

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2012-0099116 (43) 공개일자 2012년09월06일
(51) 국제특허분류(Int. Cl.) C12N 1/21 (2006.01) C12N 15/52 (2006.01) C12P 7/20 (2006.01) C12P 7/04 (2006.01) (21) 출원번호 10-2012-7017290 (22) 출원일자(국제) 2010년12월03일 심사청구일자 없음 (85) 번역문제출일자 2012년07월03일 (86) 국제출원번호 PCT/US2010/058832 (87) 국제공개번호 WO 2011/069033 국제공개일자 2011년06월09일 (30) 우선권주장 61/266,605 2009년12월04일 미국(US)		(71) 출원인 이 아이 듀폰 디 네모아 앤드 캠파니 미합중국 델라웨어주 (우편번호 19898) 월밍톤시 마아캣트 스트리트 1007 (72) 발명자 엘리엇, 앤드류, 씨. 미국 19806 델라웨어주 월밍톤 스타플러 플레이 스 1407 가텐비, 앤토니, 에이. 미국 19802 델라웨어주 월밍톤 베이나드 불러바 드 2309 반 디크, 티나, 케이. 미국 19803 델라웨어주 월밍톤 캠브리지 드라이 브 119 (74) 대리인 김영, 양영준, 양영환

전체 청구항 수 : 총 13 항

(54) 발명의 명칭 수크로스로부터 글리세롤 및 글리세롤-유래의 산물을 생산하기 위한 재조합 박테리아

(57) 요약

수크로스로부터 글리세롤 및 글리세롤-유래의 산물을 생산할 수 있는 재조합 박테리아가 기술된다. 재조합 박테리아는 그들의 게놈 내에 또는 적어도 하나의 재조합 작제물 상에, 수크로스 수송체 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열; 프럭토키나제 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열; 및 수크로스 하이드롤라제 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열을 포함한다. 이들 뉴클레오티드 서열은 동일하거나 상이한 프로모터에 각각 작동적으로 연결된다. 이들 재조합 박테리아는 수크로스를 대사시켜 글리세롤 및/또는 글리세롤-유래의 산물, 예를 들어 1,3-프로판다이올 및 3-하이드록시프로피온산을 생산할 수 있다.

특허청구의 범위

청구항 1

(a) 수크로스 수송체(sucrose transporter) 활성을 갖는 폴리펩티드 또는 폴리펩티드 복합체를 암호화하는 하나 이상의 뉴클레오티드 서열;

(b) 프럭토키나제(fructokinase) 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열; 및

(c) 수크로스 하이드롤라제(sucrose hydrolase) 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열(여기서 (a), (b) 및 (c)는 동일하거나 상이한 프로모터에 각각 작동적으로 연결됨)을 그의 게놈 내에 또는 적어도 하나의 재조합 작제물 상에 포함하며,

수크로스를 대사시켜 글리세롤, 1,3-프로판다이올 및 3-하이드록시프로피온산으로 구성된 군으로부터 선택된 산물을 생산할 수 있는 재조합 박테리아.

청구항 2

제1항에 있어서, 수크로스 수송체 활성을 갖는 폴리펩티드가, 서열 번호 24, 서열 번호 26, 또는 서열 번호 28에 제시된 아미노산 서열에 비교할 때 클러스탈 V(Clustal V) 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 재조합 박테리아.

청구항 3

제1항에 있어서, 수크로스 수송체 활성을 갖는 폴리펩티드 복합체가,

a) 서열 번호 30에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제1 서브유닛;

b) 서열 번호 32에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제2 서브유닛; 및

c) 서열 번호 34에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제3 서브유닛을 포함하는 재조합 박테리아.

청구항 4

제1항에 있어서, 수크로스 수송체 활성을 갖는 폴리펩티드 복합체가,

a) 서열 번호 36에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제1 서브유닛;

b) 서열 번호 38에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제2 서브유닛;

c) 서열 번호 40에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제3 서브유닛; 및

d) 서열 번호 42에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제4 서브유닛을 포함하는 재조합 박테리아.

청구항 5

제1항에 있어서, 프럭토키나제 활성을 갖는 폴리펩티드가, 서열 번호 44, 서열 번호 46, 서열 번호 48, 서열 번호 50, 서열 번호 52, 서열 번호 54, 서열 번호 85, 또는 서열 번호 87에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 재조합 박테리아.

청구항 6

제1항에 있어서, 수크로스 하이드롤라제 활성을 갖는 폴리펩티드가, 서열 번호 56, 서열 번호 58, 서열 번호 60, 서열 번호 62, 서열 번호 64, 서열 번호 66, 또는 서열 번호 68에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 재조합 박테리아.

청구항 7

제1항에 있어서, 수크로스 수송체 활성을 갖는 폴리펩티드가, 서열 번호 26에 제시된 서열에 실질적으로 상응하는 재조합 박테리아.

청구항 8

제1항에 있어서, 프럭토키나제 활성을 갖는 폴리펩티드가, 서열 번호 48에 제시된 서열에 실질적으로 상응하는 재조합 박테리아.

청구항 9

제1항에 있어서, 수크로스 하이드롤라제 활성을 갖는 폴리펩티드가, 서열 번호 58에 제시된 서열에 실질적으로 상응하는 재조합 박테리아.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서, 에스케리키아(Escherichia), 클렙시엘라(Klebsiella), 시트로박터(Citrobacter), 및 아에로박터(Aerobacter) 속으로 구성된 군으로부터 선택되는 재조합 박테리아.

청구항 11

제10항에 있어서, 에스케리키아 콜라이(Escherichia coli)인 재조합 박테리아.

청구항 12

- 수크로스의 존재 하에 제1항 내지 제9항 중 어느 한 항의 재조합 박테리아를 배양하는 단계; 및
- 임의로, 생산된 글리세롤, 1,3-프로판다이올 및/또는 3-하이드록시프로피온산을 회수하는 단계를 포함하는, 수크로스로부터 글리세롤, 1,3-프로판다이올 및/또는 3-하이드록시프로피온산을 제조하는 방법.

청구항 13

- 수크로스의 존재 하에 제10항의 재조합 박테리아를 배양하는 단계; 및
- 임의로, 생산된 글리세롤, 1,3-프로판다이올 및/또는 3-하이드록시프로피온산을 회수하는 단계를 포함하는, 수크로스로부터 글리세롤, 1,3-프로판다이올 및/또는 3-하이드록시프로피온산을 제조하는 방법.

명세서

기술 분야

[0001] 본 발명은 미생물학 및 분자생물학 분야에 관한 것이다. 더욱 구체적으로, 수크로스를 탄소 공급원으로 사용하여 글리세롤 및 글리세롤-유래의 산물을 생산하는 능력을 갖는 재조합 박테리아 및 이러한 재조합 박테리아를 이용하는 방법이 제공된다.

배경 기술

[0002] 다수의 상업적으로 유용한 미생물이 그들의 탄수화물 주 공급원으로서 글루코스를 사용한다. 그러나, 상업적으로 바람직한 산물의 생산을 위해 개발된 미생물에 의한 글루코스의 사용의 단점은 글루코스의 높은 가격이다. 미생물 생산 시스템에 있어서 탄수화물 공급원으로서 수크로스 및 다른 당을 함유하는 혼합 공급원료 및 수크로스를 사용하는 것은, 이들 재료가 더 낮은 가격으로 용이하게 입수가가능하기 때문에 상업적으로 더욱 바람직할 것이다.

[0003] 생산 미생물은, 그것이 혼합 공급원료 내에 존재하는 임의의 수크로스를 이용할 수 있을 경우에 더욱 효율적으로 기능할 수 있다. 그러므로, 생산 미생물이 주요 탄소 공급원으로서 수크로스를 효율적으로 이용하는 능력을 갖지 않는 경우, 그것은 효율적으로 작동할 수 없다. 예를 들어, 전형적으로 박테리아 세포는 당을 우선적으로 사용하는 것으로 나타나며, 글루코스를 가장 선호한다. 당의 혼합물을 함유하는 인공 배지에서, 전형적으로 다른 당에 앞서 글루코스가 그의 전부까지 대사된다. 더욱이, 다수의 박테리아는 수크로스를 이용하는 능력이 결여되어 있다. 예를 들어, 에스케리키아 콜라이(Escherichia coli) 변종의 50% 미만이 수크로

스를 이용하는 능력을 가지고 있다. 따라서, 생산 미생물이 탄수화물 공급원으로서 수크로스를 이용할 수 없는 경우, 미생물을 조작하여 그것이 수크로스를 이용할 수 있도록 하는 것이 바람직하다.

[0004] 수크로스 이용 유전자를 혼입함으로써 수크로스를 이용하도록 조작된 재조합 박테리아가 보고된 바 있다. 예를 들어, 미국 특허 제6,960,455호(Livshits et al.)에는 수크로스 이용을 위한 대사 경로를 암호화하는 유전자를 함유하는 에스케리키아 콜라이 변종을 사용하는 아미노산의 생산이 기술되어 있다. 부가적으로, 문헌 [Olson et al., Appl. Microbiol. Biotechnol. 74:1031-1040, 2007]에는 수크로스 분해를 담당하는 유전자를 지닌 에스케리키아 콜라이 변종이 기술되어 있으며, 이는 탄소 공급원으로서 수크로스를 사용하여 L-타이로신 또는 L-페닐알라닌을 생산한다. 그러나, 탄소 공급원으로서 수크로스를 사용하여 글리세롤 및 글리세롤-유래의 산물을 생산할 수 있는 박테리아 변종에 대한 필요성이 존재한다.

발명의 내용

[0005] 일 실시 양태에서 본 발명은, 그의 게놈 내에 또는 적어도 하나의 재조합 작제물 상에 하기의 것을 포함하는 재조합 박테리아를 제공한다:

[0006] (a) 수크로스 수송체(sucrose transporter) 활성을 갖는 폴리펩티드 또는 폴리펩티드 복합체를 암호화하는 하나 이상의 뉴클레오티드 서열;

[0007] (b) 프럭토키나제(fructokinase) 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열; 및

[0008] (c) 수크로스 하이드롤라제(sucrose hydrolase) 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열

[0009] (여기서, (a), (b) 및 (c)는 동일하거나 상이한 프로모터에 각각 작동적으로 연결되며, 여기서 추가로, 상기 재조합 박테리아는 수크로스를 대사시켜 글리세롤, 1,3-프로판다이올 및 3-하이드록시프로피온산으로 구성된 군으로부터 선택된 산물을 생산할 수 있음).

[0010] 제2 실시 양태에서 본 발명은,

[0011] a) 본 명세서에 개시된 재조합 박테리아를 수크로스의 존재 하에 배양하는 단계; 및

[0012] b) 임의로, 생산된 글리세롤, 1,3-프로판다이올 및/또는 3-하이드록시프로피온산을 회수하는 단계를 포함하는, 수크로스로부터 글리세롤, 1,3-프로판다이올 및/또는 3-하이드록시프로피온산을 제조하는 방법을 제공한다.

[0013] 서열의 간단한 설명

[0014] 하기의 서열은 37 C.F.R. 1.821 1.825("뉴클레오티드 서열 및/또는 아미노산 서열 개시를 포함하는 특허 출원에 관한 요건 - 서열 규정")를 따르며 세계지적재산권기구(WIPO: World Intellectual Property Organization) 표준 ST.25(1998), 및 EPO 및 PCT의 서열 목록 요건(규정 5.2 및 49.5(a bis), 및 시행 세칙의 섹션 208 및 부칙 C)에 부합한다. 뉴클레오티드 및 아미노산 서열 데이터에 사용된 기호 및 체제는 37 C.F.R. § 1.822에 제시된 규정을 준수한다.

[0015] [표 A]

유전자 및 단백질 서열 번호의 요약

유전자	코딩 서열 서열 번호	암호화된 단백질 서열 번호
사카로마이세스 세레비시아(<i>Saccharomyces cerevisiae</i>)로부터의 <i>GPD1</i>	1	2
사카로마이세스 세레비시아로부터의 <i>GPD2</i>	3	4
사카로마이세스 세레비시아로부터의 <i>GPP1</i>	5	6
사카로마이세스 세레비시아로부터의 <i>GPP2</i>	7	8
클렙시엘라 뉴모니아(<i>Klebsiella pneumoniae</i>)로부터의 <i>dhaB1</i>	9	10
클렙시엘라 뉴모니아로부터의 <i>dhaB2</i>	11	12
클렙시엘라 뉴모니아로부터의 <i>dhaB3</i>	13	14
에스케리키아 콜라이로부터의 <i>aldB</i>	15	16
에스케리키아 콜라이로부터의 <i>aldA</i>	17	18
에스케리키아 콜라이로부터의 <i>aldH</i>	19	20
에스케리키아 콜라이로부터의 <i>galP</i>	21	22
에스케리키아 콜라이 EC3132로부터의 <i>cscB</i>	23	24
에스케리키아 콜라이 ATCC13281로부터의 <i>cscB</i>	25	26
비피도박테리움 락티스(<i>Bifidobacterium lactis</i>)로부터의 <i>cscB</i>	27	28
스트렙토코커스 뉴모니아(<i>Streptococcus pneumoniae</i>) 변종 TIGR4로부터의 <i>susT1</i>	29	30
스트렙토코커스 뉴모니아 변종 TIGR4로부터의 <i>susT2</i>	31	32
스트렙토코커스 뉴모니아 변종 TIGR4로부터의 <i>susX</i>	33	34
스트렙토코커스 뮤탄스(<i>Streptococcus mutans</i>)로부터의 <i>malE</i>	35	36
스트렙토코커스 뮤탄스로부터의 <i>malF</i>	37	38
스트렙토코커스 뮤탄스로부터의 <i>malG</i>	39	40
스트렙토코커스 뮤탄스로부터의 <i>malK</i>	41	42
아그로박테리움 튤메파시엔스(<i>Agrobacterium tumefaciens</i>)로부터의 <i>scrK</i>	43	44
스트렙토코커스 뮤탄스로부터의	45	46

[0016]

scrK		
에스케리키아 콜라이로부터의 scrK	84	85
클렙시엘라 뉴모니아로부터의 scrK	86	87
에스케리키아 콜라이로부터의 cscK	47	48
엔테로코커스 파칼리스(Enterococcus faecalis)로부터의 cscK	49	50
사카로마이세스 세레비시아로부터의 HXK1	51	52
사카로마이세스 세레비시아로부터의 HXK2	53	54
에스케리키아 콜라이 EC3132로부터의 cscA	55	56
에스케리키아 콜라이 ATCC13281로부터의 cscA	57	58
비피도박테리움 락티스 변종 DSM 10140 ^T 로부터의 bfrA	59	60
사카로마이세스 세레비시아로부터의 SUC2	61	62
코리네박테리움 글루타미쿰(Corynebacterium glutamicum)으로부터의 scrB	63	64
류코노스톡 메센테로이데스(Leuconostoc mesenteroides) DSM 20193 으로부터의 수크로스 포스포릴라제(sucrose phosphorylase) 유전자	65	66
비피도박테리움 아돌레센티스(Bifidobacterium adolescentis) DSM 20083 으로부터의 sucP	67	68
클렙시엘라 뉴모니아로부터의 dhaT	69	70

[0017]

[0018] 서열 번호 71은 클렙시엘라 뉴모니아로부터의 dhaX 유전자의 코딩 영역의 뉴클레오티드 서열이다.

[0019] 서열 번호 72는 플라스미드 pSYCO101의 뉴클레오티드 서열이다.

[0020] 서열 번호 73은 플라스미드 pSYCO103의 뉴클레오티드 서열이다.

[0021] 서열 번호 74는 플라스미드 pSYCO106의 뉴클레오티드 서열이다.

[0022] 서열 번호 75는 플라스미드 pSYCO109의 뉴클레오티드 서열이다.

[0023] 서열 번호 76은 플라스미드 pSYCO400/AGRO의 뉴클레오티드 서열이다.

[0024] 서열 번호 77은 본 명세서의 실시예 1에 기술된 플라스미드 pScr1의 뉴클레오티드 서열이다.

[0025] 서열 번호 78은 본 명세서의 실시예 1에 기술된 플라스미드 pBHRcscBKA의 뉴클레오티드 서열이다.

[0026] 서열 번호 79는 본 명세서의 실시예 1에 기술된 플라스미드 pBHRcscBKAmutB의 뉴클레오티드 서열이다.

[0027] 서열 번호 80 내지 83은 본 명세서의 실시예 2 내지 4에 기술된 변종 TTab를 작제하기 위해 사용된 프라이머의 뉴클레오티드 서열이다.

발명을 실시하기 위한 구체적인 내용

[0028] 본 명세서에 제시된 각각의 참고 문헌의 개시는 그 전체 내용이 본 명세서에 참고로 포함된다.

[0029] 본 명세서 및 첨부된 특허청구범위에 사용되는 바와 같이, 문맥이 명백하게 달리 지시하지 않는 한, 단수 형태("a", "an", 및 "the")는 복수 언급(plural reference)을 포함한다. 따라서, 예를 들어, "세포"에 대한 언급은 하나 이상의 세포 및 당업자에게 공지된 그의 균등물 등을 포함한다.

[0030] 본 개시와 관련하여, 다수의 용어 및 약어가 사용된다. 하기의 정의가 제공된다.

[0031] "열린 해독틀"은 "ORF(Open reading frame)"로 축약한다.

[0032] "폴리머라제 연쇄 반응"은 "PCR(Polymerase chain reaction)"로 축약한다.

[0033] "미국 미생물 보존센터"는 "ATCC(American Type Culture Collection)"로 축약한다.

[0034] 용어 "재조합 글리세롤-생산 박테리아"는 1,3-프로판다이올 및 3-하이드록시프로피온산과 같은 글리세롤 및/

또는 글리세롤-유래의 산물을 생산할 수 있도록 유전적으로 조작된 박테리아를 지칭한다.

- [0035] 용어 "수크로스 수송체 활성을 갖는 폴리펩티드 또는 폴리펩티드 복합체"는 미생물 세포 내로의 수크로스 수송을 매개할 수 있는 폴리펩티드 또는 폴리펩티드 복합체를 지칭한다. 수크로스 수송체 활성을 갖는 폴리펩티드의 예는 수크로스:H⁺ 공수송체를 포함하나, 이에 한정되지 않는다. 수크로스 수송체 활성을 갖는 폴리펩티드 복합체의 예는 ABC-유형 수송체를 포함하나, 이에 한정되지 않는다. 수크로스:H⁺ 공수송체는, 예를 들어, E. 콜라이 변종, 예를 들어 EC3132(Jahreis et al., J. Bacteriol. 184:5307-5316, 2002) 또는 ATCC13281(Olson et al., Appl. Microbiol. Biotechnol. 74:1031-1040, 2007), 및 비피도박테리움 락티스 변종 DSM 10140^T(Ehrmann et al., Curr. Microbiol. 46(6):391-397, 2003)에서 발견되는 cscB 유전자에 의해 암호화된다. 수크로스에 대한 활성을 가진 ABC-유형 수송체의 일례는 스트렙토코커스 뉴모니아 변종 TIGR4의 유전자 susT1, susT2 및 susX에 의해 암호화되는 복합체이다(Iyer and Camilli, Molecular Microbiology 66:1-13, 2007). 수크로스 수송체 활성을 갖는 폴리펩티드 또는 폴리펩티드 복합체는 또한 다른 당류에 대한 활성을 가질 수 있다. 일례는 malEFGK에 의해 암호화되는 스트렙토코커스 뮤탄스의 말토스 수송체 복합체이다(Kilic et al., FEMS Microbiol Lett. 266:218, 2007).
- [0036] 용어 "프럭토키나제 활성을 갖는 폴리펩티드"는 D-프럭토스 + ATP를 프럭토스-포스페이트 + ADP로 전환하는 촉매 작용을 하는 능력을 갖는 폴리펩티드를 지칭한다. 전형적인 프럭토키나제는 EC 2.7.1.4이다. 프럭토스를 인산화하는 일부 능력을 갖는 효소는, 이 활성이 그들의 주된 활성인지 여부를 불문하고, 프럭토키나제로 지칭될 수 있다. 프럭토키나제 활성을 갖는 단백질 및 프럭토키나제를 암호화하는 유전자에 사용되는 약어는, 예를 들어, "Frk", "scrK", "cscK", "FK", 및 "KHK"를 포함한다. 프럭토키나제는 아그로박테리움 튜메파시엔스 및 스트렙토코커스 뮤탄스의 scrK 유전자; 및 소정의 에스케리키아 콜라이 변종의 cscK 유전자에 의해 암호화된다.
- [0037] 용어 "수크로스 하이드롤라제 활성을 갖는 폴리펩티드"는 수크로스를 가수분해하여 글루코스 및 프럭토스를 생산하는 촉매 작용을 하는 능력을 갖는 폴리펩티드를 지칭한다. 이러한 폴리펩티드는 흔히 "인버타제(invertase)" 또는 " β -프럭토포라노시다제"로 지칭된다. 전형적인 이들 효소는 EC 3.2.1.26이다. 수크로스 하이드롤라제 활성을 갖는 폴리펩티드를 암호화하는 유전자의 예는 E. 콜라이 변종 EC3132(Jahreis et al. supra) 또는 ATCC13281(Olson et al., supra)에서 발견되는 cscA 유전자, 비피도박테리움 락티스 변종 DSM 10140^T로부터의 bfrA 유전자, 및 사카로마이세스 세레비시아로부터의 SUC2 유전자(Carlson and Botstein, Cell 28:145, 1982)이다. 수크로스 하이드롤라제 활성을 갖는 폴리펩티드는 또한 수크로스 포스페이트 하이드롤라제 활성을 가질 수 있다. 이러한 펩티드의 일례는 코리넨박테리움 글루타미쿰의 scrB에 의해 암호화된다(Engels et al., FEMS Microbiol Lett. 289:80-89, 2008). 수크로스 하이드롤라제 활성을 갖는 폴리펩티드는 또한 수크로스 포스포릴라제 활성을 가질 수 있다. 전형적인 이러한 효소는 EC 2.4.1.7이다. 수크로스 하이드롤라제 활성을 갖는 수크로스 포스포릴라제를 암호화하는 유전자의 예는 특히 류코노스톡 메센테로이데스 DSM 20193(Goedl et al., Journal of Biotechnology 129:77-86, 2007) 및 비피도박테리움 아돌레센티스 DSM 20083(van den Broek et al., Appl. Microbiol. Biotechnol. 65:219-227, 2004)에서 발견된다.
- [0038] 용어 "글리세롤 유도체" 및 "글리세롤-유래의 산물"은 본 명세서에서 호환적으로 사용되며, 글리세롤로부터, 또는 글리세롤을 포함하는 경로에서 합성되는 화합물을 지칭한다. 이러한 산물의 예는 3-하이드록시프로피온산, 메틸글리옥살, 1,2-프로판다이올, 및 1,3-프로판다이올을 포함한다.
- [0039] 용어 "미생물 산물"은 미생물에 의해 생산되는 산물, 즉, 미생물이 물질을 대사시킨 결과를 지칭한다. 미생물에 의해 산물이 자연적으로 생산될 수 있거나, 산물을 생산하도록 미생물을 유전적으로 조작할 수 있다. 용어 "포스포에놀피루베이트-당 포스포트랜스페라제 시스템", "PTS 시스템", 및 "PTS"는 본 명세서에서 호환적으로 사용되며, 포스포에놀피루베이트-의존성 당 흡수 시스템을 지칭한다.
- [0040] 용어 "포스포캐리어 단백질 HPr" 및 "PtsH"는 E. 콜라이에서 ptsH에 의해 암호화되는 포스포캐리어 단백질을 지칭한다.
- [0041] 용어 "포스포에놀피루베이트-단백질 포스포트랜스페라제" 및 "PtsI"은 E. 콜라이에서 ptsI에 의해 암호화되는 포스포트랜스페라제, EC 2.7.3.9를 지칭한다. 용어 "글루코스-특이적 IIA 컴포넌트", 및 "Crr"은 E. 콜라이에서 crr에 의해 암호화되는, EC 2.7.1.69로 표기되는 효소를 지칭한다. PtsH, PtsI, 및 Crr은 PTS 시스템을 포함한다.
- [0042] 용어 "PTS 음성"은, 그의 천연 상태에서 PTS 시스템을 함유하지 않는 미생물, 또는 PTS 유전자의 불활성화를

통해 PTS 시스템이 불활성화된 미생물을 지칭한다.

- [0043] 용어 "글리세롤-3-포스페이트 데하이드로게나제(glycerol-3-phosphate dehydrogenase)" 및 "G3PDH"는, 다이하이드록시아세톤 포스페이트(DHAP)를 글리세롤 3-포스페이트(G3P)로 전환하는 촉매 작용을 하는 효소 활성을 담당하는 폴리펩티드를 지칭한다. 생체내 G3PDH는 NAD- 또는 NADP-의존성일 수 있다. 보조인자 특이적 글리세롤-3-포스페이트 데하이드로게나제를 특이적으로 지칭할 경우, 용어 "NAD-의존성 글리세롤-3-포스페이트 데하이드로게나제" 및 "NADP-의존성 글리세롤-3-포스페이트 데하이드로게나제"가 사용될 것이다. 일반적으로 NAD-의존성 및 NADP-의존성 글리세롤-3-포스페이트 데하이드로게나제는 NAD 및 NADP를 호환적으로 사용할 수 있는 것이 사실이므로(예를 들어 *gpsA*에 의해 암호화되는 효소에 의해), 용어 NAD-의존성 및 NADP-의존성 글리세롤-3-포스페이트 데하이드로게나제는 호환적으로 사용될 것이다. NAD-의존성 효소(EC 1.1.1.8)는, 예를 들어, 본 명세서에서 DAR1(서열 번호 1에 제시된 코딩 서열; 서열 번호 2에 제시된 암호화된 단백질 서열)로도 지칭되는 GPD1, 또는 GPD2(서열 번호 3에 제시된 코딩 서열; 서열 번호 4에 제시된 암호화된 단백질 서열), 또는 GPD3을 포함하는 몇 개의 유전자에 의해 암호화된다. NADP-의존성 효소(EC 1.1.1.94)는, 예를 들어 *gpsA*에 의해 암호화된다.
- [0044] 용어 "글리세롤 3-포스파타제(glycerol 3-phosphatase)", "sn-글리세롤 3-포스파타제", "d,l-글리세롤 포스파타제", 및 "G3P 포스파타제"는 글리세롤 3-포스페이트 및 물을 글리세롤 및 무기 포스페이트로 전환하는 촉매 작용을 할 수 있는 효소 활성을 갖는 폴리펩티드를 지칭한다. G3P 포스파타제는, 예를 들어, GPP1(서열 번호 5에 제시된 코딩 서열; 서열 번호 6에 제시된 암호화된 단백질 서열), 또는 GPP2(서열 번호 7에 제시된 코딩 서열; 서열 번호 8에 제시된 암호화된 단백질 서열)에 의해 암호화된다.
- [0045] 용어 "글리세롤 데하이드라타제" 또는 "데하이드라타제 효소"는 글리세롤 분자를 산물 3-하이드록시프로피온알데히드(3-HPA)로 전환하는 촉매 작용을 할 수 있는 효소 활성을 갖는 폴리펩티드를 지칭한다.
- [0046] 본 발명의 목적상, 데하이드라타제 효소는 글리세롤 데하이드라타제(E.C. 4.2.1.30) 및 다이올 데하이드라타제(E.C. 4.2.1.28)를 포함하며, 이들은 각각 글리세롤 및 1,2-프로판다이올의 선호되는 기질을 갖는다. 데하이드라타제 효소를 위한 유전자는, 특히 클렙시엘라 뉴모니아, 시트로박터 프론디(*Citrobacter freundii*), 클로스트리디움 파스퇴리아눔(*Clostridium pasteurianum*), 살모넬라 티피뮤리움(*Salmonella typhimurium*), 클렙시엘라 옥시토키아(*Klebsiella oxytoca*), 및 락토바실러스 루테리(*Lactobacillus reuteri*)에서 동정되었다. 각각의 경우에, 데하이드라타제는 하기 3개의 서브유닛으로 구성된다: 대형 또는 "α" 서브유닛, 중간 또는 "β" 서브유닛, 및 소형 또는 "γ" 서브유닛. 유전자는 또한, 예를 들어 문헌[Daniel et al., FEMS Microbiol. Rev. 22, 553 (1999)] 및 문헌[Toraya and Mori, J. Biol. Chem. 274, 3372 (1999)]에 기술되어 있다. 글리세롤 데하이드라타제의 대형 또는 "α" (알파) 서브유닛을 암호화하는 유전자는 *dhaB1*(서열 번호 9에 제시된 코딩 서열, 서열 번호 10에 제시된 암호화된 단백질 서열), *gldA* 및 *dhaB*를 포함하고; 중간 또는 "β" (베타) 서브유닛을 암호화하는 유전자는 *dhaB2*(서열 번호 11에 제시된 코딩 서열, 서열 번호 12에 제시된 암호화된 단백질 서열), *gldB*, 및 *dhaC*를 포함하며; 소형 또는 "γ" (감마) 서브유닛을 암호화하는 유전자는 *dhaB3*(서열 번호 13에 제시된 코딩 서열, 서열 번호 14에 제시된 암호화된 단백질 서열), *gldC*, 및 *dhaE*를 포함한다. 다이올 데하이드라타제의 대형 또는 "α" 서브유닛을 암호화하는 다른 유전자는 *pduC* 및 *pddA*를 포함하고; 중간 또는 "β" 서브유닛을 암호화하는 다른 유전자는 *pduD* 및 *pddB*를 포함하며; 소형 또는 "γ" 서브유닛을 암호화하는 다른 유전자는 *pduE* 및 *pddC*를 포함한다.
- [0047] 글리세롤 및 다이올 데하이드라타제는 글리세롤 및 일부 다른 기질에 의해 메커니즘-기초의 자멸적 불활성화(suicide inactivation)를 겪는다(Daniel et al., FEMS Microbiol. Rev. 22, 553 (1999)). 용어 "데하이드라타제 재활성화 인자"는 데하이드라타제 활성의 재활성화를 담당하는 단백질을 지칭한다. 용어 "데하이드라타제 재활성화 활성", "데하이드라타제 활성의 재활성화" 및 "데하이드라타제 활성의 재생"은 호환적으로 사용되며, 반응의 촉매가 불가능한 데하이드라타제를 반응의 촉매가 가능한 것으로 전환하는 현상 또는 데하이드라타제의 불활성화를 저해하는 현상 또는 생체내 데하이드라타제 효소의 가용 반감기를 연장하는 현상을 지칭한다. 2개의 단백질이 데하이드라타제 재활성화 인자로서 참여하는 것으로 동정되었다(예를 들어, 미국 특허 제6,013,494호 및 그 안의 참고 문헌; 문헌[Daniel et al., supra]; 문헌[Toraya and Mori, J. Biol. Chem. 274, 3372 (1999)]; 및 문헌[Tobimatsu et al., J. Bacteriol. 181, 4110 (1999)] 참조). 단백질 중 하나를 암호화하는 유전자는, 예를 들어, *orfZ*, *dhaB4*, *gdrA*, *pduG* 및 *ddrA*를 포함한다. 2개의 단백질 중 두번째 것을 암호화하는 유전자는, 예를 들어, *orfX*, *orf2b*, *gdrB*, *pduH* 및 *ddrB*를 포함한다.
- [0048] 용어 "1,3-프로판다이올 옥시도리덕타제(1,3-propanediol oxidoreductase)", "1,3-프로판다이올 데하이드로게나제" 및 "DhaT"는 본 명세서에서 호환적으로 사용되고, 3-HPA 및 1,3-프로판다이올의 상호 전환을 촉매 작용

할 수 있는 효소 활성을 갖는 폴리펩티드(들)를 지칭하며, 다만 이러한 활성을 암호화하는 유전자(들)는 그의 자연적(즉, 야생형) 설정에서 데하이드라타제 효소에 물리적으로, 또는 전사 단계에서 연결된 것으로 발견된다(예를 들어, 클렙시엘라 뉴모니아로부터의 dhaT의 경우와 같이 유전자가 dha 레귤론 내에서 발견됨). 1,3-프로판다이올 옥시도리덕타제를 암호화하는 유전자는 클렙시엘라 뉴모니아, 시트로박터 프론디, 및 클로스트리디움 파스티리아눔으로부터의 dhaT를 포함하나, 이에 한정되지 않는다. 이들 각각의 유전자는 유형 III 알코올 데하이드로게나제의 계열에 속하는 폴리펩티드를 암호화하며, 이는 보존된 철-결합 모티프를 나타내고, 3-HPA 및 1,3-프로판다이올의 NAD⁺/NADH 연결된 상호 전환에 대한 선호성을 갖는다(Johnson and Lin, J. Bacteriol. 169, 2050 (1987); Daniel et al., J. Bacteriol. 177, 2151 (1995); and Leurs et al., FEMS Microbiol. Lett. 154, 337 (1997)). 유사한 물리적 특성을 가진 효소들이 락토바실러스 브레비스(Lactobacillus brevis) 및 락토바실러스 부크네리(Lactobacillus buchneri)로부터 단리된 바 있다(Veiga da Dunha and Foster, Appl. Environ. Microbiol. 58, 2005 (1992)).

[0049] 용어 "dha 레귤론"은, 데하이드라타제 활성, 재활성화 활성, 및 1,3-프로판다이올 옥시도리덕타제를 포함하나 이에 한정되지 않는 다양한 생물학적 활성을 갖는 폴리펩티드를 암호화하는 연계된 폴리뉴클레오티드 또는 열린 해독틀의 세트를 지칭한다. 전형적으로 dha 레귤론은, 미국 특허 제7,371,558호에 기술된 바와 같이 열린 해독틀 dhaR, orfY, dhaT, orfX, orfW, dhaB1, dhaB2, dhaB3, 및 orfZ를 포함한다.

[0050] 용어 "알데히드 데하이드로게나제" 및 "Ald"는 알데히드를 카르복실산으로 전환하는 촉매 작용을 하는 폴리펩티드를 지칭한다. 알데히드 데하이드로게나제는 NAD, NADP, FAD, 또는 PQQ와 같은 산화환원 보조인자를 사용할 수 있다. 전형적인 알데히드 데하이드로게나제는 EC 1.2.1.3(NAD-의존성); EC 1.2.1.4(NADP-의존성); EC 1.2.99.3(PQQ-의존성); 또는 EC 1.2.99.7(FAD-의존성)이다. NADP-의존성 알데히드 데하이드로게나제의 일례는, E. 콜라이 유전자 aldB(서열 번호 15에 제시된 코딩 서열)에 의해 암호화된 AldB(서열 번호 16)이다. NAD-의존성 알데히드 데하이드로게나제의 예는, E. 콜라이 유전자 aldA(서열 번호 17에 제시된 코딩 서열)에 의해 암호화된 AldA(서열 번호 18); 및 E. 콜라이 유전자 aldH(서열 번호 19에 제시된 코딩 서열)에 의해 암호화된 AldH(서열 번호 20)를 포함한다.

[0051] 용어 "글루코키나제(glucokinase)" 및 "G1k"는 본 명세서에서 호환적으로 사용되며, D-글루코스 + ATP를 글루코스 6-포스페이트 + ADP로 전환하는 촉매 작용을 하는 단백질을 지칭한다. 전형적인 글루코키나제는 EC 2.7.1.2이다. 글루코키나제는 E. 콜라이에서 g1k에 의해 암호화된다.

[0052] 용어 "포스포에놀피루베이트 카르복실라제(phosphoenolpyruvate carboxylase)" 및 "Ppc"는 본 명세서에서 호환적으로 사용되며, 포스포에놀피루베이트 + H₂O + CO₂를 포스페이트 + 옥살로아세트산으로 전환하는 촉매 작용을 하는 단백질을 지칭한다. 전형적인 포스포에놀피루베이트 카르복실라제는 EC 4.1.1.31이다. 포스포에놀피루베이트 카르복실라제는 E. 콜라이에서 ppc에 의해 암호화된다.

[0053] 용어 "글리세르알데히드-3-포스페이트 데하이드로게나제" 및 "GapA"는 본 명세서에서 호환적으로 사용되며, 글리세르알데히드 3-포스페이트 + 포스페이트 + NAD⁺를 3-포스포-D-글리세로일-포스페이트 + NADH + H⁺로 전환하는 촉매 작용을 할 수 있는 효소 활성을 갖는 단백질을 지칭한다. 전형적인 글리세르알데히드-3-포스페이트 데하이드로게나제는 EC 1.2.1.12이다. 글리세르알데히드-3-포스페이트 데하이드로게나제는 E. 콜라이에서 gapA에 의해 암호화된다.

[0054] 용어 "유기 호흡 제어 단백질" 및 "ArcA"는 본 명세서에서 호환적으로 사용되며, 총체적 조절 단백질(global regulatory protein)을 지칭한다. 유기 호흡 제어 단백질은 E. 콜라이에서 arcA에 의해 암호화된다.

[0055] 용어 "메틸글리옥살 신타제(methylglyoxal synthase)" 및 "MgsA"는 본 명세서에서 호환적으로 사용되며, 다이하이드록시아세톤 포스페이트를 메틸글리옥살 + 포스페이트로 전환하는 촉매 작용을 할 수 있는 효소 활성을 갖는 단백질을 지칭한다. 전형적인 메틸글리옥살 신타제는 EC 4.2.3.3이다. 메틸글리옥살 신타제는 E. 콜라이에서 mgsA에 의해 암호화된다.

[0056] 용어 "포스포글루코네이트 데하이드라타제" 및 "Edd"는 본 명세서에서 호환적으로 사용되며, 6-포스포-글루코네이트를 2-케토-3-데옥시-6-포스포-글루코네이트 + H₂O로 전환하는 촉매 작용을 할 수 있는 효소 활성을 갖는 단백질을 지칭한다. 전형적인 포스포글루코네이트 데하이드라타제는 EC 4.2.1.12이다. 포스포글루코네이트 데하이드라타제는 E. 콜라이에서 edd에 의해 암호화된다.

[0057] 용어 "YciK"는 yciK에 의해 암호화되는 추정상의 효소를 지칭하며, 이는 E. 콜라이에서 코브(I)알라민 아테노

실트랜스페라제(Cob(I)alamin adenosyltransferase)를 암호화하는 유전자 btuR에 번역 단계에서 연결된다.

- [0058] 용어 "코브(I)알라민 아데노실트랜스페라제"는 테옥시아데노실 부분을 ATP로부터 환원된 코리노이드로 전달할 수 있는 효소를 지칭한다. 전형적인 코브(I)알라민 아데노실트랜스페라제는 EC 2.5.1.17이다. 코브(I)알라민 아데노실트랜스페라제는 E. 콜라이의 "btuR", 살모넬라 티피뮤리움의 "cobA", 및 슈도모나스 데니트리피칸스(Pseudomonas denitrificans)의 "cobO" 유전자에 의해 암호화된다.
- [0059] 용어 "갈락토스-양성자 공수송체" 및 "GalP"는 본 명세서에서 호환적으로 사용되며, 당 및 양성자를 주변 세포질로부터 세포질로 수송할 수 있는 효소 활성을 갖는 단백질을 지칭한다. D-글루코스는 GalP에 있어서 선호되는 기질이다. 갈락토스-양성자 공수송체는 에스케리키아 콜라이에서 galP에 의해 암호화된다(서열 번호 21에 제시된 코딩 서열, 서열 번호 22에 제시된 암호화된 단백질 서열).
- [0060] 용어 "비-특이적 촉매 활성"은 3-HPA 및 1,3-프로판다이올의 상호 전환을 촉매 작용할 수 있는 효소 활성을 갖는 폴리펩티드(들)를 지칭하며, 특이적으로 1,3-프로판다이올 옥시도리덕타제(들)를 배제한다. 전형적으로 이들 효소는 알코올 데하이드로게나제이다. 이러한 효소는, FAD 또는 FMN과 같은 플라빈을 포함하나 이에 한정되지 않는, NAD⁺/NADH 외의 보조인자를 이용할 수 있다. 비-특이적 알코올 데하이드로게나제(yqhD)를 위한 유전자는, 예를 들어, E. 콜라이 K-12 변종 내에서 내인성으로 암호화되고 기능적으로 발현되는 것으로 발견된다.
- [0061] 용어 "1.6 장 GI 프로모터(long GI promoter)", "1.20 단/장 GI 프로모터(short/long GI Promoter)", 및 "1.5 장 GI 프로모터"는, 미국 특허 제7,132,527호에 기술된 바와 같이 스트렙토마이세스 리비단스(Streptomyces lividans) 글루코스 아이소머라제(glucose isomerase) 유전자로부터의 프로모터를 함유하는 폴리뉴클레오티드 또는 단편을 지칭한다. 이들 프로모터 단편은 야생형 스트렙토마이세스 리비단스 글루코스 아이소머라제 유전자 프로모터에 비하여 그들의 활성을 감소시키는 돌연변이를 포함한다.
- [0062] 용어 "기능" 및 "효소 기능"은 본 명세서에서 호환적으로 사용되며, 반응에 의해 그 자신은 소모되지 않으면서 특정 화학 반응이 일어나는 속도를 변경시키는 효소의 촉매 활성을 지칭한다. 적합한 조건 하에 산물 또는 기질의 생산이 이루어질 수 있는 평형 중의 반응에 이러한 활성이 적용될 수 있다는 것이 이해된다.
- [0063] 용어 "폴리펩티드" 및 "단백질"은 본 명세서에서 호환적으로 사용된다.
- [0064] 용어 "탄소 기질" 및 "탄소 공급원"은 본 명세서에서 호환적으로 사용되며, 본 명세서에 개시된 재조합 박테리아에 의해 대사될 수 있는 탄소 공급원, 및 특히 프럭토스 및 글루코스를 포함하는 탄소 공급원을 지칭한다. 탄소 공급원은 다른 단당류; 수크로스와 같은 이당류; 올리고당류; 또는 다당류를 추가로 포함할 수 있다.
- [0065] 용어 "숙주 세포" 및 "숙주 박테리아"는 본 명세서에서 호환적으로 사용되며, 외래 또는 이종 유전자를 수용할 수 있고 이들 유전자를 발현시켜 활성 유전자 산물을 생산할 수 있는 박테리아를 지칭한다.
- [0066] 본 명세서에 사용되는 바와 같이 용어 "생산 미생물"은, 재조합된 것들을 포함하나 이에 한정되지 않으며, 1,3-프로판다이올, 글리세롤, 3-하이드록시프로피온산, 다중불포화 지방산 등과 같은 특정 산물을 제조하기 위해 사용되는 미생물을 지칭한다.
- [0067] 본 명세서에 사용되는 바와 같이, "핵산"은 폴리뉴클레오티드를 의미하며 테옥시리보뉴클레오티드 또는 리보뉴클레오티드 염기의 단일 또는 이중-가닥 중합체를 포함한다. 핵산은 또한 단편 및 변형된 뉴클레오티드를 포함할 수 있다. 따라서, 용어 "폴리뉴클레오티드", "핵산 서열", "뉴클레오티드 서열" 또는 "핵산 단편"은 본 명세서에서 호환적으로 사용되며, 합성된 비-자연적 뉴클레오티드 염기 또는 변형된 뉴클레오티드 염기를 임의로 함유하는 단일- 또는 이중-가닥인 RNA 또는 DNA의 중합체를 지칭한다. 뉴클레오티드(통상적으로 그들의 5'-모노포스페이트 형태로 발견됨)는 하기와 같이 그들의 단문자 표기에 의해 지칭된다: 아데닐레이트 또는 테옥시아데닐레이트(각각 RNA 또는 DNA에 대하여)에 대하여 "A", 시티딜레이트 또는 테옥시시티딜레이트에 대하여 "C", 구아닐레이트 또는 테옥시구아닐레이트에 대하여 "G", 유리딜레이트에 대하여 "U", 테옥시티미딜레이트에 대하여 "T", 퓨린(A 또는 G)에 대하여 "R", 피리미딘(C 또는 T)에 대하여 "Y", G 또는 T에 대하여 "K", A 또는 C 또는 T에 대하여 "H", 이노신에 대하여 "I", 그리고 임의의 뉴클레오티드에 대하여 "N".
- [0068] 폴리뉴클레오티드는 합성된, 비-자연적 뉴클레오티드 염기 또는 변형된 뉴클레오티드 염기를 임의로 함유하는 단일- 또는 이중-가닥인 RNA 또는 DNA의 중합체일 수 있다. DNA의 중합체 형태인 폴리뉴클레오티드는 cDNA, 게놈 DNA, 합성 DNA, 또는 그의 혼합물의 하나 이상의 세그먼트로 이루어질 수 있다.

- [0069] "유전자"는 특정 단백질을 발현시키는 핵산 단편을 지칭하며, 이는 코딩 영역만을 지칭하거나 코딩 서열 앞에 있는 조절 서열(5' 비-코딩 서열) 및 코딩 서열 뒤에 있는 조절 서열(3' 비-코딩 서열)을 포함할 수 있다. "천연 유전자"는 그 자신의 조절 서열과 함께 자연에서 발견되는 유전자를 지칭한다. "키메라 유전자"는 자연에서 함께 발견되지 않는 조절 및 코딩 서열을 포함하는, 천연 유전자가 아닌 임의의 유전자를 지칭한다. 따라서, 키메라 유전자는 상이한 공급원으로부터 유래된 조절 서열 및 코딩 서열, 또는 동일한 공급원으로부터 유래되었지만, 자연에서 발견되는 것과는 상이한 방식으로 배열된 조절 서열 및 코딩 서열을 포함할 수 있다. "내인성 유전자"는 유기체의 게놈에서 그의 자연적 위치에 있는 천연 유전자를 지칭한다. "외래" 유전자는 유전자 전달에 의해 숙주 유기체 내로 도입된 유전자를 지칭한다. 외래 유전자는 비-천연 유기체에 삽입된 유전자, 천연 숙주 내의 새로운 위치에 도입된 유전자, 또는 키메라 유전자를 포함할 수 있다.
- [0070] 용어 "천연 뉴클레오타이드 서열"은 숙주 미생물에서 정상적으로 발견되는 뉴클레오타이드 서열을 지칭한다.
- [0071] 용어 "비-천연 뉴클레오타이드 서열"은 숙주 미생물에서 정상적으로 발견되지 않는 뉴클레오타이드 서열을 지칭한다.
- [0072] 용어 "천연 폴리펩티드"는 숙주 미생물에서 정상적으로 발견되는 폴리펩티드를 지칭한다.
- [0073] 용어 "비-천연 폴리펩티드"는 숙주 미생물에서 정상적으로 발견되지 않는 폴리펩티드를 지칭한다.
- [0074] 용어 "암호화하는" 및 "코딩"은 본 명세서에서 호환적으로 사용되며, 유전자가 전사 및 번역의 메커니즘을 통해 아미노산 서열을 생산하는 과정을 지칭한다.
- [0075] 용어 "코딩 서열"은 특정 아미노산 서열을 코딩하는 뉴클레오타이드 서열을 지칭한다.
- [0076] "적합한 조절 서열"은 코딩 서열의 상류(5' 비-코딩 서열), 그의 내부 또는 그의 하류(3' 비-코딩 서열)에 위치하고, 전사, RNA 가공 또는 안정성 또는 연계된 코딩 서열의 번역에 영향을 미치는 뉴클레오타이드 서열을 지칭한다. 조절 서열은 프로모터, 인핸서(enhancer), 사일런서(silencer), 5' 비번역 리더(untranslated leader) 서열(예를 들어, 전사 시작 부위와 번역 개시 코돈의 사이), 인트론, 폴리아데닐화 인식 서열, RNA 가공 부위, 이펙터(effector) 결합 부위 및 스템-루프 구조를 포함할 수 있다.
- [0077] 용어 "발현 카세트는", 선택된 유전자의 코딩 서열 및 선택된 유전자 산물의 발현을 위해 필요한 코딩 서열의 앞(5' 비-코딩 서열) 및 뒤(3' 비-코딩 서열)에 있는 조절 서열을 포함하는 DNA의 단편을 지칭한다. 따라서, 발현 카세트는 전형적으로, 1) 프로모터 서열; 2) 코딩 서열(즉, ORF) 및, 3) 진핵 세포에서 통상적으로 폴리아데닐화 부위를 함유하는 3' 비번역 영역(예를 들어, 종결자(terminator))으로 구성된다. 발현 카세트(들)는 통상적으로 벡터 내에 포함되어, 클로닝 및 형질전환을 용이하게 한다. 각각의 숙주에 대해 정확한 조절 서열이 사용되는 한, 박테리아, 효모, 및 진균류를 포함하는 상이한 유기체를 상이한 발현 카세트로 형질전환시킬 수 있다.
- [0078] "형질전환"은, 숙주 유기체 내로 핵산 분자를 전달하여 유전적으로 안정한 유전성을 유발하는 것을 지칭한다. 핵산 분자는, 예를 들어, 자체적으로 복제될 수 있는 플라스미드일 수도 있고, 그것이 숙주 유기체의 게놈 내로 통합될 수도 있다. 핵산 단편으로 형질전환된 숙주 유기체는 "재조합" 또는 "형질전환된" 유기체 또는 "형질전환체"라고 지칭된다. "안정한 형질전환"은 핵 및 소기관 게놈 양자 모두를 포함하는 숙주 유기체의 게놈 내로 핵산 단편을 전달하여, 유전적으로 안정한 유전성을 유발하는 것을 지칭한다. 반대로, "일시적 형질전환"은, 숙주 유기체의 핵, 또는 DNA-함유 소기관 내로 핵산 단편을 전달하여, 통합 또는 안정한 유전성이 없는 유전자 발현을 유발하는 것을 지칭한다.
- [0079] "코돈 축퇴성(Codon degeneracy)"은 암호화된 폴리펩티드의 아미노산 서열에 영향을 주지 않으면서 뉴클레오타이드 서열의 변이를 허용하는 유전 암호의 성질을 지칭한다. 주어진 아미노산을 특정하기 위한 뉴클레오타이드 코돈의 사용에 있어서 특정 숙주 세포에서 나타나는 "코돈-바이어스(codon-bias)"가 당업자에게 주지되어 있다. 그러므로, 숙주 세포 내의 발현을 개선하기 위한 유전자를 합성할 경우에는, 그의 코돈 사용 빈도가 숙주 세포가 선호하는 코돈의 사용 빈도에 근접하도록 유전자를 고안하는 것이 바람직하다.
- [0080] 용어 "기능적으로 균등한 하위단편(subfragment)" 및 "기능적 균등 하위단편"은 본 명세서에서 호환적으로 사용된다. 이들 용어는 단리된 핵산 단편의 일부 또는 하위서열을 지칭하며, 여기에는 단편 또는 하위단편이 활성 효소를 암호화하는지 여부에 무관하게 유전자 발현을 변경하거나 소정의 표현형을 생산하는 능력이 보유된다. 키메라 유전자는, 그것이 활성 효소를 암호화하는지 여부에 무관하게, 프로모터 서열에 대해 센스 또는 안티센스 배향으로, 그의 핵산 단편 또는 하위단편을 연결함으로써 억제에 사용하기 위해 고안될 수 있다.

- [0081] 용어 "보존된 도메인" 또는 "모티프"는 진화적으로 관련된 단백질의 정렬된 서열을 따라 특정 위치에서 보존된 아미노산의 세트를 의미한다. 다른 위치의 아미노산은 상동 단백질들 사이에 변동될 수 있는 반면에, 특정 위치에서 고도로 보존된 아미노산은 단백질의 구조, 안정성, 또는 활성에 있어서 필수적인 아미노산을 나타낸다.
- [0082] 용어 "실질적으로 유사한" 및 "실질적으로 상응하는"은 본 명세서에서 호환적으로 사용된다. 그들은 유전자 발현을 매개하거나 소정의 표현형을 생산하는 핵산 단편의 능력에 하나 이상의 뉴클레오티드 염기의 변화가 영향을 미치지 않는 핵산 단편을 지칭한다. 이들 용어는 또한, 본 발명의 핵산 단편의 변형, 예를 들어 초기의 비변형된 단편에 대해 생성되는 핵산 단편의 기능적 특성을 실질적으로 변경하지 않는 하나 이상의 뉴클레오티드의 결실 또는 삽입을 지칭한다. 그러므로, 당업자가 인식할 바와 같이, 본 발명은 특정 예시적 서열을 초과하여 포괄한다는 것이 이해된다. 더욱이, 본 발명이 포괄하는 실질적으로 유사한 핵산 서열은 본 명세서에 예시된 서열, 또는 본 명세서에 개시되고 본 명세서에 개시된 핵산 서열 중 임의의 것과 기능적으로 균등한 뉴클레오티드 서열의 임의의 부분에 혼성화되는 그들의 능력(중간 정도의 엄격 조건 하에, 예를 들어, 0.5X SSC(표준 소듐 시트레이트), 0.1% SDS(소듐 도데실 설페이트), 60°C)에 의해 또한 정의된다는 것을, 당업자는 인식한다. 엄격성 조건을 조정하여, 관계가 먼 유기체로부터의 상동 서열과 같이 중간 정도로 유사한 단편 내지 밀접하게 관련된 유기체로부터의 기능성 효소를 중복시키는 유전자와 같이 고도로 유사한 단편을 스크리닝할 수 있다. 혼성화후 세척은 엄격성 조건을 결정한다.
- [0083] 용어 "선택적으로 혼성화되다"는, 엄격 혼성화 조건 하에 핵산 서열이 비-표적 핵산 서열에 대한 그의 혼성화보다 검출가능하게 더 큰 정도로(예를 들어, 백그라운드에 비해 적어도 2배) 특정된 핵산 표적 서열에 혼성화됨에 대한 언급, 및 비-표적 핵산의 실질적인 배제에 대한 언급을 포함한다. 선택적으로 혼성화되는 서열은, 뉴클레오티드 서열 중 하나의 상보체가 다른 뉴클레오티드 서열에 대해 전형적으로 적어도 약 80%의 서열 동일성, 또는 90%의 서열 동일성, 100% 이하의 서열 동일성(즉, 완전히 상보적임)을 갖는 2개의 뉴클레오티드 서열이다.
- [0084] 용어 "엄격 조건" 또는 "엄격 혼성화 조건"은, 탐침이 그의 표적 서열에 선택적으로 혼성화될 조건에 대한 언급을 포함한다. 전형적으로 탐침은, 검출하고자 하는 핵산 서열에 상보적인 단일 가닥 핵산 서열이다. 탐침은 검출하고자 하는 핵산 서열에 "혼성화가능"하다. 일반적으로, 탐침은 약 1000 뉴클레오티드 미만의 길이이며, 임의로 500 뉴클레오티드 미만의 길이이다.
- [0085] 혼성화 방법은 잘 정립되어 있다. 전형적으로, 탐침과 샘플은 핵산 혼성화를 허용할 조건 하에서 혼합된다. 이는 적절한 농도 및 온도 조건 하에 무기 또는 유기염의 존재 하에 탐침 및 시료를 접촉시키는 단계를 포함한다. 임의로, 카오트로픽제(chaotropic agent)가 첨가될 수 있다. 핵산 혼성화는 다양한 분석 체제에 적용시킬 수 있다. 가장 적합한 것 중의 하나는 샌드위치 분석(sandwich assay) 체제이다. 샌드위치-유형 분석의 주요 컴포넌트는 고체 지지체이다. 고체 지지체에는, 서열의 일부에 대해 상보적이고 표지되지 않은 고정화된 핵산 탐침이 흡착되거나 공유적으로 결합된다.
- [0086] 엄격 조건은 서열-의존성이며, 상이한 상황에서 상이할 것이다. 혼성화 및/또는 세척 조건의 엄격성을 제어함으로써, 탐침에 100% 상보적인 표적 서열을 동정할 수 있다(상동 탐침법(homologous probing)). 대안적으로, 더 낮은 정도의 유사성이 검출되도록 서열의 일부 불일치를 허용하기 위해 엄격성 조건을 조정할 수 있다(이중 탐침법(heterologous probing)).
- [0087] 전형적으로 엄격 조건은, 염 농도가 pH 7.0 내지 8.3에서 약 1.5 M 미만의 Na 이온, 전형적으로는 약 0.01 내지 1.0 M의 Na 이온 농도(또는 다른 염)이며, 온도가 짧은 탐침(예를 들어, 10 내지 50 뉴클레오티드)에 있어서는 적어도 약 30°C이고 긴 탐침(예를 들어, 50 뉴클레오티드 초과)에 있어서는 적어도 약 60°C인 것들일 것이다. 엄격 조건은 또한, 포름아미드와 같은 불안정화제의 첨가로 달성할 수 있다. 예시적인 낮은 엄격성 조건은, 37°C에서 30 내지 35% 포름아미드, 1 M NaCl, 1% SDS(소듐 도데실 설페이트)의 완충 용액을 이용하는 혼성화, 및 50 내지 55°C에서 1X 내지 2X SSC(20X SSC = 3.0 M NaCl/0.3 M 트라이소듐 시트레이트) 중의 세척을 포함한다. 예시적인 중간 정도의 엄격성 조건은, 37°C에서 40 내지 45% 포름아미드, 1 M NaCl, 1% SDS 중의 혼성화, 및 55 내지 60°C에서 0.5X 내지 1X SSC 중의 세척을 포함한다. 예시적인 높은 엄격성 조건은, 37°C에서 50% 포름아미드, 1 M NaCl, 1% SDS 중의 혼성화, 및 60 내지 65°C에서 0.1X SSC 중의 세척을 포함한다.
- [0088] 특이성은 전형적으로 혼성화후 세척의 함수이며, 결정적 인자는 최종 세척 용액의 이온 세기 및 온도이다. DNA-DNA 혼성의 경우, 문헌[Meinkoth et al., Anal. Biochem. 138:267-284 (1984)]의 하기 수학적식으로부터 얻

용융점(T_m)의 근사치를 계산할 수 있다: $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ 포름}) - 500/L$ (여기서 M 은 1가 양이온의 몰 농도이고, $\%GC$ 는 DNA 중의 구아노신 및 시토신 뉴클레오타이드의 백분율이며, $\% \text{ 포름}$ 은 혼성화 용액 중의 포름아미드의 백분율이고, L 은 염기쌍 단위의 혼성 길이임). T_m 은 50%의 상보적인 표적 서열이 완전하게 일치하는 탐침에 혼성화되는 온도이다(정의된 이온 세기 및 pH 하에). T_m 은 매 1%의 불일치마다 약 1°C 씩 감소하므로, 목적하는 동일성의 서열에 혼성화되도록 T_m , 혼성화 및/또는 세척 조건을 조정할 수 있다. 예를 들어, $\geq 90\%$ 동일성을 가진 서열을 탐색한다면, T_m 을 10°C 감소시킬 수 있다. 일반적으로, 엄격 조건은 정의된 이온 세기 및 pH에서 특정 서열 및 그의 상보체에 대한 T_m 보다 약 5°C 낮게 선택된다. 그러나, 극도의 엄격 조건은 T_m 보다 1, 2, 3, 또는 4°C 낮은 온도에서의 혼성화 및/또는 세척을 이용할 수 있고; 중간 정도의 엄격 조건은 T_m 보다 6, 7, 8, 9, 또는 10°C 낮은 온도에서의 혼성화 및/또는 세척을 이용할 수 있으며; 낮은 엄격성 조건은 T_m 보다 11, 12, 13, 14, 15, 또는 20°C 낮은 온도에서의 혼성화 및/또는 세척을 이용할 수 있다. 수확식, 혼성화 및 세척 조성물, 및 목적하는 T_m 을 사용하여, 혼성화 및/또는 세척 용액의 엄격성의 변동이 본질적으로 기술된다는 것을, 당업자는 이해할 것이다. 목적하는 불일치의 정도가 45°C (수용액) 또는 32°C (포름아미드 용액) 미만의 T_m 을 유발한다면, 더 높은 온도를 사용할 수 있도록 SSC 농도를 증가시키는 것이 바람직하다. 핵산의 혼성화에 대한 광범위한 지침은 문헌[Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993)]; 및 문헌[Current Protocols in Molecular Biology, Chapter 2, Ausubel et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995)]에서 발견된다. 혼성화 및/또는 세척 조건은 적어도 10, 30, 60, 90, 120, 또는 240 분 동안 적용될 수 있다.

[0089] 핵산 또는 폴리펩티드 서열과 관련하여 "서열 동일성" 또는 "동일성"은, 특정된 비교창 상에 최대 상응도로 정렬될 경우에 2개 서열 중의 동일한 핵산 염기 또는 아미노산 잔기를 지칭한다.

[0090] 따라서, "서열 동일성의 백분율"은, 비교창 상에 최적으로 정렬된 2개의 서열을 비교함으로써 결정된 값을 지칭하며, 여기서 비교창 내의 폴리뉴클레오타이드 또는 폴리펩티드 서열의 부분은 2개 서열의 최적 정렬을 위한 기준 서열(부가 또는 결실을 포함하지 않음)과 비교하여 부가 또는 결실(즉, 갭)을 포함할 수 있다. 양자 모두의 서열 내에서 동일한 핵산 염기 또는 아미노산 잔기가 나타나는 위치의 개수를 결정하여 일치하는 위치의 개수를 산출하고, 일치하는 위치의 개수를 비교창 내의 위치의 총 개수로 나누고, 그 결과에 100을 곱하여 서열 동일성의 백분율을 산출함으로써 백분율을 계산한다. 퍼센트 서열 동일성의 유용한 예는 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95%, 또는 50% 내지 100%의 임의의 정수 백분율을 포함하나, 이에 한정되지 않는다. 이들 동일성은 본 명세서에 기술된 임의의 프로그램을 사용하여 결정할 수 있다.

[0091] 서열 정렬 및 퍼센트 동일성 또는 유사성 계산은, 레이저진 바이오인포매틱스 컴퓨팅 스위트(LASERGENE bioinformatics computing suite)(위스콘신주 메디슨 소재의 디엔에이스타 인코포레이티드(DNASTAR Inc.))의 맥얼라인(MegAlign)(상표) 프로그램을 포함하나 이에 한정되지 않는, 상동 서열을 검출하도록 고안된 다양한 비교 방법을 사용하여 결정할 수 있다. 본 출원과 관련하여, 서열 분석 소프트웨어가 분석을 위해 사용된 경우에는, 그 분석 결과가 달리 명시되지 않는 한, 참조한 프로그램의 "내정값"에 기초할 것이라는 사실을 이해할 것이다. 본 명세서에 사용되는 바와 같이, "내정값"은 최초 초기화할 때에 소프트웨어에 원래 로딩된 값 또는 파라미터의 임의의 세트를 의미할 것이다.

[0092] "클러스탈 V(Clustal V) 정렬 방법"은 클러스탈 V(문헌[Higgins and Sharp, CABIOS. 5:151-153 (1989)]; 문헌[Higgins, D.G. et al., Comput. Appl. Biosci. 8:189-191 (1992)]에 기술됨)로 표지되고 레이저진 바이오인포매틱스 컴퓨팅 스위트(위스콘신주 메디슨 소재의 디엔에이스타 인코포레이티드)의 맥얼라인(상표) 프로그램에서 발견되는 정렬 방법에 상응한다. 다중 정렬의 경우, 내정값은 갭 페널티(GAP PENALTY)=10 및 갭 길이 페널티(GAP LENGTH PENALTY)=10에 상응한다. 클러스탈 V 방법을 사용하는 단백질 서열의 퍼센트 동일성의 계산 및 쌍정렬을 위한 내정 파라미터는 케이터플(KTUPLE)=1, 갭 페널티=3, 윈도우(WINDOW)=5 및 다이아고널스 세이브드(DIAGONALS SAVED)=5이다. 핵산의 경우, 이들 파라미터는 케이터플=2, 갭 페널티=5, 윈도우=4 및 다이아고널스 세이브드=4이다. 클러스탈 V 프로그램을 사용하여 서열들을 정렬한 후에, 동일한 프로그램 내의 "서열 거리" 표를 보면 "퍼센트 동일성"을 얻을 수 있다.

[0093] "클러스탈 W 정렬 방법"은 클러스탈 W(문헌[Higgins and Sharp, supra]; 문헌[Higgins, D.G. et al., supra])

에 기술됨)로 표시되고 레이저진 바이오인포매틱스 컴퓨팅 스위트(위스콘신주 메디슨 소재의 디엔에이스타 인코퍼레이티드)의 맥얼라인(상표) v6.1 프로그램에서 발견되는 정렬 방법에 상응한다. 다중 정렬을 위한 내정 파라미터는 갭 페널티=10, 갭 길이 페널티=0.2, 지연 이탈 서열(Delay Divergen Seqs)(%)=30, DNA 전이 중량(DNA Transition Weight)=0.5, 단백질 중량 매트릭스(Protein Weight Matrix)=고넷 시리즈(Gonnet Series), DNA 중량 매트릭스(DNA Weight Matrix)=IUB에 상응한다. 클러스탈 W 프로그램을 사용하여 서열들을 정렬한 후에, 동일한 프로그램 내의 "서열 거리" 표를 보면 "퍼센트 동일성"을 얻을 수 있다.

