2) 을 함유하는 수용액을 사용하였다. 유리 아미노산을 포스트-컬럼 유도체화를 통해 닌히드린으로 착색하고 570 nm 에서 광도법으로 검출하였다.

[0121] 표 7 은 효모 추출물 FM902 (Angel Yeast Co.,LTD, Hubei, P.R.China) 에서 결정된 유리 및 총 L-아르기닌 및 글리신의 함량 및 생성 배지 (PM) 에서의 결과량을 나타낸다.

[0122] 표 7: 효모 추출물 (YE) FM902 중의 L-아르기닌 및 글리신의 함량 및 1.5 g/l YE 를 함유하는 생산 배지 (PM) 중의 생성된 농도.

아미노산	YE 중 유리 아미노산	YE 중 총 아미노산	PM 중 산출 유리 아미노산	PM 중 산출 총 아미노산
L-아르기닌	15.1 g/kg	32.1 g/kg	22.7 mg/l	48.2 mg/l
글리신	6.9 g/kg	30.5 g/kg	10.4 mg/l	45.7 mg/l

[0123]

[0124] **GAA 의 정량화**

[0125] 샘플을 질량 분석기 "Triple Quad 6420" (Agilent Technologies Inc., Santa Clara, USA) 와 커플링된 HPLC "Infinity 1260" 으로 구성된 Agilent 사의 분석 시스템으로 분석하였다. 35°C 에서 Atlantis HILIC Silica 컬럼, 4,6 X 250mm, 5 µm (Waters Corporation, Milford, USA) 상에서 크로마토그래피 분리를 수행하였다. 이동상 A 는 10 mM 암모늄 포르메이트 및 0.2% 포름산을 갖는 물이었다. 이동상 B 는 90% 아세트 니트릴과 10% 물의 혼합물이었고, 10 mM 암모늄 포르메이트를 혼합물에 첨가하였다. HPLC 시스템은 100% B 로 시작한 후, 22 분 동안 선형 구배 및 0,6 mL/분의 일정한 유속으로 66% B 까지 이어졌다. 질량 분석기는 ESI 양성 이온화 모드로 작동되었다. GAA 의 검출을 위해, m/z 값을 MRM 단편화 [M+H] + 118 - 76 을 사용하여 모니터링하였다. GAA 에 대한 정량화 한계 (LOQ) 는 7 ppm 으로 고정되었다.

[0126] **B) 실험 결과**

[0127] **실시예 1: 다양한 유기체로부터 L-아르기닌:글리신 아미디노트랜스퍼라제 (AGAT, EC 2.1.4.1) 를 코딩하는 유전자의 합성**

[0128] 무레아 프로듀센스 (*Moorea producens*) 는 사상성 시아노박테리아이다. 무레아 프로듀센스 균주 PAL-8-15-08-1 의 계통은 Leao et al. 에 의해 공개되었다 (Leao T, Castel $\tilde{\text{o}}$ G, Korobeynikov A, Monroe EA, Podell S, Glukhov E, Allen EE, Gerwick WH, Gerwick L, Proc Natl Acad Sci U S A. 2017 Mar 21;114(12):3198-3203. doi: 10.1073/pnas.1618556114; 수탁 번호 CP017599.1). 이는 L-아르기닌:글리신 아미디노트랜스퍼라제 (AGAT, EC 2.1.4.1, 수탁 번호 BJP34_00300, SEQ ID NO:1) 를 추정적으로 코딩하는 오픈 리딩 프레임을 함유한다. SEQ ID NO:2 및 SEQ ID NO: 4 는 AGAT_Mp 로서 지정된 유래된 아미노산 서열 (수탁 번호 WP_070390602) 을 나타낸다.

[0129] 실린드로스페르몹시스 라시보르스키 (*Cylindrospermopsis raciborskii*) AWT205 (수탁 번호 EU140798.1) 로부터의 유전자 *cyrA* 는 L-아르기닌:글리신 아미디노트랜스퍼라제를 코딩한다 (Mihali TK, Kellmann R, Muenchhoff J, Barrow KD, Neilan BA (2008) "Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis.", Appl Environ Microbiol., 74(3):716-22, doi: 10.1128/AEM.01988-07; SEQ ID NO:15). SEQ ID NO:16 및 SEQ ID NO: 26 은 AGAT_cyrA 로서 지정된 유래된 아미노산 서열 (수탁 번호 ABX60160) 을 나타낸다.

[0130] 인간 L-아르기닌:글리신 아미디노트랜스퍼라제의 cDNA 서열은 Humm et al., 1994 에 의해 기술되었다 (Humm A,

Huber R, Mann K (1994) "The amino acid sequences of human and pig l-arginine:glycine amidinotransferase.", FEBS Letters, Vol. 339 (1-2), 101-107, DOI: 10.1016/0014-5793(94)80394-3; 수탁 번호 NM_001482.3, SEQ ID NO:17). 유래된 아미노산 서열 (수탁번호 NP_001473.1, SEQ ID NO:18) 은 성숙 효소에 부재하는 미토콘드리아 전이 펩티드 (아미노산 1 내지 37) 로 시작한다. 아미노산 56 에서 시작하는, 절단된 효소는 *E. 콜라이*에서 발현될 때 활성인 것으로 밝혀졌다 (Humm A, Fritsche E, Mann K, Goehl M, Huber R (1997) "Recombinant expression and isolation of human L-arginine : glycine amidinotransferase and identification of its active-site cysteine residue." Biochem. J. 322, 771-776, DOI: 10.1042/bj3220771). 7 개 아미노산 태그 (SEQ ID NO:19) 의 N-말단 융합은 *E. 콜라이*에서 단백질 발현을 개선시키는 것으로 나타났다 (Hansted JG, Pietikainen L, Hoeg F, Sperling-Petersen HU, Mortensen KK (2011) "Expressivity tag: A novel tool for increased expression in Escherichia coli." Journal of Biotechnology 155 (2011) 275- 283, DOI:10.1016/j.jbiotec.2011.07.013). 따라서, 태그 및 절두된 AGAT 로 이루어진 융합 단백질을 설계하고, 이를 AGAT_Hs (SEQ ID NO:20 및 SEQ ID NO:28) 로 지정하였다.

[0131] 라투스 노르베지쿠스 (*Rattus norvegicus*) 로부터의 L-아르기닌:글리신 아미디노트랜스퍼라제의 아미노산 서열 (수탁 번호 NP_112293.1, SEQ ID NO:21) 은 인간 효소와 매우 유사하다. 인간 효소에 대해 기재된 바와 같이, 상기 서열을 사용하여 N-말단 발현 태그 및 효소의 절단된 서열로 이루어진 융합 단백질을 설계하였다. 생성된 융합 단백질을 AGAT_Rn (SEQ ID NO:22 및 SEQ ID NO:30) 으로 지정하였다.

[0132] 순다 날개여우원숭이 (sunda flying lemur) 갈레옵테루스 바리에가투스 (*Galeopterus variegatus*) 는 예측된 L-아르기닌:글리신 아미디노트랜스퍼라제 (수탁 번호 NP_112293.1, SEQ ID NO:23) 를 갖는다. 인간 효소에 대해 기재된 바와 같이, 이의 아미노산 서열을 사용하여 N-말단 발현 태그 및 효소의 절단된 서열로 이루어진 융합 단백질을 설계하였다. 생성된 융합 단백질을 AGAT_Gv (SEQ ID NO:24 및 SEQ ID NO:32) 으로 지정하였다.

[0133] 소프트웨어 툴 "GeneOptimizer" (Geneart/ ThermoFisher Scientific, Waltham, USA) 를 사용하여, AGAT_Mp, AGAT_cyrA, AGAT_Hs, AGAT_Rn 및 AGAT_Gv 의 아미노산 서열을 DNA 서열로 다시 번역하고 *C. 글루타미쿰*의 코돈 사용에 최적화하였다. 이들의 말단은 어셈블리 클로닝을 위한 서열로 확장되었고, 샤인-달가노-서열 (Shine-Dalgarno-Sequence) 은 오픈 리딩 프레임의 업스트림에 추가되었다. 생성된 DNA 서열은 SEQ ID NO:3 (AGAT_Mp 를 코딩함), SEQ ID NO:25 (AGAT_cyrA 를 코딩함), SEQ ID NO:27 (AGAT_Hs 를 코딩함), SEQ ID NO: 29 (AGAT_Rn 을 코딩함) 및 SEQ ID NO:31 (AGAT_Gv 를 코딩함) 이다. 이들은 Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, USA) 로부터 유전자 합성을 위해 주문하였다. 합성 유전자는 pMA-T_AGAT_Mp, pMA-T_AGAT_cyrA, pMA-T_AGAT_Hs, pMA-T_AGAT_Rn 및 pMA-T_AGAT_Gv 로서 지정된 클로닝 플라스미드의 일부로서 전달되었다.

[0134] **실시예 2: AGAT_Mp 의 발현 플라스미드 pEC-XK99E 내로의 클로닝**

[0135] *E. 콜라이-C. 글루타미쿰* 서플 플라스미드 pEC-XK99E 를 제한 엔도뉴클레아제 SmaI 를 사용하여 소화시켰다. "FastAP Thermosensitive Alkaline Phosphatase" (Thermo Fisher Scientific, Waltham, USA) 를 사용하여 말단 포스페이트를 제거하였다. 이후, DNA 를 "QIAquick PCR Purification Kit" (Qiagen GmbH, Hilden, Germany) 로 정제하였다.

[0136] 클로닝 플라스미드 pMA-T_AGAT_Mp 를 MluI + AatII 로 소화시키고, 생성된 단편을 "Fast DNA End Repair Kit" (Thermo Fisher Scientific, Waltham, USA) 를 사용하여 블런트화하였다. 이들을 아가로스 겔 전기영동 (TAE 완충액 중 0,8% 아가로스) 으로 분리하고, "AGAT_Mp" (1174 bp) 에 해당하는 밴드를 잘라냈다. 이의 DNA 를 "QIAquick Gel Extraction Kit" (Qiagen GmbH, Hilden, Germany) 를 사용하여 정제하였다.

[0137] AGAT_Mp 단편 및 선형화된 pEC-XK99E 를 "Ready-To-Go T4 DNA ligase" (GE Healthcare Europe GmbH, Freiburg, Germany) 를 사용하여 라이게이션하였다. 라이게이션 생성물을 "NEB Stable Competent *E. coli* (High Efficiency)" (New England Biolabs, Ipswich, USA) 내로 형질전환시키고, 세포를 25 mg/1 카나마이신을 함유하는 LB 아가 상에서 성장시켰다. 적절한 클론은 제한 효소 소화 및 DNA 서열분석에 의해 확인되었다. 생성된 플라스미드를 pEC-XK99E_AGAT_Mp 로 명명하였다.

[0138] **실시예 2.1: AGAT_Mp 및 AGAT_cyrA 의 발현 플라스미드 pEKEx2 내로의 클로닝**

[0139] *E. 콜라이-C. 글루타미쿰* 서플 플라스미드 pEKEx2 (Eikmanns, 1991) 를 제한 엔도뉴클레아제 PstI 를 사용하여 소화시켰다. 생성된 단편을 "Fast DNA End Repair Kit" (Thermo Fisher Scientific) 를 사용하여 블런트화

하고, 말단 포스페이트를 "FastAP Thermosensitive Alkaline Phosphatase" (Thermo Fisher Scientific, Waltham, USA) 로 제거하였다. 이후, DNA 를 "QIAquick PCR Purification Kit" (Qiagen GmbH, Hilden, Germany) 로 정제하였다.

- [0140] 클로닝 플라스미드 pMA-T_AGAT_Mp 및 pMA-T_AGAT_cyrA 를 MluI + AatII 로 소화시키고, 생성된 단편을 "Fast DNA End Repair Kit" (Thermo Fisher Scientific, Waltham, USA) 를 사용하여 블린트화하였다. 이들을 아가로스 겔 전기영동 (TAE 완충액 중 0,8% 아가로스) 으로 분리하고, AGAT_Mp (1174 bp) 및 AGAT_cyrA (1204 bp) 에 해당하는 밴드를 잘라냈다. DNA 를 "QIAquick Gel Extraction Kit" (Qiagen GmbH, Hilden, Germany) 를 사용하여 정제하였다.
- [0141] 각각의 AGAT 단편을 "Ready-To-Go T4 DNA ligase" (GE Healthcare Europe GmbH, Freiburg, Germany) 를 사용하여 선형화된 pEKEEx2 로 라이게이션하였다. 라이게이션 생성물을 "NEB Stable Competent E. coli (High Efficiency)" (New England Biolabs, Ipswich, USA) 내로 형질전환시키고, 세포를 25 mg/1 카나마이신을 함유하는 LB 아가 상에서 성장시켰다. 적절한 클론은 제한 효소 소화 및 DNA 서열분석에 의해 확인되었다. 생성된 플라스미드를 pEKEEx2_AGAT_Mp 및 pEKEEx2_AGAT_cyrA 로 명명하였다.
- [0142] **실시예 2.2: AGAT_Hs, AGAT_Rn 및 AGAT_Gv 의 발현 플라스미드 pEKEEx2 내로의 클로닝**
- [0143] 클로닝 플라스미드 pMA-T_AGAT_Hs, pMA-T_AGAT_Rn 및 pMA-T_AGAT_Gv 를 Eco31I 로 소화시키고, 생성물을 "QIAquick Gel Extraction Kit" (Qiagen GmbH, Hilden, Germany) 를 사용하여 정제하였다.
- [0144] *E. 콜라이-C. 글루타미쿰* 셔틀 플라스미드 pEKEEx2 (Eikmanns BJ, Kleinertz E, Liebl W, Sahm H (1991) "A family of Corynebacterium glutamicum/Escherichia coli shuttle vectors for cloning, controlled gene expression, and promoter probing.", Gene. 1991 Jun 15;102(1):93-8) 를 제한 엔도뉴클레아제 SbfI 및 BamHI 를 사용하여 소화시켰다. DNA 를 "QIAquick PCR Purification Kit" (Qiagen GmbH, Hilden, Germany) 로 정제하였다.
- [0145] 각각의 AGAT 단편은 "NEBuilder HiFi DNA Assembly Cloning Kit" (New England Biolabs Inc., Ipswich, USA, Cat. No. E5520) 를 사용하여 소화된 pEKEEx2 와 어셈블리하였다. 어셈블리 생성물을 "NEB Stable Competent E. coli (High Efficiency)" (New England Biolabs, Ipswich, USA) 내로 형질전환시키고, 세포를 25 mg/1 카나마이신을 함유하는 LB 아가 상에서 성장시켰다. 적절한 클론은 제한 효소 소화 및 DNA 서열분석에 의해 확인되었다. 생성된 플라스미드를 각각 pEKEEx2_AGAT_Hs, pEKEEx2_AGAT_Rn 및 pEKEEx2_AGAT_Gv 로 명명하였다.
- [0146] **실시예 3: 유전자 argF 의 플라스미드 pCR-Blunt II-TOPO 내로의 클로닝**
- [0147] *argF* 유전자를 *C. 글루타미쿰* ATCC13032 의 게놈 DNA 와 올리고뉴클레오타이드 프라이머 argF_1.p (SEQ ID NO:5) 및 argF_2.p (SEQ ID NO:6) 를 사용하여 Phusion High-Fidelity DNA Polymerase Kit (New England Biolabs Inc., Ipswich, USA) 로 PCR 증폭하였다. 생성된 PCR 생성물을 플라스미드 pCR-Blunt II-TOPO (Thermo Fisher Scientific/Invitrogen, Waltham, USA) 내에 클로닝하고, 적절한 플라스미드 클론을 제한 효소 소화 및 DNA 서열분석에 의해 확인하였다. 이 플라스미드를 pCRII-argF 로 명명하였다.
- [0148] **실시예 4: 유전자 argG 및 argH 의 플라스미드 pCR-Blunt II-TOPO 내로의 클로닝**
- [0149] 유전자 *argG* 및 *argH* 를 *C. 글루타미쿰* ATCC13032 의 게놈 DNA 와 올리고뉴클레오타이드 프라이머 argG_1.p (SEQ ID NO:7) 및 argH_2.p (SEQ ID NO:8) 를 사용하여 Phusion High-Fidelity DNA Polymerase Kit (New England Biolabs Inc., Ipswich, USA) 로 PCR 증폭하였다. 생성된 PCR 생성물을 플라스미드 pCR-Blunt II-TOPO (Thermo Fisher Scientific/Invitrogen, Waltham, USA) 내에 클로닝하고, 적절한 플라스미드 클론을 제한 효소 소화 및 DNA 서열분석에 의해 확인하였다. 플라스미드를 pCRII-argGH 로 명명하였다.
- [0150] **실시예 5: 유전자 argG 및 argH 의 플라스미드 pCRII-argF 내로의 클로닝**
- [0151] pCRII-argGH 를 HpaI + AvrII 를 사용하여 절단하고, 2773 bps 의 제한 단편을 아가로스 겔로부터 단리하였다. pCRII-argF 를 SspI + AvrII 를 사용하여 절단하고, 4526 bps 의 제한 단편을 아가로스 겔로부터 단리하였다. 두 단편을 라이게이션한 후 *E. 콜라이*로 형질전환하였다. 적절한 플라스미드 클론은 제한 효소 소화 및 DNA 서열분석에 의해 확인되었다. 생성된 플라스미드를 pCRII-argFGH 로 명명하였다.
- [0152] **실시예 6: 유전자 argF, argG 및 argH 의 발현 플라스미드 pEC-XK99E 내로의 클로닝**

- [0153] pCRII-argFGH 를 HpaI + AvrII 를 사용하여 절단하고, 2773 bps 의 제한 단편을 아가로스 겔로부터 단리하였다. 플라스미드 pEC-XK99E 를 Ecl136II + XbaI 를 사용하여 절단하였다. 6999 bps 의 제한 단편을 아가로스 겔로부터 단리하였다. 두 단편을 라이게이션한 후 *E. 콜라이*로 형질전환하였다. 적절한 플라스미드 클론은 제한 소화 및 DNA 서열분석에 의해 확인되었다. 생성된 플라스미드 pEC-XK99E_argFGH 는 *C. 글루타미쿰*으로부터의 유전자 *argF*, *argG* 및 *argH* 를 함유한다.
- [0154] **실시예 7: 유전자 *argF*, *argG* 및 *argH* 의 발현 플라스미드 pEC-XK99E_AGAT_Mp 내로의 클로닝**
- [0155] pCRII-argFGH 를 XbaI + SpeI 를 사용하여 절단하고, 3868 bps 의 제한 단편을 아가로스 겔로부터 단리하였다. 플라스미드 pEC-XK99E_AGAT_Mp 를 XbaI 를 사용하여 절단하였다. 8188 bps 의 제한 단편을 아가로스 겔로부터 단리하였다. 두 단편을 라이게이션한 후 *E. 콜라이*로 형질전환하였다. 적절한 플라스미드 클론은 제한 소화 및 DNA 서열분석에 의해 확인되었다. 생성된 플라스미드 pEC-XK99E_AGAT_Mp_argFGH 는 AGAT_Mp 와 조합으로 *C. 글루타미쿰*으로부터의 유전자 *argF*, *argG* 및 *argH* 를 함유한다.
- [0156] **실시예 8: 유전자 *argF* 의 발현 플라스미드 pEC-XK99E_AGAT_Mp 내로의 클로닝**
- [0157] pCRII-argF 를 KpnI + XbaI + AseI 를 사용하여 절단하고, DNA 를 "QIAquick PCR Purification Kit" (Qiagen GmbH, Hilden, Germany) 로 정제하였다. 플라스미드 pEC-XK99E_AGAT_Mp 를 KpnI + XbaI 를 사용하여 절단하고, DNA 를 "QIAquick PCR Purification Kit" (Qiagen GmbH, Hilden, Germany) 로 정제하였다. 두 용리액을 혼합하고, DNA 단편을 라이게이션하고, 생성물을 사용하여 *E. 콜라이*를 형질전환시켰다. 적절한 플라스미드 클론은 제한 소화 및 DNA 서열분석에 의해 확인되었다. 생성된 플라스미드 pEC-XK99E_AGAT_Mp_argF 는 AGAT_Mp 와 조합으로 *C. 글루타미쿰*으로부터의 유전자 *argF* 를 함유한다.
- [0158] **실시예 9: 유전자 *argG* 의 발현 플라스미드 pEC-XK99E_AGAT_Mp 내로의 클로닝**
- [0159] pCRII-argGH 를 XbaI + SalI 를 사용하여 절단하고, 1798 bps 의 제한 단편을 아가로스 겔로부터 단리하였다. 플라스미드 pEC-XK99E_AGAT_Mp 를 XbaI + SalI 를 사용하여 절단하고, DNA 를 "QIAquick PCR Purification Kit" (Qiagen GmbH, Hilden, Germany) 로 정제하였다. DNA 단편을 라이게이션하고, 생성물을 사용하여 *E. 콜라이*를 형질전환시켰다. 적절한 플라스미드 클론은 제한 소화 및 DNA 서열분석에 의해 확인되었다. 생성된 플라스미드 pEC-XK99E_AGAT_Mp_argG 는 AGAT_Mp 와 조합으로 *C. 글루타미쿰*으로부터의 유전자 *argG* 를 함유한다.
- [0160] **실시예 10: ATCC13032 내 *carAB* 오페론의 업스트림 *sod* 프로모터의 염색체 삽입**
- [0161] 카르바모일 포스페이트 신테타제의 효소 활성은 ATCC13032 내 *carAB* 오페론의 업스트림 강한 *sod*-프로모터의 게놈 삽입에 의해 증가되었다. 따라서, 플라스미드 pK18mobsacB_Psod-carAB 는 다음과 같이 구성되었다. 플라스미드 pK18mobsacB 를 EcoRI + HindIII 를 사용하여 절단하고, 선형화된 벡터 DNA (5670 bps) 를 아가로스 겔로부터 절단하였다. DNA 를 "QIAquick Gel Extraction Kit" (QIAGEN GmbH, Hilden, Germany) 를 사용하여 추출하였다.
- [0162] 삽입체를 구성하기 위해, 3 개의 DNA 단편을 하기 프라이머 쌍 (주형으로서 ATCC13032 의 게놈 DNA) 을 사용하는 PCR 에 의해 생성하였다:
- [0163] PsodcarAB-LA-F (SEQ ID NO:9) + PsodcarAB-LA-R (SEQ ID NO:10)
- [0164] = 좌측 상동 암 (arm) (1025 bps)
- [0165] PsodcarAB-F (SEQ ID NO:11) + PsodcarAB-R (SEQ ID NO:12)
- [0166] = *sod*-프로모터 (250 bps)
- [0167] PsodcarAB-RA-F SEQ ID NO:13) + PsodcarAB-RA-R (SEQ ID NO:14)
- [0168] = 우측 상동 암 (944 bps)
- [0169] 생성물 DNA 를 "QIAquick PCR Purification Kit" (Qiagen GmbH, Hilden, Germany) 를 사용하여 정제하였다. 선형화된 플라스미드 및 PCR 생성물은 이후 "NEBuilder HiFi DNA Assembly Cloning Kit" (New England BioLabs Inc., Ipswich, USA, Cat. No. E5520) 를 사용하여 어셈블리하였다. 적절한 플라스미드 클론은 제한 소화 및 DNA 서열분석에 의해 확인되었다.

[0170] pK18mobsacB_Psod-carAB 를 사용하여 *carAB* 유전자의 업스트림 강한 *sod*-프로모터를 *C. 글루타미쿰 ATCC13032* 의 계놈 내로 통합하였다. 플라스미드를 전기천공에 의해 ATCC13032 내로 형질전환시켰다. 염색체 통합 (제 1 재조합 사건으로부터 초래됨) 을 134 g/l 소르비톨, 2.5 g/l 효모 추출물 및 25 mg/l 카나마이신이 보충된 BHI 아가 상에 플레이트함으로써 선별하였다. 아가 플레이트를 33°C 에서 48 시간 동안 인큐베이션한다.

[0171] 개별 콜로니를 신선한 아가 플레이트 (카나마이신 25 mg/l 함유) 로 옮기고 33°C 에서 24 시간 동안 인큐베이션 하였다. 이들 클론의 액체 배양물을 3 개의 배플이 있는 100 ml Erlenmeyer 플라스크에 함유된 10 ml BHI 배지에서 24 시간 동안 33°C 에서 배양하였다. 제 2 재조합 사건에 직면한 클론을 단리하기 위해, 각각의 액체 배양물로부터 분취물을 취하고, 적당하게 희석하고, 10% 사카로오스가 보충된 BHI 아가 상에 플레이트하였다 (전형적으로 100 내지 200 μ l). 이들 아가 플레이트를 33°C 에서 48 시간 동안 인큐베이션하였다. 이어서, 사카로오스 함유 아가 플레이트 상에서 성장하는 콜로니를 카나마이신 민감성에 대해 검사하였다. 이를 위해, 이쑤시개를 사용하여 콜로니로부터 세포 물질을 제거하고, 이를 25 mg/l 카나마이신을 함유하는 BHI 아가 및 10% 사카로오스를 함유하는 BHI 아가 상에 전달하였다. 아가 플레이트를 60 시간 동안 33°C 에서 인큐베이션하였다. 카나마이신에 민감하고 사카로오스에 내성인 것으로 입증된 클론을 *sod* 프로모터의 적절한 통합을 위해 PCR 및 DNA 서열분석에 의해 검사하였다. 생성된 균주를 ATCC13032_Psod-carAB 로 명명하였다.

[0172] 표 8: 균주 목록

균주	코멘트
<i>에스케리치아 콜라이</i>	
NEB® Stable	시판 클로닝 균주 (New England BioLabs Inc., Ipswich, USA)
MG1655	<i>E. 콜라이</i> 참조 균주 (Blattner <i>et al.</i> , 1997*)
<i>코리네박테리움 글루타미쿰</i>	
ATCC13032	<i>코리네박테리움 글루타미쿰</i> 야생형 균주 (Kinoshita <i>et al.</i> , 1957**)
ATCC21831	L-아르기닌 생산 <i>C. 글루타미쿰</i> 변이체 (Park <i>et al.</i> , 2014***)
ATCC13032__Psod-carAB	<i>C. 글루타미쿰</i> ATCC13032 에서 <i>carAB</i> 의 업스트림 강한 <i>sod</i> 프로모터의 염색체 통합

*) Blattner *et al.*, 1997: Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y, Science. 1997 Sep 5;277(5331):1453-62. (doi: 10.1126/science.277.5331.1453)

**) Kinoshita S, Udaka S, Shimono M., J. Gen. Appl. Microbiol. 1957; 3(3): 193-205.

***) Park *et al.*, 2014: Park SH, Kim HU, Kim TY, Park JS, Kim SS, Lee, Nat Commun. 2014 Aug 5; 5:4618. (doi: 10.1038/ncomms5618)

[0173]

[0174] 표 9: 플라스미드 목록

플라스미드	코멘트
pMA-T_AGAT_Mp	AGAT_Mp (<i>무레아 프로듀센스</i>)에 대한 유전자 합성 및 클로닝 벡터; Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, USA)의 전매 플라스미드
pMA-T_AGAT_cyrA	AGAT_cyrA (<i>실린드로스페르몐시스 라시보르스키</i> AWT205)에 대한 유전자 합성 및 클로닝 벡터; Invitrogen/Geneart (Thermo

[0175]

플라스미드	코멘트
	Fisher Scientific, Waltham, USA)의 전매 플라스미드
pMA-T_AGAT_Hs	AGAT_Hs (<i>호모 사피엔스</i>)에 대한 유전자 합성 및 클로닝 벡터; Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, USA)의 전매 플라스미드
pMA-T_AGAT_Rn	AGAT_Rn (<i>라투스 노르베지쿠스</i>)에 대한 유전자 합성 및 클로닝 벡터; Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, USA)의 전매 플라스미드
pMA-T_AGAT_Gv	AGAT_Gv (<i>Galeopterus variegatus</i>)에 대한 유전자 합성 및 클로닝 벡터; Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, USA)의 전매 플라스미드
pEKEx2	빈 <i>E. 콜라이</i> - <i>C. 글루타미쿰</i> 서플 플라스미드
pEKEx2_AGAT_Mp	AGAT_Mp (<i>무레아 프로듀센스</i>)의 발현
pEKEx2_AGAT_cyrA	AGAT_cyrA (<i>실린드로스페르몐시스 라시보르스키</i> AWT205)의 발현
pEKEx2_AGAT_Hs	AGAT_Hs (<i>호모 사피엔스</i>)의 발현
pEKEx2_AGAT_Rn	AGAT_Rn (<i>라투스 노르베지쿠스</i>)의 발현
pEKEx2_AGAT_Gv	AGAT_Gv (<i>Galeopterus variegatus</i>)의 발현
pEC-XK99E	빈 <i>E. 콜라이</i> - <i>C. 글루타미쿰</i> 서플 플라스미드
pEC-XK99E_AGAT_Mp	AGAT_Mp (<i>무레아 프로듀센스</i>)의 발현
pCR-Blunt II-TOPO	클로닝 PCR 생성물에 대한 빈 플라스미드 (Thermo Fisher Scientific/Invitrogen, Waltham, USA)
pCRII-argF	클로닝된 유전자 <i>argF</i> (<i>C. 글루타미쿰</i>)를 함유하는 pCR-Blunt II-TOPO
pCRII-argGH	클로닝된 유전자 <i>argG</i> , <i>argH</i> (<i>C. 글루타미쿰</i>)를 함유하는 pCR-Blunt II-TOPO

[0176]

플라스미드	코멘트
pCRII-argFGH	클로닝된 유전자 <i>argF</i> , <i>argG</i> , <i>argH</i> (<i>C. 글루타미쿰</i>) 를 함유하는 pCR-Blunt II-TOPO
pEC-XK99E_argFGH	<i>argF</i> , <i>argG</i> , <i>argH</i> (<i>C. 글루타미쿰</i>) 의 발현
pEC-XK99E_AGAT_Mp_argFGH	AGAT_Mp (<i>무레아 프로듀센스</i>), <i>argF</i> , <i>argG</i> , <i>argH</i> (<i>C. 글루타미쿰</i>) 의 발현
pEC-XK99E_AGAT_Mp_argGH	AGAT_Mp (<i>무레아 프로듀센스</i>), <i>argG</i> , <i>argH</i> (<i>C. 글루타미쿰</i>) 의 발현
pEC-XK99E_AGAT_Mp_argG	AGAT_Mp (<i>무레아 프로듀센스</i>), <i>argG</i> (<i>C. 글루타미쿰</i>) 의 발현
pEC-XK99E_AGAT_Mp_argF	AGAT_Mp (<i>무레아 프로듀센스</i>), <i>argF</i> (<i>C. 글루타미쿰</i>) 의 발현
pEC-XK99E_AGAT_Mp_argG	AGAT_Mp (<i>무레아 프로듀센스</i>), <i>argG</i> (<i>C. 글루타미쿰</i>) 의 발현
pK18mobsacB	<i>C. 글루타미쿰</i> 의 계놈 개질을 위한 빈 플라스미드
pK18mobsacB_Psod-carAB	<i>C. 글루타미쿰</i> 에서 <i>carAB</i> 의 업스트림 강한 <i>sod</i> 프로모터의 염색체 통합을 위한 플라스미드

[0177]

[0178] **실시예 11: 다양한 발현 플라스미드로의 *C. 글루타미쿰* 균주의 형질전환**

[0179] 하기 *C. 글루타미쿰* 균주를 다양한 플라스미드로 형질전환시켰다:

[0180] · *C. 글루타미쿰* ATCC13032 ATCC13032 (Kinoshita *et al.*, J. Gen. Appl. Microbiol. 1957; 3(3): 193-205) 가 통상 사용되는 야생형 균주이다.

[0181] · *C. 글루타미쿰* 균주 ATCC21831 (Park *et al.*, Nat Commun. 2014 Aug 5; 5:4618) 는 암모니아 및 글루코스와 같은 일차 기질로부터 L-아르기닌을 합성한다.

[0182] · ATCC13032_Psod-carAB 는 *carAB* 유전자의 업스트림 강한 *sod* 프로모터를 갖는 ATCC13032 의 변이체이다.

[0183] 균주를 다양한 플라스미드 (표 10 에 나타낸 바와 같음) 로 전기천공함으로써 형질전환시켰다. 플라스미드 함유 세포를 25 mg/l 카나마이신으로 선별하였다.

