

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(51) Int. Cl.⁶
C12N 1/00

(11) 공개번호 특2000-0037116
(43) 공개일자 2000년07월05일

| | | | |
|---------------|------------------|-------------|---------------|
| (21) 출원번호 | 10-1999-7001698 | | |
| (22) 출원일자 | 1999년03월02일 | | |
| 번역문제출일자 | 1999년03월02일 | | |
| (86) 국제출원번호 | PCT/US1997/07657 | (87) 국제공개번호 | W0 1998/08932 |
| (86) 국제출원출원일자 | 1997년05월05일 | (87) 국제공개일자 | 1998년03월05일 |

또는 그 내부로 유전공학처리될 수 있다.

명세서

관련 출원

본 출원은 1993년 5월 18일 출원된 미국 특허출원 제08/063,615호의 계속출원인 1995년 2월 28일 출원된 미국 특허출원 제08/395,947호의 계속출원인 1996년 2월 28일 출원된 미국 특허출원 제08/608,423호의 계속출원인 1996년 8월 28일 출원된 미국 특허출원 제08/705,484호의 계속출원인 1999년 11월 6일 출원된 미국 특허출원 제08/743,699호의 계속출원이다. 본 출원은 또한 1995년 11월 6일 출원된 가출원인 미국 특허출원 제60/007,255호의 계속출원이다.

기술분야

본 발명은 박테리아로부터 단리된 독소 및 살충제로서의 그의 용도에 관한 것이다.

배경기술

많은 곤충들이 주택소유자, 소풍객, 정원사 및 농부, 및 종종 발작물에 대한 곤충 피해의 결과로 파괴되거나 감소되는 농산물에 투자하는 사람들에게 광범위하게 해충으로 여겨진다. 특히, 재배철이 짧은 지역에서는, 상당한 곤충 피해는 재배자에게 모든 이익의 손실을 의미하고 농작물 수확의 극적인 감소를 의미할 수 있다. 특정 농산물의 공급이 부족하면, 반드시 식품 가공업자들이 더 높은 비용을 치르고, 따라서 식용 식물 및 그러한 식물로부터 얻은 생산물의 궁극적인 소비자까지 더 높은 비용을 치러야 한다.

농작물 및 화훼에 대한 곤충 피해를 방지하고 곤충 해충의 폐해를 제거하는 것은 전형적으로 광범위한 독성을 갖는 강한 유기 살충제 및 살충제에 의존해왔다. 이들 합성 제품은 환경과 그러한 약제에 노출되는 사람에게 너무 해로우므로 일반 대중에 의해 공격을 받게 되었다. 유사하게, 비경작 환경에서, 주택소유자는 그들의 가정 또는 야외 식사에서 곤충을 죽일 필요없이 곤충을 피하는 것에만 만족할 수 있다.

화학 살충제의 광범위한 사용은 농부, 살충제를 생산하는 회사, 정부 기관, 공공 이익 단체 및 일반적인 대중의 환경 및 건강에 대한 관심을 증가시켰다. 환경에 대한 사회적인 관심 및 곤충 관리의 메카니즘을 개발하는 생물학적 도구의 개발에 의해 덜 침해적인 곤충 관리 방법의 개발이 자극되었다. 생물학적 방제제가 화학 살충제에 대한 유망한 대체물로 제시된다.

각각의 진화 발전 단계에서 생물체는 그들 자신의 성공 및 생존을 강화시키기 위한 수단을 발견하였다. 방어 및 공격의 도구로서의 생물학적 분자의 사용이 동물계 및 식물계 전체를 통하여 공지되어 있다. 또한, 유전 공학의 상대적으로 새로운 도구에 의해 생물학적 살충제를 변형하여 특정한 문제에 대한 특정한 해결을 성취할 수 있게 되었다.

그러한 약제 중 하나인 바실루스 터린기엔시스(Bt: *Bacillus thuringiensis*)는 유효한 살충제로 상업적으로 널리 사용된다. 실제로, Bt 박테리아의 살충제는 그러한 제한된 독성을 갖는 단백질이며, 수확기에도 사람의 식용 작물에 사용될 수 있다. 비표적의 생물체에 대해, Bt 독소는 소화 가능한 비독성 단백질이다.

다른 공지된 종류의 생물학적 곤충 방제제는 살충 박테리아 공생자에 대해 전달 벡터로 공지된 특정한 속의 선충류이다. 살충성 박테리아를 포함하는 선충류는 곤충의 유충을 공격한다. 이어서, 박테리아가 유충을 죽인다. 선충류는 유충의 시체에서 번식한다. 이어서, 선충류 후대는 내부로부터 시체를 먹는다. 이어서, 이렇게 생산된 박테리아 함유 선충류 후대는 또 다른 유충을 공격할 수 있다.

과거에는, 스테이너네마 (*Steinernema*) 속 및 헤테로라피디티스 (*Heterorhabditis*) 속의 살충성 선충류가 곤충 방제제로서 사용되었다. 명백하게, 이들 속의 선충류는 각각 특정 종의 박테리아를 숙주로 한다. 헤테로라피디티스 속의 선충류에서, 공생 박테리아는 포토라프두스 루미네센스이다.

이들 선충류가 곤충 방제제로서 유효하지만, 현재 곤충 방제용 선충류를 생산하고 유지하고 분배시키는 것은 어렵고, 고가이며 비효율적이다.

나비목 및 딱정벌레목 곤충의 유충에 주사되는 경우에만 활성을 갖는 포토라프두스 루미네센스로부터 살충성 독소를 단리할 수 있음이 당업계에 공지되어 있다. 이로 인해, 선충류 또는 그의 박테리아 공생자의 살충 특성을 효과적으로 개발할 수 없었다. 전달후에 그의 생물학적 특성을 보유할, 살충성 독소의 더 실용적이고, 덜 노동집약적인 광범위한 전달 방법이 유용할 것이다. 포토라프두스 속에 의해 생산된 경구 활성을 갖는 독소를 발견하는 것이 매우 요망되었다. 효능의 이유로 이들 독소의 단리 및 사용이 요망되었다. 본 출원인의 발명에 와서야, 이들 독소가 단리되고 특성화되었다.

<발명의 요약>

천연 독소는 포토라프두스 속의 성장하는 박테리아 세포에 의해 생산되고 분비되는 단백질 복합체이며, 관심있는 것은 포토라프두스 루미네센스 종에 의해 생산된 단백질이다. 약 1,000 kDa의 분자 크기를 갖는 단백질 복합체를 SDS-PAGE 겔 분석에 의해 많은 성분 단백질로 분리할 수 있다. 독소는 헤몰리신, 리파제, 유형 C 포스포리파제 또는 뉴클레아제 활성을 갖지 않는다. 독소는 많은 곤충에 노출 투여시에 상당한 독성을 나타낸다.

본 발명은 쉽게 투여되는 살충성 단백질 뿐만 아니라 이종계내에서의 독소의 발현을 제공한다.

본 발명은 또한 많은 목의 곤충에 대하여 작용 활성이고 유효한 살충성 독소를 전달하기 위한 방법을 제공한다.

본 발명의 목적, 이점 및 특징은 하기 상세한 설명으로부터 명백해질 것이다.

도면의 간단한 설명

도 1은 본 발명의 독소를 위한 서열 유전자의 일부로서 사용된 한 쌍의 클론된 DNA 단리물의 도면이다.

도 2는 시퀀싱 방법에서 사용된 3개의 플라스미드의 지도이다.

도 3은 몇몇 부분 DNA 단편의 내부-관계를 도시하는 지도이다.

도 4는 TcbA_{ii} 및 TcaB_{ii} 단백질들의 단백질 서열 사이의 상동성 분석의 도면이다.

도 5는 포토랍두스 균주의 페노그램(phenogram)이다. 포토랍두스 균주의 관계는 rep-PCR에 의해 정의된다. 도 5의 상부축은 rep-PCR 생성물의 스코어링 (scoring)을 기준으로한 균주의 % 유사성을 측정한다(즉, 0.0 [유사성 없음] 내지 1.0 [100% 유사성]). 오른쪽 축에서, 숫자 및 문자는 시험된 다양한 균주를 나타낸다: 14=W-14, Hm=Hm, H9=H9, 7=W-7, 1=W-1, 2=W-2, 88=HP88, NC-1=NC-1, 4=W-4, 9=W-9, 8=W-8, 10=W-10, WIR=WIR, 3=W-3, 11=W-11, 5=W-5, 6=W-6, 12=W-12, 14=W-14, 15=W-15, Hb=Hb, B2=B2, 48 내지 52=ATCC 43948 내지 ATCC 43952. 수평선을 분리하는 수직선은 수평선의 기부에서 균주 또는 균주의 군 사이의 상관도(상부축과의 수직선의 외삽 교차점을 읽음)를 나타낸다(예를 들면, 균주 W-14는 균주 H9 및 Hm에 대해 약 60% 유사하다).

도 6은 W-14 균주의 계통 지도의 도면이다.

도 6A는 tca 및 tcb 좌위 및 주 유전자 산물을 나타낸다.

도 7은 rep-PCR에 의해 정의된 것과 같은 포토랍두스 균주의 페노그램 (phenogram)이다. 도 7의 상부축은 rep-PCR 생성물의 스코어링 (scoring)을 기준으로한 균주의 % 유사성을 나타낸다(즉, 0.0 [유사성 없음] 내지 1.0 [100% 유사성]). 오른쪽 축에서, 숫자 및 문자는 시험된 다양한 균주를 나타낸다. 수평선을 분리하는 수직선은 수평선의 기부에서 균주 또는 균주의 군 사이의 상관도(상부축과의 수직선의 외삽 교차점을 읽음)를 나타낸다(예를 들면, 균주 인디커스는 균주 MP1 및 HB 오스웨고에 대해 약 30% 유사하다). 페노그램상의 포토랍두스 균주는 다음과 같다: 14=W-14, Hm=Hm, H9=H9, 7=W-7, 1=W-1, 2=W-2, 88=HP88, NC1=NC-1, 4=W-4, 9=W-9, 8=W-8, 10=W-10, 30=W30, WIR=WIR, 3=W-3, 11=W-11, 5=W-5, 6=W-6, 12=W-12, 15=W-15, 14=W-14, Hb=Hb, B2=B2, 48=ATCC 43948, 49=ATCC 43949, 50=ATCC 43950, 51=ATCC 43951, 52=ATCC 43952.

발명의 상세한 설명

본 발명은 곤충에 대해 경구 독성을 갖는 포토랍두스 속 유래의 독특한 종류의 살충 단백질 독소의 발견에 관한 것이다. 포토랍두스의 독특한 특성은 그의 생물발광이다. 포토랍두스는 다양한 공급원으로부터 단리될 수 있다. 그러한 공급원 중 하나는 선충류, 더 특히 헤테로라피디스 속의 선충류이다. 그러한 공급원 중 다른 하나는 상처로부터의 사람 임상 샘플이다(파머 등의 문헌[Farmer 등, 1989, J. Clin. Microbiol. 27, pp 1594-1600]을 참조하십시오). 이들 부패 균주는 아메리칸 타입 컬처 컬렉션[American Type Culture Collection(Rockville, MD)]에서 ATCC# 43948, 43949, 43950, 43951, 및 43952로서 기탁되어 있고, 본원에서 참고로 포함하였다. 다른 공급원이 살충성 독소를 생산하는 포토랍두스 박테리아를 포함할 수 있다. 그러한 공급원의 환경은 육상이거나 수상일 수 있다.

상기 포토랍두스 속은 분류학적으로 엔테로박테리아세에 (Enterobacteriaceae) 과로 정의되지만, 이 과의 비전형적인 특정 특성을 갖는다. 예를 들면, 이 속의 균주는 질산염 환원 음성이고, 황색 및 적색 안료를 생산하며 생물발광성이다. 반면, 후자의 특성은 엔테로박테리아세에 내에서는 공지되어 있지 않다. 포토랍두스는 최근에 와서야 제노랍두스 (Xenorhabdus)로부터 분리된 속으로서 설명되었다(Boemare 등, 1993, Int. J. Syst. Bacteriol. 43, 249-255). 이러한 구별은 DNA-DNA 하이브리드화 연구, 표현형의 차이 (예를 들면, 카탈라제 및 생물발광의 존재(포토랍두스) 또는 부재(제노랍두스)), 및 선충류 숙주의 과 (제노랍두스: 스테이너네마티드(Steinernematidae), 포토랍두스: 헤테로라피디드(Heterorhabditidae))를 기준으로 한다. 세포질 지방산의 비교 분석(Janse 등, 1990, Lett. Appl. Microbiol. 10, 131-135; Suzuki 등, 1990, J. Gen. Appl. Microbiol., 36, 393-401)은 제노랍두스로부터 포토랍두스의 구별을 지지한다.

본원에 개시된 균주 컬렉션이 포토랍두스 균주로 이루어짐을 확정하기 위해, 균주를 포토랍두스를 정의하는 인식된 특성을 기준으로 특성화하고, 그를 다른 엔테로박테리아세에 및 제노랍두스 종으로부터 구별하였다. (Farmer, 1984, Bergey's Manual of Systemic Bacteriology, Vol.1, pp.510-511; Akhurst 및 Boemere 1988, J. Gen. Microbiol. 134, pp 1835-1845; Boemare 등, 1993, Int. J. Syst. Bacteriol. 43, pp 249-255, 본원에서 참고로 인용함). 연구된 특성은 하기와 같다: 그람 염색 음성 간균, 생물체 크기, 콜로니 색소형성, 봉입체, 카탈라제의 존재, 질산염 환원 능력, 생물발광, 염료 업테이크, 젤라틴 가수분해, 선택 배지 상의 성장, 성장 온도, 혐기 조건하의 생존 및 운동성. 지방산 분석은 본원의 균주가 모두 하나의 포토랍두스 속에 속함을 확정하기 위해 사용되었다.

현재, 박테리아 포토랍두스 속은 하나의 정의된 종인 포토랍두스 루미네센스 (ATCC 유형 균주 #29999, Poinar 등, 1977, Nematologica 23, 97-102)로 이루어진다. 다양한 관련 균주가 문헌에

기재되어 있다(예를 들면, Akhurst 등, 1988, J. Gen. Microbiol., 134, 1835-1845; Boemare 등, 1993, Int. J. Syst. Bacteriol. 43 pp. 249-255; Putz 등, 1990, Appl. Environ. Microbiol., 56, 181-186). 본원에서 많은 포토랍두스 균주를 특성화하였다. 그러한 균주를 실시예에서 표 18에 나열하였다. 현재 포토랍두스 속에 한 종(루미네센스)만이 정의되었으므로, 본원에서 균주를 특성화하기 위해 루미네센스 종의 특성을 사용하였다. 도 5에서 알 수 있는 바와 같이, 이들 균주는 매우 다양하다. 미래에, 루미네센스 종의 일부 속성 뿐만 아니라, 포토랍두스 루미네센스의 특성으로 현재에는 정의되지 않은 일부 다른 특성을 갖는 다른 포토랍두스 종이 있을 수 있음을 예측할 수 있다. 그러나, 본원에서 본 발명의 범위는 다른 특성 및 특성화를 고려하지 않고, 곤충 방제제로서 작용 활성을 갖는 단백질을 생산하는 모든 포토랍두스 종 또는 균주이다.

또한, 본원에서 증명된 바와 같이, 포토랍두스 속의 박테리아는 본원에서 정의된 작용 활성을 갖는 단백질을 생산한다. 포토랍두스 루미네센스 종에 의해 생산된 단백질에 특히 관심을 둔다. 본 발명은 본원에 개시된 균주로 제한되지 않아야 한다. 이들 균주는 처음으로 포토랍두스의 다양한 단리물에 의해 생산된 단백질이 곤충에 노출시에 유독함을 설명한다. 따라서, 본원에 열거된 균주 및 그의 임의의 돌연변이체 뿐만 아니라, 본원에 설명된 작용 활성을 갖는 포토랍두스 속의 임의의 균주 또는 종이 본원에서 설명된 본 발명에 포함된다.

본원에서 사용된 몇몇 용어는 하기와 같은 특정한 의미를 갖는다:

"작용 활성"은 본원에서 단백질 독소가 단백질이 경구적으로 활성인 곤충 방제제로서 작용하거나, 독성 효과를 갖거나, 음식 섭취를 방해하거나 중단시킬 수 있어, 곤충의 죽음을 유발시킬 수 있거나 유발시킬 수 없는 것을 의미한다. 곤충이 유전자전환 식물의 발현, 제형화된 단백질 조성물(들), 분무가능한 단백질 조성물(들), 미끼 매트릭스 또는 다른 전달계에 의해 전달된 독소의 유효량에 접촉하는 경우, 전형적으로 곤충의 죽음을 야기하거나, 곤충이 곤충에 대해 독소를 이용가능하게 하는 공급원을 섭취하지 못하게 한다.

본원에서 용어 "유전 물질"의 사용은 모든 유전자, 핵산, DNA 및 RNA를 포함하는 것을 의미한다.

본원에서 용어 "상동체"는 대조 W-14 독소 폴리펩티드 아미노산 서열에 대한 상동성을 갖는 것으로 식별된 아미노산 서열을 의미한다.

본원에서 용어 "상동성"은 B10sum 62 단백질 스코어링 매트릭스 (위스콘신 패키지 버전 9.0, 제너틱 컴퓨터 그룹 (GCG), 위스콘신주 매디슨)를 사용한 GAP 알고리즘을 사용하여 측정하였을 경우 대조 W-14 독소 폴리펩티드 아미노산 서열에 대하여 33 % 이상의 유사성 지수 및(또는) 26 % 이상의 동일성 지수를 갖는 아미노산 서열을 의미한다.

본원에서 용어 "동일성"은 GAP 알고리즘에 의한 대조 W-14 독소 폴리펩티드 아미노산 서열의 정렬 이후에 소정의 위치에서 동일한 잔기를 함유하는 아미노산 서열을 의미한다.

본원에서 논의된 단백질 독소는 전형적으로 "살충제"로서 지칭된다. 살충제는 본원에서 단백질 독소가 본원에서 더욱 설명될 바와 같은 "작용 활성"을 갖고 곤충 방제제로서 사용되는 것을 의미한다.

용어 "올리고뉴클레오티드"의 사용은 RNA 또는 DNA의 뉴클레오티드의 짧은 사슬로 이루어진 거대 분자를 의미한다. 그러한 길이는 적어도 하나의 뉴클레오티드일 수 있지만, 전형적으로 약 10 내지 약 12개의 뉴클레오티드의 범위내에 있다. 올리고뉴클레오티드의 길이의 측정은 숙련인의 기술 범위내에 잘 알려져 있고, 본원에서 제한되지 않는다. 따라서, 올리고뉴클레오티드는 10개 미만이거나 12개 이상일 수 있다.

용어 "포토랍두스 독소"는 곤충에 대해 작용 활성을 갖는 포토랍두스 미생물 균주에 의해 생산된 임의의 단백질을 의미하며, 포토랍두스 독소는 분무가능한 조성물로 제형화되거나, 유전자전환 식물에 의해 발현되거나, 미끼 매트릭스로서 제형화되거나, 바쿠로바이러스를 통해 전달되거나, 임의의 다른 이용가능한 숙주 또는 전달계에 의해 전달될 수 있다.

본원에서 용어 "유독" 또는 "독성"의 사용은 포토랍두스에 의해 생산된 독소가 본원에서 정의된 바와 같은 "작용 활성"을 갖는 것을 의미한다.

본원에서 용어 "절단 (truncated) 펩티드"는 작용 활성을 갖는 것으로 관찰된 펩티드의 단편인 모든 펩티드를 포함하는 것을 의미한다.

본원에서 용어 "실질적인 서열 상동성"은 유사한 생화학적 성질을 갖는 부분을 제조하기 위하여 또 다른 DNA 단편과 충분히 유사한 뉴클레오티드 서열을 갖는 DNA 단편, 또는 유사한 생화학적 성질을 나타내기 위한 또 다른 폴리펩티드와 충분히 유사한 아미노산 사열을 갖는 폴리펩티드를 의미한다.

표 20에 기록된 선택된 균주로부터의 발효 브로스는 포토랍두스 속에 의한 살충성 독소 생산의 폭 및 이들 독소의 살충 스펙트럼을 측정하기 위해, 및 독소 복합체를 정제하기 위한 공급원 물질을 제공하기 위해 사용되었다. 본원에서 특성화된 균주는 다양한 곤충 목에 대해 경구 독성을 갖는 것으로 나타났다. 그러한 곤충 목은 딱정벌레목(Coleoptera), 동시목(Homoptera), 나비목(Lepidoptera), 파리목(Diptera), 응애목(Acarina), 벌목(Hymenoptera) 및 방시목(Dictyoptera)을 포함하지만, 이에 한정되지 않는다.

다른 박테리아 독소에서와 같이, 군집내에서의 박테리아의 돌연변이율로 인해, 서열에서 약간 상이한 많은 관련 독소가 존재한다. 본원에서 관심있는 독소는 본원에서 설명된 바와 같이 노출시에 다양한 곤충에 유독한 단백질 복합체를 생산하는 것이다. 바람직하게는, 독소는 나비목, 딱정벌레목, 동시목, 파리목, 벌목, 방시목 및 응애목에 대해 활성이다. 본 발명은 본원의 균주 및 그의 임의의 변종 균주에 의해 생산된 단백질 독소에 상동성인 단백질 독소 뿐만 아니라, 포토랍두스에 의해 생산된 임의의 단백질 독소를 포함하도록 의도된다. 이들 상동성 단백질은 서열에서

는 다를 수 있지만, 작용에서는 본원에서 설명된 독소와 다르지 않다. 상동성 독소는 300 kDa 내지 2,000 kDa의 단백질 복합체를 포함하는 것을 의미하고, 적어도 2개의 서브유닛으로 이루어지며, 서브유닛은 다른 서브유닛과 동일하거나 동일하지 않은 펩티드이다. 다양한 단백질 서브유닛이 동정되었고, 본원에서 실시예에 교시되어 있다. 전형적으로, 단백질 서브유닛은 약 18 kDa 내지 약 230 kDa; 약 160 kDa 내지 약 230 kDa; 약 100 kDa 내지 약 160 kDa; 약 80 kDa 내지 약 100 kDa; 및 약 50 kDa 내지 약 80 kDa이다.

전술된 바와 같이, 일부 포토랍두스 균주가 선충류로부터 단리될 수 있다. 일부 선충류, 즉, 선충문(Nematoda)의 긴 원통형 기생충은 유리한 성장 환경으로서 곤충 유충을 이용하는 능력을 진화시켰다. 곤충 유충은 선충류를 성장시키는 먹이원 및 번식하기 위한 환경을 제공한다. 특정한 선충류가 유충을 침입함에 따른 한 극적인 효과가 유충의 죽음이다. 유충의 죽음은 특정한 선충류내의, 유충 성장을 정지시키고 먹이섭취 활동을 억제하는 살충성 독소를 생산하는 박테리아의 존재에 의한 것이다.

흥미롭게, 곤충 기생 선충류의 각각의 속은 선충류와의 공생적 성장을 위해 독특하게 적응된 특정한 종의 박테리아를 숙주로 하는 것으로 나타난다. 이러한 연구가 시작된 사이에, 박테리아 제노랍두스 속의 명칭이 제노랍두스 및 포토랍두스로 재분류되었다. 포토랍두스 속의 박테리아는 헤테로랍디투스 선충류의 공생자로서 특성화되며, 제노랍두스 종은 스테인너네마 종의 공생자이다. 명칭에서의 이러한 변경을 본 명세서에서 반영하지만, 명칭에서의 변경이 어떠한 방식으로든 본원에 설명된 발명의 범위를 변화시키지 않아야 한다.

본원에 개시된 펩티드 및 유전자는 문헌[the Journal of Bacteriology "Instructions to Authors" p. i-xii(Jan. 1996)]에서 최근에 공개된 지침에 따라 명명된다. 하기 펩티드 및 유전자는 포토랍두스 균주 W-14로부터 단리되었다.

[표 1]

| 펩티드/유전자 명칭독소 복합체 (Tc) | | | |
|---|---------------------------------|--------|-----------|
| 펩티드 명칭 | 펩티드 서열 식별 번호 | 유전자 명칭 | 유전자 서열 |
| tca 계놈 영역 | | | |
| TcaA | 34C | tcaA | 33 |
| TcaAi | 프로펩티드 | tcaA | - |
| TcaAii | [15]a, 34c | tcaA | - |
| TcaAiii | [4]a, 35c | tcaA | - |
| TcaAiv | [62]a | tcaA | - |
| TcaB | [3]a, (19,20)b, 26c | tcaB | 25 |
| TcaBi | [3]a, (19,20)b, 28c | tcaB | 27 |
| TcaBii | [5]a, 30c | tcaB | 29 |
| TcaC | [2]a, 32c | tcaC | 31 |
| tcb 계놈 영역 | | | |
| TcbA | 12c, [16]a, (21,22,23,24)b, 53c | tcbA | 11 |
| TcbAi | 프로펩티드 | tcbA | - |
| TcbAii | [1]a, (21,22,23,24)b, 53c | tcbA | 52 |
| TcbAiii | [40]a, 55c | tcbA | 54 |
| tcc 계놈 영역 | | | |
| TccA | [8]a, 57c | tccA | 56 |
| TccB | [7]a, 59c | tccB | 58 |
| TccC | 61c | tccC | 60 |
| tcd 계놈 영역 | | | |
| TcdA | (17,18,37,38,39,42,43)b, 47c | tcdA | (36)d, 46 |
| TcdAi | 프로펩티드 | tcdA | - |
| TcdAii | [13]a, (17,18,37,38,39)b, 49c | tcdA | 48 |
| TcdAiii | [41]a, (42,43)b, 51c | tcdA | 50 |
| TcdB | [14]a | tcdB | - |
| a: 대괄호안의 서열 식별번호는 펩티드 N-말단이고; b: 괄호안의 수는 내부 펩티드의 트립신 절단 단편의 N-말단이고; c: 유전자 서열로 부터의 추정이고; d: 내부 유전자 단편이다. | | | |

상기 목록의 서열은 계놈 영역에 의해 분류된다. 더욱 구체적으로는, 포토랍두스 루미네스센스 박테리아 (W-14)는 4개 이상의 구별되는 계놈 영역, tca, tcb, tcc 및 tcd를 갖는다. 표 1에서 알 수 있는 바와 같이, 펩티드 생성물은 이들 구별되는 계놈 영역에서 제조된다. 더욱이, 실시예, 구체적으로 실시예 15 및 21에서 설명되는 것처럼, 세 계놈 영역으로 부터 제조된 각각의 유전자 생성물은 조합하여 곤충 활성을 나타낸다. 또한, 이들 4개의 계놈 영역 사이에는 현저한 상동성이 있다.

실시예에서 자세하게 설명되는 것처럼, tcbA 유전자는 두 가능한 생물학적 활성 단백질 단편 (TcbA 및 TcbA_{ii/iii})으로서 이. 콜리에서 발현되었다. 또한, TcdA 유전자도 이. 콜리에서 발현되었다. 실시예 16에서 설명되는 것처럼, 천연의 비프로세싱된 TcbA 독소를 내생의 메탈로프로테아제

또는 프로테아제를 함유하는 곤충의 소화관 내용물로 처리하면, TcbA 단백질 독소는 더 작은 서브 유닛으로 프로세싱되어 천연의 펩티드보다 크기가 작아지고, 써던 옥수수 뿌리벌레 활성이 증가한다. 더 작은 독소 펩티드는 독소 복합체의 부분으로서 잔류하게 된다. 몇몇 상황에서 단백질 분해적 프로세싱 또는 절단 펩티드를 사용하여 독소의 활성화를 증가시키는 것이 소망될 수 있다. 따라서, 몇몇 용도, 즉 상업적 유전자전환 식물 용도에서 절단 펩티드를 사용하는 것이 더욱 바람직할 수 있다.

W-14 균주 이외에, 차별적인 작용 활성을 갖는 포도랍두스 속 내의 다른 종들이 있다 (구체적으로 표 20 및 36 참조). 차별적인 활성이 있을지라도, 몇몇 경우의 아미노산 서열은 실질적인 서열 상동성을 갖는다. 더욱이, 분자 프로브는 균주에 함유되어 있는 몇몇 유전자가 W-14 균주에 함유된 유전자와 상동성이 있다는 것을 나타낸다. 사실 본원에서 설명된 균주 모두는 W-14 독소 유전자에 대한 하나 이상의 상동체를 갖는다. 실시예 26에서의 항체 데이터 및 실시예 25의 N-말단 서열 데이터는 이들 균주에 의하여 생산된 단백질 독소 사이에 상동성 및 동일성 (아미노산 서열을 기준으로)이 있다는 결론을 더욱 지지해준다. 분자 수준에서, W-14 유전자 프로브는 포도랍두스 속 전체에 걸쳐 이들 상동체 또는 W-14 유전자들 (표 37, 38 및 39)이 존재한다는 것을 가리킨다. 더욱이, W-14와 상동성이 아니라, 전체적인 단백질 특성을 유지하는 다른 균주에서 새로운 독소 유전자가 존재하는 것도 가능하다 (참조. 구체적으로, 실시예 14 및 25).

포도랍두스 균주에 의하여 생성된 독소 유전자 사이에 상동성 또는 동일성이 존재할 지라도, 균주들 자체는 다양하다. 실시예 22에 설명된 폴리머라제 연쇄 증합반응 기법을 사용할 때, 본원에서 설명된 균주 대부분은 구별된다. 예를 들면, 도 5에서 알 수 있는 것처럼, HP88 및 NC-1과 같은 몇몇 균주에서의 상대 %유사성은 약 0.8로, 이는 상기 균주들이 유사하다는 것을 나타내고, HP88 및 Hb는 0.1 인데, 이는 실질적으로 다름을 나타낸다. 따라서, 곤충의 독소 유전자 또는 이들 균주가 생산하는 유전자 산물이 동일하거나 유사할지라도, 상기 균주들은 다른 종류이다.

실시예 및 본원에서 추가로 개시된 데이터면에서 신규하고 독특한 균의 살충 단백질 독소가 발견되었다는 것은 분명하다. 본원에서 이들 독소들이 포도랍두스 속의 박테리아 균주내에 넓게 존재한다는 것을 설명하고 있다. 또한, 이들 독소 유전자가 엔테로박테라카에 과내에 넓게 존재한다는 것을 보여준다. 실시예 21에 설명된 것처럼 제조된 항체 또는 실시예 25에서 설명된 것처럼 제조된 유전자 프로브가 작용 활성을 갖는 상동성의 독소를 생성하는 엔테로박테라카에 과내의 박테리아 균주를 추가로 스크리닝하기 위하여 사용될 수 있다. 또한, 포도랍두스 속 또는 엔테로박테라카에 과 내의 새로운 유전자를 식별하는 것을 용이하게 할 수 있는 특성의 프라이머가 존재하는 지를 확인하기 위하여 사용될 수 있다.

상기 설명한 것처럼, 포도랍두스 속 또는 엔테로박테라카에 과의 박테리아를 실시예 26에서 설명한 것처럼 상동성의 독소 생성물에 대해 신속히 스크리닝하기 위하여 사용될 수 있다. 본 분야의 숙련자들은 분석 또는 스크리닝 기법으로서 항체를 사용하는 것에 매우 익숙하다 (참조. 미국 특허 제5,430,137호가 본원에 참고로 도입된다). 더욱이, 폴리펩티드의 노출된 표면을 점유하는 경향이 있는 6 내지 20개의 아미노산 잔기 세그먼트에 의해 항체가 유도되는 것은 일반적으로 문헌 (Current Protocols in Immunology, Coligan et al, National Institutes of Health, John Wiley & Sons, Inc.)에서 설명된다. 일반적으로 아미노산은 인접하는 아미노산 잔기들로 이루어지지만, 특성의 경우에는 특성의 형태에 의하여 제한되는 인접하지 않는 아미노산에 의하여 형성되는 경우도 있다. 항체에 의하여 인식되는 아미노산 세그먼트는 매우 특이적이며, 일반적으로 에피토프로 간주된다. 아미노산 단편은 천연 단백질의 화학적 및(또는) 효소적 절단에 의하여, 자동화된 고체상 펩티드 합성에 의하여 또는 유전공학적으로 합성된 부서의 생성에 의하여 생성될 수 있다. 폴리펩티드 단편은 본 분야에 공지된 다양한 HPLC 또는 FPLC 크로마토그래피법 또는 그 조합에 의하여 분리될 수 있다. 폴리펩티드 단편의 선별은 예를 들면 Kyte 및 Doolittle의 1982의 문헌 (Journal of Molecular Biology 157: 105-132) 및 Chou 및 Fasman의 1974년의 문헌 (Biochemistry 13: 222-245)의 단백질의 표면에 노출된 서열을 나타내는 알고리즘에 의해 도움을 받을 수 있다. 유의한 폴리펩티드 단편을 함유하는 면역원의 제조를 위하여 일반적으로 화학적 반응물을 사용하여 유리 아미노 (라이신), 설프히드릴 (시스테인), 페놀기 (티로신) 또는 카르복실기 (아스파르트레이트 또는 글루탐레이트)를 통하여 키울 림펫 헤모시아닌과 같은 캐리어 단백질에 폴리펩티드가 공유적으로 결합된다. 보강제와 함께 면역원을 쥐 또는 토끼, 또는 면역원에 대하여 면역반응을 이끌어내는 닭에게 주사한다. 폴리펩티드 단편에 대한 주사 동물의 항혈청에서의 항체양의 분석은 ELISA 및 웨스턴 블롯과 같은 여러 면역학적 방법을 사용하여 측정될 수 있다. 이외에, 중앙 세포와의 융합이 단일 항체군을 생성하는 불멸의 히브리도마 세포를 제조하는 주입된 동물의 비장 세포를 사용하여 단클론성 항체를 제조될 수 있다. 히브리도마 세포를 면역학적 방법을 통하여 스크리닝하여 유의한 폴리펩티드 단편에 대한 특이 항체를 제조하는 세포주를 선별한다. 상이한 생물원으로 부터 항체의 정제는 다양한 항원 친화성 또는 항체 친화성 컬럼 또는 다른 크로마토그래피 HPLC 또는 FPLC 방법을 사용하여 수행될 수 있다.

본원에 설명된 독소는 독소가 작용 활성을 갖는 점에서 매우 독특하며, 이는 곤충 관리 방법을 개발하기 위한 핵심이다. 곤충 관리 방법을 개발하는데 있어서, 단백질을 소화관을 피하여 생물체에 직접 주사함으로써 단백질 분해 과정을 지연시키거나 회피할 수 있다. 그러한 경우, 생물체에 투여된 단백질은 변성되거나, 비특이적으로 분해되거나, 고등 생물체에서 면역계에 의해 제거되기 전까지 그의 기능을 유지할 것이다. 살충성 독소를 곤충에 주사하는 것은 단지 실험실에서, 및 그에 따라 쉽게 주사되는 큰 곤충에 대해서만 가능한 적용이다. 본원에 설명된 살충 단백질 독소가 경구 섭취 또는 독소와의 접촉 후에 그들의 유독 활성을 나타낸다는 관찰은 단백질 독소를 곤충의 먹이내로 혼입시키는 것에만 전적으로 기초된 곤충 관리 계획의 개발을 허용한다. 그러한 계획은 곤충 미끼의 생산을 이끌 수 있다.

포도랍두스 독소는 정제된 형태로 곤충에 투여될 수 있다. 독소는 또한 약 1 내지 약 100mg/ℓ 브로쓰의 양으로 전달될 수 있다. 이는 제형 조건, 접종 공급원의 조건, 독소의 단리 기술 등에 따라 변할 수 있다. 독소는 이종 원핵세포 또는 진핵세포 숙주에서 원래 발현된 산출 분비물 또는

는 세포내 단백질로서 투여될 수 있다. 박테리아는 전형적으로 단백질이 발현되는 숙주이다. 진핵세포 숙주는 식물, 곤충 및 효모를 포함할 수 있지만 이에 한정되지 않는다. 이외에, 독소는 박테리아 또는 들판의 유전자전환 식물에서 또는 바콜로바이러스 벡터에 의해 곤충내에서 생산될 수 있다. 전형적으로, 독소는 하나 이상의 독소를 곤충의 먹이에 혼입시킴으로써 곤충에 도입될 것이다.

완전 치사량이 곤충에 섭취시키기 위해 유용하지만, 유용한 독성을 성취하기위해서 요구되지는 않는다. 곤충이 독소를 회피하거나 먹이 섭취를 중단하면, 일부 적용에서는 효과가 거의 치사량에 가깝더라도, 그러한 회피가 유용할 것이다. 예를 들면, 내충 유전자전환 작물 식물이 요망되는 경우, 궁극적인 목적은 곤충을 죽이는 것보다 식물의 보호에 있으므로, 식물을 섭취하는 것에 대한 곤충의 거리낌이 곤충에 대한 치명적인 독성만큼 유용하다.

독소를 곤충의 먹이에 혼입할 수 있는 다른 많은 방법이 있다. 예로서, 본원에 개시된 바와 같이 단백질 용액을 먹이에 분무함으로써, 유충 먹이원을 독성 단백질로 혼합시킬 수 있다. 이외에, 정제된 단백질을 다른 무해한 박테리아내로 유전 공학적으로 처리한 다음, 배양물에서 성장시키고, 먹이원에 적용하거나 곤충 박멸이 요구되는 지역내 토양에 잔류시킬 수 있다. 또한, 단백질을 직접 곤충 먹이원내로 유전공학적으로 처리할 수 있다. 예를 들면, 많은 곤충 유충의 주요 먹이원은 식물이다.

포도랍두스 독소의 살충 특성을 코딩하는 유전 물질을 특정 곤충 해충의 먹이가 되는 식물의 계통내로 혼입시킴으로써, 성충 또는 유충이 먹이 식물을 소비한 후에 죽을 수 있다. 많은 수의 단자엽식물속 및 쌍자엽식물속이 형질전환되었다. 유전자전환 농작물 뿐만 아니라 열매 및 채소가 상업적인 대상이다. 그러한 농작물은 옥수수, 벼, 대두, 카놀라, 해바라기, 알팔파, 사탕수수, 밀, 면, 땅콩, 토마토, 감자 등을 포함하지만 이에 한정되지 않는다. 외인성 유전 물질을 식물 세포내에 도입시키기 위한 및 도입된 유전자를 안정하게 유지하고 발현하는 식물을 얻기 위한 몇몇 기술이 존재한다. 그러한 기술은 미립자에 피복된 유전 물질의 세포내로의 직접적인 도입 과정을 포함한다(미국 특허 4,945,050 (Cornell), 및 미국 특허 5,141,131 (DowElanco)). 식물은 아그로박테리움 기술을 사용하여 형질전환될 수 있다[미국 특허 5,177,010 (University of Toledo), 미국 특허 5,104,310 (Texas A&M), 유럽 특허 출원 0131624B1, 유럽 특허 출원 120516, 159418B1 및 176,112(Schilperoot), 미국 특허 5,149,645, 5,469,976, 5,464,763, 4,940,838, 및 4,693,976 (Schilperoot), 유럽 특허 출원 116718, 290799 및 320500 (모두 MaxPlanck), 유럽 특허 출원 604662 및 627752 (Japan Tobacco), 유럽 특허 출원 0267159 및 0292435, 및 미국 특허 5,231,019 (모두 Ciba Geigy), 미국 특허 5,463,174 및 4,762,785 (모두 Calgene), 및 미국 특허 5,004,863 및 5,159,135 (모두 Agracetus)를 참조하십시오]. 다른 형질전환 기술은 휘스커스 (whiskers) 기술을 포함한다[미국 특허 5,302,523 및 5,464,765 (모두 Zeneca)를 참조하십시오]. 일렉트로포레이션 기술이 또한 식물을 형질전환시키기 위해 사용되고 있다[국제 출원 공개 87/06614 (Boyce Thompson Institute), 미국 특허 5,472,869 및 5,384,253 (모두 Dekalb), 국제 출원 공개 92/09696 및 93/21335 (모두 PGS)를 참조하십시오]. 이들 형질전환 특허 및 공개를 모두 본원에서 참고로 인용하였다. 식물을 형질전환시키기 위한 많은 기술에 덧붙여, 외인성 유전자와 접촉되는 조직 유형이 또한 다양할 수 있다. 그러한 조직은 배아 조직, 칼루스 조직 유형 I 및 II, 배축 조직(hypocotyl), 분열조직 등을 포함할 것이지만 이에 한정되지는 않는다. 숙련인의 기술내의 적절한 기술을 사용하여 거의 모든 식물 조직이 탈분화 중에 형질전환될 수 있다.

다른 변수는 선택가능한 마커의 선택이다. 특정 마커에 대한 선호는 숙련인의 재량이지만, 임의의 하기 선택가능한 마커가 선택가능한 마커로서 작용할 수 있는 본원에 나열되지 않은 임의의 다른 유전자와 함께 사용될 수 있다. 그러한 선택가능한 마커는 항생제 카나마이신, 네오마이신 및 G418에 내성을 코딩하는 트랜스포존 Tn5의 아미노글리코사이드 포스포트랜스퍼라제 유전자(Aph II) 뿐만 아니라, 글리포세이트; 하이그로마이신; 메토틱세이트; 포스포노트리신(비알로포스); 이미다졸리논, 술폰일우레아 및 트리아졸로피리미딘 제초제(예를 들면, 글로로술폰; 브로멕시닐, 달라폰 등)에 대한 내성 또는 용인성을 코딩하는 유전자를 포함하지만 이에 한정되지는 않는다.

선택 마커 외에, 리포터(reporter) 유전자의 사용이 바람직할 수 있다. 몇몇 경우, 선택 마커없이 리포터 유전자를 사용할 수 있다. 리포터 유전자는 전형적으로 수용 생물체 또는 조직에 존재하거나 발현되지 않는 유전자이다. 리포터 유전자는 전형적으로 일부 표현형의 변화 또는 효소 특성을 제공하는 단백질을 코딩한다. 그러한 유전자의 예는 본원에 참고로 인용한 문헌(K. Weising 등, Ann. Rev. Genetics, 22, 421 (1988))에서 제공된다. 바람직한 리포터 유전자는 글루쿠로니다제(GUS) 유전자이다.

형질전환 기술에 상관없이, 유전자는 바람직하게는 식물 프로모터를 벡터에 포함시킴으로써 식물 세포내에서 포도랍두스 독소를 발현시키기 위해 개조된 유전자 전달 벡터내로 도입된다. 식물 프로모터 외에, 다양한 공급원으로부터의 프로모터가 외인성 유전자를 발현시키기 위해 식물 세포에서 효율적으로 사용될 수 있다. 예를 들면, 박테리아 기원의 프로모터(예를 들면, 옥토펜 신타제 프로모터, 노팔린 신타제 프로모터, 만노핀 신타제 프로모터), 바이러스 기원의 프로모터(콜리플라워 모자이크 바이러스(35S 및 19S), 35T로 알려진 재가공된 35S) 등이 사용될 수 있다 (참조. PCT/US96/16582, 1997년 4월 17일 공개된 W097/13402가 본원에 참고로 도입된다). 식물 프로모터는 리볼로스-1,6-비스포스페이트 (RUBP) 카르복실라제의 작은 서브유닛(ssu), 베타-콘글리시닌 프로모터, 파제올린 프로모터, ADH 프로모터, 열 속 프로모터 및 조직 특이성 프로모터를 포함하지만 이에 한정되지는 않는다. 프로모터는 또한 전자 효능을 개선시킬 수 있는 특정 인핸서 서열 요소를 포함할 수 있다. 전형적인 인핸서는 Adh-인드론 1 및 Adh-인드론 6을 포함하지만 이에 한정되지는 않는다. 구성 프로모터가 사용될 수 있다. 구성 프로모터는 모든 세포 유형에서 모든 시기에 연속적인 유전자 발현을 일으킨다(예를 들면, 액틴, 유비퀴틴, CaMV 35S). 조직 특이성 프로모터는 잎 또는 종자와 같은 특이 세포 또는 조직 유형에서 유전자 발현을 담당하고(예를 들면, 제인, 올레오신, 나핀, ACP), 이들 프로모터가 또한 사용될 수 있다. 프로모터는 또한 식물 발생의 특정 단계 동안 활성이며 또한 식물 조직 및 기관에서 활성일 수 있다. 그러한 프로모터

의 예는 화분 특이성, 배 특이성, 옥수수 수염 특이성, 면 섬유 특이성, 뿌리 특이성, 종자 배유 특이성 프로모터 등을 포함하지만 이에 한정되지는 않는다.

특정 환경 하에, 유도성 프로모터를 사용하는 것이 바람직할 수 있다. 유도성 프로모터는 특이 신호 (예를 들면, 물리적 자극(열 속 유전자); 광(RUBP 카르복실라제); 호르몬(Em); 대사물질; 및 스트레스)에 대해 반응하는 유전자의 발현을 담당한다. 식물에서 작용하는 다른 바람직한 전사 및 번역 요소가 사용될 수 있다. 많은 식물-특이성 유전자 전달 벡터가 당업계에 공지되어 있다.

그 외에, 식물 내에서 박테리아 유전자의 높은 발현을 수득하기 위해, 박테리아 유전자를 재공학 처리하여 식물의 세포질내에서 더 효율적으로 발현할 수 있도록 하는 것이 바람직한 것으로 공지되어 있다. 옥수수가 식물내 독소의 발현 수준을 증가시키기 위해 형질전환에 앞서 박테리아 유전자(들)를 재공학처리하는 것이 바람직한 그러한 식물 중 하나이다. 재공학처리하는 한 가지 이유는 천연 박테리아 유전자(들)의 매우 낮은 G+C 함량 (및 결과적으로 높은 A+T 함량으로의 편향) 때문이다. 이는 A+T가 매우 풍부한 것으로 공지된 식물 유전자 대조 서열을 의태하거나 복제하는 서열을 생산시킨다. 식물내로 도입된 유전자(들)의 DNA 내의 일부 A+T 풍부 서열의 존재(예를 들면, 유전자 프로모터에서 보통 발견되는 TATA 박스 영역)는 유전자(들)의 이상 전사를 일으킬 수 있다. 반면, 전사된 mRNA내에 존재하는 다른 조절 서열의 존재(예를 들면, 폴리아데닐화 신호 서열(AAUAAA), 또는 예비-mRNA 스플라이싱에 관여하는 작은 핵 RNA에 상보적인 서열)는 RNA 불안정성을 이끌 수 있다. 따라서, 더 바람직하게 식물 최적화된 유전자(들)로 지칭되는, 재공학처리된 박테리아 유전자(들)의 디자인에서 한 목적은 더 높은 G+C 함량을 갖는 DNA 서열을 생산하는 것이며, 바람직하게는 대사성 효소를 코딩하는 식물 유전자(들)의 것에 근접하는 것을 생산하는 것이다. 식물 최적화된 유전자(들)의 디자인에서 다른 목적은 더 높은 G+C 함량을 가질 뿐만 아니라, 서열 변화를 변경시킴으로써 번역이 저해되지 않도록 제조되어야 하는 DNA 서열을 생산하는 것이다.

높은 G+C 함량을 갖는 식물의 예는 옥수수이다. 하기 표 2는 옥수수내에서 G+C 함량의 높은 정도를 설명한다. 옥수수에서와 같이, 다른 식물내 G+C 함량이 또한 높을 것으로 생각된다.

[표 2]

| 옥수수 유전자의 단백질 코딩 영역의 G+C 함량의 수집 | | |
|---|-----------|-------------------------|
| 단백질 종류 ^a | %G+C의 범위 | %G+C의 평균 ^b |
| 대사성 효소 (40) | 44.4-75.3 | 59.0 (8.0) |
| 저장 단백질 | | |
| I 군 (23) | 46.0-51.9 | 48.1 (1.3) |
| II 군 (13) | 60.4-74.3 | 67.5 (3.2) |
| I+II 군 (36) | 46.0-74.3 | 55.1 (9.6) ^c |
| 구조 단백질 (18) | 48.6-70.5 | 63.6 (6.7) |
| 조절 단백질 (5) | 57.2-68.9 | 62.0 (4.9) |
| 비특성화된 단백질 (9) | 41.5-70.3 | 64.3 (7.2) |
| 총 단백질 (108) | 44.4-75.3 | 60.8 (5.2) |
| ^a 종류내 유전자의 수를 괄호안에 제공한다. | | |
| ^b 표준 편차를 괄호안에 제공한다. | | |
| ^c 합한 군의 평균은 총 평균의 계산에서 무시된다. | | |

표 2의 데이터에서, 유전자들의 코딩 영역은 진뱅크(GenBank)(발매 71) 등록으로부터 발체하고, 기초 조성은 맥벡터(MacVectorTM) 프로그램(IBM, New Haven, CT)를 사용하여 계산하였다. 인트론 서열은 계산에서 무시된다. I 군 및 II 군 저장 단백질 유전자 서열은 기초 조성에서 그들의 두드러진 차이에 의해 구별된다.

유전자 코드의 여분에 의해 제공되는 유연성으로 인해(즉, 몇몇 아미노산이 하나 이상의 코돈에 의해 지정됨), 상이한 생물체 또는 종류 또는 생물체의 진화는 여분의 코돈의 상이한 사용을 일컫는다. 이 "코돈 바이어스"는 단백질 코딩 영역의 평균 기초 조성물에서 반영된다. 예를 들면, 상대적으로 낮은 G+C 함량을 갖는 생물체는 여분의 코돈의 제 3 위치에서 A 또는 T를 갖는 코돈을 이용하는 반면, 더 높은 G+C 함량을 갖는 생물체는 제 3 위치에서 G 또는 C를 갖는 코돈을 이용한다. 유전자의 mRNA내에 "마이너(minor)" 코돈의 존재는 특히 마이너 코돈에 상응하는 충전된 tRNA의 상대적인 풍부성이 낮은 경우, mRNA의 완전한 번역율을 감소시킬 수 있는 것으로 생각된다. 이를 확장하면, 개별적인 마이너 코돈에 의한 번역율의 감소가 다중 마이너 코돈에 대해 적어도 부가적일 것이다. 따라서, 상대적으로 높은 함량의 마이너 코돈을 갖는 mRNA는 상응하게 낮은 번역율을 가질 것이다. 이러한 번역율은 코딩되는 단백질의 낮은 수준의 합성에 의해 반영될 것이다.

박테리아 유전자(들)을 재공학처리하기 위해, 식물의 코돈 바이어스를 측정한다. 코돈 바이어스는 식물이 그의 단백질을 코딩하기 위해 사용하는 통계적인 코돈 분포이다. 바이어스를 측정한 후, 관심있는 유전자(들) 내의 코돈의 빈도%를 측정한다. 식물이 선호하는 제 1 코돈은 바람직한 코돈의 제 2 및 제 3의 선택과 함께 결정되어야 한다. 관심있는 단백질의 아미노산 서열이 역번역되어 생성된 핵산 서열은 천연 박테리아 유전자에서와 동일한 단백질을 코딩하지만, 이 생성된 핵산 서열은 원하는 식물의 바람직한 제 1 코돈에 해당한다. 새로운 서열을 변경에 의해 생성될 수 있는 제한 효소 자리에 대해 분석한다. 동정된 자리는 코돈을 제 2 또는 제 3 선택의 바람직한 코돈으로 교체함으로써 추가로 변경된다. 관심있는 유전자의 전사 또는 번역에 영향을 끼칠 수 있는 서열내의 다른 자리는 엑손:인트론 5' 또는 3' 결함, 폴리 A 첨가 신호, 또는 RNA 폴리머라제 종결 신호이다. 서열은 추가로 분석되고 TA 또는 GC 더블렛(doublet)의 빈도를 감소시키기 위해 변경된다. 더블렛 외에, 약 4개 이상의 동일한 잔기를 갖는 G 또는 C 서열 블록이 서열의 전사에 영향을 끼칠 수 있다. 따라서, 이들 블록은 또한 제 1 또는 제 2 선택 등의 코돈을 다음 바람직한 선택의 코돈으로 교체함으로써 변경된다. 식물 최적화된 유전자(들)가 약 63%의 제 1 선택 코돈, 약 22% 내지 약 37%의 제 2 선택 코돈, 및 15% 내지 0%의 제 3 선택 코돈을 포함하는 것이 바람직하며, 총 %는 100%이다. 가장 바람직한 식물 최적화된 유전자(들)는 약 63%의 제 1 선택 코돈, 적어도 약 22%의 제 2 선택 코돈, 약 7.5%의 제 3 선택 코돈 및 약 7.5%의 제 4 선택 코돈을 포함하며, 총 %는 100%이다. 전술된 방법으로 당업계의 숙련인은 특정 식물에 외인성인 유전자(들)를 변경하여 유전자가 식물내에서 최적으로 발현되도록 할 수 있다. 이 방법은 1997년 4월 17일 W097/13402로 공개된 국제 특허 출원 제PCT/US96/16582호에 더 설명되어 있다.

따라서, 식물 최적화된 유전자(들)를 디자인하기 위해서는, 형질전환될 특정 식물의 유전자 DNA 서열에 대해 파악한 코돈 바이어스 프로부터 확정된 비어부의 유전자 코드를 사용하여, 독소의 아미노산 서열을 DNA 서열로 역번역시킨다. 코돈 사용에서 완전하게 상동성인, 생성된 DNA 서열은 더 높은 정도의 코돈 다양성 외에 또한 전략적으로 배치된 제한 효소 인식 자리 및 바람직한 염기 조성물을 갖고, 유전자의 전사 또는 생성물 mRNA의 번역을 방해할 수 있는 서열이 결여된 DNA 서열을 정하기 위해 추가로 변경된다.

박테리아 유전자가 색소체(plastid)내에서 발현되는 경우에는 이론상 식물내에서 그 박테리아 유전자는 식물 내에서 더 쉽게 발현될 수 있다. 이런 경우 식물 발현을 위해 유전자를 최적화시키지 않고도, 식물내에서 박테리아 유전자를 발현시킬 수 있으며, 단백질을 높은 발현율로 수득할 수 있다. 본원에서 참고로 인용한 미국 특허 제 4,762,785호, 제 5,451,513호 및 제 5,545,817호를 참조하시오.

유전자전환 식물을 상업적으로 이용하는데 관한 문제점 중 하나는 내성 관리이다. 이는 특히 바실루스 터링기엔시스(*Bacillus thuringiensis*) 독소에 있어서 그러하다. 많은 회사가 바실루스 터링기엔시스를 상업적으로 이용하며, Bt 독소가 내성이 되는 것에 대하여 많은 관심을 두고 있다. 곤충 내성 관리의 한 방법은 포도랍두스에 의해 생산된 독소를 Bt와 같은 독소, 식물 곤충 단백질(Ciba Geigy) 또는 다른 독소와 결합시키는 것이다. 이 결합물은 분무가능한 적용을 위해 제형화될 수 있거나, 분자 결합물이 될 수 있다. 식물은 곤충 독소를 생성하는 포도랍두스 유전자, 및 Bt와 같은 다른 곤충 독소 유전자로 형질전환될 수 있다.

유럽 특허 출원 0400246A1에서는 임의의 2개의 유전자일 수 있는, 식물내 2개의 Bt의 형질전환을 설명하였다. 하나 이상의 곤충 내성 유전자를 포함하는 유전자전환(transgenic) 식물을 생산하기 위한 다른 방법은 각각의 식물이 곤충 내성 유전자를 포함하는 2개의 식물을 생산하는 것이다. 이들 식물은 하나 이상의 곤충 내성 유전자를 포함하는 식물을 생산하기 위해 전통적인 식물 교배 기술을 사용하여 역교배될 수 있다.

식물 최적화된 유전자(들)를 포함하는 유전자전환 식물을 생산하는 것 외에, 박테리아 유전자(들)를 재공학처리하는 것이 요망될 수 있는 다른 전달계가 존재한다. 동일한 라인을 따라, 먹이원으로서 곤충을 유인하는 분자 및 독소의 살충 활성을 함께 융합시키는, 유전 공학처리되고 쉽게 단리된 단백질 독소를 잘 알려진 표준 기술을 사용하여 박테리아 또는 진핵세포내에 공학처리되고 발현될 수 있다. 실험실에서의 정제 후에, "조립된(built-in)" 미끼를 갖는 그러한 독성 약제를 표준 곤충 및 하우스내에 포장할 수 있다.

다른 전달계는 바쿨로바이러스 벡터내로의 독소의 유전 물질을 혼입하는 것이다. 바쿨로바이러스는 포도랍두스 독소의 바람직한 표적을 포함하는 특정한 곤충 숙주를 감염시킨다. 포도랍두스 독소를 위한 발현 구조물을 포함하는 감염성 바쿨로바이러스를 곤충 만연 지역으로 도입하여, 곤충을 감염시켜 유독화시키거나 중독시킬 수 있다.

살충 특성을 전달하기 위해서는 벡터가 거주할 숙주에 적합한 단백질 발현 벡터내로, 포도랍두스 독소의 코딩 아미노산 서열을 코딩하는 핵산 서열이 통합될 필요가 있다. 살충 특성을 갖는 단백질을 코딩하는 핵산 서열을 수득하기 위한 한 방법은 독소의 아미노산 서열로부터 추정된 정보(이의 대부분이 하기에 설명됨)를 이용하여 포도랍두스에서 독소를 생산하는 천연 유전 물질을 단리하는 것이다. 후술될 바와 같이, 독소 활성에 관여하는 단백질의 정제 방법을 또한 개시한다.

하기한 바와 같은 N-말단 아미노산 서열 데이터를 사용하여, 독소의 제 1 아미노산을 코딩하는 DNA 염기의 전부 또는 일부에 상보적인 올리고뉴클레오티드를 제작할 수 있다. 이들 올리고뉴클레오티드는 방사성표지화된 후, 포도랍두스의 균주에서 단리된 유전 물질로부터 제작한 게놈 유전자 라이브러리에서 유전 물질을 단리하기 위한 분자 프로브로서 사용될 수 있다. 유전자 라이브러리는 플라스미드, 코스미드, 파지 또는 파지미드 벡터내에 클로닝될 수 있다. 이 라이브러리에 에스케리치아 콜리(*Escherichia coli*)를 형질전환시키고 이 형질전환된 세포가 독소를 생산하는지 여부를 알기 위해 이 독소에 대해 발생된 항체를 사용하여 스크리닝하거나, 곤충 독성에 대해 직접 분석할 수 있다.

동의성 유전 코드가 수개의 3-뉴클레오티드 조합 중 임의의 것에 의해 아미노산이 DNA내에서 코딩

될 수 있도록하므로, 이러한 연구는 한벌의 올리고뉴클레오티드의 생산을 필요로한다. 예를 들면, 아미노산 아르기닌은 핵산 트리플렛 CGA, CGC, CGG, CGT, AGA 및 AGG에 의해 코딩될 수 있다. 독소 유전자내에서 그러한 위치에서 어떠한 트리플렛이 사용되는지를 예측할 수 없으므로, 제시된 각각의 가능한 트리플렛을 갖는 올리고뉴클레오티드를 제조하여야 한다. 단백질 서브유닛에 상응하는 하나 이상의 DNA 분자가 경우 독성을 성취하기 위해 필수적인 단백질 서브유닛의 전부를 회수하기 위한 충분한 수의 올리고뉴클레오티드 프로브를 제작하기 위해 필수적일 수 있다.

분자 생물학 분야의 숙련인에게 잘 공지된 몇몇 기술중 임의의 것을 사용하여 정제된 단백질의 아미노산 서열로부터, 독소의 생산을 책임지는 유전 물질을 쉽게 분리하여, 전체 또는 일부를 발현 벡터내로 클로닝시킬 수 있다. 전형적인 발현 벡터는 DNA 플라스미드이지만, 코스미드, 파지미드 및 파지를 포함하지만 이에 한정되지 않는 다른 전달 수단도 또한 계획될 수 있다. 플라스미드 복제를 위해 필요하거나 요구되는 특징, 예를 들면, 복제 기원 및 항생제 내성 또는 선택가능한 마커의 다른 형태, 예를 들면, 스트렙토마이시스 하이그로스코피쿠스(*Streptomyces hygroscopicus*) 또는 비리도크로모겐스(*viridochromogene*)의 bar 유전자 외에, 단백질 발현 벡터는 일반적으로 관심있는 유전자의 전사 및 번역에 필요한 시스-작용 서열을 혼입하는 발현 카셋트를 부가적으로 필요로한다. 원핵세포내의 발현을 위해 요구되는 시스-작용 서열은 진핵세포 및 식물에서 요구되는 것과 상이하다.

진핵세포 발현 카셋트는 관심있는 유전자에 대한 상류(5')의 전사 프로모터, 폴리-A 첨가 자리와 같은 전사 종결 영역 및 관심있는 유전자의 제 1 코돈의 상류의 리보솜 결합 자리를 필요로한다. 박테리아 세포에서, 벡터내로 포함될 수 있는 유용한 전사 프로모터는 T7 RNA 폴리머라제-결합 프로모터이다. 본원에서 전술된 바와 같은 프로모터는 mRNA의 전사를 유효하게 진행시키는 것으로 공지되어 있다. 또한, 관심있는 유전자의 상류에서, 벡터는 단백질을 세포 표면과 같은 숙주 세포의 특정 부분에 공유적으로 결합시키는 것으로 공지된 신호 서열을 코딩하는 뉴클레오티드 서열을 포함할 수 있다.

곤충 바이러스 또는 바쿨로바이러스는 특정한 곤충을 감염시켜 부작용을 끼치는 것으로 공지되어 있다. 곤충에 대한 바이러스의 영향은 느리며, 바이러스는 곤충의 먹이섭취를 중단시키지 않는다. 따라서, 바이러스는 곤충 해충 방제제로서 유용한 것으로 보이지 않는다. 포토랍두스 독소 유전자를 바쿨로바이러스 벡터내로 결합시키면, 바이러스의 치사율을 증가시키면서 독소의 전달의 유효한 방법을 제공할 수 있다. 그외에, 상이한 바쿨로바이러스가 상이한 곤충에 특이적이므로, 특히 해를 끼치는 곤충 해충을 선택적으로 표적으로하기 위해 특정 독소를 사용하는 것이 가능할 수 있다. 독소 유전자를 위해 특히 유용한 벡터는 핵 폴리헤드로시스 바이러스이다. 이 바이러스를 이용하는 전달 벡터는 이전에 기술되었고, 현재 곤충에 외부 유전자를 전달시키기 위한 선택 벡터이다. 바이러스-독소 유전자 재조합물은 경구 전달가능한 형태로 제작될 수 있다. 바쿨로바이러스는 일반적으로 곤충의 중장 장점막을 통하여 감염시킨다. 강한 바이러스의 코트 단백질 프로모터의 뒤에 삽입된 독소 유전자가 발현되어 감염된 곤충을 신속하게 죽일 것이다.

곤충 바이러스 또는 바쿨로바이러스 또는 본 발명의 단백질 독소를 위한 유전자전환 식물 전달계 외에, 단백질은 바실러스 터린기엔시스 캡슐화 기술(예를 들면, 본원에서 모두 참고로 인용한 미국 특허 제 4,695,455호, 제 4,695,462호, 제 4,861,595호에서와 같지만 이에 한정되지는 않음)을 사용하여 캡슐화될 수 있다. 본 발명의 단백질 독소를 위한 다른 전달계는 미끼 매트릭스 내로의 단백질의 제제이며, 이어서, 이 제제를 지상 또는 지하 곤충 미끼 스테이션에서 사용할 수 있다. 그러한 기술의 예는 본원에서 참고로 인용한 PCT 특허 출원 WO 93/23998을 포함하지만 이에 제한되지는 않는다.

전술된 바와 같이, 다른 숙주의 코돈 선호도가 포토랍두스와는 상이할 수 있으므로, 비천연 숙주에서 발현시키는 경우, 단백질을 코딩하는 서열을 변경하는 것이 필요할 수 있다. 그러한 경우, 코딩 서열에 대한 보충 변경을 하지 않는다면, 새로운 숙주내에서 번역은 매우 불충분할 수 있다. 부가적으로, 새로운 숙주의 단백질과의 억제 교차 반응성을 피하거나, 새로운 숙주내에서 단백질의 살충 특성을 개량하기 위해 아미노산 서열에 대한 변형이 요망될 수 있다. 유전적으로 변형된 독소 유전자는 예를 들면, 증강되거나 감소된 독성, 변환된 곤충 내성 개발, 변환된 안정성 또는 변형된 표적 종 특이성을 나타내는 독소를 코딩할 것이다.

독소를 코딩하는 포토랍두스 유전자 외에, 본 발명은 범위는 독소 단백질에 상동성인 아미노산 생중합체를 코딩하고, 곤충류에서 경구 섭취후에 포토랍두스 단백질의 독성 효과를 보유하는 관련 핵산 서열을 포함하는 것이다.

예를 들면, 본 발명에서 사용된 독소는 유충을 죽이기 전에 먼저 유충의 먹이 섭취를 억제하는 것으로 보인다. 포토랍두스 독소의 핵산 서열 또는 그의 제어 서열을 조작함으로써, 독소 유전자를 식물내에 배치시키는 유전 공학자는 예를 들면, 유충에 대한 완전한 독성을 제거하면서 먹이섭취-저해 활성을 유지시키기위해 그의 효력 또는 그의 작용 양식을 조정할 수 있다. 이러한 변화는 형질전환된 식물이 모든 표적 곤충을 생태계내에서 없애는 불필요한 극적인 효과를 갖지 않고도 추수때까지 생존할 수 있도록 한다. 독소를 코딩하는 유전자 또는 유전자에 의해 코딩되는 단백질의 모든 그러한 변경은 본 발명의 범위내에 있는 것으로 계획된다.

핵산의 다른 계획된 변경은 그의 효능을 개선하기 위해 곤충 유충의 특정 부위로 독소를 전달시키는 표적 서열의 첨가를 포함한다.

균주 W-14, ATCC 55397, 43948, 43949, 43950, 43951 및 43952는 아메리칸 타입 컬처 컬렉션(미국 20852 메릴랜드주 록크빌 파크로운 드라이브 12301)에 기탁되어 있다. W-14 천연 독소(ATCC 55397)에 대한 아미노산 및 뉴클레오티드 서열 데이터를 하기에 제시한다. 박테리아 숙주로부터의 독소에 대한 게놈 DNA의 단리를 또한 본원에서 예시한다. 본원의 다른 균주는 미농무부 (미국 일리노이주 페오리아 노쓰 유니버시티 드라이브 1815)에 기탁되었다.

표준 및 분자 생물학 기술은 본원에서 명세서에 후술하고 교시한다. 부가의 정보는 본원에 참고

로 인용한 문헌[Sambrook, J., Fritsch, E.F., 및 Maniatis, T. (1989), Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press] 및 [Current Protocols in Molecular Biology, ed. F. M. Ausubel et al., (1997)]에서 찾을 수 있다.

하기 약자를 실시예 전체에서 사용한다: Tris= 트리스(히드록시메틸) 아미노 메탄; SDS= 소듐 도데실 술페이트; EDTA= 에틸렌디아민테트라아세트산; IPTG= 이소프로필티오-β-갈락토시드; X-gal= 5-브로모-4-클로로-3-인도일-β-D-갈락토시드; CTAB= 세틸트리메틸암모늄 브로마이드; kbp= 킬로염기 쌍; dATP, dCTP, dGTP, dTTP, I= 각각 아데닌, 시토신, 구아닌, 티민 및 이노신의 2'-데옥시뉴클레오시드 5'-트리포스페이트; ATP= 아데노신 5'-트리포스페이트.

실시예

<실시예 1>

포도랍두스 루미네센스 유래의 독소의 정제 및 정제된 독소의 경구 전달 후의 독성의 증명

본 발명의 살충 단백질 독소를 포도랍두스 루미네센스 균주 W-14, ATCC 기탁 번호 55397로부터 정제하였다. 포도랍두스 루미네센스의 원배양물을 1.5% 아가 중 2% 프로테오스 펩톤 #3(즉, PP3, Difco Laboratories, Detroit MI)가 들어있는 페트리 접시 상에 유지하고, 25°C에서 배양하고 매주 옮겼다. 박테리아의 제 1 형태의 콜로니를 1ℓ 플라스크 중에서 0.5%의 폴리옥시에틸렌 소르비탄 모노스테아레이트(트윈 60(Tween 60), Sigma Chemical Company, St. Louis MO)가 보충된 200 ml의 PP3 브로쓰내로 접종하였다. 브로쓰 배양물을 회전 진탕기 상에서 30°C에서 72시간 동안 성장시켰다. 트윈의 존재 또는 부재하에 성장된 배양물로부터 독소 단백질을 회수할 수 있지만; 트윈이 부재하면 성장된 박테리아의 형태 및 박테리아에 의해 생산된 단백질의 프로파일에 영향을 끼칠 수 있다. 트윈의 부재하에, 적어도 하나의 동정된 독소 서브유닛의 분자량이 약 200 kDa에서 약 185 kDa로 이동하는 한, 변형 이동(variant shift)이 발생한다.

72 시간 배양물을 10,000×g에서 30분 동안 원심분리하여 세포 및 파쇄물을 제거하였다. 살충 활성을 포함하는 상청액 분획을 기울어 따르고 적절한 용적의 1.0M K₂HPO₄를 첨가함으로써 50mM K₂HPO₄로 조정하였다. 수산화칼륨을 첨가함으로써 pH를 8.6으로 조정하였다. 이어서, 이 상청액 분획을 50mM K₂HPO₄로 평형화된 DEAE-세파셀(Sephacel) (Pharmacia LKB Biotechnology)와 혼합시켰다. 독성 활성이 DEAE 수지에 흡착되었다. 이어서, 이 혼합물을 2.6×40cm 컬럼에 따르고, 유출액이 280nm에서 정상(steady) 베이스라인 UV 흡수에 도달할때까지 실온에서 30ml/hr의 유동속도로 50mM K₂HPO₄로 세척하였다. 이어서, 컬럼을 유출액이 다시 정상 280nm 베이스라인에 도달할때까지 150mM KCl로 세척하였다. 최종적으로, 컬럼을 300mM KCl로 세척하고, 분획을 수집하였다.

독소를 포함하는 분획을 모으고, 0.2미크론 공극 막 필터를 사용하여 여과 멸균하였다. 이어서, 독소를 농축시키고, 4°C에서 100 kDa의 분자량 컷오프를 갖는 초여과막(Centriprep 100, Amicon Division-W.R. Grace and Company)을 사용하여 100mM KPO₄, pH 6.9로 평형화시켰다. 독소 농축물의 3ml 샘플을 2.6×95cm 세파크릴(Sephacryl) S-400 HR 겔 여과 컬럼(Pharmacia LKB Biotechnology)의 상단에 적용하였다. 용출 완충액은 100mM KPO₄, pH 6.9이었고, 이를 4°C에서 17 ml/hr의 유동속도로 전개시켰다. 용출물을 280nm에서 모니터링하였다.

분획을 모으고, 독성 활성에 대하여 시험하였다. 크로마토그래피 분획의 독성을 만두카섹스타(Manduca sexta) 유충을 사용하는 생물학적 분석으로 시험하였다. 분획을 곤충 먹이(매미 나방 밀 배(germ) 먹이, ICN Biochemicals Division - ICN Biomedicals, Inc.) 상에 직접 적용하거나 5μl의 샘플을 30개이지 바늘을 사용하여 4회 또는 5회령(instar) 유충의 제 1 전각(proleg)을 통하여 인트라헤모셀릭 (intrahemocelic) 주사에 의해 투여하였다. 처리군내의 각각의 유충의 무게를 24시간 간격으로 기록하였다. 처리된 곤충 먹이를 소비한 수일내에 곤충이 먹이섭취를 중단하고 죽거나, 분획을 주사한 후 24시간 내에 죽는다면 독성이 추정된다.

독성 분획을 모으고 센트리프렙(Centriprep)-100을 사용하여 농축시킨 다음, 100mM의 인산 칼륨, pH 6.9의 용출 완충액을 0.4ml/분으로 전개하면서 7.5mm×60cm TSK-GEL G-4000 SW 겔 삼투 컬럼을 사용하여 HPLC에 의해 분석하였다. 이 분석으로 독소 단백질이 약 33.6분의 보유 시간으로 컬럼으로부터 용출된 하나의 날카로운 피크내에 포함되는 것을 밝혔다. 이 보유 시간은 1,000 kDa의 추정된 분자량에 상응하였다. 추가의 정제를 위해 피크 분획을 모으고, 이 단백질을 포함하지 않는 분획을 폐기하였다. HPLC로부터 용출된 피크는 218 및 280nm에서 UV 광을 흡수하지만, 405nm에서는 흡수하지 않았다. 405nm에서의 흡수는 제노라핀 항생제 화합물에 의한 것으로 나타났다.

모든 피크 분획을 비-변성 아가로즈겔(Metaphor Agarose, FMC BioProducts) 내에서 전기영동하면, 두 개의 단백질 복합체가 피크내에 존재함을 나타냈다. 50 mM Tris-HCl(pH 7.0)에 완충된 피크 물질을 표준 완충 조건(양극 완충액 1M Tris-HCl(pH 8.3); 음극 완충액 0.025M Tris, 0.192M 글라이신)하에 100mM Tris-HCl(pH 7.0)로 완충된 1.5% 아가로즈 스테핑 겔 및 200mM Tris-보레이트(pH8.3)으로 완충된 1.9% 아가로즈 용해 겔 상에서 분리하였다. 겔을 15°C에서 13 mA의 불변 전류에서 페놀 레드 트래킹 염료가 겔의 단부에 도달할때까지 전개시켰다. 두 개의 단백질 밴드가 쿠마지 브릴리언트 블루(Coomassie brilliant blue) 염색을 사용하여 아가로즈 겔에 가시화되었다.

느리게 이동하는 밴드를 "단백질 밴드 1"로 지칭하고, 빠르게 이동하는 밴드를 "단백질 밴드 2"로 지칭하였다. 두 개의 단백질 밴드는 대략 동일한 양으로 존재하였다. 겔의 비염색된 부분으로부터 두 개의 단백질 밴드를 정확하게 절제하기 위해, 쿠마지 염색된 아가로즈 겔을 가이드로서 사용하였다. 단백질 밴드를 포함하는 절제된 조각을 용매에 넣어 부드럽게하고, 소량의 멸균수를 첨가하였다. 대조군으로서, 단백질을 포함하지 않는 겔의 일부를 또한 절제하고, 단백질을 포함하는 겔 조각에서와 동일한 방법으로 처리하였다. 단백질을 전기용출에 의해 100V(불변 전압)에

서 겔 조각으로부터 2시간 동안 100mM Tris-보레이트(pH 8.3) 중으로 회수하였다. 대안적으로, 겔 조각에 등용적의 50mM Tris-HCl(pH 7.0)을 첨가함으로써 겔 조각으로부터 단백질을 수동적으로 용출시킨 다음, 30°C에서 16시간 동안 배양시켰다. 이에 의해, 단백질이 겔로부터 완충액 중으로 분산되며, 이어서, 이를 모았다.

HPLC-정제된 독소(33.6분, 피크) 및 아가로즈 겔 정제된 독소를 사용하는 곤충 독성 시험의 결과는 추출물의 독성을 증명하였다. HPLC 정제된 단백질 1.5µg을 주사하는 경우, 24시간 내에 곤충이 죽었다. 수동 용출 또는 전기용출에 의해 아가로즈 겔로부터 회수된 단백질 밴드 1 및 단백질 밴드 2는 모두 주사하는 경우 곤충을 치사시켰다. 이들 샘플에 대해 추정된 단백질 농도는 50ng 미만/유충이었다. 단백질 밴드 1 또는 단백질 밴드 2가 주사된 유충군 사이의 무게 증가 및 치사율을 비교하면, 단백질 밴드 1이 주사 전달에 의해 더 독성임을 나타냈다.

HPLC-정제된 독소가 7.5µg/유충의 농도로 유충 먹이에 적용되는 경우, 이는 유충 무게 증가를 중단시켰다(24마리의 유충이 시험됨). 유충이 먹이 섭취를 시작하지만, 단지 매우 소량의 독소 처리된 먹이를 소비한 후에, 이들은 다시 독소에 의해 유도된 병리 증상을 나타내기 시작하고, 먹이 섭취를 중단한다. 유충의 배설물이 퇴색되기 시작했고, 대부분의 유충은 설사 증상을 나타내었다. 수회의 5µg 독소 투여량을 7 내지 10일 기간에 걸쳐 먹이에 적용하는 경우, 상당한 곤충 치사가 발생하였다.

아가로즈-분리된 단백질 밴드 1은 200ng/유충의 투여량에서 유충의 무게 증가를 상당히 억제하였다. 유사한 농도의 단백질 밴드 2를 섭취한 유충은 억제되지 않으며, 대조 유충과 동일한 비율로 무게가 증가하였다. 12마리의 유충에 용출된 단백질을 섭취시켰고, 45마리의 유충에 단백질-항유 아가로즈 조각을 섭취시켰다. 이들 2세트의 데이터는 단백질 밴드 1이 만두카 섹스타에 대해 경구적으로 독성임을 나타낸다. 본 시험에서, 단백질 밴드 2가 만두카 섹스타에 대해 독성이 아님이 명백하다.

변성 조건하에 SDS-PAGE에 의한 단백질 밴드 1 및 2의 추가 분석은 각각의 밴드가 수개의 더 작은 단백질 서브유닛으로 이루어짐을 나타내었다. 단백질은 최대의 감도를 얻기 위해 쿠마지 브릴리언트 블루 염색에 이어 은 염색함으로써 가시화되었다.

두 개의 밴드내의 단백질 서브유닛은 매우 유사하였다. 단백질 밴드 1은 25.1, 56.2, 60.8, 65.6, 166, 171, 184 및 208 kDa의 8개의 단백질 서브유닛을 포함하였다. 단백질 밴드 2는 25.1, 60.8 및 65.6 kDa 단백질이 존재하지 않는 것을 제외하고는 동일한 프로파일을 가졌다. 56.2, 60.8, 65.6 및 184 kDa 단백질은 단백질 밴드 1의 복합체내에 대략 동일한 농도로 존재하고, 복합체의 총 단백질 함량의 80% 이상을 나타내었다.

천연 HPLC-정제된 독소를 하기와 같이 추가로 특성화하였다. 독소는 열 불안정성이며, 60°C로 15분 동안 가열한 후에 옴. 섹스타 유충에 주사하거나 섭취시키는 경우 치사시키거나 무게 증가를 억제하는 그의 능력을 잃었다. 리파제, 유형 C 포스포리파제, 뉴클레아제 또는 적혈구 용혈 활성을 검출하기 위해 분석을 디자인하여, 정제된 독소로 수행하였다. 이들 활성 중 어느 것도 존재하지 않았다. 항생 지대 억제 분석을 또한 수행하였고, 정제된 독소는 그람 음성 또는 그람 양성 박테리아, 효모 또는 필라멘트형 진균의 성장을 억제하지 못하였고, 이는 독성이 제노라핀 항생제가 아님을 나타낸다.

천연 HPLC-정제된 독소를 만두카 섹스타 이외의 곤충을 치사시키는 능력에 대해 시험하였다. 표 3에 본 연구에서 HPLC-정제된 포토랍두스 루미네센스 독소에 의해 치사되는 곤충을 나열하였다.

[표 3]

포토랍두스 루미네센스 독소에 의해 치사되는 곤충

| 일반명 | 목 | 속 및 종 | 송달 경로 |
|-----------------------------|-------|---|---------|
| 박각시나방 (tobacco hornworm) | 나비목 | 만두카 섹스타 (<i>Manduca sexta</i>) | 경구 및 주사 |
| 밀웜 (mealworm) | 딱정벌레목 | 테네브리오 몰리터 (<i>Tenebrio molitor</i>) | 경구 |
| 파라오 개미 | 벌목 | 모노모리움 파라오아니스 (<i>Monomorium pharoanis</i>) | 경구 |
| 바퀴 (German cockroach) | 방시목 | 블라텔라 게르마니카 (<i>Blattella germanica</i>) | 경구 및 주사 |
| 모기 | 파리목 | 애데스 애집티 (<i>Aedes aegypti</i>) | 경구 |

고분자량 독소 복합체의 추가 특징 연구

추가 분석에서, 독소 단백질 복합체를 W-14 성장 배지로부터 추가로 특징을 규명했다. 배양 조건 및 S-400 HR 컬럼을 통한 초기 정제 단계는 상기한 바와 동일했다. 고분자량 독소 복합체를 S-400 HR 컬럼 분획물에서 단리한 후, 독소 분획물을 10mM Tris-HCl(pH 8.6)로 평형화하고 센트리플러스 100(Amicon) 농축기에서 농축시켰다. 그 다음 단백질 독소 복합체는 약한 음이온 교환(WAX) 컬럼, Vydac 301VPH575 (Hesperia, CA)에 0.5 ml/분의 유속으로 적용시켰다. 단백질을 선

형 염화칼륨 구배 10mM Tris-HCl(pH 8.6) 중의 0 - 250 mM KCl로 50분 동안 용출했다. 280nm에서 흡광도 측정된 결과 8개의 단백질 피크가 검출되었다.

신생 써던 옥수수 뿌리벌레 (*Diabrotica undecimpunctata howardi*, SCR) 유충 및 담배 흡수 벌레 (*Manduca sexta*, THW)를 사용하여 HPLC 컬럼으로부터 용출된 모든 분획물에 대해 생검을 수행했다. THW는 낮 16시간, 밤 8시간의 주기로 25 °C에서 매미나방 밀 배(胚) 식이(ICN)를 먹여 성장시켰다. SCR은 낮 16시간/밤 8시간 주기로 25°C에서 써던 옥수수 뿌리벌레 유충의 Insecta-Diet(BioServ)를 먹여 키웠다.

SCR 및 THW 유충의 최대 치사율은 피크 6에서 관찰되었는데, 이 피크는 약 112 mM 내지 132 mM KCl로 용출했다. 피크 6를 SDS-PAGE로 분석하면 두드러진 펩티드는 170 kDa, 66 kDa, 63 kDa, 59.5 kDa 및 31 kDa로 나타났다. 피크 6 단백질 분획물에 대한 웨스턴 블롯 분석은 TcaA_{iii}-syn, TcaA_{iii}-syn, TcaB_{iii}-syn, TcaC-syn, 및 TcbA_{iii}-syn 펩티드 (실시에 21에 기재되어 있음)에 대한 다클론 항체 혼합물 및 TcbA_{iii} 펩티드에 대한 단클론성 항체인 C5F2를 사용하여 수행했다. 피크 6은 170 kDa, 90 kDa, 66 kDa, 59.5 kDa 및 31 kDa의 면역반응성 밴드를 함유하고 있었다. 이들은 TcaC (166 kDa), TcaA_{ii} + TcaA_{iii} (92 kDa), TcaA_{iii} (66 kDa), TcaB_i (60 kDa) 및 TcaA_{ii} (25 kDa)에 대해 각각 예상한 크기에 매우 근접한 것이다. 피크 6은 본 명세서에 기재된 바와 같이 천연 아가로오즈 겔 전기영동으로 추가로 분석한 밴드 1의 것과 유사한 이동도를 갖는 단일 밴드로 이동했다.

정제된 피크 6 독소 단백질의 단백질 농도는 BCA 시약 (Pierce)을 사용하여 측정했다. 이 단백질의 희석물은 10 mM Tris(pH 8.6)로 제조한 후 식이 생검에 사용했다. 240 시간 후, 식이 생검을 통해 피크 6 단백질 분획물을 450 ng 또는 그 이상 투여받은 신생 유충은 모두 죽었다. 동일한 분획물 90 ng을 섭취한 유충 군은 40%의 치사율을 나타냈다. 240 시간 후에 피크 6 단백질 분획물 90 ng 및 20 ng을 투여받고 살아남은 것은 각각 대조구 총량의 약 10% 및 70%였다.

<실시예 2>

살충제 이용성

포도랍두스 루미네센스의 이용성 및 독성을 추가로 특성화하였다. 포도랍두스 루메네센스(균주 W-14) 배양 브로쓰는 하기와 같이 생산하였다. 생산 배지는 Milli-Q(등록상표) 탈이온수 중의 2% 백토 프로테오스 펄톤(등록상표) #3(PP3, Difco Laboratories, Detroit, Michigan)이었다. 175ml의 배지로 이루어지는 종자 배양물 플라스크를 카푹(Kaput)으로 덮인 데롱 목(Delong neck)이 달린 500ml의 트리배플드 플라스크에 담고, T= 121°C(250°F)에서 20분 동안 오토클레이브시켰다. 생산 플라스크는 신-에쯔(Shin-etsu) 실리콘 포움 마개로 덮인 데롱 목이 달린 500ml 트리배플드 플라스크에서 2.8ℓ 중 500ml로 이루어진다. 이들을 T= 121°C(250°F)에서 45분 동안 오토클레이브시켰다. 종자 배양물을 6.08 cm(2인치) 작동 반경의 회전 진탕 배양기 내에서 150rpm에서 28°C에서 배양시켰다. 16시간 동안 성장시킨 후, 1%의 종자 배양물을 생산 플라스크내에 놓고 회수하기 전 24시간 동안 성장시켰다. 독소의 생산은 로그상 성장동안 나타났다. 미생물 브로쓰를 1ℓ 원심분리병에 옮기고 세포 생물량을 펠릿화하였다(2500 RPM 4°C에서 30분[R.C.F.~1600] HG-4L Rotor RC3 Sorval centrifuge, Dupont, Wilmington, Delaware). 제 1 브로쓰를 4°C에서 8 내지 16시간 동안 냉장시키고, 적어도 2시간 동안 재원심분리시켜(상기 조건) 정지시에 침전되는 추정 무코폴리사카라이드를 제거함으로써 브로쓰를 추가로 정화시켰다. (대안적인 프로세싱 방법은 두 단계를 합하고, 상기와 동일한 조건하에 16시간의 정화 원심분리의 사용을 포함한다.) 이어서, 이 브로쓰를 생분석 또는 여과하기 전에 4°C에서 저장하였다.

포도랍두스 배양 브로쓰, 및 이 브로쓰로부터 정제된 단백질 독소(들)은 많은 곤충에 대하여 활성(치사 및(또는) 성장 억제, 감소된 성충 출현)을 나타냈다. 더 구체적으로, 활성은 딱정벌레목의 곤충의 구성원인 옥수수 뿌리벌레(유충 및 성충), 콜로라도 감자 딱정벌레, 및 잔디 땅벌레(turf grub)에 대하여 나타났다. 딱정벌레목의 다른 구성원은 방아벌레(wireworm), 화분(pollen) 딱정벌레, 뽕벼룩 갑충(flea beetle), 종자 딱정벌레 및 바구미를 포함한다. 활성은 또한 동시목의 구성원인 별 매미충(aster leafhopper)에 대해서도 관찰되었다. 동시목의 다른 구성원은 열구(planthopper), 배나무이(pysllia), 사과나무이(sucker), 개각충(scale insect), 화이트플라이(whiteflies) 및 침(spittle)벌레 뿐만 아니라, 많은 숙주 특이성 진딧물 종을 포함한다. 브로쓰 및 정제된 분획은 또한 사탕무 행렬구더기(armyworm), 양배추 자벌레, 블랙 야도충(black cutworm), 회색담배나방, 유럽 곡물 천공충, 옥수수 귀벌레(earworm) 및 코들링 나방(이들은 모두 나비목의 구성원임)에 대하여 활성이었다. 이 목의 다른 전형적인 구성원은 웃줄 나방, 화랑곡나방(Indian mealmoth), 잎말이나방(leaf roller), 배추벌레, 목화다래나방(cotton ballworm), 도롱이벌레(bagworm), 이스턴 텐트 나방(Eastern tent caterpillar), 잔디 포충나방(sod webworm), 및 풀 행렬구더기(fall armyworm)를 포함한다. 활성은 또한 파리목의 구성원인 과일파리 및 모기 유충에 대해서도 나타났다. 파리목의 다른 구성원은 완두 미지(midge), 당근 파리, 배추 뿌리 파리, 순무 뿌리 파리, 양파 파리, 크레인 파리, 집파리 및 다양한 모기 종을 포함한다. 또한 볼개미, 오더러스(oderous) 집개미 및 작은 흑개미를 포함하는 목의 구성원인 카펜터 개미 및 아르헨티나 개미에 대해서도 활성이 나타났다.

브로쓰/분획은 곤충의 군집을 감소시키기 위해 유용하며, 곤충 군집을 억제하는 방법에서 사용되었다. 방법은 전술된 유효한 곤충 불활성화량의 활성제를 곤충의 지역에 적용하는 것을 포함할 수 있다. 결과를 표 4에 기록하였다.

옥수수 뿌리벌레 유충에 대한 활성을 하기와 같이 시험하였다. 포도랍두스 배양 브로쓰(여과 멸균된, 세포 없음) 또는 정제된 HPLC 분획을 30μl 분취액내 인공 먹이의 0.25ml의 표면(~1.5cm²)에 직접 적용하고, 각각 대조 배지 또는 10mM 인산나트륨 완충액(pH 7.0)으로 희석하였다. 먹이 평판을 멸균 유동 후드내에서 공기 건조시키고, 웰을 멸균된 알로부터 부화된 단일, 신생 디아브로

티카 운대심푼크타타 호워드(*Diabrotica undecimpunctata howardi*)(써던 옥수수 뿌리벌레, SCR)로 채우고, 2회령 SCR을 인공 먹이에서 성장시키거나, 2회령 디아브로티카 비르기페라 비르기페라(*Diabrotica virgifera virgifera*)(웨스턴 옥수수 뿌리벌레, WCR)을 메트로믹스(Metromix)(등록상표)에서 성장된 묘목 상에 재배하였다. 2회령 유충을 먹이를 첨가하기 전에 칭량하였다. 평판을 밀봉하고, 가슴된 성장 체임버에 놓고, 27°C에서 적절한 기간(신생 및 성충 SCR에 대해 4일, WCR 유충에 대해 2 내지 5일, 2회령 SCR에 대해 7 내지 14일). 치사율 및 무게 측정을 지시된 바와 같이 기록하였다. 일반적으로, 모든 연구에서 처리 당 16마리의 곤충을 사용하였다. 대조군 치사율은 하기와 같았다: 신생 유충: <5%, 성충 딱정벌레: 5%.

콜로라도 감자 딱정벌레에 관한 활성은 다음과 같이 시험하였다. 포도랍두스 배양 브로쓰 또는 대조군 배지를 24-웰 조직 배양 플레이트의 웰 중에 유지된 표준 인공 사료의 1.5 ml의 표면(~2.0 cm²)에 가하였다. 각각의 웰은 50 µl 씩 처리되었고 방치하여 공기 건조시켰다. 이어서, 개별적인 2회령의 콜로라도 감자 딱정벌레(*Leptinotarsa decemlineata*, CPB) 유충을 상기 사료 상에 위치시키고 치사율을 4일 후에 기록하였다. 처리 당 10 개의 유충을 모든 경우의 연구에 사용하였다. 대조군의 사망률은 3.3%이었다.

일본 딱정벌레 굽병이 및 딱정벌레에 대한 활성은 다음과 같이 시험하였다. 잔디 굽병이(*Popillia japonica*, 2-3회령)을 침입받은 잔디로부터 수집하여 추가의 사료로서 첨가된 당근 조각이 포함된 토양/토탄 중에 실험실에서 유지시켰다. 잔디 딱정벌레들은 국부적으로 페로몬으로 덮을 놀아 잡아두고 사료로서 단풍나무 잎이 첨가된 플라스틱 용기 중에 실험실에서 유지시켰다. 무희석 포도랍두스 배양 브로쓰 또는 대조군 옥수수 배지를 뿌리벌레 인공 사료(30 µl/1.54 cm²) 또는 당근 조각 (유충)에 가한 후, 두 단계들을 단독으로 사료 웰 중에 위치시키고 임의의 치사율 및 섭취를 관찰하였다. 모두의 경우에 있어서 관찰된 섭취량 (배설물 생산)에서 명백한 감소가 있었다.

모기 유충에 대한 활성은 다음과 같이 시험하였다. 분석은 96-웰 마이크로타이더 플레이트에서 수행하였다. 각각의 웰은 수용액 (포도랍두스 배양 브로쓰, 대조군 배지 또는 H₂O) 200 µl 및 대략 20 개씩, 1일령의 유충(애데스 애집티)을 포함하였다. 처리 당 6 개의 웰이 존재하였다. 결과는 침입 후 2 시간이 지나서 판독하였고 3일의 관찰 기간 동안 변화되지 않았다. 대조군 치사율은 관찰되지 않았다.