[0094] "BLASTN 정렬 방법"은 국립 생물공학 정보센터(NCBI: National Center for Biotechnology Information)에 의해 제공되는, 내정 파라미터를 사용하여 뉴클레오티드 서열을 비교하는 알고리즘이다.

[0095] 동일하거나 유사한 기능 또는 활성을 가진 폴리펩티드를 다른 종으로부터 동정함에 있어서 여러 수준의 서열 동일성이 유용하다는 것이 당업자에게 주지되어 있다. 퍼센트 동일성의 유용한 예는 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95%, 또는 50% 내지 100%의 임의의 정수 백분율을 포함하나, 이에 한정되지 않는다. 사실상, 본 발명을 기술함에 있어서, 50% 내지 100%의 임의의 정수 아미노산 동일성, 예를 들어 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99%가 유용할 수 있다. 이 단리된 뉴클레오티드 단편의 임의의 전체-길이 또는 부분 상보체 또한 관심의 대상이다.

[0096] 따라서, 본 발명은 본 명세서에 개시된 특정 예시적 뉴클레오티드 서열을 초과하여 포괄한다. 예를 들어, 유전 암호의 축퇴성을 반영하는 유전자 서열의 변경이 고려된다. 또한, 주어진 부위에 화학적으로 균등한 아미노산의 생산을 유발하면서 암호화된 단백질의 기능적 특성에 영향을 미치지 않는 유전자의 변경이 흔하다는 것은 당업계에 주지되어 있다. 본 명세서의 논의에 있어서 치환은 하기의 5개 그룹 중 하나 내에서의 교환으로 정의된다:

[0097] 1. 작은 지방족의, 비극성 또는 약한 극성의 잔기: Ala, Ser, Thr (Pro, Gly);

[0098] 2. 극성의, 음성 하전된 잔기 및 그의 아미드: Asp, Asn, Glu, Gln;

[0099] 3. 극성의, 양성 하전된 잔기: His, Arg, Lys;

[0100] 4. 큰 지방족의, 비극성 잔기: Met, Leu, Ile, Val (Cys); 및

[0101] 5. 큰 방향족 잔기: Phe, Tyr, Trp.

[0102] 따라서, 소수성 아미노산인 아미노산 알라닌에 대한 코돈은 소수성이 좀 더 낮은 다른 잔기(예를 들어, 글리신) 또는 소수성이 좀 더 높은 잔기(예를 들어, 발린, 류신 또는 아이소류신)를 암호화하는 코돈으로 치환될 수 있다. 마찬가지로, 음성 하전된 하나의 잔기로 다른 하나를 치환하거나(예를 들어 글루탐산을 아스파르트산으로) 양성 하전된 하나의 잔기로 다른 하나를 치환하는(예를 들어, 알지닌을 라이신으로) 변화도 기능적으로 균등한 산물을 생성시킬 것으로 예상할 수 있다. 많은 경우에, 단백질 분자의 N-말단 및 C-말단 부분의 변경을 유발하는 뉴클레오티드 변화도 단백질의 활성을 변경하지 않을 것으로 예상된다.

[0103] 각각의 제안된 변형은 당업계의 통상적 기술로 충분히 가능하며, 암호화된 산물의 생물학적 활성의 보존을 결정하는 것도 그러하다. 또한, 본 발명에 의해 포괄되는 실질적으로 유사한 서열은 상기 정의된 바와 같은 엄격 조건 하에 혼성화되는 그들의 능력에 의해 또한 정의된다는 것을, 당업자는 인식한다.

[0104] 본 발명의 실질적으로 유사한 바람직한 핵산 단편은, 뉴클레오티드 서열이 본 명세서에 보고된 핵산 단편의 뉴클레오티드 서열과 적어도 70% 동일한 핵산 단편들이다. 더욱 바람직한 핵산 단편은 본 명세서에 보고된 핵산 단편의 뉴클레오티드 서열과 적어도 90% 동일하다. 본 명세서에 보고된 핵산 단편의 뉴클레오티드 서열과 적어도 95% 동일한 핵산 단편이 가장 바람직하다.

[0105] 아미노산 또는 뉴클레오티드 서열의 "실질적인 부분"은, 당업자에 의한 서열의 수동 평가에 의해, 또는 BLAST(기본적 국소 정렬 검색 도구(Basic Local Alignment Search Tool); Altschul, S. F., et al., J. Mol. Biol., 215:403-410(1993))와 같은 알고리즘을 사용하는 컴퓨터-자동화 서열 비교 및 동정에 의해, 폴리펩티드 또는 유전자를 추정상으로 동정하기에 충분한 폴리펩티드의 아미노산 서열 또는 유전자의 뉴클레오티드 서열을 포함하는 부분이다. 일반적으로, 폴리펩티드 또는 핵산 서열을 공지의 단백질 또는 유전자와 상동인 것으로 추정상 동정하기 위해서는 10개 이상 연속되는 아미노산 또는 30개 이상의 뉴클레오티드의 서열이 필요하다. 또한, 뉴클레오티드 서열에 있어서는, 서열-의존적인 유전자 동정(예를 들어, 서던 혼성화(Southern

hybridization)) 및 단리(예를 들어, 박테리아 콜로니 또는 박테리오파아지 플라크의 인 시츄 혼성화(in situ hybridization)) 방법에 20 내지 30개의 연속된 뉴클레오티드를 포함하는 유전자-특이적 올리고뉴클레오티드 탐침이 사용될 수 있다. 또한, 12 내지 15 염기의 짧은 올리고뉴클레오티드를 PCR에서 증폭 프라이머로 사용하여, 프라이머를 포함하는 특정 핵산 단편을 얻을 수 있다. 따라서, 뉴클레오티드 서열의 "실질적인 부분"은, 서열을 포함하는 핵산 단편을 특이적으로 동정 및/또는 단리하기에 충분한 서열을 포함한다. 본 명세서에는, 특정 단백질을 암호화하는 완전한 아미노산 및 뉴클레오티드 서열을 교시한다. 본 명세서에 보고된 서열의 유익성을 가지고, 당업자는 개시된 서열의 전체 또는 실질적인 부분을 당업계에 공지된 목적으로 사용할 수 있게 되었다.

[0106] 용어 "상보적인"은, 역평행 배향으로 정렬될 때 왓슨-크릭(Watson-Crick) 염기-대합이 가능한 2개 서열의 뉴클레오티드 염기 사이의 관계를 기술한다. 예를 들어, DNA에 있어서, 아데노신은 티민과 염기-대합이 가능하고 시토신은 구아닌과 염기-대합이 가능하다. 따라서 본 발명은, 첨부된 서열 목록 및 명세서에 보고된 완전한 서열 뿐 아니라 실질적으로 유사한 핵산 서열에 상보적인 단리된 핵산 분자를 이용할 수 있다.

[0107] 용어 "단리된"은, 폴리펩티드 또는 뉴클레오티드 서열과 자연적으로 연계된 적어도 하나의 컴포넌트로부터 분리된 폴리펩티드 또는 뉴클레오티드 서열을 지칭한다.

[0108] "프로모터"는 코딩 서열 또는 기능성 RNA의 발현을 제어할 수 있는 DNA 서열을 지칭한다. 프로모터 서열은 근위 상류 요소 및 더욱 원위의 상류 요소로 이루어지며, 후자의 요소는 흔히 인헨서로 지칭된다. 따라서, "인헨서"는 프로모터 활성을 자극할 수 있는 DNA 서열이며, 프로모터의 고유한 요소이거나 프로모터의 조직-특이성 또는 수준을 향상시키도록 삽입된 이종 요소일 수 있다. 프로모터는 천연 유전자로부터 그 전체가 유래될 수도 있거나, 자연에서 발견되는 상이한 프로모터로부터 유래된 상이한 요소로 구성될 수도 있거나, 심지어 합성 DNA 세그먼트를 포함할 수도 있다. 상이한 프로모터는 상이한 조직 또는 세포 유형에서, 또는 상이한 발생 단계에서, 또는 상이한 환경적 조건에 반응하여, 유전자의 발현을 유도할 수 있음을 당업자는 이해할 것이다. 대부분의 경우에 조절 서열의 정확한 경계는 완벽하게 정의되지 않았으므로, 일부 변이의 DNA 단편이 동일한 프로모터 활성을 가질 수 있음이 추가로 인정된다. 대부분의 세포 유형에서 대부분의 시간에 유전자가 발현되도록 하는 프로모터는 통상적으로 "구성성 프로모터"라고 지칭된다.

[0109] "3' 비-코딩 서열", "전사 종결자" 및 "종결 서열"은 본 명세서에서 호환적으로 사용되며, 코딩 서열의 하류에 위치하는 DNA 서열, 예를 들어 폴리아데닐화 인식 서열 및 mRNA 가공 또는 유전자 발현에 영향을 미칠 수 있는 조절 신호를 암호화하는 다른 서열을 지칭한다. 폴리아데닐화 신호는 통상적으로 mRNA 전구체의 3' 말단에 폴리아데닐산 트랙트를 부가하는 데에 영향을 미치는 것을 특징으로 한다.

[0110] 용어 "작동적으로 연결된"은 하나의 기능이 나머지에 의해 영향을 받도록 하는, 단일 핵산 단편 상에서의 핵산 서열의 연계를 지칭한다. 예를 들어, 프로모터가 코딩 서열의 발현에 영향을 미칠 수 있는 경우에, 프로모터는 그 코딩 서열에 작동적으로 연결된다(즉, 코딩 서열은 프로모터의 전사 제어 하에 있다). 코딩 서열은 센스 또는 안티센스 배향으로 조절 서열에 작동적으로 연결될 수 있다. 다른 예에서, 본 발명의 상보적 RNA 영역은 직접 또는 간접적으로 표적 mRNA에 대해 5', 또는 표적 mRNA에 대해 3', 또는 표적 mRNA 내부에 작동적으로 연결될 수 있거나, 제1 상보적 영역은 표적 mRNA에 대해 5'이고 그의 상보체는 3'이다.

[0111] 본 명세서에 사용되는 표준 제조법 DNA 및 분자 클로닝 기술은 당업계에 주지되어 있으며 문헌[Sambrook, J., Fritsch, E.F. and Maniatis, T. Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989)]에 더욱 완전하게 기술되어 있다. 형질전환 방법은 당업자에게 주지되어 있으며, 하기에 기술된다.

[0112] "PCR" 또는 "폴리머라제 연쇄 반응"은 다량의 특정 DNA 세그먼트의 합성을 위한 기술이며 일련의 반복 주기로 구성된다(코네티컷주 노르워크 소재의 퍼킨 엘머 세터스 인스트루먼트). 전형적으로, 이중-가닥 DNA를 열변성시키고, 표적 세그먼트의 3' 경계에 상보적인 2개의 프라이머를 저온에서 어닐링시킨 후, 중간 온도에서 연장한다. 이들 3개의 연이은 단계의 1 세트를 "주기"라고 지칭한다.

[0113] "플라스미드" 또는 "벡터"는, 흔히 세포의 중심 대사작용의 일부가 아닌 유전자를 운반하는, 통상적으로 원형 이중-가닥 DNA 단편 형태의 염색체 외 요소이다. 이러한 요소는, 세포 내로 발현 카세트(들)를 도입할 수 있는 독특한 작제물 내로 많은 뉴클레오티드 서열이 결합되거나 재조합된, 임의의 공급원으로부터 유래된 단일-또는 이중-가닥 DNA 또는 RNA의 선형 또는 원형의 자체적으로 복제되는 서열, 게놈 통합 서열, 파아지 또는 뉴클레오티드 서열일 수 있다.

- [0114] 용어 "유전적으로 변경된"은 유전공학, 형질전환 및/또는 돌연변이에 의해 유전 물질을 변화시키는 과정을 지칭한다.
- [0115] 용어 "재조합"은, 예를 들어 유전공학 기술에 의한 단리된 핵산 세그먼트의 조작에 의하거나 화학적 합성에 의해, 그렇지 않다면 분리되어 있는 2개의 서열 세그먼트의 인공 조합을 지칭한다. "재조합"은 또한, 이중 핵산의 도입에 의해 변형된 세포 또는 벡터, 또는 그렇게 변형된 세포로부터 유래하는 세포에 대한 언급을 포함하나, 의도적인 인위적 개입 없이 발생하는 것들과 같이 자연적으로 발생하는 사건(예를 들어, 자발적 돌연변이, 자연적 형질전환, 자연적 형질도입, 자연적 전위)에 의한 세포 또는 벡터의 변경은 포괄하지 않는다.
- [0116] 용어 "재조합 작제물", "발현 작제물", "키메라 작제물", "작제물", 및 "재조합 DNA 작제물"은 본 명세서에서 호환적으로 사용된다. 재조합 작제물은 핵산 단편, 예를 들어, 자연에서는 함께 발견되지 않는 조절 및 코딩 서열의 인공 조합을 포함한다. 예를 들어, 재조합 작제물은 상이한 공급원으로부터 유래된 조절 서열 및 코딩 서열, 또는 동일한 공급원으로부터 유래되지만, 자연에서 발견되는 것과는 상이한 방식으로 배열된 조절 서열 및 코딩 서열을 포함할 수 있다. 이러한 작제물은 단독으로 사용될 수도 있고, 벡터와 함께 사용될 수도 있다. 벡터를 사용한다면, 당업자에게 주지된 바와 같이 벡터의 선택은 숙주 세포를 형질전환시키기 위해 사용될 방법에 따라 달라진다. 예를 들어, 플라스미드 벡터를 사용할 수 있다. 본 발명의 임의의 단리된 핵산 단편을 포함하는 숙주 세포를 성공적으로 형질전환시키고 선택하고 증식시키기 위하여 벡터 상에 존재해야 하는 유전 요소는 당업자에게 주지되어 있다. 상이하고 독립적인 형질전환 사건은 발현의 상이한 수준 및 양상을 유발할 수 있으며(Jones et al., EMBO J. 4:2411-2418 (1985); De Almeida et al., Mol. Gen. Genetics 218:78-86 (1989)), 따라서 목적하는 발현 수준 및 양상을 나타내는 라인을 얻기 위해서는 그러한 다중의 사건을 스크리닝할 필요가 있다는 것을, 당업자는 또한 인식할 것이다. 이러한 스크리닝은 특히 DNA의 서던 분석, mRNA 발현의 노던 분석, 단백질 발현의 면역 블로팅 분석, 또는 표현형 분석에 의해 이루어질 수 있다.
- [0117] 본 명세서에 사용되는 바와 같이, 용어 "발현"은 기능성 최종 산물(예를 들어, mRNA 또는 단백질[전구체 또는 성숙체])의 생산을 지칭한다.
- [0118] 용어 "도입된"은 핵산(예를 들어, 발현 작제물) 또는 단백질을 세포 내에 제공함을 의미한다. "도입된"은, 진핵 또는 원핵 세포 내로의 핵산의 혼입(여기서 핵산은 세포의 게놈 내로 혼입될 수 있음)에 대한 언급을 포함하며, 핵산 또는 단백질을 세포에 일시적으로 제공하는 것에 대한 언급을 포함한다. "도입된"은 유성 교배(sexually crossing)와 더불어 안정하거나 일시적인 형질전환 방법에 대한 언급을 포함한다. 따라서, 핵산 단편(예를 들어, 재조합 작제물/발현 작제물)을 세포 내로 삽입하는 단계와 관련하여, "도입된"은 "형질감염" 또는 "형질전환" 또는 "형질도입"을 의미하며, 진핵 또는 원핵 세포 내로의 핵산 단편의 혼입(여기서 핵산 단편은 세포의 게놈(예를 들어, 염색체, 플라스미드, 플라스티드 또는 미토콘드리아 DNA) 내로 혼입되거나, 자체적 레플리콘(autonomous replicon)으로 전환되거나, 일시적으로 발현될 수 있음(예를 들어, 형질감염된 mRNA))에 대한 언급을 포함한다.
- [0119] 용어 "상동"은, 유사한 촉매 기능을 가진, 공통의 진화적 기원의 단백질 또는 폴리펩티드를 지칭한다. 본 발명은 재조합 기술을 통해 상동 단백질을 생산하는 박테리아를 포함할 수 있다.
- [0120] 본 명세서에는, 그들의 게놈 내에 또는 적어도 하나의 재조합 작제물 상에 하기의 것을 포함하는 재조합 박테리아가 개시된다: 수크로스 수송체 활성을 갖는 폴리펩티드 또는 폴리펩티드 복합체를 암호화하는 하나 이상의 뉴클레오티드 서열; 프럭토키나제 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열; 및 수크로스 하이드롤라제 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열. 이들 뉴클레오티드 서열은 동일하거나 상이한 프로모터에 각각 작동적으로 연결된다. 이들 재조합 박테리아는 수크로스를 대사시켜 글리세롤 및/또는 글리세롤-유래의 산물, 예를 들어 1,3-프로판다이올 및 3-하이드록시프로피온산을 생산할 수 있다. 본 명세서에 하기 기술되는 바와 같이, 글리세롤 및/또는 글리세롤-유래의 산물을 생산할 수 있는 박테리아 변종은 고도로 조작된 변종이다.
- [0121] 본 명세서에 개시된 재조합 박테리아의 작제에 사용하기에 적합한 숙주 박테리아는 에스케리키아, 스트렙토코커스, 아그로박테리움, 바실러스, 코리네박테리움, 락토바실러스, 클로스트리디움, 글루코노박터, 시트로박터, 엔테로박터, 클렙시엘라, 아에로박터, 메틸로박터, 살모넬라, 스트렙토마이세스, 및 슈도모나스 속의 유기체를 포함하나, 이에 한정되지 않는다.
- [0122] 일 실시 양태에서 숙주 박테리아는 에스케리키아, 클렙시엘라, 시트로박터, 및 아에로박터 속으로부터 선택된다.

- [0123] 다른 실시 양태에서, 숙주 박테리아는 에스케리키아 콜라이이다.
- [0124] 일부 실시 양태에서, 숙주 박테리아는 PTS 음성이다. 이들 실시 양태에서, 숙주 박테리아는 그의 천연 상태에서 PTS 음성이거나, 하기와 같이 PTS 유전자의 불활성화를 통해 PTS 음성이 되게 할 수 있다.
- [0125] 생산 미생물에서는, 당의 수송과 수송되는 당의 인산화를 위한 포스포에놀피루베이트(PEP)의 사용을 연결하지 않는 것이 간혹 바람직하다.
- [0126] 용어 "하향 조절된"은, 야생형 단백질(들)의 활성에 비교하여 활성 단백질(들)의 활성을 감소시키거나 없애는 것을 지칭한다. 이 유형의 수송에 필요한 단백질을 암호화하는 하나 이상의 내인성 유전자의 발현을 하향 조절함으로써 PTS를 불활성화할 수 있다("PTS 음성" 유기체를 유발함). 하향 조절은 전형적으로, 하나 이상의 이들 유전자가 "파괴(disruption)"를 갖는 경우(유전자가 게놈으로부터 결실되고 단백질이 번역되지 않도록 하는 완전한 유전자 녹아웃(gene knockout), 또는 삽입, 결실, 아미노산 치환 또는 다른 표적화 돌연변이를 갖도록 번역된 단백질을 유발하는, 그 유전자의 일부 내의 삽입, 결실, 또는 표적화 돌연변이를 지칭함)에 발생한다. 단백질 내의 파괴의 위치는, 예를 들어 단백질의 N-말단 부분 내에, 또는 단백질의 C-말단 부분 내에 존재할 수 있다. 파괴된 단백질은 파괴되지 않은 단백질에 비해 손상된 활성을 가질 것이며, 비-기능성일 수 있다. 단백질 발현이 낮거나 결여되게 하는 하향 조절은 또한, 조절 서열, 전사 및 번역 인자 및/또는 신호 전달(signal transduction) 경로의 조작을 통해, 또는 센스, 안티센스 또는 RNAi 기술 등의 사용에 의해 유발될 수 있다.
- [0127] 수크로스 수송체 폴리펩티드 또는 폴리펩티드 복합체는, 미생물 세포 내로의 수크로스의 수송을 매개할 수 있는 폴리펩티드 또는 폴리펩티드 복합체이다. 수크로스 수송 폴리펩티드 및 폴리펩티드 복합체는 상기와 같이 공지되어 있다. 수크로스 수송체 활성을 갖는 폴리펩티드의 예는, 유전자 cscB(서열 번호 23에 제시된 코딩 서열)에 의해 암호화되는, E. 콜라이 야생형 변종 EC3132로부터의 CscB(서열 번호 24에 제시됨); 유전자 cscB(서열 번호 25에 제시된 코딩 서열)에 의해 암호화되는, E. 콜라이 ATCC13281로부터의 CscB(서열 번호 26에 제시됨); 및 유전자 cscB(서열 번호 27에 제시된 코딩 서열)에 의해 암호화되는, 비피도박테리움 락티스로부터의 CscB(서열 번호 28에 제시됨)를 포함하나, 이에 한정되지 않는다. 수크로스 수송체 활성을 갖는 폴리펩티드 복합체의 예는, 유전자 susT1(서열 번호 29에 제시된 코딩 서열), susT2(서열 번호 31에 제시된 코딩 서열), 및 susX(서열 번호 33에 제시된 코딩 서열)에 의해 암호화되는, 서열 번호 30, 32, 및 34에 제시된 3개의 폴리펩티드 서브유닛을 포함하는, 스트렙토코커스 뉴모니아 변종 TIGR4로부터의 수크로스 ABC-유형 수송체 복합체; 및 각각 유전자 malE(서열 번호 35에 제시된 코딩 서열), malF(서열 번호 37에 제시된 코딩 서열), malG(서열 번호 39에 제시된 코딩 서열), 및 malK(서열 번호 41에 제시된 코딩 서열)에 의해 암호화되는, 서열 번호 36, 38, 40, 및 42에 제시된 4개의 폴리펩티드 서브유닛을 포함하는, 스트렙토코커스 뮤탄스의 말토스 수송체 복합체를 포함하나, 이에 한정되지 않는다.
- [0128] 일 실시 양태에서, 수크로스 수송체 활성을 갖는 폴리펩티드는, 서열 번호 24, 서열 번호 26, 또는 서열 번호 28에 제시된 아미노산 서열에 대해, 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는다.
- [0129] 다른 실시 양태에서, 수크로스 수송체 활성을 갖는 폴리펩티드 복합체는, 서열 번호 30에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제1 서브유닛; 서열 번호 32에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제2 서브유닛; 및 서열 번호 34에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제3 서브유닛을 포함한다.
- [0130] 다른 실시 양태에서, 수크로스 수송체 활성을 갖는 폴리펩티드 복합체는, 서열 번호 36에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제1 서브유닛; 서열 번호 38에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제2 서브유닛; 및 서열 번호 40에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제3 서브유닛; 및 서열 번호 42에 제시된 아미노산 서열에 비교할 때 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는 제4 서브유닛을 포함한다.
- [0131] 다른 실시 양태에서, 수크로스 수송체 활성을 갖는 폴리펩티드는 서열 번호 26에 제시된 아미노산 서열에 실질적으로 상응한다.
- [0132] 프럭토키나제 활성을 갖는 폴리펩티드는 프럭토키나제(EC 2.7.1.4로 표기) 및 프럭토스 인산화 활성을 갖는 다양한 헥소스 키나제(hexose kinase)(EC 2.7.1.3 및 EC 2.7.1.1)를 포함한다. 프럭토스 인산화 활성은 헥소

키나제(hexokinase) 및 케토헥소키나제(ketohexokinase)에 의해 나타날 수 있다. 본 명세서에 개시된 재조합 박테리아를 작제하기 위해 사용될 수 있는, 다양한 미생물로부터의 폴리펩티드를 암호화하는 대표적인 유전자들은 표 1에 열거되어 있다. 프럭토스를 인산화할 수 있는 단백질과 실질적으로 유사한 단백질(예를 들어, 표 1에 열거된 유전자에 의해 암호화되는 것) 또한 사용할 수 있다는 것을, 당업자는 알 것이다.

표 1

프럭토키나제 활성을 가진 효소를 암호화하는 서열

공급원	유전자 명칭	EC 번호	뉴클레오타이드 서열 번호	단백질 서열 번호
아그로박테리움 튜메파시엔스 스트렙토코커스	scrK(프럭토키나제)	2.7.1.4	43	44
뮤탄스 에스케리키아 콜라이 클렙시엘라 뉴모니아 에스케리키아 콜라이 엔테로코커스 파칼리스	scrK(프럭토키나제)	2.7.1.4	45	46
사카로마이세스 세레비시아	scrK(프럭토키나제)	2.7.1.4	84	85
사카로마이세스 세레비시아	scrK(프럭토키나제)	2.7.1.4	86	87
사카로마이세스 세레비시아	scrK(프럭토키나제)	2.7.1.4	47	48
사카로마이세스 세레비시아	cscK(프럭토키나제)	2.7.1.4	49	50
사카로마이세스 세레비시아	cscK(프럭토키나제)	2.7.1.4	51	52
사카로마이세스 세레비시아	HXK1(헥소키나제)	2.7.1.1	53	54
사카로마이세스 세레비시아	HXK2(헥소키나제)	2.7.1.1		

[0133]

[0134]

일 실시 양태에서, 프럭토키나제 활성을 갖는 폴리펩티드는, 서열 번호 44, 서열 번호 46, 서열 번호 48, 서열 번호 50, 서열 번호 52, 서열 번호 54, 서열 번호 85, 또는 서열 번호 87에 제시된 아미노산 서열에 대해, 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는다.

[0135]

다른 실시 양태에서, 프럭토키나제 활성을 갖는 폴리펩티드는, 서열 번호 48에 제시된 서열에 실질적으로 상응한다.

[0136]

수크로스 하이드롤라제 활성을 갖는 폴리펩티드는, 수크로스를 가수분해하여 프럭토스 및 글루코스를 생산하는 촉매 작용을 하는 능력을 갖는다. 수크로스 하이드롤라제 활성을 갖는 폴리펩티드는 상기와 같이 공지되어 있으며, 유전자 cscA(서열 번호 55에 제시된 코딩 서열)에 의해 암호화되는, E. 콜라이 야생형 변종 EC3132로부터의 CscA(서열 번호 56에 제시됨), 유전자 cscA(서열 번호 57에 제시된 코딩 서열)에 의해 암호화되는, E. 콜라이 ATCC13821로부터의 CscA(서열 번호 58에 제시됨); 유전자 bfrA(서열 번호 59에 제시된 코딩 서열)에 의해 암호화되는, 비피도박테리움 락티스 변종 DSM 10140^T로부터의 BfrA(서열 번호 60에 제시됨); 유전자 SUC2(서열 번호 61에 제시된 코딩 서열)에 의해 암호화되는, 사카로마이세스 세레비시아로부터의 Suc2p(서열 번호 62에 제시됨); 유전자 scrB(서열 번호 63에 제시된 코딩 서열)에 의해 암호화되는, 코리네박테리움 글루타미쿰으로부터의 ScrB(서열 번호 64에 제시됨); 암호화 유전자의 코딩 서열이 서열 번호 65에 제시된, 류코노스톡 메센테로이데스 DSM 20193으로부터의 수크로스 포스포릴라제(서열 번호 66에 제시됨); 및 유전자 sucP(서열 번호 67에 제시된 코딩 서열)에 의해 암호화되는, 비피도박테리움 아돌레센티스 DSM 20083으로부터의 수크로스 포스포릴라제(서열 번호 68에 제시됨)를 포함하나, 이에 한정되지 않는다.

[0137]

일 실시 양태에서, 수크로스 하이드롤라제 활성을 갖는 폴리펩티드는, 서열 번호 56, 서열 번호 58, 서열 번호 60, 서열 번호 62, 서열 번호 64, 서열 번호 66, 또는 서열 번호 68에 제시된 아미노산 서열에 대해, 클러스탈 V 정렬 방법을 기준으로 적어도 95%의 서열 동일성을 갖는다.

[0138]

다른 실시 양태에서, 수크로스 하이드롤라제 활성을 갖는 폴리펩티드는 서열 번호 58에 제시된 아미노산 서열에 실질적으로 상응한다.

[0139]

수크로스 수송체 활성을 갖는 폴리펩티드 또는 폴리펩티드 복합체, 프럭토키나제 활성을 갖는 폴리펩티드, 및 수크로스 하이드롤라제 활성을 갖는 폴리펩티드를 암호화하는 유전자의 코딩 서열을 사용하여, 동일하거나 다른 미생물 종으로부터 상동 폴리펩티드를 암호화하는 뉴클레오타이드 서열을 단리할 수 있다. 예를 들어, BLASTN과 같은 서열 분석 소프트웨어를 사용하여 공중 이용이 가능한 핵산 서열 데이터 베이스를 검색함으로써

써, 유전자의 상동체를 동정할 수 있다. 부가적으로, 서열-의존성 프로토콜을 사용하는 상동 유전자의 단리는 당업계에 주지되어 있다. 서열-의존성 프로토콜의 예는, 핵산 혼성화 방법, 및 핵산 증폭 기술의 다양한 용도에 의해 예시되는 DNA 및 RNA 증폭 방법(예를 들어, 폴리머라제 연쇄 반응(PCR), 미국 특허 제4,683,202호(Mullis et al.); 리가아제 연쇄 반응(LCR: ligase chain reaction), 문헌[Tabor, S. et al., Proc. Acad. Sci. USA 82, 1074, (1985)]; 또는 가닥 치환 증폭(SDA: strand displacement amplification), 문헌[Walker, et al., Proc. Natl. Acad. Sci. U.S.A., 89: 392, (1992)]을 포함하나, 이에 한정되지 않는다. 예를 들어, 상기 폴리펩티드를 암호화하는 뉴클레오티드 서열을 상동체의 동정을 위한 혼성화 탐침으로 채용할 수 있다.

[0140] 다른 공급원으로부터 단리된 이들 폴리펩티드를 암호화하는 유전자 또한 본 명세서에 개시된 재조합 박테리아에 사용할 수 있다는 것을, 당업자는 인식할 것이다. 부가적으로, 코돈 축퇴성으로 인해 암호화된 폴리펩티드의 아미노산 서열에 영향을 미치지 않으면서 폴리펩티드를 암호화하는 뉴클레오티드 서열에 변이를 만들 수 있으며, 실질적으로 유사한 단백질을 생산하는 그 아미노산 치환, 결실 또는 부가는 암호화된 단백질에 포함될 수 있다.

[0141] PCR(예를 들어, 미국 특허 제4,683,202호 참조) 및 목적하는 서열에 결합하도록 고안된 프라이머(이 서열이 공지된 경우)를 사용하여, 수크로스 수송체 활성을 갖는 폴리펩티드 또는 폴리펩티드 복합체, 프럭토키나제 활성을 갖는 폴리펩티드, 및 수크로스 하이드롤라제 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열을 단리할 수 있다. 축퇴 프라이머(degenerate primer) 또는 이중 탐침 혼성화의 사용에 의한 것과 같은 유전자 단리의 다른 방법이 당업자에게 주지되어 있다. 뉴클레오티드 서열은 또한, 화학적으로 합성하거나 DNA2.0 인코포레이티드(DNA2.0 Inc.)(캘리포니아주 멘로 파크 소재)와 같은 판매원으로부터 구매할 수 있다. 부가적으로, 본 명세서의 실시예 1에 상세하게 기술된 바와 같이, E. 콜라이 변종 ATCC13281의 게놈 DNA로부터 전체 csc 오페론을 단리할 수 있다.

[0142] 당업자에게 공지된 여러 방법 중 하나를 사용하여 폴리펩티드의 발현이 이루어질 수 있다. 예를 들어, 적어도 하나의 다중 카피 플라스미드 상에, 또는 코딩 서열의 하나 이상의 카피를 숙주 게놈 내로 통합시킴으로써, 상기 폴리펩티드를 암호화하는 뉴클레오티드 서열을 박테리아 내로 도입할 수 있다. 폴리펩티드를 암호화하는 뉴클레오티드 서열은, 별도로(예를 들어, 별도의 플라스미드 상에), 또는 임의의 조합으로(예를 들어, 본 명세서의 실시예에 기술된 바와 같이, 단일 플라스미드 상에), 숙주 박테리아 내로 도입할 수 있다. 숙주 박테리아가 폴리뉴클레오티드 중 하나를 암호화하는 유전자를 함유한다면, 나머지 뉴클레오티드 서열만 박테리아 내로 도입하면 된다. 예를 들어, 숙주 박테리아가 프럭토키나제 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열을 함유한다면, 수크로스 이용을 가능하게 하기 위해서는 수크로스 수송체 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열 및 수크로스 하이드롤라제 활성을 갖는 폴리펩티드를 암호화하는 뉴클레오티드 서열만 박테리아 내로 도입하면 된다. 플라스미드(들) 상에 있거나 게놈 내에 있는 도입된 코딩 영역은 적어도 하나의 고도로 활성인 프로모터로부터 발현될 수 있다. 통합된 코딩 영역은 그 자신의 프로모터를 갖는 키메라 유전자의 일부로서 도입될 수도 있고, 게놈에 내인성인 고도로 활성인 프로모터에 인접하여, 또는 고도로 발현되는 오페론 내에, 그것이 통합될 수도 있다. 적합한 프로모터는 CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, 및 lac, ara, tet, trp, IPL, IPR, T7, tac, 및 trc(에스케리키아 콜라이에서의 발현에 유용함)와 더불어 amy, apr, npr 프로모터 및 다양한 과아지 프로모터(바실러스에서의 발현에 유용함)를 포함하나, 이에 한정되지 않는다. 프로모터는 또한, 미국 특허 제7,132,527호(Payne et al.)에 기술된 스트렙토마이세스 리비단스 글루코스 아이소머라제 프로모터 또는 그의 변이체일 수 있다.

[0143] 일 실시 양태에서, 본 명세서에 개시된 재조합 박테리아는 글리세롤을 생산할 수 있다. 탄수화물 또는 당을 사용하여 글리세롤을 제조하기 위한 생물학적 방법은 효모 및 일부 박테리아, 다른 진균류, 및 조류에서 공지되어 있다. 박테리아 및 효모 양자 모두는, 해당 작용의 프럭토스-1,6-비스포스페이트 경로를 통해 글루코스 또는 다른 탄수화물을 전환함으로써 글리세롤을 생산한다. 본 명세서에 개시된 글리세롤 생산 방법에서는, 자연적으로 글리세롤을 생산하는 숙주 박테리아를 사용할 수 있다. 또한, 글리세롤 및 글리세롤 유도체의 생산을 위해 박테리아를 조작할 수 있다. 미국 특허 제7,005,291호에 기술된 바와 같이, 다양한 기질로부터 글리세롤을 생산하는 능력은 효소 활성 글리세롤-3-포스페이트 데하이드로게나제(G3PDH) 및/또는 글리세롤-3-포스파타제의 발현을 통해 제공될 수 있다. 숙주 박테리아에서 효소 활성을 발현시키기 위해 사용될 수 있는 이들 단백질을 암호화하는 유전자는 미국 특허 제7,005,291호에 기술되어 있다. 글리세롤-3-포스페이트 데하이드로게나제 활성을 갖는 폴리펩티드를 암호화하는 유전자의 적합한 예는 사카로마이세스 세레비시아로부터의 GPD1(서열 번호 1에 제시된 코딩 서열, 서열 번호 2에 제시된 암호화된 단백질 서열) 및 사카로마이세스 세레비시아로부터의 GPD2(서열 번호 3에 제시된 코딩 서열, 서열 번호 4에 제시된 암호화된 단백질 서열)를

포함하나, 이에 한정되지 않는다. 글리세롤-3-포스파타제 활성을 갖는 폴리펩티드를 암호화하는 유전자의 적합한 예는 사카로마이세스 세레비시아로부터의 GPP1(서열 번호 5에 제시된 코딩 서열, 서열 번호 6에 제시된 암호화된 단백질 서열) 및 사카로마이세스 세레비시아로부터의 GPP2(서열 번호 7에 제시된 코딩 서열, 서열 번호 8에 제시된 암호화된 단백질 서열)를 포함하나, 이에 한정되지 않는다.

[0144] 표적 내인성 유전자의 발현을 감소시킴으로써 글리세롤의 증가된 생산을 달성할 수 있다. 미국 특허 제 7,005,291호에 기술된 바와 같이, 글리세롤 키나제 및 글리세롤 데하이드로게나제 활성을 암호화하는 내인성 유전자의 하향 조절은 글리세롤 생산을 추가로 증진한다. 미국 특허 제 7,371,558호에 기술된 바와 같이, 글리세르알데히드 3-포스페이트 데하이드로게나제를 암호화하는 내인성 유전자의 발현을 감소시킴으로써 글리세롤로의 탄소의 증가된 채널링(channeling)이 이루어질 수 있다. 당업계에 공지된 임의의 방법, 예를 들어, PTS 시스템의 유전자의 하향 조절을 위한 상기 방법을 사용함으로써 하향 조절이 이루어질 수 있다.

[0145] 글리세롤은 유용한 산물의 미생물 생산을 위한 기질을 제공한다. 이러한 산물, 즉, 글리세롤 유도체의 예는 3-하이드록시프로피온산, 메틸글리옥살, 1,2-프로판다이올, 및 1,3-프로판다이올을 포함하나, 이에 한정되지 않는다.

[0146] 다른 실시 양태에서, 본 명세서에 개시된 재조합 박테리아는 1,3-프로판다이올을 생산할 수 있다. 글리세롤 유도체 1,3-프로판다이올은 폴리에스테르 섬유의 생산 및 폴리우레탄 및 환형 화합물의 제조에 있어서 잠재적 이용성을 갖는 단량체이다. 미국 특허 제 5,686,276호에 기술된 바와 같이, 1,3-프로판다이올은 글리세롤 또는 다이하이드록시아세톤이 아닌 탄소 기질의 생물전환을 통해 단일 미생물에 의해 생산될 수 있다. 이 생물 전환에서는, 상기과 같이 탄소 기질로부터 글리세롤이 생산된다. 숙주 박테리아에 의해 암호화될 수 있거나 재조합에 의해 숙주 내로 도입될 수 있는 데하이드라타제 효소에 의해 글리세롤이 중간체 3-하이드록시프로피온알데히드로 전환된다. 데하이드라타제는 글리세롤 데하이드라타제(E.C. 4.2.1.30), 다이올 데하이드라타제(E.C. 4.2.1.28) 또는 이 전환을 촉매 작용할 수 있는 임의의 다른 효소일 수 있다. 글리세롤 데하이드라타제의 "α"(알파), "β"(베타), 및 "γ"(감마) 서브유닛을 암호화하는 유전자의 적합한 예는, 각각 클렙시엘라 뉴모니아로부터의 dhaB1(서열 번호 9에 제시된 코딩 서열), dhaB2(서열 번호 11에 제시된 코딩 서열), 및 dhaB3(서열 번호 13에 제시된 코딩 서열)을 포함하나, 이에 한정되지 않는다. 3-하이드록시프로피온알데히드의 1,3-프로판다이올로의 추가의 전환은 1,3-프로판다이올 데하이드로게나제(E.C. 1.1.1.202) 또는 다른 알코올 데하이드로게나제에 의해 촉매될 수 있다. 1,3-프로판다이올 데하이드로게나제를 암호화하는 유전자의 적합한 예는 클렙시엘라 뉴모니아로부터의 dhaT(서열 번호 69에 제시된 코딩 서열, 서열 번호 70에 제시된 암호화된 단백질 서열)이다.

[0147] 박테리아를 재조합 조작하여 글리세롤 및 글리세롤 유도체 1,3-프로판다이올의 더욱 효율적인 생산을 제공할 수 있다. 예를 들어, 미국 특허 제 7,005,291호는 내인성 활성 글리세롤 키나제 및 글리세롤 데하이드로게나제 중 하나 또는 양자 모두를 파괴하면서 글리세롤-3-포스페이트 데하이드로게나제 및 글리세롤-3-포스페이트 포스파타제 중 하나 또는 양자 모두의 외인성 활성을 발현시킴으로부터 유래하는 이점을 가진, 형질전환된 미생물 및 글리세롤 및 1,3-프로판다이올의 생산 방법을 개시한다.

[0148] 미국 특허 제 6,013,494호는 외인성 글리세롤-3-포스페이트 데하이드로게나제, 글리세롤-3-포스페이트 포스파타제, 데하이드라타제, 및 1,3-프로판다이올 옥시도리덕타제(예를 들어, dhaT)를 포함하는 단일 미생물을 사용하는, 1,3-프로판다이올의 생산 방법을 기술한다. 미국 특허 제 6,136,576호는 데하이드라타제 및 단백질 X(데하이드라타제 재활성화 인자 펩티딘인 것으로 이후에 동정됨)를 추가로 포함하는 재조합 미생물을 포함하는 1,3-프로판다이올의 생산 방법을 개시한다.

[0149] 미국 특허 제 6,514,733호는, 3-하이드록시프로피온알데히드를 1,3-프로판다이올로 전환시키는 비-특이적 촉매 활성(dhaT에 의해 암호화되는 1,3-프로판다이올 옥시도리덕타제와는 구별됨)에 의해 타이터(titer)(리터당 그램 산물)에 유의적인 증가가 얻어지는 방법에 대한 개선을 기술한다. 부가적으로, 미국 특허 제 7,132,527호는 1,3-프로판다이올의 생산에 있어서 유용한 벡터 및 플라스미드를 개시한다.

[0150] 미국 특허 제 7,371,558호에 기술된 바와 같이, 숙주 박테리아에 대한 추가의 변형, 예를 들어 일부 표적 유전자의 발현의 하향 조절 및 다른 표적 유전자의 발현의 상향 조절에 의해 1,3-프로판다이올의 증가된 생산을 달성할 수 있다. PTS 음성 숙주에서 글루코스를 탄소 공급원으로 이용하기 위해서는, 글루코키나제 활성의 발현을 증가시킬 수 있다.

[0151] 유전자의 증가된 발현 또는 상향 조절된 발현이 1,3-프로판다이올 생산을 증가시키는 부가적인 유전자는 하기의 것들을 암호화하는 유전자를 포함한다:

- [0152] ? 전형적으로 EC 4.1.1.31로 특성화되는 포스포에놀피루베이트 카르복실라제
- [0153] ? 전형적으로 EC 2.5.1.17로 특성화되는 코브(I)알라민 아데노실트랜스페라제
- [0154] ? 3-HPA 및 1,3-프로판다이올의 상호 전환을 촉매하기에 충분하며, 특이적으로 1,3-프로판다이올 옥시도리덕 타제(들)를 배제하는 비-특이적 촉매 활성(전형적으로 이들 효소는 알코올 데하이드로게나제임)
- [0155] 유전자의 감소된 발현 또는 하향 조절된 발현이 1,3-프로판다이올 생산을 증가시키는 유전자는 하기의 것들을 암호화하는 유전자를 포함한다:
- [0156] ? 유기 호흡 제어 단백질
- [0157] ? 메틸글리옥살 신타제
- [0158] ? 아세테이트 키나제
- [0159] ? 포스포트랜스아세틸라제
- [0160] ? 알데히드 데하이드로게나제 A
- [0161] ? 알데히드 데하이드로게나제 B
- [0162] ? 트라이오스포스페이트 아이소머라제
- [0163] ? 포스포글루코네이트 데하이드라타제
- [0164] 다른 실시 양태에서, 본 명세서에 개시된 재조합 박테리아는 3-하이드록시프로피온산을 생산할 수 있다. 3-하이드록시프로피온산은 특수 합성에 있어서 이용성을 가지며, 화학 산업의 공지 기술에 의해 상업적으로 중요한 중간체로 전환될 수 있다(예를 들어, 탈수에 의한 아크릴산, 산화에 의한 말론산, 알코올과의 에스테르화 반응에 의한 에스테르, 및 환원에 의한 1,3-프로판다이올). 동시계속 중인 공유의 미국 특허 출원 제 61/187476호에 기술된 바와 같이, 3-하이드록시프로피온산은 단일 미생물에 의해 발효성 탄소 공급원으로부터 생물학적으로 생산될 수 있다. 1,3-프로판다이올의 생산에 관해 상기 기술된 바와 같이, 한가지 대표적인 생합성 경로에서는 탄소 기질이 3-하이드록시프로피온알데히드로 전환된다. 3-하이드록시프로피온알데히드는 알데히드 데하이드로게나제에 의해 3-하이드록시프로피온산으로 전환된다. 알데히드 데하이드로게나제의 적합한 예는 E. 콜라이 유전자 aldB(서열 번호 15에 제시된 코딩 서열)에 의해 암호화되는 AldB(서열 번호 16); E. 콜라이 유전자 aldA(서열 번호 17에 제시된 코딩 서열)에 의해 암호화되는 AldA(서열 번호 18); 및 E. 콜라이 유전자 aldH(서열 번호 19에 제시된 바와 같은 코딩 서열)에 의해 암호화되는 AldH(서열 번호 20)를 포함하나, 이에 한정되지 않는다.
- [0165] 재조합 박테리아에 의해 1,3-프로판다이올 생산을 개선하기 위한 상기 변형 중 다수는 또한, 3-하이드록시프로피온산 생산을 개선하게 만들 수 있다. 예를 들어, 글리세롤 키나제의 제거는 G3P 포스파타제의 작용에 의해 G3P로부터 형성되는 글리세롤이 ATP를 소모하면서 G3P로 재전환되는 것을 방지한다. 또한, 글리세롤 데하이드로게나제(예를 들어, gldA)의 제거는, NAD-의존성 글리세롤-3-포스페이트 데하이드로게나제의 작용에 의해 DHAP로부터 형성되는 글리세롤이 다이하이드록시아세톤으로 전환되는 것을 방지한다. 돌연변이는, 효소 활성의 활성을 손상시키거나 개선하도록 구조 유전자를 지향할 수도 있고, 효소 활성의 발현 수준을 조정하도록 프로모터 영역 및 리보솜 결합 부위를 포함하는 조절 유전자를 지향할 수도 있다.
- [0166] 상향 조절 또는 하향 조절은 당업자에게 공지된 다양한 방법에 의해 달성될 수 있다. 유전자의 상향 조절 또는 하향 조절은, 그 유전자에 의해 암호화되는 단백질로부터 유래하는 세포 내에 존재하는 활성 수준이, 예를 들어 상응하는(또는 변경되지 않은) 야생형 유전자에 의해 암호화되는 단백질의 활성에 의한 대조군 활성 수준에 비교하여 변경됨을 지칭한다는 것이 주지되어 있다.
- [0167] 효소 경로에 관련되는 특정 유전자를 상향 조절하여 그들의 암호화된 기능(들)의 활성을 증가시킬 수 있다. 예를 들어, 선택된 유전자의 부가적인 카피를 pBR322와 같은 다중 카피 플라스미드 상에 숙주 세포 내로 도입할 수 있다. 이러한 유전자는 또한, 그들의 암호화된 기능의 증가된 활성을 유발하는 적절한 조절 서열과 함께 염색체 내로 통합될 수 있다. 비-천연 프로모터 또는 변경된 천연 프로모터의 제어 하에 있도록 표적 유전자를 변형시킬 수 있다. 돌연변이, 결실, 및/또는 치환에 의해 내인성 프로모터를 생체내 변경할 수 있다.
- [0168] 대안적으로, 소정의 유전자의 발현을 주어진 활성 수준에 대해 감소시키거나 제거하는 것이 유용할 수 있다. 유전자를 하향 조절(파괴)하는 방법은 당업자에게 공지되어 있다.

- [0169] 하향 조절은 코딩 영역 및/또는 조절(프로모터) 영역의 결실, 삽입, 또는 변경에 의해 발생할 수 있다. 무작위 돌연변이 후의 스크리닝 또는 선택에 의해, 또는 유전자 서열이 공지된 경우에는 당업자에게 공지된 분자생물학 방법에 의한 직접 개입에 의해, 특정 하향 조절을 얻을 수 있다. 하향 조절이 이루어지게 하는 특히 유용하나 배타적이지 않은 방법은 프로모터 강도를 변경하는 것이다.
- [0170] 또한, 유전자 발현의 하향 조절을 사용하여 관심의 대상인 단백질의 발현을 방지하거나 비-기능성인 단백질의 발현을 유발할 수 있다. 이는 예를 들어, 1) 코딩 영역 및/또는 조절(프로모터) 영역의 결실, 2) 코딩 영역 및/또는 조절(프로모터) 영역 내로의 외인성 핵산 서열의 삽입, 및 3) 코딩 영역 및/또는 조절(프로모터) 영역의 변경(예를 들어, DNA 염기쌍을 변화시킴으로써)에 의해 이루어질 수 있다. 무작위 돌연변이 후의 스크리닝 또는 선택에 의해 특정 파괴를 얻을 수도 있거나, 유전자 서열이 공지된 경우에는 당업자에게 공지된 분자생물학 방법을 사용하는 직접 개입에 의해 특정 파괴를 얻을 수도 있다. 특히 유용한 방법은 유의적인 양의 코딩 영역 및/또는 조절(프로모터) 영역의 결실이다.
- [0171] 재조합 단백질 발현을 변경하는 방법은 당업자에게 공지되어 있으며, 문헌[Baneyx, Curr. Opin. Biotechnol. (1999) 10:411]; 문헌[Ross, et al., J. Bacteriol. (1998) 180:5375]; 문헌[deHaseth, et al., J. Bacteriol. (1998) 180:3019]; 문헌[Smolke and Keasling, Biotechnol. Bioeng. (2002) 80:762]; 문헌[Swartz, Curr. Opin. Biotech. (2001) 12:195]; 및 문헌[Ma, et al., J. Bacteriol. (2002) 184:5733]에 부분적으로 논의되어 있다.
- [0172] 글리세롤 및 글리세롤 유도체를 포함하는 미생물 산물의 생산에 있어서 수크로스를 대사시키기 위한 유전자 발현에 필요한 변화를 포함하는 재조합 박테리아는, 상기와 같이, 당업계에 주지된 기술을 사용하여 작제할 수 있으며, 이들 중 일부는 본 명세서의 실시예에 예시되어 있다.
- [0173] 본 명세서에 개시된 재조합 박테리아의 작제는, 적합한 숙주 미생물 내에서의 글리세롤 및 그의 유도체의 생산에 수크로스를 이용하는 능력을 부여하는 코딩 영역의 클로닝, 형질전환 및 발현에 적합한 다양한 벡터 및 형질전환 및 발현 카세트를 사용하여 이루어질 수 있다. 적합한 벡터는 채용되는 박테리아와 상용성인 것들이다. 적합한 벡터는, 예를 들어 박테리아, 바이러스(예를 들어, 박테리오파지 T7 또는 M-13 유래의 파아지), 코스미드, 효모 또는 식물로부터 유래될 수 있다. 이러한 벡터를 얻고 사용하기 위한 프로토콜은 당업자에게 공지되어 있다(Sambrook et al., supra).
- [0174] 목적하는 숙주 박테리아에서 본 발명을 위한 코딩 영역의 발현을 추진하기에 유용한 개시 제어 영역, 또는 프로모터는 다수이며 당업자에게 잘 알려져 있다. 발현을 추진할 수 있는 사실상 임의의 프로모터가 본 명세서에 사용하기에 적합하다. 예를 들어, 상기 열거된 프로모터 중 임의의 것을 사용할 수 있다.
- [0175] 종결 제어 영역 또한 바람직한 숙주의 다양한 천연의 유전자로부터 유래될 수 있다. 임의로, 종결 부위는 불필요할 수 있으나; 포함된다면 가장 바람직하다.
- [0176] 본 폴리펩티드의 효과적인 발현을 위하여, 폴리펩티드를 암호화하는 뉴클레오티드 서열을 선택된 발현 제어 영역에 개시 코돈을 통해 작동적으로 연결하여, 발현이 적절한 전령 RNA의 형성을 유발하도록 한다.
- [0177] 벡터 pSYCO101, pSYCO103, pSYCO106, 및 pSYCO109(미국 특허 제7,371,558호에 기술됨), 및 pSYCO400/AGRO(미국 특허 제7,524,660호에 기술됨)가 본 발명에 특히 유용하다. 이들 벡터의 필수 요소는 클램시엘라 뉴모니아 및 사카로마이세스 세레비시아로부터 단리된 dha 레귤론으로부터 유래된다. 각각의 벡터는 열린 해독틀 dhaB1, dhaB2, dhaB3, dhaX(서열 번호 71에 제시된 코딩 서열), orfX, DAR1, 및 GPP2(3개의 별도 오페론에 배열됨)를 함유한다. pSYCO101, pSYCO103, pSYCO106, pSYCO109, 및 pSYCO400/AGRO의 뉴클레오티드 서열은 각각 서열 번호 72, 서열 번호 73, 서열 번호 74, 서열 번호 75, 및 서열 번호 76에 제시되어 있다. 벡터들 사이의 상이점은 하기 차트에 예시되어 있다[접두어 "p-"는 프로모터를 나타내고; 각각의 "()" 내에 포함된 열린 해독틀은 오페론의 구성을 나타냄]:
- [0178] pSYCO101(서열 번호 72):
- [0179] 다른 2개의 경로 오페론에 비교하여 반대 배향인 p-trc(Dar1_GPP2),
- [0180] p-1.6 장 GI(dhaB1_dhaB2_dhaB3_dhaX), 및
- [0181] p-1.6 장 GI(orfY_orfX_orfW).
- [0182] pSYCO103(서열 번호 73):

- [0183] 다른 2개의 경로 오페론에 비교하여 동일 배향인 p-trc(Dar1_GPP2),
- [0184] p-1.5 장 GI(dhaB1_dhaB2_dhaB3_dhaX), 및
- [0185] p-1.5 장 GI(orfY_orfX_orfW).
- [0186] pSYCO106(서열 번호 74):
- [0187] 다른 2개의 경로 오페론에 비교하여 동일 배향인 p-trc(Dar1_GPP2),
- [0188] p-1.6 장 GI(dhaB1_dhaB2_dhaB3_dhaX), 및
- [0189] p-1.6 장 GI(orfY_orfX_orfW).
- [0190] pSYCO109(서열 번호 75):
- [0191] 다른 2개의 경로 오페론에 비교하여 동일 배향인 p-trc(Dar1_GPP2),
- [0192] p-1.6 장 GI(dhaB1_dhaB2_dhaB3_dhaX), 및
- [0193] p-1.6 장 GI(orfY_orfX).
- [0194] pSYCO400/AGRO(서열 번호 76):
- [0195] 다른 2개의 경로 오페론에 비교하여 동일 배향인 p-trc(Dar1_GPP2),
- [0196] p-1.6 장 GI(dhaB1_dhaB2_dhaB3_dhaX), 및
- [0197] p-1.6 장 GI(orfY_orfX).
- [0198] 경로 오페론에 비교하여 반대 배향의 p-1.20 단/장 GI(scrK).
- [0199] 일단 적합한 발현 카세트가 작제되면, 그들을 사용하여 적절한 숙주 박테리아를 형질전환한다. 코딩 영역을 함유하는 카세트의 숙주 박테리아 내로의 도입은, 형질전환에 의한 것(예를 들어, 칼슘-투과화 세포, 또는 전기천공을 사용함) 또는 재조합 파아지 바이러스를 사용하는 형질감염에 의한 것(Sambrook et al., supra)과 같은 공지의 절차에 의해 이루어질 수 있다. 발현 카세트는 숙주 세포 내에서 안정한 플라스미드 상에 유지될 수 있다. 또한, 당업자에게 주지된 벡터 및 방법을 사용하는 상동 또는 무작위 재조합을 통해 발현 카세트를 숙주 박테리아의 게놈 내에 통합할 수 있다. 부위-특이적 재조합 시스템 또한 발현 카세트의 게놈 통합에 사용할 수 있다.
- [0200] 예시된 세포에 부가하여, 글리세롤 및/또는 그의 유도체를 포함하는 미생물 산물의 생산을 증진하도록 특이적으로 고안된 단일 또는 다중 돌연변이를 갖는 세포 또한 사용할 수 있다. 정상적으로 탄소 공급원료를 비-생산적 경로에 전용하거나, 유의적인 이화산물 억제를 나타내는 세포는, 이들 표현형 결함을 회피하도록 돌연변이화할 수 있다.
- [0201] 돌연변이체를 생성시키는 방법은 통상적이며 당업계에 주지되어 있다. 일부 방법의 요약이 미국 특허 제 7,371,558호에 제공되어 있다. 방사선 또는 화학 약제를 사용하여 돌연변이체를 생성시키는 특정 방법이 당업계에 잘 입증되어 있다. 예를 들어, 문헌[Thomas D. Brock, Biotechnology: A Textbook of Industrial Microbiology, Second Edition(1989) Sinauer Associates, Inc., Sunderland, MA.] 또는 문헌[Deshpande, Mukund V., Appl. Biochem. Biotechnol. 36, 227 (1992)]을 참조한다.
- [0202] 돌연변이 유발이 발생한 후에, 목적하는 표현형을 갖는 돌연변이체를 다양한 방법으로 선택할 수 있다. 무작위 스크리닝이 가장 통상적이며, 여기서 돌연변이를 일으킨 세포는 목적하는 산물 또는 중간체를 생산하는 능력에 대하여 선택된다. 대안적으로, 돌연변이를 일으킨 개체군을 저항성 콜로니만 발생할 수 있는 선택적 배지 상에서 성장시킴으로써 돌연변이체의 선택적 단리를 수행할 수 있다. 돌연변이체 선택 방법은 고도로 개발되어 산업 미생물학 분야에 주지되어 있다. 예를 들어, 문헌[Brock, Supra; DeMancilha et al., Food Chem. 14, 313 (1984)]을 참조한다.
- [0203] 본 발명에서 발효 배지는 수크로스를 탄소 기질로 포함한다. 글루코스 및 프럭토스와 같은 다른 탄소 기질 또한 존재할 수 있다.
- [0204] 적합한 발효 배지는, 탄소 기질에 부가하여, 예를 들어, 글리세롤 및 그의 유도체, 예를 들어 1,3-프로판다이올의 생산을 위해 필요한 효소 경로의 촉진 및 배양물의 성장에 적합한, 당업자에게 공지된 적합한 미네랄,

염, 보조인자, 완충제 및 다른 컴포넌트를 함유한다. 1,3-프로판다이올의 생산에 있어서, Co(II) 염 및/또는 비타민 B₁₂ 또는 그의 전구체가 특히 주목 받는다.

[0205] 아데노실-코발라민(조효소 B₁₂)은 테하이드라타제 활성에 있어서 중요한 보조인자이다. 원핵 세포에서는 조효소 B₁₂의 합성이 발견되며, 이들 중 일부는 화합물의 드 노보(de novo) 합성이 가능하나(예를 들어, 에스케리키아 블라타이(*Escherichia blattae*), 클렙시엘라 종, 시트로박터 종, 및 클로스트리디움 종), 다른 것들은 부분 반응을 수행할 수 있다. 예를 들어, E. 콜라이는 코린 고리 구조를 제작할 수 없지만, 코빈아미드를 코리노이드로 전환시키는 촉매 작용을 할 수 있으며 5'-데옥시아데노실 기를 도입할 수 있다. 따라서, E. 콜라이 발효에는 비타민 B₁₂와 같은 조효소 B₁₂ 전구체를 공급할 필요가 있다는 것이 당업계에 공지되어 있다. 비타민 B₁₂는 E. 콜라이 발효에 일정한 속도로 연속적으로 첨가할 수도 있고 세포 매스의 생성과 일치하도록 단계화하거나, 단일 또는 다중 볼루스 첨가로 첨가할 수도 있다.

[0206] 본 명세서에 기술된 형질전환된 E. 콜라이에 비타민 B₁₂가 첨가되지만, 드 노보 비타민 B₁₂ 생합성이 가능한 다른 박테리아 또한 적합한 생산 세포일 것이며 이들 박테리아에는 비타민 B₁₂의 첨가가 불필요할 것으로 예상된다.

[0207] 전형적으로 박테리아 세포는 25 내지 40℃에서 수크로스를 함유하는 적절한 배지 내에서 성장한다. 본 명세서에 사용하기에 적합한 성장 배지의 예는, 루리아 베르타니(LB: Luria Bertani) 브로스, 사보로드 텍스트로스(SD: Sabouraud Dextrose) 브로스 또는 효모 배지(YM: Yeast medium) 브로스와 같은 상업적으로 제조된 통상적인 배지이다. 그밖에 정의되거나 합성된 성장 배지도 사용할 수 있으며, 특정 박테리아의 성장에 적절한 배지는 미생물학 또는 발효과학 분야의 당업자에게 공지되어 있을 것이다. 직접적으로 또는 간접적으로 이화산물 억제제를 조정하는 것으로 공지된 약제, 예를 들어 환형 아데노신 2',3'-모노포스페이트의 사용 또한 반응 배지 내에 도입될 수 있다. 마찬가지로, 1,3-프로판다이올 생산의 증진을 유발하는 효소 활성을 조정하는 것으로 공지된 약제(예를 들어, 메틸 비올로겐)의 사용을, 1,3-프로판다이올 생산 변종을 이용하는 유전자 조작과 함께, 또는 그 대안으로 사용할 수 있다.

[0208] 발효를 위해 적합한 pH 범위는 pH 5.0 내지 pH 9.0이며, 여기서 pH 6.0 내지 pH 8.0이 초기 조건으로서 전형적이다.

[0209] 재조합 박테리아의 요구 조건에 따라, 호기성, 무산소, 또는 혐기성 조건 하에 반응을 수행할 수 있다. 유가식 발효는 한정되거나 과량인 탄소 공급, 예를 들어 탄소 기질을 사용하여 수행될 수 있다.

[0210] 회분식 발효는 통상적으로 사용되는 방법이다. 고전적인 회분식 발효는, 배지의 조성이 발효 시작시에 설정되고 발효 중에 인공적으로 변경되지 않는 폐쇄 시스템이다. 따라서, 목적하는 박테리아가 발효의 시작시에 배지에 접종되며, 시스템에 아무것도 첨가하지 않고서 발효가 일어나게 한다. 그러나, 전형적으로 "회분식" 발효는 탄소 공급원의 첨가와 관련하여 회분식이며, pH 및 산소 농도와 같은 인자를 제어하려는 시도가 흔히 이루어진다. 회분식 시스템에서, 시스템의 대사산물 및 바이오매스 조성은 발효가 중단되는 시점까지 끊임없이 변화한다. 회분식 배양 내에서, 세포는 정적 유도기(static lag phase)를 통하여 고 성장 대수기(log phase)로, 그리고 최종적으로 성장 속도가 감소되거나 정지되는 정체기(stationary phase)로 완화된다. 미처리될 경우, 정체기의 세포는 결국 사멸할 것이다. 대수기의 세포는 일반적으로 최종 산물 또는 중간체의 대량 생산을 담당한다.

[0211] 표준 회분식 시스템의 한 변형은 유가식(Fed-Batch) 시스템이다. 유가식 발효 방법 또한 본 명세서에 사용하기에 적합하며, 이는 발효가 진행됨에 따라 기질을 증분식으로 첨가한다는 점을 제외하고는 전형적인 회분식 시스템을 포함한다. 유가식 시스템은 이화산물 억제제가 세포의 대사작용을 저해하는 경향이 있을 때 그리고 배지 중에 한정된 양의 기질이 있는 것이 바람직한 경우 유용하다. 유가식 시스템 내의 실제 기질 농도의 측정은 어려우므로, pH, 용존 산소 및 CO₂와 같은 폐 가스(waste gas)의 분압과 같이 측정가능한 인자의 변화를 기초로 추산된다. 회분식 및 유가식 발효는 통상적이며 당업계에 주지되어 있고, 그 예는 문헌[Brock, supra]에서 발견할 수 있다.

[0212] 연속식 발효는 정의된 발효 배지가 연속적으로 생물 반응기에 첨가되고, 동일한 양의 적응용 배지가 프로세싱을 위하여 동시에 제거되는 개방 시스템이다. 연속식 발효는 일반적으로, 세포가 주로 대수기 성장에 존재하는 경우 일정한 고밀도에서 배양물을 유지시킨다.