[0184] 표 10: 플라스미드-함유 *C. 글루타미쿰* 균주 목록

플라스미드	클로닝된 유전자	수여 균주	산출 균주
pEKEx2	없음	ATCC21831	ATCC21831/pEKEx2
pEKEx2_AGAT_Mp	AGAT_Mp	ATCC21831	ATCC21831/pEKEx2_AGAT_Mp
pEKEx2_AGAT_Hs	AGAT_Hs	ATCC21831	ATCC21831/pEKEx2_AGAT_Hs
pEKEx2_AGAT_Rn	AGAT_Rn	ATCC21831	ATCC21831/pEKEx2_AGAT_Rn
pEKEx2_AGAT_Gv	AGAT_Gv	ATCC21831	ATCC21831/pEKEx2_AGAT_Gv
pEKEx2_AGAT_cyrA	AGAT_cyrA	ATCC21831	ATCC21831/pEKEx2_AGAT_cyrA
pEC-XK99E	없음	ATCC13032	ATCC13032/pEC-XK99E
pEC-XK99E_argFGH	<i>argF</i> , <i>argG</i> , <i>argH</i>	ATCC13032	ATCC13032/pEC-XK99E_argFGH
pEC-XK99E_AGAT_Mp	AGAT_Mp	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp
pEC-XK99E_AGAT_Mp_argG	AGAT_Mp, <i>argG</i>	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp_argG
pEC-XK99E_AGAT_Mp_argF	AGAT_Mp, <i>argF</i>	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp_argF
pEC-XK99E_AGAT_Mp_argGH	AGAT_Mp, <i>argG</i> , <i>argH</i>	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp_argGH
pEC-XK99E_AGAT_Mp_argFGH	AGAT_Mp, <i>argF</i> , <i>argG</i> , <i>argH</i>	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp_argFGH
pEC-XK99E	없음	ATCC13032_Psod-carAB	ATCC13032_Psod-carAB/pEC-XK99E
pEC-XK99E_AGAT_Mp_argFGH	AGAT_Mp, <i>argF</i> , <i>argG</i> , <i>argH</i>	ATCC13032_Psod-carAB	ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argFGH

[0185]

[0186] 실시예 12: GAA 생산에 대한 AGAT 및 기질 이용능의 영향

[0187]

균주 ATCC13032/pEC-XK99E_AGAT_Mp (*무레아 프로듀센스*로부터의 AGAT 효소에 대한 유전자를 보유함) 및 ATCC13032/pEC-XK99E (대조균을 위한 빈 벡터) 를 Wouter Duetz 의 시스템 (상기 기재됨) 을 사용하여 배치 배양에서 GAA 를 생산하는 능력에 대해 분석하였다. 생산 배지 (PM) 는 40 g/l D-글루코스를 주 탄소원으로 함유하였다. 일부 배치에는 지시된 바와 같이 L-아르기닌 및/또는 글리신을 보충하였다.

[0188] 표 11: 균주 ATCC13032/pEC-XK99E 및 ATCC13032/pEC-XK99E_AGAT_Mp 에 의한 GAA 생산

균주	보충	GAA
ATCC13032/pEC-XK99E	없음	검출 불가능
ATCC13032/pEC-XK99E	1.90 g/l Arg 및 0.818 g/l Gly	검출 불가능
ATCC13032/pEC-XK99E_AGAT_Mp	없음	25 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp	0.818 g/l Gly	31 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp	1.90 g/l Arg 및 0.818 g/l Gly	124 mg/l

[0189]

[0190]

표 11 에 나타난 바와 같이, 대조군 균주 ATCC13032/pEC-XK99E 는 전구체 L-아르기닌 및 글리신을 제공하더라도 GAA 를 생산할 수 없었다. 본 발명자들은 그것이 본질적인 AGAT 활성을 가지지 않는다고 결론지었다. 균주 ATCC13032/pEC-XK99E_AGAT_Mp 는 무레아 프로듀센스로부터의 추정 AGAT 를 코딩하는 폴리뉴클레오티드를 함유한다. 이것은 보충되지 않은 PM 에서 25 mg/l GAA 를 생성하였다. 글리신의 보충은 31 mg/l GAA 로의 작은 증가를 초래하였다. 글리신 및 L-아르기닌의 보충은 GAA 생산을 124 mg/l 로 크게 증가시켰다.

[0191]

실시예 13: 일차 기질로부터 GAA 의 생산

[0192]

산업적 GAA 생산 공정에서, L-아르기닌의 보충은 암모니아, 우레아 및 글루코스와 같은 일차 기질과 비교할 때 오히려 비용이 많이 들 것이다. 따라서, 이러한 일차 기질로부터 직접 GAA 를 생성하는 것이 바람직할 것이다.

[0193]

이를 위해, L-아르기닌 생산자 *C. 글루타미쿰* ATCC21831 을:

[0194]

- pEKEx2 (대조군을 위한 빈 벡터),

[0195]

- pEKEx2_AGAT_Mp (무레아 프로듀센스 (*Moorea producens*) 로부터의 AGAT_Mp 유전자 함유),

[0196]

- pEKEx2_AGAT_Hs (호모 사피엔스 (*Homo sapiens*) 로부터의 AGAT_Hs 유전자 함유),

[0197]

- pEKEx2_AGAT_Rn (라투스 노르베지쿠스 (*Rattus norvegicus*) 로부터의 AGAT_Rn 유전자 함유),

[0198]

- pEKEx2_AGAT_Gv (갈레옵테루스 바리에가투스 (*Galeopterus variegatus*), 및 로부터의 AGAT_Gv 유전자 함유) 및

[0199]

- pEKEx2_AGAT_cyrA (실린드로스페르모프시스 라시보르스키 (*Cylindrospermopsis raciborskii*) 로부터의 AGAT_cyrA 유전자 함유)

[0200]

로 형질전환시켰다.

[0201]

ATCC21831 은 카나바닌 내성 돌연변이체로서 단리되었고, 이는 L-아르기닌을 생산하는 것으로 밝혀졌다. 이의 계통은 Park *et al.* 에 의해 서열분석되었고 (Nat Commun. 2014 Aug 5;5:4618. doi: 10.1038/ncomms5618; accession number CP007722), 균주는 LGC 표준 (LGC Standards GmbH, Wesel, Germany) 으로부터 공개적으로 입수가능하다.

[0202]

모든 형질전환된 ATCC21831 균주를 Wouter Duetz 의 시스템 (상기 기재됨) 을 사용하여 배치 배양에서 GAA 를 생산하는 그들의 능력에 대해 분석하였다. 생산 배지 (PM) 는 40 g/l D-글루코스를 주 탄소원으로 함유하였다. 일부 배지는 글리신 및/또는 L-아르기닌으로 보충되었다.

[0203]

표 12: 1.5 g/l 효모 추출물을 갖는 CGAF 및 MOPS 배지에서 상이한 종으로부터의 AGAT 를 함유하는 pEKEx2 벡터

로 형질전환된 ATCC21831 균주에 의한 GAA 생산

균주	보충	GAA
ATCC21831/pEKEEx2	없음	검출 불가능
ATCC21831/pEKEEx2	1.90 g/l 아르기닌, 0.82 g/l 글리신	검출 불가능
ATCC21831/pEKEEx2_AGAT_Mp	없음	26.5 mg/l
ATCC21831/pEKEEx2_AGAT_Mp	0.82 g/l 글리신	166.3 mg/l
ATCC21831/pEKEEx2_AGAT_Hs	없음	3.6 mg/l
ATCC21831/pEKEEx2_AGAT_Hs	0.82 g/l 글리신	56.2 mg/l

[0204]

균주	보충	GAA
ATCC21831/pEKEEx2_AGAT_Rn	없음	0.9 mg/l
ATCC21831/pEKEEx2_AGAT_Rn	0.82 g/l 글리신	7.6 mg/l
ATCC21831/pEKEEx2_AGAT_Gv	없음	0.7 mg/l
ATCC21831/pEKEEx2_AGAT_Gv	0.82 g/l 글리신	8.5 mg/l
ATCC21831/pEKEEx2_AGAT_cyrA	없음	20.8 mg/l
ATCC21831/pEKEEx2_AGAT_cyrA	0.82 g/l 글리신	52.7 mg/l

[0205]

[0206]

표 12 에 나타난 바와 같이, ATCC21831/pEKEEx2 는 심지어 전구체 L-아르기닌 및 글리신이 존재하는 경우에도, GAA 를 생산하지 않았다. 본 발명자들은 ATCC21831/ pEKEEx2 가 본질적인 AGAT 활성을 가지지 않는다고 결론 지었다. 형질전환된 균주 ATCC21831/ pEKEEx2_AGAT_Mp, ATCC21831/pEKEEx2_AGAT_Hs, ATCC21831/pEKEEx2_AGAT_Rn, ATCC21831/pEKEEx2_AGAT_Gv, 및 ATCC21831/pEKEEx2_AGAT_cyrA 는 보충되지 않은 PM 에서 약 1 내지 26 mg/l GAA 를 생산하였다. 본 발명자들은 전구체 L-아르기닌 및 글리신이 일차 기질 D-글루코스, 암모늄 및 우레아로부터 합성되었다고 결론내렸다.

[0207]

글리신을 첨가하였을 때, ATCC21831/pEKEEx2_AGAT_Mp, ATCC21831/pEKEEx2_AGAT_Hs, ATCC21831/pEKEEx2_AGAT_Rn, ATCC21831/pEKEEx2_AGAT_Gv, 및 ATCC21831/pEKEEx2_AGAT_cyrA 의 GAA 생산은 보충되지 않은 실험에 비해 상당히 증가하였다. 균주 ATCC13032/pEC-XK99E_AGAT_Mp (표 10 참조) 와 비교할 때, AGAT 유전자를 갖는 L-아르기닌 생산자 ATCC21831 은 글리신이 제한되지 않을 때 훨씬 더 많은 GAA 를 축적한다. 본 발명자들은 L-아르기닌을 내부적으로 제공하는 능력이 GAA 생산을 향상시킨다고 결론지었다.

[0208]

실시예 14: GAA 생산에 대한 L-아르기닌 재생의 영향

[0209] L-아르기닌 생산 균주 (예를 들어, ATCC21831) 에서, 중간체 L-오르니틴을 드 노브로 합성하고, 추가로 L-아르기닌으로 전환시킨다. 이러한 균주에 AGAT 가 제공되는 경우, 효소는 동일한 몰량의 GAA 및 L-오르니틴을 생성할 것이다. L-오르니틴의 형성은 상당한 양의 일차 C- 및 N-공급원을 소비하고, 따라서 GAA 의 수율을 저하시킨다.

[0210] 본 발명자들은 L-오르니틴으로부터 L-아르기닌에 이르는 생합성 경로의 향상이, 아마도 L-오르니틴의 L-아르기닌으로의 재순환을 개선함으로써, GAA 생산을 개선한다는 것을 발견하였다.

[0211] ATCC13032 로부터 유래된 다양한 균주를 Wouter Duetz 시스템을 사용하여 배양한 다음, GAA 를 생산하는 이들의 능력에 대해 분석하였다 (표 13, 14 및 15).

[0212] 표 13: 글리신 또는 L-아르기닌 보충 없이 YE 를 포함하는 PM 에서의 GAA 생산에 대한 개선된 L-아르기닌 재생의 영향

균주	GAA
ATCC13032/pEC-XK99E	0 mg/l
ATCC13032/pEC-XK99E_argFGH	0 mg/l
ATCC13032_Psod-carAB/pEC-XK99E	0 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp	25 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argG	26 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argH	24 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argF	27 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argGH	26 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argFGH	25 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp	43 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argG	45 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argFGH	45 mg/l

[0213]

[0214] 표 14: YE 가 없는 PM 에서의 GAA 생산에 대한 개선된 L-아르기닌 재생의 영향 모든 배양물은 0.82 g/l 글리신으로 보충되었다

균주	GAA
ATCC13032/pEC-XK99E	0 mg/l
ATCC13032/pEC-XK99E_argFGH	0 mg/l
ATCC13032_Psod-carAB/pEC-XK99E	0 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp	22 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argG	24 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argH	22 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argF	23 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argGH	23 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argFGH	23 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp	53 mg/l

[0215]

[0216] 표 15: YE 를 포함하는 PM 에서의 GAA 생산에 대한 개선된 L-아르기닌 재생의 영향 모든 배양물은 0.82 g/l 글리신 및/또는 1.9 g/l L-아르기닌으로 보충되었다.

균주	GAA
ATCC13032/pEC-XK99E	검출 불가능
ATCC13032/pEC-XK99E_argFGH	검출 불가능
ATCC13032_Psod-carAB/pEC-XK99E	검출 불가능
ATCC13032/pEC-XK99E_AGAT_Mp	124 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argG	136 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argF	136 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argGH	133 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argFGH	154 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp	156 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argFGH	171 mg/l

[0217]

[0218] 표 13 내지 15 에 나타난 바와 같이, AGAT 유전자가 결여된 균주는 검출가능한 양의 GAA 를 생산하지 않았다.

[0219] ATCC13032/pEC-XK99E_AGAT_Mp 에서 AGAT_Mp 의 발현은 124 mg/l GAA 를 생성하였다. *argG* (균주 ATCC13032/pEC-XK99E_AGAT_Mp_argG), *argF* (균주 ATCC13032/pEC-XK99E_AGAT_Mp_argF) 또는 *argG+argH* (균주 ATCC13032/pEC-XK99E_AGAT_Mp_argGH) 의 추가적인 증폭은 GAA 의 생산을 개선시켰다 (표 14 참조).

[0220] 균주 ATCC13032/pEC-XK99E_AGAT_Mp_argFGH 에서, 유전자 *argF* (오르니틴 카르바모일트랜스퍼라제를 코딩함), *argG* (아르기니노숙시네이트 신테타제를 코딩함) 및 *argH* (아르기니노숙시네이트 리아제를 코딩함) 의 발현이 향상된다. 이는 GAA 의 생산을 154 mg/l 로 추가로 개선시켰다 (표 14 참조).

[0221] 오르니틴 카르바모일트랜스퍼라제에 의해 촉매되는, L-오르니틴의 L-시트룰린으로의 전환은 공동-기질 카르바모일 포스페이트의 이용가능성에 의존한다. 카르바모일 포스페이트는 유전자 *carA* 및 *carB* 에 의해 인코딩되는 카르바모일 포스페이트 신타아제에 의해 생성된다. 균주 ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp 에서, 강한 *sod*-프로모터의 게놈 삽입은 *carA* 및 *carB* 의 발현을 향상시킨다. ATCC13032/pEC-XK99E_AGAT_Mp 와 비교할 때, 이는 개선된 GAA 생산 (156 mg/l 대 124 mg/l) 을 초래하였다.

[0222] 균주 ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argFGH 에서, 개선된 L-오르니틴 전환 (*argF*, *argG* 및 *argH* 의 과발현) 을 개선된 카르바모일 포스페이트 생합성 (*carA* 및 *carB* 의 과발현) 과 조합하였다. 이러한 조합은 GAA 생산을 171 mg/l 로 추가로 개선시켰다.

[0223] **실시예 15: 무레아 프로듀센스 유전자 AGAT_Mp 를 위한 P. 푸티다 발현 벡터의 구축**

[0224] *P. 푸티다* KT2440 에서 무레아 프로듀센스 (EC 2.1.4.1, SEQ ID NO:2 및 SEQ ID NO:4) 로부터의 AGAT 의 이중 발현을 위해, 플라스미드 pACYCATH-5{PRha}[*agat_Mp*(coPp)] 를 구축하였다. 코돈 최적화된 AGAT_Mp 유전자를 램노오스 유도성 프로모터 P_{Rha} 의 제어 하에 벡터 pACYATH-5 내로 클로닝하였다. AGAT_Mp 유전자의 다운스트림에 터미네이터 서열이 위치한다. AGAT_Mp 유전자를 Eurofins Genomics Germany GmbH (Ebersberg, Germany) 로부터 유전자 합성을 위해 주문하고, 유전자 단편의 DNA 서열을 *P. 푸티다* KT2440 에서의 발현을 위해 코돈-최적화하였다 (SEQ ID NO:33). 오픈 리딩 프레임의 업스트림에는 샤인-달가노 서열이 추가되었다.

P_{Rha} 프로모터 카세트 (SEQ ID NO:34) 및 터미네이터 서열 (SEQ ID NO:35) 을 *E. 콜라이* K12 게놈 DNA 로부터 증폭시켰다. 벡터는 pACYC184 (New England BioLabs Inc., Ipswich, USA) 에 기초하며 *E. 콜라이*에 대한 p15A 복제 기원 및 *P. 푸티다* KT2440 에서의 복제를 위한 pVS1 복제 기원을 가지고 있다. pVS1 기원은 슈도모나스 플라스미드 pVS1 (Itoh Y, Watson JM, Haas D, Leisinger T, Plasmid 1984, 11(3), 206-20) 로부터 유래한다. 다음 단계에서, AGAT_Mp 유전자 단편을 프라이머 MW_20_01_fw (SEQ ID NO:36) 및 MW_20_02_rv (SEQ ID NO:37) 를 사용하는 PCR 을 통해 증폭시키고, 제한 부위 ApaI/XhoI 및 NEBuilder® HiFi DNA Assembly Cloning Kit (New England BioLabs Inc., Ipswich, USA, Cat. No. E5520 사제) 를 사용하여 벡터 pACYCATH-5 내로 클로닝하였다. 어셈블리된 생성물을 10-베타 전기적격 *E. 콜라이* 세포 (New England BioLabs Inc., Ipswich, USA, Cat. No. C3020K) 내로 형질전환시켰다. PCR 정제, 클로닝 및 형질전환 절차는 제조사의 매뉴얼에 따라 수행하였다. 표적 유전자의 정확한 삽입을 제한 분석에 의해 확인하고, 도입된 DNA 단편의 진

위를 DNA 서열분석에 의해 확인하였다. 생성된 발현 벡터를 pACYCATH-5{PRha}[agat_Mp(coPp)] (SEQ ID NO:38, 표 17 참조) 로 명명하였다.

[0225] *P. 푸티다* 균주 KT2440 을 플라스미드 pACYCATH-5{PRha}[agat_Mp(coPp)] 로 전기천공에 의해 형질전환하고, 테트라사이클린 (10 mg/1) 이 보충된 LB 아가 플레이트 상에 플레이팅하였다. 형질전환체를 플라스미드 제조 및 분석 제한 분석에 의해 정확한 플라스미드의 존재에 대해 체크하였다. 생성된 균주를 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)] 로 명명하였다 (표 18 참조).

[0226] **실시예 16: 무레아 프로듀센스 유전자 AGAT_Mp 및 *P. 푸티다* 유전자 argF, argG 및 argH 에 대한 *P. 푸티다* 발현 벡터의 구축**

[0227] 무레아 프로듀센스로부터의 AGAT_Mp 및 *P. 푸티다* KT2440 로부터의 argF (SEQ ID NO:39), argG (SEQ ID NO:41), argH (SEQ ID NO:43) 의 이중 발현을 위해, 플라스미드 pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp] 를 구축하였다. L-아르기닌:글리신 아미도노트랜스퍼라제 (AGAT, EC 2.1.4.1, SEQ ID NO:2 및 SEQ ID NO:4) 를 인코딩하는 AGAT_Mp, 오르니틴 카르바모일트랜스퍼라제를 인코딩하는 argF (ArgF, EC 2.1.3.3, SEQ ID NO:40), 아르기니노숙시네이트 신타아제를 인코딩하는 argG (ArgG, E.C. 6.3.4.5, SEQ ID NO:42) 및 아르기니노숙시네이트 리아제를 인코딩하는 argH (ArgH, E.C. 4.3.2.1, SEQ ID NO:44) 로 각각 이루어진 합성 오페론을 람노오스 유도성 프로모터 P_{rha} 의 제어 하에 벡터 pACYCATH-5 내로 클로닝하였다. 합성 오페론의 다운스트림에 터미네이터 서열이 위치한다. AGAT_Mp 유전자를 Eurofins Genomics Germany GmbH (Ebersberg, Germany) 로부터 유전자 합성을 위해 주문하고, 유전자 단편의 DNA 서열을 *P. 푸티다* KT2440 에서의 발현을 위해 코돈-최적화하였다. 유전자 argFGH 는 또한 유전자 단편 argFGH (SEQ ID NO: 45) 로서 합성되었다. P_{rha} 프로모터 카세트 (SEQ ID NO:34) 및 터미네이터 서열 (SEQ ID NO:35) 을 *E. 콜라이* K12 계놈 DNA 로부터 증폭시켰다. 벡터는 pACYC184 (New England BioLabs Inc., Ipswich, USA) 에 기초하며 *E. 콜라이*에 대한 p15A 복제 기원 및 *P. 푸티다* KT2440 에서의 복제를 위한 pVS1 복제 기원을 가지고 있다. pVS1 기원은 슈도모나스 플라스미드 pVS1 (Itoh Y, Watson JM, Haas D, Leisinger T, Plasmid 1984, 11(3), 206-20) 로부터 유래한다. 클로닝을 위해, AGAT_Mp 및 argFGH 를 PCR 을 통해 증폭시켰다. 클로닝에 사용된 프라이머가 표 16 에 열거되어 있다. 최적화된 오페론을 생성하기 위해 제한 부위 ApaI/XhoI 및 NEBuilder® HiFi DNA Assembly Cloning Kit (New England BioLabs Inc., Ipswich, USA, Cat. No. E5520 사제) 를 사용하여 PCR 생성물을 벡터 pACYCATH-5 내로 클로닝하였다. 증폭을 위해 New England Biolabs (Ipswich, USA) 로부터의 Phusion™ High-Fidelity Master Mix 를 제조사의 매뉴얼에 따라 사용하였다. 어셈블리된 생성물을 10-베타 전기적격 *E. 콜라이* 세포 (New England BioLabs Inc., Ipswich, USA, Cat. No. C3020K) 내로 형질전환시켰다. PCR 정제, 클로닝 및 형질전환 절차는 제조사의 매뉴얼에 따라 수행하였다. 표적 유전자의 정확한 삽입을 제한 분석에 의해 확인하고, 도입된 DNA 단편의 진위를 DNA 서열분석에 의해 확인하였다. 생성된 발현 벡터를 pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp] (SEQ ID NO:49, 표 17 참조) 로 명명하였다.

[0228] *P. 푸티다* 균주 KT2440 을 플라스미드 pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp] 로 전기천공에 의해 형질전환하고, 테트라사이클린 (10 mg/1) 이 보충된 LB 아가 플레이트 상에 플레이팅하였다. 형질전환체를 플라스미드 제조 및 분석 제한 분석에 의해 정확한 플라스미드의 존재에 대해 체크하였다. 생성된 균주를 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp] 로 명명하였다 (표 18 참조).

[0229] 표 16: AGAT_Mp 및 argFGH 의 pACYCATH-5 내로의 클로닝을 위해 사용된 프라이머.

프라이머	SEQ ID	주형	PCR 생성물
MW_20_01_fw	SEQ ID NO:36	AGAT_Mp 유전자 분획	AGAT_Mp
MW_20_03_rv	SEQ ID NO:46	AGAT_Mp 유전자 분획	AGAT_Mp
MW_20_04_fw	SEQ ID NO:47	argFGH 유전자 분획	argFGH
MW_20_09_rv	SEQ ID NO:48	argFGH 유전자 분획	argFGH

[0230] .

[0231] **실시예 17: *P. 푸티다* KT2440 으로부터 유전자 carAB 의 발현 벡터 pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp] 내로의 클로닝**

[0232] 무레아 프로듀센스로부터의 AGAT_Mp, *P. 푸티다* KT2440 로부터의 argF (SEQ ID NO:39), argG (SEQ ID NO:41), argH (SEQ ID NO:43), carA (SEQ ID NO:50) 및 carB (SEQ ID NO:52) 의 이중 발현을 위해, 플라스미드

pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] 를 구축하였다. 카르바모일-포스페이트 신타아제 (CarAB, EC 6.3.5.5, SEQ ID NO:51 및 SEQ ID NO:53) 를 인코딩하는 carA (SEQ ID NO:50) 및 carB 유전자 (SEQ ID NO:52) 를 프라이머 MW_20_35_fw (SEQ ID NO:54) 및 MW_20_36_rv (SEQ ID NO:55) 를 사용하여 PCR 을 통해 carAB 오페론의 고유의 프로모터를 포함하는 *P. 푸티다* KT2440 의 게놈 DNA 로부터 증폭시켰다. 증폭을 위해 New England Biolabs (Ipswich, USA) 로부터의 Phusion™ High-Fidelity Master Mix 를 제조자의 매뉴얼에 따라 사용하였다. PCR 생성물 (SEQ ID NO:56) 을 NEBuilder® HiFi DNA Assembly Cloning Kit (New England BioLabs Inc., Ipswich, USA, Cat. No. E5520 사제) 를 사용하여 *Bsu36I* 로 절단된 pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp] 내에 클로닝하였다. 어셈블리된 생성물을 10-베타 전기적격 *E. 콜라이* 세포 (New England BioLabs Inc., Ipswich, USA, Cat. No. C3020K) 내로 형질전환시켰다. PCR 정제, 클로닝 및 형질전환 절차는 제조사의 매뉴얼에 따라 수행하였다. 표적 유전자의 정확한 삽입을 제한 분석에 의해 확인하고, 도입된 DNA 단편의 진위를 DNA 서열분석에 의해 확인하였다. 생성된 발현 벡터를 pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] (SEQ ID NO:57, 표 17 참조) 로 명명하였다.

[0233] *P. 푸티다* 균주 KT2440 을 플라스미드 pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] 로 전기천공에 의해 형질전환하고, 테트라사이클린 (10 mg/l) 이 보충된 LB 아가 플레이트 상에 플레이팅하였다. 형질전환체를 플라스미드 제조 및 분석 제한 분석에 의해 정확한 플라스미드의 존재에 대해 체크하였다. 생성된 균주를 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] 로 명명하였다 (표 18 참조).

[0234] 표 17: *P. 푸티다* 발현 플라스미드 목록

플라스미드	코멘트
pACYCATH-5	빈 <i>E. 콜라이</i> - <i>P. 푸티다</i> 서플 플라스미드
pACYCATH-5{PRha}[agat_Mp(coPp)]	AGAT_Mp (<i>우레아 프로듀센스</i>) 의 발현
pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]	AGAT_Mp (<i>우레아 프로듀센스</i>), <i>argF</i> , <i>argG</i> , <i>argH</i> (<i>P. 푸티다</i>) 의 발현
pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter]	AGAT_Mp (<i>우레아 프로듀센스</i>), <i>argF</i> , <i>argG</i> , <i>argH</i> , <i>carA</i> , <i>carB</i> (<i>P. 푸티다</i>) 의 발현

[0235]

[0236] 표 18: 플라스미드-함유 *P. 푸티다* 균주 목록

플라스미드	클로닝된 유전자	수여 균주	산출 균주
pACYCATH-5	없음	<i>P. 푸티다</i> KT2440	<i>P. 푸티다</i> /KT2440/pACYCATH-5
pACYCATH-5{PRha}[agat_Mp(coPp)]	AGAT_Mp	<i>P. 푸티다</i> KT2440	<i>P. 푸티다</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)]
pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]	AGAT_Mp, argF, argG, argH	<i>P. 푸티다</i> KT2440	<i>P. 푸티다</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]
pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp][carAB_Pp][carAB_Pp][ter]	AGAT_Mp, argF, argG, argH, carA, carB	<i>P. 푸티다</i> KT2440	<i>P. 푸티다</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp][carAB_Pp][carAB_Pp][ter]

[0237]

[0238] 실시예 18: *P. 푸티다* KT2440 에서의 GAA 생산에 대한 AGAT 의 영향

[0239]

M. 프로듀센스로부터의 AGAT 효소에 대한 유전자를 운반하는 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)] 및 *P. 푸티다* KT2440/pACYCATH-5 (빈 벡터 대조군) 를 진탕 플라스크를 사용하여 배지 배양에서 GAA 를 생산하는 능력에 대해 분석하였다. 10 mg/l 테트라사이클린을 함유하는 LB 아가 플레이트 상에서, 상응하는 균주의 글리세롤 동결배양물의 접종 루프를 스트리킹하였다. 아가 플레이트를 30°C 에서 24 시간 동안 인큐베이션한다. 10 mg/l 테트라사이클린이 있는, 15 ml 의 종자 배지 (오토클레이브됨: 4.4 g/L Na₂HPO₄ x 2 H₂O, 1.5 g/L KH₂PO₄, 1 g/L NH₄Cl, 10 g/L 효모 추출물, 각각 멸균됨: 20 g/L 글루코스, 0.2 g/L MgSO₄ x 7 H₂O, 0.006 g/L FeCl₃, 0.015 g/L CaCl₂, 1 ml/L 미량 원소 용액 SL6 (멸균-여과됨: 0.3 g/L H₃BO₃, 0.2 g/L CoCl₂ x 6 H₂O, 0.1 g/L ZnSO₄ x 7 H₂O, 0.03 g/L MnCl₂ x 4H₂O, 0.01 g/L CuCl₂ x 2 H₂O, 0.03 g/L Na₂MoO₄ x 2 H₂O, 0.02 g/L NiCl₂ x 6 H₂O), pH 7) 를 함유하는 배플을 가진 100 ml 플라스크에 아가 플레이트의 단일 콜로니를 접종하고, 진탕 인큐베이터에서 18 시간 동안 30°C 및 200 rpm 에서 인큐베이션하여, 예비배양물을 생산하였다. 예비배양물을 사용하여 10 mg/l 테트라사이클린, 3.48 g/l 아르기닌 및 1.5 g/l 글리신이 있는, 40 ml 의 M12 배지 (조성: (2.2 g/L (NH₄)₂SO₄, 0.02 g/L NaCl, 0.4 g/L MgSO₄ x 7H₂O, 0.04 g/L CaCl₂ x 2H₂O, 각각 멸균됨: 2 g/L KH₂PO₄, 8.51 g/L Na₂HPO₄, 20 g/L 그루코스, 10 ml/l 미량 원소 용액 M12 (멸균-여과됨: 0.2 g/L ZnSO₄ x 7 H₂O, 0.1 g/L MnCl₂ x 4H₂O, 1.5 g/L Na₃-시트레이트 x 2 H₂O, 0.1 g/L CuSO₄ x 5 H₂O, 0.002 g/L NiCl₂ x 6 H₂O, 0.003 g/L Na₂MoO₄ x 2 H₂O, 0.03 g/L H₃BO₃, 1 g/L FeSO₄ x 7 H₂O), pH 7.4) 에 접종하고 0.1 의 OD₆₀₀ 을 시작하였다. 균주를 48 시간 동안 배양하였다. 대략 0.5-0.8 의 OD₆₀₀ 에서, 0.2% (w/v) 램노스를 첨가함으로써 유전자 발현을 유도하였다. 유도 9 시간 및 24 시간 후 1.74 g/l 아르기닌 및 0.75 g/l 글리신이 스파이킹되었다. 배양 종료시, 생성된 GAA 의 농도를 결정하기 위해 샘플을 취하였다.

[0240]

결과는 표 19 에 제시된다.

[0241]

표 19: 균주 *P. 푸티다* KT2440/pACYCATH-5 및 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)] 에 의한 GAA 생산

균주	GAA
<i>P. 푸티다</i> KT2440/pACYCATH-5	검출 불가능
<i>P. 푸티다</i> KT2440/pACYCATH-5{PRha}[<i>agat_Mp</i> (coPp)]	81.5 mg/l

[0242]

[0243]

표 19 에서 알 수 있는 바와 같이, *M. 프로듀센스*로부터의 *AGAT_Mp* 유전자가 장착된 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[*agat_Mp*(coPp)] 는 약 81.5 mg/l GAA 를 생산할 수 있었다. 대조군 균주 *P. 푸티다* KT2440/pACYCATH-5 는 어떠한 GAA 도 전혀 생산할 수 없었다.

[0244]

실시예 19: *P. 푸티다* KT2440 에서 GAA 생산에 대한 *AGAT* 및 증가된 L-아르기닌 공급의 영향

[0245]

*M. 프로듀센스*로부터의 *AGAT* 효소에 대한 유전자를 운반하는 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[*agat_Mp*(coPp)] 및 아르기닌 생합성 유전자 *argFGH* 를 부가적으로 운반하는 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[*agat_Mp*(coPp) *argFGH_Pp*] 를 진탕 플라스크를 사용하여 배치 배양에서 GAA 를 생산하는 능력에 대해 분석하였다. 10 mg/l 테트라사이클린을 함유하는 LB 아가 플레이트 상에서, 상응하는 균주의 글리세롤 동결배양물의 접종 루프를 스트리킹하였다. 아가 플레이트를 30°C 에서 24 시간 동안 인큐베이션한다. 10 mg/l 테트라사이클린이 있는, 15 ml 의 종자 배지 (오토클레이브됨: 4.4 g/L Na₂HPO₄ x 2 H₂O, 1.5 g/L KH₂PO₄, 1 g/L NH₄Cl, 10 g/L 효모 추출물, 각각 멸균됨: 20 g/L 글루코스, 0.2 g/L MgSO₄ x 7 H₂O, 0.006 g/L FeCl₃, 0.015 g/L CaCl₂, 1 ml/L 미량 원소 용액 SL6 (멸균-여과됨: 0.3 g/L H₃BO₃, 0.2 g/L CoCl₂ x 6 H₂O, 0.1 g/L ZnSO₄ x 7 H₂O, 0.03 g/L MnCl₂ x 4H₂O, 0.01 g/L CuCl₂ x 2 H₂O, 0.03 g/L Na₂MoO₄ x 2 H₂O, 0.02 g/L NiCl₂ x 6 H₂O), pH 7) 를 함유하는 배플을 가진 100 ml 플라스크에 아가 플레이트의 단일 콜로니를 접종하고, 진탕 인큐베이터에서 18 시간 동안 30°C 및 200 rpm 에서 인큐베이션하여, 예비배양물을 생산하였다. 예비배양물을 사용하여 10 mg/l 테트라사이클린, 3.48 g/l 아르기닌 및 1.5 g/l 글리신이 있는, 40 ml 의 M12 배지 (조성: (2.2 g/L (NH₄)₂SO₄, 0.02 g/L NaCl, 0.4 g/L MgSO₄ x 7H₂O, 0.04 g/L CaCl₂ x 2H₂O, 각각 멸균됨: 2 g/L KH₂PO₄, 8.51 g/L Na₂HPO₄, 20 g/L 그루코스, 10 ml/l 미량 원소 용액 M12 (멸균-여과됨: 0.2 g/L ZnSO₄ x 7 H₂O, 0.1 g/L MnCl₂ x 4H₂O, 1.5 g/L Na₃-시트레이트 x 2 H₂O, 0.1 g/L CuSO₄ x 5 H₂O, 0.002 g/L NiCl₂ x 6 H₂O, 0.003 g/L Na₂MoO₄ x 2 H₂O, 0.03 g/L H₃BO₃, 1 g/L FeSO₄ x 7 H₂O), pH 7.4) 에 접종하고 0.1 의 OD₆₀₀ 을 시작하였다. 균주를 48 시간 동안 배양하였다. 대략 0.5-0.8 의 OD₆₀₀ 에서, 0.2% (w/v) 람노스를 첨가함으로써 유전자 발현을 유도하였다. 유도 9 시간 및 24 시간 후 1.74 g/l 아르기닌 및 0.75 g/l 글리신이 스파이킹되었다. 배양 종료시, 생성된 GAA 의 농도를 결정하기 위해 샘플을 취하였다.