초파리에 대한 활성은 다음과 같이 수행하였다. 구매한 초파리 (*Drosophila melanogaster*) 배지를 50% 건조 배지와 물이나 대조군 배지나 포도랍두스 배양 배지 중 어느 하나의 50% 용액을 사용하여 제조하였다. 이는 처리 당 3 개의 사육 바이알 각각 중에 건조 배지 8 ml를 넣고, 적절한 용액 8 ml를 첨가함으로써 이루어졌다. 이어서 10회 후기 령의 초파리 구더기를 각 바이알에 첨가하였다. 바이알들은 형광 천장의 조명하에 실온에서 실험실 벤취 상에 유지시켰다. 새끼 또는 성체 계수는 노출 3, 7 및 10일 후에 행하였다. 포도랍두스 배양 브로쓰를 초파리 구더기 사료 배지로 혼입하면 10일째 성체 출현에 있어서 약간의 (17%) 감소가 나타나지만, 이는 물 및 대조군 배지를 사용한 경우(3%)에 비해서는 상당한 감소를 유발하였다.

별 매미충에 대한 활성은 다음과 같이 시험하였다. 별 매미충 (*Macrosteles severini*)에 대한 섭취 분석은 다른 외부의 접촉이 없이 적극적인 섭취를 허용하도록 설계하였다. 능동적/"사료" 용액의 저장용기는 35 x 10 mm 페트리 접시의 바닥 부분의 중앙에 2 개의 구멍을 내어 만들었다. 6.08 cm (2 인치) Parafilm M(등록상표) 스퀘어를 접시의 상부를 가로질러 올려놓고 "0" 링으로 닫았다. 이어서 28.3 g(1 온스) 플라스틱 컵에 대략 7마리의 매미충을 넣고 상기 저장용기를 컵 위에, 파라필름이 밑으로 가도록 올려놓았다. 이어서 시험 용액을 구멍을 통하여 상기 저장용기에 첨가하였다. 무희석 포도랍두스 배양 브로쓰를 사용한 시험에 있어서, 브로쓰 및 대조군 배지를 물에 대하여 투석시켜 대조군 치사율을 감소시켰다. 치사율은 2 일째에 보고하였고 여기서 26.5%의 대조군 치사율이 나타났다. 정제된 분획 (200 mg 단백질/ml)을 사용한 시험에 있어서, 5% 수크로오스의 최종 농도를 모든 처리에 사용하여 별 매미충의 생존능을 향상시켰다. 이 분석물은 배양기에서 16/8의 광조사 기간의 28 °C, 70% RH로 유지시켰다. 분석물은 72 시간 생존능에 대하여 등급을 매겼다. 대조군 치사율은 5.5%이었다.

아르헨티나 개미에 대한 활성은 다음과 같이 시험하였다. 100% 포도랍두스 배양 브로쓰, 대조군 배지 또는 물의 1.5 ml 분취량을 피펫팅하여 2.0 ml 투명 유리 바이알 내로 가하였다. 이 바이알들을 적절한 처리에 의해 젖게한 목화 덴탈 워(dental wick) 조각으로 틀어막았다. 각각의 바이알을 8 내지 12 마리의 성체 아르헨티나 개미 (*Linepithema humile*)를 담은 별도의 60 x 16 mm 페트리 접시 내에 위치시켰다. 처리 당 3회의 반복 실시했다. 생물학적 시험 플레이트는 형광 천장의 조명 하에 실온에서 실험실 벤취 상에 유지시켰다. 치사율 판독은 노출후 5일만에 행하였다. 대조군 치사율은 24%이었다.

카펜터 개미에 대한 활성은 다음과 같이 시험하였다. 블랙 카펜터 일개미(*Camponotus pennsylvanicus*)를 인디애나주 인디애나폴리스의 DowElanco 소유의 나무에서 수집하였다. 포도랍두스 배양 브로쓰를 사용한 시험을 다음과 같이 수행하였다. 각각의 플라스틱 생물학적 시험 용기 (7 1/8" x 3")는 플라스틱 단 유리 중에 15 마리의 일개미, 종이 피난처 및 브로쓰 또는 대조군 배지 10 ml를 포함하였다. 목화 워는 짧은 유리 덮개 중의 구멍을 통하여 개미에 대한 처리물에 전달되었다. 모든 처리물은 5% 수크로스를 포함하였다. 생물학적 시험물들은 실온의 암실에서 유지되었고 19 일 동안 채점되었다. 대조군 치사율은 9%이었다. 정제된 분획을 운반하는 분석물들은 플라스틱 시험관 중에 0.2 ml 처리/2.0 g 사료의 속도로 처리물(정제된 분획 또는 대조군 용액)과 혼합된 인공 개미 사료를 이용하였다. 정제 분획의 최종 단백질 농도는 10 µg/g 사료 미만이었다. 처리 당 10 마리의 개미, 물 공급원, 피난처 및 처리된 사료를 밀봉 플라스틱 용기 내에 위치시키고 습기가 있는 배양기 중의 27 °C의 암실에서 유지시켰다. 치사율은 10일에 기록하였다. 대조군 치사율은 나타나지 않았다.

다양한 나비목 유충에 대한 시험은 다음과 같이 시험하였다. 포토랍두스 배양 브로쓰 또는 정제된 분획을 각각 대조군 배지 또는 10 mM 인산나트륨 완충액, pH 7.0 중에 희석한 후 30 μ l 분취량 중의 표준 인공 사료 0.25 ml의 표면 (~1.5 cm²)의 표면에 직접적으로 가하였다. 사료 플레이트를 멸균 플로우 후드 중에서 공기 건조시키고 웰들을 단독의 신생 유충으로 섭취시켰다. 유럽 옥수수 천공충 (*Ostrinia nubilalis*) 및 옥수수 귀벌레 (*Helicoverpa zea*) 알들을 상업적인 공급원으로 부터 공급받아서 우리 중에서 부화시킨 반면, 사탕무 행렬구더기 (*Spodoptera exigua*), 양배추 자벌레 (*Trichoplusia ni*), 회색담배나방(*Heliothis virescens*), 사과 나방 (*Laspeyresia pomonella*) 및 블랙 야도충 (*Agrotis ipsilon*) 유충은 내부적으로 공급받았다. 유충으로 침입시킨 후, 사료 플레이트들을 밀봉하고, 습기가 있는 성장 챔버 중에 위치시켜서 적절한 기간 동안 27 °C의 암실에서 유지시켰다. 치사율 및 중량 측정은 포토랍두스 배양 배지에 대해서 5-7 일에 기록하였고, 정제된 분획에 대해서는 4-7일에 기록하였다. 일반적으로 처리당 16마리의 곤충이 모든 연구에서 사용되었다. 대조군 치사율은 대조군 배지에 대해서 4 - 12.5%의 범위이었고 인산염 완충액에 대해서는 10% 미만이었다.

[표 4]

포토랍두스 루미네센스 (균주 W-14) 배양 배지 및 정제된 독소 분획의 상이한 곤충목/종의 치사율 및 성장 억제에 미치는 효과

| 곤충 목/종 | | 브로쓰 | | 정제된 분획 | |
|--------|--------------|------|-------|--------|-------|
| | | 사망% | G.I.% | 사망% | G.I.% |
| 딱정벌레목 | 옥수수뿌리벌레 | | | | |
| | 써던/신생유충 | 100 | na | 100 | na |
| | 써던/2회령 | na | 38.5 | nt | nt |
| | 써던/성체 | 45 | nt | nt | nt |
| | 웨스턴/2회령 | na | 35 | nt | nt |
| | 콜로라도 감자 딱정벌레 | | | | |
| | 2회령 | 93 | nt | nt | nt |
| | 잔디 궁병이 | | | | |
| | 2회령 | na | a.f. | nt | nt |
| | 성체 | na | a.f. | nt | nt |
| 파리목 | | | | | |
| | 초파리 (성체 출현) | 17 | nt | nt | nt |
| | 모기 유충 | 100 | na | nt | nt |
| 동시목 | | | | | |
| | 별 매미충 | 96.5 | na | 100 | na |
| 벌목 | | | | | |
| | 아르헨티나 개미 | 75 | na | nt | na |
| | 카펜터 개미 | 71 | na | 100 | na |

| | | | | |
|------------|------|------|-------|------|
| 나비목 | | | | |
| 사탕무 행렬구더기 | 12.5 | 36 | 18.75 | 41.4 |
| 블랙 야도충 | nt | nt | 0 | 71.2 |
| 양배추 자벌레 | nt | nt | 21.9 | 66.8 |
| 사과 나방 | nt | nt | 6.25 | 45.9 |
| 옥수수 귀벌레 | 56.3 | 94.2 | 97.9 | na |
| 유럽 옥수수 천공충 | 96.7 | 98.4 | 100 | na |
| 담배 모충 | 13.5 | 52.5 | 19.4 | 85.6 |

G. I.= 성장 억제,
na = 적용불가, nt = 시험되지 않음, a.f. = 안티-사료화제

<실시예 3>

토양 적용시 살충제 이용성

포도랍두스 루미네센스(균주 W-14) 배양 브로쓰는 토양 또는 토양 혼합물 (메트로믹스(Metromix)(등록상표)에 직접 가하였을 때 옥수수 뿌리벌레에 대하여 활성인 것으로 나타났다. 메트로믹스(등록상표) 중의 신생 SCR 및 WCR에 대한 활성은 다음과 같이 시험하였다 (표 5). 시험은 습윤 여과지 상의 빛 중에서 6 일간 발아시킨 옥수수 묘목 (United Agriseeds brand CL614)를 사용하여 진행시켰다. 뿌리가 대략 3 - 6 cm 길이어 있을 때, 단일 인/묘목을 무수 메트로믹스(등록상표) 50g를 포함하는 투명 플라스틱 컵 591 ml 중에 식묘하였다. 이어서 20 개의 신생 SCR 또는 WCR를 묘목의 뿌리 상에 직접 위치시키고 메트로믹스(등록상표)로 덮었다. 침습시, 묘목들을 희석 브로쓰 용액의 50 ml 전체 부피로 적셨다. 적신 후, 컵들을 밀봉시키고 빛 중의 실온에서 7 일 동안 방치하였다. 그 후, 묘목들을 세척하여 메트로믹스(등록상표)를 제거하고 뿌리를 절단하여 중량을 재었다. 활성은 대조군 식물에 대하여 남아있는 옥수수 뿌리의 비율의 백분율 및 섭취에 의해 유도된 잎 손상으로서 평가하였다. 잎 손상은 육안으로 관찰하였고 -, +, ++, 또는 +++ 중의 하나로서 평가하였으며, -는 무손상을 나타내고, +++는 심한 손상을 나타낸다.

토양에서 신생 SCR에 대한 활성은 다음과 같이 시험하였다(표 6). 시험은 습윤 여과지 상의 빛 중에서 6 일간 발아시킨 옥수수 묘목 (United Agriseeds brand CL614)를 사용하여 진행시켰다. 뿌리가 대략 3 - 6 cm 길이어 있을 때, 단일 인/묘목을 전년도에 옥수수를 식묘한 레바논(Lebanon)의 야외 토양 150 g을 포함하는 투명 플라스틱 컵 591 ml 중에 식묘하였다. 이 토양은 이전에 살충제로 처리하지 않았다. 이어서 20 개의 신생 SCR을 묘목의 뿌리 상에 직접 위치시키고 토양으로 덮었다. 침습후, 묘목들을 희석 브로쓰 용액의 50 ml 전체 부피로 적셨다. 적신 후, 밀봉되지 않은 컵들을 26.7 °C(78 °F)에서 높은 상대습도 (80%)의 챔버 중에서 배양시켰다. 그 후, 묘목들을 세척하여 모든 토양을 제거하고 뿌리를 절단하여 중량을 재었다. 활성은 대조군 식물에 대하여 남아있는 옥수수 뿌리의 비율의 백분율 및 섭취에 의해 유도된 잎 손상으로서 평가하였다. 잎 손상은 육안으로 관찰하였고 -, +, ++, 또는 +++ 중의 하나로서 평가하였으며, -는 무손상을 나타내고, +++는 심한 손상을 나타낸다.

[표 5]

포도랍두스 루미네센스 (균주 W-14) 배양 브로쓰의 침입후 적신 후의 뿌리벌레 유충에 대한 효과 (메트로믹스(등록상표))

| 처리 | 유충 | 잎 손상 | 뿌리 중량 (g) | % |
|-----------------|----|------|----------------|------|
| 써던 옥수수 뿌리벌레 | | | | |
| 물 | - | - | 0.4916 ± 0.023 | 100 |
| 배지 (2.0% v/v) | - | - | 0.4416 ± 0.029 | 100 |
| 브로쓰 (6.25% v/v) | - | - | 0.4641 ± 0.081 | 100 |
| 물 | + | +++ | 0.1410 ± 0.006 | 28.7 |
| 배지 (2.0% v/v) | + | +++ | 0.1345 ± 0.028 | 30.4 |
| 브로쓰 (1.56% v/v) | + | - | 0.4830 ± 0.031 | 104 |
| 웨스턴 옥수수 뿌리벌레 | | | | |
| 물 | - | - | 0.4446 ± 0.019 | 100 |
| 브로쓰 (2.0% v/v) | - | - | 0.4069 ± 0.026 | 100 |
| 물 | + | - | 0.2202 ± 0.015 | 49 |
| 브로쓰 (2.0% v/v) | + | - | 0.3879 ± 0.013 | 95 |

[표 6]

포도랍두스 루미네센스 (균주 W-14) 배양 브로쓰의 침입후 적신 후의 써던 뿌리벌레 유충에 대한 효과 (토양)

| 처리 | 유충 | 잎 손상 | 뿌리 중량 (g) | % |
|---------------|----|------|----------------|-----|
| 물 | - | - | 0.2148 ± 0.014 | 100 |
| 배지 (2.0% v/v) | - | - | 0.2222 ± 0.013 | 100 |

등록상표)를 투명 플라스틱 컵 591 ml에 첨가하였다. 이어서, 이 메트로믹스(등록상표)를 희석 포도랍두스 브로쓰 용액 50% (v/v)의 50 ml의 전체 부피로 적셨다. 조 브로쓰의 희석은 물로 행하였고, 50 ml 전체 부피에 대해 조 브로쓰 25 ml을 25 ml에 첨가함으로써 50% 브로쓰를 제조하였다. 정상 배지 농도의 50% 희석인 프로테오스 펩톤 #3(pp3)의 1% (w/v) 용액을 브로쓰 대조군으로서 사용하였다. 적신 후, 5 개의 제2령 잔디 곰팡이 들을 습윤화된 메트로믹스(등록상표)의 상부에 위치시켰다. 건강한 잔디 곰팡이 유충은 메트로믹스(등록상표) 속으로 신속하게 굴을 파고 은신하였다. 1 시간 내에 은신하지 못한 유충들은 제거하고 새 유충으로 대체하였다. 컵을 밀봉시키고 28 °C의 암실 배양기 중에 위치시켰다. 7일 후, 유충들을 메트로믹스(등록상표)로 부터 제거하고 치사율을 기록하였다. 활성은 대조군에 대한 치사율의 백분율로 평가하였다.

[표 7]

포도랍두스 루미네센스 (균주 W-14) 배양 브로쓰의 침입후 적신 후의 잔디 곰팡이에 대한 효과 (메트로믹스(등록상표))

| 처리 | 사망* | 사망% |
|--------------------|-------|-----|
| 물 | 7/15 | 47 |
| 대조 배지(1.0 % v/v) | 12/19 | 63 |
| 브로쓰 (50% v/v) | 17/20 | 85 |
| * 치사/생존 유충의 비로 표시함 | | |

<실시에 4>

잎 적용시 살충제 이용성

유럽 옥수수 천공충에 대한 포도랍두스 브로쓰의 활성은 브로쓰를 옥수수 잎의 표면에 직접 적용했을 때 관찰하였다(표 8). 이들 분석에 있어서, 포도랍두스 브로쓰는 배양 배지로 100 배 희석하고 잘라진 옥수수 잎의 표면에 잎 표면의 ~ 6.0 µl/cm²의 속도로 손수 가하였다. 잎들을 공기 건조시키고 대략 2 x 2 인치의 동일한 크기의 조각으로 절단하였다. 잎들을 말아서 종이 클립으로 묶고 28.3 g(1 온스) 플라스틱 단 유리 중에 2%의 아가의 0.64 cm (0.25 인치)가 습기를 제공하도록 바닥 표면에 있게 위치시켰다. 이어서, 12 개의 신생 유럽 옥수수 천공충을 말린 잎 상에 위치시키고 컵을 밀봉하였다. 27 °C의 암실에서 5일 동안 배양시킨 후, 시료들을 섭취 손상 및 회복된 유충에 대하여 기록하였다.

[표 8]

포도랍두스 루미네센스 (균주 W-14) 배양 브로쓰의 침입후 절단 옥수수 잎에 적용한 후의 유럽 옥수수 천공충 유충에 대한 효과

| 처리 | 잎손상 | 회복된 유충 | 중량 (mg) |
|----------------|-----|--------|---------|
| 물 | 광범위 | 55/120 | 0.42 mg |
| 대조군 배지 | 광범위 | 40/120 | 0.50 mg |
| 브로쓰 (1.0% v/v) | 근소함 | 3/120 | 0.15 mg |

배양 브로쓰의 신생 회색담배나방(*Heliothis virescens*)에 대한 활성은 잎 침지 방법론을 사용하여 입증하였다. 프랑스 목화 잎을 식물로부터 잘라내어 잎 디스크를 18.5 mm 코르크-끝로 절단하였다. 디스크를 개별적으로 대조군 배지 (pp3) 또는 10 kDa 필터를 가진 Amicon (매사추세츠주 Beverly 소재) Profflux M12 트랜스제니탈 여과 시스템을 사용하여 대략 10배로 농축시킨 포도랍두스 루미네센스 (균주 W-14) 배양 브로쓰 중에 침지시켰다. 과량의 용액을 제거하고 똑바로 편 종이 클립을 디스크의 중앙을 통해 위치시켰다. 이어서 종이 클립을 1% 아가 대략 2.0 ml을 함유하는 플라스틱, 28.3 g(1.0 온스) 단유리 내에 밀어 넣었다. 이는 아가 상에 잎 디스크를 매달리게 하는 역할을 하였다. 잎 디스크의 건조 후, 단일 신생 회색담배나방 유충을 디스크 상에 위치시키고 컵을 클램핑시켰다. 이어서 컵들을 플라스틱 백 중에 밀봉시키고 암화된 27 °C 배양기 중에서 5 일 동안 위치시켰다. 이때 잔여 유충 및 잎 재료들을 중량을 재어 잎 손상에 대한 척도로 설정하였다 (표 9).

[표 9]

포도랍두스 루미네센스 (균주 W-14) 배양 브로쓰의 목화 잎 침지 분석에 있어서 회색담배나방 신 생물 상에 대한 효과

| 처리 | 잎 디스크 | 최종 중량(mg) 유충 |
|--------------------------------|------------|--------------|
| 대조 잎 | 55.7 ± 1.3 | na* |
| 대조 배지 | 34.0 ± 2.9 | 4.3 ± 0.91 |
| 포도랍두스 브로쓰 | 54.3 ± 1.4 | 0.0** |
| * - 적용 불가, ** - 생존 유충이 발견되지 않음 | | |

<실시에 5>

파트 A

독소 펩티드 성분의 특성화

후속되는 분석에 있어서, 실시예 1에서 분리된 밴드의 독소 단백질 소단위들을 30:0.8 (아크릴아미드:비스-아크릴아미드)의 비율의 7% SDS 폴리아크릴아미드 전기영동 겔 상에 분해시켰다. 이 겔 매트릭스는 더 큰 단백질에 대한 더 양호한 해상도를 촉진하였다. 실시예 1에서 밴드 1 및 밴드 2 소단위 분자량을 측정하는데 사용했던 겔 시스템은 38:0.8 (아크릴아미드:비스-아크릴아미드)의 18% 겔이었고, 이는 넓은 범위의 크기 분리를 허용하였으나 고분자량 성분의 해상력은 낮았다.

이 분석에 있어서, 8개 보다는 10개의 단백질 밴드가 분해되었다. 표 9는 10개의 해상된 밴드의 이론적 분자량을 나타내고 이들 조건하에서 측정된 분자량을 선행 실시예의 것에 대하여 직접적으로 비교하였다. 추가의 밴드가 이 실시예에서 사용된 상이한 분리 조건하에서 검출되었다는 것은 놀라운 것이 아니다. 분자량에 대한 선행 및 신규 측정 사이의 편차는 또한 분석 조건에 있어서 차이에 따라 예견될 수 있는 것이다. 본 실시예의 분석에 있어서 고분자량 측정물은 개선된 해상력의 결과로서 실시예 1에서의 것보다 더 정확한 것으로 사료된다. 그러나, 이들은 SDS PAGE 분석에 기초한 측정 결과이며 이는 전형적으로 분석 상 정확하지 않은 것이며 펩티드의 추정을 초래하며 후- 및 동시-번역 수식에 기인하여 더욱 변경될 수 있다.

아미노산 서열들을 10개의 해상된 펩티드 중 5 개의 N-말단 부분에 대하여 결정하였다. 표 10는 단백질의 분자량 및 동정된 서열과 상관 관계가 있다. 서열 2에 있어서, 특정 분석은 잔기 5의 프롤린이 아스파라긴 (asp)일 수 있음을 암시한다. 서열 3에 있어서, 특정 분석은 잔기 13 및 14의 아미노산 잔기가 모두 아르기닌 (arg)일 수 있음을 암시한다. 서열 4에 있어서, 특정 분석은 잔기 6의 아미노산 잔기가 알라닌(ala) 또는 세린 (ser) 중 하나 일 수 있음을 암시한다. 서열 5에 있어서, 특정 분석은 잔기 3의 아미노산 잔기가 아스파르탐산 (asp)일 수 있음을 암시한다.

[표 10]

| 실시예 1 측정 | 새로운 측정* | 서열 목록 |
|----------|-----------|-------|
| 208 | 200.2 kDa | 서열 1 |
| 184 | 175.0 kDa | 서열 2 |
| 65.6 | 68.1 kDa | 서열 3 |
| 60.8 | 65.1 kDa | 서열 4 |
| 56.2 | 58.3 kDa | 서열 5 |
| 25.1 | 23.2 kDa | 서열 15 |

* 새로운 측정은 SDS PAGE에 기초한 것이고 유전자 서열에 기초하지 않았다. SDS PAGE는 분석상 정확도가 없다.

<실시에 5>

파트 B

독소 펩티드 성분의 특성화

새로운 N-말단 서열, 서열 15, Ala Gln Asp Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr은 상기 실시예 5, 파트 A에서 기술한 천연 HPLC 정제된 독소로부터 분리된 펩티드의 추가의 N-말단 서열화에 의해 얻었다. 이 펩티드는 tcaA 유전자로부터 유래된다. 이 펩티드 표지 TcaA_{iii}는 위치 254에서 시작하여 위치 491까지 펼쳐있고, 여기서 TcaA_{iii} 펩티드, 서열 4가 시작된다. 유전자 서열에 기초한 펩티드의 추정 크기는 25,240 Da이다.

<실시에 6>

독소 펩티드 성분의 특성화

더욱 다른 분석에 있어서, 독소 단백질 복합체를 포도랍두스 루미네센스 성장 배지(Tween의 부재 하에서 배양 후)로부터 10% - 80% 암모늄 설페이트 침전에 이어 이온 교환 크로마토그래피 단계 (모노 Q) 및 두 분자 사이징 크로마토그래피 단계를 수행함으로써 분리하였다. 이들 조건들은 실시예 1에서 사용된 것과 유사하다. 처음의 분자 사이징 단계 동안, 두번째의 생물학적으로 활성인 피크는 약 100 ± 10 kDa에서 발견되었다. 단백질 측정에 기초하여 이 단편은 더 큰 또는 처음의 약 860 ± 100 kDa(천연)의 활성 피크 보다 20 - 50 배 덜 활성이었다. 이 분리 실험 동안, 출발 생물학적 활성의 상당부를 함유하였던 약 325 ± 50 kDa의 덜 활성인 피크 역시 해상되었다. 325 kDa 피크는 860 kDa 피크에 관련되거나 또는 이로부터 유도된 것으로 생각된다.

56 kDa 단백질이 이 분석에서 해상되었다. 이 단백질의 N-말단 서열은 서열 6에 나타났다. 이 단백질이 N-말단에서 서열 5와 상당한 동일성 및 보존성을 나타내는 것이 주목할만 하며, 이는 이들이 유전자 집단의 상이한 원에 의해 코딩 될 수 있고 각 유전자에 의해 생성된 단백질은 이들이 살충의 독소 복합체로서 작용하기에 서로에 대해 충분히 유사함을 암시한다.

두번째, 두드러진 185 kDa 단백질은 표 10로부터 단백질 3의 것에 비교할 만한 양으로 일관되게 존재하였으며 동일한 단백질 또는 단백질 단편일 것이다. 이 185 kDa 단백질의 N-말단 서열은 서열 7에 나타났다.

추가로 N-말단 아미노산 서열 데이터는 또한 단리된 단백질들로부터 얻었다. 결정된 N-말단 서열의 어느 것도 표 10에서 동정된 단백질에 동일하게 나타나지 않았다. 기타의 단백질들이 단리된 제제물에 존재하였다. 하나의 그러한 단백질은 108 kDa의 추정 분자량을 가졌고 서열 8에 나타낸 바와 같은 N-말단 서열을 가졌다. 두번째의 그러한 단백질은 80 kDa의 추정 분자량을 가졌고 서열 9에 나타낸 바와 같은 N-말단 서열을 가졌다.

대략 325 kDa 활성 피크에서 단백질 물질이 크기에 의해 분석되었을 때, 대략 51, 31, 28 및 22 kDa의 밴드들이 관찰되었다. 분자량이 전기영동적 이동의 분석에 의해 결정된 모든 경우에서와 같이, 이들 분자량은 완충액 이온 강도 차이, 전기영동 전력 차이 등에 의해 야기되는 오류 효과를 받았다. 당업자는 결정적인 분자량값은 이들 표준 방법에 의해 결정될 수 없으며 각각은 편차를 갖게 될 것임을 이해할 것이다. 이들 크기의 단백질은 더욱 큰 초기의 독소 복합체에서 관찰된 보다 큰 단백질 중 (대략 200 kDa 크기의)의 분해 산물인 것으로 가정되었다.

마지막으로, 몇몇 제제는 서열 10에 나타낸 N-말단 서열을 갖는 단백질을 포함하였다. 이 서열은 큰 단백질 복합체의 어셈블리에서 기능을 하는 것으로 알려진 보조 단백질인 카페로닌 단백질에 강하게 상동성이었다. 출원인은 서열 10에서 동정된 단백질에 대한 그러한 어셈블리 기능을 규정할 수는 없지만, 그러한 카페로닌 단백질이 그의 어셈블리에서 수반될 수 있는 기술된 독소 단백질 복합체의 존재와 일치하였다. 더욱이, 그러한 단백질은 독성 활성을 갖는 것으로 직접적으로 제안될 수 없으나, 이 단백질은 단백질 독소의 전체의 구조적 성질을 결정하는데 중요할 수 있으며, 따라서 독성 활성 또는 경구 전달 후 복합체의 생체내 생존력에 기여할 것이다.

프로테이나제 K에 대한 단백질 독소 복합체의 안정성에 관한 후속 분석을 수행하였다. 10몰 과량의 프로테이나제 K의 존재 중에서 복합체를 24 시간 배양 시킨 후, 활성이 궁극적으로 제거되었다(경구 적용시 치사율이 약 5%까지 떨어졌다)는 것이 확인되었다. 이들 데이터는 독소의 단백질 분해 효소적 성질을 입증한다.

독성 활성은 투석 멤브레인에 의해서도 유지되었으며 이는 재차 큰 크기의 천연 독소 복합체를 확인한다.

<실시예 7>

포도랍두스 독소의 단리, 특성화 및 부분적인 아미노산 서열화

단리 및 N-말단 아미노산 서열화: 실시예 5 및 6에 비슷하게 행해진 셋트의 실험에 있어서, 포도랍두스 단백질의 암모늄 설페이트 침전은 포도랍두스 브로쓰, 전형적으로 2 - 3 리터를 암모늄 설페이트 결정물 서서히 첨가함으로써 10% 또는 20% 중 어느 하나의 최종 농도까지 조정함으로써 수행하였다. 4°C에서 1 시간 동안 교반시킨 후, 이 재료를 12,000 x g에서 60 분 동안 원심분리시켰다. 상층액을 80% 암모늄 설페이트로 조정하고 4°C에서 1 시간 동안 교반시킨 후, 12,000 x g에서 60 분 동안 원심분리시켰다. 펠릿을 10 mM Na₂·PO₄, pH 7.0의 1/10 부피 중에 재현탁시키고 동일한 인산염 완충액에 대하여 4°C에서 밤새 투석시켰다. 투석된 재료를 이온 교환 크로마토그래피에 앞서 12,000 x g에서 1 시간 동안 원심분리시켰다.

HR 16/50 Sepharose (Pharmacia) 음이온 교환 컬럼을 10 mM Na₂PO₄, pH 7.0으로 평형시켰다. 원심분리시키고, 투석시킨 암모늄 설페이트 펠릿을 Q Sepharose 컬럼에 1.5 ml/분의 속도로 가하고 광학적 밀도 (O.D. 280)이 0.100 미만에 도달할 때까지 평형 완충액으로 3.0 ml/분의 속도로 활발하게 세척하였다. 다음에, 0 내지 0.5 M에 걸쳐 3ml/분의 60 분 동안 NaCl 구배 또는 60 분 동안 각각 0.1 M, 0.4M 및 최종적으로 1.0 NaCl을 사용하는 일련의 단계 용출물을 컬럼에 가하였다. 분획들을 모으고 Centriprep 100을 사용하여 농축시켰다. 별법으로, 단백질은 0.1 M NaCl을 사용한 사전 용출이 없이 단독의 0.4M에 의해 용출될 수도 있다.

농축된 Q Sepharose 시료의 2 ml 분취량을 0.5 ml/분의 속도로 10 mM Na₂PO₄, pH 7.0으로 평형시킨 HR 16/50 Sepharose 12 (Pharmacia) 겔 투과 컬럼 상에 부하하였다. 이 컬럼을 동일한 완충액으로 240 분 동안 0.5 ml/분의 속도로 세척하고 2 분 시료를 회수하였다. 공 부피의 재료를 수집하고 Centriprep 100을 사용하여 농축시켰다. 농축된 Sepharose 12 시료의 2 ml 분취량을 0.5 ml/분의 속도로 10 mM Na₂PO₄, pH 7.0으로 평형시킨 HR 16/50 Sepharose 4B-CL (Pharmacia) 겔 투과 컬럼 상에 부하하였다. 이 컬럼을 동일한 완충액으로 240 분 동안 0.5 ml/분의 속도로 세척하고 2 분 시료를 회수하였다.

배제된 단백질 피크는 ~30 kDa 내지 1000 kDa의 범위에서 단백질을 분리하는 Sepharose CL-4B 수지를 사용한 겔 투과 컬럼에 적용함으로써 2회째의 분획을 행하였다. 이 분획은 두 개의 피크로 분해되었는데: 공 부피 (> 1000 kDa)에서 부수적인 피크 및 약 860 kDa의 겔보기 분자량으로 용출되는 주된 피크이었다. 일주일내 걸쳐서 후속 겔 투과된 시료들은 아마도 제한된 단백질 가수분해에 의한 주된 피크로부터 형성된 것으로 보이는 세번째 피크 (대략 325 kDa)의 점진적인 출현을 나타냈다. 이들 3 개의 피크에 대하여 수행한 생물학적 시험은 공 피크는 활성이 없는 반면, 860 kDa 독소 복합체는 고도로 활성이었고, 325 kDa 피크는 아주 강력한 하였지만 덜 활성이었음을 나타낸다. 상이한 발효 생산으로부터의 Sepharose CL-4B 독소 복합체 피크의 SDS PAGE 분석은 "P" 및 "S"로 표시된 두개의 뚜렷한 펩티드 패턴을 나타냈다. 두 패턴은 그들의 분자량 및 이들의 분획 중의 펩티드 성분의 농도에 있어서 현저한 차이를 나타냈다. 가장 빈번하게 생성된 "S" 패턴은 4 개의 고분자량 펩티드 (> 150 kDa)를 가진 반면, "P" 패턴은 3 개의 고분자량 펩티드를 가졌다. 또한, "S" 펩티드 분획은 유럽 옥수수 천공충에 대하여 2-3 배 더 높은 활성을 갖는 것으로 나타났다. 이 이동은 접종물의 연령 및(또는) 오래된 배양물의 성장 변수에 기초한 다른 인자

에 기인한 단백질 발현에 있어서 변이에 관련될 수 있다.

"P" 또는 "S" 펩티드 패턴으로 결정된 피크 독소 복합체 분석의 밀리그램 정량은 예비 SDS PAGE 를 수행하였고, TRIS-글리신 (PVDF 멤브레인(ProBlott(등록상표, Applied Biosystems) 에 대한 Seprabuff(등록상표)으로 3 - 4 시간 동안 트랜스블롯(transblot)시켰다. 블롯들을 각각 Harvard MicroChem 및 Cambridge ProChem에 아미노산 분석 및 N-말단 아미노산 서열화를 위해 보냈다. "S" 패턴에서 3 개의 펩티드는 이전의 실시예에서 동정된 서열에 비하여 유일한 N-말단 아미노산 서열 을 가졌다. 후술하는 서열 13에 나타난 201 kDa (TcdA_{ii}) 펩티드는 33% 아미노산 동일성을 공유하 였고 서열 1과는 50% 유사성을 나타냈다 (TcbA_{ii}) (표 11, 표 11에서 세로선은 아미노산 동일성 및 콜론은 보존적인 아미노산 치환을 나타낸다). 서열 14에 나타난 197 kDa의 두번째 펩티드(TcdB)는 42% 동일성을 가졌고 서열 2(TcaC)와는 58% 유사성을 나타냈다(수동 계산된 유사성과 동일성). 세 번째 205 kDa의 펩티드는 TcdA_{ii}로 표시하였다. 또한, 235 kDa 이상의 펩티드의 서열 16의 제한된 N-말단 아미노산 서열(TcbA)는 서열 1 (Tcb_{ii})에 대응하는 추정된 아미노산 서열을 포함하는 클로 닝된 유전자 (tcbA), 서열 11로부터 추정된 아미노산 서열, 서열 12와 일치하였다. 이는 더 큰 235+ kDa 펩티드가 발효 동안 201 kDa 펩티드 (TcbA_{ii})(서열 1)로 단백질 가수분해적으로 가공되어 가능하게는 분자의 활성화를 초래함을 나타낸다. 또다른 서열에 있어서 상기 실시예 5에서 보고된 당초 서열 5 (TcaB_{ii})로서 보고된 서열은 세번째 위치에서 글리신 (Gly) 보다는 아스파르탐산 잔기(Asp) 및 각각 18 및 19 위치에서 두 개의 추가의 아미노산 Gly 및 Asp를 함유하는 것으로 나 타났다. 다른 두개의 기타 서열, 서열 2 (TcaC) 및 서열 3 (TcaB_i)에 있어서, 추가의 아미노산 서 열이 얻어졌다. 밀도계 정량분석을 N-말단 분석을 위해 보내졌던 "S" 제제와 동일한 시료를 사용 하여 수행하였다. 이 분석은 201 Kda 및 197 kDa 펩티드가 "S" 패턴에서 전체 쿠마시 브릴리언트 블루 염색된 단백질의 각각 7.0% 및 7.2%를 나타냈고, 기타 풍부한 펩티드에 유사한 양으로 존재 함을 보여주었다. 이들 펩티드들은 다양한 CryI Bt 독소 등의 기타 세균성 독소에서 발견되는 상 황과 유사한 단백질 동족체를 나타낼 수 있는 것으로 고려된다. 이들 단백질은 독성 단백을 포함 하여 N-말단 아미노산 서열에서 40 - 90% 까지 유사성이 다양하다.

내부 아미노산 서열화: 독소 펩티드 유전자의 클로닝을 촉진하기 위하여, 선택된 펩티드의 내부 아미노산 서열을 다음과 같이 얻었다. "P" 또는 "S" 펩티드 패턴으로 결정된 피크 2A 분석의 밀리 그램 정량분석은 예비 SDS PAGE로 수행하였고, TRIS-글리신 (PVDF 멤브레인(ProBlott(등록상표, Applied Biosystems)에 대한 Seprabuff(등록상표)으로 3 - 4 시간 동안 트랜스블롯(transblot)시 켜다. 블롯들을 각각 Harvard MicroChem 및 Cambridge ProChem에 아미노산 분석 및 N-말단 아미노 산 서열화를 위해 보냈다. TcbA_{ii} (서열 1을 포함함), TcdA_{ii}, 및 TcaB_i (서열 3을 포함함)으로서 언 급된 3개의 펩티드들을 Harvard MicroChem에 의해 트립신 분해를 하고 이어서 HPLC 크로마토그래 피를 시켜 개별 펩티드들을 분리하였다. N-말단 아미노산 분석은 선택된 트립신형 펩티드 단편에 수행하였다. 두개의 내부 펩티드들은 TcdA_{ii}-PT111 (서열 17) 및 TcdA_{ii}-PT79(서열 18)로서 언급 된 펩티드 TcdA_{ii} (205 kDa 펩티드)에 대해서 서열화하였다. 두개의 내부 펩티드들은 TcaB_i-PT158 (서열 19) 및 TcaB_i-PT108(서열 20)으로서 언급된 펩티드 TcaB_i (68 kDa 펩티드)에 대해서 서열화하 였다. 4개의 내부 펩티드들은 TcbA_{ii}-PT103 (서열 21), TcbA_{ii}-PT56(서열 22), TcbA_{ii}-PT81(a) (서열 23) 및 TcbA_{ii}-PT81(b)(서열 24)로서 언급된 펩티드 TcbA_{ii} (201 kDa 펩티드)에 대해서 서열화하였 다.

[표 11]

M-말단 아미노산 서열 (동일성 및 유사성은 수동계산됨)

| |
|-----------------------------------|
| 201 kDa (서열 1과 33% 동일성 및 50% 유사성) |
| L I G Y N N O F S G * A 서열 13 |
| : : |
| F I Q G Y S D L F G N - A 서열 1 |
| 197 kDa (서열 2와 42% 동일성 및 58% 유사성) |
| M Q N S Q T F S V G E L 서열 14 |
| : : : |
| M Q D S P E V S I T T L 서열 2 |

<실시예 8>

포도랍두스 루미네센스 (균주 W-14) 게놈 DNA의 코스미드 라이브러리 제작 및 독소 단백질 제제로 이루어진 펩티드를 코딩하는 유전자의 단리에 대한 그의 스크리닝

이형의 숙주에서 포도랍두스 곤충 독소 단백질의 생산 및 기타 용도를 위한 전제 조건으로서, 그 러한 펩티드를 코딩하는 유전자를 단리하고 특성화하는 것이 필요하다. 이 목적인 동시에 추구된 다. 후술하는 한 접근법은 정제된 독소에 대한 단클론성 클론 또는 폴리클로날 항체의 사용하여 이것을 발현 라이브러리로부터 클론을 단리하는데 사용하는 것에 기초한다. 본 실시예에 기술된 다른 접근법은 N-말단 및 내부의 아미노산 서열 데이터를 사용하여 PCR 증폭에 사용하기 위한 동 의성 올리고뉴클레오티드를 설계하는 것에 기초를 두고 있다. 어느 방법이든 각각의 유전자의 단 리 및 이들의 DNA 염기 서열의 결정을 허용하도록 펩티드-코딩 유전자를 함유하는 DNA 클론을 동

정하는데 사용될 수 있다.

게놈 DNA 단리: 포토랍두스 루미네센스 균주 W-14 (ATCC 기탁 번호 55397)을 2% 프로테오스 펩톤 #3 아가 (Difco Laboratories, 미시간주 디트로이트 소재) 상에 성장시키고, 상기 실시예 1에 기술한 방법을 사용하여 계대 후 반복 생물학적 시험에 의해 유지시켰다. 50 ml 진탕 배양물이 2% 프로테오스 펩톤 #3 배지 중의 175 ml 배플 플라스크 중에서 생성되었고, 28 °C 및 150 rpm으로 대략 24 시간 동안 성장시켰다. 이 배양물 15 ml를 펠릿화하고 이를 DNA 단리를 위해 해동시킬 때까지는 그의 배지 중에 -20 °C에서 냉동시켰다. 해동된 배양물은 원심분리 (700 x g, 30 분)시키고 부유 오렌지 유코플리사카라이드 물질을 제거하였다. 잔류 세포 물질을 원심분리시켜(25,000 x g, 15분) 세균 세포를 펠릿화하고, 배지를 제거하여 버렸다.

게놈 DNA를 문헌 [Current Protocols in Molecular Biology, Ausubel et al. eds, John Wiley & Sons, 1994의 섹션 2.4.1]에 기술된 CTAB 법(염 소크를 포함시키고 모든 부피를 10배 증가하도록 변형시킴)을 채용함으로써 단리하였다. 펠릿화된 세균 세포들을 TE 완충액(10 mM Tris-HCl, 1 mM EDTA, pH 8.0) 중에 10 ml로 최종 부피로 재현탁시키고, 이어서 5 M NaCl 12 ml를 첨가하였다; 이 혼합물은 15,000 x g에서 20분 원심분리시켰다. 펠릿을 TE 5.7 ml 중에서 재현탁시키고, 10% SDS 300 ml 및 20 mg/ml 프로테이나아제 K(Gibco BRL Products, Grand Island, NY; 멸균 증류수 중) 60 ml를 현탁액에 첨가하였다. 이 혼합물을 37 °C에서 1 시간 동안 배양시키고; 이어서 대략 10 mg 라이소자임 (Worlington Biochemical Corp., Freehold, NJ)를 첨가하였다. 추가의 45분 후, 5 M NaCl 1 ml 및 CTAB/NaCl 용액(10% w/v CTAB, 0.7 M NaCl) 800 ml를 첨가하였다. 이 제제는 65 °C에서 10 분 동안 배양시키고 이어서 부드럽게 교란시킨 다음 추가로 배양시키고 대략 20 분 동안 교란시켜 세포 물질의 정화를 거들었다. 동일 부피의 클로로포름/이소아밀 알코올 용액 (24:1 v/v)을 첨가하고, 부드럽게 혼합시키고 원심분리시켰다. 동일 부피의 PCI (페놀/클로로포름/이소아밀 알코올 용액 (50:49:1 v/v, 1 M Tris-HCl, pH 8.0으로 평형시킴; Intermountain Scientific Corporation, Kaysville, UT)로 2회 추출 후, DNA를 이소프로판올 0.6 부피로 침전시켰다. DNA 침전물을 유리 막대로 가볍게 제거하여, 70% 에탄올로 2회 세척하고, 건조시키고, STE 완충액(10 mM Tris-HCl pH 8.0, 10 mM NaCl, 1 mM EDTA) 2 ml 중에 용해시켰다. 이 제제물은 260 nm의 광학 밀도 (즉, O.D₂₆₀)에 의해 측정된 바 2.5 mg/ml DNA를 함유하였다.

단리된 게놈 DNA의 분자량 크기 범위는 라이브러리 제작에 대한 적합성에 대하여 평가하였다. CHEF 겔 분석은 0.5 x TBE 완충액(44.5 mM Tris-HCl pH 8.0, 44.5 mM H₃BO₃, 1 mM EDTA)으로 만든 1.5% 아가로스 (Seakem (등록상표) LE, FMC BioProducts, Rockland, ME) 겔 중에서 Pulsewave 760 Switcher가 구비된 BioRad CHEF-DR II 장치 (Bio-Rad Laboratories, Inc., Richmond, CA) 상에서 수행하였다. 진행 변수들은 초기 A 시간, 3초; 최종 A 시간, 12초; 200 볼트; 진행 온도 4 - 18 °C; 진행 시간, 16.5 시간이었다. 에치움 브로마이드 염색 및 자외선 광하에 겔의 검사는 DNA 가 길이에서 30 - 250 kbp에 걸쳐 나타냈다.

라이브러리의 제작: 부분적인 Sau3A 1 분해물을 이 포토랍두스 게놈 DNA 제제로 제조하였다. 이 방법은 상기 Ausbel의 섹션 3.1.3에 기초하였다. 거의 최적의 결과가 얻어질 때까지 다양한 조건 하에서 소규모 반응을 진행하여 개조하였다. 이 개조 조건을 사용하여 몇몇의 확대 규모 반응을 진행시켰으며 결과들을 CHEF 겔 상에서 분석하고, 단지 최상의 확대 규모의 제제를 계속하여 수행하였다. 최적의 경우에 있어서, 포토랍두스 게놈 DNA 200 µg을 2 ml 전체 부피의 1 X NEB 4 완충액 (제조원에 의해 10 X로 공급됨) 중에서 Sau3A 1 (New England Biolabs, "NEB", Beverly, MA) 1.5 단위를 사용하여 37 °C에서 15 분 동안 배양시켰다. 반응을 PCI 2 ml를 첨가하여 정지시키고 8000 x g에서 10 분 동안 원심분리시켰다. 상층액에 5 M NaCl 200 µl 및 빙냉 에탄올 6 ml를 첨가하였다. 이 제제를 -20 °C에서 30 분 동안 급냉시키고, 이어서 12,000 x g에서 15 분 동안 원심분리시켰다. 상층액을 제거하고 침전물을 진공 중의 40 °C에서 건조시킨 다음 STE 400 µl 중에 재현탁시켰다. 분광광도계 분석은 투입 DNA의 약 40% 회수율을 나타냈다. 분해된 DNA를 CPMB(인용 문헌 참조)의 섹션 5.3.2에 따라 수크로스 구배 상에서 크기 분획화하였다. 10% 내지 40%(w/v) 선형 수크로스 구배를 Ultra-Clear(등록상표) 튜브 (Beckman Instruments, Inc., Palo Alto, CA) 중의 구배 마커를 사용하여 제조하고 DNA 시료를 상부에 위치시켰다. 원심분리(26,000 rpm, 17 시간, Beckman SW41 로타, 20 °C) 후, 분획들 (약 750 µl)을 구배의 상부로부터 꺼내어 상기한 바의 CHEF 겔 전기영동에 의해 분석하였다. 크기 범위 20 - 40 kbp의 Sau3A 1를 포함하는 분획들을 선별하고 DNA를 상기 Ausbel의 섹션 5.3.3의 방법을 수정(모든 용액의 양을 대략 6.3배로 증가시킴) 함으로써 침전시켰다. 철야 침전 후, DNA를 원심분리 (17,000 x g, 15분)에 의해 회수하고, 건조시키고, TE 중에 재용해시키고, 80 µl의 최종 부피로 모은 다음, 3 M 소듐 아세테이트 8 µl 및 에탄올 220 µl을 첨가하여 재침전시켰다. 상기한 바와 같이 원심분리에 의해 회수한 펠릿을 TE 12 µl 중에 재현탁시켰다. DNA의 농축물은 Hoefer TK0100 플루오로미터 (Hoefer Scientific Instruments, 캘리포니아주 샌프란시스코 소재)에서 Hoechst 33258 염료 (Polyscience, Inc., Warrington, PA) 형광분석기에 의해 결정하였다. 대략 2.5 µg의 크기 분획된 DNA가 회수되었다.

코스미드 pWE15 DNA (Stratagene, 캘리포니아주 라졸라 소재) 30 µg을 제조원의 완충액(최종 부피 200 µl, 37 °C, 1 시간) 중의 제한 효소 BamH 1(NEB) 100 단위를 사용하여 분해를 완료하였다. 반응물을 PCI 100 µl로 추출하고 DNA를 3M 소듐 아세테이트 및 -20°C 무수 에탄올 550 µl을 첨가하여 수층으로부터 침전시켰다. -70 °C에서 20 분 후, DNA를 원심분리 (17,000 x g, 15분)에 의해 회수하고 진공하에 건조시키고, 10 mM Tris-HCl, pH 8.0 180µl 중에 용해시켰다. 여기에 10 X CIP 완충액(10 mM Tris-HCl, pH 8.3; 10 mM ZnCl₂; 10 mM MgCl₂) 20µl 및 1:4 희석 송아지 장 알칼린 포스파타아제(Boehringer Mannheim Corporation, Indianapolis, IN) 1 µl (0.25 단위)을 첨가하였다. 37 °C에서 30분 후, 다음을 첨가하였다: 0.5M EDTA 2 µl pH 8.0, 10% SDS 10 µl; 20 mg/ml 프로테이나아제 K(상기함) 0.5 µl. 이어서, 55 °C에서 30 분 동안 배양시켰다. PCI 100 µl 및 페놀(Intermountain Scientific Corporation, 1M Tris-HCl, pH 8.0으로 평형시킴) 100µl을 사용하여 추출한 후, 탈인산화된 DNA를 7.5 M 암모늄 아세테이트 72 µl 및 -20°C 에탄올 550 µl을 첨가하여 침전시키고, 30 분 동안 얼음 상에서 배양시키고 상기한 바와 같이 원심분리시켰다. 펠릿화

된 DNA를 -20 °C 70% 에탄올 500 μl로 1회 세척하고, 진공 중에서 건조시키고 TE 완충액 20 μl 중에 용해시켰다.

크기 분획화된 Sau3A 1 단편의 BamHI-분해 및 인산화된 pWE15 벡터에 대한 라이게이션은 Ausbel 섹션 3.33의 프로토콜을 변형시킴으로써(즉, 제조원에 의해 공급된 미리 혼합한 10 X 라이게이션 완충액을 사용함) T4 리가아제를 사용하여 수행하였다. 라이게이션은 20 μl의 전체 부피 중에 15 °C에서 밤새 수행한 후 -20°C에 저장하였다.

DNA 약 1 μl을 함유하는 코스미드 DNA 라이게이션 반응물 4 μl을 상업상의 패키징 추출물(Gigapack(등록상표) III Gold Packaging Extract, Stratagene)을 사용하여 제조원의 지시에 따라서 박테리오파아지 램다 내에 패키징시켰다. 패키징된 제조물을 사용할 때까지 4°C에 저장하였다. 패키징된 코스미드 제조물은 다음과 같이 (Gigapack(등록상표) III Gold 프로토콜("Titering the cosmid Library")에 따라서 대장균 XL1 Blue MR 세포(Stratagene)를 감염시키는 데 사용하였다. XL1 Blue MR 세포를 0.2% (w/v) 말토오스 및 10 mM MgSO₄를 함유하는 LB 배지 (g/L: 박토-트립톤, 10; 박토-이스트 엑스, 5; 박토-아카, 15, NaCl, 5(Difco Laboratories, Detroit, MI) 중의 37 °C에서 성장시켰다. 5 시간 성장 후, 세포들을 700 x g (15분)으로 펠릿화하고 10 mM MgSO₄ 6 ml 중에 재현탁시켰다. 배양 밀도는 10 mM MgSO₄를 사용하여 OD₆₀₀ = 0.5 까지 조정하였다. 패키징된 코스미드 라이브러리를 멸균 SM 배지 (0.1 M NaCl, 10 mM MgSO₄, 50 mM Tris-HCl pH 7.5, 0.01% w/v 젤라틴)를 사용하여 1:10 또는 1:20으로 희석시키고 희석된 제제 25 μl를 희석된 XL1 Blue MR 세포 25 μl과 혼합하였다. 이 혼합물을 25 °C에서 30 분 동안 진탕없이 배양시키고 이어서 LB 브로스 200 μl을 첨가하고 배양을 간혹 가볍게 진탕시키면서 대략 1 시간 동안 계속하였다. 이 배양물의 분취량 (20 - 40 μl)을 100 mg/l 앰피실린 (즉, LB-Amp₁₀₀)을 함유하는 LB 아가 플레이트 상에 뿌리고, 37 °C에서 밤새 배양시켰다. 증폭이 없는 라이브러리를 저장하기 위하여, 단일 콜로니를 뽑아서 멸균 96-웰 마이크로웰 플레이트의 개별 웰 내로 접종하였으며, 각각의 웰은 75 μl의 Terrific Broth (TB 배지: 12 g/l 박토-트립톤, 24 g/l 박토-이스트 엑스, 0.4% v/v 글리세롤, 17 mM KH₂PO₄, 72 mM K₂HPO₄) 및 100 mg/l 앰피실린 (즉, TB-Amp₁₀₀)을 함유하였으며, 진탕없이 37 °C에서 밤새 배양시켰다. 96-웰 플레이트를 복제 플레이트 내로 복제 한 후, 여과 멸균된 TB:글리세롤 (1:1, v/v: 100 mg/l 앰피실린을 포함하거나 또는 포함하지 않음)의 75 μl/웰을 플레이트에 첨가한 후, 이를 100 rpm, 37 °C에서 재빨리 진탕시킨 후, 이어서 Parafilm(등록상표)(American National Can, Greenwich, CT)로 밀봉하고 -70°C 냉동기 내에 위치시켜 저장하였다. 복제 플레이트를 마스터 플레이트에 동일하게 성장시키고 처리하였다. 총 40 개의 그러한 마스터 플레이트(및 이들의 복제물)이 제조되었다.

방사성표지된 DNA 프로브를 사용한 라이브러리의 스크리닝: 방사성 동위원소로 표지된 프로브를 사용하여 프로브시키기 위한 콜로니 필터를 제조하기 위하여, 라이브러리의 96-웰 플레이트를 25 °C에서 해동시켰다(실온에서 벤취 톱). 96 프롱(prong)을 가진 복제 플레이팅 도구를 사용하여 TB-Amp₁₀₀의 75 μl/웰을 포함하는 새로운 96-웰 복제 플레이트를 접종시켰다. 복제 플레이트를 37 °C에서 밤새(고정시킴) 성장시키고, 이어서 37 °C에서 약 30 분 동안 100 rpm으로 진탕시켰다. 일부 단리물의 비성장예 기인하여 총 800 개의 콜로니가 이들 복제 플레이트에서 나타났다. 복제 도구를 사용하여 생물 분석 플라스틱 접시(Nunc, 243 x 243 x 18 mm; Curtin Mathison Scientific, Inc., Wood Dale, IL) 중의 고체 LB-Amp₁₀₀ (100 ml/접시) 상에 위치한 Magna NT (MSI, Westboro, MA) 나일론 멤브레인 (0.45 미크론, 220 x 250 mm) 상에 96-웰 어레이의 복제 함침부를 접종시켰다. 이 콜로니들을 37 °C에서 약 3 시간 동안 멤브레인 상에서 성장시켰다.

양성 대조군 콜로니(GZ4 서열 삽입물을 함유하는 세균 콜로니, 이하 참조)들은 35 mg/l 클로람페니콜을 보충한 LB 배지(즉, LB-Cam₃₅) 상의 개별적인 Magna NT 멤브레인 (Nunc, 0.45 미크론, 82 mm 원) 상에서 성장시키고, 라이브러리 콜로니 멤브레인 측면에 나란히 처리하였다. 멤브레인 상의 세균 콜로니들을 Genius(등록상표) 시스템 사용자 가이드 버전 2.0 (Boehringer Mannheim Corporation, Indianapolis, IN)으로부터 취한 프로토콜에 따라 용해시키고 DNA를 변성시키고 중화시켰다. 멤브레인들을 0.5 N NaOH 및 1.5 M NaCl로 적신 여과지 상의 콜로니 측면에 15 분 동안 위치시켜 변성시키고, 1M Tris-HCl pH 8.0, 1.5 M NaCl로 적신 여과지 상에 15 분 동안 중화시켰다. 자동 가교(crosslink)로 설정해 놓은 Stratagene UV Stratalinker를 사용하여 UV 가교를 시킨 후, 멤브레인들을 25 °C에 사용할 때까지 저장하였다. 멤브레인들을 단일 96-웰 플레이트의 복제 함침부를 함유하는 스트립으로 다듬은 후, CPMB(문헌 참조)의 섹션 6.4.1의 방법에 의해 대규모로 세척하고 (3 X SSC, 0.1% (w/v) SDS 중의 25 °C에서 3 시간에 이어서, 동일 용액 중에서 65 °C에서 1 시간), 이어서 하이브리드화 단계용 제제(20 X SSC = 3 M NaCl, 0.3 M 소듐 시트레이트, pH 7.0) 중의 2 X SSC 중에서 행구었다.

TcaC 유전자의 특정 게놈 단편의 증폭: 정제된 TcaC 펩티드 단편에 대해 결정된 N-말단 아미노산 서열(본 명세서에 서열 2로서 기재함)에 기초하여, 동의성 올리고뉴클레오티드의 풀(pool S4Psh)를 Applied Biosystem ABI394 DNA/RNA 합성기(Perkin Elmer, Foster City, CA) 상의 가닥 β-시아노에틸 화학에 의해 합성하였다. 이 올리고뉴클레오티드들을 55 °C에서 8 시간 동안 탈보호시키고, 물 중에 용해시키고, 분광광도계 측정에 의해 정량화하고, 희석하여 사용하였다. 이 풀은 TcaC 펩티드의 결정된 N-말단 아미노산 서열에 해당한다. 결정된 아미노산 서열 및 대응하는 동의성 DNA 서열은 이하에 나타냈고, 여기서 A, C, G 및 T는 표준 DNA 염기이고, I는 이노신을 나타낸다.

| | | | | | | | |
|----------|-----|---------|---------|-----------------|-----|---------|-------|
| 아미노산 | Met | Gln | Asp | Ser | Pro | Glu | Val |
| S4Psh 5' | ATG | CA(A/G) | GA(T/C) | (T/A)(C/G)(T/A) | CCI | GA(A/G) | GT 3' |

또다른 세트의 동의성 올리고뉴클레오티드를 합성하였으며(pool P2.3.5R), 서열 17의 결정된 아미

노산 서열에 대한 코딩 가닥의 상보체를 나타낸다:

| | | | | | | | |
|---------|----------------|---------|---------|-----------|---------|---------|-------|
| 아미노산 | Ala | Phe | Asn | Ile | Asp | Asp | Val |
| 코돈 | 5' GCN | TT(T/C) | AA(T/C) | AT(A/T/C) | GA(T/C) | GA(T/C) | GT 3' |
| P2.3.5R | 3' CG(A/C/G/T) | AA(A/G) | TT(A/G) | TA(T/A/G) | CT(A/G) | CT(A/G) | CA 5' |

이들 뉴클레오티드들은 폴리머라제 사슬 반응(PCT(등록상표, Roche Molecular System, Branchburg, NJ)에 사용하여 포토랍두스 균주 W-14로부터 제조된 게놈 DNA로부터의 특정 DNA 단편(상기 함)을 증폭시켰다. 대표적인 반응물 (50 μ l)은 각 프라이머 풀 P2Psh 및 P2.3.5R 125 pmol, 게놈 주형 DNA 253 ng, dATP, dCTP, dGTP 및 dTTP 각각 10 pmol, 1X GeneAmp(등록상표) PCR 완충액 및 AmpTaq(등록상표) 2.5 단위(이 모두는 Roche Molecular Systems로부터 시판됨; 10X GeneAmp(등록상표) 완충액은 100 mM Tris-HCl pH 8.3, 500 mM KCl, 0.01% w/v 젤라틴을 함유하였다). 증폭은 Perkin Elmer Cetus DNA Thermal Cycler(Perkin Elmer, Foster City, CA) 중에서 94 °C (1.0분), 55 °C (2.0분), 72 °C (3.0분)의 35 주기에 이어서 72 °C에서 7.0분의 확장 기간을 사용하여 수행하였다. 증폭 절차는 TEA 완충액 (40 mM Tris-아세테이트, 2 mM EDTA, pH 8.0) 중에서 2% w/v NuSieve(등록상표) 3:1 아가로스 (FMC BioProducts)를 통한 전기영동에 의해 분석하였다. 추정 크기 250 bp의 특정 생성물이 에치움 브로마이드 (0.5 μ g/ml) 염색 및 자외선 광 하의 검사에 의해 기타 증폭 생성물 중에서 관찰되었다.

대략 250 bp 생성물을 함유하는 겔의 영역을 자르고, 작은 플러그 (0.5 mm 직경)를 제거하여 PCR 증폭 (40 사이클)에 대한 주형을 공급하는데 사용하였다. 반응물 (50 μ l)은 게놈 주형 DNA를 빼고 상기와 동일한 성분을 함유하였다. 증폭 다음, 단편들의 말단을 평활말단이 되게 하고 25 °C에서 20 분 동안 T4 DNA 리가아제 (NEB), 1 nmol ATP 및 T4 키나아제 (Pharmacia Biotec Inc., Piscataway, NJ) 2.15 단위를 사용하여 배양시킴으로써 인산화시켰다.

DNA 단편들을 TEA 중의 1% GTG(등록상표) 아가로스 (FMC)를 통해 전기영동시킴으로써 잔여 프라이머로부터 분리하였다. 겔보기 크기 250 bp의 단편을 함유하는 겔 슬라이스를 자르고, DNA를 Qiaex 키트(Qiaex Inc., Chatsworth, CA)를 사용하여 추출하였다.

추출된 DNA 단편들을 제한 효소 Sma I으로 분해를 완료한 후 상기 pWE15 DNA에 대해 기술한 바와 유사한 방식으로 추출한 플라스미드 벡터 pBC KS(+) (Stratagene)에 라이게이션시켰다. 대표적인 라이게이션 반응물 (16.3 μ l)은 1X 라이게이션 완충액 (50 mM Tris-HCl, pH 7.4; 10 mM MgCl₂; 10 mM 디티오트레이톨; 1mM 스퍼미딘, 1mM ATP; 100 mg/ml 소 혈청 알부민) 중에 분해된 pBC KS(+) DNA 100 ng, 250 bp 단편 DNA 70 ng, 1 nmol [Co(NH₃)₆]Cl, 및 T4 DNA 리가아제 (Collaborative Biomedical Products, Bedford, MA) 3.9 Weiss 단위를 포함하였다. 14 °C에서 밤새 배양시킨 후, 라이게이션된 생성물을 공급자의 지침에 따라서 열린, 컴피턴트 대장균 DH5 α 세포 (Gibco BRL) 내로 형질전환시키고, IPTG 119 μ g/ml 및 X-gal 50 μ g/ml을 함유하는 LB-Cam₃₅ 플레이트 상에 위치시켰다. 독립적인 백색 콜로니들을 뽑아서 플라스미드 DNA를 변형된 알칼린-용해/PEG 침전법 [PRISM(등록상표) Ready Reaction DyeDeoxy(등록상표) Terminator Cycle Sequencing Kit Protocols; ABI/Perkin Elmer]에 의하여 제조하였다. 삽입 DNA의 두 가닥의 뉴클레오티드 서열을 T7 프라이머[pBC KS(+)] 염기 601-623: TAAACGACGGCCAGTGAGCGCG] 및 LacZ 프라이머 [pBC KS(+)] 염기 792-816: ATGACCATGATTACGCCAAGCGC GC) 및 PRISM(등록상표) 서열화 키트 [ABI/Perkin Elmer]와 함께 제공된 프로토콜을 사용하여 결정하였다. 비혼입된 염료 종결인자 디데옥시리보뉴클레오티드들은 제조원의 지시에 따라 Centri-Sep 100 컬럼 (Princeton Separation, Inc., Adelphia, NJ)을 통해 통과시킴으로써 제거하였다. DNA 서열은 ABI 모델 373 A DNA 서열화기 (ABI/Perkin Elmer) 상에 시료를 분석함으로써 얻었다. 두 단리물, GZ4 및 HB14의 DNA 서열은 도 1에 설명된 바와 같이 발견되었다.

이 서열은 다음의 특성들을 나타내었다: i) 염기 1-20은 S4Psh 동의성 올리고뉴클레오티드의 64개의 가능한 서열 중 하나를 나타내었고, ii) 아미노산 서열 1-3 및 6-12의 서열은 TcaC (서열 2에 기술됨)의 N-말단의 것에 정확히 대응하였고, iii) 코딩된 네번째 아미노산은 세린 보다는 시스테인이었다. 이 차이는 세린 코돈에 대한 동의성 내에서 코딩된다(상기 참조), iv) 다섯번째 아미노산은 프롤린으로, 서열 2에 주어진 TcaC N-말단 서열에 대응하고, v) 염기 257-276은 동의성 풀 내로 디자인된 192 개의 가능한 서열 중 하나를 코딩하고, vi) 염기 268-270에서 도입된 TGA 종결 코돈은 대응하는 위치에서 뉴클레오티드 풀 내로 조립된 동의성에 상보성의 결과이며, 대응 유전자에 대한 짧은 해독 프레임을 제공하지 않는다.

<TcaC 펩티드 유전자-특이성 프로브의 표지화>

상기 276 염기에 대응하는 DNA 단편들을 P2Psh 및 P2.3.5R 프라이머 각각 100 pmol, 주형으로서 플라스미드 GZ4 또는 HB104 10 ng, dATP, dCTP, dGTP 및 dTTP 각각 20 nmol, AmpliTaq (등록상표) DNA 폴리머라제 5 단위, 및 1X 농도의 GeneAmp (등록상표) 완충액을 상기한 바와 동일한 온도 관리 하에서 PCR (등록상표)에 의해 100 μ l 반응 부피로 증폭시켰다 (35 주기). 증폭 생성물은 Qiaex 키트에 의해 1 % GTG (등록상표) 아가로스 겔로부터 추출하고 형광계에 의해 정량화하였다.

플라스미드 HB14 주형 (대략 400 ng)으로부터 추출된 증폭 생성물을 5 개의 분취량으로 나누고 High Prime Labeling Mix (Boehringer Mannheim 제품)을 사용하여 제조원의 지시에 따라서 ³²P-dCTP로 표지화하였다. 비혼입된 방사성 동위원소들은 제조원의 지시에 따라 NucTrap (등록상표) 프로브 정제 컬럼 (Stratagene)을 통해 통과시킴으로써 제거하였다. 표지화된 DNA 생성물의 특이 활성화는 액체 섬광계에 의해 3.11 $\times 10^8$ dpm/ μ g인 것으로 측정하였다. 이 표지화된 DNA는 게놈

라이브러리의 800 원으로부터 제조된 멤브레인을 프로브시키는데 사용하였다.

<TcaC-펩티드 유전자 특이성 프로브를 사용한 스크리닝>

방사성 표지화된 HB14 프로브를 대략 10 분 동안 비등시키고, 이어서 "최소 혼성 (minimal hyb)" 용액에 가하였다. [주: "최소 혼성" 방법은 CERES 프로토콜; "Restriction Fragment Length Polymorphism Laboratory Manual 버전 4.0", 섹션 4-40 및 4-47; CERES/NPI, 유타주 솔트레이크 시티 소재로부터 입수한 것임, NPI는 현존하지 않고, 그의 승계인이 Linkage Genetics로서 경영하고 있음]. "최소 혼성" 용액은 10 % w/v PEG (폴리에틸렌 글리콜), 분자량 대략 8000), 7 % w/v SDS; 0.6X SSC, 10 mM 인산 나트륨 완충액 (95 g/l NaH₂PO₄ · 1H₂O 및 84.5 g/l Na₂HPO₄ · 7H₂O 를 함유하는 1 M 원액으로부터 제조), 5 mM EDTA 및 100 mg/ml 변성 연어 정자 DNA를 함유한다. 멤브레인들을 신속하게 지압 건조시키고, 이어서 예비 하이브리드 없이 멤브레인의 5 스트립을 "최소 혼성" 75 ml 및 방사성 표지화된 HB14 프로브 2.6 ng/ml을 함유하는 2 개의 플라스틱 상자 각각에 위치시켰다. 이들을 60 °C에서 서서히 진탕시키면서 (50 rpm) 밤새 배양시켰다. 필터를 각각 25 °C에서 "최소 혼성 세척 용액" (0.25X SSC, 0.2% SDS) 중에 대략 10 분 동안 3회 세척하고, 이어서 동일 용액 중의 60 °C에서 서서히 진탕시키면서 30 분 동안 2회 세척하였다. 필터들을 차광 방사선 사진 카세트 중의 Saran Wrap (등록 상표) (Dow Brands 제조, 인디애나주 인디애나폴리스 소재)으로 덮힌 종이 위에 위치시키고, 두 개의 DuPont Cronex Lightening-Plus C1 강화기 (Sigma Chemical Co. 제조, 미주리주 세인트 루이스 소재)를 구비한 X-Omat X-선 필름 (Kodak 제조, 뉴욕주 로케스터 소재)에 -70 °C에서 4 시간 동안 노출시켰다. 현상시 (표준 사진 공정), 중요한 신호들은 더 약하고, 더욱 불규칙적인 신호의 높은 피이크 중에서 두 복제물에서 두드러졌다. 필터들을 다시 "최소 혼성 세척 용액" 중의 68 °C에서 약 4 시간 동안 세척하고나서, 다시 카세트에 위치시키고, 필름을 -70 °C에서 밤새 노출시켰다. 12 개의 가능한 양성들이 복제물 96-웰 콜로니 함유부 모두에서의 강한 신호로 인하여 확인되었다. 음성 대조군 멤브레인 (pWE15를 함유하는 XL1 Blue MR 세포의 콜로니)에서는 신호가 나타나지 않았고, 동시에 실험적인 시료를 사용하여 처리한 양성 대조군 멤브레인 (PCR 생성물의 GZ4 단리물을 함유하는 DH5α 세포)에서는 매우 강한 신호가 나타났다.

12 개로 추정되는 하이브리다이제이션-양성 콜로니들을 동결된 96-웰 라이브러리 플레이트로부터 회수하여 LB-Amp₁₀₀ 배지 중에서 37 °C에서 밤새 성장시켰다. 이어서, 이들을 고상 LB-Amp₁₀₀ 상에 패칭하였다 (3/플레이트, + 3 음성 대조군: pWE15 벡터를 함유하는 XL Blue MR 세포). 두 세트의 멤브레인 (Magna NT 나일론, 0.45 마이크론)을 하이브리다이제이션용으로 제조하였다. 처음 세트는 필터를 패치 플레이트 위의 콜로니 상에 직접 위치시키고나서, 이것을 부착 박테리아 세포와 함께 제거하고, 아래와 같이 처리함으로써 제조하였다. 두번째 세트의 필터들은 LB-Amp₁₀₀ 배지를 포함하는 플레이트 상에 위치시키고, 이어서 패치 플레이트로부터 필터 위로 세포들을 이동시킴으로써 접종하였다. 37 °C에서 밤새 성장시킨 후에, 필터들을 플레이트로부터 제거하여 처리하였다.

필터 상의 박테리아 세포들을 용해시키고, DNA를 각 필터의 콜로니 측부를 플라스틱 플레이트 중의 0.5 N NaOH의 풀 (1.0 ml) 상에 3 분 동안 위치시킴으로써 변성시켰다. 필터들을 종이 타월 상에서 지압 건조시키고나서, 이 공정을 새로운 0.5 N NaOH로 반복하였다. 지압 건조시킨 후, 필터들을 각각 1 M Tris-HCl pH 7.5의 1.0 ml 풀 상에 3 분 동안 위치시킴으로써 중화시키고, 지압 건조시키고, 새로운 완충액으로 다시 중화시켰다. 계속하여, 0.5 M Tris-HCl pH 7.5 및 1.5 M NaCl의 풀 상에서 두 유사한 침지를 (각각 5분) 수행하였다. 지압 건조 후, DNA를 상기한 바와 같이 필터에 자외선 가교화시키고, 필터들을 3X SSC 및 0.1 % (w/v) SDS 약 100 ml 중에서 세척하였다 (25 °C, 100 rpm) (4회, 매회 세척시 새로운 용액으로 각각 30분). 이어서, 이들을 예비 하이브리다이제이션 용액 [6X SSC 및 Ficoll 400 (Pharmacia 제품), 폴리비닐피롤리돈 (평균 분자량 360,000; Sigma 제품) 및 소혈청 알부민 단편 V; (Sigma 제품) 각각 1 % w/v]의 최소 부피 중에 65 °C, 50 rpm에서 2 시간 동안 위치시켰다. 예비 하이브리다이제이션 용액을 제거하고, 라이브러리 멤브레인의 이전 하이브리다이제이션로부터 저장되었고, 95 °C에서 5 분 동안 변성시킨 HB14³²P-표지화된 프로브로 대체하였다. 하이브리다이제이션은 60 °C에서 16 시간 동안 50 rpm으로 진탕시키면서 수행하였다.

표지화된 프로브 용액의 제거에 이어서, 멤브레인들을 3X SSC 중에서 25 °C (50 rpm, 15 분)에서 3회 세척하였다 (매회 세척시 약 150 ml). 이어서, 이들을 0.25X SSC 및 0.2% SDS (최소 혼성 세척 용액) 중에서 68 °C에서 3 시간 동안 (50 rpm) 세척하고, 상기한 바와 같이 25 °C에서 1.5 시간 동안 X-선 필름에 노출시켰다 (강화기 스크린이 없음). 이 노출은 코스미드 단리물 22G12, 25A10, 26A5 및 26B10에 대하여 매우 강한 하이브리다이제이션 신호를 나타내었고, 코스미드 단리물 8B10에 대하여 매우 약한 신호를 나타내었다. 음성 대조군 (pWE15) 콜로니에서는 신호가 나타나지 않았고, 동시에 실험적인 시료를 사용하여 처리한 양성 대조군 멤브레인 (PCR 생성물의 GZ4 단리물을 함유하는 DH5α 세포)에서는 매우 강한 신호가 나타났다.

<TcaB 유전자의 특정 게놈 단편의 증폭>

정제된 TcaB_i 펩티드 단편에 대해 결정된 N-말단 아미노산 서열 (본 명세서에 서열 3으로서 기재함)에 기초하여, 동의성 올리고뉴클레오티드의 풀 (pool P8F)을 펩티드 TcaC에 대하여 기재된 바와 같이 합성하였다. 결정된 아미노산 서열 및 대응하는 동의성 DNA 서열은 이하에 나타나 있고, 여기서 A, C, G 및 T는 표준 DNA 염기이고, I는 이노신을 나타낸다.

아미노산 Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg
 P8F 5' TTT ACI CA(A/G) ACI (C/T)TI AAA GAA GCI (A/C)G 3'
 (C/T)TI

또다른 세트의 동의성 올리고뉴클레오티드를 합성하였으며 (pool P8.108.3R), 이것은 TcaB_i-PT108

내부 펩티드 (서열 20)의 결정된 아미노산 서열에 대한 코딩 가닥의 상보체를 나타낸다:

| | | | | | | | | |
|-----------|-----|---------|---------|-----------|---------|-------------|---------|---------|
| 아미노산 | Met | Tyr | Tyr | Ile | Gln | Ala | Gln | Gln |
| 코돈 | ATG | TA(T/C) | TA(T/C) | AT(T/C/A) | CA(A/G) | GC(A/C/G/T) | CA(A/G) | CA(A/G) |
| P8.108.3R | 3' | AT(A/G) | AT(A/G) | TA(A/G/T) | GT(T/C) | CGI | GT(T/C) | GT 5' |
| | TAC | | | | | | | |

이들 올리고뉴클레오티드들은 HotStart 50 Tubes (등록 상표) (Molecular Bio-Products, Inc. 제품, 캘리포니아주 샌디에고 소재)를 이용하는 PCR (등록 상표)용 프라이머로서 사용하여 포도랍두스 균주 W-14 (상기 기재됨)로부터 제조된 게놈 DNA로부터 특정 DNA 단편을 증폭시켰다. 대표적인 반응물 (50 μ l)은 1X GeneAmp (등록 상표) PCR 완충액 중에 각 프라이머 풀 P8F 및 P8.108.3R 25 pmol, dATP, dCTP, dGTP 및 dTTP 각각 2 nmol을 함유하였고 (바닥층), 1X GeneAmp (등록 상표) PCR 완충액 중에 게놈 주형 DNA 230 ng, dATP, dCTP, dGTP 및 dTTP 각각 8 nmol, 및 AmpTaq (등록 상표) DNA 폴리머라제 2.5 단위를 함유하였다 (상부층). 증폭은 TCaC 펩티드에 대해 기재된 바와 같이 35 주기까지 수행하였다. 증폭 생성물은 TEA 완충액 중의 0.7 % w/v SeaKem (등록 상표) LE 아가로스 (FMC)를 통한 전기영동에 의해 분석하였다. 추정된 크기 1600 bp의 특정 생성물이 관찰되었다.

4 개의 이러한 반응물을 풀링하고, 증폭된 DNA를 상기 TcaC 펩티드에 대해 기재된 바와 같이 Qiaex 키트에 의해 1.0 % SeaKem (등록 상표) LE 아가로스 (FMC)로부터 추출하였다. 추출된 DNA를 P8F 및 P8.108.3R 프라이머 풀을 사용하여 서열 결정용 주형 [PRISM (등록 상표) 서열화 키트]으로 직접 사용하였다. 각각의 반응물은 약 100 ng 주형 DNA 및 하나의 프라이머 풀 25 pmol을 함유하였고, TcaC 펩티드에 대해 기재된 바와 같은 표준 프로토콜에 따라 처리하였다. P8F 프라이머의 연장으로부터 유도된 서열 분석은 서열 3 (TcaB_i)의 N-말단 펩티드 서열의 일부에 대응하는 짧은 DNA 서열 (및 코딩된 아미노산 서열)을 나타내었다:

GAT GCA TTG NTT GCT
Asp Ala Leu (Val) Ala

<TcaB_i 펩티드 유전자-특이성 프로브의 표지화>

겔-정제된 TcaB_i DNA 단편 대략 50 ng을 상기한 바와 같이 ³²P-dCTP로 표지화하고, 비혼입된 방사성 동위원소들은 NICK 컬럼 (등록 상표) (Pharmacia 제품)을 통해 통과시켜 제거하였다. 표지화된 DNA의 특이 활성은 6 × 10⁹ dpm/ μ g인 것으로 측정되었다. 이 표지화된 DNA는 TcaC-펩티드 특이성 프로브에 혼성된 게놈 라이브러리의 성분들로부터 제조된 콜로니 멤브레인을 프로빙하는데 사용하였다.

TcaC-프로브 라이브러리 스크린 (상기 참조)에서 동정된 12 개의 콜로니를 함유하는 멤브레인을 0.1X SSC 및 0.1 % SDS 1l 중에서 매회 대략 30 분 동안 2회 비등시킴으로써 방사성 TcaC-특이성 표지를 제거하였다. 방사성 표지의 제거는 6 시간 동안 필름 노출시켜 확인하였다. 이어서, 제거된 멤브레인은 상기에서 제조한 TcaB_i 펩티드 특이성 프로브로 배양시켰다. 표지화된 DNA를 10 분 동안 비등시켜 변성시키고나서, "최소 혼성" 용액 100 ml 중에서 60 °C에서, 1 시간 동안 배양시킨 필터에 가하였다. 이 온도에서 방새 하이브리다이제이션한 후, 프로브 용액을 제거하고, 여액을 다음과 같이 세척하였다 (모두 0.3X SSC 및 0.1 % SDS 중에서): 25 °C에서 5 분 동안 1 회, 새로운 용액 중에서 60 °C에서 1 시간 동안 1 회, 및 새로운 용액 중에서 63 °C에서 1 시간 동안 1회. 표준 방법에 의해 1.5 시간 동안 X-선 필름에 노출시킨 후, 4 개의 강하게 하이브리드된 콜로니가 관찰되었다. 이들은 TcaC-특이성 프로브와 마찬가지로 단리물, 22G12, 25A10, 26A5 및 26B10 이었다.

상기 TcaB_i 프로브 용액을 동일 부피 (약 100 ml)의 "최소 혼성" 용액으로 희석하고나서, 게놈 라이브러리의 800 원을 함유하는 멤브레인들을 스크리닝하는데 사용하였다. 상기한 바와 같이 하이브리다이제이션, 세척 및 X-선 필름에 노출시킨 후, 단지 4 개의 코스미드 클론 22G12, 25A10, 26A5 및 26B10이 이 프로브에 강하게 하이브리드되는 것으로 밝혀졌다.

<TcaC 및 YcaB_i 펩티드를 코딩하는 유전자를 함유하는 서브클론의 단리 및 그의 DNA 염기 서열의 결정>

균주 XL1 Blue MR 중의 3 개의 하이브리다이제이션-양성 코스미드들을 TB-Amp₁₀₀ 중에서 30 °C에서 밤새 진탕(200 rpm)시키면서 성장시켰다. 세포들을 원심분리에 의해 수확한 후, 코스미드 DNA를 제조원의 프로토콜에 따라 시판되는 키트 (BIGprep (등록 상표), 5 Prime 3 Prime, Inc. 제품, 콜로라도주 볼더 소재)를 사용하여 제조하였다. 단 하나의 코스미드, 26A5가 이 절차에 의해 성공적으로 단리되었다. 제한 효소 EcoR 1 (NEB)로 분해하고 전기영동에 의해 분석했을 때, 대략적인 크기 14, 10, 8 (백터), 5, 3.3, 2.9 및 1.5 kbp의 단편들이 검출되었다. 동일한 3 개의 균주 (8 ml 배양물; TB-Amp₁₀₀, 30 °C)로부터 코스미드 DNA를 단리하려는 두번째 시도는 boiling miniprep 방법 [Evans G. and G. Wahl., 1987, "Cosmid vectors for genomic walking and rapid restriction mapping." in Guide to Molecular Cloning Techniques. Meth. Enzymology, Vol. 152, S. Breger and A. Kimmel, eds., pgs. 604-610]을 이용하였다. 단 하나의 코스미드, 25A10이 이 방법에 의해 성공적으로 단리되었다. 제한 효소 EcoR 1 (NEB)로 분해하고 겔 전기영동에 의해 분석했을 때, 이 코스미드는 코스미드 26A5에서 이전에 나타났던 것과 동일한 단편화 패턴을 나타내었다.

26A5 코스미드 DNA의 시료 0.15 μ g을 공급자의 프로토콜에 따라 대장균 DH5 α 세포 (Gibco BRL 제품) 50 ml을 형질전환시키는데 사용하였다. 그 균주의 단일 콜로니 단리물을 TB-Amp₁₀₀ 4 ml 중에 접종시키고, 37 °C에서 8 시간 동안 성장시켰다. 클로람페니콜을 최종 농도 225 μ g/ml로 첨가하

고, 배양을 24 시간 동안 더 지속시키고나서, 세포들을 원심분리에 의해 수확하고, -20 °C에서 동결시켰다. 26A5 코스미드 DNA의 단리는 표준 알칼리 용균 미니프렘 (Maniatis et al., 문헌 참조, p. 382)에 의해 수행하고, 모든 단계에서 와동시키기 보다는 교반하거나 또는 가볍게 혼합시키면서 모든 부피를 50 %까지 증가시켜 변형시켰다. DNA 펠릿을 70 % 에탄올 중에서 세척한 후, 이것을 25 µg/ml 리보뉴클레아제 (Boehringer Mannheim 제품)를 포함한 TE에 용해시켰다.

<GZ4-유도 및 TcaBi-프로브에 하이브리드되는 EcoR I 단편의 동정>

코스미드 25A10 (XL1 Blue MR 세포로부터 추출) 대략 0.4 µg 및 코스미드 26A5 (클로람페니콜-중독된 DH5α 세포로부터 추출) 약 0.5 µg를 각각 EcoR I (NEB) 약 15 단위로 85 분 동안 분해하고, 방새 동결시키고, 이어서 65 °C에서 5 분 동안 가열시키고, 0.7 % 아가로스 겔 중에서 전기영동시켰다 [Seakem (등록 상표) LE, 1X TEA, 80 볼트, 90분]. DNA를 상기한 바와 같이 에티뮴 브로마이드로 염색시키고, 자외선 하에 사진을 찍었다. 코스미드 25A10의 EcoR I 분해물은 완전히 분해된 것이었으나, 코스미드 26A5의 시료는 이들 조건 하에서 단지 부분적으로 분해되었다. DNA 단편을 함유하는 아가로스 겔을 탈퓨린화하고, 변형시키고, 중화시킨 다음, Ausubel et al (CPMB, 문헌 참조)의 섹션 2.9에 기재된 바와 같이 고염 (20X SSC) 프로토콜을 사용하여 Magna NT 나일론 멤브레인 상에 써던 블로팅 (Southern Blotting) 시켰다. 이어서, 전이된 DNA들을 상기한 바와 같이 나일론 멤브레인에 UV-가교화시켰다.

플라스미드 단리물 GZ4의 삽입체에 대응하는 TcaC-펩티드 특이적 DNA 단편을 상기한 바와 같은 100 ml 반응 부피 중에서 PCR (등록 상표)에 의해 증폭하였다. 이러한 3회의 반응으로부터의 증폭 생성물을 풀링하고, 상기한 바와 같이 Qiaex 키트에 의해 1 % GTG (등록 상표) 아가로스 겔로부터 추출하고, 형광계에 의해 정량하였다. 겔 정제된 DNA 100 ng을 상기한 바와 같이 High Prime Labeling Mix (Boehringer Mannheim 제품)를 사용하여 ³²P-dCTP로 6.34 × 10⁸ dpm/µg의 특이 활성으로 표지화하였다.

³²P-표지화된 GZ4 프로브를 10분 동안 비등시킨 후, "최소 혼성" 완충액 1 ng/ml에 첨가하고, 분해된 코스미드 DNA 단편을 포함하는 써던 블롯 멤브레인을 첨가하고, 50 rpm에서 부드럽게 진탕하면서 60 °C에서 4 시간 동안 배양하였다. 이어서, 멤브레인을 25 °C에서 각각 약 5 분 동안 3 회 세척한 후 (최소 혼성 세척 용액), 60 °C에서 각각 30 분 동안 2 회 세척하였다. 블롯을 -70 °C에서 약 30 분 동안 (강화 스크린과 함께) 필름에 노출시켰다. GZ4 프로브는 2 개의 코스미드 26A5 및 25A10의 5.0 kbp (외관상 크기) EcoR I 단편에 강하게 하이브리다이제이션하였다.

멤브레인을 0.1X SSC + 0.1 % SDS 중에서 약 30 분 동안 비등시켜 방사성을 제거하고, 필름에 노출시켜 방사성 표지의 부재를 확인하였다. 이어서, 이것을 상기 콜로니 멤브레인의 스크리닝에 미리 사용된 "최소 혼성" 완충액 중에서 변형 TcaBi 프로브를 사용하여 60 °C에서 3.5 시간 동안 하이브리다이제이션시키고, 상기한 바와 같이 세척하고, 2 개의 강화 스크린과 함께 -70 °C에서 40 분 동안 필름에 노출시켰다. 코스미드를 모두 사용할 때, TcaBi 프로브는 약 5.0 kbp EcoR I 단편과 약하게 하이브리다이제이션하고 약 2.9 kbp의 단편과는 강하게 하이브리다이제이션하였다.

(DH5α 세포로부터의) 상기한 코스미드 26A5 DNA의 시료를 원하는 밴드를 서브클로닝하기 위한 DNA 공급원으로서 사용하였다. 이 DNA 2.5 µg를 총 부피 30 µl 중에서 EcoR I (NEB) 약 3 단위로 1.5 시간 동안 분해하여 부분 분해물을 수득하였으며, 겔 전기영동에 의해 확인하였다. pBC KS (+) DNA (Stratagene 제품) 10 µg를 총 부피 20 µl 중에서 EcoR I 20 단위로 1.5 시간 동안 분해하여 완전히 분해하고, 전기영동에 의해 확인하였다. EcoR I 및 절단된 DNA 제제 모두를 물을 사용하여 50 µl까지 희석하고, 각각에 동일 부피의 PCI를 가하고, 현탁액을 부드럽게 혼합하고, 마이크로원심분리기로 회전시키고, 수성 상등액을 수거하였다. 에탄올 150 µl로 DNA를 침전시키고, 혼합물을 -20 °C에서 방새 방치시켰다. 원심분리하고 건조시킨 후, EcoR I 및 분해된 pBC KS (+)를 TE 100 µl에 용해시키고, 부분 분해된 26A5를 TE 20 µl에 용해시켰다. DNA 회수를 형광계로 확인하였다.

별개의 반응에서, 약 60 ng의 EcoR I 및 분해된 pBC KS(+) DNA를 부분 분해된 코스미드 26A5 DNA 약 180 ng 또는 270 ng과 라이게이션시켰다. 라이게이션은 T4 리가제 및 New England BioLabs 제품의 완충액을 사용하여 15 °C에서 부피 20 µl 중에서 5 시간 동안 수행하였다. 멸균 TE를 사용하여 100 µl로 희석시킨 라이게이션 혼합물을 사용하여 공급자의 지시에 따라 냉동된 적격의 DH5α 세포 (Gibco BRL 제품)를 형질전환시켰다. 상이한 양 (25 내지 200 µl)의 형질전환된 세포를 1 mM IPTG 및 50 mg/l X-gal을 갖는 새로 제조한 고상 LB-Cam₃₅ 배지 위에 플레이팅하였다. 플레이트를 37 °C에서 약 20 시간 배양한 후, 약 3 시간 동안 암실에서 냉각시켜 삽입체 선별을 위한 색상을 강화시켰다. 백색 콜로니를 동일한 조성의 패치 플레이트 상에 플레이팅하고 37 °C에서 철야 배양하였다.

각각의 선별된 패치 플레이트의 2 개의 콜로니 리프트를 다음과 같이 제조하였다. 백색 콜로니를 새로운 플레이트에 선택한 후, 구형의 Magna NT 나일론 멤브레인을 패치 플레이트 상에 압착하고, 멤브레인을 들어올리고, 라이브러리 콜로니 막에 대하여 상기한 바와 같이 변형, 중화 및 UV 가교화되게 하였다. 가교화된 콜로니 리프트를 격렬하게 세척하고, 티슈로 과량의 세포 파쇄물을 부드럽게 닦아내었다. "최소 혼성" 프로토콜에 따라 한 세트를 상기한 GZ4 (TcaC) 프로브 용액과 하이브리다이제이션시키고, 다른 한 세트는 상기한 TcaBi 프로브 용액과 하이브리다이제이션시킨 후, 세척하고 라이브러리 콜로니 멤브레인에 대해서 기술한 바와 같이 필름에 노출시켰다.

단지 GZ4 프로브와, GZ4 및 TcaBi 모두와, 또는 단지 TcaBi 프로브와만 하이브리다이제이션 신호를 나타내는 콜로니를 추가 연구를 위하여 선별하고, 상기와 같이 단일 콜로니의 단리를 위해 IPTG 및 X-gal을 갖는 LB-Cam₃₅ 배지 상에 세포를 스트리킹하였다. 16 개의 상이한 단리물로부터의 약 35 개의 단일 콜로니를 액체 LB-Cam₃₅ 배지에 넣고, 37 °C에서 철야 성장시키고, 세포를 원심분리

에 의해 수거하고, 플라스미드 DNA를 문헌 [Maniatis et al., 문헌 참조, p. 368]에 따라 표준 알칼리 용균 미니프랩에 의해 단리하였다. DNA 펠렛을 TE + 25 µg/ml 리보뉴클레아제 A에 용해시키고, 형광계로 DNA 농도를 측정하였다. EcoR I 분해 패턴을 겔 전기영동에 의해 분석하였다. 다음과 같은 단리물을 유용하게 사용하였다. 단리물 A17.2는 재라이게이션된 pBC KS(+)¹만을 포함하며, (음성) 대조군으로 사용하였다. 단리물 D38.3 및 C44.1은 각각 pBC KS(+)¹에 삽입된 2.9 kbp의 TcaB_i - 하이브리다이제이션 EcoR I 단편만을 포함한다. 이들 플라스미드 pDAB2000 및 pDAB2001를 각각 도 2에 도시하였다.

단리물 A35.3은 pBC KS(+)¹에 삽입된 약 5 kbp의 GZ4 - 하이브리다이제이션 EcoR I 단편만을 포함한다. 이 플라스미드는 pDAB2002 (도 2 참조)로 명명되었다. 이들 단리물은 DNA 서열 분석을 위한 주형을 제공하였다.

플라스미드 pDAB2000 및 pDAB2001을 상기 BIGprep (등록 상표) 키트를 사용하여 제조하였다. 배양물 30 ml를 OD₆₀₀이 2가 될 때까지 TB-Cam₃₅ 중에서 철야 성장시킨 후, 플라스미드를 제조자의 지시에 따라 단리하였다. DNA 펠렛을 각각 TE 100 µl에 재용해시키고, 시료의 보전을 EcoR I 분해 및 겔 전기영동 분석법에 의해 확인하였다.

주형으로서 pDAB2000 DNA를 사용하는 하나의 복제물과 주형으로서 pDAB2001 DNA를 사용하는 다른 하나의 복제물을 사용하여 이중으로 서열화 반응을 실시하였다. 반응은 GZ4/HB14 DNA의 서열 분석에 대해 상기한 바와 같은 디데옥시 염료 종결자 주기 서열 분석법을 사용하여 수행하였다. 초기 서열화는 프라이머로서 상기 LacZ 및 T7 프라이머, 및 TcaB_i PCR 증폭 생성물 (TH1 = ATTGCAGACTGCCAATCGCTTCGG, TH12 = GAGAGTATCCAGACCGCGGATGATCTG)의 결정된 서열을 기초로 한 프라이머를 사용하였다.