- [0213] 연속식 발효는 세포 성장 또는 최종 산물 농도에 영향을 주는 하나의 인자 또는 임의의 개수의 인자의 조정을 허용한다. 예를 들어, 한 가지 방법은 한정 영양소, 예를 들어 탄소 공급원 또는 질소의 수준을 고정된 비율로 유지하고 다른 모든 파라미터는 조정되게 할 것이다. 다른 시스템에서는, 배지의 탁도로 측정되는 세포 농도를 일정하게 유지하는 가운데, 성장에 영향을 주는 다수의 인자를 연속적으로 변경할 수 있다. 연속식 시스템은 안정 상태 성장 조건을 유지하도록 노력하므로, 배출되는 배지로 인한 세포 손실은 발효에서 세포 성장 속도에 대하여 균형이 맞추어져야 한다. 연속식 발효 방법에 있어서 영양소 및 성장 인자를 조정하는 방법뿐만 아니라, 산물 형성 속도를 최대화하는 기술도 산업 미생물학 분야에서 주지되어 있으며, 다양한 방법이 문헌[Brock, supra]에 상술되어 있다.
- [0214] 본 발명은 회분식, 유가식 또는 연속식 방법을 사용하여 실시할 수 있으며 임의의 공지 발효 모드가 적합할 것으로 예상된다. 부가적으로, 세포를 기재 상에 전체 세포 촉매로서 고정시키고 글리세롤 및 글리세롤 유도체, 예를 들어 1,3-프로판다이올의 생산을 위한 발효 조건을 적용할 수 있을 것으로 예상된다.
- [0215] 일 실시 양태에는, 수크로스로부터 글리세롤, 1,3-프로판다이올, 및/또는 3-하이드록시프로피온산을 제조하는 방법이 제공된다. 본 방법은, 상기와 같이 수크로스의 존재 하에 재조합 박테리아를 배양하는 단계, 및 임의로 생산된 글리세롤, 1,3-프로판다이올, 및/또는 3-하이드록시프로피온산을 회수하는 단계를 포함한다. 당업계에 공지된 방법을 사용하여 산물을 회수할 수 있다. 예를 들어, 원심분리, 여과, 경사분리(decantation) 등에 의해 발효 배지로부터 고체를 제거할 수 있다. 이어서, 증류, 액-액 추출 또는 막-기체의 분리와 같은 방법을 사용하여, 상기와 같이 고체를 제거하기 위해 처리된 발효 배지로부터 산물을 단리할 수 있다.
- [0216] 실시예
- [0217] 본 발명은 추가로 하기 실시예에서 정의된다. 이들 실시예는 본 발명의 바람직한 실시 양태를 나타내는 한편, 단지 예시로 주어짐을 이해해야 한다. 당업자라면, 상기 논의 및 이러한 실시예로부터 본 발명의 본질적인 특징을 확인할 수 있으며, 본 발명의 사상 및 범주를 벗어나지 않고서도 본 발명을 다양하게 변경하고 수정하여 본 발명을 다양한 용도 및 조건에 적합하게 할 수 있다.
- [0218] 일반 방법
- [0219] 실시예에 기재된 표준 재조합 DNA 및 분자 클로닝 기술은 당업계에 주지되어 있고, 문헌[Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, (1989) (Maniatis)] 및 문헌[T. J. Silhavy, M. L. Bannan, and L. W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984)] 및 문헌[Ausubel, F. M. et al., Current Protocols in Molecular Biology, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987)]에 기술되어 있다.
- [0220] 박테리아 배양물의 유지 및 성장을 위해 적합한 재료 및 방법은 당업계에 주지되어 있다. 하기 실시예에서 사용하기에 적합한 기술들은 문헌[Manual of Methods for General Bacteriology, Philipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds, American Society for Microbiology, Washington, DC.(1994)]; 또는 문헌[Thomas D. Brock, Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Sinauer Associates, Inc., Sunderland, MA(1989)]에 정리된 바와 같이 찾을 수 있다. 박테리아 세포의 유지 및 성장에 관해 기술된 모든 시약, 제한 효소 및 재료는 알드리치 케미칼스(Aldrich Chemicals)(위스콘신주 밀워키 소재), BD 다이아그노스틱 시스템즈(BD Diagnostic Systems)(메릴랜드주 스팅스 소재), 라이프 테크놀로지스(Life Technologies)(메릴랜드주 락빌 소재), 뉴 잉글랜드 바이오랩스(New England Biolabs)(메사추세츠주 베벌리 소재), 또는 시그마 케미칼 컴퍼니(Sigma Chemical Company)(미주리주 세인트 루이스 소재)로부터 입수할 수 있다.
- [0221] 약어의 의미는 하기와 같다: "s"는 초(들)를 의미하고, "min"은 분(들)을 의미하며, "h"는 시간(들)을 의미하고, "nm"은 나노미터를 의미하며, " μ l"은 마이크로리터(들)를 의미하고, "ml"은 밀리리터(들)를 의미하며, "L"은 리터(들)를 의미하고, "mM"은 밀리몰 농도를 의미하며, "M"은 몰 농도를 의미하고, "g"는 그램(들)을 의미하며, " μ g"는 마이크로그램(들)을 의미하고, "bp"는 염기쌍(들)을 의미하며, "kbp"는 킬로염기쌍(들)을 의미하고, "rpm"은 분당 회전을 의미하며, "ATCC"는 버지니아주 머내서스 소재의 미국 미생물 보존센터를 의미하고, "dH₂O"는 증류수를 의미한다.

[0222] 실시예 1

- [0223] csc 오페론 발현 플라스미드의 작제
- [0224] 본 실시예는 csc 오페론 발현 플라스미드 pBHRcscBKA 및 pBHRcscBKAmutB의 작제를 예시한다.
- [0225] E. 콜라이 변종 ATCC13281로부터 게놈 DNA를 단리하고 EcoRI 및 BamHI으로 절단하였다. 트리스-보레이트-EDTA 아가로스 겔 전기영동으로 길이가 대략 4 kbp인 단편을 단리하고, EcoRI 및 BamHI으로 절단한 플라스미드 벡터 pLitmus28(뉴 잉글랜드 바이오랩스, 메사추세츠주 베벌리 소재)과 함께 결합하였다. 생성된 플라스미드를 사용하여 E. 콜라이 변종 DH5알파(캘리포니아주 캄즈배드 소재의 인비트로젠(Invitrogen))를 형질전환시키고, 100 µg/ml 암피실린을 함유하는 맥콘키(MacConkey) 수크로스 아가(메릴랜드주 스팅스 소재의 디프코(Difco)로부터의 맥콘키 아가 베이스) 상에서의 성장에 의해 수크로스 이용에 필요한 유전자를 함유하는 형질전환체를 동정하였다. 수크로스를 대사시키는 능력을 획득한 콜로니로부터 플라스미드 DNA를 단리하고, 플라스미드(pScr1로 표기; 서열 번호 77에 제시됨)를 서열분석하여 수크로스 이용에 필요한 DNA 영역을 동정하였다. 삽입체는 길이가 4140 bp였으며, 공지된 E. 콜라이 수크로스 이용 유전자 cscB, cscK, 및 cscA(Jahreis et al., J. Bacteriol. 184:5307-5316, 2002)에 대해 상동인 추정상의 열린 해독틀을 함유하였다.
- [0226] 이어서, 하기의 절차를 사용하여 csc 오페론을 플라스미드 pBHR1(독일 괴팅겐 소재의 모비텍 GmbH(MobiTec GmbH))으로 옮겼다. 플라스미드 pScr1을 XhoI으로 절단하고 클레나우 단편(Klenow fragment)으로 처리하여 평활 단부를 생성시킨 후, AgeI으로 절단하였다. 생성된 csc 유전자를 함유하는 4175 bp 단편을 겔 정제에 의해 단리하였다. 플라스미드 pBHR1을 AgeI 및 NaeI으로 절단하고, 생성된 5142 bp 단편을 겔 정제에 의해 단리하였다. 이어서, 겔 정제된 2개의 단편을 결합하고, 생성된 플라스미드를 사용하여 E. 콜라이 변종 DH5알파를 형질전환시켰다. 50 µg /ml 가나마이신을 함유하는 루리아 베르타니(LB) 아가 상에서의 성장에 의해 형질전환체를 선택하였다. 형질전환체로부터 플라스미드 DNA를 단리하고, 플라스미드의 서열을 확인하였다. 플라스미드를 pBHRcscBKA(서열 번호 78에 제시됨)로 표기하였다.
- [0227] 스트래터진 퀵체인지(Stratagene QuikChange)(등록상표) 부위-지정 돌연변이 유발 키트(Site-Directed Mutagenesis Kit)(캘리포니아주 라 호야 소재의 스트래터진)를 사용하여 pBHRcscBKA 내에 단일 염기쌍 치환이 일어나게 함으로써 다른 발현 플라스미드를 생성시켰다. 위치 4263의 티민 염기를 구아노신으로 대체하고, 생성된 플라스미드를 pBHRcscBKAmutB(서열 번호 79에 제시됨)로 표기하였다. 이 치환은 cscB에 의해 암호화되는 폴리펩티드에서 글루타민 잔기를 히스티딘으로 대체시켰으며, 이 변화는 E. 콜라이 변종 EC3132로부터의 상동 단백질의 수송 능력을 변경하는 것으로 보고되었다(Jahreis et al., supra).
- [0228] 실시예 2 내지 4
- [0229] csc 오페론을 포함하는 재조합 E. 콜라이 변종의 작제
- [0230] 이들 실시예는 csc 오페론을 포함하는 플라스미드로 형질전환된 재조합 E. 콜라이 변종의 작제를 예시한다. 이들 재조합 변종에 의한 수크로스의 소모 및 수크로스로부터의 최종 산물 1,3-프로판다이올(PDO) 및 글리세롤의 생산을 설명하였다.
- [0231] E. 콜라이 변종 TTab pSYC0109
- [0232] 미국 특허 제7,371,558호(실시예 17)에 기술된 변종 TT aldA로부터 aldB 유전자를 결실시킴으로써 변종 TTab를 생성시켰다. 약술하면, 먼저 E. 콜라이 변종 MG1655(ATCC 번호 700926으로서 미국 미생물 보존센터로부터 입수가가능함) 내의 aldB의 코딩 영역의 1.5 kbp를 pKD3 플라스미드의 FRT-CmR-FRT 카세트(Datsenko and Wanner, Proc. Natl. Acad. Sci. USA 97:6640-6645, 2000)로 대체함으로써 aldB 결실을 일으켰다. pKD3를 주형으로 사용하여 프라이머 쌍 서열 번호 80 및 서열 번호 81로 대체 카세트를 증폭하였다. 프라이머 서열 번호 80은 aldB의 5'-단부에 대해 80 bp의 상동성, 및 pKD3에 대해 20 bp의 상동성을 포함한다. 프라이머 서열 번호 81은 aldB의 3' 단부에 대해 80 bp의 상동성, 및 pKD3에 대해 20 bp 상동성을 포함한다. PCR 산물을 겔-정제하고 MG1655/pKD46 수용 세포(competent cell)(미국 특허 제7,371,558호) 내로 전기천공시켰다. 재조합 변종을 12.5 mg/L의 클로로암페니콜을 가진 LB 플레이트 상에서 선택하였다. 프라이머 쌍 서열 번호 82 및 서열 번호 83을 사용하여, PCR에 의해 aldB 유전자의 결실을 확인하였다. 야생형 변종은 1.5 kbp PCR 산물을 제공한 반면에, 재조합 변종은 특징적인 1.1 kbp PCR 산물을 제공하였다. 돌연변이를 TT aldA 변종으로 이동시켜 TT aldAΔaldB::Cm 변종을 형성시키기 위하여 P1 용해물을 제조하고 사용하였다. 돌연변이가 존재함을 보장하기 위하여, 프라이머 쌍 서열 번호 82 및 서열 번호 83을 이용하는 게놈 PCR에 의해 클로람페니콜-내성 클론을 확인하였다. 변종 TTab를 생성시키기 위하여, FLP 리코비나제(FLP recombinase)를 사용하여 클로람페니콜 내성 표지자를 제거하였다(Datsenko and Wanner, supra). 이어서, 변종 TTab를 미국 특허 제7,371,558

호에 기술된 pSYC0109(서열 번호 75에 제시됨)로 형질전환시켜, 변종 TTab pSYC0109를 생성시켰다.

- [0233] 인용된 참고 문헌에 기술된 바와 같이, 변종 TTab는 하기의 변형을 포함하는 E. 콜라이 변종 FM5(ATCC No. 53911)의 유도체이다:
- [0234] glpK, gldA, ptsHI, crr, edd, arcA, mgsA, qor, ackA, pta, aldA 및 aldB 유전자의 결실;
- [0235] galP, glk, btuR, ppc, 및 yqhD 유전자의 상향 조절; 및
- [0236] gapA 유전자의 하향 조절.
- [0237] 플라스미드 pSYC0109는 글리세롤 생산 경로(DAR1 및 GPP2)를 암호화하는 유전자 및 글리세롤 데하이드라타제 및 연계된 재활성화 인자를 암호화하는 유전자(dhaB123, dhaX, orfX, orfY)를 함유한다.
- [0238] 실시예 1에 기술된 2개의 csc 오페론 과발현 플라스미드 pBHRcscBKA 및 pBHRcscBKAmutB 각각으로 변종 TTab/pSYC0109를 형질전환시켰다. 50 µg/ml의 스펙티노마이신 및 50 µg/ml의 가나마이신을 함유하는 LB 아가 상의 성장에 의해 형질전환체를 선택하였다. 개별적인 콜로니를 취하고 동일한 항생제를 가진 LB 브로스 내에 34°C에서 진탕하면서(250 rpm) 하룻밤 성장시켰다. 가나마이신을 제외하고는 동일한 조건 하에서 대조군 변종 TTab/pSYC0109를 성장시켰다.
- [0239] 이들 하룻밤 배양물을 10.5 g/L 수크로스를 함유하는 TM3 배지 내에 0.01 단위의 광학 밀도(550 nm에서 측정 함)까지 희석시켰다. TM3는 13.6 g/L KH₂PO₄, 2.04 g/L 시트르산 다이하이드레이트, 2 g/L 마그네슘 설페이트 헵타하이드레이트, 0.33 g/L 시트르산 암모늄 제2절, 0.5 g/L 효모 추출물, 3 g/L 암모늄 설페이트, 0.2 g/L CaCl₂·2H₂O, 0.03 g MnSO₄·2H₂O, 0.01 g/L NaCl, 1 mg/L FeSO₄·7H₂O, 1 mg/L, CoCl₂·6H₂O, 1 mg/L ZnSO₄·7H₂O, 0.1 mg/L CuSO₄·5H₂O, 0.1 mg/L H₃BO₃, 0.1 mg/L NaMoO₄·2H₂O 및 최종 pH 6.8을 제공하기에 충분한 NH₄OH를 함유하는 최소 배지이다. 비타민 B₁₂를 0.1 mg/L의 농도까지 배지에 첨가하였다. 배양물을 34°C에서 진탕하면서 (225 rpm) 24 시간 동안 인큐베이션하였다. 접종후 0, 5, 8, 11, 14, 17, 20 및 23 시간에 분취물을 수거하고, 브로스 내의 수크로스, 글리세롤 및 1,3-프로판다이올(PDO)의 농도를 고속 액체 크로마토그래피로 결정하였다.
- [0240] 아미넥스(Aminex) HPX-87P 컬럼(캘리포니아주 허클레스 소재의 바이오-라드(Bio-Rad))을 사용하여 dH₂O의 등 용매 이동상으로 0.5 ml/min의 유속 및 60°C의 컬럼 온도에서 크로마토그래피 분리를 달성하였다. TM3 배지에 기지 농도로 용해된 상업적으로 구매한 순수 화합물로부터 얻은 표준 곡선을 기준으로 굴절률 검출에 의해 용출된 화합물을 정량화하였다. 체류 시간은 12.2 min에 수크로스, 17.9 min에 1,3-프로판다이올, 그리고 23.6 min에 글리세롤이었다.
- [0241] 양자 모두의 csc 발현 플라스미드(실시예 3 및 4)가 수크로스의 대사작용 및 PDO 및 글리세롤의 생산을 유발한 반면에, 모체 대조군 변종(실시예 2, 비교)은 이들 조건(표 2 내지 4 참조) 하에서 수크로스를 대사시키거나 PDO 또는 글리세롤을 생산하지 못하였다. 표에 주어진 데이터 점들은 2개의 이중실험 배양물에 대해 이루어진 측정의 평균을 나타낸다.

표 2

수크로스 소모		수크로스 (g/L)	
시 간 (h)	실시예 2, 비교 대조군 변종	실시예 3 +pBHRcscBKA	실시예 4 +pBHRcscBKAmutB
0	10.48	10.48	10.48
6	10.14	10.05	10.08
12	10.34	9.87	10.17
18	10.28	7.31	10.17
24	10.32	0.65	10.13
30	10.37	0.00	8.44
36	10.36	0.00	3.32
42	10.33	0.00	0.00

[0242]

표 3

PDO 생산

시 간 (h)	P D O (g / L)		
	실 시 예 2, 비 교 대 조 군 변 종	실 시 예 3 + p B H R c s c B K A	실 시 예 4 + p B H R c s c B K A m u t B
0	0.00	0.00	0.00
6	0.00	0.00	0.00
12	0.00	0.00	0.00
18	0.00	0.41	0.00
24	0.00	2.20	0.02
30	0.00	3.15	0.24
36	0.00	3.15	1.35
42	0.00	3.06	2.82

[0243]

표 4

글리세롤 생산

시 간 (h)	글 리 세 롤 (g / L)		
	실 시 예 2, 비 교 대 조 군 변 종	실 시 예 3 + p B H R c s c B K A	실 시 예 4 + p B H R c s c B K A m u t B
0	0.00	0.00	0.00
6	0.00	0.00	0.00
12	0.00	0.08	0.00
18	0.00	0.96	0.00
24	0.00	3.24	0.00
30	0.00	2.54	0.60
36	0.00	2.52	2.33
42	0.00	2.49	2.98

[0244]

서 열 목 록

SEQUENCE LISTING

<110> E.I. du Pont de Nemours and Co.

Eliot, Andrew

Van Dyk, Tina

Gatenby, Anthony

<120> Recombinant Bacteria for Producing Glycerol and Glycerol-Derived
Products from Sucrose

<130> CL4246

<160> 87

<170> PatentIn version 3.4

<210> 1

<211> 1176

<212> DNA

<213> Saccharomyces cerevisiae

<400> 1

atgtctgctg ctgctgatag attaaactta acttccggcc acttgaatgc tggtagaaag 60

agaagttcct ctctgtttc ttggaaggct gccgaaaagc ctttcaaggt tactgtgatt 120

ggatctggta actgggggtac tactattgcc aagggtggtg ccgaaaattg taagggatac 180

ccagaagttt tcgctccaat agtacaaatg tgggtgttcg aagaagagat caatggtgaa 240

aaattgactg aaatcataaa tactagacat caaaacgtga aatacttgcc tggcatcact 300

ctaccgaca atttggttgc taatccagac ttgattgatt cagicaagga tgcgacatc 360

atcgttttca acattccaca tcaatttttg ccccgatatc gtagccaatt gaaaggtcat 420

gttgattcac acgtcagagc tatctcctgt ctaaagggtt ttgaagttgg tgctaaaggt 480

gtccaattgc tatcctctta catcactgag gaactaggta ttcaatgtgg tgctctatct 540

ggtgctaaca ttgccaccga agtcgctcaa gaacactggt ctgaaacaac agttgcttac 600

cacattccaa aggatttcag aggcgagggc aaggacgtcg accataaggt tctaaaggcc 660

ttgttccaca gaccttactt ccacgttagt gtcacgaag atgttgctgg tatctccatc 720

tgtggtgctt tgaagaacgt tgttgcctta ggttgtggtt tcgtcgaagg tctaggctgg 780

ggtaacaacg ctctgctgc catccaaaga gtcggtttgg gtgagatcat cagattcggg 840

caaatgtttt tcccagaatc tagagaagaa acatactacc aagagtctgc tgggtgttgc 900

gatttgatca ccacctgcgc tggtggtaga aacgtcaagg ttgctaggct aatggctact 960

tctggttaagg acgcctggga atgtgaaaag gagttgttga atggccaatc cgctcaaggt 1020

ttaattacct gcaaagaagt tcacgaatgg ttggaaacat gtggtctctg cgaagacttc 1080

ccattatttg aagccgtata ccaaatcgtt tacaacaact acccaatgaa gaacctgccg 1140

gacatgattg aagaattaga tctacatgaa gattag 1176

<210> 2

<211> 391

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 2

Met Ser Ala Ala Ala Asp Arg Leu Asn Leu Thr Ser Gly His Leu Asn

1 5 10 15

Ala Gly Arg Lys Arg Ser Ser Ser Val Ser Leu Lys Ala Ala Glu

20 25 30

Lys Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr Thr

35 40 45

Ile Ala Lys Val Val Ala Glu Asn Cys Lys Gly Tyr Pro Glu Val Phe

50 55 60
 Ala Pro Ile Val Gln Met Trp Val Phe Glu Glu Glu Ile Asn Gly Glu
 65 70 75 80

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

 Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Val Gly Ala Lys Gly
 145 150 155 160
 Val Gln Leu Leu Ser Ser Tyr Ile Thr Glu Glu Leu Gly Ile Gln Cys
 165 170 175
 Gly Ala Leu Ser Gly Ala Asn Ile Ala Thr Glu Val Ala Gln Glu His
 180 185 190
 Trp Ser Glu Thr Thr Val Ala Tyr His Ile Pro Lys Asp Phe Arg Gly
 195 200 205

 Glu Gly Lys Asp Val Asp His Lys Val Leu Lys Ala Leu Phe His Arg
 210 215 220
 Pro Tyr Phe His Val Ser Val Ile Glu Asp Val Ala Gly Ile Ser Ile
 225 230 235 240
 Cys Gly Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu
 245 250 255
 Gly Leu Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly
 260 265 270

 Leu Gly Glu Ile Ile Arg Phe Gly Gln Met Phe Phe Pro Glu Ser Arg
 275 280 285
 Glu Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr
 290 295 300

Thr Cys Ala Gly Gly Arg Asn Val Lys Val Ala Arg Leu Met Ala Thr
 305 310 315 320
 Ser Gly Lys Asp Ala Trp Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln
 325 330 335

Ser Ala Gln Gly Leu Ile Thr Cys Lys Glu Val His Glu Trp Leu Glu
 340 345 350
 Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln
 355 360 365
 Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu
 370 375 380
 Glu Leu Asp Leu His Glu Asp
 385 390

<210> 3

<211> 1323

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 3

atgcttgctg tcagaagatt aacaagatac acatttcctta agcgaacgca tccgggtgta	60
tatactcgtc gtgcatataa aattttgcct tcaagatcta ctttcctaag aagatcatta	120
ttacaaacac aactgcactc aaagatgact gtcatacta atatcaaaca gcacaaacac	180
tgtcatgagg accatcctat cagaagatcg gactctgccg tgtcaattgt acatttgaaa	240
cgtgcgccct tcaagggttac agtgattggt tctggtaact gggggaccac catcgccaaa	300
gtcattgcgg aaaacacaga attgcattcc catatcttcg agccagaggt gagaatgtgg	360
gtttttgatg aaaagatcgg cgacgaaaat ctgacggata tcataaatat aagacaccag	420
aacgttaaat atctaccaa tattgacctg ccccataatc tagtggccga tctgatctt	480
ttacactcca tcaagggtgc tgacatcctt gttttcaaca tccctcatca atttttacca	540
aacatagtca aacaattgca aggccacgtg gccctcatg taagggccat ctcgtgtcta	600
aaagggttcg agttgggctc caagggtgtg caattgctat cctctatgt tactgatgag	660
ttaggaatcc aatgtggcgc actatctggt gcaaacttgg caccggaagt ggccaaggag	720
cattggtccg aaaccaccgt ggcttaccaa ctaccaaagg attatcaagg tgatggcaag	780
gatgtagatc ataagatttt gaaattgctg ttccacagac cttacttcca cgtcaatgtc	840

atcgatgatg ttgctggat atccattgcc ggtgccttga agaacgtcgt ggcacttgca 900
 tgtggtttcg tagaaggat gggatggggt aacaatgcct cgcagccat tcaaaggctg 960
 ggttttagtg aaattatcaa gticggtaga atgtttttcc cagaatccaa agtcgagacc 1020
 tactatcaag aatccgctgg tggttcagat ctgacacca cctgctcagg cggtagaaac 1080
 gtcaaggttg ccacatacat ggccaagacc ggtaagtcag ccttgaagc agaaaaggaa 1140
 ttgcttaacg gtcaatccgc ccaagggata atcacatgca gagaagttca cgagtggcta 1200
 caaacatgtg agttgaccca agaattccca ttattcgagg cagtctacca gatagtctac 1260

aacaacgtcc gcatggaaga cctaccggag atgattgaag agctagacat cgatgacgaa 1320
 tag 1323

<210> 4

<211> 440

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 4

Met Leu Ala Val Arg Arg Leu Thr Arg Tyr Thr Phe Leu Lys Arg Thr

1 5 10 15

His Pro Val Leu Tyr Thr Arg Arg Ala Tyr Lys Ile Leu Pro Ser Arg

20 25 30

Ser Thr Phe Leu Arg Arg Ser Leu Leu Gln Thr Gln Leu His Ser Lys

35 40 45

Met Thr Ala His Thr Asn Ile Lys Gln His Lys His Cys His Glu Asp

50 55 60

His Pro Ile Arg Arg Ser Asp Ser Ala Val Ser Ile Val His Leu Lys

65 70 75 80

Arg Ala Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr

85 90 95

Thr Ile Ala Lys Val Ile Ala Glu Asn Thr Glu Leu His Ser His Ile

100 105 110

Phe Glu Pro Glu Val Arg Met Trp Val Phe Asp Glu Lys Ile Gly Asp

115 120 125

Glu Asn Leu Thr Asp Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr

130 135 140

Leu Pro Asn Ile Asp Leu Pro His Asn Leu Val Ala Asp Pro Asp Leu
 145 150 155 160
 Leu His Ser Ile Lys Gly Ala Asp Ile Leu Val Phe Asn Ile Pro His

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

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

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

 355 360 365
 Lys Thr Gly Lys Ser Ala Leu Glu Ala Glu Lys Glu Leu Leu Asn Gly
 370 375 380
 Gln Ser Ala Gln Gly Ile Ile Thr Cys Arg Glu Val His Glu Trp Leu

385 390 395 400
 Gln Thr Cys Glu Leu Thr Gln Glu Phe Pro Leu Phe Glu Ala Val Tyr
 405 410 415
 Gln Ile Val Tyr Asn Asn Val Arg Met Glu Asp Leu Pro Glu Met Ile

 420 425 430
 Glu Glu Leu Asp Ile Asp Asp Glu

 435 440

<210> 5

<211> 816

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 5

atgaaacgtt tcaatgtttt aaaatatatc agaacaacaa aagcaaatat acaaaccatc 60

gcaatgcctt tgaccacaaa acctttatct ttgaaaatca acgccgctct attcgatgtt 120

gacggtagca tcatcatctc tcaaccagcc attgctgctt tctggagaga tticggtaaa 180

gacaagcctt acttcgatgc cgaacacgtt attcacatct ctcacggttg gagaacttac 240

gatgccattg ccaagttcgc tccagacttt gctgatgaag aatacgttaa caagctagaa 300

ggtgaaatcc cagaaaaagta cggtgaacac tccatcgaag ttccaggtgc tgtcaagttg 360

tgtaatgctt tgaacgcctt gccaaaggaa aaatgggctg tcgccacctc tggtagccgt 420

gacatggcca agaaatgggt cgacattttg aagatcaaga gaccagaata ctcatcacc 480

gccaatgatg tcaagcaagg taagcctcac ccagaacat acttaaaggg tagaaacggt 540

ttgggtttcc caattaatga acaagacca tccaaatcta aggttggtgt ctttgaagac 600

gcaccagctg gtattgtgc tggttaaggct gctggctgta aaatcggttg tattgctacc 660

actttcgatt tggacttctt gaaggaaaag ggttgtgaca tcattgtcaa gaaccacgaa 720

tctatcagag tcggtgaata caacgtgaa accgatgaag tcgaattgat ctttgatgac 780

tacttatacg ctaaggatga cttgttgaaa tggtaa 816

<210> 6

<211> 271

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 6

Met Lys Arg Phe Asn Val Leu Lys Tyr Ile Arg Thr Thr Lys Ala Asn

1 5 10 15
 Ile Gln Thr Ile Ala Met Pro Leu Thr Thr Lys Pro Leu Ser Leu Lys
 20 25 30

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

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

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

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

Ile Phe Asp Asp Tyr Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp

260 265 270

<210> 7

<211> 753

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 7

atgggattga ctactaaacc tctatctttg aaagttaacg ccgctttgtt cgacgtcgac 60

ggtaccatta tcctctctca accagccatt gctgcattct ggagggattt cggttaaggac 120

aaaccttatt tcgatgctga acacgttatt caagtctcgc atggttggag aacgtttgat 180

gccattgcta agttcgctcc agactttgcc aatgaagagt atgttaacaa attagaagct 240

gaaattccgg tcaagtacgg tgaataatcc attgaagtcc cagggtcagt taagctgtgc 300

aacgtttga acgtctacc aaaagagaaa tgggtgtgtg caacttccgg taccctgat 360

atggcacaaa aatggttga gcctctggga atcaggagac caagtactt cattaccgt 420

aatgatgtca aacagggtaa gcctcatcca gaaccatatt tgaagggcag gaatggctta 480

ggatatccga tcaatgagca agacccttcc aaatctaagg tagtagtatt tgaagacgct 540

ccagcaggta ttgccgccgg aaaagccgcc ggttgaaga tcattggtat tgccactact 600

ttcgacttgg acttctctaaa ggaaaaaggc tgtgacatca ttgtcaaaaa ccacgaatcc 660

atcagagttg gcggtctaca tgccgaaaca gacgaagttg aattcatttt tgacgactac 720

ttatatgcta aggacgatct gttgaaatgg taa 753

<210> 8

<211> 250

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 8

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

1 5 10 15

Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln Pro Ala Ile Ala Ala

20 25 30

Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr Phe Asp Ala Glu His

35 40 45

Val Ile Gln Val Ser His Gly Trp Arg Thr Phe Asp Ala Ile Ala Lys

50 55 60
 Phe Ala Pro Asp Phe Ala Asn Glu Glu Tyr Val Asn Lys Leu Glu Ala

 65 70 75 80
 Glu Ile Pro Val Lys Tyr Gly Glu Lys Ser Ile Glu Val Pro Gly Ala

 85 90 95
 Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro Lys Glu Lys Trp Ala

 100 105 110
 Val Ala Thr Ser Gly Thr Arg Asp Met Ala Gln Lys Trp Phe Glu His

 115 120 125
 Leu Gly Ile Arg Arg Pro Lys Tyr Phe Ile Thr Ala Asn Asp Val Lys

 130 135 140
 Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys Gly Arg Asn Gly Leu

 145 150 155 160
 Gly Tyr Pro Ile Asn Glu Gln Asp Pro Ser Lys Ser Lys Val Val Val

 165 170 175
 Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly Lys Ala Ala Gly Cys

 180 185 190
 Lys Ile Ile Gly Ile Ala Thr Thr Phe Asp Leu Asp Phe Leu Lys Glu

 195 200 205
 Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu Ser Ile Arg Val Gly

 210 215 220
 Gly Tyr Asn Ala Glu Thr Asp Glu Val Glu Phe Ile Phe Asp Asp Tyr

 225 230 235 240
 Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp

 245 250

<210> 9

<211> 1668

<212> DNA

<213> *Klebsiella pneumoniae*

<220><221> CDS

<222> (1)..(1668)

<400> 9

atg aaa aga tca aaa cga ttt gca gta ctg gcc cag cgc ccc gtc aat	48
Met Lys Arg Ser Lys Arg Phe Ala Val Leu Ala Gln Arg Pro Val Asn	
1 5 10 15	
cag gac ggg ctg att ggc gag tgg cct gaa gag ggg ctg atc gcc atg	96
Gln Asp Gly Leu Ile Gly Glu Trp Pro Glu Glu Gly Leu Ile Ala Met	
20 25 30	
gac agc ccc ttt gac ccg gtc tct tca gta aaa gtg gac aac ggt ctg	144
Asp Ser Pro Phe Asp Pro Val Ser Ser Val Lys Val Asp Asn Gly Leu	
35 40 45	
atc gtc gaa ctg gac ggc aaa cgc cgg gac cag ttt gac atg atc gac	192
Ile Val Glu Leu Asp Gly Lys Arg Arg Asp Gln Phe Asp Met Ile Asp	
50 55 60	
cga ttt atc gcc gat tac gcg atc aac gtt gag cgc aca gag cag gca	240
Arg Phe Ile Ala Asp Tyr Ala Ile Asn Val Glu Arg Thr Glu Gln Ala	
65 70 75 80	
atg cgc ctg gag gcg gtg gaa ata gcc cgt atg ctg gtg gat att cac	288
Met Arg Leu Glu Ala Val Glu Ile Ala Arg Met Leu Val Asp Ile His	
85 90 95	
gtc agc cgg gag gag atc att gcc atc act acc gcc atc acg ccg gcc	336
Val Ser Arg Glu Glu Ile Ile Ala Ile Thr Thr Ala Ile Thr Pro Ala	
100 105 110	
aaa gcg gtc gag gtg atg gcg cag atg aac gtg gtg gag atg atg atg	384
Lys Ala Val Glu Val Met Ala Gln Met Asn Val Val Glu Met Met Met	
115 120 125	
gcg ctg cag aag atg cgt gcc cgc cgg acc ccc tcc aac cag tgc cac	432
Ala Leu Gln Lys Met Arg Ala Arg Arg Thr Pro Ser Asn Gln Cys His	
130 135 140	
gtc acc aat ctc aaa gat aat ccg gtg cag att gcc gct gac gcc gcc	480
Val Thr Asn Leu Lys Asp Asn Pro Val Gln Ile Ala Ala Asp Ala Ala	
145 150 155 160	

gag gcc ggg atc cgc ggc ttc tca gaa cag gag acc acg gtc ggt atc	528
Glu Ala Gly Ile Arg Gly Phe Ser Glu Gln Glu Thr Thr Val Gly Ile	
165 170 175	
gcg cgc tac gcg ccg ttt aac gcc ctg gcg ctg ttg gtc ggt tcg cag	576
Ala Arg Tyr Ala Pro Phe Asn Ala Leu Ala Leu Leu Val Gly Ser Gln	
180 185 190	
tgc ggc cgc ccc ggc gtg ttg acg cag tgc tcg gtg gaa gag gcc acc	624
Cys Gly Arg Pro Gly Val Leu Thr Gln Cys Ser Val Glu Glu Ala Thr	
195 200 205	
gag ctg gag ctg ggc atg cgt ggc tta acc agc tac gcc gag acg gtg	672
Glu Leu Glu Leu Gly Met Arg Gly Leu Thr Ser Tyr Ala Glu Thr Val	
210 215 220	
tcg gtc tac ggc acc gaa gcg gta ttt acc gac ggc gat gat acg ccg	720
Ser Val Tyr Gly Thr Glu Ala Val Phe Thr Asp Gly Asp Asp Thr Pro	
225 230 235 240	
tgg tca aag gcg ttc ctc gcc tcg gcc tac gcc tcc cgc ggg ttg aaa	768
Trp Ser Lys Ala Phe Leu Ala Ser Ala Tyr Ala Ser Arg Gly Leu Lys	
245 250 255	
atg cgc tac acc tcc ggc acc gga tcc gaa gcg ctg atg ggc tat tcg	816
Met Arg Tyr Thr Ser Gly Thr Gly Ser Glu Ala Leu Met Gly Tyr Ser	
260 265 270	
gag agc aag tcg atg ctc tac ctc gaa tcg cgc tgc atc ttc att act	864
Glu Ser Lys Ser Met Leu Tyr Leu Glu Ser Arg Cys Ile Phe Ile Thr	
275 280 285	
aaa ggc gcc ggg gtt cag gga ctg caa aac ggc gcg gtg agc tgt atc	912
Lys Gly Ala Gly Val Gln Gly Leu Gln Asn Gly Ala Val Ser Cys Ile	
290 295 300	
ggc atg acc ggc gct gtg ccg tcg ggc att cgg gcg gtg ctg gcg gaa	960
Gly Met Thr Gly Ala Val Pro Ser Gly Ile Arg Ala Val Leu Ala Glu	
305 310 315 320	
aac ctg atc gcc tct atg ctc gac ctc gaa gtg gcg tcc gcc aac gac	1008

Asn Leu Ile Ala Ser Met Leu Asp Leu Glu Val Ala Ser Ala Asn Asp	
325 330 335	
cag act ttc tcc cac tcg gat att cgc cgc acc gcg cgc acc ctg atg	1056
Gln Thr Phe Ser His Ser Asp Ile Arg Arg Thr Ala Arg Thr Leu Met	
340 345 350	
cag atg ctg ccg ggc acc gac ttt att ttc tcc ggc tac agc gcg gtg	1104
Gln Met Leu Pro Gly Thr Asp Phe Ile Phe Ser Gly Tyr Ser Ala Val	
355 360 365	
ccg aac tac gac aac atg ttc gcc ggc tcg aac ttc gat gcg gaa gat	1152
Pro Asn Tyr Asp Asn Met Phe Ala Gly Ser Asn Phe Asp Ala Glu Asp	
370 375 380	
ttt gat gat tac aac atc ctg cag cgt gac ctg atg gtt gac ggc ggc	1200
Phe Asp Asp Tyr Asn Ile Leu Gln Arg Asp Leu Met Val Asp Gly Gly	
385 390 395 400	
ctg cgt ccg gtg acc gag gcg gaa acc att gcc att cgc cag aaa gcg	1248
Leu Arg Pro Val Thr Glu Ala Glu Thr Ile Ala Ile Arg Gln Lys Ala	
405 410 415	
gcg cgg gcg atc cag gcg gtt ttc cgc gag ctg ggg ctg ccg cca atc	1296
Ala Arg Ala Ile Gln Ala Val Phe Arg Glu Leu Gly Leu Pro Pro Ile	
420 425 430	
gcc gac gag gag gtg gag gcc gcc acc tac gcg cac ggc agc aac gag	1344
Ala Asp Glu Glu Val Glu Ala Ala Thr Tyr Ala His Gly Ser Asn Glu	
435 440 445	
atg ccg ccg cgt aac gtg gtg gag gat ctg agt gcg gtg gaa gag atg	1392
Met Pro Pro Arg Asn Val Val Glu Asp Leu Ser Ala Val Glu Glu Met	
450 455 460	
atg aag cgc aac atc acc ggc ctc gat att gtc ggc gcg ctg agc cgc	1440
Met Lys Arg Asn Ile Thr Gly Leu Asp Ile Val Gly Ala Leu Ser Arg	
465 470 475 480	
agc ggc ttt gag gat atc gcc agc aat att ctc aat atg ctg cgc cag	1488

Ser Gly Phe Glu Asp Ile Ala Ser Asn Ile Leu Asn Met Leu Arg Gln
485 490 495
cgg gtc acc ggc gat tac ctg cag acc tcg gcc att ctc gat cgg cag 1536
Arg Val Thr Gly Asp Tyr Leu Gln Thr Ser Ala Ile Leu Asp Arg Gln
500 505 510
ttc gag gtg gtg agt gcg gtc aac gac atc aat gac tat cag ggg ccg 1584
Phe Glu Val Val Ser Ala Val Asn Asp Ile Asn Asp Tyr Gln Gly Pro
515 520 525
ggc acc ggc tat cgc atc tct gcc gaa cgc tgg gcg gag atc aaa aat 1632
Gly Thr Gly Tyr Arg Ile Ser Ala Glu Arg Trp Ala Glu Ile Lys Asn
530 535 540
att ccg ggc gtg gtt cag ccc gac acc att gaa taa 1668
Ile Pro Gly Val Val Gln Pro Asp Thr Ile Glu
545 550 555

<210> 10

<211> 555

<212> PRT

<213> Klebsiella pneumoniae

<400> 10

Met Lys Arg Ser Lys Arg Phe Ala Val Leu Ala Gln Arg Pro Val Asn
1 5 10 15
Gln Asp Gly Leu Ile Gly Glu Trp Pro Glu Glu Gly Leu Ile Ala Met
20 25 30
Asp Ser Pro Phe Asp Pro Val Ser Ser Val Lys Val Asp Asn Gly Leu
35 40 45
Ile Val Glu Leu Asp Gly Lys Arg Arg Asp Gln Phe Asp Met Ile Asp
50 55 60
Arg Phe Ile Ala Asp Tyr Ala Ile Asn Val Glu Arg Thr Glu Gln Ala
65 70 75 80
Met Arg Leu Glu Ala Val Glu Ile Ala Arg Met Leu Val Asp Ile His
85 90 95
Val Ser Arg Glu Glu Ile Ile Ala Ile Thr Thr Ala Ile Thr Pro Ala

100 105 110
 Lys Ala Val Glu Val Met Ala Gln Met Asn Val Val Glu Met Met Met

 115 120 125
 Ala Leu Gln Lys Met Arg Ala Arg Arg Thr Pro Ser Asn Gln Cys His
 130 135 140
 Val Thr Asn Leu Lys Asp Asn Pro Val Gln Ile Ala Ala Asp Ala Ala
 145 150 155 160
 Glu Ala Gly Ile Arg Gly Phe Ser Glu Gln Glu Thr Thr Val Gly Ile
 165 170 175
 Ala Arg Tyr Ala Pro Phe Asn Ala Leu Ala Leu Leu Val Gly Ser Gln

 180 185 190
 Cys Gly Arg Pro Gly Val Leu Thr Gln Cys Ser Val Glu Glu Ala Thr
 195 200 205
 Glu Leu Glu Leu Gly Met Arg Gly Leu Thr Ser Tyr Ala Glu Thr Val
 210 215 220
 Ser Val Tyr Gly Thr Glu Ala Val Phe Thr Asp Gly Asp Asp Thr Pro
 225 230 235 240
 Trp Ser Lys Ala Phe Leu Ala Ser Ala Tyr Ala Ser Arg Gly Leu Lys

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

 305 310 315 320
 Asn Leu Ile Ala Ser Met Leu Asp Leu Glu Val Ala Ser Ala Asn Asp
 325 330 335
 Gln Thr Phe Ser His Ser Asp Ile Arg Arg Thr Ala Arg Thr Leu Met
 340 345 350

Gln Met Leu Pro Gly Thr Asp Phe Ile Phe Ser Gly Tyr Ser Ala Val

355 360 365

Pro Asn Tyr Asp Asn Met Phe Ala Gly Ser Asn Phe Asp Ala Glu Asp

370 375 380

Phe Asp Asp Tyr Asn Ile Leu Gln Arg Asp Leu Met Val Asp Gly Gly

385 390 395 400

Leu Arg Pro Val Thr Glu Ala Glu Thr Ile Ala Ile Arg Gln Lys Ala

405 410 415

Ala Arg Ala Ile Gln Ala Val Phe Arg Glu Leu Gly Leu Pro Pro Ile

420 425 430

Ala Asp Glu Glu Val Glu Ala Ala Thr Tyr Ala His Gly Ser Asn Glu

435 440 445

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

450 455 460

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

465 470 475 480

Ser Gly Phe Glu Asp Ile Ala Ser Asn Ile Leu Asn Met Leu Arg Gln

485 490 495

Arg Val Thr Gly Asp Tyr Leu Gln Thr Ser Ala Ile Leu Asp Arg Gln

500 505 510

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

515 520 525

Gly Thr Gly Tyr Arg Ile Ser Ala Glu Arg Trp Ala Glu Ile Lys Asn

530 535 540

Ile Pro Gly Val Val Gln Pro Asp Thr Ile Glu

545 550 555

<210> 11

<211> 585

<212> DNA

<213> Klebsiella pneumoniae

<220><221> CDS

<

222> (1)..(585)

<400> 11

gtg caa cag aca acc caa att cag ccc tct ttt acc ctg aaa acc cgc 48

Val Gln Gln Thr Thr Gln Ile Gln Pro Ser Phe Thr Leu Lys Thr Arg

1 5 10 15

gag ggc ggg gta gct tct gcc gat gaa cgc gcc gat gaa gtg gtg atc 96

Glu Gly Gly Val Ala Ser Ala Asp Glu Arg Ala Asp Glu Val Val Ile

20 25 30

ggc gtc ggc cct gcc ttc gat aaa cac cag cat cac act ctg atc gat 144

Gly Val Gly Pro Ala Phe Asp Lys His Gln His His Thr Leu Ile Asp

35 40 45

atg ccc cat ggc gcg atc ctc aaa gag ctg att gcc ggg gtg gaa gaa 192

Met Pro His Gly Ala Ile Leu Lys Glu Leu Ile Ala Gly Val Glu Glu

50 55 60

gag ggg ctt cac gcc cgg gtg gtg cgc att ctg cgc acg tcc gac gtc 240

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

65 70 75 80

tcc ttt atg gcc tgg gat gcg gcc aac ctg agc ggc tcg ggg atc ggc 288

Ser Phe Met Ala Trp Asp Ala Ala Asn Leu Ser Gly Ser Gly Ile Gly

85 90 95

atc ggt atc cag tcg aag ggg acc acg gtc atc cat cag cgc gat ctg 336

Ile Gly Ile Gln Ser Lys Gly Thr Thr Val Ile His Gln Arg Asp Leu

100 105 110

ctg ccg ctc agc aac ctg gag ctg ttc tcc cag gcg ccg ctg ctg acg 384

Leu Pro Leu Ser Asn Leu Glu Leu Phe Ser Gln Ala Pro Leu Leu Thr

115 120 125

ctg gag acc tac cgg cag att ggc aaa aac gct gcg cgc tat gcg cgc 432

Leu Glu Thr Tyr Arg Gln Ile Gly Lys Asn Ala Ala Arg Tyr Ala Arg

130 135 140

aaa gag tca cct tcg ccg gtg ccg gtg gtg aac gat cag atg gtg cgg 480

Lys Glu Ser Pro Ser Pro Val Pro Val Val Asn Asp Gln Met Val Arg

145	150	155	160	
ccg aaa ttt atg gcc aaa gcc gcg cta ttt cat atc aaa gag acc aaa				528
Pro Lys Phe Met Ala Lys Ala Ala Leu Phe His Ile Lys Glu Thr Lys				
	165	170	175	
cat gtg gtg cag gac gcc gag ccc gtc acc ctg cac atc gac tta gta				576
His Val Val Gln Asp Ala Glu Pro Val Thr Leu His Ile Asp Leu Val				
	180	185	190	
agg gag tga				585
Arg Glu				

<210> 12

<211> 194

<212> PRT

<213> *Klebsiella pneumoniae*

<400> 12

Val Gln Gln Thr Thr Gln Ile Gln Pro Ser Phe Thr Leu Lys Thr Arg
1 5 10 15

Glu Gly Gly Val Ala Ser Ala Asp Glu Arg Ala Asp Glu Val Val Ile
20 25 30

Gly Val Gly Pro Ala Phe Asp Lys His Gln His His Thr Leu Ile Asp
35 40 45

Met Pro His Gly Ala Ile Leu Lys Glu Leu Ile Ala Gly Val Glu Glu

50 55 60

Glu Gly Leu His Ala Arg Val Val Arg Ile Leu Arg Thr Ser Asp Val
65 70 75 80

Ser Phe Met Ala Trp Asp Ala Ala Asn Leu Ser Gly Ser Gly Ile Gly

85 90 95

Ile Gly Ile Gln Ser Lys Gly Thr Thr Val Ile His Gln Arg Asp Leu
100 105 110

Leu Pro Leu Ser Asn Leu Glu Leu Phe Ser Gln Ala Pro Leu Leu Thr
115 120 125

Leu Glu Thr Tyr Arg Gln Ile Gly Lys Asn Ala Ala Arg Tyr Ala Arg

130 135 140

Lys Glu Ser Pro Ser Pro Val Pro Val Val Asn Asp Gln Met Val Arg

145 150 155 160

Pro Lys Phe Met Ala Lys Ala Ala Leu Phe His Ile Lys Glu Thr Lys

165 170 175

His Val Val Gln Asp Ala Glu Pro Val Thr Leu His Ile Asp Leu Val

180 185 190

Arg Glu

<210> 13

<211> 426

<212> DNA

<213> *Klebsiella pneumoniae*

<220><221> CDS

<222> (1)..(426)

<400> 13

atg agc gag aaa acc atg cgc gtg cag gat tat ccg tta gcc acc cgc 48

Met Ser Glu Lys Thr Met Arg Val Gln Asp Tyr Pro Leu Ala Thr Arg

1 5 10 15

tgc ccg gag cat atc ctg acg cct acc ggc aaa cca ttg acc gat att 96

Cys Pro Glu His Ile Leu Thr Pro Thr Gly Lys Pro Leu Thr Asp Ile

20 25 30

acc ctc gag aag gtg ctc tct ggc gag gtg ggc ccg cag gat gtg cgg 144

Thr Leu Glu Lys Val Leu Ser Gly Glu Val Gly Pro Gln Asp Val Arg

35 40 45

atc tcc cgc cag acc ctt gag tac cag gcg cag att gcc gag cag atg 192

Ile Ser Arg Gln Thr Leu Glu Tyr Gln Ala Gln Ile Ala Glu Gln Met

50 55 60

cag cgc cat gcg gtg gcg cgc aat ttc cgc cgc gcg gcg gag ctt atc 240

Gln Arg His Ala Val Ala Arg Asn Phe Arg Arg Ala Ala Glu Leu Ile

65 70 75 80

gcc att cct gac gag cgc att ctg gct atc tat aac gcg ctg cgc ccg 288
Ala Ile Pro Asp Glu Arg Ile Leu Ala Ile Tyr Asn Ala Leu Arg Pro
85 90 95

ttc cgc tcc tcg cag gcg gag ctg ctg gcg atc gcc gac gag ctg gag 336
Phe Arg Ser Ser Gln Ala Glu Leu Leu Ala Ile Ala Asp Glu Leu Glu
100 105 110

cac acc tgg cat gcg aca gtg aat gcc gcc ttt gtc cgg gag tcg gcg 384
His Thr Trp His Ala Thr Val Asn Ala Ala Phe Val Arg Glu Ser Ala
115 120 125

gaa gtg tat cag cag cgg cat aag ctg cgt aaa gga agc taa 426
Glu Val Tyr Gln Gln Arg His Lys Leu Arg Lys Gly Ser
130 135 140

<210> 14
<211> 141
<212> PRT
<213> *Klebsiella pneumoniae*

<400> 14
Met Ser Glu Lys Thr Met Arg Val Gln Asp Tyr Pro Leu Ala Thr Arg
1 5 10 15
Cys Pro Glu His Ile Leu Thr Pro Thr Gly Lys Pro Leu Thr Asp Ile
20 25 30
Thr Leu Glu Lys Val Leu Ser Gly Glu Val Gly Pro Gln Asp Val Arg
35 40 45
Ile Ser Arg Gln Thr Leu Glu Tyr Gln Ala Gln Ile Ala Glu Gln Met
50 55 60

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

115 120 125

Glu Val Tyr Gln Gln Arg His Lys Leu Arg Lys Gly Ser

130 135 140

<210> 15

<211> 1539

<212> DNA

<213> Escherichia coli

<220><221> CDS

<222> (1)..(1539)

<400> 15

atg acc aat aat ccc cct tca gca cag att aag ccc ggc gag tat ggt 48

Met Thr Asn Asn Pro Pro Ser Ala Gln Ile Lys Pro Gly Glu Tyr Gly

1 5 10 15

ttc ccc ctc aag tta aaa gcc cgc tat gac aac ttt att ggc ggc gaa 96

Phe Pro Leu Lys Leu Lys Ala Arg Tyr Asp Asn Phe Ile Gly Gly Glu

20 25 30

tgg gta gcc cct gcc gac ggc gag tat tac cag aat ctg acg ccg gtg 144

Trp Val Ala Pro Ala Asp Gly Glu Tyr Tyr Gln Asn Leu Thr Pro Val

35 40 45

acc ggg cag ctg ctg tgc gaa gtg gcg tct tcg ggc aaa cga gac atc 192

Thr Gly Gln Leu Leu Cys Glu Val Ala Ser Ser Gly Lys Arg Asp Ile

50 55 60

gat ctg gcg ctg gat gct gcg cac aaa gtg aaa gat aaa tgg gcg cac 240

Asp Leu Ala Leu Asp Ala Ala His Lys Val Lys Asp Lys Trp Ala His

65 70 75 80

acc tcg gtg cag gat cgt gcg gcg att ctg ttt aag att gcc gat cga 288

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

85 90 95

atg gaa caa aac ctc gag ctg tta gcg aca gct gaa acc tgg gat aac 336

Met Glu Gln Asn Leu Glu Leu Leu Ala Thr Ala Glu Thr Trp Asp Asn

100 105 110

ggc aaa ccc att cgc gaa acc agt gct gcg gat gta ccg ctg gcg att	384
Gly Lys Pro Ile Arg Glu Thr Ser Ala Ala Asp Val Pro Leu Ala Ile	
115 120 125	
gac cat ttc cgc tat ttc gcc tcg tgt att cgg gcg cag gaa ggt ggg	432
Asp His Phe Arg Tyr Phe Ala Ser Cys Ile Arg Ala Gln Glu Gly Gly	
130 135 140	
atc agt gaa gtt gat agc gaa acc gtg gcc tat cat ttc cat gaa ccg	480
Ile Ser Glu Val Asp Ser Glu Thr Val Ala Tyr His Phe His Glu Pro	
145 150 155 160	
tta ggc gtg gtg ggg cag att atc ccg tgg aac ttc ccg ctg ctg atg	528
Leu Gly Val Val Gly Gln Ile Ile Pro Trp Asn Phe Pro Leu Leu Met	
165 170 175	
gcg agc tgg aaa atg gct ccc gcg ctg gcg gcg ggc aac tgt gtg gtg	576
Ala Ser Trp Lys Met Ala Pro Ala Leu Ala Ala Gly Asn Cys Val Val	
180 185 190	
ctg aaa ccc gca cgt ctt acc ccg ctt tct gta ctg ctg cta atg gaa	624
Leu Lys Pro Ala Arg Leu Thr Pro Leu Ser Val Leu Leu Leu Met Glu	
195 200 205	
att gtc ggt gat tta ctg ccg ccg ggc gtg gtg aac gtg gtc aat ggc	672
Ile Val Gly Asp Leu Leu Pro Pro Gly Val Val Asn Val Val Asn Gly	
210 215 220	
gca ggt ggg gta att ggc gaa tat ctg gcg acc tcg aaa cgc atc gcc	720
Ala Gly Gly Val Ile Gly Glu Tyr Leu Ala Thr Ser Lys Arg Ile Ala	
225 230 235 240	
aaa gtg gcg ttt acc ggc tca acg gaa gtg ggc caa caa att atg caa	768
Lys Val Ala Phe Thr Gly Ser Thr Glu Val Gly Gln Gln Ile Met Gln	
245 250 255	
tac gca acg caa aac att att ccg gtg acg ctg gag ttg ggc ggt aag	816
Tyr Ala Thr Gln Asn Ile Ile Pro Val Thr Leu Glu Leu Gly Gly Lys	
260 265 270	
tcg cca aat atc ttc ttt gct gat gtg atg gat gaa gaa gat gcc ttt	864

Ser Pro Asn Ile Phe Phe Ala Asp Val Met Asp Glu Glu Asp Ala Phe

275 280 285

ttc gat aaa gcg ctg gaa ggc ttt gca ctg ttt gcc ttt aac cag ggc 912

Phe Asp Lys Ala Leu Glu Gly Phe Ala Leu Phe Ala Phe Asn Gln Gly

290 295 300

gaa gtt tgc acc tgt ccg agt cgt gct tta gtg cag gaa tct atc tac 960

Glu Val Cys Thr Cys Pro Ser Arg Ala Leu Val Gln Glu Ser Ile Tyr

305 310 315 320

gaa cgc ttt atg gaa cgc gcc atc cgc cgt gtc gaa agc att cgt agc 1008

Glu Arg Phe Met Glu Arg Ala Ile Arg Arg Val Glu Ser Ile Arg Ser

325 330 335

ggc aac ccg ctc gac agc gtg acg caa atg ggc gcg cag gtt tct cac 1056

Gly Asn Pro Leu Asp Ser Val Thr Gln Met Gly Ala Gln Val Ser His

340 345 350

ggg caa ctg gaa acc atc ctc aac tac att gat atc ggt aaa aaa gag 1104

Gly Gln Leu Glu Thr Ile Leu Asn Tyr Ile Asp Ile Gly Lys Lys Glu

355 360 365

ggc gct gac gtg ctc aca ggc ggg cgg cgc aag ctg ctg gaa ggt gaa 1152

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

370 375 380

ctg aaa gac ggc tac tac ctc gaa ccg acg att ctg ttt ggt cag aac 1200

Leu Lys Asp Gly Tyr Tyr Leu Glu Pro Thr Ile Leu Phe Gly Gln Asn

385 390 395 400

aat atg cgg gtg ttc cag gag gag att ttt ggc ccg gtg ctg gcg gtg 1248

Asn Met Arg Val Phe Gln Glu Glu Ile Phe Gly Pro Val Leu Ala Val

405 410 415

acc acc ttc aaa acg atg gaa gaa gcg ctg gag ctg gcg aac gat acg 1296

Thr Thr Phe Lys Thr Met Glu Glu Ala Leu Glu Leu Ala Asn Asp Thr

420 425 430

caa tat ggc ctg ggc gcg ggc gtc tgg agc cgc aac ggt aat ctg gcc 1344

Gln Tyr Gly Leu Gly Ala Gly Val Trp Ser Arg Asn Gly Asn Leu Ala
435 440 445
tat aag atg ggg cgc ggc ata cag gct ggg cgc gtg tgg acc aac tgt 1392
Tyr Lys Met Gly Arg Gly Ile Gln Ala Gly Arg Val Trp Thr Asn Cys
450 455 460
tat cac gct tac ccg gca cat gcg gcg ttt ggt ggc tac aaa caa tca 1440
Tyr His Ala Tyr Pro Ala His Ala Ala Phe Gly Gly Tyr Lys Gln Ser
465 470 475 480
ggg atc ggt cgc gaa acc cac aag atg atg ctg gag cat tac cag caa 1488
Gly Ile Gly Arg Glu Thr His Lys Met Met Leu Glu His Tyr Gln Gln
485 490 495
acc aag tgc ctg ctg gtg agc tac tcg gat aaa ccg ttg ggg ctg ttc 1536
Thr Lys Cys Leu Leu Val Ser Tyr Ser Asp Lys Pro Leu Gly Leu Phe
500 505 510
tga 1539
<210> 16
<211> 512
<212> PRT
<213> Escherichia coli
<400> 16
Met Thr Asn Asn Pro Pro Ser Ala Gln Ile Lys Pro Gly Glu Tyr Gly
1 5 10 15
Phe Pro Leu Lys Leu Lys Ala Arg Tyr Asp Asn Phe Ile Gly Gly Glu
20 25 30
Trp Val Ala Pro Ala Asp Gly Glu Tyr Tyr Gln Asn Leu Thr Pro Val
35 40 45
Thr Gly Gln Leu Leu Cys Glu Val Ala Ser Ser Gly Lys Arg Asp Ile
50 55 60
Asp Leu Ala Leu Asp Ala Ala His Lys Val Lys Asp Lys Trp Ala His
65 70 75 80
Thr Ser Val Gln Asp Arg Ala Ala Ile Leu Phe Lys Ile Ala Asp Arg
85 90 95

Met Glu Gln Asn Leu Glu Leu Leu Ala Thr Ala Glu Thr Trp Asp Asn

100 105 110

Gly Lys Pro Ile Arg Glu Thr Ser Ala Ala Asp Val Pro Leu Ala Ile

115 120 125

Asp His Phe Arg Tyr Phe Ala Ser Cys Ile Arg Ala Gln Glu Gly Gly

130 135 140

Ile Ser Glu Val Asp Ser Glu Thr Val Ala Tyr His Phe His Glu Pro

145 150 155 160

Leu Gly Val Val Gly Gln Ile Ile Pro Trp Asn Phe Pro Leu Leu Met

165 170 175

Ala Ser Trp Lys Met Ala Pro Ala Leu Ala Ala Gly Asn Cys Val Val

180 185 190

Leu Lys Pro Ala Arg Leu Thr Pro Leu Ser Val Leu Leu Leu Met Glu

195 200 205

Ile Val Gly Asp Leu Leu Pro Pro Gly Val Val Asn Val Val Asn Gly

210 215 220

Ala Gly Gly Val Ile Gly Glu Tyr Leu Ala Thr Ser Lys Arg Ile Ala

225 230 235 240

Lys Val Ala Phe Thr Gly Ser Thr Glu Val Gly Gln Gln Ile Met Gln

245 250 255

Tyr Ala Thr Gln Asn Ile Ile Pro Val Thr Leu Glu Leu Gly Gly Lys

260 265 270

Ser Pro Asn Ile Phe Phe Ala Asp Val Met Asp Glu Glu Asp Ala Phe

275 280 285

Phe Asp Lys Ala Leu Glu Gly Phe Ala Leu Phe Ala Phe Asn Gln Gly

290 295 300

Glu Val Cys Thr Cys Pro Ser Arg Ala Leu Val Gln Glu Ser Ile Tyr

305 310 315 320

Glu Arg Phe Met Glu Arg Ala Ile Arg Arg Val Glu Ser Ile Arg Ser

325 330 335

Gly Asn Pro Leu Asp Ser Val Thr Gln Met Gly Ala Gln Val Ser His

340 345 350
Gly Gln Leu Glu Thr Ile Leu Asn Tyr Ile Asp Ile Gly Lys Lys Glu

355 360 365
Gly Ala Asp Val Leu Thr Gly Gly Arg Arg Lys Leu Leu Glu Gly Glu

370 375 380
Leu Lys Asp Gly Tyr Tyr Leu Glu Pro Thr Ile Leu Phe Gly Gln Asn

385 390 395 400
Asn Met Arg Val Phe Gln Glu Glu Ile Phe Gly Pro Val Leu Ala Val

405 410 415
Thr Thr Phe Lys Thr Met Glu Glu Ala Leu Glu Leu Ala Asn Asp Thr

420 425 430
Gln Tyr Gly Leu Gly Ala Gly Val Trp Ser Arg Asn Gly Asn Leu Ala

435 440 445
Tyr Lys Met Gly Arg Gly Ile Gln Ala Gly Arg Val Trp Thr Asn Cys

450 455 460
Tyr His Ala Tyr Pro Ala His Ala Ala Phe Gly Gly Tyr Lys Gln Ser

465 470 475 480
Gly Ile Gly Arg Glu Thr His Lys Met Met Leu Glu His Tyr Gln Gln

485 490 495
Thr Lys Cys Leu Leu Val Ser Tyr Ser Asp Lys Pro Leu Gly Leu Phe

500 505 510
<210> 17

<211> 1440

<212> DNA

<213> Escherichia coli

<220><221> CDS

<222> (1)..(1440)

<400> 17

atg tca gta ccc gtt caa cat cct atg tat atc gat gga cag ttt gtt 48

Met Ser Val Pro Val Gln His Pro Met Tyr Ile Asp Gly Gln Phe Val

1 5 10 15

acc tgg cgt gga gac gca tgg att gat gtg gta aac cct gct aca gag	96
Thr Trp Arg Gly Asp Ala Trp Ile Asp Val Val Asn Pro Ala Thr Glu	
20 25 30	
gct gtc att tcc cgc ata ccc gat ggt cag gcc gag gat gcc cgt aag	144
Ala Val Ile Ser Arg Ile Pro Asp Gly Gln Ala Glu Asp Ala Arg Lys	
35 40 45	
gca atc gat gca gca gaa cgt gca caa cca gaa tgg gaa gcg ttg cct	192
Ala Ile Asp Ala Ala Glu Arg Ala Gln Pro Glu Trp Glu Ala Leu Pro	
50 55 60	
gct att gaa cgc gcc agt tgg ttg cgc aaa atc tcc gcc ggg atc cgc	240
Ala Ile Glu Arg Ala Ser Trp Leu Arg Lys Ile Ser Ala Gly Ile Arg	
65 70 75 80	
gaa cgc gcc agt gaa atc agt gcg ctg att gtt gaa gaa ggg ggc aag	288
Glu Arg Ala Ser Glu Ile Ser Ala Leu Ile Val Glu Glu Gly Gly Lys	
85 90 95	
atc cag cag ctg gct gaa gtc gaa gtg gct ttt act gcc gac tat atc	336
Ile Gln Gln Leu Ala Glu Val Glu Val Ala Phe Thr Ala Asp Tyr Ile	
100 105 110	
gat tac atg gcg gag tgg gca cgg cgt tac gag ggc gag att att caa	384
Asp Tyr Met Ala Glu Trp Ala Arg Arg Tyr Glu Gly Glu Ile Ile Gln	
115 120 125	
agc gat cgt cca gga gaa aat att ctt ttg ttt aaa cgt gcg ctt ggt	432
Ser Asp Arg Pro Gly Glu Asn Ile Leu Leu Phe Lys Arg Ala Leu Gly	
130 135 140	
gtg act acc ggc att ctg ccg tgg aac ttc ccg ttc ttc ctc att gcc	480
Val Thr Thr Gly Ile Leu Pro Trp Asn Phe Pro Phe Phe Leu Ile Ala	
145 150 155 160	
cgc aaa atg gct ccc gct ctt ttg acc ggt aat acc atc gtc att aaa	528
Arg Lys Met Ala Pro Ala Leu Leu Thr Gly Asn Thr Ile Val Ile Lys	
165 170 175	
cct agt gaa ttt acg cca aac aat gcg att gca ttc gcc aaa atc gtc	576