[0246]

결과는 표 20 에 제시된다.

[0247]

표 20: 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[*agat_Mp*(coPp)] 및 *P. 푸티다* KT2440/pACYCATH-5{PRha}[*agat_Mp*(coPp) *argFGH_Pp*] 에 의한 GAA 생산

균주	GAA
<i>P. 푸티다</i> KT2440/pACYCATH-5	검출 불가능
<i>P. 푸티다</i> KT2440/pACYCATH-5{PRha}[<i>agat_Mp</i> (coPp)]	81.5 mg/l
<i>P. 푸티다</i> KT2440/pACYCATH-5{PRha}[<i>agat_Mp</i> (coPp) <i>argFGH_Pp</i>]	169.5 mg/l

[0248]

[0249]

표 20 에서 알 수 있는 바와 같이, *M. 프로듀센스*로부터의 *AGAT_Mp* 유전자가 장착된 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[*agat_Mp*(coPp)] 는 약 81.5 mg/l GAA 를 생산할 수 있었다. 부가적으로 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[*agat_Mp*(coPp) *argFGH_Pp*] 에서 *argF*, *argG* 및 *argH* 의 구현은 GAA 생산을 169.5 mg/l 로 개선시켰다.

[0250]

실시예 20: *P. 푸티다* KT2440 에서 GAA 생산에 대한 *AGAT* 및 L-아르기닌 재생의 영향

[0251]

*M. 프로듀센스*로부터의 *AGAT* 효소에 대한 유전자를 운반하는 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[*agat_Mp*(coPp)], 아르기닌 생합성 유전자 *argFGH* 를 부가적으로 운반하는 균주 *P. 푸티다*

KT2440/pACYCATH-5{PRha}[agat_Mp argFGH_Pp] 및 카르바모일-포스페이트 신타아제 유전자 *carAB* 를 부가적으로 운반하는 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] 를 진탕 플라스크를 사용하여 배지 배양에서 GAA 를 생산하는 능력에 대해 분석하였다. 오르니틴 카르바모일트랜스퍼라제에 의해 촉매되는, L-오르니틴의 L-시트룰린으로의 전환은 공동-기질 카르바모일 포스페이트의 이용가능성에 의존한다. 카르바모일 포스페이트는 유전자 *carA* 및 *carB* 에 의해 인코딩되는 카르바모일 포스페이트 신타아제에 의해 생성된다. 10 mg/l 테트라사이클린을 함유하는 LB 아가 플레이트 상에서, 상응하는 균주의 글리세롤 동결배양물의 접종 루프를 스트리킹하였다. 아가 플레이트를 30°C 에서 24 시간 동안 인큐베이션한다. 10 mg/l 테트라사이클린이 있는, 15 ml 의 종자 배지 (오토클레이브됨: 4.4 g/L Na₂HPO₄ x 2 H₂O, 1.5 g/L KH₂PO₄, 1 g/L NH₄Cl, 10 g/L 효모 추출물, 각각 멸균됨: 20 g/L 글루코스, 0.2 g/L MgSO₄ x 7 H₂O, 0.006 g/L FeCl₃, 0.015 g/L CaCl₂, 1 ml/L 미량 원소 용액 SL6 (멸균-여과됨: 0.3 g/L H₃BO₃, 0.2 g/L CoCl₂ x 6 H₂O, 0.1 g/L ZnSO₄ x 7 H₂O, 0.03 g/L MnCl₂ x 4H₂O, 0.01 g/L CuCl₂ x 2 H₂O, 0.03 g/L Na₂MoO₄ x 2 H₂O, 0.02 g/L NiCl₂ x 6 H₂O), pH 7) 를 함유하는 배플을 가진 100 ml 플라스크에 아가 플레이트의 단일 콜로니를 접종하고, 진탕 인큐베이터에서 18 시간 동안 30°C 및 200 rpm 에서 인큐베이션하여, 예비배양물을 생산하였다. 예비배양물을 사용하여 10 mg/l 테트라사이클린, 3.48 g/l 아르기닌 및 1.5 g/l 글리신이 있는, 40 ml 의 M12 배지 (조성: (2.2 g/L (NH₄)₂SO₄, 0.02 g/L NaCl, 0.4 g/L MgSO₄ x 7H₂O, 0.04 g/L CaCl₂ x 2H₂O, 각각 멸균됨: 2 g/L KH₂PO₄, 8.51 g/L Na₂HPO₄, 20 g/L 그루코스, 10 ml/l 미량 원소 용액 M12 (멸균-여과됨: 0.2 g/L ZnSO₄ x 7 H₂O, 0.1 g/L MnCl₂ x 4H₂O, 1.5 g/L Na₃-시트레이트 x 2 H₂O, 0.1 g/L CuSO₄ x 5 H₂O, 0.002 g/L NiCl₂ x 6 H₂O, 0.003 g/L Na₂MoO₄ x 2 H₂O, 0.03 g/L H₃BO₃, 1 g/L FeSO₄ x 7 H₂O), pH 7.4) 에 접종하고 0.1 의 OD₆₀₀ 을 시작하였다. 균주를 48 시간 동안 배양하였다. 대략 0.5-0.8 의 OD₆₀₀ 에서, 0.2% (w/v) 람노스를 첨가함으로써 유전자 발현을 유도하였다. 유도 후 4 시간 / 18 시간 / 23 시간에, 6.97 g/l Arg/1.5 g/l Gly, 2.34 g/l Arg/0.75 g/l Gly 및 6.97 g/l Arg/1.5 g/l Gly 가 스파이킹되었다. 배양 종료 시, 생성된 GAA 의 농도를 결정하기 위해 샘플을 취하였다.

[0252] 결과는 표 21 에 제시된다.

[0253] 표 21: 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)], *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp] 및 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] 에 의한 GAA 생산

균주	GAA
<i>P. 푸티다</i> /KT2440/pACYCATH-5	검출 불가능
<i>P. 푸티다</i> /KT2440/pACYCATH-5{PRha}[agat_Mp argFGH_Pp]	589 mg/l
<i>P. 푸티다</i> /KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter]	693 mg/l

[0254]

[0255] 표 21 에서 알 수 있는 바와 같이, M. 프로듀센스로부터의 AGAT_Mp 유전자 및 *P. 푸티다*로부터의 argFGH 유전자가 장착된 균주는 589 mg/l GAA 를 생산할 수 있었다. 부가적으로 균주 *P. 푸티다* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] 에서 *carAB* 의 구현은 GAA 생산을 693 mg/l 로 개선시켰다.

서열 목록

SEQUENCE LISTING

<110> Evonik Operations GmbH

<120> Method for the fermentative production of guanidinoacetic acid

<130> 000
 <160> 57
 <170> PatentIn version 3.5
 <210> 1
 <211> 1146
 <212> DNA
 <213> Moorea producens strain PAL-8-15-08-1
 <220><221> CDS
 <222> (1)..(1143)
 <223> Genbank accession Number WP_070390602
 <400> 1

atg tcg gaa aaa att gtt aat tcc tgg aat gaa tgg gat gaa ttg gaa 48
 Met Ser Glu Lys Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Glu

1 5 10 15
 gaa atg gtg gta gga att gca gac tat gct agc ttt gaa cca aaa gaa 96
 Glu Met Val Val Gly Ile Ala Asp Tyr Ala Ser Phe Glu Pro Lys Glu

20 25 30
 cca ggg aat cat ccg aaa tta aga aat caa aat tta gcg gaa atc att 144
 Pro Gly Asn His Pro Lys Leu Arg Asn Gln Asn Leu Ala Glu Ile Ile

35 40 45

cct ttc ccc agt gga cct aaa gac cct aaa gtc ctt gaa aaa gct aat 192
 Pro Phe Pro Ser Gly Pro Lys Asp Pro Lys Val Leu Glu Lys Ala Asn

50 55 60
 gaa gaa tta aat gga ctg gct tat tta tta aaa gac cac gat gtg ata 240
 Glu Glu Leu Asn Gly Leu Ala Tyr Leu Leu Lys Asp His Asp Val Ile

65 70 75 80
 gta aga aga ccc gaa aaa att gat ttt act aaa tct cta aaa aca cct 288

Val Arg Arg Pro Glu Lys Ile Asp Phe Thr Lys Ser Leu Lys Thr Pro

85 90 95
 tac ttt gaa gtt gca aat caa tac tgt gga gtc tgt cct cgg gat gtc 336
 Tyr Phe Glu Val Ala Asn Gln Tyr Cys Gly Val Cys Pro Arg Asp Val

100 105 110

atg att acc ttt ggg aat gaa atc atg gaa gcg act atg tcg aag aga 384
 Met Ile Thr Phe Gly Asn Glu Ile Met Glu Ala Thr Met Ser Lys Arg

115 120 125
 gct aga ttt ttt gaa tac tta cct tac cgg aaa ttg gtc tat gaa tat 432
 Ala Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr

130 135 140
 tgg aat aaa gac gag cat atg att tgg aat gct gcg cct aaa ccg act 480
 Trp Asn Lys Asp Glu His Met Ile Trp Asn Ala Ala Pro Lys Pro Thr

145 150 155 160
 atg cag gat agt atg tat cta gag aat ttc tgg gag ctg tct tta gaa 528
 Met Gln Asp Ser Met Tyr Leu Glu Asn Phe Trp Glu Leu Ser Leu Glu

165 170 175
 gaa cga ttt aag cgt atg cat gat ttt gaa ttt tgt att aca caa gat 576
 Glu Arg Phe Lys Arg Met His Asp Phe Glu Phe Cys Ile Thr Gln Asp

180 185 190
 gaa gta att ttt gat gcg gct gac tgt agc aga tta gga aag gat ata 624

Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Leu Gly Lys Asp Ile

195 200 205
 tta gtt cag gaa tcg atg aca aca aat aga aca gga att cgg tgg tta 672
 Leu Val Gln Glu Ser Met Thr Thr Asn Arg Thr Gly Ile Arg Trp Leu

210 215 220
 aaa aag cac cta gaa cca aga gga ttt cgg gtt cac cct gtt cat ttt 720
 Lys Lys His Leu Glu Pro Arg Gly Phe Arg Val His Pro Val His Phe

225 230 235 240
 ccc ctt gat ttt ttc ccc tca cac att gac tgt acg ttt gtt cct ttg 768
 Pro Leu Asp Phe Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu

245 250 255
 cga cca ggt ctt att ttg aca aac cct gaa aga cct ata cgg gaa gag 816
 Arg Pro Gly Leu Ile Leu Thr Asn Pro Glu Arg Pro Ile Arg Glu Glu

260	265	270	
gag gag aag att ttt aaa gag aat ggc tgg gag ttg atc aca gtt cct			864
Glu Glu Lys Ile Phe Lys Glu Asn Gly Trp Glu Leu Ile Thr Val Pro			
275	280	285	
caa ccg act tgc tcg aat gat gaa atg cca atg ttt tgc cag tcc agt			912
Gln Pro Thr Cys Ser Asn Asp Glu Met Pro Met Phe Cys Gln Ser Ser			
290	295	300	
aag tgg ttg tca atg aat gtt ctg agt ata tca ccg aca aag gtt atc			960
Lys Trp Leu Ser Met Asn Val Leu Ser Ile Ser Pro Thr Lys Val Ile			
305	310	315	320
tgt gag gaa aga gaa aaa cct ctc caa gaa ttg ttg gat aag cat gga			1008
Cys Glu Glu Arg Glu Lys Pro Leu Gln Glu Leu Leu Asp Lys His Gly			
325	330	335	
ttt gag gtt ttt cct tta ccc ttt aga cat gtc ttt gaa ttt ggg ggg			1056
Phe Glu Val Phe Pro Leu Pro Phe Arg His Val Phe Glu Phe Gly Gly			
340	345	350	
tct ttt cat tgt gca act tgg gat att cgc cga aaa ggt gag tgt gaa			1104
Ser Phe His Cys Ala Thr Trp Asp Ile Arg Arg Lys Gly Glu Cys Glu			
355	360	365	
gat tat tta cca aat tta aac tat caa ccg att tgt ggt taa			1146
Asp Tyr Leu Pro Asn Leu Asn Tyr Gln Pro Ile Cys Gly			
370	375	380	
<210> 2			
<211> 381			
<212> PRT			
<213> Moorea producens strain PAL-8-15-08-1			
<400> 2			
Met Ser Glu Lys Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Glu			
1	5	10	15
Glu Met Val Val Gly Ile Ala Asp Tyr Ala Ser Phe Glu Pro Lys Glu			
20	25	30	

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

ccg ggt aat cat ccg aaa ctg cgt aat cag aat ctg gca gaa att att 201
 Pro Gly Asn His Pro Lys Leu Arg Asn Gln Asn Leu Ala Glu Ile Ile

 35 40 45
 ccg ttt ccg agc ggt ccg aaa gat ccg aaa gtt ctg gaa aaa gca aat 249
 Pro Phe Pro Ser Gly Pro Lys Asp Pro Lys Val Leu Glu Lys Ala Asn
 50 55 60
 gaa gaa ctg aat ggt ctg gcc tat ctg ctg aaa gat cat gat gtt att 297
 Glu Glu Leu Asn Gly Leu Ala Tyr Leu Leu Lys Asp His Asp Val Ile
 65 70 75 80

 gtt cgc cgt ccg gaa aaa atc gac ttt acc aaa agc ctg aaa acc ccg 345
 Val Arg Arg Pro Glu Lys Ile Asp Phe Thr Lys Ser Leu Lys Thr Pro
 85 90 95
 tat ttc gaa gtt gcc aat cag tat tgt ggt gtt tgt ccg cgt gat gtt 393
 Tyr Phe Glu Val Ala Asn Gln Tyr Cys Gly Val Cys Pro Arg Asp Val
 100 105 110
 atg att acc ttt ggc aac gaa att atg gaa gcc acc atg agc aaa cgt 441

 Met Ile Thr Phe Gly Asn Glu Ile Met Glu Ala Thr Met Ser Lys Arg
 115 120 125
 gcc cgt ttt ttt gaa tat ctg ccg tat cgt aaa ctg gtg tat gag tat 489
 Ala Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140
 tgg aac aaa gat gag cat atg atc tgg aat gca gca ccg aaa ccg acc 537
 Trp Asn Lys Asp Glu His Met Ile Trp Asn Ala Ala Pro Lys Pro Thr

 145 150 155 160
 atg cag gat agc atg tat ctg gaa aac ttt tgg gaa ctg agc ctg gaa 585
 Met Gln Asp Ser Met Tyr Leu Glu Asn Phe Trp Glu Leu Ser Leu Glu
 165 170 175
 gaa cgt ttt aaa cgt atg cac gat ttt gag ttt tgc atc acc cag gat 633
 Glu Arg Phe Lys Arg Met His Asp Phe Glu Phe Cys Ile Thr Gln Asp

180	185	190	
gaa gtg att ttt gat gca gca gat tgt agc cgt ctg ggt aaa gat att			681
Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Leu Gly Lys Asp Ile			
195	200	205	
ctg gtt caa gaa agc atg acc acc aat cgt acc ggt att cgt tgg ctg			729
Leu Val Gln Glu Ser Met Thr Thr Asn Arg Thr Gly Ile Arg Trp Leu			
210	215	220	
aaa aaa cat ctg gaa ccg cgt ggt ttt cgt gtt cat ccg gtt cat ttt			777
Lys Lys His Leu Glu Pro Arg Gly Phe Arg Val His Pro Val His Phe			
225	230	235	240
ccg ctg gat ttt ttt ccg agc cat att gat tgt acc ttt gtt ccg ctg			825
Pro Leu Asp Phe Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu			
245	250	255	
cgt ccg ggt ctg att ctg acc aat ccg gaa cgt ccg att cgt gaa gaa			873
Arg Pro Gly Leu Ile Leu Thr Asn Pro Glu Arg Pro Ile Arg Glu Glu			
260	265	270	
gaa gag aaa atc ttc aaa gag aat ggc tgg gag ctg att acc gtt ccg			921
Glu Glu Lys Ile Phe Lys Glu Asn Gly Trp Glu Leu Ile Thr Val Pro			
275	280	285	
cag ccg acc tgt agc aat gat gaa atg ccg atg ttt tgt cag agc agc			969
Gln Pro Thr Cys Ser Asn Asp Glu Met Pro Met Phe Cys Gln Ser Ser			
290	295	300	
aaa tgg ctg agc atg aat gtt ctg agc att agc ccg acc aaa gtt att			1017
Lys Trp Leu Ser Met Asn Val Leu Ser Ile Ser Pro Thr Lys Val Ile			
305	310	315	320
tgt gaa gaa cgt gaa aaa ccg ctg caa gaa ctg ctg gat aaa cat ggt			1065
Cys Glu Glu Arg Glu Lys Pro Leu Gln Glu Leu Leu Asp Lys His Gly			
325	330	335	
ttt gaa gtg ttt ccg ctg ccg ttt cgt cat gtt ttt gaa ttt ggt ggt			1113

Phe Glu Val Phe Pro Leu Pro Phe Arg His Val Phe Glu Phe Gly Gly
 340 345 350
 agc ttt cat tgt gcc acc tgg gat att cgt cgt aaa ggt gaa tgt gaa 1161
 Ser Phe His Cys Ala Thr Trp Asp Ile Arg Arg Lys Gly Glu Cys Glu
 355 360 365
 gat tat ctg ccg aat ctg aat tat cag ccg att tgt ggt taataagacg 1210
 Asp Tyr Leu Pro Asn Leu Asn Tyr Gln Pro Ile Cys Gly
 370 375 380
 tccgcgaggg ccgtgttgcc ggtttcttca gagagacg 1248
 <210> 4
 <211> 381
 <212> PRT
 <213> synthetic DNA
 <400> 4
 Met Ser Glu Lys Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Glu
 1 5 10 15
 Glu Met Val Val Gly Ile Ala Asp Tyr Ala Ser Phe Glu Pro Lys Glu
 20 25 30
 Pro Gly Asn His Pro Lys Leu Arg Asn Gln Asn Leu Ala Glu Ile Ile
 35 40 45
 Pro Phe Pro Ser Gly Pro Lys Asp Pro Lys Val Leu Glu Lys Ala Asn
 50 55 60
 Glu Glu Leu Asn Gly Leu Ala Tyr Leu Leu Lys Asp His Asp Val Ile
 65 70 75 80
 Val Arg Arg Pro Glu Lys Ile Asp Phe Thr Lys Ser Leu Lys Thr Pro
 85 90 95
 Tyr Phe Glu Val Ala Asn Gln Tyr Cys Gly Val Cys Pro Arg Asp Val
 100 105 110
 Met Ile Thr Phe Gly Asn Glu Ile Met Glu Ala Thr Met Ser Lys Arg
 115 120 125
 Ala Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140

Trp Asn Lys Asp Glu His Met Ile Trp Asn Ala Ala Pro Lys Pro Thr
 145 150 155 160
 Met Gln Asp Ser Met Tyr Leu Glu Asn Phe Trp Glu Leu Ser Leu Glu

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

 225 230 235 240
 Pro Leu Asp Phe Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu
 245 250 255
 Arg Pro Gly Leu Ile Leu Thr Asn Pro Glu Arg Pro Ile Arg Glu Glu
 260 265 270
 Glu Glu Lys Ile Phe Lys Glu Asn Gly Trp Glu Leu Ile Thr Val Pro
 275 280 285
 Gln Pro Thr Cys Ser Asn Asp Glu Met Pro Met Phe Cys Gln Ser Ser

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

 355 360 365
 Asp Tyr Leu Pro Asn Leu Asn Tyr Gln Pro Ile Cys Gly
 370 375 380

<210> 5

<211> 30
 <212> DNA
 <213> synthetic oligonucleotide
 <400> 5
 ccgttaactg ccgagacaat cgcataaagg 30
 <210> 6
 <211> 34
 <212> DNA
 <213> synthetic oligonucleotide
 <400> 6
 ctctaggct aatattctta cctcggctgg ttgg 34
 <210> 7
 <211> 27

 <212> DNA
 <213> synthetic oligonucleotide
 <400> 7
 ccgttaacac ggctggcaag gaactta 27
 <210> 8
 <211> 41
 <212> DNA
 <213> synthetic oligonucleotide
 <400> 8
 gccctaggtc aatattacag gccataaact aatgcttate g 41
 <210> 9
 <211> 46
 <212> DNA
 <213> synthetic oligonucleotide
 <400> 9
 ggaaacagct atgacatgat tacgcgggta tcgcggaatc cgtatg 46
 <210> 10
 <211> 25
 <212> DNA
 <213> PsodcarAB-LA-R

<400> 10
ttaaagcgttt tgtgcaactc cgtct 25

<210> 11
<211> 50
<212> DNA
<213> synthetic oligonucleotide

<400> 11
agacggagtt gcacaaaacg cttaaacct acttagctgc caattattcc 50

<210> 12
<211> 50
<212> DNA
<213> synthetic oligonucleotide

<400> 12
ggtagtggt ggtgtcttta ctcatgggta aaaaatcctt tcgtaggttt 50

<210> 13
<211> 25
<212> DNA
<213> synthetic oligonucleotide

<400> 13
atgagtaaag acaccaccac ctacc 25

<210> 14
<211> 46
<212> DNA
<213> synthetic oligonucleotide

<400> 14
gttgtaaac gacggccagt gccaccggtg atgtggttct tcaactg 46

<210> 15
<211> 1176
<212> DNA
<213> *Cylindrospermopsis raciborskii* AWT205

<220><221> CDS
<222> (1)..(1176)
<223> genbank accession number EU140798.1
<400> 15

atg caa aca aga att gta aat agc tgg aat gag tgg gat gaa cta aag 48
 Met Gln Thr Arg Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Lys
 1 5 10 15

gag atg gtt gtc ggg att gca gat ggt gct tat ttt gaa cca act gag 96
 Glu Met Val Val Gly Ile Ala Asp Gly Ala Tyr Phe Glu Pro Thr Glu
 20 25 30

cca ggt aac cgc cct gct tta cgc gat aag aac att gcc aaa atg ttc 144
 Pro Gly Asn Arg Pro Ala Leu Arg Asp Lys Asn Ile Ala Lys Met Phe
 35 40 45

tct ttt ccc agg ggt ccg aaa aag caa gag gta aca gag aaa gct aat 192

Ser Phe Pro Arg Gly Pro Lys Lys Gln Glu Val Thr Glu Lys Ala Asn
 50 55 60

gag gag ttg aat ggg ctg gta gcg ctt cta gaa tca cag ggc gta act 240
 Glu Glu Leu Asn Gly Leu Val Ala Leu Leu Glu Ser Gln Gly Val Thr
 65 70 75 80

gta cgc cgc cca gag aaa cat aac ttt ggc ctg tct gtg aag aca cca 288
 Val Arg Arg Pro Glu Lys His Asn Phe Gly Leu Ser Val Lys Thr Pro
 85 90 95

ttc ttt gag gta gag aat caa tat tgt gcg gtc tgc cca cgt gat gtt 336
 Phe Phe Glu Val Glu Asn Gln Tyr Cys Ala Val Cys Pro Arg Asp Val
 100 105 110

atg atc acc ttt ggg aac gaa att ctc gaa gca act atg tca cgg cgg 384
 Met Ile Thr Phe Gly Asn Glu Ile Leu Glu Ala Thr Met Ser Arg Arg
 115 120 125

tca cgc ttc ttt gag tat tta ccc tat cgc aaa cta gtc tat gaa tat 432
 Ser Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140

tgg cat aaa gat cca gat atg atc tgg aat gct gcg cct aaa ccg act 480
 Trp His Lys Asp Pro Asp Met Ile Trp Asn Ala Ala Pro Lys Pro Thr
 145 150 155 160

atg caa aat gcc atg tac cgc gaa gat ttc tgg gag tgt ccg atg gaa 528

Met Gln Asn Ala Met Tyr Arg Glu Asp Phe Trp Glu Cys Pro Met Glu
 165 170 175

gat cga ttt gag agt atg cat gat ttt gag ttc tgc gtc acc cag gat 576

Asp Arg Phe Glu Ser Met His Asp Phe Glu Phe Cys Val Thr Gln Asp
 180 185 190

gag gtg att ttt gac gca gca gac tgt agc cgc ttt ggc cgt gat att 624

Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Phe Gly Arg Asp Ile
 195 200 205

ttt gtg cag gag tca atg acg act aat cgt gca ggg att cgc tgg ctc 672

Phe Val Gln Glu Ser Met Thr Thr Asn Arg Ala Gly Ile Arg Trp Leu
 210 215 220

aaa cgg cat tta gag ccg cgt cgc ttc cgc gtg cat gat att cac ttc 720

Lys Arg His Leu Glu Pro Arg Arg Phe Arg Val His Asp Ile His Phe
 225 230 235 240

cca cta gat att ttc cca tcc cac att gat tgt act ttt gtc ccc tta 768

Pro Leu Asp Ile Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu
 245 250 255

gca cct ggg gtt gtg tta gtg aat cca gat cgc ccc atc aaa gag ggt 816

Ala Pro Gly Val Val Leu Val Asn Pro Asp Arg Pro Ile Lys Glu Gly
 260 265 270

gaa gag aaa ctc ttc atg gat aac ggt tgg caa ttc atc gaa gca ccc 864

Glu Glu Lys Leu Phe Met Asp Asn Gly Trp Gln Phe Ile Glu Ala Pro
 275 280 285

ctc ccc act tcc acc gac gat gag atg cct atg ttc tgc cag tcc agt 912

Leu Pro Thr Ser Thr Asp Asp Glu Met Pro Met Phe Cys Gln Ser Ser
 290 295 300

aag tgg ttg gcg atg aat gtg tta agc att tcc ccc aag aag gtc atc 960

Lys Trp Leu Ala Met Asn Val Leu Ser Ile Ser Pro Lys Lys Val Ile

305 310 315 320
 tgt gaa gag caa gag cat ccg ctt cat gag ttg cta gat aaa cac ggc 1008
 Cys Glu Glu Gln Glu His Pro Leu His Glu Leu Leu Asp Lys His Gly
 325 330 335
 ttt gag gtc tat cca att ccc ttt cgc aat gtc ttt gag ttt ggc ggt 1056
 Phe Glu Val Tyr Pro Ile Pro Phe Arg Asn Val Phe Glu Phe Gly Gly
 340 345 350

 tcg ctc cat tgt gcc acc tgg gat atc cat cgc acg gga acc tgt gag 1104
 Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu
 355 360 365
 gat tac ttc cct aaa cta aac tat acg ccg gta act gca tca acc aat 1152
 Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn
 370 375 380
 ggc gtt tct cgc ttc atc att tag 1176

Gly Val Ser Arg Phe Ile Ile

385 390

<210> 16

<211> 391

<212> PRT

<213> *Cylindrospermopsis raciborskii* AWT205

<400> 16

Met Gln Thr Arg Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Lys

1 5 10 15

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

20 25 30

Pro Gly Asn Arg Pro Ala Leu Arg Asp Lys Asn Ile Ala Lys Met Phe

35 40 45

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

50 55 60

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

65 70 75 80

Val Arg Arg Pro Glu Lys His Asn Phe Gly Leu Ser Val Lys Thr Pro

Phe Glu Val Tyr Pro Ile Pro Phe Arg Asn Val Phe Glu Phe Gly Gly
 340 345 350

Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu
 355 360 365

Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn
 370 375 380

Gly Val Ser Arg Phe Ile Ile
 385 390

<210> 17

<211> 1272

<212> DNA

<213> Homo sapiens

<220><221> CDS

<222> (1)..(1272)

<223> codes for NP_001473.1

<220><221> transit_peptide

<222> (1)..(111)

<223>

mitochondrial localisation signal peptide

<400> 17

atg ctg cgg gtg cgg tgt ctg cgc ggc ggg agc cgc ggc gcc gag gcg 48

Met Leu Arg Val Arg Cys Leu Arg Gly Gly Ser Arg Gly Ala Glu Ala

1 5 10 15

gtg cac tac atc gga tct cgg ctt gga cga acc ttg aca gga tgg gtg 96

Val His Tyr Ile Gly Ser Arg Leu Gly Arg Thr Leu Thr Gly Trp Val

20 25 30

cag cga act ttc cag agc acc cag gca gct acg gct tcc tcc cgg aac 144

Gln Arg Thr Phe Gln Ser Thr Gln Ala Ala Thr Ala Ser Ser Arg Asn

35 40 45

tcc tgt gca gct gac gac aaa gcc act gag cct ctg ccc aag gac tgc 192

Ser Cys Ala Ala Asp Asp Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys

50 55 60

cct gtc tct tct tac aac gaa tgg gac ccc tta gag gaa gtg ata gtg 240

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 65 70 75 80

ggc aga gca gaa aac gcc tgt gtt cca ccg ttc acc atc gag gtg aag 288

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Ile Glu Val Lys
 85 90 95

gcc aac aca tat gaa aag tac tgg cca ttt tac cag aag caa gga ggg 336

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly

100 105 110

cat tat ttt ccc aaa gat cat ttg aaa aag gct gtt gct gaa att gaa 384

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
 115 120 125

gaa atg tgc aat att tta aaa acg gaa gga gtg aca gta agg agg cct 432

Glu Met Cys Asn Ile Leu Lys Thr Glu Gly Val Thr Val Arg Arg Pro
 130 135 140

gac ccc att gac tgg tca ttg aag tat aaa act cct gat ttt gag tct 480

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 145 150 155 160

acg ggt tta tac agt gca atg cct cga gac atc ctg ata gtt gtg ggc 528

Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly
 165 170 175

aat gag att atc gag gct ccc atg gca tgg cgt tca cgc ttc ttt gag 576

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 180 185 190

tac cga gcg tac agg tca att atc aaa gac tac ttc cac cgt ggc gcc 624

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 195 200 205

aag tgg aca aca gct cct aag ccc aca atg gct gat gag ctt tat aac 672

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asn

210	215	220	
cag gat tat ccc atc cac tct gta gaa gac aga cac aaa ttg gct gct			720
Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala			
225	230	235	240
cag gga aaa ttt gtg aca act gag ttt gag cca tgc ttt gat gct gct			768
Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala			
	245	250	255
gac ttc att cga gct gga aga gat att ttt gca cag aga agc cag gtt			816
Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val			
	260	265	270
aca aac tac cta ggc att gaa tgg atg cgt agg cat ctt gct cca gac			864
Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp			
	275	280	285
tac aga gtg cat atc atc tcc ttt aaa gat ccc aat ccc atg cat att			912
Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile			
	290	295	300
gat gct acc ttc aac atc att gga cct ggt att gtg ctt tcc aac cct			960
Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Ile Val Leu Ser Asn Pro			
305	310	315	320
gac cga cca tgt cac cag att gat ctt ttc aag aaa gca gga tgg act			1008
Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr			
	325	330	335
atc att act cct cca aca cca atc atc cca gac gat cat cca ctc tgg			1056
Ile Ile Thr Pro Pro Thr Pro Ile Ile Pro Asp Asp His Pro Leu Trp			
	340	345	350
atg tca tcc aaa tgg ctt tcc atg aat gtc tta atg cta gat gaa aaa			1104
Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys			
	355	360	365
cgt gtt atg gtg gat gcc aat gaa gtt cca att caa aag atg ttt gaa			1152
Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu			
	370	375	380

aag ctg ggt atc act acc att aaa gtt aac att cgt aat gcc aat tcc 1200
Lys Leu Gly Ile Thr Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
385 390 395 400
ctg gga gga ggc ttc cat tgc tgg acc tgc gat gtc cgg cgc cga ggc 1248