각각의 서열화 결과의 배열 및 편집 후, 각각의 서열을 Perkin Elmer Applied Biosynthesis Division SeqEd 675 소프트웨어에 의해 번역한 염색체 데이터의 보전에 따라 250 내지 350 염기로 절단하였다. 후속하는 서열화 "단계"는 새로운 프라이머에 적합한 서열을 선택하여 실시하였다. 몇가지의 예외를 제외하고, 프라이머 (상기와 같이 합성됨)는 50 % G+C 조성을 갖는 24 염기의 길이를 가졌다. 이러한 방법에 의한 서열화는 약 2.9 kbp EcoR I 단편의 가닥 모두에 대해 수행하였다.

DNA 서열화를 위한 주형으로서 사용하기 위해, 단리물 pDAB2002로부터의 플라스미드 DNA를 BIGprep (등록 상표) 키트에 의해 제조하였다. 서열화 반응을 수행하고 상기한 바와 같이 분석하였다. 초기에는, T3 프라이머 (pBS KS(+)¹ 염기 774-796: CGCGCAATTAACCCTCACTAAAG) 및 T7 프라이머 (pBS KS(+)¹ 염기 621-643: GCGCGTAATACGACTCACTATAG)를 사용하여 인접하는 벡터 서열로부터 서열화 반응물을 프라이밍하여 삽입 DNA로 판독하였다. 다른 프라이머 세트 (GZ4F: GTATCGATTACAACGCTGTCACTTCCC; TH13: GGGAAAGTGACAGCGTTGTAATCGATAC; TH14: ATGTTGGGTGCGTGGCTAATGGACATAAC; 및 LW1-204: GGGAAAGTGACAGCGTTGTAATCGATAC)를 제조하여 서브클로닝된 TcaC-펩티드 PCR 생성물의 동작성 올리고뉴클레오티드-매개 서열화에 의해 미리 결정된 내부 서열로부터 프라이밍하였다. 서열화의 초기 순환 동안 생성된 데이터로부터 새로운 세트의 프라이머를 고안하여 약 5 kbp 단편의 총 길이를 프라이밍하기 위해 사용하였다. 총 55 개의 올리고 프라이머를 사용하여 총 4832 bp의 연속 서열을 확인할 수 있었다.

pDAB2002의 EcoR I 단편 삽입체의 DNA 서열을 pDAB2000/pDAB2001 단리물의 결정 서열의 일부와 합할 때, 총 6005 bp의 연속 서열이 생성되었다 (서열 25로서 표시됨). 긴 개방 해독 프레임을 대응하는 아미노산으로 번역할 때, 서열은 메티오닌 잔기 (번역 개시부) 바로 전의 염기 68-124에 의해 코딩되는 TcaB_i N-말단 펩티드 (서열 3으로 표시됨)를 분명하게 나타내었다. 상류는 잠재적인 리보솜 결합 부위 (염기 51-58)이고, 하류는 염기 215-277에서 TcaB_i-PT158 내부 펩티드 (서열 19로 표시됨)가 코딩된다. 또한, 동일한 해독 프레임에서 하류의 염기 1787-1822에는 TcaB_i-PT108 내부 펩티드 (서열 20으로 표시됨)를 코딩하는 서열이 존재한다. 또한, 동일한 해독 프레임의 염기 1946-1972에서는 TcaB_{ii} N-말단 펩티드 (서열 5)가 코딩되고, 해독 프레임은 뉴클레오티드 3632-3634에서 번역 종결 코돈까지 방해받지 않고 계속된다.

TcaB_i-PT108을 코딩하는 서열의 말단과 TcaB_{ii} 코딩 영역의 개시부 사이의 프레임 내 (in-frame) 정지 코돈의 결여, 및 TcaB_{ii} 코딩 영역 바로 상류의 식별할 수 있는 리보솜 결합 부위의 결여는 펩티드 TcaB_{ii} 및 TcaB_i가 서열 25의 염기쌍 65에서 시작하는 3567 bp의 단일 개방 해독 프레임에 의해 코딩됨을 나타내고, 후-번역 분해에 의해 1189 아미노산 (131,586 달톤; 서열 26으로 표시됨)의 단일 1차 유전자 생성물 TcaB로부터 유래할 가능성이 가장 크다. TcaB_{ii} N-말단 펩티드 바로 앞의 아미노산이 펩티드 TcaB_i의 C-말단 아미노산을 나타내는 경우, TcaB_{ii} (627 개의 아미노산)의 예상 질량은 70,814 달톤 (서열 28로 표시됨)이고, SDS-PAGE에 의해 관찰된 크기 (68 kDa) 보다 다소 크다. 이 펩티드는 1881 염기쌍 (서열 27로 표시됨)의 연속 스트레치에 의해 코딩될 것이다. TcaB_i의 천연 C-말단은 TcaB_i-PT108의 C-말단에 보다 근접해 있는 것으로 생각된다.

PT108의 분자량 (3.438 kDa; 이 펩티드의 N-말단 아미노산 서열 분석 동안 측정됨)을 통하여 30개의 아미노산의 크기를 예상할 수 있다. TcaB_i 코딩 영역의 C-말단 (서열 28의 604 위치에 있는 Glu)을 나타내기 위하여 이 펩티드의 크기를 사용하여, TcaB_i의 유도된 크기는 604 개의 아미노산 또는 68,463 달톤인 것으로 측정되어 실험적 관찰 결과와 보다 일치하였다.

1686 염기쌍 (서열 29로 표시됨)의 TcaB_{ii} 펩티드 코딩 영역의 번역은 예상 질량이 60,789 달톤인 562 아미노산 (서열 30으로 표시됨)의 단백질을 생성하며, 이것은 관찰치 61 kDa와 잘 일치한다.

잠재적인 리보솜 결합 부위 (염기 3682-3687)는 tcaB 개방 해독 프레임의 정지 코돈의 48 bp 하류

에서 발견된다. 염기 3694-3726에서 펩티드 TcaC (서열 2)의 N-말단을 코딩하는 서열이 발견된다. 이 N-말단 펩티드에 의해 개시되는 개방 해독 프레임은 염기 6005까지 방해되지 않고 계속된다 (2361 염기쌍, 서열 31의 처음 2361 염기쌍으로 표시됨). 전체 TcaC 펩티드 (외관상 크기 약 165 bp; 약 1500 아미노산)를 코딩하는 유전자 (tcaC)는 약 4500 bp를 포함할 것이다.

또한, 코스미드 26A5의 클로닝된 EcoR I 단편을 포함하는 다른 단리물 E20.6을 앞에서 언급한 GZ4 및 TcaB_i 프로브와의 상동성에 의해 확인하였다. 이 균주에 포함된 플라스미드 (pDAB2004, 도 2)의 DNA의 EcoR I 분해물의 아가로스 겔 분석은 추정된 크기 2.9, 5 및 3.3 kbp의 삽입 단편을 나타내었다. 플라스미드 pDAB2002의 서열로부터 고안된 프라이머로부터 개시된 DNA 서열 분석은 pDAB2004의 3.3 kbp EcoR I 단편이 pDAB2002에 나타난 5 kbp EcoR I 단편에 인접하여 존재함을 나타내었다. pDAB2002에서 발견된 2361 염기쌍 개방 해독 프레임은 pDAB2004 내의 다른 2094 염기 (서열 31의 염기쌍 2362 내지 4458로 표시됨)로 방해받지 않고 계속된다. 주형으로서 모(母) 코스미드 26A5 DNA를 사용한 DNA 서열 분석을 통하여 개방 해독 프레임의 연속성을 확인하였다. 또한, 개방 해독 프레임 (TcaC 서열 31)은 4455 개의 염기쌍을 포함하고, 1485 개 아미노산의 단백질 (TcaC) (서열 32로 표시됨)을 코딩한다. 분자 크기의 계산치 166,214 달톤은 TcaC 펩티드의 추정 크기 (165 kDa)와 일치하고, 유도된 아미노산 서열은 TcaC N-말단 서열 (서열 2)에 개시된 서열과 정확하게 일치하였다.

동작성 올리고뉴클레오티드 프라이머 풀을 고안하기 위해 사용된 서열 17에 대응하는 아미노산 서열이 발견된 서열 내에 존재하지 않는 것은 초기 라이브러리 스크린에서 프로브로서 사용된 단리물 GZ4 및 HB14에서 발견된 PCR (등록 상표) 생성물의 생성이 동작성 풀의 프라이머 중 하나에 의해 역-가닥 프라이밍 (reverse-strand priming)에 의해 우연하게 생성되었음을 나타낸다. 또한, 유도된 단백질 서열은 서열 18로서 표시된 내부 단편을 포함하지 않는다. 이들 서열은 플라스미드 pDAB2004가 TcaC 펩티드의 상보체 코딩 영역을 포함한다는 것을 나타낸다.

서열 25을 추가로 분석한 결과는, 본 명세서에서 서열 35로서 개시된 TcaA_{iii}의 최종 13 개의 아미노산을 코딩하는 개방 해독 프레임 (염기 1-43)의 말단을 나타낸다. 단지 24 개의 염기가 TcaA_{iii} 코딩 영역의 말단과 TcaB_i 코딩 영역의 개시를 분리한다. 리보솜 결합 자리로서 작용할 수 있는 서열이 24 염기 내에 포함된다. 가능하기는 하지만, 포토랍두스 유전자 프로모터가 이러한 짧은 영역 내에서 코딩될 것 같지는 않다. 세 개의 긴 개방 해독 프레임 [단지 59 염기에 의해 tcaB 말단으로부터 분리되는 tcaA (서열 33), tcaB (서열 25, 염기 65-36334) 및 tcaC (서열 31)]을 포함하는 게놈 영역 tca를 tcaA 유전자 (서열 33)의 개시부의 초기 상류를 전사하고 폴리시스트론 메신저 RNA를 생성하면서 오페론으로서 조절한다.

<실시에 9>

포토랍두스 게놈 라이브러리에서의 TcbA_{ii} 펩티드 코딩 유전자 스크리닝

본 실시예는 TcbA_{ii} 펩티드 코딩 유전자를 포함하는 DNA 클론을 확인하는데 사용되는 방법, 상기 유전자의 단리 및 그의 부분 DNA 염기 서열의 결정을 기술한다.

프라이머 및 PCR 반응

곤충 활성 제제의 TcbA_{ii} 폴리펩티드는 약 206 kDa이다. 이 펩티드의 N-말단 아미노산 서열은 서열 1로서 나타나 있다. 실시예 8에서 기술한 바와 같이, 이 아미노산 서열의 일부를 코딩하기 위해 동작성 올리고뉴클레오티드 프라이머의 4 개의 풀 ("전향 프라이머": TH-4, TH-5, TH-6 및 TH-7)을 합성하였고, 아래에 나타내었다.

[표 12]

| 아미노산 | Phe | Ile | Gln | Gly | Tyr | Ser | Asp | Leu | Phe |
|------|------------|-----|---------|-----|---------|---------|---------|---------|-------|
| TH-4 | 5'-TT(T/C) | ATI | CA(A/G) | GGI | TA(T/C) | TCT | GA(T/C) | CTI | TT-3' |
| TH-5 | 5'-TT(T/C) | ATI | CA(A/G) | GGI | TA(T/C) | AG(T/C) | GA(T/C) | CTI | TT-3' |
| TH-6 | 5'-TT(T/C) | ATI | CA(A/G) | GGI | TA(T/C) | TCT | GA(T/C) | TT(A/G) | TT-3' |
| TH-7 | 5'-TT(T/C) | ATI | CA(A/G) | GGI | TA(T/C) | AG(T/C) | GA(T/C) | TT(A/G) | TT-3' |

또한, 내부 펩티드 제제 (TcbA_{ii}-PT81)의 1차 ("a") 및 2차 ("b") 서열을 결정하고, 서열 23 및 서열 24로서 각각 표시하였다. 동작성 올리고뉴클레오티드 ("역방향 프라이머": TH-8, TH-9, TH-10 및 TH-11)의 4 개의 풀을 유사하게 고안하고, 서열 23의 펩티드의 일부를 코딩하는 서열의 역방향 상보체를 코딩하기 위해 하기와 같이 합성하였다.

[표 13]

| | | | | | | | | | |
|-------|-------------------|---------|----------------|------------|------------|--------------------|--------------------|--------------------|------------|
| 아미노산 | Thr | Tyr | Leu | Thr | Ser | Phe | Glu | Gln | Val |
| TH-8 | 3' TGI AT(A/G) | AT(A/G) | GAI TT(A/G) | TGI TGI | AGI AGI | AA(A/G) AA(A/G) | CT(T/C) CT(T/C) | GT(T/C) GT(T/C) | CAI CAI |
| TH-9 | 3' TGI AT(A/G) | AT(A/G) | GAI TT(A/G) | TGI TGI | AGI AGI | AA(A/G) AA(A/G) | CT(T/C) CT(T/C) | GT(T/C) GT(T/C) | CAI CAI |
| TH-10 | 3' TGI AT(A/G) | AT(A/G) | GAI TT(A/G) | TGI TGI | AGI AGI | AA(A/G) AA(A/G) | CT(T/C) CT(T/C) | GT(T/C) GT(T/C) | CAI CAI |
| TH-11 | 3' TGI AT(A/G) | AT(A/G) | GAI TT(A/G) | TGI TGI | AGI AGI | AA(A/G) AA(A/G) | CT(T/C) CT(T/C) | GT(T/C) GT(T/C) | CAI CAI |

상기 프라이머의 세트를 PCR 반응에 사용하여 실시예 6에서 제조한 게놈 포토랩두스 루미네센스 W-14 DNA로부터의 TcbA_{ii} 코딩 유전자 단편을 증폭하였다. 모든 PCR 반응은 AmpliWax (등록 상표) gems 및 다른 Perkin Elmer 시약 및 프로토콜을 사용하여 "핫 스타트 (Hot Start)" 기술로 실시하였다. 일반적으로, 총 부피 11 μ l의 MgCl₂, dNTP's, 10X GeneAmp (등록 상표) PCR 완충액 II 및 프라이머의 혼합물을 단일 왁스 비드를 포함하는 튜브에 가하였다 [10X GeneAmp (등록 상표) PCR 완충액 II는 100 mM Tris-HCl, pH 8.3; 및 500 mM KCl로 구성됨]. 튜브를 2분 동안 80 °C로 가열하고 냉각시켰다. 왁스 밀봉의 상부에 10X GeneAmp (등록 상표) PCR 완충액 II, DNA 주형 및 AmpliTaq (등록 상표) DNA 폴리머라제를 포함하는 용액을 가하였다. 왁스 밀봉을 용융시키고 열순환시켜 성분을 혼합한 후의 최종 반응 조건 (부피 50 μ l)은 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 2.5 mM MgCl₂; 각각 200 μ M의 dATP, dCTP, dGTP, dTTP; 1.25 mM 단일 전향 프라이머 풀; 1.25 μ M 단일 역방향 프라이머 풀, AmpliTaq (등록 상표) DNA 폴리머라제 1.25 단위 및 주형 DNA 170 ng이었다.

반응물을 실시예 8에서와 같은 열 순환기 중에 놓고 하기 프로그램으로 가동시켰다.

[표 14]

| 온도 | 시간 | 주기 반복 |
|------------|------|-------|
| 94 °C | 2 분 | 1X |
| 94 °C | 15 초 | 30X |
| 55 - 65 °C | 30 초 | 30X |
| 72 °C | 1 분 | 30X |
| 72 °C | 7 분 | 1X |
| 15 °C | 일정 | |

동형성 프라이머 풀을 사용하여 3 개의 상이한 어닐링 온도 (55, 60, 65 °C)에서 일련의 증폭을 실시하였다. 65 °C에서 어닐링시킨 반응물은 아가로스 겔 전기영동 후에 가시적인 증폭 생성물을 갖지 않았다. 60 °C에서 어닐링시키고 프라이머 TH-5+TH-10을 포함하는 반응물은 2.9 kbp에 대응하는 이동성을 갖는 증폭 생성물을 생성하였다. 상기 조건에서 프라이머 TH-7+TH-10을 사용하면 보다 소량의 2.9 kbp 생성물이 생성되었다. 반응물을 55 °C에서 어닐링시킬 때, 상기 프라이머쌍은 2.9 kbp 생성물을 보다 많이 생성하였고, 이 생성물은 또한 프라이머쌍 TH-5+TH-8 및 TH-5+TH-11에 의해서도 생성되었다. 추가의 매우 희미한 2.9 kbp 밴드가 프라이머쌍 TH-7 + TH-8, TH-9,

TH-10 또는 TH-11로부터의 증폭 생성물을 포함하는 레인에서 발견되었다.

클로닝 및 DNA 서열 결정을 위한 충분한 PCR 증폭 생성물을 얻기 위해서 프라이머 TH-5+TH-10를 사용하여 10 개의 별개의 PCR 반응을 준비하고, 55 °C의 어닐링 온도에서 상기 조건을 사용하여 실시하였다. 모든 반응물을 모아서 상기한 바와 같이 아가로스 겔로부터 Qiaex 추출에 의해 2.9 kbp 생성물을 정제하였다.

TcbA_{ij} 내부 펩티드에 대해 결정한 또다른 서열을 서열 21 및 서열 22로 표시하였다. 상기한 바와 같이, 이 펩티드의 아미노산 서열의 일부를 코딩하는 서열의 역방향 상보체에 대응하도록 동의성 올리고뉴클레오티드 (역방향 프라이머 TH-17 및 TH-18)를 제조하였다.

[표 15]

서열 21로부터

| | | | | | | | | | |
|-------|--------|-------|-----|-------|-------|-----|-------|-------|-------|
| 아미노산 | Met | Glu | Thr | Gln | Asn | Ile | Gln | Glu | Pro |
| TH-17 | 3'-TAC | CTT/C | TGI | GTT/C | TTA/G | TAI | GTT/C | GTT/C | GG-5' |

[표 16]

서열 22로부터

| | | | | | | | | | | |
|-------|------------|-----|-----|---------|-----|---------|-----|-----|-----|------------|
| 아미노산 | Asn | Pro | Ile | Asn | Ile | Asn | Thr | Gly | Ile | Asp |
| TH-18 | 3'-TT(A/G) | GGI | TAI | TT(A/G) | TAI | TT(A?G) | TGI | CCI | TAI | CT(A/G)-5' |

주형으로서 포토랍두스 루미네센스 W-14 DNA를 사용하고 5'-(전향) 프라이머로서 프라이머 TH-4, TH-5, TH-6 또는 TH-7을 사용한 증폭 실험에 동의성 올리고뉴클레오티드 TH-18 및 TH-17을 사용하였다. 이들 반응은 각각 약 4 kbp 및 4.5 kbp의 생성물을 증폭시켰다. 이들 DNA를 아가로스 겔로부터 나일론 멤브레인에 전사하여 TH-5+TH-10 프라이머 쌍에 의해 증폭된 2.9 kbp 생성물로부터 제조한 ³²P-표지화된 프로브 (상기한 바와 같음)와 하이브리다이제이션시켰다. 4 kbp 및 4.5 kbp 증폭 생성물은 모두 2.9 kbp 프로브와 강하게 하이브리다이제이션하였다. 이러한 결과를 사용하여 도 3에 도시한 TcbA_{ij} 내부 펩티드 서열을 배열한 지도를 제작하였다. 프라이머들간의 대략적인 거리는 도 3의 뉴클레오티드에 나타나 있다.

2.9 kbp TcbA_{ij} 코딩 단편의 DNA 서열

상기에서 제조된 정제 2.9 kbp 단편 약 200 ng을 에탄올을 사용하여 침전시키고 물 17 ml에 용해시켰다. 이 용액의 반을 프라이머로서의 TH-5 풀 25 pmol과 함께 서열화 주형으로서 사용하고, 나머지 반은 TH-10 프라이밍을 위한 주형으로서 사용하였다. 서열화 반응은 실시예 8에 나타난 바와 같이 실시하였다. TH-10 프라이머 풀을 사용한 경우에는 신뢰할 수 있는 서열이 생성되지 않았지만, TH-5 프라이머 풀을 사용한 반응은 아래와 같은 서열을 생성시켰다.

```

1  AATCGTGTG ATCCCTATGC CGNGCCGGT TCGGTGGAAT CGATGTCCTC ACCGGGGGTT
61  TATTNGAGGG ANTNGTCCCG TGAGGCCAAA AANTGGAATG AAGAAGTTC AATTTNTTAC
121 CTAGATAAAC GTCCGCCGGN TTTAGAAAGN TTANTGNTCA GCCAGAAAT TTGTTTGAG
181 GAAATTCAC CGNTGGTTCT CTCTATTGAT TNGGCCCTGG CCGGTTCCGA ANNAACNA
241 GAAATTCAC AAGTTGAGGT GATGGNTTGG TNGCMANCTT NTCGTTTAGG TGGGAGAAA
301 CCTTNTCANC ACGNTTNTGA AACTGTCCGG GAAATCGTCC ATGANCGTGA NCCAGGNTTN
361 CGCCATTGG
    
```

상기 서열을 기초로, 서열화 프라이머 (TH-21, 5'-CCGGGCGACGTTTATCTAGG-3')를 역방향 상보체 염기 120-139로 고안하여 겔-정제된 2.9 kbp TcbA_{ij} 코딩 PCR 단편의 5' 말단 (즉, TH-5 말단) 쪽으로 중합을 개시하였다. 결정된 서열은 아래와 같고, TcbA_{ij} (서열 1)의 생화학적으로 결정된 N-말단 펩티드 서열과 비교하였다.

TcbA_{ij} 2.9 kbp PCR 단편 서열 확인

(밑줄그은 아미노산 = 동의성 올리고뉴클레오티드에 의해 코딩됨)

```

서열번호 : 1      F I O G Y S D L F G - - A
2.9 kbp seq GC ATG CAG GGG TAT AGT GAC CTG TTT GGT AAT CGT GCT
                M Q G Y S D L F G N R A >
    
```

유도된 아미노산 서열과 생화학적으로 결정된 서열과의 상동성으로부터 2.9 kbp PCR 단편이 TcbA 코딩 영역을 나타낸다는 것을 명백하게 알 수 있다. 이어서, 이 2.9 kbp 단편을 혼성 프로브로서 사용하여 TcbA_{ij} 코딩 유전자를 포함하는 코스미드에 대해 실시예 8에서 제조된 포토랍두스 W-14 게놈 라이브러리를 스크리닝하였다.

포토랍두스 코스미드 라이브러리의 스크리닝

2.9 kb 겔 정제된 PCR 단편을 실시예 8에서 기술한 바와 같이 보링거 만하임 High Prime 표지화 키트를 사용하여 ³²P로 표지화하였다. 코스미드 라이브러리로부터 약 800 개의 콜로니의 잔류물을 포함하는 필터를 실시예 8에서 상기한 바와 같이 스크리닝하고, 양성 클론을 단리 콜로니를 위해

스트리킹하고 재스크리닝하였다. 3 개의 클론 (8A11, 25G8 및 26D1)가 수 회의 스크리닝 및 특성화 단계를 통하여 양성 결과를 보였다. TcbA_{ii} 특이 프로브는 실시예 8에서 확인된, tcaB 및 tcaC 유전자를 포함하는 4 개의 코스미드 중 어느 것과도 하이브리다이제이션하지 않았다. 코스미드 8A11, 25G8 및 26D1으로부터의 DNA를 제한 효소 Bgl II, EcoR I 또는 Hind III (단독으로 또는 서로 혼합됨)로 분해하고, 단편을 아가로스 겔 상에서 분리하고, 실시예 8에서 기술한 바와 같이 나일론 멤브레인으로 옮겼다. 멤브레인을 4.5 kbp 단편 (포도랍두스 게놈 DNA를 프라이머 TH-5+TH-17로 증폭하여 생성됨)으로부터 제조된 ³²P-표지화된 프로브와 하이브리다이제이션시켰다. 코스미드 DNA 8A11 및 26D1로부터 생성된 패턴은 동일한 멤브레인 상에 유사하게 절단된 게놈 DNA로 생성된 것과 동일하였다. 코스미드 8A11 및 26D1은 게놈 TcbA_{ii} 코딩 로커스를 정확하게 나타낸다는 결론을 얻었다. 그러나, 코스미드 25G8은 게놈 DNA 보다 약간 더 큰 하나의 Bgl II 단편을 갖는다. 이것은 벡터 내에 존재하는 삽입체로 인한 것일 수 있다.

tcbA 코딩 유전자의 DNA 서열

코스미드 26D1의 멤브레인 하이브리다이제이션 분석은 4.5 kbp 프로브가 하나의 큰 EcoR I 단편 (9 kbp보다 큼)에 하이브리다이제이션한다는 것을 보여주었다. 이 단편을 겔 정제하여 실시예 8에서 기술한 바와 같이 pBC KS(+)₁의 EcoR I 부위에 라이게이션시켜 플라스미드 pBC-S1/R1을 생성시켰다. 이 플라스미드의 삽입 DNA의 부분 DNA 서열을 실시예 8에서 기술한 과정을 사용하여 인접 벡터 서열로부터 "프라이머 워킹 (primer walking)"에 의해 결정하였다. 미리 결정된 서열로부터 고안된 새로운 올리고뉴클레오티드를 신장시켜 다른 서열을 생성시켰다. 다른 방법으로 확인된 tcbA 유전자의 결정된 DNA 서열 (실시예 12에 기술된 바와 같이 서열 11로 표시됨)과 비교시에 완전한 상동성이 뉴클레오티드 1-272, 319-826, 2578-3036 및 3068-3540 (총 염기 1712)에서 발견되었다. 두 개의 방법을 사용하여 TcbA_{ii} 펩티드를 코딩하는 DNA 단편을 확인할 수 있다는 결론을 얻었다.

tcbA 유전자의 유도된 아미노산 서열의 분석

서열 11로 확인된 DNA 단편의 서열은 유도된 아미노산 서열이 서열 12로 표시된 단백질을 코딩한다. 몇 개의 특징을 통하여 TcbA_{ii} 단백질을 코딩하는 유전자의 정체를 확인할 수 있다. TcbA_{ii} N-말단 펩티드 (서열 1; Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn Arg Ala)는 아미노산 88-100으로서 코딩된다. TcbA_{ii} 내부 펩티드 TcbA_{ii}-PT81(a) (서열 23)는 아미노산 1065-1077로서 코딩되고 TcbA_{ii}-PT81(b) (서열 24)는 아미노산 1571-1592로서 코딩된다. 또한, 내부 펩티드 TcbA_{ii}-PT56 (서열 22)는 아미노산 1474-1488로서 코딩되고 내부 펩티드 TcbA_{ii}-103 (서열 21)은 아미노산 1614-1639로서 코딩된다. 이 유전자는 포도랍두스 루미네센스 균주 W-14의 살충 단백질 제제로부터 단리된 TcbA_{ii} 펩티드를 코딩하는 진정한 클론임이 분명하다.

펩티드 TcbA_{ii}로서 단리된 단백질은 보다 긴 펩티드의 분해로부터 유도된다. 이에 대한 증거는 TcbA_{ii} N-말단 펩티드 (서열 1)를 코딩하는 뉴클레오티드가 보다 긴 개방 해독 프레임 (서열 11)의 261 개의 염기 (87 개의 N-말단-부근 아미노산을 코딩함) 뒤에 온다는 사실에 의해 제공된다. 이 해독 프레임은 큰 펩티드 TcbA의 N-말단 서열에 대응하고 서열 16으로 표시된 아미노산 서열 Met Gln Asn Ser Leu를 코딩하는 뉴클레오티드로 시작한다. TcbA는 TcbA_{ii}의 전구체 단백질인 것으로 생각된다.

tcbA, tcaB 및 tcaC 유전자의 관계

tcbB 및 tcaC 유전자는 밀접하게 연관되어 있고 하나의 mRNA로서 전사될 수 있다 (실시예 8). tcbA 유전자는, 각각의 게놈 라이브러리가 확인된 상이한 코스미드를 스크리닝하기 때문에 tcaB 및 tcaC 클러스터를 포함하는 것과 분명히 중복되지 않는 코스미드 상에 존재한다. 그러나, tcaB 및 tcaC 유전자와 tcbA 유전자에 의해 코딩되는 아미노산 서열을 비교하면 실질적인 정도의 상동성이 존재함을 알 수 있다. 아미노산 보존 (Protein Alignment Mode of MacVector (등록 상표) 서열 분석 소프트웨어, scoring matrix pam250, hash value = 2; 미국 캘리포니아주 캠벨 소재 Oxford Molecular Group)은 도 4에 도시하였다. 도 4에서 각 패널의 스코어 라인에서 상부 캐럿 (^)은 상동성 또는 보존적 아미노산 변화를 나타내고, 하부 캐럿 (v)은 비상동성을 나타낸다.

이 분석은 TcbA 펩티드 잔기 1739 내지 1894의 아미노산 서열이 TcaB_i 펩티드의 아미노산 441 내지 603 (TcaB의 총 627 아미노산 중 162 개; 서열 28)과 상동성이 매우 높다. 또한, TcbA 아미노산 1932 내지 2459의 서열은 펩티드 TcaB_{ii}의 아미노산 12 내지 531 (총 562 개의 아미노산 중 520 개; 서열 30)과 상동성이 매우 높다. TcbA 펩티드 (서열 12)가 2505 개의 아미노산을 포함한다는 것을 고려할 때, 그의 C-말단부 근처에서 총 684 개의 아미노산 (27 %)이 TcaB_i 또는 TcaB_{ii} 펩티드에 상동성이고, 이 상동부는 추정 TcaB 전구단백질 (서열 26)의 배열에 직선상으로 배열된다. TcbA 상동부 내의 상당한 갭은 TcaB 전구단백질의 TcaB_i와 TcaB_{ii} 부분 사이의 연결부에 대응한다. 분명한 것은, TcbA 및 TcaB 유전자 생성물은 진화상 관련되어 있고, 이것은 이들이 포도랍두스에서 몇몇의 공통적인 기능(들)을 공유한다는 것을 시사한다.

<실시예 10>

포도랍두스 브로스 (broth)에서 아연-메탈로프로테아제의 특성화; 프로테아제 억제, 분류 및 정제

프로테아제 억제 및 분류 분석: 프로테아제 분석은 기질로서 물에 용해시킨 FITC-카제인을 사용하여 실시하였다 (최종 분석 농도 0.08 %). 적합한 완충액 중에서 포도랍두스 브로스 25 μl (총 반을 부피 150 μl)로 25 °C에서 1 시간 동안 단백질 분해 반응을 실시하였다. 또한, 시료를 디티오 트레이틀의 존재 또는 부재 하에 분석하였다. 배양한 후, 동량의 12 % 트리클로로아세트산을 가

하여 분해되지 않은 단백질을 침전시켰다. 0.5 시간 동안 침전시키고 원심분리한 후, 상등액 100 μ l를 96웰 미세적정 플레이트 내에 넣고 동량의 4N NaOH를 가하여 용액의 pH를 조절하였다. 이어서, 각각 485 및 538 nm의 여기 및 방출 파장에서 Fluoroskan II 형광 플레이트 판독기를 사용하여 단백질 분해를 정량하였다. 프로테아제 활성을 pH 5.0 - 10.0에 걸쳐서 0.5 단위로 시험하였다. 하기 완충액을 최종 농도 50 mM에서 사용하였다: 아세트산나트륨 (pH 5.0-6.5), Tris-HCl (pH 7.0-8.0), 및 bis-Tris 프로판 (pH 8.5-10.0). 관찰된 프로테아제(들)의 종류를 확인하기 위해서, 조 브로쓰를 상이한 프로테아제 억제제 (최종 농도 0.5 μ g/ μ l)로 처리한 후 상기 기질을 사용하여 pH 8.0에서 프로테아제 활성을 조사하였다. 사용된 프로테아제 억제제로는 E-64 (L-트랜스-에폭시사키닐류실아미도[4-, -구아니디노]-부탄), 3,4-디클로로이소쿠마린, 류펩틴, 펩스타틴, 아마스타틴, 에틸렌디아민테트라아세트산 (EDTA) 및 1,10-페난트롤린을 들 수 있다.

특정 pH 범위에 걸쳐 실시한 프로테아제 분석은 pH 약 8.0에서 최대 활성을 보이는 프로테아제(들)이 존재한다는 것을 보여주었다 (표 17). DTT의 첨가는 프로테아제 활성에 아무런 영향을 주지 않았다. 이어서, 조 브로쓰를 상이한 프로테아제 억제제로 처리하였다 (표 18). 상기한 억제제로 조 브로쓰를 처리하면 1,10-페난트롤린은 최종 농도 50 μ g으로 첨가했을 경우 2 mg/ml의 조 브로쓰 용액 100 μ l에서 IC_{50} = 5 μ g으로서 모든 프로테아제 활성을 완전히 억제한다는 것을 알 수 있었다. 이들 데이터는 포토랩두스 브로쓰에서 발견된 대부분의 풍부한 프로테아제(들)이 아연-메탈로프로테아제 부류의 효소라는 것을 나타낸다.

[표 17]

포토랩두스 루미네센스 (균주 W-14)의 1일 생성에서 발견된 프로테아제 활성에 대한 pH의 효과

| pH | 형광 단위 ^a | 백분율 활성 ^b |
|------|--------------------|---------------------|
| 5.0 | 3013 ± 78 | 17 |
| 5.5 | 7994 ± 448 | 45 |
| 6.0 | 12965 ± 483 | 74 |
| 6.5 | 14390 ± 1291 | 82 |
| 7.0 | 14386 ± 1287 | 82 |
| 7.5 | 14135 ± 198 | 80 |
| 8.0 | 17582 ± 831 | 100 |
| 8.5 | 16183 ± 953 | 92 |
| 9.0 | 16795 ± 760 | 96 |
| 9.5 | 16279 ± 1022 | 93 |
| 10.0 | 15225 ± 210 | 87 |

a 형광 단위 = 형광 단위 (최대 = 약 28,000; 배경 = 약 2200)
 b pH 8.0에서의 최대치에 대한 백분율 활성

[표 18]

포토랩두스 루미네센스 (균주 W-14)의 1일 생성에서 발견된 pH 8에서의 프로테아제 활성에 대한 상이한 프로테아제 억제제의 효과

| 억제제 | 보정된 형광 단위 ^a | 백분율 억제 ^b |
|----------------------------|------------------------|---------------------|
| 대조군 | 13053 | 0 |
| E-64 | 14259 | 0 |
| 1,10 페난트롤린 ^c | 15 | 99 |
| 3,4 디클로로이소쿠마린 ^d | 7956 | 39 |
| 류펩틴 | 13074 | 0 |
| 펩스타틴 ^c | 13441 | 0 |
| 아마스타틴 | 12474 | 4 |
| DMSO 대조군 | 12005 | 8 |
| 메탄올 대조군 | 12125 | 7 |

a 보정된 형광 단위 = 형광 단위 - 배경 (2200 형광 단위)
 b pH 8.0에서의 프로테아제 활성에 대한 백분율 억제
 c 억제제를 메탄올에 용해시킴.
 d 억제제를 DMSO에 용해시킴.

투석된 10 내지 80 % 암모늄 술페이트 펠렛을 포토랍두스 독소에 대해 실시예 5에서 기술한 바와 같이, 50 mM Na₂PO₄, pH 7.0에서 평형화된 Q 세파로스 컬럼에 가하여 아연-메탈로프로테아제를 분리하였다. 강하게 세척한 후, 0 내지 0.5 M NaCl 구배를 사용하여 독소 단백질을 용출시켰다. 대부분의 생물학적 활성 및 단백질을 0.15 내지 0.45 M NaCl로부터 용출시켰다. 그러나, 대부분의 단백질 분해 활성이 0.25 - 0.35 M NaCl 분획에 존재하고 소량의 활성이 0.15 - 0.25 M NaCl 분획에 존재하는 것이 관찰되었다. 0.25 - 0.35 M NaCl 분획에 대한 SDS-PAGE 분석은 약 60 kDa의 펩티드 주밴드를 보였다. 0.15 - 0.25 M NaCl 분획은 유사한 60 kDa 밴드를 포함하였으나 보다 넓은 상대적인 단백질 농도를 가졌다. 이어서, 이 분획을 수퍼로스 12 HR 16/50컬럼을 사용하여 겔 여과함으로써 SDS-PAGE 분석에 의한 미세한 (총 염색된 단백질의 90 % 초과) 58.5 kDa 밴드를 포함하는, 57.5 kDa에서 이동하는 주피크를 얻었다. 상술한 바와 같은 각종 프로테아제 억제제를 사용한 상기 분획에 대한 추가의 분석을 통하여 프로테아제가 아연-메탈로프로테아제임을 결정하였다. 발효 1일째에 포토랍두스 브로쓰에 존재하는 거의 모든 프로테아제 활성은 약 58 kDa 아연-메탈로프로테아제에 대응하였다.

아연-메탈로프로테아제(들)의 제2 단리에서 3 일 동안 성장한 W-14 포토랍두스 브로쓰를 사용하고, 문헌 [Schmidt, T.M., Bleakley, B. and Nealson, K.M., 1988]에 기재된 바와 같이 젤라틴이 결합된 소용 도데실 술페이트 폴리아크릴아미드 겔 전기영동 (SDS-PAGE)을 사용하여 프로테아제 활성을 가시화하였다. SDS 전개 겔 (5.5 × 8 cm)은, 물 중에 용해시에 최종 농도 0.1 % 젤라틴 (Biorad EIA 그레이드 시약; 미국 캘리포니아주 리치몬드 소재)이 혼입되는 12.5 % 폴리아크릴아미드 (아크릴아미드/비스-아크릴아미드의 40 % 원액; 미국 미주리주 세인트루이스 소재의 Sigma Chemical Co.)로 제조하였다. SDS-스태킹 겔 (1.0 × 8 cm)를 0.1 % 젤라틴이 결합된 5 % 폴리아크릴아미드를 사용하여 제조하였다. 일반적으로, 시험 단백질 2.5 μg을 디티오트레이톨 (DTT)이 없는 SDS-PAGE 로딩 완충액 0.03 ml 중에 희석시키고 겔 상에 로딩하였다. 0 °C 및 8 mA에서 SDS 전개 완충액 (Laemmli, U.K. 1970. Nature 227, 680) 중에서 단백질을 전기영동시켰다. 전기영동이 종결된 후, 겔을 2.5 % (v/v) 트리톤 (Triton) X-100 중에서 2 시간 동안 세척하였다. 이어서, 겔을 0.1 M 글리신 (pH 8.0) 중에서 37 °C에서 1 시간 동안 배양하였다. 배양한 후, 겔을 고정시키고 메탄올-아세트산-물 (30:10:60, v/v/v; Sigma Chemical Co.) 중의 0.1 % 아미도 블랙으로 방색 염색하였다. 단백질 분해 및 후속한 결합 젤라틴의 확산에 의해 아미도 블랙 염색된 어두운 배경에 대하여 밝은 영역으로서 단백질 분해 활성이 가시화되었다. 58, 41 및 38 kDa에서 단백질 분해 활성에 의해 생성된 3 개 이상의 특이 밴드가 관찰되었다.

기질로서 물에 용해시킨 FITC-카제인을 사용하여 W-14 3일 배양 브로쓰 중에서의 상이한 프로테아제의 활성을 분석하였다 (최종 분석 농도 0.02 %). 단백질 분해 실험은 총 부피 0.15 ml로 상이한 단백질 분획을 사용하여 0.1 M Tris-HCl (pH 8.0) 중에서 0 내지 0.5 시간 동안 37 °C에서 실시하였다. 물에 용해시킨 동일 부피의 12 % 트리클로로아세트산 (TCA)을 가하여 반응을 종결시켰다. 실온에서 0.25 시간 동안 배양한 후, 시료를 10,000 × g에서 0.25 시간 동안 원심분리하고, 분취액 0.10 ml를 96-웰 미세적정 플레이트에 넣었다. 이어서, 동일 부피의 2 N 수산화나트륨을 가하여 용액을 중화시킨 후, Fluoroskan II 형광 플레이트 판독기를 사용하여 각각 485 및 538 nm의 여기 및 방출 파장으로 정량하였다. 상이한 프로테아제 농도로 FITC-카제인을 사용하여 37 °C에서 0 내지 10 분 동안 활성을 측정하였다. 활성의 단위는 1000 형광 단위/분을 생성시키는데 필요한 효소의 양으로 임의로 규정하고, 특이 활성은 단위/프로테아제 mg으로 규정하였다.

각각 100 % 에탄올 및 물 중에 용해시킨 억제제의 원액과 함께 2 개의 아연-메탈로프로테아제 억제제, 즉 1,10-페난트롤린 및 N-(α-람노피라노실옥시히드록시포스포닐)-Leu-Trp (포스포르아미돈)을 사용하여 억제 연구를 실시하였다. 원액 농도는 1,10-페난트롤린 및 포스포르아미돈에 대해 각각 일반적으로 10 mg/ml 및 5 mg/ml이었고, 억제제의 최종 농도는 반응당 0.5 내지 1.0 mg/ml이었다. 3일째의 W-14 조 브로쓰를 모든 아연-메탈로프로테아제의 억제제인 1,10-페난트롤린으로 처리하면 모든 프로테아제 활성이 완전히 제거되었고, 써모리신 (thermolysine) 유사 프로테아제의 억제제 [Weaver, L.H., Kester, W.R., and Matthews, B.W. 1977. J. Mol. Biol. 114, 119-132]인 포스포르아미돈으로 처리하면 프로테아제 활성이 약 56 % 감소하였다. 잔여 단백질 분해 활성은 추가의 포스포르아미돈으로 처리하여도 더 감소시킬 수 없을 것이다.

3일째의 W-14 포토랍두스 브로쓰의 프로테아제를 다음과 같이 정제하였다: Amicon M-12 여과 장치에 부착된 Amicon 나선형 한외여과 카트리지 Type S1Y100을 사용하여 브로쓰 4.0 l를 농축하였다. Amicon M-12 여과 장치에 부착된 Amicon 나선형 한외여과 카트리지 Type S1Y100을 사용하여 크기가 100 kDa 미만인 천연 단백질을 갖는 통과 물질 3.8 l를 0.375 l까지 농축하였다. 잔류 물질은 10 내지 100 kDa 크기의 단백질을 포함하였다. 이 물질을 10 mM 인산나트륨 완충액 (pH 7.0) 중에 평형화시킨 PerSeptive Biosystem (미국 매릴랜드주 프라임턴 소재) Poros (등록 상표) 50 HQ 강 음이온 교환 충전재로 충전된 Pharmacia HR16/10 컬럼 상에 로딩하였다. 단백질을 5 ml/분의 유속으로 컬럼에 로딩한 후 A₂₈₀ = 0.001 될 때까지 결합하지 않은 단백질을 완충액으로 세척하였다. 이어서, 7.5 ml/분의 유속으로 40분 내에 0 내지 1.0 M NaCl의 NaCl 구배를 사용하여 단백질을 용출시켰다. 상기와 같이, 프로테아제 활성에 대해 분획을 분석하고 활성 분획을 모았다. 단백질 분해 활성을 보이는 분획을 50 % (v/v) 10 mM 인산나트륨 완충액 (pH 7.0)으로 희석하고 10 mM 인산나트륨 중에 평형화시킨 Pharmacia HR 10/10 Mono Q 컬럼에 로딩하였다. A₂₈₀ = 0.001 될 때까지 완충액으로 컬럼을 세척한 후, 2.0 ml/분의 유속으로 1 시간 동안 0 내지 0.5 M NaCl의 NaCl 구배를 사용하여 단백질을 용출시켰다. 분획을 프로테아제 활성에 대하여 분석하였다. 포스포르아미돈 감수성 프로테아제 활성은 41/38 kDa 프로테아제에 의한 것으로 (하기 참조) 가장 높은 포스포르아미돈 감수성 프로테아제 활성을 갖는 분획을 모았다. 이 분획은 0.15 - 0.25 M NaCl에서 용출되는 것으로 판명되었다. 포스포르아미돈 불감성 프로테아제 활성, 58 kDa 프로테아제를 주로 포함하는 분획도 모았다. 이 분획은 0.25 - 0.35 M NaCl에서 용출되는 것으로 판명되었다. 이어서, 포스포르아미돈 감수성 프로테아제 분획을 Millipore Ultrafree (등록 상표)-15 원심분리 여과 장치 Biomax-5K NMWL 멤브레인을 사용하여 최종 부피 0.75 ml까지 농축하였

다. 이 물질을 10 mM 인산나트륨 완충액 (pH 7.0)/0.1 M NaCl 중에 평형화시킨 Pharmacia Sephadex G-50으로 충전된 Pharmacia HR10/30 컬럼 상에 0.5 ml/분의 유속으로 로딩하였다. 이어서, 최대 포스포라미돈 감수성 프로테아제 활성을 갖는 분획을 모아서 Millipore Ultrafree (등록 상표)-15 원심분리 여과 장치 Biomax-50K NMWL 멤브레인을 사용하여 원심분리하였다. 상기와 같은 단백질 분해 활성 분석은 이 물질이 포스포라미돈 감수성 프로테아제 활성만을 갖는다는 것을 나타내었다. 포스포라미돈 불감성 프로테아제, 58 kDa 단백질을 모은 후, Millipore Ultrafree (등록 상표)-15 원심분리 여과 장치 Biomax-50K NMWL 멤브레인 중에서 농축하고, Pharmacia Superdex-75 컬럼 상에서 추가로 분리하였다. 이 프로테아제를 포함하는 분획을 모았다.

정제된 58 및 41/38 kDa 프로테아제의 분석은 두 종류의 프로테아제가 1,10-페난트롤린으로 완전히 억제되지만, 포스포라미돈으로는 41/38 kDa 프로테아제만 억제되었음을 보여주었다. 조 브로쓰에 대한 추가의 분석은 1일째의 W-14 프로쓰의 프로테아제 활성이 41/38 kDa 프로테아제에 기인하여 전체 프로테아제 활성의 23 %이고, 3일째의 W-14 브로쓰에서는 44 %까지 증가한다는 것을 보여주었다.

단백질 순도를 조사하고 아미노 말단 서열을 얻기 위해서 미국 매사추세츠주 나티크 소재의 Integrated Separation Systems에서 시판하는 4 내지 20 % 구배 MiniPlus SeptraGels를 사용하여 표준 SDS-PAGE 분석을 실시하였다. 아미노 말단에서 서열화될 단백질을 정제 후에 PVDF 멤브레인 [미국 캘리포니아주 포스터 시티 소재의 Applied Biosystems사의 ProBlott (등록 상표) 멤브레인]에 블로팅하고, 0.1 % 아미노 블랙으로 가시화시키고 절개하여 서열화를 위해 미국 매사추세츠주 캠브리지 소재의 Cambridge Prochem에 보냈다.

3일된 W-14 브로쓰로부터의 58 (서열 45) 및 41/38 kDa (서열 44) 프로테아제의 추정된 아미노 말단 서열은 각각 DV-GSEKANEKLIK (서열 45) 및 DSGDDDKVTNTDIHR (서열 44)이었다.

41/38 kDa 프로테아제의 서열화는 각각 추가의 아미노산이 단백질 분해에 의해 제거된 몇 개의 아미노 말단을 보여 주었다. 38 및 41 kDa 폴리펩티드에 대한 1차, 2차, 3차 및 4차의 서열 조사는 상기 서열의 추정을 가능하게 하였고 이들 2 개의 프로테아제가 상동성이라는 것을 보여주었다.

<실시에 11a>

TcbA 펩티드 코딩 유전자에 대한 포도랍두스 게놈 라이브러리의 항체를 사용한 스크리닝

상술한 서열화와 동시에 TcbA_{ii} 펩티드 (서열 1)를 기초로 하여 적합한 프로빙 및 서열화를 실시하였다. 이 서열화는 박테리아 배양 브로쓰를 제조하고 상기 실시예 1 및 2에서 기술한 독소를 정제하여 실시하였다.

게놈 DNA를 Grace의 곤충 조직 배양 배지에서 성장한 포도랍두스 루미네센스 균주 W-14로부터 분리하였다. 박테리아를 250 ml 삼각 플라스크 중의 배양 배지 5 ml 중에서 28 °C 및 250 rpm에서 약 24 시간 동안 성장시켰다. 배양 배지 100 ml로부터 박테리아 세포를 5000 × g에서 10 분 동안 펠렛화시켰다. 상등액을 버리고나서, 세포 펠렛을 게놈 DNA 단리에 사용하였다.

게놈 DNA는 Ausubel의 상기 문헌 섹션 2.4.3에 기술된 CTAB 방법을 변형하여 사용하여 분리하였다. 이어서, "Large Scale CsCl prep of bacterial genomic DNA"라는 제목의 섹션을 단계 6을 통하여 실시하였다. 이 때, 추가의 클로로포름/이소아밀 알콜 (24:1) 추출을 실시한 후 페놀/클로로포름/이소아밀 (25:24:1) 추출 단계 및 최종적으로 클로로포름/이소아밀/알콜 (24:1) 추출을 실시하였다. 0.6 부피의 이소프로판올을 첨가하여 DNA를 침전시켰다. 침전된 DNA를 굵은 유리 막대 끝에 감아서 70 % 에탄올에 잠시 담가 최종적으로 세척하고, TE 완충액 3 ml에 용해시켰다.

280/260 nm에서의 광학 밀도에 의해 추정된 DNA 농도는 약 2 mg/ml이었다.

상기 게놈 DNA를 사용하여 라이브러리를 제조하였다. 게놈 DNA 약 50 µg을 Sau3 A1으로 부분 분해하였다. 이어서, NaCl 밀도 구배 원심분리를 사용하여 부분 분해 DNA 단편을 크기 분별하였다. 아가로스 겔 전기영동에 의해 측정된 평균 크기가 12 kb 이상인 DNA 단편을 포함하는 분획을 미국 캘리포니아주 라 줄라 소재의 Stratagene사의 플라스미드 Bluscript에 라이게이션시키고 E. coli DH5α 또는 DHB10 균주로 형질전환시켰다.

이와 별도로, 단백질의 정제 분취액을 단백질에 대한 단클론성 항체의 생성을 위해 미국 매디슨 소재의 위스콘신 대학교의 바이오테크놀로지 히브리도마 센터에 보냈다. 이 센터에 보낸 물질은 65 °C에서 변성된 천연 밴드 1 및 2를 포함하는 HPLC 정제 분획으로, 그 중 20 µg을 4 마리의 마우스에게 각각 주사하였다. 안정한 단클론성 항체 생성 히브리도마 세포주를 비면역된 마우스의 비장 세포를 안정한 골수종 세포주와 융합시킨 후 회수하였다. 단클론성 항체를 하이브리도마 세포로부터 회수하였다.

이와 별개로, 천연 아가로스 겔 정제 밴드 1 (실시에 1 참조) 단백질을 뉴질랜드 흰토끼를 면역시키는데 사용하여 폴리클로날 항체를 생성시켰다. 천연 아가로스 겔로부터 밴드를 절개하고 겔 조각을 65 °C로 가열하여 아가로스 용융시키고, 그 직후 보조제로 유화시켜 단백질을 제조하였다. 프로인트 (Freund) 완전 보조제를 1차 면역을 위해 사용하고, 프로인트 불완전 보조제를 사용하여 1 달 간격으로 추가로 3 회 주사하였다. 각각의 주사에 대해 단백질 50 내지 100 µg을 포함하는 유화 밴드 1 약 0.2 ml를 토끼의 등에 다중 피하주사하여 전달하였다. 최종 주사 10 일 후에 혈청을 얻고, 3 주 후에 다시 채혈을 실시하였다. 혈청 보충물을 56 °C까지 15 분 동안 가열하여 불활성화시킨 후 -20 °C에서 보관하였다.

이어서, 단클론성 항체 및 폴리클로날 항체를 사용하여 에피토프에 의해 검출될 수 있는 항원의 발현에 대하여 게놈 라이브러리를 스크리닝하였다. 양성 클론은 니트로셀룰로스 필터 콜로니 리프트 상에서 검출되었다. 양성 클론에 대한 면역블롯 분석을 실시하였다.

면역블롯 및 써던 분석에 의해 규명된 클론의 분석을 통하여 4 개의 계능 영역을 실험적으로 확인하였다.

제1 영역의 클론에는 TcbA_{ii}로 명명된 펩티드를 코딩하는 유전자가 존재하였다. 이 유전자 (TcbA)의 전체 DNA 서열을 얻고 서열 11로서 나타내었다. 이 서열이 서열 1의 내부 서열을 코딩한다는 것의 확인은 서열 11의 개방 해독 프레임에 의해 생성된 추정 아미노산 서열로부터 아미노산 번호 88에서 서열 1이 존재한다는 것에 의해 입증된다. 이것은 서열 11에 의해 생성되는 추정 아미노산 서열인 서열 12를 통하여 확인될 수 있다.

제2 영역의 독소 펩티드는 TcaB_i, TcaB_{ii} 및 TcaC로서 언급한 세그먼트를 포함한다. 폴리클로날 항형체를 사용하여 라이브리리를 스크리닝한 후에, 모두가 면역블롯 상에서 폴리클로날 항체와 교차 반응하고 또한 써던 블롯 상에서 DNA 상동성을 공유하는 것으로 밝혀진, 상이한 크기의 단백질을 생성하는 몇 개의 클론에 의해 상기 제2 영역의 독소 유전자를 확인하였다. 서열 비교를 통하여 이들은 상기 TcaB 및 TcaC로 명명된 유전자 복합체에 속한다는 것을 알았다.

2 개의 다른 영역의 항체 독소 클론을 폴리클로날 스크린에서 단리하였다. 이들 영역은 폴리클로날 항체와 교차반응하는 단백질을 생성하고, 써던 블로팅에 의해 측정된 영역과 DNA 상동성을 공유하였다. 따라서, 포토랩두스 루미네센스 세포외 단백질 유전자는 진화상 관련된 일군의 유전지를 나타내는 것으로 보인다.

이러한 생물체 내부에 포함된 독소 펩티드에 진화상 관련된 변이가 존재할 수도 있다는 개념을 추가로 확인하기 위해, 관련 단백질의 존재 여부에 대해 포토랩두스 루미네센스의 다른 균주를 검사하는 2 가지의 방법을 실시하였다. 이것은 계능 DNA의 PCR 증폭 및 폴리클로날 및 단클론성 항체를 사용하는 면역 블롯 분석 모두에 의해 실시되었다.

이 결과는 관련 단백질이 포토랩두스 루미네센스 균주 WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-11, WX-12, WX-15 및 WX-14에 의해 생성된다는 것을 나타낸다.

<실시예 11b>

tcc 독소 클론의 서열화 및 분석

실시예 11a에서 기술한 E. coli 클론에서 단리한 플라스미드에 대해 다른 DNA 서열화를 실시하였다. 상기 E. coli 클론의 제3 영역으로부터의 뉴클레오티드 서열은 상기 계능 로커스에서 3 개의 밀접하게 연관된 개방 해독 프레임인 것으로 밝혀졌다. 이 로커스는 tccA (서열 56), tccB (서열 58) 및 tccC (서열 60)로 명명된 3 개의 개방 해독 프레임을 갖는 tcc로 명명되었다. 이러한 개방 해독 프레임들 간의 유사한 연관성은 93 bp가 tccb (염기 2992-2994, 서열 56)의 개시 코돈으로부터 tccA의 정지 코돈을 분리하는 서열 56을 시험하거나, 131 염기가 tccB 및 tccC (염기 4930-4932, 서열 58)의 정지 코돈을 분리하는 서열 58을 시험하여 확인한다. 천연 지도가 도 6B에 나타나 있다.

tccA 개방 해독 프레임으로부터 추정된 아미노산 서열은 105,459 Da의 단백질을 코딩하는 유전자를 나타낸다. 이 단백질을 TccA (서열 57)로서 명명하였다. 이 단백질의 처음 12 개의 아미노산은 독소 복합체의 일부로서 앞에서 확인된 108 kDa 단백질 (서열 8)로부터 얻은 N-말단 서열과 일치한다.

tccB 개방 해독 프레임으로부터 추정된 아미노산은 175,716 Da의 단백질을 코딩하는 유전자를 나타낸다. 이 단백질은 TccB (서열 59)로 명명되었다. 이 단백질의 처음 11 개의 아미노산은 추정 분자량이 185 kDa인 단백질 (서열 7)로부터 얻은 N-말단 서열과 일치한다. 유사한 분석 결과, TccB 단백질은 TcbA (서열 12, 37 % 유사 및 28 % 동일), TcdA (서열 47, 35 % 유사 및 28 % 동일) 및 TcaB (서열 26, 32 % 유사 및 26 % 동일)로서 확인된 단백질과 관련된다는 것을 나타내었다 (위스콘신주 매디슨 소재의 Genetics Computer Group (GCG)의 GAP 알고리즘 위스콘신 패키지 버전 9.0 사용).

tccC의 추정 아미노산 서열은 이 개방 해독 프레임이 111,694 Da의 단백질을 코딩하는 것을 나타내고 생성 단백질은 TccC (서열 61)로 명명되었다.

<실시예 12>

포토랩두스 균주의 특성화

본 명세서에 기재된 컬렉션이 포토랩두스 균주로 이루어졌음을 확인하기 위하여, 본 발명의 균주를 포토랩두스에 특징적이고 다른 엔테로박테리아세 (Enterobacteriaceae) 및 제노랍두스 (Xenorhabdus) 종과 구별하게 하는, 식별되는 미생물학적 특징의 측면에서 평가하였다 [Farmer, J.J. 1984. Bergey's Manual of Systemic Bacteriology, vol 1. pp. 510-511. (ed. Kreig N.R. and Holt, J.G.). Williams & Wilkins, Baltimore.; Akhurst and Boemare, 1988, Boemare et al., 1993). 이러한 특유의 특징으로는 그람 음성 간균 균주, 생물체 크기는 폭 0.5 내지 2 μm, 길이 2 내지 10 μm, 적색/황색 콜로니 착색, 결정 봉입체의 존재, 카탈라제의 존재, 질산염 환원능, 생체 발광성 존재, 성장 배지로부터 영양의 흡수능 보유, 양성의 프로테아제 생성, 성장 온도 범위 37 °C 미만, 혐기성 조건 하에 생존 및 활발한 운동성 보유 등이 있다 (표 20). 참고용의 Escherichia coli, 제노랍두스 및 포토랩두스 균주가 비교용으로 모든 시험에 포함되었다. 전체적인 결과는 엔테로박테리아세 과 및 포토랩두스 속의 일부인 모든 균주와 일치하였다.

발광계를 사용하여 각 균주의 생체 발광을 확인하고, 광 생성을 정량하고 상대적으로 측정하였다. 상대적인 발광 단위 측정을 위해, 각 균주 (세포 및 배지)의 브로스를 액체 배양물에 접종한 후 세 간격 (6, 12 및 24 시간)에서 측정하여 배경 발광 (비접종 배지 및 물)과 비교하였다. 각종 브로스로부터의 발광을 측정하기 전에, 시퍼 (sipper) 세포를 사용하여 Gilford Systems (오하이

오주 소재 Oberlin 제품) 분광계 중에서 흡광도 (560 nm)를 측정하여 세포 밀도를 확인하였다. 이어서, 발광도를 측정하기 전에 (광학 밀도를 1.0 단위로 표준화하기 위해) 적합한 희석액을 제조하였다. 계속하여, 희석 브로쓰의 분취액을 큐벳에 넣고 (각각 300 μ l) Bio-Orbit 1251 발광계 (핀란드 트위쿠 소재의 Bio-Orbit Oy 제품)로 판독하였다. 각 시료에 대한 총 시간은 45 초이었다. 시료를 연속적으로 혼합하면서 (배플드 큐벳에서 회전시킴) 산소 이용 가능성을 제공하기 위하여 판독하였다. 양성 시험은 배경 발광 (약 5 내지 10 단위)의 5 배 이상인 것으로 정하였다. 또한, 콜로니 발광도는 암실에 유지시킨 후 사진 필름 오버레이를 사용하여 가시적으로 측정하였다. 각 균주의 그람 염색 특성은 시판되는 그람 염색 조절 슬라이드 (펜실바니아주 피츠버그 소재의 Fisher Scientific 제품)와 함께 사용되는 그람 염색 키트 (메릴랜드 주의 콕스빌 소재의 BBL 제품)로 확인하였다. 이어서, Zeiss 현미경 (독일 Carl Zeiss 제품) 100X 오일 침지 대물렌즈 (10X 접안렌즈 및 2X 체 배율을 가짐)를 사용하여 현미경 평가를 실시하였다. 생물체 크기, 세포 내부 및 봉입체 (로그상 성장 후)에 대한 각 균주의 현미경 검사는 오일 침지 및 상 대조 현미경을 갖는 습식 마운트 슬라이드 (10X 접안렌즈, 2X 체 배율, 40X 대물렌즈)를 사용하여 실시하였다 [Akhurst, R.J. and Boemare, N.E. 1990. Entomopathogenic Nematodes in Biological Control (ed. Gaugler, R. and Kaya, H.). pp. 75-90. CRC Press, Boca Raton, USA.; Baghdiguian S., Boyer-Giglio M.H., Thaler, J.O., Bonnot G., Boemare N. 1993. Biol. Cell 79, 177-185]. 콜로니 착색은 제조자의 지시에 따라 제조된 Bacto 영양 아가 (미시간주 디트로이트 소재의 Difco Laboratories 제품) 상에 접종한 후에 관찰하였다. 28 °C에서 배양하였고, 5 내지 7 일 후에 발생하였다. 효소 카탈라제의 존재를 시험하기 위해서, 시험 생물체의 콜로니를 영양 아가 플레이트로부터 작은 플러그 상에서 제거하고, 유리 시험관의 바닥에 넣었다. 가정용 과산화수소 용액 1 ml를 부드럽게 시험관의 측면을 따라 첨가하였다. 기체 (산소로 추정됨)의 기포가 즉시 또는 5 초 내에 발생할 때 양성 반응을 기록하였다. 비점종 영양 아가 및 과산화수소 용액의 대조군도 조사하였다. 질산염 환원을 시험하기 위해서, 각 배양물을 Bacto 질산염 브로쓰 (미시간주 디트로이트 소재의 Difco Laboratories 제품) 10 ml에 접종하였다. 28 °C에서 24 시간 배양한 후, 솔 파닐산 시약 2 액적 및 알파-나프틸아민 시약 2 액적 (Difco Manual, 10th edition, 미시간주 디트로이트 소재의 Difco Laboratories, 1984 참조)을 가하여 아질산염 생성을 시험하였다. 특유의 핑크 또는 적색의 생성은 질산염으로부터 아질산염이 형성되었음을 나타낸다. 각 균주가 성장 배지로부터 영양을 흡수하는 능력은 영양 중성 레드를 포함하는 Bacto MacConkey 아가; 영양 브로모 티몰 블루를 포함하는 Bacto Tergitol-7 아가 및 영양 에오신-Y를 포함하는 Bacto EMB 아가 (미시간주 디트로이트 소재의 Difco Laboratories 제품, 모두 제조자의 지시에 따라 제조함)를 사용하여 시험하였다. 상기 배지 상에서 접종한 후, 28 °C에서 5 일 동안 배양하여 영양 흡수를 기록하였다. 이들 배지 상에서의 성장은 엔테로박테리아세과의 특징이다. 각 균주의 운동성은 제조자의 지시에 따라 제조한 Bacto Motility 시험 배지 (미시간주 디트로이트 소재의 Difco Laboratories 제품)의 용액을 사용하여 시험하였다. 버트-스탑 (butt-stab) 접종은 각 균주를 사용하여 실시하고, 운동성은 접종 라인으로부터 퍼지는 성장 확산 영역에 의해 현미경으로 판단하였다. 많은 경우에, 운동성은 습식 마운트 슬라이드 하에 액체 배양물로부터 현미경으로 관찰하였다. 각 균주에 대한 생화학적 영양 평가를 BBL Enterotube II (독일 디킨슨 소재의 Benton 제품)를 사용하여 실시하였다. 제조자 지시에 따르되, 배양은 28 °C에서 5 일 동안 실시하였다. 그 결과는 앞에서 인용한 포토랩두스에 대한 기록과 일치하였다. 프로테아제 생성은 Bacto 젤라틴 (미시간주 디트로이트 소재의 Difco Laboratories 제품) 플레이트를 제조자의 지시에 따라 제조하여 젤라틴의 가수분해를 함으로써 시험하였다. 배양물을 접종하고 플레이트를 28 °C에서 5 일 동안 배양하였다. 상이한 온도에서의 성장을 평가하기 위해서, 아가 플레이트 [탈이온수 중의 2 % Bacto-Agar를 갖는 2 % 프로테오스 펩톤 3번 (미시간주 디트로이트 소재의 Difco 제품)를 통상의 접종물 공급원으로 스트리킹하였다. 플레이트를 Nesco (등록 상표) 필름으로 밀봉하고, 3 주 이하 동안 20, 28 및 37 °C에서 배양하였다. 37 °C에서 성장을 보이지 않는 플레이트는 1 주 동안 28 °C 배양기에 이송한 후 세포 생존성을 보이지 않았다. 포토랩두스 균주에 대한 산소 요구량을 다음과 같이 시험하였다. 유체 티오글리콜레이트 브로쓰 배지 (미시간주 디트로이트 소재의 Difco 제품)로 버트-스탑 접종을 실시하였다. 튜브를 실온에서 1 주 동안 배양한 후, 배양물을 종류 및 성장 정도에 대해 조사하였다. 인디케이터 레사주린은 배지 산화 수준 또는 호기성 생활 영역을 나타낸다 (Difco Manual, 10th edition, Difco Laboratories, 미시간주 디트로이트 소재). 시험된 포토랩두스 균주에 대해 얻은 성장 영역 결과는 통성 혐기성 미생물에 대한 결과와 일치하였다.

[표 19]

포토랍두스 균주의 분류학적 특징

(조사원 특성*)

| 균주 | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|-------|---|---|---|------|---|---|---|---|---|----|---|---|---|---|---|---|---|
| W-14 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| WX-1 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| WX-2 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| WX-3 | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| WX-4 | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| WX-5 | - | ± | ± | rd S | ± | ± | ± | ± | ± | LO | ± | ± | ± | ± | ± | ± | ± |
| WX-6 | - | ± | ± | rd S | ± | ± | ± | ± | ± | LY | ± | ± | ± | ± | ± | ± | ± |
| WX-7 | - | ± | ± | rd S | ± | ± | ± | ± | ± | R | ± | ± | ± | ± | ± | ± | ± |
| WX-8 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| WX-9 | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| WX-10 | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| WX-11 | - | ± | ± | rd S | ± | ± | ± | ± | ± | RO | ± | ± | ± | ± | ± | ± | ± |
| WX-12 | - | ± | ± | rd S | ± | ± | ± | ± | ± | RO | ± | ± | ± | ± | ± | ± | ± |
| WX-14 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| WX-15 | - | ± | ± | rd S | ± | ± | ± | ± | ± | LR | ± | ± | ± | ± | ± | ± | ± |
| H9 | - | ± | ± | rd S | ± | ± | ± | ± | ± | LY | ± | ± | ± | ± | ± | ± | ± |
| H5 | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| Hm | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| HP88 | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| NC-1 | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| W30 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| WIR | - | ± | ± | rd S | ± | ± | ± | ± | ± | YI | ± | ± | ± | ± | ± | ± | ± |
| B2 | - | ± | ± | rd S | ± | ± | ± | ± | ± | RO | ± | ± | ± | ± | ± | ± | ± |
| 43948 | - | ± | ± | rd S | ± | ± | ± | ± | ± | R | ± | ± | ± | ± | ± | ± | ± |
| 43949 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| 43950 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| 43951 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |
| 43952 | - | ± | ± | rd S | ± | ± | ± | ± | ± | Q | ± | ± | ± | ± | ± | ± | ± |

* - A = 그람 균주, B = 결정 봉입체, C = 생체 발광, D = 세포 형태, E = 운동성, F = 질산염 환원, G = 카탈라제의 존재, H = 젤라틴 가수분해, I = 염료 흡수, J = 착색, K = EMB 아가 상에서의 성장, L = MacConkey 아가 상에서의 성장, M = Tergitol-7 아가 상에서의 성장, N = 통성 혐기성균, O = 20 °C에서의 성장, P = 28 °C에서의 성장, Q = 37 °C에서의 성장, + - +/- = 특징의 양성 또는 음성, rd = 간균, S = 속 기술자 범위 내의 크기, RO = 레드-오렌지색, LR = 담적색, R = 적색, O = 오렌지색, Y = 황색, T = 황갈색, LY = 담황색, YI = 담황갈색, 및 LO = 담오렌지색.

세포 지방산 분석은 속 및 종 수준에서 박테리아의 특성을 파악하기 위한 식별 수단이고 (Tornabene, T.G. 1985. Lipid Analysis and the Relationship to Chemotaxonomy in Methods in Microbiology, Vol 18, 209-234.; Goodfellow, M. and O'Donnell, A.G. 1993. Roots of Bacterial Systematics in Handbook of New Bacterial Systematics (ed. Goodfellow, M. & O'Donnell, A.G.) pp.3-54. London: Academic Press Ltd., 이들 참고 문헌의 개시 내용은 본 명세서에 참고로 포함됨), 본 발명의 컬렉션이 속 수준에서 관련되었음을 확인하기 위해 사용하였다. Microbial ID (MIDI 제품, 미국 델라웨어주 뉴와크 소재) Microbial Identification System (MIS)을 사용하는 지방산 메틸 에스테르 분석 (FAME)을 위해 배양물을 외부의 계약 실험실에 보냈다. MIS 시스템은 25 mm × 0.2 mm 5 % 메틸페닐 실리콘 융합 실리카 모세관 컬럼을 갖는 Hewlett Packard HP5890A 기체 크로마토그래피로 구성된다. 수소를 운반 기체로 사용하고 화염-이온화 검출기가 자동화 샘플러, 적분기 및 컴퓨터와 함께 기능한다. 컴퓨터는 시료의 지방산 메틸 에스테르를 미생물 지방산 라이브러리 및 공지의 지방산 검정 보정 혼합물과 비교한다. 계약 실험실에서 선별한 균주를 분석 전에 트리피카제 대두 아가 상에서 28 °C에서 24 시간 동안 성장시켰다. 시료 추출은 표준 FAME 방법에 의해 계약 실험실에서 실시하였다. 포토랍두스 외의 다른 발광 박테리아 균에 대한 균주의 직접적인 확인은 없었다. 1균의 단리물의 지방산 프로파일을 비교하는 클러스터 분석을 실시할 때, 균주 지방산 프로파일은 속 수준에서 관련되었다.

본 발명의 컬렉션에서의 포토랍두스 균주의 진화적 다양성을 각 균주의 게놈 DNA를 사용하여 PCR (폴리머라제 연쇄 반응) 매개 게놈 지문 분석에 의해 측정하였다. 이 기술은 다양한 박테리아 종의 게놈 전반에 존재하는 일군의 반복 DNA 서열을 기초로 한 것이다 (Versalovic, J., Schneider, M., DE Bruijn, F.J. and Lupski, J.R. 1994. Methods Mol. Cell. Biol., 5, 25-40 참고). 이들 서열 중 반복 유전자외 회문 서열 (REP), 장내 박테리아 반복 유전자내 공통 서열 (ERIC) 및 BOX 성분은 박테리아 게놈의 조직에 중요한 역할을 수행하는 것으로 생각된다. 게놈 조직은 선별에 의해 향상되는 것으로 생각되고, 밀접하게 관련된 박테리아 균주의 게놈 내의 이들 성분의 상이한 분산은 이들 균주의 식별에 사용될 수 있다 (예를 들면, Louws, F.J., Fulbright, D.W., Stephens, C.T. and DE Bruijn, F.J. 1994. Appl. Environ. Micro. 60, 2286-2295). Rep-PCR은 반복 서열 사이에 존재하는 크기가 상이한 DNA 단편을 증폭하기 위해 상기 반복 서열에 상보성인 올리고뉴클레오타이드 프라이머를 이용한다. 생성물은 각 균주에 대한 DNA "지문"을 확립하기 위해 전기영동에 의해 분리한다.

본 발명의 균주로부터 게놈 DNA를 단리하기 위해서, 세포 펠렛을 최종 부피가 10 ml가 되도록 TE 완충액 (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) 중에 재현탁시키고나서, 5 M NaCl 12 ml를 가하였다. 이 혼합물을 15,000 × g에서 20 분 동안 원심분리하였다. 생성 펠렛을 TE 5.7 ml 및 10 % SDS 300 μl 중에 재현탁시키고 20 mg/ml 프로테이나제 K 60 μl (Gibco BRL 제품, 뉴욕주 그랜드 아일랜드 소재)을 가하였다. 이 혼합물을 37 °C에서 1 시간 동안 배양한 후, 리소자임 약 10 mg을 가하고, 혼합물을 추가로 45 분 동안 배양하였다. 계속하여, 5M NaCl 1 ml 및 CTAB/NaCl 용액 800 μl (10 % w/v CTAB, 0.7 M NaCl)을 가하고, 혼합물을 65 °C에서 10 분 동안 배양하고, 부드럽

게 진탕시킨 후 배양하고, 추가로 20 분 동안 진탕시켜 세포 물질의 세정을 보조하였다. 동일 부피의 클로로포름/이소아밀 알콜 용액 (24:1, v/v)을 가하고, 부드럽게 혼합한 후 원심분리하였다. 이어서, 동일 부피의 페놀/클로로포름/이소아밀 알콜 (50:49:1)을 사용하여 2 회 추출하였다. 게놈 DNA를 이소프로판올 0.6 부피를 사용하여 침전시켰다. 침전된 DNA를 유리 막대기로 수거하여 70 % 에탄올로 2 회 세척하고, 건조시키고, STE (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 1 mM EDTA) 2 ml에 용해시켰다. 이어서, 260 nm에서의 광학 밀도에 의해 DNA를 정량하였다. 포도랍두스 게놈 DNA의 rep-PCR 분석을 실시하기 위해서, 다음과 같은 프라이머를 사용하였다: REP1R-1; 5'-1111CGICGICATCIGGC-3' 및 REP2-1; 5'-ICGICTTATCIGGCCTAC-3'. 다음과 같은 25 μ l 반응물을 사용하여 PCR을 실시하였다: 7.75 μ l H₂O, 2.5 μ l 10X LA 완충액 (PanVera Corp. 제품, 위스콘신주 매디슨 소재), 16 μ l dNTP 혼합물 (각각 2.5 mM), 각각의 프라이머 1 μ l (50 pM/ μ l), DMSO 1 μ l, 게놈 DNA 1.5 μ l (농도 0.075 내지 0.480 μ g/ μ l) 및 0.25 μ l TaKaRa EX Taq (PanVera Corp. 제품, 위스콘신 매디슨 소재). Perkin Elmer DNA 열 순환기 (Norwalk, CT) 중에서 다음과 같은 조건을 사용하여 PCR 증폭을 실시하였다: 95 °C/7 분, 이어서 94 °C/1 분, 44 °C/1 분, 65 °C/8 분의 순환 35회 반복, 이어서 65 °C에서 15 분. 순환 후, 반응물 25 μ l를 6X 겔 로딩 완충액 (물 중의 0.25 % 브로모페놀 블루, 40 % w/v 수크로스) 5 μ l에 가하였다. 이어서, 15 × 20 cm 1 % 아가로스 겔을 각각의 반응액 8 μ l를 사용하여 TBE 완충액 (0.09 M Tris-보레이트, 0.002 M EDTA) 중에서 전개시켰다. 겔을 45 V에서 약 16 시간 동안 전개시켰다. 이어서, 겔을 에티뮴 브로마이드 20 μ g/ml 중에서 1 시간 동안 염색시키고, TBE 완충액 중에서 약 3 시간 동안 탈염색시켰다. 겔의 폴라로이드 사진을 UV 조명 하에 인화하였다.

각 염색에 특이적인 크기에서의 밴드의 존재 여부를 사진으로부터 기록하고, 수치 분류 소프트웨어 프로그램인 NTSYS-pc (Exeter Software, 뉴욕주 세타우켓 소재)에 유사 매트릭스로서 입력하였다. 동시에 분석되는 E. coli 균주 HB101 및 잔토모나스 오리재 피브이, 오리재 (Xanthomonas oryzae pv. oryzae)의 대조군은 출판된 문헌에 따른 "지문"을 생성시켰다 (Versalovic, J., Koeuth, T. and Lupski, J.R. 1991. Nucleic Acids Res. 19, 6823-6831; Vera Cruz, C.M., Haldal-Alija, L., Louws, F., Skinner, D.Z., George, M.L., Nelson, R.J., DE Bruijn, F.J., Rice, C. and Leach, J.E. 1995. Int. Rice Res. Notes, 20, 23-24.; Vera Cruz, C.M., Ardales, E.Y., Skinner, D.Z., Talag, J., Nelson, R.J., Louws, F.J., Leung, H., Mew, T.W. and Leach, J.E. 1996. Phytopathology). 이어서, NTSYS-pc 내의 일련의 프로그램; 매트릭스의 유사성 계수 (Jaccard 계수 사용)를 생성하기 위한 SIMQUAL (정량 데이터에 대한 유사성) 및 관련 균주를 한데 모으고 페노그램 (phenogram) (도 5)으로 표현될 수 있는 SAHN (연속적인, 응집성의, 하이라키칼 (Heirarchical) 및 네스티드 (Nested)) 클러스터링 [UPGMA (산술평균을 사용하는 비계량 페어-그룹 (Unweighted Pair-Group) 방법을 사용)]을 사용하여 포도랍두스 균주로부터의 데이터를 분석하였다. COPH (코페네틱 (cophenetic) 값) 및 MXCOMP (매트릭스 비교) 프로그램을 사용하여 코페네틱 값 매트릭스를 생성시키고, 이것과 클러스터링의 기초가 되는 본래의 매트릭스 사이의 관계를 비교하였다. 클러스터 분석에 적합한 양호함의 척도인 (r = 0.8 내지 0.9가 매우 양호하게 적합함을 나타냄) 생성되는 정상화 만델 (Mantel) 통계치 (r)을 생성시켰다. 본 발명의 경우에 r은 0.919 이었다. 따라서, 본 발명의 컬렉션은 포도랍두스 속의 대표적인 다양한 균의 용이하게 식별가능한 균주로 이루어진 것이다.

<실시예 13>

다양한 포도랍두스 균주에 의해 생산되는 독소(들)의 살충 효용

카팍으로 덮여진 델롱 목을 가진 500 ml 트리배플드 플라스크중의, 2% 프로테오스 펩톤 #3 (PP3, 미시간주 디트로이트 소재의 디프코 래버러토리즈 (Difco Laboratories) 제조) 액체 배지 175 ml에 일차 변종 서브클론을 접종하여, 다양한 포도랍두스 균주의 초기 종자 배양물을 제조하였다. 각 종자 배양을 위한 접종물은 오일이 덮여진 아가 사면 배양물 또는 플레이트 배양물로부터 유래되었다. 접종 후, 상기 플라스크를 회전식 진탕기 상에서, 150 rpm에서 16 시간 동안 28°C에서 배양하였다. 이어서 상기 종자 배양물은 각 균주의 정해진 발효를 위한 균일한 접종원으로서 사용되었다. 또한, 로그 후상 (post-log) 종자 배양물을 멸균의 미네랄 오일로 덮고, 멸균의 자석 교반 막대를 장래의 재현택을 위하여 첨가하며, 배양물을 암실, 실온에서 보관함으로써, 접종물을 독소 생산 상태로 장기간 보존하였다. 활발하게 성장하는 종자 배양물 1% (예를 들어, 신선 배양물 175 ml 당 1.75 ml)를 2% PP3 신선 배지에 첨가함으로써 생산 브로쓰를 접종하였다. 브로쓰는, 실리콘 포음 마개로 밀폐된 500 ml 트리배플드 플라스크 (상기 참조) 또는 2800 ml 배플드 요철면 바닥 플라스크 (500 ml 용량)에서 생산하였다. 생산 플라스크를 상기 조건 하에서 24 내지 48시간 동안 배양하였다. 배양 후, 브로쓰를 멸균의 1 L 폴리에틸렌 병 내로 분배시켜 10°C에서 1 시간 동안 2600 × g에서 회전시킴으로써, 세포 및 파쇄물 펠렛으로부터 분리시켰다. 액체 브로쓰는 그 후 왓만(Whatman) GF/D (2.7 μ m 보유) 및 GF/B (1.0 μ m 보유) 유리 여과기를 통하여 신포 여과시켜 파쇄물을 제거하였다. 0.5 μ m 개방 및 닫힌 여과기를 사용, 접선 유출 극소 여과 장치 (매사추세츠주 노스보로 소재의 폴 필터론 (Pall Filtron) 제조)로 브로쓰를 보다 더 정화하였다. 필요시, 브로쓰를 냉각시켜(4°C까지) 2600 × g에서 여러 시간 동안 원심분리시킴으로써 추가로 정화시킬 수 있다. 상기 과정 후, 0.2 μ m 니트로셀룰로스 막 여과기를 사용하여 브로쓰를 여과 살균시켰다. 무균 브로쓰는 그 후, 생물학적 분석, 생화학적 분석용으로 직접 사용되었거나 분자량 10000 컷오프 M12 환외여과 장치 (매사추세츠주 베벌리 소재 아미콘 (Amicon) 제조) 또는 분자량 10000의 구멍 크기를 갖는 원심 분리 농축기 (매사추세츠주 베드포드 소재 밀리포어 (Millipore) 제조 및 매사추세츠주 노스보로 소재 폴 필터론 제조)를 사용하여 농축 (15배까지)시켰다. 원심 분리 농축기의 경우, 브로쓰를 2000 × g에서 약 2시간 동안 원심분리시켰다. 분자량 10000 투과액을, 상응하는 잔류물에 첨가하여 분자량 10000 초과 성분의 원하는 농축액을 수득하였다. 모래를 채운 히트 블록 (heat block)에서 10분간, 100°C에서 샘플을 가열함으로써, 처리되는 브로쓰 샘플을 가열 불활성화시켰다.

서로 다른 포도랍두스 균주로부터의 브로쓰(들) 및 독소 복합체(들)은 곤충 집단을 감소시키는 데

유용하며, 기술된 활성물질은, 곤충 불활성화에 효과적인 양만큼 곤충 서식지에 사용하는 것으로 이루어지는 곤충 집단 억제 방법에 사용되었다. 상기와 같이 발효시킨 포토랍두스 균주의 선별군의 브로쓰로부터 관찰된 살충 활성의 범위의 증거는 표 20에 나타나 있다. 브로쓰 농도를 증가시키거나 다른 발효 방법을 사용함으로써 상기 균주에서 부가적인 살충 활성의 검출이 가능하다. 단백질과 관계된 활성과 일관되게, 시험된 모든 균주들의 살충 활성은 열에 불안정하였다 (상기 참고).

다양한 포토랍두스 균주로부터의 배양 브로쓰(들)은 곤충 수에 대하여 차별적인 살충 활성 (치사율 및(또는) 성장 저해율, 성체 출현의 감소)을 나타낸다. 보다 구체적으로, 활성은 딱정벌레 (Coleoptera) 곤충 목의 일원인 옥수수 뿌리벌레 유충 및 목화 바구미 유충에 대하여 나타난다. 딱정벌레 목의 다른 일원으로는 방아벌레, 화분 딱정벌레, 뽕벼룩 갑충, 종자 갑충 및 콜로라도 감자 딱정벌레가 있다. 활성은 별 매미충 및 옥수수 열구에 대해서도 관찰되는데, 이들은 동시목의 일원이다. 동시목의 다른 일원으로는 열구, 배나무이, 사과나무이, 개각충, 화이트플라이, 침벌레 및 수많은 숙주 특이 진딧물 균주들이 있다. 브로쓰 및 정제된 독소 복합체(들)은 또한 나비목의 일원인 회색담배나방, 박각시나방 및 유럽 옥수수 천공충에 대해서도 활성을 갖는다. 상기 목의 다른 전형적인 일원으로는 사탕무 행렬구더기, 양배추 자벌레, 블랙 야도충, 옥수수 귀벌레, 사과 나방, 반대종, 화랑곡나방, 잎말이나방, 배추 벌레, 목화다래나방, 동홍이벌레, 이스턴 텐트 나방, 잔디 포충나방 및 풀 행렬구더기가 있다. 활성은 파리목의 일원인 과일파리 및 모기 유충에 대해서도 나타난다. 파리목의 다른 일원으로는 완두콩 벌레, 당근 파리, 양배추 뿌리 파리, 순무 뿌리 파리, 양파 파리, 학파리 및 집파리와 다양한 모기 종들이 있다. 브로쓰(들) 및 독소 복합체(들)은 또한, 딸기 개미 진드기, 브로드 진드기, 감귤류 적색 진드기, 유럽형 적색 진드기, 배 녹빛 진드기 및 토마토 적갈색 진드기를 포함하는 응애목의 일원인 점박이 거미 진드기에 대해 활성을 나타낸다.

옥수수 뿌리벌레 유충에 대한 활성은 하기와 같이 시험되었다. 포토랍두스 배양 브로쓰(들)(0 내지 15배 농축, 여과 살균), 2% 프로테오스 펩톤 #3, 정제 독소 복합체(들), pH 7.0의 10 mM 인산 나트륨 완충액을 인공 먹이의 표면 (약 1.5 cm²)에 40 μl씩 직접 도포하였다[Rose, R. I. 및 McCabe, J. M. (1973). J. Econ. Entomol., 66, (389-400) 참조]. 독소 복합체를 pH 7.0의 10 mM 인산 나트륨 완충액에 희석시켰다. 먹이 플레이트를 무균 플로우 후드에서 공기 건조시키고, 표면 살균된 알로부터 부화시킨 단일, 신생 디아브로티카 운대심프크타타 호와르디 (Diabrotica undecimpunctata howardi) (써던 옥수수 뿌리벌레, SCR)로 웰을 만연시켰다. 상기 플레이트를 봉인하여 습윤 성장 챔버에 두고 적당한 기간 (3 내지 5일)동안 27°C에서 유지시켰다. 그 후 치사율 및 유충 무게 측정치를 기록하였다. 일반적으로, 모든 연구에서 처리당 16마리의 곤충이 사용되었다. 대조군 치사율은 일반적으로 5% 미만이었다.

목화 바구미 (Anthonomus grandis)에 대한 활성은 하기와 같이 시험되었다. 농축 (1 내지 10배) 포토랍두스 브로쓰, 대조군 배지 (2% 프로테오스 펩톤 #3), 정제 독소 복합체(들)(0.23 g/ml) 또는 pH 7.0의 10 mM 인산 나트륨 완충액을 0.35 g의 인공 먹이 (스톤빌 옐로우 나비목 (Stoneville Yellow lepidopteran) 먹이) 표면에 60 μl씩 도포시켜 건조시켰다. 12 내지 24시간 된 목화 바구미 유충 한마리를 먹이에 놓고, 웰을 봉인시켜 50% 상대 습도하 25°C에서 5일간 유지시켰다. 그 후, 치사율 및 유충 무게를 평가하였다. 대조군 치사율은 0 내지 13%였다.

모기 유충에 대한 활성은 하기와 같이 시험되었다. 분석은 96 웰 마이크로타이더 플레이트에서 행해졌다. 각 웰에는, 200 μl의 수용액 (10배 농축 포토랍두스 배양 브로쓰(들)), 대조군 배지 (2% 프로테오스 펩톤 #3), 10 mM 인산 나트륨 완충액, 독소 복합체(들)(0.23 mg/ml) 또는 물 및 약 20개의, 1일된 유충 (Aedes aegypti)이 포함되었다. 처리당 6웰을 사용하였다. 결과는 만연시킨 후 3 내지 4일에 해석하였다. 대조군 치사율은 0 내지 20%이었다.

과일파리에 대한 활성은 하기와 같이 시험하였다. 구입한 드로소필라 멜라노가스터 (Drosophila melanogaster) 배지는, 50% 건조 배지와, 물, 대조군 배지 (2% 프로테오스 펩톤 #3), 10배 농축 포토랍두스 배양 브로쓰(들), 정제 독소 복합체(들)(0.23 mg/ml) 또는 pH 7.0의 10 mM 인산 나트륨 완충액 중 어느 하나의 50% 액체를 사용하여 제조하였다. 이것은 처리당 각각 3개의 리어링 바이알 (rearing vial)에 4.0 ml의 건조 배지를 넣고 4.0 ml의 적절한 액체를 첨가함으로써 성취되었다. 그 후 각 25 ml 바이알에 10개의 후기 영 드로소필라 멜라노가스터 구더기를 첨가하였다. 상기 바이알을 형광등하, 상온에서 실험실 벤치상에 놓아 두었다. 노출 15일 후에, 새끼 또는 성충수를 계산하였다. 물 및 대조 배지와 비교해서 성충 출현은 감소하였다(0 내지 16% 감소).

다른 외부 접촉 없이 활성물을 섭취하도록 고안된 섭취 분석으로, 별 매미충 성충 (Macrosteles severini) 및 옥수수 식물벌레 애벌레 (Peregrinus maidis)에 대한 활성을 시험하였다. 활성물/먹이 용액의 저장기는 35 × 10 mm 페트리 디쉬의 바닥 중심에 2개의 흡을 뚫어 만든다. 정사각형의 2인치 파라필름 엠 (등록상표 Parafilm M)을 디쉬 상부를 가로질러 놓고 "0" 링으로 고정시킨다. 그 후, 1 온즈 플라스틱 컵에 약 7마리의 메뚜기를 넣고, 파라필름이 아래에 오도록 저장기를 컵의 상부에 놓는다. 이어서, 시험 용액을 흡을 통하여 저장기에 첨가한다. 10배 농축 포토랍두스 배양 브로쓰(들)를 사용하는 시험에서는, 브로쓰 및 대조군 배지 (2% 프로테오스 펩톤 #3)를 pH 7.0의 10 mM 인산 나트륨 완충액에 대하여 투석시켰으며, 그 결과로서 생기는 용액에 자당 (5%까지)을 첨가하여 대조 치사율을 감소시켰다. 정제 독소 복합체(들)(0.23 mg/ml) 또는 pH 7.0의 10 mM 인산 나트륨 완충액을 또한 시험하였다. 치사율은 3일째에 기록한다. 분석물은 16/8의 광주기로 상대 습도 70%하 28°C에서 배양기 내에 보존하였다. 분석물은 72시간째에 치사율을 기록하였다. 대조군 치사율은 6% 미만이었다.

나비목 유충에 대한 활성은 하기와 같이 시험하였다. 농축(10배) 포토랍두스 배양 브로쓰(들), 대조군 배지 (2% 프로테오스 펩톤 #3), 정제 독소 복합체(들)(0.23 mg/ml) 또는 pH 7.0의 10 mM 인산 나트륨 완충액을, 표준 인공 나비목 먹이(스톤빌 옐로우 먹이, 약 1.5 cm²) 표면에 40 μl씩

직접적 도포시켰다. 먹이 플레이트를 무균 플로우 후드내에서 공기 건조시켰고, 각 웰을 신생 유충 1마리로 만연시켰다. 유럽 옥수수 천공충 (*Ostrinia nubilalis*) 및 박각시나방 (*Manduca sexta*) 알을 구입하여 부화시켰고, 회색담배나방 (*Heliothis virescens*) 유충은 내부적으로 공급하였다. 유충으로 만연시킨 후, 먹이 플레이트를 봉인하여 습윤 성장 챔버내에 두고 적절한 기간 동안 27°C 암실에서 유지시켰다. 치사율 및 중량 측정은 5일째에 실시하였다. 일반적으로, 모든 연구에서 처리당 16마리의 곤충이 사용되었다. 대조군 배지의 대조군 치사율은 일반적으로 4 내지 약 12.5%였고, 인산 완충액의 대조군 치사율은 10% 미만이었다.

점박이 거미 진드기 (*Tetranychus urticae*)에 대한 활성은 하기와 같이 측정되었다. 어린 호박 식물을 단일 자엽으로 다듬어 10배 농축 브로쓰(들), 대조군 배지 (2% 프로테오스 펩톤 #3), 정제 독소 복합체(들), pH 7.0의 10 mM 인산 나트륨 완충액을 흘려내릴 정도로 분무시켰다. 건조 후, 식물을 거미 진드기의 혼합 집단으로 만연시키고 실험실 온도 및 습도에서 72시간동안 유지시켰다. 그 후, 대조군의 수준을 결정하기 위하여 살아있는 진드기의 수를 세었다.