Pro Ser Glu Phe Thr Pro Asn Asn Ala Ile Ala Phe Ala Lys Ile Val	
180 185 190	
gat gaa ata ggc ctt ccg cgc ggc gtg ttt aac ctt gta ctg ggg cgt	624
Asp Glu Ile Gly Leu Pro Arg Gly Val Phe Asn Leu Val Leu Gly Arg	
195 200 205	
ggg gaa acc gtt ggg caa gaa ctg gcg ggt aac cca aag gtc gca atg	672
Gly Glu Thr Val Gly Gln Glu Leu Ala Gly Asn Pro Lys Val Ala Met	
210 215 220	
gtc agt atg aca ggc agc gtc tct gca ggt gag aag atc atg gcg act	720
Val Ser Met Thr Gly Ser Val Ser Ala Gly Glu Lys Ile Met Ala Thr	
225 230 235 240	
gcg gcg aaa aac atc acc aaa gtg tgt ctg gaa ttg ggg ggt aaa gca	768
Ala Ala Lys Asn Ile Thr Lys Val Cys Leu Glu Leu Gly Gly Lys Ala	
245 250 255	
cca gct atc gta atg gac gat gcc gat ctt gaa ctg gca gtc aaa gcc	816
Pro Ala Ile Val Met Asp Asp Ala Asp Leu Glu Leu Ala Val Lys Ala	
260 265 270	
atc gtt gat tca cgc gtc att aat agt ggg caa gtg tgt aac tgt gca	864
Ile Val Asp Ser Arg Val Ile Asn Ser Gly Gln Val Cys Asn Cys Ala	
275 280 285	
gaa cgt gtt tat gta cag aaa ggc att tat gat cag ttc gtc aat cgg	912
Glu Arg Val Tyr Val Gln Lys Gly Ile Tyr Asp Gln Phe Val Asn Arg	
290 295 300	
ctg ggt gaa gcg atg cag gcg gtt caa ttt ggt aac ccc gct gaa cgc	960
Leu Gly Glu Ala Met Gln Ala Val Gln Phe Gly Asn Pro Ala Glu Arg	
305 310 315 320	
aac gac att gcg atg ggg ccg ttg att aac gcc gcg gcg ctg gaa agg	1008
Asn Asp Ile Ala Met Gly Pro Leu Ile Asn Ala Ala Ala Leu Glu Arg	
325 330 335	
gtc gag caa aaa gtg gcg cgc gca gta gaa gaa ggg gcg aga gtg gcg	1056
Val Glu Gln Lys Val Ala Arg Ala Val Glu Glu Gly Ala Arg Val Ala	

340	345	350	
ttc ggt ggc aaa gcg gta gag ggg aaa gga tat tat tat ccg ccg aca			1104
Phe Gly Gly Lys Ala Val Glu Gly Lys Gly Tyr Tyr Tyr Pro Pro Thr			
355	360	365	
ttg ctg ctg gat gtt cgc cag gaa atg tgc att atg cat gag gaa acc			1152
Leu Leu Leu Asp Val Arg Gln Glu Met Ser Ile Met His Glu Glu Thr			
370	375	380	
ttt ggc ccg gtg ctg cca gtt gtc gca ttt gac acg ctg gaa gat gct			1200
Phe Gly Pro Val Leu Pro Val Val Ala Phe Asp Thr Leu Glu Asp Ala			
385	390	395	400
atc tca atg gct aat gac agt gat tac ggc ctg acc tca tca atc tat			1248
Ile Ser Met Ala Asn Asp Ser Asp Tyr Gly Leu Thr Ser Ser Ile Tyr			
405	410	415	
acc caa aat ctg aac gtc gcg atg aaa gcc att aaa ggg ctg aag ttt			1296
Thr Gln Asn Leu Asn Val Ala Met Lys Ala Ile Lys Gly Leu Lys Phe			
420	425	430	
ggt gaa act tac atc aac cgt gaa aac ttc gaa gct atg caa ggc ttc			1344
Gly Glu Thr Tyr Ile Asn Arg Glu Asn Phe Glu Ala Met Gln Gly Phe			
435	440	445	
cac gcc gga tgg cgt aaa tcc ggt att ggc ggc gca gat ggt aaa cat			1392
His Ala Gly Trp Arg Lys Ser Gly Ile Gly Gly Ala Asp Gly Lys His			
450	455	460	
ggc ttg cat gaa tat ctg cag acc cag gtg gtt tat tta cag tct taa			1440
Gly Leu His Glu Tyr Leu Gln Thr Gln Val Val Tyr Leu Gln Ser			
465	470	475	
<210> 18			
<211> 479			
<212> PRT			
<213> Escherichia coli			
<400> 18			
Met Ser Val Pro Val Gln His Pro Met Tyr Ile Asp Gly Gln Phe Val			

1	5	10	15
Thr Trp Arg Gly Asp Ala Trp Ile Asp Val Val Asn Pro Ala Thr Glu			
	20	25	30
Ala Val Ile Ser Arg Ile Pro Asp Gly Gln Ala Glu Asp Ala Arg Lys			
	35	40	45
Ala Ile Asp Ala Ala Glu Arg Ala Gln Pro Glu Trp Glu Ala Leu Pro			
	50	55	60
Ala Ile Glu Arg Ala Ser Trp Leu Arg Lys Ile Ser Ala Gly Ile Arg			
65	70	75	80
Glu Arg Ala Ser Glu Ile Ser Ala Leu Ile Val Glu Glu Gly Gly Lys			
	85	90	95
Ile Gln Gln Leu Ala Glu Val Glu Val Ala Phe Thr Ala Asp Tyr Ile			
	100	105	110
Asp Tyr Met Ala Glu Trp Ala Arg Arg Tyr Glu Gly Glu Ile Ile Gln			
	115	120	125
Ser Asp Arg Pro Gly Glu Asn Ile Leu Leu Phe Lys Arg Ala Leu Gly			
	130	135	140
Val Thr Thr Gly Ile Leu Pro Trp Asn Phe Pro Phe Phe Leu Ile Ala			
145	150	155	160
Arg Lys Met Ala Pro Ala Leu Leu Thr Gly Asn Thr Ile Val Ile Lys			
	165	170	175
Pro Ser Glu Phe Thr Pro Asn Asn Ala Ile Ala Phe Ala Lys Ile Val			
	180	185	190
Asp Glu Ile Gly Leu Pro Arg Gly Val Phe Asn Leu Val Leu Gly Arg			
	195	200	205
Gly Glu Thr Val Gly Gln Glu Leu Ala Gly Asn Pro Lys Val Ala Met			
	210	215	220
Val Ser Met Thr Gly Ser Val Ser Ala Gly Glu Lys Ile Met Ala Thr			
225	230	235	240
Ala Ala Lys Asn Ile Thr Lys Val Cys Leu Glu Leu Gly Gly Lys Ala			
	245	250	255

Pro Ala Ile Val Met Asp Asp Ala Asp Leu Glu Leu Ala Val Lys Ala
260 265 270

Ile Val Asp Ser Arg Val Ile Asn Ser Gly Gln Val Cys Asn Cys Ala
275 280 285

Glu Arg Val Tyr Val Gln Lys Gly Ile Tyr Asp Gln Phe Val Asn Arg
290 295 300

Leu Gly Glu Ala Met Gln Ala Val Gln Phe Gly Asn Pro Ala Glu Arg
305 310 315 320

Asn Asp Ile Ala Met Gly Pro Leu Ile Asn Ala Ala Ala Leu Glu Arg
325 330 335

Val Glu Gln Lys Val Ala Arg Ala Val Glu Glu Gly Ala Arg Val Ala
340 345 350

Phe Gly Gly Lys Ala Val Glu Gly Lys Gly Tyr Tyr Tyr Pro Pro Thr
355 360 365

Leu Leu Leu Asp Val Arg Gln Glu Met Ser Ile Met His Glu Glu Thr
370 375 380

Phe Gly Pro Val Leu Pro Val Val Ala Phe Asp Thr Leu Glu Asp Ala
385 390 395 400

Ile Ser Met Ala Asn Asp Ser Asp Tyr Gly Leu Thr Ser Ser Ile Tyr
405 410 415

Thr Gln Asn Leu Asn Val Ala Met Lys Ala Ile Lys Gly Leu Lys Phe
420 425 430

Gly Glu Thr Tyr Ile Asn Arg Glu Asn Phe Glu Ala Met Gln Gly Phe
435 440 445

His Ala Gly Trp Arg Lys Ser Gly Ile Gly Gly Ala Asp Gly Lys His
450 455 460

Gly Leu His Glu Tyr Leu Gln Thr Gln Val Val Tyr Leu Gln Ser
465 470 475

<210> 19

<211> 1488

<212> DNA

<213> Escherichia coli

<220><221> CDS

<222> (1)..(1488)

<400> 19

atg aat ttt cat cat ctg gct tac tgg cag gat aaa gcg tta agt ctc 48

Met Asn Phe His His Leu Ala Tyr Trp Gln Asp Lys Ala Leu Ser Leu

1 5 10 15

gcc att gaa aac cgc tta ttt att aac ggt gaa tat act gct gcg gcg 96

Ala Ile Glu Asn Arg Leu Phe Ile Asn Gly Glu Tyr Thr Ala Ala Ala

20 25 30

gaa aat gaa acc ttt gaa acc gtt gat ccg gtc acc cag gca ccg ctg 144

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

35 40 45

gcg aaa att gcc cgc ggc aag agc gtc gat atc gac cgt gcg atg agc 192

Ala Lys Ile Ala Arg Gly Lys Ser Val Asp Ile Asp Arg Ala Met Ser

50 55 60

gca gca cgc ggc gta ttt gaa cgc ggc gac tgg tca ctc tct tct ccg 240

Ala Ala Arg Gly Val Phe Glu Arg Gly Asp Trp Ser Leu Ser Ser Pro

65 70 75 80

gct aaa cgt aaa gcg gta ctg aat aaa ctc gcc gat tta atg gaa gcc 288

Ala Lys Arg Lys Ala Val Leu Asn Lys Leu Ala Asp Leu Met Glu Ala

85 90 95

cac gcc gaa gag ctg gca ctg ctg gaa act ctc gac acc ggc aaa ccg 336

His Ala Glu Glu Leu Ala Leu Leu Glu Thr Leu Asp Thr Gly Lys Pro

100 105 110

att cgt cac agt ctg cgt gat gat att ccc ggc gcg gcg cgc gcc att 384

Ile Arg His Ser Leu Arg Asp Asp Ile Pro Gly Ala Ala Arg Ala Ile

115 120 125

cgc tgg tac gcc gaa gcg atc gac aaa gtg tat ggc gaa gtg gcg acc 432

Arg Trp Tyr Ala Glu Ala Ile Asp Lys Val Tyr Gly Glu Val Ala Thr

130 135 140

acc agt agc cat gag ctg gcg atg atc gtg cgt gaa ccg gtc ggc gtg	480
Thr Ser Ser His Glu Leu Ala Met Ile Val Arg Glu Pro Val Gly Val	
145 150 155 160	
att gcc gcc atc gtg ccg tgg aac ttc ccg ctg ttg ctg act tgc tgg	528
Ile Ala Ala Ile Val Pro Trp Asn Phe Pro Leu Leu Leu Thr Cys Trp	
165 170 175	
aaa ctc ggc ccg gcg ctg gcg gcg gga aac agc gtg att cta aaa ccg	576
Lys Leu Gly Pro Ala Leu Ala Ala Gly Asn Ser Val Ile Leu Lys Pro	
180 185 190	
tct gaa aaa tca ccg ctc agt gcg att cgt ctc gcg ggg ctg gcg aaa	624
Ser Glu Lys Ser Pro Leu Ser Ala Ile Arg Leu Ala Gly Leu Ala Lys	
195 200 205	
gaa gca ggc ttg ccg gat ggt gtg ttg aac gtg gtg acg ggt ttt ggt	672
Glu Ala Gly Leu Pro Asp Gly Val Leu Asn Val Val Thr Gly Phe Gly	
210 215 220	
cat gaa gcc ggg cag gcg ctg tcg cgt cat aac gat atc gac gcc att	720
His Glu Ala Gly Gln Ala Leu Ser Arg His Asn Asp Ile Asp Ala Ile	
225 230 235 240	
gcc ttt acc ggt tca acc cgt acc ggg aaa cag ctg ctg aaa gat gcg	768
Ala Phe Thr Gly Ser Thr Arg Thr Gly Lys Gln Leu Leu Lys Asp Ala	
245 250 255	
ggc gac agc aac atg aaa cgc gtc tgg ctg gaa gcg ggc ggc aaa agc	816
Gly Asp Ser Asn Met Lys Arg Val Trp Leu Glu Ala Gly Gly Lys Ser	
260 265 270	
gcc aac atc gtt ttc gct gac tgc ccg gat ttg caa cag gcg gca agc	864
Ala Asn Ile Val Phe Ala Asp Cys Pro Asp Leu Gln Gln Ala Ala Ser	
275 280 285	
gcc acc gca gca ggc att ttc tac aac cag gga cag gtg tgc atc gcc	912
Ala Thr Ala Ala Gly Ile Phe Tyr Asn Gln Gly Gln Val Cys Ile Ala	
290 295 300	
gga acg cgc ctg ttg ctg gaa gag agc atc gcc gat gaa ttc tta gcc	960

Gly Thr Arg Leu Leu Leu Glu Glu Ser Ile Ala Asp Glu Phe Leu Ala
305 310 315 320
ctg tta aaa cag cag gcg caa aac tgg cag ccg ggc cat cca ctt gat 1008
Leu Leu Lys Gln Gln Ala Gln Asn Trp Gln Pro Gly His Pro Leu Asp
325 330 335
ccc gca acc acc atg ggc acc tta atc gac tgc gcc cac gcc gac tcg 1056
Pro Ala Thr Thr Met Gly Thr Leu Ile Asp Cys Ala His Ala Asp Ser
340 345 350
gtc cat agc ttt att cgg gaa ggc gaa agc aaa ggg caa ctg ttg ttg 1104
Val His Ser Phe Ile Arg Glu Gly Glu Ser Lys Gly Gln Leu Leu Leu
355 360 365
gat ggc cgt aac gcc ggg ctg gct gcc gcc atc ggc ccg acc atc ttt 1152
Asp Gly Arg Asn Ala Gly Leu Ala Ala Ala Ile Gly Pro Thr Ile Phe
370 375 380
gtg gat gtg gac ccg aat gcg tcc tta agt cgc gaa gag att ttc ggt 1200
Val Asp Val Asp Pro Asn Ala Ser Leu Ser Arg Glu Glu Ile Phe Gly
385 390 395 400
ccg gtg ctg gtg gtc acg cgt ttc aca tca gaa gaa cag gcg cta cag 1248
Pro Val Leu Val Val Thr Arg Phe Thr Ser Glu Glu Gln Ala Leu Gln
405 410 415
ctt gcc aac gac agc cag tac ggc ctt ggc gcg gcg gta tgg acg cgc 1296
Leu Ala Asn Asp Ser Gln Tyr Gly Leu Gly Ala Ala Val Trp Thr Arg
420 425 430
gac ctc tcc cgc gcg cac cgc atg agc cga cgc ctg aaa gcc ggt tcc 1344
Asp Leu Ser Arg Ala His Arg Met Ser Arg Arg Leu Lys Ala Gly Ser
435 440 445
gtc ttc gtc aat aac tac aac gac ggc gat atg acc gtg ccg ttt ggc 1392
Val Phe Val Asn Asn Tyr Asn Asp Gly Asp Met Thr Val Pro Phe Gly
450 455 460
ggc tat aag cag agc ggc aac ggt cgc gac aaa tcc ctg cat gcc ctt 1440
Gly Tyr Lys Gln Ser Gly Asn Gly Arg Asp Lys Ser Leu His Ala Leu

465	470	475	480	
gaa aaa ttc act gaa ctg aaa acc atc tgg ata agc ctg gag gcc tga				1488
Glu Lys Phe Thr	Glu Leu Lys Thr	Ile Trp Ile Ser Leu Glu Ala		
	485	490	495	
<210>	20			
<211>	495			
<212>	PRT			
<213>	Escherichia coli			
<400>	20			
Met Asn Phe His His Leu Ala Tyr Trp Gln Asp Lys Ala Leu Ser Leu				
1	5	10	15	
Ala Ile Glu Asn Arg Leu Phe Ile Asn Gly Glu Tyr Thr Ala Ala Ala				
	20	25	30	
Glu Asn Glu Thr Phe Glu Thr Val Asp Pro Val Thr Gln Ala Pro Leu				
	35	40	45	
Ala Lys Ile Ala Arg Gly Lys Ser Val Asp Ile Asp Arg Ala Met Ser				
50	55	60		
Ala Ala Arg Gly Val Phe Glu Arg Gly Asp Trp Ser Leu Ser Ser Pro				
65	70	75	80	
Ala Lys Arg Lys Ala Val Leu Asn Lys Leu Ala Asp Leu Met Glu Ala				
	85	90	95	
His Ala Glu Glu Leu Ala Leu Leu Glu Thr Leu Asp Thr Gly Lys Pro				
	100	105	110	
Ile Arg His Ser Leu Arg Asp Asp Ile Pro Gly Ala Ala Arg Ala Ile				
	115	120	125	
Arg Trp Tyr Ala Glu Ala Ile Asp Lys Val Tyr Gly Glu Val Ala Thr				
130	135	140		
Thr Ser Ser His Glu Leu Ala Met Ile Val Arg Glu Pro Val Gly Val				
145	150	155	160	
Ile Ala Ala Ile Val Pro Trp Asn Phe Pro Leu Leu Leu Thr Cys Trp				
	165	170	175	

Lys Leu Gly Pro Ala Leu Ala Ala Gly Asn Ser Val Ile Leu Lys Pro
 180 185 190
 Ser Glu Lys Ser Pro Leu Ser Ala Ile Arg Leu Ala Gly Leu Ala Lys
 195 200 205
 Glu Ala Gly Leu Pro Asp Gly Val Leu Asn Val Val Thr Gly Phe Gly
 210 215 220

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

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

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

 Leu Ala Asn Asp Ser Gln Tyr Gly Leu Gly Ala Ala Val Trp Thr Arg

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

Glu Lys Phe Thr Glu Leu Lys Thr Ile Trp Ile Ser Leu Glu Ala
 485 490 495

<210> 21

<211> 1395

<212> DNA

<213> Escherichia coli

<400> 21

atgcctgacg ctaaaaaaca ggggcggtca aacaaggcaa tgacgttttt cgtctgcttc 60
 ctigccgctc tggcgggatt actctttggc ctggatatcg gtgtaattgc tggcgccactg 120
 ccgtttattg cagatgaatt ccagattact tcgcacacgc aagaatgggt cgtaagctcc 180
 atgatgttcg gtgcggcagt cgggtcgggtg ggccagcggct ggctctcctt taaactcggg 240
 cgcaaaaaga gcctgatgat cggcgcaatt ttgtttgttg ccggttcgct gttctctgcg 300

 gctgcgccaa acgttgaagt actgattctt tcccgcgttc tactggggct ggccgtgggt 360
 gtggcctctt ataccgcacc gctgtacctc tctgaaattg cgccggaaaa aattcgtggc 420
 agtatgatct cgatgtatca gttgatgac actatcggga tcctcgggtc ttatctttct 480
 gataccgcct tcagctacac cgggtcatgg cgtcggatgc tgggtgtgat tatcatcccg 540
 gcaattttgc tgctgattgg tgtcttcttc ctgccagaca gccacgttg gtttgccgcc 600
 aaacgccgtt ttgttgatgc cgaacgcgtg ctgctacgcc tgcgtgacac cagcgcggaa 660
 gcgaaacgcg aactggatga aatccgtgaa agtttcagg ttaaacagag tggctgggcg 720

 ctgtttaaag agaacagcaa cttccgccgc gcggtgttcc ttggcgact gttgcaggta 780
 atgcagcaat tcaccgggat gaacgtcatc atgtattacg cgccgaaaat cttcgaactg 840
 gcgggttata ccaacactac cgagcaaatg tgggggaccg tgattgtcgg cctgaccaac 900
 gtacttgcca cttttatcgc aatcggcctt gttgaccgt ggggacgtaa accaacgcta 960
 acgctgggct tcctggatga ggctgctggc atgggcgtac tcggtacaat gatgcatatc 1020
 ggtattcact ctccgtcgcc gcagtatttc gccatcgcca tgctgctgat gtttattgtc 1080

ggttttgccca tgagtgccgg tccgtgatt tgggtactgt gtcgccaaat tcagccgctg 1140

aaaggccgcg attttggcat cacctgctcc actgccacca actggattgc caacatgac 1200

gttggcgcaa cgttcctgac catgctcaac acgctgggta acgccaacac cttctgggtg 1260

tatgcggctc tgaacgtact gtttatcctg ctgacattgt ggctgggtacc ggaaacaaaa 1320

cacgtttcgc tggaacatat tgaacgtaat ctgatgaaag gtcgtaaact gcgcgaaata 1380

ggcgctcacg attaa 1395

<210> 22

<211> 464

<212> PRT

<213> Escherichia coli

<400> 22

Met Pro Asp Ala Lys Lys Gln Gly Arg Ser Asn Lys Ala Met Thr Phe

1 5 10 15
Phe Val Cys Phe Leu Ala Ala Leu Ala Gly Leu Leu Phe Gly Leu Asp

20 25 30
Ile Gly Val Ile Ala Gly Ala Leu Pro Phe Ile Ala Asp Glu Phe Gln

35 40 45
Ile Thr Ser His Thr Gln Glu Trp Val Val Ser Ser Met Met Phe Gly

50 55 60
Ala Ala Val Gly Ala Val Gly Ser Gly Trp Leu Ser Phe Lys Leu Gly

65 70 75 80
Arg Lys Lys Ser Leu Met Ile Gly Ala Ile Leu Phe Val Ala Gly Ser

85 90 95
Leu Phe Ser Ala Ala Ala Pro Asn Val Glu Val Leu Ile Leu Ser Arg

100 105 110
Val Leu Leu Gly Leu Ala Val Gly Val Ala Ser Tyr Thr Ala Pro Leu

115 120 125
Tyr Leu Ser Glu Ile Ala Pro Glu Lys Ile Arg Gly Ser Met Ile Ser

130 135 140
Met Tyr Gln Leu Met Ile Thr Ile Gly Ile Leu Gly Ala Tyr Leu Ser

145 150 155 160

Asp Thr Ala Phe Ser Tyr Thr Gly Ala Trp Arg Trp Met Leu Gly Val
 165 170 175
 Ile Ile Ile Pro Ala Ile Leu Leu Leu Ile Gly Val Phe Phe Leu Pro
 180 185 190
 Asp Ser Pro Arg Trp Phe Ala Ala Lys Arg Arg Phe Val Asp Ala Glu
 195 200 205
 Arg Val Leu Leu Arg Leu Arg Asp Thr Ser Ala Glu Ala Lys Arg Glu
 210 215 220
 Leu Asp Glu Ile Arg Glu Ser Leu Gln Val Lys Gln Ser Gly Trp Ala
 225 230 235 240
 Leu Phe Lys Glu Asn Ser Asn Phe Arg Arg Ala Val Phe Leu Gly Val
 245 250 255
 Leu Leu Gln Val Met Gln Gln Phe Thr Gly Met Asn Val Ile Met Tyr
 260 265 270
 Tyr Ala Pro Lys Ile Phe Glu Leu Ala Gly Tyr Thr Asn Thr Thr Glu
 275 280 285
 Gln Met Trp Gly Thr Val Ile Val Gly Leu Thr Asn Val Leu Ala Thr
 290 295 300
 Phe Ile Ala Ile Gly Leu Val Asp Arg Trp Gly Arg Lys Pro Thr Leu
 305 310 315 320
 Thr Leu Gly Phe Leu Val Met Ala Ala Gly Met Gly Val Leu Gly Thr
 325 330 335
 Met Met His Ile Gly Ile His Ser Pro Ser Ala Gln Tyr Phe Ala Ile
 340 345 350
 Ala Met Leu Leu Met Phe Ile Val Gly Phe Ala Met Ser Ala Gly Pro
 355 360 365
 Leu Ile Trp Val Leu Cys Ser Glu Ile Gln Pro Leu Lys Gly Arg Asp
 370 375 380
 Phe Gly Ile Thr Cys Ser Thr Ala Thr Asn Trp Ile Ala Asn Met Ile
 385 390 395 400
 Val Gly Ala Thr Phe Leu Thr Met Leu Asn Thr Leu Gly Asn Ala Asn

405 410 415
 Thr Phe Trp Val Tyr Ala Ala Leu Asn Val Leu Phe Ile Leu Leu Thr
 420 425 430
 Leu Trp Leu Val Pro Glu Thr Lys His Val Ser Leu Glu His Ile Glu
 435 440 445
 Arg Asn Leu Met Lys Gly Arg Lys Leu Arg Glu Ile Gly Ala His Asp

450 455 460
 <210> 23
 <211> 1248
 <212> DNA
 <213> Escherichia coli
 <400> 23
 atggcactga atattccatt cagaaatgcg tactatcggt ttgcatccag ttactcattt 60
 ctctttttta tttcctggtc gctgtgggtg tcgttatacg ctatttggct gaaaggacat 120
 ctaggattaa cagggacgga attaggtaca ctttatcgg tcaaccagtt taccagcatt 180
 ctatttatga tgttctacgg catcggttcag gataaactcg gtctgaagaa accgctcatc 240
 tgggtgatga gtttcattct ggtcttgacc ggaccgttta tgatttacgt ttatgaaccg 300
 ttactgcaaa gcaatttttc tgtaggtcta attctggggg cgctcttttt tggcctgggg 360
 tatctggcgg gatgcggttt gcttgacagc ttaccgaaa aaatggcgcg aaattttcat 420
 ttcgaatatg gaacagcgcg cgctgggga tcttttggct atgctattgg cgcgttcttt 480
 gccggtatat tttttagtat cagtcccat atcaacttct ggttgggtctc gctatttggc 540
 gctgtattta tgatgatcaa catgcgtttt aaagataagg atcaccagtg catagcggcg 600
 gatgcgggag gggtaaaaaa agaggatttt atcgagttt tcaaggatcg aaacttctgg 660
 gttttcgtca tatttatgtt ggggacgtgg tctttctata acatttttga tcaacaactc 720
 tttcctgtct tttatgcagg tttattcgaa tcacacgatg taggaacgcg cctgtatggt 780
 tatctcaact cattccaggt ggtactcgaa gcgctgtgca tggcgattat tcctttcttt 840
 gtgaatcggg tagggccaaa aaatgcatta cttatcgggtg ttgtgattat ggcgttgcgt 900
 atcctttcct gcgcgttgtt cgtaaacccc tggattatct cattagttaa gctgttacat 960
 gccattgagg ttccactttg tgtcatatcc gtcttcaaat acagcgtggc aaactttgat 1020
 aagcgctgt cgtcgacgat ctttctgatt ggttttcaaa ttgccagttc gcttgggatt 1080
 gtgctgcttt caacgccgac tgggatactc ttgaccacg caggctacca gacagttttc 1140

```

ttcgcaattt cgggtattgt ctgcctgatg ttgctatttg gcattttctt cctgagtaaa 1200
aaacgcgagc aaatagttat ggaaacgcct gtaccttcag caatatag 1248
<210> 24
<211> 415
<212> PRT
<213> Escherichia coli
<400> 24
Met Ala Leu Asn Ile Pro Phe Arg Asn Ala Tyr Tyr Arg Phe Ala Ser
1          5          10          15
Ser Tyr Ser Phe Leu Phe Phe Ile Ser Trp Ser Leu Trp Trp Ser Leu
          20          25          30
Tyr Ala Ile Trp Leu Lys Gly His Leu Gly Leu Thr Gly Thr Glu Leu

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

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

          165          170          175
Ser Leu Phe Gly Ala Val Phe Met Met Ile Asn Met Arg Phe Lys Asp
          180          185          190
Lys Asp His Gln Cys Ile Ala Ala Asp Ala Gly Gly Val Lys Lys Glu

```

195 200 205
 Asp Phe Ile Ala Val Phe Lys Asp Arg Asn Phe Trp Val Phe Val Ile
 210 215 220
 Phe Ile Val Gly Thr Trp Ser Phe Tyr Asn Ile Phe Asp Gln Gln Leu

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

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

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

<210> 25

<211> 1248

<212> DNA

<213> Escherichia coli

<400> 25

atggcactga atattccatt cagaaatgcg tactatcggt ttgcatccag ttactcattt	60
ctctttttta tttcctgggtc gctgtgggtgg tcgttatatcg ctatttggct gaaaggacat	120
ctagggttga caggagacga attaggtaca ctttattcgg tcaaccagtt taccagcatt	180
ctattttatga tgttctacgg catcggttcag gataaaactcg gtctgaagaa accgctcatc	240
tgggtgatga gtttcatcct ggtcttgacc ggaccgttta tgatttacgt ttatgaaccg	300
ttactgcaaa gcaatttttc tgtaggtcta attctggggg cgctattttt tggcttgggg	360
tatctggcgg gatgcggttt gcttgatagc ttcaccgaaa aaatggcgcg aaattttcat	420
ttcgaatatg gaacagcgcg cgcctgggga tcttttggct atgctattgg cgcgttcttt	480
gccggcatat tttttagtat cagtcccat atcaacttct ggttgggtctc gctatttggc	540
gctgtattta tgatgatcaa catgcgtttt aaagataagg atcaccagtg cgtagcggca	600
gatgcgggag gggtaaaaaa agaggatttt atcgagttt tcaaggatcg aaacttctgg	660
gttttcgtca tatttattgt ggggacgtgg tctttctata acatttttga tcaacaactt	720
tttcctgtct tttattcagg tttattcgaa tcacacgatg taggaacgcg cctgtatggt	780
tatctcaact cattccaggt ggtactcgaa gcgctgtgca tggcgattat tcctttcttt	840
gtgaatcggg tagggcctaa aaatgcattt cttatcggag ttgtgattat ggcgttgcgt	900
atcctttcct gcgcgtgtt cgtaacccc tggattattt cattagttaa gtgtttacat	960
gccattgagg ttccactttg tgtcatatcc gtcttcaaat acagcgtggc aaactttgat	1020
aagcgctgt cgtcgacgat ctttctgatt ggttttcaaa ttgccagttc gcttgggatt	1080
gtgctgcttt caacgccgac tgggatactc ttgaccacg caggetacca gacagttttc	1140
ttcgcaattt cgggtattgt ctgcctgatg ttgctatttg gcattttctt cttgagtaaa	1200
aaacgcgagc aaatagttaa ggaaacgcct gtaccttcag caatatag	1248

<210> 26

<211> 415

<212> PRT

<213> Escherichia coli

<400> 26

Met	Ala	Leu	Asn	Ile	Pro	Phe	Arg	Asn	Ala	Tyr	Tyr	Arg	Phe	Ala	Ser
1				5					10					15	
Ser	Tyr	Ser	Phe	Leu	Phe	Phe	Ile	Ser	Trp	Ser	Leu	Trp	Trp	Ser	Leu
				20					25					30	
Tyr	Ala	Ile	Trp	Leu	Lys	Gly	His	Leu	Gly	Leu	Thr	Gly	Thr	Glu	Leu

35 40 45
 Gly Thr Leu Tyr Ser Val Asn Gln Phe Thr Ser Ile Leu Phe Met Met

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

 115 120 125
 Asp Ser Phe Thr Glu Lys Met Ala Arg Asn Phe His Phe Glu Tyr Gly
 130 135 140
 Thr Ala Arg Ala Trp Gly Ser Phe Gly Tyr Ala Ile Gly Ala Phe Phe
 145 150 155 160
 Ala Gly Ile Phe Phe Ser Ile Ser Pro His Ile Asn Phe Trp Leu Val
 165 170 175
 Ser Leu Phe Gly Ala Val Phe Met Met Ile Asn Met Arg Phe Lys Asp

 180 185 190
 Lys Asp His Gln Cys Val Ala Ala Asp Ala Gly Gly Val Lys Lys Glu
 195 200 205
 Asp Phe Ile Ala Val Phe Lys Asp Arg Asn Phe Trp Val Phe Val Ile
 210 215 220
 Phe Ile Val Gly Thr Trp Ser Phe Tyr Asn Ile Phe Asp Gln Gln Leu
 225 230 235 240
 Phe Pro Val Phe Tyr Ser Gly Leu Phe Glu Ser His Asp Val Gly Thr

 245 250 255
 Arg Leu Tyr Gly Tyr Leu Asn Ser Phe Gln Val Val Leu Glu Ala Leu
 260 265 270
 Cys Met Ala Ile Ile Pro Phe Phe Val Asn Arg Val Gly Pro Lys Asn
 275 280 285

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

290 295 300

Ala Leu Phe Val Asn Pro Trp Ile Ile Ser Leu Val Lys Leu Leu His

305 310 315 320

Ala Ile Glu Val Pro Leu Cys Val Ile Ser Val Phe Lys Tyr Ser Val

325 330 335

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

340 345 350

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

355 360 365

Ile Leu Phe Asp His Ala Gly Tyr Gln Thr Val Phe Phe Ala Ile Ser

370 375 380

Gly Ile Val Cys Leu Met Leu Leu Phe Gly Ile Phe Phe Leu Ser Lys

385 390 395 400

Lys Arg Glu Gln Ile Val Met Glu Thr Pro Val Pro Ser Ala Ile

405 410 415

<210> 27

<211> 1326

<212> DNA

<213> Bifidobacterium lactis

<400> 27

atggcaacaa ccacgaaggt gtggaggaa cctcctacc tgcaaagctc aaccggcatc 60

ttcctgttct tctgctcctg gggcatctgg tggctgttct tccagcgtg gctcaactcg 120

atgggactca acggcgcgaa agtgggcacg atctattcga tcaactcgct ggccacgctc 180

atcctcatgt tcgggtacgg cctcatccag gacaatctcg gactcaagcg ccgtcttgtg 240

ctcgtcatct cggcgcgcgc cgcactcgtc ggacccttcg tgcagttcgt gtacgcgccg 300

ctgatgagga cgaacatgat ggccgccgca ctcgtgggct ccgtcgttct ctccgcgggc 360

ttcatggcag gctgctcgct catagagccc gtgaccgaac ggtacagccg ccgtttcaac 420

ttagagtacg gccaatcccg cgcatggggt tccttcggat atgccattgt ggcgcttgtc 480

gccggttcg tgttcaacat caaccgatg atcaacttct ggctcggtc cgcattcggc 540

gtgggcatgc tcatcgtgta cctcacctgg tatccggccg agcagcgcgga agcgctcaag 600

gaagccgccg atccgaatgc cgcgccaact aaccgacca tcaaagacat gctcggcgtg 660
 ctcaagatgc ccacgtgtg ggtgctcatc gtgttcatgc tgctaccaa cacgttttac 720
 accgtattcg accagcagat gtccccacc tactacgcct cgctcttccc gaatgaggcc 780
 accggcaacg ccgtctacgg cacgtcaac tcggtgcagg tgttctgcga atccgcgatg 840
 atgggcgtcg tgccgatcat catgcgcaag gtaggtgtgc gcaacgcgtt gctgctcgga 900
 tccacggtga tgttccttcg catcgggctg tgcggcatct tccacgatcc ggtgtccatc 960

tcgatcgtca aaatgttcca cgccattgaa gtcccgctgt tctgcctgcc ggcgttccgc 1020
 tacttcacgc tccattcaa tccgaagctc tccgcgacgc tctacatggt cggcttccag 1080
 attgcctcac agatcggcca ggtcgtcttc tccacccgc tcggcatgct gcatgaccgc 1140
 atgggcgacc gcacgacgtt cctgacgac tccgccatcg tgcttgctgc caccgtctac 1200
 ggattcttcg tgatcaagcg cgacgacgag caggtggatg gcatccgtt catcccgcat 1260
 tcgaagaagc tgccgtcgtc cgccaccgac gaggcgatcc tctccgcgga ttccgaggat 1320
 atgtaa 1326

<210> 28

<211> 441

<212> PRT

<213> Bifidobacterium lactis

<400> 28

Met Ala Thr Thr Thr Lys Val Trp Arg Asn Pro Ser Tyr Leu Gln Ser

1 5 10 15

Ser Thr Gly Ile Phe Leu Phe Phe Cys Ser Trp Gly Ile Trp Trp Ser

20 25 30

Phe Phe Gln Arg Trp Leu Asn Ser Met Gly Leu Asn Gly Ala Lys Val

35 40 45

Gly Thr Ile Tyr Ser Ile Asn Ser Leu Ala Thr Leu Ile Leu Met Phe

50 55 60

Gly Tyr Gly Leu Ile Gln Asp Asn Leu Gly Leu Lys Arg Arg Leu Val

65 70 75 80

Leu Val Ile Ser Ala Ile Ala Ala Leu Val Gly Pro Phe Val Gln Phe

85 90 95

Val Tyr Ala Pro Leu Met Arg Thr Asn Met Met Ala Ala Ala Leu Val

100 105 110

Gly Ser Val Val Leu Ser Ala Gly Phe Met Ala Gly Cys Ser Leu Ile

115 120 125

Glu Pro Val Thr Glu Arg Tyr Ser Arg Arg Phe Asn Leu Glu Tyr Gly

130 135 140

Gln Ser Arg Ala Trp Gly Ser Phe Gly Tyr Ala Ile Val Ala Leu Val

145 150 155 160

Ala Gly Phe Val Phe Asn Ile Asn Pro Met Ile Asn Phe Trp Leu Gly

165 170 175

Ser Ala Phe Gly Val Gly Met Leu Ile Val Tyr Leu Thr Trp Tyr Pro

180 185 190

Ala Glu Gln Arg Glu Ala Leu Lys Glu Ala Ala Asp Pro Asn Ala Ala

195 200 205

Pro Thr Asn Pro Thr Ile Lys Asp Met Leu Gly Val Leu Lys Met Pro

210 215 220

Thr Leu Trp Val Leu Ile Val Phe Met Leu Leu Thr Asn Thr Phe Tyr

225 230 235 240

Thr Val Phe Asp Gln Gln Met Phe Pro Thr Tyr Tyr Ala Ser Leu Phe

245 250 255

Pro Asn Glu Ala Thr Gly Asn Ala Val Tyr Gly Thr Leu Asn Ser Val

260 265 270

Gln Val Phe Cys Glu Ser Ala Met Met Gly Val Val Pro Ile Ile Met

275 280 285

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

290 295 300

Phe Leu Arg Ile Gly Leu Cys Gly Ile Phe His Asp Pro Val Ser Ile

305 310 315 320

Ser Ile Val Lys Met Phe His Ala Ile Glu Val Pro Leu Phe Cys Leu

325 330 335

Pro Ala Phe Arg Tyr Phe Thr Leu His Phe Asn Pro Lys Leu Ser Ala

340 345 350

Thr Leu Tyr Met Val Gly Phe Gln Ile Ala Ser Gln Ile Gly Gln Val

355 360 365
Val Phe Ser Thr Pro Leu Gly Met Leu His Asp Arg Met Gly Asp Arg

370 375 380
Thr Thr Phe Leu Thr Ile Ser Ala Ile Val Leu Ala Ala Thr Val Tyr
385 390 395 400
Gly Phe Phe Val Ile Lys Arg Asp Asp Glu Gln Val Asp Gly Asp Pro
405 410 415
Phe Ile Arg Asp Ser Lys Lys Leu Pro Ser Leu Ala Thr Asp Glu Ala
420 425 430
Ile Leu Ser Ala Asp Ser Glu Asp Met
435 440

<210> 29

<211> 858

<212> DNA

<213> Streptococcus pneumoniae

<400> 29

ttattgatga ctgtccccgg tttagtttta acctttatct ttaaatacat ccctatgtat 60
gggggttttaa tcgcatttaa agattacaat cctttaaaag gaattttagg gagtgtattgg 120
attgggtttt ctgagtttac aaaattcata tcctctccca actttgggtat ctgtttagcc 180
aacacattaa aattaagtat ctatggttta ttgcttggct tttaccacc aatcattctc 240
gcgattatgc tcaatcaact cttagtgtaa aaagtcaaaa aacgaattca gtcattttta 300
tacgcaccaa actttatctc agtcgttggt attgtcggta tgattttcct cttcttttca 360

gtgggaggac caatcaacaa ttttctttct atgtttggaa tgaaggctga ctcttgaca 420
aatccagact tcttagacc tttatacatc ttagtggtta tctggcaagg aatgggctgg 480
gcttcaacgc tctacacggc aacattggta aatgtagatc cagccttagt agaagcagcc 540
cgactggatg gagccaatat cttccaacga atctggcaca ttgatattcc agctcttaag 600
cctattatgg ttatccaatt tgttttagct gcaggtggaa ttatgaatgt cggatatgaa 660
aaagcattct tgatgcagac atcgttaaat ttgccaactt ctgaaattat ctgcacatat 720
gtctataaag ttggtcttgt atcaggagac tattcttact caacagcggg ttggtttgtt 780

aatgcagtga ttaacgtagt attgcttgtt gcagttaacc aaatcgtaa acgcatgaat 840
aatggtgaag gaatttaa 858

<210> 30

<211> 305

<212> PRT

<213> Streptococcus pneumoniae

<400> 30

Met Asn Ser Lys Ala Lys Gln Val Ser Leu Trp Glu Arg Ile Lys Lys

1 5 10 15

Gln Lys Leu Leu Leu Met Thr Val Pro Gly Leu Val Leu Thr Phe

20 25 30

Ile Phe Lys Tyr Ile Pro Met Tyr Gly Val Leu Ile Ala Phe Lys Asp

35 40 45

Tyr Asn Pro Leu Lys Gly Ile Leu Gly Ser Asp Trp Ile Gly Phe Ser

50 55 60

Glu Phe Thr Lys Phe Ile Ser Ser Pro Asn Phe Gly Ile Leu Leu Ala

65 70 75 80

Asn Thr Leu Lys Leu Ser Ile Tyr Gly Leu Leu Leu Gly Phe Leu Pro

85 90 95

Pro Ile Ile Leu Ala Ile Met Leu Asn Gln Leu Leu Ser Glu Lys Val

100 105 110

Lys Lys Arg Ile Gln Leu Ile Leu Tyr Ala Pro Asn Phe Ile Ser Val

115 120 125

Val Val Ile Val Gly Met Ile Phe Leu Phe Phe Ser Val Gly Gly Pro

130 135 140

Ile Asn Asn Phe Leu Ser Met Phe Gly Met Lys Ala Asp Phe Leu Thr

145 150 155 160

Asn Pro Asp Phe Phe Arg Pro Leu Tyr Ile Phe Ser Gly Ile Trp Gln

165 170 175

Gly Met Gly Trp Ala Ser Thr Leu Tyr Thr Ala Thr Leu Val Asn Val

180 185 190

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

195 200 205

Gln Arg Ile Trp His Ile Asp Ile Pro Ala Leu Lys Pro Ile Met Val

210 215 220
Ile Gln Phe Val Leu Ala Ala Gly Gly Ile Met Asn Val Gly Tyr Glu

225 230 235 240
Lys Ala Phe Leu Met Gln Thr Ser Leu Asn Leu Pro Thr Ser Glu Ile

245 250 255
Ile Ser Thr Tyr Val Tyr Lys Val Gly Leu Val Ser Gly Asp Tyr Ser

260 265 270
Tyr Ser Thr Ala Val Gly Leu Phe Asn Ala Val Ile Asn Val Val Leu

275 280 285
Leu Val Ala Val Asn Gln Ile Val Lys Arg Met Asn Asn Gly Glu Gly

290 295 300
Ile

305

<210> 31

<211> 918

<212> DNA

<213> Streptococcus pneumoniae

<400> 31

atggtgaagg aatttaagga ggaaagtatg aaaaattcga ttatggatac aaaatttgat 60
agacgtatct tactcttaaa taaatcatt attgtcttta tcgttttgat gactttgctt 120
cctttacttt atatcgctgt agcatccttt atggatccta aggttctggt tagtagaggg 180
attagcttta atccagccga ttggactgta gaaggttacc agcgtgtatt cagtgaccaa 240
tctattctaa gaggttttat caattctcta ctatactctt ttggatttgc agctttaaca 300

gtcttgctat ctgtgtttac agcttatcct ctttctaaga aagacttggt tggacgtcgt 360
tggattaact acttcttgat tgtaactatg ttctttggtg gtggtttagt cccaacttac 420
ttgctcgtaa aagaattggg aatgctcaat actccatggg ctatcattgt tccaggtgct 480
gttaacgttt ggaatattat tcttgctagg gcctatttcc aaggattgcc tgaagaatta 540
gttgaagctg ctgtcattga tggtgcaa atgattacaga ttttcttcaa aatcatgctt 600
cctcttgcaa aaccaattat gtttgttctc ttcttttatg cttttgtagg acagtggaac 660
tcatactttg atgcaatgat ttatatcaag gatccaaact tggaaccatt gcaacttgta 720

cttcgtaaaa ttctcattca gagccaacca ggtcaagaca tgattggagc acaagcggct 780

atgaatgaaa tgaacggtt agctgaattg attaaatagc caactattgt catttccagc 840
 ttgccattga ttgttatgta tccattcttc caaaaatact ttgataaagg aattatggct 900
 ggttcactta aaggataa 918

<210> 32

<211> 305

<212> PRT

<213> Streptococcus pneumoniae

<400> 32

Met Val Lys Glu Phe Lys Glu Glu Ser Met Lys Asn Ser Ile Met Asp

1 5 10 15

Thr Lys Phe Asp Arg Arg Ile Leu Leu Leu Asn Lys Ile Ile Ile Val

20 25 30

Phe Ile Val Leu Met Thr Leu Leu Pro Leu Leu Tyr Ile Val Val Ala

35 40 45

Ser Phe Met Asp Pro Lys Val Leu Val Ser Arg Gly Ile Ser Phe Asn

50 55 60

Pro Ala Asp Trp Thr Val Glu Gly Tyr Gln Arg Val Phe Ser Asp Gln

65 70 75 80

Ser Ile Leu Arg Gly Phe Ile Asn Ser Leu Leu Tyr Ser Phe Gly Phe

85 90 95

Ala Ala Leu Thr Val Leu Leu Ser Val Phe Thr Ala Tyr Pro Leu Ser

100 105 110

Lys Lys Asp Leu Val Gly Arg Arg Trp Ile Asn Tyr Phe Leu Ile Val

115 120 125

Thr Met Phe Phe Gly Gly Gly Leu Val Pro Thr Tyr Leu Leu Val Lys

130 135 140

Glu Leu Gly Met Leu Asn Thr Pro Trp Ala Ile Ile Val Pro Gly Ala

145 150 155 160

Val Asn Val Trp Asn Ile Ile Leu Ala Arg Ala Tyr Phe Gln Gly Leu

165 170 175

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

180 185 190

Gln Ile Phe Phe Lys Ile Met Leu Pro Leu Ala Lys Pro Ile Met Phe
195 200 205

Val Leu Phe Leu Tyr Ala Phe Val Gly Gln Trp Asn Ser Tyr Phe Asp
210 215 220

Ala Met Ile Tyr Ile Lys Asp Pro Asn Leu Glu Pro Leu Gln Leu Val
225 230 235 240

Leu Arg Lys Ile Leu Ile Gln Ser Gln Pro Gly Gln Asp Met Ile Gly
245 250 255

Ala Gln Ala Ala Met Asn Glu Met Lys Arg Leu Ala Glu Leu Ile Lys
260 265 270

Tyr Ala Thr Ile Val Ile Ser Ser Leu Pro Leu Ile Val Met Tyr Pro
275 280 285

Phe Phe Gln Lys Tyr Phe Asp Lys Gly Ile Met Ala Gly Ser Leu Lys
290 295 300

Gly

305

<210> 33

<211> 1617

<212> DNA

<213> Streptococcus pneumoniae

<400> 33

atgaaattca aaacattctc aaaatcagca gttttgttga cagctagttt agcagtactt 60

gcagcctgtg gctcaaaaaa tacagcttca agtccagatt ataagttgga aggtgtaaca 120

ttcccgttc aagaaaagaa aacattgaag tttatgacag ccagttcacc gttatctcct 180

aaagacccaa atgaaaagtt aattttgcaa cgtttgaga aggaaactgg cgttcattatt 240

gactggacca actaccaatc cgactttgca gaaaaacgta acttgatat ttctagtgg 300

gatttaccag atgctatcca caacgacgga gcttcagatg tggacttgat gaactgggct 360

aaaaaagggt ttattattcc agttgaagat ttgattgata aatacatgcc aaatcttaag 420

aaaattttgg atgagaaacc agagtacaag gccttgatga cagcacctga tgggcacatt 480

tactcatttc catggattga agagcttgga gatggtaaag agtctattca cagtgtcaac 540

gatatggcct ggattaacaa agattggcctt aagaaacttg gtcttgaaat gccaaaaact 600

actgatgatt tgattaaagt cctagaagct ttcaaaaacg gggatccaaa tggaaatgga 660
gaggctgatg aaattccatt ttcatTTtatt agtggtaacg gaaacgaaga ttttaaattc 720
ctatttTctg catttggTat aggggataac gatgatcatt tagtagtagg aaatgatggc 780
aaagttgact tcacagcaga taacgataac tataaagaag gtgtcaaatt tatccgtcaa 840
ttgcaagaaa aaggcctgat tgataaagaa gctttcgaac atgattggaa tagttacatt 900
gctaaaggTc atgatcagaa atttggTgtt tactttacat gggataagaa taatgttact 960

ggaagtaacg aaagttatga tgttttacca gtacttTctg gaccaagtgg tcaaaaacac 1020
gtagctcgta caaacggTat gggatttTca cgtgacaaga tggttattac cagtgtaaac 1080
aaaaacctag aattgacagc taaatggatt gatgcacaat acgctccact ccaatctgtg 1140
caaaataact ggggaactta cggagatgac aaacaacaaa acatctttga attggatcaa 1200
gcgtcaaata gtctaaaaca cttaccactia aacggaactg caccagcaga acttcgtcaa 1260
aagactgaag taggaggacc actagctatc ctagattcat actatggtaa agtaacaacc 1320
atgcctgatg atgccaaatg gcgtttggat cttatcaaag aatattatgt tccttacatg 1380

agcaatgtca ataactatcc aagagtcttt atgacacagg aagatttTga caagattgcc 1440
catatcgaag cagatatgaa tgactatatc taccgtaaac gtgctgaatg gattgtaaat 1500
ggcaatattg atactgagtg ggatgattac aagaaagaac ttgaaaaata cggactttct 1560
gattacctcg ctattaaaca aaaatactac gaccaatacc aagcaaacaa aaactag 1617

<210> 34

<211> 538

<212> PRT

<213> Streptococcus pneumoniae

<400> 34

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

1 5 10 15

Leu Ala Val Leu Ala Ala Cys Gly Ser Lys Asn Thr Ala Ser Ser Pro

20 25 30

Asp Tyr Lys Leu Glu Gly Val Thr Phe Pro Leu Gln Glu Lys Lys Thr

35 40 45

Leu Lys Phe Met Thr Ala Ser Ser Pro Leu Ser Pro Lys Asp Pro Asn

50 55 60

Glu Lys Leu Ile Leu Gln Arg Leu Glu Lys Glu Thr Gly Val His Ile

65 70 75 80

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

 Glu Lys Pro Glu Tyr Lys Ala Leu Met Thr Ala Pro Asp Gly His Ile
 145 150 155 160
 Tyr Ser Phe Pro Trp Ile Glu Glu Leu Gly Asp Gly Lys Glu Ser Ile
 165 170 175
 His Ser Val Asn Asp Met Ala Trp Ile Asn Lys Asp Trp Leu Lys Lys
 180 185 190
 Leu Gly Leu Glu Met Pro Lys Thr Thr Asp Asp Leu Ile Lys Val Leu
 195 200 205

 Glu Ala Phe Lys Asn Gly Asp Pro Asn Gly Asn Gly Glu Ala Asp Glu
 210 215 220
 Ile Pro Phe Ser Phe Ile Ser Gly Asn Gly Asn Glu Asp Phe Lys Phe
 225 230 235 240
 Leu Phe Ala Ala Phe Gly Ile Gly Asp Asn Asp Asp His Leu Val Val
 245 250 255
 Gly Asn Asp Gly Lys Val Asp Phe Thr Ala Asp Asn Asp Asn Tyr Lys
 260 265 270

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

325 330 335
 Gly Gln Lys His Val Ala Arg Thr Asn Gly Met Gly Phe Ala Arg Asp
 340 345 350
 Lys Met Val Ile Thr Ser Val Asn Lys Asn Leu Glu Leu Thr Ala Lys
 355 360 365
 Trp Ile Asp Ala Gln Tyr Ala Pro Leu Gln Ser Val Gln Asn Asn Trp
 370 375 380
 Gly Thr Tyr Gly Asp Asp Lys Gln Gln Asn Ile Phe Glu Leu Asp Gln
 385 390 395 400

 Ala Ser Asn Ser Leu Lys His Leu Pro Leu Asn Gly Thr Ala Pro Ala
 405 410 415
 Glu Leu Arg Gln Lys Thr Glu Val Gly Gly Pro Leu Ala Ile Leu Asp
 420 425 430
 Ser Tyr Tyr Gly Lys Val Thr Thr Met Pro Asp Asp Ala Lys Trp Arg
 435 440 445
 Leu Asp Leu Ile Lys Glu Tyr Tyr Val Pro Tyr Met Ser Asn Val Asn
 450 455 460

 Asn Tyr Pro Arg Val Phe Met Thr Gln Glu Asp Leu Asp Lys Ile Ala
 465 470 475 480
 His Ile Glu Ala Asp Met Asn Asp Tyr Ile Tyr Arg Lys Arg Ala Glu
 485 490 495
 Trp Ile Val Asn Gly Asn Ile Asp Thr Glu Trp Asp Asp Tyr Lys Lys
 500 505 510
 Glu Leu Glu Lys Tyr Gly Leu Ser Asp Tyr Leu Ala Ile Lys Gln Lys
 515 520 525

 Tyr Tyr Asp Gln Tyr Gln Ala Asn Lys Asn
 530 535
 <210> 35
 <211> 1248
 <212> DNA
 <213> Streptococcus mutans

<400> 35

atgaaaacat ggcaaaaaat cgtcgttggc ggtgcaggcc ttatgcttgc aagcagtatt	60
cttgttgcct gtggatcaaa ggattcaaaa tcaagttcat ctgatcccaa aaccattaaa	120
ctttgggttc caacaggagc caagaaatct tatcaaagta ttgttcacaa atttgaaaag	180
gattctaact ataaagtaaa gattattgaa tctgaagacc caaaagctca ggaaaagatc	240
aaaaaagatc ctagtactgc tgcagatgtt ttctcgtgc cgcatgatca gctgggccag	300
ttagttgact ctggtgttat ccaagagatt cctcaaaaat attcaaaaga aataaataaa	360
aatgaaacac agcaggctgc aacaggagct atgtacaaag gtaagactta tgcttttcct	420
tttggaatcg agtctcaagt actttactat aataaatcaa aactctcagc tgatgatgtc	480
acatcatatg agactattac cagcaaggca actttcggag caaaattcaa acaagttaat	540
gcctatgcga ctgcaccact tttctattca gtaggtgata cactctttgg taaaaatggc	600
gaagatgcc aaggaactaa ctggggaaat gatgctggtg tatctgtttt gaaatggatt	660
gccagtcaaa aaggtaacgc tggctttgtc aatcttgacg ataacaatgt catgtctaaa	720
tttggtgatg gtctctgtagc ttcttttgaa tcaggtcctt gggattatga agccgcacaa	780
aaggcagttg gcaaaaacaa cctcgggtgtt acggtttatc caacaataaa tattaatggt	840
caagaagttc aacagaaagc tttcttaggt gttaaactct acgctgttaa tcaagctcct	900
tctaaaggaa ataccaaacg tattgctgct agttataaat tagcttctta cttacaagt	960
gctgaaagcc aagaaaatca atttaagaca aaaggacgca acatcatccc atctaataag	1020
accgttcaaa actctgatac agtcaaaaat catgaactcg cacaggctgt tatccaatg	1080
ggatcttctt cagattatac tgttgttatg cctaaactca accaaatgtc aacattctgg	1140
acggaaagcg cagctattct tagtgatact tacaatggta aaattaaaga aagtgattac	1200
cttgctaatt taaaacaatt tgataaagat ttagcagctg ctaaataa	1248

<210> 36

<211> 415

<212> PRT

<213> Streptococcus mutans

<400> 36

Met	Lys	Thr	Trp	Gln	Lys	Ile	Val	Val	Gly	Gly	Ala	Gly	Leu	Met	Leu
1					5				10				15		
Ala	Ser	Ser	Ile	Leu	Val	Ala	Cys	Gly	Ser	Lys	Asp	Ser	Lys	Ser	Ser
					20				25				30		
Ser	Ser	Asp	Pro	Lys	Thr	Ile	Lys	Leu	Trp	Val	Pro	Thr	Gly	Ala	Lys

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

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

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

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

Leu Gly Val Lys Leu Tyr Ala Val Asn Gln Ala Pro Ser Lys Gly Asn

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

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

<210> 37

<211> 1362

<212> DNA

<213> Streptococcus mutans

<400> 37

atgattcagt catcttctca tgatcagtta tctgtacttg aaacttttaa aaagggcggg	60
atagatatca aattatcggtt tgtcatcatg ggatttgcca atttgatgaa taagcaattc	120
ataaaaggcc tctcttttct attaagttag atagcttttc taattgcttt tgtcacacag	180
gttattccag ctttttcagg ctacttcaact ctcggtacta aaacacaagg gatgcaagaa	240
aaaattgtgg atggcggttaa attacagggtg gcagttgaag gcgataattc gatgctgatg	300
ctcatttttg gattagcctc actaatcttt tgtttggttt ttgcctacat ttattgggtg	360
aatcttaaaa gtgccagaaa tctctatatg ttaaaaaaag agggacgtca cattccatct	420
ttcaaagaag attttatgac tttggcaaac ggccgattcc atatgacttt gatgtttatt	480
cctttgattg gtgttcttct ttttaccatt ttgccactcg tttatatgat ttgcctggcc	540
ttaccaatt atgatcacia tcatcttccg cctaaatccc tttttgattg ggtaggggtg	600

gctaattttg gtaatgtttt gaatggccgc atggctggaa cttctttccc tgcctttct 660
 tggacactta tctgggcigt ttctgcaact gtgacaaact ttctttttgg agtcacttg 720
 gcacttatta tcaatgctaa gggattaaaa ttgaaaaaaa tgtggcggac tatctttgtt 780
 attaccattg ctgtgccgca gttcatttca cttttgctga tgagaaattt ccttaatgat 840

caaggtccgc tcaatgcttt cctagaaaaa attggcctga tttctcattc tctgccattt 900
 ctatcagatc ctacttgggc aaaattttca attatcttcg ttaatatgtg ggttggtatt 960
 ccttttacca tgtagtcgc aacaggaatt atcatgaatc ttccgagtga gcaaattgag 1020
 gctgcagaaa ttgacggcgc tagtaagttc caaattttta aatccatcac ttccccgag 1080
 attcttttaa ttatgatgcc atctttaatc cagcaattta ttggaaatat caataatttt 1140
 aatgtcatct accttttaac cggtaggcgga ccaactaatt cacaattcta tcaagcaggc 1200
 agcacagact tattggtcac ttggctttat aaactaacia tgaatgctgc agactataat 1260

ttagcttctg ttattggtat ctttatcttt gccatttcag ctatcttcag tcttttagct 1320
 tatacgcata cagcatcata caaggaagga gctgttaaata aa 1362

<210> 38

<211> 453

<212> PRT

<213> Streptococcus mutans

<400> 38

Met Ile Gln Ser Ser Ser His Asp Gln Leu Ser Val Leu Glu Thr Phe

1 5 10 15

Lys Lys Gly Gly Ile Asp Ile Lys Leu Ser Phe Val Ile Met Gly Phe

20 25 30

Ala Asn Leu Met Asn Lys Gln Phe Ile Lys Gly Leu Leu Phe Leu Leu

35 40 45

Ser Glu Ile Ala Phe Leu Ile Ala Phe Val Thr Gln Val Ile Pro Ala

50 55 60

Phe Ser Gly Leu Leu Thr Leu Gly Thr Lys Thr Gln Gly Met Gln Glu

65 70 75 80

Lys Ile Val Asp Gly Val Lys Leu Gln Val Ala Val Glu Gly Asp Asn

85 90 95

Ser Met Leu Met Leu Ile Phe Gly Leu Ala Ser Leu Ile Phe Cys Leu

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

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

420 425 430
 Ser Ala Ile Phe Ser Leu Leu Ala Tyr Thr His Thr Ala Ser Tyr Lys
 435 440 445
 Glu Gly Ala Val Lys

450

<210> 39

<211> 837

<212> DNA

<213> Streptococcus mutans

<400> 39

atgaaaagaa aaaacaact tcagatcggc tctatctatg ctttactgat tctcttatcc	60
ttcatttggc tatttccgat catttgggtt atactgacga gttttcgcgg tgaaggcaca	120
gcttatgttc cttatattat tccaaaaacg tggacttttag ataattatat taaattattt	180
accaattctt ctttccatt tggacgtgg tttttaaata ccttaatcgt ttcaacagcc	240
acttgtgttc tgicaacttc tatcacagt gcaatggctt attcgcttag ccgtattaaa	300
tttaaacacc gtaacggctt tttaaaatta gctcttggtc tgaatatgtt tccgggattt	360
atgagtatga ttgcagttta ctacattcta aaagcactca atctcaccca aacattaaca	420
tctcttggtt tggctctatc ttcaggagct gccttaactt tctatatcgc taaaggcttt	480
tttgatacga ttccttattc attggatgaa tcagctatga ttgatggggc tacgcgtaaa	540
gatattttct taaaaatcac tctgccgcta tctaagccca tcatcgttta tacggccctg	600
ttggcattta ttgcccttg gattgacttt atttttgctc aggttattct tggagatgcc	660
accagcaa ataccgtagc gattggactc ttctctatgc ttcaagctga taccattaat	720
aattggttca tggcctttgc agcaggttct gtactgatcg ccattccaat cacgatactt	780

tttatcttca tgcaaaagta ttacgttgaa ggcattactg gcggatctgt taaataa 837

<210> 40

<211> 278

<212> PRT

<213> Streptococcus mutans

<400> 40

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

1 5 10 15

Ile Leu Leu Ser Phe Ile Trp Leu Phe Pro Ile Ile Trp Val Ile Leu

20 25 30

Thr Ser Phe Arg Gly Glu Gly Thr Ala Tyr Val Pro Tyr Ile Ile Pro

35 40 45

Lys Thr Trp Thr Leu Asp Asn Tyr Ile Lys Leu Phe Thr Asn Ser Ser

50 55 60

Phe Pro Phe Gly Arg Trp Phe Leu Asn Thr Leu Ile Val Ser Thr Ala

65 70 75 80

Thr Cys Val Leu Ser Thr Ser Ile Thr Val Ala Met Ala Tyr Ser Leu

85 90 95

Ser Arg Ile Lys Phe Lys His Arg Asn Gly Phe Leu Lys Leu Ala Leu

100 105 110

Val Leu Asn Met Phe Pro Gly Phe Met Ser Met Ile Ala Val Tyr Tyr

115 120 125

Ile Leu Lys Ala Leu Asn Leu Thr Gln Thr Leu Thr Ser Leu Val Leu

130 135 140

Val Tyr Ser Ser Gly Ala Ala Leu Thr Phe Tyr Ile Ala Lys Gly Phe

145 150 155 160

Phe Asp Thr Ile Pro Tyr Ser Leu Asp Glu Ser Ala Met Ile Asp Gly

165 170 175

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

180 185 190

Pro Ile Ile Val Tyr Thr Ala Leu Leu Ala Phe Ile Ala Pro Trp Ile

195 200 205

Asp Phe Ile Phe Ala Gln Val Ile Leu Gly Asp Ala Thr Ser Lys Tyr
210 215 220

Thr Val Ala Ile Gly Leu Phe Ser Met Leu Gln Ala Asp Thr Ile Asn
225 230 235 240

Asn Trp Phe Met Ala Phe Ala Ala Gly Ser Val Leu Ile Ala Ile Pro
245 250 255

Ile Thr Ile Leu Phe Ile Phe Met Gln Lys Tyr Tyr Val Glu Gly Ile
260 265 270

Thr Gly Gly Ser Val Lys
275

<210> 41

<211> 1134

<212> DNA

<213> Streptococcus mutans

<400> 41

atgacaactt taaaacttga taacatctac aaaagatatc ccaatgcaaa gcattattcc 60
gttgaaaatt ttaatcttga cattcatgat aaagaattta ttgtctttgt cggtccttca 120
ggatgcggaa agtcaaccac tcttcgcatg attgctgggc tggaagatat tacagaaggc 180
aacctttata ttgatgataa actcatgaat gatgcctctc ctaaagatcg cgatattgct 240
atggtttttc aaaattatgc tctttatcct catatgagcg tttatgaaaa tatggctttt 300
ggcctaaaac ttcgtaata caaaaaagat gatattaata aacgtgtaca cgaagctgct 360

gaaattcttg gactgacaga atttcttgaa agaaagcctg cggacctctc tggcggacag 420
cggcagcggg ttgctatggg acgtgctatt gtccgagatg ctaaggtctt cttaatggac 480
gaacctttgt caaattdaga tgccaaactt cgagttgcca tgcgagccga aatcgctaaa 540
attcaccgcc gcattggggc aacgactatc tatgttacc atgaccaaac agaagccatg 600
accttagcag atcgatttgt tatcatgagc gctactccaa acccagataa aaccggctct 660
atcggtcgta ttgagcagat tggaacacca caggaactct acaatgaacc tgctaataaa 720
tttgttgctg gcttcatcgg aagccccgct atgaatttct ttgaagtga cgttgaaaaa 780

gagcgtttgg ttaaccaaga tggcttaagc cttgcgcttc ctcagggaca ggaaaaaatt 840
cttgaggaga aaggttatct tggtaaaaaa gtcactttag gtattcgacc agaagacatc 900
tcaagtgatc aaattgtcca cgagactttc ccaaatgcca gtgttacagc tgacatacta 960

gtatcagaac ttttaggcag cgaaagcatg ttatatgtca aatttggcag tactgaattt 1020
acagctcgcg tcaatgctcg tgactctcac agtcccgag aaaaagtaca attaaccttt 1080
aatattgcta agggacactt ctttgattta gagactgaaa aacgaatcaa ttaa 1134