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
405 410 415
acc tta cag tcc tac ttg gac tga 1272
Thr Leu Gln Ser Tyr Leu Asp
420
<210> 18
<211> 423
<212> PRT
<213> Homo sapiens
<400> 18
Met Leu Arg Val Arg Cys Leu Arg Gly Gly Ser Arg Gly Ala Glu Ala

1 5 10 15
Val His Tyr Ile Gly Ser Arg Leu Gly Arg Thr Leu Thr Gly Trp Val
20 25 30
Gln Arg Thr Phe Gln Ser Thr Gln Ala Ala Thr Ala Ser Ser Arg Asn
35 40 45
Ser Cys Ala Ala Asp Asp Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys
50 55 60
Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val

65 70 75 80
Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Ile Glu Val Lys
85 90 95
Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly
100 105 110
His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
115 120 125
Glu Met Cys Asn Ile Leu Lys Thr Glu Gly Val Thr Val Arg Arg Pro

130 135 140
 Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 145 150 155 160
 Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly
 165 170 175
 Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 180 185 190
 Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala

 195 200 205
 Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asn
 210 215 220
 Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
 225 230 235 240
 Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 245 250 255
 Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val

 260 265 270
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 275 280 285
 Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile
 290 295 300
 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Ile Val Leu Ser Asn Pro
 305 310 315 320
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr

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

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

385 390 395 400

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly

405 410 415

Thr Leu Gln Ser Tyr Leu Asp

420

<210> 19

<211> 7

<212> PRT

<213> Escherichia coli

<400> 19

Met Thr Asp Val Thr Ile Lys

1 5

<210> 20

<211> 375

<212> PRT

<213> Homo sapiens

<400> 20

Met Thr Asp Val Thr Ile Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys

1 5 10 15

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val

20 25 30

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Ile Glu Val Lys

35 40 45

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly

50 55 60

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu

65 70 75 80

Glu Met Cys Asn Ile Leu Lys Thr Glu Gly Val Thr Val Arg Arg Pro

85 90 95

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser

100 105 110

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

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala

195 200 205

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp

210 215 220

Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala

225 230 235 240

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala

245 250 255

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val

260 265 270

Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp

275 280 285

Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile

290 295 300

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

305 310 315 320

Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr

325 330 335

Ile Val Thr Pro Pro Thr Pro Val Ile Pro Asp Asp His Pro Leu Trp

340 345 350

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys

355 360 365

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu

370 375 380

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

385 390 395 400

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly

405 410 415

Thr Leu Gln Ser Tyr Phe Asp

420

<210> 22

<211> 375

<212> PRT

<213> Rattus norvegicus

<400> 22

Met Thr Asp Val Thr Ile Lys Ala Thr His Pro Leu Pro Lys Asp Cys

1 5 10 15

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val

20 25 30

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys

35 40 45

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Asn Gly Gly

50 55 60

Leu Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Val Glu

65 70 75 80

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

85 90 95

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser

100 105 110

Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Met Val Val Gly

115 120 125

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu

130 135 140

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala

145 150 155 160

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp

165 170 175

Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala

180 185 190

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala

195 200 205

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val

Ser Cys Ala Ala Glu Asp Lys Ala Thr His Pro Leu Pro Lys Asp Cys
 50 55 60
 Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 65 70 75 80
 Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
 85 90 95
 Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Asn Gly Gly
 100 105 110
 Leu Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Val Glu
 115 120 125
 Glu Met Cys Asn Ile Leu Ser Met Glu Gly Val Thr Val Lys Arg Pro
 130 135 140
 Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 145 150 155 160
 Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Met Val Val Gly
 165 170 175
 Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 180 185 190
 Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 195 200 205
 Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp
 210 215 220
 Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
 225 230 235 240
 Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 245 250 255
 Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 260 265 270
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 275 280 285
 Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile

Glu Met Cys Asn Ile Leu Lys Met Glu Gly Val Thr Val Arg Arg Pro
 85 90 95
 Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 100 105 110
 Thr Gly Leu Tyr Gly Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly
 115 120 125
 Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ala Arg Phe Phe Glu
 130 135 140
 Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 145 150 155 160
 Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asn Glu Leu Tyr Asp
 165 170 175
 Gln Asp Tyr Pro Ile Tyr Thr Val Glu Asp Arg His Lys Leu Ala Ala
 180 185 190
 Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 195 200 205
 Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 210 215 220
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 225 230 235 240
 Tyr Arg Val His Ile Val Ser Phe Lys Asp Pro Asn Pro Met His Ile
 245 250 255
 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
 260 265 270
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285
 Ile Ile Thr Pro Pro Ile Pro Val Ile Pro Asp Asp His Pro Leu Trp
 290 295 300
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Gly

65	70	75	80	
gtt cgt cgt ccg gaa aaa cac aat ttt ggt ctg agc gtt aaa acc ccg				345
Val Arg Arg Pro Glu Lys His Asn Phe Gly Leu Ser Val Lys Thr Pro				
	85	90	95	
ttt ttt gaa gtg gaa aat cag tat tgt gca gtg tgt ccg cgt gat gtt				393
Phe Phe Glu Val Glu Asn Gln Tyr Cys Ala Val Cys Pro Arg Asp Val				
	100	105	110	
atg att acc ttt ggt aac gaa att ctg gaa gca acc atg agc cgt cgt				441
Met Ile Thr Phe Gly Asn Glu Ile Leu Glu Ala Thr Met Ser Arg Arg				
	115	120	125	
agc cgt ttt ttt gaa tat ctg ccg tat cgt aaa ctg gtg tat gag tat				489
Ser Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr				
	130	135	140	
tgg cat aaa gat ccg gat atg att tgg aat gca gca ccg aaa ccg acc				537
Trp His Lys Asp Pro Asp Met Ile Trp Asn Ala Ala Pro Lys Pro Thr				
145	150	155	160	
atg cag aat gca atg tat cgt gaa gat ttt tgg gaa tgc ccg atg gaa				585
Met Gln Asn Ala Met Tyr Arg Glu Asp Phe Trp Glu Cys Pro Met Glu				
	165	170	175	
gat cgt ttt gaa agc atg cac gat ttt gaa ttt tgc gtg acc cag gat				633
Asp Arg Phe Glu Ser Met His Asp Phe Glu Phe Cys Val Thr Gln Asp				
	180	185	190	
gaa gtg att ttt gat gca gca gat tgc agc cgt ttt ggt cgt gat att				681
Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Phe Gly Arg Asp Ile				
	195	200	205	
ttt gtt caa gaa agc atg acc acc aat cgt gca ggt att cgc tgg ctg				729
Phe Val Gln Glu Ser Met Thr Thr Asn Arg Ala Gly Ile Arg Trp Leu				
	210	215	220	
aaa cgt cat ctg gaa ccg cgt cgt ttt cgt gtt cat gat att cat ttt				777

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

225 230 235 240

ccg ctg gat att ttt ccg agc cat att gat tgt acc ttt gtt ccg ctg 825

Pro Leu Asp Ile Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu

 245 250 255

gca ccg ggt gtt gtt ctg gtt aat ccg gat cgt ccg att aaa gaa ggt 873

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

 260 265 270

gaa gaa aaa ctg ttt atg gac aac ggc tgg cag ttt att gaa gca ccg 921

Glu Glu Lys Leu Phe Met Asp Asn Gly Trp Gln Phe Ile Glu Ala Pro

 275 280 285

ctg ccg acc agc acc gat gat gaa atg ccg atg ttt tgt cag agc agc 969

Leu Pro Thr Ser Thr Asp Asp Glu Met Pro Met Phe Cys Gln Ser Ser

 290 295 300

aaa tgg ctg gcc atg aat gtt ctg agc att agc ccg aaa aaa gtg atc 1017

Lys Trp Leu Ala Met Asn Val Leu Ser Ile Ser Pro Lys Lys Val Ile

305 310 315 320

tgt gaa gaa caa gaa cat ccg ctg cat gaa ctg ctg gat aaa cat ggt 1065

Cys Glu Glu Gln Glu His Pro Leu His Glu Leu Leu Asp Lys His Gly

 325 330 335

ttt gag gtt tat ccg att ccg ttt cgc aac gtg ttt gaa ttt ggt ggt 1113

Phe Glu Val Tyr Pro Ile Pro Phe Arg Asn Val Phe Glu Phe Gly Gly

 340 345 350

agc ctg cat tgt gca acc tgg gat att cat cgt acc ggc acc tgt gaa 1161

Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu

 355 360 365

gat tat ttt ccg aaa ctg aat tat aca ccg gtt acc gca agc acc aat 1209

Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn

 370 375 380

ggt gtg agc cgt ttt atc att taataagacg tccgcgaggg ccgtgttgcc 1260
 Gly Val Ser Arg Phe Ile Ile
 385 390
 ggtttcttca gagagacg 1278
 <210> 26
 <211> 391
 <212> PRT
 <213> *Cylindrospermopsis raciborskii* AWT205
 <400> 26
 Met Gln Thr Arg Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Lys
 1 5 10 15

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

 Val Arg Arg Pro Glu Lys His Asn Phe Gly Leu Ser Val Lys Thr Pro
 85 90 95
 Phe Phe Glu Val Glu Asn Gln Tyr Cys Ala Val Cys Pro Arg Asp Val
 100 105 110
 Met Ile Thr Phe Gly Asn Glu Ile Leu Glu Ala Thr Met Ser Arg Arg
 115 120 125
 Ser Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140

 Trp His Lys Asp Pro Asp Met Ile Trp Asn Ala Ala Pro Lys Pro Thr
 145 150 155 160
 Met Gln Asn Ala Met Tyr Arg Glu Asp Phe Trp Glu Cys Pro Met Glu
 165 170 175
 Asp Arg Phe Glu Ser Met His Asp Phe Glu Phe Cys Val Thr Gln Asp

180 185 190
 Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Phe Gly Arg Asp Ile
 195 200 205

Phe Val Gln Glu Ser Met Thr Thr Asn Arg Ala Gly Ile Arg Trp Leu
 210 215 220

Lys Arg His Leu Glu Pro Arg Arg Phe Arg Val His Asp Ile His Phe
 225 230 235 240

Pro Leu Asp Ile Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu
 245 250 255

Ala Pro Gly Val Val Leu Val Asn Pro Asp Arg Pro Ile Lys Glu Gly
 260 265 270

Glu Glu Lys Leu Phe Met Asp Asn Gly Trp Gln Phe Ile Glu Ala Pro
 275 280 285

Leu Pro Thr Ser Thr Asp Asp Glu Met Pro Met Phe Cys Gln Ser Ser
 290 295 300

Lys Trp Leu Ala Met Asn Val Leu Ser Ile Ser Pro Lys Lys Val Ile
 305 310 315 320

Cys Glu Glu Gln Glu His Pro Leu His Glu Leu Leu Asp Lys His Gly
 325 330 335

Phe Glu Val Tyr Pro Ile Pro Phe Arg Asn Val Phe Glu Phe Gly Gly
 340 345 350

Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu
 355 360 365

Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn
 370 375 380

Gly Val Ser Arg Phe Ile Ile
 385 390

<210> 27

<211> 1313

<212> DNA

<213> Homo sapiens

<220><221> RBS

<222> (55)..(58)

<223> Shine-Dalgarno-Sequence

<220><221> CDS

<222> (62)..(1189)

<400> 27

ggtctcaaga tatgaccatg attacgcca gcttgcatgc ctgcaggaaa ggagaggatt 60

g atg aca gat gta acg att aaa gcg acg gag ccg ctg ccg aag gat tgc 109

Met Thr Asp Val Thr Ile Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys

1 5 10 15

ccg gtt agc tct tat aat gag tgg gat ccg ttg gag gag gtt atc gtg 157

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val

20 25 30

ggc gcg gcg gag aac gcc tgt gtt cct cca ttc act atc gag gtg aag 205

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Ile Glu Val Lys

35 40 45

gcc aac act tat gaa aag tac tgg ccg ttc tac cag aag caa gga ggt 253

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly

50 55 60

cat tac ttc cct aag gat cat ctg aaa aaa gcg gta gcc gaa att gag 301

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu

65 70 75 80

gag atg tgc aac atc ctg aaa acg gaa ggc gtg acc gtc cgt cgc ccg 349

Glu Met Cys Asn Ile Leu Lys Thr Glu Gly Val Thr Val Arg Arg Pro

85 90 95

gat ccc att gac tgg tca ttg aag tat aaa acc ccg gac ttc gaa agc 397

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser

100 105 110

acg ggt ctt tat tca gcc atg ccg cgt gat atc ctg atc gtt gtt ggt 445

Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly

115 120 125

aat gaa att att gaa gcc cct atg gcc tgg cgg agc cgc ttc ttc gag 493

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu

130 135 140

tac cgt get tat egg tct att atc aaa gat tac ttc cat cgt ggt gca 541

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala

145 150 155 160

aaa tgg acg act get cct aag cca acc atg gca gac gag ttg tac aac 589

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asn

165 170 175

cag gat tac cct atc cat tca gtc gaa gac cgc cac aag ctg gca gca 637

Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala

180 185 190

cag ggt aaa ttt gtc acg act gag ttc gaa ccc tgc ttc gat gct gcc 685

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala

195 200 205

gac ttt att cgt gcc gga cgg gat att ttc gca caa cgc tgc caa gta 733

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val

210 215 220

act aac tac ctc ggc att gag tgg atg cgt cgg cat ctg gcc cct gat 781

Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp

225 230 235 240

tac cgt gtc cac atc att agc ttc aaa gac ccg aac ccg atg cac att 829

Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile

245 250 255

gac gcg act ttt aat atc atc gga ccc ggt att gtg ctg tgc aat cct 877

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

260 265 270

gac gcg ccg tgt cac cag att gac ctt ttt aag aaa gct gga tgg acg 925

Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr

275 280 285

att atc acc cct cca act ccg att atc cct gat gac cac cca ctc tgg 973
 Ile Ile Thr Pro Pro Thr Pro Ile Ile Pro Asp Asp His Pro Leu Trp
 290 295 300

atg tcc tct aaa tgg ctt tcc atg aac gtt ctg atg ctt gac gag aag 1021
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320

cgg gtt atg gtt gac gcc aac gaa gtc ccc att cag aaa atg ttt gaa 1069

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 325 330 335

aag ctt ggc atc act acg att aag gtc aat atc cgt aac gcg aac agc 1117
 Lys Leu Gly Ile Thr Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350

ttg ggc ggt ggt ttt cat tgc tgg acg tgc gat gtg cgc cgc cgg gga 1165
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365

acc ctt cag tcc tat ctc gac taa taagacgtct aaaaaaaaaac cccgccctg 1219
 Thr Leu Gln Ser Tyr Leu Asp
 370 375

acagggcggg gtttttttta gtcgactcta gaggatcccc gggtaccgag ctcgaattca 1279
 ctggccgtcg ttttacagcc aagcttggga gacc 1313

<210> 28
 <211> 375
 <212> PRT
 <213> Homo sapiens

<400> 28
 Met Thr Asp Val Thr Ile Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys
 1 5 10 15
 Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 20 25 30
 Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Ile Glu Val Lys
 35 40 45
 Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 325 330 335

Lys Leu Gly Ile Thr Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365

Thr Leu Gln Ser Tyr Leu Asp
 370 375

<210> 29

<211> 1313

<212> DNA

<213> Rattus norvegicus

<220><221> RBS

<222> (55)..(58)

<220><221> CDS

<222> (62)..(1189)

<400> 29

ggtctcaaga tatgaccatg attacgcaa gcttgcatgc ctgcaggaaa ggagaggatt 60

g atg aca gat gta acg att aaa gct act cat ccg ctt cct aaa gat tgc 109

Met Thr Asp Val Thr Ile Lys Ala Thr His Pro Leu Pro Lys Asp Cys

1 5 10 15

cct gtg tca agc tat aac gaa tgg gac cct ttg gag gag gtg atc gtg 157

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val

20 25 30

ggc cgt gct gag aac gct tgt gta ccc ccc ttt act gtc gaa gta aaa 205

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys

35 40 45

gcc aac act tac gag aag tac tgg ccg ttt tac cag aaa aat ggt ggc 253

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Asn Gly Gly

50	55	60	
ctt tat ttc ccc aaa gat cac ctg aag aaa gcc gtt gca gag gtc gaa			301
Leu Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Val Glu			
65	70	75	80
gag atg tgc aat att ctg agc atg gaa ggt gtg act gtg aaa cgc cca			349
Glu Met Cys Asn Ile Leu Ser Met Glu Gly Val Thr Val Lys Arg Pro			
85	90	95	
gat ccg att gac tgg tca ctt aag tat aag acg ccc gac ttt gag tcg			397
Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser			
100	105	110	
act gga ctt tac agc gcg atg ccg cgc gac att ctg atg gta gtg gga			445
Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Met Val Val Gly			
115	120	125	
aat gag atc att gag gcg ccc atg gcg tgg cgg tcc cgc ttt ttt gag			493
Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu			
130	135	140	
tat cgc gcg tac cgg tcg att atc aag gat tac ttt cac cgg ggc gcg			541
Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala			
145	150	155	160
aag tgg acc acc gct cca aag cct acg atg gct gac gag ctt tac gat			589
Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp			
165	170	175	
caa gat tac ccc att cat tcg gtt gaa gac cgt cac aaa ttg gcg gca			637
Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala			
180	185	190	
caa gga aaa ttc gtt acg act gag ttc gag cct tgc ttt gac gcc gcc			685
Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala			
195	200	205	
gat ttt att cgg gct ggc cgc gac atc ttc gca cag cgt tca caa gta			733

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 210 215 220
 acc aac tac ttg gga atc gag tgg atg cgt cgg cac ctt gcc cca gat 781
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 225 230 235 240
 tac cgg gtc cat atc att tca ttt aaa gat ccc aac ccg atg cac att 829
 Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile
 245 250 255
 gac gct acc ttc aat att att gga cca gga ctg gta ctc tca aac cct 877
 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
 260 265 270
 gat cgg ccc tgt cac cag att gat ctc ttc aag aag gct ggt tgg acg 925
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285
 att gtc acg cca ccg act cct gta atc ccg gat gat cat ccc ctc tgg 973
 Ile Val Thr Pro Pro Thr Pro Val Ile Pro Asp Asp His Pro Leu Trp
 290 295 300
 atg tct tcc aaa tgg ctc tct atg aac gta ctg atg ctc gat gaa aaa 1021
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320
 cgc gtg atg gta gat gcg aat gag gtc cca att caa aaa atg ttc gag 1069
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 325 330 335
 aag ctg ggc atc tct act att aag gtg aac att cgg aat gca aac tcg 1117
 Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350
 ctc gga gga ggt ttc cac tgt tgg act tgt gac gta cgg cgg cgt gga 1165
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365
 act ctc cag agc tac ttc gat taa taagacgtct aaaaaaaaaac cccgccctg 1219
 Thr Leu Gln Ser Tyr Phe Asp

<220><221> CDS

<222> (62)..(1189)

<400> 31

ggtctcaaga taigaccatg attacgccaa gcttgcatgc ctgcaggaaa ggagaggatt 60

g atg aca gat gta acg att aaa gcg acc gat cct ttg cct aaa gat tgc 109

Met Thr Asp Val Thr Ile Lys Ala Thr Asp Pro Leu Pro Lys Asp Cys

1 5 10 15

ccc gtg tcc tcg tat aat gaa tgg gac ccg ctc gag gag gta atc gtt 157

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val

20 25 30

ggt cgg gct gaa aat gcc tgc gtt ccg ccc ttc act gtt gaa gtc aaa 205

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys

35 40 45

gcg aat act tat gaa aag tac tgg ccg ttc tat caa aaa cac ggt ggc 253

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys His Gly Gly

50 55 60

cac tat ttc ccg aaa gac cac ctg aaa aaa gct gtt gct gag att gaa 301

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu

65 70 75 80

gaa atg tgt aac atc ctg aag atg gaa ggc gtt acg gtt cgt cgg cca 349

Glu Met Cys Asn Ile Leu Lys Met Glu Gly Val Thr Val Arg Arg Pro

85 90 95

gac cca atc gat tgg tca ctg aag tac aag acc ccc gac ttt gag tcg 397

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser

100 105 110

act gga ctc tac gga gct atg ccc cgt gat atc ctc atc gtt gtt ggc 445

Thr Gly Leu Tyr Gly Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly

115 120 125

aat gaa atc atc gaa gcc cct atg gcc tgg cgg gca cgc ttc ttc gaa 493

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ala Arg Phe Phe Glu

130 135 140

tac cgg gct tat cgc agc atc atc aag gat tat ttt cac cgg gga gcc 541
 Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 145 150 155 160
 aaa tgg acc acg gcc ccg aaa ccc act atg gct aac gag ttg tat gat 589
 Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asn Glu Leu Tyr Asp
 165 170 175

 caa gac tat ccg atc tat acg gtc gag gac cgg cat aaa ctt gca gcg 637
 Gln Asp Tyr Pro Ile Tyr Thr Val Glu Asp Arg His Lys Leu Ala Ala
 180 185 190
 caa ggt aaa ttt gtg act acg gag ttc gaa cct tgt ttc gat gcc gct 685
 Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 195 200 205
 gat ttt att cgt gca gga cgc gat atc ttt gcg caa cgg agc caa gta 733

 Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 210 215 220
 act aat tac ctc ggt atc gag tgg atg cgg cgt cat ctg gca cca gat 781
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 225 230 235 240
 tac cgt gtt cac atc gtt tcg ttc aaa gat ccg aac ccg atg cac att 829
 Tyr Arg Val His Ile Val Ser Phe Lys Asp Pro Asn Pro Met His Ile
 245 250 255
 gat gcc acc ttc aac atc att ggt ccc ggc ctc gtg ctc agc aat cca 877
 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
 260 265 270
 gac cgg cct tgc cac caa att gac ttg ttc aaa aaa gct ggt tgg acc 925
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285

 att att act cct cca att ccc gtg att cca gac gac cac ccg ctt tgg 973
 Ile Ile Thr Pro Pro Ile Pro Val Ile Pro Asp Asp His Pro Leu Trp
 290 295 300
 atg agc tca aaa tgg ttg agc atg aat gtc ctt atg ctt gat gag aag 1021

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320
 cgt gtc atg gtt gat gct aac gag gta cca atc caa aag atg ttc ggc 1069

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Gly
 325 330 335
 aag ctg ggt att tca acg atc aaa gtc aac att cgg aac gcc aat tca 1117

Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350
 ctg ggc ggt ggc ttt cac tgc tgg act tgt gat gtt cgt cgg cgg ggc 1165

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365
 acg ctt caa agc tat ttc gac taa taagactct aaaaaaaaaac cccgccctg 1219

Thr Leu Gln Ser Tyr Phe Asp
 370 375

acagggcggg gtttttttta gtcgactcta gaggatcccc gggtagccgag ctccaattca 1279
 ctggccgtcg ttttacagcc aagcttggga gacc 1313

<210> 32

<211> 375

<212> PRT

<213> Galeopterus variegatus

<400> 32

Met Thr Asp Val Thr Ile Lys Ala Thr Asp Pro Leu Pro Lys Asp Cys
 1 5 10 15

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 20 25 30

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
 35 40 45

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys His Gly Gly
 50 55 60

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
 65 70 75 80

Glu Met Cys Asn Ile Leu Lys Met Glu Gly Val Thr Val Arg Arg Pro
 85 90 95
 Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 100 105 110
 Thr Gly Leu Tyr Gly Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly
 115 120 125

 Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ala Arg Phe Phe Glu
 130 135 140
 Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 145 150 155 160
 Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asn Glu Leu Tyr Asp
 165 170 175
 Gln Asp Tyr Pro Ile Tyr Thr Val Glu Asp Arg His Lys Leu Ala Ala
 180 185 190

 Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 195 200 205
 Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 210 215 220
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 225 230 235 240
 Tyr Arg Val His Ile Val Ser Phe Lys Asp Pro Asn Pro Met His Ile
 245 250 255

 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
 260 265 270
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285
 Ile Ile Thr Pro Pro Ile Pro Val Ile Pro Asp Asp His Pro Leu Trp
 290 295 300
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320

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

gacgcggatc c 1211

<210> 34

<211> 2028

<212> DNA

<213> Escherichia coli

<400> 34

ttaatctttc tgcgaattga gatgacgcca ctggctgggc gtcaccccgg tttcccgggt 60

aaacaccacc gaaaaatagt tactatcttc aaagccacat tcggtcgaaa taccactgat 120

taacaggcgg ctatgctgga gaagatattg cgcatgacac actctgacct gtcgcagata 180

ttgattgatg gtcattccag tctgctggcg aaattgctga cgcaaacgc gtcactgca 240

cgatgctca tcacaaaatt tatccagcgc aaagggactt ttcaggctag ccgccagccg 300

ggtaatcagc ttatccagca acgtttcgct ggatgttggc ggcaacgaat cactggtgta 360

acgatggcga ttcagcaaca tcaccaactg cccgaacagc aactcagcca tttcgttagc 420

aaacggcaca tgetgaetac tttcatgctc aagctgaccg ataacctgcc gcgcctgcgc 480

catccccatg ctacctaagc gccagtgtgg ttgccctgcg ctggcgtaa atcccgaat 540

cgccccctgc cagtcaagat tcagcttcag acgctccggg caataaataa tattctgcaa 600

aaccagatcg ttaacggaag cgtaggagtg tttatcgtca gcatgaatgt aaaagagatc 660

gccacgggta atgcgataag ggcgatcgtt gattacatgc aggccattac cgcgccagac 720

aatcaccagc tcacaaaaat catgtgtatg ttcagcaaag acatcttgcg gataacggtc 780

agccacagcg actgcctgct ggtcgtggc aaaaaaatca tctttgagaa gttttaactg 840

atgcgccacc gtggctacct cggccagaga acgaagtga ttattcgaat tatggcgtac 900

aaatacgttg agaagattcg cgttattgca gaaagccatc ccgtccctgg cgaatatcac 960

gcggtgacca gttaaactct cggcgaaaaa gcgtcgaaaa gtggttactg tcgctgaatc 1020

cacagcgata ggcgatgtca gtaacgctgg cctcgtctgtg gcgtagcaga tgtcgggctt 1080

tcatcagtcg caggcggttc aggtatcgtt gaggcgtcag tcccgtttgc tgcttaagct 1140

gccgatgtag cgtacgcagt gaaagagaaa attgatccgc cacggcatcc caattcacct 1200

catcggcaaa atggctctcc agccaggcca gaagcaagtt gagacgtgat gcgctgtttt 1260

ccaggttctc ctgcaaacctg cttttacgca gcaagagcag taattgcata aacaagatct 1320

cgcgactggc ggtcagggtt aatcatttt cccttctctg ctgttccatc tgtgcaacca 1380

gctgtcgcac ctgctgcaat acgctgtggt taacgcgcca gtgagacgga tactgcccac 1440

ccagctcttg tggcagcaac tgattcagcc cggcgagaaa ctgaaatcga tccggcgagc 1500

gatacagcac attggtcaga cacagattat cggatatgttc atacagatgc cgatcatgat 1560
 cgcgtacgaa acagaccgtg ccaccggtga tggatataggg ctgcccatta aacacatgaa 1620

taccctgccc atgttcgaca atcacaattt catgaaaatc atgatgatgt tcaggaaaat 1680
 ccgcctgccc gagccggggt tctatcgcca cggacgcgtt accagacgga aaaaaatcca 1740
 cactatgtaa tacggtcata ctggcctcct gatgtcgtca acacggcgaa atagtaatca 1800
 cgaggtcagg ttcttacctt aaattttcga cggaaaacca cgtaaaaaac gtcgattttt 1860
 caagatacag cgigaatttt caggaaatgc ggtgagcatc acatcaccac aattcagcaa 1920
 attgtgaaca tcatcacgtt catctttccc tggttgccaa tggcccattt tctgtcagt 1980
 aacgagaagg tcgcgaattc aggcgctttt tagactggtc gtaatgaa 2028

<210> 35

<211> 107

<212> DNA

<213> Escherichia coli

<400> 35

caaataaac gaaaggctca gtcgaaagac tgggcctttc gttttatctg ttgtttgtcg 60
 gtgaacgctc tctgagtag gacaaatccg ccgggagcgg atttgaa 107