[표 20]

서로 다른 포도랍두스 균주로부터의 브로쓰의 살충 범위 관측치

| 포도랍두스 종 | 감수성* 곤충 종 |
|--|----------------------------|
| WX-1 | 3**, 4, 5, 6, 7, 8 |
| WX-2 | 2, 4 |
| WX-3 | 1, 4 |
| WX-4 | 1, 4 |
| WX-5 | 4 |
| WX-6 | 4 |
| WX-7 | 3, 4, 5, 6, 7, 8 |
| WX-8 | 1, 2, 4 |
| WX-9 | 1, 2, 4 |
| WX-10 | 4 |
| WX-11 | 1, 2, 4 |
| WX-12 | 2, 4, 5, 6, 7, 8 |
| WX-14 | 1, 2, 4 |
| WX-15 | 1, 2, 4 |
| W30 | 3, 4, 5, 8 |
| NC-1 | 1, 2, 3, 4, 5, 6, 7, 8, 9. |
| WIR | 2, 3, 5, 6, 7, 8 |
| HP88 | 1, 3, 4, 5, 7, 8 |
| Hb | 3, 4, 5, 7, 8 |
| Hm | 1, 2, 3, 4, 5, 7, 8 |
| H9 | 1, 2, 3, 4, 5, 6, 7, 8 |
| W-14 | 1, 2, 3, 4, 5, 6, 7, 8, 10 |
| ATCC 43948 | 4 |
| ATCC 43949 | 4 |
| ATCC 43950 | 4 |
| ATCC 43951 | 4 |
| ATCC 43952 | 4 |
| * = 대조군과 비교하여 25% 이상의 치사율 및(또는) 성장 저해 | |
| ** = 1; 담배 모충, 2; 유럽형 옥수수 천공충, 3; 담배 호른웜, 4; 서던 옥수수 뿌리벌레, 5; 목화 바구미, 6; 모기, 7; 과일파리, 8; 애스터 열구, 9; 옥수수 식물메뚜기, 10; 점박이 거미 진드기 | |

실시에 14

비 W-14 포도랍두스 균주: 정제, 특성화 및 활성 범위

정제

하기와 같은 실험 방법은 W-14의 정제를 위해 개발된 것과 유사하며, 생물학적 정량 (실시에 13 참조)에서 결정된 바와 같이 써던 옥수수 뿌리벌레 (SCR)에 대하여 가장 큰 활성을 갖는 분획만을 정제하는 것을 기초로 하여 확립되었다. 일반적으로, 실시에 13에 기술된 바와 같이, 여과시킨 4 내지 20 L의 브로쓰를 수취하여, 아미콘 M-12 여과 장치에 부착되어 있는 아미콘 나선형 한외여과 카트리지 타입 S1Y100을 사용하여 농축시켰다. 잔류물에는 100 kDa을 초과하는 분자 크기로 이루어

어진 천연 단백질들을 함유하고 있었고, 투과액은 크기 100 kDa 미만의 천연 단백질을 함유하고 있었다. SCR에 대한 활성의 대부분은 100 kDa 잔류물에 포함되어 있었다. 그 후, 잔류물은 10 mM 인산 나트륨 완충액 (pH 7.0)으로 여과물의 A₂₈₀이 0.100 미만일 때까지 계속해서 다이아필터링 (diafiltering)시켰다. 다른 언급이 없는 한, 지금부터의 모든 처리는 10 mM 인산 나트륨 (pH 7.0)으로 한정된 완충액에서 수행하였다. 잔류물은 그 후, 약 0.20 L의 최종 부피까지 농축시켜 낱진 필터웨어 (등록상표 Nalgene Filterware)의 0.45 mm 무균 여과 단위를 사용하여 여과시켰다. 여과된 물질을, 퍼셉티브 바이오시스템 스프린트 (등록상표 PerSeptive Biosystem Sprint) HPLC 시스템을 사용하여 완충액으로 평형시킨 퍼셉티브 바이오시스템 포로스 (등록상표 PerSeptive Biosystem Poros) 50 HQ 강력 음이온 교환 매트릭스로 충전시킨 파마시아 (Pharmacia) HR16/10 컬럼에 분당 7.5 ml로 로딩시켰다. 로딩후, A₂₈₀이 0.100이 될 때까지, 컬럼을 완충액으로 세척하였다. 이어서 0.4 M NaCl을 갖는 완충액을 사용하여 단백질을 분당 2.5 ml로 20분간 전체 용량 50 ml에 대해 컬럼으로부터 용출 (elution)시켰다. 그 후 컬럼은 추가로 20분간 동일한 유출 속도로 1.0 M NaCl을 갖는 완충액을 사용하여 세척시켰다 (최종 용량 50 ml). 0.4 M 및 1.0 M NaCl로 용출시킨 단백질을 별개의 투석 백 (등록상표 스펙트라/포 (Spectra/Por) 막 MWC0:2000)에 넣고, 12 L 완충액에서 4°C에서 철야 투석시켰다. SCR에 대한 활성을 가지는 대부분은 0.4 M 분획에 포함되어 있었다. 세파로스 (Sephacrose) CL4B (파마시아 제품)로 선 충전시킨 파마시아 XK 26/100 컬럼에 0.4 M 분획 20 ml을, 분당 0.75 ml의 유출 속도를 사용하여 로딩시킴으로써 한 층 더 정제시켰다. 분획들은 A₂₈₀ 피크 프로파일을 기초로 하여 모아 밀리포아 울트라프리-15 (등록상표 Millipore Ultrafree-15) 원심 분리 여과 장치 바이오맥스-50K (Biomax-50K) NMWL 막을 사용하여 최종 용량 0.75 ml까지 농축시켰다. 단백질 농도는, 소과의 감마 글로불린을 표준물질로 하는 바이오라드 단백질 분석 키트 (Biorad Protein Assay Kit)를 사용하여 결정하였다.

특성화

SCR 독소 복합체의 천연 분자량은 완충액에 있는 세파로스 CL4B로 선 충전시킨 파마시아 HR 16/50을 사용하여 결정되었다. 그 후, 컬럼을 독소의 대략적인 천연 분자 크기를 계산하기 위하여 공지된 분자 크기의 단백질을 사용하여 캘리브레이션 (calibration)시켰다. 표 21에 나타낸 것과 같이, 독소 복합체의 분자 크기는 균주 Hb 경우의 777 kDa 내지 균주 WX-14 경우의 1900 kDa의 범위였다. 독소 복합체의 수율 또한 다양하였는데, 균주 WX-12 경우의 리터 당 0.8 mg 내지 균주 Hb 경우의 리터당 7.0 mg을 생산하였다.

독소 복합체에서 발견되는 단백질을 SDS-PAGE 분석을 사용하여 개개의 폴리펩티드 크기를 조사하였다. 일반적으로, 각 균주로부터의 20 mg의 독소 복합체 단백질을 2-15% 폴리아크릴아미드 겔 (인테그레이티드 세퍼레이션 시스템즈 (Integration Separation Systems)) 상에 로딩시켜, 바이오라드 SDS-PAGE 완충액에서 20 mA로 전기 영동시켰다. 전기 영동 완료 후, 바이오라드 쿠마지 블루 R-250 (Coomassie blue R-250, 메탄올: 아세트산: 물 (용량비 40: 10: 40) 혼합 용액에서 2% 임)에서 철야 착색시켰다. 그 후 겔을 메탄올: 아세트산: 물 (용량비 40: 10: 40)에서 탈색시켰다. 이어서, 겔을 물로 15분간 행구고 개인용 분자 역학 레이저 농도계 (등록상표 Molecular Dynamics Personal Laser Densitometer)를 사용하여 주사시켰다. 레인을 정량화하여 200 내지 45 kDa 범위의 바이오라드 고분자량 표준물에 비교하여 분자 크기를 추정하였다.

각 균주로부터의 SCR 독소 복합체를 구성하는 개개의 폴리펩티드들의 크기를 표 22에 나타내었다. 개개의 폴리펩티드들의 크기는, 균주 WX-1 경우의 230 kDa 내지 균주 WX-7 경우의 16 kDa 범위였다. 균주 Hb를 제외한 모든 균주는 160 내지 230 kDa, 100 내지 160 kDa 및 50 내지 80 kDa의 독소 복합체를 포함하는 폴리펩티드를 가지고 있었다. 상기 데이터는, 독소 복합체가 펩티드 조성 및 성분 면에서 균주에 따라 다를 수 있지만, 모든 경우에 있어서 독소의 특성이 대형 올리고머 단백질 복합체로 이루어짐을 나타낸다.

[표 21]

비 W-14 포토랍두스 균주로부터의 독소 복합체의 특성화

| 균주 | 천연 분자량 근사치 ^a | 활성 분획 수득량 (mg/L) ^b |
|---|-------------------------|-------------------------------|
| H9 | 972.000 | 1.8 |
| Hb | 777.000 | 7.0 |
| Hm | 1.400.000 | 1.1 |
| HP88 | 813.000 | 2.5 |
| NC1 | 1.092.000 | 3.3 |
| WIR | 979.000 | 1.0 |
| WX-1 | 973.000 | 0.8 |
| WX-2 | 951.000 | 2.2 |
| WX-7 | 1.000.000 | 1.5 |
| WX-12 | 898.000 | 0.4 |
| WX-14 | 1.900.000 | 1.9 |
| W-14 | 860.000 | 7.5 |
| a 세파로스 CL4B로 충전시킨 파마시아 HR 16/50 컬럼을 사용하여 결정된 천연 분자량 | | |
| b 배양 브로쓰로부터 회수된 독소 복합체의 양 | | |

활성 범위

표 23에 나타난 바와 같이, 균주 Hm 및 H9으로부터 정제된 독소 복합체들을 균주 W-14로부터의 독소 복합체와 비교하여 각종 곤충들에 대한 활성을 시험하였다. 분석은 실시예 13에 기술된 바와 같이 수행하였다. 세가지 모든 균주로부터의 독소 복합체는 담배 버드 웜, 유럽 옥수수 천공충, 씨던 옥수수 뿌리 벌레, 및 벌 매미충에 대하여 활성을 나타내었다. 또한, 균주 Hm 및 W-14로부터의 독소 복합체는 점박이 거미 진드기에 대한 활성을 나타내었다. 게다가, W-14로부터의 독소 복합체는 모기 유충에 대한 활성도 나타내었다. 상기 데이터는, 독소 복합체가 몇몇 곤충 중 사 이에서 유사한 활성을 가지고 있는 반면, 다른 곤충 종에 대한 차별적인 활성을 또한 나타낼 수 있음을 뜻한다.

[표 22]

비 W-14 포토랍두스로부터의 정제 독소 복합체에서의 펩티드 크기의 근사치 (kDa)

| H9 | Hb | Hm | HP88 | NC-1 | WIR | WX-1 | WX-2 | WX-7 | WX-12 | WX-14 | W-14 |
|-----|-----|-----|------|------|-----|------|------|------|-------|-------|------|
| 180 | 150 | 170 | 170 | 180 | 170 | 230 | 200 | 200 | 180 | 210 | 190 |
| 170 | 140 | 140 | 160 | 170 | 160 | 190 | 170 | 180 | 160 | 180 | 180 |
| 160 | 139 | 100 | 140 | 140 | 120 | 170 | 150 | 110 | 140 | 160 | 170 |
| 140 | 130 | 81 | 130 | 110 | 110 | 160 | 120 | 87 | 139 | 120 | 160 |
| 120 | 120 | 72 | 129 | 44 | 89 | 110 | 110 | 75 | 130 | 110 | 150 |
| 98 | 100 | 68 | 110 | 16 | 79 | 98 | 82 | 43 | 110 | 100 | 130 |
| 87 | 98 | 49 | 100 | | 74 | 76 | 64 | 33 | 92 | 95 | 120 |
| 84 | 88 | 46 | 86 | | 62 | 58 | 37 | 28 | 87 | 80 | 110 |
| 79 | 81 | 30 | 81 | | 51 | 53 | 30 | 26 | 80 | 69 | 93 |
| 72 | 75 | 22 | 77 | | 40 | 41 | | 23 | 73 | 49 | 90 |
| 68 | 69 | 20 | 73 | | 39 | 35 | | 22 | 59 | 41 | 77 |
| 60 | 60 | 19 | 60 | | 37 | 31 | | 21 | 56 | 33 | 69 |
| 57 | 57 | | 58 | | 33 | 28 | | 19 | 51 | | 65 |
| 52 | 54 | | 45 | | 30 | 24 | | 18 | 37 | | 63 |
| 46 | 49 | | 39 | | 28 | 22 | | 16 | 33 | | 60 |
| 40 | 44 | | 35 | | 27 | | | | 32 | | 51 |
| 37 | 39 | | | | 25 | | | | 26 | | 46 |
| | 37 | | | | 23 | | | | | | 40 |
| | 35 | | | | | | | | | | 39 |
| | | | | | | | | | | | 29 |

[표 23]

포도랍두스 균주들로부터의 정제 독소 복합체의 살충 범위 관측치

| 포도랍두스 균주 | 감수성* 곤충 종 |
|---|------------------------|
| Hm 독소 복합체 | 1**, 2, 3, 5, 6, 7, 8 |
| H9 독소 복합체 | 1, 2, 3, 6, 7, 8 |
| W-14 독소 복합체 | 1, 2, 3, 4, 5, 6, 7, 8 |
| * = >25% 치사율 또는 성장 저해율 | |
| * = >25% 치사율 또는 성장 저해율 | |
| ** = 1; 담배 버드 웜, 2; 유럽형 옥수수 천공충, 3; 서던 옥수수 뿌리벌레, 4; 모기, 5; 점박이 거미 진드기, 6; 애스터 열구, 7; 과일 파리, 8; 목화 바구미 | |

<실시에 15>

포도랍두스 단백질 독소 복합체의 하위 분획

포도랍두스 단백질 독소 복합체를 실시예 14에 기술된 바와 같이 분리시켰다. 다음, 독소 약 10 mg을 pH 7.0의 20 mM Tris-HCl로 평형시킨 모노큐 (MonoQ) 5/5 컬럼에 1 ml/분의 유출 속도로 로딩시켰다. 280 nm에서의 광학상의 농도가 기준선 흡광도로 되돌아갈 때까지 컬럼을 pH 7.0의 20 mM Tris-HCl로 세척시켰다. 컬럼에 부착되어 있는 단백질은, pH 7.0의 20 mM Tris-HCl에 녹아 있는 0 내지 1.0 M 선형 농도 구배 NaCl을 사용, 1 ml/분의 속도로 30분간 용출시켰다. 1 ml의 분획을 모아 써던 옥수수 뿌리벌레 (SCR) 생물학적 정량을 실시하였다(실시예 13 참조). 활성 피크들은 SCR 생물학적 정량에서 각 분획을 차례로 희석시킴으로써 결정하였다. SCR에 대한 두 개의 활성 피크를 관찰하였는데 이들을 A (약 0.2 내지 0.3 M NaCl에서 용출시킴.) 및 B (0.3 내지 0.4 M NaCl에서 용출시킴)로 명명하였다. 활성 피크 A 및 B를 개별적으로 모아 하기 3 단계 처리를 하여 두 피크를 한층 더 정제하였다.

고형 (NH₄)₂SO₄를 상기 단백질 분획에 첨가하여 최종 농도가 1.7 M이 되게 하였다. 그 후, 단백질을 pH 7의 50 mM 인산 칼륨 완충액에 녹아 있는 1.7 M (NH₄)₂SO₄로 평형시킨 페닐-수퍼로스 (phenyl-Suprose) 5/5 컬럼에 1 ml/분으로 로딩시켰다. 컬럼에 부착되어 있는 단백질들은 1.7 M (NH₄)₂SO₄, 0% 에틸렌 글리콜, pH 7.0의 50 mM 인산 나트륨 내지 25% 에틸렌 글리콜, pH 7.0의 25 mM 인산 나트륨 ((NH₄)₂SO₄ 부재)의 선형 농도 구배로 0.5 ml/분의 속도에서 용출시켰다. 분획은, pH 7.0의 10 mM 인산 나트륨에 대하여 철야 투석시켰다. SCR에 대한 각 분획에서의 활성을 생물학적 정량으로 결정하였다.

가장 높은 활성을 가지는 분획을 모아 pH 7.0의 20 mM Tris-HCl로 평형시킨 모노큐 5/5 컬럼에 1 ml/분으로 로딩시켰다. 컬럼에 부착되어 있는 단백질들은 pH 7.0의 20 mM Tris-HCl에 녹아 있는, 0 내지 1 M 선형 농도 구배의 NaCl로 1 ml/분의 속도에서 용출시켰다.

정제의 최종 단계를 위하여, 상기 가장 높은 활성을 나타내는 분획을 모아 두 번째 페닐-수퍼로스 5/5 컬럼에 로딩시켰다. 고형 (NH₄)₂SO₄를 최종 농도가 1.7 M이 되게 첨가하였다. 그 후, 용액을 pH 7.0의 50 mM 인산 칼륨 완충액에 녹아 있는 1.7 M (NH₄)₂SO₄로 평형시킨 컬럼상에 1 ml/분으로 로딩시켰다. 컬럼에 부착되어 있는 단백질들은 1.7 M (NH₄)₂SO₄, pH 7.0의 50 mM 인산 칼륨 내지 pH 7.0의 10 mM 인산 칼륨 선형 농도 구배로 0.5 ml/분으로 용출시켰다. 분획들을 pH 7.0의 10 mM 인산 나트륨 완충액에 대하여 철야 투석시켰다. SCR에 대한 각 분획에서의 활성은 생물학적 정량으로 결정하였다.

상기 3단계 처리로, 피크 A로부터 최종 정제된 단백질을 독소 A라 명명하였고, 피크 B로부터 최종 정제된 단백질을 독소 B라 명명하였다.

독소 A 및 독소 B의 특성화 및 아미노산 서열 결정

SDS-PAGE에서, 독소 A 및 독소 B는 모두 192 kDa (각각 A1 및 B1으로 명명) 및 58 kDa (각각 A2 및 B2로 명명)인 두 개의 주요 펩티드 (전체 쿠마지 착색 단백질의 >90%)를 포함하였다. 독소 A 및 독소 B는 모두 천연 PAGE 상에서는 1개의 주요 밴드만을 나타내었는데, 이는 A1 및 A2가 한 단백질 복합체의 서브유닛이며, B1 및 B2는 한 단백질 복합체의 서브유닛이라는 것을 나타낸다. 또한, 독소 A 및 독소 B 모두의 천연 분자량은 겔 여과 크로마토그래피에 의해 860 kDa으로 결정되었다. A1 대 A2의 상대 몰 농도는 SDS-PAGE 겔의 농도계 분석으로 결정한 바, 1 대 1 등량인 것으로 판단되었다. 유사하게, B1 및 B2 펩티드는 동일 몰 농도로 존재하였다.

독소 A 및 독소 B를 10% SDS-PAGE에서 전기 영동시켰으며, PVDF 막으로 이동시켜 블로팅시켰다. 아미노산 분석 및 N-말단 아미노산 서열 결정을 위하여, 블롯을 하버드 마이크로켄 (Harvard MicroChem) 및 캠브리지 프로켄 (Cambridge ProChem)으로 각각 보냈다. B1의 N-말단 아미노 서열은 서열 1인 tcbA 유전자의 TcbA₁₁ 영역 (서열 12, 87 내지 99 위치)와 동일하다는 것이 밝혀졌다. 펩티드 B2의 N-말단 특이 서열이 획득되었다(서열: 40). 펩티드 B2의 N-말단 아미노산 서열은 tcbA 유전자로부터 추정된 아미노산 서열의 TcbA₁₁₁ 영역 (서열: 12, 1935 내지 1945 위치)와 동일하였다. 그러므로, B 독소는 동일 유전자 생성물인 TcbA로부터 유래된 것으로 관찰된 두 종의 펩

티드 TcbA_{ii} 및 TcbA_{iii}를 주로 포함하고 있다.

A2의 N-말단 서열 (서열: 41)은 TcbA_{iii} 펩티드 및 다른 펩티드와 비교시 특유하다. A2 펩티드는 TcdA_{iii}로 나타내었다(실시에 17 참조). 서열 6은 서열 40 및 41의 혼합 아미노산 서열로 밝혀졌다.

펩티드 A1 및 A2에 대하여 내부 아미노산 서열 결정을 또한 실시하였다. 내부 아미노산 서열 결정을 위하여 10 µg의 독소 A를 10% SDS-PAGE에서 전기 영동시키고 PVDF 막에 이동시켜 블로팅시켰다. 블롯을 아미도 블랙으로 착색시킨 후, 각각 TcdA_{ii} 및 TcdA_{iii}로 나타낸 펩티드 A1 및 A2를 블롯으로부터 오려내어 하바드 마이크로캠 및 캠브리지 프로캠으로 보냈다. 펩티드를 트립신으로 분해시킨 후 HPLC 크로마토그래피를 실시하여, 개개의 펩티드를 분리시켰다. 선택된 트립신 처리 펩티드 단편에 대해 N-말단 아미노산 분석을 수행하였다. 펩티드 A1의 두 가지 내부 아미노산 서열 (TcdA_{ii}-PK71, 서열: 38 및 TcdA_{ii}-PK44, 서열: 39)은 tcbA 유전자 (서열: 12) TcbA_{ii} 영역의 추정 아미노산 서열과 현저한 상동성을 가지고 있음이 판명되었다. 유사하게, 펩티드 A2의 N-말단 서열 (서열: 41) 및 두 가지 내부 서열 (TcdA_{iii}-PK57, 서열: 42 및 TcdA_{iii}-PK20, 서열: 43)은 또한 tcbA 유전자 (서열: 12) TcbA_{iii} 영역의 추정 아미노산 서열과의 현저한 상동성을 보였다.

상기 결과들을 요약해 보면, 독소 복합체는 SCR에 대하여 독소 A 및 독소 B의 두 가지 이상의 활성 단백질 독소 복합체들을 가지고 있다. 독소 A 및 독소 B는 천연 분자량 및 서브유닛 분자량에 있어서 유사하지만, 이들 펩티드의 조성은 다르다. 독소 A는 펩티드 TcdA_{ii} 및 TcdA_{iii}를 주요 펩티드로서 포함하고 있으며, 독소 B는 펩티드 TcbA_{ii} 및 TcbA_{iii}를 주요 펩티드로서 포함하고 있다.

독소 C, Tca 펩티드의 정제 및 특성화

포도람두스 단백질 독소 복합체를 상기 설명한 바와 같이 분리하였다. 이 다음에, 독소 약 50mg을 2ml/분의 유속에서 pH 7.0의 20mM Tris-HCl로 평형시킨 MonoQ 10/10 칼럼에 가하였다. 280nm에서의 광학 밀도가 기준선 수준으로 돌아갈 때까지 이 칼럼을 pH 7.0의 20mM Tris-HCl로 세척하였다. 이 칼럼에 결합되어 있는 단백질을 60분 동안 2ml/분에서 pH 7.0의 20mM Tris-HCl 중의 0 내지 1M NaCl의 선형 구배를 사용하여 용출시켰다. 2ml의 분획을 모아서 1차 항체로서 pAb TcaB_{ii}-syn 항체 (실시에 21 참조)를 사용하여 웨스턴 분석을 하였다. pAb TcaB_{ii}-syn 항체와 반응시킨 분획을 모아서 최종 농도가 1.7M이 될때까지 고체 (NH₄)₂SO₄를 첨가하였다. 이어서, 단백질을 1ml/분에서 pH 7의 50mM 인산 칼륨 완충액 중의 1.7M (NH₄)₂SO₄로 평형시킨 페닐-수퍼로스 10/10 칼럼에 가하였다. 이 칼럼에 결합되어 있는 단백질을 120분 동안 1ml/분에서 1.7M (NH₄)₂SO₄, pH 7.0의 50mM 인산 칼륨 내지 pH 7.0의 10mM 인산 칼륨의 선형 구배를 사용하여 용출시켰다. 2ml의 분획을 모아서 pH 7.0의 10mM 인산 나트륨 완충액으로 하룻밤 동안 투석시키고, 1차 항체로서 pAb TcaB_{ii}-syn 항체를 사용하여 웨스턴 블롯에 의해 분석하였다.

상기 항체와 교차 반응시킨 분획을 모아서 1ml/분에서 pH 7.0의 20 mM Tris-HCl로 평형시킨 MonoQ 5/5 칼럼에 가하였다. 이 칼럼에 결합되어 있는 단백질을 30분 동안 pH 7.0의 20 mM Tris-HCl 중의 0 내지 1M NaCl의 선형 구배에 의해 1ml/분에서 용출시켰다.

상기에서 pAb TcaB_{ii}-syn 항체와 반응시킨 분획을 모아서 페닐-수퍼로스 5/5 칼럼에 가하였다. 최종 농도가 1.7M이 될때까지 고체 (NH₄)₂SO₄를 첨가하였다. 이어서, 이 용액을 1ml/분에서 pH 7.0의 50mM 인산 칼륨 완충액 중의 1.7M (NH₄)₂SO₄로 평형시킨 칼럼에 가하였다. 이어서, 이 칼럼에 결합되어 있는 단백질을 60분 동안 1.7M (NH₄)₂SO₄, pH 7.0의 50mM 인산 칼륨 내지 pH 7.0의 10mM 인산 칼륨의 선형 구배를 사용하여 용출시켰다. 분획은 pH 7.0의 10mM 인산 나트륨 완충액으로 하룻밤 동안 투석시켰다.

최종 정제 단계에서는, 웨스턴 분석에 의해 상기에서 측정된 pAb TcaB_{ii}-syn 항체와 반응시킨 분획을 모아서 1ml/분에서 pH 7.0의 20mM Tris-HCl로 평형시킨 MonoQ 5/5 칼럼에 가하였다. 이 칼럼에 결합되어 있는 단백질은 30분 동안 pH 7.0의 20mM Tris-HCl 중의 0 내지 1M NaCl의 선형 구배에 의해 1ml/분에서 용출시켰다.

최종 정제된 단백질 분획은 SDS-PAGE에 의해 시험된 6개의 주요 펩티드, 165kDa, 90kDa, 64kDa, 62kDa, 58kDa 및 22kDa를 함유하였다. 이와 같이 정제된 분획의 살충 활성의 LD50은 SCR 및 ECB 각각에 대해 100ng 및 500ng으로 측정되었다.

상기 펩티드를 PVDF 멤브레인으로 블롯시키고, 블롯들을 각각 Harvard MicroChem 및 Cambridge ProChem에 아미노산 분석 및 5개의 아미노산 N-말단 서열화를 위해 보냈다. 165kDa 펩티드의 N-말단 아미노산 서열은 펩티드 TcaC (서열 2, 1 내지 5 위치)와 동일한 것으로 결정되었다. 90kDa 펩티드의 N-말단 아미노산 서열은 tcaA 유전자 (서열 33, 254 내지 258 위치)의 아미노산 서열에서 유도된 TcaA_{ii} 영역인 것으로 결정되었다. 64kDa 펩티드의 N-말단 아미노산 서열은 펩티드 TcaB_i (서열 3, 1 내지 5 위치)와 동일한 것으로 결정되었다. 62kDa 펩티드의 N-말단 아미노산 서열은 tcaA 유전자 (서열 33, 489 내지 493 위치)의 아미노산 서열에서 유도된 TcaA_{ii} 영역인 것으로 결정되었다. 58kDa 펩티드의 N-말단 아미노산 서열은 펩티드 TcaB_{ii} (서열 5, 1 내지 5 위치)와 동일한 것으로 결정되었다. 22kDa 펩티드 (서열 62)의 N-말단 아미노산 서열은 tcaA 유전자 (서열 34, 98 내지 102 위치)의 아미노산 서열에서 유도된 TcaA_{iv}로 표시된 TcaA_i 영역인 것으로 결정되었다. 모든 tcaA, tcaB 및 tcaC 유전자가 동일한 tca 오페론에 있다는 것이 주목된다 (도 6A).

정제된 Tca 분획, 정제된 독소 A, 및 정제된 독소 B 5 µg을 1차 항체로서 pAb TcaB_{ii}-syn 항체, mAb CF52 항체, pAb TcdA_{ii}-syn 항체, 및 pAb Tcd_{iii}-syn 항체와 같은 항체를 개별적으로 사용하여 웨스턴 블롯에 의해 분석하였다 (실시에 21). pAb TcaB_{ii}-syn 항체와는 정제된 Tca 펩티드 분획만을 반응시키고, 독소 A 및 독소 B는 반응시키지 않는다. mAb CF52 항체와는 독소 B만을 반응시키고, Tca 펩티드 분획 또는 독소 A는 반응시키지 않는다. pAb TcdA_{ii}-syn 항체 또는 pAb Tcd_{iii}-syn 항체와는 독소 A만을 반응시키고, Tca 펩티드 분획 또는 독소 B는 반응시키지 않는다. 이는, 정제된 Tca 펩티드 분획에서 관찰된 살충 활성이 독소 A 및 독소 B와는 무관하다는 것을 증명한다. 정제된 Tca 펩티드 분획은 독소 C로 표시되는 제3의 독특한 단백질 독소이다.

<실시에 16>

TcbA 펩티드의 분해 및 활성화

독소 B 복합체에서 펩티드 TcbA_{ii} 및 TcbA_{iii}는 단일 유전자 생성물인 TcbA로부터 유래한다(실시에 15). TcbA 펩티드가 TcbA_{ii} 및 TcbA_{iii}로 프로세싱되는 것은 아마도 포토랍두스 단백질 분해효소(들), 가장 가능성 있게는, 실시에 10에서 기술한 메탈로프로테아제의 작용에 의한 것으로 추정된다. 몇몇 경우에서, 포토랍두스 W-14 브로쓰가 프로세싱되었을 때 펩티드 TcbA_{ii} 및 TcbA_{iii} 뿐만 아니라 TcbA 펩티드도 독소 B 복합체에서 주요 성분으로 존재한다는 것을 알았다. W-14 브로쓰의 독소 복합체 분획으로부터 펩티드 TcbA를 모으기 위해, 독소 B 복합체의 정제에서 기술된 것과 (실시에 15) 동일한 방법을 사용하였다. 최종 정제 물질을 4-20% 농도구배 SDS-PAGE에서 분석하였으며, 주요 펩티드는 농도계로 정량하였다. TcbA, TcbA_{ii} 및 TcbA_{iii}는 각각 전체 단백질의 58%, 36% 및 6%이었다. 상기 펩티드들의 본질은 SDS-PAGE에서의 각각의 분자 크기와 단일특이 항체를 사용한 웨스턴 블롯 분석으로 확인하였다. 상기 분획의 천연 분자량은 860 kDa으로 결정되었다.

상기 정제 물질에, 38 kDa 및 58 kDa인 정제 W-14 포토랍두스 메탈로프로테아제로 처리하고 (실시에 10) 대조군 효소로서 트립신 (미주리주 소재, 시그마 (Sigma))을 처리함으로써 TcbA의 분해를 평가하였다. 표준 반응물에는, 전체 용량 100 µl 중에 상기 정제 분획 17.5 µg, 1.5 단위 프로테아제, 및 pH 8.0의 0.1 M Tris 완충액이 포함되어 있다. 대조군 반응에서는 프로테아제가 없다. 반응 혼합물을 37°C에서 90분간 방치하였다. 반응 종결시에 20 µl를 취하였으며, 4-20% 농도구배 SDS-PAGE에서의 전기 영동 분석을 위하여 SDS-PAGE 샘플 완충액으로 즉시 끓였다. 38 kDa 및 58 kDa 프로테아제 처리 모두에서, 펩티드 TcbA_{ii} 및 TcbA_{iii}의 양은 약 3배 증가한 반면, TcbA 펩티드의 양은 비례적으로 감소하였음이 SDS-PAGE로부터 결정되었다 (표 24). 선택된 펩티드의 상대적인 감소 및 증가는 웨스턴 블롯 분석으로 확인하였다. 또한, 분해된 물질을 겔 여과시켜 본 결과, 상기 복합체의 천연 분자 크기가 변화하지 않았음을 알 수 있었다. 트립신 처리시, 펩티드 TcbA 및 TcbA_{ii}가 작은 펩티드들로 비특이적으로 분해되었다. 이는, 38 kDa 및 58 kDa 포토랍두스 프로테아제가 펩티드 TcbA를 펩티드 TcbA_{ii} 및 TcbA_{iii}로 특이적으로 프로세싱시킬 수 있음을 나타내는 것이었다. 남아있는 80 µl 반응 혼합물의 프로테아제 처리 및 비처리 대조군을 pH 7.0의 10 mM 인산 나트륨 완충액으로 연속적으로 희석하여 SCR 생물학적 정량으로 분석하였다. 여러 희석물에서의 활성을 비교함으로써 38 kDa 프로테아제 처리시 SCR 살충 활성이 약 3 내지 4 배 증가되었음이 나타났다. 또한, 프로테아제 처리에서, 남아있는 곤충들의 성장 저해율은 대조군보다 심각하였다 (표 24).

[표 24]

프로테아제 처리에 의한 펩티드 TcbA의 펩티드 TcbA_{ii} 및 TcbA_{iii}로의 전환 및 활성화.

| | 대조군 | 38 kDa 프로테아제 처리 |
|---|-----|-----------------|
| S0(전체 단백질에 대한 %) | 58 | 18 |
| S1(전체 단백질에 대한 %) | 36 | 64 |
| S9(전체 단백질에 대한 %) | 6 | 18 |
| LD50 (µg 단백질) | 2.1 | 0.52 |
| SCR 중량 (mg/곤충)* | 0.2 | 0.1 |
| *: 본 분석에서, 규정식 5일 후 살아있는 곤충의 평균 중량의 측정에 의한 성장 저해율을 나타냄. | | |

SCR 장 프로테아제에 의한 독소 B의 활성화 및 처리

단백질 가수분해 활성화에 대한 두 번째 증명에 있어서, W-14 독소가 곤충에 의해 처리되는지의 여부를 시험하였다. 포토랍두스 W-14 브로쓰로부터 정제된 독소 B (실시에 15 참조)는 단클론성 항체를 사용하는 SDS-PAGE 및 웨스턴 블롯 분석에 의해 판정된 완전한 TcbA 펩티드로 주로 이루어져 있었다. SCR에 대한 이 분획의 LD50은 약 700ng인 것으로 측정되었다.

SCR 유충은 제4기의 영(instar) 단계 (각 곤충의 총 중량 약 100 내지 125mg)에 이를때까지 콜레옵테라(coleopteran) 먹이로 성장하였다. SCR 장 성분은 절개 가위 및 핀셋을 사용해서 제거하는 방법으로 모았다. 장 내벽을 감싸고 있는 다량의 지방질을 제거한 후, 약 40개의 장을 100µl의 무균수가 들어 있는 마이크로 원심 분리기 튜브에서 균질화하였다. 이어서, 이 튜브를 10분 동안 14,000 rpm에서 원심 분리하고, 펠렛은 버렸다. 사용할 때까지 상층액은 -70°C 동결기에서

저장하였다.

상기에서 정제된 독소 B를 모아진 SCR 장 성분으로 처리함으로써 곤충의 장에 의한 독소 B의 처리를 평가하였다. 이 반응은 100 μ l의 총부피 중의 40 μ g의 독소 B (1mg/ml), 50 μ l의 SCR 장 성분, 및 pH 8.0의 0.1M Tris 완충액을 포함하였다. 대조 반응에서는 SCR 장 성분을 뺐다. 이 반응 혼합물을 37 $^{\circ}$ C에서 하룻밤 동안 배양하였다. 반응 말렵에는, 10 μ l를 빼내서 SDS-PAGE 분석을 위해 동일한 부피의 2 \times SDS-PAGE 시료 완충액과 함께 끓였다. 나머지 90 μ l의 반응 혼합물은 pH 7.0의 10mM 인산 나트륨 완충액으로 연속 희석하고, SCR 성분 분석에 의해 분석하였다. SDS-PAGE 분석은 SCR 장 성분 처리에서 펩티드 TcbA가 보다 작은 펩티드로 완전히 소화되었다는 것을 증명하였다. 변성되지 않은 독소 분석의 분석은 더 큰 펩티드를 절단한다 하더라도 원래의 크기인 약 860kDa가 그대로 남아 있다는 것을 증명하였다. SCR 성분 분석에서는, 독소 B로 처리된 SCR 장의 LD50이 10배 증가된 약 70ng으로 밝혀졌다. 별도의 실험에서, 프로테아제 K 처리는 독소 활성을 완전히 제거하였다.

<실시예 17>

TcdA_{II} 펩티드를 코딩하고 있는 유전자 라이브러리의 스크리닝

서열 17 (내부 펩티드 TcdA_{II}-PT111 N-말단 서열) 및 서열 18 (내부 펩티드 TcdA_{II}-PT79 N-말단 서열)에서 기술된, TcdA_{II} 펩티드를 코딩하는 유전자의 클로닝 및 특성화가 완료되었다. 서열 17 (표 25) 및 서열 18 (표 26)의, 아미노산 서열을 코딩하도록 설계된 두 가지 폴의 동의성 올리고뉴클레오티드 및 그러한 서열의 역방향 상보체들을 실시예 8에 기술된 바와 같이 합성하였다. 올리고뉴클레오티드들의 DNA 서열을 하기에 나타내었다.

[표 25]

서열 17을 위한 동의성 올리고뉴클레오티드

| P2-PT111 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------|----------------|-------------|-------------|---------------|-------------|-------------|--------------------|--------|
| 아미노산 | Ala | Phe | Asn | Ile | Asp | Asp | Val | Ser |
| 코돈 | 5' GCN | TT(T/C) | AA(T/C) | AT(T/C/A) | GA(T/C) | GA(T/C) | GTN 3' | |
| P2.3.6.CB | 5' GC(A/C/G/T) | TT(T/C) | AAT | ATT | GAT | GAT | GT 3' | |
| P2.3.5 | 5' GC(A/C/G/T) | TT(T/C) | AA(T/C) | AT(T/C/A) | GA(T/C) | GA(T/C) | GT 3' | |
| P2.3.5.R | 5' AC | (G/A)T C | (G/A)T C | (T/G/A)A T | (G/A)T T | (G/A)A A | (A/C/G/T) GC 3' | |
| P2.3.5.R1 | 5' ACI | TCI | TCI | ATI | TTI | AAI | GC 3' | |
| P2.3R.CB | 5' CAG | (A/G)C T | (A/C)A C | ATC | ATC | AAT | ATT | AAA 3' |

[표 26]

서열 18을 위한 동의성 올리고뉴클레오티드

| P2-PT79 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|-----------|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------|
| 아미노산 | Phe | Ile | Val | Tyr | Thr | Ser | Leu | Gly | Val | Asn | Pro | Asn | Asn |
| 코돈* | 5' TTY | ATH | GTN | TAY | ACN | 6 | 6 | GGN | GTN | AAV | CCN | AAV | AAV 3' |
| P2.79.2 | 5' TTY | ATY | GTK | TAT | ACY | TCI | YTR | GGY | GTK | AAT | CCR | AAT | AAT 3' |
| P2.79.3 | 5' TTT | ATT | GTK | TAT | ACY | AGY | YTR | GGY | GTK | AAT | CCR | AAT | AAT 3' |
| P2.79.R.1 | 5' ATT | ATT | YGG | ATT | MAC | RCC | YAR | RCT | RGT | ATA | MAC | AAT | AAA 3' |
| P2.79R.CB | 5' ATT | ATT | YGG | ATT | MAC | ACC | CAG | RCT | GGT | ATA | MAC | AAT | AAA 3' |

* 뉴클레오티드의 IUPAC-IUB 코드에 따라서, Y=C 또는 T, H=A, C 또는 T, N=A, C, G 또는 T, K=G 또는 T, R=A 또는 G, 및 M=A 또는 C

모든 전향/역방향 조합에서, 전향 프라이머로서 P2.3.6CB 또는 P2.3.5 및 역방향 프라이머로서 P2.79.R.1 또는 P2.79R.CB를 사용하고, 주형으로서 포토랩두스 W-14 게놈 DNA를 사용, 본질적으로 실시예 8에 기술된 바와 같이, 폴리머라제 연쇄 반응을 수행하였다. 다른 반응 세트에서는, 모든 전향/역방향 조합에서, 전향 프라이머로서 프라이머 P2.79.2 또는 P2.79.3을 사용하였고, 역방향 프라이머로서 P2.3.5R, P2.3.5R1, 및 P2.3R.CB를 사용하였다. 전향 프라이머로서 P2.3.6.CB, 역

방향 프라이머로서 P2.79.R.1 또는 P2.79R.CB를 포함하는 반응에서만 2500 염기쌍의 추정 크기 (아가로스 겔 상의 이동성)를 가지는 비 유사 증폭 생성물이 관찰되었다. 상기 증폭 생성물을 수득하기 위해 사용된 프라이머의 서열은, 펩티드 단편 TcdA_{ii}-PT111이 펩티드 단편 TcdA_{ii}-PT79의 아미노-인접부위에 있음을 나타낸다.

2500 bp PCR 생성물은 제조업자의 지시에 따라 플라스미드 벡터 pCR™11 (미국 캘리포니아주 샌 디에고 소재의 인비트로젠 (Invitrogen) 제조)에 라이게이션시켰고, 두가지 단리물 (HS24 및 HS27)의 삽입체 단편의 DNA 서열은 제조자가 제시한 상기 프라이머 및 서열 분석 방법을 사용하여 결정하였다. 두가지 단리물 모두의 서열은 동일하였다. 신규 프라이머는 결정된 서열을 기초로 하여 합성시켰고, 전체 2557 염기의 삽입체 (서열 36)의 서열을 수득하기 위한 추가 서열 분석 반응을 프라이밍하기 위해 상기 신규 프라이머를 사용하였다. 서열 36에 의해 코딩되는 펩티드의 부분해독으로 서열 37에 공개된 845 아미노산 서열을 얻는다. TcdA_{ii} 펩티드 단편의 상기 부위의 단백질 상동성 분석에는, 상당한 아미노산 상동성(위스콘신주 매디슨 소재의 제너틱스 컴퓨터 그룹 (Genetics Computer Group, GCG)사의 위스콘신 패키지 버전 8.0을 사용하여 단백질 TcbA (서열 12)의 542 내지 1390 잔기와 68% 유사성, 및 53% 동일성, 또는 위스콘신주 매디슨 소재의 제너틱스 컴퓨터 그룹 (GCG)사의 위스콘신 패키지 버전 9.0을 사용하여 567 내지 1389 잔기와 60% 유사성, 및 54% 동일성)이 나타난다. 따라서, 서열 36에 부분적으로 나타낸 유전자는 TcbA 단백질과 유사하지만 동일하지는 않은 아미노산 서열을 나타내고, TcbA 단백질과 유사하지만 동일하지는 않은 생물학적 활성을 가지고 있다는 것이 명백하다.

다른 경우에는, 서열 39(내부 펩티드 TcdA_{ii}-PK44 서열) 및 서열 41(TcdA_{iii} 58 kDa N-말단 펩티드 서열)에 기술되어 있는 펩티드 TcdA_{ii}-PK44 및 TcdA_{iii} 58 kDa N-말단 펩티드를 코딩하는 유전자가 단리되었다. 서열 39 (표 28) 및 서열 ID 번호 41 (표 27)에 기술된 바와 같은 아미노산 서열을 코딩하기 위하여 설계된, 두 가지 풀의 동의성 올리고뉴클레오티드 및 그 서열의 역방향 상보체를 실시예 8 및 그들의 DNA 서열에 기술되어 있는 것과 같이 합성하였다.

[표 27]

서열 41을 위한 동의성 올리고뉴클레오티드

| 코돈 # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|--------|--------|-------|-----|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------|
| 아미노산 | Leu | Arg | Ser | Ala | Asn | Thr | Leu | Thr | Asp | Leu | Phe | Leu | Pro | Gln |
| A2.1 | 5' YTR | CGY | AGY | GCI | AAT | ACY | YTR | ACY | GAT | YTR | TTT | YTR | CCR | CA 3' |
| A2.2 | | | | GCI | AAT | ACI | YTR | ACI | GAY | YTR | TTY | YTR | CCI | CA 3' |
| A2.3.R | | 5' TG | YGG | YAR | AAA | YAR | RTC | RGT | YAR | RGT | RTT | IGC | RCT | RCG 3' |
| A2.4.R | | | | 5' TG | IGG | CAG | AAA | CAG | RTC | IGT | CAG | IGT | ATT | ICG 3' |

[표 28]

서열 39를 위한 동의성 올리고뉴클레오티드

| 아미노산 # | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|----------|--------|-----|------|------|------|------|------|------|-------|
| 코돈 # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 아미노산 | Gly | Pro | Val | Glu | Ile | Asn | Thr | Ala | Ile |
| A1.44.1 | 5' GGY | CCR | GTK | GAA | ATT | AAT | ACC | GCI | AT 3' |
| A1.44.1R | 5' ATI | GCG | GTA | TTA | ATT | TCM | ACY | GGR | CC 3' |
| A1.44.2 | 5' GGI | CCI | GTI | GAR | ATY | AAY | ACI | GCI | AT 3' |
| A1.44.2R | 5' ATI | GCI | GTR | TTR | ATY | TCI | ACI | GGI | CC 3' |

모든 전향/역방향 조합에서, 전향 프라이머로서 A1.44.1 또는 A1.44.2 및 역방향 프라이머로서 A2.3R 또는 A2.4R을 사용하고, 주형으로서 포토랩투스 W-14 게놈 DNA를 사용하여, 본질적으로 실시예 8에 기술된 바와 같이 폴리머라제 연쇄 반응을 수행하였다. 다른 반응 세트에서는, 모든 전향/역방향 조합에서, 전향 프라이머로서 프라이머 A2.1 또는 A2.2를 사용하였고, 역방향 프라이머로서 A1.44.1R 및 A1.44.2R을 사용하였다. 전향 프라이머로서 A1.44.1 또는 A1.44.2, 역방향 프라이머로서 A2.3R을 포함하는 반응에서만 1400 염기쌍의 추정 크기 (아가로스 겔 상의 이동성)를 가지는 비 유사 증폭 생성물이 관찰되었다. 상기 증폭 생성물을 수득하기 위해 사용된 프라이머의 서열은 펩티드 단편 TcdA_{ii}-PK44가 TcdA_{iii}의 58 kDa 펩티드 단편의 아미노-인접부위에 있음을

나타낸다.

1400 bp PCR 생성물은 제조자의 지시에 따라 플라스미드 벡터 pCR™ II 라이게이션시켰다. 네가지 단리물의 삽입체 단편의 DNA 서열은, 제조자가 제시한 프라이머와 유사한 서열의 프라이머 및 상기 서열 분석 방법을 사용하여 결정하였다. 모든 단리물들의 핵산 서열은 예상했던 바와 같이 동의성 프라이머 서열에 상응하는 부위에서 달랐으나, 상기 데이터로부터 추정된 아미노산 서열은 상기 결정된 펩티드의 실제 아미노산 서열과 동일하였다 (서열 41 및 39).

실시에 8에 기술된 바와 같이, 상기 제조된 DNA (서열 36)로 이루어진 식별용 동위원소로 표지된 프로브를 사용하여, W-14 계능 코스미드 라이브러리의 스크리닝을 실시하여 5개의 하이브리드화 코스미드 단리물, 즉 1709, 20B10, 21D2, 27B10, 및 26D1을 동정하였다. 상기 코스미드들은, 서열 11 또는 서열 25에 기술된 바와 같은 유전자에 대응하는 프로브로 이전에 동정되었던 코스미드들과는 다른 것들이었다. 제한 효소 분석 및 DNA 블롯 하이브리드화를 실시하여 서열 36의 DNA를 포함하는 부위인, 약 3.7, 3.7, 및 1.1 kbp의 크기를 갖는 3개의 EcoRI 단편을 동정하였다. 본 실시예에서 제조된, 동위 원소 표지된 1.4 kbp DNA 단편을 프로브로 사용하여 W-14 계능 코스미드 라이브러리를 스크리닝하여, 동일한 5개의 코스미드 (1709, 20B10, 21D2, 27B10, 및 26D1)를 동정하였다. EcoRI으로 분해된 코스미드 DNA에 대한 DNA 블롯 하이브리드화는 또한, 2.5 kbp TcdA_{II} 유전자 프로브에서 보여지는 것과 같이 EcoRI 단편의 동일 서브세트에 대한 하이브리드화를 나타내었는데, 이는 두 단편 모두가 계능 DNA 상에 코딩되어 있음을 나타내는 것이다.

클로닝된 EcoRI 단편의 DNA 서열 결정으로 2516 아미노산 (서열 47)의 282.9 kDa 단백질을 코딩하는 7551 염기쌍 (서열 46)의 비단속 해독 프레임이 밝혀졌다. 상기 단백질의 아미노산 서열 분석으로, 펩티드 TcdA_{II}의 예상되었던 모든 내부 단편 (서열 17, 18, 37, 38 및 39)와 TcdA_{III} 펩티드 N-말단 (서열 41) 및 모든 TcdA_{III} 내부 펩티드 (서열 42 및 43)가 밝혀졌다. TcdA_{II} 및 TcdA_{III}로 단리 및 동정된 펩티드는 서열 46으로 나타낸, tcdA로 표시된 개방 해독 프레임의 각각의 생성물이다. 또한, 서열 47은, 89 위치에서 출발하여 약 201 kDa 크기 펩티드의 N-말단 서열인 서열 13에 개시된 서열을 예시하는데, 이는 서열 46으로부터 생성된 초기 단백질이 앞에서 밝힌 서열 12에서와 유사한 방식으로 프로세싱됨을 나타낸다. 또한, 상기 단백질은 더한층 분해되어 서열 48에 의해 코딩되고 서열 49 (TcdA_{II} 펩티드)로서 개시된 209.2 kDa 크기의 생성물과, 서열 50에 의해 코딩되고 서열 51 (TcdA_{III} 펩티드)로서 개시된 63.6 kDa 크기의 생성물을 생기게 한다. 따라서, 서열 47로 개시된 282.9 kDa 크기의 전질이 단백질로 예시되는 바와 같이, 서열 46의 생성물로부터 유래되는 독소 A (실시에 15)로 동정되는 살충 활성은 프로세싱되어 서열 49 및 51로 개시된 펩티드를 생산하는 것으로 생각된다. 독소 B (실시에 15)로 동정되는 살충 활성은, 서열 12로 개시된 280.6 kDa 단백질로 예시되는 서열 11의 생성물로부터 유래하는 것으로 생각된다. 상기 단백질은 단백질 가수 분해에 의해 프로세싱되어, 서열 52에 의해 코딩되는, 서열 53으로 개시된 207.6 kDa 펩티드와, 서열 54에 의해 코딩되는, 서열 40 및 서열 55로 개시된 N-말단 서열을 가지고 있는 62.9 kDa 펩티드를 생산한다.

서열 12로 개시된 단백질과 서열 47로 개시된 단백질의 아미노산 서열을 비교해 보면, 위스콘신주 매드슨 소재의 제너텍스 컴퓨터 그룹 (GCG)사의 위스콘신 패키지 버전 8.0을 사용하여 69% 유사성 및 54% 동일성, 또는 이 프로그램의 버전 9.0을 사용하여 60% 유사성 및 54% 동일성을 나타낸다. 상기 고도의 진화론적인 관계는 상기 펩티드들의 전체 아미노산 서열을 통하여 일률적이지는 않지만, 단백질의 카르복시-말단 끝으로 갈수록 그 정도가 더 커지는데, 이는 서열 51 (서열 47로부터 유래)과 서열 55(서열 12로부터 유래)로 개시된 펩티드들이 위스콘신주 매드슨 소재의 제너텍스 컴퓨터 그룹 (GCG)사의 위스콘신 패키지 버전 8.0을 사용하여 76% 유사성 및 64% 동일성, 또는 이 프로그램의 버전 9.0을 사용하여 71% 유사성 및 64% 동일성을 가지고 있기 때문이다.

<실시에 18>

포토라프두스 (균주 W-14) 브로쓰의 분무 사용에 의한, 유럽 옥수수 천공충에 의해 야기된 옥수수 식물 잎 손상의 제어

곤충 유충에 의해 야기되는 식물 손상을 감소시키는 포토라프두스 독소(들)의 능력을, 포토라프두스 브로쓰로 처리한 옥수수 식물상에 만연시킨 유럽 옥수수 천공충 (*Ostrinia nubilalis*)에 의해 야기된 잎 손상을 측정함으로써 증명하였다. 포토라프두스 균주 W-14로부터의 발효 브로쓰를 만들어 실시에 13에 기술된 바와 같이 한외여과 (10000 MW 구멍 크기)를 사용하여 약 10배 농축시켰다. 그 결과 생성된 농축 브로쓰를 0.2 마이크론 니트로셀룰로스 막 여과기를 사용하여 여과 살균시켰다. 유사하게 만들어진, 접종시키지 않은 2% 프로테오스 펩톤 #3 샘플을 대조군 목적으로 사용하였다. 옥수수 식물 (번식 계열)은, 온실 내 흙을 사용하지 않는 혼합물을 포함하는 화분에서 종자로부터 식물 단계 7 또는 8까지 성장시켰다(낮 27°C, 밤 22°C, 약 50%의 상대 습도, 낮길이 14 시간, 필요시 물/비료 공급). 시험 식물들은, 낮 온도 약 22°C, 밤 온도 약 18°C, 인공 조명 무 및 부분 차광, 약 50%의 상대 습도 및 필요시 물/비료 공급하는 온실에서, 랜덤한 완전 차단 설계 (3 렘/처리, 6 식물/처리)로 준비하였다. 처리 (접종시키지 않은 배지 및 농축 포토라프두스 브로쓰)는 주사기 분무기를 사용하여 나선부 위 (약 6인치)로부터 직접적으로 2.0 ml을 적용하고, 추가로 2.0 ml을 나선부 위 약 30 cm (약 1 걸음)에서 원형 이동으로 도포시켰다. 또한, 식물의 한 집단은 아무 처리도 하지 않았다. 처리물이 마른 후 (약 30분), 12개의 신생 유럽 옥수수 천공충 유충 (알은 구매원으로부터 구입하여 부화시켰음)을 나선부에 직접적으로 적용하였다. 1주일 후, 변형된 구트리 스케일 (Guthrie Scale)을 사용하여 잎에 대한 손상을 기록하였고 [Koziel, M. G., Beland, G. L., Bowman, C., Carozzi, N. B., Crenshaw, R., Crossland, L., Dawson, J., Desai, N., Hill, M., Kadwell, S., Launis, K., Lewis, K., Maddox, D., McPherson, K., Meghji, M. Z., Merlin, E., Rhodes, R., Warren, G. W., Wright, M. 및 Evola, S. V., 1993, *Bio/Technology*, 11, 194-195 참조], 그 기록을 통계적으로 비교하였다 (T-시험 (LSD) p < 0.05 및 투키의 연구원화 범위 (Tukey's Studentized Range (HSD) 시험 p < 0.1). 결과는 표 29에 예시되어 있다. 참고로,

스코어 1은 손상이 없음을 나타내며, 스코어 2는 바늘구멍이 없이 퍼지지 않은 잎 위에 미세한 "윈도 페인 (window pain)" 손상을 나타내고, 스코어 5는 길다란 손상의 잎 관통 및(또는) 3개 이상의 잎에 명확한 중앙 엽맥 섭식 (손상 < 1인치)을 나타낸다. 상기 데이터는 브로쓰 또는 다른 단백질 포함 분획이 분무 가능 제제로 전달되거나 단백질 또는 그의 일부분을 코딩하는 유전자 또는 그의 유도체가 유전자전환된 식물 또는 세균을 통하여 전달될 때, 특정 해충에 대한 방어력을 제공할 수 있음을 나타낸다.

[표 29]

유럽 옥수수 천공충에 의해 야기된 옥수수 상의 잎 손상에 대한 포토랩두스 배양 브로쓰의 효과

| 처리 | 평균 구트리 스코어 |
|--|-------------------|
| 비처리 | 5.02 ^a |
| 비접종 배지 | 5.15 ^a |
| 포토랩두스 브로쓰 | 2.24 ^b |
| 상이한 문자로 된 평균은 통계학적으로 서로 다름(p < 0.05 또는 p < 0.1). | |

<실시예 19>

대장균에서의 발현을 위한 유전자들의 유전 공학

구조물의 요약

일련의 플라스미드들을 포토랩두스 W-14의 tcbA 유전자를 대장균에서 발현시키기 위하여 제작하였다. 플라스미드의 목록은 표 30에 나타내었다. 각 구조물의 간단한 설명 및 획득된 대장균 발현 데이터의 요약도 나타내었다.

[표 30]

tcbA 유전자 발현용 플라스미드

| 플라스미드 | 유전자 | 벡터/선발표지 | 구획 |
|-------------------------------------|------|--------------|--------------|
| pDAB2025 | tcbA | pBC/ChI | 세포내 |
| pDAB2026 | tcbA | pAcGP67B/Amp | 배클로바이러스, 분비됨 |
| pDAB2027 | tcbA | pET27b/Kan | 페리플라즘 |
| pDAB2028 | tcbA | pET15-tcbA | 세포내 |
| 약어: Kan=카나마이신, ChI=클로람페니콜, Amp=암피실린 | | | |

pDAB2025의 제작

TcbA_{II} 프로브에 하이브리드화하는 큰 EcoRI 단편이 실시예 9에 기술되어 있다. 상기 단편을 pBC (캘리포니아주 라 졸라 소재의 스트라타젠 (Stratagene) 제조) 내로 서브클로닝시켜서 pDAB2025를 만들었다. 서열 분석을 해 보면 상기 단편이 8816 염기쌍이라는 것이 나타난다. 상기 단편은 571 위치에서 개시 ATG를, 8086 위치에서 종결 TAA를 가지고 있는 tcbA 유전자를 코딩한다. 그러므로 상기 단편은, ATG의 상류에 570 염기쌍의 포토랩두스 DNA 및 TAA의 하류에 730 염기쌍을 가지고 있다.

플라스미드 pDAB2026의 제작

5' 프라이머 (S1Ac51)인 5' TTT AAA CCA TGG GAA ACT CAT TAT CAA GCA CTA TC 3' 및 3' 프라이머 (S1Ac31)인 5' TTT AAA GCC GCT TAA CGG ATG GTA TAA CGA ATA TG 3'을 사용하여 tcbA 유전자를 플라스미드 pDAB2025로부터 PCR로 증폭시켰다. 물 57.5 ml, 10X LA 완충액 10 ml, dNTPs 16 ml (각각이 2.5 mM인 원액), 10 pmoles/ml로 각 프라이머 20 ml, W-14 tcbA 유전자를 포함하고 있는 플라스미드 pDAB2025를 300 ng, TaKaRa LA Taq 폴리머라제 1 ml의 반응 조건에서 판베라 (PanVera, 위스콘신주 메디슨 소재)로부터 구입한 TaKaRa LA PCR 키트를 사용하여 PCR을 수행하였다. 사이클링 조건은 98°C/20초, 68°C/5분, 72°C/10분의 30회 사이클이었다. 약 7526 bp로 예상되는 PCR 생성물을 TBE 완충액 (100 mM Tris, 90 mM 붕산 (Boric acid), 1 mM EDTA)를 사용한 0.8% 아가로스 겔에서 단리시켜 퀴아젠 (Qiagen, 캘리포니아주 챗스워쓰 소재)으로부터 구입한 Qiaex II 키트를 사용하여 정제하였다. 정제된 tcbA 유전자를 Nco I 및 Not I으로 잘라 배클로바이러스 이동 벡터인 pAcGP67B (캘리포니아주 산 디에고 소재의 파밍젠 (PharMingen) 제조) 내로 라이게이션시키고, 이를 사용하여 DH5α 대장균을 형질전환시켰다. 그 결과 생성되는 재조합형을 pDAB2026이라 명명하였다. 그 후 tcbA 유전자를 pDAB2026으로부터 잘라내어 pET27b로 옮겨 플라스미드 pDAB2027을 만들었다. tcbA 유전자 내의 미스센스 돌연 변이는 pDAB2027에서 복구되었다.

복구된 tcbA 유전자는 서열 11에 나타난 서열로부터 212에서의 A>G 변환은 아스파라긴 71을 세린 71로 변화시키고 229에서의 G>A 변환은 알라닌 77을 트레오닌 77로 변화시키는 두 가지 변화를

포함하고 있다. 상기 변화 모두는 제안된 TcbA_{iii} N-말단의 상류에서 일어난다.

pDAB2028의 제작

pDAB2027의 tcbA 코딩 부위를 벡터 pET15b로 옮겼다. 이는 쏘건 (shotgun) 라이게이션법을 사용하여 실시하였으며, DNA는 제한 효소 Nco I 및 Xho I으로 분해하였다. 그 결과 생성되는 재조합형을 pDAB2028라 명명하였다.

대장균에서의 플라스미드 pDAB2028로부터의 TcbA 발현

대장균에서의 tcbA의 발현은 Studier 등에 의해 종래에 기술된 방법을 변형시킴으로써 성취되었다 (Studier, F.W., Rosenberg, A., Dunn, J., 및 Dubendorff, J., (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol., 185: 60-89). 수용성 대장균 세포 균주 BL21 (DE3)를 플라스미드 pDAB2028로 형질전환시켜 100 µg/ml의 암피실린 및 40 mM의 포도당을 함유하는 LB 아가상에 도말하였다. 형질전환시킨 세포들은, 수백개의 단리된 콜로니/플레이트의 밀도가 되도록 도말하였다. 37°C에서 철야 방치한 후 세포를 플레이트로부터 긁어 모아 100 µg/ml의 암피실린을 포함하는 LB 브로쓰에 현탁시켰다. 일반적인 배양 용량은 200 내지 250 ml 이었다. 시간 0에서, 배양물 밀도 (OD600)는 실험에 따라 0.05 내지 0.15였다. 배양물은 세 종류의 온도 (22°C, 30°C 또는 37°C) 중 어느 하나의 온도에서 0.15 내지 0.5의 밀도가 획득될 때까지 흔들며 주었고, 상기 밀도가 획득되었을 때에 1 mM 이소프로필티오-β-갈락토시드 (IPTG)로 유도시켰다. 배양물은 지정된 온도에서 4 내지 5시간 동안 배양시킨 후 처리를 할 때까지 (12 내지 72 시간) 4°C로 옮겨서 방치하였다.

대장균에서 발현시킨 플라스미드 pDAB2028로부터의 TcbA의 정제 및 특성화

TcbA 펩티드를 발현하는 대장균 배양물을 하기와 같이 처리하였다. 17000 × g에서 원심분리하여 세포를 수확하였으며 배지를 따라 별도의 용기에 모아 두었다.

배지는 M12 (매사추세츠주 베벌리 소재의 아미콘 (Amicon) 제조) 여과 시스템 및 100 kD 분자 크기 컷 오프 여과기를 사용하여 약 8× 농축시켰다. 농축시킨 배지를 음이온 교환 컬럼상에 로딩시켰으며, 부착된 단백질을 1.0 M NaCl로 용출시켰다. 1.0 M NaCl 용출 피크는 써던 옥수수 뿌리 벌레 (SCR) 유충에 대하여 살충 효과를 야기하는 것으로 밝혀졌다 (표 30). 1.0 M NaCl 분획을 pH 7.0의 10 mM 인산 나트륨 완충액에 대하여 투석시켜 농축시켰으며, 세파로스 CL-4B (뉴저지주 피스카타웨이 소재의 파마시아 제조) 상에서 겔 여과시켰다. 천연 W-14 독소 복합체의 추정 분자량 (약 900 kDa)에 대응하는 CL-4B 용출 프로파일의 영역을 모아 농축시켰으며 유충에 대하여 생물학적 정량을 실시하였다. 모아진 900 kDa 분획은 천연 W-14 독소 복합체에 의해 야기되는 것과 유사한 징후를 갖는 살충 활성을 가지는 것으로 밝혀졌다 (하기 31 참조). 이 분획은 프로테이나제 K 및 열처리에 민감하였고, 두 경우 모두에서 활성이 소멸되거나 감소하였는데, 이는 활성이 사실상 단백질성에 의한 것이라는 증거를 제공한다. 또한, 활성 분획은 항-TcbA_{iii} 단클론성 항체로 시험시, 면역블롯 분석에서 TcbA 및 TcbA_{iii} 펩티드에 대해 면역학적 양성 여부를 시험하였다 (표 31).

[표 31]

면역블롯 및 SCR 생물학적 정량 결과

| 분획 | SCR 활성 | | 면역블롯 | 천연 크기 |
|--------------------------|--------|----------|---------------------------|----------|
| | % 치사율 | % 성장 저해율 | | |
| TcbA 배지 1.0 M | +++ | +++ | TcbA | |
| 이온 교환 | | | | |
| TcbA 배지 CL-4B | +++ | +++ | TcbA, TcbA _{iii} | ~900 kDa |
| TcbA 배지 CL-4B + 프로테이나제 K | ++ | +++ | NT | |
| TcbA 배지 CL-4B + 열처리 | - | - | NT | |
| TcbA 세포 상등액 CL-4B | - | +++ | NT | ~900 kDa |

PK = 프로테이나제 K 처리 2시간, 열처리 = 100°C에서 10분, ND = 검출되지 않음, NT = 시험되지 않음. 대조군 샘플과 비교한 치사율 및 성장 저해율의 스코어링 시스템; 5-24% = +, 25-49% = ++, 50-100% = +++.

세포 펠렛을 pH 7.0의 10 mM 인산 나트륨 완충액에 재현탁시켰으며 등록상표 Bio-Neb 세포 네블라이저 (nebulizer, 인디애나주 테라 호트 소재의 글라스-콜 인크. (Glas-Col Inc.) 제조)에 통과시켜 용균시켰다. 펠렛을 DNase로 처리하여 DNA를 제거하였으며, 17000 × g에서 원심분리하여 세포 상등액과 세포 펠렛을 분리시켰다. 상등액 분획을 따라내어 0.2 마이크론 여과기를 통하여 여과시켜 큰 입자를 제거하였으며, 음이온 교환 크로마토그래피를 실시하였다. 부착된 단백질을 1.0 M NaCl로 용출시켜 투석시켰으며, 50000 달톤의 분자 크기 컷 오프를 가지는 등록상표 바이오맥스 농축기 (매사추세츠주 베드포드 소재의 밀리포아 코포레이션 제조)를 사용하여 농축시켰다. 농축된 분획은 세파로스 CL-4B 비드로 된 매트릭스를 사용하여 겔 여과 크로마토그래피를 실시하였다.

상기 방식으로 제조된 물질의 생물학적 정량 데이터는 표 30에 예시되어 있으며, TcbA 세포 상등액으로 나타내었다.

많은 양의 물질을 다루는 다른 방법에서, 세포 펠렛을 pH 7.0의 10 mM 인산 나트륨 완충액에 재현탁시켰으며, 콘테스 글래스 캠페니 (Kontes Glass Company, 뉴저지주 바인랜드 소재)의 40 ml 조직 분쇄기를 사용하여 완전히 균질화시켰다. 세포 파쇄물은 25000 × g에서 원심분리하여 펠렛화하였고, 세포 상등액을 따라내어 0.2 마이크론 여과기를 통과시켰으며, 다공성 HQ 50 비드로 충전시킨 파마시아 10/10 컬럼을 사용하여 음이온 교환 크로마토그래피를 실시하였다. 부착된 단백질은 0.0 내지 1.0 M의 NaCl 농도구배를 실시하여 용출시켰다. TcbA 단백질을 포함하는 분획을 조합시켜 50 kDa 농축기를 사용하여 농축시켰으며, 파마시아 CL-4B 비드로 된 매트릭스를 사용하는 겔 여과 크로마토그래피를 실시하였다. 약 900 kDa의 분자 크기를 가지는 TcbA 올리고머를 포함하는 분획을 모았으며, 상기 분획을 pH 7.3의 20 mM Tris 완충액으로 평형시킨 파마시아 모노 Q 10/10 컬럼을 사용하여 음이온 교환 크로마토그래피를 실시하였다. 재조합 TcbA 단백질을 용출시키기 위하여 0.0 내지 1.0 M 선형 농도 구배의 NaCl이 사용되었다. 재조합 TcbA는 천연 TcbA 올리고머가 모노 Q 10/10 컬럼으로부터 용출되는 것과 동일한 몰 농도인 약 0.3 내지 0.4 M NaCl 염 농도에서 컬럼으로부터 용출되었다. 재조합 TcbA 분획은 표 31에서의 실험과 유사한 생물학적 정량 실험에서 SCR 치사를 유도하는 것으로 밝혀졌다.

발현 구조의 두번째 세트를 제조하고 TcbA 단백질 독소의 발현에 대해 검사하였다.

pDAB2030의 구성: tcbA 코딩 영역에 대한 발현 플라스미드

플라스미드 pDAB2028 (본원 참조)는 상업용 벡터 pET15 (Novagen, 위스콘신주 메디슨) 중에 tcbA 코딩 영역을 함유하고, 앰피실린 선택 마커를 코딩한다. 플라스미드 pDAB2030은 카나마이신 선택 마커를 코딩하는 플라스미드로부터 tcbA 코딩 영역을 발현시키도록 제조하였다. 이 제조는 pET27 (Novagen, 위스콘신 메디슨) 카나마이신 선택 플라스미드 및 pDAB2028를 Xba I 및 Xho I로 절단하여 실시한다. 이로써, 플라스미드 pDAB2028로부터 tcbA 코딩 영역을 포함한 전체 다중 클로닝 부위를 얻는다. 두개의 절단된 플라스미드를 혼합하고 라이게이션시켰다. 재조합 플라스미드를 카나마이신 상에서 선택하고 pDAB2028 단편을 함유한 것들을 제한 분석에 의해 확인하였다. 새로운 재조합 플라스미드를 pDAB2030으로 칭한다.

플라스미드 pDAB2031의 구성: tcbA_i에서 돌연변이의 보정

실시에 19에 기재된 바와 같이 tcbA 코딩 영역의 N-말단에서 두개의 돌연변이 (서열 ID No: 11: 212에서 A>G가 아스파라긴 71을 세린 71로 변화시킴; 229에서 G>A가 알라닌 77을 트레오닌 77로 변화시킴)를 다음과 같이 보정하였다. PCR 생성물을 프라이머 TH50 (5' ACC GTC TTC TTT ACG ATC AGT G 3') 및 S1Ac51 (5' TTT AAA CCA TGG GAA ACT CAT TAT CAA GCA CTA TC 3')과 템플릿으로서 pDAB2025를 사용하여 발생시킴으로써 1778 bp 생성물을 얻었다. 이 PCR 생성물을 플라스미드 pCR2.1 (Invitrogen, 캘리포니아주 샌 디에고)로 클로닝하고 클론을 단리하고 서열화하였다. 클론을 Nco I 및 Pin AI로 절단시키고 1670bp 단편을 1% 아가로스겔로부터 정제하였다. 돌연변이된 tcbA 코딩 영역을 함유한 플라스미드 (pDAB 2030)를 NcoI 및 Not I로 절단시키고 Qiaex II (Qiagen, 캘리포니아주 챗스워쓰)를 갖는 0.8% 아가로스 내에서 1670 bp 단편으로부터 정제하였다. 이어서, 보정된 Nco I/Pin AI 단편을 pDAB2030으로 라이게이션시켰다. 라이게이션된 DNA를 DH5α E. coli로 형질변환시켰다. 클론을 단리하고, 시퀀싱하고 보정되었음을 확인하였다. 이 보정된 tcbA 코딩 영역을 함유한 플라스미드를 pDAB2031로 칭한다.

pDAB2033 및 pDAB2034의 구성: tcbA에 대한 발현 플라스미드

발현 플라스미드 pDAB2025 및 pDAB2027은 모두 박테리오파지 T7 발현 시스템에 좌우된다. 추가 벡터 시스템을 tcbA 유전자 및 그의 유도체의 발테리아성 발현에 사용하였다. 발현 벡터 Trc99a (Pharmacia Biotech, 뉴저지주 핏츠카타웨이)는 pDAB2030 및 2031로부터의 tcbA 코딩 영역과 상용성인 5' Nco I 부위와 다중 클로닝 부위의 강한 trc 프로모터 상류를 함유한다. 그러나, 플라스미드는 상용성 3' 부위를 갖지 않는다. 그러므로, Trc99a의 Hind III 부위를 절단하고 T4 DNA 폴리머라제 (Boehringer Mannheim, 인디애나주 인디애나폴리스)로 처리하여 융통하게 만들었다. 이어서, 벡터 플라스미드를 Nco I로 절단하고 알칼리성 포스파타아제로 처리하였다. 플라스미드 pDAB2030 및 pDAB2031을 각각 Xho I로 절단하고 (tcbA 코딩 영역의 3' 말단에서 절단) T4 DNA 폴리머라제로 처리하여 말단을 융통하게 만들었다. 이어서, 플라스미드를 Nco I로 절단하고, DNA를 페놀로 추출하고, 에탄올 침전시키고, 완충액에 재현탁시켰다. Trc99a과 pDAB2030 및 pDAB2031 플라스미드를 따로따로 혼합하고, 라이게이션하고, DH5α 세포로 형질전환시키고, 앰피실린 및 50 mM 글루코스를 함유한 LB 매질 상에 플레이팅하였다. 재조합 플라스미드를 제한 절단에 의해 확인하였다. 새로운 플라스미드를 pDAB2033 (tcbA_i 중의 두개의 돌연변이를 갖는 tcbA 코딩 서열을 함유함) 및 pDAB2034 (pDAB2031로부터의 tcbA의 보정된 버전을 함유함)로 칭한다.

플라스미드 pDAB2032의 구성: tcbA_iA_{iii}에 대한 발현 플라스미드

TcbA의 tcbA_iA_{iii} 부분을 코딩하는 플라스미드를 플라스미드 pDAB2031과 같은 방식으로 제조하였다. PCR 생성물을 TH42 (5' TAG GTC TCC ATG GCT TTT ATA CAA GGT TAT AGT GAT CTG 3') 및 TH50 (5' ACC GTC TTC TTT ACG ATC AGT G 3') 프라이머 및 템플릿으로서 pDAB2025를 사용하여 발생시켰다. 이로써 tcbA_i의 코딩 서열의 개시부에 초기 코돈을 갖는 1521 bp의 생성물을 얻었다. 이 PCR 생성물을 1% 아가로스겔 중에 단리하여 정제하였다. 정제된 생성물을 상기와 같이 pCR2.1로 클로닝하고 보정된 클론을 DNA 서열 분석에 의해 확인하였다. 이 클론을 Nco I 및 Pin AI로 절단시키고, 1414 bp 단편 1% 아가로스 중에서 단리하고, 플라스미드 pDAB2030의 Nco I 및 Pin AI 부위로 라이게이션시키고 DH5α E. coli로 형질변환시켰다. TcbA_iA_{iii}를 발현하도록 지정된 이 새로운 플라스미드를 pDAB2032로 칭한다.

플라스미드 pDAB2030, pDAB2031 및 pDAB2032로부터 tcbA 및 tcbA_{ii}A_{iii}의 발현

플라스미드 pDAB2030, pDAB2031 및 pDAB2032로부터 E. coli에 tcbA를 발현시키는 것을 tcbA_{ii}A_{iii}를 E. coli 균주 HMS174 (DE3) (Novagen, 위스콘신주 메디슨)에서 실시한 것을 제외하고는 본원에 기재한 바와 같이 실시하였다.

플라스미드 pDAB2033으로부터 tcbA의 발현

플라스미드 pDAB2033을 BL21 세포 (novagen 위스콘신주 메디슨)로 형질변환시키고, 100 µg/ml 앰피실린 및 50 mM 글루코스를 함유하는 LB 상에 플레이팅하였다. 각 플레이트 상에 수백개의 웰의 분리된 군체가 존재하도록 플레이트를 스프레딩한 후 밤새 30°C 또는 37°C에서 인큐베이션시켰다. 군체를 플레이트로부터 긁어내고 100 µg/ml 앰피실린을 함유하지만 글루코스를 함유하지 않은 LB 중에 현탁시켰다. 전형적인 배양액 부피는 단일 1 ℓ 배플 바닥 플라스크 중에 250 ml였다. 배양액의 밀도가 0.3 내지 0.6 OD600 nm에 이르렀을 때 배양액을 유도시켰다. 이 농도는 플레이트로부터 세포를 현탁시킨 직후 도달하는 경우가 가장 빈번하고 액상 매질 중에서 성장 기간을 필요로 하지 않았다. 2 개의 유도 방법을 사용하였다. 방법 1: 세포를 37°C에서 1 mM IPTG로 유도시켰다. 배양액을 플랫폼용 진탕기 상에서 200 rpm으로 5 시간 동안 진탕하고 수확하였다. 방법 2: 배양액을 30°C에서 25 µM IPTG로 유도시키고 20°C 또는 30°C에서 15 시간 동안 200 rpm으로 진탕하였다. 배양액을 정제를 위해 사용할 때까지 4°C에 보관하였다.

E. coli로부터 TcbA의 정제

TcbA 및 TcbA_{ii}A_{iii}의 정제, 생물 검정법 및 면역블롯 분석을 본원에 기재된 바와 같이 실시하였다. 몇몇 대표적인 E. coli 발현 실험의 결과를 표 32에 제시하였다. 표 32에 제시된 모든 물질은 배양 배지 분획으로부터 정제하였다. 예측된 고유 분자량은 본원에 기재된 바와 같이 대략 900 kD 이다. 샘플의 순도, 오염 단백질에 대한 TcbA의 양은 각각의 제조법에 따라 달랐다.

[표 32]

| 이. 콜리로 부터 생산되고, 배양 배지로부터 정제된 TcbA 및 유도체의 생물 검정 활성 및 면역블롯 분석 | | | | | | |
|--|-------------------------------------|-------------|----------------------|-------|---|-----------------|
| 플라스미드 | 코딩 부위 | 이.콜리 균주 | 써던 옥수수 뿌리벌레 생물 분석 활성 | | 면역 블롯에 의하여 검출된 펩티드 | 막이에 투여된 단백질(µg) |
| | | | 성장억제 % | 치사율 % | | |
| pDAB2030 | tcbA | BL21(DE3) | - | +++ | TcbA+TcbA _{ii} | 1-8 |
| pDAB2031 | tcbA | BL21(DE3) | - | +++ | TcbA+TcbA _{ii} | 1-10 |
| pDAB2033 | tcbA | BL21 | - | +++ | TcbA+TcbA _{ii} | 1-2 |
| pDAB2032 | tcbA _{ii} A _{iii} | HMS174(DE3) | +++ | + | TcbA _{ii} A _{iii} +TcbA _{ii} | 13-27 |
| 대조 시료와 비교할 때, 써던 옥수수 뿌리벌레에서의 치사율 및 성장억제에 대한 스코어링 시스템: 5-24 %=+, 25-49%=++, 50-100%=+++ | | | | | | |

<실시에 20>

매트릭스-보조 레이저 흡수 이온화 타임-오브-플라이트 질량 스펙트로스코피를 이용한 독소 펩타이드의 특성화

W-14 브로쓰로부터 분리한 독소를 실시에 15에 기재한 바와 같이 분리하였다. 몇몇 경우에 TcaB 단백질 독소를 앞서 기재한 바와 같이 (실시에 15) W-14 브로쓰로부터 분리한 프로테아제 (실시에 16)으로 예비처리하였다. 단백질 분자 질량을 VOYAGER BIOSPECTROMETRY 워크스테이션에서 지연 추출 기술로 매트릭스-보조 레이저 흡수 이온화 타임-오브-플라이트 질량 스펙트로스코피 (이후 MALDI-TOF로 칭함)를 사용하여 결정하였다. 대개 주요 단백질 (5 µl 중의 100 - 500 pmoles)을 아세트니트릴 1 µl과 혼합하고 공극 크기가 0.025 µm인 밀리포어 VS 필터 (Millipore Corp. 메사추세츠주 베드포드) 상에서 0.5 내지 1 시간 동안 투석하였다. 투석은 물에 대해 필터를 부유시켜 수행하고 (광택면이 위로 가도록) 단백질-아세트니트릴 혼합물을 필터 표면 상에 적가하였다. 투석 후, 투석된 단백질을 피펫을 사용하여 제거하고 시나핀산 및 트리플루오로아세트산으로 이루어진 매트릭스와 엷자 지시에 따라 혼합하였다. 단백질 및 매트릭스를 약 3 cm² 금 도금 샘플 플레이트 (PerSeptive Corp.) 상에서 공결정화시켰다. 하기 조건을 사용하여 결정을 여기시키고 후속 질량 분석을 실시하였다. 레이저 셋팅 3050; 압력 4.55e-07; 저 질량 게이트 1500.0; 음이온 오프; 가속 전압 25,000; 그리드 전압; 90.0%; 가이드 와이어 전압 0.010%; 선형 모드; 및 펄스

지연 시간 350 ns.

단백질 질량 분석 데이터를 표 33에 제시하였다. MALDI-TOF로부터 얻은 데이터를 유전자 서열 정보로부터 가정한 것과 SDS-PAGE로 앞서 결정한 것과 비교하였다.

[표 33]

| 유전자 서열을 기초로하여 MALDI-TOF, SDS-PAGE 및 예측 측정에 의한 펩티드의 분자 분석 | | | |
|--|------------|------------|--|
| 펩티드 | 예측 (유전자) | SDS PAGE | MALDI-TOF |
| TcbA | 80,634 Da | 240,000 Da | 281,040 Da |
| TcbA _{i/ii} | 217,710 Da | 해석되지 않음 | 216,812 Da |
| TcbA _{ii} | 207,698 Da | 201,000 Da | 206,473 Da |
| TcbA _{iii} | 62,943 Da | 58,000 Da | 63,520 Da |
| TcbA _{ii} | 209,218 Da | 280,634 Da | 280,634 Da |
| TcbA _{iii} | 63,943 Da | 280,634 Da | 280,634 Da |
| TcbA _{ii} | 생성된 프로테아제 | 280,634 Da | 280,634 Da 280,634 Da 280,634 Da 280,634 Da |
| TcbA _{iii} | 생성된 프로테아제 | 280,634 Da | 280,634 Da |

*TcbA_{i/ii}에서 관찰된 다중 단편인 일반화된 TcbA 데이터

<실시예 21>

펩티드 특이성 다클론성 항체의 생산

W-14 독소 복합체의 9 개의 펩티드 성분, 즉 TcaA, TcaA_{iii}, TcaB, TcaB_{ii}, TcaC, TcbA_{ii}, TcbA_{iii}, TcdA_{ii} 및 TcdA_{iii}를 표적으로 선택하고 이에 대해 항체를 생산하였다. 이들 펩티드에 대한 상보성 DNA 및 유도된 아미노산 서열 데이터는 펩티드 중 몇몇 사이의 서열 동일성이 실제함을 나타내었다. 만일 전체 펩티드가 항체 생산을 유도하는 임유노겐으로서 사용되었다면, 생성되는 항체는 독소 제조에서 여러 펩티드에 결합할 것이다. 이러한 문제점을 피하기 위해 각각의 원하는 펩티드의 고유 영역에 특이적으로 결합하는 항체를 발생시켰다. 각각의 표적 펩티드의 고유 영역 (서브펩티드)는 하기 기재된 분석을 기초로하여 선택하였다.

각각의 전체 펩티드 서열을 MacVector (상표명) Protein Analysis Tool (IBI Wequence Analysis Software, International Biotechnologies, Inc., 코네티컷주 뉴헤이브 피.오. 박스 9558)을 사용하여 분석하고 그의 항원 지수를 결정하였다. 이 프로그램은 가능한 외부에 위치한 아미노산 서열, 즉 항원 부위일 수 있는 영역의 위치를 지정하도록 고안되었다. 이 방법은 친수성, 표면 가능성 및 2차 구조 예측에 따른 골격 가요성 예측으로부터 정보를 조합하여 단백질의 표면 윤곽에 대한 복합적인 예측안을 낸다. 각각의 분석에 대한 점수는 -1.0과 +1.0 사이의 값으로 표준화하였다 (MacVector (상표명) Manual). 항원 지수값을 표적 펩티드의 전체 서열에 대해 얻었다. 각각의 펩티드로부터, 원래 서열로부터 높은 항원 지수를 보인 19개 이상의 아미노산을 커버하는 영역을 재분석하여 측면 잔기 없이 서브펩티드의 항원 지수를 결정하였다. 이 재분석은 펩티드의 항원 지수가 측면 아미노산 잔기에 의해 영향을 받을 수 있기 때문에 필수적이다. 만일 단리된 서브펩티드 서열이 고 항원 지수를 유지하지 않는다면, 새로운 영역을 선택하고 분석을 반복하였다.

각각의 선택된 서브펩티드 서열을 정렬하고 MacVector (상표명) 정렬 프로그램을 사용하여 모든 7 개 표적 펩티드 서열에 대해 비교하였다. 만일 선택된 서브펩티드 서열이 다른 표적 펩티드에 대해 동일성 (20% 이상)을 보이면, 새로운 19 개 이상의 아미노산 영역을 단리하여 재분석하였다. 고 항원 지수를 보이는 19 개 이상의 아미노산을 커버하는 고유의 서브펩티드 서열을 모든 표적 펩티드로부터 선택하였다.

7개 서브펩티드 서열을 제네미드 바이오테크놀로지 인크. (Genemed Biotechnology Inc.)에 보냈다. 각각의 서브펩티드 상의 마지막 아미노산 잔기는 항원 지수에 뚜렷한 영향을 보이지 않기 때문에 삭제하였다. 시스테인 잔기를 내부 시스테인 잔기를 함유한 TcaBi-syn을 제외하고 각각의 서브펩티드 서열의 N-말단에 첨가하였다. 시스테인 잔기가 존재하면 담체 단백질 (KLH)의 컨주게이션이 촉진된다. 적당한 독소 펩티드에 대응한 최종 펩티드 생성물 및 서열 ID NO.를 표 34에 제시하였다.

[표 34]

합성 펩티드의 아미노산 서열

| 서열 번호 | 펩티드 아미노산 서열 |
|-------|--|
| 63 | TcaA _{iii} -syn NH ₂ -(C)LRGNSPTNPDKDGI FAQVA |
| 64 | TcaA _{iii} -syn NH ₂ -(C)YTPDQTPSFYETA FR SADG |
| 65 | TcaB _i -syn NH ₂ -HGQSYNDNNYCNFTLSINT |
| 66 | TcaB _{iii} -syn NH ₂ -(C)VDPKTLQRQQAGGDGTGSS |
| 67 | TcaC-syn NH ₂ -(C)YKAPQRQEDGDSNAVTYDK |
| 68 | TcbA _{iii} -syn NH ₂ -(C)YNENPSSSEDKKWFSSKDD |
| 69 | TcbA _{iii} -syn NH ₂ -(C)FDSYSQLYEENINAGEQRA |
| 70 | TcdA _{iii} -syn NH ₂ -(C)NPNNSSNKLMFPVYQYSGNT |
| 71 | TcdA _{iii} -syn NH ₂ -(C)VSQGSAGSGNNNLAFGAG |

각각의 컨주게이트된 합성 펩티드를 제네미드 (Genemed) 가속화 프로그램에 따라 2마리의 토끼에 주입하였다. 1개월 후 검사를 위한 프리 및 포스트 면역 혈청을 얻을 수 있었다.

각각의 토끼로부터 얻은 프리 및 포스트 면역 혈청 모두의 1차 검사는 제네미드 바이오테크놀로지 인크.가 수행하였다. 제네미드 (Genemed)는 ELISA 및 웨스턴 블롯 기술을 사용하여 각각의 합성 펩티드에 대한 포스트 면역 혈청의 반응을 검색하였다고 보고하였다. 이어서, 혈청을 웨스턴 블롯 분석에 따라 전체 표적 펩티드로 검사하였다. 부분적으로 정제된 포도라드부스 균주 W-14 독소 복합체의 2개 밴치를 항원으로 사용하였다. 2개의 샘플은 썬던 옥수수 뿌리 벌레에 대해 활성을 보였다. SDS-PAGE 겔에 대한 그들의 패턴은 약간 달랐다.

4-20% 구배를 갖는 프리-캐스트 SDS-폴리아크릴아미드 겔 (Integrated Separation Systems, 매사추세츠주 01760 나틱)을 사용하였다. 1 내지 8 µg의 단백질을 각각의 겔 웰에 적용하였다. 전기영동을 실시하고 단백질을 Hybond-ECL (상표명) 니트로셀룰로오즈 막 (Amersham International) 상에 일렉트로블롯팅하였다. 막을 실온에서 1 시간 동안 TBST (25 mM Tris HCl, pH 7.4, 136 mM NaCl, 2.7 mM KCl, 0.1% Tween 20) 중의 10% 우유로 블로킹하였다. 각각의 토끼의 혈청을 10% 우유/TBST 중에서 1:500으로 희석시켰다. 1:50 내지 1:1000의 다른 희석도를 또한 사용하였다. 혈청을 막에 첨가하고, 1 시간 이상 플렛트포움 진탕기 상에 두었다. 막을 블로킹 용액 또는 TBST로 완전히 세척하였다. 10% 우유/TBST 중의 2차 항원 (호스 래디쉬 퍼옥시다아제에 컨주게이트된 염소 항-쥐 IgG)을 1 시간 동안 플렛트포움 진탕기 상에 둔 막에 적용하였다. 이어서 막을 과량의 TBST로 세척하였다. 단백질의 검색은 ECL (증강 화학형광) 검색 키트 (Amersham International)을 사용하여 실시하였다.

웨스턴 블롯 분석을 실시하여 각각의 항-합성 펩티드 항체의 결합 특이성을 확인하였다. 모든 합성 다클론성 항체는 포도라드스 배양 브로쓰로부터 유도된 단백질 분획으로부터 가공된 표적 펩티드 및 가공되지 않은 표적 펩티드 (사용가능한 경우)에 대해 특이성을 보였다. 다양한 항체가 바쿨로스바이러스 발현 구조를 사용하여 박테리아 또는 곤충 세포와 같은 이중 발현 시스템으로부터 유도된 가공되거나 가공되지 않은 재조합 단백질을 인지하는 것으로 보인다. 어느 경우에는 항-TcbA_{iii}-syn 항체가 항-TcdA_{iii} 펩티드에 대해 약간의 교차 반응성을 보였다. 두번째 경우에는, 항-TcaC-syn 항체가 W-14 독소 복합체 분획 중의 확인되지 않은 190 kDa 펩티드를 인지하였다.

<실시예 22>

포도라드스 균주의 특성화

본원에 기재된 수집물이 포도라드스 균주를 포함함을 확인하기 위해 본원의 균주를 박테리아 속 포도라드스의 특성이고 엔테로박테리아네 및 제노라드스 에스피피.와 다른 인지된 미생물학적 특징 면에서 평가했다 (Farmer, J.J. 1984. Bergey's Manual of Systemic Bacteriology, Vol 1, pp. 510-511. (ed. Kreig N. R. and Holt, J.G.). Williams & Wilkins, Baltimore.; Akhurst and Boemare, 1988, J. Gen. Microbiol. 134, 1835-1845; Forst and Nealon, 1996. Microbiol. Rev. 60, 21-43). 이들 특징적인 성질은 다음과 같다: 그람 염색 음성 로드, 생물 크기, 폭 0.3 - 2 µm 및 길이 2 - 10 µm [필라멘트 (15 - 50 µm)와 구형형상이 나타남], 영양 한천 배양기 상의 황색 내지 오렌지색/적색 군체 착색, 결정질 봉입체 존재, 촉매 존재, 질산염 환원 불가능, 생물발광 존재, 생장 배지로부터 염료 흡수 가능, 프로테아제 생성에 대해 양성, 37°C 이하 온도에서 생장, 혐기성 조건 하에서 생존, 자력 생존 가능 (표 33). 검사 방법은 에스케리키아 콜리, 제노라드스 및 포도라드스 균주를 사용하여 체크하였다. 전체 결과는 장박테리아 과 및 포도라드스 속 에 속하는 모든 균주에 대해 일정하였다. DEP1, DEP2 및 DEP301 American Type Culture Collection 12301 Parklawn Drive, Rockville, MD 20852 USA (#29304, 29999 및 51583)으로부터 얻은 포도라드스 균주를 참고함을 주목해야 한다.