<210> 42

<211> 377

<212> PRT

<213> Streptococcus mutans

<400> 42

Met Thr Thr Leu Lys Leu Asp Asn Ile Tyr Lys Arg Tyr Pro Asn Ala

1 5 10 15

Lys His Tyr Ser Val Glu Asn Phe Asn Leu Asp Ile His Asp Lys Glu

20 25 30

Phe Ile Val Phe Val Gly Pro Ser Gly Cys Gly Lys Ser Thr Thr Leu

35 40 45

Arg Met Ile Ala Gly Leu Glu Asp Ile Thr Glu Gly Asn Leu Tyr Ile

50 55 60

Asp Asp Lys Leu Met Asn Asp Ala Ser Pro Lys Asp Arg Asp Ile Ala

65 70 75 80

Met Val Phe Gln Asn Tyr Ala Leu Tyr Pro His Met Ser Val Tyr Glu

85 90 95

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

100 105 110

Asn Lys Arg Val His Glu Ala Ala Glu Ile Leu Gly Leu Thr Glu Phe

115 120 125

Leu Glu Arg Lys Pro Ala Asp Leu Ser Gly Gly Gln Arg Gln Arg Val

130 135 140

Ala Met Gly Arg Ala Ile Val Arg Asp Ala Lys Val Phe Leu Met Asp

145 150 155 160

Glu Pro Leu Ser Asn Leu Asp Ala Lys Leu Arg Val Ala Met Arg Ala

165 170 175

Glu Ile Ala Lys Ile His Arg Arg Ile Gly Ala Thr Thr Ile Tyr Val

180 185 190

Thr His Asp Gln Thr Glu Ala Met Thr Leu Ala Asp Arg Ile Val Ile
195 200 205

Met Ser Ala Thr Pro Asn Pro Asp Lys Thr Gly Ser Ile Gly Arg Ile
210 215 220

Glu Gln Ile Gly Thr Pro Gln Glu Leu Tyr Asn Glu Pro Ala Asn Lys
225 230 235 240

Phe Val Ala Gly Phe Ile Gly Ser Pro Ala Met Asn Phe Phe Glu Val
245 250 255

Thr Val Glu Lys Glu Arg Leu Val Asn Gln Asp Gly Leu Ser Leu Ala
260 265 270

Leu Pro Gln Gly Gln Glu Lys Ile Leu Glu Glu Lys Gly Tyr Leu Gly
275 280 285

Lys Lys Val Thr Leu Gly Ile Arg Pro Glu Asp Ile Ser Ser Asp Gln
290 295 300

Ile Val His Glu Thr Phe Pro Asn Ala Ser Val Thr Ala Asp Ile Leu
305 310 315 320

Val Ser Glu Leu Leu Gly Ser Glu Ser Met Leu Tyr Val Lys Phe Gly
325 330 335

Ser Thr Glu Phe Thr Ala Arg Val Asn Ala Arg Asp Ser His Ser Pro
340 345 350

Gly Glu Lys Val Gln Leu Thr Phe Asn Ile Ala Lys Gly His Phe Phe
355 360 365

Asp Leu Glu Thr Glu Lys Arg Ile Asn
370 375

<210> 43

<211> 927

<212> DNA

<213> Agrobacterium tumefaciens

<400> 43

atgatcctgt gttgtggtga agccctgata gacatgctgc cccggcagac gacgctgggt 60

gaggcgggct ttgccctta cgcaggcgga gcggtcttca acacggcaat tgcgctgggg 120

cgctcttgcg tcccttcagc cttttttacc ggtctttccg acgacatgat gggcgatata 180
ctgcgggaga ccttcggggc cagcaagggtg gatttcagct attgcgccac cctgtcgcgc 240
cccaccacca ttgcgttcgt taagctgggt gatggccatg cgacctacgc tttttacgac 300
gagaacaccg ccggccggat gatcaccgag gccgaacttc cggccttggg agcggattgc 360
gaagcgtgc atttcggcgc catcagcctt attcccgaac cctgcggcag cacctatgag 420

gcgctgatga cgcgcgagca tgagacccgc gtcactctgc tcgatccgaa cattcgtccc 480
ggcttcatcc agaacaagca gtcgcacatg gccgcgcatcc gccgcattgc ggcgatgtct 540
gacatcgtca agttctcgga tgaggacctg gctggttctg gtctggaagg cgacgaggac 600
acgcttgccc gccactggct gcaccacggt gcaaaactcg tcgttgtcac ccgtggcgcc 660
aagggtgccg tgggttacag cgccaatctc aaggtggaag tggcctccga gcgcgtcgaa 720
gtggtcgata cggtcggcgc cggcgatacg ttcatgccc gcatcttgc ttcgctgaaa 780
atgcagggcc tgcagacaa agcgcagggt gcttcgctga gcgaagagca gatcagaaaa 840

gctttggcgc ttggcgcgaa agccgctgcg gtcactgtct cgcggtctgg cgcaaatccg 900
cctttcgcgc atgaaatcgg tttgtga 927

<210> 44

<211> 308

<212> PRT

<213> Agrobacterium tumefaciens

<400> 44

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

1 5 10 15

Thr Thr Leu Gly Glu Ala Gly Phe Ala Pro Tyr Ala Gly Gly Ala Val

20 25 30

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

35 40 45

Phe Thr Gly Leu Ser Asp Asp Met Met Gly Asp Ile Leu Arg Glu Thr

50 55 60

Leu Arg Ala Ser Lys Val Asp Phe Ser Tyr Cys Ala Thr Leu Ser Arg

65 70 75 80

Pro Thr Thr Ile Ala Phe Val Lys Leu Val Asp Gly His Ala Thr Tyr

85 90 95

Ala Phe Tyr Asp Glu Asn Thr Ala Gly Arg Met Ile Thr Glu Ala Glu

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

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

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

290 295 300
 Glu Ile Gly Leu

305

<210> 45

<211> 1404

<212> DNA

<213> Streptococcus mutans

<400> 45

cagctgatta tgcgtcagtt gaaaccctcg cttcttcagg aactgttgct gtaggtgata	60
gcttacttga agttaaaaa taagaatat tatcagaaag accgtaaggt ctttttgact	120
gcttaaaaga ttcagtaaca atagtattaa agccttttgg ctaactaata ctgaaattt	180
agcaaattat gatataatgt taagtagtcc ttaagggtag attaagggtt tcaaatcca	240
aaaattgatt tggttaagtt agtaaaatat aagaggttta ttatgtctaa attatatggc	300
agcatcgaag ctggcggaac aaaatttgtc tgtgctgtag gtgatgaaaa ttttcaaatt	360
ttagaaaaag ttcagttccc aacaacaaca ctttatgaaa caatagaaaa aacagttgct	420
ttctttaaaa aatttgaagc tgatttagcc agtgttgcca ttggttcttt tggccctatt	480
gatattgatc aaaattcaga cacttatggt tacattactt caacaccaaa gccaaactgg	540
gctaactgtg attttgcgg ctttaatttct aaagatttta aaattccatt ttactttacg	600
acagatgtta attcttctgc ttatggggaa acaattgctc gttcaaatgt taaaagtctg	660
gtttattata ctattggaac aggcattgga gcaggggcta ttcaaaatgg cgaattcatt	720
ggcggtatgg gacatacga agctggacac gtttcatgg ctccgcatcc caatgatgtt	780
catcatggtt ttgtaggcac ctgtcctttc cataaaggct gtttagaagg acttgcagcg	840
ggtcctagct tagaggctcg tactggtatt cgtggtgagt taattgagca aaactcagaa	900
gtttgggata ttcaggcata ctacattgct caggcggcta ttcaagcgac tgtcctttat	960
cgcccgcaag tcattgtatt tggcggaggc gttatggcac aagaacatat gctcaatcgg	1020
gttcgtgaaa aatttacttc acttttgaat gactatcttc cagttccaga tgttaaagat	1080
tatattgtga caccagctgt tgcagaaaat ggttcagcaa cattgggaaa tctcgttta	1140
gctaaaaaga tagcagcgcg ttaattaaaa atgaattgga agattaaagc accttcta	1200
attcaatatt aaactgttag aatttacttg aacgaaattt tcattttatg aggataatga	1260
agtgaatata attactcttg atttcctctg aaactagata gtggtatatt gaaaaacaga	1320
aaggagaaca ctatggaagg acctttgttt ttacaatcac aaatgcataa aaaaatctgg	1380
ggcggcaatc ggctcagaaa agaa	1404

<210> 46

<211> 293

<212> PRT

<213> Streptococcus mutans

<400> 46

Met Ser Lys Leu Tyr Gly Ser Ile Glu Ala Gly Gly Thr Lys Phe Val

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

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

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

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

260	265	270	
Val Ala Glu Asn Gly Ser Ala Thr Leu Gly Asn Leu Ala Leu Ala Lys			
275	280	285	
Lys Ile Ala Ala Arg			
290			
<210> 47			
<211> 915			
<212> DNA			
<213> Escherichia coli			
<400> 47			
atgtcagcca aagtatgggt tttaggggat gcggtcgtag atctcttgcc agaatcagac	60		
gggcgcctac tgccttgccc tggcggcgcg ccagctaacg ttgcggtggg aatcgccaga	120		
ttaggcggaa caagtgggtt tataggtcgg gtgggggatg atccttttgg tgcgttaatg	180		
caaagaacgc tgctaactga gggagtcgat atcacgtatc tgaagcaaga tgaatggcac	240		
cggacatcca cggctgcttg cgatctgaac gatcaagggg aacgttcatt tacgtttatg	300		
gtccgcccc gtgccgatct ttttttagag acgacagact tgccctgctg gcgacatggc	360		
gaatggttac atctctgttc aattgcgttg tctgccgagc cttcgcgtac cagcgcattt	420		
actgcgatga cggcgatccg gcatgccgga ggttttgtca gcttcgatcc taatattcgt	480		
gaagatctat ggcaagacga gcatttgctc cgcttgtgtt tgcggcaggc gctacaactg	540		
gcggatgtcg tcaagctctc ggaagaagaa tggcgactta tcagtggaaa aacacagaac	600		
gatcaggata tatgcgcctt ggcaaaagag tatgagatcg ccatgctgtt ggtgactaaa	660		
ggtgcagaag ggggtggtgt ctgttatcga ggacaagtgc accatcttgc tggaatgtct	720		
gtgaattgtg tcgatagcac gggggcggga gatgcgttcg ttgccgggtt actcacaggt	780		
ctgtcctcta cgggattatc tacagatgag agagaaatgc gacgaattat cgatctcgct	840		
caacgttgcg gagcgcttgc agtaacggcg aaaggggcaa tgacagcgct gccatgtcga	900		
caagaactgg aatag	915		
<210> 48			
<211> 304			
<212> PRT			
<213> Escherichia coli			
<400> 48			

Met Ser Ala Lys Val Trp Val Leu Gly Asp Ala Val Val Asp Leu Leu
1 5 10 15
Pro Glu Ser Asp Gly Arg Leu Leu Pro Cys Pro Gly Gly Ala Pro Ala
20 25 30
Asn Val Ala Val Gly Ile Ala Arg Leu Gly Gly Thr Ser Gly Phe Ile
35 40 45
Gly Arg Val Gly Asp Asp Pro Phe Gly Ala Leu Met Gln Arg Thr Leu
50 55 60
Leu Thr Glu Gly Val Asp Ile Thr Tyr Leu Lys Gln Asp Glu Trp His
65 70 75 80
Arg Thr Ser Thr Val Leu Val Asp Leu Asn Asp Gln Gly Glu Arg Ser
85 90 95
Phe Thr Phe Met Val Arg Pro Ser Ala Asp Leu Phe Leu Glu Thr Thr
100 105 110
Asp Leu Pro Cys Trp Arg His Gly Glu Trp Leu His Leu Cys Ser Ile
115 120 125
Ala Leu Ser Ala Glu Pro Ser Arg Thr Ser Ala Phe Thr Ala Met Thr
130 135 140
Ala Ile Arg His Ala Gly Gly Phe Val Ser Phe Asp Pro Asn Ile Arg
145 150 155 160
Glu Asp Leu Trp Gln Asp Glu His Leu Leu Arg Leu Cys Leu Arg Gln
165 170 175
Ala Leu Gln Leu Ala Asp Val Val Lys Leu Ser Glu Glu Glu Trp Arg
180 185 190
Leu Ile Ser Gly Lys Thr Gln Asn Asp Gln Asp Ile Cys Ala Leu Ala
195 200 205
Lys Glu Tyr Glu Ile Ala Met Leu Leu Val Thr Lys Gly Ala Glu Gly
210 215 220
Val Val Val Cys Tyr Arg Gly Gln Val His His Phe Ala Gly Met Ser
225 230 235 240
Val Asn Cys Val Asp Ser Thr Gly Ala Gly Asp Ala Phe Val Ala Gly
245 250 255

Leu Leu Thr Gly Leu Ser Ser Thr Gly Leu Ser Thr Asp Glu Arg Glu
260 265 270

Met Arg Arg Ile Ile Asp Leu Ala Gln Arg Cys Gly Ala Leu Ala Val
275 280 285

Thr Ala Lys Gly Ala Met Thr Ala Leu Pro Cys Arg Gln Glu Leu Glu
290 295 300

<210> 49

<211> 879

<212> DNA

<213> Enterococcus faecalis

<400> 49

atgacagaaa aacttttagg aagtatcgaa gccggtggca caaaatttgt atgtggcggtt 60

gggacagatg atttgaccat cgtagaacgt gtcagttttc ccacaacaac cccagaagaa 120

acaatgaaaa aagtaataga atttttccaa caatatcctt taaaagcgat tgggatttgt 180

tcatttggtc cgattgatat tcacgttgat tctctacgt atggttatat cacttctaca 240

ccaaaattag ctiggcgtaa ctttgacttg ttaggaacta tgaacaaca ttttgatgtg 300

ccaatggctt ggacaacgga tgtgaatgct gcggcatatg gtgagtatgt tgctggaaat 360

gggcaacata catctagtgt tgtatattat acaattggaa ctgggtgttg cgctggagcg 420

attcaaaacg gtgagtttat tgaaggcttt agccaccag aatggggca tgcgttagtt 480

cgctgcatc ctgaagatac gtatgcagga aattgtcctt atcatggaga ttgtttagaa 540

gggattgcag caggaccagc agttgaaggt cgttctggta aaaaaggaca tttattggaa 600

gaggatcata aaacttggga attagaagct tattatttag cgcaagcggc gtacaatacg 660

actttattat tagcgccaga agtgcattt ttaggtggcg gcgtcatgaa acaacgtcat 720

ttgatgccga aagttcgtga aaaatttgct gaattagtca atggatatgt ggaaacaccg 780

cctttagaaa aatacttggg gacgcctctt ttagaagata atccaggaac aatcggttgc 840

tttgccttgg caaaaaaagc tttaatggct caaaaataa 879

<210> 50

<211> 292

<212> PRT

<213> Enterococcus faecalis

<400> 50

Met Thr Glu Lys Leu Leu Gly Ser Ile Glu Ala Gly Gly Thr Lys Phe
1 5 10 15
Val Cys Gly Val Gly Thr Asp Asp Leu Thr Ile Val Glu Arg Val Ser
20 25 30
Phe Pro Thr Thr Thr Pro Glu Glu Thr Met Lys Lys Val Ile Glu Phe
35 40 45
Phe Gln Gln Tyr Pro Leu Lys Ala Ile Gly Ile Gly Ser Phe Gly Pro
50 55 60
Ile Asp Ile His Val Asp Ser Pro Thr Tyr Gly Tyr Ile Thr Ser Thr
65 70 75 80
Pro Lys Leu Ala Trp Arg Asn Phe Asp Leu Leu Gly Thr Met Lys Gln
85 90 95
His Phe Asp Val Pro Met Ala Trp Thr Thr Asp Val Asn Ala Ala Ala
100 105 110
Tyr Gly Glu Tyr Val Ala Gly Asn Gly Gln His Thr Ser Ser Cys Val
115 120 125
Tyr Tyr Thr Ile Gly Thr Gly Val Gly Ala Gly Ala Ile Gln Asn Gly
130 135 140
Glu Phe Ile Glu Gly Phe Ser His Pro Glu Met Gly His Ala Leu Val
145 150 155 160
Arg Arg His Pro Glu Asp Thr Tyr Ala Gly Asn Cys Pro Tyr His Gly
165 170 175
Asp Cys Leu Glu Gly Ile Ala Ala Gly Pro Ala Val Glu Gly Arg Ser
180 185 190
Gly Lys Lys Gly His Leu Leu Glu Glu Asp His Lys Thr Trp Glu Leu
195 200 205
Glu Ala Tyr Tyr Leu Ala Gln Ala Ala Tyr Asn Thr Thr Leu Leu Leu
210 215 220
Ala Pro Glu Val Ile Ile Leu Gly Gly Gly Val Met Lys Gln Arg His
225 230 235 240
Leu Met Pro Lys Val Arg Glu Lys Phe Ala Glu Leu Val Asn Gly Tyr

245	250	255	
Val Glu Thr Pro Pro Leu Glu Lys Tyr Leu Val Thr Pro Leu Leu Glu			
260	265	270	
Asp Asn Pro Gly Thr Ile Gly Cys Phe Ala Leu Ala Lys Lys Ala Leu			
275	280	285	
Met Ala Gln Lys			
290			
<210> 51			
<211> 1458			
<212> DNA			
<213> Saccharomyces cerevisiae			
<400> 51			
ttaagcgcca atgataccaa gagacttacc ttcggaatt cttttttcgg acaatgcagc	60		
aataacagca gcacctgcac ctgaaccatc ctgagctgga acaatcgtaa ttggatcttt	120		
gcttgcgta ccagtcacat catagatatc tctcaaacc tttagcgcggt cttccttgaa	180		
acctgggtat ttgttataga cagaaccgtc agcggcaatg tgaccagtct tgtaacctct	240		
cttttgcaa atagcggaac taccacaaac agctaacta gcagctctgg taccgatcaa	300		
ttcacaagt cttctaata actttacgtt ttgagagtg gtcttgacac caaagtcctt	360		
ttggaagatg tcatcagtat cttcaagtt ttcaaatgga tcatcctga ttcttgctgg	420		
gtaggaggta tccatgatgt atggttggtt caacttgctt agatcttgat cttcaacat	480		
caagcccttc tcgtttaatt caagtaacac tagacgcaac aattcaccca agtagtaacc	540		
ggaggtcatc ttttcaaaag cttgttgacc aggtcttgga gattgttcgt cgacagcaac	600		
atcgtaactg gttcttgga agacaaatg ttcatatcg aaggaacct attcacaatt	660		
gatagccatt ggagagttac ttggaatgc gtctgctaatt ttgcctcca acttttcgat	720		
atcggaaca acatcataga aagcaccgtt gacaccagta ccgaaaatca cacccatctt	780		
agtctctggg tcagtgtagt atgaggcaat taaagtacca acagtatcat taatcaatgc	840		
tacaatttca ataggcaact ctctcttgga aatttcgttt ttagcaatg ggacgacatc	900		
gtggccttcg acatttgga tctgaaacc cttggtccat ctttgcaaaa taccttcgtt	960		
aatcttggtt tgggaagctg ggtacagaa ggtgaaacct aatggtgaagg tgccttggt	1020		
gtttagcaat tcttgctga ccataaagtc cttcaaagag tcggcaataa aggaccataa	1080		
ctcctcttg tgccttagtg ttctcatgtc atgtggtagt ttatcttgg attgagtgg	1140		
gtcaaaggta tggttaccgc tcaacttgac caacacgact ctttaagttg taccaccaa	1200		

atcaatggcc aaatagttac cagattcttt acctgttggg aattccatga cccaaccggg 1260
aatcattgga atgttacctc cttctttgtt caaaccttta ttcaattcgt cgataaagtg 1320

cttaacaacc tttctcaagg tctcgctgtc aactgtaaac atatcttcca actgatgaat 1380
ttcatccatc aattccttgg gcacatcagc catggaaccc tttctagcct gtggtttctt 1440
tggacctaaa tgaaccat 1458

<210> 52

<211> 485

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 52

Met Val His Leu Gly Pro Lys Lys Pro Gln Ala Arg Lys Gly Ser Met

1 5 10 15

Ala Asp Val Pro Lys Glu Leu Met Asp Glu Ile His Gln Leu Glu Asp

20 25 30

Met Phe Thr Val Asp Ser Glu Thr Leu Arg Lys Val Val Lys His Phe

35 40 45

Ile Asp Glu Leu Asn Lys Gly Leu Thr Lys Lys Gly Gly Asn Ile Pro

50 55 60

Met Ile Pro Gly Trp Val Met Glu Phe Pro Thr Gly Lys Glu Ser Gly

65 70 75 80

Asn Tyr Leu Ala Ile Asp Leu Gly Gly Thr Asn Leu Arg Val Val Leu

85 90 95

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

100 105 110

Lys Leu Pro His Asp Met Arg Thr Thr Lys His Gln Glu Glu Leu Trp

115 120 125

Ser Phe Ile Ala Asp Ser Leu Lys Asp Phe Met Val Glu Gln Glu Leu

130 135 140

Leu Asn Thr Lys Asp Thr Leu Pro Leu Gly Phe Thr Phe Ser Tyr Pro

145 150 155 160

Ala Ser Gln Asn Lys Ile Asn Glu Gly Ile Leu Gln Arg Trp Thr Lys

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

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

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

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

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

Gly Ile Ile Gly Ala
 485

<210> 53

<211> 1461

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 53

atggttcatt taggtccaaa aaaaccacaa gccagaaagg gttccatggc cgatgtgcca	60
aaggaattga tgcaacaaat tgagaatttt gaaaaaattt tcactgttcc aactgaaact	120
ttacaagccg ttaccaagca cttcatittcc gaattggaaa agggtttgtc caagaagggt	180
ggtaacattc caatgattcc aggttgggtt atggatttcc caactggtaa ggaatccggt	240
gatttcttgg ccattgattt ggggtgtacc aacttgagag ttgtcttagt caagttgggc	300
ggtgaccgta ctttgacac cactcaatct aagtacagat taccagatgc tatgagaact	360
actcaaaatc cagacgaatt gtgggaattt attgccgact ctttgaaagc ttttattgat	420
gagcaattcc cacaaggat ctctgagcca attccattgg gtttcacctt ttctttccca	480
gcttctcaaa acaaaatcaa tgaaggtatc ttgcaaagat ggactaaagg ttttgatatt	540
ccaaacattg aaaaccacga tgttgttcca atgttgcaaa agcaaatcac taagaggaat	600
atcccaattg aagtgtgtgc ttgataaac gacactaccg gtacttttgt tgcttttac	660
tacactgacc cagaaactaa gatgggtgtt atcttcggta ctggigtcaa tgggtgcttac	720
tacgatgttt gttccgatat cgaaaagcta caaggaaaac tatctgatga cattccacca	780
tctgctccaa tggccatcaa ctgtgaatac ggttccttcg ataatgaaca tgtcgttttg	840
ccaagaacta aatacgatat caccattgat gaagaatctc caagaccagg ccaacaaacc	900
tttgaaaaaa tgcittctgg ttactactta ggtgaaattt tgcgtttggc ctgatggac	960
atgtacaaac aaggtttcat cttaagaac caagacttgt ctaagttcga caagcctttc	1020

gtcatggaca ctctttaccc agccagaatc gaggaagatc cattcgagaa cctagaagat 1080
accgatgact tgttcacaaa tgagttcggc atcaacacta ctgttcaaga acgtaaattg 1140
atcagacgtt tatctgaatt gattggtgct agagctgcta gattgtccgt ttgtggtatt 1200

gctgctatct gtcaaaagag aggttacaag accggtcaca tcgctgcaga cggttccgtt 1260
tacaacagat acccaggttt caaagaaaag gctgccaatg ctttgaagga catttacggc 1320
tggactcaaa cctcactaga cgactacca atcaagattg ttcctgctga agatggttcc 1380
ggtgctggtg ccgctgttat tgctgctttg gcccaaaaaa gaattgctga aggtaagtcc 1440
gttggtatca tcggtgctta a 1461

<210> 54

<211> 486

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 54

Met Val His Leu Gly Pro Lys Lys Pro Gln Ala Arg Lys Gly Ser Met

1 5 10 15
Ala Asp Val Pro Lys Glu Leu Met Gln Gln Ile Glu Asn Phe Glu Lys

20 25 30
Ile Phe Thr Val Pro Thr Glu Thr Leu Gln Ala Val Thr Lys His Phe

35 40 45
Ile Ser Glu Leu Glu Lys Gly Leu Ser Lys Lys Gly Gly Asn Ile Pro

50 55 60
Met Ile Pro Gly Trp Val Met Asp Phe Pro Thr Gly Lys Glu Ser Gly

65 70 75 80
Asp Phe Leu Ala Ile Asp Leu Gly Gly Thr Asn Leu Arg Val Val Leu

85 90 95
Val Lys Leu Gly Gly Asp Arg Thr Phe Asp Thr Thr Gln Ser Lys Tyr

100 105 110
Arg Leu Pro Asp Ala Met Arg Thr Thr Gln Asn Pro Asp Glu Leu Trp

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

130

135

140

Gln Gly Ile Ser Glu Pro Ile Pro Leu Gly Phe Thr Phe Ser Phe Pro
 145 150 155 160
 Ala Ser Gln Asn Lys Ile Asn Glu Gly Ile Leu Gln Arg Trp Thr Lys
 165 170 175
 Gly Phe Asp Ile Pro Asn Ile Glu Asn His Asp Val Val Pro Met Leu
 180 185 190
 Gln Lys Gln Ile Thr Lys Arg Asn Ile Pro Ile Glu Val Val Ala Leu
 195 200 205
 Ile Asn Asp Thr Thr Gly Thr Leu Val Ala Ser Tyr Tyr Thr Asp Pro
 210 215 220
 Glu Thr Lys Met Gly Val Ile Phe Gly Thr Gly Val Asn Gly Ala Tyr
 225 230 235 240
 Tyr Asp Val Cys Ser Asp Ile Glu Lys Leu Gln Gly Lys Leu Ser Asp
 245 250 255
 Asp Ile Pro Pro Ser Ala Pro Met Ala Ile Asn Cys Glu Tyr Gly Ser
 260 265 270
 Phe Asp Asn Glu His Val Val Leu Pro Arg Thr Lys Tyr Asp Ile Thr
 275 280 285
 Ile Asp Glu Glu Ser Pro Arg Pro Gly Gln Gln Thr Phe Glu Lys Met
 290 295 300
 Ser Ser Gly Tyr Tyr Leu Gly Glu Ile Leu Arg Leu Ala Leu Met Asp
 305 310 315 320
 Met Tyr Lys Gln Gly Phe Ile Phe Lys Asn Gln Asp Leu Ser Lys Phe
 325 330 335
 Asp Lys Pro Phe Val Met Asp Thr Ser Tyr Pro Ala Arg Ile Glu Glu
 340 345 350
 Asp Pro Phe Glu Asn Leu Glu Asp Thr Asp Asp Leu Phe Gln Asn Glu
 355 360 365
 Phe Gly Ile Asn Thr Thr Val Gln Glu Arg Lys Leu Ile Arg Arg Leu
 370 375 380
 Ser Glu Leu Ile Gly Ala Arg Ala Ala Arg Leu Ser Val Cys Gly Ile

385 390 395 400
 Ala Ala Ile Cys Gln Lys Arg Gly Tyr Lys Thr Gly His Ile Ala Ala
 405 410 415
 Asp Gly Ser Val Tyr Asn Arg Tyr Pro Gly Phe Lys Glu Lys Ala Ala
 420 425 430
 Asn Ala Leu Lys Asp Ile Tyr Gly Trp Thr Gln Thr Ser Leu Asp Asp
 435 440 445
 Tyr Pro Ile Lys Ile Val Pro Ala Glu Asp Gly Ser Gly Ala Gly Ala

450 455 460
 Ala Val Ile Ala Ala Leu Ala Gln Lys Arg Ile Ala Glu Gly Lys Ser
 465 470 475 480
 Val Gly Ile Ile Gly Ala

485

<210> 55

<211> 1374

<212> DNA

<213> Escherichia coli

<400> 55

ggtaacactt tctatcccca ttttcacctc gcgcctcctg ccgggtggat gaacgatcca 60
 aacggcctga tctggtttaa cgatcgttat caccgctttt atcaacatca cccgatgagc 120
 gaacactggg ggccaatgca ctggggacat gccaccagcg acgatatgat ccactggcag 180

 catgagccta ttgcgctagc gccaggagac gagaatgaca aagacgggtg tttttcaggt 240
 agtgctgtcg atgacaatgg tgctctctca cttatctaca ccggacacgt ctggctcgat 300
 ggtgcaggta atgacgatgc aattcgcgaa gtacaatgtc tggctaccag tcgggatggt 360
 attcatttcg agaaacaggg tgtgatactc actccaccag aaggcatcat gcacttccgc 420
 gatcctaag tgtagcgatga agccgacaca tggtaggatgg tagtcggggc gaaagaccca 480
 ggcaacacgg ggcagatcct gctttatcgc ggcagttcat tgcgtgaatg gactttcgat 540
 cgcgtactgg cccacgtga tgcgggtgaa agctatatgt gggaatgtcc ggactttttc 600

 agccttggcg atcagcatta tctgatgttt tccccgagg gaatgaatgc cgagggatac 660
 agttatcgaa atcgctttca aagtggcgta ataccggaa tgtggtcgcc aggacgactt 720
 ttgacacaat ccgggcattt tactgaactt gataacgggc atgactttta tgcaccacaa 780
 agctttgtag cgaaggatgg tcggcgatatt gttatcggtt ggatggatat gtgggaatcg 840

ccaatgccct caaaacgtga aggctgggca ggctgcatga cgctggcgcg cgagctatca 900
gagagcaatg gcaaactcct acaacgcccgtgtacacgaag ctgagtcgtt acgccagcag 960
catcaatcta tctctccccg cacaatcagc aataaatatg ttttcagga aaacgcgcaa 1020

gcagttgaga ttcagttgca gtgggcgctg aagaacagtg atgccgaaca ttacggatta 1080
cagctcggcg ctggaatcgg gctgtatatt gataaccaat ctgagcgact tgttttgtgg 1140
cgggtattacc cacacgagaa tttagatggc tacgtagta ttccctccc gcagggtgac 1200
atgctgccc taaggatatt tatcgataca tcattccgtg aagtatttat taacgacggg 1260
gaggcggtaga tgagtagccg aatatatccg cagccagaag aacgggaact gtcgctctat 1320
gcctcccacg gattggctgt gctgaacat ggagcactct ggcaactggg ttaa 1374

<210> 56

<211> 477

<212> PRT

<213> Escherichia coli

<400> 56

Met Thr Gln Ser Arg Leu His Ala Ala Gln Asn Ala Leu Ala Lys Leu

1 5 10 15

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

20 25 30

Pro Ala Gly Trp Met Asn Asp Pro Asn Gly Leu Ile Trp Phe Asn Asp

35 40 45

Arg Tyr His Ala Phe Tyr Gln His His Pro Met Ser Glu His Trp Gly

50 55 60

Pro Met His Trp Gly His Ala Thr Ser Asp Asp Met Ile His Trp Gln

65 70 75 80

His Glu Pro Ile Ala Leu Ala Pro Gly Asp Glu Asn Asp Lys Asp Gly

85 90 95

Cys Phe Ser Gly Ser Ala Val Asp Asp Asn Gly Val Leu Ser Leu Ile

100 105 110

Tyr Thr Gly His Val Trp Leu Asp Gly Ala Gly Asn Asp Asp Ala Ile

115 120 125

Arg Glu Val Gln Cys Leu Ala Thr Ser Arg Asp Gly Ile His Phe Glu

130	135	140	
Lys Gln Gly Val Ile Leu Thr Pro Pro Glu Gly Ile Met His Phe Arg			
145	150	155	160
Asp Pro Lys Val Trp Arg Glu Ala Asp Thr Trp Trp Met Val Val Gly			
	165	170	175
Ala Lys Asp Pro Gly Asn Thr Gly Gln Ile Leu Leu Tyr Arg Gly Ser			
	180	185	190
Ser Leu Arg Glu Trp Thr Phe Asp Arg Val Leu Ala His Ala Asp Ala			
	195	200	205
Gly Glu Ser Tyr Met Trp Glu Cys Pro Asp Phe Phe Ser Leu Gly Asp			
	210	215	220
Gln His Tyr Leu Met Phe Ser Pro Gln Gly Met Asn Ala Glu Gly Tyr			
225	230	235	240
Ser Tyr Arg Asn Arg Phe Gln Ser Gly Val Ile Pro Gly Met Trp Ser			
	245	250	255
Pro Gly Arg Leu Phe Ala Gln Ser Gly His Phe Thr Glu Leu Asp Asn			
	260	265	270
Gly His Asp Phe Tyr Ala Pro Gln Ser Phe Val Ala Lys Asp Gly Arg			
	275	280	285
Arg Ile Val Ile Gly Trp Met Asp Met Trp Glu Ser Pro Met Pro Ser			
	290	295	300
Lys Arg Glu Gly Trp Ala Gly Cys Met Thr Leu Ala Arg Glu Leu Ser			
305	310	315	320
Glu Ser Asn Gly Lys Leu Leu Gln Arg Pro Val His Glu Ala Glu Ser			
	325	330	335
Leu Arg Gln Gln His Gln Ser Ile Ser Pro Arg Thr Ile Ser Asn Lys			
	340	345	350
Tyr Val Leu Gln Glu Asn Ala Gln Ala Val Glu Ile Gln Leu Gln Trp			
	355	360	365
Ala Leu Lys Asn Ser Asp Ala Glu His Tyr Gly Leu Gln Leu Gly Ala			
	370	375	380

Gly Met Arg Leu Tyr Ile Asp Asn Gln Ser Glu Arg Leu Val Leu Trp
 385 390 395 400
 Arg Tyr Tyr Pro His Glu Asn Leu Asp Gly Tyr Arg Ser Ile Pro Leu
 405 410 415
 Pro Gln Gly Asp Met Leu Ala Leu Arg Ile Phe Ile Asp Thr Ser Ser
 420 425 430
 Val Glu Val Phe Ile Asn Asp Gly Glu Ala Val Met Ser Ser Arg Ile
 435 440 445

Tyr Pro Gln Pro Glu Glu Arg Glu Leu Ser Leu Tyr Ala Ser His Gly
 450 455 460
 Val Ala Val Leu Gln His Gly Ala Leu Trp Gln Leu Gly
 465 470 475

<210> 57

<211> 1434

<212> DNA

<213> Escherichia coli

<400> 57

atgacgaat ctcgattgca tgcggcgcaa aacgccctag caaaacttca tgagcaccgg	60
ggtaacactt tctatcccca ttttcacctc gcgcctcctg ccgggtggat gaacgatcca	120
aacggcctga tctggtttaa cgatcgttat caccggtttt atcaacatca tccgatgagc	180
gaacactggg ggccaatgca ctggggacat gccaccagcg acgatatgat ccactggcag	240
catgagccta ttgcgctagc gccaggagac gataatgaca aagacgggtg tttttcaggt	300
agtgtgtgtc atgacaatgg tgtctcttca cttatctaca ccggacacgt ctggctcgat	360
ggtgcaggta atgacgatgc aattcgcgaa gtacaatgtc tggctaccag tcgggatggt	420
attcatttcg agaaacaggg tgtgatectc actccaccag aaggaatcat gcacttccgc	480
gatacctaag tgtggcgtga agccgacaca tgggtggatgg tagtcggggc gaaagatcca	540
ggcaacacgg ggcagatcct gctttatcgc ggcagttcgt tgcgtgaatg gaccttcgat	600
cgcggtactgg cccacgtga tgcgggtgaa agctatatgt gggaatgtcc ggactttttc	660
agccttggcg atcagcatta tctgatgttt tccccgagg gaatgaatgc cgagggatac	720
agttaccgaa atcgctttca aagtggcgta ataccggaa tgtggtcgcc aggacgactt	780
tttgacaat ccgggcattt tactgaactt gataacgggc atgactttta tgcaccacaa	840
agcttttttag cgaaggatgg tcggcgattt gttatcggct ggatggatat gtgggaatcg	900

ccaatgccct caaaacgtga aggatgggca ggctgcatga cgctggcgcg cgagctatca 960
gagagcaatg gcaaacttct acaacgcccc gtacacgaag ctgagtcgtt acgccagcag 1020

catcaatctg tctctccccg cacaatcagc aataaatatg ttttcagga aaacgcgcaa 1080
gcagttgaga ttcagttgca gtggcgctg aagaacagtg atgccgaaca ttacggatta 1140
cagctcgga ctggaatcg gctgtatatt gataaccaat ctgagcgact tgtttgtgg 1200
cggattacc cacacgagaa tttagacggc taccgtagta ttccctccc gcagcgtgac 1260
acgctgccc taaggatatt tatcgtatac tcatacgtgg aagtatttat taacgacggg 1320
gaagcggtag tgagtagtcg aatctatccg cagccagaag aacgggaact gtcgctttat 1380
gcctcccacg gactggctgt gctgaacat ggagcactct ggctactggg ttaa 1434

<210> 58

<211> 477

<212> PRT

<213> Escherichia coli

<400> 58

Met Thr Gln Ser Arg Leu His Ala Ala Gln Asn Ala Leu Ala Lys Leu

1 5 10 15

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

20 25 30

Pro Ala Gly Trp Met Asn Asp Pro Asn Gly Leu Ile Trp Phe Asn Asp

35 40 45

Arg Tyr His Ala Phe Tyr Gln His His Pro Met Ser Glu His Trp Gly

50 55 60

Pro Met His Trp Gly His Ala Thr Ser Asp Asp Met Ile His Trp Gln

65 70 75 80

His Glu Pro Ile Ala Leu Ala Pro Gly Asp Asp Asn Asp Lys Asp Gly

85 90 95

Cys Phe Ser Gly Ser Ala Val Asp Asp Asn Gly Val Leu Ser Leu Ile

100 105 110

Tyr Thr Gly His Val Trp Leu Asp Gly Ala Gly Asn Asp Asp Ala Ile

115 120 125

Arg Glu Val Gln Cys Leu Ala Thr Ser Arg Asp Gly Ile His Phe Glu

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

 180 185 190
 Ser Leu Arg Glu Trp Thr Phe Asp Arg Val Leu Ala His Ala Asp Ala
 195 200 205
 Gly Glu Ser Tyr Met Trp Glu Cys Pro Asp Phe Phe Ser Leu Gly Asp
 210 215 220
 Gln His Tyr Leu Met Phe Ser Pro Gln Gly Met Asn Ala Glu Gly Tyr
 225 230 235 240
 Ser Tyr Arg Asn Arg Phe Gln Ser Gly Val Ile Pro Gly Met Trp Ser

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

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

 370 375 380

Gly Met Arg Leu Tyr Ile Asp Asn Gln Ser Glu Arg Leu Val Leu Trp
 385 390 395 400
 Arg Tyr Tyr Pro His Glu Asn Leu Asp Gly Tyr Arg Ser Ile Pro Leu
 405 410 415
 Pro Gln Arg Asp Thr Leu Ala Leu Arg Ile Phe Ile Asp Thr Ser Ser
 420 425 430
 Val Glu Val Phe Ile Asn Asp Gly Glu Ala Val Met Ser Ser Arg Ile

435 440 445
 Tyr Pro Gln Pro Glu Glu Arg Glu Leu Ser Leu Tyr Ala Ser His Gly
 450 455 460
 Val Ala Val Leu Gln His Gly Ala Leu Trp Leu Leu Gly
 465 470 475

<210> 59

<211> 1599

<212> DNA

<213> Bifidobacterium lactis

<400> 59

atggcaaccc ttccaccaa tattcccgcc aacggcattc tgacccccga cccggcgctc 60
 gaccctgtgc tcacgccgat ctcgacccat gccgagcagc tgtcactcgc cgaagcaggc 120

gtgtcggcac tggaaaccac ccgcaacgac cgttggtacc cgaagttcca cattgcctcc 180
 aatggcgggt ggatcaacga cccgaacggc ctgtgccgct acaacggacg ctggcacgtg 240
 ttctaccagc tgcattccca cggcacacag tggggcccga tgcattgggg ccacgtctcc 300
 tccgacaaca tggtcgactg gcaccgcgaa cccatcgctt tcgcgccaag cctcgaacag 360
 gaagccacg gtgtgttctc cggttccgcc gtgattggcg acgacggcaa gccgtggatt 420
 ttctacaccg gccaccgctg ggccaacggc aaggacaaca ccggaggcga ctggcagggtg 480
 cagatgctcg ccaagccgaa cgacgacgaa ctgaagacct tcacgaagga gggcatgata 540

atcgactgcc ccaccgacga ggtggaccac cacttccgcg acccgaaggt gtggaagacc 600
 ggtgacacct ggtatatgac cttcgggtgc tcgtcgaagg agcatcgtgg ccagatgtgg 660
 ctgtacacgt cgagcgacat ggtgcactgg agcttcgata ggggtctgtt cgagcatccg 720
 gatccgaacg tgttcatgct tgaatgcccc gatctcttcc cgatccgcga tgcgcggggc 780
 aacgagaaat gggtcacatcg cttctccgcg atgggtgcc aagccaaatgg cttcatgaac 840
 cgcaacgtga acaatgccgg ctacatggtg ggcacatgga agccaggcga gagcttcaag 900

ccggagaccg agttccgcct gtgggacgaa ggccataact tctatgcacc acagtcgttc 960

aacaccgaag ggcgccagat catgtacggc tggatgagcc cgttcgtcgc ccccatcccg 1020

atggaggagg acggctggtg cggcaacctc accctccccc gcgagatcac gctgggcgat 1080

gacggtgacc tggtcaccgc cccaccatc gaaatggagg ggctgcgcga gaataccata 1140

ggcttcgact cgtcgcacct tggtagaac cagacctcca cgtcctcga cgtgacggc 1200

ggcgccctgg aaatcgagat gagactcgat ctgaacaaaa ccaccgccga acgcgccgga 1260

ctgcatgtgc atgccacaag cgacggccac tacacggcaa tcgtattcga cgcgcagatc 1320

ggcggcgctg tcatcgaccg gcagaacgtg gcgaacggag acaaaggcta ccgggtggcc 1380

aagctcagcg acaccgagct cgcagccgat acgcttgact tgcgcgtgtt catcgaccgc 1440

ggatgcgtcg aggtctacgt cgacggcggc aagcatgcga tgagctcgta ctgcttcct 1500

ggcgtggcg cagcgccgt cgaactcgtg agcgaatccg gcaccacgca catcgacacc 1560

ctcaccatgc actcgtcaa gtccatcgga ctcgagtga 1599

<210> 60

<211> 532

<212> PRT

<213> Bifidobacterium lactis

<400> 60

Met Ala Thr Leu Pro Thr Asn Ile Pro Ala Asn Gly Ile Leu Thr Pro

1 5 10 15

Asp Pro Ala Leu Asp Pro Val Leu Thr Pro Ile Ser Asp His Ala Glu

20 25 30

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

35 40 45

Asn Asp Arg Trp Tyr Pro Lys Phe His Ile Ala Ser Asn Gly Gly Trp

50 55 60

Ile Asn Asp Pro Asn Gly Leu Cys Arg Tyr Asn Gly Arg Trp His Val

65 70 75 80

Phe Tyr Gln Leu His Pro His Gly Thr Gln Trp Gly Pro Met His Trp

85 90 95

Gly His Val Ser Ser Asp Asn Met Val Asp Trp His Arg Glu Pro Ile

100 105 110

Ala Phe Ala Pro Ser Leu Glu Gln Glu Arg His Gly Val Phe Ser Gly
115 120 125

Ser Ala Val Ile Gly Asp Asp Gly Lys Pro Trp Ile Phe Tyr Thr Gly
130 135 140

His Arg Trp Ala Asn Gly Lys Asp Asn Thr Gly Gly Asp Trp Gln Val
145 150 155 160

Gln Met Leu Ala Lys Pro Asn Asp Asp Glu Leu Lys Thr Phe Thr Lys
165 170 175

Glu Gly Met Ile Ile Asp Cys Pro Thr Asp Glu Val Asp His His Phe
180 185 190

Arg Asp Pro Lys Val Trp Lys Thr Gly Asp Thr Trp Tyr Met Thr Phe
195 200 205

Gly Val Ser Ser Lys Glu His Arg Gly Gln Met Trp Leu Tyr Thr Ser
210 215 220

Ser Asp Met Val His Trp Ser Phe Asp Arg Val Leu Phe Glu His Pro
225 230 235 240

Asp Pro Asn Val Phe Met Leu Glu Cys Pro Asp Phe Phe Pro Ile Arg
245 250 255

Asp Ala Arg Gly Asn Glu Lys Trp Val Ile Gly Phe Ser Ala Met Gly
260 265 270

Ala Lys Pro Asn Gly Phe Met Asn Arg Asn Val Asn Asn Ala Gly Tyr
275 280 285

Met Val Gly Thr Trp Lys Pro Gly Glu Ser Phe Lys Pro Glu Thr Glu
290 295 300

Phe Arg Leu Trp Asp Glu Gly His Asn Phe Tyr Ala Pro Gln Ser Phe
305 310 315 320

Asn Thr Glu Gly Arg Gln Ile Met Tyr Gly Trp Met Ser Pro Phe Val
325 330 335

Ala Pro Ile Pro Met Glu Glu Asp Gly Trp Cys Gly Asn Leu Thr Leu
340 345 350

Pro Arg Glu Ile Thr Leu Gly Asp Asp Gly Asp Leu Val Thr Ala Pro

355 360 365
 Thr Ile Glu Met Glu Gly Leu Arg Glu Asn Thr Ile Gly Phe Asp Ser
 370 375 380
 Leu Asp Leu Gly Thr Asn Gln Thr Ser Thr Ile Leu Asp Asp Asp Gly
 385 390 395 400

Gly Ala Leu Glu Ile Glu Met Arg Leu Asp Leu Asn Lys Thr Thr Ala
 405 410 415
 Glu Arg Ala Gly Leu His Val His Ala Thr Ser Asp Gly His Tyr Thr
 420 425 430
 Ala Ile Val Phe Asp Ala Gln Ile Gly Gly Val Val Ile Asp Arg Gln
 435 440 445
 Asn Val Ala Asn Gly Asp Lys Gly Tyr Arg Val Ala Lys Leu Ser Asp
 450 455 460

Thr Glu Leu Ala Ala Asp Thr Leu Asp Leu Arg Val Phe Ile Asp Arg
 465 470 475 480
 Gly Cys Val Glu Val Tyr Val Asp Gly Gly Lys His Ala Met Ser Ser
 485 490 495
 Tyr Ser Phe Pro Gly Asp Gly Ala Arg Ala Val Glu Leu Val Ser Glu
 500 505 510
 Ser Gly Thr Thr His Ile Asp Thr Leu Thr Met His Ser Leu Lys Ser
 515 520 525

Ile Gly Leu Glu

530

<210> 61

<211> 1599

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 61

atgcttttgc aagcttttct tttccttttg gctggttttg cagccaaaat atctgcatca 60
 atgacaaacg aaactagcga tagacctttg gtccacttca cacccaacaa gggctggatg 120
 aatgacccaa atgggttttg gtacgatgaa aaagatgcc aatggcatct gtactttcaa 180
 tacaacccaa atgacaccgt atggggtacg ccattgtttt ggggccatgc tacttccgat 240

gatttgacta attgggaaga tcaacccatt gctatcgctc ccaagcgtaa cgattcaggt 300
gctttctctg gctccatggg ggttgattac aacaacacga gtgggttttt caatgatact 360

attgatccaa gacaaagatg cgttgcgatt tggacttata acactcctga aagtgaagag 420
caatacatta gctattctct tgatgggtggg tacactttta ctgaatacca aaagaaccct 480
gttttagctg ccaactccac tcaattcaga gatccaaagg tgttctggta tgaaccttct 540
caaaaatgga ttatgacggc tgccaaatca caagactaca aaattgaaat ttactcctct 600
gatgacttga agtcctggaa gctagaatct gcatttgcca atgaaggttt cttaggtctac 660
caatacgaat gtccaggttt gattgaagtc ccaactgagc aagatccttc caaatcttat 720
tgggtcatgt ttatttctat caaccagggt gcacctgctg gcggttcctt caaccaatat 780

tttgttggat ctttcaatgg tactcatttt gaagcgtttg acaatcaatc tagagtggta 840
gattttggta aggactacta tgccttgcaa actttcttca acactgaccc aacctacggt 900
tcagcattag gtattgcctg ggcttcaaac tgggagtaca gtgcctttgt cccaactaac 960
ccatggagat catccatgct tttggtccgc aagttttctt tgaacactga atatcaagct 1020
aatccagaga ctgaattgat caatttgaat gccgaaccaa tattgaacat tagtaatgct 1080
ggtccttggg ctctgttttg tactaacaca actctaacta aggccaatc ttacaatgct 1140
gatttgagca actcgactgg taccctagag tttagattgg ttacgctgt taacaccaca 1200

caaaccatat ccaaatccgt ctttgccgac ttatcacttt ggttcaaggg tttagaagat 1260
cctgaagaat atttgagaat gggttttgaa gtcagtgcct cttccttctt ttggaccgt 1320
ggtaactcta aggtcaagtt tgtcaaggag aacccatatt tcacaaacag aatgtctgtc 1380
aacaaccaac cattcaagtc tgagaacgac ctaagttact ataaagtgt cggcctactg 1440
gatcaaaaca tcttggatt gtacttcaac gatggagatg tggtttctac aaatactac 1500
ttcatgacca cggtaacgc tctaggatct gtgaacatga ccactggtgt cgataatttg 1560
ttctacattg acaagttcca agtaaggga gtaaaatag 1599

<210> 62

<211> 532

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 62

Met Leu Leu Gln Ala Phe Leu Phe Leu Leu Ala Gly Phe Ala Ala Lys

1 5 10 15

Ile Ser Ala Ser Met Thr Asn Glu Thr Ser Asp Arg Pro Leu Val His

20 25 30
 Phe Thr Pro Asn Lys Gly Trp Met Asn Asp Pro Asn Gly Leu Trp Tyr
 35 40 45
 Asp Glu Lys Asp Ala Lys Trp His Leu Tyr Phe Gln Tyr Asn Pro Asn

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

 115 120 125
 Ala Ile Trp Thr Tyr Asn Thr Pro Glu Ser Glu Glu Gln Tyr Ile Ser
 130 135 140
 Tyr Ser Leu Asp Gly Gly Tyr Thr Phe Thr Glu Tyr Gln Lys Asn Pro
 145 150 155 160
 Val Leu Ala Ala Asn Ser Thr Gln Phe Arg Asp Pro Lys Val Phe Trp
 165 170 175
 Tyr Glu Pro Ser Gln Lys Trp Ile Met Thr Ala Ala Lys Ser Gln Asp

 180 185 190
 Tyr Lys Ile Glu Ile Tyr Ser Ser Asp Asp Leu Lys Ser Trp Lys Leu
 195 200 205
 Glu Ser Ala Phe Ala Asn Glu Gly Phe Leu Gly Tyr Gln Tyr Glu Cys
 210 215 220
 Pro Gly Leu Ile Glu Val Pro Thr Glu Gln Asp Pro Ser Lys Ser Tyr
 225 230 235 240
 Trp Val Met Phe Ile Ser Ile Asn Pro Gly Ala Pro Ala Gly Gly Ser

 245 250 255
 Phe Asn Gln Tyr Phe Val Gly Ser Phe Asn Gly Thr His Phe Glu Ala
 260 265 270

Phe Asp Asn Gln Ser Arg Val Val Asp Phe Gly Lys Asp Tyr Tyr Ala
 275 280 285
 Leu Gln Thr Phe Phe Asn Thr Asp Pro Thr Tyr Gly Ser Ala Leu Gly
 290 295 300
 Ile Ala Trp Ala Ser Asn Trp Glu Tyr Ser Ala Phe Val Pro Thr Asn
 305 310 315 320
 Pro Trp Arg Ser Ser Met Ser Leu Val Arg Lys Phe Ser Leu Asn Thr
 325 330 335
 Glu Tyr Gln Ala Asn Pro Glu Thr Glu Leu Ile Asn Leu Lys Ala Glu
 340 345 350
 Pro Ile Leu Asn Ile Ser Asn Ala Gly Pro Trp Ser Arg Phe Ala Thr
 355 360 365
 Asn Thr Thr Leu Thr Lys Ala Asn Ser Tyr Asn Val Asp Leu Ser Asn
 370 375 380
 Ser Thr Gly Thr Leu Glu Phe Glu Leu Val Tyr Ala Val Asn Thr Thr
 385 390 395 400
 Gln Thr Ile Ser Lys Ser Val Phe Ala Asp Leu Ser Leu Trp Phe Lys
 405 410 415
 Gly Leu Glu Asp Pro Glu Glu Tyr Leu Arg Met Gly Phe Glu Val Ser
 420 425 430
 Ala Ser Ser Phe Phe Leu Asp Arg Gly Asn Ser Lys Val Lys Phe Val
 435 440 445
 Lys Glu Asn Pro Tyr Phe Thr Asn Arg Met Ser Val Asn Asn Gln Pro
 450 455 460
 Phe Lys Ser Glu Asn Asp Leu Ser Tyr Tyr Lys Val Tyr Gly Leu Leu
 465 470 475 480
 Asp Gln Asn Ile Leu Glu Leu Tyr Phe Asn Asp Gly Asp Val Val Ser
 485 490 495
 Thr Asn Thr Tyr Phe Met Thr Thr Gly Asn Ala Leu Gly Ser Val Asn
 500 505 510
 Met Thr Thr Gly Val Asp Asn Leu Phe Tyr Ile Asp Lys Phe Gln Val

515	520	525	
Arg Glu Val Lys			
530			
<210> 63			
<211> 1302			
<212> DNA			
<213> Corynebacterium glutamicum			
<400> 63			
gtgtgtgggg	ctatgcacac	agaactttcc	agtttgcgcc
ctgcgtacca	tgtgactcct	60	
ccgcagggca	ggctcaatga	tccaacgga	atgtacgtcg
atggcgatac	cctccacgtc	120	
tactaccagc	acgateccagg	tttccccttc	gcaccaaagc
gcaccggctg	ggctcacacc	180	
accacgccgt	tgaccggacc	gcagcgattg	cagtggacgc
acctgcccga	cgctctttac	240	
ccggatgcat	cctatgacct	ggatggatgc	tattccgggtg
gagccgtatt	tactgacggc	300	
acacttaaac	ttttctacac	cggcaacctia	aaaattgacg
gcaagcgccg	cgccacccaa	360	
aacctcgtcg	aagtcgagga	cccaactggg	ctgatgggcg
gcattcatcg	ccgttcgcct	420	
aaaaatccgc	ttatcgacgg	acccgccagc	ggtttcacac
cccattaccg	cgatcccatg	480	
atcagccctg	atggatgatg	ttggaaaatg	gttcttgggg
cccaacgcga	aaacctcacc	540	
ggtgcagcgg	ttctataccg	ctcgacagat	cttgaaaact
gggaattctc	cggtgaaatc	600	
acctttgacc	tcagtgatgc	acaacctggt	tctgctcctg
atctcgttcc	cggtggctac	660	
atgtgggaat	gccccaaact	ttttacgctt	cgcgatgaag
aaactggcga	agatctcgac	720	
gtgctgattt	tcigtccaca	aggattggac	cgaatccacg
atgaggttac	tcactacgca	780	
agctctgacc	agtgcggata	tgctgctggc	aagcttgaag
gaacgacctt	ccgcgtcttg	840	
cgaggattca	gcgagctgga	tttcggccat	gaattctacg
caccgcaggt	tgcagtaaac	900	
ggttctgatg	ccitggctcgt	gggctggatg	gggctgcccc
cgaggatga	tcaccaaca	960	
gttgcacggg	aaggatgggt	gcaactgcctg	actgtgcccc
gcaagcttca	tttgcgcaac	1020	
cacgcgatct	atcaagagct	tctttctcca	gagggggagt
caggggtaat	cagatctgta	1080	
ttaggttctg	aacctgtccg	agtagacatc	cgaggcaata
tttccctcga	gtgggatggt	1140	
gtccgtttgt	ctgtggatcg	tggtggatgat	cgctcgctag
ctgaggtaaa	acctggcgaa	1200	
ttagtgatcg	cggacgataa	tacagccatt	gagataactg
caggtgatgg	acaggtttca	1260	
ttcgttttcc	gggctttcaa	aggtgacact	attgagagat
aa		1302	
<210> 64			
<211> 433			

<212> PRT

<213> *Corynebacterium glutamicum*

<400> 64

Met Cys Gly Ala Met His Thr Glu Leu Ser Ser Leu Arg Pro Ala Tyr

1 5 10 15
His Val Thr Pro Pro Gln Gly Arg Leu Asn Asp Pro Asn Gly Met Tyr

20 25 30
Val Asp Gly Asp Thr Leu His Val Tyr Tyr Gln His Asp Pro Gly Phe

35 40 45
Pro Phe Ala Pro Lys Arg Thr Gly Trp Ala His Thr Thr Thr Pro Leu

50 55 60
Thr Gly Pro Gln Arg Leu Gln Trp Thr His Leu Pro Asp Ala Leu Tyr

65 70 75 80
Pro Asp Ala Ser Tyr Asp Leu Asp Gly Cys Tyr Ser Gly Gly Ala Val

85 90 95
Phe Thr Asp Gly Thr Leu Lys Leu Phe Tyr Thr Gly Asn Leu Lys Ile

100 105 110
Asp Gly Lys Arg Arg Ala Thr Gln Asn Leu Val Glu Val Glu Asp Pro

115 120 125
Thr Gly Leu Met Gly Gly Ile His Arg Arg Ser Pro Lys Asn Pro Leu

130 135 140
Ile Asp Gly Pro Ala Ser Gly Phe Thr Pro His Tyr Arg Asp Pro Met

145 150 155 160
Ile Ser Pro Asp Gly Asp Gly Trp Lys Met Val Leu Gly Ala Gln Arg

165 170 175
Glu Asn Leu Thr Gly Ala Ala Val Leu Tyr Arg Ser Thr Asp Leu Glu

180 185 190
Asn Trp Glu Phe Ser Gly Glu Ile Thr Phe Asp Leu Ser Asp Ala Gln

195 200 205
Pro Gly Ser Ala Pro Asp Leu Val Pro Gly Gly Tyr Met Trp Glu Cys

210 215 220

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

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

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

 385 390 395 400
 Leu Val Ile Ala Asp Asp Asn Thr Ala Ile Glu Ile Thr Ala Gly Asp
 405 410 415
 Gly Gln Val Ser Phe Ala Phe Arg Ala Phe Lys Gly Asp Thr Ile Glu
 420 425 430
 Arg

<210> 65

<211> 1473

<212> DNA

<213> Leuconostoc mesenteroides

<400> 65

atggaaattc aaaacaaagc aatgttgatc acttatgctg attcgttggg caaaaactta	60
aaagatgttc atcaagtctt gaaagaagat attggagatg cgattggtgg ggttcatttg	120
ttgcctttct tcccttcaac aggtgatcgc ggttttgcgc cagccgatta tactcgtgtt	180
gatgccgcat ttggtgattg ggcagatgtc gaagcattgg gtgaagaata ctatttgatg	240
tttgacttca tgattaacca tatttctcgt gaatcagtga tgtatcaaga ttttaagaag	300
aatcatgacg attcaaagta taaagatttc tttattcgtt gggaaaagtt ctgggcaaag	360
gccggcgaaa accgtccaac acaagccgat gttgacttaa tttacaagcg taaagataag	420
gcaccaacgc aagaaatcac ttttgatgat ggcacaacag aaaactttgt gaatactttt	480
ggtgaagaac aaattgacat tgatgttaat tcagccattg ccaaggaatt tattaagaca	540
acccttgaag acatggtaaa acatggtgct aacttgattc gtttggatgc ctttgcgtat	600
gcagttaaaa aagttgacac aaatgacttc ttcgttgagc cagaaatctg ggacactttg	660
aatgaagtac gtgaaatttt gacaccatta aaggctgaaa ttttaccaga aattcatgaa	720
cattactcaa tccctaaaaa gatcaatgat catggttact tcacctatga ctttgcatta	780
ccaatgacaa cgctttacac attgtattca ggtaagacaa atcaattggc aaagtggttg	840
aagatgtcac caatgaagca attcacaaca ttggacacgc atgatggatg tgggtgctgt	900
gatgcccgtg atattctaac tgatgatgaa attgactacg cttctgaaca actttacaag	960
gttggcgcga atgtcaaaaa gacatattca tctgcttcat acaacaacct tgatatttac	1020
caaattaact caacttatta ttcagcattg ggaaatgatg atgcagcata cttgttgagt	1080
cgtgtcttcc aagtctttgc gcctggaatt ccacaaattt attacgttgg tttgttgga	1140
ggtgaaaacg atatcgcgct tttggagtca actaaagaag gtcgtaatat taaccgtcat	1200
tactatacgc gtgaagaagt taagtcagaa gttaagcgac cagttgttgc taacttattg	1260
aagctattgt catggcgtaa tgaaagccct gcatttgatt tggttggtc aatcacagtt	1320
gacacgcaa ctgataaac aattgtggtg acacgtcaag atgaaatgg tcaaaacaaa	1380
gctgtattaa cagccgatgc ggccaacaaa acttttgaaa tcgttgagaa tggtaaact	1440
gttatgagca gtgataattt gactcagaac taa	1473