<210> 36

<211> 24

<212> DNA

<213> synthetic oligonucleotide

<400> 36

tctagagatc cgcgggggcc cagg 24

<210> 37

<211> 46

<212> DNA

<213> synthetic oligonucleotide

<400> 37

agttccctac tctcgcgtgc tcgagggatc cgcgtcagcc gcaaat 46

<210> 38

<211> 10235

<212> DNA

<213> various

<400> 38

tcaggcgggc aagaatgta ataaaggccg gataaaactt gtgcttattt ttctttacgg 60
 tctttaaaaa ggccgtaata tccagctgaa cggctctggt ataggtacat tgagcaactg 120
 actgaaatgc ctcaaaatgt tctttacgat gccattggga tataatcaacg gtggtatatac 180
 cagtgatattt tttctcatt ttagcttctt tagctcctga aaatctcgat aactcaaaaa 240
 atacgcccg tagtgatctt atttcattat ggtgaaagt ggaacctctt acgtgccgat 300
 caacgtctca ttttcgcaa aagttggccc agggcttccc ggtatcaaca gggacaccag 360

gatttattta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt 420
 cgggtgatgc tgccaactta ctgatttagt gtatgatggt gttttgagg tgctccagtg 480
 gcttctgttt ctatcagctg tcctcctgt tcagctactg acggggtggt gcgtaacggc 540
 aaaagcaccg ccggacatca gcgctagcgg agtgatact ggcttactat gttggcactg 600
 atgagggtgt cagtgaagt cttcatgtgg caggagaaaa aaggctgcac cggtgctca 660
 gcagaatatg tgatacagga tatattccgc ttctcctc actgactcgc tacgctcgg 720
 cgttcgactg cggcgagcgg aatggctta cgaacggggc ggagatttcc tggaaatgc 780

caggaagata cttaacaggg aagtgagagg gcccgccaa agccgttttt ccataggctc 840
 cgccccctg acaagcatca cgaaatctga cgtcaaatc agtggtagcg aaaccgaca 900
 ggactataaa gataccaggc gtttccccct ggccgctccc tcgtcgcctc tectgttct 960
 gcctttcggg ttaccgggtg cattccgctg ttatggccgc gtttgtctca ttccacgct 1020
 gacactcagt tccggtagg cagttcctc caagctggac tgtatgcacg aacccccgt 1080
 tcagtccgac cgctgcgctt tatccggtaa ctatcgtctt gagtccaacc cggaaagaca 1140
 tgcaaaagca cactggcag cagccactgg taattgattt agaggagtta gtcttgaagt 1200

catgcgccgg ttaaggetaa actgaaagga caagttttgg tgactgcgct cetccaagcc 1260
 agttacctcg gttcaaagag ttggtagctc agagaacctt cgaaaaaccg cctgcaagg 1320
 cggtttttcc gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc 1380
 atcttattaa tcagataaaa tatttctaga tttcagtga atttatctt tcaaatgtag 1440
 cacctgaagt cagccccata cgatataagt tgaattctc atgtttgaca gcttatcatc 1500
 gataagcttt aatcggttag tttatcacag ttaaatgct aacgcagtca ggcaccgtgt 1560
 atgaaatcta acaatgcgct catcgtcatc ctccggcaccg tcaccctgga tgctgtaggc 1620

ataggcttgg ttatgccggt actgccgggc ctcttgccgg atatcgtcca ttccgacagc 1680
 atcggcagtc actatggcgt gctgctagcg ctatatcgtg tgatgcaatt tetatgcgca 1740
 cccgttctcg gagcactgtc cgaccgcttt ggccgcccgc cagtctgct cgcttcgcta 1800

cttggagcca ctatcgacta cgcgatcatg gcgaccacac ccgtcctgtg gatcctctac 1860
 gccggacgca tcgtggccgg catcaccggc gccacaggtg cggttgctgg cgctatatac 1920
 gccgacatca ccgatgggga agatcgggct cgcacttcg ggctcatgag cgcttgtttc 1980
 ggcgtaggta tggtagcagg ccccgtaggc gggggactgt tgggcgcat ctcttgcat 2040

 gcaccattcc ttgcggcggc ggtgctcaac ggcctcaacc tactactggg ctgcttcta 2100
 atgcaggagt cgcataaggg agagcgtcga ccgatgcct tgagagcctt caaccagtc 2160
 agctccttc ggtagggcgg gggcatgact atcgtcggc cacttatgac tgtctcttt 2220
 atcatgcaac tcgtaggaca ggtgcccga ggcctctggg tcattttcgg cgaggaccgc 2280
 tttcgtgga gcgcgacgat gatcggcctg tcgcttggg tattcggaat cttgcacgcc 2340
 ctgctcaag ccttcgtcac tggtagggc accaaactt tcggcgagaa gcaggcatt 2400
 atcggcgga tggcgccga cgcgctggg tacgtctgc tggcgttcgc gacgcgaggc 2460

 tggatggcct tccccattat gattcttctc gcttccggcg gcatcgggat gcccgcgttg 2520
 caggccatgc tgtccaggca ggtagatgac gaccatcagg gacagcttca aggatcgtc 2580
 gcggtcttta ccagcctaac ttgatcatt ggaccgctga tcgtcacggc gatttatgcc 2640
 gcctcggcga gcacatggaa cgggttggca tggattgtag gcgcccctt atacctgtc 2700
 tgctccccg cgttgcgtcg cggtagcagg agccgggcca cctcgacctg aatggaagcc 2760
 ggcgaccct cgtaacgga ttaccactc caagaattgg agccaatcaa ttcttgcgga 2820
 gaactgtgaa tgcgcaaac aacccttggc agaacatata catcgcgtcc gccatctcca 2880

 gcagccgca cggcgcatc tcgggcagcg ttgggtcctg gccacgggtg cgcgatgatc 2940
 tgctcctgtc gttgaggacc cggctaggct ggcggggttg ccttactggt tagcagaatg 3000
 aatcaccgat acgcgagcga acgtgaagcg actgctgctg caaacgtct gcgacctgag 3060
 caacaacatg aatggtcttc ggtttccgtg tttcgtaaag tctggaaac cggaggtccc 3120
 ctactgctg ctgaagtgc ccgcaacaga gagggaacc aaccggtgat accacgatac 3180
 tatgactgag agtcaacgcc atgagcggc tcatttctta ttctgagta caacagtccg 3240
 caccgtgctc cggtagctcc ttccggtggg cgcggggcat gactatctc gccgactta 3300

 tgactgtctt cttatcatg caactcgtag gacaggtgcc ggcagcggc aacagtcctc 3360
 cggccacggg gctgccacc ataccacgc cgaacaagc gccctgcacc attatgttcc 3420
 ggatctgcat cgcaggatgc tgctggctac cctgtggaac acctacatct gtattaacga 3480
 agcctaacc gttttatca ggctctggga ggcagaataa atgatcatat cgtcaattat 3540
 tacctccagc gggagagcct gagcaaacg gcctcaggca tttgagaagc acacggtcac 3600
 actgcttccg gtagtcaata aaccggtaaa ccagcaatag acataagcgg ctatttaacg 3660

accctgccct gaaccgacga cgggctgaa tttgctttcg aatttctgcc attcatccgc 3720

ttattatcac ttattcaggc gtagcaccag gcgtttaagg gcaccaataa ctgccttaaa 3780

aaaattacgc cccgceetgc cactcategc agtactgttg taattcatta agcattctgc 3840

cgacatggaa gccatcacia acggcatgat gaacctgaat cggcagcggc atcagcacct 3900

tgctgccttg cgtataatat ttgcccattg atttaaattt aatctttctg cgaattgaga 3960

tgacgccact ggctgggcgt cateccggtt tcccgggtaa acaccaccga aaaatagtta 4020

ctatcttcaa agccacattc ggtcgaaata tcaactgatta acaggcggct atgctggaga 4080

agatattgcg catgacacac tctgacctgt cgcagatatt gattgatggt cattccagtc 4140

tgctggcgaa attgctgacg caaaacgcgc tcaactgcacg atgcctcatc acaaaattta 4200

tccagcgcaa agggactttt caggctagcc gccagccggg taatcagctt atccagcaac 4260

gtttcgtcgg atgttggcgg caacgaatca ctggtgtaac gatggcgatt cagcaacatc 4320

accaactgcc cgaacagcaa ctcagccatt tcgttagcaa acggcacatg ctgactactt 4380

tcatgctcaa gctgaccgat aacctgccgc gctcgcgcca tccccatgct acctaagcgc 4440

cagtgtgggt gccctgcgct ggcgttaaat cccggaatcg cccctgcca gtcaagattc 4500

agcttcagac gctccgggca ataaataata tctgcaaaa ccagatcgtt aacggaagcg 4560

taggagtgtt tatcgtcagc atgaatgtaa aagagatcgc cacgggtaat gcgataaggg 4620

cgatcgttga gtacatgcag gccattaccg cgccagacia tcaccagctc acaaaaatca 4680

tgtgtatgtt cagcaaagac atcttgcgga taacggctcag ccacagcgac tgctcgtcgg 4740

tcgctggcaa aaaaatcatc tttgagaagt tttaaactgat gcgccaccgt ggctacctcg 4800

gccagagaac gaagttagatt attcgcaata tggcgtacia atacgttag aagattcgcg 4860

ttattgcaga aagccatccc gtccctggcg aatatacgc ggtgaccagt taaactctcg 4920

gcgaaaaage gtcgaaaagt ggttactgtc gctgaatcca cagcgatagg cgatgtcagt 4980

aacgctggcc tcgctgtggc gtagcagatg tgggctttc atcagtcgca ggcggttcag 5040

gtatcgtgta ggcgtcagtc ccgtttgctg cttaaactgc cgatgtagcg tacgcagtga 5100

aagagaaaat tgatccgcca cggcatccca attcacctca tcggcaaaaat ggtcctccag 5160

ccaggccaga agcaagttag gacgtgatgc gctgttttcc aggttctcct gcaaactgct 5220

tttacgcagc aagagcagta attgcataaa caagatctcg cgactggcgg tcgagggtaa 5280

atcattttcc ccttctgct gttccatctg tgcaaccagc tgtcgacct gctgcaatac 5340

gctgtggtta acgcccagc gagacggata ctgcccaccc agctcttctg gcagcaactg 5400

attcagcccc gcgagaaact gaaatcgatc cggcgagcga tacagcacat tggtcagaca 5460
 cagattatcg gtatgttcat acagatgccg atcatgatcg cgtacgaaac agaccgtgcc 5520
 accggtgatg gtatagggtc gccattaaa cacatgaata cccgtgccat gttcgacaat 5580
 cacaatttca tgaaaatcat gatgatgttc aggaaaaatcc gcctgcggga gccggggttc 5640
 tatcgccacg gacgcgttac cagacgaaa aaaatccaca ctatgtaata cggtcatact 5700
 ggctcctga tgtcgtcaac acggcgaaat agtaatcacg aggtcaggtt ctacacctaa 5760
 attttcgacg gaaaaccacg taaaaaacgt cgatttttca agatacagcg tgaattttca 5820

 ggaaatgccc tgagcatcac atcaccacaa ttcagcaaat tgtgaacatc atcacgttca 5880
 tctttccctg gttgccaatg gccattttc ctgtcagtaa cgagaaggtc gcgaattcag 5940
 gcgcttttta gactggtcgt aatgaacatt taatgaatt ccttgggac tctagagatc 6000
 cgcgggggcc caggaggggg gatctggcat ttttgggagg tgtgaaatga gtgagaagat 6060
 cgtcaactcg tggaacgaat gggatgagct cgaggagatg gtggtcggca ttgcggacta 6120
 tgccagcttc gaaccgaaag agccaggcaa ccatcccaa ctgcgcaacc agaacctggc 6180
 cgaaatcacc cctttcccaa gcggcccaaa ggacccgaag gtgctggaga aagcgaacga 6240

 agagctgaat gggctggctt acctgctgaa ggaccacgat gtgatcgtc gtcgtcccga 6300
 gaagatcgac ttcaccaaga gcctgaaaac cccgtatttc gaggttgcca accagtactg 6360
 cggcgtttgt cctcgcgacg tgatgateac gtttggcaac gaaatcatgg aagcggacct 6420
 gtccaaacgt gcacgcttct tcaataacct ccctatcgg aagctggtct acgagtactg 6480
 gaacaaggac gagcacatga tctggaacgc agccccgaaa ccgacctgc aggatagcat 6540
 gtacctgtaa aacttctggg agctctcgtt ggaagaacgc ttcaagcgga tgcacgactt 6600
 cgaattctgc atacccaag acgaggtgat ctctgatgcc gccgattgct cccgcttggg 6660

 taaggacatc ctggtgcagg aaagcatgac caccaatcgc actggcatcc getggetgaa 6720
 gaagcatctc gaaccacgcg gctttcgcgt ccatccggtg cacttcccgt tggacttctt 6780
 ccctagccac atcgactgca cgttcgtacc gttgcgtccg ggtctgatcc tgaccaatcc 6840
 ggaaccccc attcgcgagg aagaggagaa gatcttcaag gagaatggct gggagctgat 6900
 caccgtaccg cagcctacct gctcgaacga cgagatgcc atgttctgcc agagctcгаа 6960
 atggctgtcc atgaacgtc tgagcattag tcccaccaag gtgatctgcg aagaacggga 7020
 aaagccgctg caagaactgc tggacaagca cgggttcгаа gtctttccct tgcctttccg 7080

 ccatgtgttt gatttcggtg gcagctttca ctgtgccacg tgggatattc gccgcaaggg 7140
 cgagtgcgag gactacctgc cgaacctgaa ctaccagccg atttgcggtc gacgcggatc 7200
 cctcagcac gcgagagtag ggaactgcca ggcatcaaat aaaacgaaag gctcagtcga 7260

aagactgggc ctttcgtttt atctgttgtt tgtcggtgaa cgctctcctg agtaggacaa 7320
atccgccggg agcggatttg aacgatgata agctgtcaaa catgagaatt cttgaagacg 7380
aaaggcctc gtgtgtacaa acgttcgtca aaagggcgac acaaaattcc tgcaggggcc 7440
ggcccagcgc cggcggtcga gtggcgacgg cgcggcttgt ccgcgccctg gtagattgcc 7500

tggccgtagg ccagccattt ttgagcggcc agcggcccgcg ataggccgac gcgaagcggc 7560
ggggcgtagg gagcgcagcg accgaagggt aggcgctttt tgcagctctt cggctgtgcg 7620
ctggccagac agttatgcac aggccaggcg ggttttaaga gttttaataa gttttaaaga 7680
gttttagcgc gaaaaatcgc cttttttctc ttttatatca gtcacttaca tgtgtgaccg 7740
gttccaatg tacggctttg ggttccaat gtacgggttc cggttcccaa tgtacggctt 7800
tgggttccca atgtacgtgc tatccacagg aaagagacct tttcgacctt tttcccctgc 7860
tagggcaatt tgcctagca tctgtccgt acattaggaa ccggcggatg cttegcctc 7920

gatcaggttg cggtagcga tgactaggat cgggccagcc tgccccgcct cctccttcaa 7980
atcgtactcc ggcaggatcat ttgaccgat cagcttgcgc acggtgaaac agaacttctt 8040
gaactctccg gcgctgccac tgcgttcgta gatcgtcttg aacaaccatc tggcttctgc 8100
cttgcttgcg gcgcggcgtg ccaggcggta gagaaaacgg ccgatgccgg gatcgatcaa 8160
aaagtaatcg gggatgaaccg tcagcacgtc cgggttcttg ctttctgtga tctcgcggtg 8220
catccaatca actagctcga tctcgatgta ctccggccgc ccggtttcgc tctttacgat 8280
ctttagcgg ctaatcaagg cttcacctc ggataccgtc accaggcggc cgttcttggc 8340

cttcttcgta cgetgcatgg caactgctg ggtgtttaac cgaatgcagg tttctaccag 8400
gtcgtctttc tgctttccgc catcggctcg ccggcagaac ttgagtacgt ccgcaacgtg 8460
tggacggaac acgcggccgg gcttgtctcc cttcccttcc ccggtatcgg tcatggatc 8520
ggtttagatg gaaaccgcca tcagtaccag gtcgtaatcc cacacactgg ccatgccggc 8580
cggccctgcg gaaacctcta cgtgccctc tggaaactcg tagcggatca cctcgccagc 8640
tcgtcggtea cgtttcgaca gacgaaaac ggccacgtcc atgatgtgc gactatcgcg 8700
ggtgcccacg teatagagca tcggaacgaa aaaatctggt tgctcgtcgc cettgggagg 8760

cttctaate gacggcgcac cggctgccgg cggttgccgg gattctttgc ggattcgatc 8820
agcggccgct tgcacgatt caccggggcg tctttctgcc tcgatgcgtt gccgctgggc 8880
ggcctgcgcg gccttcaact tctccaccag gtcacaccc agcgcgcgc cgattgtac 8940
cgggccggat ggtttgcgac cgctcacgc gattcctcgg gcttgggggt tccagtcca 9000
ttgcagggcc ggagacaac ccagccgctt acgcctggcc aaccgcccgt tectccacac 9060
atggggcatt ccacggcgtc ggtgcctggt tgttcttgat tttccatgcc gcctccttta 9120

gccgctaaaa ttcactact catttattca ttgtctcatt tactctggta gctgcgcgat 9180

gtattcagat agcagctcgg taatggtctt gccttggcgt accgcgtaca tcttcagctt 9240

ggtgtgatcc tccgceggca actgaaagt gacccegttc atggtcggcg tgtctgccag 9300

gctggccaac gtigcagcct tgctgctgcg tgcgctcgga cggccggcac ttagcgtgtt 9360

tgtgcttttg ctcatcttct cttttacctca ttaactcaaa tgagttttga ttttaattca 9420

gcggccagcg cctggacctc gcgggcagcg tgcacctcgg gttctgattc aagaacggtt 9480

gtgccggcgg cggcagtgcc tgggtagctc acgcgctgcg tgatcggga ctcaagaatg 9540

ggcagctcgt acccggccag cgctcggca acctcaccgc cgatgcgcgt gcctttgatc 9600

gccccgaca cgacaaaggc cgctttagc cttccatccg tgacctaat gcgctgctta 9660

accagctcca ccaggtcggc ggtggcccat atgtcgtaag ggcttggctg caccggaatc 9720

agcacgaagt cgctgcctt gatcgcggac acagccaagt ccgcccctg gggcgtcctg 9780

tcgatcacta cgaagtccg ccggccgatg gccttcacgt cgcggtcaat cgtcgggcgg 9840

tcgatgccga caacggttag cggttgatct tcccgcacgg ccgccaatc gcgggcactg 9900

ccctggggat cggaatcgac taacagaaca tggccccgg cgagttgcag ggcgcccggc 9960

agatgggttg cgatggtcgt cttgctgac ccgctttct ggttaagtac agcgataacc 10020

ttcatgcgtt cccttgcgt atttgtttat ttactcatcg catcatatac gcagcgaccg 10080

catgacgcaa gcigttttac tcaaatacac atcaccttt tagacggcgg cgctcggttt 10140

cttcagcggc caagctggcc ggccagggcg ccagcttggc atcagacaaa ccggccagga 10200

tttcatgcag ccgcacggtt ccggatgagc attca 10235

<210> 39

<211> 921

<212> DNA

<213> Pseudomonas putida

<220><221> CDS

<222> (1)..(921)

<400> 39

atg agc gct agg cac ttt ctc tcc ctg ctg gac ttc acc acc gac gaa 48

Met Ser Ala Arg His Phe Leu Ser Leu Leu Asp Phe Thr Thr Asp Glu

1 5 10 15

ttg ctc ggg gtg atc cgc cac ggc atc gag ctg aag gac ctg cgc aag 96

Leu Leu Gly Val Ile Arg His Gly Ile Glu Leu Lys Asp Leu Arg Lys

20	25	30	
cga ggc gtg ctg ttc gaa ccg ctg aag aac cgt gtg cta ggc atg atc			144
Arg Gly Val Leu Phe Glu Pro Leu Lys Asn Arg Val Leu Gly Met Ile			
35	40	45	
ttc gaa aag tcc tcg acc cgt acc cgt gtg tcg ttc gag gcc ggc atg			192
Phe Glu Lys Ser Ser Thr Arg Thr Arg Val Ser Phe Glu Ala Gly Met			
50	55	60	
atc cag ctc ggc ggc cag gcc atc ttc ctg tcg ccc cgc gac acc cag			240
Ile Gln Leu Gly Gly Gln Ala Ile Phe Leu Ser Pro Arg Asp Thr Gln			
65	70	75	80
ctg ggc cgc ggc gag cca att ggt gac agc gcc atc gtg ctg tcg agc			288
Leu Gly Arg Gly Glu Pro Ile Gly Asp Ser Ala Ile Val Leu Ser Ser			
85	90	95	
atg gtt gat gtg gtg atg atc cgg acc cac gcc cac agc acc ctg acc			336
Met Val Asp Val Val Met Ile Arg Thr His Ala His Ser Thr Leu Thr			
100	105	110	
gag ttc gcc gcc aag tcg cgt gtg ccc gtg atc aac ggc ctg tcc gac			384
Glu Phe Ala Ala Lys Ser Arg Val Pro Val Ile Asn Gly Leu Ser Asp			
115	120	125	
gaa tcg cac ccg tgc caa ctg ctg gcc gac atg cag acc ttc gtt gaa			432
Glu Ser His Pro Cys Gln Leu Leu Ala Asp Met Gln Thr Phe Val Glu			
130	135	140	
cac cgc ggc tcg att cag ggc aag acc gtg acc tgg atc ggc gat ggc			480
His Arg Gly Ser Ile Gln Gly Lys Thr Val Thr Trp Ile Gly Asp Gly			
145	150	155	160
ttc aac atg tgc aac tcc tat atc gaa gcc gcc agg cag ttc gat ttc			528
Phe Asn Met Cys Asn Ser Tyr Ile Glu Ala Ala Arg Gln Phe Asp Phe			
165	170	175	
cag ctg cgc atc gcc tgc ccc gaa ggc tat gag ccg gat caa cgc ttc			576
Gln Leu Arg Ile Ala Cys Pro Glu Gly Tyr Glu Pro Asp Gln Arg Phe			
180	185	190	

atg gca ctg ggc ggc gac cgc gtg cag atc atc cgg gat gcc agg gaa 624

Met Ala Leu Gly Gly Asp Arg Val Gln Ile Ile Arg Asp Ala Arg Glu

195 200 205

gct gtg cgt gat gca cac ctg gtg gtc acc gat gtc tgg act tcc atg 672

Ala Val Arg Asp Ala His Leu Val Val Thr Asp Val Trp Thr Ser Met

210 215 220

ggt cag gag gag gaa act gca cgg cgc ctg gcg cat ttc gcg cct tac 720

Gly Gln Glu Glu Glu Thr Ala Arg Arg Leu Ala His Phe Ala Pro Tyr

225 230 235 240

cag gtc acc cgc gaa ctg ctc gac ctg gct gca ccc gat gtc ctc ttc 768

Gln Val Thr Arg Glu Leu Leu Asp Leu Ala Ala Pro Asp Val Leu Phe

245 250 255

atg cac tgc ctg ccc gcc cac cgt ggc gag gaa atc agc cag gac ctg 816

Met His Cys Leu Pro Ala His Arg Gly Glu Glu Ile Ser Gln Asp Leu

260 265 270

ctc gac gac cca cgt tgc gtc gcc tgg gac gag gct gaa aac cgc ctg 864

Leu Asp Asp Pro Arg Ser Val Ala Trp Asp Glu Ala Glu Asn Arg Leu

275 280 285

cat gca cag aag gcg ctt ctc gaa ttc ctt gta gaa ccg gct tac cac 912

His Ala Gln Lys Ala Leu Leu Glu Phe Leu Val Glu Pro Ala Tyr His

290 295 300

cac gca tga 921

His Ala

305

<210> 40

<211> 306

<212> PRT

<213> Pseudomonas putida

<400> 40

Met Ser Ala Arg His Phe Leu Ser Leu Leu Asp Phe Thr Thr Asp Glu

1 5 10 15

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

85	90	95	
ccg ctg atc gcc aag cgc ctg atc gaa atc gcc aac gaa acc ggc gct			336
Pro Leu Ile Ala Lys Arg Leu Ile Glu Ile Ala Asn Glu Thr Gly Ala			
100	105	110	
gac gcc att tcc cat ggc gcc acc ggc aag ggt aac gac cag gtg cgc			384
Asp Ala Ile Ser His Gly Ala Thr Gly Lys Gly Asn Asp Gln Val Arg			
115	120	125	
ttc gag ctg ggt gcc tat gcc ctg aag cca ggc gtc aag gtc atc gct			432
Phe Glu Leu Gly Ala Tyr Ala Leu Lys Pro Gly Val Lys Val Ile Ala			
130	135	140	
cca tgg cgc gag tgg gac ctg ctg tcc cgc gaa aag ctg atg gac tac			480
Pro Trp Arg Glu Trp Asp Leu Leu Ser Arg Glu Lys Leu Met Asp Tyr			
145	150	155	160
gcc gag aag cac ggc atc ccg atc gag cgc cac ggc aag aag aag tcg			528
Ala Glu Lys His Gly Ile Pro Ile Glu Arg His Gly Lys Lys Lys Ser			
165	170	175	
ccg tac tcg atg gac gcc aac ctg ctg cac atc tcc tac gag ggc ggt			576
Pro Tyr Ser Met Asp Ala Asn Leu Leu His Ile Ser Tyr Glu Gly Gly			
180	185	190	
gtc ctg gaa gat acc tgg acc gag cac gaa gaa gac atg tgg cgc tgg			624
Val Leu Glu Asp Thr Trp Thr Glu His Glu Glu Asp Met Trp Arg Trp			
195	200	205	
agt gtc tcg cct gag aat gcc ccg gac cag gct acc tac atc gag ctg			672
Ser Val Ser Pro Glu Asn Ala Pro Asp Gln Ala Thr Tyr Ile Glu Leu			
210	215	220	
acc tac cgc aat ggt gac atc gtt gcc atc gac ggc gtc gag aaa tcc			720
Thr Tyr Arg Asn Gly Asp Ile Val Ala Ile Asp Gly Val Glu Lys Ser			
225	230	235	240
ccg gcc acc gtc ctg gca gac ctg aac cgt atc ggt ggt gcc aac ggc			768

Pro Ala Thr Val Leu Ala Asp Leu Asn Arg Ile Gly Gly Ala Asn Gly
 245 250 255
 atc ggc cgt ctg gac atc gtc gaa aac cgt tac gtc ggc atg aag tcg 816
 Ile Gly Arg Leu Asp Ile Val Glu Asn Arg Tyr Val Gly Met Lys Ser
 260 265 270
 cgc ggt tgc tac gaa acg cct ggc ggt acc atc atg ctc aag gca cac 864
 Arg Gly Cys Tyr Glu Thr Pro Gly Gly Thr Ile Met Leu Lys Ala His
 275 280 285
 cgt gcc atc gag tcg atc acc ctg gac cgc gaa gtc gct cac ctg aaa 912
 Arg Ala Ile Glu Ser Ile Thr Leu Asp Arg Glu Val Ala His Leu Lys
 290 295 300
 gat gag ctg atg cca aag tat gcc agc ctg atc tac acc ggc tac tgg 960
 Asp Glu Leu Met Pro Lys Tyr Ala Ser Leu Ile Tyr Thr Gly Tyr Trp
 305 310 315 320
 tgg agc ccg gag cgt ctg atg ctg caa cag atg atc gat gct tcg cag 1008
 Trp Ser Pro Glu Arg Leu Met Leu Gln Gln Met Ile Asp Ala Ser Gln
 325 330 335
 gtc aac gtg aat ggt gtg gtg cgc ctg aaa ctg tac aag ggc aac gtg 1056
 Val Asn Val Asn Gly Val Val Arg Leu Lys Leu Tyr Lys Gly Asn Val
 340 345 350
 acc gtg gtt ggc cgc aag tcg gac gat tcg ctg ttc gat gcc aac atc 1104
 Thr Val Val Gly Arg Lys Ser Asp Asp Ser Leu Phe Asp Ala Asn Ile
 355 360 365
 gcc acc ttt gaa gaa gat ggt ggt gcc tac aac cag gca gat gct gct 1152
 Ala Thr Phe Glu Glu Asp Gly Gly Ala Tyr Asn Gln Ala Asp Ala Ala
 370 375 380
 ggc ttc atc aag ctc aat gca ctg cgt atg cgc att gcc gcc aac aag 1200
 Gly Phe Ile Lys Leu Asn Ala Leu Arg Met Arg Ile Ala Ala Asn Lys
 385 390 395 400
 ggc cgt tcg ctg ctc tga 1218
 Gly Arg Ser Leu Leu

405

<210> 42

<211> 405

<212> PRT

<213> *Pseudomonas putida*

<400> 42

Met Ala Asp Val Lys Lys Val Val Leu Ala Tyr Ser Gly Gly Leu Asp
 1 5 10 15

Thr Ser Val Ile Leu Lys Trp Leu Gln Asp Thr Tyr Asn Cys Glu Val
 20 25 30

Val Thr Phe Thr Ala Asp Leu Gly Gln Gly Glu Glu Val Glu Pro Ala
 35 40 45

Arg Ala Lys Ala Gln Ala Met Gly Val Lys Glu Ile Tyr Ile Asp Asp
 50 55 60

Leu Arg Glu Glu Phe Val Arg Asp Phe Val Phe Pro Met Phe Arg Ala
 65 70 75 80

Asn Thr Val Tyr Glu Gly Glu Tyr Leu Leu Gly Thr Ser Ile Ala Arg
 85 90 95

Pro Leu Ile Ala Lys Arg Leu Ile Glu Ile Ala Asn Glu Thr Gly Ala
 100 105 110

Asp Ala Ile Ser His Gly Ala Thr Gly Lys Gly Asn Asp Gln Val Arg
 115 120 125

Phe Glu Leu Gly Ala Tyr Ala Leu Lys Pro Gly Val Lys Val Ile Ala
 130 135 140

Pro Trp Arg Glu Trp Asp Leu Leu Ser Arg Glu Lys Leu Met Asp Tyr
 145 150 155 160

Ala Glu Lys His Gly Ile Pro Ile Glu Arg His Gly Lys Lys Lys Ser
 165 170 175

Pro Tyr Ser Met Asp Ala Asn Leu Leu His Ile Ser Tyr Glu Gly Gly
 180 185 190

Val Leu Glu Asp Thr Trp Thr Glu His Glu Glu Asp Met Trp Arg Trp

<220><221> CDS

<222> (1)..(1407)

<400> 43

atg agt gaa tcc atg agc acc gag aag acc aat cag tcc tgg ggc ggc	48
Met Ser Glu Ser Met Ser Thr Glu Lys Thr Asn Gln Ser Trp Gly Gly	
1 5 10 15	
cgc ttc agt gag ccc gtc gac gcc ttc gtc gcc cgt ttc acc gcc tcg	96
Arg Phe Ser Glu Pro Val Asp Ala Phe Val Ala Arg Phe Thr Ala Ser	
20 25 30	
gta gat ttc gac aag cgc ctg tac cgt cac gac atc atg ggt tcg att	144
Val Asp Phe Asp Lys Arg Leu Tyr Arg His Asp Ile Met Gly Ser Ile	
35 40 45	
gcc cat gcc acc atg ctg gcg cag gtc ggc gtg ctc agt gat gcc gag	192
Ala His Ala Thr Met Leu Ala Gln Val Gly Val Leu Ser Asp Ala Glu	
50 55 60	
cgc gac acc atc atc gat ggc ctg aaa acc atc cag ggc gag att gaa	240
Arg Asp Thr Ile Ile Asp Gly Leu Lys Thr Ile Gln Gly Glu Ile Glu	
65 70 75 80	
gcc ggc aac ttc gac tgg cgt gtc gac ctc gaa gac gtg cac atg aac	288
Ala Gly Asn Phe Asp Trp Arg Val Asp Leu Glu Asp Val His Met Asn	
85 90 95	
atc gaa gca cgc ctg acc gac cgc atc ggc atc acc ggc aag aag ctg	336
Ile Glu Ala Arg Leu Thr Asp Arg Ile Gly Ile Thr Gly Lys Lys Leu	
100 105 110	
cat act ggg cgt agc cgc aac gac cag gtg gcc acc gac atc cgc ctt	384
His Thr Gly Arg Ser Arg Asn Asp Gln Val Ala Thr Asp Ile Arg Leu	
115 120 125	
tgg ctg cgc gac gaa atc gac ctg atc ctg ggc gaa atc acc cgc ctg	432
Trp Leu Arg Asp Glu Ile Asp Leu Ile Leu Gly Glu Ile Thr Arg Leu	
130 135 140	
cag cag ggc ctg ctg gag cag gca gag cgt gaa gcc gaa acc atc atg	480

Gln Gln Gly Leu Leu Glu Gln Ala Glu Arg Glu Ala Glu Thr Ile Met
 145 150 155 160
 cct ggt ttc acc cac ctg cag acg gcg cag ccg gtc acc ttt ggc cac 528
 Pro Gly Phe Thr His Leu Gln Thr Ala Gln Pro Val Thr Phe Gly His
 165 170 175

 cac ctg ctg gcg tgg ttc gaa atg ctc agc cgc gac tat gag cgc ctg 576
 His Leu Leu Ala Trp Phe Glu Met Leu Ser Arg Asp Tyr Glu Arg Leu
 180 185 190
 gtc gac tgc cgc aag cgc acc aac cgc atg cca ctg ggc agc gcc gcg 624
 Val Asp Cys Arg Lys Arg Thr Asn Arg Met Pro Leu Gly Ser Ala Ala
 195 200 205
 ctg gcc ggc acc acc tac ccg atc gac cgt gaa ctg acc tgc aag ctg 672

 Leu Ala Gly Thr Thr Tyr Pro Ile Asp Arg Glu Leu Thr Cys Lys Leu
 210 215 220
 ctg ggc ttt gaa gcc gtg gcc ggc aac tcg ctg gat ggc gtg tcg gac 720
 Leu Gly Phe Glu Ala Val Ala Gly Asn Ser Leu Asp Gly Val Ser Asp
 225 230 235 240
 cgt gat ttc gcc atc gaa ttc tgc gcc gct gcc agc gtg gcg atg atg 768
 Arg Asp Phe Ala Ile Glu Phe Cys Ala Ala Ala Ser Val Ala Met Met

 245 250 255
 cac ctt tcg cgc ttc tcc gaa gag ctg gtg ctg tgg acc agc gcg cag 816
 His Leu Ser Arg Phe Ser Glu Glu Leu Val Leu Trp Thr Ser Ala Gln
 260 265 270
 ttc cag ttc atc gac ctt ccg gac cgc ttc tgc act ggc agc tcg atc 864
 Phe Gln Phe Ile Asp Leu Pro Asp Arg Phe Cys Thr Gly Ser Ser Ile
 275 280 285

 atg ccg cag aaa aag aac ccg gac gtg cca gag ctg gta cgt ggc aag 912
 Met Pro Gln Lys Lys Asn Pro Asp Val Pro Glu Leu Val Arg Gly Lys
 290 295 300
 agc ggc cgc gtg ttc ggc gcc ctg acc ggc ctg ctg acc ctg atg aaa 960
 Ser Gly Arg Val Phe Gly Ala Leu Thr Gly Leu Leu Thr Leu Met Lys

Leu Ala Ser Arg

465

<210> 44

<211> 468

<212> PRT

<213> Pseudomonas putida

<400> 44

Met Ser Glu Ser Met Ser Thr Glu Lys Thr Asn Gln Ser Trp Gly Gly

1 5 10 15

Arg Phe Ser Glu Pro Val Asp Ala Phe Val Ala Arg Phe Thr Ala Ser

20 25 30

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

35 40 45

Ala His Ala Thr Met Leu Ala Gln Val Gly Val Leu Ser Asp Ala Glu

50 55 60

Arg Asp Thr Ile Ile Asp Gly Leu Lys Thr Ile Gln Gly Glu Ile Glu

65 70 75 80

Ala Gly Asn Phe Asp Trp Arg Val Asp Leu Glu Asp Val His Met Asn

85 90 95

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

100 105 110

His Thr Gly Arg Ser Arg Asn Asp Gln Val Ala Thr Asp Ile Arg Leu

115 120 125

Trp Leu Arg Asp Glu Ile Asp Leu Ile Leu Gly Glu Ile Thr Arg Leu

130 135 140

Gln Gln Gly Leu Leu Glu Gln Ala Glu Arg Glu Ala Glu Thr Ile Met

145 150 155 160

Pro Gly Phe Thr His Leu Gln Thr Ala Gln Pro Val Thr Phe Gly His

165 170 175

His Leu Leu Ala Trp Phe Glu Met Leu Ser Arg Asp Tyr Glu Arg Leu

180 185 190

Val Asp Cys Arg Lys Arg Thr Asn Arg Met Pro Leu Gly Ser Ala Ala
 195 200 205
 Leu Ala Gly Thr Thr Tyr Pro Ile Asp Arg Glu Leu Thr Cys Lys Leu
 210 215 220
 Leu Gly Phe Glu Ala Val Ala Gly Asn Ser Leu Asp Gly Val Ser Asp
 225 230 235 240

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

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

 Ala Leu Arg Gly Phe Ser Thr Ala Thr Asp Leu Ala Asp Tyr Leu Val
 370 375 380
 Arg Arg Gly Leu Pro Phe Arg Asp Cys His Glu Ile Val Gly His Ala
 385 390 395 400
 Val Lys Tyr Gly Val Asp Thr Gly Lys Asp Leu Ala Glu Met Ser Leu
 405 410 415
 Asp Glu Leu Arg Gln Phe Ser Asp Gln Ile Glu Gln Asp Val Phe Ala
 420 425 430