발광계를 사용하여 이들 포도라드스 균주와 관련된 생물 발광을 확인하였다. 상대 발광 단위의 존재 또는 부재를 측정하기 위해 각각의 균주 (세포 및 배지)로부터 얻은 브로쓰를 액상 배양액 중에 접종한 3회 간격을 두고 측정하고 (24, 48, 72 시간) 배경 발광 (접종하지 않은 배지)에 비교하였다. 몇몇 제노라드스 균주를 발광의 음성 대조군으로 검사하였다. 다양한 브로쓰로부터의 발광을 측정하기 전에, 시퍼 세포를 사용하여 길포드 시스템스 (Gilford Systems, Oberlin, 오하이오주) 스펙트로포토미터에서 흡광을 측정하여 (560 nm) 세포 밀도를 확인하였다. 이어서, 생성된 발광 단위를 세포의 밀도에 대해 표준화하였다. 브로쓰의 분취량을 96-웰 미세 적정기 플레이트 (100 µl 각각)에 두고 Packard Lumicount (상표명) 발광계 (Packard Instrument Co., 코네티컷주 메리덴)에서 판독하였다. 각 샘플에 대한 측정 기간은 0.1 내지 1.0초였다. 샘플을 판독하

기 전에 10초 동안 발광기에서 교반하였다. 양성 검사는 5-배 배경 발광으로 결정하였다 (약 1-15 상대광 단위). 그 밖에, 군체 발광도를 포토그래픽 필름 오버레이와 암실에서 적응된 후의 육안으로 확인하였다. 각 균주의 그램스 착색 특성화를 그램스 착색 대조 슬라이드 (Fisher Scientific, 펜실베니아주 피츠버그)와 함께 사용하는 시판중인 그램스 착색 키트 (BBL, 메릴랜드주 코키스빌)로 확인하였다. 이어서, Zeiss 현미경 (Carl Zeiss, 독일) 100X 오일 침지 대물 렌즈 (10X 대안 및 2X 대물 확대)를 사용하여 현미경 평가를 실시하였다. 생물 크기, 세포 형태 및 봉입체에 대한 개별 균주의 현미경 검사 (세포 형태와 봉입체는 대수적 성장 후 관찰)는 습식 장착 슬라이드 (10X 대안, 2X 대물 및 40X 대물 확대) 및 파아지 대조 현미경법을 사용하여 μ m로 실시하였다 (Akhurst, R. J. and Boemare, N.E. 1990. *Entomopathogenic Nematodes in Biological Control* (ed. Gaugler, R. and Kaya, H.). pp 75-90. CRC Press, Boca Raton, USA.; Baghdiguin S., Boyer-Giglio M. H., Thaler, J.O., Bonnot G., Boemare N. 1993. *Biol. Cell* 79, 177-185).

군체 착색은 라벨 지시에 따라 제조한 Bacto 영양 한천 배지 (Difco Laboratories, 미시간주 디트로이트) 상에 접종한 후 관찰하였다. 28°C에서 인큐베이션시키고 5일에 품목을 생산하였다. 효소 촉매의 존재를 검사하기 위해 검사 생물의 군체를 영양 한천 배지 플레이트로부터의 작은 플러그 상에 제거하고 유리 검사관의 바닥에 두었다. 가정용 과산화수소 용액 1 ml를 튜브 벽을 따라 부드럽게 첨가하였다. 양성 반응을 기포 (산소로 추정)가 나타나는 직후 또는 5초 내에 기록하였다. 접종하지 않은 영양 한천 배지 및 과산화수소 용액의 대조군을 또한 검사하였다. 니트레이트 환원을 검사하기 위해 각각의 배양액을 Bacto Nitrate 브로쓰 (Difco Laboratories, 미시간주 디트로이트) 10 ml에 접종하였다. 28°C에서 부드럽게 교반하며 24 시간 동안 인큐베이션시킨 후, 황산 시약 2 방울 및 알파-나프틸아민 시약 2 방울을 첨가하여 니트레이트 생성을 검사하였다 (Difco Manual, 10th edition, Difco Laboratories, 미시간주 디트로이트, 1984 참조). 뚜렷한 분홍색 또는 적색이 발생하는 것은 니트레이트로부터 니트릴이 형성된 것을 나타내는 한편 색 형성이 부족한 것은 균주가 니트레이트 환원 음성을 나타낸다. 나중 경우에, 미분말 아연을 첨가하여 환원되지 않은 니트레이트의 존재를 추가 확인하고; 니트레이트의 형성 및 적색을 확인하였다. 각 균주가 성장 배지로부터 영양을 흡수하는 능력을 천연 적색 영양을 함유한 Bacto MacConkey 한천 배지; 브로모티몰 청색 영양을 함유한 Bacto Tergitol-7 한천 배지 및 eosin-Y 영양을 함유한 Bacto EMB 한천 배지 (Difco Laboratories (미시간주 디트로이트)로부터 배합한 한천 배지, 모두 라벨 지시에 따라 제조)로 검사하였다. 이들 배지 상에 접종한 후, 28°C에서 5 시간 동안 인큐베이션한 후 영양 흡수를 기록하였다. 이들 배지 상에서의 생장은 장 박테리아 과의 생물의 특징이다. 각 균주의 자동력은 라벨 지시에 따라 제조된 Bacto 자동력 검사 매질 (Difco Laboratories, 미시간주 디트로이트)의 용액을 사용하여 검사하였다. 각각의 균주로 버트-스태브 접종을 실시하고 접종 선으로부터 확산된 생장의 전파 영역으로 자동력을 거시적으로 판정하였다. 프로테아제의 생성은 라벨 지시에 따라 제조된 Bacto 젤라틴 (Difco Laboratories, 미시간주 디트로이트)를 사용하여 젤라틴의 가수분해를 관찰함으로써 검사하였다. 배양액을 접종하고 튜브 또는 플레이트를 28°C에서 5 일 동안 인큐베이션시켰다. 이어서, 젤라틴 가수분해를 실온, 즉 22°C 미만에서 체크하였다. 여러가지 온도에서 생장을 평가하기 위해, 한천 배지 플레이트 [타리온수 내 Bacto-Agar (Difco, 미시간주 디트로이트) 2%를 갖는 2% 프로테아제 펄론 #3]를 통상적인 접종 원으로부터 스트리킹하였다. 플레이트를 20, 28 및 37°C에서 3 일 이하로 인큐베이션시켰다. 인큐베이터 온도 수준을 전기 열전쌍으로 체크하고 확실한 온도 세팅을 위해 계측하였다. 포도람두스 균주를 위한 산소 조건은 하기 방식으로 검사하였다. 유체 티오글리콜레이트 브로쓰 배지 (Difco, 미시간주 디트로이트)내로 버트-스태브 접종하였다. 튜브를 실온에서 1주 동안 인큐베이션하고 배양액을 성장 유형 정도에 대해 시험하였다. 인디케이터 레사주린은 배지 산절단 또는 호기성 구역이 존재함을 입증한다 (Difco Manual, 10th edition, Difco Laboratories, 미시간주 디트로이트). 검사한 포도람두스 균주에 대해 얻어진 성장 구역 결과는 성장가능한 혐기성 미생물의 것과 일치하였다. 결과가 불확실한 경우, 유체 티오글리콜레이트 브로쓰 배지의 최종 한천 배지 농도를 0.75%로 올리고 성장 특성을 다시 체크하였다.

[표 35]

포토라투스 균주의 분류학적 특징

| 균주 | A* | B | C | D | E | F | G | H | I | J [†] | K | L | M | N | O | P | Q |
|----------|----|---|---|------|---|---|---|---|---|----------------|---|---|---|---|---|---|---|
| 외. 제알란디카 | -† | + | + | rd S | + | - | + | + | + | PO | + | + | + | + | + | + | - |
| 외. 헤피알투스 | - | + | + | rd S | + | - | + | + | + | Y | - | + | + | + | + | + | - |
| HB-Arg | - | + | + | rd S | + | - | + | + | + | W | + | + | + | + | + | + | - |
| HB 오스웨고 | - | + | + | rd S | + | - | + | + | + | W | + | + | + | + | + | + | - |
| HB 레위스톤 | - | + | + | rd S | + | - | + | + | + | I | + | + | + | + | + | + | - |
| K-122 | - | + | + | rd S | + | - | + | + | + | Y | + | + | + | + | + | + | - |
| HMGD | - | + | + | rd S | + | - | + | + | + | Kd | + | + | + | + | + | + | - |
| 인디커스 | - | + | + | rd S | + | - | + | + | + | W | - | + | + | + | + | + | - |
| GD | - | + | + | rd S | + | - | + | + | + | YI | + | + | + | + | + | + | - |
| PWH-5 | - | + | + | rd S | + | - | + | + | + | Y | + | + | + | + | + | + | - |
| 메기타스 | - | + | + | rd S | + | - | + | + | + | R | - | + | + | + | + | + | - |
| RF-85 | - | + | + | rd S | + | - | + | + | + | R | + | + | + | + | + | + | - |
| 에이 카우스 | - | + | + | rd S | + | - | + | + | + | PR | + | + | + | + | + | + | - |
| MF1 | - | + | + | rd S | + | - | + | + | + | I | - | + | + | + | + | + | - |
| MF2 | - | + | + | rd S | + | - | + | + | + | I | - | + | + | + | + | + | - |
| MF3 | - | + | + | rd S | + | - | + | + | + | R | - | + | + | + | + | + | - |
| MF4 | - | + | + | rd S | + | - | + | + | + | Y | - | + | + | + | + | + | - |
| MF5 | - | + | + | rd S | + | - | + | + | + | PR | + | + | + | + | + | + | - |
| GL98 | - | + | + | rd S | + | - | + | + | + | W | + | + | + | + | + | + | - |
| GL101 | - | + | + | rd S | + | - | + | + | + | W | + | + | + | + | + | + | - |
| GL138 | - | + | + | rd S | + | - | + | + | + | W | + | + | + | + | + | + | - |
| GL155 | - | + | + | rd S | + | - | + | + | + | W | + | + | + | + | + | + | - |
| GL217 | - | + | + | rd S | + | - | + | + | + | Y | - | + | + | + | + | + | - |
| GL257 | - | + | + | rd S | + | - | + | + | + | O | + | + | + | + | + | + | - |
| DEP1 | - | + | + | rd S | + | - | + | + | + | W | + | + | + | + | + | + | - |
| DEP2 | - | + | + | rd S | + | - | + | + | + | PR | - | + | + | + | + | + | - |
| DEP3 | - | + | + | rd S | + | - | + | + | + | CR | + | + | + | + | + | + | - |

* - A=그람 균주, B=결정 봉입체, C=생체 발광, D=세포 형태, E=운동성, F=질산염 환원, G=카탈라제의 존재, H=젤라틴 가수분해, I=영양 흡수, J=착색, K=EMB 아가 상에서의 성장, L=MacConKey 아가 상에서의 성장, M=Tergitol-7 아가 상에서의 성장, N=동성 혐기성균, O=20°C에서의 성장, P=28°C에서의 성장, Q=37°C에서의 성장, † - +/- = 특징의 양성 또는 음성, rd=간균, S=속 범위 내의 크기, W=백색, CR= 크림색, Y=황색, YI=담황갈색, T=갈색, PO=엷은 오렌지색, O=오렌지색, PR=엷은 적색, R=적색.

본 발명의 컬렉션에서의 포토라투스 균주의 진화적 다양성을 각 균주의 게놈 DNA를 사용하여 PCR (폴리머라제 연쇄 반응) 매개 게놈 핑거프린팅의 분석에 의해 측정하였다. 이 기술은 다양한 세균 종의 게놈 전반에 존재하는 일군의 반복 DNA 서열을 기초로 한 것이다 (Versalovic, J., Schneider, M., DE Bruijn, F.J. and Lupski, J.R. 1994. Methods Mol. Cell. Biol., 5, 25-40.). 이들 서열 중 반복 유전자와 회문 서열 (REP), 장내 세균 반복 유전자내 공통 서열 (ERIC) 및 BOX 성분은 세균 게놈의 조직에 중요한 역할을 수행하는 것으로 생각된다. 게놈 조직은 선별에 의해 형성화되는 것으로 생각되고 밀접하게 관련된 세균 균주의 게놈 내의 이들 성분의 상이한 분산은 이들 균주의 식별에 사용될 수 있다 (예를 들면, Louws, F.J., Fulbright, D.W., Stephens, C.T. and DE Bruijn, F.J. 1994. Appl. Environ. Micro. 60, 2286-2295.). Rep-PCR은 반복서열 사이에 존재하는 크기가 상이한 DNA 단편을 증폭하기 위해 상기 반복 서열에 상보적인 올리고뉴클레오티드 프라이머를 이용한다. 생성물은 각 균주에 대한 DNA "핑거프린트"를 확립하기 위해 전기영동에 의해 분리한다.

본 발명의 균주로부터 게놈 DNA를 단리하기 위해서 세포 펠렛을 최종 용량 10 ml이 되도록 TE 완충액 (10 mM Tris-HCl, 1 mM EDTA, pH8.0) 중에 재현탁시키고 5 M NaCl 12 ml를 첨가하였다. 이 혼합물을 15,000 x g에서 20분 동안 원심분리하였다. 생성 펠렛을 TE 5.7 ml 및 10% SDS 300 µl 중에 재현탁시키고 20 mg/ml 프로테이나제 K 60 µl (Gibco BRL 제품, Grand Island, NY)을 첨가하였다. 이 혼합물을 37°C에서 1시간 동안 배양한 후 리소자임 약 10 mg을 첨가하고 혼합물을 추가로 45분 동안 배양하였다. 5 M NaCl 1 ml 및 CTAB/NaCl 800 µl 용액 (10% w/v CTAB, 0.7 M NaCl)을 첨가하고 혼합물을 65°C에서 10분 동안 배양하고, 부드럽게 진탕시킨 후 배양하고 추가로 20분 동안 진탕시켜 세포 물질의 클리어링을 도왔다. 동량의 클로로포름/이소아밀 알콜 용액 (24:1, v/v)을 첨가하고 부드럽게 혼합한 후 원심분리하였다. 이어서, 동용량의 페놀/클로로포름/이소아밀 알콜 (50:49:1)을 사용하여 2회 추출하였다. 게놈 DNA를 이소프로판올 0.6 용량을 사용하여 침전시켰다. 침전된 DNA를 유리 막대기로 수거하여 70% 에탄올로 2회 세척하고 건조시켜 STE 2 ml (10 mM Tris-HCl pH8.0, 10 mM NaCl, 1 mM EDTA) 중에 용해시켰다. 이어서, 260 nm에서의 광학 밀도에 의해 DNA를 정량하였다. 포토라투스 게놈 DNA의 rep-PCR 분석을 실시하기 위해서 다음과 같은 프라이머를 사용하였다: REP1R-1; 5'-IIICGICGICATCIGGC-3' 및 REP2-1; 5'-ICGICTTATCIGGCCTAC-3'. 다음과 같은 25 µl 반응물을 사용하여 PCR을 실시하였다: 7.75 µl H₂O, 2.5 µl 10X LA 완충액 (PanVera Corp., Madison, WI), 16 µl dNTP 믹스 (각각 2.5 mM), 각각의 프라이머 1 µl (50 pM/µl), DMSO µl, 게놈 DNA 1.5 µl (농도 0.075-0.480 µg/µl) 및 0.25 µl TaKaRa EX Taq (PanVera Corp., Madison, WI). Perkin Elmer DNA 열 순환기 (Norwalk, CT) 중에서 다음과 같은 조건을 사용하여 PCR 증폭을 실시하였다: 95°C/7분, 이어서 94°C/1분, 44°C/1분, 65°C/8분의 싸이클의 35회 반복, 이어서 65°C에서 15분. 순환 후에 반응물 25 µl를 6X 겔 로딩 완충액 (물 중의 0.25% 브로모페놀 블루, 40% 수크로스) 5 µl에 첨가하였다. 이어서, 15x20 cm 1% 아가로스 겔을 각각의 반응액 8 µl를 사용하여 TBE 완충액 (0.09 M Tris-보레이트, 0.002 M EDTA) 중에서 전개시켰다. 겔을 45V에서 약 16시간 동안 전개시켰다. 이어서, 겔을 에티듐 브로마이드 20 µg/ml 중에서 1시간 동안 염색시키고 TBE 완충액 중에서 약

3시간 동안 탈염색시켰다. 겔의 플라로이드 사진을 UV 조명 하에 인화하였다.

각 염색에 특이적인 크기에서의 밴드의 존재 여부를 사진으로부터 기록하고 수치 분류 소프트웨어 프로그램인 NTSYS-pc (Exeter Software, Setauket, NY)에 유사성 매트릭스로서 입력하였다. 동시에 분석되는 이. 콜라이 균주 HB101 및 잔토모나스 오리재 피브이. 오리재 (*Xanthomonas oryzae* pv. *oryzae*)의 대조군은 출판된 문헌에 따른 "핑거프린트"를 생성시켰다 (Versalovic, J., Koeuth, T. and Lupski, J.R. 1991. *Nucleic Acids Res.* 19, 6823-6831; Vera Cruz, C.M., Haldal-Alija, L., Louws, F., Skinner, D.Z., George, M.L., Nelson, R.J., DE Bruijn, F.J., Rice, C. and Leach, J.E. 1995. *Int. Rice Res. Notes*, 20, 23-24.; Vera Cruz, C.M., Ardales, E.Y., Skinner, D.Z., Talag, J., Nelson, R.J., Louws, F.J., Leung, H., Mew, T.W. and Leach, J.E. 1996. *Phytopathology*). 이어서, NTSYS-pc 내의 일련의 프로그램; 매트릭스의 유사성 계수 (Jaccard 계수를 사용)를 생성하기 위한 SIMQUAL (정량 데이터에 대한 유사성) 및 관련 균주를 한 데 모으고 페노그램 (phenogram) (도 5)으로 표현될 수 있는 SAHN (연속적인, 응집성의, 헤라키칼 (Heirarchical) 및 네스티드 (Nested)) 클러스터링 [UPGMA (산술평균을 사용하는 비계량 페어-그룹 (Unweighted Pair-Group) 방법을 사용) 방법을 사용하여 포토랍두스 균주로부터의 데이터를 분석하였다. COPH (코페네틱 (cophenetic) 값) 및 MXCOMP (매트릭스 비교) 프로그램을 사용하여 코페네틱 값 매트릭스를 생성시켜 이것과 클러스터링의 기초가 되는 본래의 매트릭스 사이의 관계를 비교하였다. 클러스터 분석에 적합한 양호함의 척도인 ($r=0.8-0.9$ 가 매우 양호하게 적합함을 나타냄) 생성되는 정상화 만텔 (Mantel) 통계치 (r)을 생성시켰다. 본 발명의 경우에 r 은 0.919이 었다. 따라서, 본 발명의 컬렉션은 포토랍두스 속의 대표적인 용이하게 식별가능한 균주의 다양 한 군으로 이루어진 것이다.

<실시예 23>

다양한 포토랍두스 균주에 의해 생산되는 독소(들)의 살충 효용

카뮷으로 덮여진 델롱 목을 가진 500 ml 트리배플드 플라스크중의, 2% 프로테오스 펩톤 #3 (PP3, 미시간주 디트로이트 소재의 디프코 래버러토리즈 (Difco Laboratories) 제조) 액체 배지 175 ml 에 일차 변종 서브클론을 접종하여, 다양한 포토랍두스 균주의 초기 종자 배양물을 제조하였다. 각 종자 배양을 위한 접종물은 오일이 덮여진 아가 사면 배양물 또는 플레이트 배양물로부터 유래 되었다. 접종 후, 상기 플라스크를 회전식 진탕기 상에서, 150 rpm에서 16 시간 동안 28°C에서 배양하였다. 이어서 상기 종자 배양물은 각 균주의 정해진 발효를 위한 균일한 접종원으로서 사용되었다. 또한, 로그 후상 (post-log) 종자 배양물을 멸균의 미세할 오일로 덮고, 멸균의 자석 교반 막대를 장래의 재현탁을 위하여 첨가하며, 배양물을 암실, 실온에서 보관함으로써, 접종물을 독소 생산 상태로 장기간 보존하였다. 활발하게 성장하는 종자 배양물 1% (예를 들어, 신선 배양 물 175 ml 당 1.75 ml)를 2% PP3 신선 배지에 첨가함으로써 생산 브로쓰를 접종하였다. 브로쓰 는, 실리콘 담용 마개와 밀폐된 500 ml 트리배플드 플라스크 (상기 참조) 또는 2800 ml 배플드 요 칠면 바닥 플라스크 (500 ml 용량)에서 생산하였다. 생산 플라스크를 상기 조건 하에서 24 내지 72시간 동안 배양하였다. 배양 후, 브로쓰를 멸균의 1 L 폴리에틸렌 병 내로 분배시켜 10°C에서 1 시간 동안 2600 × g에서 회전시킴으로써, 세포 및 파쇄물 펠렛으로부터 분리시켰다. 액체 브로 쓰는 그 후 와트만(Whatman) GF/D (2.7 μm 보유) 및 GF/B (1.0 μm 보유) 유리 여과기를 통하여 진공 여과시켜 파쇄물을 제거하였다. 0.5 μm 개방 도관 여과기를 사용, 점선 유출 극소 여과 장 치 (폴 필트론 (Pall Filtron), 매사추세츠주 노스보로)로 브로쓰를 보다 더 정화하였다. 필요 시, 브로쓰를 냉각시켜(4°C까지) 2600 × g에서 여러 시간 동안 원심분리시킴으로써 추가로 정화 시킬 수 있다. 상기 과정 후, 0.2 μm 니트로셀룰로스 막 여과기를 사용하여 브로쓰를 여과 살균 시켰다. 무균 브로쓰는 그 후, 생물학적 분석, 생화학적 분석용으로 직접 사용되었거나 분자량 10000 컷오프 M12 한외여과 장치 (Amicon, 매사추세츠주 베벌리) 또는 분자량 10000의 구멍 크기를 갖는 원심 분리 농축기 (Millipore, 매사추세츠주 베드포드)를 사용하여 농축 (10배까지)시켰다. 원심 분리 농축기의 경우, 브로쓰를 2000 × g에서 약 2시간 동안 원심분리시켰다. 분자량 10000 투과액을, 상응하는 잔류물에 첨가하여 분자량 10000 초과 성분의 원하는 농축액을 수득하였 다. 모래를 채운 히트 블록 (heat block)에서 10분간, 100°C에서 샘플을 가열함으로써, 처리되 는 브로쓰 샘플을 가열 불활성화시켰다.

서로 다른 포토랍두스 균주로부터의 브로쓰(들) 및 독소 복합체(들)은 곤충 집단을 감소시키는 데 유용하며, 기술된 활성물질들, 곤충 불활성화에 효과적인 양만큼 곤충 서식지에 사용하는 것으로 이루어지는 곤충 집단 억제 방법에 사용되었다. 상기와 같이 발효시킨 포토랍두스 균주의 선별균 의 브로쓰로부터 관찰된 살충 활성의 범위의 증거는 표 36에 나타나 있다. 브로쓰 농도를 증가시 키거나 다른 발효 방법을 사용함으로써 상기 균주에서 부가적인 살충 활성의 검출이 가능하다. 단백질과 관련된 활성과 연관되게, 시험된 모든 균주들의 살충 활성은 열에 불안정하였다.

다양한 포토랍두스 균주로부터의 배양 브로쓰(들)은 곤충 수에 대하여 차별적인 살충 활성 (치사 율 및(또는) 성장 저해율, 성체 출현의 감소)을 나타낸다. 보다 구체적으로, 활성은 딱정벌레 (Coleoptera) 곤충 목의 일원인 옥수수 뿌리벌레 유충에 대하여 나타난다. 딱정벌레 목의 다른 일원으로는 목화 바구미, 노래기, 방아벌레, 화분 딱정벌레, 웹벼룩 갑충, 종자 갑충 및 콜로라도 감자 딱정벌레가 있다. 브로쓰 및 정제된 독소 복합체(들)은 또한 나비목의 일원인 회색담배나 일 방, 박각시나방 및 유럽 옥수수 천공충에 대해서도 활성을 갖는다. 상기 목의 다른 전형적인 일 원으로는 사탕무 행렬구더기, 양배추 자벌레, 블랙 야도충, 옥수수 귀벌레, 사과 나방, 반대충, 화랑곡나방, 잎말이나방, 배추 벌레, 목화다래나방, 도롱이벌레, 이스턴 텐트 나방, 잔디 포충나 방 및 풀 행렬구더기가 있다. 활성은 디티오프테라 (또는 블라토데아) 목의 일원인 독일 바퀴벌 레에 대해서도 관찰된다. 이 목의 다른 일원은 동양 바퀴벌레 및 어메리칸 바퀴벌레가 있다.

옥수수 뿌리벌레 유충에 대한 활성은 하기와 같이 시험되었다. 포토랍두스 배양 브로쓰(들)(0 내 지 15배 농축, 여과 살균), 2% 프로테오스 펩톤 #3, 정제 독소 복합체(들)(0.23 mg/ml) 또는 pH 7.0의 10 mM 인산 나트륨 완충액을 인공 먹이의 표면 (약 1.5 cm²)에 40 μl씩 직접

도포하였다[Rose, R. I. 및 McCabe, J. M. (1973). J. Econ. Entomol. 66, (389-400) 참조]. 독소 복합체를 pH 7.0의 10 mM 인산 나트륨 완충액에 희석시켰다. 먹이 플레이트를 무균 플로우 후드에서 공기 건조시키고, 표면 살균된 알로부터 부화시킨 단일, 신생 디아브로티카 운데심푼크타 호와르디 (*Diabrotica undecimpunctata howardi*) (써든 옥수수 뿌리벌레, SCR)로 웰을 만연시켰다. 상기 플레이트를 봉인하여 습윤 성장 챔버에 두고 적절한 기간 (3 내지 5일)동안 27°C에서 유지시켰다. 그 후 치사율 및 유충 무게 측정치를 기록하였다. 일반적으로, 모든 연구에서 처리당 16마리의 곤충이 사용되었다. 대조군 치사율은 일반적으로 5% 미만이었다.

나비목 유충에 대한 활성은 하기와 같이 시험하였다. 농축(10배) 포토랍두스 배양 브로쓰(들), 대조군 배지 (2% 프로테오스 펩톤 #3), 정제 독소 복합물(들)(0.23 mg/ml) 또는 pH 7.0의 10 mM 인산 나트륨 완충액을, 표준 인공 나비목 먹이(스톤빌 옐로우 먹이, 약 1.5 cm²) 표면에 40 µl씩 직접 도포시켰다. 먹이 플레이트를 무균 플로우 후드내에서 공기 건조시켰고, 각 웰을 신생 유충 1마리로 만연시켰다. 유럽 옥수수 천공충 (*Ostrinia nubilalis*) 및 박각시나방 (*Manduca sexta*) 알을 구입하여 부화시켰고, 회색담배나방 (*Heliothis virescens*) 유충은 내부적으로 공급하였다. 유충으로 만연시킨 후, 먹이 플레이트를 봉인하여 습윤 성장 챔버내에 두고 적절한 기간 동안 27°C 암실에서 유지시켰다. 치사율 및 중량 측정은 5일째에 실시하였다. 일반적으로, 모든 연구에서 처리당 16마리의 곤충이 사용되었다. 대조군 배지의 대조군 치사율은 일반적으로 0 내지 12.5%였고, 인산 완충액의 대조군 치사율은 10% 미만이었다.

바퀴벌레에 대한 활성은 하기와 같이 시험하였다. 농축(10배) 포토랍두스 배양 브로쓰(들), 대조군 배지 (2% 프로테오스 펩톤 #3)을, 표준 인공 나비목 먹이(스톤빌 옐로우 먹이, 약 1.5 cm²) 표면에 40 µl씩 직접 도포시켰다. 먹이 플레이트를 무균 플로우 후드내에서 공기 건조시켰고, 각 웰을 CO₂ 마취시킨 제1 인스타 독일 바퀴벌레 (*Blattella Germanica*)로 만연시켰다. 유충으로 만연시킨 후, 먹이 플레이트를 봉인하여 습윤 성장 챔버내에 두고 적절한 기간 동안 27°C 암실에서 유지시켰다. 치사율 및 중량 측정은 5일째에 실시하였다. 일반적으로, 모든 연구에서 처리당 16마리의 곤충이 사용되었다. 대조군 치사율은 10% 미만이었다.

[표 36]

상이한 포토랍두스 균주로 부터 브로쓰의 살충 스펙트럼 관찰

| 포도랍두스 균주 | 감수성*의 곤충종 |
|----------|-----------|
| 피. 제알란디카 | 1**, 2, 4 |
| 피. 해피알루스 | 1, 2, 4 |
| HB-Arg | 1, 2, 4 |
| HB 오스웨고 | 1, 2, 4 |
| HB 레위스톤 | 1, 2, 4 |
| K-122 | 1, 4 |
| HMGD | 1, 4 |
| 인디커스 | 1, 2, 4 |
| GD | 2, 4 |
| PWH-5 | 1, 2, 4 |
| 메기디스 | 1, 2, 4 |
| HF-85 | 1, 2, 4 |
| 에이. 카우스 | 1, 4 |
| MP1 | 1, 2, 4 |
| MP2 | 1, 2, 4 |
| MP3 | 4 |
| MP4 | 1, 4 |
| MP5 | 4 |
| GL98 | 1, 4 |
| GL101 | 1, 4, 5 |
| GL138 | 1, 2, 4 |
| GL155 | 1, 4 |
| GL 217 | 1, 2, 4 |

키고, 기포형 물질이 맑아지도록 약 20분 동안 교반시켰다. 클로로포름/이소아밀 알콜 용액 (24:1 v/v)의 등가 용적을 첨가하고, 매우 완만하게 혼합하고, 12,000 x g에서 15분 동안 원심분리하여 상층을 분리하였다. 상층 (수성 상)을 넓은 보어 피펫으로 완만하게 제거하고, PCI (페놀/클로로포름.이소아밀 알콜; 50:49:1, v/v/v; 1 M 트리스-HCl (pH 8.0, Intermountain Scientific Corporation, Kayville, UT)으로 평형시킴)의 등가 용적으로 상기와 같이 2회 추출하였다. 0.6 용적의 이소프로판올로 침전시킨 DNA를 유리막대로 완만하게 제거하고, 70 % 에탄올로 2회 세척하고, 2 ml STE (10 mM 트리스-HCl, 10 mM NaCl, 1 mM EDTA, pH 8) 중에 용해시켰다. 이 침전물은 260 nm에서 광학밀도로 측정할 때 2.5 mg/ml를 함유하였다.

[tc-유전자 특이적 프로브로 하이브리드되는 Bgl II/Hind III 단편의 동정]

게놈 DNA 약 10 µg를 Bgl II/Hind III (NEB)의 각각의 약 30 단위로 완결되게 180분 동안 분해시키고, 밤새 동결시키고, 이어서 65 °C에서 5분 동안 가열하고, 0.8 %의 아가로제 젤 (등록상표 Seakem LE, 1X TEA, 80 볼트, 90분) 중에 전기영동시켰다. DNA를 전술한 바와 같이 브롬화에티듐 (50 µg/ml)으로 염색시키고, 자외선하에 사진을 찍었다. 아가로제 젤 중의 DNA 단편을 정화시키고 (0.2 M HCl 중에서 5분), 변성시키고 (0.5 M NaOH, 1.5 M NaCl 중에서 15분), 중화시키고 (0.5 M 트리스-HCl, pH 8.0, 1.5 M NaCl에서 15분), 각 단계 사이에 증류수로 3회 행구었다. DNA를 오수벨 (Ausubel, CPMB, op.) 등의 책션 2.9에 기재된 바와 같이 고급 염 (20X SSC) 프로토콜을 사용하여 써던 블롯으로 겔에서 NYTRAN 나일론 막 (Amersham, Arlington Heights, IL)로 옮겼다. 이어서, 이동시킨 DNA를 자동 가교에 맞추어진 스트라타겐 (Stratagene) UV 스트라타링커(Stratalinker)를 사용하여 나일론 막과 UV 가교시켰다. 막들을 사용할 때까지 25 °C에서 건조 저장시켰다.

하이브리다이제이션을 ECL (상표명) 디렉트 (Amersham, Arlington Heights, IL) 라벨링 및 제조자에 의해 제공된 프로토콜을 따라 검출 시스템을 사용하여 수행하였다. 간단하게는, 프로브를 효소 호스 래디쉬 퍼옥시다제에 변성된 DNA를 공유 결합시켜 제조하였다. 일단 라벨링된 프로브를 효소 활성을 유지하는 하이브리다이제이션 조건하에 사용하였다. 0.5xSSC, 0.4 % SDS 및 6 M 우레아 중에 42 °C에서 각각 20분 동안 2회 부드럽게 세척하여 하이브리드되지 않은 프로브를 제거하였다. 이어서, 이것을 2xSSC 중에 실온에서 각각 5분 동안 2회 세척하였다. 제조자에 의해 지시된 바와 같이, ECL 시약을 사용하여 하이브리드된 DNA 밴드를 검출하였다. 각종 포도랍두스 균주 및 균주 W-14 사이에 유전자 관계를 검출하기 위한 능력에 영향을 주는 몇몇 인자들이 있다. 첫째, 매우 엄격한 조건을 이들 하이브리다이제이션에 사용하였다. 하이브리다이제이션의 엄격성 및 세척 조건을 변화시키는 것은 하이브리다이제이션 밴드의 패턴 및 강도에 영향을 주는 것으로 공지되어 있다. 둘째, 써던 블롯의 블롯 대 블롯 변형은 하이브리다이제이션 밴드 및 분자량 측정치의 유동성에 영향을 준다. 따라서, W-14는 모든 써던 블롯 상의 표준물로서 포함되었다.

W-14 독성 유전자로부터 유도된 유전자 특이적 프로브를 이들 하이브리다이제이션에 사용하였다. 하기에 프로브가 상응하는 각각의 유전자 서열내에서 특이적 배위를 나열하였다. *tcaBi/B₁₁*에 특이적인 프로브: 서열 ID #25의 1174 내지 3642, *tcaC*에 특이적인 프로브: 서열 ID #25의 3637 내지 6005, *tcbA*에 특이적인 프로브: 서열 ID #11의 2097 내지 4964, *tcdA₁*에 특이적인 프로브: 서열 ID #46의 1660 내지 4191. 하기 표에 포도랍두스 균주의 써던 블롯 분석을 요약하였다. 프로브의 하이브리다이제이션이 일어나는 경우, 하이브리드된 단편은 W-14 균주에 대해 관찰한 패턴과 동일하거나 또는 상이한 것으로 나타났다.

[표 37]

| 포토랩두스 균주의 씨던 분석 | | | | |
|-----------------|------|------|------|----------------------|
| 균주 | tcdA | tcbA | tcaC | tcaB _{i/ii} |
| WX-1 | D | D | D | D |
| WX-2 | D | D | - | D |
| WX-3 | D | D | D | D |
| WX-4 | D | D | ND | D |
| WX-5 | D | D | D | D |
| WX-6 | D | D | D | D |
| WX-7 | D | D | ND | D |
| WX-8 | D | D | D | D |
| WX-9 | ND | D | D | D |
| WX-10 | ND | D | D | D |
| WX-11 | ND | D | D | D |
| WX-12 | D | D | D | D |
| WX-14 | D | D | D | D |
| WX-15 | D | D | D | D |
| HP88 | D | - | D | D |
| Hm | D | - | D | D |
| Hb | D | - | D | - |
| H9 | D | - | I | D |
| B2 | D | - | D | - |
| NC-1 | D | - | D | D |
| WIR | D | - | D | D |
| W30 | D | D | D | D |
| W-14 | I | I | I | I |

ND=측정없음; - = 하이브리다이즈 산물을 검출할 수 없음;
I = 동일한 단편 패턴; D = 상이한 단편 패턴

[표 38]

| |
|-----------------|
| 포토랩두스 균주의 씨던 분석 |
|-----------------|

| 균주 | tcdA | tcbA | tcaC | tcaB _{i/ii} |
|---------|----------|------|------|----------------------|
| K-122 | 3.3, 2.8 | D | - | ND |
| PWH-5 | + | D | D | - |
| 인디커스 | D | D | 3.0 | I |
| 메기디스 | D | D | D | - |
| GD | D | D | D | - |
| HF-85 | D | D | D | - |
| MP3 | D | - | D | - |
| MP1 | D | + | D | - |
| 에이. 카우스 | D | + | D | - |
| HB-Arg | D | ND | D | - |
| HMGD | D | D | D | - |
| HB 레위스톤 | D | D | D | - |
| HB 오스웨고 | D | D | D | - |
| W-14 | I | I | I | I |

ND=측정없음; - = 하이브리다이즈 산물을 검출할 수 없음;
I = 동일한 단편 패턴; D = 상이한 단편 패턴

[표 39]

| 포토랩두스 균주의 씨던 분석 | | | | |
|-----------------|------|--------|------|----------------------|
| 균주 | tcdA | tcbA | tcaC | tcaB _{i/ii} |
| GL98 | + | + | D | |
| GL101 | - | + | D | |
| GL138 | - | + | D | |
| GL155 | - | - | - | |
| GL217 | + | - | D | |
| GL257 | + | + | D | |
| MP4 | - | + | - | |
| MP5 | - | - | - | |
| 피. 해피알루스 | + | - | D | |
| 피. 제알란디아 | + | - | 11.0 | |
| DEP1 | | 237-64 | | |
| DEP2 | | | | |
| DEP3 | | | | |

상이한 포토랍두스 균주로부터의 독소 착물 펩티드의 N-말단 아미노산 서열

실시에 14에 기재한 바와 같이, 상이한 포토랍두스 균주로부터 단리한 펩티드의 관계를 N-말단 아미노산 서열화에 도입하였다. 몇몇 균주에서의 독소 펩티드 N-말단 아미노산 서열을 W-14 독소와 비교하였다. 표 40에서, 지금까지 비교한 독소 펩티드의 비교는 동일하거나 유사한 (W-14 유전자/펩티드와 40 % 이상의 유사성) 독소 펩티드가 모든 균주에서 존재한다는 것을 보여주었다. 예를 들어, TcaC, SEQ ID NO: 2의 N-말단 아미노산 서열은 HP88에서 160 kDa 펩티드에 대한 것과 동일한 것으로 밝혀졌고, 또한 유사체가 균주 WIR, H9, Hb, WX-1 및 Hm에서 존재하였다. 몇몇 W-14 펩티드 또는 유사체는 다른 균주에서 발견되지 않았지만, 모든 펩티드는 N-말단 차단 또는 낮은 어번던스로 인하여 다른 균주로부터 독소 착물에 대해 서열화되지 않았다. 또한, 많은 다른 N-말단 아미노산 서열 (SEQ ID NOS: 82 내지 88)은 W-14로부터의 펩티드와 유사성이 없는 다른 균주로부터 독소 착물 펩티드에 대해 얻어졌고, 몇몇 경우에는 서로 동일하였다. 예를 들어, 동일한 아미노산 서열, SEQ ID NO: 82는 HP88 및 Hb 균주에 존재하는 64 kDa 펩티드에 대해 얻었고, NC-1 균주 (SEQ ID NO: 83) 중의 70 kDa 펩티드에 대해 유사한 서열을 얻었다.

[표 40]

| 비W-14 균주로부터 단리한 단백질 사이의 아미노 말단 서열 상동성 비교 | | | | | | |
|--|----------|---------|----|-------|---------|---------|
| W-14 펩티드 | W-14 유전자 | W-14 서열 | 서열 | 균주 | 동일성 | 상동성 |
| TcaAii | tcaA | 15 | | | | |
| TcaAiii | tcaA | 4 | | | | |
| TcaBi | tcaB | 3 | 76 | H9 | - | 74 KDa |
| | | | 76 | Hm | - | 71 KDa |
| TcaBii | tcaA | 5 | | H9 | 61 KDa | - |
| | | | | Hm | 61 KDa | - |
| TcaC | tcaA | 2 | 72 | Hb | | 160 KDa |
| | | | | HP88 | 160 KDa | - |
| | | | 73 | WIR | - | 170 KDa |
| | | | 74 | H9 | - | 180 KDa |
| | | | 75 | Hm | - | 170 KDa |
| | | | 80 | WX-1 | - | 170 KDa |
| TcbAii | tcbA | 1 | | | | |
| TcbAiii | tcbA | 40 | | | | |
| TccA | tccA | 8 | 77 | Hb | - | 81 KDa |
| TccB | tccB | 7 | | WX-1 | 170 KDa | - |
| | | | | WX-2 | 180 KDa | - |
| | | | | WX-14 | 180 KDa | - |
| | | | | WIR | 170 KDa | - |
| | | | 78 | H9 | - | 170 KDa |
| | | | | NC-1 | 140 KDa | - |
| | | | 79 | Hm | - | 190 KDa |
| TcdAii | tcdA | | | | | |
| TcdAiii | tcdA | 41 | | Hb | 57 KDa | - |
| | | | 81 | H9 | - | 69 KDa |
| ? | ? | 9 | | Hb | 86 KDa | - |
| | | | | HP88 | 86 KDa | - |

상동성은 W14 유전자/펩티드에 적어도 40%의 유사성이 있는 아미노산 서열을 의미한다. 유사한 잔기는 하기 5개 군 중 하나로 특성화된다 (P,A,G,S,T); (Q,N,E,B,D,Z), (H,K,R), (L,I,V,M), 및 (F,Y,W)

<실시에 26>

포토라프두스 균주의 면역학적 분석

포토라프두스 균주의 배양 브로쓰를 센트리프럽-10 한외여과 장치 (Ameicon, Inc. Beverly, MA 01915)를 사용하여 10 내지 15배로 농축하였다. 단백질의 농도는 ml 당 0.3 내지 3.0 mg의 범위이다. 전체 단백질 10 내지 20 μg을 프리캐스트 4-20 % 폴리아크릴아미드 겔 (Integrated Separation Systems, Natick, MA 01760)의 웰 각각에 충전하였다. 겔 전기영동법을 겔 당 25 ma에 맞춘 일정한 전류를 사용하여 1.25시간 동안 수행하였다. 겔을 반건조 전기 블로터 (Pharmacia Biotech Inc., Piscataway, NJ 08854)를 사용하여 Hybond-ECL (상표명) 니트로셀룰로오스 막 (Amersham Corporation, Arlington Hts, IL 60005)에 전기 블롯시켰다. 일정한 전류를 2.5시간 동안 cm 당 0.75 ma으로 인가하였다. 막을 TBST (25 mM 트리스 HCl pH 7.4, 136 mM NaCl, 2.7 mM KCl, 0.1 % 트윈 20) 중의 10 % 밀크로 실온에서 1시간 동안 차단시켰다. 각 1차 항체를 10 % 밀/TBST 중에서 1:500으로 희석하였다. 1:50 내지 1:1000 사이의 다른 희석액을 또한 사용하였다. 이어서, 이것을 차단 용액 또는 TBST로 철저히 세척하였다. 10 % 밀/TBST 중의 2차 항체 (호르서라디쉬 퍼옥시다제에 공액된 염소 항-쥐 IgG 또는 염소 항-토끼 TgG; BioRad Laboratories, Hercules, CA 94547)의 1:2000 희석액을 막에 도포시키고 이것을 1시간 동안 평면형 진탕기 위에 두었다. 후속해서 막을 과량의 TBST로 세척하였다. 단백질의 검출을 ECL (Enhanced Chemiluminescence) 검출 장치 (Amersham International)를 사용하여 세척하였다.

W-14 펩티드에 대해 생성된 펩티드 특이적 항체의 패널을 사용하고, 웨스턴 블롯 분석을 사용하여 9개의 비-W-14 포토라프두스 균주로부터 브로쓰의 단백질 조성을 특성화시켰다. 또한, W-14 유도 독소 착물 중의 TcbA_{iii} 단백질을 인식하는 하나의 다클론성 항체 (MAb-C5F2)를 사용하였다. 결과 (표 39)는 이들 브로쓰 중의 단백질의 일부에 대한 항체의 상호 인식을 보여주었다. 몇몇 경우에서, 항체에 의해 인식되는 단백질은 W-14 표적 펩티드와 동일한 크기이었다. 다른 경우에는 항체에 의해 인식되는 단백질은 W-14 표적 펩티드 보다 작았다. 이 데이터는 비-W-14 포토라프두스의 일부가 W-14 균주와 유사한 단백질을 생성할 수 있다는 것을 나타낸다. 이들 차이는 검출 또는 단백질 가공 또는 분해 과정에 인한 것이다. 균주의 일부는 몇몇 항체에 의해 인식될 수 있는 단백질을 함유하지 않았지만, 농도가 W-14 펩티드에 대해 관찰된 것 보다 상당히 낮았다. 각종 독소 펩티드 유사체에 대해 비교할 때, 이들 결과는 포토라프두스 균주 사이의 펩티드 다양성을 보여주었다.

[표 41]

선택된 비-W-14 포토라프두스의 브로쓰 중의 W-14 펩티드 내지 단백질에 대해 생성된

다클론성 또는 다클론성 항체에 의한 상호 인식

| 포토라프두스 균주 | MAb C5F2 | PAB TcdA ii-syn | PAB TcdA iii-syn | PAB TcaC -syn | PAB TcaB ii-syn | PAB TcbA iii-syn | PAB TcaB i-syn | PAB TcaA ii-syn | PAB TcaA iii-syn |
|-----------|----------|-----------------|------------------|---------------|-----------------|------------------|----------------|-----------------|------------------|
| MP1 | - | + | + | + | - | + | + | + | + |
| MP2 | + | + | + | + | - | + | + | + | + |
| MP3 | - | + | + | + | - | NT | + | + | - |
| 에이. 카우스 | - | + | + | + | - | NT | + | + | + |
| Hb-osw | - | - | NT | + | + | NT | + | + | + |
| H-Arg | - | + | + | + | - | NT | + | + | + |
| Hb-leu | - | + | + | + | - | NT | + | + | + |
| 인디키스 | + | + | + | + | + | NT | + | + | + |
| HF85 | - | + | + | + | - | + | + | + | + |
| W-14 | + | + | + | + | + | + | + | + | + |

+: 양성 반응; -: 음성 반응; NT: 시험하지 않음

추가로 비-W-14 포토라프두스 균주를 항체로서 배양 브로쓰 및(또는) 부분 정제된 단백질 분획을 사용하여 웨스턴 블롯 분석으로 특성화하였다. 항체의 패널은 MAb-C5F2, MAb-DE1 (인식 TcdA_{ii}), PAb-DE2 (인식 TcaB), PAb-TcbA-syn, PAb-, TcaC-syn, PAb TcaB_{iii}-syn, PAb-TcbA_{iii}-syn, PAb-TcaB_i-syn을 포함한다. 이들 항체는 브로쓰 및 비-W-14 균주의 부분 정제된 분획 중의 단백질과 교차 반응을 보여주었다.

데이터는 항체를 사용하여 브로쓰 뿐만 아니라 부분 정제된 단백질 분획 중에서 단백질을 동정할 수 있다는 것을 나타낸다.

[표 42]

신변된 비 W14 포도알두스의 브로솝 및(또는) 부분 절제된 단백질 분획에서의 단백질에 대한 W-14 펩티드에 대해 생성된 단백질 생성 항체 또는 단백질 생성 항체에 의한 상호 인식

| 포도알두스 균주 | 단백질 생성 항체 | | 다량단백질 항체 | | | | | |
|----------|-----------|---------|----------|------------------------------|--------------|------------------------------|------------------------------|------------------------------|
| | Maδ-CSF2 | Maδ-DE1 | PAδ-DE2 | PAδ-TcαA ₁₁₁ -syn | PAδ-TcαC-syn | PAδ-TcαB ₁₁₁ -syn | PAδ-TcαA ₁₁₁ -syn | PAδ-TcαB ₁₁₁ -syn |
| WX-1 | + | + | + | + | + | + | + | + |
| WX-2 | + | + | + | + | + | + | NT | + |
| WX-3 | + | NT | + | NT | NT | NT | NT | + |
| WX-5 | + | NT | + | NT | NT | NT | NT | NT |
| WX-6 | + | NT | NT | NT | NT | NT | NT | NT |
| WX-7 | + | + | + | + | + | + | NT | + |
| WX-8 | + | NT | NT | NT | NT | NT | NT | NT |
| WX-9 | + | NT | NT | NT | NT | NT | NT | NT |
| WX-10 | - | NT | NT | NT | NT | NT | NT | NT |
| WX-12 | + | + | + | + | + | + | + | + |
| WX-14 | + | + | + | + | NT | + | NT | + |
| WX-15 | + | NT | NT | NT | NT | NT | NT | NT |
| W30 | + | + | + | NT | NT | NT | NT | NT |
| HD | - | NT | + | NT | + | NT | - | + |
| H9 | - | - | - | NT | + | + | NT | NT |
| Hm | - | NT | + | + | + | + | NT | ++ |
| HP58 | - | NT | + | - | + | - | - | + |
| NC-1 | + | - | + | + | + | + | NT | + |
| WIK | - | NT | + | + | + | + | + | + |
| W-14 | + | + | + | + | - | - | - | + |

-: 음성 반응; +: 양성 반응; NT: 시험하지 않음

<실시예 27>

tcdA 코딩 영역의 세균 발현

세균 발현에 대한 tcdA 유전자의 엔지니어링

tcdA 코딩 영역 (SEQ ID NO: 46)의 5 및 3 말단을 변형시켜 불균질 발현 벡터로 세그먼트를 삽입하기 위해 유용한 클로닝 부위를 추가하였다. 말단들을 실시예 8에서 기재한 바와 같이 기본적으로 수행한 폴리머라제 사슬 반응 (PCR) 중의 단일 프라이머를 사용하여 변형시켰다. 하기 기재한 바와 같은 프라이머 세트를 템플레이트로서 코스미드 21D2.4와 결합시켜 사용하여 적절하게 변형된 말단을 갖는 생성물을 생성하였다.

제1 프라이머 세트를 사용하여 유전자의 5 말단을 변형시켜 선행 프라이머 AOF1 (5' GAT CGA TCG ATC CAT GGC CAA CGA CGA GTC TGT AAA AGA GAT ACC TGA TG TAT TAA AAA GCC AGT GTG 3')를 사용하여 개시제 코돈에서 단일 nCO i 부위를 삽입하고 역 프라이머 AOR1 (5' GAT CGA TCG TAC GCG GCC GCT GAT TCG ATC GTC GAC CCA TTG ATT TGA GAT CTG GGC GGC GGG TAT CCA GAT AAT AAA CGG AGT CAC 3')를 사용하여 유전자의 나머지의 삽입을 용이하게 하기 위해 단일 Bgl II, Sal I 및 Not I를 추가하였다.

다른 PCR 반응을 추가의 중단 코돈을 추가함으로써 유전자의 3 말단을 변형시키도록 고안하였다. 선행 프라이머 AOF2 (5' ACT GGC TGC GTC GTC GAC TGG CGG CGA TTT ACT 3')를 사용하여 유전자 중의 단일 Sal I 부위를 가로질러 증폭시키고 후자를 사용하여 변형된 3 말단을 클로닝하였다. 역 프라이머 AOR2 (5' CGA TGC ATG CTG CGG CCG CAG GCC TTC CTC GAG TCA TTA TTT AAT GGT GTA GCG AAT ATG CAA AAT 3')를 사용하여 제2 중단 코돈 (TGA) 및 클로닝 부위 Xho I, Stu I 및 Not I를 삽입하였다. 세균 발현 벡터 pET27b (Novagen, Madison, WI)를 변형시켜 표준 분자 생물학 기술을 사용하여 446 위치에서 Bgl II 부위를 제거하였다.

유전자의 5 말단을 변형시키기 위한 제1 증폭 반응 (AOF1 + AOR1)로부터의 497 bp PCR 생성물을 공급자의 지시에 따라 변형된 pET27b에 라이게이션시켰다. 3개의 단리물의 증폭된 부분의 DNA 서열을 공급자의 추천된 프라이머를 사용하고 전술한 방법을 배열하여 측정하였다. 모든 단리물의 서열은 동일하였다.

이어서, 하나의 단리물을 클로닝 벡터로 사용하여 6341 bp Bgl II 내지 Sal I 단편 상에서 tcdA 유전자의 중간 부분을 삽입하였다. 생성된 클론을 MC4로 부르며 모두 그러나 tcdA 코딩 서열의 3'의 대부분을 함유하였다. 마지막으로, 전체 길이의 코딩 영역을 완성하기 위해, 유전자의 3 말단을 변형시키기 위한 제2 PCR 증폭 (AOF2 + AOR2)으로부터의 832 bp PCR 생성물을 표준 분자 생물학 기술에 따라 라이게이션시켜 Sal I 내지 Not I 단편 상에 MC4를 단리하였다. tcdA 코딩 영역은 서열화되었고 완성된 것으로 밝혀졌고, 생성된 플라스미드는 pDAB2035로 칭한다.

tcdA의 세균 발현에 대한 플라스미드 pDAB2036, pDAB2037 및 pDAB2038의 구성

tcdA 코딩 영역을 제한 효소 Nco I 및 Xho I로 플라스미드 pDAB2035로부터 절단시키고 겔을 정제

시켰다. 단편을 발현 벡터 pET15의 Nco I 및 Xho I 부위로 라이게이션시켜 플라스미드 pDAB2036을 생성하였다. 또한, pDAB2035를 Nco I 및 Not I로 절단하여 tcdA 코딩 영역을 방출시키고 이것을 발현 벡터 pET28b의 Nco I 및 Not I 위치로 라이게이션시켜 플라스미드 pDAB2037를 생성하였다. 마지막으로, 플라스미드 pDAB2035를 Nco I 및 Stu I로 절단하여 tcdA 영역을 방출시켰다. 이 단편을 발현 벡터 Trc99a로 라이게이션시키고, Hind III로 절단시킨 다음 T4 DNA 폴리머라제 처리하여 말단을 뭉뚱하게 하였다. 이어서, 벡터를 Nco I로 절단시키고 tcdA 단편을 절단한 Nco I/Stu I로 라이게이션시켰다. 생성된 플라스미드는 pDAB2038로 칭한다.

플라스미드 pDAB2038로부터의 tcdA의 발현

플라스미드 pDAB2038를 BL21 세포로 전사시키고 실시예 19에서 플라스미드 pDAB2033에 대해 전술한 바와 같이 발현시켰다.

E. 콜라이로부터의 tcdA의 정제

발현 배양물을 10,300 g에서 30분 동안 원심분리하고 상청액을 수집했다. 이것을 2 용적의 H₂O로 희석시키고, 10 mM 인산나트륨 완충액 (pH 7.0, 완충액 A)로 평형시킨 공극 50 HQ (Perspective System, MA) 칼럼 (1.6 cm x 10 cm)에 7.5 ml/분의 유속으로 걸었다. 칼럼을 280 nm에서 광학밀도가 기본선으로 되돌아 갈때까지 완충액 A로 세척하였다. 이어서, 칼럼에 결합된 단백질을 완충액 A 중의 1 M NaCl로 용출시켰다.

분획을 완충액 A로 평형시킨 겔 여과 칼럼, 세파로스 CL-4B (2.6 x 100 cm)에 분취액 20 ml로 충전시켰다. 단백질을 0.75 ml/분의 유속으로 완충액 A로 용출시켰다. 체류시간이 260분 및 460분 사이인 분획을 모으고 20 mM 트리스-HCl (pH 7.0, 완충액 B)로 평형시킨 모노 Q 5/5 칼럼에 1 ml/분으로 걸었다. 칼럼을 280 nm에서 광학밀도가 기본선으로 되돌아 갈때까지 완충액 A로 세척하였다. 이어서, 칼럼에 결합된 단백질을 완충액 B 중의 1 M NaCl로 용출시켰다. 칼럼에 결합한 단백질을 30분 동안 1 ml/분으로 완충액 B 중의 0 내지 1 M NaCl의 선형 구배로 용출하였다. 1 밀리리터 분획을 수집하고, 연속으로 희석하고 SCR 생분석에 도입하였다. 0.1 및 0.3 M NaCl 사이에 용출된 분획은 가장 높은 살충 활성을 갖는 것으로 밝혀졌다. pAb TcdA_{III}-syn 항체 및 Tcd_{III}-syn 항체를 사용하는 활성인 분획의 웨스턴 분석에서 펩티드 TcdA_{III} 및 TcdA_{III}의 존재가 확인되었다.

DNA SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Ensign, Jerald C

Bowen, David J

Petell, James

Fatig, Raymond

Schoonover, Sue

French-Constant, Richard

Orr, Gregory L

Merlo, Donald J

Roberts, Jean L

Rocheleau, Thomas A

(ii) TITLE OF INVENTION: Insecticidal Protein Toxins from

Photorhabdus

(iii) NUMBER OF SEQUENCES: 88

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: DowElanco

(B) STREET: 9330 Zionsville Road

(C) CITY: Indianapolis

(D) STATE: IN

(E) COUNTRY: US

(F) ZIP: 46268

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/063,615
- (B) FILING DATE: 18-MAY-1993

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/395,497
- (B) FILING DATE: 28-FEB-1995

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 60/007,255
- (B) FILING DATE: 06-NOV-1995

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/608,423
- (B) FILING DATE: 28-FEB-1996

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/705,484
- (B) FILING DATE: 28-AUG-1996

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/743,699
- (B) FILING DATE: 06-NOV-1996

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Borucki, Andrea T.
- (B) REGISTRATION NUMBER: 33651
- (C) REFERENCE/DOCKET NUMBER: 50301E

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 317-337-4846
- (B) TELEFAX: 317-337-4847

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1 (TcbAii N-terminus):

Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn
 1 5 10

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2 (TcaC N-terminus):

Met Gln Asp Ser Pro Glu Val Ser Ile Thr Thr Trp
 1 5 10

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3 (TcaBi N-terminus):

Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp Ala
 1 5 10 15
 Leu Val Ala

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4 (TcaAiii N-terminus):

Ala Ser Pro Leu Ser Thr Ser Glu Leu Thr Ser Lys Leu Asn
 1 5 10

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5 (TcaBii N-terminus):

Ala Gly Asp Thr Ala Asn Ile Gly Asp
 1 5

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Leu Gly Gly Ala Ala Thr Leu Leu Asp Leu Leu Leu Pro Gln Ile
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7 (TccB N-terminus):

Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu
 1 5 10

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8 (TccA N-terminus):

Met Asn Leu Ala Ser Pro Leu Ile Ser
 1 5

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Ile Asn Leu Asp Ile Asn Glu Gln Asn Lys Ile Met Val Val Ser
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

| | | | |
|---|-----|-----|------|
| 130 | 135 | 140 | |
| TTA GCA AGC TTA ATG CTC AGC CAG AAA AAT ATG GAT GAG GAA ATT TCA | | | 480 |
| Leu Ala Ser Leu Met Leu Ser Gln Lys Asn Met Asp Glu Glu Ile Ser | | | |
| 145 | 150 | 155 | 160 |
| ACG CTG GCT CTC TCT AAT GAA TTG TGC CTT GCC GGG ATC GAA ACA AAA | | | 528 |
| Thr Leu Ala Leu Ser Asn Glu Leu Cys Leu Ala Gly Ile Glu Thr Lys | | | |
| 165 | 170 | 175 | |
| ACA GGA AAA TCA CAA GAT GAA GTG ATG GAT ATG TTG TCA ACT TAT CGT | | | 576 |
| Thr Gly Lys Ser Gln Asp Glu Val Met Asp Met Leu Ser Thr Tyr Arg | | | |
| 180 | 185 | 190 | |
| TTA AGT GGA GAG ACA CCT TAT CAT CAC GCT TAT GAA ACT GTT CGT GAA | | | 624 |
| Leu Ser Gly Glu Thr Pro Tyr His His Ala Tyr Glu Thr Val Arg Glu | | | |
| 195 | 200 | 205 | |
| ATC GTT CAT GAA CGT GAT CCA GGA TTT CGT CAT TTG TCA CAG GCA CCC | | | 672 |
| Ile Val His Glu Arg Asp Pro Gly Phe Arg His Leu Ser Gln Ala Pro | | | |
| 210 | 215 | 220 | |
| ATT GTT GCT GCT AAG CTC GAT CCT GTG ACT TTG TTG GGT ATT AGC TCC | | | 720 |
| Ile Val Ala Ala Lys Leu Asp Pro Val Thr Leu Leu Gly Ile Ser Ser | | | |
| 225 | 230 | 235 | 240 |
| CAT ATT TCG CCA GAA CTG TAT AAC TTG CTG ATT GAG GAG ATC CCG GAA | | | 768 |
| His Ile Ser Pro Glu Leu Tyr Asn Leu Leu Ile Glu Glu Ile Pro Glu | | | |
| 245 | 250 | 255 | |
| AAA GAT GAA GCC GCG CTT GAT ACG CTT TAT AAA ACA AAC TTT GGC GAT | | | 816 |
| Lys Asp Glu Ala Ala Leu Asp Thr Leu Tyr Lys Thr Asn Phe Gly Asp | | | |
| 260 | 265 | 270 | |
| ATT ACT ACT GCT CAG TTA ATG TCC CCA AGT TAT CTG GCC CGG TAT TAT | | | 864 |
| Ile Thr Thr Ala Gln Leu Met Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr | | | |
| 275 | 280 | 285 | |
| GGC GTC TCA CCG GAA GAT ATT GCC TAC GTG ACG ACT TCA TTA TCA CAT | | | 912 |
| Gly Val Ser Pro Glu Asp Ile Ala Tyr Val Thr Thr Ser Leu Ser His | | | |
| 290 | 295 | 300 | |
| GTT GGA TAT AGC AGT GAT ATT CTG GTT ATT CCG TTG GTC GAT GGT GTG | | | 960 |
| Val Gly Tyr Ser Ser Asp Ile Leu Val Ile Pro Leu Val Asp Gly Val | | | |
| 305 | 310 | 315 | 320 |
| GGT AAG ATG GAA GTA GTT CGT GTT ACC CGA ACA CCA TCG GAT AAT TAT | | | 1008 |
| Gly Lys Met Glu Val Val Arg Val Thr Arg Thr Pro Ser Asp Asn Tyr | | | |
| 325 | 330 | 335 | |
| ACC AGT CAG ACG AAT TAT ATT GAG CTG TAT CCA CAG GGT GGC GAC AAT | | | 1056 |
| Thr Ser Gln Thr Asn Tyr Ile Glu Leu Tyr Pro Gln Gly Gly Asp Asn | | | |
| 340 | 345 | 350 | |
| TAT TTG ATC AAA TAC AAT CTA AGC AAT AGT TTT GGT TTG GAT GAT TTT | | | 1104 |
| Tyr Leu Ile Lys Tyr Asn Leu Ser Asn Ser Phe Gly Leu Asp Asp Phe | | | |
| 355 | 360 | 365 | |
| TAT CTG CAA TAT AAA GAT GGT TCC GCT GAT TGG ACT GAG ATT GCC CAT | | | 1152 |
| Tyr Leu Gln Tyr Lys Asp Gly Ser Ala Asp Trp Thr Glu Ile Ala His | | | |

370 375 380
 AAT CCC TAT CCT GAT ATG GTC ATA AAT CAA AAG TAT GAA TCA CAG GCG 1200
 Asn Pro Tyr Pro Asp Met Val Ile Asn Gln Lys Tyr Glu Ser Gln Ala
 385 390 395 400
 ACA ATC AAA CGT AGT GAC TCT GAC AAT ATA CTC AGT ATA GGG TTA CAA 1248
 Thr Ile Lys Arg Ser Asp Ser Asp Asn Ile Leu Ser Ile Gly Leu Gln
 405 410 415
 AGA TGG CAT AGC GGT AGT TAT AAT TTT GCC GCC GCC AAT TTT AAA ATT 1296
 Arg Trp His Ser Gly Ser Tyr Asn Phe Ala Ala Ala Asn Phe Lys Ile
 420 425 430
 GAC CAA TAC TCC CCG AAA GCT TTC CTG CTT AAA ATG AAT AAG GCT ATT 1344
 Asp Gln Tyr Ser Pro Lys Ala Phe Leu Leu Lys Met Asn Lys Ala Ile
 435 440 445
 CGG TTG CTC AAA GCT ACC GGC CTC TCT TTT GCT ACG TTG GAG CGT ATT 1392
 Arg Leu Leu Lys Ala Thr Gly Leu Ser Phe Ala Thr Leu Glu Arg Ile
 450 455 460
 GTT GAT AGT GTT AAT AGC ACC AAA TCC ATC ACG GTT GAG GTA TTA AAC 1440
 Val Asp Ser Val Asn Ser Thr Lys Ser Ile Thr Val Glu Val Leu Asn
 465 470 475 480
 AAG GTT TAT CGG GTA AAA TTC TAT ATT GAT CGT TAT GGC ATC AGT GAA 1488
 Lys Val Tyr Arg Val Lys Phe Tyr Ile Asp Arg Tyr Gly Ile Ser Glu
 485 490 495
 GAG ACA GCC GCT ATT TTG GCT AAT ATT AAT ATC TCT CAG CAA GCT GTT 1536
 Glu Thr Ala Ala Ile Leu Ala Asn Ile Asn Ile Ser Gln Gln Ala Val
 500 505 510
 GGC AAT CAG CTT AGC CAG TTT GAG CAA CTA TTT AAT CAC CCG CCG CTC 1584
 Gly Asn Gln Leu Ser Gln Phe Glu Gln Leu Phe Asn His Pro Pro Leu
 515 520 525
 AAT GGT ATT CGC TAT GAA ATC AGT GAG GAC AAC TCC AAA CAT CTT CCT 1632
 Asn Gly Ile Arg Tyr Glu Ile Ser Glu Asp Asn Ser Lys His Leu Pro
 530 535 540
 AAT CCT GAT CTG AAC CTT AAA CCA GAC AGT ACC GGT GAT GAT CAA CGC 1680
 Asn Pro Asp Leu Asn Leu Lys Pro Asp Ser Thr Gly Asp Asp Gln Arg
 545 550 555 560
 AAG GCG GTT TTA AAA CGC GCG TTT CAG GTT AAC GCC AGT GAG TTG TAT 1728
 Lys Ala Val Leu Lys Arg Ala Phe Gln Val Asn Ala Ser Glu Leu Tyr
 565 570 575
 CAG ATG TTA TTG ATC ACT GAT CGT AAA GAA GAC GGT GTT ATC AAA AAT 1776
 Gln Met Leu Leu Ile Thr Asp Arg Lys Glu Asp Gly Val Ile Lys Asn
 580 585 590
 AAC TTA GAG AAT TTG TCT GAT CTG TAT TTG GTT AGT TTG CTG GCC CAG 1824
 Asn Leu Glu Asn Leu Ser Asp Leu Tyr Leu Val Ser Leu Leu Ala Gln
 595 600 605
 ATT CAT AAC CTG ACT ATT GCT GAA TTG AAC ATT TTG TTG GTG ATT TGT 1872
 Ile His Asn Leu Thr Ile Ala Glu Leu Asn Ile Leu Leu Val Ile Cys

610 615 620
 GGC TAT GGC GAC ACC AAC ATT TAT CAG ATT ACC GAC GAT AAT TTA GCC 1920
 Gly Tyr Gly Asp Thr Asn Ile Tyr Gln Ile Thr Asp Asp Asn Leu Ala
 625 630 635 640
 AAA ATA GTG GAA ACA TTG TTG TGG ATC ACT CAA TGG TTG AAG ACC CAA 1968
 Lys Ile Val Glu Thr Leu Leu Trp Ile Thr Gln Trp Leu Lys Thr Gln
 645 650 655
 AAA TGG ACA GTT ACC GAC CTG TTT CTG ATG ACC ACG GCC ACT TAC AGC 2016
 Lys Trp Thr Val Thr Asp Leu Phe Leu Met Thr Thr Ala Thr Tyr Ser
 660 665 670
 ACC ACT TTA ACG CCA GAA ATT AGC AAT CTG ACG GCT ACG TTG TCT TCA 2064
 Thr Thr Leu Thr Pro Glu Ile Ser Asn Leu Thr Ala Thr Leu Ser Ser
 675 680 685
 ACT TTG CAT GGC AAA GAG AGT CTG ATT GGG GAA GAT CTG AAA AGA GCA 2112
 Thr Leu His Gly Lys Glu Ser Leu Ile Gly Glu Asp Leu Lys Arg Ala
 690 695 700
 ATG GCG CCT TGC TTC ACT TCG GCT TTG CAT TTG ACT TCT CAA GAA GTT 2160
 Met Ala Pro Cys Phe Thr Ser Ala Leu His Leu Thr Ser Gln Glu Val
 705 710 715 720
 GCG TAT GAC CTG CTG TTG TGG ATA GAC CAG ATT CAA CCG GCA CAA ATA 2208
 Ala Tyr Asp Leu Leu Leu Trp Ile Asp Gln Ile Gln Pro Ala Gln Ile
 725 730 735
 ACT GTT GAT GGG TTT TGG GAA GAA GTG CAA ACA ACA CCA ACC AGC TTG 2256
 Thr Val Asp Gly Phe Trp Glu Glu Val Gln Thr Thr Pro Thr Ser Leu
 740 745 750
 AAG GTG ATT ACC TTT GCT CAG GTG CTG GCA CAA TTG AGC CTG ATC TAT 2304
 Lys Val Ile Thr Phe Ala Gln Val Leu Ala Gln Leu Ser Leu Ile Tyr
 755 760 765
 CGT CGT ATT GGG TTA AGT GAA ACG GAA CTG TCA CTG ATC GTG ACT CAA 2352
 Arg Arg Ile Gly Leu Ser Glu Thr Glu Leu Ser Leu Ile Val Thr Gln
 770 775 780
 TCT TCT CTG CTA GTG GCA GGC AAA AGC ATA CTG GAT CAC GGT CTG TTA 2400
 Ser Ser Leu Leu Val Ala Gly Lys Ser Ile Leu Asp His Gly Leu Leu
 785 790 795 800
 ACC CTG ATG GCC TTG GAA GGT TTT CAT ACC TGG GTT AAT GGC TTG GGG 2448
 Thr Leu Met Ala Leu Glu Gly Phe His Thr Trp Val Asn Gly Leu Gly
 805 810 815
 CAA CAT GCC TCC TTG ATA TTG GCG GCG TTG AAA GAC GGA GCC TTG ACA 2496
 Gln His Ala Ser Leu Ile Leu Ala Ala Leu Lys Asp Gly Ala Leu Thr
 820 825 830
 GTT ACC GAT GTA GCA CAA GCT ATG AAT AAG GAG GAA TCT CTC CTA CAA 2544
 Val Thr Asp Val Ala Gln Ala Met Asn Lys Glu Glu Ser Leu Leu Gln
 835 840 845
 ATG GCA GCT AAT CAG GTG GAG AAG GAT CTA ACA AAA CTG ACC AGT TGG 2592
 Met Ala Ala Asn Gln Val Glu Lys Asp Leu Thr Lys Leu Thr Ser Trp

1090 1095 1100
 ACG TAT TAC TGG CGT AGT GTT GAT CAC AGC AAA TGT GAA AAT GGC AAG 3360
 Thr Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Cys Glu Asn Gly Lys
 1105 1110 1115 1120
 TTT GCC GCT AAT GCT TGG GGT GAG TGG AAT AAA ATT ACC TGT GCT GTC 3408
 Phe Ala Ala Asn Ala Trp Gly Glu Trp Asn Lys Ile Thr Cys Ala Val
 1125 1130 1135
 AAT CCT TGG AAA AAT ATC ATC CGT CCG GTT GTT TAT ATG TCC CGC TTA 3456
 Asn Pro Trp Lys Asn Ile Ile Arg Pro Val Val Tyr Met Ser Arg Leu
 1140 1145 1150
 TAT CTG CTA TGG CTG GAG CAG CAA TCA AAG AAA AGT GAT GAT GGT AAA 3504
 Tyr Leu Leu Trp Leu Glu Gln Gln Ser Lys Lys Ser Asp Asp Gly Lys
 1155 1160 1165
 ACC ACG ATT TAT CAA TAT AAC TTA AAA CTG GCT CAT ATT CGT TAC GAC 3552
 Thr Thr Ile Tyr Gln Tyr Asn Leu Lys Leu Ala His Ile Arg Tyr Asp
 1170 1175 1180
 GGT AGT TGG AAT ACA CCA TTT ACT TTT GAT GTG ACA GAA AAG GTA AAA 3600
 Gly Ser Trp Asn Thr Pro Phe Thr Phe Asp Val Thr Glu Lys Val Lys
 1185 1190 1195 1200
 AAT TAC ACG TCG AGT ACT GAT GCT GCT GAA TCT TTA GGG TTG TAT TGT 3648
 Asn Tyr Thr Ser Ser Thr Asp Ala Ala Glu Ser Leu Gly Leu Tyr Cys
 1205 1210 1215
 ACT GGT TAT CAA GGG GAA GAC ACT CTA TTA GTT ATG TTC TAT TCG ATG 3696
 Thr Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Ser Met
 1220 1225 1230
 CAG AGT AGT TAT AGC TCC TAT ACC GAT AAT AAT GCG CCG GTC ACT GGG 3744
 Gln Ser Ser Tyr Ser Ser Tyr Thr Asp Asn Asn Ala Pro Val Thr Gly
 1235 1240 1245
 CTA TAT ATT TTC GCT GAT ATG TCA TCA GAC AAT ATG ACG AAT GCA CAA 3792
 Leu Tyr Ile Phe Ala Asp Met Ser Ser Asp Asn Met Thr Asn Ala Gln
 1250 1255 1260
 GCA ACT AAC TAT TGG AAT AAC AGT TAT CCG CAA TTT GAT ACT GTG ATG 3840
 Ala Thr Asn Tyr Trp Asn Asn Ser Tyr Pro Gln Phe Asp Thr Val Met
 1265 1270 1275 1280
 GCA GAT CCG GAT AGC GAC AAT AAA AAA GTC ATA ACC AGA AGA GTT AAT 3888
 Ala Asp Pro Asp Ser Asp Asn Lys Lys Val Ile Thr Arg Arg Val Asn
 1285 1290 1295
 AAC CGT TAT GCG GAG GAT TAT GAA ATT CCT TCC TCT GTG ACA AGT AAC 3936
 Asn Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Thr Ser Asn
 1300 1305 1310
 AGT AAT TAT TCT TGG GGT GAT CAC AGT TTA ACC ATG CTT TAT GGT GGT 3984
 Ser Asn Tyr Ser Trp Gly Asp His Ser Leu Thr Met Leu Tyr Gly Gly
 1315 1320 1325
 AGT GTT CCT AAT ATT ACT TTT GAA TCG GCG GCA GAA GAT TTA AGG CTA 4032
 Ser Val Pro Asn Ile Thr Phe Glu Ser Ala Ala Glu Asp Leu Arg Leu

1330 1335 1340
 TCT ACC AAT ATG GCA TTG AGT ATT ATT CAT AAT GGA TAT GCG GGA ACC 4080
 Ser Thr Asn Met Ala Leu Ser Ile Ile His Asn Gly Tyr Ala Gly Thr
 1345 1350 1355 1360
 CGC CGT ATA CAA TGT AAT CTT ATG AAA CAA TAC GCT TCA TTA GGT GAT 4128
 Arg Arg Ile Gln Cys Asn Leu Met Lys Gln Tyr Ala Ser Leu Gly Asp
 1365 1370 1375
 AAA TTT ATA ATT TAT GAT TCA TCA TTT GAT GAT GCA AAC CGT TTT AAT 4176
 Lys Phe Ile Ile Tyr Asp Ser Ser Phe Asp Asp Ala Asn Arg Phe Asn
 1380 1385 1390
 CTG GTG CCA TTG TTT AAA TTC GGA AAA GAC GAG AAC TCA GAT GAT AGT 4224
 Leu Val Pro Leu Phe Lys Phe Gly Lys Asp Glu Asn Ser Asp Asp Ser
 1395 1400 1405
 ATT TGT ATA TAT AAT GAA AAC CCT TCC TCT GAA GAT AAG AAG TGG TAT 4272
 Ile Cys Ile Tyr Asn Glu Asn Pro Ser Ser Glu Asp Lys Lys Trp Tyr
 1410 1415 1420
 TTT TCT TCG AAA GAT GAC AAT AAA ACA GCG GAT TAT AAT GGT GGA ACT 4320
 Phe Ser Ser Lys Asp Asp Asn Lys Thr Ala Asp Tyr Asn Gly Gly Thr
 1425 1430 1435 1440
 CAA TGT ATA GAT GCT GGA ACC AGT AAC AAA GAT TTT TAT TAT AAT CTC 4368
 Gln Cys Ile Asp Ala Gly Thr Ser Asn Lys Asp Phe Tyr Tyr Asn Leu
 1445 1450 1455
 CAG GAG ATT GAA GTA ATT AGT GTT ACT GGT GGG TAT TGG TCG AGT TAT 4416
 Gln Glu Ile Glu Val Ile Ser Val Thr Gly Gly Tyr Trp Ser Ser Tyr
 1460 1465 1470
 AAA ATA TCC AAC CCG ATT AAT ATC AAT ACG GGC ATT GAT AGT GCT AAA 4464
 Lys Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
 1475 1480 1485
 GTA AAA GTC ACC GTA AAA GCG GGT GGT GAC GAT CAA ATC TTT ACT GCT 4512
 Val Lys Val Thr Val Lys Ala Gly Gly Asp Asp Gln Ile Phe Thr Ala
 1490 1495 1500
 GAT AAT AGT ACC TAT GTT CCT CAG CAA CCG GCA CCC AGT TTT GAG GAG 4560
 Asp Asn Ser Thr Tyr Val Pro Gln Gln Pro Ala Pro Ser Phe Glu Glu
 1505 1510 1515 1520
 ATG ATT TAT CAG TTC AAT AAC CTG ACA ATA GAT TGT AAG AAT TTA AAT 4608
 Met Ile Tyr Gln Phe Asn Asn Leu Thr Ile Asp Cys Lys Asn Leu Asn
 1525 1530 1535
 TTC ATC GAC AAT CAG GCA CAT ATT GAG ATT GAT TTC ACC GCT ACG GCA 4656
 Phe Ile Asp Asn Gln Ala His Ile Glu Ile Asp Phe Thr Ala Thr Ala
 1540 1545 1550
 CAA GAT GGC CGA TTC TTG GGT GCA GAA ACT TTT ATT ATC CCG GTA ACT 4704
 Gln Asp Gly Arg Phe Leu Gly Ala Glu Thr Phe Ile Ile Pro Val Thr
 1555 1560 1565
 AAA AAA GTT CTC GGT ACT GAG AAC GTG ATT GCG TTA TAT AGC GAA AAT 4752
 Lys Lys Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn

1570 1575 1580
 AAC GGT GTT CAA TAT ATG CAA ATT GGC GCA TAT CGT ACC CGT TTG AAT 4800
 Asn Gly Val Gln Tyr Met Gln Ile Gly Ala Tyr Arg Thr Arg Leu Asn
 1585 1590 1595 1600
 ACG TTA TTC GCT CAA CAG TTG GTT AGC CGT GCT AAT CGT GGC ATT GAT 4848
 Thr Leu Phe Ala Gln Gln Leu Val Ser Arg Ala Asn Arg Gly Ile Asp
 1605 1610 1615
 GCA GTG CTC AGT ATG GAA ACT CAG AAT ATT CAG GAA CCG CAA TTA GGA 4896
 Ala Val Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly
 1620 1625 1630
 GCG GGC ACA TAT GTG CAG CTT GTG TTG GAT AAA TAT GAT GAG TCT ATT 4944
 Ala Gly Thr Tyr Val Gln Leu Val Leu Asp Lys Tyr Asp Glu Ser Ile
 1635 1640 1645
 CAT GGC ACT AAT AAA AGC TTT GCT ATT GAA TAT GTT GAT ATA TTT AAA 4992
 His Gly Thr Asn Lys Ser Phe Ala Ile Glu Tyr Val Asp Ile Phe Lys
 1650 1655 1660
 GAG AAC GAT AGT TTT GTG ATT TAT CAA GGA GAA CTT AGC GAA ACA AGT 5040
 Glu Asn Asp Ser Phe Val Ile Tyr Gln Gly Glu Leu Ser Glu Thr Ser
 1665 1670 1675 1680
 CAA ACT GTT GTG AAA GTT TTC TTA TCC TAT TTT ATA GAG GCG ACT GGA 5088
 Gln Thr Val Val Lys Val Phe Leu Ser Tyr Phe Ile Glu Ala Thr Gly
 1685 1690 1695
 AAT AAG AAC CAC TTA TGG GTA CGT GCT AAA TAC CAA AAG GAA ACG ACT 5136
 Asn Lys Asn His Leu Trp Val Arg Ala Lys Tyr Gln Lys Glu Thr Thr
 1700 1705 1710
 GAT AAG ATC TTG TTC GAC CGT ACT GAT GAG AAA GAT CCG CAC GGT TGG 5184
 Asp Lys Ile Leu Phe Asp Arg Thr Asp Glu Lys Asp Pro His Gly Trp
 1715 1720 1725
 TTT CTC AGC GAC GAT CAC AAG ACC TTT AGT GGT CTC TCT TCC GCA CAG 5232
 Phe Leu Ser Asp Asp His Lys Thr Phe Ser Gly Leu Ser Ser Ala Gln
 1730 1735 1740
 GCA TTA AAG AAC GAC AGT GAA CCG ATG GAT TTC TCT GGC GCC AAT GCT 5280
 Ala Leu Lys Asn Asp Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ala
 1745 1750 1755 1760
 CTC TAT TTC TGG GAA CTG TTC TAT TAC ACG CCG ATG ATG ATG GCT CAT 5328
 Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Met Met Met Ala His
 1765 1770 1775
 CGT TTG TTG CAG GAA CAG AAT TTT GAT GCG GCG AAC CAT TGG TTC CGT 5376
 Arg Leu Leu Gln Glu Gln Asn Phe Asp Ala Ala Asn His Trp Phe Arg
 1780 1785 1790
 TAT GTC TGG AGT CCA TCC GGT TAT ATC GTT GAT GGT AAA ATT GCT ATC 5424
 Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val Asp Gly Lys Ile Ala Ile
 1795 1800 1805
 TAC CAC TGG AAC GTG CGA CCG CTG GAA GAA GAC ACC AGT TGG AAT GCA 5472
 Tyr His Trp Asn Val Arg Pro Leu Glu Glu Asp Thr Ser Trp Asn Ala

1810 1815 1820
 CAA CAA CTG GAC TCC ACC GAT CCA GAT GCT GTA GCC CAA GAT GAT CCG 5520
 Gln Gln Leu Asp Ser Thr Asp Pro Asp Ala Val Ala Gln Asp Asp Pro
 1825 1830 1835 1840
 ATG CAC TAC AAG GTG GCT ACC TTT ATG GCG ACG TTG GAT CTG CTA ATG 5568
 Met His Tyr Lys Val Ala Thr Phe Met Ala Thr Leu Asp Leu Leu Met
 1845 1850 1855
 GCC CGT GGT GAT GCT GCT TAC CGC CAG TTA GAG CGT GAT ACG TTG GCT 5616
 Ala Arg Gly Asp Ala Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Ala
 1860 1865 1870
 GAA GCT AAA ATG TGG TAT ACA CAG GCG CTT AAT CTG TTG GGT GAT GAG 5664
 Glu Ala Lys Met Trp Tyr Thr Gln Ala Leu Asn Leu Leu Gly Asp Glu
 1875 1880 1885
 CCA CAA GTG ATG CTG AGT ACG ACT TGG GCT AAT CCA ACA TTG GGT AAT 5712
 Pro Gln Val Met Leu Ser Thr Thr Trp Ala Asn Pro Thr Leu Gly Asn
 1890 1895 1900
 GCT GCT TCA AAA ACC ACA CAG CAG GTT CGT CAG CAA GTG CTT ACC CAG 5760
 Ala Ala Ser Lys Thr Thr Gln Gln Val Arg Gln Gln Val Leu Thr Gln
 1905 1910 1915 1920
 TTG CGT CTC AAT AGC AGG GTA AAA ACC CCG TTG CTA GGA ACA GCC AAT 5808
 Leu Arg Leu Asn Ser Arg Val Lys Thr Pro Leu Leu Gly Thr Ala Asn
 1925 1930 1935
 TCC CTG ACC GCT TTA TTC CTG CCG CAG GAA AAT AGC AAG CTC AAA GGC 5856
 Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn Ser Lys Leu Lys Gly
 1940 1945 1950
 TAC TGG CGG ACA CTG GCG CAG CGT ATG TTT AAT TTA CGT CAT AAT CTG 5904
 Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn Leu Arg His Asn Leu
 1955 1960 1965
 TCG ATT GAC GGC CAG CCG CTC TCC TTG CCG CTG TAT GCT AAA CCG GCT 5952
 Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu Tyr Ala Lys Pro Ala
 1970 1975 1980
 GAT CCA AAA GCT TTA CTG AGT GCG GCG GTT TCA GCT TCT CAA GGG GGA 6000
 Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser Ala Ser Gln Gly Gly
 1985 1990 1995 2000
 GCC GAC TTG CCG AAG GCG CCG CTG ACT ATT CAC CGC TTC CCT CAA ATG 6048
 Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His Arg Phe Pro Gln Met
 2005 2010 2015
 CTA GAA GGG GCA CGG GGC TTG GTT AAC CAG CTT ATA CAG TTC GGT AGT 6096
 Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu Ile Gln Phe Gly Ser
 2020 2025 2030
 TCA CTA TTG GGG TAC AGT GAG CGT CAG GAT GCG GAA GCT ATG AGT CAA 6144
 Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala Glu Ala Met Ser Gln
 2035 2040 2045
 CTA CTG CAA ACC CAA GCC AGC GAG TTA ATA CTG ACC AGT ATT CGT ATG 6192
 Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu Thr Ser Ile Arg Met

2050 2055 2060
 CAG GAT AAC CAA TTG GCA GAG CTG GAT TCG GAA AAA ACC GCC TTG CAA 6240
 Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu Lys Thr Ala Leu Gln
 2065 2070 2075 2080
 GTC TCT TTA GCT GGA GTG CAA CAA CGG TTT GAC AGC TAT AGC CAA CTG 6288
 Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp Ser Tyr Ser Gln Leu
 2085 2090 2095
 TAT GAG GAG AAC ATC AAC GCA GGT GAG CAG CGA GCG CTG GCG TTA CGC 6336
 Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg Ala Leu Ala Leu Arg
 2100 2105 2110
 TCA GAA TCT GCT ATT GAG TCT CAG GGA GCG CAG ATT TCC CGT ATG GCA 6384
 Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln Ile Ser Arg Met Ala
 2115 2120 2125
 GGC GCG GGT GTT GAT ATG GCA CCA AAT ATC TTC GGC CTG GCT GAT GGC 6432
 Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe Gly Leu Ala Asp Gly
 2130 2135 2140
 GGC ATG CAT TAT GGT GCT ATT GCC TAT GCC ATC GCT GAC GGT ATT GAG 6480
 Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile Ala Asp Gly Ile Glu
 2145 2150 2155 2160
 TTG AGT GCT TCT GCC AAG ATG GTT GAT GCG GAG AAA GTT GCT CAG TCG 6528
 Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu Lys Val Ala Gln Ser
 2165 2170 2175
 GAA ATA TAT CGC CGT CGC CGT CAA GAA TGG AAA ATT CAG CGT GAC AAC 6576
 Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys Ile Gln Arg Asp Asn
 2180 2185 2190
 GCA CAA GCG GAG ATT AAC CAG TTA AAC GCG CAA CTG GAA TCA CTG TCT 6624
 Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln Leu Glu Ser Leu Ser
 2195 2200 2205
 ATT CGC CGT GAA GCC GCT GAA ATG CAA AAA GAG TAC CTG AAA ACC CAG 6672
 Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu Tyr Leu Lys Thr Gln
 2210 2215 2220
 CAA GCT CAG GCG CAG GCA CAA CTT ACT TTC TTA AGA AGC AAA TTC AGT 6720
 Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu Arg Ser Lys Phe Ser
 2225 2230 2235 2240
 AAT CAA GCG TTA TAT AGT TGG TTA CGA GGG CGT TTG TCA GGT ATT TAT 6768
 Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg Leu Ser Gly Ile Tyr
 2245 2250 2255
 TTC CAG TTC TAT GAC TTG GCC GTA TCA CGT TGC CTG ATG GCA GAG CAA 6816
 Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys Leu Met Ala Glu Gln
 2260 2265 2270
 TCC TAT CAA TGG GAA GCT AAT GAT AAT TCC ATT AGC TTT GTC AAA CCG 6864
 Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile Ser Phe Val Lys Pro
 2275 2280 2285
 GGT GCA TGG CAA GGA ACT TAC GCC GGC TTA TTG TGT GGA GAA GCT TTG 6912
 Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu Cys Gly Glu Ala Leu

2290 2295 2300
 ATA CAA AAT CTG GCA CAA ATG GAA GAG GCA TAT CTG AAA TGG GAA TCT 6960
 Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr Leu Lys Trp Glu Ser
 2305 2310 2315 2320
 CGC GCT TTG GAA GTA GAA CGC ACG GTT TCA TTG GCA GTG GTT TAT GAT 7008
 Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu Ala Val Val Tyr Asp
 2325 2330 2335
 TCA CTG GAA GGT AAT GAT CGT TTT AAT TTA GCG GAA CAA ATA CCT GCA 7056
 Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala Glu Gln Ile Pro Ala
 2340 2345 2350
 TTA TTG GAT AAG GGG GAG GGA ACA GCA GGA ACT AAA GAA AAT GGG TTA 7104
 Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr Lys Glu Asn Gly Leu
 2355 2360 2365
 TCA TTG GCT AAT GCT ATC CTG TCA GCT TCG GTC AAA TTG TCG GAC TTG 7152
 Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val Lys Leu Ser Asp Leu
 2370 2375 2380
 AAA CTG GGA ACG GAT TAT CCA GAC AGT ATC GTT GGT AGC AAC AAG GTT 7200
 Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val Gly Ser Asn Lys Val
 2385 2390 2395 2400
 CGT CGT ATT AAG CAA ATC AGT GTT TCG CTA CCT GCA TTG GTT GGG CCT 7248
 Arg Arg Ile Lys Lys Ile Ser Val Ser Leu Pro Ala Leu Val Gly Pro
 2405 2410 2415
 TAT CAG GAT GTT CAG GCT ATG CTC AGC TAT GGT GGC AGT ACT CAA TTG 7296
 Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly Gly Ser Thr Gln Leu
 2420 2425 2430
 CCG AAA GGT TGT TCA GCG TTG GCT GTG TCT CAT GGT ACC AAT GAT AGT 7344
 Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His Gly Thr Asn Asp Ser
 2435 2440 2445
 GGT CAG TTC CAG TTG GAT TTC AAT GAC GGC AAA TAC CTG CCA TTT GAA 7392
 Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys Tyr Leu Pro Phe Glu
 2450 2455 2460
 GGT ATT GCT CTT GAT GAT CAG GGT ACA CTG AAT CTT CAA TTT CCG AAT 7440
 Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn Leu Gln Phe Pro Asn
 2465 2470 2475 2480
 GCT ACC GAC AAG CAG AAA GCA ATA TTG CAA ACT ATG AGC GAT ATT ATT 7488
 Ala Thr Asp Lys Lys Lys Ala Ile Leu Gln Thr Met Ser Asp Ile Ile
 2485 2490 2495
 TTG CAT ATT CGT TAT ACC ATC CGT TAA 7515
 Leu His Ile Arg Tyr Thr Ile Arg *
 2500 2505

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2504 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12 (TcbA protein):

Met Gln Asn Ser Leu Ser Ser Thr Ile Asp Thr Ile Cys Gln Lys Leu
 1 5 10 15
 Gln Leu Thr Cys Pro Ala Glu Ile Ala Leu Tyr Pro Phe Asp Thr Phe
 20 25 30
 Arg Glu Lys Thr Arg Gly Met Val Asn Trp Gly Glu Ala Lys Arg Ile
 35 40 45
 Tyr Glu Ile Ala Gln Ala Glu Gln Asp Arg Asn Leu Leu His Glu Lys
 50 55 60
 Arg Ile Phe Ala Tyr Ala Asn Pro Leu Leu Lys Asn Ala Val Arg Leu
 65 70 75 80
 Gly Thr Arg Gln Met Leu Gly Phe Ile Gln Gly Tyr Ser Asp Leu Phe
 85 90 95
 Gly Asn Arg Ala Asp Asn Tyr Ala Ala Pro Gly Ser Val Ala Ser Met
 100 105 110
 Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Lys Asn
 115 120 125
 Leu His Asp Ser Ser Ser Ile Tyr Tyr Leu Asp Lys Arg Arg Pro Asp
 130 135 140
 Leu Ala Ser Leu Met Leu Ser Gln Lys Asn Met Asp Glu Glu Ile Ser
 145 150 155 160
 Thr Leu Ala Leu Ser Asn Glu Leu Cys Leu Ala Gly Ile Glu Thr Lys
 165 170 175
 Thr Gly Lys Ser Gln Asp Glu Val Met Asp Met Leu Ser Thr Tyr Arg
 180 185 190
 Leu Ser Gly Glu Thr Pro Tyr His His Ala Tyr Glu Thr Val Arg Glu
 195 200 205
 Ile Val His Glu Arg Asp Pro Gly Phe Arg His Leu Ser Gln Ala Pro
 210 215 220
 Ile Val Ala Ala Lys Leu Asp Pro Val Thr Leu Leu Gly Ile Ser Ser
 225 230 235 240
 His Ile Ser Pro Glu Leu Tyr Asn Leu Leu Ile Glu Glu Ile Pro Glu
 245 250 255
 Lys Asp Glu Ala Ala Leu Asp Thr Leu Tyr Lys Thr Asn Phe Gly Asp
 260 265 270
 Ile Thr Thr Ala Gln Leu Met Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr
 275 280 285
 Gly Val Ser Pro Glu Asp Ile Ala Tyr Val Thr Thr Ser Leu Ser His
 290 295 300
 Val Gly Tyr Ser Ser Asp Ile Leu Val Ile Pro Leu Val Asp Gly Val
 305 310 315 320
 Gly Lys Met Glu Val Val Arg Val Thr Arg Thr Pro Ser Asp Asn Tyr
 325 330 335
 Thr Ser Gln Thr Asn Tyr Ile Glu Leu Tyr Pro Gln Gly Gly Asp Asn