<210> 66

<211> 490

<212> PRT

<213> Leuconostoc mesenteroides

<400> 66

Met Glu Ile Gln Asn Lys Ala Met Leu Ile Thr Tyr Ala Asp Ser Leu
1 5 10 15
Gly Lys Asn Leu Lys Asp Val His Gln Val Leu Lys Glu Asp Ile Gly
20 25 30
Asp Ala Ile Gly Gly Val His Leu Leu Pro Phe Phe Pro Ser Thr Gly
35 40 45
Asp Arg Gly Phe Ala Pro Ala Asp Tyr Thr Arg Val Asp Ala Ala Phe
50 55 60
Gly Asp Trp Ala Asp Val Glu Ala Leu Gly Glu Glu Tyr Tyr Leu Met
65 70 75 80
Phe Asp Phe Met Ile Asn His Ile Ser Arg Glu Ser Val Met Tyr Gln
85 90 95
Asp Phe Lys Lys Asn His Asp Asp Ser Lys Tyr Lys Asp Phe Phe Ile
100 105 110
Arg Trp Glu Lys Phe Trp Ala Lys Ala Gly Glu Asn Arg Pro Thr Gln
115 120 125
Ala Asp Val Asp Leu Ile Tyr Lys Arg Lys Asp Lys Ala Pro Thr Gln
130 135 140
Glu Ile Thr Phe Asp Asp Gly Thr Thr Glu Asn Leu Trp Asn Thr Phe
145 150 155 160
Gly Glu Glu Gln Ile Asp Ile Asp Val Asn Ser Ala Ile Ala Lys Glu
165 170 175
Phe Ile Lys Thr Thr Leu Glu Asp Met Val Lys His Gly Ala Asn Leu
180 185 190
Ile Arg Leu Asp Ala Phe Ala Tyr Ala Val Lys Lys Val Asp Thr Asn
195 200 205
Asp Phe Phe Val Glu Pro Glu Ile Trp Asp Thr Leu Asn Glu Val Arg
210 215 220
Glu Ile Leu Thr Pro Leu Lys Ala Glu Ile Leu Pro Glu Ile His Glu
225 230 235 240
His Tyr Ser Ile Pro Lys Lys Ile Asn Asp His Gly Tyr Phe Thr Tyr

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

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

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

 420 425 430
 Asp Leu Ala Gly Ser Ile Thr Val Asp Thr Pro Thr Asp Thr Thr Ile
 435 440 445
 Val Val Thr Arg Gln Asp Glu Asn Gly Gln Asn Lys Ala Val Leu Thr
 450 455 460
 Ala Asp Ala Ala Asn Lys Thr Phe Glu Ile Val Glu Asn Gly Gln Thr
 465 470 475 480
 Val Met Ser Ser Asp Asn Leu Thr Gln Asn
 485 490

<210> 67

<211> 1515

<212> DNA

<213> Bifidobacterium adolescentis

<400> 67

```

atgaaaaaca aggtgcagct catcacttac gccgaccgcc ttggcgacgg caccatcaag    60
tcgatgaccg acattctgcg caccgccttc gacggcgtgt acgacggcgt tcacatcctg    120
ccgtttcttc ccccgcttca cggcgccgac gcaggcttcg acccgatcga ccaccaacag    180
gtcgcagAAC gtctcggcag ctgggacgac gtcgccgaac tctccaagac ccacaacatc    240
atggtcgacg ccatcgtcaa ccacatgagt tgggaatcca agcagttcca ggacgtgctg    300
gccaaaggcg aggagtccga atactatccg atgttcttca ccatgagctc cgtgttcccg    360

aacggcgcca ccgaagagga cctggccggc atctaccgtc cgcgtccggg cctgccgttc    420
accactaca agttcgccgg caagaccgcg ctctgtgtggg tcagcttcac cccgcagcag    480
gtggacatcg acaccgattc cgacaagggt tgggaatacc tcatgtcgat ttctgaccag    540
atggccgcct ctacgtcag ctacatccgc ctgcacgcg tcggctatgg cgccaaggaa    600
gccggcacca gctgcttcac gaccccgaa accttcaagc tgatctccg tctgcgtgag    660
gaaggcgtca agcgcggtct ggaaatctc atcgaagtgc actcctacta caagaagcag    720
gtcgaaatcg catccaaggt ggaccgcgtc tacgacttcg cctgcctcc gctgctgctg    780

cacgcgtga gcaccggcca cgtcgagccc gtcgccact ggaccgacat acgccgaac    840
aacgccgtca ccgtgctcga tacgcacgac ggcatcggcg tgatcgacat cggtccgac    900
cagctcgacc gctcgctcaa gggctctgtg ccggatgagg acgtggacaa cctcgtcaac    960
accatccag ccaacaccca cggcgaatcc caggcagcca ctggcgccgc cgcattcaat    1020
ctcgacctct accaggtcaa cagcacctac tattcggcgc tcgggtgcaa cgaccagcac    1080
tacatcgccg cccgcgcggt gcagttcttc ctgccggcg tgccgcaagt ctactacgtc    1140
ggcgcgctcg cggcgaagaa gcacatggag ctgctgcgta agacgaataa cggccgcgac    1200

atcaatcgcc attactactc caccgcggaa atcgacgaga acctcaagcg tccggtcgtc    1260
aaggccctga acgcgtcgc caagttccgc aacgagctcg acgcgttcga cggcacgttc    1320
tcgtacacca ccgatgacga cagtcctatc agcttcacct ggcgcggcga aaccagccag    1380
gccacgtga cgttcgacgc gaagcgcggt ctcggtgtgg acaacactac gccggtcgcc    1440
atgttgaat gggaggattc cgcgggagac caccgttcgg atgatctgat cgccaatccg    1500
cctgtcgtcg cctga

```

<210> 68

<211> 504

<212> PRT

<213> Bifidobacterium adolescentis

<400> 68

Met Lys Asn Lys Val Gln Leu Ile Thr Tyr Ala Asp Arg Leu Gly Asp

1 5 10 15

Gly Thr Ile Lys Ser Met Thr Asp Ile Leu Arg Thr Arg Phe Asp Gly

20 25 30

Val Tyr Asp Gly Val His Ile Leu Pro Phe Phe Thr Pro Phe Asp Gly

35 40 45

Ala Asp Ala Gly Phe Asp Pro Ile Asp His Thr Lys Val Asp Glu Arg

50 55 60

Leu Gly Ser Trp Asp Asp Val Ala Glu Leu Ser Lys Thr His Asn Ile

65 70 75 80

Met Val Asp Ala Ile Val Asn His Met Ser Trp Glu Ser Lys Gln Phe

85 90 95

Gln Asp Val Leu Ala Lys Gly Glu Glu Ser Glu Tyr Tyr Pro Met Phe

100 105 110

Leu Thr Met Ser Ser Val Phe Pro Asn Gly Ala Thr Glu Glu Asp Leu

115 120 125

Ala Gly Ile Tyr Arg Pro Arg Pro Gly Leu Pro Phe Thr His Tyr Lys

130 135 140

Phe Ala Gly Lys Thr Arg Leu Val Trp Val Ser Phe Thr Pro Gln Gln

145 150 155 160

Val Asp Ile Asp Thr Asp Ser Asp Lys Gly Trp Glu Tyr Leu Met Ser

165 170 175

Ile Phe Asp Gln Met Ala Ala Ser His Val Ser Tyr Ile Arg Leu Asp

180 185 190

Ala Val Gly Tyr Gly Ala Lys Glu Ala Gly Thr Ser Cys Phe Met Thr

195 200 205

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

210 215 220
 Arg Gly Leu Glu Ile Leu Ile Glu Val His Ser Tyr Tyr Lys Lys Gln
 225 230 235 240
 Val Glu Ile Ala Ser Lys Val Asp Arg Val Tyr Asp Phe Ala Leu Pro
 245 250 255

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

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

 Gly Lys Asn Asp Met Glu Leu Leu Arg Lys Thr Asn Asn Gly Arg Asp
 385 390 395 400
 Ile Asn Arg His Tyr Tyr Ser Thr Ala Glu Ile Asp Glu Asn Leu Lys
 405 410 415
 Arg Pro Val Val Lys Ala Leu Asn Ala Leu Ala Lys Phe Arg Asn Glu
 420 425 430
 Leu Asp Ala Phe Asp Gly Thr Phe Ser Tyr Thr Thr Asp Asp Asp Thr
 435 440 445

 Ser Ile Ser Phe Thr Trp Arg Gly Glu Thr Ser Gln Ala Thr Leu Thr
 450 455 460

Phe Glu Pro Lys Arg Gly Leu Gly Val Asp Asn Thr Thr Pro Val Ala
 465 470 475 480
 Met Leu Glu Trp Glu Asp Ser Ala Gly Asp His Arg Ser Asp Asp Leu
 485 490 495
 Ile Ala Asn Pro Pro Val Val Ala

500

<210> 69

<211> 1164

<212> DNA

<213> Klebsiella pneumoniae

<400> 69

tcagaatgcc tggcggaaaa tcgcggcaat ctctgctcg ttgcctttac gcgggttcga	60
gaacgcattg ccgtctttca gagccatctc cgccatgtag gggaagtcgg cctcttttac	120
tcccagatcg cgcatgtct gcggaatacc gatatccatc gacagacgcg tgatagcggc	180
gatggctttt tccgccgctg cgagagtggg cagtccggtg atattttcgc ccatcagttc	240
agcgatatcg gcgaatttct ccgggttggc gatcaggttg tagcgggcca catgcggcag	300
caggacacg ttggccacgc cgtgcggcat gtcgtacagg ccgccagct ggtgcgcat	360
ggcgtgcacg tagccgaggt tggcgttatt gaaagccatc ccggccagca gagaggcata	420
ggccatgttt tcccgccct gcagattgct gccgagggcc acggcctggc gcaggttgcg	480
ggcgaatgagg cggatcgct gcattggcgc ggctccgct accgggttag cgtctttgga	540
gatataggcc tctacggcgt gggtcagggc atccatcccg gtcgcccgcg tcagggcggc	600
cgtttaccg atcatcagca gcgcatcgt gatagagacc gacggcaggt tgcgccagct	660
gacgatcaca aacttactt tggtttcggt gttggtcagg acgcagtggc gggtgacctc	720
gctggcgggt ccggcgggtg tattgaccgc gacgataggc ggacgcgggt tggtcagggt	780
ctcgattccg gcatactggt acagatcgcc ctcatgggtg gcggcgatgc cgatgccttt	840
gccgcaatcg tgcgggctgc cgcccccac ggtgacgatg atgtcgact gttcgcgcg	900
aaacacggcg aggcgctgc gcacgttggg gtctttcggg ttcggtcga cgccgtcaaa	960
gatgccacc tcgatccgg cctcccgag ataatgcagg gttttgtcca ctgcgccatc	1020
tttaattgcc cgcaggcctt tgctggtgac cagcagggt tttttcccc ccagcagctg	1080
gcagcgttcg ccgactacgg aaatggcgtt ggggccaaaa aagttaacgt ttggcaccag	1140
ataatcaaac atacgatagc tcat	1164

<210> 70

<211> 387

<212> PRT

<213> *Klebsiella pneumoniae*

<400> 70

Met Ser Tyr Arg Met Phe Asp Tyr Leu Val Pro Asn Val Asn Phe Phe

1 5 10 15

Gly Pro Asn Ala Ile Ser Val Val Gly Glu Arg Cys Gln Leu Leu Gly

20 25 30

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

35 40 45

Asp Gly Ala Val Asp Lys Thr Leu His Tyr Leu Arg Glu Ala Gly Ile

50 55 60

Glu Val Ala Ile Phe Asp Gly Val Glu Pro Asn Pro Lys Asp Thr Asn

65 70 75 80

Val Arg Asp Gly Leu Ala Val Phe Arg Arg Glu Gln Cys Asp Ile Ile

85 90 95

Val Thr Val Gly Gly Gly Ser Pro His Asp Cys Gly Lys Gly Ile Gly

100 105 110

Ile Ala Ala Thr His Glu Gly Asp Leu Tyr Gln Tyr Ala Gly Ile Glu

115 120 125

Thr Leu Thr Asn Pro Leu Pro Pro Ile Val Ala Val Asn Thr Thr Ala

130 135 140

Gly Thr Ala Ser Glu Val Thr Arg His Cys Val Leu Thr Asn Thr Glu

145 150 155 160

Thr Lys Val Lys Phe Val Ile Val Ser Trp Arg Asn Leu Pro Ser Val

165 170 175

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

180 185 190

Ala Ala Thr Gly Met Asp Ala Leu Thr His Ala Val Glu Ala Tyr Ile

195 200 205

Ser Lys Asp Ala Asn Pro Val Thr Asp Ala Ala Ala Met Gln Ala Ile

210 215 220
 Arg Leu Ile Ala Arg Asn Leu Arg Gln Ala Val Ala Leu Gly Ser Asn
 225 230 235 240
 Leu Gln Ala Arg Glu Asn Met Ala Tyr Ala Ser Leu Leu Ala Gly Met
 245 250 255

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

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

Gln Ala Phe

385

<210> 71

<211> 1824

<212> DNA

<213> Klebsiella pneumoniae

<400> 71

atgccgttaa tagccgggat tgatatcggc aacgccacca ccgaggtggc gctggcgtcc 60
 gactaccgcg aggcgagggc gtttgttgcc agcgggatcg tcgacgacgac gggcatgaaa 120
 gggacgcggg acaatatcgc cgggaccctc gccgcgtgg agcaggccct ggcgaaaaca 180
 ccgtggtcga tgagcgatgt ctctcgcatc tatcttaacg aagccgcgcc ggtgattggc 240

gatgtggcga tggagacat caccgagacc attatcacccg aatcgacat gatcggtcat 300
aaccgcaga cgccggcggt ggtgggcgtt ggcgtgggga cgactatcgc cctcggcggt 360

ctggcgacgc tgcggcggtc gcagtatgcc gaggggtgga tcgtactgat tgacgacgcc 420
gtcgatttcc ttgacgccgt gtggtggctc aatgaggcgc tcgaccgggg gatcaacgtg 480
gtggcgcgga tctcaaaaa ggacgacggc gtgctggtga acaaccgcct gcgtaaaacc 540
ctgccggtgg tggatgaagt gacgctgctg gagcaggtcc ccgagggggt aatggcggtg 600
gtggaagtgg ccgcgccggg ccaggtggtg cggatcctgt cgaatcccta cgggatcgcc 660
accttcttcg ggctaagccc ggaagagacc caggccatcg tccccatcgc ccgcgccctg 720
attggcaacc gtccgcggt ggtgctcaag accccgcagg gggatgtgca gtcgcgggtg 780

atcccgcggt gcaacctcta cattagcggc gaaaagcgcc gcggagaggc cgatgtcgcc 840
gaggcgcggt aagccatcat gcaggcgatg agcgctgctg ctccggtacg cgacatccgc 900
ggcgaaccgg gcaaccacgc cggcggtatg ctgagcggtg tgcgaaggt aatggcgctc 960
ctgaccggcc atgagatgag cgcatatac atccaggatc tgcgtggcgt ggatacgttt 1020
attccgcga aggtgcaggg cgggatggcc ggcgagtgct ccatggagaa tgccgtcggg 1080
atggcgcgga tggatgaaagc ggatcgtctg caaatgcagg ttatcgcccg cgaactgagc 1140
gcccgaactgc agaccgaggt ggtggtgggc ggcgtggagg ccaacatggc catcgccggg 1200

gcgttaacca ctcccggtg tgcggcgccg ctggcgatcc tcgacctcgg cgccggtcgt 1260
acggatgcgg cgatcgtcaa cgccgagggg cagataacgg cggatccatc cgccggggcg 1320
gggaatatgg tcagcctgtt gattaaaacc gagctgggccc tcgaggatct ttcgctggcg 1380
gaagcgataa aaaaataccc gctggccaaa gtggaaagcc tgttcagtat tcgtcacgag 1440
aatggcgcggt tggagtcttt tcgggaagcc ctgagcccg cgggtgttcgc caaagtgggtg 1500
tacatcaagg agggcgaaat ggtgccgatc gataacgcca gcccgtgga aaaaattcgt 1560
ctcgtgcgcc ggcaggcgaa agagaaagtg ttgtcacca actgcctcgc gcgctgcgc 1620

caggtctcac ccggcggttc cattcgcat atcgctttg tggctggtt gggcggtca 1680
tcgtggact ttgatgcc gcagcttatc acggaagcct tgcgcaacta tggcgtggtc 1740
gccggcgagg gcaatatcgc gggaacagaa gggccgcga atcggtcgc caccgggctg 1800
ctactggcgc gtcaggcgaa ttaa 1824

<210> 72

<211> 13669

<212> DNA

<213> Artificial Sequence

<220><223> Plasmid

<400> 72

tagtaaagcc ctcgctagat tttaatgcgg atgttgcgat tacttcgcca actattgcga	60
taacaagaaa aagccagcct ttcattgatat atctccaat ttgttaggg cttattatgc	120
acgcttaaaa ataataaaag cagacttgac ctgatatgtt ggctgtgagc aattatgtgc	180
ttagtgcatc taacgcttga gtttaagccgc gccgcgaagc ggctgcggct tgaacgaatt	240
gttagacatt atttgccgac taccttgggt atctcgctt tcacgtagt gacaaattct	300
tccaactgat ctgcgcgcga ggccaagcga tcttcttctt gtccaagata agcctgtcta	360
gcttcaagta tgacgggctg atactgggccc ggccaggcgt ccattgccca gtcggcagcg	420
acatccttcg gcgcgatttt gccggttact gcgctgtacc aaatgcggga caacgtaagc	480
actacatttc gtcacatgcc agcccagtcg ggccggcgagt tccatagcgt taaggtttca	540
tttagcgctt caaatagatc ctgttcagga accggatcaa agagtctctc cgccgctgga	600
cctaccaagg caacgctatg ttctcttctt ttgttcagca agatagccag atcaatgtcg	660
atcgtggctg gctcgaagat acctgcaaga atgtcattgc gctgccattc tccaaattgc	720
agttcgcgtt tagctggata acgccacgga atgatgtcgt cgtgcacaac aatgggtgact	780
tctacagcgc ggagaatctc gctctctcca ggggaagccg aagtttcaa aaggctgttg	840
atcaaagctc gccgcgttgt ttcattcaagc cttacggtca ccgtaaccag caaatcaata	900
tactgtgtg gcttcaggcc gccatccact gcggagccgt acaaatgtac ggccagcaac	960
gtcggttcga gatggcgctc gatgacgcca actacctctg atagttgagt cgatacttcg	1020
gcgatcaccg cttccctcat gatgtttaac ttgttttag ggcgactgcc ctgctgcgta	1080
acatcgttgc tgctcataa catcaaacat cgaccacgg cgtaacgcgc ttgctgcttg	1140
gatgcccag gcatagactg taccctaaaa aaacagtcac aacaagccat gaaaaccgcc	1200
actgcgccgt taccaccgct gcgttcggtc aaggttctgg accagttgcg tgagcgcata	1260
cgctacttgc attacagctt acgaaccgaa caggcttatg tccactgggt tcgtgccttc	1320
atcgttttc acggtgtgcg taccgggca accttgggca gcagcgaagt cgaggcattt	1380
ctgtcttggc tggcgaacga gcgcaagggt tcggtctcca cgcatcgtca ggcatggcg	1440
gccttgcgtt tcttctacgg caaggtgctg tgcacggatc tgccttggct tcaggagatc	1500
ggaagacctc ggccgtcgcg gcgcttgccg gtggtgctga ccccgatga agtggttcgc	1560
atcctcggtt tcttggaagg cgagcatcgt ttgttcgccc agcttctgta tggaaacggc	1620
atgcggatca gtgagggttt gcaactgcgg gtcaaggatc tggatttca tcacggcacg	1680
atcatcgtgc gggagggcaa gggctccaag gatcgggcct tgatgttacc cgagagcttg	1740

gcacccagcc tgcgcgagca ggggaattaa ttcccacggg ttttgctgcc cgcaaacggg 1800

ctgttctggt gttgctagtt tgttatcaga atcgagatc cggcttcagc cggtttgccg 1860

gctgaaagcg ctattttctc cagaattgcc atgatttttt ccccacggga ggcgtcactg 1920

gctcccggtg tgcggcagc tttgattcga taagcagcat cgcctgttcc aggcgtgcta 1980

tgtgtgactg ttgagctgta acaagtgtc tcaggtgttc aatttcattg tctagtgtct 2040

ttgttttact ggtttcacct gttctattag gtgttacatg ctgttcattc gttacattgt 2100

cgatctgttc atggatgaac gctttgaatg caccataaac tcgtaaaagc tctgatgtat 2160

ctatcttttt tacaccgttt tcattctgtc atatggacag ttttcctttt gatatgtaac 2220

ggatgaacagt tgttctactt ttgtttgtta gtcttgatgc ttcactgata gatacaagag 2280

ccataagaac ctcatatcct tccgtattta gccagtatgt tctctagtgt ggttcgttgt 2340

ttttgcgtga gccatgagaa cgaaccattg agatcatact tactttgcat gtcactcaaa 2400

aattttgcct caaaactggt gagctgaatt ttgcagttta aagcatcgtg tagtgttttt 2460

cttagtccgt tatgtaggta ggaatctgat gtaatgggtg ttggtatttt gtcaccattc 2520

atttttatct gggtgttctc aagttcgggt acgagatcca ttgtctatc tagttcaact 2580

tggaaaatca acgtatcagt cgggcggcct cgttatcaa ccaccaattt catattgctg 2640

taagtgttta aatctttact tattggtttc aaaaccattt ggtaagcct tttaaactca 2700

tggtagttaa tttcaagcat taacatgaac ttaaattcat caaggctaat ctctatattt 2760

gccttgtagg ttttcttttg tgttagttct ttttaatacc actcataaat cctcatagag 2820

tatttgtttt caaaagactt aacatgttcc agattatatt ttatgaattt ttttaactgg 2880

aaaagataag gcaatatctc ttactataaa actaattcta atttttcgtc tgagaacttg 2940

gcatagtttg tccactggaa aatctcaaag ctttaacca aaggattcct gatttccaca 3000

gttctcgtca tcagctctct gggtgcttta gctaatacac cataagcatt ttcctactg 3060

atgttcacca tctgagcgta ttggttataa gtgaacgata ccgtccgttc tttccttgta 3120

gggttttcaa tcgtgggggt gagtagtgcc acacagcata aaattagctt ggtttcatgc 3180

tccgttaagt catagcgact aatcgctagt tcatttgctt tgaaaacaac taattcagac 3240

atacatctca attggcttag gtgattttta tcactatacc aattgagatg ggctagtcaa 3300

tgataattac tagtcttttt cttttgagtt gtgggtatct gtaaatctg ctagaccttt 3360

gctggaaaac ttgtaaattc tgctagacc cctgtaaatt ccgctagacc tttgtgtgtt 3420

ttttttgttt atattcaagt gggtataatt tatagaataa agaaagaata aaaaaagata 3480

aaaagaatag atcccagccc tgtgtataac tcactacttt agtcagttcc gcagtattac 3540

aaaaggatgt cgcaaacgct gtttgctcct ctacaaaaca gaccttaaaa ccctaaggc 3600
 ttaagtagca cctcgcagc ctggggcaaa tcgctgaata ttccttttgt ctccgacat 3660
 caggcacctg agtcgctgtc tttttcgtga cattcagttc gctgcgctca cggctctggc 3720
 agtgaatggg ggtaaatggc actacaggcg ccttttatgg attcatgcaa ggaaactacc 3780
 cataatacaa gaaaagcccg tcacgggctt ctacgggcgt tttatggcgg gtctgctatg 3840
 tgggtgctatc tgactttttg ctgttcagca gttcctgccc tctgattttc cagtctgacc 3900

 atttcggatt atcccgtgac aggtcattca gactggctaa tgcacccagt aaggcagcgg 3960
 tatcatcaac aggtttaccc gtcttactgt cgggaattca tttaaatagt caaaagcctc 4020
 cgaccggagg cttttgactg ctaggcgatc tgtgctgttt gccacggtat gcagcaccag 4080
 cgcgagatta tgggctcgca cgctcgactg tcggacgggg gcactggaac gagaagtcag 4140
 gcgagccgtc acgcccctga caatgccaca tctgagcaa ataattcaac cactaaacaa 4200
 atcaaccgcg tttcccgagg gtaaccaagc ttgcgggaga gaatgatgaa caagagccaa 4260
 caagttcaga caatcacctt ggccgcccgc cagcaaatgg cggcggcggg ggaaaaaaaa 4320

 gccactgaga tcaacgtggc ggtggtgttt tccgtagtgt accgcggagg caacacgtg 4380
 cttatccagc ggatggacga ggctctctgc tccagctgcg atatttcct gaataaagcc 4440
 tggagcgctt gcagccigaa gcaaggtacc catgaaatta cgtcagcggg ccagccagga 4500
 caatctctgt acggtctgca gctaaccaac caacagcgaa ttattatatt tggcggcggc 4560
 ctgccagtta tttttaatga gcaggttaatt ggccgctcgc gcgttagcgg cggtagcgtc 4620
 gagcaggatc aattattagc ccagtgcgcc ctggattgtt tttccgatt ataacctgaa 4680
 gcgagaaggt atattatgag ctatcgtatg ttccgccagg cattctgagt gttaacgagg 4740

 ggaccgtcat gtcgctttca ccgccaggcg tacgcctgtt ttacgatccg cgcgggcacc 4800
 atgccggcgc catcaatgag ctgtgctggg ggctggagga gcaggggggc cctgcccaga 4860
 ccataacctt tgacggaggc ggtgacgccg ctgcgctggg cgccctggcg gccagaagct 4920
 cgccctcgcg ggtgggtatc gggctcagcg cgtccggcga gatagccctc actcatgcc 4980
 agctgccggc ggacgcgcgc ctggctaccg gacacgtcac cgatagcgac gatcaactgc 5040
 gtacgctcgg cgccaacgcc gggcagctgg ttaaagtcct gccgttaagt gagagaaact 5100
 gaatgtatcg tatctatacc cgcaccgggg ataaaggcac caccgccctg tacggcggca 5160

 gccgcacga gaaagacat attcgcgtcg aggcctacgg caccgtcgat gaactgatat 5220
 cccagctggg cgtctgctac gccacgacc gcgacgccgg gctgcgggaa agcctgcacc 5280
 atattcagca gacgctgttc gtgctggggg ctgaactggc cagcgatgcg cggggcctga 5340
 cccgcctgag ccagacgac ggcaagagg agatcaccgc cctggagcgg ctatcgacc 5400

gcaatatggc cgagagcggc ccgttaaaac agttcgtgat cccggggagg aatctcgctt 5460
ctgcccgact gcacgtggcg cgcaccagct cccgtcggct cgaacgcctg ctgacggcca 5520
tggaccgcg ccatccgtg cgcgacgcg tcaaacgcta cagcaatcgc ctgtcggatg 5580

ccctgttctc catggcgga atcgaagaga ctaggcctga tgcttgcgct tgaactggcc 5640
tagcaaacac agaaaaaagc ccgcacctga cagtgcgggc ttttttttc ctaggcgatc 5700
tgtgctgttt gccacggtat gcagaccag cgcgagatta tgggctcgca cgctcgactg 5760
tcggacgggg gcaactggaac gagaagtcag gcgagccgct acgcccttga caatgccaca 5820
tcctgagcaa ataattcaac cactaaacaa atcaaccgcg tttcccgag gtaaccaagc 5880
ttcacctttt gagccgatga acaatgaaa gatcaaacg atttcagta ctggcccagc 5940
gccccgtcaa tcaggacggg ctgattggcg agtggcctga agaggggctg atcgccatgg 6000

acagccctt tgaccggctc tcttcagtaa aagtggacaa cggctctgac gtcgaactgg 6060
acggcaaagc ccgggaccag tttgacatga tcgaccgatt tatcgccgat tacgcatca 6120
acgttgagcg cacagagcag gcaatgcgcc tggaggcggt ggaaatagcc cgtatgctgg 6180
tggatattca cgtcagccgg gaggagatca ttgccatcac taccgcatc acgccggcca 6240
aagcggtcga ggtgatggcg cagatgaacg tggtaggat gatgatggcg ctgcagaaga 6300
tgcgtgccc cggaccccc tccaaccagt gccacgtcac caatctaaa gataatccgg 6360
tgcagattgc cgtgacgcc gccaggccg ggtccgagg cttctcagaa caggagacca 6420

cggtcgggat cgcgcgtac gcgccgttta acgccctggc gctgttggc ggttcgcagt 6480
gcggccgcc cggcgtgtg acgcagtgt cggtaggaaga ggccaccgag ctggagctgg 6540
gcatcgtgg cttaaccagc tacgccgaga cgggtgcggt ctacggcacc gaagcggat 6600
ttaccgacgg cgatgatag ccgtggtcaa aggcgttct cgcctcgcc tacgcctccc 6660
gcgggttgaa atgcgctac acctccgca ccggtaccga agcgtgatg ggctattcgg 6720
agagcaagtc gatgctctac ctgcaatcgc gctgcatctt cattactaaa ggcccgggg 6780
ttcagggact gcaaacggc gcggtgagct gtatcgcat gaccggcgt gtgccgtcgg 6840

gcattcgggc ggtgctggcg gaaaacctga tcgcctctat gctcgacct gaagtggcgt 6900
ccgccaagca ccagacttc tccactcgg atattcgcc caccgcgcg acctgatgc 6960
agatgctgcc gggcaccgac tttatcttct ccggtacag cgcggtgcc aactacgaca 7020
acatgttcgc cggtcgaac ttcatgacgg aagattttga tgattacaac atcctgcagc 7080
gtgacctgat ggttgacggc ggcctcgctc cggtagcca ggcggaacc attgccattc 7140
gccagaaagc ggcgcgggag atccaggcgg tttccgca gctggggctg ccgccaatcg 7200
ccgacgagga ggtggaggcc gccacctacg cgcacggcag caacgagatg ccgcccgta 7260

acgtggtgga ggatctgagt gcggtggaag agatgatgaa gcgcaacatc accggcctcg 7320
atattgtcgg cgcgctgagc cgcagcggct ttgaggatat cgccagcaat attctcaata 7380
tgctgcgcca gggggtcacc ggcgattacc tgcagacctc ggccattctc gatcggcagt 7440
tcgaggtggt gagtgcggtc aacgacatca atgactatca ggggccgggc accggctatc 7500
gcatctctgc cgaacgtcgg gcggagatca aaaatatcc gggcgtggtt cagccccgaca 7560
ccattgaata aggcggtiatt cctgtgcaac agacaacca aattcagccc tcttttacc 7620
tgaaaacccg cgagggcggg gtagcttctg ccgatgaacg cgccgatgaa gtggtgatcg 7680

gcgtcggccc tgccttcgat aaacaccagc atcacactct gatcgatatg ccccatggcg 7740
cgatcctcaa agagctgatt gccggggtgg aagaagaggg gcttcacgcc cgggtggtgc 7800
gcattctcgc cagctccgac gtctccttta tggcctggga tgcggccaac ctgagcggct 7860
cggggatcgg catcggtatc cagtcaagg ggaccacggg catccatcag cgcgatctgc 7920
tgccgctcag caacctggag ctgttctccc aggcgccgct gctgacgtg gagacctacc 7980
ggcagattgg caaaaacgt gcgcgctatg cgcgcaaaga gtcaccttcg ccggtgccgg 8040
tggtgaacga tcagatggtg cggccgaaat ttatggccaa agccgcgcta tttcatatca 8100

aagagaccaa acatgtggtg caggacgccg agcccgtcac cctgcacatc gacttagtaa 8160
gggagtgacc atgagcgaga aaacatgcg cgtgcaggat tatccgttag ccaccgctg 8220
cccggagcat atcctgacgc ctaccggcaa accattgacc gatattacc tcgagaaggt 8280
gtctctctggc gaggtgggcc cgcaggatgt gcggatctcc cgccagacc ttgagtacca 8340
ggcgcagatt gccgagcaga tgcagcgcca tgcggtggcg cgcaatttc gccgcgcggc 8400
ggagcttacc gccattcctg acgagcgcat tctggctatc tataacgcgc tgcgccgtt 8460
ccgctcctcg caggcggagc tgctggcgat cgccgacgag ctggagcaca cctggcatgc 8520

gacagtgaat gccgcctttg tccgggagtc ggcggaagtg tatcagcagc ggcataagct 8580
gcgtaaagga agctaagcgg aggtcagcat gccgttaata gccgggattg atatcggcaa 8640
cgccaccacc gaggtggcgc tggcgtccga ctaccgcag gcgagggcgt ttgttgccag 8700
cgggatcgtc gcgacgacgg gcatgaaagg gacgcgggac aatatcgccg ggacctcgc 8760
cgcgctggag caggccctgg cgaaaacacc gtggtcgatg agcgatgtct ctcgcatcta 8820
tcttaacgaa gccgcgccgg tgattggcga tgtggcgatg gagaccatca ccgagacat 8880
tataccgaa tcgaccatga tcggtcataa cccgcagacg ccgggcgggg tgggcgttgg 8940

cgtaggggacg actatcgccc tcgggcggct ggcgacgtg ccggcggcgc agtatgccga 9000
ggggtggatc gtactgattg acgacgccgt cgatttcctt gacgccgtgt ggtggctcaa 9060

ttaggcgctc gaccggggga tcaacgtggt ggccggcgatc ctcaaaaagg acgacggcgt 9120
 gctggtgaac aaccgcctgc gtaaaacct gccggtggtg gatgaagtga cgctgctgga 9180
 gcaggtcccc gagggggtaa tggcggcggt ggaagtggcc gcgccgggcc aggtggtgcg 9240
 gatcctgtcg aatccctacg ggatcgccac ctctctcggg ctaagcccgg aagagacca 9300
 ggccatcgtc cccatcgccc gcgccctgat tggcaaccgt tccgcggtgg tgctcaagac 9360

 cccgcagggg gatgtgcagt cgcggtgat cccggcgggc aacctctaca ttagcggcga 9420
 aaagcggcg ggagaggccg atgtcgccga ggccgcggaa gccatcatgc aggcgatgag 9480
 cgctgctgct ccggtacgag acatccgagg cgaaccgggc acccacgccc gcggcatgct 9540
 tgagcgggtg cgcaaggtaa tggcgtccct gaccggccat gagatgagcg cgataacat 9600
 ccaggatctg ctggcggtgg atacgtttat tccgcgcaag gtgcagggcg ggatggccgg 9660
 cgagtgcgcc atggagaatg ccgtcgggat ggccggcgatg gtgaaagcgg atcgtctgca 9720
 aatgcaggtt atcgcccgcg aactgagcgc ccgactgcag accgaggtgg tgggtggcgg 9780

 cgtaggaggc aacatggcca tcgccggggc gttaccact cccggtgtg cgccgcccgt 9840
 ggcgatcctc gacctggcg ccggtcgac ggatgcggcg atcgtcaac cggaggggca 9900
 gataacggcg gtccatctcg ccggggcggg gaatatggtc agcctgttga ttaaaaccga 9960
 gctgggcctc gaggatcttt cgctggcgga agcgataaaa aaataccgc tggccaaagt 10020
 ggaaagcctg ttcagtattc gtcacgagaa tggcgcggtg gattctttc gggaagccct 10080
 cagcccggcg gtgttcgcca aagtgggtga catcaaggag ggcgaaactg tgccgatcga 10140
 taacgccagc ccgtggaaa aaattcgtct cgtgcgccgg caggcgaaag agaaagtgtt 10200

 tgtcaccaac tgctgcgag cgctgcgcca ggtctcacc ggcggttcca ttcgcgatat 10260
 cgctttgtg gtgctgggtg gcggtcatc gctggacttt gagatcccgc agcttatcac 10320
 ggaagcctt tgcaactatg gcgtggtcgc cgggcagggc aatattcggg gaacagaagg 10380
 gccgcgaat gcggtcgcca ccgggtgct actggccggt caggcgaatt aaacggcgc 10440
 tcgcgccagc ctctaggtac aaataaaaaa ggcacgtcag atgacgtgcc tttttcttg 10500
 tctaggtac tggcgaaagg gggatgtgct gcaaggcgat taagtgggt aacgccaggg 10560
 ttttccagc cagcaggtg taaaacgacg gccagtgaat tcgagctcgg taccggggc 10620

 ggccgcgcta gcgccgac cagctggagt ttgtagaaac gcaaaaaggc catccgtcag 10680
 gatggccttc tgcttaattt gatgcctggc agtttatggc gggcgctctg cccgccacc 10740
 tccgggccgt tgcttcgcaa gtttcaaatc cgtcccgcc ggatttgtcc tactcaggag 10800
 agcgttcacc gacaaacaac agataaaacg aaaggccag tcttcgact gagcctttcg 10860
 ttttatttga tgctggcag ttcctactc tcgcatgggg agaccccaca ctaccatcg 10920

cgctacggcg tticacttct gagttcggca tggggtcagg tgggaccacc gcgctactgc 10980
cgccaggcaa attctgtttt atcagaccgc ttctgcgttc tgatttaatc tgtatcaggc 11040

tgaaaatctt ctctcatccg ccaaaacagc caagcttgca tgcctgcagc ccgggttacc 11100
atttcaacag atcgtcctta gcatataagt agtcgtcaaa aatgaattca acttcgtctg 11160
tttcggcatt gtagccgcca actctgatgg attcgtgggt tttgacaatg atgtcacagc 11220
ctttttcctt taggaagtcc aagtcgaaag tagtggcaat accaatgac ttacaaccgg 11280
cggtttttcc ggcggaataa cctgctggag cgtcttcaaa tactactacc ttagatttgg 11340
aagggtcttg ctcatcgatc ggatatacta agccattcct gcccttcaga tatggttctg 11400
gatgaggctt accctgtttg acatcattag cggtaatgaa gtactttggt ctctgattc 11460

ccagatgctc gaaccatttt tgtgccatat cacgggtacc ggaagttgcc acagccatt 11520
tctcttttgg tagagcgttc aaagcgttgc acagcttaac tgcacctggg acttcaatgg 11580
atttttcacc gtacttgacc ggaatttcag ctctcaattt gttaacatac tcttcattgg 11640
caaagtctgg agcgaactta gcaatggcat caaacgttct ccaaccatgc gagacttga 11700
taacgtgttc agcatgaaa taaggtttgi ccttaccgaa atccctccag aatgcagcaa 11760
tggctggttg agagatgata atggtagcgt cgacgtcgaa caaagcggcg ttaactttca 11820
aagatagagg ttiagtagtc aatcccataa ttctagtctg tttcctggat ccaataaatc 11880

taatcttcat gtagatctaa ttcttcaatc atgtccggca ggttcttcat tgggtagttg 11940
ttgtaaacga ttigtatata ggcttcaaata aatgggaagt cttcgacaga gccacatgtt 12000
tccaaccatt cgtgaacttc ttgacagta attaaacctt gagcggattg gccattcaac 12060
aactcctttt cacattccca ggcttcctta ccagaagtag ccattagcct agcaaccttg 12120
acgtttctac caccagcgca ggtggtgatc aaatcagcaa caccagcaga ctcttggtag 12180
tatgtttctt ctctagattc tgggaaaaac atttgaccga atctgatgat ctacccaaa 12240
ccgactcttt ggatggcagc agaagcgttg ttacccagc ctagaccttc gacgaaacca 12300

caacctaagg caacaacgtt cttcaaagca ccacagatgg agataccagc aacatcttcg 12360
atgacactaa cgtggaagta aggtctgttg aacaaggcct ttagaacctt atggtcgacg 12420
tccttgccct cgcctctgaa atcctttgga atgtggtgaa caactgttgt ttcagaccag 12480
tgttcttgag cgacttcggt ggcaatgtta gcaccagata gagcaccaca ttgaatacct 12540
agttcctcag tgatgtaaga ggatagcaat tggacacctt tagcaccaac ttcaaaacct 12600
tttagacagg agatagctct gacgtgtgaa tcaacatgac ctttcaattg gctacagata 12660
cggggcaaaa attgatgtgg aatgttgaaa acgatgatgt cgacatcctt gactgaatca 12720

atcaagtctg gattagcaac caaatgtctg ggtagagtga tgccaggcaa gtatttcacg 12780
 ttttgatgtc tagtatttat gatttcagtc aatttttcac cattgatctc ttcttcgaac 12840
 acccacattt gtactattgg agcgaaaact tctgggtatc cttacaatt ttcggcaacc 12900
 accttggcaa tagtagtacc ccagttacca gatccaatca cagtaacctt gaaaggcttt 12960
 tcggcagcct tcaaagaaac agaagaggaa cttctctttc taccagcatt caagtggccg 13020
 gaagttaagt ttaatctatc agcagcagca gccatggaat tgtcctcctt actagtcatg 13080
 gtctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacattata cgagccggat 13140

gattaattgt caacagctca ttccagaata ttgccagaa ccgttatgat gtcggcgcaa 13200
 aaaacattat ccagaacggg agtgcgccct gagcgacacg aattatgcag tgatttacga 13260
 cctgcacagc cataccacag cttccgatgg ctgcctgacg ccagaagcat tggcgcacgc 13320
 tagccagtac atttaaatgg taccctctag tcaaggcctt aagtgagtcg tattacggac 13380
 tggccgtcgt ttacaacgt cgtgactggg aaaaccctgg cgttaccaa cttaatcgcc 13440
 ttgcagcaca tcccccttc gccagctggc gtaatagcga agaggcccg accgatcgcc 13500
 cttccaaca gttgcgcagc ctgaatggcg aatggcgcc gatgcggtat ttctctctta 13560

cgcatctgtg cggtatttca caccgcataat ggtgcactct cagtacaatc tgctctgatg 13620
 ccgatagtt aagccagccc cgacacccgc caacacccgc tgacgagct 13669

<210> 73

<211> 13543

<212> DNA

<213> Artificial Sequence

<220><223> Plasmid

<400> 73

tagtaaagcc ctcgctagat tttaatgcgg atgttgcgat tacttcgcca actattgcga 60
 taacaagaaa aagccagcct ttcatgatat atctccaat ttgtgtaggg cttattatgc 120
 acgcttaaaa ataataaaag cagacttgac ctgatatgtt ggctgtgagc aattatgtgc 180
 ttagtgcatc taacgcttga gttaagccgc gccgcgaagc ggctgcggt tgaacgaatt 240

gttagacatt atttgcgcac taccttgggtg atctcgccct tcacgtagt gacaaattct 300
 tccaactgat ctgcgcgcga ggccaagcga tcttcttctt gtccaagata agcctgtcta 360
 gcttcaagta tgacgggctg atactgggcc ggcaggcgct ccattgccca gtcggcagcg 420
 acatccttcg gcgcgatttt gccggttact gcgctgtacc aaatgcggga caacgtaagc 480
 actacatttc gtcacatgcc agcccagtcg ggccggcgagt tccatagcgt taaggtttca 540
 ttttagcgct caaatagatc ctgttcagga accggatcaa agagtctctc cgccgctgga 600

cctaccaagg caacgctatg ttctcttgcg tttgtcagca agatagccag atcaatgtcg 660

atcgtggctg gctcgaagat acctgcaaga atgtcattgc gctgccattc tccaaattgc 720

agttcgcgct tagctggata acgccacgga atgatgtcgt cgtgcacaac aatgggtgact 780

tctacagcgc ggagaatctc gctctctcca ggggaagccg aagtttcaa aaggtcgttg 840

atcaaagctc gccgcgttgt ttcatcaagc cttacggtea ccgtaaccag caaatcaata 900

tcactgtgtg gcttcaggcc gccatccact gcggagccgt acaaatgtac ggccagcaac 960

gtcggttcga gatggcgctc gatgacgcca actacctctg atagttgagt cgatacttcg 1020

gcgatcaccg cttccctcat gatgtttaac tttgttttag ggcgactgcc ctgctgcgta 1080

acatcgttgc tgctccataa catcaaacaat cgacccacgg cgtaacgcgc ttgctgcttg 1140

gatgcccag gcatagactg taccceaaaa aaacagtcac aacaagccat gaaaaccgcc 1200

actgcgccgt taccaccgct gcgttcggtc aaggttctgg accagttgcg tgagcgcata 1260

cgctacttgc attacagctt acgaaccgaa caggcttatg tccactgggt tcgtgccttc 1320

atccgtttcc acggtgtgcg tcaccgggca accttgggca gcagcgaagt cgaggcattt 1380

ctgtcctggc tggcgaacga gcgcaagggt tcggtctcca cgcatcgtca ggcattggcg 1440

gccttgctgt tcttctacgg caaggtgctg tgcacggatc tgccttggtc tcaggagatc 1500

ggaagacctc ggccgtcgcg gcgcttgccg gtggtgctga ccccgatga agtggttcgc 1560

atcctcggtt ttctggaagg cgagcatcgt ttgttcgcc agcttctgta tggaacgggc 1620

atgcggatca gtgagggttt gcaactgcgg gtcaaggatc tggatttcga tcacggcacg 1680

atcatcgtgc gggagggcaa gggctccaag gatcgggcct tgatgttacc cgagagcttg 1740

gcaccagacc tgcgcgagca ggggaattaa ttcccacggg ttttgcgtcc cgcaaacggg 1800

ctgttctggt gttagctagt ttgtatcaga atcgagatc cggttcagc cggtttgccg 1860

gctgaaagcg ctatttcttc cagaattgcc atgatttttt ccccacggga ggcgtcactg 1920

gtcccgctgt tgicggcagc ttgtattcga taagcagcat cgcctgtttc aggctgtcta 1980

tgtgtgactg ttgagctgta acaagttgtc tcagggtgtc aatttcattg tctagtgtct 2040

ttgttttact ggtttcacct gttctattag gtgttacatg ctgttcattc gttacattgt 2100

cgatctgttc atggtgaaca gctttgaatg caccaaaaac tcgtaaaagc tctgatgtat 2160

ctatcttttt tacaccgttt tcatctgtgc atatggacag ttttcccttt gatatgtaac 2220

ggtgaacagt tgttctactt ttgtttgtta gtcttgatgc ttactgata gatacaagag 2280

ccataagaac ctcagatcct tccgtattta gccagtatgt tctctagtgt gggttcgttg 2340

ttttgcgtga gccatgagaa cgaaccattg agatcatact tactttgcat gtcactcaaa 2400

aattttgcct caaaactggg gagctgaatt ttgacagtta aagcatcgtg tagtggtttt 2460
cttagtccgt tatgttagta ggaatctgat gtaatgggtg ttggtatttt gtcaccattc 2520
atTTTTatct ggttggtctc aagttcgggt acgagatcca ttgtctatc tagttcaact 2580
tgaaaaatca acgtatcagt cgggcggcct cgcttatcaa ccaccaattt catattgctg 2640
taagtgttta aatctttact tattggtttc aaaaccattt ggtaagcct tttaaactca 2700
tggtagttaa ttcaagcat taacatgaac ttaaattcat caaggctaatt ctctatattt 2760

gccttgtagg ttttcttttg tgtagttct ttaataaacc actcataaat cctcatagag 2820
tatttgtttt caaaagactt aacatgttcc agattatatt ttatgaattt ttttaactgg 2880
aaaagataag gcaatatctc ttcactaaaa actaattcta attttcgct tgagaacttg 2940
gcatagtttg tccactggaa aatctcaaag cctttaacca aaggattcct gatttccaca 3000
gttctcgtca tcagctctct ggttgcttta gctaatacac cataagcatt ttcctactg 3060
atgttcatca tctgagcgta ttggttataa gtgaacgata ccgtccgttc tttccttgta 3120
gggttttcaa tctgtggggt gagtagtgcc acacagcata aaattagctt ggtttcatgc 3180

tccgttaagt catagcgact aatcgctagt tcatttgctt tgaaaacaac taattcagac 3240
atacatctca attggctctg gtgattttta tctactatcc aattgagatg ggctagtcaa 3300
tgataattac tagtccittt cctttgagtt gtgggtatct gtaaatctg ctagaccttt 3360
gctggaaaac ttgtaaattc tgctagacc tctgtaaatt ccgctagacc tttgtgtgtt 3420
ttttttgttt atattcaagt ggttataatt tatagaataa agaaagaata aaaaaagata 3480
aaaagaatag atcccagccc tgtgtataac tctactttt agtcagttcc gcagtattac 3540
aaaaggatgt cgcaaagct gtttgcctt ctacaaaaca gaccttaaaa ccctaaaggc 3600

ttaagtagca cctcgcgaag ctgggcaaa tcgctgaata ttccttttgt ctccgacat 3660
caggcacctg agtcgctgtc ttttctgta cattcagttc gctgcgtca cggtctggc 3720
agtgaatggg ggtaaatggc actacaggcg ccttttatgg attcatgcaa ggaaactacc 3780
cataatacaa gaaaagcccc tcacgggctt ctacggcggt tttatggcgg gtctgctatg 3840
tggtgctatc tgactttttg ctgttcagca gttcctgccc tctgatttcc cagtctgacc 3900
acttcggatt atcccgtagc aggtcattca gactggctaa tgcacccagt aaggcagcgg 3960
tatcatcaac aggtttaccg gtcttactgt cgggaattca tttaaatagt caaaagcctc 4020

cgaccggagg cttttgactg ctaggcgatc tgtgctgttt gccacggtat gcagcaccag 4080
cgcgagatta tgggctcgca cgctcgactg tcggacgggg gcactggaac gagaagtcag 4140
gcgagccgtc acgcccttga ctatgccaca tctgagcaa ataattcaac cactaaacaa 4200
atcaaccgag tttcccgag gtaaccaagc ttgcgggaga gaatgatgaa caagagccaa 4260

caagttcaga caatcacctt ggccgcccgc cagcaaatgg cggcggcggt ggaaaaaaa 4320
 gccactgaga tcaacgtggc ggtggtgttt tccgtagtgt accgcggagg caacacgtg 4380
 ctatccagc ggatggacga ggccttcgtc tccagctgcg atatttcct gaataaagc 4440

 tggagcgctt gcagcctgaa gcaaggtacc catgaaatta cgtcagcggt ccagccagga 4500
 caatctctgt acggtctgca gctaaccaac caacagcgaa ttattatatt tggcggcggc 4560
 ctgccagtta ttttaatga gcagtaatt ggccgctcgc gcgttagcgg cggtagcggtc 4620
 gagcaggatc aattattagc ccagtgcgcc ctggtattgt tttccgatt ataacctgaa 4680
 gcgagaaggt atattatgag ctatcgtatg ttccgccagg cattctgagt gttaacgagg 4740
 ggaccgtcat gtcgcttca ccgccaggcg tacgcctgtt ttacgatccg cgcgggcacc 4800
 atgccggcgc catcaatgag ctgtgctggg ggctggagga gcagggggtc ccctgccaga 4860

 ccataaccta tgacggaggc ggtgacgccg ctgcgtggg cgccttgcg gccagaagct 4920
 cgccttcgc ggtgggtatc gggctcagcg cgtccgcga gatagccctc actcatgcc 4980
 agctccggc ggacgcgccg ctggctaccg gacacgtcac cgtatgcgc gatcaactgc 5040
 gtacgtcgg cgccaacgc gggcagctgg ttaaagtcct gccgttaagt gagagaaact 5100
 gaatgtatc tatctatacc cgcaccgggg ataaaggcac caccgccctg tacggcgcca 5160
 gccgcatcga gaaagacat attcgcgtcg aggcctacgg caccgtcgat gaactgatat 5220
 ccagctggg cgtctgtac gccacgacc gcgacgccg gctgcgggaa agcctgcacc 5280

 atattcagca gacgtgttc gtgctggggg ctgaactggc cagcgatgcg cggggcctga 5340
 cccgctgag ccagacgac ggcaagagg agatcacgc cctggagcgg ctatcgacc 5400
 gcaatatggc cgagagcggc ccgttaaac agttcgtgat cccggggagg aatctgcct 5460
 ctgccagct gcacgtggcg cgcaccagc cccgtcggct cgaacgcctg ctgacggcca 5520
 tggaccgcgc gcatccgtg cgcgacgcgc tcaaacgcta cagcaatcgc ctgtcggatg 5580
 ccctgttctc catggcgga atcgaagaga ctaggcctga tgcttgcgt tgaactggcc 5640
 tagcaaacac agaaaaagc ccgcacctga cagtgcgggc ttttttttc ctaggcgatc 5700

 tgtgtgttt gccacggtat gcagcaccag cgcgagatta tgggctcgca cgctcgactg 5760
 tcggacgggg gactggaac gagaagtcag gcgagccgtc acgcccttga ctatgccaca 5820
 tcttgagcaa ataattcaac cactaataca atcaaccgcg tttcccgag gtaaccaagc 5880
 ttacatttt gagccgatga acaatgaaaa gatcaaacg atttgcagta ctggccagc 5940
 gccccgtcaa tcaggacggg ctgattggcg agtggcctga agaggggctg atcgccatgg 6000
 acagccctt tgaccgggtc tcttcagtaa aagtggacaa cggctcgatc gtcgaactgg 6060
 acggcaaac cggggaccag tttgacatga tcgaccgatt tatcgccgat tacgcgatca 6120

acgttgagcg cacagagcag gcaatgcgcc tggaggcggg ggaatagcc cgtatgctgg 6180
 tggatattca cgtcagccgg gaggagatca ttgccatcac taccgccatc acgccggcca 6240
 aagcggtcga ggtgatggcg cagatgaacg tggaggagat gatgatggcg ctgcagaaga 6300
 tgcgtgcccc ccggaccccc tccaaccagt gccacgtcac caatctcaa gataatccgg 6360
 tgcagattgc cgtgacgcc gccgagccgg ggatccggcg cttctcagaa caggagacca 6420
 cggtcggtat cgcgcgtac gcgcgcttta acgccctggc gctgttggc ggttcgcagt 6480
 gcggccgccc cggcgtgttg acgcagtgtc cggtggaaga ggccaccgag ctggagctgg 6540

gcatgcgtgg cttaccagc tacgccgaga cgggtgctgg ctacggcacc gaagcggat 6600
 ttaccgacgg cgatgatacg ccgtggtcaa aggcgttct cgcctcggcc tacgcctccc 6660
 gcgggttgaa aatgcgtac acctccggca ccggatccga agcgtgatg ggctattcgg 6720
 agagcaagtc gatgctctac ctgaatcgc gctgcatctt cattactaaa ggccgcccggg 6780
 ttcagggaact gcaaacggc gcggtgagct gtatcgcat gaccggcgct gtgccgtcgg 6840
 gcattcgggc ggtgctggcg gaaaacctga tcgcctctat gctcgacct gaagtggcgt 6900
 ccgccaacga ccagacttc tccactcgg atattcgcc caccgcgcgc acctgatgc 6960

agatgctgcc gggcaccgac tttattttct ccggctacag cgcggtgcc aactacgaca 7020
 acatgttcgc cggtcgaac ttcgatgcgg aagattttga tgattacaac atcctgcagc 7080
 gtgacctgat ggttgacggc ggctgcgtc cggtgaccga ggccgaaacc attgccattc 7140
 gccagaaagc ggccgcccgg atccaggcgg tttccgcga gctggggctg ccgccaatcg 7200
 ccgacgagga ggtggaggcc gccacctacg cgcacggcag caacgagatg ccgccgcgta 7260
 acgtggtgga ggatctgagt gcggtggaag agatgatgaa gcgcaacatc accggcctcg 7320
 atattgtcgg cgcgctgagc cgcagcggct ttgaggatat cgccagcaat attctcaata 7380

tgctgcgcca gcgggtcacc ggcgattacc tgcagacctc ggccattctc gatcggcagt 7440
 tcgaggtggt gagtgcggtc aacgacatca atgactatca ggggccgggc accggctatc 7500
 gcatctctgc cgaacgtcgg gcggagatca aaaatatcc gggcgtggtt cagcccgaca 7560
 ccattgaata aggcggtatt cctgtgcaac agacaacca aattcagccc tcttttacc 7620
 tgaaaacccg cgaggcgagg gtagcttctg ccgatgaacg cgccgatgaa gtggtgatcg 7680
 gcgtcggccc tgccttcgat aaacaccagc atcacactct gatcgatatg ccccatggcg 7740
 cgatcctcaa agagctgatt gccggggtgg aagaagagg gcttcacgcc cgggtggtgc 7800

gcattctgcg cagtcgcac gtctcttta tggcctggga tgcggccaac ctgagcggct 7860
 cggggatcgg catcggtatc cagtccaagg ggaccacgg catccatcag cgcgatctgc 7920

tgccgctcag caacctggag ctgttctccc aggcgccgct gctgacgctg gagacctacc 7980
 ggcagattgg caaaaacgct gcgcgctatg cgcgcaaaga gtcaccttcg ccggtgccgg 8040
 tggatgaacga tcagatgggtg cggccgaaat ttatggccaa agccgcgcta ttcatatca 8100
 aagagaccaa acatgtgggtg caggacgccg agcccgtcac cctgcacatc gacttagtaa 8160
 gggagtgacc atgagcgaga aaaccatgcg cgtgcaggat tatccgttag ccacccgctg 8220

 cccggagcat atcctgacgc ctaccggcaa accattgacc gatattacc tcgagaaggt 8280
 gctctctggc gaggtgggccc cgcaggatgt gcggatctcc cgccagaccc ttgagtacca 8340
 ggcgagcagatt gccgagcaga tgcagcgcca tgcggtggcg cgcaatttcc gccgcgcggc 8400
 ggagcttctc gccattctg acgagcgcat tctggctatc tataacgcgc tgcgcccgtt 8460
 ccgctcctcg caggcggagc tgctggcgat cgcgcagcag ctggagcaca cctggcatgc 8520
 gacagtgaat gccgcctttg tccgggagtc ggcggaagtg tatcagcagc ggcataagct 8580
 gcgtaaagga agctaagcgg aggtcagcat gccgttaata gccgggattg atatcggcaa 8640

 cgccaccacc gaggtggcgc tggcgctcga ctaccgcag gcgagggcgt ttgttgccag 8700
 cgggatcgtc gcgacgacgg gcatgaaagg gacgcgggac aatatcgccg ggaccctcgc 8760
 cgcgctggag caggccctgg cgaaaacacc gtggtcgatg agcgatgtct ctgcacatca 8820
 tcttaacgaa gccgcgccgg tgattggcga tgtggcgatg gagacatca ccgagacat 8880
 tatcaccgaa tcgacctga tcggtcataa cccgcagacg ccgggcgggg tggcgcttgg 8940
 cgtggggacg actatcgccc tcggcgggct ggcgacgctg ccggcgggcg agtatgccga 9000
 ggggtggatc gtactgattg acgacgccgt cgatttcctt gacgccgtgt ggtggctcaa 9060

 tgaggcgctc gaccggggga tcaactggtt ggccggcgatc ctcaaaaagg acgacggcgt 9120
 gctggtgaac aaccgcctgc gtaaaacctt gccggtgggtg gatgaagtga cgctgctgga 9180
 gcaggtcccc gagggggtaa tggcggcggt ggaagtggcc gcgccgggcc aggtggtgctg 9240
 gatcctgtcg aatccctacg ggatcgccac cttcttcggg ctaagcccgg aagagaccca 9300
 ggccatcgtc cccatcgccc gcgcctgat tggcaaccgt tccgcggtgg tgctcaagac 9360
 cccgcagggg gatgtgcagt cgcggtgat cccggcgggc aacctctaca ttagcggcga 9420
 aaagcggcg gcgagggccg atgtcgccga gggcgcgga gccatcatgc aggcgatgag 9480

 cgctcgctc ccggtacgcg acatccgcgg cgaaccgggc acccagccg gcggcatgct 9540
 tgagcgggtg cgcaaggtaa tggcgctcct gaccggccat gagatgagcg cgatatacat 9600
 ccaggatctg ctggcggtgg atacgtttat tccgcgcaag gtgcagggcg gtagggccgg 9660
 cgagtgcgcc atggagaatg ccgtcgggat ggccggcgatg gtgaaagcgg atcgtctgca 9720
 aatgcaggtt atcgcccgcg aactgagcgc ccgactgcag accgaggtgg tgggtggcg 9780

cgtggaggcc aacatggcca tcgccggggc gttaccact cccggctgtg cggcgccgt 9840
 ggcgatcctc gacctggcg ccggctcgac ggatgcggcg atcgtaacg cggaggggca 9900

 gataacggcg gtccatctcg ccggggcggg gaatatggc agcctgttg ttaaaaccga 9960
 gctgggcctc gaggatcttt cgctggcgga agcgataaaa aaataccgc tggccaaagt 10020
 ggaaagcctg ttcagtattc gtcacgagaa tggcgcggtg gagttcttc gggaagccct 10080
 cagcccgcg gtgttcgcca aagtgggtg catcaaggag ggcgaactgg tgccgatcga 10140
 taacgccagc ccgtggaaa aaattcgtct cgtgcgccg caggcgaaag agaaagtgtt 10200
 tgtcaccaac tgcctgcg cgctgcgcca ggtctcacc ggcggttcca ttcgcgatat 10260
 cgctttgtg gtgtgtgtg gcggctcatc gctggacttt gagatccgc agcttatcac 10320

 ggaagcctg tcgcactatg gcgtggtcgc cgggcagggc aatattcggg gaacagaagg 10380
 gccgcgaat gcggtcgcca ccgggtgct actggccggt caggcgaatt aaacggcgcg 10440
 tcgcgccagc ctctaggtac aaataaaaaa ggcacgtcag atgacgtgcc ttttttctg 10500
 tctagcgtg accaatgctt ctggcgtag gcagccatcg gaagctgtg tatggctgtg 10560
 caggctgtaa atcactgcat aattcgtgtc gctcaaggcg cactcccgtt ctggataatg 10620
 ttttttgcg cgacatcata acggttctg caaatattct gaaatgagct gttgacaatt 10680
 aatcatccg ctcgtataat gtgtggaatt gtgagcggat aacaatttca cacaggaaac 10740

 agaccatgac tagtaaggag gacaattcca tggtgtgtc tgctgataga ttaaaactaa 10800
 ctccggcca ctigaatgct ggtagaaaga gaagtcttc ttcgtttct ttgaaggctg 10860
 ccgaaaagcc tttaagggt actgtgattg gatctggtaa ctggggtact actattgcca 10920
 aggtggttgc cgaaaattgt aagggatacc cagaagtitt cgctccaata gtacaaatgt 10980
 gggtgttga agaagagatc aatggtgaaa aattgactga aatcataaat actagacatc 11040
 aaaacgtgaa atacttgctt ggcacactc tacccgacaa tttggttgc aatccagact 11100
 tgattgattc agtcaaggat gtcgacatca tcgttttcaa cattccacat caattttgc 11160

 cccgtatctg tagccaattg aaaggtcatg ttgattcaca cgtcagagct atctcctgtc 11220
 taaagggttt tgaagtgtt gctaaagggt tccaattgct atctcttac atcactagg 11280
 aactaggtat tcaatgtgtt gctctatctg gtgctaact tgccaccgaa gtcgctcaag 11340
 aacactggtc tgaacaaca gttgcttacc acattccaaa ggatttcaga ggcgagggca 11400
 aggacgtcga ccataagggt ctaaaggcct tgttcacag acctacttc cacgttagtg 11460
 tcatcgaaga tgttctgtt atctccatct gtggtgcttt gaagaacgtt gttgccttag 11520
 gttgtggttt cgtcgaagg ctaggctggg gtaacaacgc ttctgctgcc atccaaagag 11580

tcggtttggg tgagatcatc agattcggtc aaatgttttt cccagaatct agagaagaaa 11640
catactacca agagtctgct ggtgttgctg atttgatcac cacctgcgct ggtggtagaa 11700
acgtcaaggt tgctaggeta atggctactt ctggttaagga cgcctgggaa tgtgaaaagg 11760
agttgttgaa tggccaatcc gctcaagggt taattacctg caaagaagtt cacgaatggt 11820
tggaaacatg tggctctgct gaagacttcc cattatttga agcgtatac caaatcgttt 11880
acaacaacta cccaatgaag aacctgccgg acatgattga agaattagat ctacatgaag 11940
attagattta ttggatccag gaaacagact agaattatgg gattgactac taaacctcta 12000

tctttgaaag ttaacgccgc tttgttcgac gtcgacggta ccattatcat ctctcaacca 12060
gccattgctg cattctggag ggatttcggt aaggacaaac cttatttcga tgctgaacac 12120
gttatccaag tctcgcatgg ttggagaacg ttgatgccca ttgctaagtt cgctccagac 12180
tttgccaatg aagagtatgt taacaaattia gaagctgaaa ttccgggtcaa gtacggtgaa 12240
aatccattg aagtcacagg tgcagttaag ctgtgcaacg ctttgaacgc tctacaaaaa 12300
gagaaatggg ctgtggcaac ttccgggtacc cgtgatatgg cacaaaaatg gtctgagcat 12360
ctgggaatca ggagacaaa gtacttcatt accgctaag atgtcaaaca gggtaacctt 12420

catccagaac catatctgaa gggcaggaat ggcttaggat atccgatcaa tgagcaagac 12480
ccttccaaat ctaaggtagt agtatttgaa gacgctccag caggtattgc cgccggaaaa 12540
gccgccggtt gtaagatcat tggatttgcc actactttcg acttggaactt cctaaaggaa 12600
aaaggctgtg acatcattgt caaaaaccac gaatccatca gagtggcggt ctacaatgcc 12660
gaaacagacg aagtigaatt catttttgac gactacttat atgctaagga cgatctgttg 12720
aatggtaac ccgggctgca ggcatgcaag ctgggctgtt ttggcggatg agagaagatt 12780
ttcagcctga tacagattaa atcagaacgc agaagcggtc tgataaaaca gaatttgcct 12840