 Val Leu Thr Leu Glu Gly Ser Val Asn Ala Arg Asp His Ile Gly Gly

aatcgccaac gaaaccggcg ctgacgcat ttcccatggc gccaccggca agggtaacga 1380

ccaggtgcgc ttcgagctgg gtgcctatgc cctgaagcca ggctcaagg tcatcgtcc 1440
atggcgcgag tgggacctgc tgtcccgcga aaagctgatg gactacgccg agaagcacgg 1500
catcccgatc gagcggccag gcaagaagaa gtcgccgtac tcgatggacg ccaacctgct 1560
gcacatctcc tacgagggcg gtgtcctgga agatacctgg accgagcacg aagaagacat 1620
gtggcgctgg agtgtctcgc ctgagaatgc cccggaccag gctacctaca tcgagctgac 1680
ctaccgcaat ggtgacatcg ttgccatcga cggcgtcgag aaatccccgg ccacctcct 1740
ggcagacctg aaccgtatcg gtggtgcaa cggcatcggc cgtctggaca tcgtcgaaaa 1800

ccgttacgtc ggcatgaagt cgcgcggttg ctacgaaacg cctggcggta ccatcatgct 1860
caaggcacac cgtgccatcg agtcgatcac cctggaccgc gaagtcgctc acctgaaaga 1920
tgagctgatg ccaaagtatg ccagcctgat ctacaccggc tactggtgga gcccgagcg 1980
tctgatgctg caacagatga tcgatgcttc gcaggtcaac gtgaatggtg tggcgcct 2040
gaaactgtac aagggaacg tgaccgtggt tggccgcaag tcggacgatt cgctgttcga 2100
tgccaacatc gccaccttg aagaagatgg tggcgcctac aaccaggcag atgctgctgg 2160
cttcatcaag ctcaatgcac tgcgtatgcg cattgccgcc aacaagggcc gttcgtgct 2220

ctgattgcta tcgacgccac ttttcgttc acgctgcaa tgagtgaatc catgagcacc 2280
gagaagacca atcagtctg gggcggccgc ttcagtgagc ccgtcgacgc cttcgtcgcc 2340
cgtttaccg cctcggtaga tttcgacaag cgcctgtacc gtcacgacat catgggttcg 2400
attgcccattg ccacctgct ggcgcaggtc ggctgctca gtgatgccga gcgcgacacc 2460
atcatcgatg gcctgaaaac catccagggc gagattgaag ccggcaactt cgactggcgt 2520
gtcgacctc aagacgtgca catgaacatc gaagcacgcc tgaccgaccg catcgccatc 2580
accggcaaga agctgcatac tgggcgtagc cgcaacgacc aggtggccac cgacatccgc 2640

ctttggctgc gcgacgaaat cgacctgatc ctggcgaaa tcaccgcct gcagcagggc 2700
ctgctggagc aggcagagcg tgaagccgaa accatcatgc ctggtttcac ccacctgcag 2760
acggcgcagc cggcacctt tggccaccac ctgctggcgt ggttcgaaat gctcagccgc 2820
gactatgagc gcctggtcga ctgccgcaag cgcaccaacc gcatgccact gggcagcgcc 2880
gcgctggcgc gcaccacctc cccgatcgac cgtgaactga cctgcaagct gctgggcttt 2940
gaagccgtgg ccggcaactc gctggatggc gtgtcggacc gtgatttcgc catcgaattc 3000
tgcgccgctg ccagcgtggc gatgatgcac ctttcgcgt tctccgaaga gctggtgctg 3060

tggaccagcg cgcagttcca gttcatcgac cttccggacc gtttctgcac tggcagctcg 3120
 atcatgccg agaaaaagaa cccggacgtg ccagagctgg tacgtggcaa gagcggccgc 3180
 gtgttcggcg cctgaccgg cctgctgacc ctgatgaaag gccaacctct ggcctacaac 3240
 aaggacaacc aggaagacaa ggaaccgctg ttcgacgccg ccgatacct gcgcgactcg 3300
 ctgcggcct tcctgacat gatcccggcg atcaagccca agcagccat catgcgtgaa 3360
 gcggccctgc gcggtttctc caccgtacc gacctggctg actatctgtg tcgccgtggc 3420
 ctgccgttcc gtgactgcca cgagatcgtt ggccacgcgg tgaagtatgg tgtggacact 3480

ggcaaggacc tggccgagat gagcctggac gaactgcgcc aattcagcga ccagatcgag 3540
 caggacgtgt ttgccgtgct gacgctggaa ggctcggatga atgcgcgtga ccacattggt 3600
 ggtacggcgc cggcgcaggt gcgtgctgcc gtcgttcgtg gcaaggccct gttggcgtct 3660
 cgctaatccc ccaaggctcg agcagcgcag agtagggaac t 3701

<210> 46

<211> 23

<212> DNA

<213> synthetic oligonucleotide

<400> 46

ggatccgctg cagccgcaaa tcg 23

<210> 47

<211> 45

<212> DNA

<213> synthetic oligonucleotide

<400> 47

tttgcggctg acgcgatcc cgacctgcat accgttcag ataag 45

<210> 48

<211> 22

<212> DNA

<213> synthetic oligonucleotide

<400> 48

agttccctac tctcgcgtgc tc 22

<210> 49

<211> 13891

<212> DNA

<213> various

<400> 49

tcaggcgggc aagaatgtga ataaaggccg gataaaactt gtgcttattt ttctttacgg 60
 tctttaaaaa ggccgtaata tccagctgaa cggctctggtt ataggtacat tgagcaactg 120
 actgaaatgc ctcaaaatgt tctttacgat gccattggga tataatcaacg gtggatatatc 180

cagtgatttt tttctccatt ttagcttctt tagctcctga aaatctcgat aactcaaaaa 240
 atacgcccg tagtgatctt atttcattat ggtgaaagt ggaacctctt acgtgccgat 300
 caacgtctca ttttcgcaa aagtggccc agggcttccc ggtatcaaca gggacaccag 360
 gatttattta tctcgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt 420
 cgggtgatgc tgccaactta ctgatttagt gtatgatggt gttttgagg tgctccagtg 480
 gcttctgttt ctatcagctg tcctcctgt tcagctactg acggggtggt gcgtaacggc 540
 aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg 600

atgaggggtg cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgctca 660
 gcagaatatg tgatacagga tataatccgc ttctcctc actgactcgc tacgctcgggt 720
 cgttcactg cggcgagcgg aaatggctta cgaacggggc ggagatttcc tggaaagtgc 780
 caggaagata cttaacaggg aagtgagagg gccgcggcaa agccgttttt ccataggctc 840
 cgccccctg acaagcatca cgaaatctga cgtcaaatc agtggtagcg aaaccgaca 900
 ggactataaa gataccaggc gtttcccctt ggccgctccc tcgtcgcctc tectgttctt 960
 gcctttcgggt ttaccgggtg cattccgctg ttatggccgc gtttgtctca ttccacgctt 1020

gacactcagt tccgggtagg cagttcctc caagctggac tgtatgcacg aacccccctt 1080
 tcagtccgac cgctgcgctt tatccggtaa ctatcgtctt gaggccaacc cggaaagaca 1140
 tgcaaaagca ccaactggcag cagccactgg taattgattt agaggagtta gtcttgaagt 1200
 catgcgccgg ttaaggetaa actgaaagga caagttttgg tgactgcgct cetccaagcc 1260
 agttacctcg gttcaaagag ttggtagctc agagaacctt cgaaaaaccg cctgcaagg 1320
 cggtttttct gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc 1380
 atcttattaa tcagataaaa tatttctaga tttcagtga atttatctt tcaaatgtag 1440

cacctgaagt cagccccata cgatataagt tgaattctc atgtttgaca gcttatcatc 1500
 gataagcttt aatcggttag tttatcacag ttaaatgct aacgcagtca ggcaccgtgt 1560
 atgaaatcta acaatgcgct catcgtcatc ctcgccaccg tcaccctgga tgctgtaggc 1620
 ataggcttgg ttatgccggt actgccgggc ctcttcgggg atatcgtcca ttccgacagc 1680
 atcggcagtc actatggcgt gctgctagcg ctatatgcgt tgatgcaatt tetatgcgca 1740
 cccgttctcg gagcactgtc cgaccgcttt ggccgcccgc cagtctgct cgcttcgcta 1800

cttggagcca ctatcgacta cgcgatcatg gcgaccacac ccgtcctgtg gatcctctac 1860

 gccggacgca tcgtggccgg catcaccggc gccacaggtg cggttgctgg cgcctatatac 1920
 gccgacatca ccgatgggga agatcgggct cgcacattcg ggctcatgag cgcttgtttc 1980
 ggctgtggta tggtagcagg ccccgtaggc gggggactgt tgggcgcat ctcttgcatt 2040
 gcaccattcc ttgcggcggc ggtgtcaac ggcctcaacc tactactggg ctgcttctca 2100
 atgcaggagt cgcataaggg agagcgtcga ccgatgcctt tgagagcctt caaccagtc 2160
 agctccttcc ggtgggcggc gggcatgact atcgtcggc cacttatgac tgtcttcttt 2220
 atcatgcaac tcgtaggaca ggtgccggca gcgctctggg tcattttcgg cgaggaccgc 2280

 ttctcctgga gcgcgacgat gatcggcctg tcgcttgcgg tattcggaat cttgcacgcc 2340
 ctctctcaag ccttcgtcac tggtagcggc accaaacgtt tcggcgagaa gcaggccatt 2400
 atcgcggca tggcggcga cgcgctgggc tacgtcttgc tggcgttcgc gacgcgaggc 2460
 tggatggcct tccccattat gattcttctc gcttccggcg gcatcgggat gcccgcttg 2520
 caggccatgc tgtccaggca ggtagatgac gaccatcagg gacagcttca aggatcgtc 2580
 gcggtcttta ccagcctaac ttgatcatt ggaccgtga tcgtcacggc gatttatgcc 2640
 gcctcggcga gcacatggaa cgggttggca tggattgtag gcgccgcctt atacctgtc 2700

 tgctccccg cgttgcgtcg cggtagcatg agccgggcca cctcgacctg aatggaagcc 2760
 ggccggcacct cgtaacgga ttaccactc caagaattgg agccaatcaa ttcttgcgga 2820
 gaactgtgaa tgcgcaaac aacccttggc agaacatatac catcgcgtcc gccatctcca 2880
 gcagccgca cggcggcctc tcgggcagcg ttgggtcctg gccacgggtg cgcattgatc 2940
 tgctcctgtc gttgaggacc cggctaggct ggcggggttg ccttactggt tagcagaatg 3000
 aatcaccgat acgcgagcga acgtgaagcg actgctgctg caaacgtct gcgacctgag 3060
 caacaacatg aatggtcttc ggtttccgtg tttcgtaaag tctggaaac cggaagtccc 3120

 ctactgtctg ctgaagttgc ccgcaacaga gattggaacc aaccggtgat accacgatac 3180
 tatgactgag agtcaacgcc atgagcggcc tcatttctta ttctgagtta caacagtccg 3240
 caccgtgtc cggtagctcc ttccggtggg cgcggggcat gactatctc gccgactta 3300
 tgactgtctt cttatcatg caactcgtag gacaggtgcc ggcagcgcgc aacagtcccc 3360
 cggccacggg gctgccacc ataccacgc cgaacaaga gccctgcacc attatgttcc 3420
 ggatctgcat cgcaggatgc tgcctggctac cctgtggaac acctacatct gtattaacga 3480
 agcctaacc gttttatca ggctctggga ggcagaataa atgatcatat cgtcaattat 3540

tacctccacg gggagagcct gagcaaacctg gcctcaggca tttgagaagc acacggtcac 3600
 actgcttccg gtagtcaata aaccggtaaa ccagcaatag acataagcgg ctatttaacg 3660
 accctgccct gaaccgacga ccgggtcgaa tttgctttcg aatttctgcc attcatccgc 3720
 ttattatcac ttattcaggc gtagcaccag gcgtttaagg gcaccaataa ctgccttaaa 3780
 aaaattacgc cccgccctgc cactcatcgc agtactgttg taattcatta agcattctgc 3840
 cgacatggaa gccatcaaaa acggcatgat gaacctgaat cggcagcggc atcagcacct 3900
 tgtcgccctg cgtataatat ttgcccatgg atttaaattt aatctttctg cgaattgaga 3960

 tgacgccact ggctgggctg catcccgggt tcccgggtaa acaccaccga aaaatagtta 4020
 ctatcttcaa agccacattc ggtcgaaaata tcaactgatta acaggcggct atgctggaga 4080
 agatattcgc catgacacac tctgacctgt cgcagatatt gattgatggt cattccagtc 4140
 tgctggcgaa attgctgacg caaaacgcgc tcaactgcag atgcctcatc acaaaattta 4200
 tccagcgcga agggactttt caggctagcc gccagccggg taatcagctt atccagcaac 4260
 gtttctgctg atgttggcgg caacgaatca ctggtgtaac gatggcgatt cagcaacatc 4320
 accaactgcc cgaacagcaa ctacagccatt tcgttagcaa acggcacatg ctgactactt 4380

 tcatgctcaa gctgaccgat aacctgccgc gcctgcgcca tccccatgct acctaagcgc 4440
 cagtgtggtt gccctgcgct ggcgttaaat cccggaatcg ccccctgcca gtcaagattc 4500
 agcttcagac gctccgggca ataaataata tttcgcaaaa ccagatcgtt aacggaagcg 4560
 taggagtgtt tatcgtcagc atgaatgtaa aagagatcgc cacgggtaat gcgataaggg 4620
 cgatcgttga gtacatgcag gccattaccg cggcagacaa tcaccagctc acaaaaatca 4680
 tgtgtatgtt cagcaaagac atcttgcgga taacggtcag ccacagcgac tgcctgctgg 4740
 tcgctggcaa aaaaatcatc tttgagaagt ttttaactgat gcgccaccgt ggctacctcg 4800

 gccagagaac gaagttgatt attcgcaata tggcgtacaa atacgttgag aagattcgcg 4860
 ttattgcaga aagccatccc gtccttggcg aatatacgc ggtgaccagt taaactctcg 4920
 gcgaaaaagc gtcgaaaagt ggttactgtc gctgaatcca cagcgatagg cgatgtcagt 4980
 aacgctggcc tcgctgtggc gtagcagatg tcgggctttc atcagtcgca ggcggttcag 5040
 glatcgtga ggcgtcagtc ccgtttgctg cttaagctgc cgatgtagcg tacgcagtga 5100
 aagagaaaat tgatccgcca cggcatccca attcacctca tcggcaaaat ggtcctccag 5160
 ccaggccaga agcaagttga gacgtgatgc gctgttttcc aggttctcct gcaaaactgct 5220

 tttacgcage aagagcagta attgcataaa caagatctcg cgactggcgg tcgagggtaa 5280
 atcattttcc ctttctgct gttccatctg tgcaaccagc tgtcgacct getgcaatac 5340
 gctgtggtta acgcgccagt gagacggata ctgccatcc agctcttgtg gcagcaactg 5400

attcagcccc gcgagaaact gaaatcgatc cggcgagcga tacagcacat tggtcagaca 5460
 cagattatcg gtatgttcat acagatgccg atcatgatcg cgtacgaaac agaccgtgcc 5520
 accggtgatg gtatagggtc gccattaaa cacatgaata cccgtgccat gttcgacaat 5580
 cacaatttca tgaaaatcat gatgatgttc aggaaaatcc gcctgcggga gccggggttc 5640

 tatcgccacg gacgcgttac cagacggaaa aaaatccaca ctatgtaata cggtcatact 5700
 ggccctctga tgtcgtcaac acggcgaaat agtaatcacg aggtcaggtt cttaccttaa 5760
 attttcgacg gaaaaccacg taaaaaacgt cgatttttca agatacagcg tgaattttca 5820
 ggaaatgcgg tgagcatcac atcaccacaa ttcagcaaat tgtgaacatc atcacgttca 5880
 tctttccctg gttgccaatg gcccattttc ctgtcagtaa cgagaaggtc gcgaattcag 5940
 gcgcttttta gactggtcgt aatgaacatt taatgaatt cccttgggac tctagagatc 6000
 cgcggggggc caggaggggg gatctggcat ttttgggagg tgtgaaatga gtgagaagat 6060

 cgtcaactcg tggaacgaat gggatgagct cgaggagatg gtggtcggca ttgcggacta 6120
 tgccagcttc gaaccgaaag agccaggcaa ccatccaaa ctgcgcaacc agaacctggc 6180
 cgaaatcate cccttcccaa gcggcccaaa ggacccgaag gtgctggaga aagcgaacga 6240
 agagctgaat gggctggctt acctctgaa ggaccacgat gtgatcgtgc gtcgtcccga 6300
 gaagatcgac ttcaccaaga gcctgaaaac cccgtatttc gaggttgcca accagtactg 6360
 cggcgtttgt cctcgcgacg tgatgateac gtttggcaac gaaatcatgg aagcgacat 6420
 gtccaaacgt gcacgcttct tcaataacct ccctatcgg aagctggtct acgagtactg 6480

 gaacaaggac gagcacatga tctggaacgc agccccgaaa ccgacatgc aggatagcat 6540
 gtacctgga aacttctggg agctctcgct ggaagaacgc ttcaacggga tgcacgactt 6600
 cgaattctgc atcaccaag acgaggtgat ctctgatgcc gccgattgct cccgcttggg 6660
 taaggacatc ctggtgcagg aaagcatgac caccaatgc actggcatcc getggetgaa 6720
 gaagcatctc gaaccacgcg gctttcgcgt ccatccggtg cacttcccgt tggacttctt 6780
 ccctagccac atcgactgca cgttcgtacc gttgcgtccg ggtctgatcc tgaccaatcc 6840
 ggaacgcccg attcgcgagg aagaggagaa gatcttcaag gagaatggct gggagctgat 6900

 caccgtaccg cagcctacct gctcgaacga cgagatgcc atgttctgcc agagctcгаа 6960
 atggctgtcc atgaacgtcc tgagcattag tcccaccaag gtgatctgcg aagaacggga 7020
 aaagccgctg caagaactgc tggacaagca cgggttcгаа gtctttccct tgcctttccg 7080
 ccatgtgttt gatttcggtg gcagctttca ctgtgccacg tgggatattc gccgcaaggg 7140
 cgagtgcgag gactacctgc cgaacctgaa ctaccagccg atttgcggtc gacgcggatc 7200
 ccgacctgca taccgttgca gataaggtag tcaactcatga gcgctaggca ctttctctcc 7260

ctgctggact tcaccaccga cgaattgctc ggggtgatcc gccacggcat cgagctgaag 7320

gacctgcga agcgaggcgt gctgttcgaa ccgctgaaga accgtgtgct aggcattgatc 7380

ttcgaaaagt cctcgaccgg taccctgtg tcgttcgagg ccggcatgat ccagctcggc 7440

ggccaggcca tcttctgtc gccccgcgac acccagctgg gccgcggcga gccaattggt 7500

gacagcgcca tcgtgctgtc gagcatggtt gatgtggtga tgatccggac ccacgcccac 7560

agcacctga ccgagttcgc cgccaagtgc cgtgtgcccc tgatcaacgg cctgtccgac 7620

gaatcgcacc cgtgccaaact gctggccgac atgcagacct tcgttgaaca ccgcggctcg 7680

attcagggca agaccgtgac ctggatcggc gatggcttca acatgtgcaa ctctatatac 7740

gaagccgcca ggcagttcga tttccagctg cgcattgcct gccccgaagg ctatgagccg 7800

gatcaacgct tcatggcact gggcggcgac cgcgtgcaga tcatccggga tgccagggaa 7860

gctgtgctgt atgcacacct ggtggtcacc gatgtctgga cttccatggg tcaggaggag 7920

gaaactgcac ggcgcctggc gcatttcgcg ccttaccagg tcaccgcga actgctcgac 7980

ctggctgcac ccgatgtcct ctccatgcac tgcttccccg cccaccgtgg cgaggaaatc 8040

agccaggacc tgctcgacga cccacgttcg gtgcctggg acgaggtga aaaccgctg 8100

catgcacaga aggcgcttct cgaattcctt gtagaacgg cttaccacca cgcattgagtc 8160

aaccgtaca ccccgtggag tgatggcatg gcggacgtaa aaaaggtcgt actggcgtat 8220

tccggcggcc ttgatactc ggtgattctc aagtggctgc aggataccta caactgcgaa 8280

gtggtgacct tcaccgtga cctggggcag ggcgaagagg tcgaaccggc ccgtgccaaag 8340

gcccaggcaa tggcggttaa agagatctac atcgacgacc tgcgcgaaga attcgtgctg 8400

gatttcgtgt tccgatgtt ccgcgccaac accgtctacg aaggcgagta cctgctgggt 8460

acttccatcg cccgtccgt gatcgccaag cgcctgatcg aaatcgcaa cgaaaccggc 8520

gctgacgcca tttccatgg cgccaccggc aagggtaacg accaggtgcg cttcgagctg 8580

ggtgcctatg ccctgaagcc aggcgtcaag gtcattcctc catggcgcga gtgggacctg 8640

ctgtcccgcg aaaagctgat ggactacgc gagaagcacg gcatccgat cgagcggcac 8700

ggcaagaaga agtcgccgta ctcgatggac gccaacctgc tgcacatctc ctacgagggc 8760

ggtgtcctgg aagatacctg gaccgagcac gaagaagaca tgtggcgtg gactgtctcg 8820

cctgagaatg ccccggacca ggctacctac atcgagctga cctaccgcaa tggtgacatc 8880

gttgccatcg acggcgtcga gaaatccccg gccaccgtcc tggcagacct gaaccgtatc 8940

ggtggtgcca acggcatcgg ccgtctggac atcgtcgaac accgttacgt cggcatgaag 9000

tcgcgcggtt gctacgaaac gcctggcggg accatcatgc tcaaggcaca ccgtgccatc 9060
gagtcgatca ccctggaccg cgaagtcgct cacctgaaag atgagctgat gccaaagtat 9120
gccagcctga tctacaccgg ctactggtgg agcccggagc gtctgatgct gcaacagatg 9180
atcgatgctt cgcagggtcaa cgtgaatggt gtggtgcgcc tgaactgta caagggcaac 9240
gtgaccgtgg ttggccgcaa gtcggacgat tcgctgttcg atgccaacat cgccaccttt 9300
gaagaagatg gtggtgccta caaccaggca gatgctgctg gcttcatcaa gctcaatgca 9360
ctgcgatatg gcattgccgc caacaagggc cgttcgctgc tctgattgct atcgacgcca 9420

ctttttcggt cagcctgca atgagtgaat ccatgagcac cgagaagacc aatcagtctt 9480
ggggcggccg cticagtgag cccgtcgacg ccttcgtcgc ccgtttcacc gcctcggtag 9540
atttcgacaa ggcctgtac cgtcacgaca tcatgggttc gattgcccat gccaccatgc 9600
tggcgcaggt cggcgtgctc agtgatgccg agcgcgacac catcatcgat ggctgaaaa 9660
ccatccaggg cgagattgaa gccggcaact tcgactggcg tgtcgacctc gaagacgtgc 9720
acatgaacat cgaagcacgc ctgaccgacc gcatcgcat caccggcaag aagctgcata 9780
ctgggcgtag ccgcaacgac caggtggcca ccgacatccg cctttggctg cgcgacgaaa 9840

tcgacctgat ccggggcгаа atcacccgcc tgcagcaggg cctgctggag caggcagagc 9900
gtgaagccga aaccatcatg cctggtttca cccacctgca gacggcgcag ccggtcacct 9960
ttggccacca ctgctggcg tggttcgaaa tgcctagccg cgactatgag cgcctggtcg 10020
actgccgcaa ggcaccaaac cgcatgccac tgggcagcgc cgcgctggcc ggcaccacct 10080
accgatcga ccgtgaactg acctgcaagc tgcctggctt tgaagccgtg gccggcaact 10140
cgctggatgg cgtgtcggac cgtgatcttc ccatcgaatt ctgcgccgt gccagcgtgg 10200
cgatgatgca cctttcgcgc ttctccgaag agctggtgct gtggaccagc gcgcagttcc 10260

agttcatcga ctttcggac cgcttctgca ctggcagctc gatcatgccg cagaaaaaga 10320
accggacgt gccagagctg gtactggca agagcggccg cgtgttcggc gcctgaccg 10380
gcctgtgac cctgatgaaa ggccaaccgc tggcctaaa caaggacaac caggaagaca 10440
aggaaccgt gttcgacgcc gccgataacc tgcgcgactc gctgcgggcc ttcgctgaca 10500
tgatcccggc gatcaagccc aagcacgcca tcatcgtga agcggccctg cgcggtttct 10560
ccaccgtac cgacctggt gactatctgg ttcgccgtgg cctgccgttc cgtgactgcc 10620
acgagatcgt tggccacgcg gtgaagtatg gtgtggacac tggcaaggac ctggccgaga 10680

tgagcctgga cgaactgcgc caattcagcg accagatcga gcaggacgtg tttgccgtgc 10740
tgacgtgga aggtcgggtg aatgcgcgtg accacattgg tggtagggc ccggcgcagg 10800
tgctgctgc cgtcgttctg ggcaaggccc tgttggcgtc tcgctaatcc cccaaggctc 10860

gagcacgcga gaggtaggaa ctgccaggca tcaataaaaa cgaaaggctc agtcgaaaga 10920
ctgggccttt cgttttatct gttgtttgtc ggtgaacgct ctctgagta ggacaaatcc 10980
gccgggagcg gatttgaacg atgataagct gtcaaacatg agaattcttg aagacgaaag 11040
ggcctcgtgt gtacaaacgt tctgcaaaaag ggcgacacaa aattcctgca ggggccggcc 11100

cagcgccggc ggtcgagtgg cgacggcgcg gcttgtccgc gccctggtag attgcctggc 11160
cgtaggccag ccatttttga gcggccagcg gcccgatag gccgacgcga agcggcgggg 11220
cgtagggagc gcagcgaccg aaggtaggc gctttttgca gctcttcggc tgtgcctgg 11280
ccagacagtt atgcacagcg caggcgggtt ttaagagttt taataagttt taaagagttt 11340
taggcggaaa aatcgctttt tttctttttt atatcagtca cttacatgtg tgaccggttc 11400
ccaatgtacg gctttgggtt cccaatgtac gggttccggt tccaatgta cgctttggg 11460
ttccaatgt acgtgctatc cacaggaaag agaccttttc gaccttttc cctgctagg 11520

gcaatttgc ctagcatctg ctccgtacat taggaaccgg cggatgcttc gcctcgatc 11580
aggttgcggt agcgcgatgac taggatcggg ccagcctgcc ccgcctctc cttcaaatcg 11640
tactccggca ggtcatttga cccgatcagc ttgcgcacgg tgaaacagaa cttcttgaac 11700
tctccggcgc tgccactcgc ttcgtagatc gtcttgaaca accatctggc tctgccttg 11760
cctgcggcgc ggcgtgccag gcggtagaga aaacggccga tgccgggatc gatcaaaaag 11820
taatcgggtt gaacctcag cacgtccggg tcttgcctt ctgtgatctc gcggtacatc 11880
caatcaacta gctcgatctc gatgtactcc ggccgcccgg tttcgtctt tacgatctt 11940

tagcggctaa tcaaggettc accctcggat accgteacca ggcggcgtt ctggccttc 12000
ttcgtacgct gcatggcaac gtgcgtggig tttaacgaa tgcaggttc taccagtcg 12060
tctttctgct ttcgccatc ggctcggcgg cagaactga gtactcgc aacgtgtgga 12120
cggaacacgc ggcgggctt gtctccctc ccttccggg atcggttcat ggattcgggt 12180
agatgggaaa ccgccatcag taccagtcg taatcccaca cactggccat gccggccggc 12240
cctgcggaaa cctctacgtg cccgtctgga agctcgtagc ggatcacctc gccagctcgt 12300
cggtcacgct tegacagacg gaaaacggc acgtccatga tctcgcgact atcgggggtg 12360

cccacgtcat agagcatcgg aacgaaaaaa tctggttctc cgtcgccctt gggcgcttc 12420
ctaatcgacg gcgcaccggc tgccggcggg tgcgggatt ctttgcggat tcatcagcg 12480
gccgcttgc acgattcacc gggcgtgct tctgcctcga tgcgttgcg ctggcgcc 12540
tgccggcct tcaacttctc caccaggtca taccagcg ccgcgccgat ttgtaccggg 12600
ccgatgggt tgcgaccgt cacgccgatt cctcggcctt gggggttcca gtgceattgc 12660
agggccggca gacaaccag ccgcttacgc ctggccaacc gcccttctc ccacatgg 12720

ggcattccac ggcgtcggtg cctggttggtt cttgattttc catgccgcct cctttagccg 12780

ctaaaaattca tctactcatt tattcatttg ctcatftact ctggtagctg cgcgatgtat 12840

tcagatagca gtcggtaat ggtcttgcct tggcgtaccg cgtacatctt cagcttgggtg 12900

tgatcctccg ccggcaactg aaagttgacc cgcttcatgg ctggcgtgtc tgccaggctg 12960

gccaacgttg cagccttgtt gctgcgtgcg ctgggacggc cggcacttag cgtgtttgtg 13020

cttttgctca ttttctcttt acctcattaa ctcaaatgag ttttgattta atttcagcgg 13080

ccagcgcctg gacctcgcgg gcagcgtcgc cctcgggttc tgattcaaga acggttgtgc 13140

cggcggcggc agtgcttggg tagctcacgc gctgcgtgat acgggactca agaatgggca 13200

gctcgtaccg ggcagcgc tgcgcaacct caccgccgat gcgcgtgcct ttgatcgcgc 13260

gcgacacgac aaaggccgct tgtagccttc catccgtgac ctcaatgcgc tgcttaacca 13320

gctccaccag gtcggcgggt gcccatatgt cgtaaaggct tggetgcacc ggaatcagca 13380

cgaagtcggc tgccttgatc gcggacacag ccaagtccgc cgcctggggc gctccgtcga 13440

tactacgaa gtcgcccgg ccgatggcct tcacgtcgcg gtcaatcgtc gggcggtcga 13500

tgccgacaac ggttagcggg tgatcttccc gcacggccgc ccaatcgcgg gcaactgccct 13560

gggatcggga atcgactaac agaacatcgg ccccgccgag ttgcagggcg cgggctagat 13620

gggttgcgat ggtcgtcttg cctgaccgc ctttctggtt aagtacagcg ataaccttca 13680

tgcgttcccc ttgcgtatft gtttatttac tcacgcac atatacgcag cgaccgatg 13740

acgcaagctg ttttactcaa atacacatca cttttttaga cggcggcgtc cggtttcttc 13800

agcggccaag ctggccggcc aggcgccag cttggcatca gacaaaccgg ccaggatttc 13860

atgcagccgc acggttccgg atgagcattc a 13891

<210> 50

<211> 1137

<212> DNA

<213> Pseudomonas putida

<220><221> CDS

<222> (1)..(1137)

<400> 50

ttg aca aag cca gcc ata ctc gcc ctt gcc gac ggc agt att ttc cgc 48

Leu Thr Lys Pro Ala Ile Leu Ala Leu Ala Asp Gly Ser Ile Phe Arg

1 5 10 15

ggt gaa gcc atc ggt gcc gac ggt cag acc gtt ggt gag gtg gta ttc 96

Gly Glu Ala Ile Gly Ala Asp Gly Gln Thr Val Gly Glu Val Val Phe
 20 25 30
 aac acc gct atg acc ggc tac cag gaa atc ctt aca gac cct tcc tac 144
 Asn Thr Ala Met Thr Gly Tyr Gln Glu Ile Leu Thr Asp Pro Ser Tyr
 35 40 45
 gcg cag caa atc gtt acc ctg acc tac ccg cac atc ggc aac acc ggt 192
 Ala Gln Gln Ile Val Thr Leu Thr Tyr Pro His Ile Gly Asn Thr Gly
 50 55 60
 act acc ccg gaa gac gcc gag tcg agc cgc gtc tgg tcc gct ggc ctg 240
 Thr Thr Pro Glu Asp Ala Glu Ser Ser Arg Val Trp Ser Ala Gly Leu
 65 70 75 80
 gtc atc cgt gac ctg ccg ctg ctg gcc agc aac tgg cgt aac acc cag 288
 Val Ile Arg Asp Leu Pro Leu Leu Ala Ser Asn Trp Arg Asn Thr Gln
 85 90 95
 tcg ctg cct gag tac ctc aag gcc aac aac gtc gtc gcc atc gcc ggc 336
 Ser Leu Pro Glu Tyr Leu Lys Ala Asn Asn Val Val Ala Ile Ala Gly
 100 105 110
 atc gac acc cgt cgc ctg acc cgt atc ctg cgt gaa aag ggc gcc cag 384
 Ile Asp Thr Arg Arg Leu Thr Arg Ile Leu Arg Glu Lys Gly Ala Gln
 115 120 125
 aac ggc tgc att ctg gcg ggt gac aac atc agc gaa gaa gct gcc atc 432
 Asn Gly Cys Ile Leu Ala Gly Asp Asn Ile Ser Glu Glu Ala Ala Ile
 130 135 140
 gct gct gcc cgc ggc ttc ccg ggc ctg aag ggc atg gac ctg gcc aag 480
 Ala Ala Ala Arg Gly Phe Pro Gly Leu Lys Gly Met Asp Leu Ala Lys
 145 150 155 160
 gtc gtc tcc acc aag gaa cgt tac gag tgg cgc tcc agc gtg tgg gag 528
 Val Val Ser Thr Lys Glu Arg Tyr Glu Trp Arg Ser Ser Val Trp Glu
 165 170 175
 ctg aaa acc gac agc cac ccg acc atc gac gct gcc gac ctg ccg tac 576
 Leu Lys Thr Asp Ser His Pro Thr Ile Asp Ala Ala Asp Leu Pro Tyr