340 345 350
 Tyr Leu Ile Lys Tyr Asn Leu Ser Asn Ser Phe Gly Leu Asp Asp Phe
 355 360 365
 Tyr Leu Gln Tyr Lys Asp Gly Ser Ala Asp Trp Thr Glu Ile Ala His
 370 375 380
 Asn Pro Tyr Pro Asp Met Val Ile Asn Gln Lys Tyr Glu Ser Gln Ala
 385 390 395 400
 Thr Ile Lys Arg Ser Asp Ser Asp Asn Ile Leu Ser Ile Gly Leu Gln
 405 410 415
 Arg Trp His Ser Gly Ser Tyr Asn Phe Ala Ala Ala Asn Phe Lys Ile
 420 425 430
 Asp Gln Tyr Ser Pro Lys Ala Phe Leu Leu Lys Met Asn Lys Ala Ile
 435 440 445
 Arg Leu Leu Lys Ala Thr Gly Leu Ser Phe Ala Thr Leu Glu Arg Ile
 450 455 460
 Val Asp Ser Val Asn Ser Thr Lys Ser Ile Thr Val Glu Val Leu Asn
 465 470 475 480
 Lys Val Tyr Arg Val Lys Phe Tyr Ile Asp Arg Tyr Gly Ile Ser Glu
 485 490 495
 Glu Thr Ala Ala Ile Leu Ala Asn Ile Asn Ile Ser Gln Gln Ala Val
 500 505 510
 Gly Asn Gln Leu Ser Gln Phe Glu Gln Leu Phe Asn His Pro Pro Leu
 515 520 525
 Asn Gly Ile Arg Tyr Glu Ile Ser Glu Asp Asn Ser Lys His Leu Pro
 530 535 540
 Asn Pro Asp Leu Asn Leu Lys Pro Asp Ser Thr Gly Asp Asp Gln Arg
 545 550 555 560
 Lys Ala Val Leu Lys Arg Ala Phe Gln Val Asn Ala Ser Glu Leu Tyr
 565 570 575
 Gln Met Leu Leu Ile Thr Asp Arg Lys Glu Asp Gly Val Ile Lys Asn
 580 585 590
 Asn Leu Glu Asn Leu Ser Asp Leu Tyr Leu Val Ser Leu Leu Ala Gln
 595 600 605
 Ile His Asn Leu Thr Ile Ala Glu Leu Asn Ile Leu Leu Val Ile Cys
 610 615 620
 Gly Tyr Gly Asp Thr Asn Ile Tyr Gln Ile Thr Asp Asp Asn Leu Ala
 625 630 635 640
 Lys Ile Val Glu Thr Leu Leu Trp Ile Thr Gln Trp Leu Lys Thr Gln
 645 650 655
 Lys Trp Thr Val Thr Asp Leu Phe Leu Met Thr Thr Ala Thr Tyr Ser
 660 665 670
 Thr Thr Leu Thr Pro Glu Ile Ser Asn Leu Thr Ala Thr Leu Ser Ser
 675 680 685
 Thr Leu His Gly Lys Glu Ser Leu Ile Gly Glu Asp Leu Lys Arg Ala
 690 695 700

Met Ala Pro Cys Phe Thr Ser Ala Leu His Leu Thr Ser Gln Glu Val
 705 710 715 720
 Ala Tyr Asp Leu Leu Leu Trp Ile Asp Gln Ile Gln Pro Ala Gln Ile
 725 730 735
 Thr Val Asp Gly Phe Trp Glu Glu Val Gln Thr Thr Pro Thr Ser Leu
 740 745 750
 Lys Val Ile Thr Phe Ala Gln Val Leu Ala Gln Leu Ser Leu Ile Tyr
 755 760 765
 Arg Arg Ile Gly Leu Ser Glu Thr Glu Leu Ser Leu Ile Val Thr Gln
 770 775 780
 Ser Ser Leu Leu Val Ala Gly Lys Ser Ile Leu Asp His Gly Leu Leu
 785 790 795 800
 Thr Leu Met Ala Leu Glu Gly Phe His Thr Trp Val Asn Gly Leu Gly
 805 810 815
 Gln His Ala Ser Leu Ile Leu Ala Ala Leu Lys Asp Gly Ala Leu Thr
 820 825 830
 Val Thr Asp Val Ala Gln Ala Met Asn Lys Glu Glu Ser Leu Leu Gln
 835 840 845
 Met Ala Ala Asn Gln Val Glu Lys Asp Leu Thr Lys Leu Thr Ser Trp
 850 855 860
 Thr Gln Ile Asp Ala Ile Leu Gln Trp Leu Gln Met Ser Ser Ala Leu
 865 870 875 880
 Ala Val Ser Pro Leu Asp Leu Ala Gly Met Met Ala Leu Lys Tyr Gly
 885 890 895
 Ile Asp His Asn Tyr Ala Ala Trp Gln Ala Ala Ala Ala Ala Leu Met
 900 905 910
 Ala Asp His Ala Asn Gln Ala Gln Lys Lys Leu Asp Glu Thr Phe Ser
 915 920 925
 Lys Ala Leu Cys Asn Tyr Tyr Ile Asn Ala Val Val Asp Ser Ala Ala
 930 935 940
 Gly Val Arg Asp Arg Asn Gly Leu Tyr Thr Tyr Leu Leu Ile Asp Asn
 945 950 955 960
 Gln Val Ser Ala Asp Val Ile Thr Ser Arg Ile Ala Glu Ala Ile Ala
 965 970 975
 Gly Ile Gln Leu Tyr Val Asn Arg Ala Leu Asn Arg Asp Glu Gly Gln
 980 985 990
 Leu Ala Ser Asp Val Ser Thr Arg Gln Phe Phe Thr Asp Trp Glu Arg
 995 1000 1005
 Tyr Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Glu Leu Val Tyr
 1010 1015 1020
 Tyr Pro Glu Asn Tyr Val Asp Pro Thr Gln Arg Ile Gly Gln Thr Lys
 1025 1030 1035 1040
 Met Met Asp Ala Leu Leu Gln Ser Ile Asn Gln Ser Gln Leu Asn Ala
 1045 1050 1055
 Asp Thr Val Glu Asp Ala Phe Lys Thr Tyr Leu Thr Ser Phe Glu Gln

Phe Ser Ser Lys Asp Asp Asn Lys Thr Ala Asp Tyr Asn Gly Gly Thr
 1425 1430 1435 1440
 Gln Cys Ile Asp Ala Gly Thr Ser Asn Lys Asp Phe Tyr Tyr Asn Leu
 1445 1450 1455
 Gln Glu Ile Glu Val Ile Ser Val Thr Gly Gly Tyr Trp Ser Ser Tyr
 1460 1465 1470
 Lys Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
 1475 1480 1485
 Val Lys Val Thr Val Lys Ala Gly Gly Asp Asp Gln Ile Phe Thr Ala
 1490 1495 1500
 Asp Asn Ser Thr Tyr Val Pro Gln Gln Pro Ala Pro Ser Phe Glu Glu
 1505 1510 1515 1520
 Met Ile Tyr Gln Phe Asn Asn Leu Thr Ile Asp Cys Lys Asn Leu Asn
 1525 1530 1535
 Phe Ile Asp Asn Gln Ala His Ile Glu Ile Asp Phe Thr Ala Thr Ala
 1540 1545 1550
 Gln Asp Gly Arg Phe Leu Gly Ala Glu Thr Phe Ile Ile Pro Val Thr
 1555 1560 1565
 Lys Lys Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn
 1570 1575 1580
 Asn Gly Val Gln Tyr Met Gln Ile Gly Ala Tyr Arg Thr Arg Leu Asn
 1585 1590 1595 1600
 Thr Leu Phe Ala Gln Gln Leu Val Ser Arg Ala Asn Arg Gly Ile Asp
 1605 1610 1615
 Ala Val Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly
 1620 1625 1630
 Ala Gly Thr Tyr Val Gln Leu Val Leu Asp Lys Tyr Asp Glu Ser Ile
 1635 1640 1645
 His Gly Thr Asn Lys Ser Phe Ala Ile Glu Tyr Val Asp Ile Phe Lys
 1650 1655 1660
 Glu Asn Asp Ser Phe Val Ile Tyr Gln Gly Glu Leu Ser Glu Thr Ser
 1665 1670 1675 1680
 Gln Thr Val Val Lys Val Phe Leu Ser Tyr Phe Ile Glu Ala Thr Gly
 1685 1690 1695
 Asn Lys Asn His Leu Trp Val Arg Ala Lys Tyr Gln Lys Glu Thr Thr
 1700 1705 1710
 Asp Lys Ile Leu Phe Asp Arg Thr Asp Glu Lys Asp Pro His Gly Trp
 1715 1720 1725
 Phe Leu Ser Asp Asp His Lys Thr Phe Ser Gly Leu Ser Ser Ala Gln
 1730 1735 1740
 Ala Leu Lys Asn Asp Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ala
 1745 1750 1755 1760
 Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Met Met Met Ala His
 1765 1770 1775
 Arg Leu Leu Gln Glu Gln Asn Phe Asp Ala Ala Asn His Trp Phe Arg

| | | |
|---|------|------|
| 1780 | 1785 | 1790 |
| Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val Asp Gly Lys Ile Ala Ile | | |
| 1795 | 1800 | 1805 |
| Tyr His Trp Asn Val Arg Pro Leu Glu Glu Asp Thr Ser Trp Asn Ala | | |
| 1810 | 1815 | 1820 |
| Gln Gln Leu Asp Ser Thr Asp Pro Asp Ala Val Ala Gln Asp Asp Pro | | |
| 1825 | 1830 | 1835 |
| 1840 | | |
| Met His Tyr Lys Val Ala Thr Phe Met Ala Thr Leu Asp Leu Leu Met | | |
| 1845 | 1850 | 1855 |
| Ala Arg Gly Asp Ala Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Ala | | |
| 1860 | 1865 | 1870 |
| Glu Ala Lys Met Trp Tyr Thr Gln Ala Leu Asn Leu Leu Gly Asp Glu | | |
| 1875 | 1880 | 1885 |
| Pro Gln Val Met Leu Ser Thr Thr Trp Ala Asn Pro Thr Leu Gly Asn | | |
| 1890 | 1895 | 1900 |
| Ala Ala Ser Lys Thr Thr Gln Gln Val Arg Gln Gln Val Leu Thr Gln | | |
| 1905 | 1910 | 1915 |
| 1920 | | |
| Leu Arg Leu Asn Ser Arg Val Lys Thr Pro Leu Leu Gly Thr Ala Asn | | |
| 1925 | 1930 | 1935 |
| Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn Ser Lys Leu Lys Gly | | |
| 1940 | 1945 | 1950 |
| Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn Leu Arg His Asn Leu | | |
| 1955 | 1960 | 1965 |
| Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu Tyr Ala Lys Pro Ala | | |
| 1970 | 1975 | 1980 |
| Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser Ala Ser Gln Gly Gly | | |
| 1985 | 1990 | 1995 |
| 2000 | | |
| Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His Arg Phe Pro Gln Met | | |
| 2005 | 2010 | 2015 |
| Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu Ile Gln Phe Gly Ser | | |
| 2020 | 2025 | 2030 |
| Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala Glu Ala Met Ser Gln | | |
| 2035 | 2040 | 2045 |
| Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu Thr Ser Ile Arg Met | | |
| 2050 | 2055 | 2060 |
| Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu Lys Thr Ala Leu Gln | | |
| 2065 | 2070 | 2075 |
| 2080 | | |
| Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp Ser Tyr Ser Gln Leu | | |
| 2085 | 2090 | 2095 |
| Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg Ala Leu Ala Leu Arg | | |
| 2100 | 2105 | 2110 |
| Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln Ile Ser Arg Met Ala | | |
| 2115 | 2120 | 2125 |
| Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe Gly Leu Ala Asp Gly | | |
| 2130 | 2135 | 2140 |

Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile Ala Asp Gly Ile Glu
 2145 2150 2155 2160
 Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu Lys Val Ala Gln Ser
 2165 2170 2175
 Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys Ile Gln Arg Asp Asn
 2180 2185 2190
 Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln Leu Glu Ser Leu Ser
 2195 2200 2205
 Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu Tyr Leu Lys Thr Gln
 2210 2215 2220
 Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu Arg Ser Lys Phe Ser
 2225 2230 2235 2240
 Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg Leu Ser Gly Ile Tyr
 2245 2250 2255
 Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys Leu Met Ala Glu Gln
 2260 2265 2270
 Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile Ser Phe Val Lys Pro
 2275 2280 2285
 Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu Cys Gly Glu Ala Leu
 2290 2295 2300
 Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr Leu Lys Trp Glu Ser
 2305 2310 2315 2320
 Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu Ala Val Val Tyr Asp
 2325 2330 2335
 Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala Glu Gln Ile Pro Ala
 2340 2345 2350
 Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr Lys Glu Asn Gly Leu
 2355 2360 2365
 Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val Lys Leu Ser Asp Leu
 2370 2375 2380
 Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val Gly Ser Asn Lys Val
 2385 2390 2395 2400
 Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro Ala Leu Val Gly Pro
 2405 2410 2415
 Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly Gly Ser Thr Gln Leu
 2420 2425 2430
 Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His Gly Thr Asn Asp Ser
 2435 2440 2445
 Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys Tyr Leu Pro Phe Glu
 2450 2455 2460
 Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn Leu Gln Phe Pro Asn
 2465 2470 2475 2480
 Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr Met Ser Asp Ile Ile
 2485 2490 2495
 Leu His Ile Arg Tyr Thr Ile Arg *

2500

2505

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13 (TcdAii N-terminus):

Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Xaa Ala
 1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14 (TcdB N-terminus):

Met Gln Asn Ser Gln Thr Phe Ser Val Gly Glu Leu
 1 5 10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15 (TcaAii N-terminus):

Ala Gln Asp Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr
 1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16 (TcbA N-terminus):

Met Gln Asn Ser Leu
 1 5

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17 (TcdAii-PTIII internal peptide):

Ala Phe Asn Ile Asp Asp Val Ser Leu Phe
 1 5 10

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18 (TcdAii- PT79 internal peptide):

Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19 (TcaBi- PT158 internal peptide):

Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile Gly Ser
 1 5 10 15
 Leu Gln Leu Phe Ile
 20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20 (TcaBi- PT 108 internal peptide):

Met Tyr Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro
 1 5 10

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21 (TcbAii- PT103

internal peptide):

Gly Ile Asp Ala Val Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro
 1 5 10 15
 Gln Leu Gly Ala Gly Thr Tyr Val Gln Leu
 20 25

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22 (TcbAii- PT56 internal

peptide):

Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23 (TcbA- PT81 (a)

internal peptide):

Thr Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys
 1 5 10

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24 (TcbAii- PT81 (b)

internal peptide):

Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn Asn Gly
 1 5 10 15
 Val Gln Tyr Met Gln Ile
 20

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6054 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..43
- (D) OTHER INFORMATION: /product= "end of TcaAiii"

(ix) FEATURE:

- (A) NAME/KEY: RBS
- (B) LOCATION: 51 58

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 65 3634
- (D) OTHER INFORMATION: /product= "TcaBi"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```

A GTA GCC CAA AAC TTA AGT GCC GCA ATC AGC AAT CGT CAG TAACCGGATA 50
  Val Ala Gln Asn Leu Ser Ala Ala Ile Ser Asn Arg Gln ???
AAGAAGGAAT TGATT ATG TCT GAA TCT TTA TTT ACA CAA ACG TTG AAA GAA 100
      Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu
          1           5           10
GCG CGC CGT GAT GCA TTG GTT GCT CAT TAT ATT GCT ACT CAG GTG CCC 148
Ala Arg Arg Asp Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro
      15           20           25
GCA GAT TTA AAA GAG AGT ATC CAG ACC GCG GAT GAT CTG TAC GAA TAT 196
Ala Asp Leu Lys Glu Ser Ile Gln Thr Ala Asp Asp Leu Tyr Glu Tyr
      30           35           40
CTG TTG CTG GAT ACC AAA ATT AGC GAT CTG GTT ACT ACT TCA CCG CTG 244
Leu Leu Leu Asp Thr Lys Ile Ser Asp Leu Val Thr Thr Ser Pro Leu
      45           50           55           60
TCC GAA GCG ATT GGC AGT CTG CAA TTG TTT ATT CAT CGT GCG ATA GAG 292
Ser Glu Ala Ile Gly Ser Leu Gln Leu Phe Ile His Arg Ala Ile Glu
          65           70           75
GGC TAT GAC GGC ACG CTG GCA GAC TCA GCA AAA CCC TAT TTT GCC GAT 340
Gly Tyr Asp Gly Thr Leu Ala Asp Ser Ala Lys Pro Tyr Phe Ala Asp
          80           85           90
GAA CAG TTT TTA TAT AAC TGG GAT AGT TTT AAC CAC CGT TAT AGC ACT 388
Glu Gln Phe Leu Tyr Asn Trp Asp Ser Phe Asn His Arg Tyr Ser Thr
          95           100          105
TGG GCT GGC AAG GAA CGG TTG AAA TTC TAT GCC GGG GAT TAT ATT GAT 436
Trp Ala Gly Lys Glu Arg Leu Lys Phe Tyr Ala Gly Asp Tyr Ile Asp
      110           115           120

```

CCA ACA TTG CGA TTG AAT AAG ACC GAG ATA TTT ACC GCA TTT GAA CAA 484
 Pro Thr Leu Arg Leu Asn Lys Thr Glu Ile Phe Thr Ala Phe Glu Gln
 125 130 135 140

GGT ATT TCT CAA GGG AAA TTA AAA AGT GAA TTA GTC GAA TCT AAA TTA 532
 Gly Ile Ser Gln Gly Lys Leu Lys Ser Glu Leu Val Glu Ser Lys Leu
 145 150 155

CGT GAT TAT CTA ATT AGT TAT GAC ACT TTA GCC ACC CTT GAT TAT ATT 580
 Arg Asp Tyr Leu Ile Ser Tyr Asp Thr Leu Ala Thr Leu Asp Tyr Ile
 160 165 170

ACT GCC TGC CAA GGC AAA GAT AAT AAA ACC ATC TTC TTT ATT GGC CGT 628
 Thr Ala Cys Gln Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg
 175 180 185

ACA CAG AAT GCA CCC TAT GCA TTT TAT TGG CGA AAA TTA ACT TTA GTC 676
 Thr Gln Asn Ala Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val
 190 195 200

ACT GAT GGC GGT AAG TTG AAA CCA GAT CAA TGG TCA GAG TGG CGA GCA 724
 Thr Asp Gly Gly Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala
 205 210 215 220

ATT AAT GCC GGG ATT AGT GAG GCA TAT TCA GGG CAT GTC GAG CCT TTC 772
 Ile Asn Ala Gly Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe
 225 230 235

TGG GAA AAT AAC AAG CTG CAC ATC CGT TGG TTT ACT ATC TCG AAA GAA 820
 Trp Glu Asn Asn Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu
 240 245 250

GAT AAA ATA GAT TTT GTT TAT AAA AAC ATC TGG GTG ATG AGT AGC GAT 868
 Asp Lys Ile Asp Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp
 255 260 265

TAT AGC TGG GCA TCA AAG AAA AAA ATC TTG GAA CTT TCT TTT ACT GAC 916
 Tyr Ser Trp Ala Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp
 270 275 280

TAC AAT AGA GTT GGA GCA ACA GGA TCA TCA AGC CCG ACT GAA GTA GCT 964
 Tyr Asn Arg Val Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala
 285 290 295 300

TCA CAA TAT GGT TCT GAT GCT CAG ATG AAT ATT TCT GAT GAT GGG ACT 1012
 Ser Gln Tyr Gly Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr
 305 310 315

GTA CTT ATT TTT CAG AAT GCC GGC GGA GCT ACT CCC AGT ACT GGA GTG 1060
 Val Leu Ile Phe Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val
 320 325 330

ACG TTA TGT TAT GAC TCT GGC AAC GTG ATT AAG AAC CTA TCT AGT ACA 1108
 Thr Leu Cys Tyr Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr
 335 340 345

GGA AGT GCA AAT TTA TCG TCA AAG GAT TAT GCC ACA ACT AAA TTA CGC 1156
 Gly Ser Ala Asn Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg
 350 355 360

ATG TGT CAT GGA CAA AGT TAC AAT GAT AAT AAC TAC TGC AAT TTT ACA 1204
 Met Cys His Gly Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr
 365 370 375 380
 CTC TCT ATT AAT ACA ATA GAA TTC ACC TCC TAC GGC ACA TTC TCA TCA 1252
 Leu Ser Ile Asn Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser
 385 390 395
 GAT GGA AAA CAA TTT ACA CCA CCT TCT GGT TCT GCC ATT GAT TTA CAC 1300
 Asp Gly Lys Gln Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His
 400 405 410
 CTC CCT AAT TAT GTA GAT CTC AAC GCG CTA TTA GAT ATT AGC CTC GAT 1348
 Leu Pro Asn Tyr Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp
 415 420 425
 TCA CTA CTT AAT TAT GAC GTT CAG GGG CAG TTT GGC GGA TCT AAT CCG 1396
 Ser Leu Leu Asn Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro
 430 435 440
 GTT GAT AAT TTC AGT GGT CCC TAT GGT ATT TAT CTA TGG GAA ATC TTC 1444
 Val Asp Asn Phe Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe
 445 450 455 460
 TTC CAT ATT CCG TTC CTT GTT ACG GTC CGT ATG CAA ACC GAA CAA CGT 1492
 Phe His Ile Pro Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg
 465 470 475
 TAC GAA GAC GCG GAC ACT TGG TAC AAA TAT ATT TTC CGC AGC GCC GGT 1540
 Tyr Glu Asp Ala Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly
 480 485 490
 TAT CGC GAT GCT AAT GGC CAG CTC ATT ATG GAT GGC AGT AAA CCA CGT 1588
 Tyr Arg Asp Ala Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg
 495 500 505
 TAT TGG AAT GTG ATG CCA TTG CAA CTG GAT ACC GCA TGG GAT ACC ACA 1636
 Tyr Trp Asn Val Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr
 510 515 520
 CAG CCC GCC ACC ACT GAT CCA GAT GTG ATC GCT ATG GCG GAC CCG ATG 1684
 Gln Pro Ala Thr Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met
 525 530 535 540
 CAT TAC AAG CTG GCG ATA TTC CTG CAT ACC CTT GAT CTA TTG ATT GCC 1732
 His Tyr Lys Leu Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala
 545 550 555
 CGA GGC GAC AGC GCT TAC CGT CAA CTT GAA CGC GAT ACT CTA GTC GAA 1780
 Arg Gly Asp Ser Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu
 560 565 570
 GCC AAA ATG TAC TAC ATT CAG GCA CAA CAG CTA CTG GGA CCG CGC CCT 1828
 Ala Lys Met Tyr Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro
 575 580 585
 GAT ATC CAT ACC ACC AAT ACT TGG CCA AAT CCC ACC TTG AGT AAA GAA 1876
 Asp Ile His Thr Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu
 590 595 600

GCT GGC GCT ATT GCC ACA CCG ACA TTC CTC AGT TCA CCG GAG GTG ATG 1924
Ala Gly Ala Ile Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met
605 610 615 620

ACG TTC GCT GCC TGG CTA AGC GCA GGC GAT ACC GCA AAT ATT GGC GAC 1972
Thr Phe Ala Ala Trp Leu Ser Ala Gly Asp Thr Ala Asn Ile Gly Asp
625 630 635

GGT GAT TTC TTG CCA CCG TAC AAC GAT GTA CTA CTC GGT TAC TGG GAT 2020
Gly Asp Phe Leu Pro Pro Tyr Asn Asp Val Leu Leu Gly Tyr Trp Asp
640 645 650

AAA CTT GAG TTA CGC CTA TAC AAC CTG CGC CAC AAT CTG AGT CTG GAT 2068
Lys Leu Glu Leu Arg Leu Tyr Asn Leu Arg His Asn Leu Ser Leu Asp
655 660 665

GGT CAA CCG CTA AAT CTG CCA CTG TAT GCC ACG CCG GTA GAC CCG AAA 2116
Gly Gln Pro Leu Asn Leu Pro Leu Tyr Ala Thr Pro Val Asp Pro Lys
670 675 680

ACC CTG CAA CGC CAG CAA GCC GGA GGG GAC GGT ACA GGC AGT AGT CCG 2164
Thr Leu Gln Arg Gln Gln Ala Gly Gly Asp Gly Thr Gly Ser Ser Pro
685 690 695 700

GCT GGT GGT CAA GGC AGT GTT CAG GGC TGG CGC TAT CCG TTA TTG GTA 2212
Ala Gly Gly Gln Gly Ser Val Gln Gly Trp Arg Tyr Pro Leu Leu Val
705 710 715

GAA CGC GCC CGC TCT GCC GTG AGT TTG TTG ACT CAG TTC GGC AAC AGC 2260
Glu Arg Ala Arg Ser Ala Val Ser Leu Leu Thr Gln Phe Gly Asn Ser
720 725 730

TTA CAA ACA ACG TTA GAA CAT CAG GAT AAT GAA AAA ATG ACG ATA CTG 2308
Leu Gln Thr Thr Leu Glu His Gln Asp Asn Glu Lys Met Thr Ile Leu
735 740 745

TTG CAG ACT CAA CAG GAA GCC ATC CTG AAA CAT CAG CAC GAT ATA CAA 2356
Leu Gln Thr Gln Gln Glu Ala Ile Leu Lys His Gln His Asp Ile Gln
750 755 760

CAA AAT AAT CTA AAA GGA TTA CAA CAC AGC CTG ACC GCA TTA CAG GCT 2404
Gln Asn Asn Leu Lys Gly Leu Gln His Ser Leu Thr Ala Leu Gln Ala
765 770 775 780

AGC CGT GAT GGC GAC ACA TTG CGG CAA AAA CAT TAC AGC GAC CTG ATT 2452
Ser Arg Asp Gly Asp Thr Leu Arg Gln Lys His Tyr Ser Asp Leu Ile
785 790 795

AAC GGT GGT CTA TCT GCG GCA GAA ATC GCC GGT CTG ACA CTA CGC AGC 2500
Asn Gly Gly Leu Ser Ala Ala Glu Ile Ala Gly Leu Thr Leu Arg Ser
800 805 810

ACC GCC ATG ATT ACC AAT GGC GTT GCA ACG GGA TTG CTG ATT GCC GGC 2548
Thr Ala Met Ile Thr Asn Gly Val Ala Thr Gly Leu Leu Ile Ala Gly
815 820 825

GGA ATC GCC AAC GCG GTA CCT AAC GTC TTC GGG CTG GCT AAC GGT GGA 2596
Gly Ile Ala Asn Ala Val Pro Asn Val Phe Gly Leu Ala Asn Gly Gly
830 835 840

TCG GAA TGG GGA GCG CCA TTA ATT GGC TCC GGG CAA GCA ACC CAA GTT 2644
 Ser Glu Trp Gly Ala Pro Leu Ile Gly Ser Gly Gln Ala Thr Gln Val
 845 850 855 860

GGC GCC GGC ATC CAG GAT CAG AGC GCG GGC ATT TCA GAA GTG ACA GCA 2692
 Gly Ala Gly Ile Gln Asp Gln Ser Ala Gly Ile Ser Glu Val Thr Ala
 865 870 875

GGC TAT CAG CGT CGT CAG GAA GAA TGG GCA TTG CAA CGG GAT ATT GCT 2740
 Gly Tyr Gln Arg Arg Gln Glu Glu Trp Ala Leu Gln Arg Asp Ile Ala
 880 885 890

GAT AAC GAA ATA ACC CAA CTG GAT GCC CAG ATA CAA AGC CTG CAA GAG 2788
 Asp Asn Glu Ile Thr Gln Leu Asp Ala Gln Ile Gln Ser Leu Gln Glu
 895 900 905

CAA ATC ACG ATG GCA CAA AAA CAG ATC ACG CTC TCT GAA ACC GAA CAA 2836
 Gln Ile Thr Met Ala Gln Lys Gln Ile Thr Leu Ser Glu Thr Glu Gln
 910 915 920

GCG AAT GCC CAA GCG ATT TAT GAC CTG CAA ACC ACT CGT TTT ACC GGG 2884
 Ala Asn Ala Gln Ala Ile Tyr Asp Leu Gln Thr Thr Arg Phe Thr Gly
 925 930 935 940

CAG GCA CTG TAT AAC TGG ATG GCC GGT CGT CTC TCC GCG CTC TAT TAC 2932
 Gln Ala Leu Tyr Asn Trp Met Ala Gly Arg Leu Ser Ala Leu Tyr Tyr
 945 950 955

CAA ATG TAT GAT TCC ACT CTG CCA ATC TGT CTC CAG CCA AAA GCC GCA 2980
 Gln Met Tyr Asp Ser Thr Leu Pro Ile Cys Leu Gln Pro Lys Ala Ala
 960 965 970

TTA GTA CAG GAA TTA GGC GAG AAA GAG AGC GAC AGT CTT TTC CAG GTT 3028
 Leu Val Gln Glu Leu Gly Glu Lys Glu Ser Asp Ser Leu Phe Gln Val
 975 980 985

CCG GTG TGG AAT GAT CTG TGG CAA GGG CTG TTA GCA GGA GAA GGT TTA 3076
 Pro Val Trp Asn Asp Leu Trp Gln Gly Leu Leu Ala Gly Glu Gly Leu
 990 995 1000

AGT TCA GAG CTA CAG AAA CTG GAT GCC ATC TGG CTT GCA CGT GGT GGT 3124
 Ser Ser Glu Leu Gln Lys Leu Asp Ala Ile Trp Leu Ala Arg Gly Gly
 1005 1010 1015 1020

ATT GGG CTA GAA GCC ATC CGC ACC GTG TCG CTG GAT ACC CTG TTT GGC 3172
 Ile Gly Leu Glu Ala Ile Arg Thr Val Ser Leu Asp Thr Leu Phe Gly
 1025 1030 1035

ACA GGG ACG TTA AGT GAA AAT ATC AAT AAA GTG CTT AAC GGG GAA ACG 3220
 Thr Gly Thr Leu Ser Glu Asn Ile Asn Lys Val Leu Asn Gly Glu Thr
 1040 1045 1050

GTA TCT CCA TCC GGT GGC GTC ACT CTG GCG CTG ACA GGG GAT ATC TTC 3268
 Val Ser Pro Ser Gly Gly Val Thr Leu Ala Leu Thr Gly Asp Ile Phe
 1055 1060 1065

CAA GCA ACA CTG GAT TTG AGT CAG CTA GGT TTG GAT AAC TCT TAC AAC 3316
 Gln Ala Thr Leu Asp Leu Ser Gln Leu Gly Leu Asp Asn Ser Tyr Asn
 1070 1075 1080

TTG GGT AAC GAG AAG AAA CGT CGT ATT AAA CGT ATC GCC GTC ACC CTG 3364
 Leu Gly Asn Glu Lys Lys Arg Arg Ile Lys Arg Ile Ala Val Thr Leu
 1085 1090 1095 1100

CCA ACA CTT CTG GGG CCA TAT CAA GAT CTT GAA GCC ACA CTG GTA ATG 3412
 Pro Thr Leu Leu Gly Pro Tyr Gln Asp Leu Glu Ala Thr Leu Val Met
 1105 1110 1115

GGT GCG GAA ATC GCC GCC TTA TCA CAC GGT GTG AAT GAC GGA GGC CGG 3460
 Gly Ala Glu Ile Ala Ala Leu Ser His Gly Val Asn Asp Gly Gly Arg
 1120 1125 1130

TTT GTT ACC GAC TTT AAC GAC AGC CGT TTT CTG CCT TTT GAA GGT CGA 3508
 Phe Val Thr Asp Phe Asn Asp Ser Arg Phe Leu Pro Phe Glu Gly Arg
 1135 1140 1145

GAT GCA ACA ACC GGC ACA CTG GAG CTC AAT ATT TTC CAT GCG GGT AAA 3556
 Asp Ala Thr Thr Gly Thr Leu Glu Leu Asn Ile Phe His Ala Gly Lys
 1150 1155 1160

GAG GGA ACG CAA CAC GAG TTG GTC GCG AAT CTG AGT GAC ATC ATT GTG 3604
 Glu Gly Thr Gln His Glu Leu Val Ala Asn Leu Ser Asp Ile Ile Val
 1165 1170 1175 1180

CAT CTG AAT TAC ATC ATT CGA GAC GCG TAA ATTTCTTTTC TTTGTCGATT 3654
 His Leu Asn Tyr Ile Ile Arg Asp Ala *
 1185 1190

ACAGGTCCCT ATCAGGGGCC TGTTATTAAG GAGTACTTTA TGCAGGATTC ACCAGAAGTA 3714
 TCGATTACAA CGCTGTCACT TCCCAAAGGT GGCAGTGCTA TCAATGGCAT GGGAGAAGCA 3774
 CTGAATGCTG CCGGCCCTGA TGAATGGCC TCCCTATCTC TGCCATTACC CCTTTCGACC 3834
 GGCAGAGGGA CGGCTCCTGG ATTATCGCTG ATTTACAGCA ACAGTGCAGG TAATGGGCCT 3894
 TTCGGCATCG GCTGGCAATG CGGTGTTATG TCCATTAGCC GACGCACCCA ACATGGCATT 3954
 CCACAATACG GTAATGACGA CACGTTCTTA TCCCACAAG GCGAGGTCAT GAATATCGCC 4014
 CTGAATGACC AAGGGCAACC TGATATCCGT CAAGACGTTA AAACGCTGCA AGGCGTTACC 4074
 TTGCCAATTT CCTATACCGT GACCCGCTAT CAAGCCCGCC AGATCCTGGA TTTCAGTAAA 4134
 ATCGAATACT GGCAACCTGC CTCGGTCAA GAAGGACGCG CTTTCTGGCT GATATCGACA 4194
 CCGGACGGGC ATCTACACAT CTTAGGGAAA ACCGCGCAGG CTTGTCTGGC AAATCCGCAA 4254
 AATGACCAAC AAATCGCCCA GTGGTTGTG GAAGAACTG TGACGCCAGC CGGTGAACAT 4314
 GTCAGCTATC AATATCGAGC CGAAGATGAA GCCCATTGTG ACGACAATGA AAAAACCCT 4374
 CATCCCAATG TTACCGCACA GCGCTATCTG GTACAGGTGA ACTACAGGCA ACATCAAACC 4434
 ACAAGCCAGC CTGTTCTGAC TGGATAACGC ACCTCCCGCA CCGGAAGAGT GGCTGTTTCA 4494
 TCTGGTCTTT GACCACGGTG AGCGCGTACC TCACTTCATA CCGTGCCAAC ATGGGATGCA 4554
 GGTACAGCGC AATGGTCTGT ACGCCCGGAT ATCTTCTCTC GCTATGAATA TGGTTTTGAA 4614
 GTGGTACTC GCCGTTATG TCAACAAGTG CTGATGTTTC ACCGCACCGC GCTCATGGCC 4674
 GGAGAAGCCA GTACCAATGA CGCCCCGAA CTGGTTGGAC GCTTAATACT GGAATATGAC 4634
 AAAAACGCCA GCGTACCAC GTTGATTACC ATCCGTCAAT TAAGCCATGA ATCGGACGGG 4794
 AGGCCAGTCA CCCAGCCACC ACTAGAACTA GCCTGGCAAC GGTGTTGATCT GGAGAAAATC 4854
 CCGCATGGC AACGCTTTGA CGCACTAGAT AATTTTAACT CGCAGCAACG TTATCAACTG 4865
 GTTGATCTGC GGGGAGAAGG GTTGCCAGGT ATGCTGTATC AAGATCGAGG CGCTTGGTGG 4914
 TATAAAGCTC CGCAACGTCA GGAAGACGGA GACAGCAATG CCGTCACTTA CGACAAAATC 4974
 GCCCCTACTG CTACCTACC CAATTTGCAG GATAATGCCT CATTGATGGA TATCAACGGA 5034

GACGGCCAAC TGGATTGGGT TGTTACCGCC TCCGGTATTC GCGGATACCA TAGTCAGCAA 5094
 CCCGATGGAA AGTGGACGCA CTTTACGCCA ATCAATGCCT TGCCCGTGA ATATTTTCAT 5214
 CCAAGCATCC AGTTCGCTGA CCTTACCGGG GCAGGCTTAT CTGATTTAGT GTTGATCGGG 5274
 CCGAAAAGCG TGCCTCTATA TGCCAACCAG CGAAACGGCT GCGGTAAAGG AGAAGATGTC 5334
 CCCCAATCCA CAGGTATCAC CCTGCCTGTC ACAGGGACCG ATGCCCGCAA ACTGGTGGCT 5394
 TTCAGTGATA TGCTCGGTTT CCGTCAACAA CATCTGGTGG AAATCAAGGG TAATCGCGTC 5454
 ACCTGTTGGC CGAATCTAGG GCATGGCCGT TTCGGTCAAC CACTAACTCT GTCAGGATTT 5514
 AGCCAGCCCG AAAATAGCTT CAATCCCGAA CGGCTGTTTC TGGCGGATAT CGACGGCTCC 5574
 GGCACCACCG ACCTTATCTA TGCGAATCC GGCTCTTTGC TCATTTATCT CAACCAAAGT 5634
 GGTAATCAGT TTGATGCCCC GTTGACATTA GCGTTGCCAG AAGGCGTACA ATTTGACAAC 5694
 ACTTGCCAAC TTCAAGTCGC CGATATTCAG GGATTAGGGA TAGCCAGCTT GATTCTGACT 5754
 GTGCCACATA TCGCGCCACA TCACTGGCGT TGTGACCTGT CACTGACCAA ACCCTGGTTG 5814
 TTGAATGTAA TGAACAATAA CCGGGGCGCA CATCACACGC TACATTATCG TAGTCCGCG 5874
 CAATTCTGGT TGGATGAAAA ATTACAGCTC ACCAAAGCAG GCAAATCTCC GGCTTGTAT 5934
 CTGCCGTTTC CAATGCATTT GCTATGGTAT ACCGAAATTC AGGATGAAAT CAGCGGCAAC 5994
 CGGCTCACCA GTGAAGTCAA CTACAGCCAC GCGCTCTGGG ATGGTAAAGA GCGGGAATTC 6054

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1189 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26 (TcaB protein):

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

Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg Thr Gln Asn Ala
 180 185 190
 Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val Thr Asp Gly Gly
 195 200 205
 Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala Ile Asn Ala Gly
 210 215 220
 Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe Trp Glu Asn Asn
 225 230 235 240
 Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu Asp Lys Ile Asp
 245 250 255
 Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp Tyr Ser Trp Ala
 260 265 270
 Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp Tyr Asn Arg Val
 275 280 285
 Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala Ser Gln Tyr Gly
 290 295 300
 Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr Val Leu Ile Phe
 305 310 315 320
 Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val Thr Leu Cys Tyr
 325 330 335
 Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr Gly Ser Ala Asn
 340 345 350
 Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg Met Cys His Gly
 355 360 365
 Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr Leu Ser Ile Asn
 370 375 380
 Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser Asp Gly Lys Gln
 385 390 395 400
 Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His Leu Pro Asn Tyr
 405 410 415
 Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp Ser Leu Leu Asn
 420 425 430
 Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe
 435 440 445
 Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro
 450 455 460
 Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala
 465 470 475 480
 Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala
 485 490 495
 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val
 500 505 510
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr
 515 520 525
 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu

530 535 540
 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser
 545 550 555 560
 Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu Ala Lys Met Tyr
 565 570 575
 Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr
 580 585 590
 Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile
 595 600 605
 Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala
 610 615 620
 Trp Leu Ser Ala Gly Asp Thr Ala Asn Ile Gly Asp Gly Asp Phe Leu
 625 630 635 640
 Pro Pro Tyr Asn Asp Val Leu Leu Gly Tyr Trp Asp Lys Leu Glu Leu
 645 650 655
 Arg Leu Tyr Asn Leu Arg His Asn Leu Ser Leu Asp Gly Gln Pro Leu
 660 665 670
 Asn Leu Pro Leu Tyr Ala Thr Pro Val Asp Pro Lys Thr Leu Gln Arg
 675 680 685
 Gln Gln Ala Gly Gly Asp Gly Thr Gly Ser Ser Pro Ala Gly Gly Gln
 690 695 700
 Gly Ser Val Gln Gly Trp Arg Tyr Pro Leu Leu Val Glu Arg Ala Arg
 705 710 715 720
 Ser Ala Val Ser Leu Leu Thr Gln Phe Gly Asn Ser Leu Gln Thr Thr
 725 730 735
 Leu Glu His Gln Asp Asn Glu Lys Met Thr Ile Leu Leu Gln Thr Gln
 740 745 750
 Gln Glu Ala Ile Leu Lys His Gln His Asp Ile Gln Gln Asn Asn Leu
 755 760 765
 Lys Gly Leu Gln His Ser Leu Thr Ala Leu Gln Ala Ser Arg Asp Gly
 770 775 780
 Asp Thr Leu Arg Gln Lys His Tyr Ser Asp Leu Ile Asn Gly Gly Leu
 785 790 795 800
 Ser Ala Ala Glu Ile Ala Gly Leu Thr Leu Arg Ser Thr Ala Met Ile
 805 810 815
 Thr Asn Gly Val Ala Thr Gly Leu Leu Ile Ala Gly Gly Ile Ala Asn
 820 825 830
 Ala Val Pro Asn Val Phe Gly Leu Ala Asn Gly Gly Ser Glu Trp Gly
 835 840 845
 Ala Pro Leu Ile Gly Ser Gly Gln Ala Thr Gln Val Gly Ala Gly Ile
 850 855 860
 Gln Asp Gln Ser Ala Gly Ile Ser Glu Val Thr Ala Gly Tyr Gln Arg
 865 870 875 880
 Arg Gln Glu Glu Trp Ala Leu Gln Arg Asp Ile Ala Asp Asn Glu Ile
 885 890 895

Thr Gln Leu Asp Ala Gln Ile Gln Ser Leu Gln Glu Gln Ile Thr Met
 900 905 910
 Ala Gln Lys Gln Ile Thr Leu Ser Glu Thr Glu Gln Ala Asn Ala Gln
 915 920 925
 Ala Ile Tyr Asp Leu Gln Thr Thr Arg Phe Thr Gly Gln Ala Leu Tyr
 930 935 940
 Asn Trp Met Ala Gly Arg Leu Ser Ala Leu Tyr Tyr Gln Met Tyr Asp
 945 950 955 960
 Ser Thr Leu Pro Ile Cys Leu Gln Pro Lys Ala Ala Leu Val Gln Glu
 965 970 975
 Leu Gly Glu Lys Glu Ser Asp Ser Leu Phe Gln Val Pro Val Trp Asn
 980 985 990
 Asp Leu Trp Gln Gly Leu Leu Ala Gly Glu Gly Leu Ser Ser Glu Leu
 995 1000 1005
 Gln Lys Leu Asp Ala Ile Trp Leu Ala Arg Gly Gly Ile Gly Leu Glu
 1010 1015 1020
 Ala Ile Arg Thr Val Ser Leu Asp Thr Leu Phe Gly Thr Gly Thr Leu
 1025 1030 1035 1040
 Ser Glu Asn Ile Asn Lys Val Leu Asn Gly Glu Thr Val Ser Pro Ser
 1045 1050 1055
 Gly Gly Val Thr Leu Ala Leu Thr Gly Asp Ile Phe Gln Ala Thr Leu
 1060 1065 1070
 Asp Leu Ser Gln Leu Gly Leu Asp Asn Ser Tyr Asn Leu Gly Asn Glu
 1075 1080 1085
 Lys Lys Arg Arg Ile Lys Arg Ile Ala Val Thr Leu Pro Thr Leu Leu
 1090 1095 1100
 Gly Pro Tyr Gln Asp Leu Glu Ala Thr Leu Val Met Gly Ala Glu Ile
 1105 1110 1115 1120
 Ala Ala Leu Ser His Gly Val Asn Asp Gly Gly Arg Phe Val Thr Asp
 1125 1130 1135
 Phe Asn Asp Ser Arg Phe Leu Pro Phe Glu Gly Arg Asp Ala Thr Thr
 1140 1145 1150
 Gly Thr Leu Glu Leu Asn Ile Phe His Ala Gly Lys Glu Gly Thr Gln
 1155 1160 1165
 His Glu Leu Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr
 1170 1175 1180
 Ile Ile Arg Asp Ala *
 1185 1190

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1881 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1881

(D) OTHER INFORMATION: tcaBi

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27 (tcaBi coding region):