ggcggcagta gcgcggtggt cccacctgac cccatgccga actcagaagt gaaacgccgt 12900
agcgcgatg gtagtgtggg gtctcccat gcgagagtag ggaactgcc aagcatcaaat 12960
aaaacgaaag gctcagtga aagactgggc ctttcgtttt atctgttgtt tgcggtgaa 13020
cgctctcctg agtaggacaa atccgccggg agcggatttg aacgttgca agcaacggcc 13080
cggagggtgg cgggcaggac gcccgcata aactgccagg catcaaatta agcagaaggc 13140
catcctgacg gatggccttt ttgcgtttct acaaaactcca gctggatcgg gcgctagagt 13200
atacatttaa atggtaccct ctagtcaagg ccttaagtga gtcgtattac ggactggccg 13260

tcgttttaca acgtcgtgac tgggaaaacc ctggcggttac ccaacttaat cgcttgcag 13320
cacatcccc tttcgccagc tggcgtaata gcgaagaggc ccgcaccgat cgcccttccc 13380
aacagttgcg cagcctgaat ggcgaaatggc gcctgatgcg gtattttctc ctacgcac 13440

tgtgcggtat ttcacaccgc atatggtgca ctctcagtac aatctgctct gatgccgcat 13500
 agttaagcca gccccgacac ccgccaacac ccgctgacga gct 13543
 <210> 74
 <211> 13543
 <212> DNA
 <213> Artificial Sequence
 <220><223> Plasmid
 <400> 74
 tagtaaagcc ctcgctagat tttaatgcgg atgttgcgat tacttcgcca actattgcga 60
 taacaagaaa aagccagcct ttcattgatat atctcccaat ttgtgtaggg cttattatgc 120
 acgcttaaaa ataataaaag cagacttgac ctgatatgtt ggctgtgagc aattatgtgc 180
 ttagtgcatc taacgcttga gttaagccgc gccgcgaagc ggctgcggct tgaacgaatt 240
 gttagacatt atttgccgac taccttggtg atctcgccct tcacgtagtg gacaaattct 300
 tccaactgat ctgcgcgcga ggccaagcga tcttcttctt gtccaagata agcctgtcta 360
 gcttcaagta tgacgggctg atactgggcc ggccaggcgt ccattgccca gtcggcagcg 420
 acatccttcg gcgcgatttt gccggttact gcgctgtacc aaatgcggga caacgtaagc 480
 actacatttc gctcatcgcc agcccagtcg ggccggcaggt tccatagcgt taaggtttca 540
 tttagcgcct caaatagatc ctgttcagga accggatcaa agagttcttc cgccgctgga 600
 cctaccaagg caacgctatg ttctcttgct ttgttcagca agatagccag atcaatgtcg 660
 atcgtggctg gctcgaagat acctgcaaga atgtcattgc gctgccattc tccaaattgc 720
 agttcgcgt tagctggata acgccacgga atgatgtcgt cgtgcacaac aatggtgact 780
 tctacagcgc ggagaatctc gctctctcca ggggaagccg aagtttcaa aaggtcgttg 840
 atcaaagctc gccgcgttgt ttcattcaagc cttacggtca ccgtaaccag caaatcaata 900
 tctactgttg gcttcaggcc gccatccact gcggagccgt acaaatgtac ggccagcaac 960
 gtcggttcga gatggcgctc gatgacgcca actacctctg atagttgagt cgatacttcg 1020
 gcgatcaccg ctccctcat gatgtttaac ttgttttag ggccactgcc ctgctgcgta 1080
 acatcgttgc tgcctcataa catcaaacat cgaccacgg cgtaacgcgc ttgctgcttg 1140
 gatgcccag gcatagactg taccctaaaa aaacagtcac aacaagccat gaaaaccgcc 1200
 actgcgccgt taccaccgt gcgttcggtc aaggttctgg accagttgcg tgagcgcata 1260
 cgctacttgc attacagctt acgaaccgaa caggcttatg tccactgggt tcgtgccttc 1320
 atccgtttcc acggtgtgcg tcaccgggca accttgggca gcagcgaagt cgaggcattt 1380

ctgtcctggc tggcgaacga gcgcaagggt tcggtctcca cgcatcgta gccattggcg	1440
gccttctgtg tcttctacga caagggtgtg tgcacggatc tgccctggct tcaggagatc	1500
ggaagacctc ggccgtcgcg gcgcttgccg gtggtgtctga ccccgatga agtggttcgc	1560
atcctcgggt tcttgaagg cgagcatcgt ttgttcgcc agcttctgta tggaaacggc	1620
atgcggatca gtgagggttt gcaactgcgg gtcaaggatc tggatttcga tcacggcacg	1680
atcatcgtgc gggagggcaa gggctccaag gatcgggcct tgatgttacc cgagagcttg	1740
gcaccagcc tgcgcgagca ggggaattaa ttcccacggg ttttctgcc cgcaaacggg	1800
ctgttctggt gttgctagt ttgtatcaga atcgagatc cgcttcagc cggtttgccg	1860
gctgaaacgc ctatttcttc cagaattgcc atgatttttt cccacggga ggcgtcactg	1920
gtccccgtgt tgcggcagc ttgtattcga taagcagcat gcctgttgc aggtgtctta	1980
tgtgtgactg ttgagctgta acaagtgtc tcaggtgttc aatttcattg tctagtgtg	2040
ttgttttact ggtttcacct gttctattag gtgttacatg ctgttcattc gttacattgt	2100
cgatctgttc atggtgaaca gctttgaatg caccaaaaac tcgtaaaagc tctgatgtat	2160
ctatcttttt tacaccgttt tcatctgtgc atatggacag tttcccttt gatatgtaac	2220
ggtgaacagt ttttctact ttgtttgtta gtcttgatgc ttactgata gatacaagag	2280
ccataagaac ctcatcct tccgtattta gccagtatgt tctctagtgt ggttcgttgt	2340
ttttgcgtga gccatgagaa cgaaccattg agatcatact tactttgcat gtcactcaaa	2400
aattttgcct caaaactggt gagctgaatt ttgacgtta aagcatcgtg tagtgttttt	2460
cttagtccgt tatgtaggta ggaatctgat gtaatggttg ttggtatttt gtcaccattc	2520
atttttatct ggttggtctc aagtccggtt acgagatcca ttgtctatc tagttcaact	2580
tggaaaatca acgtatcagt cgggcggcct cgcttatcaa ccaccaattt catattgctg	2640
taagtgttta aatctttact tattggtttc aaaaccattt ggttaagcct tttaaaactca	2700
tggtagtatt ttcaagcat taacatgaac ttaaattcat caaggctaata ctctatattt	2760
gccttgtagg ttttcttttg tgttagttct ttaataacc actcataaat cctcatagag	2820
tatttgtttt caaaagactt aacatgttcc agattatatt ttatgaattt tttaactgg	2880
aaaagataag gcaatatctc ttactaaaa actaattcta attttcgct tgagaacttg	2940
gcatagtttg tccactggaa aatctcaaag ccttaacca aaggattcct gatttcaca	3000
gttctcgtca tcagctctct ggttgcttta gctaatacac cataagcatt ttcctactg	3060
atgttatca tctgagcgt ttggttataa gtgaacgata ccgtccgttc tttccttgta	3120
gggttttcaa tcgtggggtt gagtagtgcc acacagcata aaattagctt ggtttcatgc	3180
tccgttaagt catagcgact aatcgctagt tcatttgctt tgaaaacaac taattcagac	3240

atacatctca attggtctag gtgattttaa tcactatacc aattgagatg ggctagtcaa 3300
 tgataattac tagtcctttt cctttgagtt gtgggtatct gtaaattctg cttagacctt 3360
 gctggaaaac ttgtaaattc tgctagaccc tctgtaaatt ccgctagacc ttigtgtgtt 3420

 ttttttgttt atattcaagt gggtataatt tatagaataa agaaagaata aaaaaagata 3480
 aaaagaatag atcccagccc tgtgtataac tcactacttt agtcagttcc gcagtattac 3540
 aaaaggatgt cgcaaacgct gtttctcct ctacaaaaca gaccttaaaa ccctaaggc 3600
 ttaagtagca cctcgcgaag ctggggcaaa tcgctgaata ttccttttgt ctccgacct 3660
 caggcacctg agtcgctgtc ttttctgtga cattcagttc gctgcgctca cggctctggc 3720
 agtgaatggg ggtaaatggc actacaggcg ccttttatgg attcatgcaa ggaaactacc 3780
 cataatacaa gaaaagcccc tcacgggctt ctgaggcgt tttatggcgg gtctgctatg 3840

 tgggtgctatc tgactttttg ctgttcagca gtctctgccc tctgattttc cagtctgacc 3900
 acttcggatt atcccgtagc aggtcattca gactggctaa tgcacccagt aaggcagcgg 3960
 tatcatcaac aggcctaccc gtcttactgt cgggaattca tttaaatagt caaaagcctc 4020
 cgaccggagg cttttgactg ctaggcgatc tgtgctgttt gccacggtat gcagcaccag 4080
 cgcgagatta tgggctcgca cgctcgactg tcggacgggg gcaactggaac gagaagttag 4140
 gcgagccgtc acgcccttga caatgccaca tctgagcaa ataattcaac cactaataca 4200
 atcaaccgcg tttcccgag gtaaccaagc ttgcgggaga gaatgatgaa caagagccaa 4260

 caagttcaga caatcacctt ggccgcccgc cagcaaatgg cggcggcggg ggaaaaaaaa 4320
 gccactgaga tcaacgtggc ggtggtgttt tccgtagtgt accgcggagg caacacgctg 4380
 ctatccagc ggatggacga ggccttcgtc tccagctgcg atatttcct gaataaagcc 4440
 tggagcgcct gcagcctgaa gcaaggtacc catgaaatta cgtcagcggg ccagccagga 4500
 caatctctgt acggtctgca gctaaccaac caacagcgaa ttattatttt tggcggcggc 4560
 ctgccagtta ttttaatga gcaggttaatt ggcgccgtcg gcgttagcgg cggtagcgtc 4620
 gagcaggatc aattattagc ccagtgcgcc ctggattgtt tttccgatt ataacctgaa 4680

 gcgagaaggt atattatgag ctatcgtagt ttcgccagg cattctgagt gttaacgagg 4740
 ggaccgtcat gtcgctttca ccgccaggcg tacgcctgtt ttacgatccg cgcgggcacc 4800
 atgccggcgc catcaatgag ctgtgctggg ggctggagga gcagggggtc ccctgccaga 4860
 ccataaccta tgacggaggg ggtgacgccg ctgcgctggg cgccctggcg gccagaagct 4920
 cggccctgcg ggtgggtatc gggctcagcg cgtccggcga gatagccctc actcatgcc 4980
 agctgccggc ggacgcgccg ctggctaccg gacacgtcac cgatagcgac gatcaactgc 5040
 gtacgctcgg cgccaacgcc gggcagctgg ttaaagtcct gccgttaagt gagagaaact 5100

gaatgtatcg tatctatacc cgcaccgggg ataaaggcac caccgccctg tacggcggca 5160
 gccgcatcga gaaagaccat attcgcgtcg aggcttacgg caccgtcgat gaactgatat 5220
 cccagctggg cgtctgtac gccacgacc gcgacgccgg gctgcgggaa agcctgcacc 5280
 atattcagca gacgtgttc gtgctggggg ctgaactggc cagcgatgcg cggggcctga 5340
 cccgcctgag ccagacgac ggcaagagg agatcacgc cctggagcgg ctatcgacc 5400
 gcaatatggc cgagagcggc ccgttaaac agttcgtgat cccggggagg aatctgcct 5460
 ctgcccagct gcacgtggcg cgcaccagc cccgtcggct cgaacgcctg ctgacggcca 5520

 tggaccgcgc gcatccgtg cgcgacgcgc tcaaacgcta cagcaatcgc ctgtcggatg 5580
 ccctgttctc catggcgca atcgaagaga ctaggcctga tgcttgcgt tgaactggcc 5640
 tagcaaacac agaaaaagc ccgcacctga cagtgcgggc ttttttttc ctaggcgatc 5700
 tgtgtgttt gccacggtat gcagaccag cgcgagatta tgggctcgca cgctcgactg 5760
 tcggacgggg gcaactggaac gagaagtcag gcgagccgtc acgcccttga caatgccaca 5820
 tcctgagcaa ataattcaac cactaataca atcaaccgcg tttcccgag gtaaccaagc 5880
 ttcaccttt gagccgatga acaatgaaa gatcaaacg atttgagta ctggcccagc 5940

 gccccgtcaa tcaggacggg ctgattggcg agtggcctga agaggggctg atcgccatgg 6000
 acagccctt tgaccggtc tcttcagtaa aagtggacaa cggctctgac gtcgaactgg 6060
 acggcaaacg cgggaccag tttgacatga tcgaccgatt tatcgccgat tacgcatca 6120
 acgttagcgc cacagagcag gcaatgcgcc tggaggcggg ggaaatagcc cgtatgctgg 6180
 tggatattca cgtcagccgg gaggagatca ttgccatcac taccgccatc acgccggcca 6240
 aagcgtcga ggtgatggcg cagatgaacg tggaggagat gatgatggcg ctgcagaaga 6300
 tgcgtgcccg ccggaccccc tccaaccagt gccacgtcac caatctcaa gataatccgg 6360

 tgcagattgc cgtgacgcc gccgaggcgg ggatccgagg cttctcagaa caggagacca 6420
 cggctcgtat cgcgcgtac gcgccgttta acgccctggc gctgttggtc ggttcgcagt 6480
 gcggccgcc cggcgtgttg acgcagtgt cggtggaaga ggccaccgag ctggagctgg 6540
 gcatgcgtgg ctttaaccagc tacgccgaga cgggtgtcgg ctacggcacc gaagcgtat 6600
 ttaccgacgg cgatgatacg ccgtgggtcaa agcggttct cgcctcggcc tacgcctccc 6660
 gcgggttgaa aatgcgtac acctccggca ccggatccga agcgtgatg ggctattcgg 6720
 agagcaagtc gatgctctac ctggaatcgc gctgcatctt cattactaaa ggcccgggg 6780

 ttcagggact gcaaacggc gcggtgagct gtatcgcat gaccggcgt gtgccgtcgg 6840
 gcattcgggc ggtgctggcg gaaaacctga tcgcctctat gctcgacct gaagtggcgt 6900

ccgccaacga ccagactttc tcccactcgg atattcgccg caccgcgcgc accctgatgc 6960
 agatgctgcc gggcaccgac tttatcttct ccggctacag cgcggtgccg aactacgaca 7020
 acatgttcgc cggctcgaac ttcatgacgg aagattttga tgattacaac atcctgcagc 7080
 gtgacctgat ggttgacggc ggcctgcgtc cggtgaccga ggcggaaacc attgccattc 7140
 gccagaaagc ggcgcgggcg atccaggcgg ttttccgcga gctggggctg ccgccaatcg 7200

 ccgacgagga ggtggaggcc gccacctacg cgcacggcag caacgagatg ccgccgcgta 7260
 acgtggtgga ggatctgagt gcggtggaag agatgatgaa gcgcaacatc accggcctcg 7320
 atattgtcgg cgcgctgagc cgcagcggct ttgaggatat cgccagcaat attctcaata 7380
 tgctgcgcca gcggtcacc ggcgattacc tgcagacctc ggccattctc gatcggcagt 7440
 tcgagtggt gagtgcggtc aacgacatca atgactatca ggggccgggc accggctatc 7500
 gcatctctgc cgaacgctgg gcggagatca aaaatatcc gggcgtggtt cagcccgaca 7560
 ccattgaata aggcggtatt cctgtgcaac agacaacca aattcagccc tcttttacc 7620

 tgaaaacccg cgagggcggg gtagcttctg ccgatgaacg cgccgatgaa gtggtgatcg 7680
 gcgtcggccc tgccttcgat aaacaccagc atcacactct gatcgatatg ccccatggcg 7740
 cgatcctcaa agagctgatt gccggggtgg aagaagaggg gcttcacgcc cgggtggtgc 7800
 gcattctgcg cactgccgac gtctccttta tggcctggga tgcggccaac ctgagcggct 7860
 cggggatcgg catcggtatc cagtccaagg ggaccacggc catccatcag cgcgatctgc 7920
 tgccgctcag caacctggag ctgttctccc aggcgccgct gctgacgctg gagacctacc 7980
 ggcagattgg caaaaacgct gcgcgctatg cgcgcaaaga gtcaccttcg ccggtgccgg 8040

 tggatgaacga tcagatggtg cggccgaaat ttatggccaa agccgcgcta tttcatatca 8100
 aagagaccaa acatgtggtg caggacgccg agcccgctac cctgcacatc gacttagtaa 8160
 gggagtgacc atgagcgaga aaacctgcg cgtgcaggat tatccgttag ccaccgctg 8220
 cccggagcat atcctgacgc ctaccggcaa accattgacc gatattacc tcgagaaggt 8280
 gctctctggc gaggtgggccc cgcaggatgt gcggtatctc cgccagaccc ttgagtacca 8340
 ggcgagatt gccgagcaga tgcagcgcca tgcggtggcg cgcaatttcc gccgcgcggc 8400
 ggagcttacc gccattctg acgagcgcat tctggctatc tataacgcgc tgcgcccggt 8460

 ccgtctctcg caggcggagc tgctggcgat cgcgcagcag ctggagcaca cctggcatgc 8520
 gacagtgaat gccgcctttg tccgggagtc ggcggaagtg tatcagcagc ggcataagct 8580
 gcgtaaagga agctaagcgg aggtcagcat gccgttaata gccgggattg atatcgga 8640
 cgccaccacc gaggtggcgc tggcgtccga ctaccgcag gcgagggcgt ttgttgccag 8700
 cgggatcgtc gcgacgacgg gcatgaaagg gacgcgggac aatatcgccg ggaccctcg 8760

cgcgctggag caggccctgg cgaaaacacc gtggtcgatg agcgatgtct ctcgcatcta 8820
 tcttaacgaa gccgcgccgg tgattggcga tgtggcgatg gagaccatca ccgagaccat 8880

 tatcaccgaa tcgaccatga tcggtcataa cccgcagacg ccgggcgggg tgggcgttgg 8940
 cgtggggacg actatcgccc tcgggcggct ggcgacgctg ccggcgggcg agtatgccga 9000
 ggggtggatc gtactgattg acgacgccgt cgatttcctt gacgccgtgt ggtggctcaa 9060
 tgaggcgctc gaccggggga tcaacgttgt ggcgcgatc ctcaaaaagg acgacggcgt 9120
 gcttgtgaac aaccgcctgc gtaaaacct gccggttgtg gatgaagtga cgctgctgga 9180
 gcaggtcccc gagggggtaa tggcgccggt ggaagtggcc gcgccgggcc aggtggtgcg 9240
 gatcctgtcg aatccctacg ggatcgccac ctcttcggg ctaagcccg aagagacca 9300

 ggccatcgtc cccatcgccc gcgccctgat tggcaaccgt tccgcggtgg tgctcaagac 9360
 cccgcagggg gatgtgcagt cgcggtgat ccggcgggc aacctctaca ttagcgcgca 9420
 aaagcggcg ggagaggccg atgtcgccga ggcgcgga gccatcatgc aggcgatgag 9480
 cgctcgctc ccggtacgac acatccgagg cgaaccggc acccacgccc gcggcatgct 9540
 tgagcgggtg cgcaaggtaa tggcgctcct gaccggccat gagatgagcg cgatatacat 9600
 ccaggtatct ctggcggtg atacgtttat tccgcgcaag gtgcaggcg ggatggccg 9660
 cgagtgcgcc atggagaatg ccgtcgggat ggcgcgatg gtgaaagcg atcgctgca 9720

 aatgcaggtt atcgcccgcg aactgagcgc ccgactgcag accgaggtgg tgggtggcgg 9780
 cgtggaggcc aacatggcca tcgccggggc gttaccact cccggctgtg cggcgccgct 9840
 ggcatcctc gacctggcg ccggtcgac ggatggcg atcgtaac cgaggggca 9900
 gataacggcg gtccatctcg ccggggcggg gaatatggc agcctgttga ttaaaaccga 9960
 gctgggcctc gaggatcttt cgctggcgga agcgataaaa aaataccgc tggccaaagt 10020
 ggaaagcctg ttcagtattc gtcacgagaa tggcgcggtg gattcttcc gggaagccct 10080
 cagcccgcg gtgttcgcca aagtgggtga catcaaggag ggcgaaactg tgccgatcga 10140

 taacgccagc ccgctgga aaattcgtct cgtgcgccg caggcgaaag agaaagtgtt 10200
 tgtcaccaac tgctgcgcg cgctgcgcca ggtctaccc ggcggttcca ttcgcatat 10260
 cgcctttgtg gtgtgtgtg gcggtcatc gctggacttt gagatccgc agcttatcac 10320
 ggaagcctt tcgcactatg gcgtggtcgc cgggcagggc aatattcggg gaacagaagg 10380
 gccgcgaat gcggtcgcca ccgggtgct actggccggt caggcgaatt aaacggcgcg 10440
 tcgcgccagc ctctaggta aaataaaaaa ggcacgtcag atgacgtgcc tttttcttg 10500
 tctagcgtgc accaatgctt ctggcgtcag gcagccatcg gaagctgtgg tatggctgtg 10560

caggtcgtaa atcactgcat aattcgtgtc gctcaaggcg cactcccgtt ctggataatg 10620
 ttttttgcgc cgacatcata acggttcttg caaatattct gaaatgagct gttgacaatt 10680
 aatcatccgg ctcgataaat gtgtggaatt gtgagcggat aacaatttca cacaggaaac 10740
 agacatgac tagtaaggag gacaattcca tggctgctgc tgcigataga ttaaacttaa 10800
 cttccggcca cttgaatgct ggtagaaaga gaagttcctc ttctgtttct ttgaaggctg 10860
 ccgaaaagcc tticaaggtt actgtgattg gatctggtaa ctgggggtact actattgccca 10920
 aggtggttgc cgaaaattgt aagggatacc cagaagtttt cgctccaata gtacaaatgt 10980

gggtgttcga agaagagatc aatggtgaaa aattgactga aatcataaat actagacatc 11040
 aaaacgtgaa atacttgctt ggcatcactc tacccgacaa tttggttgct aatccagact 11100
 tgattgattc agtcaaggat gtcgacatca tcgttttcaa cattccacat caatttttgc 11160
 cccgtatctg tagccaattg aaaggatcatg ttgattcaca cgtcagagct atctcctgtc 11220
 taaagggttt tgaagttggt gctaaagggt tccaattgct atcctcttac atcactgagg 11280
 aactaggtat tcaatgtggt gctctatctg gtgctaacat tgccaccgaa gtcgctcaag 11340
 aacactggtc tgaacaaca gttgcttacc acattccaaa ggatttcaga ggcgagggca 11400

aggacgtcga ccataaggtt ctaaaggcct tgttccacag acctacttc cactttagtg 11460
 tcatcgaaga tgttgctggt atctccatct gtggtgcttt gaagaacgtt gttgccttag 11520
 gttgtggttt cgtcgaaggt ctaggctggg gtaacaacgc ttctgctgcc atccaaagag 11580
 tcggtttggg tgagatcatc agattcggtc aaatgttttt ccagaatct agagaagaaa 11640
 catactacca agagtctgct ggtgttgctg atttgatcac cacctgcgct ggtggtagaa 11700
 acgtcaaggt tgctaggcta atggctactt ctggtaagga cgctgggaa tgtgaaaagg 11760
 agttgttgaa tggccaatcc gctcaaggtt taattacctg caaagaagtt cacgaatggt 11820

tggaacatg tggtctgtc gaagacttcc cattatttga agccgtatac caaatcgttt 11880
 acaacaacta cccaatgaag aacctgccgg acatgattga agaattagat ctacatgaag 11940
 attagattta ttggatccag gaaacagact agaattatgg gattgactac taaacctcta 12000
 tctttgaaag ttaacgccgc ttgttctgac gtcgacggta ccattatcat ctctcaacca 12060
 gccattgctg cattctggag ggatttcggt aaggacaaac cttatttcga tgctgaacac 12120
 gttatccaag tctcgcatgg ttggagaacg ttgatgccca ttgctaagtt cgctccagac 12180
 ttgccaatg aagagtatgt taacaaattia gaagctgaaa ttccgggtcaa gtacggtgaa 12240

aaatccattg aagtcccagg tgcagttaag ctgtgcaacg ctttgaacgc tctacaaaa 12300
 gagaaatggg ctgtggcaac ttccggtacc cgtgatatgg cacaaaaatg gttcgagcat 12360
 ctgggaatca ggagacaaa gtacttcatt accgctaag atgtcaaaca gggtaacct 12420

catccagaac catatctgaa gggcaggaat ggcttaggat atccgatcaa tgagcaagac 12480
 ccttcctaat ctaaggtagt agtatttgaa gacgtccag caggtattgc cgccggaaaa 12540
 gccgccggtt gtaagatcat tggtagtgcc actactttcg acttggactt cctaaaggaa 12600
 aaaggctgtg acatcattgt caaaaaccac gaatccatca gagttggcgg ctacaatgcc 12660

gaaacagacg aagtigaatt ctttttgac gactacttat atgctaagga cgatctgttg 12720
 aaatggtaac ccgggctgca ggcatgcaag cttggctgtt ttggcggatg agagaagatt 12780
 ttcagcctga tacagattaa atcagaacgc agaagcggtc tgataaaaca gaatttgcct 12840
 ggccggcagta gcgcggtggt cccacctgac cccatgccga actcagaagt gaaacgccgt 12900
 agcgcgatg gtagtgtggg gtctcccat gcgagagtag ggaactgcca ggcatcaaat 12960
 aaaacgaaag gctcagtcga aagactgggc ctttcgtttt atctgttgtt tgcggtgaa 13020
 cgctctcctg agtaggacaa atccgccggg agcggatttg aacgttgca agcaacggcc 13080

cggagggtgg cgggcaggac gcccgcata aactgccagg catcaaatta agcagaaggc 13140
 catcctgacg gatggccttt ttgcgtttct acaaaactcca gctggatcgg gcgctagagt 13200
 atacatttaa atggtaccct ctagtcaagg ccttaagtga gtcgtattac ggactggccg 13260
 tcgttttaca acgtcgtgac tgggaaaacc ctggcggttac ccaacttaat cgccttgacg 13320
 cacatcccc tttcgccagc tggcgtaata gcgaagaggc ccgcaccgat cgcccttccc 13380
 aacagttgcg cagcctgaat ggcgatggc gcctgatgcg gtattttctc cttacgcac 13440
 tgtgcggtat ttacaccgc atatggtgca ctctcagtac aatctgctct gatgccgat 13500

agttaagcca gccccgacac ccgccaacac ccgctgacga gct 13543

<210> 75

<211> 13402

<212> DNA

<213> Artificial Sequence

<220><223> Plamid

<400> 75

tagtaaagcc ctcgctagat tttaatgcgg atgttgcat tacttcgcca actattgcga 60
 taacaagaaa aagccagcct ttcatgatat atctcccaat ttgtgtaggg cttattatgc 120
 acgttataaa ataataaaag cagacttgac ctgatatgtt ggctgtgagc aattatgtgc 180
 ttagtgcatc taacgcttga gttaagccgc gccgcgaagc ggctgcggct tgaacgaatt 240
 gttagacatt atttgccgac taccttggtg atctcgctt tcacgtagt gacaaattct 300
 tccaactgat ctgcgcgca ggccaagcga tcttcttctt gtccaagata agcctgtcta 360

gcttcaagta tgacgggctg atactgggcc ggcaggcgct ccattgccca gtcggcagcg 420
 acatccttcg gcgcgatttt gccggttact gcgctgtacc aaatgcggga caacgtaagc 480
 actacatttc gctcatcgcc agcccagtcg ggccggcaggt tccatagcgt taaggtttca 540
 tttagcgctt caaatagatc ctgttcagga accggatcaa agagttcctc cgccgctgga 600
 cctaccaagg caacgctatg ttctcttgct ttgttcagca agatagccag atcaatgtcg 660
 atcgtggctg gctcgaagat acctgcaaga atgtcattgc gctgccattc tccaaattgc 720

agttcgcgct tagctggata acgccacgga atgatgtcgt cgtgcacaac aatggtgact 780
 tctacagcgc ggagaatctc gctctctcca ggggaagccg aagtttccaa aaggtcgttg 840
 atcaaagctc gccgcgttgt ttcataagc cttacggta ccgtaaccag caaatcaata 900
 tcaactgttg gcttcaggcc gccatccact gcggagccgt acaaatgtac ggccagcaac 960
 gtcggttcga gatggcgctc gatgacgcca actacctctg atagttgagt cgatacttcg 1020
 gcgatcaccg ctccctcat gatgtttaac ttgttttag ggcgactgcc ctgctgcgta 1080
 acatcgttgc tgcctcataa catcaaacat cgacccacgg cgtaacgcgc ttgctgcttg 1140

gatgcccag gcatagactg taccctaaaa aaacagtcac aacaagccat gaaaaccgcc 1200
 actgcgcgct taccaccgct gcgttcggtc aaggttctgg accagttgcg tgagcgcata 1260
 cgctacttgc attacagctt acgaaccgaa caggcttatg tccactgggt tcgtgccttc 1320
 atccgtttcc acggtgtgcg taccaccgca accttgggca gcagcgaagt cgaggcattt 1380
 ctgtcctggc tggcgaacga gcgcaagggt tcggcttcca cgcatcgtca ggcatcggcg 1440
 gccttctgtg tcttctacga caaggtgtcg tgcacggatc tgccctggct tcaggagatc 1500
 ggaagacctc ggccgtcgcg gcgcttgccg gtggtgtctga ccccgatga agtggttcgc 1560

atcctcgggt ttctggaagg cgagcatcgt ttgttcgccc agcttctgta tggaaacgggc 1620
 atgcggatca gtgagggttt gcaactgcgg gtcaaggatc tggatttcga tcacggcacg 1680
 atcatcgtgc gggagggcaa gggctccaag gatcgggcct tgatgttacc cgagagcttg 1740
 gcaccagcc tgcgcgagca ggggaattaa ttcccacggg ttttctgcc cgcaaacggg 1800
 ctgttctggt gttgctagtt tggtatcaga atcgagatc cggttcagc cggtttgccg 1860
 gctgaaagcg ctatttcttc cagaattgcc atgatttttt cccacggga ggctcactg 1920
 gctcccgtgt tgtcggcagc ttgtattcga taagcagcat cgcctgttcc aggtgtctta 1980

tgtgtgactg ttgagctgta acaagttgtc tcagggtgtc aatttcattg tctagttgct 2040
 ttgttttact ggtttcacct gttctattag gtgttacatg ctgttcattc gttacattgt 2100
 cgatctgttc atggtgaaca gctttgaatg caccaaaaac tcgtaaaagc tctgatgtat 2160
 ctatcttttt tacaccgttt tcatctgtgc atatggacag tttcccttt gatatgtaac 2220

ggatgaacagt tgttctactt ttgtttgtta gtcttgatgc ttactgata gatacaagag 2280
 ccataagaac ctcagatcct tccgtattta gccagtatgt tctctagtgt gggttcgttgt 2340
 ttttgcgtga gccatgagaa cgaaccattg agatcatact tactttgcat gtcactcaaa 2400

 aattttgcct caaaactggg gagctgaatt ttgacagtta aagcatcgtg tagtggtttt 2460
 cttagtccgt tatgtaggta ggaatctgat gtaatggttg ttggtatttt gtcaccattc 2520
 atttttatct gggtgttctc aagttcgggt acgagatcca ttgtctatc tagttcaact 2580
 tggaaaatca acgtatcagt cgggcggcct cgcttatcaa ccaccaattt catattgctg 2640
 taagtgttta aatctttact tattggtttc aaaaccatt gggttaagcct tttaaactca 2700
 tggtagttat ttcaagcat taacatgaac ttaaattcat caaggctaatt ctctatattt 2760
 gccttgtagg ttttcttttg tgtagttct ttaataaacc actcataaat cctcatagag 2820

 tatttggttt caaaagactt aacatgttcc agattatatt ttaagaattt ttttaactgg 2880
 aaaagataag gcaatatctc ttactataaa actaattcta attttctgct tgagaacttg 2940
 gcatagtttg tccactggaa aatctcaaag ctttaacca aaggattcct gatttccaca 3000
 gttctcgtca tcagctctct ggttgcttta gctaatacac cataagcatt ttcctactg 3060
 atgttcatca tctgagcgtt ttggttataa gtgaacgata ccgtccgttc tttccttgta 3120
 ggggtttcaa tcgtgggggt gagtagtgcc acacagcata aaattagctt ggtttcatgc 3180
 tccgttaagt catagcgact aatcgctagt tcatttgctt tgaaaacaac taattcagac 3240

 atacatctca attggtctag gtgattttta tcactatacc aattgagatg ggctagtcaa 3300
 tgataattac tagtcccttt ctttgagtt gtgggtatct gtaaattctg ctagaccttt 3360
 gctggaaaac ttgtaaattc tgctagacc cctgtaaatt ccgctagacc ttgtgtgttt 3420
 tttttgttt atattcaagt ggttataatt tatagaataa agaaagaata aaaaaagata 3480
 aaaagaatag atcccgccc tgtgtataac tcactacttt agtcagttcc gcagtattac 3540
 aaaaggatgt cgaaacgct gtttctcct ctacaaaaca gaccttaaaa ccctaaaggc 3600
 ttaagtagca cctcgcgaag ctggggcaaa tcgctgaata ttccttttgt ctccgaccat 3660

 caggcacctg agtcgtgtc ttttctgta cattcagttc gctgcgtca cggctctggc 3720
 agtgaatggg ggtaaatggc actacaggcg ctttttatgg attcatgcaa ggaaactacc 3780
 cataataca gaaaagcccg tcacgggctt ctacggcgt tttatggcgg gtctgtatg 3840
 tgggtctatc tgactttttg ctgttcagca gttcctgccc tctgattttc cagtctgacc 3900
 acttcggatt atcccgtag aggtcattca gactggctaa tgcacccagt aaggcagcgg 3960
 tatcatcaac aggtttacc gtcttactgt cgggaattca tttaatagt caaaagcctc 4020
 cgaccggagg cttttgactg ctaggcgatc tgtgctgttt gccacggtat gcagcaccag 4080

cgcgagatta tgggctcgca cgctcgactg tcggacgggg gcactggaac gagaagtcag 4140
 gcgagccgtc acgcccttga caatgccaca tcctgagcaa ataattcaac cactaaacaa 4200
 atcaaccgcg tttcccgag gtaaccaagc ttgcgggaga gaatgatgaa caagagccaa 4260
 caagttcaga caatcacctt ggccgccgcc cagcaaatgg cggcggcggt ggaaaaaaa 4320
 gccactgaga tcaacgtggc ggtggtgttt tccgtagtgt accgcggagg caacacgtg 4380
 ctatccagc ggatggacga ggccttcgtc tccagctgcg atatttcct gaataaagcc 4440
 tggagcgct gcagcctgaa gcaaggtacc catgaaatta cgtcagcggt ccagccagga 4500

 caatctctgt acggtctgca gctaaccaac caacagcgaa ttattatttt tggcggcggc 4560
 ctgccagtta ttttaatga gcagtaatt ggcccgctcg gcgttagcgg cggtacggtc 4620
 gagcaggatc aattattagc ccagtgcgcc ctggattgtt tttccgcat ataacctgaa 4680
 gcgagaaggt atattatgag ctatcgatg ttcgccagg catctgagt gttaacgagg 4740
 ggaccgtcat gtgctttca ccgccaggcg tacgctgtt ttacgatccg cgcgggcacc 4800
 atgccggcg catcaatgag ctgtgctggg ggctggagga gcagggggtc ccctgccaga 4860
 ccataaccta tgacggaggc ggtgacgccg ctgctgtggg cgccctggcg gccagaagct 4920

 cgccctgcg ggtgggtatc gggctcagcg cgtccggcga gatagccctc actcatgcc 4980
 agctgccggc ggacgcgccg ctggctaccg gacacgtcac cगतagcgac gatcaactgc 5040
 gtacgtcgg cgccaacgcc gggcagctgg ttaaagtcct gccgttaagt gagagaaact 5100
 gaatgtatcg tatctatacc cgcaccgggg ataaaggcac caccgccctg tacggcggca 5160
 gccgcatcg gaaagacct attcgcgtcg aggcctacgg caccgtcgat gaactgatat 5220
 ccagctggg cgtctgtac gccacgacc gcgacgccgg gctgcgggaa agcctgcacc 5280
 atattcagca gacgtgttc gtgctggggg ctgaactggc cagcgatcg cggggcctga 5340

 cccgcctgag ccagacgac ggcaagagg agatcacgc cctggagcgg ctatcgacc 5400
 gcaatatggc cgagagcggc ccgttaaaac agttcgtgat cccggggagg aatctgcct 5460
 ctgccagct gcacctgat gcttgcgtt gaactggcct agcaaacaca gaaaaagcc 5520
 cgcacctgac agtgcgggt tttttttcc taggcgatct gtgctgttg ccacggtatg 5580
 cagcaccagc gcgagattat gggctcgac gctcgactgt cggacggggg cactggaacg 5640
 agaagtcagg cgagccgtca cgccctgac aatgccacat cctgagcaaa taattcaacc 5700
 actaaacaaa tcaaccgct ttcgggagg taaccaagct tcacctttg agccgatgaa 5760

 caatgaaaag atcaaacga ttgcagtac tggccagcg cccgtcaat caggacgggc 5820
 tgattggcga gtggcctgaa gaggggctga tcgcatgga cagcccttt gaccgggtct 5880

cttcagtaaa agtggacaac ggtctgatcg tcgaactgga cggcaaacgc cgggaccagt 5940
 ttgacatgat cgaccgattt atcgccgatt acgcgatcaa cgttgagcgc acagagcagg 6000
 caatgcgcct ggaggcgggtg gaaatagccc gtatgctggt ggatattcac gtcagccggg 6060
 aggagatcat tgccatcact accgccatca cgccggccaa agcgggtcgag gtgatggcgc 6120
 agatgaacgt ggtggagatg atgatggcgc tgcagaagat gcgtgcccgc cggaccccct 6180

ccaaccagtg ccacgtcacc aatctcaaag ataatccggt gcagattgcc gctgacgccg 6240
 ccgaggccgg gatccgcggc ttctcagaac aggagaccac ggtcggatc gcgcgctacg 6300
 cgccgtttaa cgccctggcg ctgttggtcg gttcgcagtg cggccgcccc ggcgtgttga 6360
 cgcagtgctc ggtggaagag gccaccgagc tggagctggg catgcgtggc ttaaccagct 6420
 acgccgagac ggtgtcggtc tacggcaccg aagcggattt taccgacggc gatgatacgc 6480
 cgtggtaaaa ggcgttcctc gcctcggcct acgcctcccg cgggttgaaa atgcgctaca 6540
 cctccggcac cggatccgaa gcgctgatgg gctattcgga gagcaagtgc atgctctacc 6600

tcgaatcgcg ctgcatcttc attactaaag gcgccggggt tcagggactg caaacggcg 6660
 cggtagagctg tatcgcatg accggcgctg tgccgtcggg cattcgggcg gtgctggcgg 6720
 aaaacctgat cgctctatg ctcgacctcg aagtggcgtc cgccaacgac cagactttct 6780
 cccactcgga tattcgccgc accgcgcgca cctgatgca gatgctgccg ggcaccgact 6840
 ttattttctc cgctacagc gcggtgccga actacgaaa catgttcgcc ggctcgaact 6900
 tcgatcgga agattttgat gattacaaca tctgcagcg tgacctgatg gttgacggcg 6960
 gcctcgctcc ggtgaccgag gcggaacca ttgccattcg ccagaaagcg gcgcgggcga 7020

tccaggcgggt ttccgcgag ctggggctgc cgccaatgc cgacgaggag gtggaggccg 7080
 ccacctacgc gcacggcagc aacgagatgc cgccgcgtaa cgtggtggag gatctgagtg 7140
 cgggtggaaga gatgatgaag gcgaacatca ccggcctcga tattgtcggc gcgtgagcc 7200
 gcagcggctt tgaggatc gccaagaata ttctcaatat gctgcgccag cgggtcaccg 7260
 gcgattacct gcagacctcg gccattctcg atcggcagtt cgaggtggtg agtgcggtca 7320
 acgacatcaa tgactatcag gggccgggca ccggctatcg catctctgcc gaacgtggg 7380
 cggagatcaa aaatattccg ggcgtgggtc agcccacac cattgaataa ggcggtattc 7440

ctgtgcaaca gacaaccaa attcagccct cttttaccct gaaaaccgc gagggcgggg 7500
 tagctttctg cgatgaacgc gccgatgaag tggatgatcg cgtcggccct gccttcgata 7560
 aacaccagca tcacactctg atcgatatgc ccatggcgc gatcctcaa gagctgattg 7620
 ccggggtgga agaagagggg cttcacgcc gggtgggtcg cattctgcgc acgtccgacg 7680
 tctcctttat ggcttgggat gcggccaacc tgagcggtc ggggatcggc atcggtatcc 7740

agtcgaaggg gaccacggtc atccatcagc gcgatctgct gccgctcagc aacctggagc 7800
 tgtttctccca ggccgcccgtg ctgacgctgg agacctaccg gcagattggc aaaaacgctg 7860

 cgcgctatgc gcgcaaagag tcaccttcgc cggcgccggt ggtgaacgat cagatggtgc 7920
 ggccgaaatt tatggccaaa gccgcgctat ttcatatcaa agagacccaaa catgtggtgc 7980
 aggacgccga gcccgctacc ctgcacatcg acttagtaag ggagtgaaca tgagcgagaa 8040
 aacctgctgc gtgcaggatt atccgttagc caccgctgc ccggagcata tcctgacgcc 8100
 taccggcaaa ccattgaccg atattaccct cgagaagggt ctctctggcg aggtgggccc 8160
 gcaggatgtg cggatctccc gccagacct tgagtaccag gcgcagattg ccgagcagat 8220
 gcagcgccat gcggtggcgc gcaatttcg ccgcgcggcg gagcttatcg ccattcctga 8280

 cgagcgcatt ctggctatct ataacgcgt gcgcccgttc cgctcctcgc aggcggagct 8340
 gctggcgatc gccgacgagc tggagcacac ctggcatgcg acagtgaatg ccgcctttgt 8400
 ccgggagtcg gcggaagtgt atcagcagcg gcataagctg cgtaaaggaa gctaagcgga 8460
 ggtcagcatg ccgttaatag ccgggattga tatcggaac gccaccaccg aggtggcgct 8520
 ggcgctccgac taccgcagg cgagggcggt tgttgccagc gggatcgctg cgacgacggg 8580
 catgaaaggg acgcgggaca atatcgccgg gacctcgcc gcgctggagc aggcctggc 8640
 gaaaacaccg tggtcgatga gcgatgtctc tcgcatctat cttaacgaag ccgcgcgggt 8700

 gattggcgat gtggcgatgg agaccatcac cgagaccatt atcaccgaat cgacatgat 8760
 cggtcataac ccgcagacgc cgggcggggt ggcggttggc gtggggacga ctatcgccct 8820
 cgggcggctg gcgacgctgc cggcgccgca gtatgccgag ggggtgatcg tactgattga 8880
 cgacgccgtc gatttccctg acgcctgtg gtggctcaat gagcgctcg accgggggat 8940
 caacgtgggt gcggcgatcc tcaaaaagga cgacggcgtg ctggtgaaca accgcctgcg 9000
 taaaaccctg ccggtggtgg atgaagtac gctgctggag caggtccccg agggggtaat 9060
 ggcggcgggt gaagtggcgc cgccgggcca ggtggtgcgg atcctgtcga atccctacgg 9120

 gatcgccacc ttcttcgggc taagcccga agagaccag gccatcgtcc ccatcgccc 9180
 cgccctgatt ggcaaccgtt ccgcggtggt gctcaagacc ccgcaggggg atgtgcagtc 9240
 gcgggtgatc ccggcgggca acctctacat tagcggcgaa aagcgccgcg gagaggccga 9300
 tgtcgccgag ggcgcggaag ccatcatgca ggcgatgagc gcctgcgtc cggtacgca 9360
 catccgggc gaaccgggca ccacgccgg cggcatgctt gagcgggtgc gcaagtaat 9420
 ggcgctccctg accggccatg agatgagcgc gatatacatc caggatctgc tggcggtgga 9480
 tacgtttatt ccgcgaagg tgcaggcgcg gatggccggc gagtgcgcca tggagaatgc 9540

cgtcgggatg gcggcgatgg tgaaagcgga tcgtctgcaa atgcaggtta tcgcccgcga 9600
 actgagcgcc cgactgcaga ccgaggtggt ggtgggcggc gtggaggcca acatggccat 9660
 cgccggggcg ttaaccactc ccggctgtgc ggccgctg gcgacctcg acctcggcgc 9720
 cggctcgacg gatgcggcga tcgtcaacgc ggaggggcag ataacggcgg tccatctcgc 9780
 cggggcgggg aatatggta gcctgttgat taaaaccgag ctgggcctcg aggatctttc 9840
 gctggcggaa gcgataaaaa aatacccgct ggccaaagt gaaagcctgt tcagtattcg 9900
 tcacgagaat ggcgcggtgg agttctttcg ggaagccctc agcccggcgg tgttcgcaa 9960

 agtgggtgac atcaaggagg gcgaactggt gccgatcgat aacgccagcc cgctggaaaa 10020
 aattcgtctc gtgcgccggc aggcgaaaga gaaagtgttt gtcaccaact gcctgcgcgc 10080
 gctgcgccag gtctcaccgc gcggttccat tcgcgatatc gcctttgtgg tctgtgtggg 10140
 cggctcatcg ctggactttg agatcccga gcttatcacg gaagccttgt cgcactatgg 10200
 cgtggtcgcc ggcgaggcca atattcgggg aacagaaggg ccgcgcaatg cggtcgccac 10260
 cgggctgcta ctggccggtc aggcgaatta aacgggcgct cgcccagcc tctaggtaca 10320
 aataaaaaag gcacgtcaga tgacgtgcct tttttcttgt ctacgtgca ccaatgcttc 10380

 tggcgtcagg cagccatcgg aagctgtggt atggctgtgc aggtcgtaaa tctactgcata 10440
 attcgtgtcg ctcaaggcgc actcccgctt tggataatgt tttttgcgc gacatcataa 10500
 cggttctggc aaatattctg aaatgagctg ttgacaatta atcatccggc tcgtataatg 10560
 tgtggaattg tgagcggata acaatttcac acaggaaaca gaccatgact agtaaggagg 10620
 acaattccat ggctgtgtct gctgatagat taaacttaac ttccggccac ttgaatgctg 10680
 gtagaaagag aagtccctct tctgtttctt tgaaggctgc cgaaaagcct ttcaaggtta 10740
 ctgtgattgg atctggtaac tggggtacta ctattgcaa ggtggttgcc gaaaattgta 10800

 aggatacc accaagtcttc gctccaatag tacaatgtg ggtgttcgaa gaagagatca 10860
 atggtgaaaa attgactgaa atcataaata ctagacatca aaacgtgaaa tacttgccctg 10920
 gcatcactct acccgacaat ttggttgcta atccagactt gattgattca gtcaaggatg 10980
 tcgacatcat cgttttcaac attccacatc aatttttgcc ccgtatctgt agccaattga 11040
 aaggatcatg tgattcacac gtcagagcta tctcctgtct aaagggtttt gaagtgggtg 11100
 ctaaagggtg ccaattgcta tctctttaca tctactagga actaggtatt caatgtggtg 11160
 ctctatctgg tgctaacatt gccaccgaag tcgctcaaga aactggtct gaaacaacag 11220

 ttgcttacca cattccaaag gatttcagag gcgagggcaa ggacgtcgac cataaggttc 11280
 taaaggcctt gttccacaga cttacttcc acgttagtgt catcgaagat gttgctggtg 11340
 tctccatctg tgggtgcttg aagaacgttg ttgccttagg ttgtggttc gtcgaaggtc 11400

taggctgggg taacaacgct tctgctgcca tccaaagagt cggtttgggt gagatcatca 11460
gattcggta aatgtttttt ccagaatcta gagaagaaac atactaccaa gagtctgctg 11520
gtgttgctga tttgatcacc acctgcgctg gtggtagaaa cgtcaagggt gctaggctaa 11580
tggctacttc tggttaaggac gcctgggaat gtgaaaagga gttgttgaat ggccaatccg 11640

ctcaaggttt aattacctgc aaagaagttc acgaatgggt ggaaacatgt ggctctgtcg 11700
aagacttccc attatttgaa gccgtatacc aaatcggtta caacaactac ccaatgaaga 11760
acctgccgga catgattgaa gaattagatc tacatgaaga ttagatttat tggatccagg 11820
aaacagacta gaattatggg attgactact aaacctctat ctttgaaagt taacgccgct 11880
ttgttcgacg tcgacggta c cattatcatc tctcaaccag ccattgctgc attctggagg 11940
gatttcggta aggacaaacc ttatttcgat gctgaacacg ttatccaagt ctgcatggt 12000
tggagaacgt ttgatgcat tgctaagttc gctccagact ttgccaatga agagtatgtt 12060

aacaaattag aagctgaaat tccggtcaag tacggtgaaa aatccattga agtcccagg 12120
gcagttaagc tgtgcaacgc tttgaacgct ctacaaaaag agaaatgggc tgtggcaact 12180
tccggtaccc gtgatatggc acaaaaatgg ttcgagcatc tgggaatcag gagaccaaag 12240
tacttcatta ccgctaata tgtaaacag ggtaagcctc atccagaacc atatctgaag 12300
ggcaggaatg gcttaggata tccgatcaat gagcaagacc cttccaaatc taaggtagta 12360
gtatttgaag acgctccagc aggtattgcc gccggaaaag ccgccggttg taagatcatt 12420
ggtattgcca ctactttcga cttggacttc ctaaaggaaa aaggctgtga catcattgtc 12480

aaaaaccacg aatccatcag agttggcggc tacaatgccg aaacagacga agttgaattc 12540
atttttgacg actacttata tgctaaggac gatctgttga aatggtaacc cgggctgcag 12600
gcatgcaagc ttggctgttt tggcggatga gagaagattt tcagcctgat acagattaaa 12660
tcagaacgca gaagcggctc gataaaacag aatttgctg gcggcagtag cgcggtggtc 12720
ccacctgacc ccatgccgaa ctgagaagtg aaacgccgta gcgccgatgg tagtgtgggg 12780
tctcccatg cgagagtagg gaactgccag gcatcaata aaacgaaagg ctgagtcgaa 12840
agactgggcc tttcgtttta tctgttgttt gtcggtgaac gctctctga gtaggacaaa 12900

tccgccggga gcgatttga acgttgcgaa gcaacggccc ggagggtggc gggcaggacg 12960
cccgcataa actgccagc atcaaatata gcagaaggcc atcctgacgg atggcctttt 13020
tgcgtttcta caaactccag ctggatcggg cgctagagta tacatttaaa tggtagcctc 13080
tagtcaagc ctttaagtga tcgtattacg gactggccgt cgttttaca cgtcgtgact 13140
gggaaaaccc tggcgttacc caacttaatc gccttgacg acatccccct ttcgccagct 13200
ggcgtaatat cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg 13260

gcgaatggcg cctgatgcgg tattttctcc ttacgcatct gtgcggtatt tcacaccgca 13320

tatgggtgcac tctcagtaca atctgctctg atgccgcata gtttaagccag ccccgacacc 13380

cgccaacacc cgctgacgag ct 13402

<210> 76

<211> 14443

<212> DNA

<213> Artificial Sequence

<220><223> Plasmid

<400> 76

ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgtggga 60

attaattccc ctgctcgcgc aggctgggtg ccaagctctc gggtaacatc aaggcccgat 120

ccttgagacc cttcttacag agatgaaaaa caaacccgca cgccaggcgg catcgcggtc 180

tcagagatat gtttacgtag atcgaagagc accggtgttt aaacgccttt gacgatgcca 240

catcctgagc aaataattca accactaac aaatcaaccg cgtttcccgg aggtaaccga 300

gctcatgac ctgtgttgtg gtgaagccct gatcgacatg ctgccccggc agacgacgt 360

gggtgaggcg ggctttgccc cttacgcagg cggagcggtc ttcaacacgg caattgcgct 420

ggggcgctct ggcgctccct cagccttttt taccggtctt tccgacgaca tgatgggcca 480

tatcctgagg gagacctgc gggccagcaa ggtggatttc agctattgcg ccacctgtc 540

gcgccccacc accattgcgt tcgttaagct ggttgatggc catgcgacct acgcttttta 600

cgacgagaac accgccggcc ggatgatcac cgaggccgaa cttccggcct tgggagcgga 660

ttgcgaagcg ctgcatttcg gcgccatcag ccttattccc gaaccctgcg gcagcaccta 720

tgaggcgctg atgacgcgag agcatgagac ccgcgtcacc tcgctcgatc cgaacattcg 780

tcccggcttc atccagaaca agcagtcgca catggcccg atccgccgca tggcggcgat 840

gtctgacatc gtcaagtctt cggatgagga cctggcgtgg ttcggtctgg aaggcgacga 900

ggacacgctt gcccgccact gggtgcacca cgttgcaaaa ctgctcgttg tcacctgtgg 960

cgccaagggg gccgtgggtt acagcgccaa tctcaagggt gaagtggcct ccgagcgct 1020

cgaagtggtc gatacggtcg gcgccggcga tacgttcgat gccggcattc ttgcttcgct 1080

gaaaatgcag ggctgtctga ccaaagcgca ggtggcttcg ctgagcgaag agcagatcag 1140

aaaagctttg gcgcttgagg cgaaagccgc tgcggtcact gtctcgcggg ctggcgcaaa 1200

tccgcctttc gcgcatgaaa tcggtttgtg attaatataa gcacgcagtc aaacaaaaaa 1260

cccgcgccat tgcgcgggtt tttttatgcc cgaaggcgcg ccagcacgca gtcaaaaaa 1320

aaaccgcgc cattgcgcgg gtttttttat gcccgaacgg ccgaggtctt ccgatctcct 1380
 gaagccaggg agatccgtg cacagcacct tgccgtagaa gaacagcaag gccgccaatg 1440
 cctgacgatg cgtggagacc gaaaccttgc gctcgttcgc cagccaggac agaaatgcct 1500

 cgacttcgct gctgccaag gttgccgggt gacgcacacc gtggaaacgg atgaaggcac 1560
 gaaccagtg gacataagcc tgttcgggtc gtaagctgta atgcaagtag cgtatgcgct 1620
 cacgcaactg gtccagaacc ttgaccgaac gcagcgggtg taacggcgca gtggcggttt 1680
 tcatggcttg ttatgactgt ttttttgggg tacagtctat gcctcgggca tccaagcagc 1740
 aagcgcgtta cgccgtgggt cgatgtttga tgttatggag cagcaacgat gttacgcagc 1800
 agggcagtcg ccctaaaaca aagttaaaca tcatgaggga agcgggtgatc gccgaagtat 1860
 cgactcaact atcagaggta gttggcgta tgcagcgcca tctcgaaccg acgttgctgg 1920

 ccgtacattt gtacggctcc gcagtggatg gcggcctgaa gccacacagt gatattgatt 1980
 tgctgggttac ggtgaccgta aggccttgatg aaacaacgcg gcgagctttg atcaacgacc 2040
 ttttgaaac ttcggttcc cctggagaga gcgagattct ccgcgctgta gaagtcacca 2100
 ttgttgtgca cgacgacatc attccgtggc gttatccagc taagcgcgaa ctgcaatttg 2160
 gagaatggca gcgcaatgac attcttgag gtatcttcga gccagccacg atcgacattg 2220
 atctggctat ctgtctgaca aaagcaagag aacatagcgt tgccttggtg ggtccagcgg 2280
 cggaggaact cttgatccg gttcctgaac aggatctatt tgaggcgcta aatgaaacct 2340

 taacgctatg gaactcgccg cccgactggg ctggcgatga gcgaaatgta gtgcttacgt 2400
 tgtccgcat ttgttacagc gcagtaaccg gcaaaatcgc gccgaaggat gtcgctgccg 2460
 actgggcaat ggagcgctg ccggcccagt atcagcccgt catacttgaa gctagacagg 2520
 cttatcttgg acaagaagaa gatcgcttgg cctcgcgcg agatcagttg gaagaatttg 2580
 tccactacgt gaaaggcgag atcaccaagg tagtcggcaa ataatgtcta acaattcggt 2640
 caagccgacg ccgcttcgag gcgcggctta actcaagcgt tagatgcact aagcacataa 2700
 ttgtcacag ccaaactatc aggtcaagtc tgcttttatt attttaagc gtgcataata 2760

 agccctacac aaattgggag atatatcatg aaaggctggc ttttcttgt tatcgcaata 2820
 gttggcgaag taatcgcaac atccgcatta aaatctagcg agggctttac taagctcgtc 2880
 agcgggtgtt ggcgggtgtc ggggctggct taactatgag gcatcagagc agattgtact 2940
 gagagtgcac catatgcggt gtgaaatacc gcacagatgc gtaaggagaa aataccgcat 3000
 caggcgccat tcgccattca ggctgcgcaa ctgttgggaa gggcgatcgg tgcgggcctc 3060
 ttcgctatta cgccagctgg cgaaaggggg atgtgctgca aggcgattaa gttgggtaac 3120
 gccagggttt tccagtcac gacgttgtaa aacgacggcc agtccgtaat acgactcact 3180

taaggccttg actagagggt accatttaaa tgtatactct agcgcccgat ccagctggag 3240
 tttgtagaaa cgcaaaaagg ccatccgtca ggatggcctt ctgcttaatt tgatgcctgg 3300
 cagtttatgg cgggcgtcct gcccgccacc ctccgggccg ttgcttcgca acgttcaaat 3360
 ccgctcccgg cggatttgc ctactcagga gagcgttcac cgacaaaca cagataaac 3420
 gaaaggccca gtctttcgac tgagcctttc gttttatttg atgcctggca gtccctact 3480
 ctcgcatggg gagacccac actaccatcg gcgctacggc gtttcacttc tgagttcggc 3540
 atggggtcag gtgggaccac cgcgtactg ccgccaggca aattctgttt tatcagaccg 3600

 cttctcggtt ctgatttaat ctgtatcagg ctgaaaatct tctctcatcc gccaaaacag 3660
 ccaagcttgc atgcctgcag cccgggttac catttcaaca gatcgtcctt agcatataag 3720
 tagtcgtcaa aaatgaattc aacttcgtct gtttcggcat tgtagccgcc aactctgatg 3780
 gattcgtggt ttttgacaat gatgtcacag ctttttctt ttaggaagtc caagtcgaaa 3840
 gtatggcga taccaatgat cttacaaccg gcggcttttc cggcggcaat acctgctgga 3900
 gcgtcttcaa atactactac cttagatttg gaagggtctt gctcattgat cggatatact 3960
 aagccattcc tgcccttcag atatggttct ggatgaggct taccctgttt gacatcata 4020

 gcgtaaatga agtactttgg tctctgatt ccagatgct cgaaccattt ttgtgccata 4080
 tcacgggtac cggaagtgc cacagcccat ttctcttttg gtagagcgtt caaagcgttg 4140
 cacagcttaa ctgcacctgg gacttcaatg gatttttcac cgtacttgac cggaatttca 4200
 gcttctaatt tgtaacata ctcttcattg gcaaagtcg gagcgaactt agcaatggca 4260
 tcaaagcttc tccaaccat cgagacttgg ataacgtgtt cagcatcgaa ataaggtttg 4320
 tccttaccga aatccctcca gaatgcagca atggctgggt gagagatgat aatggtaccg 4380
 tcgacgtcga acaaagcggc gtttaactttc aaagatagag gtttagtagt caatccata 4440

 attctagtct gtttctgga tccaataaat ctaatcttca tgtagatcta attcttcaat 4500
 catgtccggc aggttcttca ttgggtagt gttgtaaagc atttggtata cggttcaaa 4560
 taatgggaag tcttcgacag agccacatgt ttccaaccat tcgtgaactt ctttgcaggt 4620
 aattaaacct tgagcggatt ggccattcaa caactccttt tcacattccc aggcgtcctt 4680
 accagaagta gccattagcc tagcaacctt gacgtttcta ccaccagcg aggtggtgat 4740
 caaatcagca acaccagcag actcttggtg gtatgtttct tctctagatt ctgggaaaaa 4800
 catttgaccg aatctgatga tctacccaa accgactctt tggatggcag cagaagcgtt 4860

 gttacccag cctagacctt cgacgaaacc acaacctaag gcaacaacgt tcttcaaagc 4920
 accacagatg gagataccag caacatcttc gatgacacta acgtggaagt aaggtctgtg 4980

gaacaaggcc tttagaacct tatggctgac gtccttgccc tcgcctctga aatcctttgg 5040
aatgtggtaa gcaactgttg tttcagacca gtgttcttga gcgacttcgg tggcaatgtt 5100
agcaccagat agagcaccac attgaatacc tagttcctca gtgatgtaag aggatagcaa 5160
ttggacacct ttagcaccaa cttcaaaacc ctttagacag gagatagctc tgacgtgtga 5220
atcaacatga cctttcaatt ggctacagat acggggcaaa aattgatgtg gaatgttgaa 5280

aacgatgatg tcgacatcct tgactgaatc aatcaagtct ggattagcaa ccaaattgtc 5340
gggtagagtg atgccaggca agtatcttcc gttttgatgt ctagtattta tgatttcagt 5400
caatctttca ccattgatct cttcttcgaa caccacatt tgtactattg gagcgaaaac 5460
ttctgggtat cccttacaat tttcggcaac caccttggca atagtagtac ccagttacc 5520
agatccaatc acagtaacct tgaaaggctt ttcggcagcc ttcaaagaaa cagaagagga 5580
acttctcttt ctaccagcat tcaagtggcc ggaagttaag tttaatctat cagcagcagc 5640
agccatggaa ttgtcctcct tactagtcac ggtctgtttc ctgtgtgaaa ttgttatccg 5700

ctcacaattc cacacattat acgagccgga tgattaattg tcaacagctc atttcagaat 5760
atttgccaga accgttatga tgtcggcgca aaaaacatta tccagaacgg gagtgcgcct 5820
tgagcgacac gaattatgca gtgatttacg acctgcacag ccataccaca gcttccgatg 5880
gtgccttgac gccagaagca ttgggtgcacg ctgacaaga aaaaaggcac gtcacttgac 5940
gtgccttttt tatttgtacc tagaggctgg cgcgagcgcc cgtttaattc gcctgaccgg 6000
ccagtagcag cccggtggcg accgcattgc gcggcccttc tgttccccga atattgccct 6060
gccccgagac cagccatag tgcgacaagg ctcccgatg aagctgcggg atctcaaagt 6120

ccagcgatga gccgcccacc agcaccacaa aggcgatatc gcgaatggaa ccgccgggtg 6180
agacctggcg cagcgcgcg caggcagttgg tgacaaacac tttctcttc gcctgcccgc 6240
gcacgagacg aatcttttcc agcgggctgg cgttatcgat cggcaccagt tcgcctcct 6300
tgatgtacac cactttggcg aacaccgccg ggctgagggc ttcccgaag aactccaccg 6360
cgccattctc gtgacgaata ctgaacaggc ttccacttt ggccagcggg tattttttta 6420
tcgttccgc cagcgaaaga tcctcgaggc ccagctcggt tttaatcaac aggctgacca 6480
tattccccgc cccggcgaga tggaccgccg ttatctgccc ctcccggtg acgatcgccg 6540

catccgtcga gccggcgccg aggtcgagga tcgccagcgg cgcgcacag ccgggagtgg 6600
ttaacgcccc ggcgatggcc atgttggcct ccacgccgc caccaccacc tcggtctgca 6660
gtcgggcgct cagttcggc gcgataacct gcatttgag acgatccgct ttcaccatcg 6720
ccgccatccc gacggcattc tccatggcgc actcgccggc catcccgccc tgcaccttgc 6780
gcggaataaa cgtatccacc gccagcagat cctggatgta tatcgcgctc atctcatggc 6840

cggtcaggga cgccattacc ttgcgcaccc gctcaagcat gccgccggcg tgggtgcccc 6900
gttcgccgcg gatgtcgcgt accggagcgc aggcgctcat cgcctgcatg atggcttccg 6960