180	185	190	
cac gtg gtt gcc ttc gac tat ggc gtc aag ctg aac atc ctg cgc atg			624
His Val Val Ala Phe Asp Tyr Gly Val Lys Leu Asn Ile Leu Arg Met			
195	200	205	
ctg gtg gcc cgc ggc tgc cgc gtg acc gtg gta cca gcc cag acc ccg			672
Leu Val Ala Arg Gly Cys Arg Val Thr Val Val Pro Ala Gln Thr Pro			
210	215	220	
gcc agc gaa gta ctg gca ctc aac ccg gac ggc gtg ttc ctg tcc aac			720
Ala Ser Glu Val Leu Ala Leu Asn Pro Asp Gly Val Phe Leu Ser Asn			
225	230	235	240
ggc cct ggt gac cct gag ccg tgc gac tac gcg atc cag gcg atc aag			768
Gly Pro Gly Asp Pro Glu Pro Cys Asp Tyr Ala Ile Gln Ala Ile Lys			
245	250	255	
gaa atc ctc gaa acc gag atc ccg gta ttc ggc atc tgc ctc ggc cac			816
Glu Ile Leu Glu Thr Glu Ile Pro Val Phe Gly Ile Cys Leu Gly His			
260	265	270	
cag ctg ctg gcc ctg gcg tcc ggc gcc aag acc gtg aaa atg ggc cac			864
Gln Leu Leu Ala Leu Ala Ser Gly Ala Lys Thr Val Lys Met Gly His			
275	280	285	
ggc cac cac ggt gcc aac cac ccg gtc cag gac ctg gat act ggt gtg			912
Gly His His Gly Ala Asn His Pro Val Gln Asp Leu Asp Thr Gly Val			
290	295	300	
gtc atg atc acc agc cag aac cac ggt ttc gcc gtt gac gag gcg acc			960
Val Met Ile Thr Ser Gln Asn His Gly Phe Ala Val Asp Glu Ala Thr			
305	310	315	320
ctg ccg ggc aac gtt cgc gcc att cac aag tcg ctg ttc gac ggc acc			1008
Leu Pro Gly Asn Val Arg Ala Ile His Lys Ser Leu Phe Asp Gly Thr			
325	330	335	
ctg cag ggt atc gag cgt acc gac aag agc gcg ttc agc ttc cag ggc			1056

Leu Gln Gly Ile Glu Arg Thr Asp Lys Ser Ala Phe Ser Phe Gln Gly
 340 345 350
 cac cct gaa gcg agc ccg ggc ccg acc gac gtc gcg cct ctg ttc gat 1104
 His Pro Glu Ala Ser Pro Gly Pro Thr Asp Val Ala Pro Leu Phe Asp
 355 360 365
 cgt ttc acc gat gcc atg gcc aag cgc cgc tga 1137
 Arg Phe Thr Asp Ala Met Ala Lys Arg Arg

370 375
 <210> 51
 <211> 378
 <212> PRT
 <213> Pseudomonas putida
 <400> 51

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

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

Ile Asp Thr Arg Arg Leu Thr Arg Ile Leu Arg Glu Lys Gly Ala Gln
 115 120 125
 Asn Gly Cys Ile Leu Ala Gly Asp Asn Ile Ser Glu Glu Ala Ala Ile
 130 135 140
 Ala Ala Ala Arg Gly Phe Pro Gly Leu Lys Gly Met Asp Leu Ala Lys

145 150 155 160
 Val Val Ser Thr Lys Glu Arg Tyr Glu Trp Arg Ser Ser Val Trp Glu
 165 170 175

Leu Lys Thr Asp Ser His Pro Thr Ile Asp Ala Ala Asp Leu Pro Tyr
 180 185 190

His Val Val Ala Phe Asp Tyr Gly Val Lys Leu Asn Ile Leu Arg Met
 195 200 205

Leu Val Ala Arg Gly Cys Arg Val Thr Val Val Pro Ala Gln Thr Pro
 210 215 220

Ala Ser Glu Val Leu Ala Leu Asn Pro Asp Gly Val Phe Leu Ser Asn
 225 230 235 240

Gly Pro Gly Asp Pro Glu Pro Cys Asp Tyr Ala Ile Gln Ala Ile Lys
 245 250 255

Glu Ile Leu Glu Thr Glu Ile Pro Val Phe Gly Ile Cys Leu Gly His
 260 265 270

Gln Leu Leu Ala Leu Ala Ser Gly Ala Lys Thr Val Lys Met Gly His
 275 280 285

Gly His His Gly Ala Asn His Pro Val Gln Asp Leu Asp Thr Gly Val
 290 295 300

Val Met Ile Thr Ser Gln Asn His Gly Phe Ala Val Asp Glu Ala Thr
 305 310 315 320

Leu Pro Gly Asn Val Arg Ala Ile His Lys Ser Leu Phe Asp Gly Thr
 325 330 335

Leu Gln Gly Ile Glu Arg Thr Asp Lys Ser Ala Phe Ser Phe Gln Gly
 340 345 350

His Pro Glu Ala Ser Pro Gly Pro Thr Asp Val Ala Pro Leu Phe Asp
 355 360 365

Arg Phe Thr Asp Ala Met Ala Lys Arg Arg
 370 375

<210> 52

<211> 3222

<212> DNA

<213> Pseudomonas putida

<220><221> CDS

<222> (1)..(3222)

<400> 52

atg cca aaa cgt aca gac atc aaa agc atc ctg att ctc ggc gct ggc 48

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

1 5 10 15

ccg atc gtg atc ggc cag gcc tgc gaa ttc gac tac tcc ggc gcc cag 96

Pro Ile Val Ile Gly Gln Ala Cys Glu Phe Asp Tyr Ser Gly Ala Gln

20 25 30

gcc tgt aaa gcc ctg cgc gag gaa ggt ttc cgc gtc atc ctg gtg aac 144

Ala Cys Lys Ala Leu Arg Glu Glu Gly Phe Arg Val Ile Leu Val Asn

35 40 45

tcc aac cca gcc acc atc atg acc gac ccg gcc atg gct gac gcc acc 192

Ser Asn Pro Ala Thr Ile Met Thr Asp Pro Ala Met Ala Asp Ala Thr

50 55 60

tac atc gag ccg atc aag tgg caa tcg gtg gcc aag atc atc gag aaa 240

Tyr Ile Glu Pro Ile Lys Trp Gln Ser Val Ala Lys Ile Ile Glu Lys

65 70 75 80

gag cgc ccg gac gcc gtc ctg ccg acc atg ggt ggc cag acc gcc ctg 288

Glu Arg Pro Asp Ala Val Leu Pro Thr Met Gly Gly Gln Thr Ala Leu

85 90 95

aac tgc gcc ctg gac ctg gag cgc cac ggc gtt ctg gag aag ttc ggc 336

Asn Cys Ala Leu Asp Leu Glu Arg His Gly Val Leu Glu Lys Phe Gly

100 105 110

gtg gag atg atc ggt gcc aac gct gac acc atc gac aag gcc gaa gac 384

Val Glu Met Ile Gly Ala Asn Ala Asp Thr Ile Asp Lys Ala Glu Asp

115 120 125

cgt tcg cgc ttc gac aag gcc atg aag gac atc ggc ctg gag tgc ccg 432

Arg Ser Arg Phe Asp Lys Ala Met Lys Asp Ile Gly Leu Glu Cys Pro

tcg cgt tcg tcc gcc ctg gcc tcc aag gcc acc ggc ttc ccg atc gcc 960
 Ser Arg Ser Ser Ala Leu Ala Ser Lys Ala Thr Gly Phe Pro Ile Ala
 305 310 315 320
 aag atc gcc gcc aag ctg gcc att ggt tac acc ctc gac gag ctg cag 1008

 Lys Ile Ala Ala Lys Leu Ala Ile Gly Tyr Thr Leu Asp Glu Leu Gln
 325 330 335
 aac gac atc act ggc ggt cgc acc cca gcg tcc ttc gaa cct tcg atc 1056
 Asn Asp Ile Thr Gly Gly Arg Thr Pro Ala Ser Phe Glu Pro Ser Ile
 340 345 350
 gac tac gtc gtc acc aag ctg cca cgc ttc gcc ttc gag aaa ttc ccg 1104
 Asp Tyr Val Val Thr Lys Leu Pro Arg Phe Ala Phe Glu Lys Phe Pro
 355 360 365
 aaa gcc gac gcc cgc ctg acc acc cag atg aaa tcc gtg ggt gaa gtc 1152
 Lys Ala Asp Ala Arg Leu Thr Thr Gln Met Lys Ser Val Gly Glu Val
 370 375 380
 atg gcc atc ggc cgt act ttc cag gaa tcc ctg cag aaa gcc ctg cgc 1200
 Met Ala Ile Gly Arg Thr Phe Gln Glu Ser Leu Gln Lys Ala Leu Arg
 385 390 395 400

 ggc ctg gaa gtc ggc gcc tgc ggc ctc gac ccg aaa gtc gac ctg gcc 1248
 Gly Leu Glu Val Gly Ala Cys Gly Leu Asp Pro Lys Val Asp Leu Ala
 405 410 415
 agc ccg gaa gcc gcc age atc ctc aag cgc gaa ctg acc gtg ccg ggt 1296
 Ser Pro Glu Ala Ala Ser Ile Leu Lys Arg Glu Leu Thr Val Pro Gly
 420 425 430
 gcc gag cgt atc tgg tac gtg gct gac gcc atg cgt tcg ggc atg acc 1344

 Ala Glu Arg Ile Trp Tyr Val Ala Asp Ala Met Arg Ser Gly Met Thr
 435 440 445
 tgc gaa gaa atc ttc aat ctg acc ggc atc gac atg tgg ttc ctg gtg 1392
 Cys Glu Glu Ile Phe Asn Leu Thr Gly Ile Asp Met Trp Phe Leu Val
 450 455 460
 cag atg gaa gac ctg atc aag gaa gaa gag aag gtc aag acc ctg gcc 1440

Gln Met Glu Asp Leu Ile Lys Glu Glu Glu Lys Val Lys Thr Leu Ala

465 470 475 480

ctg tcg gca atc gac aag gac tac atg ctg cgc ctc aag cgc aag ggc 1488

Leu Ser Ala Ile Asp Lys Asp Tyr Met Leu Arg Leu Lys Arg Lys Gly

 485 490 495

ttc tcg gac cag cgc ctg gca gta ctg ctg ggt atc acc gac aag aac 1536

Phe Ser Asp Gln Arg Leu Ala Val Leu Leu Gly Ile Thr Asp Lys Asn

 500 505 510

ctg cgt cgc cac cgc cac aag ctg gaa gtg ttc ccg gtg tac aag cgc 1584

Leu Arg Arg His Arg His Lys Leu Glu Val Phe Pro Val Tyr Lys Arg

 515 520 525

gtc gac acc tgc gcc gcc gag ttc gcc acc gac acc gcc tac ctg tac 1632

Val Asp Thr Cys Ala Ala Glu Phe Ala Thr Asp Thr Ala Tyr Leu Tyr

 530 535 540

tcc acc tac gag gaa gag tgc gag gcc aac ccg tcg acc cgc gac aag 1680

Ser Thr Tyr Glu Glu Glu Cys Glu Ala Asn Pro Ser Thr Arg Asp Lys

545 550 555 560

atc atg atc ctg ggt ggc ggc ccg aac cgt atc ggc caa ggt atc gag 1728

Ile Met Ile Leu Gly Gly Gly Pro Asn Arg Ile Gly Gln Gly Ile Glu

 565 570 575

ttc gac tac tgc tgc gta cac gcc gcc ctg gcg ctg cgt gaa gac ggt 1776

Phe Asp Tyr Cys Cys Val His Ala Ala Leu Ala Leu Arg Glu Asp Gly

 580 585 590

tac gag acc atc atg gtc aac tgc aac ccg gaa acc gtc tcc acc gac 1824

Tyr Glu Thr Ile Met Val Asn Cys Asn Pro Glu Thr Val Ser Thr Asp

 595 600 605

tac gac act tcc gac cgc ctg tac ttc gag ccg ctg acc ctg gaa gac 1872

Tyr Asp Thr Ser Asp Arg Leu Tyr Phe Glu Pro Leu Thr Leu Glu Asp

 610 615 620

gtg ctg gaa gtc tgc cgc gtc gag aag ccg aag ggc gtg atc gtt cac 1920
 Val Leu Glu Val Cys Arg Val Glu Lys Pro Lys Gly Val Ile Val His
 625 630 635 640

tac ggc ggc cag acc ccg ctg aag ctg gcc cgc gct ctg gaa gag gct 1968
 Tyr Gly Gly Gln Thr Pro Leu Lys Leu Ala Arg Ala Leu Glu Glu Ala
 645 650 655

ggc gtg ccg atc atc ggt acc agc cct gac gcc atc gac cgc gcc gaa 2016

Gly Val Pro Ile Ile Gly Thr Ser Pro Asp Ala Ile Asp Arg Ala Glu
 660 665 670

gac cgc gag cgc ttc cag cag atg gtt cag cgc ctg agc ctg ctg cag 2064
 Asp Arg Glu Arg Phe Gln Gln Met Val Gln Arg Leu Ser Leu Leu Gln
 675 680 685

ccg cca aac gcc acc gtg cgc agc gaa gaa gaa gcc atc cgt gct gcg 2112
 Pro Pro Asn Ala Thr Val Arg Ser Glu Glu Glu Ala Ile Arg Ala Ala

690 695 700

ggc agc atc ggc tac ccg ctg gtc gtg cgt ccg tcc tac gta ctg ggc 2160
 Gly Ser Ile Gly Tyr Pro Leu Val Val Arg Pro Ser Tyr Val Leu Gly
 705 710 715 720

ggc cgt gcc atg gag atc gtc tac gag ctg gac gag ctc aag cgc tac 2208
 Gly Arg Ala Met Glu Ile Val Tyr Glu Leu Asp Glu Leu Lys Arg Tyr
 725 730 735

ctg cgt gaa gcg gta caa gtg teg aac gac agc ccg gta ctg etc gac 2256
 Leu Arg Glu Ala Val Gln Val Ser Asn Asp Ser Pro Val Leu Leu Asp
 740 745 750

cac ttc ctc aac tgc gcc atc gag atg gac gtg gat gcg gtg tgc gac 2304
 His Phe Leu Asn Cys Ala Ile Glu Met Asp Val Asp Ala Val Cys Asp
 755 760 765

ggc acc gac gtg gtc atc ggc gcg atc atg cag cac atc gag cag gcc 2352

Gly Thr Asp Val Val Ile Gly Ala Ile Met Gln His Ile Glu Gln Ala
 770 775 780

ggc gta cac tcc ggc gac tcg gcg tgc tcg ctg cca cct tac tcg ctg 2400

Gly Val His Ser Gly Asp Ser Ala Cys Ser Leu Pro Pro Tyr Ser Leu
785 790 795 800
agc aag gaa gtg cag gac gaa gtc cgc gtt cag gtc aag aaa atg gcg 2448
Ser Lys Glu Val Gln Asp Glu Val Arg Val Gln Val Lys Lys Met Ala
805 810 815
ctg gag ctg ggt gta gtc ggc ctg atg aac gtg cag ctg gcc ctg cag 2496
Leu Glu Leu Gly Val Val Gly Leu Met Asn Val Gln Leu Ala Leu Gln
820 825 830
ggc gac aag atc tac gtg atc gaa gtc aac ccg cgt gcc tcg cgt acc 2544
Gly Asp Lys Ile Tyr Val Ile Glu Val Asn Pro Arg Ala Ser Arg Thr
835 840 845
gta ccg ttc gtg tcc aag tgc atc ggc acg tcc ctg gcg atg atc gca 2592
Val Pro Phe Val Ser Lys Cys Ile Gly Thr Ser Leu Ala Met Ile Ala
850 855 860
gcc cgt gtc atg gcg ggt aaa acc ctg aaa gag ctg ggc ttc acc cag 2640
Ala Arg Val Met Ala Gly Lys Thr Leu Lys Glu Leu Gly Phe Thr Gln
865 870 875 880
gaa atc atc ccg aac ttc tac agc gtg aag gaa gcc gtc ttc ccg ttc 2688
Glu Ile Ile Pro Asn Phe Tyr Ser Val Lys Glu Ala Val Phe Pro Phe
885 890 895
gcc aag ttc cca ggg gtt gac ccg atc ctc ggc cct gag atg aaa tcg 2736
Ala Lys Phe Pro Gly Val Asp Pro Ile Leu Gly Pro Glu Met Lys Ser
900 905 910
acc ggt gaa gtg atg ggt gtc ggt gac agc ttc ggt gaa gcc ttc gcc 2784
Thr Gly Glu Val Met Gly Val Gly Asp Ser Phe Gly Glu Ala Phe Ala
915 920 925
aaa gcc cag atg ggt gcc agc gaa gtg ctg ccg act ggc ggt acc gcg 2832
Lys Ala Gln Met Gly Ala Ser Glu Val Leu Pro Thr Gly Gly Thr Ala
930 935 940
ttc atc agc gtg cgc gac gac gac aag cca caa gtg gcc ggc gtt gcc 2880
Phe Ile Ser Val Arg Asp Asp Asp Lys Pro Gln Val Ala Gly Val Ala

1 5 10 15
 Pro Ile Val Ile Gly Gln Ala Cys Glu Phe Asp Tyr Ser Gly Ala Gln
 20 25 30

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

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

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

 Lys Asp Asn Cys Ile Ile Val Cys Ser Ile Glu Asn Phe Asp Pro Met
 225 230 235 240
 Gly Val His Thr Gly Asp Ser Ile Thr Val Ala Pro Ala Gln Thr Leu
 245 250 255

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

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

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

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

 Leu Ser Ala Ile Asp Lys Asp Tyr Met Leu Arg Leu Lys Arg Lys Gly
 485 490 495
 Phe Ser Asp Gln Arg Leu Ala Val Leu Leu Gly Ile Thr Asp Lys Asn

His Phe Leu Asn Cys Ala Ile Glu Met Asp Val Asp Ala Val Cys Asp
 755 760 765
 Gly Thr Asp Val Val Ile Gly Ala Ile Met Gln His Ile Glu Gln Ala
 770 775 780
 Gly Val His Ser Gly Asp Ser Ala Cys Ser Leu Pro Pro Tyr Ser Leu
 785 790 795 800

 Ser Lys Glu Val Gln Asp Glu Val Arg Val Gln Val Lys Lys Met Ala
 805 810 815
 Leu Glu Leu Gly Val Val Gly Leu Met Asn Val Gln Leu Ala Leu Gln
 820 825 830
 Gly Asp Lys Ile Tyr Val Ile Glu Val Asn Pro Arg Ala Ser Arg Thr
 835 840 845
 Val Pro Phe Val Ser Lys Cys Ile Gly Thr Ser Leu Ala Met Ile Ala
 850 855 860

 Ala Arg Val Met Ala Gly Lys Thr Leu Lys Glu Leu Gly Phe Thr Gln
 865 870 875 880
 Glu Ile Ile Pro Asn Phe Tyr Ser Val Lys Glu Ala Val Phe Pro Phe
 885 890 895
 Ala Lys Phe Pro Gly Val Asp Pro Ile Leu Gly Pro Glu Met Lys Ser
 900 905 910
 Thr Gly Glu Val Met Gly Val Gly Asp Ser Phe Gly Glu Ala Phe Ala
 915 920 925

 Lys Ala Gln Met Gly Ala Ser Glu Val Leu Pro Thr Gly Gly Thr Ala
 930 935 940
 Phe Ile Ser Val Arg Asp Asp Asp Lys Pro Gln Val Ala Gly Val Ala
 945 950 955 960
 Arg Asp Leu Ile Ala Leu Gly Phe Glu Val Val Ala Thr Ala Gly Thr
 965 970 975
 Ala Lys Val Ile Glu Ala Ala Gly Leu Lys Val Arg Arg Val Asn Lys
 980 985 990

 Val Thr Glu Gly Arg Pro His Val Val Asp Met Ile Lys Asn Asp Glu

995	1000	1005	
Val Ser Leu Ile Ile Asn Thr Thr Glu Gly Arg Gln Ser Ile Ala			
1010	1015	1020	
Asp Ser Tyr Ser Ile Arg Arg Asn Ala Leu Gln His Lys Ile Tyr			
1025	1030	1035	
Cys Thr Thr Thr Ile Ala Ala Gly Glu Ala Ile Cys Glu Ala Leu			
1040	1045	1050	
Lys Phe Gly Pro Glu Lys Thr Val Arg Arg Leu Gln Asp Leu His			
1055	1060	1065	
Ala Gly Leu Lys Ala			
1070			
<210> 54			
<211> 44			
<212> DNA			
<213> synthetic oligonucleotide			
<400> 54			
agagcctgag caaactggcc gactgatgtg gtgttgctt cttc		44	
<210> 55			
<211> 45			
<212> DNA			
<213> synthetic oligonucleotide			
<400> 55			
ccgtgtgctt ctcaaagcc tatttggtaa tgcctatgct ttcag		45	
<210> 56			
<			
211> 4841			
<212> DNA			
<213> Pseudomonas putida			
<400> 56			
agagcctgag caaactggcc gactgatgtg gtgttgctt cttcgcgggc ttgcccgctc		60	
ccacaggtgc ggcgttgctc ttgggctggt gcgatccctg tgggggcgca ccagcccgcg		120	
cagaagccaa cacagatgtc agtctgtcgc attgatgtgt tttagagaca catccgcggt		180	
agaccgaaaa cacacttttc tgtaagctac agcttttagtg tgtccactaa aagcgcgcag		240	
cgaattgaat tcagcacatg aaagcggggt gacgtgtcca tacgtcactc cgcttttttg		300	

caacctgcga tcgccctttc atgcttgatt tacgggaggt cttcttgaca aagccagcca 360

 tactgcacct tgcgcacggc agtatTTTTc gcggtgaagc catcgggtgcc gacggtcaga 420
 ccgttgggtga ggtgggtattc aacaccgcta tgaccggcta ccaggaaatc cttacagacc 480
 cttcctacgc gcagcaaatc gttaccctga cctaccgca catcggcaac accggtacta 540
 ccccgaaga cgccgagtgc agcccgctct ggtccgctgg cctggtcac cgtgacctgc 600
 cgctgctggc cagcaactgg cgtaacacc agtcgctgcc tgagtacctc aaggccaaca 660
 acgtcgtcgc catcgccggc atcgacacc gtcgcctgac ccgtatcctg cgtgaaaagg 720
 gcgccagaa cggctgcatt ctggcgggtg acaacatcag cgaagaagct gccatcgctg 780

 ctgcccgcg cttcccgggc ctgaagggca tggacctggc caaggtcgtc tccaccaagg 840
 aacgttacga gtggcgtcc agcgtgtggg agctgaaaac cgacagccac ccgacctcg 900
 acgtgcccga cctgccgtac cacgtggttg ccttcgacta tggcgtcaag ctgaacatcc 960
 tgcgcatgct ggiggccgc ggctgccgcg tgacctgggt accagcccag accccggcca 1020
 gcgaagtact ggcactcaac ccggacggcg tgttcctgtc caacggccct ggtgacctg 1080
 agccgtgcga ctacgcgatc cagcgatca aggaaatcct cgaaaccgag atcccgtat 1140
 tcggcatctg cctcgccac cagctgctgg cctggcgtc cggcccaag accgtgaaaa 1200

 tggccacgg ccaccacgtt gccaacacc cgtccagga cctggatact ggtgtggtca 1260
 tgatcaccag ccagaaccac ggtttcgccg ttgacgaggc gacctgccg ggcaacgttc 1320
 gcgccattca caagtcgctt ttcagcggca cctgcaggg tatcgagcgt accgacaaga 1380
 gcgcttcag cttccagggc cacctgaag cgagcccggg cccgaccgac gtcgcctc 1440
 tgttcgatcg tttcaccgat gccatggcca agcggcctg agcatcctgc ttcaaggccc 1500
 cggccggca cctgcccgc ccggcagcgc ctgaccaga ttgttcaaag cggcttgccg 1560
 actgaccccg gatttgagtg accaccatgc caaaacgtac agacatcaa agcatcctga 1620

 ttctcggcgc tggcccgatc gtgatcggc agcctgcga attcgactac tccggcggcc 1680
 aggcctgtaa agcctgcgc gaggaaggtt tccgctcat cctggtgaac tccaaccag 1740
 ccaccatcat gaccgaccg gccatggctg acgccaccta catcgagccg atcaagtggc 1800
 aatcgggtgc caagatcacc gagaaagagc gcccgacgc cgtcctgcc accatgggtg 1860
 gccagaccg cctgaactgc gccctggacc tggagcggca cggcgttctg gagaagtccg 1920
 gcgtggagat gatcgggtcc aacgtgaca ccatcgaca ggccgaagac cgttcgcct 1980
 tcgacaaggc catgaaggac atcgccctgg agtcccgcg ctccggtatc gcccacagca 2040

tggaaagaggc caatgCGGtc ctcgagaagc tcggcttccc gtgcatcatt cGcccgtcgt 2100
 tcacCATggg cggcaccggc ggcggtatcg cttacaaccg tgaagagttc gaagaaatct 2160
 gcaccCGtgg tctggacctg tcgCCgacca aagagctgct gatcgacgaa tcgctgatcg 2220
 gctggaagga atacgagatg gaggtgggtcc gcgacaagaa ggacaactgc atcatcgtct 2280
 gctcgatcga gaacttcgac cCGatgggtg tgcacaccgg tgactcgatc actgttGCC 2340
 cggcacagac cctgaccgac aaggaatacc agatcatgCG caacgcctcg ctggcgggtgc 2400
 tcgctgaaat cgggtgtggaa accggcgggt ccaacgtcca gttcggcatt tgcccgaaca 2460

 ccggccgcat ggttgtcAtc gagatgaacc cgcgcgtgtc gcgttcgtcc gccttggcct 2520
 ccaagccac cggcttcccg atcCCaaga tcCCgcca gctggccatt ggttacacc 2580
 tcgacgagct gcagaacgac atcactggcg gtCGacccc agcgtccttc gaacctcga 2640
 tcgactacgt cgtcaccaag ctgCCacgt tcgcttcga gaaattccg aaagccgacg 2700
 cccgctgac caccagatg aaatccgtgg gtgaagtcAt ggccatcggc cgtactttcc 2760
 aggaatccct gcagaaagcc ctgCGcgGCC tggaaGtcgg cgcctgCGgc ctCGaccga 2820
 aagtcgacct ggccagcccg gaagccgcca gcatcctcaa gcgcgaaGtc accgtgCCgg 2880

 gtGCCgagcg tatctggtac gtggctgacg ccatgcgttc gggcatgacc tgCGaagaaa 2940
 tcttcaatct gaccggcAtc gacatgtggt tcctggtgca gatggaagac ctgatcaagg 3000
 aagaagagaa ggtcaagacc ctggccctgt cggcaatcga caaggactac atgctgCGcc 3060
 tcaagCGaa gggcttctcg gaccagCGcc tggcagtact gctgggtatc accgacaaga 3120
 acctgCGtcg ccaccGCCac aagctggaag tgttcccggt gtacaagCGc gtcgacacct 3180
 gcgCCcgga gttCGccacc gacaccgctt acctgtactc cacctacgag gaagagtCG 3240
 aggccaacc gtcgaccCG gacaagatca tgatcctggg tggcggcccg aaccgtatcg 3300

 gccaaGgtat cGagttCGac tactgctgCG tacacCGcg cctggcgtg cgtgaagacg 3360
 gttacgagac catcatggtc aactgcaacc cggaaaccgt ctccaccgac tacgacactt 3420
 ccgaccgctt gtacttcgag cCGtGacc tggaaagcgt gctggaagtc tgccgctcg 3480
 agaagccgaa gggcgtgAtc gttcactacg gcggccagac cccgctgaag ctggcccgcg 3540
 ctctggaaga ggctggcgtg cCGatcatcg gtaccagccc tgacccatc gaccgCGccg 3600
 aagaccgCGa gcgcttccag cagatggttc agcgcctgag cctgctgCag ccgcaaacg 3660
 ccaccgtCG cagCGaagaa gaagccatcc gtgctgCGgg cagcatCGgc taccgctgg 3720

 tcgtgCGtcc gtCctacgta ctggcCGGCC gtCCatgga gatcgtctac gagctggacg 3780
 agctcaagcg ctacctgCGt gaagcGGtac aagtgtCGaa cGacagcccg gtactgctcg 3840
 accacttctt caactgCGcc atCGagatgg acgtggatgc ggtgtgCGac ggaccCGacg 3900

tggatcatcg cgcgatcatg cagcacatcg agcaggccgg cgtacactcc ggcgactcgg 3960
 cgtgctcgt gccaccttac tcgctgagca aggaagtgca ggacgaagtc cgcgttcagg 4020
 tcaagaaaat ggcgctggag ctgggtgtag tcggcctgat gaacgtgcag ctggccctgc 4080
 agggcgacaa gatctacgtg atcgaagtca acccgctgac ctcgctacc gtaccgttcg 4140

 tgtccaagtg catcggcagc tcctggcga tgatcgagc ccgtgtcatg ggggtaaaa 4200
 ccctgaaaga gctgggcttc acccaggaaa tcatcccgaa cttctacagc gtgaaggaag 4260
 ccgtcttccc gttcgccaag ttcccagggg ttgaccgat cctcggccct gagatgaaat 4320
 cgaccgtga agtgatgggt gtcggtgaca gcttcggtga agccttcgcc aaagcccaga 4380
 tgggtgccag cgaagtgtg ccgactggcg gtaccgctt catcagcgtg cgcgacgacg 4440
 acaagccaca agtggccggc gttgcccgcg acctgatcg cctgggcttc gaagtggttg 4500
 ccactgccgg caccgccaag gttatcgagg cggctggcct gaaagtgcgc cgtgtgaaca 4560

 aggtgaccga aggtcgccct cacgtggctg acatgatcaa gaacgacgaa gtgtcgtga 4620
 tcatcaacac caccgaaggt cgccagtcga tcgccactc ctactcgatt cgtcgcaatg 4680
 cgctgcagca caagatttac tgcaccacta ccattgcggc tggtaagcc atctgcaag 4740
 cgctgaaatt cggtcggaa aagaccgttc gtcgcttga ggatctgat gcaggactga 4800
 aagcatgagc attaccaat aggcatttga gaagcacacg g 4841