```

ATG TCT GAA TCT TTA TTT ACA CAA ACG TTG AAA GAA GCG CGC CGT GAT 48
Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp
  1           5           10          15
GCA TTG GTT GCT CAT TAT ATT GCT ACT CAG GTG CCC GCA GAT TTA AAA 96
Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro Ala Asp Leu Lys
  20          25          30
GAG AGT ATC CAG ACC GCG GAT GAT CTG TAC GAA TAT CTG TTG CTG GAT 144
Glu Ser Ile Gln Thr Ala Asp Asp Leu Tyr Glu Tyr Leu Leu Leu Asp
  35          40          45
ACC AAA ATT AGC GAT CTG GTT ACT ACT TCA CCG CTG TCC GAA GCG ATT 192
Thr Lys Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile
  50          55          60
GGC AGT CTG CAA TTG TTT ATT CAT CGT GCG ATA GAG GGC TAT GAC GGC 240
Gly Ser Leu Gln Leu Phe Ile His Arg Ala Ile Glu Gly Tyr Asp Gly
  65          70          75          80
ACG CTG GCA GAC TCA GCA AAA CCC TAT TTT GCC GAT GAA CAG TTT TTA 288
Thr Leu Ala Asp Ser Ala Lys Pro Tyr Phe Ala Asp Glu Gln Phe Leu
  85          90          95
TAT AAC TGG GAT AGT TTT AAC CAC CGT TAT AGC ACT TGG GCT GGC AAG 336
Tyr Asn Trp Asp Ser Phe Asn His Arg Tyr Ser Thr Trp Ala Gly Lys
 100         105         110
GAA CGG TTG AAA TTC TAT GCC GGG GAT TAT ATT GAT CCA ACA TTG CGA 384
Glu Arg Leu Lys Phe Tyr Ala Gly Asp Tyr Ile Asp Pro Thr Leu Arg
 115         120         125
TTG AAT AAG ACC GAG ATA TTT ACC GCA TTT GAA CAA GGT ATT TCT CAA 432
Leu Asn Lys Thr Glu Ile Phe Thr Ala Phe Glu Gln Gly Ile Ser Gln
 130         135         140
GGG AAA TTA AAA AGT GAA TTA GTC GAA TCT AAA TTA CGT GAT TAT CTA 480
Gly Lys Leu Lys Ser Glu Leu Val Glu Ser Lys Leu Arg Asp Tyr Leu
 145         150         155         160
ATT AGT TAT GAC ACT TTA GCC ACC CTT GAT TAT ATT ACT GCC TGC CAA 528
Ile Ser Tyr Asp Thr Leu Ala Thr Leu Asp Tyr Ile Thr Ala Cys Gln
 165         170         175
GGC AAA GAT AAT AAA ACC ATC TTC TTT ATT GGC CGT ACA CAG AAT GCA 576
Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg Thr Gln Asn Ala
 180         185         190
CCC TAT GCA TTT TAT TGG CGA AAA TTA ACT TTA GTC ACT GAT GGC GGT 624
Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val Thr Asp Gly Gly
 195         200         205
AAG TTG AAA CCA GAT CAA TGG TCA GAG TGG CGA GCA ATT AAT GCC GGC 672

```

Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala Ile Asn Ala Gly
 210 215 220

ATT AGT GAG GCA TAT TCA GGG CAT GTC GAG CCT TTC TGG GAA AAT AAC 720
 Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe Trp Glu Asn Asn
 225 230 235 240

AAG CTG CAC ATC CGT TGG TTT ACT ATC TCG AAA GAA GAT AAA ATA GAT 768
 Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu Asp Lys Ile Asp
 245 250 255

TTT GTT TAT AAA AAC ATC TGG GTG ATG AGT AGC GAT TAT AGC TGG GCA 816
 Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp Tyr Ser Trp Ala
 260 265 270

TCA AAG AAA AAA ATC TTG GAA CTT TCT TTT ACT GAC TAC AAT AGA GTT 864
 Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp Tyr Asn Arg Val
 275 280 285

GGA GCA ACA GGA TCA TCA AGC CCG ACT GAA GTA GCT TCA CAA TAT GGT 912
 Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala Ser Gln Tyr Gly
 290 295 300

TCT GAT GCT CAG ATG AAT ATT TCT GAT GAT GGG ACT GTA CTT ATT TTT 960
 Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr Val Leu Ile Phe
 305 310 315 320

CAG AAT GCC GGC GGA GCT ACT CCC AGT ACT GGA GTG ACG TTA TGT TAT 1008
 Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val Thr Leu Cys Tyr
 325 330 335

GAC TCT GGC AAC GTG ATT AAG AAC CTA TCT AGT ACA GGA AGT GCA AAT 1056
 Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr Gly Ser Ala Asn
 340 345 350

TTA TCG TCA AAG GAT TAT GCC ACA ACT AAA TTA CGC ATG TGT CAT GGA 1104
 Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg Met Cys His Gly
 355 360 365

CAA AGT TAC AAT GAT AAT AAC TAC TGC AAT TTT ACA CTC TCT ATT AAT 1152
 Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr Leu Ser Ile Asn
 370 375 380

ACA ATA GAA TTC ACC TCC TAC GGC ACA TTC TCA TCA GAT GGA AAA CAA 1200
 Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser Asp Gly Lys Gln
 385 390 395 400

TTT ACA CCA CCT TCT GGT TCT GCC ATT GAT TTA CAC CTC CCT AAT TAT 1248
 Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His Leu Pro Asn Tyr
 405 410 415

GTA GAT CTC AAC GCG CTA TTA GAT ATT AGC CTC GAT TCA CTA CTT AAT 1296
 Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp Ser Leu Leu Asn
 420 425 430

TAT GAC GTT CAG GGG CAG TTT GGC GGA TCT AAT CCG GTT GAT AAT TTC 1344
 Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe
 435 440 445

AGT GGT CCC TAT GGT ATT TAT CTA TGG GAA ATC TTC TTC CAT ATT CCG 1392

Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro
 450 455 460
 TTC CTT GTT ACG GTC CGT ATG CAA ACC GAA CAA CGT TAC GAA GAC GCG 1440
 Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala
 465 470 475 480
 GAC ACT TGG TAC AAA TAT ATT TTC CGC AGC GCC GGT TAT CGC GAT GCT 1488
 Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala
 485 490 495
 AAT GGC CAG CTC ATT ATG GAT GGC AGT AAA CCA CGT TAT TGG AAT GTG 1536
 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val
 500 505 510
 ATG CCA TTG CAA CTG GAT ACC GCA TGG GAT ACC ACA CAG CCC GCC ACC 1584
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr
 515 520 525
 ACT GAT CCA GAT GTG ATC GCT ATG GCG GAC CCG ATG CAT TAC AAG CTG 1632
 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu
 530 535 540
 GCG ATA TTC CTG CAT ACC CTT GAT CTA TTG ATT GCC CGA GGC GAC AGC 1680
 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser
 545 550 555 560
 GCT TAC CGT CAA CTT GAA CGC GAT ACT CTA GTC GAA GCC AAA ATG TAC 1728
 Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu Ala Lys Met Tyr
 565 570 575
 TAC ATT CAG GCA CAA CAG CTA CTG GGA CCG CGC CCT GAT ATC CAT ACC 1776
 Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr
 580 585 590
 ACC AAT ACT TGG CCA AAT CCC ACC TTG AGT AAA GAA GCT GGC GCT ATT 1824
 Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile
 595 600 605
 GCC ACA CCG ACA TTC CTC AGT TCA CCG GAG GTG ATG ACG TTC GCT GCC 1872
 Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala
 610 615 620
 TGG CTA AGC 1881
 Trp Leu Ser
 625

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 627 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28 (TcaBi protein):

Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp
 1 5 10 15
 Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro Ala Asp Leu Lys

Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser Asp Gly Lys Gln
 385 390 395 400
 Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His Leu Pro Asn Tyr
 405 410 415
 Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp Ser Leu Leu Asn
 420 425 430
 Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe
 435 440 445
 Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro
 450 455 460
 Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala
 465 470 475 480
 Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala
 485 490 495
 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val
 500 505 510
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr
 515 520 525
 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu
 530 535 540
 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser
 545 550 555 560
 Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu Ala Lys Met Tyr
 565 570 575
 Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr
 580 585 590
 Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile
 595 600 605
 Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala
 610 615 620
 Trp Leu Ser
 625

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1689 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1689
- (D) OTHER INFORMATION: tcaBii

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29 (tcaBii coding region):

GCA GGC GAT ACC GCA AAT ATT GGC GAC GGT GAT TTC TTG CCA CCG TAC 48
 Ala Gly Asp Thr Ala Asn Ile Gly Asp Gly Asp Phe Leu Pro Pro Tyr
 1 5 10 15

AAC GAT GTA CTA CTC GGT TAC TGG GAT AAA CTT GAG TTA CGC CTA TAC 96
 Asn Asp Val Leu Leu Gly Tyr Trp Asp Lys Leu Glu Leu Arg Leu Tyr
 20 25 30

AAC CTG CGC CAC AAT CTG AGT CTG GAT GGT CAA CCG CTA AAT CTG CCA 144
 Asn Leu Arg His Asn Leu Ser Leu Asp Gly Gln Pro Leu Asn Leu Pro
 35 40 45

CTG TAT GCC ACG CCG GTA GAC CCG AAA ACC CTG CAA CGC CAG CAA GCC 192
 Leu Tyr Ala Thr Pro Val Asp Pro Lys Thr Leu Gln Arg Gln Gln Ala
 50 55 60

GGA GGG GAC GGT ACA GGC AGT AGT CCG GCT GGT GGT CAA GGC AGT GTT 240
 Gly Gly Asp Gly Thr Gly Ser Ser Pro Ala Gly Gly Gln Gly Ser Val
 65 70 75 80

CAG GGC TGG CGC TAT CCG TTA TTG GTA GAA CGC GCC CGC TCT GCC GTG 288
 Gln Gly Trp Arg Tyr Pro Leu Leu Val Glu Arg Ala Arg Ser Ala Val
 85 90 95

AGT TTG TTG ACT CAG TTC GGC AAC AGC TTA CAA ACA ACG TTA GAA CAT 336
 Ser Leu Leu Thr Gln Phe Gly Asn Ser Leu Gln Thr Thr Leu Glu His
 100 105 110

CAG GAT AAT GAA AAA ATG ACG ATA CTG TTG CAG ACT CAA CAG GAA GCC 384
 Gln Asp Asn Glu Lys Met Thr Ile Leu Leu Gln Thr Gln Gln Glu Ala
 115 120 125

ATC CTG AAA CAT CAG CAC GAT ATA CAA CAA AAT AAT CTA AAA GGA TTA 432
 Ile Leu Lys His Gln His Asp Ile Gln Gln Asn Asn Leu Lys Gly Leu
 130 135 140

CAA CAC AGC CTG ACC GCA TTA CAG GCT AGC CGT GAT GGC GAC ACA TTG 480
 Gln His Ser Leu Thr Ala Leu Gln Ala Ser Arg Asp Gly Asp Thr Leu
 145 150 155 160

CGG CAA AAA CAT TAC AGC GAC CTG ATT AAC GGT GGT CTA TCT GCG GCA 528
 Arg Gln Lys His Tyr Ser Asp Leu Ile Asn Gly Gly Leu Ser Ala Ala
 165 170 175

GAA ATC GCC GGT CTG ACA CTA CGC AGC ACC GCC ATG ATT ACC AAT GGC 576
 Glu Ile Ala Gly Leu Thr Leu Arg Ser Thr Ala Met Ile Thr Asn Gly
 180 185 190

GTT GCA ACG GGA TTG CTG ATT GCC GGC GGA ATC GCC AAC GCG GTA CCT 624
 Val Ala Thr Gly Leu Leu Ile Ala Gly Gly Ile Ala Asn Ala Val Pro
 195 200 205

AAC GTC TTC GGG CTG GCT AAC GGT GGA TCG GAA TGG GGA GCG CCA TTA 672
 Asn Val Phe Gly Leu Ala Asn Gly Gly Ser Glu Trp Gly Ala Pro Leu
 210 215 220

ATT GGC TCC GGG CAA GCA ACC CAA GTT GGC GCC GGC ATC CAG GAT CAG 720
 Ile Gly Ser Gly Gln Ala Thr Gln Val Gly Ala Gly Ile Gln Asp Gln
 225 230 235 240

AGC GCG GGC ATT TCA GAA GTG ACA GCA GGC TAT CAG CGT CGT CAG GAA 768
 Ser Ala Gly Ile Ser Glu Val Thr Ala Gly Tyr Gln Arg Arg Gln Glu
 245 250 255

GAA TGG GCA TTG CAA CGG GAT ATT GCT GAT AAC GAA ATA ACC CAA CTG 816
 Glu Trp Ala Leu Gln Arg Asp Ile Ala Asp Asn Glu Ile Thr Gln Leu
 260 265 270

GAT GCC CAG ATA CAA AGC CTG CAA GAG CAA ATC ACG ATG GCA CAA AAA 864
 Asp Ala Gln Ile Gln Ser Leu Gln Glu Gln Ile Thr Met Ala Gln Lys
 275 280 285

CAG ATC ACG CTC TCT GAA ACC GAA CAA GCG AAT GCC CAA GCG ATT TAT 912
 Gln Ile Thr Leu Ser Glu Thr Glu Gln Ala Asn Ala Gln Ala Ile Tyr
 290 295 300

GAC CTG CAA ACC ACT CGT TTT ACC GGG CAG GCA CTG TAT AAC TGG ATG 960
 Asp Leu Gln Thr Thr Arg Phe Thr Gly Gln Ala Leu Tyr Asn Trp Met
 305 310 315 320

GCC GGT CGT CTC TCC GCG CTC TAT TAC CAA ATG TAT GAT TCC ACT CTG 1008
 Ala Gly Arg Leu Ser Ala Leu Tyr Tyr Gln Met Tyr Asp Ser Thr Leu
 325 330 335

CCA ATC TGT CTC CAG CCA AAA GCC GCA TTA GTA CAG GAA TTA GGC GAG 1056
 Pro Ile Cys Leu Gln Pro Lys Ala Ala Leu Val Gln Glu Leu Gly Glu
 340 345 350

AAA GAG AGC GAC AGT CTT TTC CAG GTT CCG GTG TGG AAT GAT CTG TGG 1104
 Lys Glu Ser Asp Ser Leu Phe Gln Val Pro Val Trp Asn Asp Leu Trp
 355 360 365

CAA GGG CTG TTA GCA GGA GAA GGT TTA AGT TCA GAG CTA CAG AAA CTG 1152
 Gln Gly Leu Leu Ala Gly Glu Gly Leu Ser Ser Glu Leu Gln Lys Leu
 370 375 380

GAT GCC ATC TGG CTT GCA CGT GGT GGT ATT GGG CTA GAA GCC ATC CGC 1200
 Asp Ala Ile Trp Leu Ala Arg Gly Gly Ile Gly Leu Glu Ala Ile Arg
 385 390 395 400

ACC GTG TCG CTG GAT ACC CTG TTT GGC ACA GGG ACG TTA AGT GAA AAT 1248
 Thr Val Ser Leu Asp Thr Leu Phe Gly Thr Gly Thr Leu Ser Glu Asn
 405 410 415

ATC AAT AAA GTG CTT AAC GGG GAA ACG GTA TCT CCA TCC GGT GGC GTC 1296
 Ile Asn Lys Val Leu Asn Gly Glu Thr Val Ser Pro Ser Gly Gly Val
 420 425 430

ACT CTG GCG CTG ACA GGG GAT ATC TTC CAA GCA ACA CTG GAT TTG AGT 1344
 Thr Leu Ala Leu Thr Gly Asp Ile Phe Gln Ala Thr Leu Asp Leu Ser
 435 440 445

CAG CTA GGT TTG GAT AAC TCT TAC AAC TTG GGT AAC GAG AAG AAA CGT 1392
 Gln Leu Gly Leu Asp Asn Ser Tyr Asn Leu Gly Asn Glu Lys Lys Arg
 450 455 460

CGT ATT AAA CGT ATC GCC GTC ACC CTG CCA ACA CTT CTG GGG CCA TAT 1440
 Arg Ile Lys Arg Ile Ala Val Thr Leu Pro Thr Leu Leu Gly Pro Tyr
 465 470 475 480

CAA GAT CTT GAA GCC ACA CTG GTA ATG GGT GCG GAA ATC GCC GCC TTA 1488
 Gln Asp Leu Glu Ala Thr Leu Val Met Gly Ala Glu Ile Ala Ala Leu
 485 490 495
 TCA CAC GGT GTG AAT GAC GGA GGC CGG TTT GTT ACC GAC TTT AAC GAC 1536
 Ser His Gly Val Asn Asp Gly Gly Arg Phe Val Thr Asp Phe Asn Asp
 500 505 510
 AGC CGT TTT CTG CCT TTT GAA GGT CGA GAT GCA ACA ACC GGC ACA CTG 1584
 Ser Arg Phe Leu Pro Phe Glu Gly Arg Asp Ala Thr Thr Gly Thr Leu
 515 520 525
 GAG CTC AAT ATT TTC CAT GCG GGT AAA GAG GGA ACG CAA CAC GAG TTG 1632
 Glu Leu Asn Ile Phe His Ala Gly Lys Glu Gly Thr Gln His Glu Leu
 530 535 540
 GTC GCG AAT CTG AGT GAC ATC ATT GTG CAT CTG AAT TAC ATC ATT CGA 1680
 Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr Ile Ile Arg
 545 550 555 560
 GAC GCG TAA 1689
 Asp Ala *

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 562 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30 (TcaBii protein):

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

Glu Leu Asn Ile Phe His Ala Gly Lys Glu Gly Thr Gln His Glu Leu
 530 535 540
 Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr Ile Ile Arg
 545 550 555 560
 Asp Ala *

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4458 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..4458

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31 (tcac gene):

ATG CAG GAT TCA CCA GAA GTA TCG ATT ACA ACG CTG TCA CTT CCC AAA 48
 Met Gln Asp Ser Pro Glu Val Ser Ile Thr Thr Leu Ser Leu Pro Lys
 1 5 10 15
 GGT GGC GGT GCT ATC AAT GGC ATG GGA GAA GCA CTG AAT GCT GCC GGC 96
 Gly Gly Gly Ala Ile Asn Gly Met Gly Glu Ala Leu Asn Ala Ala Gly
 20 25 30
 CCT GAT GGA ATG GCC TCC CTA TCT CTG CCA TTA CCC CTT TCG ACC GGC 144
 Pro Asp Gly Met Ala Ser Leu Ser Leu Pro Leu Pro Leu Ser Thr Gly
 35 40 45
 AGA GGG ACG GCT CCT GGA TTA TCG CTG ATT TAC AGC AAC AGT GCA GGT 192
 Arg Gly Thr Ala Pro Gly Leu Ser Leu Ile Tyr Ser Asn Ser Ala Gly
 50 55 60
 AAT GGG CCT TTC GGC ATC GGC TGG CAA TGC GGT GTT ATG TCC ATT AGC 240
 Asn Gly Pro Phe Gly Ile Gly Trp Gln Cys Gly Val Met Ser Ile Ser
 65 70 75 80
 CGA CGC ACC CAA CAT GGC ATT CCA CAA TAC GGT AAT GAC GAC ACG TTC 288
 Arg Arg Thr Gln His Gly Ile Pro Gln Tyr Gly Asn Asp Asp Thr Phe
 85 90 95
 CTA TCC CCA CAA GGC GAG GTC ATG AAT ATC GCC CTG AAT GAC CAA GGC 336
 Leu Ser Pro Gln Gly Glu Val Met Asn Ile Ala Leu Asn Asp Gln Gly
 100 105 110
 CAA CCT GAT ATC CGT CAA GAC GTT AAA ACG CTG CAA GGC GTT ACC TTG 384
 Gln Pro Asp Ile Arg Gln Asp Val Lys Thr Leu Gln Gly Val Thr Leu
 115 120 125
 CCA ATT TCC TAT ACC GTG ACC CGC TAT CAA GCC CGC CAG ATC CTG GAT 432
 Pro Ile Ser Tyr Thr Val Thr Arg Tyr Gln Ala Arg Gln Ile Leu Asp
 130 135 140
 TTC AGT AAA ATC GAA TAC TGG CAA CCT GCC TCC GGT CAA GAA GGA CGC 480
 Phe Ser Lys Ile Glu Tyr Trp Gln Pro Ala Ser Gly Gln Glu Gly Arg

145 150 155 160
 GCT TTC TGG CTG ATA TCG ACA CCG GAC GGG CAT CTA CAC ATC TTA GGG 528
 Ala Phe Trp Leu Ile Ser Thr Pro Asp Gly His Leu His Ile Leu Gly
 165 170 175
 AAA ACC GCG CAG GCT TGT CTG GCA AAT CCG CAA AAT GAC CAA CAA ATC 576
 Lys Thr Ala Gln Ala Cys Leu Ala Asn Pro Gln Asn Asp Gln Gln Ile
 180 185 190
 GCC CAG TGG TTG CTG GAA GAA ACT GTG ACG CCA GCC GGT GAA CAT GTC 624
 Ala Gln Trp Leu Leu Glu Glu Thr Val Thr Pro Ala Gly Glu His Val
 195 200 205
 AGC TAT CAA TAT CGA GCC GAA GAT GAA GCC CAT TGT GAC GAC AAT GAA 672
 Ser Tyr Gln Tyr Arg Ala Glu Asp Glu Ala His Cys Asp Asp Asn Glu
 210 215 220
 AAA ACC GCT CAT CCC AAT GTT ACC GCA CAG CGC TAT CTG GTA CAG GTG 720
 Lys Thr Ala His Pro Asn Val Thr Ala Gln Arg Tyr Leu Val Gln Val
 225 230 235 240
 AAC TAC GGC AAC ATC AAA CCA CAA GCC AGC CTG TTC GTA CTG GAT AAC 768
 Asn Tyr Gly Asn Ile Lys Pro Gln Ala Ser Leu Phe Val Leu Asp Asn
 245 250 255
 GCA CCT CCC GCA CCG GAA GAG TGG CTG TTT CAT CTG GTC TTT GAC CAC 816
 Ala Pro Pro Ala Pro Glu Glu Trp Leu Phe His Leu Val Phe Asp His
 260 265 270
 GGT GAG CGC GAT ACC TCA CTT CAT ACC GTG CCA ACA TGG GAT GCA GGT 864
 Gly Glu Arg Asp Thr Ser Leu His Thr Val Pro Thr Trp Asp Ala Gly
 275 280 285
 ACA GCG CAA TGG TCT GTA CGC CCG GAT ATC TTC TCT CGC TAT GAA TAT 912
 Thr Ala Gln Trp Ser Val Arg Pro Asp Ile Phe Ser Arg Tyr Glu Tyr
 290 295 300
 GGT TTT GAA GTG CGT ACT CGC CGC TTA TGT CAA CAA GTG CTG ATG TTT 960
 Gly Phe Glu Val Arg Thr Arg Arg Leu Cys Gln Gln Val Leu Met Phe
 305 310 315 320
 CAC CGC ACC GCG CTC ATG GCC GGA GAA GCC AGT ACC AAT GAC GCC CCG 1008
 His Arg Thr Ala Leu Met Ala Gly Glu Ala Ser Thr Asn Asp Ala Pro
 325 330 335
 GAA CTG GTT GGA CGC TTA ATA CTG GAA TAT GAC AAA AAC GCC AGC GTC 1056
 Glu Leu Val Gly Arg Leu Ile Leu Glu Tyr Asp Lys Asn Ala Ser Val
 340 345 350
 ACC ACG TTG ATT ACC ATC CGT CAA TTA AGC CAT GAA TCG GAC GGG AGG 1104
 Thr Thr Leu Ile Thr Ile Arg Gln Leu Ser His Glu Ser Asp Gly Arg
 355 360 365
 CCA GTC ACC CAG CCA CCA CTA GAA CTA GCC TGG CAA CGG TTT GAT CTG 1152
 Pro Val Thr Gln Pro Pro Leu Glu Leu Ala Trp Gln Arg Phe Asp Leu
 370 375 380
 GAG AAA ATC CCG ACA TGG CAA CGC TTT GAC GCA CTA GAT AAT TTT AAC 1200
 Glu Lys Ile Pro Thr Trp Gln Arg Phe Asp Ala Leu Asp Asn Phe Asn

385 390 395 400
 TCG CAG CAA CGT TAT CAA CTG GTT GAT CTG CGG GGA GAA GGG TTG CCA 1248
 Ser Gln Gln Arg Tyr Gln Leu Val Asp Leu Arg Gly Glu Gly Leu Pro
 405 410 415
 GGT ATG CTG TAT CAA GAT CGA GGC GCT TGG TGG TAT AAA GCT CCG CAA 1296
 Gly Met Leu Tyr Gln Asp Arg Gly Ala Trp Trp Tyr Lys Ala Pro Gln
 420 425 430
 CGT CAG GAA GAC GGA GAC AGC AAT GCC GTC ACT TAC GAC AAA ATC GCC 1344
 Arg Gln Glu Asp Gly Asp Ser Asn Ala Val Thr Tyr Asp Lys Ile Ala
 435 440 445
 CCA CTG CCT ACC CTA CCC AAT TTG CAG GAT AAT GCC TCA TTG ATG GAT 1392
 Pro Leu Pro Thr Leu Pro Asn Leu Gln Asp Asn Ala Ser Leu Met Asp
 450 455 460
 ATC AAC GGA GAC GGC CAA CTG GAT TGG GTT GTT ACC GCC TCC GGT ATT 1440
 Ile Asn Gly Asp Gly Gln Leu Asp Trp Val Val Thr Ala Ser Gly Ile
 465 470 475 480
 CGC GGA TAC CAT AGT CAG CAA CCC GAT GGA AAG TGG ACG CAC TTT ACG 1488
 Arg Gly Tyr His Ser Gln Gln Pro Asp Gly Lys Trp Thr His Phe Thr
 485 490 495
 CCA ATC AAT GCC TTG CCC GTG GAA TAT TTT CAT CCA AGC ATC CAG TTC 1536
 Pro Ile Asn Ala Leu Pro Val Glu Tyr Phe His Pro Ser Ile Gln Phe
 500 505 510
 GCT GAC CTT ACC GGG GCA GGC TTA TCT GAT TTA GTG TTG ATC GGG CCG 1584
 Ala Asp Leu Thr Gly Ala Gly Leu Ser Asp Leu Val Leu Ile Gly Pro
 515 520 525
 AAA AGC GTG CGT CTA TAT GCC AAC CAG CGA AAC GGC TGG CGT AAA GGA 1632
 Lys Ser Val Arg Leu Tyr Ala Asn Gln Arg Asn Gly Trp Arg Lys Gly
 530 535 540
 GAA GAT GTC CCC CAA TCC ACA GGT ATC ACC CTG CCT GTC ACA GGG ACC 1680
 Glu Asp Val Pro Gln Ser Thr Gly Ile Thr Leu Pro Val Thr Gly Thr
 545 550 555 560
 GAT GCC CGC AAA CTG GTG GCT TTC AGT GAT ATG CTC GGT TCC GGT CAA 1728
 Asp Ala Arg Lys Leu Val Ala Phe Ser Asp Met Leu Gly Ser Gly Gln
 565 570 575
 CAA CAT CTG GTG GAA ATC AAG GGT AAT CGC GTC ACC TGT TGG CCG AAT 1776
 Gln His Leu Val Glu Ile Lys Gly Asn Arg Val Thr Cys Trp Pro Asn
 580 585 590
 CTA GGG CAT GGC CGT TTC GGT CAA CCA CTA ACT CTG TCA GGA TTT AGC 1824
 Leu Gly His Gly Arg Phe Gly Gln Pro Leu Thr Leu Ser Gly Phe Ser
 595 600 605
 CAG CCC GAA AAT AGC TTC AAT CCC GAA CGG CTG TTT CTG GCG GAT ATC 1872
 Gln Pro Glu Asn Ser Phe Asn Pro Glu Arg Leu Phe Leu Ala Asp Ile
 610 615 620
 GAC GGC TCC GGC ACC ACC GAC CTT ATC TAT GCG CAA TCC GGC TCT TTG 1920
 Asp Gly Ser Gly Thr Thr Asp Leu Ile Tyr Ala Gln Ser Gly Ser Leu

625 630 635 640
 CTC ATT TAT CTC AAC CAA AGT GGT AAT CAG TTT GAT GCC CCG TTG ACA 1968
 Leu Ile Tyr Leu Asn Gln Ser Gly Asn Gln Phe Asp Ala Pro Leu Thr
 645 650 655
 TTA GCG TTG CCA GAA GGC GTA CAA TTT GAC AAC ACT TGC CAA CTT CAA 2016
 Leu Ala Leu Pro Glu Gly Val Gln Phe Asp Asn Thr Cys Gln Leu Gln
 660 665 670
 GTC GCC GAT ATT CAG GGA TTA GGG ATA GCC AGC TTG ATT CTG ACT GTG 2064
 Val Ala Asp Ile Gln Gly Leu Gly Ile Ala Ser Leu Ile Leu Thr Val
 675 680 685
 CCA CAT ATC GCG CCA CAT CAC TGG CGT TGT GAC CTG TCA CTG ACC AAA 2112
 Pro His Ile Ala Pro His His Trp Arg Cys Asp Leu Ser Leu Thr Lys
 690 695 700
 CCC TGG TTG TTG AAT GTA ATG AAC AAT AAC CGG GGC GCA CAT CAC ACG 2160
 Pro Trp Leu Leu Asn Val Met Asn Asn Asn Arg Gly Ala His His Thr
 705 710 715 720
 CTA CAT TAT CGT AGT TCC GCG CAA TTC TGG TTG GAT GAA AAA TTA CAG 2208
 Leu His Tyr Arg Ser Ser Ala Gln Phe Trp Leu Asp Glu Lys Leu Gln
 725 730 735
 CTC ACC AAA GCA GGC AAA TCT CCG GCT TGT TAT CTG CCG TTT CCA ATG 2256
 Leu Thr Lys Ala Gly Lys Ser Pro Ala Cys Tyr Leu Pro Phe Pro Met
 740 745 750
 CAT TTG CTA TGG TAT ACC GAA ATT CAG GAT GAA ATC AGC GGC AAC CGG 2304
 His Leu Leu Trp Tyr Thr Glu Ile Gln Asp Glu Ile Ser Gly Asn Arg
 755 760 765
 CTC ACC AGT GAA GTC AAC TAC AGC CAC GGC GTC TGG GAT GGT AAA GAG 2352
 Leu Thr Ser Glu Val Asn Tyr Ser His Gly Val Trp Asp Gly Lys Glu
 770 775 780
 CGG GAA TTC AGA GGA TTT GGC TGC ATC AAA CAG ACA GAT ACC ACA ACG 2400
 Arg Glu Phe Arg Gly Phe Gly Cys Ile Lys Gln Thr Asp Thr Thr Thr
 785 790 795 800
 TTT TCT CAC GGC ACC GCC CCC GAA CAG GCG GCA CCG TCG CTG AGT ATT 2448
 Phe Ser His Gly Thr Ala Pro Glu Gln Ala Ala Pro Ser Leu Ser Ile
 805 810 815
 AGC TGG TTT GCC ACC GGC ATG GAT GAA GTA GAC AGC CAA TTA GCT ACG 2496
 Ser Trp Phe Ala Thr Gly Met Asp Glu Val Asp Ser Gln Leu Ala Thr
 820 825 830
 GAA TAT TGG CAG GCA GAC ACG CAA GCT TAT AGC GGA TTT GAA ACC CGT 2544
 Glu Tyr Trp Gln Ala Asp Thr Gln Ala Tyr Ser Gly Phe Glu Thr Arg
 835 840 845
 TAT ACC GTC TGG GAT CAC ACC AAC CAG ACA GAC CAA GCA TTT ACC CCC 2592
 Tyr Thr Val Trp Asp His Thr Asn Gln Thr Asp Gln Ala Phe Thr Pro
 850 855 860
 AAT GAG ACA CAA CGT AAC TGG CTG ACG CGA GCG CTT AAA GGC CAA CTG 2640
 Asn Glu Thr Gln Arg Asn Trp Leu Thr Arg Ala Leu Lys Gly Gln Leu

865 870 875 880
 CTA CGC ACT GAG CTC TAC GGT CTG GAC GGA ACA GAT AAG CAA ACA GTG 2688
 Leu Arg Thr Glu Leu Tyr Gly Leu Asp Gly Thr Asp Lys Gln Thr Val
 885 890 895
 CCT TAT ACC GTC AGT GAA TCG CGC TAT CAG GTA CGC TCT ATT CCC GTA 2736
 Pro Tyr Thr Val Ser Glu Ser Arg Tyr Gln Val Arg Ser Ile Pro Val
 900 905 910
 AAT AAA GAA ACT GAA TTA TCT GCC TGG GTG ACT GCT ATT GAA AAT CGC 2784
 Asn Lys Glu Thr Glu Leu Ser Ala Trp Val Thr Ala Ile Glu Asn Arg
 915 920 925
 AGC TAC CAC TAT GAA CGT ATC ATC ACT GAC CCA CAG TTC AGC CAG AGT 2832
 Ser Tyr His Tyr Glu Arg Ile Ile Thr Asp Pro Gln Phe Ser Gln Ser
 930 935 940
 ATC AAG TTG CAA CAC GAT ATC TTT GGT CAA TCA CTG CAA AGT GTC GAT 2880
 Ile Lys Leu Gln His Asp Ile Phe Gly Gln Ser Leu Gln Ser Val Asp
 945 950 955 960
 ATT GCC TGG CCG CGC CGC GAA AAA CCA GCA GTG AAT CCC TAC CCG CCT 2928
 Ile Ala Trp Pro Arg Arg Glu Lys Pro Ala Val Asn Pro Tyr Pro Pro
 965 970 975
 ACC CTG CCG GAA ACG CTA TTT GAC AGC AGC TAT GAT GAT CAA CAA CAA 2976
 Thr Leu Pro Glu Thr Leu Phe Asp Ser Ser Tyr Asp Asp Gln Gln Gln
 980 985 990
 CTA TTA CGT CTG GTG AGA CAA AAA AAT AGC TGG CAT CAC CTG ACT GAT 3024
 Leu Leu Arg Leu Val Arg Gln Lys Asn Ser Trp His His Leu Thr Asp
 995 1000 1005
 GGG GAA AAC TGG CGA TTA GGT TTA CCG AAT GCA CAA CGC CGT GAT GTT 3072
 Gly Glu Asn Trp Arg Leu Gly Leu Pro Asn Ala Gln Arg Arg Asp Val
 1010 1015 1020
 TAT ACT TAT GAC CGG AGC AAA ATT CCA ACC GAA GGG ATT TCC CTT GAA 3120
 Tyr Thr Tyr Asp Arg Ser Lys Ile Pro Thr Glu Gly Ile Ser Leu Glu
 1025 1030 1035 1040
 ATC TTG CTG AAA GAT GAT GGC CTG CTA GCA GAT GAA AAA GCG GCC GTT 3168
 Ile Leu Leu Lys Asp Asp Gly Leu Leu Ala Asp Glu Lys Ala Ala Val
 1045 1050 1055
 TAT CTG GGA CAA CAA CAG ACG TTT TAC ACC GCC GGT CAA GCG GAA GTC 3216
 Tyr Leu Gly Gln Gln Gln Thr Phe Tyr Thr Ala Gly Gln Ala Glu Val
 1060 1065 1070
 ACT CTA GAA AAA CCC ACG TTA CAA GCA CTG GTC GCG TTC CAA GAA ACC 3264
 Thr Leu Glu Lys Pro Thr Leu Gln Ala Leu Val Ala Phe Gln Glu Thr
 1075 1080 1085
 GCC ATG ATG GAC GAT ACC TCA TTA CAG GCG TAT GAA GGC GTG ATT GAA 3312
 Ala Met Met Asp Asp Thr Ser Leu Gln Ala Tyr Glu Gly Val Ile Glu
 1090 1095 1100
 GAG CAA GAG TTG AAT ACC GCG CTG ACA CAG GCC GGT TAT CAG CAA GTC 3360
 Glu Gln Glu Leu Asn Thr Ala Leu Thr Gln Ala Gly Tyr Gln Gln Val

1105 1110 1115 1120
 GCG CGG TTG TTT AAT ACC AGA TCA GAA AGC CCG GTA TGG GCG GCA CGG 3408
 Ala Arg Leu Phe Asn Thr Arg Ser Glu Ser Pro Val Trp Ala Ala Arg
 1125 1130 1135
 CAA GGT TAT ACC GAT TAC GGT GAC GCC GCA CAG TTC TGG CGG CCT CAG 3456
 Gln Gly Tyr Thr Asp Tyr Gly Asp Ala Ala Gln Phe Trp Arg Pro Gln
 1140 1145 1150
 GCT CAG CGT AAC TCG TTG CTG ACA GGG AAA ACC ACA CTG ACC TGG GAT 3504
 Ala Gln Arg Asn Ser Leu Leu Thr Gly Lys Thr Thr Leu Thr Trp Asp
 1155 1160 1165
 ACC CAT CAT TGT GTA ATA ATA CAG ACT CAA GAT GCC GCT GGA TTA ACG 3552
 Thr His His Cys Val Ile Ile Gln Thr Gln Asp Ala Ala Gly Leu Thr
 1170 1175 1180
 ACG CAA GCC CAT TAC GAT TAT CGT TTC CTT ACA CCG GTA CAA CTG ACA 3600
 Thr Gln Ala His Tyr Asp Tyr Arg Phe Leu Thr Pro Val Gln Leu Thr
 1185 1190 1195 1200
 GAT ATT AAT GAT AAT CAA CAT ATT GTG ACT CTG GAC GCG CTA GGT CGC 3648
 Asp Ile Asn Asp Asn Gln His Ile Val Thr Leu Asp Ala Leu Gly Arg
 1205 1210 1215
 GTA ACC ACC AGC CGG TTC TGG GGC ACA GAG GCA GGA CAA GCC GCA GGC 3696
 Val Thr Thr Ser Arg Phe Trp Gly Thr Glu Ala Gly Gln Ala Ala Gly
 1220 1225 1230
 TAT TCC AAC CAG CCC TTC ACA CCA CCG GAC TCC GTA GAT AAA GCG CTG 3744
 Tyr Ser Asn Gln Pro Phe Thr Pro Pro Asp Ser Val Asp Lys Ala Leu
 1235 1240 1245
 GCA TTA ACC GGC GCA CTC CCT GTT GCC CAA TGT TTA GTC TAT GCC GTT 3792
 Ala Leu Thr Gly Ala Leu Pro Val Ala Gln Cys Leu Val Tyr Ala Val
 1250 1255 1260
 GAT AGC TGG ATG CCG TCG TTA TCT TTG TCT CAG CTT TCT CAG TCA CAA 3840
 Asp Ser Trp Met Pro Ser Leu Ser Leu Ser Gln Leu Ser Gln Ser Gln
 1265 1270 1275 1280
 GAA GAG GCA GAA GCG CTA TGG GCG CAA CTG CGT GCC GCT CAT ATG ATT 3888
 Glu Glu Ala Glu Ala Leu Trp Ala Gln Leu Arg Ala Ala His Met Ile
 1285 1290 1295
 ACC GAA GAT GGG AAA GTG TGT GCG TTA AGC GGG AAA CGA GGA ACA AGC 3936
 Thr Glu Asp Gly Lys Val Cys Ala Leu Ser Gly Lys Arg Gly Thr Ser
 1300 1305 1310
 CAT CAG AAC CTG ACG ATT CAA CTT ATT TCG CTA TTG GCA AGT ATT CCC 3984
 His Gln Asn Leu Thr Ile Gln Leu Ile Ser Leu Leu Ala Ser Ile Pro
 1315 1320 1325
 CGT TTA CCG CCA CAT GTA CTG GGG ATC ACC ACT GAT CGC TAT GAT AGC 4032
 Arg Leu Pro Pro His Val Leu Gly Ile Thr Thr Asp Arg Tyr Asp Ser
 1330 1335 1340
 GAT CCG CAA CAG CAG CAC CAA CAG ACG GTG AGC TTT AGT GAC GGT TTT 4080
 Asp Pro Gln Gln Gln His Gln Gln Thr Val Ser Phe Ser Asp Gly Phe

1345 1350 1355 1360
GGC CGG TTA CTC CAG AGT TCA GCT CGT CAT GAG TCA GGT GAT GCC TGG 4128
Gly Arg Leu Leu Gln Ser Ser Ala Arg His Glu Ser Gly Asp Ala Trp
 1365 1370 1375
CAA CGT AAA GAG GAT GGC GGG CTG GTC GTG GAT GCA AAT GGC GTT CTG 4176
Gln Arg Lys Glu Asp Gly Gly Leu Val Val Asp Ala Asn Gly Val Leu
 1380 1385 1390
GTC AGT GCC CCT ACA GAC ACC CGA TGG GCC GTT TCC GGT CGC ACA GAA 4224
Val Ser Ala Pro Thr Asp Thr Arg Trp Ala Val Ser Gly Arg Thr Glu
 1395 1400 1405
TAT GAC GAC AAA GGC CAA CCT GTG CGT ACT TAT CAA CCC TAT TTT CTA 4272
Tyr Asp Asp Lys Gly Gln Pro Val Arg Thr Tyr Gln Pro Tyr Phe Leu
 1410 1415 1420
AAT GAC TGG CGT TAC GTT AGT GAT GAC AGC GCA CGA GAT GAC CTG TTT 4320
Asn Asp Trp Arg Tyr Val Ser Asp Asp Ser Ala Arg Asp Asp Leu Phe
1425 1430 1435 1440
GCC GAT ACC CAC CTT TAT GAT CCA TTG GGA CGG GAA TAC AAA GTC ATC 4368
Ala Asp Thr His Leu Tyr Asp Pro Leu Gly Arg Glu Tyr Lys Val Ile
 1445 1450 1455
ACT GCT AAG AAA TAT TTG CGA GAA AAG CTG TAC ACC CCG TGG TTT ATT 4416
Thr Ala Lys Lys Tyr Leu Arg Glu Lys Leu Tyr Thr Pro Trp Phe Ile
 1460 1465 1470
GTC AGT GAG GAT GAA AAC GAT ACA GCA TCA AGA ACC CCA TAG 4458
Val Ser Glu Asp Glu Asn Asp Thr Ala Ser Arg Thr Pro *

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1485 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32 (TcaC protein):

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

Ile Asn Gly Asp Gly Gln Leu Asp Trp Val Val Thr Ala Ser Gly Ile
 465 470 475 480
 Arg Gly Tyr His Ser Gln Gln Pro Asp Gly Lys Trp Thr His Phe Thr
 485 490 495
 Pro Ile Asn Ala Leu Pro Val Glu Tyr Phe His Pro Ser Ile Gln Phe
 500 505 510
 Ala Asp Leu Thr Gly Ala Gly Leu Ser Asp Leu Val Leu Ile Gly Pro
 515 520 525
 Lys Ser Val Arg Leu Tyr Ala Asn Gln Arg Asn Gly Trp Arg Lys Gly
 530 535 540
 Glu Asp Val Pro Gln Ser Thr Gly Ile Thr Leu Pro Val Thr Gly Thr
 545 550 555 560
 Asp Ala Arg Lys Leu Val Ala Phe Ser Asp Met Leu Gly Ser Gly Gln
 565 570 575
 Gln His Leu Val Glu Ile Lys Gly Asn Arg Val Thr Cys Trp Pro Asn
 580 585 590
 Leu Gly His Gly Arg Phe Gly Gln Pro Leu Thr Leu Ser Gly Phe Ser
 595 600 605
 Gln Pro Glu Asn Ser Phe Asn Pro Glu Arg Leu Phe Leu Ala Asp Ile
 610 615 620
 Asp Gly Ser Gly Thr Thr Asp Leu Ile Tyr Ala Gln Ser Gly Ser Leu
 625 630 635 640
 Leu Ile Tyr Leu Asn Gln Ser Gly Asn Gln Phe Asp Ala Pro Leu Thr
 645 650 655
 Leu Ala Leu Pro Glu Gly Val Gln Phe Asp Asn Thr Cys Gln Leu Gln
 660 665 670
 Val Ala Asp Ile Gln Gly Leu Gly Ile Ala Ser Leu Ile Leu Thr Val
 675 680 685
 Pro His Ile Ala Pro His His Trp Arg Cys Asp Leu Ser Leu Thr Lys
 690 695 700
 Pro Trp Leu Leu Asn Val Met Asn Asn Asn Arg Gly Ala His His Thr
 705 710 715 720
 Leu His Tyr Arg Ser Ser Ala Gln Phe Trp Leu Asp Glu Lys Leu Gln
 725 730 735
 Leu Thr Lys Ala Gly Lys Ser Pro Ala Cys Tyr Leu Pro Phe Pro Met
 740 745 750
 His Leu Leu Trp Tyr Thr Glu Ile Gln Asp Glu Ile Ser Gly Asn Arg
 755 760 765
 Leu Thr Ser Glu Val Asn Tyr Ser His Gly Val Trp Asp Gly Lys Glu
 770 775 780
 Arg Glu Phe Arg Gly Phe Gly Cys Ile Lys Gln Thr Asp Thr Thr Thr
 785 790 795 800
 Phe Ser His Gly Thr Ala Pro Glu Gln Ala Ala Pro Ser Leu Ser Ile
 805 810 815
 Ser Trp Phe Ala Thr Gly Met Asp Glu Val Asp Ser Gln Leu Ala Thr

| | | | | | |
|-------------|-----------------|-----------------|-------------|-------------|------|
| | 820 | | 825 | | 830 |
| Glu Tyr Trp | Gln Ala Asp Thr | Gln Ala Tyr Ser | Gly Phe | Glu Thr Arg | |
| | 835 | | 840 | | 845 |
| Tyr Thr Val | Trp Asp His Thr | Asn Gln Thr Asp | Gln Ala Phe | Thr Pro | |
| | 850 | | 855 | | 860 |
| Asn Glu Thr | Gln Arg Asn Trp | Leu Thr Arg Ala | Leu Lys Gly | Gln Leu | |
| 865 | | 870 | | 875 | 880 |
| Leu Arg Thr | Glu Leu Tyr Gly | Leu Asp Gly Thr | Asp Lys Gln | Thr Val | |
| | 885 | | 890 | | 895 |
| Pro Tyr Thr | Val Ser Glu Ser | Arg Tyr Gln Val | Arg Ser Ile | Pro Val | |
| | 900 | | 905 | | 910 |
| Asn Lys Glu | Thr Glu Leu Ser | Ala Trp Val Thr | Ala Ile Glu | Asn Arg | |
| | 915 | | 920 | | 925 |
| Ser Tyr His | Tyr Glu Arg Ile | Ile Thr Asp Pro | Gln Phe Ser | Gln Ser | |
| 930 | | 935 | | 940 | |
| Ile Lys Leu | Gln His Asp Ile | Phe Gly Gln Ser | Leu Gln Ser | Val Asp | |
| 945 | | 950 | | 955 | 960 |
| Ile Ala Trp | Pro Arg Arg Glu | Lys Pro Ala Val | Asn Pro Tyr | Pro Pro | |
| | 965 | | 970 | | 975 |
| Thr Leu Pro | Glu Thr Leu Phe | Asp Ser Ser Tyr | Asp Asp Gln | Gln Gln | |
| | 980 | | 985 | | 990 |
| Leu Leu Arg | Leu Val Arg Gln | Lys Asn Ser Trp | His His Leu | Thr Asp | |
| | 995 | | 1000 | | 1005 |
| Gly Glu Asn | Trp Arg Leu Gly | Leu Pro Asn Ala | Gln Arg Arg | Asp Val | |
| | 1010 | | 1015 | | 1020 |
| Tyr Thr Tyr | Asp Arg Ser Lys | Ile Pro Thr Glu | Gly Ile Ser | Leu Glu | |
| 1025 | | 1030 | | 1035 | 1040 |
| Ile Leu Leu | Lys Asp Asp Gly | Leu Leu Ala Asp | Glu Lys Ala | Ala Val | |
| | 1045 | | 1050 | | 1055 |
| Tyr Leu Gly | Gln Gln Gln Thr | Phe Tyr Thr Ala | Gly Gln Ala | Glu Val | |
| | 1060 | | 1065 | | 1070 |
| Thr Leu Glu | Lys Pro Thr Leu | Gln Ala Leu Val | Ala Phe Gln | Glu Thr | |
| | 1075 | | 1080 | | 1085 |
| Ala Met Met | Asp Asp Thr Ser | Leu Gln Ala Tyr | Glu Gly Val | Ile Glu | |
| | 1090 | | 1095 | | 1100 |
| Glu Gln Glu | Leu Asn Thr Ala | Leu Thr Gln Ala | Gly Tyr Gln | Gln Val | |
| 1105 | | 1110 | | 1115 | 1120 |
| Ala Arg Leu | Phe Asn Thr Arg | Ser Glu Ser Pro | Val Trp Ala | Ala Arg | |
| | 1125 | | 1130 | | 1135 |
| Gln Gly Tyr | Thr Asp Tyr Gly | Asp Ala Ala Gln | Phe Trp Arg | Pro Gln | |
| | 1140 | | 1145 | | 1150 |
| Ala Gln Arg | Asn Ser Leu Leu | Thr Gly Lys Thr | Thr Leu Thr | Trp Asp | |
| | 1155 | | 1160 | | 1165 |
| Thr His His | Cys Val Ile Ile | Gln Thr Gln Asp | Ala Ala Gly | Leu Thr | |
| | 1170 | | 1175 | | 1180 |

Thr Gln Ala His Tyr Asp Tyr Arg Phe Leu Thr Pro Val Gln Leu Thr
 1185 1190 1195 1200
 Asp Ile Asn Asp Asn Gln His Ile Val Thr Leu Asp Ala Leu Gly Arg
 1205 1210 1215
 Val Thr Thr Ser Arg Phe Trp Gly Thr Glu Ala Gly Gln Ala Ala Gly
 1220 1225 1230
 Tyr Ser Asn Gln Pro Phe Thr Pro Pro Asp Ser Val Asp Lys Ala Leu
 1235 1240 1245
 Ala Leu Thr Gly Ala Leu Pro Val Ala Gln Cys Leu Val Tyr Ala Val
 1250 1255 1260
 Asp Ser Trp Met Pro Ser Leu Ser Leu Ser Gln Leu Ser Gln Ser Gln
 1265 1270 1275 1280
 Glu Glu Ala Glu Ala Leu Trp Ala Gln Leu Arg Ala Ala His Met Ile
 1285 1290 1295
 Thr Glu Asp Gly Lys Val Cys Ala Leu Ser Gly Lys Arg Gly Thr Ser
 1300 1305 1310
 His Gln Asn Leu Thr Ile Gln Leu Ile Ser Leu Leu Ala Ser Ile Pro
 1315 1320 1325
 Arg Leu Pro Pro His Val Leu Gly Ile Thr Thr Asp Arg Tyr Asp Ser
 1330 1335 1340
 Asp Pro Gln Gln Gln His Gln Gln Thr Val Ser Phe Ser Asp Gly Phe
 1345 1350 1355 1360
 Gly Arg Leu Leu Gln Ser Ser Ala Arg His Glu Ser Gly Asp Ala Trp
 1365 1370 1375
 Gln Arg Lys Glu Asp Gly Gly Leu Val Val Asp Ala Asn Gly Val Leu
 1380 1385 1390
 Val Ser Ala Pro Thr Asp Thr Arg Trp Ala Val Ser Gly Arg Thr Glu
 1395 1400 1405
 Tyr Asp Asp Lys Gly Gln Pro Val Arg Thr Tyr Gln Pro Tyr Phe Leu
 1410 1415 1420
 Asn Asp Trp Arg Tyr Val Ser Asp Asp Ser Ala Arg Asp Asp Leu Phe
 1425 1430 1435 1440
 Ala Asp Thr His Leu Tyr Asp Pro Leu Gly Arg Glu Tyr Lys Val Ile
 1445 1450 1455
 Thr Ala Lys Lys Tyr Leu Arg Glu Lys Leu Tyr Thr Pro Trp Phe Ile
 1460 1465 1470
 Val Ser Glu Asp Glu Asn Asp Thr Ala Ser Arg Thr Pro *
 1475 1480 1485

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3288 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33 (tcaA gene):

ATG GTG ACT GTT ATG CAA AAT AAA ATA TCA TTT TTA TCA GGT ACA TCC 48
Met Val Thr Val Met Gln Asn Lys Ile Ser Phe Leu Ser Gly Thr Ser
1 5 10 15

GAA CAG CCC CTG CTT GAC GCC GGT TAT CAA AAC GTA TTT GAT ATC GCA 96
Glu Gln Pro Leu Leu Asp Ala Gly Tyr Gln Asn Val Phe Asp Ile Ala
20 25 30

TCA ATC AGC CGG GCT ACT TTC GTT CAA TCC GTT CCC ACC CTG CCC GTT 144
Ser Ile Ser Arg Ala Thr Phe Val Gln Ser Val Pro Thr Leu Pro Val
35 40 45

AAA GAG GCT CAT ACC GTC TAT CGT CAG GCG CGG CAA CGT GCG GAA AAT 192
Lys Glu Ala His Thr Val Tyr Arg Gln Ala Arg Gln Arg Ala Glu Asn
50 55 60

CTG AAA TCC CTC TAC CGA GCC TGG CAA TTG CGT CAG GAG CCG GTT ATT 240
Leu Lys Ser Leu Tyr Arg Ala Trp Gln Leu Arg Gln Glu Pro Val Ile
65 70 75 80

AAA GGG CTG GCT AAA CTT AAC CTA CAA TCC AAC GTT TCT GTG CTT CAA 288
Lys Gly Leu Ala Lys Leu Asn Leu Gln Ser Asn Val Ser Val Leu Gln
85 90 95

GAT GCT TTG GTA GAG AAT ATT GGC GGT GAT GGG GAT TTC AGC GAT TTA 336
Asp Ala Leu Val Glu Asn Ile Gly Gly Asp Gly Asp Phe Ser Asp Leu
100 105 110

ATG AAC CGT GCC AGT CAA TAT GCT GAC GCT GCC TCT ATT CAA TCC CTA 384
Met Asn Arg Ala Ser Gln Tyr Ala Asp Ala Ala Ser Ile Gln Ser Leu
115 120 125

TTT TCA CCG GGC CGT TAT GCT TCC GCA CTC TAC AGA GTT GCT AAA GAT 432
Phe Ser Pro Gly Arg Tyr Ala Ser Ala Leu Tyr Arg Val Ala Lys Asp
130 135 140

CTG CAT AAA TCA GAT TCC AGT TTG CAT ATT GAT AAT CGC CGC GCT GAT 480
Leu His Lys Ser Asp Ser Ser Leu His Ile Asp Asn Arg Arg Ala Asp
145 150 155 160

CTG AAG GAT CTG ATA TTA AGC GAA ACG ACG ATG AAT AAA GAG GTC ACT 528
Leu Lys Asp Leu Ile Leu Ser Glu Thr Thr Met Asn Lys Glu Val Thr
165 170 175

TCC CTT GAT ATC TTG TTG GAT GTG CTA CAA AAA GGC GGT AAA GAT ATT 576
Ser Leu Asp Ile Leu Leu Asp Val Leu Gln Lys Gly Gly Lys Asp Ile
180 185 190

ACT GAG CTG TCC GGC GCA TTC TTC CCA ATG ACG TTA CCT TAT GAC GAT 624
Thr Glu Leu Ser Gly Ala Phe Phe Pro Met Thr Leu Pro Tyr Asp Asp
195 200 205

CAT CTG TCG CAA ATC GAT TCC GCT TTA TCG GCA CAA GCC AGA ACG CTG 672
His Leu Ser Gln Ile Asp Ser Ala Leu Ser Ala Gln Ala Arg Thr Leu
210 215 220

AAC GGT GTG TGG AAT ACT TTG ACA GAT ACC ACG GCA CAA GCG GTT TCA 720
Asn Gly Val Trp Asn Thr Leu Thr Asp Thr Thr Ala Gln Ala Val Ser

225 230 235 240
 GAA CAA ACC AGT AAT ACG AAT ACA CGC AAA CTG TTC GCT GCC CAA GAT 768
 Glu Gln Thr Ser Asn Thr Asn Thr Arg Lys Leu Phe Ala Ala Gln Asp
 245 250 255
 GGT AAT CAA GAT ACA TTT TTT TCC GGA AAC ACT TTT TAT TTC AAA GCG 816
 Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr Phe Tyr Phe Lys Ala
 260 265 270
 GTG GGA TTC AGC GGG CAA CCT ATG GTT TAC CTG TCA CAG TAC ACC AGC 864
 Val Gly Phe Ser Gly Gln Pro Met Val Tyr Leu Ser Gln Tyr Thr Ser
 275 280 285
 GGG AAC GGC ATT GTC GGC GCA CAA TTG ATT GCA GGT AAT CCA GAC CAA 912
 Gly Asn Gly Ile Val Gly Ala Gln Leu Ile Ala Gly Asn Pro Asp Gln
 290 295 300
 GCC GCC GCC GCA ATA GTC GCA CCG TTG AAA CTC ACT TGG TCA ATG GCA 960
 Ala Ala Ala Ala Ile Val Ala Pro Leu Lys Leu Thr Trp Ser Met Ala
 305 310 315 320
 AAA CAG TGT TAC TAC CTC GTC GCT CCC GAT GGT ACA ACG ATG GGA GAC 1008
 Lys Gln Cys Tyr Tyr Leu Val Ala Pro Asp Gly Thr Thr Met Gly Asp
 325 330 335
 GGT AAT GTT CTG ACC GGC TGT TTC TTA AGA GGC AAC AGC CCA ACT AAC 1056
 Gly Asn Val Leu Thr Gly Cys Phe Leu Arg Gly Asn Ser Pro Thr Asn
 340 345 350
 CCG GAT AAA GAC GGT ATT TTT GCT CAG GTA GCC AAC AAA TCA GGC AGT 1104
 Pro Asp Lys Asp Gly Ile Phe Ala Gln Val Ala Asn Lys Ser Gly Ser
 355 360 365
 ACT CAG CCT TTG CCA AGC TTC CAT CTG CCG GTC ACA CTG GAA CAC AGC 1152
 Thr Gln Pro Leu Pro Ser Phe His Leu Pro Val Thr Leu Glu His Ser
 370 375 380
 GAG AAT AAA GAT CAG TAC TAT CTG AAA ACA GAG CAG GGT TAT ATC ACG 1200
 Glu Asn Lys Asp Gln Tyr Tyr Leu Lys Thr Glu Gln Gly Tyr Ile Thr
 385 390 395 400
 GTA GAT AGT TCC GGA CAG TCA AAT TGG AAA AAC GCG CTG GTT ATC AAT 1248
 Val Asp Ser Ser Gly Gln Ser Asn Trp Lys Asn Ala Leu Val Ile Asn
 405 410 415
 GGG ACA AAA GAC AAG GGG CTG TTA TTA ACC TTT TGC AGC GAT AGC TCA 1296
 Gly Thr Lys Asp Lys Gly Leu Leu Leu Thr Phe Cys Ser Asp Ser Ser
 420 425 430
 GGC ACT CCG ACA AAC CCT GAT GAT GTG ATT CCT CCC GCT ATC AAT GAT 1344
 Gly Thr Pro Thr Asn Pro Asp Asp Val Ile Pro Pro Ala Ile Asn Asp
 435 440 445
 ATT CCA TCG CCG CCA GCC CGC GAA ACA CTG TCA CTG ACG CCG GTC AGT 1392
 Ile Pro Ser Pro Pro Ala Arg Glu Thr Leu Ser Leu Thr Pro Val Ser
 450 455 460
 TAT CAA TTG ATG ACC AAT CCG GCA CCG ACA GAA GAT GAT ATT ACC AAC 1440
 Tyr Gln Leu Met Thr Asn Pro Ala Pro Thr Glu Asp Asp Ile Thr Asn

465 470 475 480
 CAT TAT GGT TTT AAC GGC GCT AGC TTA CGG GCT TCT CCA TTG TCA ACC 1488
 His Tyr Gly Phe Asn Gly Ala Ser Leu Arg Ala Ser Pro Leu Ser Thr
 485 490 495
 AGC GAG TTG ACC AGC AAA CTG AAT TCT ATC GAT ACT TTC TGT GAG AAG 1536
 Ser Glu Leu Thr Ser Lys Leu Asn Ser Ile Asp Thr Phe Cys Glu Lys
 500 505 510
 ACC CGG TTA AGC TTC AAT CAG TTA ATG GAT TTG ACC GCT CAG CAA TCT 1584
 Thr Arg Leu Ser Phe Asn Gln Leu Met Asp Leu Thr Ala Gln Gln Ser
 515 520 525
 TAC AGT CAA AGC AGC ATT GAT GCG AAA GCA GCC AGC CGC TAT GTT CGT 1632
 Tyr Ser Gln Ser Ser Ile Asp Ala Lys Ala Ala Ser Arg Tyr Val Arg
 530 535 540
 TTT GGG GAA ACC ACC CCA ACC CGC GTC AAT GTC TAC GGT GCC GCT TAT 1680
 Phe Gly Glu Thr Thr Pro Thr Arg Val Asn Val Tyr Gly Ala Ala Tyr
 545 550 555 560
 CTG AAC AGC ACA CTG GCA GAC GCG GCT GAT GGT CAA TAT CTG TGG ATT 1728
 Leu Asn Ser Thr Leu Ala Asp Ala Ala Asp Gly Gln Tyr Leu Trp Ile
 565 570 575
 CAG ACT GAT GGC AAG AGC CTA AAT TTC ACT GAC GAT ACG GTA GTC GCC 1776
 Gln Thr Asp Gly Lys Ser Leu Asn Phe Thr Asp Asp Thr Val Val Ala
 580 585 590
 TTA GCC GGT CGC GCT GAA AAG CTG GTA CGT TTA TCA TCC CAG ACC GGG 1824
 Leu Ala Gly Arg Ala Glu Lys Leu Val Arg Leu Ser Ser Gln Thr Gly
 595 600 605
 CTA TCA TTT GAA GAA TTG GAC TGG CTG ATT GCC AAT GCC AGT CGT AGT 1872
 Leu Ser Phe Glu Glu Leu Asp Trp Leu Ile Ala Asn Ala Ser Arg Ser
 610 615 620
 GTG CCG GAC CAC CAC GAC AAA ATT GTG CTG GAT AAG CCG GTC CTT GAA 1920
 Val Pro Asp His His Asp Lys Ile Val Leu Asp Lys Pro Val Leu Glu
 625 630 635 640
 GCA CTG GCA GAG TAT GTC AGC CTA AAA CAG CGC TAT GGG CTT GAT GCC 1968
 Ala Leu Ala Glu Tyr Val Ser Leu Lys Gln Arg Tyr Gly Leu Asp Ala
 645 650 655
 AAT ACC TTT GCG ACC TTC ATT AGT GCA GTA AAT CCT TAT ACG CCA GAT 2016
 Asn Thr Phe Ala Thr Phe Ile Ser Ala Val Asn Pro Tyr Thr Pro Asp
 660 665 670
 CAG ACA CCC AGT TTC TAT GAA ACC GCT TTC CGC TCT GCC GAC GGT AAT 2064
 Gln Thr Pro Ser Phe Tyr Glu Thr Ala Phe Arg Ser Ala Asp Gly Asn
 675 680 685
 CAT GTC ATT GCG CTA GGT ACA GAG GTG AAA TAT GCA GAA AAT GAG CAG 2112
 His Val Ile Ala Leu Gly Thr Glu Val Lys Tyr Ala Glu Asn Glu Gln
 690 695 700
 GAT GAG TTA GCC GCC ATA TGC TGC AAA GCA TTG GGT GTC ACC AGT GAT 2160
 Asp Glu Leu Ala Ala Ile Cys Cys Lys Ala Leu Gly Val Thr Ser Asp

705 710 715 720
 GAA CTG CTC CGT ATT GGT CGC TAT TGC TTC GGT AAT GCA GGC AGT TTT 2208
 Glu Leu Leu Arg Ile Gly Arg Tyr Cys Phe Gly Asn Ala Gly Ser Phe
 725 730 735
 ACC TTG GAT GAA TAT ACC GCC AGT CAG TTG TAT CGC TTC GGC GCC ATT 2256
 Thr Leu Asp Glu Tyr Thr Ala Ser Gln Leu Tyr Arg Phe Gly Ala Ile
 740 745 750
 CCC CGT TTG TTT GGG CTG ACA TTT GCC CAA GCC GAA ATT TTA TGG CGT 2304
 Pro Arg Leu Phe Gly Leu Thr Phe Ala Gln Ala Glu Ile Leu Trp Arg
 755 760 765
 CTG ATG GAA GGC GGA AAA GAT ATC TTA TTG CAA CAG TTA GGT CAG GCA 2352
 Leu Met Glu Gly Gly Lys Asp Ile Leu Leu Gln Gln Leu Gly Gln Ala
 770 775 780
 AAA TCC CTG CAA CCA CTG GCT ATT TTA CGC CGT ACC GAG CAG GTG CTG 2400
 Lys Ser Leu Gln Pro Leu Ala Ile Leu Arg Arg Thr Glu Gln Val Leu
 785 790 795 800
 GAT TGG ATG TCG TCC GTA AAT CTA AGT CTG ACT TAT CTG CAA GGG ATG 2448
 Asp Trp Met Ser Ser Val Asn Leu Ser Leu Thr Tyr Leu Gln Gly Met
 805 810 815
 GTA AGT ACG CAA TGG AGC GGT ACC GCC ACC GCT GAG ATG TTC AAT TTC 2496
 Val Ser Thr Gln Trp Ser Gly Thr Ala Thr Ala Glu Met Phe Asn Phe
 820 825 830
 TTG GAA AAC GTT TGT GAC AGC GTG AAT AGT CAA GCT GCC ACT AAA GAA 2544
 Leu Glu Asn Val Cys Asp Ser Val Asn Ser Gln Ala Ala Thr Lys Glu
 835 840 845
 ACA ATG GAT TCG GCG TTA CAG CAG AAA GTG CTG CGG GCG CTA AGC GCC 2592
 Thr Met Asp Ser Ala Leu Gln Gln Lys Val Leu Arg Ala Leu Ser Ala
 850 855 860
 GGT TTC GGC ATT AAG AGC AAT GTG ATG GGT ATC GTC ACC TTC TGG CTG 2640
 Gly Phe Gly Ile Lys Ser Asn Val Met Gly Ile Val Thr Phe Trp Leu
 865 870 875 880
 GAG AAA ATC ACA ATC GGT AGT GAT AAT CCT TTT ACA TTG GCA AAC TAC 2688
 Glu Lys Ile Thr Ile Gly Ser Asp Asn Pro Phe Thr Leu Ala Asn Tyr
 885 890 895
 TGG CAT GAT ATT CAA ACC CTG TTT AGC CAT GAC AAT GCC ACG TTA GAG 2736
 Trp His Asp Ile Gln Thr Leu Phe Ser His Asp Asn Ala Thr Leu Glu
 900 905 910
 TCC TTA CAA ACC GAC ACT TCT CTG GTA ATT GCT ACT CAG CAA CTT AGC 2784
 Ser Leu Gln Thr Asp Thr Ser Leu Val Ile Ala Thr Gln Gln Leu Ser
 915 920 925
 CAG CTA GTG TTA ATT GTG AAA TGG CTG AGC CTG ACC GAG CAG GAT CTG 2832
 Gln Leu Val Leu Ile Val Lys Trp Leu Ser Leu Thr Glu Gln Asp Leu
 930 935 940
 CAA TTA CTG ACA ACC TAT CCC GAA CGT TTA ATC AAC GGC ATC ACG AAT 2880
 Gln Leu Leu Thr Thr Tyr Pro Glu Arg Leu Ile Asn Gly Ile Thr Asn

945 950 955 960
 GTT CCT GTA CCC AAT CCG GAG CTA TTA CTC ACG CTA TCA CGT TTT AAG 2928
 Val Pro Val Pro Asn Pro Glu Leu Leu Leu Thr Leu Ser Arg Phe Lys
 965 970 975
 CAG TGG GAA ACT CAA GTC ACC GTT TCC CGT GAT GAA GCG ATG CGC TGT 2976
 Gln Trp Glu Thr Gln Val Thr Val Ser Arg Asp Glu Ala Met Arg Cys
 980 985 990
 TTC GAT CAA TTA AAT GCC AAT GAT ATG ACG ACT GAA AAT GCA GGT TCA 3024
 Phe Asp Gln Leu Asn Ala Asn Asp Met Thr Thr Glu Asn Ala Gly Ser
 995 1000 1005
 CTG ATC GCC ACA TTG TAT GAG ATG GAT AAA GGT ACG GGA GCG CAA GTT 3072
 Leu Ile Ala Thr Leu Tyr Glu Met Asp Lys Gly Thr Gly Ala Gln Val
 1010 1015 1020
 AAT ACC TTG CTA TTA GGT GAA AAT AAC TGG CCG AAA AGT TTT ACC TCT 3120
 Asn Thr Leu Leu Leu Gly Glu Asn Asn Trp Pro Lys Ser Phe Thr Ser
 1025 1030 1035 1040
 CTC TGG CAA CTT CTG ACC TGG TTA CGC GTC GGG CAA AGA CTG AAT GTC 3168
 Leu Trp Gln Leu Leu Thr Trp Leu Arg Val Gly Gln Arg Leu Asn Val
 1045 1050 1055
 GGT AGT ACC ACT CTG GGC AAT CTG TTG TCC ATG ATG CAA GCA GAC CCT 3216
 Gly Ser Thr Thr Leu Gly Asn Leu Leu Ser Met Met Gln Ala Asp Pro
 1060 1065 1070
 GCT GCC GAG AGT AGC GCT TTA TTG GCA TCA GTA GCC CAA AAC TTA AGT 3264
 Ala Ala Glu Ser Ser Ala Leu Leu Ala Ser Val Ala Gln Asn Leu Ser
 1075 1080 1085
 GCC GCA ATC AGC AAT CGT CAG TAA 3288
 Ala Ala Ile Ser Asn Arg Gln v ?
 1090 1095

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1095 amino acids
- (B) TYPE: amino acids
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34 (TcaA protein):

| Features | From | To | Description |
|----------|------|-----|----------------|
| | 254 | 267 | SEQ ID NO:15 |
| | 254 | 492 | TcaAii peptide |

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

50 55 60
 Leu Lys Ser Leu Tyr Arg Ala Trp Gln Leu Arg Gln Glu Pro Val Ile
 65 70 75 80
 Lys Gly Leu Ala Lys Leu Asn Leu Gln Ser Asn Val Ser Val Leu Gln
 85 90 95
 Asp Ala Leu Val Glu Asn Ile Gly Gly Asp Gly Asp Phe Ser Asp Leu
 100 105 110
 Met Asn Arg Ala Ser Gln Tyr Ala Asp Ala Ala Ser Ile Gln Ser Leu
 115 120 125
 Phe Ser Pro Gly Arg Tyr Ala Ser Ala Leu Tyr Arg Val Ala Lys Asp
 130 135 140
 Leu His Lys Ser Asp Ser Ser Leu His Ile Asp Asn Arg Arg Ala Asp
 145 150 155 160
 Leu Lys Asp Leu Ile Leu Ser Glu Thr Thr Met Asn Lys Glu Val Thr
 165 170 175
 Ser Leu Asp Ile Leu Leu Asp Val Leu Gln Lys Gly Gly Lys Asp Ile
 180 185 190
 Thr Glu Leu Ser Gly Ala Phe Phe Pro Met Thr Leu Pro Tyr Asp Asp
 195 200 205
 His Leu Ser Gln Ile Asp Ser Ala Leu Ser Ala Gln Ala Arg Thr Leu
 210 215 220
 Asn Gly Val Trp Asn Thr Leu Thr Asp Thr Thr Ala Gln Ala Val Ser
 225 230 235 240
 Glu Gln Thr Ser Asn Thr Asn Thr Arg Lys Leu Phe Ala Ala Gln Asp
 245 250 255
 Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr Phe Tyr Phe Lys Ala
 260 265 270
 Val Gly Phe Ser Gly Gln Pro Met Val Tyr Leu Ser Gln Tyr Thr Ser
 275 280 285
 Gly Asn Gly Ile Val Gly Ala Gln Leu Ile Ala Gly Asn Pro Asp Gln
 290 295 300
 Ala Ala Ala Ala Ile Val Ala Pro Leu Lys Leu Thr Trp Ser Met Ala
 305 310 315 320
 Lys Gln Cys Tyr Tyr Leu Val Ala Pro Asp Gly Thr Thr Met Gly Asp
 325 330 335
 Gly Asn Val Leu Thr Gly Cys Phe Leu Arg Gly Asn Ser Pro Thr Asn
 340 345 350
 Pro Asp Lys Asp Gly Ile Phe Ala Gln Val Ala Asn Lys Ser Gly Ser
 355 360 365
 Thr Gln Pro Leu Pro Ser Phe His Leu Pro Val Thr Leu Glu His Ser
 370 375 380
 Glu Asn Lys Asp Gln Tyr Tyr Leu Lys Thr Glu Gln Gly Tyr Ile Thr
 385 390 395 400
 Val Asp Ser Ser Gly Gln Ser Asn Trp Lys Asn Ala Leu Val Ile Asn
 405 410 415

Gly Thr Lys Asp Lys Gly Leu Leu Leu Thr Phe Cys Ser Asp Ser Ser
 420 425 430

Gly Thr Pro Thr Asn Pro Asp Asp Val Ile Pro Pro Ala Ile Asn Asp
 435 440 445

Ile Pro Ser Pro Pro Ala Arg Glu Thr Leu Ser Leu Thr Pro Val Ser
 450 455 460

Tyr Gln Leu Met Thr Asn Pro Ala Pro Thr Glu Asp Asp Ile Thr Asn
 465 470 475 480

His Tyr Gly Phe Asn Gly Ala Ser Leu Arg Ala Ser Pro Leu Ser Thr
 485 490 W4 >> 495

Ser Glu Leu Thr Ser Lys Leu Asn Ser Ile Asp Thr Phe Cys Glu Lys
 500 505 510

Thr Arg Leu Ser Phe Asn Gln Leu Met Asp Leu Thr Ala Gln Gln Ser
 515 520 525

Tyr Ser Gln Ser Ser Ile Asp Ala Lys Ala Ala Ser Arg Tyr Val Arg
 530 535 540

Phe Gly Glu Thr Thr Pro Thr Arg Val Asn Val Tyr Gly Ala Ala Tyr
 545 550 555 560

Leu Asn Ser Thr Leu Ala Asp Ala Ala Asp Gly Gln Tyr Leu Trp Ile
 565 570 575

Gln Thr Asp Gly Lys Ser Leu Asn Phe Thr Asp Asp Thr Val Val Ala
 580 585 590

Leu Ala Gly Arg Ala Glu Lys Leu Val Arg Leu Ser Ser Gln Thr Gly
 595 600 605

Leu Ser Phe Glu Glu Leu Asp Trp Leu Ile Ala Asn Ala Ser Arg Ser
 610 615 620

Val Pro Asp His His Asp Lys Ile Val Leu Asp Lys Pro Val Leu Glu
 625 630 635 640

Ala Leu Ala Glu Tyr Val Ser Leu Lys Gln Arg Tyr Gly Leu Asp Ala
 645 650 655

Asn Thr Phe Ala Thr Phe Ile Ser Ala Val Asn Pro Tyr Thr Pro Asp
 660 665 670

Gln Thr Pro Ser Phe Tyr Glu Thr Ala Phe Arg Ser Ala Asp Gly Asn
 675 680 685

His Val Ile Ala Leu Gly Thr Glu Val Lys Tyr Ala Glu Asn Glu Gln
 690 695 700

Asp Glu Leu Ala Ala Ile Cys Cys Lys Ala Leu Gly Val Thr Ser Asp
 705 710 715 720

Glu Leu Leu Arg Ile Gly Arg Tyr Cys Phe Gly Asn Ala Gly Ser Phe
 725 730 735

Thr Leu Asp Glu Tyr Thr Ala Ser Gln Leu Tyr Arg Phe Gly Ala Ile
 740 745 750

Pro Arg Leu Phe Gly Leu Thr Phe Ala Gln Ala Glu Ile Leu Trp Arg
 755 760 765

Leu Met Glu Gly Gly Lys Asp Ile Leu Leu Gln Gln Leu Gly Gln Ala

770 775 780
 Lys Ser Leu Gln Pro Leu Ala Ile Leu Arg Arg Thr Glu Gln Val Leu
 785 790 795 800
 Asp Trp Met Ser Ser Val Asn Leu Ser Leu Thr Tyr Leu Gln Gly Met
 805 810 815
 Val Ser Thr Gln Trp Ser Gly Thr Ala Thr Ala Glu Met Phe Asn Phe
 820 825 830
 Leu Glu Asn Val Cys Asp Ser Val Asn Ser Gln Ala Ala Thr Lys Glu
 835 840 845
 Thr Met Asp Ser Ala Leu Gln Gln Lys Val Leu Arg Ala Leu Ser Ala
 850 855 860
 Gly Phe Gly Ile Lys Ser Asn Val Met Gly Ile Val Thr Phe Trp Leu
 865 870 875 880
 Glu Lys Ile Thr Ile Gly Ser Asp Asn Pro Phe Thr Leu Ala Asn Tyr
 885 890 895
 Trp His Asp Ile Gln Thr Leu Phe Ser His Asp Asn Ala Thr Leu Glu
 900 905 910
 Ser Leu Gln Thr Asp Thr Ser Leu Val Ile Ala Thr Gln Gln Leu Ser
 915 920 925
 Gln Leu Val Leu Ile Val Lys Trp Leu Ser Leu Thr Glu Gln Asp Leu
 930 935 940
 Gln Leu Leu Thr Thr Tyr Pro Glu Arg Leu Ile Asn Gly Ile Thr Asn
 945 950 955 960
 Val Pro Val Pro Asn Pro Glu Leu Leu Leu Thr Leu Ser Arg Phe Lys
 965 970 975
 Gln Trp Glu Thr Gln Val Thr Val Ser Arg Asp Glu Ala Met Arg Cys
 980 985 990
 Phe Asp Gln Leu Asn Ala Asn Asp Met Thr Thr Glu Asn Ala Gly Ser
 995 1000 1005
 Leu Ile Ala Thr Leu Tyr Glu Met Asp Lys Gly Thr Gly Ala Gln Val
 1010 1015 1020
 Asn Thr Leu Leu Leu Gly Glu Asn Asn Trp Pro Lys Ser Phe Thr Ser
 1025 1030 1035 1040
 Leu Trp Gln Leu Leu Thr Trp Leu Arg Val Gly Gln Arg Leu Asn Val
 1045 1050 1055
 Gly Ser Thr Thr Leu Gly Asn Leu Leu Ser Met Met Gln Ala Asp Pro
 1060 1065 1070
 Ala Ala Glu Ser Ser Ala Leu Leu Ala Ser Val Ala Gln Asn Leu Ser
 1075 1080 1085
 Ala Ala Ile Ser Asn Arg Gln v?
 1090 1095

(2) INFORMATION FOR SEQ ID NO:35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 603 amino acids
- (B) TYPE: amino acid

(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35 (TcaAiii protein):

Pro Leu Ser Thr Ser Glu Leu Thr Ser Lys Leu Asn Ser Ile Asp Thr
 1 5 10 15
 Phe Cys Glu Lys Thr Arg Leu Ser Phe Asn Gln Leu Met Asp Leu Thr
 20 25 30
 Ala Gln Gln Ser Tyr Ser Gln Ser Ser Ile Asp Ala Lys Ala Ala Ser
 35 40 45
 Arg Tyr Val Arg Phe Gly Glu Thr Thr Pro Thr Arg Val Asn Val Tyr
 50 55 60
 Gly Ala Ala Tyr Leu Asn Ser Thr Leu Ala Asp Ala Ala Asp Gly Gln
 65 70 75 80
 Tyr Leu Trp Ile Gln Thr Asp Gly Lys Ser Leu Asn Phe Thr Asp Asp
 85 90 95
 Thr Val Val Ala Leu Ala Gly Arg Ala Glu Lys Leu Val Arg Leu Ser
 100 105 110
 Ser Gln Thr Gly Leu Ser Phe Glu Glu Leu Asp Trp Leu Ile Ala Asn
 115 120 125
 Ala Ser Arg Ser Val Pro Asp His His Asp Lys Ile Val Leu Asp Lys
 130 135 140
 Pro Val Leu Glu Ala Leu Ala Glu Tyr Val Ser Leu Lys Gln Arg Tyr
 145 150 155 160
 Gly Leu Asp Ala Asn Thr Phe Ala Thr Phe Ile Ser Ala Val Asn Pro
 165 170 175
 Tyr Thr Pro Asp Gln Thr Pro Ser Phe Tyr Glu Thr Ala Phe Arg Ser
 180 185 190
 Ala Asp Gly Asn His Val Ile Ala Leu Gly Thr Glu Val Lys Tyr Ala
 195 200 205
 Glu Asn Glu Gln Asp Glu Leu Ala Ala Ile Cys Cys Lys Ala Leu Gly
 210 215 220
 Val Thr Ser Asp Glu Leu Leu Arg Ile Gly Arg Tyr Cys Phe Gly Asn
 225 230 235 240
 Ala Gly Arg Phe Thr Leu Asp Glu Tyr Thr Ala Ser Gln Leu Tyr Arg
 245 250 255
 Phe Gly Ala Ile Pro Arg Leu Phe Gly Leu Thr Phe Ala Gln Ala Glu
 260 265 270
 Ile Leu Trp Arg Leu Met Glu Gly Gly Lys Asp Ile Leu Leu Gln Gln
 275 280 285
 Xxx Gly Gln Ala Lys Ser Leu Gln Pro Leu Ala Ile Leu Arg Arg Thr
 290 295 300
 Glu Gln Val Leu Asp Trp Met Ser Pro Val Asn Leu Ser Leu Thr Tyr
 305 310 315 320
 Leu Gln Gly Met Val Ser Thr Gln Trp Ser Gly Thr Ala Thr Ala Glu
 325 330 335

Met Phe Asn Phe Leu Glu Asn Val Cys Asp Ser Val Asn Ser Gln Ala
 340 345 350
 Xxx Thr Lys Glu Thr Met Asp Ser Ala Leu Gln Gln Lys Val Leu Arg
 355 360 365
 Ala Leu Ser Ala Gly Phe Gly Ile Lys Ser Asn Val Met Gly Ile Val
 370 375 380
 Thr Phe Trp Leu Glu Lys Ile Thr Ile Gly Arg Asp Asn Pro Phe Thr
 385 390 395 400
 Leu Ala Asn Tyr Trp His Asp Ile Gln Thr Leu Phe Ser His Asp Asn
 405 410 415
 Ala Thr Leu Glu Ser Leu Gln Thr Asp Thr Ser Leu Val Ile Ala Thr
 420 425 430
 Gln Gln Leu Ser Gln Leu Val Leu Ile Val Lys Trp Val Ser Leu Thr
 435 440 445
 Glu Gln Asp Leu Gln Leu Leu Thr Thr Tyr Pro Glu Arg Leu Ile Asn
 450 455 460
 Gly Ile Thr Asn Val Pro Val Pro Asn Pro Glu Leu Leu Leu Thr Leu
 465 470 475 480
 Ser Arg Phe Lys Gln Trp Glu Thr Gln Val Thr Val Ser Arg Asp Glu
 485 490 495
 Ala Met Arg Cys Phe Asp Gln Leu Asn Ala Asn Asp Met Thr Thr Glu
 500 505 510
 Asn Ala Gly Ser Leu Ile Ala Thr Leu Tyr Glu Met Asp Lys Gly Thr
 515 520 525
 Gly Ala Gln Val Asn Thr Leu Leu Leu Gly Glu Asn Asn Trp Pro Lys
 530 535 540
 Ser Phe Thr Ser Leu Trp Gln Leu Leu Thr Trp Leu Arg Val Gly Gln
 545 550 555 560
 Arg Leu Asn Val Gly Ser Thr Thr Leu Gly Asn Leu Leu Ser Met Met
 565 570 575
 Gln Ala Asp Pro Ala Ala Glu Ser Ser Ala Leu Leu Ala Ser Val Ala
 580 585 590
 Gln Asn Leu Ser Ala Ala Ile Ser Asn Arg Gln *

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2557 base pairs
- (B) TYPE: nucleic acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36 (tcdA internal

fragment):

GAATTGGGCT TGGCTTAAAT ATTGATGATG TCTCGCTCTT CCGCCTGCTT AAAATTACCG 60
 ACCATGATAA TAAAGATGGA AAAATTAATA ATACCTAAA GAATCTTTCC AATTTATATA 120
 TTGAAAATT ACTGGCAGAT ATTCATCAAT TAACCATTGA TGAAGTGGAT TTATTACTGA 180

TTGCCGTAGG TGAAGGAAAA ACTAATTTAT CCGCTATCAG TGATAAGCAA TTGGCTACCC 240
 TGATCAGAAA ACTCAATACT ATTACCAGCT GGCTACATAC ACAGAAGTGG AGTGTATTCC 300
 AGCTATTTAT CATGACCTCC ACCAGCTATA ACAAACGCT AACGCCTGAA ATTAAGAATT 360
 TGCTGGATAC CGTCTACCAC GGTTTACAAG GTTTTGATAA AGACAAAGCA GATTTGCTAC 420
 ATGTCATGGC GCCCTATATT GCGGCCACCT TGCAATTATC ATCGGAAAAT GTCGCCCACT 480
 CGGTACTCCT TTGGGCAGAT AAGTTACAGC CCGGCGACGG CGCAATGACA GCAGAGGGAN 540
 TCTGGGACTG GTTGAATACT AAGTATACGC CGGGTTTCATC GGAAGCCGTA GAAACGCAGG 600
 AACATATCGT TCAGTATTGT CAGGCTCTGG CACAATTGGA AATGGTTTAC CATTCCACCG 660
 GCATCAACGA AAACGCCTTC CGTCTATTTG TGACAAAACC AGAGATGTTT GGCCTGCAA 720
 CTGGAGCAGC GCCCGCGCAT GATGCCCTTT CACTGATTAT GCTGACACGT TTTGCGGATT 780
 GGGTGAACGC ACTAGGCGAA AAAGCGTCCT CGGTGCTAGC GGCATTTGAA GCTAACTCGT 840
 TAACGGCAGA ACAACTGGCT GATGCCATGA ATCTTGATGC TAATTTGCTG TTGCAAGCCA 900
 GTATTCAAGC ACAAATCAT CAACATCTTC CCCAGTAAC TCCAGAAAAT GCGTTCTCCT 960
 GTTGACATC TATCAATACT ATCCTGCAAT GGGTTAATGT CGCACACAA TTGAAATGTC 1020
 GCCCCACAGG GCGTTTCCGC TTTGGTCGGG CTGGATTATA TTCAATCAAT GAAAGAGACA 1080
 CCGACCTATG CCCAGTGGGA AAACGCGGCA GCGTATTAA CCGCCGGGT GAATTCAACA 1140
 ACAGGCTAAT ACATTACAAC GCTTTTCTGG ATGAATCTCG CAGTGCCGCA TTAAGCACCT 1200
 ACTATATCCG TCAAGTCGCC AAGGCAGCGG CGGCTATTAA AAGCCGTGAT GACTTGTATC 1260
 AATACTTACT GATTGATAAT CAGGTTTCTG CGGCAATAAA AACCACCCGG ATCGCCGAAG 1320
 CCATTGCCAG TATTCAACTG TACGTCAACC GGGCATTGGA AAATGTGGAA GAAAATGCCA 1380
 ATTCGGGGGT TATCAGCCGC CAATTCTTTA TCGACTGGGA CAAATACAAT AAACGCTACA 1440
 GCACTTGGGC GGGTGTCTT CAATTAGTTT ACTACCCGGA AACTATATT GATCCGACCA 1500
 TGCATATCGG ACAAACAAA ATGATGGACG CATTACTGCA ATCCGTCAGC CAAAGCCAAT 1560
 TAAACGCCGA TACCGTCGAA GATGCCTTTA TGTCTTATCT GACATCGTTT GAACAAGTGG 1620
 CTAATCTTAA AGTTATTAGC GCATATCAGC ATAATATTA TAACGATCAA GGGCTGACCT 1680
 ATTTTATCGG ACTCAGTGAA ACTGATGCCG GTGAATATTA TTGGCGCAGT GTCGATCACA 1740
 GTAAATTCOA CGACGGTAAA TTCGCGGCTA ATGCCTGGAG TGAATGGCAT AAAATTGATT 1800
 GTCCAAATTA CCTTATAAA AGCACTATCC GTCCAGTGAT ATATAAATCC CGCCTGTATC 1860
 TGCTCTGGTT GGAACAAAAG GAGATCACC AACAGACAGG AAATAGTAAA GATGGCTATC 1920
 AACTGAAAC GGATTATCGT TATGAACTAA AATTGGCGCA TATCCGCTAT GATGGCACTT 1980
 GGAATACGCC AATCACCTTT GATGTCAATA AAAAAATATC CGAGCTAAAA CTGGAACAAA 2040
 ATAGAGCGCC CGGACTCTAT TGTGCCGTT ATCAAGGTGA AGATACGTTG CTGGTGATGT 2100
 TTTATAACCA ACAAGACACA CTAGATAGTT ATAAAAACGC TTCAATGCAA GACTATATA 2160
 TCTTTGCTGA TATGGCATCC AAAGATATGA CCCAGAACA GAGCAATGTT TATCGGATA 2220
 ATAGCTATCA ACAATTTGAT ACCAATAATG TCAGAAGAGT GAATAACCGC TATGCAGAGG 2280
 ATTATGAGAT TCCTTCTCG GTAAGTAGCC GTAAAGACTA TGGTTGGGA GATTATTACC 2340
 TCAGCATGGT ATATAACGGA GATATTCCAA CTATCAATTA CAAAGCCGCA TCAAGTATT 2400
 TAAAAATTA TATTCACCA AAATTAAGAA TTATTCATA TGGATATGAA GGACAGAAGC 2460
 GCAATCAATG CAATTTGATG AATAAATATG GCAAACCTAGG TGATAAATTT ATTGTGTATA 2520
 CCAGCCTGGG CGTTAATCCG AATAATAAGC CGAATTC 2557

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 845 amino acids
- (B) TYPE: amino acids
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein (partial)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37 (TcdA internal peptide):

Ala Phe Asn Ile Asp Asp Val Ser Leu Phe Arg Leu Leu Lys Ile Thr
 1 5 10 15
 Asp His Asp Asn Lys Asp Gly Lys Ile Lys Asn Asn Leu Lys Asn Leu
 20 25 30
 Ser Asn Leu Tyr Ile Gly Lys Leu Leu Ala Asp Ile His Gln Leu Thr
 35 40 45
 Ile Asp Glu Leu Asp Leu Leu Leu Ile Ala Val Gly Glu Gly Lys Thr
 50 55 60
 Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Thr Leu Ile Arg Lys
 65 70 75 80
 Leu Asn Thr Ile Thr Ser Trp Leu His Thr Gln Lys Trp Ser Val Phe
 85 90 95
 Gln Leu Phe Ile Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr Pro
 100 105 110
 Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly Phe
 115 120 125
 Asp Lys Asp Lys Ala Asp Leu Leu His Val Met Ala Pro Tyr Ile Ala
 130 135 140
 Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu Leu
 145 150 155 160
 Trp Ala Asp Lys Leu Gln Pro Gly Asp Gly Ala Met Thr Ala Glu Gly
 165 170 175
 Phe Trp Asp Trp Leu Asn Thr Lys Tyr Thr Pro Gly Ser Ser Glu Ala
 180 185 190
 Val Glu Thr Gln Glu His Ile Val Gln Tyr Cys Gln Ala Leu Ala Gln
 195 200 205
 Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe Arg
 210 215 220
 Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ala Ala Thr Gly Ala Ala
 225 230 235 240
 Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala Asp
 245 250 255
 Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala Phe
 260 265 270
 Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn Leu
 275 280 285
 Asp Ala Asn Leu Leu Leu Gln Ala Ser Ile Gln Ala Gln Asn His Gln
 290 295 300
 His Leu Pro Pro Val Thr Pro Glu Asn Ala Phe Ser Cys Trp Thr Ser
 305 310 315 320
 Ile Asn Thr Ile Leu Gln Trp Val Asn Val Ala Gln Gln Leu Lys Cys
 325 330 335

Arg Pro Thr Gly Arg Phe Arg Phe Gly Arg Ala Gly Leu Tyr Ser Ile
 340 345 350

Asn Glu Arg Asp Thr Asp Leu Cys Pro Val Gly Lys Arg Gly Arg Arg
 355 360 365

Ile Asn Arg Arg Val Glu Phe Asn Asn Arg Leu Ile His Tyr Asn Ala
 370 375 380

Phe Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr Tyr Ile Arg
 385 390 395 400

Gln Val Ala Lys Ala Ala Ala Ala Ile Lys Ser Arg Asp Asp Leu Tyr
 405 410 415

Gln Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Ala Ile Lys Thr Thr
 420 425 430

Arg Ile Ala Glu Ala Ile Ala Ser Ile Gln Leu Tyr Val Asn Arg Ala
 435 440 445

Leu Glu Asn Val Glu Glu Asn Ala Asn Ser Gly Val Ile Ser Arg Gln
 450 455 460

Phe Phe Ile Asp Trp Asp Lys Tyr Asn Lys Arg Tyr Ser Thr Trp Ala
 465 470 475 480

Gly Val Ser Gln Leu Val Tyr Tyr Pro Glu Asn Tyr Ile Asp Pro Thr
 485 490 495

Met Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln Ser Val
 500 505 510

Ser Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe Met Ser
 515 520 525

Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile Ser Ala
 530 535 540

Tyr His Asp Asn Ile Asn Asn Asp Gln Gly Leu Thr Tyr Phe Ile Gly
 545 550 555 560

Leu Ser Glu Thr Asp Ala Gly Glu Tyr Tyr Trp Arg Ser Val Asp His
 565 570 575

Ser Lys Phe Asn Asp Gly Lys Phe Ala Ala Asn Ala Trp Ser Glu Trp
 580 585 590

His Lys Ile Asp Cys Pro Ile Asn Pro Tyr Lys Ser Thr Ile Arg Pro
 595 600 605

Val Ile Tyr Lys Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln Lys Glu
 610 615 620

Ile Thr Lys Gln Thr Gly Asn Ser Lys Asp Gly Tyr Gln Thr Glu Thr
 625 630 635 640

Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Thr
 645 650 655

Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Lys Lys Ile Ser Glu Leu
 660 665 670

Lys Leu Glu Lys Asn Arg Ala Pro Gly Leu Tyr Cys Ala Gly Tyr Gln
 675 680 685

Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln Asp Thr Leu

Ile Ser Pro Ala Lys

20

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40 (TcbAiii N-terminus):

Ala Asn Ser Leu Thr Ala Leu Phe Leu Pro Gln

1 5 10

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41 (TcdAiii N-terminus):

Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln

1 5 10

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42 (TcdA-pk57 internal peptide):

Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu Ala Glu Val Tyr

1 5 10 15

Ala Gly Leu Glu

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43 (TcdAiii-pK20

internal peptide):

Ile Arg Glu Asp Tyr Pro Ala Ser Leu Gly Lys

1 5 10

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Asp Asp Ser Gly Asp Asp Asp Lys Val Thr Asn Thr Asp Ile His Arg

1 5 10 15

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Asp Val Xaa Gly Ser Glu Lys Ala Asn Glu Lys Leu Lys

1 5 10

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7551 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46 (tcdA):

ATG AAC GAG TCT GTA AAA GAG ATA CCT GAT GTA TTA AAA AGC CAG TGT 48

Met Asn Glu Ser Val Lys Glu Ile Pro Asp Val Leu Lys Ser Gln Cys

1 5 10 15

GGT TTT AAT TGT CTG ACA GAT ATT AGC CAC AGC TCT TTT AAT GAA TTT 96

Gly Phe Asn Cys Leu Thr Asp Ile Ser His Ser Ser Phe Asn Glu Phe

| | |
|---|-----|
| CGC CAG CAA GTA TCT GAG CAC CTC TCC TGG TCC GAA ACA CAC GAC TTA | 144 |
| Arg Gln Gln Val Ser Glu His Leu Ser Trp Ser Glu Thr His Asp Leu | |
| 35 40 45 | |
| TAT CAT GAT GCA CAA CAG GCA CAA AAG GAT AAT CGC CTG TAT GAA GCG | 192 |
| Tyr His Asp Ala Gln Gln Ala Gln Lys Asp Asn Arg Leu Tyr Glu Ala | |
| 50 55 60 | |
| CGT ATT CTC AAA CGC GCC AAT CCC CAA TTA CAA AAT GCG GTG CAT CTT | 240 |
| Arg Ile Leu Lys Arg Ala Asn Pro Gln Leu Gln Asn Ala Val His Leu | |
| 65 70 75 80 | |
| GCC ATT CTC GCT CCC AAT GCT GAA CTG ATA GGC TAT AAC AAT CAA TTT | 288 |
| Ala Ile Leu Ala Pro Asn Ala Glu Leu Ile Gly Tyr Asn Asn Gln Phe | |
| 85 90 95 | |
| AGC GGT AGA GCC AGT CAA TAT GTT GCG CCG GGT ACC GTT TCT TCC ATG | 336 |
| Ser Gly Arg Ala Ser Gln Tyr Val Ala Pro Gly Thr Val Ser Ser Met | |
| 100 105 110 | |
| TTC TCC CCC GCC GCT TAT TTG ACT GAA CTT TAT CGT GAA GCA CGC AAT | 384 |
| Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Arg Asn | |
| 115 120 125 | |
| TTA CAC GCA AGT GAC TCC GTT TAT TAT CTG GAT ACC CGC CGC CCA GAT | 432 |
| Leu His Ala Ser Asp Ser Val Tyr Tyr Leu Asp Thr Arg Arg Pro Asp | |
| 130 135 140 | |
| CTC AAA TCA ATG GCG CTC AGT CAG CAA AAT ATG GAT ATA GAA TTA TCC | 480 |
| Leu Lys Ser Met Ala Leu Ser Gln Gln Asn Met Asp Ile Glu Leu Ser | |
| 145 150 155 160 | |
| ACA CTC TCT TTG TCC AAT GAG CTG TTA TTG GAA AGC ATT AAA ACT GAA | 528 |
| Thr Leu Ser Leu Ser Asn Glu Leu Leu Leu Glu Ser Ile Lys Thr Glu | |
| 165 170 175 | |
| TCT AAA CTG GAA AAC TAT ACT AAA GTG ATG GAA ATG CTC TCC ACT TTC | 576 |
| Ser Lys Leu Glu Asn Tyr Thr Lys Val Met Glu Met Leu Ser Thr Phe | |
| 180 185 190 | |
| CGT CCT TCC GGC GCA ACG CCT TAT CAT GAT GCT TAT GAA AAT GTG CGT | 624 |
| Arg Pro Ser Gly Ala Thr Pro Tyr His Asp Ala Tyr Glu Asn Val Arg | |
| 195 200 205 | |
| GAA GTT ATC CAG CTA CAA GAT CCT GGA CTT GAG CAA CTC AAT GCA TCA | 672 |
| Glu Val Ile Gln Leu Gln Asp Pro Gly Leu Glu Gln Leu Asn Ala Ser | |
| 210 215 220 | |
| CCG GCA ATT GCC GGG TTG ATG CAT CAA GCC TCC CTA TTG GGT ATT AAC | 720 |
| Pro Ala Ile Ala Gly Leu Met His Gln Ala Ser Leu Leu Gly Ile Asn | |
| 225 230 235 240 | |
| GCT TCA ATC TCG CCT GAG CTA TTT AAT ATT CTG ACG GAG GAG ATT ACC | 768 |
| Ala Ser Ile Ser Pro Glu Leu Phe Asn Ile Leu Thr Glu Glu Ile Thr | |
| 245 250 255 | |
| GAA GGT AAT GCT GAG GAA CTT TAT AAG AAA AAT TTT GGT AAT ATC GAA | 816 |
| Glu Gly Asn Ala Glu Glu Leu Tyr Lys Lys Asn Phe Gly Asn Ile Glu | |

| | | | |
|---|-----|-----|------|
| 260 | 265 | 270 | |
| CCG GCC TCA TTG GCT ATG CCG GAA TAC CTT AAA CGT TAT TAT AAT TTA | | | 864 |
| Pro Ala Ser Leu Ala Met Pro Glu Tyr Leu Lys Arg Tyr Tyr Asn Leu | | | |
| 275 | 280 | 285 | |
| AGC GAT GAA GAA CTT AGT CAG TTT ATT GGT AAA GCC AGC AAT TTT GGT | | | 912 |
| Ser Asp Glu Glu Leu Ser Gln Phe Ile Gly Lys Ala Ser Asn Phe Gly | | | |
| 290 | 295 | 300 | |
| CAA CAG GAA TAT AGT AAT AAC CAA CTT ATT ACT CCG GTA GTC AAC AGC | | | 960 |
| Gln Gln Glu Tyr Ser Asn Asn Gln Leu Ile Thr Pro Val Val Asn Ser | | | |
| 305 | 310 | 315 | 320 |
| AGT GAT GGC ACG GTT AAG GTA TAT CGG ATC ACC CGC GAA TAT ACA ACC | | | 1008 |
| Ser Asp Gly Thr Val Lys Val Tyr Arg Ile Thr Arg Glu Tyr Thr Thr | | | |
| 325 | 330 | 335 | |
| AAT GCT TAT CAA ATG GAT GTG GAG CTA TTT CCC TTC GGT GGT GAG AAT | | | 1056 |
| Asn Ala Tyr Gln Met Asp Val Glu Leu Phe Pro Phe Gly Gly Glu Asn | | | |
| 340 | 345 | 350 | |
| TAT CGG TTA GAT TAT AAA TTC AAA AAT TTT TAT AAT GCC TCT TAT TTA | | | 1104 |
| Tyr Arg Leu Asp Tyr Lys Phe Lys Asn Phe Tyr Asn Ala Ser Tyr Leu | | | |
| 355 | 360 | 365 | |
| TCC ATC AAG TTA AAT GAT AAA AGA GAA CTT GTT CGA ACT GAA GGC GCT | | | 1152 |
| Ser Ile Lys Leu Asn Asp Lys Arg Glu Leu Val Arg Thr Glu Gly Ala | | | |
| 370 | 375 | 380 | |
| CCT CAA GTC AAT ATA GAA TAC TCC GCA AAT ATC ACA TTA AAT ACC GCT | | | 1200 |
| Pro Gln Val Asn Ile Glu Tyr Ser Ala Asn Ile Thr Leu Asn Thr Ala | | | |
| 385 | 390 | 395 | 400 |
| GAT ATC AGT CAA CCT TTT GAA ATT GGC CTG ACA CGA GTA CTT CCT TCC | | | 1248 |
| Asp Ile Ser Gln Pro Phe Glu Ile Gly Leu Thr Arg Val Leu Pro Ser | | | |
| 405 | 410 | 415 | |
| GGT TCT TGG GCA TAT GCC GCC GCA AAA TTT ACC GTT GAA GAG TAT AAC | | | 1296 |
| Gly Ser Trp Ala Tyr Ala Ala Ala Lys Phe Thr Val Glu Glu Tyr Asn | | | |
| 420 | 425 | 430 | |
| CAA TAC TCT TTT CTG CTA AAA CTT AAC AAG GCT ATT CGT CTA TCA CGT | | | 1344 |
| Gln Tyr Ser Phe Leu Leu Lys Leu Asn Lys Ala Ile Arg Leu Ser Arg | | | |
| 435 | 440 | 445 | |
| GCG ACA GAA TTG TCA CCC ACG ATT CTG GAA GGC ATT GTG CGC AGT GTT | | | 1392 |
| Ala Thr Glu Leu Ser Pro Thr Ile Leu Glu Gly Ile Val Arg Ser Val | | | |
| 450 | 455 | 460 | |
| AAT CTA CAA CTG GAT ATC AAC ACA GAC GTA TTA GGT AAA GTT TTT CTG | | | 1440 |
| Asn Leu Gln Leu Asp Ile Asn Thr Asp Val Leu Gly Lys Val Phe Leu | | | |
| 465 | 470 | 475 | 480 |
| ACT AAA TAT TAT ATG CAG CGT TAT GCT ATT CAT GCT GAA ACT GCC CTG | | | 1488 |
| Thr Lys Tyr Tyr Met Gln Arg Tyr Ala Ile His Ala Glu Thr Ala Leu | | | |
| 485 | 490 | 495 | |
| ATA CTA TGC AAC GCG CCT ATT TCA CAA CGT TCA TAT GAT AAT CAA CCT | | | 1536 |
| Ile Leu Cys Asn Ala Pro Ile Ser Gln Arg Ser Tyr Asp Asn Gln Pro | | | |

| | | | |
|---|-----|-----|------|
| 500 | 505 | 510 | |
| AGC CAA TTT GAT CGC CTG TTT AAT AGC CCA TTA CTG AAC GGA CAA TAT | | | 1584 |
| Ser Gln Phe Asp Arg Leu Phe Asn Thr Pro Leu Leu Asn Gly Gln Tyr | | | |
| 515 | 520 | 525 | |
| TTT TCT ACC GGC GAT GAG GAG ATT GAT TTA AAT TCA GGT AGC ACC GGC | | | 1632 |
| Phe Ser Thr Gly Asp Glu Glu Ile Asp Leu Asn Ser Gly Ser Thr Gly | | | |
| 530 | 535 | 540 | |
| GAT TGG CGA AAA ACC ATA CTT AAG CGT GCA TTT AAT ATT GAT GAT GTC | | | 1680 |
| Asp Trp Arg Lys Thr Ile Leu Lys Arg Ala Phe Asn Ile Asp Asp Val | | | |
| 545 | 550 | 555 | 560 |
| TCG CTC TTC CGC CTG CTT AAA ATT ACC GAC CAT GAT AAT AAA GAT GGA | | | 1728 |
| Ser Leu Phe Arg Leu Leu Lys Ile Thr Asp His Asp Asn Lys Asp Gly | | | |
| 565 | 570 | 575 | |
| AAA ATT AAA AAT AAC CTA AAG AAT CTT TCC AAT TTA TAT ATT GGA AAA | | | 1776 |
| Lys Ile Lys Asn Asn Leu Lys Asn Leu Ser Asn Leu Tyr Ile Gly Lys | | | |
| 580 | 585 | 590 | |
| TTA CTG GCA GAT ATT CAT CAA TTA ACC ATT GAT GAA CTG GAT TTA TTA | | | 1824 |
| Leu Leu Ala Asp Ile His Gln Leu Thr Ile Asp Glu Leu Asp Leu Leu | | | |
| 595 | 600 | 605 | |
| CTG ATT GCC GTA GGT GAA GGA AAA ACT AAT TTA TCC GCT ATC AGT GAT | | | 1872 |
| Leu Ile Ala Val Gly Glu Gly Lys Thr Asn Leu Ser Ala Ile Ser Asp | | | |
| 610 | 615 | 620 | |
| AAG CAA TTG GCT ACC CTG ATC AGA AAA CTC AAT ACT ATT ACC AGC TGG | | | 1920 |
| Lys Gln Leu Ala Thr Leu Ile Arg Lys Leu Asn Thr Ile Thr Ser Trp | | | |
| 625 | 630 | 635 | 640 |
| CTA CAT ACA CAG AAG TGG AGT GTA TTC CAG CTA TTT ATC ATG ACC TCC | | | 1968 |
| Leu His Thr Gln Lys Trp Ser Val Phe Gln Leu Phe Ile Met Thr Ser | | | |
| 645 | 650 | 655 | |
| ACC AGC TAT AAC AAA ACG CTA ACG CCT GAA ATT AAG AAT TTG CTG GAT | | | 2016 |
| Thr Ser Tyr Asn Lys Thr Leu Thr Pro Glu Ile Lys Asn Leu Leu Asp | | | |
| 660 | 665 | 670 | |
| ACC GTC TAC CAC GGT TTA CAA GGT TTT GAT AAA GAC AAA GCA GAT TTG | | | 2064 |
| Thr Val Tyr His Gly Leu Gln Gly Phe Asp Lys Asp Lys Ala Asp Leu | | | |
| 675 | 680 | 685 | |
| CTA CAT GTC ATG GCG CCC TAT ATT GCG GCC ACC TTG CAA TTA TCA TCG | | | 2112 |
| Leu His Val Met Ala Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser | | | |
| 690 | 695 | 700 | |
| GAA AAT GTC GCC CAC TCG GTA CTC CTT TGG GCA GAT AAG TTA CAG CCC | | | 2160 |
| Glu Asn Val Ala His Ser Val Leu Leu Trp Ala Asp Lys Leu Gln Pro | | | |
| 705 | 710 | 715 | 720 |
| GGC GAC GGC GCA ATG ACA GCA GAA AAA TTC TGG GAC TGG TTG AAT ACT | | | 2208 |
| Gly Asp Gly Ala Met Thr Ala Glu Lys Phe Trp Asp Trp Leu Asn Thr | | | |
| 725 | 730 | 735 | |
| AAG TAT ACG CCG GGT TCA TCG GAA GCC GTA GAA ACG CAG GAA CAT ATC | | | 2256 |
| Lys Tyr Thr Pro Gly Ser Ser Glu Ala Val Glu Thr Gln Glu His Ile | | | |

| | | | |
|---|-----|-----|------|
| 740 | 745 | 750 | |
| GTT CAG TAT TGT CAG GCT CTG GCA CAA TTG GAA ATG GTT TAC CAT TCC | | | 2304 |
| Val Gln Tyr Cys Gln Ala Leu Ala Gln Leu Glu Met Val Tyr His Ser | | | |
| 755 | 760 | 765 | |
| ACC GGC ATC AAC GAA AAC GCC TTC CGT CTA TTT GTG ACA AAA CCA GAG | | | 2352 |
| Thr Gly Ile Asn Glu Asn Ala Phe Arg Leu Phe Val Thr Lys Pro Glu | | | |
| 770 | 775 | 780 | |
| ATG TTT GGC GCT GCA ACT GGA GCA GCG CCC GCG CAT GAT GCC CTT TCA | | | 2400 |
| Met Phe Gly Ala Ala Thr Gly Ala Ala Pro Ala His Asp Ala Leu Ser | | | |
| 785 | 790 | 795 | 800 |
| CTG ATT ATG CTG ACA CGT TTT GCG GAT TGG GTG AAC GCA CTA GGC GAA | | | 2448 |
| Leu Ile Met Leu Thr Arg Phe Ala Asp Trp Val Asn Ala Leu Gly Glu | | | |
| 805 | 810 | 815 | |
| AAA GCG TCC TCG GTG CTA GCG GCA TTT GAA GCT AAC TCG TTA ACG GCA | | | 2496 |
| Lys Ala Ser Ser Val Leu Ala Ala Phe Glu Ala Asn Ser Leu Thr Ala | | | |
| 820 | 825 | 830 | |
| GAA CAA CTG GCT GAT GCC ATG AAT CTT GAT GCT AAT TTG CTG TTG CAA | | | 2544 |
| Glu Gln Leu Ala Asp Ala Met Asn Leu Asp Ala Asn Leu Leu Leu Gln | | | |
| 835 | 840 | 845 | |
| GCC AGT ATT CAA GCA CAA AAT CAT CAA CAT CTT CCC CCA GTA ACT CCA | | | 2592 |
| Ala Ser Ile Gln Ala Gln Asn His Gln His Leu Pro Pro Val Thr Pro | | | |
| 850 | 855 | 860 | |
| GAA AAT GCG TTC TCC TGT TGG ACA TCT ATC AAT ACT ATC CTG CAA TGG | | | 2640 |
| Glu Asn Ala Phe Ser Cys Trp Thr Ser Ile Asn Thr Ile Leu Gln Trp | | | |
| 865 | 870 | 875 | 880 |
| GTT AAT GTC GCA CAA CAA TTG AAT GTC GCC CCA CAG GGC GTT TCC GCT | | | 2688 |
| Val Asn Val Ala Gln Gln Leu Asn Val Ala Pro Gln Gly Val Ser Ala | | | |
| 885 | 890 | 895 | |
| TTG GTC GGG CTG GAT TAT ATT CAA TCA ATG AAA GAG ACA CCG ACC TAT | | | 2736 |
| Leu Val Gly Leu Asp Tyr Ile Gln Ser Met Lys Glu Thr Pro Thr Tyr | | | |
| 900 | 905 | 910 | |
| GCC CAG TGG GAA AAC GCG GCA GGC GTA TTA ACC GCC GGG TTG AAT TCA | | | 2784 |
| Ala Gln Trp Glu Asn Ala Ala Gly Val Leu Thr Ala Gly Leu Asn Ser | | | |
| 915 | 920 | 925 | |
| CAA CAG GCT AAT ACA TTA CAC GCT TTT CTG GAT GAA TCT CGC AGT GCC | | | 2832 |
| Gln Gln Ala Asn Thr Leu His Ala Phe Leu Asp Glu Ser Arg Ser Ala | | | |
| 930 | 935 | 940 | |
| GCA TTA AGC ACC TAC TAT ATC CGT CAA GTC GCC AAG GCA GCG GCG GCT | | | 2880 |
| Ala Leu Ser Thr Tyr Tyr Ile Arg Gln Val Ala Lys Ala Ala Ala Ala | | | |
| 945 | 950 | 955 | 960 |
| ATT AAA AGC CGT GAT GAC TTG TAT CAA TAC TTA CTG ATT GAT AAT CAG | | | 2928 |
| Ile Lys Ser Arg Asp Asp Leu Tyr Gln Tyr Leu Leu Ile Asp Asn Gln | | | |
| 965 | 970 | 975 | |
| GTT TCT GCG GCA ATA AAA ACC ACC CGG ATC GCC GAA GCC ATT GCC AGT | | | 2976 |
| Val Ser Ala Ala Ile Lys Thr Thr Arg Ile Ala Glu Ala Ile Ala Ser | | | |

| | | | |
|---|------|------|------|
| 980 | 985 | 990 | |
| ATT CAA CTG TAC GTC AAC CGG GCA TTG GAA AAT GTG GAA GAA AAT GCC | | | 3024 |
| Ile Gln Leu Tyr Val Asn Arg Ala Leu Glu Asn Val Glu Glu Asn Ala | | | |
| 995 | 1000 | 1005 | |
| AAT TCG GGG GTT ATC AGC CGC CAA TTC TTT ATC GAC TGG GAC AAA TAC | | | 3072 |
| Asn Ser Gly Val Ile Ser Arg Gln Phe Phe Ile Asp Trp Asp Lys Tyr | | | |
| 1010 | 1015 | 1020 | |
| AAT AAA CGC TAC AGC ACT TGG GCG GGT GTT TCT CAA TTA GTT TAC TAC | | | 3120 |
| Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Gln Leu Val Tyr Tyr | | | |
| 1025 | 1030 | 1035 | 1040 |
| CCG GAA AAC TAT ATT GAT CCG ACC ATG CGT ATC GGA CAA ACC AAA ATG | | | 3168 |
| Pro Glu Asn Tyr Ile Asp Pro Thr Met Arg Ile Gly Gln Thr Lys Met | | | |
| 1045 | 1050 | 1055 | |
| ATG GAC GCA TTA CTG CAA TCC GTC AGC CAA AGC CAA TTA AAC GCC GAT | | | 3216 |
| Met Asp Ala Leu Leu Gln Ser Val Ser Gln Ser Gln Leu Asn Ala Asp | | | |
| 1060 | 1065 | 1070 | |
| ACC GTC GAA GAT GCC TTT ATG TCT TAT CTG ACA TCG TTT GAA CAA GTG | | | 3264 |
| Thr Val Glu Asp Ala Phe Met Ser Tyr Leu Thr Ser Phe Glu Gln Val | | | |
| 1075 | 1080 | 1085 | |
| GCT AAT CTT AAA GTT ATT AGC GCA TAT CAC GAT AAT ATT AAT AAC GAT | | | 3312 |
| Ala Asn Leu Lys Val Ile Ser Ala Tyr His Asp Asn Ile Asn Asn Asp | | | |
| 1090 | 1095 | 1100 | |
| CAA GGG CTG ACC TAT TTT ATC GGA CTC AGT GAA ACT GAT GCC GGT GAA | | | 3360 |
| Gln Gly Leu Thr Tyr Phe Ile Gly Leu Ser Glu Thr Asp Ala Gly Glu | | | |
| 1105 | 1110 | 1115 | 1120 |
| TAT TAT TGG CGC AGT GTC GAT CAC AGT AAA TTC AAC GAC GGT AAA TTC | | | 3408 |
| Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Phe Asn Asp Gly Lys Phe | | | |
| 1125 | 1130 | 1135 | |
| GCG GCT AAT GCC TGG AGT GAA TGG CAT AAA ATT GAT TGT CCA ATT AAC | | | 3456 |
| Ala Ala Asn Ala Trp Ser Glu Trp His Lys Ile Asp Cys Pro Ile Asn | | | |
| 1140 | 1145 | 1150 | |
| CCT TAT AAA AGC ACT ATC CGT CCA GTG ATA TAT AAA TCC CGC CTG TAT | | | 3504 |
| Pro Tyr Lys Ser Thr Ile Arg Pro Val Ile Tyr Lys Ser Arg Leu Tyr | | | |
| 1155 | 1160 | 1165 | |
| CTG CTC TGG TTG GAA CAA AAG GAG ATC ACC AAA CAG ACA GGA AAT AGT | | | 3552 |
| Leu Leu Trp Leu Glu Gln Lys Glu Ile Thr Lys Gln Thr Gly Asn Ser | | | |
| 1170 | 1175 | 1180 | |
| AAA GAT GGC TAT CAA ACT GAA ACG GAT TAT CGT TAT GAA CTA AAA TTG | | | 3600 |
| Lys Asp Gly Tyr Gln Thr Glu Thr Asp Tyr Arg Tyr Glu Leu Lys Leu | | | |
| 1185 | 1190 | 1195 | 1200 |
| GCG CAT ATC CGC TAT GAT GGC ACT TGG AAT ACG CCA ATC ACC TTT GAT | | | 3648 |
| Ala His Ile Arg Tyr Asp Gly Thr Trp Asn Thr Pro Ile Thr Phe Asp | | | |
| 1205 | 1210 | 1215 | |
| GTC AAT AAA AAA ATA TCC GAG CTA AAA CTG GAA AAA AAT AGA GCG CCC | | | 3696 |
| Val Asn Lys Lys Ile Ser Glu Leu Lys Leu Glu Lys Asn Arg Ala Pro | | | |

| | | | |
|---|------|------|------|
| 1220 | 1225 | 1230 | |
| GGA CTC TAT TGT GCC GGT TAT CAA GGT GAA GAT ACG TTG CTG GTG ATG | | | 3744 |
| Gly Leu Tyr Cys Ala Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met | | | |
| 1235 | 1240 | 1245 | |
| TTT TAT AAC CAA CAA GAC ACA CTA GAT AGT TAT AAA AAC GCT TCA ATG | | | 3792 |
| Phe Tyr Asn Gln Gln Asp Thr Leu Asp Ser Tyr Lys Asn Ala Ser Met | | | |
| 1250 | 1255 | 1260 | |
| CAA GGA CTA TAT ATC TTT GCT GAT ATG GCA TCC AAA GAT ATG ACC CCA | | | 3840 |
| Gln Gly Leu Tyr Ile Phe Ala Asp Met Ala Ser Lys Asp Met Thr Pro | | | |
| 1265 | 1270 | 1275 | 1280 |
| GAA CAG AGC AAT GTT TAT CGG GAT AAT AGC TAT CAA CAA TTT GAT ACC | | | 3888 |
| Glu Gln Ser Asn Val Tyr Arg Asp Asn Ser Tyr Gln Gln Phe Asp Thr | | | |
| 1285 | 1290 | 1295 | |
| AAT AAT GTC AGA AGA GTG AAT AAC CGC TAT GCA GAG GAT TAT GAG ATT | | | 3936 |
| Asn Asn Val Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr Glu Ile | | | |
| 1300 | 1305 | 1310 | |
| CCT TCC TCG GTA AGT AGC CGT AAA GAC TAT GGT TGG GGA GAT TAT TAC | | | 3984 |
| Pro Ser Ser Val Ser Ser Arg Lys Asp Tyr Gly Trp Gly Asp Tyr Tyr | | | |
| 1315 | 1320 | 1325 | |
| CTC AGC ATG GTA TAT AAC GGA GAT ATT CCA ACT ATC AAT TAC AAA GCC | | | 4032 |
| Leu Ser Met Val Tyr Asn Gly Asp Ile Pro Thr Ile Asn Tyr Lys Ala | | | |
| 1330 | 1335 | 1340 | |
| GCA TCA AGT GAT TTA AAA ATC TAT ATC TCA CCA AAA TTA AGA ATT ATT | | | 4080 |
| Ala Ser Ser Asp Leu Lys Ile Tyr Ile Ser Pro Lys Leu Arg Ile Ile | | | |
| 1345 | 1350 | 1355 | 1360 |
| CAT AAT GGA TAT GAA GGA CAG AAG CGC AAT CAA TGC AAT CTG ATG AAT | | | 4128 |
| His Asn Gly Tyr Glu Gly Gln Lys Arg Asn Gln Cys Asn Leu Met Asn | | | |
| 1365 | 1370 | 1375 | |
| AAA TAT GGC AAA CTA GGT GAT AAA TTT ATT GTT TAT ACT AGC TTG GGG | | | 4176 |
| Lys Tyr Gly Lys Leu Gly Asp Lys Phe Ile Val Tyr Thr Ser Leu Gly | | | |
| 1380 | 1385 | 1390 | |
| GTC AAT CCA AAT AAC TCG TCA AAT AAG CTC ATG TTT TAC CCC GTC TAT | | | 4224 |
| Val Asn Pro Asn Asn Ser Ser Asn Lys Leu Met Phe Tyr Pro Val Tyr | | | |
| 1395 | 1400 | 1405 | |
| CAA TAT AGC GGA AAC ACC AGT GGA CTC AAT CAA GGG AGA CTA CTA TTC | | | 4272 |
| Gln Tyr Ser Gly Asn Thr Ser Gly Leu Asn Gln Gly Arg Leu Leu Phe | | | |
| 1410 | 1415 | 1420 | |
| CAC CGT GAC ACC ACT TAT CCA TCT AAA GTA GAA GCT TGG ATT CCT GGA | | | 4320 |
| His Arg Asp Thr Thr Tyr Pro Ser Lys Val Glu Ala Trp Ile Pro Gly | | | |
| 1425 | 1430 | 1435 | 1440 |
| GCA AAA CGT TCT CTA ACC AAC CAA AAT GCC GCC ATT GGT GAT GAT TAT | | | 4368 |
| Ala Lys Arg Ser Leu Thr Asn Gln Asn Ala Ala Ile Gly Asp Asp Tyr | | | |
| 1445 | 1450 | 1455 | |
| GCT ACA GAC TCT CTG AAT AAA CCG GAT GAT CTT AAG CAA TAT ATC TTT | | | 4416 |
| Ala Thr Asp Ser Leu Asn Lys Pro Asp Asp Leu Lys Gln Tyr Ile Phe | | | |

1460 1465 1470
 ATG ACT GAC AGT AAA GGG ACT GCT ACT GAT GTC TCA GGC CCA GTA GAG 4464
 Met Thr Asp Ser Lys Gly Thr Ala Thr Asp Val Ser Gly Pro Val Glu
 1475 1480 1485
 ATT AAT ACT GCA ATT TCT CCA GCA AAA GTT CAG ATA ATA GTC AAA GCG 4512
 Ile Asn Thr Ala Ile Ser Pro Ala Lys Val Gln Ile Ile Val Lys Ala
 1490 1495 1500
 GGT GGC AAG GAG CAA ACT TTT ACC GCA GAT AAA GAT GTC TCC ATT CAG 4560
 Gly Gly Lys Glu Gln Thr Phe Thr Ala Asp Lys Asp Val Ser Ile Gln
 1505 1510 1515 1520
 CCA TCA CCT AGC TTT GAT GAA ATG AAT TAT CAA TTT AAT GCC CTT GAA 4608
 Pro Ser Pro Ser Phe Asp Glu Met Asn Tyr Gln Phe Asn Ala Leu Glu
 1525 1530 1535
 ATA GAC GGT TCT GGT CTG AAT TTT ATT AAC AAC TCA GCC AGT ATT GAT 4656
 Ile Asp Gly Ser Gly Leu Asn Phe Ile Asn Asn Ser Ala Ser Ile Asp
 1540 1545 1550
 GTT ACT TTT ACC GCA TTT GCG GAG GAT GGC CGC AAA CTG GGT TAT GAA 4704
 Val Thr Phe Thr Ala Phe Ala Glu Asp Gly Arg Lys Leu Gly Tyr Glu
 1555 1560 1565
 AGT TTC AGT ATT CCT GTT ACC CTC AAG GTA AGT ACC GAT AAT GCC CTG 4752
 Ser Phe Ser Ile Pro Val Thr Leu Lys Val Ser Thr Asp Asn Ala Leu
 1570 1575 1580
 ACC CTG CAC CAT AAT GAA AAT GGT GCG CAA TAT ATG CAA TGG CAA TCC 4800
 Thr Leu His His Asn Glu Asn Gly Ala Gln Tyr Met Gln Trp Gln Ser
 1585 1590 1595 1600
 TAT CGT ACC CGC CTG AAT ACT CTA TTT GCC CGC CAG TTG GTT GCA CGC 4848
 Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Arg Gln Leu Val Ala Arg
 1605 1610 1615
 GCC ACC ACC GGA ATC GAT ACA ATT CTG AGT ATG GAA ACT CAG AAT ATT 4896
 Ala Thr Thr Gly Ile Asp Thr Ile Leu Ser Met Glu Thr Gln Asn Ile
 1620 1625 1630
 CAG GAA CCG CAG TTA GGC AAA GGT TTC TAT GCT ACG TTC GTG ATA CCT 4944
 Gln Glu Pro Gln Leu Gly Lys Gly Phe Tyr Ala Thr Phe Val Ile Pro
 1635 1640 1645
 CCC TAT AAC CTA TCA ACT CAT GGT GAT GAA CGT TGG TTT AAG CTT TAT 4992
 Pro Tyr Asn Leu Ser Thr His Gly Asp Glu Arg Trp Phe Lys Leu Tyr
 1650 1655 1660
 ATC AAA CAT GTT GTT GAT AAT AAT TCA CAT ATT ATC TAT TCA GGC CAG 5040
 Ile Lys His Val Val Asp Asn Asn Ser His Ile Ile Tyr Ser Gly Gln
 1665 1670 1675 1680
 CTA ACA GAT ACA AAT ATA AAC ATC ACA TTA TTT ATT CCT CTT GAT GAT 5088
 Leu Thr Asp Thr Asn Ile Asn Ile Thr Leu Phe Ile Pro Leu Asp Asp
 1685 1690 1695
 GTC CCA TTG AAT CAA GAT TAT CAC GCC AAG GTT TAT ATG ACC TTC AAG 5136
 Val Pro Leu Asn Gln Asp Tyr His Ala Lys Val Tyr Met Thr Phe Lys

| | | | |
|---|------|------|------|
| 1700 | 1705 | 1710 | |
| AAA TCA CCA TCA GAT GGT ACC TGG TGG GGC CCT CAC TTT GTT AGA GAT | | | 5184 |
| Lys Ser Pro Ser Asp Gly Thr Trp Trp Gly Pro His Phe Val Arg Asp | | | |
| 1715 | 1720 | 1725 | |
| GAT AAA GGA ATA GTA ACA ATA AAC CCT AAA TCC ATT TTG ACC CAT TTT | | | 5232 |
| Asp Lys Gly Ile Val Thr Ile Asn Pro Lys Ser Ile Leu Thr His Phe | | | |
| 1730 | 1735 | 1740 | |
| GAG AGC GTC AAT GTC CTG AAT AAT ATT AGT AGC GAA CCA ATG GAT TTC | | | 5280 |
| Glu Ser Val Asn Val Leu Asn Asn Ile Ser Ser Glu Pro Met Asp Phe | | | |
| 1745 | 1750 | 1755 | 1760 |
| AGC GGC GCT AAC AGC CTC TAT TTC TGG GAA CTG TTC TAC TAT ACC CCG | | | 5328 |
| Ser Gly Ala Asn Ser Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro | | | |
| 1765 | 1770 | 1775 | |
| ATG CTG GTT GCT CAA CGT TTG CTG CAT GAA CAG AAC TTC GAT GAA GCC | | | 5376 |
| Met Leu Val Ala Gln Arg Leu Leu His Glu Gln Asn Phe Asp Glu Ala | | | |
| 1780 | 1785 | 1790 | |
| AAC CGT TGG CTG AAA TAT GTC TGG AGT CCA TCC GGT TAT ATT GTC CAC | | | 5424 |
| Asn Arg Trp Leu Lys Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val His | | | |
| 1795 | 1800 | 1805 | |
| GGC CAG ATT CAG AAC TAC CAG TGG AAC GTC CGC CCG TTA CTG GAA GAC | | | 5472 |
| Gly Gln Ile Gln Asn Tyr Gln Trp Asn Val Arg Pro Leu Leu Glu Asp | | | |
| 1810 | 1815 | 1820 | |
| ACC AGT TGG AAC AGT GAT CCT TTG GAT TCC GTC GAT CCT GAC GCG GTA | | | 5520 |
| Thr Ser Trp Asn Ser Asp Pro Leu Asp Ser Val Asp Pro Asp Ala Val | | | |
| 1825 | 1830 | 1835 | 1840 |
| GCA CAG CAC GAT CCA ATG CAC TAC AAA GTT TCA ACT TTT ATG CGT ACC | | | 5568 |
| Ala Gln His Asp Pro Met His Tyr Lys Val Ser Thr Phe Met Arg Thr | | | |
| 1845 | 1850 | 1855 | |
| TTG GAT CTA TTG ATA GCA CGC GGC GAC CAT GCT TAT CGC CAA CTG GAA | | | 5616 |
| Leu Asp Leu Leu Ile Ala Arg Gly Asp His Ala Tyr Arg Gln Leu Glu | | | |
| 1860 | 1865 | 1870 | |
| CGA GAT ACA CTC AAC GAA GCG AAG ATG TGG TAT ATG CAA GCG CTG CAT | | | 5664 |
| Arg Asp Thr Leu Asn Glu Ala Lys Met Trp Tyr Met Gln Ala Leu His | | | |
| 1875 | 1880 | 1885 | |
| CTA TTA GGT GAC AAA CCT TAT CTA CCG CTG AGT ACG ACA TGG AGT GAT | | | 5712 |
| Leu Leu Gly Asp Lys Pro Tyr Leu Pro Leu Ser Thr Thr Trp Ser Asp | | | |
| 1890 | 1895 | 1900 | |
| CCA CGA CTA GAC AGA GCC GCG GAT ATC ACT ACC CAA AAT GCT CAC GAC | | | 5760 |
| Pro Arg Leu Asp Arg Ala Ala Asp Ile Thr Thr Gln Asn Ala His Asp | | | |
| 1905 | 1910 | 1915 | 1920 |
| AGC GCA ATA GTC GCT CTG CGG CAG AAT ATA CCT ACA CCG GCA CCT TTA | | | 5808 |
| Ser Ala Ile Val Ala Leu Arg Gln Asn Ile Pro Thr Pro Ala Pro Leu | | | |
| 1925 | 1930 | 1935 | |
| TCA TTG CGC AGC GCT AAT ACC CTG ACT GAT CTC TTC CTG CCG CAA ATC | | | 5856 |
| Ser Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile | | | |

| | | | |
|---|------|------|------|
| 1940 | 1945 | 1950 | |
| AAT GAA GTG ATG ATG AAT TAC TGG CAG ACA TTA GCT CAG AGA GTA TAC | | | 5904 |
| Asn Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr | | | |
| 1955 | 1960 | 1965 | |
| AAT CTG CGT CAT AAC CTC TCT ATC GAC GGC CAG CCG TTA TAT CTG CCA | | | 5952 |
| Asn Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Tyr Leu Pro | | | |
| 1970 | 1975 | 1980 | |
| ATC TAT GCC ACA CCG GCC GAT CCG AAA GCG TTA CTC AGC GCC GCC GTT | | | 6000 |
| Ile Tyr Ala Thr Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val | | | |
| 1985 | 1990 | 1995 | 2000 |
| GCC ACT TCT CAA GGT GGA GGC AAG CTA CCG GAA TCA TTT ATG TCC CTG | | | 6048 |
| Ala Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu | | | |
| 2005 | 2010 | 2015 | |
| TGG CGT TTC CCG CAC ATG CTG GAA AAT GCG CGC GGC ATG GTT AGC CAG | | | 6096 |
| Trp Arg Phe Pro His Met Leu Glu Asn Ala Arg Gly Met Val Ser Gln | | | |
| 2020 | 2025 | 2030 | |
| CTC ACC CAG TTC GGC TCC ACG TTA CAA AAT ATT ATC GAA CGT CAG GAC | | | 6144 |
| Leu Thr Gln Phe Gly Ser Thr Leu Gln Asn Ile Ile Glu Arg Gln Asp | | | |
| 2035 | 2040 | 2045 | |
| GCG GAA GCG CTC AAT GCG TTA TTA CAA AAT CAG GCC GCC GAG CTG ATA | | | 6192 |
| Ala Glu Ala Leu Asn Ala Leu Leu Gln Asn Gln Ala Ala Glu Leu Ile | | | |
| 2050 | 2055 | 2060 | |
| TTG ACT AAC CTG AGC ATT CAG GAC AAA ACC ATT GAA GAA TTG GAT GCC | | | 6240 |
| Leu Thr Asn Leu Ser Ile Gln Asp Lys Thr Ile Glu Glu Leu Asp Ala | | | |
| 2065 | 2070 | 2075 | 2080 |
| GAG AAA ACG GTG TTG GAA AAA TCC AAA GCG GGA GCA CAA TCG CGC TTT | | | 6288 |
| Glu Lys Thr Val Leu Glu Lys Ser Lys Ala Gly Ala Gln Ser Arg Phe | | | |
| 2085 | 2090 | 2095 | |
| GAT AGC TAC GGC AAA CTG TAC GAT GAG AAT ATC AAC GCC GGT GAA AAC | | | 6336 |
| Asp Ser Tyr Gly Lys Leu Tyr Asp Glu Asn Ile Asn Ala Gly Glu Asn | | | |
| 2100 | 2105 | 2110 | |
| CAA GCC ATG ACG CTA CGA GCG TCC GCC GCC GGG CTT ACC ACG GCA GTT | | | 6384 |
| Gln Ala Met Thr Leu Arg Ala Ser Ala Ala Gly Leu Thr Thr Ala Val | | | |
| 2115 | 2120 | 2125 | |
| CAG GCA TCC CGT CTG GCC GGT GCG GCG GCT GAT CTG GTG CCT AAC ATC | | | 6432 |
| Gln Ala Ser Arg Leu Ala Gly Ala Ala Ala Asp Leu Val Pro Asn Ile | | | |
| 2130 | 2135 | 2140 | |
| TTC GGC TTT GCC GGT GGC GGC AGC CGT TGG GGG GCT ATC GCT GAG GCG | | | 6480 |
| Phe Gly Phe Ala Gly Gly Gly Ser Arg Trp Gly Ala Ile Ala Glu Ala | | | |
| 2145 | 2150 | 2155 | 2160 |
| ACA GGT TAT GTG ATG GAA TTC TCC GCG AAT GTT ATG AAC ACC GAA GCG | | | 6528 |
| Thr Gly Tyr Val Met Glu Phe Ser Ala Asn Val Met Asn Thr Glu Ala | | | |
| 2165 | 2170 | 2175 | |
| GAT AAA ATT AGC CAA TCT GAA ACC TAC CGT CGT CGC CGT CAG GAG TGG | | | 6576 |
| Asp Lys Ile Ser Gln Ser Glu Thr Tyr Arg Arg Arg Arg Gln Glu Trp | | | |

| | | | |
|---|------|------|------|
| 2180 | 2185 | 2190 | |
| GAG ATC CAG CGG AAT AAT GCC GAA GCG GAA TTG AAG CAA ATC GAT GCT | | | 6624 |
| Glu Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala | | | |
| 2195 | 2200 | 2205 | |
| CAG CTC AAA TCA CTC GCT GTA CGC CGC GAA GCC GCC GTA TTG CAG AAA | | | 6672 |
| Gln Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys | | | |
| 2210 | 2215 | 2220 | |
| ACC AGT CTG AAA ACC CAA CAA GAA CAG ACC CAA TCT CAA TTG GCC TTC | | | 6720 |
| Thr Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe | | | |
| 2225 | 2230 | 2235 | 2240 |
| CTG CAA CGT AAG TTC AGC AAT CAG GCG TTA TAC AAC TGG CTG CGT GGT | | | 6768 |
| Leu Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly | | | |
| 2245 | 2250 | 2255 | |
| CGA CTG GCG GCG ATT TAC TTC CAG TTC TAC GAT TTG GCC GTC GCG CGT | | | 6816 |
| Arg Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg | | | |
| 2260 | 2265 | 2270 | |
| TGC CTG ATG GCA GAA CAA GCT TAC CGT TGG GAA CTC AAT GAT GAC TCT | | | 6864 |
| Cys Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser | | | |
| 2275 | 2280 | 2285 | |
| GCC CGC TTC ATT AAA CCG GGC GCC TGG CAG GGA ACC TAT GCC GGT CTG | | | 6912 |
| Ala Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu | | | |
| 2290 | 2295 | 2300 | |
| CTT GCA GGT GAA ACC TTG ATG CTG AGT CTG GCA CAA ATG GAA GAC GCT | | | 6960 |
| Leu Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala | | | |
| 2305 | 2310 | 2315 | 2320 |
| CAT CTG AAA CGC GAT AAA CGC GCA TTA GAG GTT GAA CGC ACA GTA TCG | | | 7008 |
| His Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser | | | |
| 2325 | 2330 | 2335 | |
| CTG GCC GAA GTT TAT GCA GGA TTA CCA AAA GAT AAC GGT CCA TTT TCC | | | 7056 |
| Leu Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser | | | |
| 2340 | 2345 | 2350 | |
| CTG GCT CAG GAA ATT GAC AAG CTG GTG AGT CAA GGT TCA GGC AGT GCC | | | 7104 |
| Leu Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala | | | |
| 2355 | 2360 | 2365 | |
| GGC AGT GGT AAT AAT AAT TTG GCG TTC GGC GCC GGC ACG GAC ACT AAA | | | 7152 |
| Gly Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys | | | |
| 2370 | 2375 | 2380 | |
| ACC TCT TTG CAG GCA TCA GTT TCA TTC GCT GAT TTG AAA ATT CGT GAA | | | 7200 |
| Thr Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu | | | |
| 2385 | 2390 | 2395 | 2400 |
| GAT TAC CCG GCA TCG CTT GGC AAA ATT CGA CGT ATC AAA CAG ATC AGC | | | 7248 |
| Asp Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser | | | |
| 2405 | 2410 | 2415 | |
| GTC ACT TTG CCC GCG CTA CTG GGA CCG TAT CAG GAT GTA CAG GCA ATA | | | 7296 |
| Val Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile | | | |

2420 2425 2430
 TTG TCT TAC GGC GAT AAA GCC GGA TTA GCT AAC GGC TGT GAA GCG CTG 7344
 Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu
 2435 2440 2445
 GCA GTT TCT CAC GGT ATG AAT GAC AGC GGC CAA TTC CAG CTC GAT TTC 7392
 Ala Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe
 2450 2455 2460
 AAC GAT GGC AAA TTC CTG CCA TTC GAA GGC ATC GCC ATT GAT CAA GGC 7440
 Asn Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly
 2465 2470 2475 2480
 ACG CTG ACA CTG AGC TTC CCA AAT GCA TCT ATG CCG GAG AAA GGT AAA 7488
 Thr Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys
 2485 2490 2495
 CAA GCC ACT ATG TTA AAA ACC CTG AAC GAT ATC ATT TTG CAT ATT CGC 7536
 Gln Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg
 2500 2505 2510
 TAC ACC ATT AAA TAA 7551
 Tyr Thr Ile Lys v?
 2516

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2516 amino acids
- (B) TYPE: amino acids
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47 (TcdA):

| Features | From | To | Description |
|----------|------|------|--------------------------------------|
| Peptide | 1 | 2516 | TcdA proteins |
| Peptide | 89 | 1937 | TcdAii peptide |
| Fragment | | 89 | 100 TcdAii N-terminus (SEQ ID NO:13) |
| Fragment | | 284 | 299 (SEQ ID NO:38) |
| Fragment | | 554 | 563 (SEQ ID NO:17) |
| Fragment | | 1080 | 1092 (SEQ ID NO:23; 12/13) |
| Fragment | | 1385 | 1400 (SEQ ID NO:18) |
| Fragment | | 1478 | 1497 (SEQ ID NO:39) |
| Fragment | | 1620 | 1642 (SEQ ID NO:21; 19/23) |
| Fragment | | 1938 | 1948 (SEQ ID NO:41) |
| Peptide | 1938 | 2516 | TcdAiii peptide |
| Fragment | | 2327 | 2345 (SEQ ID NO:42) |
| Fragment | | 2398 | 2408 (SEQ ID NO:43) |

Met Asn Glu Ser Val Lys Glu Ile Pro Asp Val Leu Lys Ser Gln Cys

1 5 10 15
 Gly Phe Asn Cys Leu Thr Asp Ile Ser His Ser Ser Phe Asn Glu Phe
 20 25 30
 Arg Gln Gln Val Ser Glu His Leu Ser Trp Ser Glu Thr His Asp Leu
 35 40 45
 Tyr His Asp Ala Gln Gln Ala Gln Lys Asp Asn Arg Leu Tyr Glu Ala
 50 55 60
 Arg Ile Leu Lys Arg Ala Asn Pro Gln Leu Gln Asn Ala Val His Leu
 65 70 75 80
 Ala Ile Leu Ala Pro Asn Ala Glu Leu Ile Gly Tyr Asn Asn Gln Phe
 85 90 95
 Ser Gly Arg Ala Ser Gln Tyr Val Ala Pro Gly Thr Val Ser Ser Met
 100 105 110
 Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Arg Asn
 115 120 125
 Leu His Ala Ser Asp Ser Val Tyr Tyr Leu Asp Thr Arg Arg Pro Asp
 130 135 140
 Leu Lys Ser Met Ala Leu Ser Gln Gln Asn Met Asp Ile Glu Leu Ser
 145 150 155 160
 Thr Leu Ser Leu Ser Asn Glu Leu Leu Leu Glu Ser Ile Lys Thr Glu
 165 170 175
 Ser Lys Leu Glu Asn Tyr Thr Lys Val Met Glu Met Leu Ser Thr Phe
 180 185 190
 Arg Pro Ser Gly Ala Thr Pro Tyr His Asp Ala Tyr Glu Asn Val Arg
 195 200 205
 Glu Val Ile Gln Leu Gln Asp Pro Gly Leu Glu Gln Leu Asn Ala Ser
 210 215 220
 Pro Ala Ile Ala Gly Leu Met His Gln Ala Ser Leu Leu Gly Ile Asn
 225 230 235 240
 Ala Ser Ile Ser Pro Glu Leu Phe Asn Ile Leu Thr Glu Glu Ile Thr
 245 250 255
 Glu Gly Asn Ala Glu Glu Leu Tyr Lys Lys Asn Phe Gly Asn Ile Glu
 260 265 270
 Pro Ala Ser Leu Ala Met Pro Glu Tyr Leu Lys Arg Tyr Tyr Asn Leu
 275 280 285
 Ser Asp Glu Glu Leu Ser Gln Phe Ile Gly Lys Ala Ser Asn Phe Gly
 290 295 300
 Gln Gln Glu Tyr Ser Asn Asn Gln Leu Ile Thr Pro Val Val Asn Ser
 305 310 315 320
 Ser Asp Gly Thr Val Lys Val Tyr Arg Ile Thr Arg Glu Tyr Thr Thr
 325 330 335
 Asn Ala Tyr Gln Met Asp Val Glu Leu Phe Pro Phe Gly Gly Glu Asn
 340 345 350
 Tyr Arg Leu Asp Tyr Lys Phe Lys Asn Phe Tyr Asn Ala Ser Tyr Leu
 355 360 365

Ser Ile Lys Leu Asn Asp Lys Arg Glu Leu Val Arg Thr Glu Gly Ala
 370 375 380
 Pro Gln Val Asn Ile Glu Tyr Ser Ala Asn Ile Thr Leu Asn Thr Ala
 385 390 395 400
 Asp Ile Ser Gln Pro Phe Glu Ile Gly Leu Thr Arg Val Leu Pro Ser
 405 410 415
 Gly Ser Trp Ala Tyr Ala Ala Ala Lys Phe Thr Val Glu Glu Tyr Asn
 420 425 430
 Gln Tyr Ser Phe Leu Leu Lys Leu Asn Lys Ala Ile Arg Leu Ser Arg
 435 440 445
 Ala Thr Glu Leu Ser Pro Thr Ile Leu Glu Gly Ile Val Arg Ser Val
 450 455 460
 Asn Leu Gln Leu Asp Ile Asn Thr Asp Val Leu Gly Lys Val Phe Leu
 465 470 475 480
 Thr Lys Tyr Tyr Met Gln Arg Tyr Ala Ile His Ala Glu Thr Ala Leu
 485 490 495
 Ile Leu Cys Asn Ala Pro Ile Ser Gln Arg Ser Tyr Asp Asn Gln Pro
 500 505 510
 Ser Gln Phe Asp Arg Leu Phe Asn Thr Pro Leu Leu Asn Gly Gln Tyr
 515 520 525
 Phe Ser Thr Gly Asp Glu Glu Ile Asp Leu Asn Ser Gly Ser Thr Gly
 530 535 540
 Asp Trp Arg Lys Thr Ile Leu Lys Arg Ala Phe Asn Ile Asp Asp Val
 545 550 555 560
 Ser Leu Phe Arg Leu Leu Lys Ile Thr Asp His Asp Asn Lys Asp Gly
 565 570 575
 Lys Ile Lys Asn Asn Leu Lys Asn Leu Ser Asn Leu Tyr Ile Gly Lys
 580 585 590
 Leu Leu Ala Asp Ile His Gln Leu Thr Ile Asp Glu Leu Asp Leu Leu
 595 600 605
 Leu Ile Ala Val Gly Glu Gly Lys Thr Asn Leu Ser Ala Ile Ser Asp
 610 615 620
 Lys Gln Leu Ala Thr Leu Ile Arg Lys Leu Asn Thr Ile Thr Ser Trp
 625 630 635 640
 Leu His Thr Gln Lys Trp Ser Val Phe Gln Leu Phe Ile Met Thr Ser
 645 650 655
 Thr Ser Tyr Asn Lys Thr Leu Thr Pro Glu Ile Lys Asn Leu Leu Asp
 660 665 670
 Thr Val Tyr His Gly Leu Gln Gly Phe Asp Lys Asp Lys Ala Asp Leu
 675 680 685
 Leu His Val Met Ala Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser
 690 695 700
 Glu Asn Val Ala His Ser Val Leu Leu Trp Ala Asp Lys Leu Gln Pro
 705 710 715 720
 Gly Asp Gly Ala Met Thr Ala Glu Lys Phe Trp Asp Trp Leu Asn Thr

| | | | | | |
|---|------|------|------|------|------|
| | 725 | | 730 | | 735 |
| Lys Tyr Thr Pro Gly Ser Ser Glu Ala Val Glu Thr Gln Glu His Ile | | | | | |
| | 740 | | 745 | | 750 |
| Val Gln Tyr Cys Gln Ala Leu Ala Gln Leu Glu Met Val Tyr His Ser | | | | | |
| | 755 | | 760 | | 765 |
| Thr Gly Ile Asn Glu Asn Ala Phe Arg Leu Phe Val Thr Lys Pro Glu | | | | | |
| | 770 | | 775 | | 780 |
| Met Phe Gly Ala Ala Thr Gly Ala Ala Pro Ala His Asp Ala Leu Ser | | | | | |
| 785 | | 790 | | 795 | 800 |
| Leu Ile Met Leu Thr Arg Phe Ala Asp Trp Val Asn Ala Leu Gly Glu | | | | | |
| | 805 | | 810 | | 815 |
| Lys Ala Ser Ser Val Leu Ala Ala Phe Glu Ala Asn Ser Leu Thr Ala | | | | | |
| | 820 | | 825 | | 830 |
| Glu Gln Leu Ala Asp Ala Met Asn Leu Asp Ala Asn Leu Leu Leu Gln | | | | | |
| | 835 | | 840 | | 845 |
| Ala Ser Ile Gln Ala Gln Asn His Gln His Leu Pro Pro Val Thr Pro | | | | | |
| | 850 | | 855 | | 860 |
| Glu Asn Ala Phe Ser Cys Trp Thr Ser Ile Asn Thr Ile Leu Gln Trp | | | | | |
| 865 | | 870 | | 875 | 880 |
| Val Asn Val Ala Gln Gln Leu Asn Val Ala Pro Gln Gly Val Ser Ala | | | | | |
| | 885 | | 890 | | 895 |
| Leu Val Gly Leu Asp Tyr Ile Gln Ser Met Lys Glu Thr Pro Thr Tyr | | | | | |
| | 900 | | 905 | | 910 |
| Ala Gln Trp Glu Asn Ala Ala Gly Val Leu Thr Ala Gly Leu Asn Ser | | | | | |
| | 915 | | 920 | | 925 |
| Gln Gln Ala Asn Thr Leu His Ala Phe Leu Asp Glu Ser Arg Ser Ala | | | | | |
| | 930 | | 935 | | 940 |
| Ala Leu Ser Thr Tyr Tyr Ile Arg Gln Val Ala Lys Ala Ala Ala Ala | | | | | |
| 945 | | 950 | | 955 | 960 |
| Ile Lys Ser Arg Asp Asp Leu Tyr Gln Tyr Leu Leu Ile Asp Asn Gln | | | | | |
| | 965 | | 970 | | 975 |
| Val Ser Ala Ala Ile Lys Thr Thr Arg Ile Ala Glu Ala Ile Ala Ser | | | | | |
| | 980 | | 985 | | 990 |
| Ile Gln Leu Tyr Val Asn Arg Ala Leu Glu Asn Val Glu Glu Asn Ala | | | | | |
| | 995 | | 1000 | | 1005 |
| Asn Ser Gly Val Ile Ser Arg Gln Phe Phe Ile Asp Trp Asp Lys Tyr | | | | | |
| | 1010 | | 1015 | | 1020 |
| Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Gln Leu Val Tyr Tyr | | | | | |
| 1025 | | 1030 | | 1035 | 1040 |
| Pro Glu Asn Tyr Ile Asp Pro Thr Met Arg Ile Gly Gln Thr Lys Met | | | | | |
| | 1045 | | 1050 | | 1055 |
| Met Asp Ala Leu Leu Gln Ser Val Ser Gln Ser Gln Leu Asn Ala Asp | | | | | |
| | 1060 | | 1065 | | 1070 |
| Thr Val Glu Asp Ala Phe Met Ser Tyr Leu Thr Ser Phe Glu Gln Val | | | | | |
| | 1075 | | 1080 | | 1085 |

Ala Asn Leu Lys Val Ile Ser Ala Tyr His Asp Asn Ile Asn Asn Asp
 1090 1095 1100

Gln Gly Leu Thr Tyr Phe Ile Gly Leu Ser Glu Thr Asp Ala Gly Glu
 1105 1110 1115 1120

Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Phe Asn Asp Gly Lys Phe
 1125 1130 1135

Ala Ala Asn Ala Trp Ser Glu Trp His Lys Ile Asp Cys Pro Ile Asn
 1140 1145 1150

Pro Tyr Lys Ser Thr Ile Arg Pro Val Ile Tyr Lys Ser Arg Leu Tyr
 1155 1160 1165

Leu Leu Trp Leu Glu Gln Lys Glu Ile Thr Lys Gln Thr Gly Asn Ser
 1170 1175 1180

Lys Asp Gly Tyr Gln Thr Glu Thr Asp Tyr Arg Tyr Glu Leu Lys Leu
 1185 1190 1195 1200

Ala His Ile Arg Tyr Asp Gly Thr Trp Asn Thr Pro Ile Thr Phe Asp
 1205 1210 1215

Val Asn Lys Lys Ile Ser Glu Leu Lys Leu Glu Lys Asn Arg Ala Pro
 1220 1225 1230

Gly Leu Tyr Cys Ala Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met
 1235 1240 1245

Phe Tyr Asn Gln Gln Asp Thr Leu Asp Ser Tyr Lys Asn Ala Ser Met
 1250 1255 1260

Gln Gly Leu Tyr Ile Phe Ala Asp Met Ala Ser Lys Asp Met Thr Pro
 1265 1270 1275 1280

Glu Gln Ser Asn Val Tyr Arg Asp Asn Ser Tyr Gln Gln Phe Asp Thr
 1285 1290 1295

Asn Asn Val Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr Glu Ile
 1300 1305 1310

Pro Ser Ser Val Ser Ser Arg Lys Asp Tyr Gly Trp Gly Asp Tyr Tyr
 1315 1320 1325

Leu Ser Met Val Tyr Asn Gly Asp Ile Pro Thr Ile Asn Tyr Lys Ala
 1330 1335 1340

Ala Ser Ser Asp Leu Lys Ile Tyr Ile Ser Pro Lys Leu Arg Ile Ile
 1345 1350 1355 1360

His Asn Gly Tyr Glu Gly Gln Lys Arg Asn Gln Cys Asn Leu Met Asn
 1365 1370 1375

Lys Tyr Gly Lys Leu Gly Asp Lys Phe Ile Val Tyr Thr Ser Leu Gly
 1380 1385 1390

Val Asn Pro Asn Asn Ser Ser Asn Lys Leu Met Phe Tyr Pro Val Tyr
 1395 1400 1405

Gln Tyr Ser Gly Asn Thr Ser Gly Leu Asn Gln Gly Arg Leu Leu Phe
 1410 1415 1420

His Arg Asp Thr Thr Tyr Pro Ser Lys Val Glu Ala Trp Ile Pro Gly
 1425 1430 1435 1440

Ala Lys Arg Ser Leu Thr Asn Gln Asn Ala Ala Ile Gly Asp Asp Tyr

Gly Gln Ile Gln Asn Tyr Gln Trp Asn Val Arg Pro Leu Leu Glu Asp
 1810 1815 1820

Thr Ser Trp Asn Ser Asp Pro Leu Asp Ser Val Asp Pro Asp Ala Val
 1825 1830 1835 1840

Ala Gln His Asp Pro Met His Tyr Lys Val Ser Thr Phe Met Arg Thr
 1845 1850 1855

Leu Asp Leu Leu Ile Ala Arg Gly Asp His Ala Tyr Arg Gln Leu Glu
 1860 1865 1870

Arg Asp Thr Leu Asn Glu Ala Lys Met Trp Tyr Met Gln Ala Leu His
 1875 1880 1885

Leu Leu Gly Asp Lys Pro Tyr Leu Pro Leu Ser Thr Thr Trp Ser Asp
 1890 1895 1900

Pro Arg Leu Asp Arg Ala Ala Asp Ile Thr Thr Gln Asn Ala His Asp
 1905 1910 1915 1920

Ser Ala Ile Val Ala Leu Arg Gln Asn Ile Pro Thr Pro Ala Pro Leu
 1925 1930 1935

Ser Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile
 1940 1945 1950

Asn Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr
 1955 1960 1965

Asn Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Tyr Leu Pro
 1970 1975 1980

Ile Tyr Ala Thr Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val
 1985 1990 1995 2000

Ala Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu
 2005 2010 2015

Trp Arg Phe Pro His Met Leu Glu Asn Ala Arg Gly Met Val Ser Gln
 2020 2025 2030

Leu Thr Gln Phe Gly Ser Thr Leu Gln Asn Ile Ile Glu Arg Gln Asp
 2035 2040 2045

Ala Glu Ala Leu Asn Ala Leu Leu Gln Asn Gln Ala Ala Glu Leu Ile
 2050 2055 2060

Leu Thr Asn Leu Ser Ile Gln Asp Lys Thr Ile Glu Glu Leu Asp Ala
 2065 2070 2075 2080

Glu Lys Thr Val Leu Glu Lys Ser Lys Ala Gly Ala Gln Ser Arg Phe
 2085 2090 2095

Asp Ser Tyr Gly Lys Leu Tyr Asp Glu Asn Ile Asn Ala Gly Glu Asn
 2100 2105 2110

Gln Ala Met Thr Leu Arg Ala Ser Ala Ala Gly Leu Thr Thr Ala Val
 2115 2120 2125

Gln Ala Ser Arg Leu Ala Gly Ala Ala Ala Asp Leu Val Pro Asn Ile
 2130 2135 2140

Phe Gly Phe Ala Gly Gly Gly Ser Arg Trp Gly Ala Ile Ala Glu Ala
 2145 2150 2155 2160

Thr Gly Tyr Val Met Glu Phe Ser Ala Asn Val Met Asn Thr Glu Ala

| | | | |
|---|------|------|------|
| | 2165 | 2170 | 2175 |
| Asp Lys Ile Ser Gln Ser Glu Thr Tyr Arg Arg Arg Arg Gln Glu Trp | | | |
| | 2180 | 2185 | 2190 |
| Glu Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala | | | |
| | 2195 | 2200 | 2205 |
| Gln Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys | | | |
| | 2210 | 2215 | 2220 |
| Thr Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe | | | |
| 2225 | 2230 | 2235 | 2240 |
| Leu Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly | | | |
| | 2245 | 2250 | 2255 |
| Arg Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg | | | |
| | 2260 | 2265 | 2270 |
| Cys Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser | | | |
| | 2275 | 2280 | 2285 |
| Ala Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu | | | |
| | 2290 | 2295 | 2300 |
| Leu Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala | | | |
| 2305 | 2310 | 2315 | 2320 |
| His Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser | | | |
| | 2325 | 2330 | 2335 |
| Leu Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser | | | |
| | 2340 | 2345 | 2350 |
| Leu Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala | | | |
| | 2355 | 2360 | 2365 |
| Gly Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys | | | |
| | 2370 | 2375 | 2380 |
| Thr Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu | | | |
| 2385 | 2390 | 2395 | 2400 |
| Asp Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser | | | |
| | 2405 | 2410 | 2415 |
| Val Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile | | | |
| | 2420 | 2425 | 2430 |
| Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu | | | |
| | 2435 | 2440 | 2445 |
| Ala Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe | | | |
| | 2450 | 2455 | 2460 |
| Asn Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly | | | |
| 2465 | 2470 | 2475 | 2480 |
| Thr Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys | | | |
| | 2485 | 2490 | 2495 |
| Gln Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg | | | |
| | 2500 | 2505 | 2510 |
| Tyr Thr Ile Lys | | | |
| | 2516 | | |

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5547 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48 (tcdAii coding region):