cgccctcggc gacatcgccc tctccgggc gcttttcgcc gctaagttag aggttgcccc 7020
ccgggatcac ccgcgactgc acatccccct gcgggggtctt gagcaccacc gcggaacggt 7080
tgccaatcag ggcgcgggcg atggggacga tggcctgggt ctcttcgggg cttagccccga 7140
agaaggtggc gatccccgtag ggattcgaca ggatccgcac cacctggccc gccgcggcca 7200
cttcaccgc cgccattacc cctcgggga cctgctccag cagcgtcact tcatccacca 7260
ccggcagggt tttagcagg cggttggttca ccagcacgcc gtgctcttt ttgaggatcg 7320
ccgccaccac gttagtcccc cggtcgagcg cctcattgag ccaccacag gcgtcaagga 7380

aatcgacggc gtgctcaatc agtacgatcc acccctcggc ataactgcgc gccggcagcg 7440
tcgccagccg ccgagggcg atagtcgtcc ccacgccaac gccaccccc cccggcgtct 7500
gcgggttatg accgatcatg gtgattcgg tgataatggt ctcggtgatg gtctccatcg 7560
ccacatgcc aatcacggc gcggcttctg taagatagat gcgagagaca tcgtcatcg 7620
accacggtgt tttagccagg gcctgtcca gcgcggcgag ggtccggcg atattgtccc 7680
gcgtcccttt catgccgtc gtgcgacga tccgctggc aacaaacgcc ctgcctgcg 7740
ggtagtcgga cgccagcgc acctcgggtg tggcgttgcc gatatcaatc ccggctatta 7800

acggcatgct gacctccgt tagcttctt tacgcagctt atgccgtgc tgataactt 7860
ccgccactc ccggacaaag gcggcattca ctgtcgcatg ccaggtgtgc tccagctct 7920
cggcgatcg cagcagctcc gcctcgagg agcggaacgg gcgcagcgc ttatagatag 7980
ccagaatcg ctgctcagga atggcgataa gctccgccgc gcggcgaaa ttgcgcgcca 8040
ccgcatggcg ctgcatctgc tcggcaatct gcgcctggta ctcaagggtc tggcgggaga 8100
tccgcacatc ctgcgggcc acctgccag agagcacctt ctgagggtg atatcggtca 8160
atggtttgcc gtaggcgtc aggatatgct ccgggcagcg ggtggctaac ggataatct 8220

gcacgcgat ggttttctc ctcatggtca ctcccttact aagtcgatgt gcagggtgac 8280
gggctcggcg tctgcacca catgtttgt ctctttgata tgaaatagcg cggctttggc 8340
cataaatttc ggccgacca tctgatctt caccaccggc accggcgaag gtgactcttt 8400
gcgcgcatag cgcgcagcgt ttttgccaat ctgccgtag gtctccagcg tcagcagcgg 8460
gcctgggag aacagctcca ggttgctgag cggcagcaga tcgcgctgat ggatgaccgt 8520
gttcccttc gactggatc cgatccgat cccgagccg ctgaggttg ccgcatccca 8580
ggccataaag gagacgtcg acgtgcgcag aatgcgcacc acccgggctg gaagccctc 8640

ttcttccacc ccggcaatca gctctttgag gatcgcgcca tggggcatat cgatcagagt 8700
 gtgatgctgg tgtttatcga aggcagggcc gacgccgatc accacttcat cggcgcgttc 8760
 atcggcagaa gctaccccg cctcgcgggt ttccagggta aaagagggt gaatttgggt 8820
 tgtctgttgc acaggaatac cgccttggtc aatggtgtcg ggctgaacca cggccggaat 8880
 atttttgata tccgccagc gtccggcaga gatgcgatag ccggtgcccg gccctgata 8940
 gtcatgtatg tcgttgaccg cactcaccac ctggaactgc cgatcgaaaa tggccgaggt 9000
 ctgcaggtaa tcgccgtga cccgtggcg cagcatattg agaatttgc tggcgatgc 9060

 ctcaaagccg ctgcggctca gcgcgccgac aatatcgagg ccggtgatgt tgcgcttcat 9120
 catctcttc accgcactca gatctccac cacgttacgc ggcggcatct cgttgctgcc 9180
 gtgcgcgtag gtggcgccct ccacctctc gtccggcatt ggccgcagcc ccagctcgcg 9240
 gaaaaccgcc tggatcgccc gcgcgccttt ctggcgaatg gcaatggttt ccgcctcgggt 9300
 caccggacgc agcccgccgt caaccatcag gtcacgtgc aggatgttgt aatcatcaaa 9360
 atcttccgca tcgaagtgc agccggcgaa catgttgtcg tagttcgga ccgcgtgta 9420
 gccggagaaa ataaagtcgg tgcccgccag catctgcac aggggtgcgc cgggtgcggcg 9480

 aatatccgag tgggagaaa tctggtcgtt ggccgacgcc acttcgaggt cgagcataga 9540
 ggcgatcagg ttttccgcca gcaccgccg aatgccgac ggcacagcg cggatcatgcc 9600
 gatacagctc accgcgccgt ttgacgtcc ctgaacccg gcgcctttag taatgaagat 9660
 gcagcgcat tcgaggtaga gcatcgactt gctctccgaa tagcccatca gcgcttcgga 9720
 tccgtgccc gaggtgtagc gcattttcaa cccgcgggag gcgtaggccg aggcgaggaa 9780
 cgctttgac caccggctat catcgccgtc ggtaaatacc gcttcggtgc cgtagaccga 9840
 caccgtctcg gcgtagctgg ttaagccag catgccagc tccagctcgg tggcctcttc 9900

 caccgagcac tgcgtcaaca cggcgggcg gccgcactgc gaaccgacca acagcgccag 9960
 ggctttaaac ggccgctagc gcgcgatacc gaccgtggtc tctgtttctg agaagccgcg 10020
 gatccggcc tggcgccgt cagcggcaat ctgcaccgga ttatctttga gattggtgac 10080
 gtggcactgg ttggaggggg tccggcgggc acgcatcttc tgcagcgcca tcatcatctc 10140
 caccagttc atctgcgcca tcacctgac cgctttggcc ggctgatgg cggtagtgat 10200
 ggcaatgata tcttccggc tgacgtgaat atccaccagc atacgggcta tttccaccgc 10260
 ctccaggcgc attgcctgct ctgtgcgtc aacgttgatc gcgtaatcgg cgataaatcg 10320

 gtgatcatg taaaactggt cccggcgttt gccgtccagt tcgacgatca gaccgtgtc 10380
 cacttttact gaagagaccg ggtcaaagg gctgtccatg gcgatcagcc cctcttcagg 10440
 ccaactcgca atcagccgt cctgattgac gggcgctgg gccagtactg caaatcgttt 10500

tgatcttttc attgttcatc ggctcaaaag gtgaagcttg gttacctccg ggaaacgcgg 10560
 ttgatttggt tagtgggtga attatttgct caggatgtgg cattgtcaag ggcgtgacgg 10620
 ctgcctgac ttctgttcc agtgcctccg tccgacagtc gagcgtgcga gcccataatc 10680
 tcgcgtggt gctgcatacc gtggcaaaca gcacagatcg cctaggaaaa aaaaagcccg 10740

 cactgtcagg tgcgggcttt tttctgtgtt tgctaggcca gttcaagcgc aagcatcagg 10800
 gtgcagctgg gcagaggcga gattcctccc cgggatcacg aactgtttta acgggcccgt 10860
 ctcgccata ttgcgttga taagccgtc cagggcgggtg atctcctctt cgccgatcgt 10920
 ctggctcagg cgggtcaggc cccgcgcacg gctggccagt tcagcccca gcacgaacag 10980
 cgtctgctga atatggtgca ggctttccc cagcccggtg tcgcgggtcg tggcgtagca 11040
 gagccccagc tgggataca gttcatcgac ggtgccgtag gcctcgacgc gaatatggc 11100
 tttctcgatg cggctgccgc cgtacagggc ggtggtgcct ttatccccgg tgcgggtata 11160

 gatacgatac attcagtttc tctcacttaa cggcaggact ttaaccagct gcccgcgctt 11220
 ggccgcgagc gtacgcagtt gatcgtcgt atcggtgacg tgtccggtag ccagcgcgcc 11280
 gtccgccggc agctgggcat gagtgagggc tatctcgccg gacgcgtga gcccgatacc 11340
 caccgcagg ggcgagcttc tggccgccag ggccgccagc gcagcgcggt caccgcctcc 11400
 gtcatagggt atggtctggc aggggacccc ctgctcctcc agccccagc acagtcatt 11460
 gatggcgccg gcatggtgcc cgcgcggatc gtaaaacagg cgtacgcctg gcggtgaaag 11520
 cgacatgacg gtccctcgt taacactcag aatgcctggc ggaacatacg atagtcata 11580

 atataccttc tcgttcagg ttataatgcg gaaaaacaat ccagggcgca ctgggctaata 11640
 aattgacct gtcgaccgt accgccgta acgccgacgg cgccaattac ctgctcatta 11700
 aaaataactg gcaggccgcc gccaaaaata ataattcgt gttggttggt tagctgcaga 11760
 ccgtacagag attgtcctgg ctggaccgt gacgtaattt catgggtacc ttgcttcagg 11820
 ctgcaggcgc tccaggcttt attcaggga atatcgacg tggagacgaa ggcctcgtcc 11880
 atccgtgga taagcagcgt gttgcctccg cggtaacta cggaaaacac caccgccag 11940
 ttgatctcag tggctttttt ttccaccgcc gccgccatit gctgggcggc ggccagggtg 12000

 attgtctgaa ctgttggtt ctgttcatc attctctccc gcaagcttgg ttacctccg 12060
 gaaacgcggt tgatttggt agtgggtgaa ttatttgctc aggatgtggc attgtcaagg 12120
 gcgtgacggc tcgctgact tctcgttcca gtccccctg ccgacagtcg agcgtgcgag 12180
 cccataatct cgcgtggtg ctgcataccg tggcaaacag cacagatcgc ctacagtc 12240
 aaagcctccg gtcggaggct ttgactatt taaatgaatt cccgacagta agacgggtaa 12300
 gcctgttgat gataccgctg ccttactggg tgcattagcc agtctgaatg acctgtcacg 12360

ggataatccg aagtggtcag actggaaaat cagagggcag gaactgctga acagcaaaaa 12420

gtcagatagc accacatagc agaccgcga taaaacgccc tgagaagccc gtgacgggct 12480

tttcttgat tatgggtagt ttcttgcag gaatccataa aaggcgctg tagtgccatt 12540

tacccccatt cactgccaga gccgtgagcg cagcgaactg aatgtcacga aaaagacagc 12600

gactcaggtg cctgatggc ggagacaaaa ggaatatcca gcatgttgc cgagcttgcg 12660

agggtgctac ttaagccttt agggttttta ggtctgttt gtagaggagc aaacagcgtt 12720

tgcgacatcc ttttgtaata ctgcggaact gactaaagta gtgagttata cacagggtg 12780

ggatctattc tttttatctt tttttattct ttctttattc tataaattat aaccacttga 12840

atataaaca aaaaaacaca caaaggtcta gcggaattta cagagggtct agcagaattt 12900

acaagtttc cagcaaaggt ctagcagaat ttacagatac ccacaactca aaggaaaagg 12960

actagtaatt atcattgact agcccatctc aattggata gtgattaaaa tcacctagac 13020

caattgagat gtatgtctga attagtgtt ttcaaagcaa atgaactagc gattagtgcg 13080

tatgacttaa cggagcatga aaccaagcta attttatgct gtgtggcact actcaacccc 13140

acgattgaaa accctacaag gaaagaacgg acggtatcgt tcacttataa ccaatacgt 13200

cagatgatga acatcagtag ggaaaatgct tatggtgtat tagctaaagc aaccagagag 13260

ctgatgacga gaactgtgga aatcaggaat cctttggta aaggctttga gattttccag 13320

tggacaaact atgccaagtt ctcaagcgaa aaattagaat tagtttttag tgaagagata 13380

ttgccttacc tttccagtt aaaaaattc ataaaatata atctggaaca tgtaagtct 13440

tttgaaaaca aatactctat gaggatttat ggtggttat taaaagaact aacacaaaag 13500

aaaactcaca aggcaaatat agagattagc cttgatgaat ttaagttcat gttaatgctt 13560

gaaaataact accatgagtt taaaaggctt aaccaatggg ttttgaaacc aataagtaaa 13620

gatttaaaaa cttacagcaa tatgaaattg gtggttgata agcgaggccg cccgactgat 13680

acgttgattt tccaagttga actagataga caaatggatc tcgtaaccga acttgagaac 13740

aaccagataa aaatgaatgg tgacaaaata ccaacaacca ttacatcaga ttctaccta 13800

cataacggac taagaaaaac actacacgat gctttaactg caaaaattca gtcaccagt 13860

tttgaggcaa aatttttgag tgacatgcaa agtaagtatg atctcaatgg ttcgttctca 13920

tggctcacgc aaaaacaacg aaccacacta gagaacatac tggctaaata cggaaggatc 13980

tgaggttctt atggctcttg tatctatcag tgaagcatca agactaaca acaaaagtag 14040

aacaactgtt caccgttaca tatcaaaggg aaaactgtcc atatgcacag atgaaaacgg 14100

tgtaaaaaag atagatacat cagagctttt acgagttttt ggtgcattca aagctgttca 14160

ccatgaacag atcgacaatg taacagatga acagcatgta acacctaata gaacaggtga 14220
aaccagtaaa acaaagcaac tagaacatga aattgaacac ctgagacaac ttgttacagc 14280
tcaacagtca cacatagaca gcctgaaaca ggcatgctg cttatcgaat caaagctgcc 14340
gacaacacgg gagccagtga cgcctcccg tgggaaaaaa tcatggcaat tctggaagaa 14400
atagcgcttt cagccggcaa accggctgaa gccggatctg cga 14443

<210> 77

<211> 6944

<212> DNA

<213> Artificial Sequence

<220><223> Plasmid

<400> 77

aattcgagg accgtgatac acgggacagg tcaactgaatg acgacaatgt cctggaatc 60
agcgaaccgc gcatctgaag tacatttgag cgactgtacc agaacatgaa tgaggcgctt 120
ggattaggcg attattagca gggctaagca tttactatt attattttcc ggttgaggga 180
tatagagcta tcgacaacaa ccggaataag tttactgcta tattgtctgaa ggtacaggcg 240
tttcataac tatttgctcg cgttttttac tcaagaagaa aatgccaaat agcaacatca 300
ggcagacaat acccgaaatt gcgaagaaaa ctgtctggta gcctgcgtgg tcaaagagta 360
tcccagtcgg cgttgaaagc agcacaatcc caagcgaact ggcaatttga aaaccaatca 420

gaaagatcgt cgacgacagg cgcttatcaa agtttgccac gctgtatttg aagacggata 480
tgacacaaag tggaacctca atggcatgta acaacttcac taatgaaata atccaggggt 540
taacgaacag cgcgaggaa aggatagca acgcataat cacaactccg ataagtaatg 600
catttttttg ccctaccga ttacacaaaga aaggaataat cgccatgcac agcgcttcga 660
gtaccacctg gaatgagttg agataacat acaggcgctg tctacatcg tgtgattcga 720
ataaacctga ataaaagaca ggaaaaagt gtgatcaaa aatgttatag aaagaccacg 780
tccccacaat aaatatgacg aaaaccaga agtttcgac cttgaaaact gcgataaaat 840

cctctttttt taccctccc gcatctgccg ctacgcactg gtgatcctta tctttaaacc 900
gcatgttgat catcataat acagcgccaa atagcgagac caaccagaag ttgatatggg 960
gactgatact aaaaaatat ccggcaaaga acgcgccaat agcatagcca aaagatcccc 1020
aggcgcgcg tgttccat tcaaatgaa aatttcgcg cattttttcg gtgaagctat 1080
caagcaaacc gcatccgcc agataccca agccaaaaa tagcgcccc agaattagac 1140
ctacagaaaa attgctttgc agtaacggt cataaacgta aatcataaac ggtccggtca 1200
agaccaggat gaaactcata caccagatga gcggtttctt cagaccgagt ttatcctgaa 1260

cgatgccgta gaacatcata aatagaatgc tggtaaactg gttgaccgaa taaagtgtac 1320
 ctaattccgt cctgtcaac cctagatgtc ctttcagcca aatagcgtat aacgaccacc 1380
 acagcgacca ggaataaaaa aagagaaatg agtaactgga tgcaaaacga tagtacgcat 1440
 ttctgaatgg aatattcagt gccataatta cctgcctgtc gttaaaaaat tcacgtccta 1500
 ttttagagata agagcgactt cgccgtttac ttctcactat tccagtcttt gtcgacatgg 1560
 cagcgtgtc attgccccctt tcgccgttac tgcaagcgct ccgcaacgtt gagcgagatc 1620
 gataattcgt cgcatttctc tctcatctgt agataatccc gtagaggaca gacctgtgag 1680

 taaccggca acgaacgcat ctccgcccc cgtgctatcg acacaattca cagacattcc 1740
 agcaaatgg tgaactgtc ctcgataaca gaccaccacc cttctgcac ctttagtcac 1800
 caacagcatg gcgactcat actcttttgc cagggcgcat atatcctgat cgttctgtgt 1860
 ttttccactg ataagtcgcc attcttcttc cgagagcttg acgacatccg ccagttgtag 1920
 cgctgccgc aaacacaagc ggagcaaatg ctgctcttgc catagatctt cacgaatatt 1980
 aggatcgaag ctgacaaaac ctccggcatg ccggatcgcc gtcacgcag taaatgcgct 2040
 ggtacgcgaa ggctcggcag acaacgcaat tgaacagaga tgtaaccatt cgccatgtcg 2100

 ccagcagggc aagtctgtcg tctctaaaaa aagatcggca ctggggcgga ccataaacgt 2160
 aaatgaacgt tccccttgat cgttcagatc gacaagcacc gtggatgtcc ggtgccattc 2220
 atcttgcttc agatactga tatcgactcc ctgagttagc agcgttcttt gcattaacgc 2280
 accaaaagga tcatccccc cccgacctat aaaccactt gttccgccta atctggcgat 2340
 tcccaccgca acgttagctg gcgcgccgcc aggacaaggc agtaggcgcc cgtctgattc 2400
 tggcaagaga tctacgaccg catcccctaa aaccatact ttggtgaca ttttttccc 2460
 ttaaattcat ctgagttacg catagtataaacctctttt tcgcaaaatc gtcattggatt 2520

 tactaaaaca tgcatattcg atcacaaaac gtcatagtta acgttaacat ttgtgatatt 2580
 catcgcatth atgaaagtaa gggactttat ttttataaaa gttaacgtta acaattcacc 2640
 aaatttgctt aaccaggatg attaaaatga cgcaatctcg attgcatgcg gcgcaaacg 2700
 ccctagcaaa acttcatgag caccggggta acactttcta tcccatttt cacctcgcgc 2760
 ctctgccgg gtggatgaac gatccaaacg gcctgatctg gtttaacgat cgttatcacg 2820
 cgttttatca acatcatccg atgagcgaac actgggggcc aatgcactgg ggacatgcca 2880
 ccagcgacga tatgatccac tggcagcatg agcctattgc gctagcgcca ggagacgata 2940

 atgacaaaga cgggtgtttt tcaggtagtg ctgtcgatga caatggtgtc ctctcactta 3000
 tctacaccgg acacgtctgg ctcgatggtg caggtaatga cgatgcaatt cgcaagtag 3060

aatgtctggc taccagtcgg gatggtattc atttcgagaa acaggggtgtg atcctcactc 3120
caccagaagg aatcatgcac ttccgcgatc ctaaagtgtg gcgtgaagcc gacacatggt 3180
ggatggtagt cggggcgaaa gatccaggca acacggggca gatcctgctt tatcgcggca 3240
gttcgttgcg tgaatggacc ttcgatcgcg tactggccca cgctgatgcg ggtgaaagct 3300
atatgtggga atgtccggac tttttcagcc ttggcgatca gcattatctg atgttttccc 3360

cgcagggaat gaatgccgag ggatacagtt accgaaatcg ctttcaaagt ggcgtaatac 3420
ccggaatgtg gtcgccagga cgactttttg cacaatccgg gcattttact gaacttgata 3480
acgggcatga cttttatgca ccacaaagct ttttagcgaa ggatggtcgg cgtattgtta 3540
tcggctggat ggatatgtgg gaatcgccaa tgccctcaaa acgtgaagga tgggcaggct 3600
gcatgacgct ggcgcgcgag ctatcagaga gcaatggcaa acttctacaa cgcccggtag 3660
acgaagctga gtcgttacgc cagcagcatc aatctgtctc tccccgcaca atcagcaata 3720
aatatgtttt gcaggaaaac gcgcaagcag ttgagattca gttgcagtgg gcgctgaaga 3780

acagtgatgc cgaacattac ggattacagc tcggcactgg aatgcggctg tatattgata 3840
accaatctga gcgacttgtt ttgtggcggg attaccaca cgagaattta gacggctacc 3900
gtagtattcc cctcccgcag cgtgacacgc tcgccctaag gatattttatc gatacatcat 3960
ccgtggaagt atttatiaac gacggggaag cggatgatgag tagtcgaatc tatccgcagc 4020
cagaagaacg ggaactgtcg ctttatgcct cccacggagt ggctgtgctg caacatggag 4080
cactctggct actgggttaa cataatatca ggtggaacaa cggatcaaca gcgggcaagg 4140
gatccacgaa gcttcccatg gtgacgtcac cggttctaga tacctagggtg agctctggta 4200

ccctctagtc aaggccttaa gtgagtcgta ttacggactg gccgtcggtt tacaacgtcg 4260
tgactgggaa aaccctggcg ttaccaact taatcgctt gcagcacatc cccctttcgc 4320
cagctggcgt aatagcgaag agggccgcac cgatcgccct tcccaacagt tgcgcagcct 4380
gaatggcgaa tggcgcttcg cttggtaata aagcccgtt cggcgggctt tttttgtta 4440
actacgtcag gtggcacitt tcggggaaat gtgcgcggaa cccctatttg tttatttttc 4500
taaatacatt caaatatgta tccgtcatg agacaataac cctgataaat gcttcaataa 4560
tattgaaaaa ggaagagtat gagtatcaa catttccgtg tcgcccttat tccctttttt 4620

gcggcatttt gccttccgtt ttttctcac ccagaaacgc tggtgaaagt aaaagatgct 4680
gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc 4740
cttgagagtt ttgcgccga agaactttct ccaatgatga gcacttttaa agttctgcta 4800
tgtggcgagg tattatcccg tgttgacgcc gggcaagagc aactcggtcg ccgcatacac 4860
tattctcaga atgacttggt tgagtactca ccagtcacag aaaagcatct tacggatggc 4920

atgacagtaa gagaattatg cagtgtgcc ataacatga gtgataacac tgcggccaac 4980
 ttacttctga caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg 5040

gatcatgtaa ctgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac 5100
 gagcgtgaca ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact attaactggc 5160
 gaactactta ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt 5220
 gcaggaccac ttctgcgtc ggcccttccg gctggctggt ttattgctga taaatctgga 5280
 gccggtgagc gtgggtctcg cggatcatt gcagcactgg ggccagatgg taagccctcc 5340
 cgtatcgtag ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag 5400
 atcgctgaga taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca 5460

tatatacttt agattgattt accccggttg ataatcagaa aagcccaaa aacaggaaga 5520
 ttgtataagc aaatatitaa attgtaaacg ttaatatit tttaaaattc gcgttaaatt 5580
 tttgttaaat cagctcattt tttaaccaat aggccgaaat cggcaaaatc cttataaat 5640
 caaaagaata gcccgagata gggttgagtg ttgttccagt ttggaacaag agtcactat 5700
 taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc gatggccac 5760
 tacgtgaacc atcacccaaa tcaagttttt tggggtcgag gtgccgtaaa gcactaaatc 5820
 ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagcgaacg tggcgagaaa 5880

ggaagggaag aaagcgaag gagcgggcgc tagggcgctg gcaagtgtag cggtcacgct 5940
 gcgcgttaacc accacacccg ccgcgcttaa tgcgccgcta caggcgcgct aaaaggatct 6000
 aggtgaagat cttttttagt aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc 6060
 actgagcgtc agaccccgta gaaaagatca aaggatcttc ttgagatcct tttttctgc 6120
 gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc agcggtggtt tgtttgccgg 6180
 atcaagact accaactctt ttccgaagg taactggctt cagcagagcg cagataccaa 6240
 atactgttct tctagtgtag ccgtagttag gccaccactt caagaactct gtagcaccgc 6300

ctacatacct cgctctgcta atcctgttac cagtggctgc tgccagtggc gataagtcgt 6360
 gtcttaccgg gttagactca agacgatagt taccggataa ggcgagcgg tcgggctgaa 6420
 cggggggttc gtgcacacag ccagcttgg agcgaacgac ctacaccgaa ctgagatacc 6480
 tacagcgtga gctatgaaa agcgccacgc ttcccgaagg gaaaaaggcg gacaggtatc 6540
 cggtaagcgg cagggtcggg acaggagagc gcacgaggga gcttccaggg ggaaacgcct 6600
 ggtatcttta tagtctgtc gggtttcgcc acctctgact tgagcgtcga tttttgtgat 6660
 gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa cgcgcccttt ttacggttcc 6720

tggccttttg ctggcctttt gctcacatgt aatgtgagtt agtcactca ttaggcaccc 6780
caggctttac actttatgct tccggtcgt atgtttgttg gaattgtgag cggataacaa 6840
tttcacacag gaaacageta tgacatgat tacgccaagc tacgtaatac gactcactag 6900
tgggcagatc ttcgaatgca tcgcgcgcac cgtacgtctc gagg 6944

<210> 78

<211> 9317

<212> DNA

<213> Artificial Sequence

<220><223> Plasmid

<400> 78

tcgaggaatt cgcaggaccg tgatacacgg gacaggtcac tgaatgacga caatgtcctg 60
gaaatcagcg aaccgcgcat ctgaagtaca ttgagcgac tgtaccagaa catgaatgag 120

gcgtttggat taggcgatta ttagcagggc taagcatitt actattatta ttttcggtt 180
gagggatata gagctatcga caacaaccgg aaaaagtta cgtctatatt gctgaaggta 240
caggcgtttc cataactatt tgctcgcgtt ttttactcaa gaagaaaatg ccaaatagca 300
acatcaggca gacaataccc gaaattgcga agaaaactgt ctggtagcct gcgtggtcaa 360
agagtatccc agtcggcggtt gaaagcagca caatccaag cgaactggca atttgaanaac 420
caatcagaaa gatcgtcgac gacaggcgct tatcaaagtt tgccacgtg tatttgaaga 480
cggatatgac acaaagtggg acctcaatgg catgtaacaa cttcactaat gaaataatcc 540

aggggttaac gaacagcgcg caggaaagga tacgcaacgc cataatcaca actccgataa 600
gtaatgcatt ttttggccct acccgattca caaagaaagg aataatcgcc atgcacagcg 660
cttcgagtac cacctggaat gaggttgagat aaccatacag gcgcgttcct acatcgtgtg 720
attcgaataa acctgaataa aagacaggaa aaagttgttg atcaaaaatg ttatagaaag 780
accacgtccc cacaataaat atgacgaaaa ccagaagtt tcgatccttg aaaactgcga 840
taaaatctc tttttttacc cctcccgcat ctgccgttac gcaactggtga tccttatctt 900
taaaacgcat gttgatcatc ataaatacag cgccaaatag cgagaccaac cagaagttga 960

tatggggact gatactaaaa aatatgccgg caaagaacgc gccaatagca tagccaaaag 1020
atccccaggc gcgcgtgttt ccatattcga aatgaaaatt tcgcgccatt ttttcggtga 1080
agctatcaag caaacgcgat cccgccagat accccaagcc aaaaaatagc gccccagaa 1140
ttagacctac agaaaaattg ctttcagta acggttcata aacgtaaatc ataaacggtc 1200
cggatcaagc caggatgaaa ctatacacc agatgagcgg tttcttcaga ccgagtttat 1260
cctgaacgat gccgtagaac atcataaata gaatgctggt aaactggttg accgaataaa 1320

gtgtacctaa ttccgtccct gtcaacccta gatgtccttt cagccaaata gcgtataacg 1380

accaccacag cgaccaggaa ataaaaaaga gaaatgagta actggatgca aaacgatagt 1440

acgcattttct gaatggaata ttcagtgccca taattacctg cctgtcgtta aaaaattcac 1500

gtcctattta gagataagag cgacttcgcc gtttacttct cactattcca gttcttgtcg 1560

acatggcagc gctgtcattg cccctttcgc cgttactgca agcgtccgc aacgttgagc 1620

gagatcgata attcgtcgca tttctctctc atctgtagat aatcccgtag aggacagacc 1680

tgtgagtaac ccggcaacga acgcatctcc cgtcccggtg ctatcgacac aattcacaga 1740

cattccagca aaatggtgaa cttgtcctcg ataacagacc accaccctt ctgcaccttt 1800

agtcaccaac agcatggcga tctcatactc ttttgccagg gcgcataat cctgatcgtt 1860

ctgtgttttt ccaactgataa gtgcgcatc ttttccgag agcttgacga catccgccag 1920

ttgtagcgcc tgcgcgaac acaagcggag caaatgctcg tcttgccata gatcttcacg 1980

aatattagga tcgaagctga caaacctcc ggcatgccgg atcgccgtca tcgcagtaaa 2040

tgcgctggta cgcgaaggct cggcagacaa cgcaattgaa cagagatgta accattcgcc 2100

atgtcgccag cagggaagt ctgtcgtctc taaaaaaga tcggcactgg ggcggaccat 2160

aaacgtaaat gaacgttccc cttgatcgtt cagatcgaca agcaccgtgg atgtccggtg 2220

ccattcatct tgcttcagat acgtgatac gactccctca gttagcagcg tttttgcat 2280

taacgcacca aaaggatcat cccccaccg acctataaac ccacttgctc cgcctaactc 2340

ggcgattccc accgcaacgt tagctggcgc gccgccagga caaggcagta ggcgcccgtc 2400

tgattctggc aagagatcta cgaccgcac ccctaaaacc catactttgg ctgacatttt 2460

tttcccttaa attcatctga gttacgcata gtgataaacc tctttttcgc aaaatcgta 2520

tggatttact aaaacatgca tattcgatca caaacgtca tagttaacgt taacatttgt 2580

gatattcatc gcatttatga aagtaaggga ctttattttt ataaaagtta acgttaacaa 2640

ttcaccaaat ttgcttaacc aggatgattia aaatgacgca atctcgattg catgcggcgc 2700

aaaacgccct agcaaaactt catgagcacc ggggtaacac tttctatccc cattttcacc 2760

tcgcgcctcc tgcgggtgg atgaacgatc caaacgccct gatctggttt aacgatcgtt 2820

atcacgcgtt ttatcaacat catccgatga gcgaacactg ggggccaatg cactggggac 2880

atgccaccag cgacgatatg atccactggc agcatgagcc tattgcgcta gcgccaggag 2940

acgataatga caaagacggg tgtttttcag gtagtgctgt cgatgacaat ggtgtcctct 3000

catttatcta caccggacac gtctggctcg atggtgcagg taatgacgat gcaattcgcg 3060

aagtacaatg tctggctacc agtcgggatg gtattcattt cgagaaacag ggtgtgatcc 3120

tcactccacc agaaggaatc atgcacttcc gcgatcctaa agtgtggcgt gaagccgaca 3180
 catgttggat ggtagtcggg gcgaaagatc caggcaacac ggggcagatc ctgctttatc 3240
 gcggcagttc gttgcgtgaa tggaccttcg atcgctact ggcccacgct gatgcgggtg 3300
 aaagctatat gtgggaatgt ccggactttt tcagccttgg cgatcagcat tatctgatgt 3360
 tttccccgca gggaatgaat gccgagggat acagttaccg aaatcgcttt caaagtggcg 3420
 taatacccgg aatgtggtcg ccaggacgac tttttgcaca atccgggcat tttactgaac 3480

 ttgataacgg gcatgacttt tatgcaccac aaagcttttt agcgaaggat ggtcggcgta 3540
 ttgttatcgg ctggatggat atgtgggaat cgccaatgcc ctcaaaacgt gaaggatggg 3600
 caggctgcat gacgctggcg cgcgagctat cagagagcaa tggcaaactt ctacaacgcc 3660
 cggtagacga agctgagtcg ttacgccagc agcatcaatc tgtctctccc cgcacaatca 3720
 gcaataaata tgttttcgag gaaaacgcgc aagcagttga gattcagttg cagtgggcgc 3780
 tgaagaacag tgatgccgaa cattacggat tacagctcgg cactggaatg cggctgtata 3840
 ttgataacca atctgagcga cttgttttgt ggcggtatta cccacacgag aatttagacg 3900

 gctaccgtag tattcccctc ccgcagcgtg acacgctcgc cctaaggata tttatcgata 3960
 catcatccgt ggaagtatit attaacgacg gggaagcggg gatgagtagt cgaatctatc 4020
 cgagccaga agaacgggaa ctgtcgcttt atgcctccca cggagtggct gtgctgcaac 4080
 atggagcact ctggctactg ggttaacata atatcaggtg gaacaacgga tcaacagcgg 4140
 gcaagggatc cacgaagctt cccatgggtga cgtcacgggt aaaccagcaa tagacataag 4200
 cggctattta acgaccctgc cctgaaccga cgaccgggtc gaatttgctt tcgaatttct 4260
 gccattcacc cgcttattat acttattcag gcgtagcacc aggcgtttta gggcaccaat 4320

 aactgcctta aaaaaattac gccccgcctt gccactcacc gcagtactgt tgtaattcat 4380
 taagcattct gccgacatgg aagccatcac agacggcatg atgaacctga atcgccagcg 4440
 gcatcagcac ctgtcgctt tgcgtataat atttgcccat ggtgaaaacg ggggcgaaga 4500
 agttgtccat attggccacg tttaaatcaa aactggtgaa actcaccag ggattggctg 4560
 agacgaaaaa catattctca ataaacctt tagggaaata ggccaggttt tcaccgtaac 4620
 acgccacatc ttgcgaatat atgtgtagaa actgccgga atcgtcgtgg tattcactcc 4680
 agagcgatga aaacgtttca gtttctcat ggaaaacggg gtaacaagg tgaacactat 4740

 cccatatcac cagctcaccg tctttcattg ccatacggaa ttccggatga gcattcatca 4800
 ggccggcgaag aatgtgaata aaggccggat aaaacttggt cttatttttc tttacggtct 4860
 ttaaaaaggc cgtaatatcc agctgaacgg tctggttata ggtacattga gcaactgact 4920
 gaaatgcctc aaaatgttct ttacgatgcc attgggatat atcaacggtg gtatatccag 4980

tgattttttt ctccatttta gcttccttag ctccgaaaa tctcgataac tcaaaaaata 5040
 cgcccgtag tgatcttatt tcattatggt gaaagttaga acctcttacg tgccgatcaa 5100
 cgtctcattt tcgcaaaaag ttggcccagg gcttcccgt atcaacaggg acaccaggat 5160

 ttattttattc tgcaagtga tcttcctca caggtattta ttcggcgcaa agggcctcgt 5220
 gatacgcta tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtag 5280
 cacttttcgg ggaatgtgc gcgcccgcgt tctgctggc gctgggcctg tttctggcgc 5340
 tggacttccc gctgttcgt cagcagcttt tcgcccacgg ccttgatgat cgcggcgcc 5400
 ttggcctgca tatcccgatt caacggcccc agggcgctca gaacgggctt caggcgctcc 5460
 cgaaggtctc gggccgtctc ttgggcttga tcggccttct tgcgcatctc acgcgtcct 5520
 gcggcggcct gtagggcagg ctcatacccc tgccgaaccg cttttgtcag ccggtcggcc 5580

 acggcttcg gcgtctcaac gcgctttgag attcccagct tttcgccaa tccctgcggt 5640
 gcataggcgc gtggctcgac cgcttgcggg ctgatggtga cgtggccac tggcggccgc 5700
 tccagggcct cgtagaacgc ctgaatgcgc gtgtgacgtg ccttgcctgcc ctcatgccc 5760
 cgttgagcc ctagatcggc cacagcgcc gcaaactgg tctggtcgcg ggtcatctgc 5820
 gctttgttc cgtgaactc cttggccgac agcctgccgt cctgcgtcag cggcaccacg 5880
 aacgcgtca tgtcgggct ggtttcgtca cggcggatgc tggccgtcac gatgcgatcc 5940
 gccccgtact tgtccgcag ccaattgtgc gccttctga agaaccgcgc ctgctgttct 6000

 tggttgccg acttccacca ttccgggctg gccgtcatga cgtactcgac cgccaacaca 6060
 gcgtccttc gccgttctc ttgcagcaac tcgcgcagtc ggcccatcgc ttcacgggtg 6120
 ctgctggccg ccagtgctc gttctctggc gtctgctgg cgtcagcgtt gggcgtctcg 6180
 cgctcgggt aggcgtgctt gagactggcc gccacgttc ccattttcgc cagcttcttg 6240
 catcgatga tcgctatgc gccatgcct gccctccct tttggtgtcc aaccggtcgc 6300
 acgggggcag cgcaaggcgg tgcctccgc gggccactca atgcttgagt atactcacta 6360
 gactttgctt cgcaaagtc tgaccgcta cggcggctgc ggcgccctac gggttgctc 6420

 tccgggcttc gccctgcgc gtcgtgcgc tcccttgcca gccgtggat atgtggacga 6480
 tggccgcgag cggccaccgc ctggctcgt tcgctcggcc cgtggacaac cctgctggac 6540
 aagctgatgg acaggctgc cctgcccacg agcttgacca cagggttgcc ccaccggcta 6600
 cccagccttc gaccacatac ccaccggctc caactgcgc gcctgcggcc ttgcccac 6660
 aattttttt attttctct gggaaaagcc tccggcctgc ggctgcgcg cttcgttg 6720
 cggttggaca ccaagtggaa ggccgggtcaa ggctcgcgc gcgaccgcgc agcggttg 6780
 ccttgacgc cctggaacga cccaagccta tgcgagtggt ggagtcgaa ggcaagccc 6840

gcccgctgc cccccgagac ctgcaggggg gggggggcgc tgaggtctgc ctctgaaga 6900
 aggtgttgc gactcatacc aggcctgaat cggcccatca tccagccaga aagtgaggga 6960
 gccacggttg atgagagctt tgtttaggt ggaccagttg gtgatttga acttttgctt 7020
 tgccacggaa cggctctcgt tgtcgggaag atgcgtgac tgatccttca actcagcaaa 7080
 agttcgattt attcaacaaa gccgcgtcc cgtcaagtc gcgtaatgct ctgccagtgt 7140
 tacaaccaat taaccaattc tgattagaaa aactcatcga gcatcaaatg aaactgcaat 7200
 ttattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga 7260

 gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg 7320
 actcgtcaa catcaatata acctattaat tccccctgt caaaaataag gttatcaagt 7380
 gagaaatcac catgagtac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct 7440
 ttccagactt gttaacagg ccagccattt cgtcgtcat caaaatcact cgcataacc 7500
 aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgac gctgttaaaa 7560
 ggacaattac aaacaggaat cgaatgcaac cggcgcagga aactgccag cgcataaca 7620
 atattttcac ctgaatcagg atattcttct aatacctgga atgctgtttt cccggggatc 7680

 gcagtgtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcgggaaga 7740
 ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 7800
 ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttccata caatcgatag 7860
 attgtcgac ctgattgcc gacattatcg cgagccatt tataccata taaatcagca 7920
 tccatgttg aatttaacg cggcctcgag caagacgttt cccgttgaat atggctcata 7980
 acaccccttg tattactgtt tatgtaagca gacagtttta ttgttcatga tgatatattt 8040
 ttatcttgt caatgtaaca tcagagattt tgagacacaa cgtggcttcc cccccccc 8100

 ctgcaggtcc cgagcctcac ggcgcgaggt gcgggggttc caagggggca gcgccacctt 8160
 gggcaaggcc gaaggccgag cagtcgatca acaagccccg gagggggcac ttttgccgg 8220
 agggggagcc gcgccgaagg cgtgggggaa cccgcaggg gtgcccttct ttgggcacca 8280
 aagaactaga tataggcgaa aatgcgaaag acttaaaaat caacaactta aaaaaggggg 8340
 gtacgcaaca gtcattcgc gcacccccg caatagctca ttgcgtaggt taaagaaat 8400
 ctgtaattga ctgccacttt tacgcaacgc ataattgttg tcgcgtgcc gaaaagttgc 8460
 agctgattgc gcatggtgcc gcaaccgtgc ggcaccctac cgcatggaga taagcatggc 8520

 cagcagtc agagaaatcg gcattcaagc caagaacaag cccggtcact ggtgcaaac 8580
 ggaacgcaaa gcgcatgagg cgtgggccgg gcttattgcg aggaaccca cggcggaat 8640

gctgctgcat cacctcgtgg cgcagatggg ccaccagaac gccgtggtgg tcagccagaa 8700
gacactttcc aagctcatcg gacgttcttt gcggacggtc caatacgag tcaaggactt 8760
ggtggccgag cgctggatct ccgtcgtgaa gctcaacggc cccggcaccc tgtcggccta 8820
cgtggtaaat gaccgcgtgg cgtggggcca gccccgcgac cagttgcgcc tgtcgggtgt 8880
cagtgccgcc gtggtggttg atcacgacga ccaggacgaa tcgtgtttgg ggcattggcga 8940

cctgcgccgc atcccgaccc tgtatccggg cgagcagcaa ctaccgaccg gccccggcga 9000
ggagccgccc agccagcccg gcattccggg catggaacca gacctgccag ccttgaccga 9060
aacggaggaa tgggaacggc gcgggcagca ggcctgccg atgccgatg agccgtgttt 9120
tctggacgat ggagagccgt tggagccgcc gacacgggtc acgtgccgc gccggtagca 9180
cttgggttgc gcagcaaccc gtaagtgcgc tgttccagac tatcggtgt agccgcctcg 9240
ccgcctata ccttgtctgc ctccccgct tgcgtcgcgg tgcatggagc cgggccacct 9300
cgacctgaat ggaagcc 9317

<210> 79

<211> 9317

<212> DNA

<213> Artificial Sequence

<220><223> Plasmid

<400> 79

tcgaggaatt cgcaggaccg tgatacacgg gacaggtcac tgaatgacga caatgtcctg 60
gaaatcagcg aaccgcgcat ctgaagtaca tttagcgac tgtaccagaa catgaatgag 120
gcgtttggat taggcgatta ttagcagggc taagcathtt actattatta ttttcgggtt 180
gagggatata gagctatcga caacaaccgg aaaaagtta cgtctatatt gctgaaggta 240
caggcgtttc cataactatt tgctcgcgtt ttttactcaa gaagaaaatg ccaaatagca 300
acatcaggca gacaataccc gaaattgcga agaaaactgt ctggtagcct gcgtggtcaa 360

agagtatccc agtcggcgtt gaaagcagca caatccaag cgaactggca atgtgaaaac 420
caatcagaaa gatcgtcgac gacaggcgct tatcaaagt ttgccacgtg tatttgaaga 480
cggatatgac acaaagtga acctcaatgg catgtaacaa cttcactaat gaaataatcc 540
aggggttaac gaacagcgcg caggaaagga tacgcaacgc cataatcaca actccgataa 600
gtaatgcatt ttttggccct acccgattca caaagaaagg aataatcgcc atgcacagcg 660
cttcgagtac cacctggaat gagttgagat aaccatacag gcgcgttcct acatcgtgtg 720
attcgaataa acctgaataa aagacaggaa aaagtgttg atcaaaaatg ttatagaaag 780

accacgtccc cacaataaat atgacgaaaa ccagaaagtt tcgatccttg aaaactgcga 840
 taaaatcctc tttttttacc cctcccgcac ctgccgttac gcactgggtga tccttatctt 900
 taaaacgcac gttgatcatc ataaatacag cgccaaatag cgagaccaac cagaagtiga 960
 tatggggact gatactaaaa aatatgccgg caaagaacgc gccaatagca tagccaaaag 1020
 atccccaggc gcgcgtgtt ccatattcga aatgaaaatt tcgcgccatt ttttcggtga 1080
 agctatcaag caaacgcac cccgccagat accccaagcc aaaaaatagc gccccagaa 1140
 ttagacctac agaaaaattg ctttcagta acggttcata aacgtaaatc ataaacggtc 1200

cggtcaagac caggatgaaa ctcatacacc agatgagcgg tttcttcaga ccgagtttat 1260
 cctgaacgat gccgtagaac atcataaata gaatgctggg aaactgggtg accgaataaa 1320
 gtgtacctaa ttccgtccct gtcaacccta gatgtccttt cagccaaata gcgtataacg 1380
 accaccacag cgaccaggaa ataaaaaga gaaatgagta actggatgca aaacgatagt 1440
 acgcatttct gaattgaata ttcagtcca taattacctg cctgtcgtta aaaaattcac 1500
 gtctatatta gagataagag cgacttcgcc gtttacttct cactattcca gtctttgtcg 1560
 acatggcagc gctgtcattg cccctttcgc cgttactgca agcgtccgc aacgttgagc 1620

gagatcgata attcgtcgca tttctctctc atctgtagat aatcccgtag aggacagacc 1680
 tgtgagtaac ccggcaacga acgcattctc cgccccgtg ctatcgacac aattcacaga 1740
 cattccagca aaatggtgaa cttgtcctcg ataacagacc accaccctt ctgcaccttt 1800
 agtcaccaac agcatggcga tctcatactc ttttgccagg gcgcataat cctgatcgtt 1860
 ctgtgttttt cactgataa gtgcgccattc ttttccgag agcttgacga catccgccag 1920
 ttgtagcgc tgcgcgaac acaagcggag caaatgctcg tcttgccata gatcttcacg 1980
 aatattagga tcgaagctga caaacctcc ggcatgccgg atcgccgtca tcgcagtaaa 2040

tgcgtggta cgcaaggct cggcagacaa cgcaattgaa cagagatgta accattcgcc 2100
 atgtcgccag cagggaagt ctgtcgtctc taaaaaaga tcggcactgg ggccgacat 2160
 aaacgtaaat gaacgttccc cttgatcgtt cagatcgaca agcaccgtgg atgtccggtg 2220
 ccattcatct tcttcagat acgtgatac gactccctca gttagcagcg ttttttgcac 2280
 taacgcacca aaaggatcat cccccccg acctataaac ccactgttc cgcctaattc 2340
 ggcgattccc accgcaacgt tagctggcgc gccgccagga caaggcagta ggccccgtc 2400
 tgattctggc aagagatcta cgaccgcac ccctaaaacc catactttgg ctgacatttt 2460

tttcccttaa attcatcga gttacgcata gtgataaacc tctttttcgc aaaatcgta 2520
 tggatttact aaaacatgca tattcgatca caaacgtca tagttaacgt taacatttgt 2580
 gatattcatc gcatttatga aagtaaggga ctttatTTTT ataaaagta acgttaacaa 2640

ttcaccaaatt ttgcttaacc aggatgattt aaatgacgca atctcgattg catgcggcgc 2700
 aaaacgccct agcaaaactt catgagcacc ggggtaacac tttctatccc cattttcacc 2760
 tcgcgccctc tgcgggtgg atgaacgatc caaacgccct gatctggttt aacgatcggt 2820
 atcacgcgtt ttatcaacat catccgatga gcgaacactg ggggccaatg cactggggac 2880

 atgccaccag cgacgatatg atccactggc agcatgagcc tattgcgcta gcgccaggag 2940
 acgataatga caaagacggg tggttttcag gtagtgctgt cgatgacaat ggtgtcctct 3000
 cacttatcta caccggacac gtctggctcg atggtgcagg taatgacgat gcaattcgcg 3060
 aagtacaatg tctggctacc agtcgggatg gtattcatit cgagaaacag ggtgtgatcc 3120
 tcactccacc agaaggaatc atgcacttcc gcgacctaag agtgtggcgt gaagccgaca 3180
 catggtggat ggtagtcggg gcgaaagatc caggcaacac ggggcagatc ctgctttatc 3240
 gcggcagttc gttagcgtgaa tggacattcg atcgctact ggcccacgct gatgcgggtg 3300

 aaagctatat gtgggaatgt ccggactttt tcagccttgg cgatcagcat tatctgatgt 3360
 tttccccga gggaatgaat gccgagggat acagttaccg aaatcgcttt caaagtggcg 3420
 taatacccg aaatgtgtcg ccaggacgac tttttgcaca atccgggcat tttactgaac 3480
 ttgataacgg gcatgacttt tatgcaccac aaagcttttt agcgaaggat ggtcggcgta 3540
 ttgttatcgg ctggatggat atgtgggaat cgccaatgcc ctcaaaacgt gaaggatggg 3600
 caggctgcat gacgctggcg cgcgagctat cagagagcaa tggcaaacct ctacaacgcc 3660
 cggtagacga agctgagtcg ttacgccagc agcatcaatc tgtctctccc cgcaaatca 3720

 gcaataaata tgttttgcag gaaaacgcgc aagcagttga gattcagttg cagtgggcgc 3780
 tgaagaacag tgatgccgaa cattacggat tacagctcgg cactggaatg cggctgtata 3840
 ttgataacca atctgagcga ctgtttttgt ggcggtatta cccacacgag aatttagacg 3900
 gctaccgtag tattccctc ccgcagcgtg acacgctcgc cctaaggata tttatcgata 3960
 catcatccgt ggaagtattt attaacgacg gggaagcggg gatgagtagt cgaatctatc 4020
 cgagccaga agaacgggaa ctgtcgcttt atgcctcca cgagtggtgt gtgctgcaac 4080
 atggagcact ctggctactg ggttaacata atatcaggtg gaacaacgga tcaacagcgg 4140

 gcaagggatc cacgaagctt cccatggtga cgtcaccggt aaaccagcaa tagacataag 4200
 cggctattta acgacctgc cctgaaccga cgaccgggtc gaatttgctt tcgaatttct 4260
 gccattcatc cgcttattat acttattcag gcgtagcacc aggcgtttta gggcaccaat 4320
 aactgcctta aaaaaattac gccccgccct gccactcatc gcagtactgt tgtaattcat 4380
 taagcattct gccgacatgg aagccatcac agacggcatg atgaacctga atcgccagcg 4440
 gcatcagcac ctgtgcct tgcgtataat atttgcccat ggtgaaaacg gggcggaaga 4500

agttgtccat attggccacg tttaaatcaa aactggtgaa actcaccag ggattggctg 4560

agacgaaaaa catattctca ataaaccctt tagggaaata ggccaggttt tcaccgtaac 4620

acgccacatc ttgcgaatat atgtgtagaa actgccggaa atcgtcgtgg tattcactcc 4680

agagcgatga aaacgtttca gtttgcctcat ggaaaacggg gtaacaaggg tgaacactat 4740

cccatatcac cagctcaccg tctttcattg ccatacggaa ttccggatga gcattcatca 4800

ggcgggcaag aatgtgaata aaggccggat aaaacttggt cttatttttc ttacgggtct 4860

ttaaaaaggc cgtaatatcc agctgaacgg tctggttata ggtacattga gcaactgact 4920

gaaatgcctc aaaatgttct ttacgatgcc attgggatat atcaacgggtg gtatatccag 4980

tgattttttt ctccatttta gcttccttag ctctgaaaa tctcgataac tcaaaaaata 5040

cgcccggtag tgatcttatt tcattatggt gaaagttgga acctcttacg tgccgatcaa 5100

cgtctcattt tcgcaaaaag ttggcccagg gcttcccggg atcaacaggg acaccaggat 5160

ttattttatc tgcgaagtga tcttcctca caggatatta ttccggcgaagg ggcctcgt 5220

gatacgcta tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtag 5280

cacttttcgg ggaatgtgc gcgcccggt tctgtctggc gctgggcctg tttctggcgc 5340

tggacttccc gctgttccgt cagcagcttt tcgccacagg ccttgatgat cgccggcgcc 5400

ttggcctgca tatcccgatt caacggcccc agggcgcca gaacgggctt caggcgctcc 5460

cgaaggcttc gggccgtctc ttgggcttga tcggccttct tgcgcatctc acgcgtcct 5520

gcggcgccct gtagggcagg ctcatacccc tgccgaaccg cttttgtcag ccggtcggcc 5580

acggcttcgc gcgtctcaac gcgctttgag attccagct tttcgccaa tccctgcggt 5640

gcataggcgc gtggctcgac cgcttgcggg ctgatggtga cgtggccac tggtagccgc 5700

tccaggcct cgtagaacgc ctgaatgcgc gtgtgacgtg ccttgctgcc ctgatgcc 5760

cgttgagcc ctagatcgcc cacagcgcc gcaaactgg tctggtcgcg ggtcatctgc 5820

gctttgttc cgatgaactc cttggccgac agcctgccgt cctgcgtcag cggcaccacg 5880

aacgcggtca tgtcggggt ggtttcgtca cgttgatgc tggccgtcac gatgcgatcc 5940

gccccgtact tgtcggccag ccacttgtgc gccttctga agaagccgc ctgctgttct 6000

tggctggcgc acttccacca ttccgggctg gccgtcatga cgtactcgac cgccaacaca 6060

gcgtccttgc gccgttctc tggcagcaac tcgcgcagtc ggcccatcgc ttcacgggtg 6120

ctgctggcgc ccagtgctc gttctctggc gtctgtctgg cgtcagcgtt gggcgtctcg 6180

cgctcgggt aggcgtgctt gagactggcc gccacgttc ccattttcgc cagcttcttg 6240

catcgcatga tcgcgtatgc cgccatgcct gccctccct tttggtgtcc aaccggctcg 6300

acgggggcag cgcaaggcgg tgcctccggc gggccactca atgcttgagt atactcacta 6360
gactttgctt cgcaaagtcg tgaccgccta cggcggctgc ggcgccttac gggcttgctc 6420
tccgggcttc gccctgcgcg gtgcctgcgc tcccttgcca gcccgtagat atgtggacga 6480
tggccgcgag cggccaccgg ctggctcgct tcgctcggcc cgtggacaac cctgctggac 6540
aagctgatgg acaggctgcg cctgcccacg agcttgacca cagggtatgc ccaccggcta 6600
cccagccttc gaccacatac ccaccggctc caactgcgcg gcctgcggcc ttgccccatc 6660

aattttttta attttctctg gggaaaagcc tccggcctgc ggctgcgcg ctctgcttgc 6720
cggttggaca ccaagtggaa ggcgggtcaa ggctgcgca ggcaccgcgc agcggcttgg 6780
ccttgacgcg cctggaacga cccaagccta tgcgagtggg ggagtcgaa ggcaagccc 6840
gcccgcctgc ccccgagac ctgcaggggg gggggggcgc tgaggtctgc ctctgaaga 6900
aggtgttctt gactcatacc aggctgaat cgcctcatca tccagccaga aagtgggga 6960
gccacggttg atgagagctt tgtttaggt ggaccagttg gtgatttga acttttctt 7020
tgccacggaa cggctctcgt tgtcgggaag atcgctgac tgatccttca actcagcaaa 7080

agttcgattt attcaacaaa gccgccgtcc cgtcaagtca gcgtaatgct ctgccagtgt 7140
tacaaccaat taaccaatc tgattagaaa aactcatcga gcatcaaag aaactgcaat 7200
ttattcatat caggattatc aataccatat ttttgaaaa gccgtttctg taatgaagga 7260
gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg 7320
actcgtccaa catcaatata acctattaat tccccctcgt caaaaataag gttatcaagt 7380
gagaaatcac catgagtac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct 7440
ttccagactt gttcaacagg ccagccatta cgtcgtcat caaaatcact cgcacaaacc 7500

aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgac gctgttaaaa 7560
ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcacaaaca 7620
atattttcac ctgaatcagg atattcttct aatactgga atgctgtttt cccggggatc 7680
gcagtgggtg gtaacctgc atcatcagga gtacggataa aatgcttgat ggtcggaaga 7740
ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg 7800
ctaccttgc catgtttcag aaacaactct ggcgcacggt gcttccata caatcgatag 7860
attgtcgac ctgattgccc gacattatcg cgagccatt tataccata taaatcagca 7920

tccatgttgg aatttaacg cggcctcgag caagacgttt cccgttgaat atggctcata 7980
acacccttg tattactgtt tatgtaagca gacagtttta ttgttcatga tgatatattt 8040
ttatcttgtg caatgtaaca tcagagattt tgagacacaa cgtggcttcc ccccccccc 8100
ctgcaggtcc cgagcctcac ggcggcgagt gcgggggttc caagggggca gcgccacctt 8160

gggcaaggcc gaaggccgcg cagtcgatca acaagcccg gaggggccac tttttgccgg 8220
agggggagcc gcgccgaagg cgtgggggaa ccccgaggg gtgcccttct ttgggcacca 8280
aagaactaga tatagggcga aatgcgaaag acttaaaaat caacaactta aaaaaggggg 8340

gtacgcaaca gctcattgcg gcaccccccg caatagctca ttgcgtaggt taaagaaaat 8400
ctgtaattga ctgccacttt tacgcaacgc ataattgttg tcgcgtgcc gaaaagtgc 8460
agctgattgc gcatggtgcc gcaaccgtgc ggcaccctac cgcatggaga taagcatggc 8520
cacgcagtcc agagaaatcg gcattcaagc caagaacaag cccggtcact ggggtgcaaac 8580
ggaacgcaaa gcgcatgagg cgtggggcgg gcttattgcg aggaaacca cggcggcaat 8640
gctgctgcat cacctcgtgg cgcagatggg ccaccagaac gccgtggtgg tcagccagaa 8700
gacactttcc aagctcatcg gacgttcttt gcggacggtc caatacgag tcaaggactt 8760

ggtggccgag cgctggatct ccgtcgtgaa gctcaacggc cccggcaccg tgcggccta 8820
cgtggtcaat gaccgcgtgg cgtggggcca gccccgcgac cagttgcgcc tgcggtgtt 8880
cagtgccgcc gtggtggttg atcacgacga ccaggacgaa tcgctgttgg ggcattggcga 8940
cctgcgccgc atcccgacc tgtatccggg cgagcagcaa ctaccgaccg gccccggcga 9000
ggagccgccc agccagcccc gcattccggg catggaacca gacctgccag ccttgaccga 9060
aacggaggaa tgggaacggc gcgggcagca gcgcctgccg atgcccgatg agccgtgttt 9120
tctggacgat ggcgagccgt tggagccgcc gacacgggtc acgtgccgc gccggtagca 9180

cttgggttgc gcagcaacc gtaagtgcgc tgttccagac tatcggtgt agccgcctcg 9240
ccgcctata ccttgtctgc ctccccgct tgcgtcgcgg tgcatggagc cgggccacct 9300
cgacctgaat ggaagcc 9317

<210> 80

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 80

cgtctaccct tgttatacct cacaccgcaa ggagacgatc atgaccaata atcccccttc 60

agcacagatt aagccccggc gtgtaggctg gagctgcttc 100

<210> 81

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 81

gcatcaggca atgaataccc aatgcgacca gcttcttata tcagaacagc cccaacggtt 60
tatccgagta gctcaccagc catatgaata tcctccttag 100

<210> 82

<211> 22

<212> DNA

<213> Artificial sequence

<220><223> Primer

<400> 82

atgaccaata atcccccttc ag 22

<210> 83

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 83

gcttcttata tcagaacagc c 21

<210> 84

<211> 921

<212> DNA

<213> Escherichia coli

<400> 84

atgagcgcaa gagtatgggt actcggatgat gcggttggtg atttattacc cgaaagccag 60
gggagactac tacagtgtcc tggcggggcg cctgctaata ttgcagtcgg tatcgcaagg 120
ctggggggga aaagtgcctt tattggcaaa gttggcgatg atcctttcgg tcgctttatg 180
tatcagacac tgagtacaga aaatgttgat acacattata tgtctcttga tcctcaacaa 240
cgcacctcaa ttgtggctgt aggacttgat gagcaaggag aaagaaactt tacctttatg 300
gtacgcccaa gtgccgatct ttttttacia cctggtgacc ttcctgcatt tgggccgggt 360

gaatggctcc atctttgttc cattgcgctc agtgcagaac cttcccgaag taccgcattt 420
ctggctatgg agaaaatagc tcaggctggc ggaaacatca gttttgatcc caatatccgc 480
agcgatctct ggcagagtga agcgtatata aggaaatacc ttgatcgcgc actttcgctg 540
gcgaatatcg ctaaattgtc cgaagaagag ttgctattca tcagtggcga aagccagggt 600

cagcaaggcg catattcatt agtacaacgt tattcgttga ctttattgct tattacacaa 660
 ggaaaaaatg gcgtacttgt gtattttcag gggcagttta tccactatcc cgccaaacct 720
 gtttctgtcg tcgatacgac cggggcagga gatgcttttg tcgctggatt acttgcaggt 780

ctggctgatt ctggaatacc aacaaatacc agacagcttg aacgaatcat tgcacaagct 840
 cagatttgtg gtgctctggc gaccacggct aaaggcgcga taaccgcctt accccgacaa 900
 cacgatctcc cttcacaata g 921

<210> 85

<211> 306

<212> PRT

<213> Escherichia coli

<400> 85

Met Ser Ala Arg Val Trp Val Leu Gly Asp Ala Val Val Asp Leu Leu
 1 5 10 15
 Pro Glu Ser Gln Gly Arg Leu Leu Gln Cys Pro Gly Gly Ala Pro Ala
 20 25 30

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

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

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

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

Thr Ala Lys Gly Ala Ile Thr Ala Leu Pro Arg Gln His Asp Leu Pro
290 295 300

Ser Gln

305

<210> 86

<211> 924

<212> DNA

<213> *Klebsiella pneumoniae*

<400> 86

atgaatggaa aaatctgggt actcggcgat gcggtcgtcg atctcctgcc cgatggagag	60
ggccgcctgc tgcaatgccc cggcggcgcg ccggccaacg tggcggtcgg cgtggcgcg	120
ctcggcggtg acagcgggtt tatcggccgc gtcggcgacg atcccttcgg cgttttatg	180
cgtcacacc tggcgaggaa gcaagtggat gtgaactata tgcgcctcga tgcggcgag	240
cgcacctcca cggtggtggt cgatctcgat agccacgggg agcgacatt tacctttatg	300
gtccgtccga gcgccgacct gtctcttcag cccgaggatc tcccgcggtt tgccgcgggt	360
cagtggctgc acgtctgctc catcgtctc agcgcggagc cgagccgcag cagcacattc	420

gcggcgatgg aggcgataaa gcgcgccggg ggctatgtca gcttcgaccc caatatccgc 480
 agcgacctgt ggcaggatcc gcaggacctt cgcgactgtc tcgaccgggc gctggccctc 540
 gccgacgcca taaaactttc ggaagaggag ctggcgttta tcagcggcag cgacgacatc 600
 gtcagcggca ccgcccggct gaacgccgcg ticcagccga cgctactgct ggtgaccag 660

ggtaaagcgg gggtcaggc cgccctgcgc gggcaggtta gccacttccc tgcccggccg 720
 gtggtggccg tcgataccac cggcgccggc gatgcctttg tcgccgggct actcgccggc 780
 ctgcccgcc acggtatccc ggacaacctc gcagccctgg ctcccgacct cgcgctggcg 840
 caaacctgcg gcgcccctggc caccacggcc aaaggcgcca tgaccgcct gccctacagg 900
 gacgatcttc agcgctcgct gtga 924

<210> 87

<211> 307

<212> PRT

<213> Klebsiella pneumoniae

<400> 87

Met Asn Gly Lys Ile Trp Val Leu Gly Asp Ala Val Val Asp Leu Leu

1 5 10 15
 Pro Asp Gly Glu Gly Arg Leu Leu Gln Cys Pro Gly Gly Ala Pro Ala

20 25 30
 Asn Val Ala Val Gly Val Ala Arg Leu Gly Gly Asp Ser Gly Phe Ile

35 40 45
 Gly Arg Val Gly Asp Asp Pro Phe Gly Arg Phe Met Arg His Thr Leu

50 55 60
 Ala Gln Glu Gln Val Asp Val Asn Tyr Met Arg Leu Asp Ala Ala Gln

65 70 75 80
 Arg Thr Ser Thr Val Val Val Asp Leu Asp Ser His Gly Glu Arg Thr

85 90 95
 Phe Thr Phe Met Val Arg Pro Ser Ala Asp Leu Phe Leu Gln Pro Glu

100 105 110
 Asp Leu Pro Pro Phe Ala Ala Gly Gln Trp Leu His Val Cys Ser Ile

115 120 125
 Ala Leu Ser Ala Glu Pro Ser Arg Ser Thr Thr Phe Ala Ala Met Glu

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

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

 260 265 270
 Leu Ala Pro Asp Leu Ala Leu Ala Gln Thr Cys Gly Ala Leu Ala Thr
 275 280 285
 Thr Ala Lys Gly Ala Met Thr Ala Leu Pro Tyr Arg Asp Asp Leu Gln
 290 295 300
 Arg Ser Leu
 305