 <210> 57
 <211> 18689
 <212> DNA
 <213> various
 <400> 57

 tcaggcgggc aagaatgtga ataaaggccg gataaaactt gtgcttattt ttctttacgg 60

 tctttaaaaa ggccgtaata tccagctgaa cggctcgtt ataggtacat tgagcaactg 120
 actgaaatgc ctcaaatgt tctttacgat gccattggga tatatcaacg gtggtatc 180
 cagtgatfff tttctcatt ttagcttct tagctcctga aaatctgat aactcaaaaa 240
 atacgccgg tagtgatctt atttcattat ggtgaaagt ggaacctctt acgtgccgat 300
 caacgtctca ttttcgcaa aagttggccc agggcttccc ggtatcaaca gggacaccag 360
 gatttattta ttctgcaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt 420
 cgggtgatgc tgccaacta ctgatttagt gtatgatggt gttttgagg tgctccagtg 480

 gcttctgttt ctatcagctg tcctcctgt tcagctactg acggggtggt gcgtaacggc 540
 aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg 600

atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtagctca 660
 gcagaatatg tgatacagga tatattccgc ttctcgcctc actgactcgc tacgctcggc 720
 cgttcgactg cggcgcagcgg aaatggctta cgaacggggc ggagatttcc tggaaatgc 780
 caggaagata cttaacaggg aagtgagagg gccgcggcaa agccgttttt ccataggctc 840
 cgccccctg acaagcatca cgaaatctga cgctcaaatc agtggtggcg aaaccgaca 900

 ggactataaa gataccaggc gtttcccctt ggccggctccc tcgtgcgctc tectgttctt 960
 gcctttcggc ttaccgggtg cattccgctg ttatggccgc gtttgtctca ttccacgctt 1020
 gacactcagt tccgggtagg cagttcgcctc caagctggac tgtatgcacg aacccccctt 1080
 tcagtccgac cgctgcgctt tatccggtaa ctatcgtctt gagtccaacc cggaaagaca 1140
 tgcaaaagca ccaactggcag cagccactgg taattgattt agaggagtta gtcttgaagt 1200
 catgcgccgg ttaaggctaa actgaaagga caagttttgg tgactgcgct cctccaagcc 1260
 agttacctcg gttcaaagag ttggtagctc agagaacctt cgaaaaaccg cctgcaagg 1320

 cggtttttcc gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc 1380
 atcttattaa tcagataaaa tatttctaga ttccagtca atttatctct tcaaatgtag 1440
 cacctgaagt cagccccata cgatataagt tgtaattctc atgtttgaca gcttatcatc 1500
 gataagcttt aatgcggtag tttatcacag ttaaattgct aacgcagtca ggcaccgtgt 1560
 atgaaatcta acaatgcgct catcgtcatc ctccggcaccg tcaccctgga tgctgtaggc 1620
 ataggcttgg ttatgccgtt actgccgggc ctcttgcggg atatcgtcca ttccgacagc 1680
 atcgccagtc actatggcgt gctgctagcg ctatatgcgt tgatgcaatt tctatgcgca 1740

 cccgttctcg gagcactgtc cgaccgcttt ggccgccgcc cagtcctgct cgcttcgcta 1800
 cttggagcca ctatcacta cgcgatcatg gcgaccacac ccgtcctgtg gatcctctac 1860
 gccggacgca tcgtggccgg catcaccggc gccacaggtg cggttgctgg cgcctatata 1920
 gccgacatca ccgatgggga agatcgggct cgccacttcc ggctcatgag cgcttgtttc 1980
 ggcttgggta ttgtggcagg ccccgtaggc gggggactgt tgggcgcat ctcttgcata 2040
 gcaccattcc ttgcggcggc ggtgctcaac ggcctcaacc tactactggg ctgcttctca 2100
 atgcaggagt cgcataaggg agagcgtcga ccgatgcctt tgagagcctt caaccagtc 2160

 agctccttcc ggtgggcggg gggcatgact atcgtcggcg cacttatgac tgtcttcttt 2220
 atcatgcaac tcgtaggaca ggtgccggca gcgctctggg tcattttcgg cgaggaccgc 2280
 tttcgtgga gcgcgacgat gatcggcctg tcgcttgcgg tattcggaaat cttgcacgcc 2340
 ctgctcaag ccttctgcaac ttgtccggcc accaaacgtt tcggcgagaa gcaggccatt 2400
 atcgccggca ttggcggcca cgcgctgggc tacgtcttgc tggcgttccg gacgcgaggc 2460

tggatggcct tcccattat gattcttctc gcttccggcg gcatcgggat gcccgcgttg 2520
 caggccatgc tgtccaggca ggtagatgac gaccatcagg gacagcttca aggatcgctc 2580

 gcggctctta ccagcctaac ttcgatcatt ggaccgetga tcgtcacggc gatttatgcc 2640
 gcctcggcga gcacatggaa cgggttggca tggattgtag gcgccgcct ataccttgtc 2700
 tgctccccg cgttgcgtcg cgggtcatgg agccgggcca cctcgacctg aatggaagcc 2760
 ggccggcacct cgtaacgga ttcaccactc caagaattgg agccaatcaa ttcttgcgga 2820
 gaactgtgaa tgcgcaaacc aacccttggc agaacatata catcgcgtcc gccatctcca 2880
 gcagccgcac gcggcgcatc tcgggcagcg ttgggtctcg gccacgggtg cgcgatgatcg 2940
 tgctctgtc gttgaggacc cggctaggct ggccgggttg ccttactggt tagcagaatg 3000

 aatcacccgat acgcgagcga acgtgaagcg actgctgctg caaaacgtct gcgacctgag 3060
 caacaacatg aatggtcttc ggtttccgtg tttcgtaaag tctggaaacg cgggaagtccc 3120
 ctacgtgctg ctgaagtggc ccgcaacaga gagtggaaacc aaccggtgat accacgatac 3180
 tatgactgag agtcaacgcc atgagcggcc tcatttctta ttctgagtta caacagtccg 3240
 caccgtgtc cggtagctcc ttccgggtggg cgcggggcat gactatcgtc gccgcaacta 3300
 tgactgtctt ctttatcatg caactcgtag gacaggtgcc ggccagcggc aacagtcccc 3360
 cggccacggg gcctgccacc ataccacgc cgaacaagc gccctgcacc attatgttcc 3420

 ggatctgcat cgcaggatgc tgctggctac cctgtggaac acctacatct gtattaacga 3480
 agcgtcaacc gttttatca ggctctggga ggcagaataa atgatcatat cgtcaattat 3540
 tacctccacg gggagagcct gagcaaacctg gccgactgat gtggtgttgg ctcttccgcg 3600
 ggcttgcccc ctcccacagg tgcggcgttg ctcttgggct ggtgcatcc ctgtgggggc 3660
 gcaccagccc gcgcagaagc caacacagat gtcagtctgt cgcattgatg tgttttagag 3720
 acacatccgc ggtagaccga aaacacactt tctgtaagc tacagcttta gtgtgtccac 3780
 taaaagcgcg cagcgaattg aattcagcac atgaaagcgg ggtgacgtgt ccatacgtca 3840

 ctccgctttt ttgcaacctg cgatgcctt ttcattgctt atttacggga ggtcttcttg 3900
 acaaagccag ccatactcgc ccttgccgac ggcagtatit tccgcggtga agccatcggc 3960
 gccgacggtc agaccgttgg tgaggtggta ttcaacaccg ctatgaccgg ctaccaggaa 4020
 atcttacag accttctca cgcgcagcaa atcgttacce tgacctacce gcacatcggc 4080
 aacaccggtg ctaccccgga agacgccgag tcgagccgcg tctggtccgc tggcctggtc 4140
 atccgtgacc tgccgtgct gccagcaac tggcgttaaca cccagtcgct gcctgagtac 4200

ctcaaggcca acaacgtcgt cgccatcgcc ggcatcgaca cccgtcgcct gacccgtatc 4260

ctgctgaaa agggcgccca gaacggctgc attctggcgg gtgacaacat cagcgaagaa 4320

gtgccatcg ctgctgcccc cggttcccc ggctgaagg gcatggacct ggccaaggtc 4380

gtctccacca aggaacgtta cgagtggcgc tccagcgtgt gggagctgaa aaccgacagc 4440

caccgacca tcgacgtcgc cgacctgccg taccacgtgg ttgccttcga ctatggcgtc 4500

aagctgaaca tectgcgcat gctggtggcc cgcggctgcc gcgtgacct ggtaccagcc 4560

cagaccccg ccagcgaagt actggcactc aaccggacg gcgtgttct gtccaacggc 4620

cctggtgacc ctgagccgtg cgactacgcg atccaggcga tcaaggaaat cctcgaacc 4680

gagatcccgg tattcgcat ctgcctcggc caccagctgc tggccctggc gtccggcgcc 4740

aagaccgtga aaatgggcca cggccaccac ggtgccaacc acccggcca ggacctggat 4800

actggtgtgg tcatgatcac cagccagaac cacggtttcg ccgttgacga ggcgacctg 4860

ccgggcaacg ttcgcgcat tcacaagtgc ctgttcgacg gcacctgca ggtatcgag 4920

cgtagcaca agagcgcgtt cagcttcag ggccacctg aagcgagccc gggcccgacc 4980

gacgtcgcgc ctctgttcga tcgtttcacc gatgcatgg ccaagcgcg ctgagcatec 5040

tgcttaagg ccccgggcgg gcacctgccg cggccggcag cgcctgacct agattgttca 5100

aagcggcttg ccgactgacc ccgatttga gtgaccacca tgccaaaacg tacagacatc 5160

aaaagcatcc tgattctcgg cgctggcccc atcgtgatcg gccaggcctg cgaattcgac 5220

tactccggcg cccaggcctg taaagccctg cgcgaggaag gtttccgct catcctggtg 5280

aactccaacc cagccaccat catgaccgac cggccatgg ctgacgccac ctacatcgag 5340

ccgatcaagt ggcaatcggg ggccaagatc atcgagaaag agcggccgga gccctcctg 5400

ccgacctggt gtggccagac gcacctgaac tgcacctgg acctggagcg ccacggcgtt 5460

ctggagaagt tcggcgtgga gatgatcggg gccaacgctg acaccatcga caaggccgaa 5520

gaccgttcgc gtttcgaca ggccatgaag gacatcggcc tggagtgcc gcgctccggt 5580

atcgcccaca gcatggaaga ggccaatcgc gtctcagaga agctcgctt cccgtgcatc 5640

attcggcctg cgttaccat gggcggcacc ggcggcggta tcgcttaca cgtgaagag 5700

ttcgaagaaa tctgacccc tggctcggac ctgtcggca ccaaagagct gctgatcgac 5760

gaatcgtga tcggctggaa ggaatcagag atggaggtgg tccgcgaca gaaggacaac 5820

tgcatcatcg tctgctgat cgagaacttc gaccgatgg gtgtgcacac cggtgactcg 5880

atcactgttg ccccgccaca gacctgacc gacaaggaat accagatcat gcgcaacgcc 5940

tcgctggcgg tgctgctga aatcgggtg gaaaccggcg gttccaacgt ccagttcggc 6000
 atttgcccga acaccggccg catggttgc atcgagatga acccgcgcgt gtcgcttcg 6060
 tccgccctgg cctccaaggc caccggcttc ccgatcgcca agatcgccgc caagctggcc 6120
 attggttaca ccctcgacga gctgcagaac gacatcactg gcggtcgcac cccagcgtcc 6180
 ttcgaacctt cgatcgacta cgtcgtcacc aagctgccac gcttcgcctt cgagaaattc 6240
 ccgaaagccg acgcccgcct gaccaccag atgaaatccg tgggtgaagt catggccatc 6300
 ggccgtactt tccaggaatc cctgcagaaa gcctgcgcg gcctggaagt cggcgcctgc 6360

 ggccctgacc cgaagtcga cctggccagc ccggaagccg ccagcatcct caagcgcgaa 6420
 ctgaccgtgc cgggtgccga gcgtatctgg tacgtggctg acgcatgcg ttcgggcatg 6480
 acctgcaag aatcttcaa tctgaccggc atcgacatgt ggttcctggt gcagatggaa 6540
 gacctgatca aggaagaaga gaaggtcaag accctggccc tgtcggcaat cgacaaggac 6600
 tacatgctgc gcctcaagcg caaggcttc tcggaccagc gcctggcagt actgctgggt 6660
 atcaccgaca agaacctgcg tcgccaccgc cacaagctgg aagtgttccc ggtgtacaag 6720
 cgcgtcgaca cctgcgcccg cgagttcgcc accgacaccg cctacctgta ctccacctac 6780

 gaggaagagt gcgaggccaa cccgtcgacc cgcgacaaga tcatgatcct ggggtggcggc 6840
 ccgaaccgta tcggccaagg tatcgagttc gactactgct gcgtacacgc cgcctggcg 6900
 ctgctggaag acggttacga gaccatcatg gtcaactgca acccgaaac cgtctccacc 6960
 gactacgaca ctccgaccg cctgtacttc gagccgctga ccctggaaga cgtgctggaa 7020
 gtctgccgcg tcgagaagcc gaaggcgctg atcgttact acggcggcca gaccccgctg 7080
 aagctggccc gcgctctgga agaggctggc gtgccgatca tcggtaccag ccctgacgcc 7140
 atcgaccgcg ccgaagaccg cgagcgcttc cagcagatgg ttcagcgctt gacctgctg 7200

 cagccgcaa acgccaccgt gcgcagcga gaagaagcca tccgtgctgc gggcagcatc 7260
 ggctaccgcg tggtcgtgcg tccgtcctac gtactggcg gccgtgcat ggagatcgtc 7320
 tacgagctgg acgagctcaa gcgctacctg cgtgaagcgg tacaagtgtc gaacgacagc 7380
 ccggtactgc tegaccactt cctcaactgc gccatcgaga tggacgtgga tgcggtgtgc 7440
 gacggcaccg acgtggtcat cggcgcgatc atgcagcaca tcgagcagc cggcgtacac 7500
 tccggcgact cggcgtgctc gctgccacct tactcgtga gcaaggaagt gcaggacgaa 7560
 gtcccgcttc aggtcaagaa aatggcgctg gagctgggtg tagtcggcct gatgaacgtg 7620

 cagctggccc tgcaggcgca caagatctac gtgatcgaag tcaaccgcg tgcctcgcgt 7680
 accgtaccgt tcgtgtcaa gtgcatcggc acgtccctgg cgatgatcgc agcccgtgtc 7740
 atggcgggta aaacctgaa agagctgggc ttcaccaggg aatcatccc gaacttctac 7800

agcgtgaagg aagccgtctt cccgttcgcc aagttcccag gggttgacce gatcctcggc 7860
 cctgagatga aatgcaccgg tgaagtgatg ggtgtcggtg acagcttcgg tgaagccttc 7920
 gccaaagccc agatgggtgc cagcgaagtg ctgccgactg gcggtaccgc gttcatcagc 7980
 gtgcgcgacg acgacaagcc acaagtggcc ggcgttgccc gcgacctgat cgccctgggc 8040

 ttcgaagtgg ttgccactgc cggcaccgcc aagttatcg aggcggctgg cctgaaagtg 8100
 cgccgtgtga acaaggtgac cgaaggtcgc cctcacgtgg tcgacatgat caagaacgac 8160
 gaagtgtcgc tgatcatcaa caccaccgaa ggtcgccagt cgatcgccga ctctactcgc 8220
 attcgtcgca atgcgctgca gcacaagatt tactgcacca ctaccattgc ggctggtgaa 8280
 gccatctcgc aagcgtgaa attcggctcg gaaaagaccg ttcgtcgtt gcaggatctg 8340
 catgcaggac tgaagcatg agcattacca aataggcatt tgagaagcac acggtcacac 8400
 tgcttcgggt agtcaataaa ccggtaaacc agcaatagac ataagcggtt atttaacgac 8460

 cctgccctga accgacgacc gggtcgaatt tgctttcgaa tttctgcat tcacccgctt 8520
 attatcactt attcaggcgt agcaccaggc gtttaagggc accaataact gccttaaaaa 8580
 aattacgccc cgcctgcca ctcatcgcag tactgttgta attcattaag cattctgccc 8640
 acatggaagc catcacaaac ggcatgatga acctgaatcg ccagcggcat cagcaccttg 8700
 tcgccttgcg tataatattt gcccatggat ttaaatttaa tctttctgcg aattgagatg 8760
 acgccactgg ctgggcgtca tcccggtttc ccgggtaaac accaccgaaa aatagtact 8820
 atcttcaaag ccacattcgg tcgaaatc actgattaac aggcggctat gctggagaag 8880

 atattgcgca tgacacactc tgacctgtcg cagatattga ttgatggtca ttccagtctg 8940
 ctggcgaaat tgctgacgca aaacgcgctc actgcacgat gcctcatcac aaaattatc 9000
 cagcgcaaag ggacttttca ggctagccgc cagccgggta atcagcttat ccagcaacgt 9060
 ttcgctggat gttggcggca acgaatcact ggtgtaacga tggcgattca gcaacatcac 9120
 caactgcccg aacagcaact cagccatttc gttagcaaac ggacatgct gactacttc 9180
 atgctcaage tgaccgataa cctgccgcgc ctgcgccatc cccatgctac ctaagcgcca 9240
 gtgtggttgc cctgcgctgg cgtaaatcc cggaatcgcc ccctgccagt caagattcag 9300

 cttcagacgc tccgggcaat aaataatatt ctgcaaaacc agatcgtaa cggaagcgt 9360
 ggagtgttta tcgtcagcat gaatgtaaaa gagatcgcca cgggtaatgc gataagggcg 9420
 atcgttgagt acatgcaggc cattaccgcg ccagacaatc accagctcac aaaaatcatg 9480
 tgtatgttca gcaaagacat cttgcggata acggtcagcc acagcactg cctgctggtc 9540
 gctggcaaaa aatcatctt tgagaagttt taactgatgc gccaccgtgg ctacctcggc 9600
 cagagaacga agttgattat tcgcaatatg gcgtacaaat acgttgagaa gattcgcgtt 9660

attgcagaaa gccatcccg tccctggcgaa taccacgcgg tgaccagtta aactctcggc 9720

 gaaaaagcgt cgaaaagtgg ttactgtcgc tgaatccaca gcgataggcg atgtcagtaa 9780
 cgctggcctc gctgtggcgt agcagatgtc gggctttcat cagtgcgagg cggttcaggt 9840
 atcgctgagg cgtcagtccc gtttctgtct taagctgccg atgtagcgta cgcagtgaaa 9900
 gagaaaattg atccgccacg gcatcccaat tcacctcacc ggcaaatgg tcctccagcc 9960
 aggccagaag caagttgaga cgtgatgcgc tgtttccag gttctcctgc aaactgcttt 10020
 tacgcagcaa gagcagtaat tgataaaca agatctcgcg actggcggtc gagggtaaat 10080
 cattttcccc ttctgtctgt tccatctgtg caaccagctg tcgcacctgc tgcaatagc 10140

 tgtggtaac gcgccagtga gacggatact gccatccag ctcttgtggc agcaactgat 10200
 tcagcccggc gagaaactga aatcgatccg gcgagcgata cagcacattg gtcagacaca 10260
 gattatcggt atgttcatac agatgccgat catgatcgcg tacgaaacag accgtgccac 10320
 cggatgatgt atagggctgc ccattaaaca catgaatacc cgtgccatgt tcgacaatca 10380
 caatttcattg aaaatcatga tgatgttcag gaaaatccgc ctgcgggagc cggggttcta 10440
 tcgccacgga cgcgttacca gacggaaaaa aatccacact atgtaatacg gtcatactgg 10500
 cctcctgatg tcgtcaacac ggcgaaatag taatcacgag gtcaggttct taccttaaat 10560

 tttcgacgga aaaccacgta aaaaacgtcg atttttcaag atacagcgtg aattttcagg 10620
 aatgcggtg agcatcacat caccacaatt cagcaaatg tgaacatcat cacgttcacc 10680
 tttcctggt tgccaatggc ccattttcct gtcagtaacg agaaggtcgc gaattcagc 10740
 gctttttaga ctggtcgtaa tgaacattta aatgaattcc cttgggactc tagagatccg 10800
 cgggggcccc ggagggggga tctggcattt ttgggaggtg tgaatgagt gagaagatcg 10860
 tcaactcgtg gaacgaatgg gatgagctcg aggagatggt ggtcggcatt gcggactatg 10920
 ccagcttcga accgaaagag ccaggcaacc atcccaaact gcgcaaccag aacctggcgg 10980

 aatcatccc ctcccgaag gccccaaagg acccgaaggt gctggagaaa gcgaacgaag 11040
 agctgaatgg gctggcttac ctgctgaagg accacgatgt gatcgtcgt cgtcccgaga 11100
 agatcgactt caccaagagc ctgaaaacc cgtatttcga ggttgccaac cagtactgcg 11160
 gcgtttgtcc tcgcgacgtg atgatcacgt ttggcaacga aatcatggaa gcgacctgt 11220
 ccaaactgac acgtttcttc gaatacctcc cctatcggaa gctggtctac gactactgga 11280
 acaaggacga gcacatgac tggaacgcag ccccgaacc gacctgcag gatagcatgt 11340
 acctgaaaa ctctgggag ctctcgtggt aagaacgctt caagcggatg cactactcg 11400

aattctgcat cacccaagac gaggtgatct tcgatgccgc cgattgctcc cgcttgggta 11460
aggacatcct ggtgcaggaa agcatgacca ccaatcgac tggcatccgc tggctgaaga 11520
agcatctcga accacgcggc ttctcgtcc atccggtgca cttcccgttg gacttcttcc 11580
ctagccacat cgactgcacg ttcgtaccgt tgcgtccggg tctgatcctg accaatccgg 11640
aacgcccgat tcgcgaggaa gaggagaaga tcttcaagga gaatggctgg gagctgatca 11700
ccgtaccgea gcttacctgc tcgaacgacg agatgcccac gttctgccag agctcgaat 11760
ggctgtccat gaacgtcctg agcattagtc ccaccaaggt gatctgcgaa gaacgggaaa 11820

agccgctgca agaactgctg gacaagcacg ggttcgaagt ctttccttg ctttccgcc 11880
atgtgtttga gttcggtagc agctttcact gtgccacgtg ggatattcgc cgcaaggcgc 11940
agtgcgagga ctacctgccg aacctgaact accagccgat ttgcggctga cgcggatccc 12000
gacctgcata ccgttcgaga taaggtagtc actcatgagc gctaggcact ttctctcct 12060
gctggacttc accaccgacg aattgctcgg ggtgatccgc cacggcatcg agctgaagga 12120
cctgcgcaag cgaggcgtgc tgttcgaacc gctgaagaac cgtgtgctag gcatgatctt 12180
cgaaaagtcc tcgaccgta cccgtgtgtc gttcgaggcc ggcatgatcc agctcggcgg 12240

ccaggccatc ttctgtgcg cccgcgacac ccagctgggc cgcggcgcgac caattggtga 12300
cagcgccatc gtgctgtcga gcatggttga tgtggtgatg atccggacc acgcccacag 12360
cacctgacc gagttcgccg ccaagtgcg tgtgcccgtg atcaacggcc tgtcgcgaca 12420
atcgcacccg tgccaactgc tggccgacat gcagaccttc gttgaacacc gcggctcgat 12480
tcagggcaag accgtgacct ggatcggcga tggcttcaac atgtgcaact cctatatcga 12540
agcccaggcag cagttcgatt tccagctgcg catcgctgc cccgaaggct atgagccgga 12600
tcaacgcttc atggcactgg gcggcgaccg cgtgcagatc atccgggatg ccagggaagc 12660

tgtgctgat gcacacctgg tggtcaccga tgtctggact tccatgggtc aggaggagga 12720
aactgcacgg cgctggcgc atttcgcgc ttaccaggtc acccgcaac tgctcgacct 12780
ggctgcacc gatgtctct tcatgcaact cctgcccgcc caccgtggcg aggaaatcag 12840
ccaggacctg ctgcagacc cacgttcggt gcctggggac gaggtgaaa accgctgca 12900
tgcacagaag gccttctcg aattccttgt agaaccggct taccaccag catgagtcaa 12960
ccgtacaacc ccgtggagt atggcatggc ggacgtaaaa aaggtcgtac tggcgtattc 13020
cggcggcctt gatacttcgg tgattctcaa gtggctgcag gatactaca actgcgaagt 13080

ggtgaccttc accgctgacc tgggacaggc cgaagaggtc gaaccggccc gtccaaggc 13140
ccaggcaatg ggcgttaaag agatctacat cgacgacctg cgcaagaat tcgtgcgtga 13200
ttctgtgttc ccgatgttc gcgccaacac cgtctacgaa ggcgagtacc tgctgggtac 13260

ttccatcgcc cgtccgctga tcgccaagcg cctgatcgaa atcgccaacg aaaccggcgc 13320
 tgacgccatt tcccatggcg ccaccggcaa ggtaacgac caggtgctgc tcgagctggg 13380
 tgccatgccc ctgaagccag gcgtcaaggt catcgctcca tggcgcgagt gggacctgct 13440
 gtcccgcgaa aagctgatgg actacgccga gaagcacggc atcccgatcg agcgccacgg 13500

 caagaagaag tcgccgtact cgatggacgc caacctgctg cacatctcct acgagggcgg 13560
 tgtcctggaa gataacctga ccgagcacga agaagacatg tggcgcctga gtgtctcgcc 13620
 tgagaatgcc ccggaccagg ctacctacat cgagctgacc taccgcaatg gtgacatcgt 13680
 tgccatcgac ggcgtcgaga aatccccggc caccgtctcg gcagacctga accgtatcgg 13740
 tgggtccaac ggcatcgccc gtctggacat cgtcgaaaac cgttacgtcg gcatgaagtc 13800
 gcgcggttgc tacgaaacgc ctggcggtag catcatgctc aaggcacacc gtgcatcga 13860
 tgcgatcacc ctggaccgag aagtcgctca cctgaaagat gagctgatgc caaagtatgc 13920

 cagcctgac tacaccgct actggtggag cccggagcgt ctgatgctgc aacagatgat 13980
 cgatgcttcg caggtcaacg tgaatggtgt ggtgcgctg aaactgtaca agggcaacgt 14040
 gaccgtggtt ggccgcaagt cggacgattc gctgttcgat gccaacatcg ccaccttga 14100
 agaagatggt ggtgcctaca accagcgaga tctgtctggc ttcatcaagc tcaatgact 14160
 gcgtatgcgc attgccgcca acaagggcgg ttcgtgctc tgattgctat cgacgccact 14220
 ttttcgttca cgctgcaat gagtgaatcc atgagcaccg agaagaccaa tcagtctcgg 14280
 ggcggccgct tcagtgagcc cgtcgacgcc ttcgtgccc gtttcaccgc ctcggtagat 14340

 ttcgacaage gctgtaccg tcacgacatc atgggttcga ttgcccatgc caccatgctg 14400
 gcgcaggtcg gcgtgctcag tgatgccgag cgcgacacca tcatcgatgg cctgaaaacc 14460
 atccaggcgc agattgaagc cggcaacttc gactggcgtg tcgacctga agacgtgcac 14520
 atgaacatcg aagcacgct gaccgaccgc atcggcacca ccggcaagaa getgcatact 14580
 gggcgtagcc gcaacgacca ggtggccacc gacatccgcc tttggctgcg cgacgaaatc 14640
 gacctgatcc tgggcgaaat caccgcctg cagcagggcc tgctggagca ggcagagcgt 14700
 gaagccgaaa ceatcatgcc tggtttcacc cacctgcaga cggcgcagcc ggtcaccttt 14760

 gccaccacc tctggcgtg gttcgaaatg ctgagccgag actatgagcg cctggtcgac 14820
 tgccgcaagc gcaccaaccg catgccactg ggcagcgcg cgctggccgg caccacctac 14880
 ccgatcgacc gtgaactgac ctgcaagctg ctgggctttg aagccgtggc cggcaactcg 14940
 ctggatggcg tctcggaccg tgatttcgcc atcgaattct gcgccgctgc cagcgtggcg 15000
 atgatgcacc tttcgcgctt ctccgaagag ctggtgctgt ggaccagcgc gcagttccag 15060
 ttcacatgacc ttcggaccg cttctgcaact ggcagctcga tcatgccga gaaaaagaac 15120

ccggacgtgc cagagctggt acgtggcaag agcggccgcg tgttcggcgc cctgaccggc 15180

 ctgctgacce tgatgaaagg ccaaccgctg gcctacaaca aggacaacca ggaagacaag 15240
 gaaccgctgt tcgacgccgc cgataccctg cgcgactcgc tgcgggcctt cgctgacatg 15300
 atccccgcga tcaagcccaa gcacgccatc atgcgtgaag cggccctgcg cggtttctcc 15360
 accgtaccg acctggctga ctatctggtt cgcgctggcc tgccttccg tgactgccac 15420
 gagatcgttg gccacgcggt gaagtatggt gtggacactg gcaaggacct ggcccagatg 15480
 agcctggacg aactgcgcca attcagcgac cagatcgagc aggacgtggt tgcctgctg 15540
 acgtggaag gctcggtgaa tgcgctgac cacattggtg gtacggcgcc ggcgcaggtg 15600

 cgtgctgccg tcgttcgtgg caaggccctg ttggcgtctc gctaateccc caaggctcga 15660
 gcacgcgaga gtagggaact gccaggcatc aaataaacg aaaggctcag tcgaaagact 15720
 gggcctttcg ttttatctgt tgtttgtcgg tgaacgctct cctgagtagg acaaatccgc 15780
 cgggagcgga ttigaacgat gataagctgt caaacatgag aattctttaa gacgaaaggg 15840
 cctcgtgtgt acaaacgttc gtcaaaaggg cgacacaaaa ttcttcaggg ggccggccca 15900
 gcgccggcgg tcgagtggcg acggcgccgc ttgtccgcgc cctggtagat tgcctggcgc 15960
 taggccagcc atttttgagc ggccagcggc cgcgataggc cgacgcgaag cggcggggcg 16020

 tagggagcgc agcgaccgaa gggtaggcgc tttttgcagc tcttcggctg tgcgctggcc 16080
 agacagttat gcacaggcca ggcgggtttt aagagtttta ataagtttta aagagtttta 16140
 ggccgaaaaa tcgccttttt tctcttttat atcagtcact tacatgtgtg accggttccc 16200
 aatgtacggc ttgggttcc caatgtacgg gttccggttc ccaatgtacg gctttgggtt 16260
 cccaatgtac gtgctatcca caggaaagag accttttcga cctttttccc ctgctagggc 16320
 aatttgccct agcatctgct ccgtacatta ggaaccggcg gatgcttcgc cctcgatcag 16380
 gttgcggtag cgcgatgacta ggatcgggcc agcctgeccc gcctcctcct tcaaatcgta 16440

 ctccggcagg tcatttgacc cgatcagctt gcgcacggtg aaacagaact tcttgaactc 16500
 tccggcgtg ccaactgcgtt cgtagatcgt ctgaacaac catctggctt ctgccttgcc 16560
 tgcggcggc cgigccaggc ggtagagaaa acggccgatg ccgggatcga tcaaaaagta 16620
 atcgggtgta accgtcagca cgtccgggtt ctfgccttct gtgatctcgc ggtacatcca 16680
 atcaactagc tcgatctcga tgtaactcgg ccgcccgggt tcgctcttta cgatcttcta 16740
 gcggctaate aaggcttcac cctcggatac cgtcaccagg cggccgttct tggccttctt 16800
 cgtacgtgc atggcaactg gcgtggtgtt taaccgaatg caggtttcta ccaggtcgtc 16860

tttctgcttt ccgccatcgg ctcgccggca gaacttgagt acgtccgcaa cgtgtggacg 16920
 gaacacgcgg cgggcttgt ctcccttccc ttcccggat cggttcatgg attcggttag 16980
 atgggaaacc gccatcagta ccaggtcgtat atcccacaca ctggccatgc cggccggccc 17040
 tgcggaaacc tctacgtgcc cgtctggaag ctctagcgg atcacctcgc cagctcgtcg 17100
 gtcacgttc gacagacgga aaacggccac gtccatgatg ctgcgactat cgcgggtgcc 17160
 cacgtcatag agcatcggaa cgaaaaaatc tggttgctcg tcgcccttgg gcggcttct 17220
 aatcagcggc gcaccggctg ccggcgggtg ccgggattct ttgcggattc gatcagcggc 17280

 cgcttgccac gattaccgg ggcgtgcttc tgcctcgtat cgcttccgct gggcggcctg 17340
 cgcggccttc aacttctcca ccaggtcatt acccagcgc gcgccgattt gtaccgggcc 17400
 ggatggtttg cgaccgctca cgccgattcc tcgggcttgg gggttccagt gccattgcag 17460
 ggccgcgaga caaccagcc gcttacgctt ggccaaccgc ccgttctctc acacatgggg 17520
 cattccacgg cgtcgggtcc tggttgttct tgattttcca tgcgcctcc tttagccgct 17580
 aaaattcatt tactcattta ttcatattgct ctttactct ggtagctgcg cgatgtattc 17640
 agatagcagc tcgtaattgg tcttgccctg gcgtaccgcg tacatcttca gcttgggtgtg 17700

 atcctccgcc ggcaactgaa agttgaccgg ctcatggct ggctgtctg ccaggctggc 17760
 caacgttga gccttctgc tgcgtgcgt cggacggccg gcacttagcg tgtttgtgct 17820
 tttgctcatt ttctctttc ctcatctaac caaatgagtt ttgatttaac ttcagcggcc 17880
 agcgcctgga cctcgcgggc agcgtcgcct tcgggttctg attcaagaac ggttgtgccg 17940
 gcggcggcag tgctgggta gctcacgcgc tgcgtgatac gggactcaag aatgggcagc 18000
 tcgtaccgg ccagcgcctc ggcaacctca ccgccgatgc gcgtgccttt gatcgcgccg 18060
 gacacgaaa aggccgctt tagccttcca tccgtgacct caatgcgctg cttaccagc 18120

 tccaccaggt cggcgggtgc ccatatgtcg taagggttgg gctgcaccgg aatcagcaag 18180
 aagtcggctg ccttgatcgc ggacacagcc aagtcgccg cctggggcgc tccgtcgatc 18240
 actacgaagt cgcgccggcc gatggccttc acgtcgcggt caatcgtcgg gcggtcgatg 18300
 ccgacaacgg ttagcgggtg atcttccgc acggccgcc aatcgcgggc actgcctgg 18360
 ggatcggaat cgactaacag aacatcggcc ccggcgagtt gcaggcgcg ggctagatgg 18420
 gttgcatgg tcgtcttgc tgaccgcct tcttggttaa gtacagcgt aacctcatg 18480
 cgttccctt gcgtatttgt ttatttactic atcgcacat atacgcagcg accgcatgac 18540

 gcaagctgtt ttactcaat acacatcacc ttttagacg gcggcgtcgc gtttcttcag 18600
 cggccaagct ggccggccag gccgccagct tggcatcaga caaacggcc aggatttcat 18660
 gcagccgcac ggttccggat gagcattca 18689