```

CTG ATA GGC TAT AAC AAT CAA TTT AGC GGT AGA GCC AGT CAA TAT GTT 48
Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln Tyr Val
1           5           10          15
GCG CCG GGT ACC GTT TCT TCC ATG TTC TCC CCC GCC GCT TAT TTG ACT 96
Ala Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr Leu Thr
          20          25          30
GAA CTT TAT CGT GAA GCA CGC AAT TTA CAC GCA AGT GAC TCC GTT TAT 144
Glu Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser Val Tyr
          35          40          45
TAT CTG GAT ACC CGC CGC CCA GAT CTC AAA TCA ATG GCG CTC AGT CAG 192
Tyr Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu Ser Gln
          50          55          60
CAA AAT ATG GAT ATA GAA TTA TCC ACA CTC TCT TTG TCC AAT GAG CTG 240
Gln Asn Met Asp Ile Glu Leu Ser Thr Leu Ser Leu Ser Asn Glu Leu
65          70          75          80
TTA TTG GAA AGC ATT AAA ACT GAA TCT AAA CTG GAA AAC TAT ACT AAA 288
Leu Leu Glu Ser Ile Lys Thr Glu Ser Lys Leu Glu Asn Tyr Thr Lys
          85          90          95
GTG ATG GAA ATG CTC TCC ACT TTC CGT CCT TCC GGC GCA ACG CCT TAT 336
Val Met Glu Met Leu Ser Thr Phe Arg Pro Ser Gly Ala Thr Pro Tyr
          100         105         110
CAT GAT GCT TAT GAA AAT GTG CGT GAA GTT ATC CAG CTA CAA GAT CCT 384
His Asp Ala Tyr Glu Asn Val Arg Glu Val Ile Gln Leu Gln Asp Pro
          115         120         125
GGA CTT GAG CAA CTC AAT GCA TCA CCG GCA ATT GCC GGG TTG ATG CAT 432
Gly Leu Glu Gln Leu Asn Ala Ser Pro Ala Ile Ala Gly Leu Met His
          130         135         140
CAA GCC TCC CTA TTG GGT ATT AAC GCT TCA ATC TCG CCT GAG CTA TTT 480
Gln Ala Ser Leu Leu Gly Ile Asn Ala Ser Ile Ser Pro Glu Leu Phe
145         150         155         160
AAT ATT CTG ACG GAG GAG ATT ACC GAA GGT AAT GCT GAG GAA CTT TAT 528
Asn Ile Leu Thr Glu Glu Ile Thr Glu Gly Asn Ala Glu Glu Leu Tyr
          165         170         175
AAG AAA AAT TTT GGT AAT ATC GAA CCG GCC TCA TTG GCT ATG CCG GAA 576
Lys Lys Asn Phe Gly Asn Ile Glu Pro Ala Ser Leu Ala Met Pro Glu
          180         185         190

```

TAC CTT AAA CGT TAT TAT AAT TTA AGC GAT GAA GAA CTT AGT CAG TTT 624
Tyr Leu Lys Arg Tyr Tyr Asn Leu Ser Asp Glu Glu Leu Ser Gln Phe
195 200 205

ATT GGT AAA GCC AGC AAT TTT GGT CAA CAG GAA TAT AGT AAT AAC CAA 672
Ile Gly Lys Ala Ser Asn Phe Gly Gln Gln Glu Tyr Ser Asn Asn Gln
210 215 220

CTT ATT ACT CCG GTA GTC AAC AGC AGT GAT GGC ACG GTT AAG GTA TAT 720
Leu Ile Thr Pro Val Val Asn Ser Ser Asp Gly Thr Val Lys Val Tyr
225 230 235 240

CGG ATC ACC CGC GAA TAT ACA ACC AAT GCT TAT CAA ATG GAT GTG GAG 768
Arg Ile Thr Arg Glu Tyr Thr Thr Asn Ala Tyr Gln Met Asp Val Glu
245 250 255

CTA TTT CCC TTC GGT GGT GAG AAT TAT CGG TTA GAT TAT AAA TTC AAA 816
Leu Phe Pro Phe Gly Gly Glu Asn Tyr Arg Leu Asp Tyr Lys Phe Lys
260 265 270

AAT TTT TAT AAT GCC TCT TAT TTA TCC ATC AAG TTA AAT GAT AAA AGA 864
Asn Phe Tyr Asn Ala Ser Tyr Leu Ser Ile Lys Leu Asn Asp Lys Arg
275 280 285

GAA CTT GTT CGA ACT GAA GGC GCT CCT CAA GTC AAT ATA GAA TAC TCC 912
Glu Leu Val Arg Thr Glu Gly Ala Pro Gln Val Asn Ile Glu Tyr Ser
290 295 300

GCA AAT ATC ACA TTA AAT ACC GCT GAT ATC AGT CAA CCT TTT GAA ATT 960
Ala Asn Ile Thr Leu Asn Thr Ala Asp Ile Ser Gln Pro Phe Glu Ile
305 310 315 320

GGC CTG ACA CGA GTA CTT CCT TCC GGT TCT TGG GCA TAT GCC GCC GCA 1008
Gly Leu Thr Arg Val Leu Pro Ser Gly Ser Trp Ala Tyr Ala Ala Ala
325 330 335

AAA TTT ACC GTT GAA GAG TAT AAC CAA TAC TCT TTT CTG CTA AAA CTT 1056
Lys Phe Thr Val Glu Glu Tyr Asn Gln Tyr Ser Phe Leu Leu Lys Leu
340 345 350

AAC AAG GCT ATT CGT CTA TCA CGT GCG ACA GAA TTG TCA CCC ACG ATT 1104
Asn Lys Ala Ile Arg Leu Ser Arg Ala Thr Glu Leu Ser Pro Thr Ile
355 360 365

CTG GAA GGC ATT GTG CGC AGT GTT AAT CTA CAA CTG GAT ATC AAC ACA 1152
Leu Glu Gly Ile Val Arg Ser Val Asn Leu Gln Leu Asp Ile Asn Thr
370 375 380

GAC GTA TTA GGT AAA GTT TTT CTG ACT AAA TAT TAT ATG CAG CGT TAT 1200
Asp Val Leu Gly Lys Val Phe Leu Thr Lys Tyr Tyr Met Gln Arg Tyr
385 390 395 400

GCT ATT CAT GCT GAA ACT GCC CTG ATA CTA TGC AAC GCG CCT ATT TCA 1248
Ala Ile His Ala Glu Thr Ala Leu Ile Leu Cys Asn Ala Pro Ile Ser
405 410 415

CAA CGT TCA TAT GAT AAT CAA CCT AGC CAA TTT GAT CGC CTG TTT AAT 1296
Gln Arg Ser Tyr Asp Asn Gln Pro Ser Gln Phe Asp Arg Leu Phe Asn
420 425 430

ACG CCA TTA CTG AAC GGA CAA TAT TTT TCT ACC GGC GAT GAG GAG ATT 1344
 Thr Pro Leu Leu Asn Gly Gln Tyr Phe Ser Thr Gly Asp Glu Glu Ile
 435 440 445

GAT TTA AAT TCA GGT AGC ACC GGC GAT TGG CGA AAA ACC ATA CTT AAG 1392
 Asp Leu Asn Ser Gly Ser Thr Gly Asp Trp Arg Lys Thr Ile Leu Lys
 450 455 460

CGT GCA TTT AAT ATT GAT GAT GTC TCG CTC TTC CGC CTG CTT AAA ATT 1440
 Arg Ala Phe Asn Ile Asp Asp Val Ser Leu Phe Arg Leu Leu Lys Ile
 465 470 475 480

ACC GAC CAT GAT AAT AAA GAT GGA AAA ATT AAA AAT AAC CTA AAG AAT 1488
 Thr Asp His Asp Asn Lys Asp Gly Lys Ile Lys Asn Asn Leu Lys Asn
 485 490 495

CTT TCC AAT TTA TAT ATT GGA AAA TTA CTG GCA GAT ATT CAT CAA TTA 1536
 Leu Ser Asn Leu Tyr Ile Gly Lys Leu Leu Ala Asp Ile His Gln Leu
 500 505 510

ACC ATT GAT GAA CTG GAT TTA TTA CTG ATT GCC GTA GGT GAA GGA AAA 1584
 Thr Ile Asp Glu Leu Asp Leu Leu Leu Ile Ala Val Gly Glu Gly Lys
 515 520 525

ACT AAT TTA TCC GCT ATC AGT GAT AAG CAA TTG GCT ACC CTG ATC AGA 1632
 Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Thr Leu Ile Arg
 530 535 540

AAA CTC AAT ACT ATT ACC AGC TGG CTA CAT ACA CAG AAG TGG AGT GTA 1680
 Lys Leu Asn Thr Ile Thr Ser Trp Leu His Thr Gln Lys Trp Ser Val
 545 550 555 560

TTC CAG CTA TTT ATC ATG ACC TCC ACC AGC TAT AAC AAA ACG CTA ACG 1728
 Phe Gln Leu Phe Ile Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr
 565 570 575

CCT GAA ATT AAG AAT TTG CTG GAT ACC GTC TAC CAC GGT TTA CAA GGT 1776
 Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly
 580 585 590

TTT GAT AAA GAC AAA GCA GAT TTG CTA CAT GTC ATG GCG CCC TAT ATT 1824
 Phe Asp Lys Asp Lys Ala Asp Leu Leu His Val Met Ala Pro Tyr Ile
 595 600 605

GCG GCC ACC TTG CAA TTA TCA TCG GAA AAT GTC GCC CAC TCG GTA CTC 1872
 Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu
 610 615 620

CTT TGG GCA GAT AAG TTA CAG CCC GGC GAC GGC GCA ATG ACA GCA GAA 1920
 Leu Trp Ala Asp Lys Leu Gln Pro Gly Asp Gly Ala Met Thr Ala Glu
 625 630 635 640

AAA TTC TGG GAC TGG TTG AAT ACT AAG TAT ACG CCG GGT TCA TCG GAA 1968
 Lys Phe Trp Asp Trp Leu Asn Thr Lys Tyr Thr Pro Gly Ser Ser Glu
 645 650 655

GCC GTA GAA ACG CAG GAA CAT ATC GTT CAG TAT TGT CAG GCT CTG GCA 2016
 Ala Val Glu Thr Gln Glu His Ile Val Gln Tyr Cys Gln Ala Leu Ala
 660 665 670

CAA TTG GAA ATG GTT TAC CAT TCC ACC GGC ATC AAC GAA AAC GCC TTC 2064
 Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe
 675 680 685

CGT CTA TTT GTG ACA AAA CCA GAG ATG TTT GGC GCT GCA ACT GGA GCA 2112
 Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ala Ala Thr Gly Ala
 690 695 700

GCG CCC GCG CAT GAT GCC CTT TCA CTG ATT ATG CTG ACA CGT TTT GCG 2160
 Ala Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala
 705 710 715 720

GAT TGG GTG AAC GCA CTA GGC GAA AAA GCG TCC TCG GTG CTA GCG GCA 2208
 Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala
 725 730 735

TTT GAA GCT AAC TCG TTA ACG GCA GAA CAA CTG GCT GAT GCC ATG AAT 2256
 Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn
 740 745 750

CTT GAT GCT AAT TTG CTG TTG CAA GCC AGT ATT CAA GCA CAA AAT CAT 2304
 Leu Asp Ala Asn Leu Leu Leu Gln Ala Ser Ile Gln Ala Gln Asn His
 755 760 765

CAA CAT CTT CCC CCA GTA ACT CCA GAA AAT GCG TTC TCC TGT TGG ACA 2352
 Gln His Leu Pro Pro Val Thr Pro Glu Asn Ala Phe Ser Cys Trp Thr
 770 775 780

TCT ATC AAT ACT ATC CTG CAA TGG GTT AAT GTC GCA CAA CAA TTG AAT 2400
 Ser Ile Asn Thr Ile Leu Gln Trp Val Asn Val Ala Gln Gln Leu Asn
 785 790 795 800

GTC GCC CCA CAG GGC GTT TCC GCT TTG GTC GGG CTG GAT TAT ATT CAA 2448
 Val Ala Pro Gln Gly Val Ser Ala Leu Val Gly Leu Asp Tyr Ile Gln
 805 810 815

TCA ATG AAA GAG ACA CCG ACC TAT GCC CAG TGG GAA AAC GCG GCA GGC 2496
 Ser Met Lys Glu Thr Pro Thr Tyr Ala Gln Trp Glu Asn Ala Ala Gly
 820 825 830

GTA TTA ACC GCC GGG TTG AAT TCA CAA CAG GCT AAT ACA TTA CAC GCT 2544
 Val Leu Thr Ala Gly Leu Asn Ser Gln Gln Ala Asn Thr Leu His Ala
 835 840 845

TTT CTG GAT GAA TCT CGC AGT GCC GCA TTA AGC ACC TAC TAT ATC CGT 2592
 Phe Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr Tyr Ile Arg
 850 855 860

CAA GTC GCC AAG GCA GCG GCG GCT ATT AAA AGC CGT GAT GAC TTG TAT 2640
 Gln Val Ala Lys Ala Ala Ala Ala Ile Lys Ser Arg Asp Asp Leu Tyr
 865 870 875 880

CAA TAC TTA CTG ATT GAT AAT CAG GTT TCT GCG GCA ATA AAA ACC ACC 2688
 Gln Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Ala Ile Lys Thr Thr
 885 890 895

CGG ATC GCC GAA GCC ATT GCC AGT ATT CAA CTG TAC GTC AAC CGG GCA 2736
 Arg Ile Ala Glu Ala Ile Ala Ser Ile Gln Leu Tyr Val Asn Arg Ala
 900 905 910

TTG GAA AAT GTG GAA GAA AAT GCC AAT TCG GGG GTT ATC AGC CGC CAA 2784
 Leu Glu Asn Val Glu Glu Asn Ala Asn Ser Gly Val Ile Ser Arg Gln
 915 920 925

TTC TTT ATC GAC TGG GAC AAA TAC AAT AAA CGC TAC AGC ACT TGG GCG 2832
 Phe Phe Ile Asp Trp Asp Lys Tyr Asn Lys Arg Tyr Ser Thr Trp Ala
 930 935 940

GGT GTT TCT CAA TTA GTT TAC TAC CCG GAA AAC TAT ATT GAT CCG ACC 2880
 Gly Val Ser Gln Leu Val Tyr Tyr Pro Glu Asn Tyr Ile Asp Pro Thr
 945 950 955 960

ATG CGT ATC GGA CAA ACC AAA ATG ATG GAC GCA TTA CTG CAA TCC GTC 2928
 Met Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln Ser Val
 965 970 975

AGC CAA AGC CAA TTA AAC GCC GAT ACC GTC GAA GAT GCC TTT ATG TCT 2976
 Ser Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe Met Ser
 980 985 990

TAT CTG ACA TCG TTT GAA CAA GTG GCT AAT CTT AAA GTT ATT AGC GCA 3024
 Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile Ser Ala
 995 1000 1005

TAT CAC GAT AAT ATT AAT AAC GAT CAA GGG CTG ACC TAT TTT ATC GGA 3072
 Tyr His Asp Asn Ile Asn Asn Asp Gln Gly Leu Thr Tyr Phe Ile Gly
 1010 1015 1020

CTC AGT GAA ACT GAT GCC GGT GAA TAT TAT TGG CGC AGT GTC GAT CAC 3120
 Leu Ser Glu Thr Asp Ala Gly Glu Tyr Tyr Trp Arg Ser Val Asp His
 1025 1030 1035 1040

AGT AAA TTC AAC GAC GGT AAA TTC GCG GCT AAT GCC TGG AGT GAA TGG 3168
 Ser Lys Phe Asn Asp Gly Lys Phe Ala Ala Asn Ala Trp Ser Glu Trp
 1045 1050 1055

CAT AAA ATT GAT TGT CCA ATT AAC CCT TAT AAA AGC ACT ATC CGT CCA 3216
 His Lys Ile Asp Cys Pro Ile Asn Pro Tyr Lys Ser Thr Ile Arg Pro
 1060 1065 1070

GTG ATA TAT AAA TCC CGC CTG TAT CTG CTC TGG TTG GAA CAA AAG GAG 3264
 Val Ile Tyr Lys Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln Lys Glu
 1075 1080 1085

ATC ACC AAA CAG ACA GGA AAT AGT AAA GAT GGC TAT CAA ACT GAA ACG 3312
 Ile Thr Lys Gln Thr Gly Asn Ser Lys Asp Gly Tyr Gln Thr Glu Thr
 1090 1095 1100

GAT TAT CGT TAT GAA CTA AAA TTG GCG CAT ATC CGC TAT GAT GGC ACT 3360
 Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Thr
 1105 1110 1115 1120

TGG AAT ACG CCA ATC ACC TTT GAT GTC AAT AAA AAA ATA TCC GAG CTA 3408
 Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Lys Lys Ile Ser Glu Leu
 1125 1130 1135

AAA CTG GAA AAA AAT AGA GCG CCC GGA CTC TAT TGT GCC GGT TAT CAA 3456
 Lys Leu Glu Lys Asn Arg Ala Pro Gly Leu Tyr Cys Ala Gly Tyr Gln
 1140 1145 1150

GGT GAA GAT ACG TTG CTG GTG ATG TTT TAT AAC CAA CAA GAC ACA CTA 3504
 Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln Asp Thr Leu
 1155 1160 1165

GAT AGT TAT AAA AAC GCT TCA ATG CAA GGA CTA TAT ATC TTT GCT GAT 3552
 Asp Ser Tyr Lys Asn Ala Ser Met Gln Gly Leu Tyr Ile Phe Ala Asp
 1170 1175 1180

ATG GCA TCC AAA GAT ATG ACC CCA GAA CAG AGC AAT GTT TAT CGG GAT 3600
 Met Ala Ser Lys Asp Met Thr Pro Glu Gln Ser Asn Val Tyr Arg Asp
 1185 1190 1195 1200

AAT AGC TAT CAA CAA TTT GAT ACC AAT AAT GTC AGA AGA GTG AAT AAC 3648
 Asn Ser Tyr Gln Gln Phe Asp Thr Asn Asn Val Arg Arg Val Asn Asn
 1205 1210 1215

CGC TAT GCA GAG GAT TAT GAG ATT CCT TCC TCG GTA AGT AGC CGT AAA 3696
 Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Ser Ser Arg Lys
 1220 1225 1230

GAC TAT GGT TGG GGA GAT TAT TAC CTC AGC ATG GTA TAT AAC GGA GAT 3744
 Asp Tyr Gly Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp
 1235 1240 1245

ATT CCA ACT ATC AAT TAC AAA GCC GCA TCA AGT GAT TTA AAA ATC TAT 3792
 Ile Pro Thr Ile Asn Tyr Lys Ala Ala Ser Ser Asp Leu Lys Ile Tyr
 1250 1255 1260

ATC TCA CCA AAA TTA AGA ATT ATT CAT AAT GGA TAT GAA GGA CAG AAG 3840
 Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Lys
 1265 1270 1275 1280

CGC AAT CAA TGC AAT CTG ATG AAT AAA TAT GGC AAA CTA GGT GAT AAA 3888
 Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys
 1285 1290 1295

TTT ATT GTT TAT ACT AGC TTG GGG GTC AAT CCA AAT AAC TCG TCA AAT 3936
 Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn
 1300 1305 1310

AAG CTC ATG TTT TAC CCC GTC TAT CAA TAT AGC GGA AAC ACC AGT GGA 3984
 Lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Ser Gly Asn Thr Ser Gly
 1315 1320 1325

CTC AAT CAA GGG AGA CTA CTA TTC CAC CGT GAC ACC ACT TAT CCA TCT 4032
 Leu Asn Gln Gly Arg Leu Leu Phe His Arg Asp Thr Thr Tyr Pro Ser
 1330 1335 1340

AAA GTA GAA GCT TGG ATT CCT GGA GCA AAA CGT TCT CTA ACC AAC CAA 4080
 Lys Val Glu Ala Trp Ile Pro Gly Ala Lys Arg Ser Leu Thr Asn Gln
 1345 1350 1355 1360

AAT GCC GCC ATT GGT GAT GAT TAT GCT ACA GAC TCT CTG AAT AAA CCG 4128
 Asn Ala Ala Ile Gly Asp Asp Tyr Ala Thr Asp Ser Leu Asn Lys Pro
 1365 1370 1375

GAT GAT CTT AAG CAA TAT ATC TTT ATG ACT GAC AGT AAA GGG ACT GCT 4176
 Asp Asp Leu Lys Gln Tyr Ile Phe Met Thr Asp Ser Lys Gly Thr Ala
 1380 1385 1390

ACT GAT GTC TCA GGC CCA GTA GAG ATT AAT ACT GCA ATT TCT CCA GCA 4224
 Thr Asp Val Ser Gly Pro Val Glu Ile Asn Thr Ala Ile Ser Pro Ala
 1395 1400 1405

AAA GTT CAG ATA ATA GTC AAA GCG GGT GGC AAG GAG CAA ACT TTT ACC 4272
 Lys Val Gln Ile Ile Val Lys Ala Gly Gly Lys Glu Gln Thr Phe Thr
 1410 1415 1420

GCA GAT AAA GAT GTC TCC ATT CAG CCA TCA CCT AGC TTT GAT GAA ATG 4320
 Ala Asp Lys Asp Val Ser Ile Gln Pro Ser Pro Ser Phe Asp Glu Met
 1425 1430 1435 1440

AAT TAT CAA TTT AAT GCC CTT GAA ATA GAC GGT TCT GGT CTG AAT TTT 4368
 Asn Tyr Gln Phe Asn Ala Leu Glu Ile Asp Gly Ser Gly Leu Asn Phe
 1445 1450 1455

ATT AAC AAC TCA GCC AGT ATT GAT GTT ACT TTT ACC GCA TTT GCG GAG 4416
 Ile Asn Asn Ser Ala Ser Ile Asp Val Thr Phe Thr Ala Phe Ala Glu
 1460 1465 1470

GAT GGC CGC AAA CTG GGT TAT GAA AGT TTC AGT ATT CCT GTT ACC CTC 4464
 Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro Val Thr Leu
 1475 1480 1485

AAG GTA AGT ACC GAT AAT GCC CTG ACC CTG CAC CAT AAT GAA AAT GGT 4512
 Lys Val Ser Thr Asp Asn Ala Leu Thr Leu His His Asn Glu Asn Gly
 1490 1495 1500

GCG CAA TAT ATG CAA TGG CAA TCC TAT CGT ACC CGC CTG AAT ACT CTA 4560
 Ala Gln Tyr Met Gln Trp Gln Ser Tyr Arg Thr Arg Leu Asn Thr Leu
 1505 1510 1515 1520

TTT GCC CGC CAG TTG GTT GCA CGC GCC ACC ACC GGA ATC GAT ACA ATT 4608
 Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile Asp Thr Ile
 1525 1530 1535

CTG AGT ATG GAA ACT CAG AAT ATT CAG GAA CCG CAG TTA GGC AAA GGT 4656
 Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly Lys Gly
 1540 1545 1550

TTC TAT GCT ACG TTC GTG ATA CCT CCC TAT AAC CTA TCA ACT CAT GGT 4704
 Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Leu Ser Thr His Gly
 1555 1560 1565

GAT GAA CGT TGG TTT AAG CTT TAT ATC AAA CAT GTT GTT GAT AAT AAT 4752
 Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val Asp Asn Asn
 1570 1575 1580

TCA CAT ATT ATC TAT TCA GGC CAG CTA ACA GAT ACA AAT ATA AAC ATC 4800
 Ser His Ile Ile Tyr Ser Gly Gln Leu Thr Asp Thr Asn Ile Asn Ile
 1585 1590 1595 1600

ACA TTA TTT ATT CCT CTT GAT GAT GTC CCA TTG AAT CAA GAT TAT CAC 4848
 Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln Asp Tyr His
 1605 1610 1615

GCC AAG GTT TAT ATG ACC TTC AAG AAA TCA CCA TCA GAT GGT ACC TGG 4896
 Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Gly Thr Trp
 1620 1625 1630

TGG GGC CCT CAC TTT GTT AGA GAT GAT AAA GGA ATA GTA ACA ATA AAC 4944
 Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val Thr Ile Asn
 1635 1640 1645

CCT AAA TCC ATT TTG ACC CAT TTT GAG AGC GTC AAT GTC CTG AAT AAT 4992
 Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val Leu Asn Asn
 1650 1655 1660

ATT AGT AGC GAA CCA ATG GAT TTC AGC GGC GCT AAC AGC CTC TAT TTC 5040
 Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser Leu Tyr Phe
 1665 1670 1675 1680

TGG GAA CTG TTC TAC TAT ACC CCG ATG CTG GTT GCT CAA CGT TTG CTG 5088
 Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln Arg Leu Leu
 1685 1690 1695

CAT GAA CAG AAC TTC GAT GAA GCC AAC CGT TGG CTG AAA TAT GTC TGG 5136
 His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys Tyr Val Trp
 1700 1705 1710

AGT CCA TCC GGT TAT ATT GTC CAC GGC CAG ATT CAG AAC TAC CAG TGG 5184
 Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn Tyr Gln Trp
 1715 1720 1725

AAC GTC CGC CCG TTA CTG GAA GAC ACC AGT TGG AAC AGT GAT CCT TTG 5232
 Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser Asp Pro Leu
 1730 1735 1740

GAT TCC GTC GAT CCT GAC GCG GTA GCA CAG CAC GAT CCA ATG CAC TAC 5280
 Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro Met His Tyr
 1745 1750 1755 1760

AAA GTT TCA ACT TTT ATG CGT ACC TTG GAT CTA TTG ATA GCA CGC GGC 5328
 Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile Ala Arg Gly
 1765 1770 1775

GAC CAT GCT TAT CGC CAA CTG GAA CGA GAT ACA CTC AAC GAA GCG AAG 5376
 Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn Glu Ala Lys
 1780 1785 1790

ATG TGG TAT ATG CAA GCG CTG CAT CTA TTA GGT GAC AAA CCT TAT CTA 5424
 Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys Pro Tyr Leu
 1795 1800 1805

CCG CTG AGT ACG ACA TGG AGT GAT CCA CGA CTA GAC AGA GCC GCG GAT 5472
 Pro Leu Ser Thr Thr Trp Ser Asp Pro Arg Leu Asp Arg Ala Ala Asp
 1810 1815 1820

ATC ACT ACC CAA AAT GCT CAC GAC AGC GCA ATA GTC GCT CTG CGG CAG 5520
 Ile Thr Thr Gln Asn Ala His Asp Ser Ala Ile Val Ala Leu Arg Gln
 1825 1830 1835 1840

AAT ATA CCT ACA CCG GCA CCT TTA TCA 5547
 Asn Ile Pro Thr Pro Ala Pro Leu Ser
 1845 1849

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1849 amino acids

(B) TYPE: amino acids

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49 (TcdAii):

| Features | From | To | Description |
|---|---|--------|----------------------------------|
| Peptide1 | 1849 | TcdAii | peptide |
| Fragment | 1 | 12 | TcdAii N-terminus (SEQ ID NO:13) |
| Fragment | 196 | 211 | (SEQ ID NO:38) |
| Fragment | 466 | 475 | (SEQ ID NO:17) |
| Fragment | 993 | 1004 | (SEQ ID NO:23; 12/13) |
| Fragment | 1297 | 1312 | (SEQ ID NO:18) |
| Fragment | 1390 | 1409 | (SEQ ID NO:39) |
| Fragment | 1532 | 1554 | (SEQ ID NO:21; 19/23) |
| Leu Ile Gly Tyr | Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln Tyr Val | | |
| 1 | 5 | 10 | 15 |
| Ala Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr Leu Thr | | | |
| | 20 | 25 | 30 |
| Glu Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser Val Tyr | | | |
| | 35 | 40 | 45 |
| Tyr Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu Ser Gln | | | |
| | 50 | 55 | 60 |
| Gln Asn Met Asp Ile Glu Leu Ser Thr Leu Ser Leu Ser Asn Glu Leu | | | |
| 65 | 70 | 75 | 80 |
| Leu Leu Glu Ser Ile Lys Thr Glu Ser Lys Leu Glu Asn Tyr Thr Lys | | | |
| | 85 | 90 | 95 |
| Val Met Glu Met Leu Ser Thr Phe Arg Pro Ser Gly Ala Thr Pro Tyr | | | |
| | 100 | 105 | 110 |
| His Asp Ala Tyr Glu Asn Val Arg Glu Val Ile Gln Leu Gln Asp Pro | | | |
| | 115 | 120 | 125 |
| Gly Leu Glu Gln Leu Asn Ala Ser Pro Ala Ile Ala Gly Leu Met His | | | |
| | 130 | 135 | 140 |
| Gln Ala Ser Leu Leu Gly Ile Asn Ala Ser Ile Ser Pro Glu Leu Phe | | | |
| 145 | 150 | 155 | 160 |
| Asn Ile Leu Thr Glu Glu Ile Thr Glu Gly Asn Ala Glu Glu Leu Tyr | | | |
| | 165 | 170 | 175 |
| Lys Lys Asn Phe Gly Asn Ile Glu Pro Ala Ser Leu Ala Met Pro Glu | | | |
| | 180 | 185 | 190 |
| Tyr Leu Lys Arg Tyr Tyr Asn Leu Ser Asp Glu Glu Leu Ser Gln Phe | | | |
| | 195 | 200 | 205 |
| Ile Gly Lys Ala Ser Asn Phe Gly Gln Gln Glu Tyr Ser Asn Asn Gln | | | |
| | 210 | 215 | 220 |
| Leu Ile Thr Pro Val Val Asn Ser Ser Asp Gly Thr Val Lys Val Tyr | | | |
| 225 | 230 | 235 | 240 |
| Arg Ile Thr Arg Glu Tyr Thr Thr Asn Ala Tyr Gln Met Asp Val Glu | | | |

Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu
 610 615 620
 Leu Trp Ala Asp Lys Leu Gln Pro Gly Asp Gly Ala Met Thr Ala Glu
 625 630 635 640
 Lys Phe Trp Asp Trp Leu Asn Thr Lys Tyr Thr Pro Gly Ser Ser Glu
 645 650 655
 Ala Val Glu Thr Gln Glu His Ile Val Gln Tyr Cys Gln Ala Leu Ala
 660 665 670
 Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe
 675 680 685
 Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ala Ala Thr Gly Ala
 690 695 700
 Ala Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala
 705 710 715 720
 Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala
 725 730 735
 Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn
 740 745 750
 Leu Asp Ala Asn Leu Leu Leu Gln Ala Ser Ile Gln Ala Gln Asn His
 755 760 765
 Gln His Leu Pro Pro Val Thr Pro Glu Asn Ala Phe Ser Cys Trp Thr
 770 775 780
 Ser Ile Asn Thr Ile Leu Gln Trp Val Asn Val Ala Gln Gln Leu Asn
 785 790 795 800
 Val Ala Pro Gln Gly Val Ser Ala Leu Val Gly Leu Asp Tyr Ile Gln
 805 810 815
 Ser Met Lys Glu Thr Pro Thr Tyr Ala Gln Trp Glu Asn Ala Ala Gly
 820 825 830
 Val Leu Thr Ala Gly Leu Asn Ser Gln Gln Ala Asn Thr Leu His Ala
 835 840 845
 Phe Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr Tyr Ile Arg
 850 855 860
 Gln Val Ala Lys Ala Ala Ala Ala Ile Lys Ser Arg Asp Asp Leu Tyr
 865 870 875 880
 Gln Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Ala Ile Lys Thr Thr
 885 890 895
 Arg Ile Ala Glu Ala Ile Ala Ser Ile Gln Leu Tyr Val Asn Arg Ala
 900 905 910
 Leu Glu Asn Val Glu Glu Asn Ala Asn Ser Gly Val Ile Ser Arg Gln
 915 920 925
 Phe Phe Ile Asp Trp Asp Lys Tyr Asn Lys Arg Tyr Ser Thr Trp Ala
 930 935 940
 Gly Val Ser Gln Leu Val Tyr Tyr Pro Glu Asn Tyr Ile Asp Pro Thr
 945 950 955 960
 Met Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln Ser Val

| | | | |
|---|------|------|------|
| | 965 | 970 | 975 |
| Ser Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe Met Ser | | | |
| | 980 | 985 | 990 |
| Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile Ser Ala | | | |
| | 995 | 1000 | 1005 |
| Tyr His Asp Asn Ile Asn Asn Asp Gln Gly Leu Thr Tyr Phe Ile Gly | | | |
| | 1010 | 1015 | 1020 |
| Leu Ser Glu Thr Asp Ala Gly Glu Tyr Tyr Trp Arg Ser Val Asp His | | | |
| | 1025 | 1030 | 1035 |
| Ser Lys Phe Asn Asp Gly Lys Phe Ala Ala Asn Ala Trp Ser Glu Trp | | | |
| | 1045 | 1050 | 1055 |
| His Lys Ile Asp Cys Pro Ile Asn Pro Tyr Lys Ser Thr Ile Arg Pro | | | |
| | 1060 | 1065 | 1070 |
| Val Ile Tyr Lys Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln Lys Glu | | | |
| | 1075 | 1080 | 1085 |
| Ile Thr Lys Gln Thr Gly Asn Ser Lys Asp Gly Tyr Gln Thr Glu Thr | | | |
| | 1090 | 1095 | 1100 |
| Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Thr | | | |
| | 1105 | 1110 | 1115 |
| Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Lys Lys Ile Ser Glu Leu | | | |
| | 1125 | 1130 | 1135 |
| Lys Leu Glu Lys Asn Arg Ala Pro Gly Leu Tyr Cys Ala Gly Tyr Gln | | | |
| | 1140 | 1145 | 1150 |
| Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln Asp Thr Leu | | | |
| | 1155 | 1160 | 1165 |
| Asp Ser Tyr Lys Asn Ala Ser Met Gln Gly Leu Tyr Ile Phe Ala Asp | | | |
| | 1170 | 1175 | 1180 |
| Met Ala Ser Lys Asp Met Thr Pro Glu Gln Ser Asn Val Tyr Arg Asp | | | |
| | 1185 | 1190 | 1195 |
| Asn Ser Tyr Gln Gln Phe Asp Thr Asn Asn Val Arg Arg Val Asn Asn | | | |
| | 1205 | 1210 | 1215 |
| Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Ser Ser Arg Lys | | | |
| | 1220 | 1225 | 1230 |
| Asp Tyr Gly Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp | | | |
| | 1235 | 1240 | 1245 |
| Ile Pro Thr Ile Asn Tyr Lys Ala Ala Ser Ser Asp Leu Lys Ile Tyr | | | |
| | 1250 | 1255 | 1260 |
| Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Lys | | | |
| | 1265 | 1270 | 1275 |
| Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys | | | |
| | 1285 | 1290 | 1295 |
| Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn | | | |
| | 1300 | 1305 | 1310 |
| Lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Ser Gly Asn Thr Ser Gly | | | |
| | 1315 | 1320 | 1325 |

Leu Asn Gln Gly Arg Leu Leu Phe His Arg Asp Thr Thr Tyr Pro Ser
 1330 1335 1340
 Lys Val Glu Ala Trp Ile Pro Gly Ala Lys Arg Ser Leu Thr Asn Gln
 1345 1350 1355 1360
 Asn Ala Ala Ile Gly Asp Asp Tyr Ala Thr Asp Ser Leu Asn Lys Pro
 1365 1370 1375
 Asp Asp Leu Lys Gln Tyr Ile Phe Met Thr Asp Ser Lys Gly Thr Ala
 1380 1385 1390
 Thr Asp Val Ser Gly Pro Val Glu Ile Asn Thr Ala Ile Ser Pro Ala
 1395 1400 1405
 Lys Val Gln Ile Ile Val Lys Ala Gly Gly Lys Glu Gln Thr Phe Thr
 1410 1415 1420
 Ala Asp Lys Asp Val Ser Ile Gln Pro Ser Pro Ser Phe Asp Glu Met
 1425 1430 1435 1440
 Asn Tyr Gln Phe Asn Ala Leu Glu Ile Asp Gly Ser Gly Leu Asn Phe
 1445 1450 1455
 Ile Asn Asn Ser Ala Ser Ile Asp Val Thr Phe Thr Ala Phe Ala Glu
 1460 1465 1470
 Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro Val Thr Leu
 1475 1480 1485
 Lys Val Ser Thr Asp Asn Ala Leu Thr Leu His His Asn Glu Asn Gly
 1490 1495 1500
 Ala Gln Tyr Met Gln Trp Gln Ser Tyr Arg Thr Arg Leu Asn Thr Leu
 1505 1510 1515 1520
 Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile Asp Thr Ile
 1525 1530 1535
 Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly Lys Gly
 1540 1545 1550
 Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Leu Ser Thr His Gly
 1555 1560 1565
 Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val Asp Asn Asn
 1570 1575 1580
 Ser His Ile Ile Tyr Ser Gly Gln Leu Thr Asp Thr Asn Ile Asn Ile
 1585 1590 1595 1600
 Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln Asp Tyr His
 1605 1610 1615
 Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Gly Thr Trp
 1620 1625 1630
 Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val Thr Ile Asn
 1635 1640 1645
 Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val Leu Asn Asn
 1650 1655 1660
 Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser Leu Tyr Phe
 1665 1670 1675 1680
 Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln Arg Leu Leu

1685 1690 1695
 His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys Tyr Val Trp
 1700 1705 1710
 Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn Tyr Gln Trp
 1715 1720 1725
 Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser Asp Pro Leu
 1730 1735 1740
 Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro Met His Tyr
 1745 1750 1755 1760
 Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile Ala Arg Gly
 1765 1770 1775
 Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn Glu Ala Lys
 1780 1785 1790
 Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys Pro Tyr Leu
 1795 1800 1805
 Pro Leu Ser Thr Thr Trp Ser Asp Pro Arg Leu Asp Arg Ala Ala Asp
 1810 1815 1820
 Ile Thr Thr Gln Asn Ala His Asp Ser Ala Ile Val Ala Leu Arg Gln
 1825 1830 1835 1840
 Asn Ile Pro Thr Pro Ala Pro Leu Ser
 1845 1849

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1740 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50 (tcdAiii coding region):

TTG CGC AGC GCT AAT ACC CTG ACT GAT CTC TTC CTG CCG CAA ATC AAT 48
 Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile Asn
 1 5 10 15
 GAA GTG ATG ATG AAT TAC TGG CAG ACA TTA GCT CAG AGA GTA TAC AAT 96
 Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr Asn
 20 25 30
 CTG CGT CAT AAC CTC TCT ATC GAC GGC CAG CCG TTA TAT CTG CCA ATC 144
 Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Tyr Leu Pro Ile
 35 40 45
 TAT GCC ACA CCG GCC GAT CCG AAA GCG TTA CTC AGC GCC GCC GTT GCC 192
 Tyr Ala Thr Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ala
 50 55 60
 ACT TCT CAA GGT GGA GGC AAG CTA CCG GAA TCA TTT ATG TCC CTG TGG 240
 Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu Trp
 65 70 75 80

CGT TTC CCG CAC ATG CTG GAA AAT GCG CGC GGC ATG GTT AGC CAG CTC 288
 Arg Phe Pro His Met Leu Glu Asn Ala Arg Gly Met Val Ser Gln Leu
 85 90 95

ACC CAG TTC GGC TCC ACG TTA CAA AAT ATT ATC GAA CGT CAG GAC GCG 336
 Thr Gln Phe Gly Ser Thr Leu Gln Asn Ile Ile Glu Arg Gln Asp Ala
 100 105 110

GAA GCG CTC AAT GCG TTA TTA CAA AAT CAG GCC GCC GAG CTG ATA TTG 384
 Glu Ala Leu Asn Ala Leu Leu Gln Asn Gln Ala Ala Glu Leu Ile Leu
 115 120 125

ACT AAC CTG AGC ATT CAG GAC AAA ACC ATT GAA GAA TTG GAT GCC GAG 432
 Thr Asn Leu Ser Ile Gln Asp Lys Thr Ile Glu Glu Leu Asp Ala Glu
 130 135 140

AAA ACG GTG TTG GAA AAA TCC AAA GCG GGA GCA CAA TCG CGC TTT GAT 480
 Lys Thr Val Leu Glu Lys Ser Lys Ala Gly Ala Gln Ser Arg Phe Asp
 145 150 155 160

AGC TAC GGC AAA CTG TAC GAT GAG AAT ATC AAC GCC GGT GAA AAC CAA 528
 Ser Tyr Gly Lys Leu Tyr Asp Glu Asn Ile Asn Ala Gly Glu Asn Gln
 165 170 175

GCC ATG ACG CTA CGA GCG TCC GCC GCC GGG CTT ACC ACG GCA GTT CAG 576
 Ala Met Thr Leu Arg Ala Ser Ala Ala Gly Leu Thr Thr Ala Val Gln
 180 185 190

GCA TCC CGT CTG GCC GGT GCG GCG GCT GAT CTG GTG CCT AAC ATC TTC 624
 Ala Ser Arg Leu Ala Gly Ala Ala Ala Asp Leu Val Pro Asn Ile Phe
 195 200 205

GGC TTT GCC GGT GGC GGC AGC CGT TGG GGG GCT ATC GCT GAG GCG ACA 672
 Gly Phe Ala Gly Gly Gly Ser Arg Trp Gly Ala Ile Ala Glu Ala Thr
 210 215 220

GGT TAT GTG ATG GAA TTC TCC GCG AAT GTT ATG AAC ACC GAA GCG GAT 720
 Gly Tyr Val Met Glu Phe Ser Ala Asn Val Met Asn Thr Glu Ala Asp
 225 230 235 240

AAA ATT AGC CAA TCT GAA ACC TAC CGT CGT CGC CGT CAG GAG TGG GAG 768
 Lys Ile Ser Gln Ser Glu Thr Tyr Arg Arg Arg Arg Gln Glu Trp Glu
 245 250 255

ATC CAG CGG AAT AAT GCC GAA GCG GAA TTG AAG CAA ATC GAT GCT CAG 816
 Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala Gln
 260 265 270

CTC AAA TCA CTC GCT GTA CGC CGC GAA GCC GCC GTA TTG CAG AAA ACC 864
 Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys Thr
 275 280 285

AGT CTG AAA ACC CAA CAA GAA CAG ACC CAA TCT CAA TTG GCC TTC CTG 912
 Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe Leu
 290 295 300

CAA CGT AAG TTC AGC AAT CAG GCG TTA TAC AAC TGG CTG CGT GGT CGA 960
 Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly Arg
 305 310 315 320

CTG GCG GCG ATT TAC TTC CAG TTC TAC GAT TTG GCC GTC GCG CGT TGC 1008
 Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg Cys
 325 330 335

CTG ATG GCA GAA CAA GCT TAC CGT TGG GAA CTC AAT GAT GAC TCT GCC 1056
 Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser Ala
 340 345 350

CGC TTC ATT AAA CCG GGC GCC TGG CAG GGA ACC TAT GCC GGT CTG CTT 1104
 Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu
 355 360 365

GCA GGT GAA ACC TTG ATG CTG AGT CTG GCA CAA ATG GAA GAC GCT CAT 1152
 Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala His
 370 375 380

CTG AAA CGC GAT AAA CGC GCA TTA GAG GTT GAA CGC ACA GTA TCG CTG 1200
 Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu
 385 390 395 400

GCC GAA GTT TAT GCA GGA TTA CCA AAA GAT AAC GGT CCA TTT TCC CTG 1248
 Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser Leu
 405 410 415

GCT CAG GAA ATT GAC AAG CTG GTG AGT CAA GGT TCA GGC AGT GCC GGC 1296
 Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala Gly
 420 425 430

AGT GGT AAT AAT AAT TTG GCG TTC GGC GCC GGC ACG GAC ACT AAA ACC 1344
 Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys Thr
 435 440 445

TCT TTG CAG GCA TCA GTT TCA TTC GCT GAT TTG AAA ATT CGT GAA GAT 1392
 Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu Asp
 450 455 460

TAC CCG GCA TCG CTT GGC AAA ATT CGA CGT ATC AAA CAG ATC AGC GTC 1440
 Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser Val
 465 470 475 480

ACT TTG CCC GCG CTA CTG GGA CCG TAT CAG GAT GTA CAG GCA ATA TTG 1488
 Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile Leu
 485 490 495

TCT TAC GGC GAT AAA GCC GGA TTA GCT AAC GGC TGT GAA GCG CTG GCA 1536
 Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu Ala
 500 505 510

GTT TCT CAC GGT ATG AAT GAC AGC GGC CAA TTC CAG CTC GAT TTC AAC 1584
 Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn
 515 520 525

GAT GGC AAA TTC CTG CCA TTC GAA GGC ATC GCC ATT GAT CAA GGC ACG 1632
 Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly Thr
 530 535 540

CTG ACA CTG AGC TTC CCA AAT GCA TCT ATG CCG GAG AAA GGT AAA CAA 1680
 Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys Gln
 545 550 555 560

GCC ACT ATG TTA AAA ACC CTG AAC GAT ATC ATT TTG CAT ATT CGC TAC 1728
 Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg Tyr
 565 570 575
 ACC ATT AAA TAA 1740
 Thr Ile Lys v?
 579

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 579 amino acids
- (B) TYPE: amino acids
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51 (TcdAiii):

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

245 250 255
 Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala Gln
 260 265 270
 Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys Thr
 275 280 285
 Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe Leu
 290 295 300
 Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly Arg
 305 310 315 320
 Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg Cys
 325 330 335
 Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser Ala
 340 345 350
 Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu
 355 360 365
 Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala His
 370 375 380
 Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu
 385 390 395 400
 Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser Leu
 405 410 415
 Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala Gly
 420 425 430
 Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys Thr
 435 440 445
 Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu Asp
 450 455 460
 Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser Val
 465 470 475 480
 Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile Leu
 485 490 495
 Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu Ala
 500 505 510
 Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn
 515 520 525
 Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly Thr
 530 535 540
 Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys Gln
 545 550 555 560
 Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg Tyr
 565 570 575
 Thr Ile Lys v?
 579

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5532 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52 (tcbAii coding

region):

```

TTT ATA CAA GGT TAT AGT GAT CTG TTT GGT AAT CGT GCT GAT AAC TAT 48
Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn Arg Ala Asp Asn Tyr
1           5           10          15
GCC GCG CCG GGC TCG GTT GCA TCG ATG TTC TCA CCG GCG GCT TAT TTG 96
Ala Ala Pro Gly Ser Val Ala Ser Met Phe Ser Pro Ala Ala Tyr Leu
           20          25          30
ACG GAA TTG TAC CGT GAA GCC AAA AAC TTG CAT GAC AGC AGC TCA ATT 144
Thr Glu Leu Tyr Arg Glu Ala Lys Asn Leu His Asp Ser Ser Ser Ile
           35          40          45
TAT TAC CTA GAT AAA CGT CGC CCG GAT TTA GCA AGC TTA ATG CTC AGC 192
Tyr Tyr Leu Asp Lys Arg Arg Pro Asp Leu Ala Ser Leu Met Leu Ser
           50          55          60
CAG AAA AAT ATG GAT GAG GAA ATT TCA ACG CTG GCT CTC TCT AAT GAA 240
Gln Lys Asn Met Asp Glu Glu Ile Ser Thr Leu Ala Leu Ser Asn Glu
65          70          75          80
TTG TGC CTT GCC GGG ATC GAA ACA AAA ACA GGA AAA TCA CAA GAT GAA 288
Leu Cys Leu Ala Gly Ile Glu Thr Lys Thr Gly Lys Ser Gln Asp Glu
           85          90          95
GTG ATG GAT ATG TTG TCA ACT TAT CGT TTA AGT GGA GAG ACA CCT TAT 336
Val Met Asp Met Leu Ser Thr Tyr Arg Leu Ser Gly Glu Thr Pro Tyr
           100          105          110
CAT CAC GCT TAT GAA ACT GTT CGT GAA ATC GTT CAT GAA CGT GAT CCA 384
His His Ala Tyr Glu Thr Val Arg Glu Ile Val His Glu Arg Asp Pro
           115          120          125
GGA TTT CGT CAT TTG TCA CAG GCA CCC ATT GTT GCT GCT AAG CTC GAT 432
Gly Phe Arg His Leu Ser Gln Ala Pro Ile Val Ala Ala Lys Leu Asp
           130          135          140
CCT GTG ACT TTG TTG GGT ATT AGC TCC CAT ATT TCG CCA GAA CTG TAT 480
Pro Val Thr Leu Leu Gly Ile Ser Ser His Ile Ser Pro Glu Leu Tyr
145          150          155          160
AAC TTG CTG ATT GAG GAG ATC CCG GAA AAA GAT GAA GCC GCG CTT GAT 528
Asn Leu Leu Ile Glu Glu Ile Pro Glu Lys Asp Glu Ala Ala Leu Asp
           165          170          175
ACG CTT TAT AAA ACA AAC TTT GGC GAT ATT ACT ACT GCT CAG TTA ATG 576
Thr Leu Tyr Lys Thr Asn Phe Gly Asp Ile Thr Thr Ala Gln Leu Met
           180          185          190
TCC CCA AGT TAT CTG GCC CGG TAT TAT GGC GTC TCA CCG GAA GAT ATT 624
Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr Gly Val Ser Pro Glu Asp Ile

```

195 200 205
 GCC TAC GTG ACG ACT TCA TTA TCA CAT GTT GGA TAT AGC AGT GAT ATT 672
 Ala Tyr Val Thr Thr Ser Leu Ser His Val Gly Tyr Ser Ser Asp Ile
 210 215 220
 CTG GTT ATT CCG TTG GTC GAT GGT GTG GGT AAG ATG GAA GTA GTT CGT 720
 Leu Val Ile Pro Leu Val Asp Gly Val Gly Lys Met Glu Val Val Arg
 225 230 235 240
 GTT ACC CGA ACA CCA TCG GAT AAT TAT ACC AGT CAG ACG AAT TAT ATT 768
 Val Thr Arg Thr Pro Ser Asp Asn Tyr Thr Ser Gln Thr Asn Tyr Ile
 245 250 255
 GAG CTG TAT CCA CAG GGT GGC GAC AAT TAT TTG ATC AAA TAC AAT CTA 816
 Glu Leu Tyr Pro Gln Gly Gly Asp Asn Tyr Leu Ile Lys Tyr Asn Leu
 260 265 270
 AGC AAT AGT TTT GGT TTG GAT GAT TTT TAT CTG CAA TAT AAA GAT GGT 864
 Ser Asn Ser Phe Gly Leu Asp Asp Phe Tyr Leu Gln Tyr Lys Asp Gly
 275 280 285
 TCC GCT GAT TGG ACT GAG ATT GCC CAT AAT CCC TAT CCT GAT ATG GTC 912
 Ser Ala Asp Trp Thr Glu Ile Ala His Asn Pro Tyr Pro Asp Met Val
 290 295 300
 ATA AAT CAA AAG TAT GAA TCA CAG GCG ACA ATC AAA CGT AGT GAC TCT 960
 Ile Asn Gln Lys Tyr Glu Ser Gln Ala Thr Ile Lys Arg Ser Asp Ser
 305 310 315 320
 GAC AAT ATA CTC AGT ATA GGG TTA CAA AGA TGG CAT AGC GGT AGT TAT 1008
 Asp Asn Ile Leu Ser Ile Gly Leu Gln Arg Trp His Ser Gly Ser Tyr
 325 330 335
 AAT TTT GCC GCC GCC AAT TTT AAA ATT GAC CAA TAC TCC CCG AAA GCT 1056
 Asn Phe Ala Ala Ala Asn Phe Lys Ile Asp Gln Tyr Ser Pro Lys Ala
 340 345 350
 TTC CTG CTT AAA ATG AAT AAG GCT ATT CGG TTG CTC AAA GCT ACC GGC 1104
 Phe Leu Leu Lys Met Asn Lys Ala Ile Arg Leu Leu Lys Ala Thr Gly
 355 360 365
 CTC TCT TTT GCT ACG TTG GAG CGT ATT GTT GAT AGT GTT AAT AGC ACC 1152
 Leu Ser Phe Ala Thr Leu Glu Arg Ile Val Asp Ser Val Asn Ser Thr
 370 375 380
 AAA TCC ATC ACG GTT GAG GTA TTA AAC AAG GTT TAT CGG GTA AAA TTC 1200
 Lys Ser Ile Thr Val Glu Val Leu Asn Lys Val Tyr Arg Val Lys Phe
 385 390 395 400
 TAT ATT GAT CGT TAT GGC ATC AGT GAA GAG ACA GCC GCT ATT TTG GCT 1248
 Tyr Ile Asp Arg Tyr Gly Ile Ser Glu Glu Thr Ala Ala Ile Leu Ala
 405 410 415
 AAT ATT AAT ATC TCT CAG CAA GCT GTT GGC AAT CAG CTT AGC CAG TTT 1296
 Asn Ile Asn Ile Ser Gln Gln Ala Val Gly Asn Gln Leu Ser Gln Phe
 420 425 430
 GAG CAA CTA TTT AAT CAC CCG CCG CTC AAT GGT ATT CGC TAT GAA ATC 1344
 Glu Gln Leu Phe Asn His Pro Pro Leu Asn Gly Ile Arg Tyr Glu Ile

435 440 445
 AGT GAG GAC AAC TCC AAA CAT CTT CCT AAT CCT GAT CTG AAC CTT AAA 1392
 Ser Glu Asp Asn Ser Lys His Leu Pro Asn Pro Asp Leu Asn Leu Lys
 450 455 460
 CCA GAC AGT ACC GGT GAT GAT CAA CGC AAG GCG GTT TTA AAA CGC GCG 1440
 Pro Asp Ser Thr Gly Asp Asp Gln Arg Lys Ala Val Leu Lys Arg Ala
 465 470 475 480
 TTT CAG GTT AAC GCC AGT GAG TTG TAT CAG ATG TTA TTG ATC ACT GAT 1488
 Phe Gln Val Asn Ala Ser Glu Leu Tyr Gln Met Leu Leu Ile Thr Asp
 485 490 495
 CGT AAA GAA GAC GGT GTT ATC AAA AAT AAC TTA GAG AAT TTG TCT GAT 1536
 Arg Lys Glu Asp Gly Val Ile Lys Asn Asn Leu Glu Asn Leu Ser Asp
 500 505 510
 CTG TAT TTG GTT AGT TTG CTG GCC CAG ATT CAT AAC CTG ACT ATT GCT 1584
 Leu Tyr Leu Val Ser Leu Leu Ala Gln Ile His Asn Leu Thr Ile Ala
 515 520 525
 GAA TTG AAC ATT TTG TTG GTG ATT TGT GGC TAT GGC GAC ACC AAC ATT 1632
 Glu Leu Asn Ile Leu Leu Val Ile Cys Gly Tyr Gly Asp Thr Asn Ile
 530 535 540
 TAT CAG ATT ACC GAC GAT AAT TTA GCC AAA ATA GTG GAA ACA TTG TTG 1680
 Tyr Gln Ile Thr Asp Asp Asn Leu Ala Lys Ile Val Glu Thr Leu Leu
 545 550 555 560
 TGG ATC ACT CAA TGG TTG AAG ACC CAA AAA TGG ACA GTT ACC GAC CTG 1728
 Trp Ile Thr Gln Trp Leu Lys Thr Gln Lys Trp Thr Val Thr Asp Leu
 565 570 575
 TTT CTG ATG ACC ACG GCC ACT TAC AGC ACC ACT TTA ACG CCA GAA ATT 1776
 Phe Leu Met Thr Thr Ala Thr Tyr Ser Thr Thr Leu Thr Pro Glu Ile
 580 585 590
 AGC AAT CTG ACG GCT ACG TTG TCT TCA ACT TTG CAT GGC AAA GAG AGT 1824
 Ser Asn Leu Thr Ala Thr Leu Ser Ser Thr Leu His Gly Lys Glu Ser
 595 600 605
 CTG ATT GGG GAA GAT CTG AAA AGA GCA ATG GCG CCT TGC TTC ACT TCG 1872
 Leu Ile Gly Glu Asp Leu Lys Arg Ala Met Ala Pro Cys Phe Thr Ser
 610 615 620
 GCT TTG CAT TTG ACT TCT CAA GAA GTT GCG TAT GAC CTG CTG TTG TGG 1920
 Ala Leu His Leu Thr Ser Gln Glu Val Ala Tyr Asp Leu Leu Leu Trp
 625 630 635 640
 ATA GAC CAG ATT CAA CCG GCA CAA ATA ACT GTT GAT GGG TTT TGG GAA 1968
 Ile Asp Gln Ile Gln Pro Ala Gln Ile Thr Val Asp Gly Phe Trp Glu
 645 650 655
 GAA GTG CAA ACA ACA CCA ACC AGC TTG AAG GTG ATT ACC TTT GCT CAG 2016
 Glu Val Gln Thr Thr Pro Thr Ser Leu Lys Val Ile Thr Phe Ala Gln
 660 665 670
 GTG CTG GCA CAA TTG AGC CTG ATC TAT CGT CGT ATT GGG TTA AGT GAA 2064
 Val Leu Ala Gln Leu Ser Leu Ile Tyr Arg Arg Ile Gly Leu Ser Glu

675 680 685
 ACG GAA CTG TCA CTG ATC GTG ACT CAA TCT TCT CTG CTA GTG GCA GGC 2112
 Thr Glu Leu Ser Leu Ile Val Thr Gln Ser Ser Leu Leu Val Ala Gly
 690 695 700
 AAA AGC ATA CTG GAT CAC GGT CTG TTA ACC CTG ATG GCC TTG GAA GGT 2160
 Lys Ser Ile Leu Asp His Gly Leu Leu Thr Leu Met Ala Leu Glu Gly
 705 710 715 720
 TTT CAT ACC TGG GTT AAT GGC TTG GGG CAA CAT GCC TCC TTG ATA TTG 2208
 Phe His Thr Trp Val Asn Gly Leu Gly Gln His Ala Ser Leu Ile Leu
 725 730 735
 GCG GCG TTG AAA GAC GGA GCC TTG ACA GTT ACC GAT GTA GCA CAA GCT 2256
 Ala Ala Leu Lys Asp Gly Ala Leu Thr Val Thr Asp Val Ala Gln Ala
 740 745 750
 ATG AAT AAG GAG GAA TCT CTC CTA CAA ATG GCA GCT AAT CAG GTG GAG 2304
 Met Asn Lys Glu Glu Ser Leu Leu Gln Met Ala Ala Asn Gln Val Glu
 755 760 765
 AAG GAT CTA ACA AAA CTG ACC AGT TGG ACA CAG ATT GAC GCT ATT CTG 2352
 Lys Asp Leu Thr Lys Leu Thr Ser Trp Thr Gln Ile Asp Ala Ile Leu
 770 775 780
 CAA TGG TTA CAG ATG TCT TCG GCC TTG GCG GTT TCT CCA CTG GAT CTG 2400
 Gln Trp Leu Gln Met Ser Ser Ala Leu Ala Val Ser Pro Leu Asp Leu
 785 790 795 800
 GCA GGG ATG ATG GCC CTG AAA TAT GGG ATA GAT CAT AAC TAT GCT GCC 2448
 Ala Gly Met Met Ala Leu Lys Tyr Gly Ile Asp His Asn Tyr Ala Ala
 805 810 815
 TGG CAA GCT GCG GCG GCT GCG CTG ATG GCT GAT CAT GCT AAT CAG GCA 2496
 Trp Gln Ala Ala Ala Ala Ala Leu Met Ala Asp His Ala Asn Gln Ala
 820 825 830
 CAG AAA AAA CTG GAT GAG ACG TTC AGT AAG GCA TTA TGT AAC TAT TAT 2544
 Gln Lys Lys Leu Asp Glu Thr Phe Ser Lys Ala Leu Cys Asn Tyr Tyr
 835 840 845
 ATT AAT GCT GTT GTC GAT AGT GCT GCT GGA GTA CGT GAT CGT AAC GGT 2592
 Ile Asn Ala Val Val Asp Ser Ala Ala Gly Val Arg Asp Arg Asn Gly
 850 855 860
 TTA TAT ACC TAT TTG CTG ATT GAT AAT CAG GTT TCT GCC GAT GTG ATC 2640
 Leu Tyr Thr Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Asp Val Ile
 865 870 875 880
 ACT TCA CGT ATT GCA GAA GCT ATC GCC GGT ATT CAA CTG TAC GTT AAC 2688
 Thr Ser Arg Ile Ala Glu Ala Ile Ala Gly Ile Gln Leu Tyr Val Asn
 885 890 895
 CGG GCT TTA AAC CGA GAT GAA GGT CAG CTT GCA TCG GAC GTT AGT ACC 2736
 Arg Ala Leu Asn Arg Asp Glu Gly Gln Leu Ala Ser Asp Val Ser Thr
 900 905 910
 CGT CAG TTC TTC ACT GAC TGG GAA CGT TAC AAT AAA CGT TAC AGT ACT 2784
 Arg Gln Phe Phe Thr Asp Trp Glu Arg Tyr Asn Lys Arg Tyr Ser Thr

915 920 925
 TGG GCT GGT GTC TCT GAA CTG GTC TAT TAT CCA GAA AAC TAT GTT GAT 2832
 Trp Ala Gly Val Ser Glu Leu Val Tyr Tyr Pro Glu Asn Tyr Val Asp
 930 935 940
 CCC ACT CAG CGC ATT GGG CAA ACC AAA ATG ATG GAT GCG CTG TTG CAA 2880
 Pro Thr Gln Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln
 945 950 955 960
 TCC ATC AAC CAG AGC CAG CTA AAT GCG GAT ACG GTG GAA GAT GCT TTC 2928
 Ser Ile Asn Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe
 965 970 975
 AAA ACT TAT TTG ACC AGC TTT GAG CAG GTA GCA AAT CTG AAA GTA ATT 2976
 Lys Thr Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile
 980 985 990
 AGT GCT TAC CAC GAT AAT GTG AAT GTG GAT CAA GGA TTA ACT TAT TTT 3024
 Ser Ala Tyr His Asp Asn Val Asn Val Asp Gln Gly Leu Thr Tyr Phe
 995 1000 1005
 ATC GGT ATC GAC CAA GCA GCT CCG GGT ACG TAT TAC TGG CGT AGT GTT 3072
 Ile Gly Ile Asp Gln Ala Ala Pro Gly Thr Tyr Tyr Trp Arg Ser Val
 1010 1015 1020
 GAT CAC AGC AAA TGT GAA AAT GGC AAG TTT GCC GCT AAT GCT TGG GGT 3120
 Asp His Ser Lys Cys Glu Asn Gly Lys Phe Ala Ala Asn Ala Trp Gly
 1025 1030 1035 1040
 GAG TGG AAT AAA ATT ACC TGT GCT GTC AAT CCT TGG AAA AAT ATC ATC 3168
 Glu Trp Asn Lys Ile Thr Cys Ala Val Asn Pro Trp Lys Asn Ile Ile
 1045 1050 1055
 CGT CCG GTT GTT TAT ATG TCC CGC TTA TAT CTG CTA TGG CTG GAG CAG 3216
 Arg Pro Val Val Tyr Met Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln
 1060 1065 1070
 CAA TCA AAG AAA AGT GAT GAT GGT AAA ACC ACG ATT TAT CAA TAT AAC 3264
 Gln Ser Lys Lys Ser Asp Asp Gly Lys Thr Thr Ile Tyr Gln Tyr Asn
 1075 1080 1085
 TTA AAA CTG GCT CAT ATT CGT TAC GAC GGT AGT TGG AAT ACA CCA TTT 3312
 Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Ser Trp Asn Thr Pro Phe
 1090 1095 1100
 ACT TTT GAT GTG ACA GAA AAG GTA AAA AAT TAC ACG TCG AGT ACT GAT 3360
 Thr Phe Asp Val Thr Glu Lys Val Lys Asn Tyr Thr Ser Ser Thr Asp
 1105 1110 1115 1120
 GCT GCT GAA TCT TTA GGG TTG TAT TGT ACT GGT TAT CAA GGG GAA GAC 3408
 Ala Ala Glu Ser Leu Gly Leu Tyr Cys Thr Gly Tyr Gln Gly Glu Asp
 1125 1130 1135
 ACT CTA TTA GTT ATG TTC TAT TCG ATG CAG AGT AGT TAT AGC TCC TAT 3456
 Thr Leu Leu Val Met Phe Tyr Ser Met Gln Ser Ser Tyr Ser Ser Tyr
 1140 1145 1150
 ACC GAT AAT AAT GCG CCG GTC ACT GGG CTA TAT ATT TTC GCT GAT ATG 3504
 Thr Asp Asn Asn Ala Pro Val Thr Gly Leu Tyr Ile Phe Ala Asp Met

1155 1160 1165
 TCA TCA GAC AAT ATG ACG AAT GCA CAA GCA ACT AAC TAT TGG AAT AAC 3552
 Ser Ser Asp Asn Met Thr Asn Ala Gln Ala Thr Asn Tyr Trp Asn Asn
 1170 1175 1180
 AGT TAT CCG CAA TTT GAT ACT GTG ATG GCA GAT CCG GAT AGC GAC AAT 3600
 Ser Tyr Pro Gln Phe Asp Thr Val Met Ala Asp Pro Asp Ser Asp Asn
 1185 1190 1195 1200
 AAA AAA GTC ATA ACC AGA AGA GTT AAT AAC CGT TAT GCG GAG GAT TAT 3648
 Lys Lys Val Ile Thr Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr
 1205 1210 1215
 GAA ATT CCT TCC TCT GTG ACA AGT AAC AGT AAT TAT TCT TGG GGT GAT 3696
 Glu Ile Pro Ser Ser Val Thr Ser Asn Ser Asn Tyr Ser Trp Gly Asp
 1220 1225 1230
 CAC AGT TTA ACC ATG CTT TAT GGT GGT AGT GTT CCT AAT ATT ACT TTT 3744
 His Ser Leu Thr Met Leu Tyr Gly Gly Ser Val Pro Asn Ile Thr Phe
 1235 1240 1245
 GAA TCG GCG GCA GAA GAT TTA AGG CTA TCT ACC AAT ATG GCA TTG AGT 3792
 Glu Ser Ala Ala Glu Asp Leu Arg Leu Ser Thr Asn Met Ala Leu Ser
 1250 1255 1260
 ATT ATT CAT AAT GGA TAT GCG GGA ACC CGC CGT ATA CAA TGT AAT CTT 3840
 Ile Ile His Asn Gly Tyr Ala Gly Thr Arg Arg Ile Gln Cys Asn Leu
 1265 1270 1275 1280
 ATG AAA CAA TAC GCT TCA TTA GGT GAT AAA TTT ATA ATT TAT GAT TCA 3888
 Met Lys Gln Tyr Ala Ser Leu Gly Asp Lys Phe Ile Ile Tyr Asp Ser
 1285 1290 1295
 TCA TTT GAT GAT GCA AAC CGT TTT AAT CTG GTG CCA TTG TTT AAA TTC 3936
 Ser Phe Asp Asp Ala Asn Arg Phe Asn Leu Val Pro Leu Phe Lys Phe
 1300 1305 1310
 GGA AAA GAC GAG AAC TCA GAT GAT AGT ATT TGT ATA TAT AAT GAA AAC 3984
 Gly Lys Asp Glu Asn Ser Asp Asp Ser Ile Cys Ile Tyr Asn Glu Asn
 1315 1320 1325
 CCT TCC TCT GAA GAT AAG AAG TGG TAT TTT TCT TCG AAA GAT GAC AAT 4032
 Pro Ser Ser Glu Asp Lys Lys Trp Tyr Phe Ser Ser Lys Asp Asp Asn
 1330 1335 1340
 AAA ACA GCG GAT TAT AAT GGT GGA ACT CAA TGT ATA GAT GCT GGA ACC 4080
 Lys Thr Ala Asp Tyr Asn Gly Gly Thr Gln Cys Ile Asp Ala Gly Thr
 1345 1350 1355 1360
 AGT AAC AAA GAT TTT TAT TAT AAT CTC CAG GAG ATT GAA GTA ATT AGT 4128
 Ser Asn Lys Asp Phe Tyr Tyr Asn Leu Gln Glu Ile Glu Val Ile Ser
 1365 1370 1375
 GTT ACT GGT GGG TAT TGG TCG AGT TAT AAA ATA TCC AAC CCG ATT AAT 4176
 Val Thr Gly Gly Tyr Trp Ser Ser Tyr Lys Ile Ser Asn Pro Ile Asn
 1380 1385 1390
 ATC AAT ACG GGC ATT GAT AGT GCT AAA GTA AAA GTC ACC GTA AAA GCG 4224
 Ile Asn Thr Gly Ile Asp Ser Ala Lys Val Lys Val Thr Val Lys Ala

1395 1400 1405
 GGT GGT GAC GAT CAA ATC TTT ACT GCT GAT AAT AGT ACC TAT GTT CCT 4272
 Gly Gly Asp Asp Gln Ile Phe Thr Ala Asp Asn Ser Thr Tyr Val Pro
 1410 1415 1420
 CAG CAA CCG GCA CCC AGT TTT GAG GAG ATG ATT TAT CAG TTC AAT AAC 4320
 Gln Gln Pro Ala Pro Ser Phe Glu Glu Met Ile Tyr Gln Phe Asn Asn
 1425 1430 1435 1440
 CTG ACA ATA GAT TGT AAG AAT TTA AAT TTC ATC GAC AAT CAG GCA CAT 4368
 Leu Thr Ile Asp Cys Lys Asn Leu Asn Phe Ile Asp Asn Gln Ala His
 1445 1450 1455
 ATT GAG ATT GAT TTC ACC GCT ACG GCA CAA GAT GGC CGA TTC TTG GGT 4416
 Ile Glu Ile Asp Phe Thr Ala Thr Ala Gln Asp Gly Arg Phe Leu Gly
 1460 1465 1470
 GCA GAA ACT TTT ATT ATC CCG GTA ACT AAA AAA GTT CTC GGT ACT GAG 4464
 Ala Glu Thr Phe Ile Ile Pro Val Thr Lys Lys Val Leu Gly Thr Glu
 1475 1480 1485
 AAC GTG ATT GCG TTA TAT AGC GAA AAT AAC GGT GTT CAA TAT ATG CAA 4512
 Asn Val Ile Ala Leu Tyr Ser Glu Asn Asn Gly Val Gln Tyr Met Gln
 1490 1495 1500
 ATT GGC GCA TAT CGT ACC CGT TTG AAT ACG TTA TTC GCT CAA CAG TTG 4560
 Ile Gly Ala Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Gln Gln Leu
 1505 1510 1515 1520
 GTT AGC CGT GCT AAT CGT GGC ATT GAT GCA GTG CTC AGT ATG GAA ACT 4608
 Val Ser Arg Ala Asn Arg Gly Ile Asp Ala Val Leu Ser Met Glu Thr
 1525 1530 1535
 CAG AAT ATT CAG GAA CCG CAA TTA GGA GCG GGC ACA TAT GTG CAG CTT 4656
 Gln Asn Ile Gln Glu Pro Gln Leu Gly Ala Gly Thr Tyr Val Gln Leu
 1540 1545 1550
 GTG TTG GAT AAA TAT GAT GAG TCT ATT CAT GGC ACT AAT AAA AGC TTT 4704
 Val Leu Asp Lys Tyr Asp Glu Ser Ile His Gly Thr Asn Lys Ser Phe
 1555 1560 1565
 GCT ATT GAA TAT GTT GAT ATA TTT AAA GAG AAC GAT AGT TTT GTG ATT 4752
 Ala Ile Glu Tyr Val Asp Ile Phe Lys Glu Asn Asp Ser Phe Val Ile
 1570 1575 1580
 TAT CAA GGA GAA CTT AGC GAA ACA AGT CAA ACT GTT GTG AAA GTT TTC 4800
 Tyr Gln Gly Glu Leu Ser Glu Thr Ser Gln Thr Val Val Lys Val Phe
 1585 1590 1595 1600
 TTA TCC TAT TTT ATA GAG GCG ACT GGA AAT AAG AAC CAC TTA TGG GTA 4848
 Leu Ser Tyr Phe Ile Glu Ala Thr Gly Asn Lys Asn His Leu Trp Val
 1605 1610 1615
 CGT GCT AAA TAC CAA AAG GAA ACG ACT GAT AAG ATC TTG TTC GAC CGT 4896
 Arg Ala Lys Tyr Gln Lys Glu Thr Thr Asp Lys Ile Leu Phe Asp Arg
 1620 1625 1630
 ACT GAT GAG AAA GAT CCG CAC GGT TGG TTT CTC AGC GAC GAT CAC AAG 4944
 Thr Asp Glu Lys Asp Pro His Gly Trp Phe Leu Ser Asp Asp His Lys

1635 1640 1645
 ACC TTT AGT GGT CTC TCT TCC GCA CAG GCA TTA AAG AAC GAC AGT GAA 4992
 Thr Phe Ser Gly Leu Ser Ser Ala Gln Ala Leu Lys Asn Asp Ser Glu
 1650 1655 1660
 CCG ATG GAT TTC TCT GGC GCC AAT GCT CTC TAT TTC TGG GAA CTG TTC 5040
 Pro Met Asp Phe Ser Gly Ala Asn Ala Leu Tyr Phe Trp Glu Leu Phe
 1665 1670 1675 1680
 TAT TAC ACG CCG ATG ATG ATG GCT CAT CGT TTG TTG CAG GAA CAG AAT 5088
 Tyr Tyr Thr Pro Met Met Met Ala His Arg Leu Leu Gln Glu Gln Asn
 1685 1690 1695
 TTT GAT GCG GCG AAC CAT TGG TTC CGT TAT GTC TGG AGT CCA TCC GGT 5136
 Phe Asp Ala Ala Asn His Trp Phe Arg Tyr Val Trp Ser Pro Ser Gly
 1700 1705 1710
 TAT ATC GTT GAT GGT AAA ATT GCT ATC TAC CAC TGG AAC GTG CGA CCG 5184
 Tyr Ile Val Asp Gly Lys Ile Ala Ile Tyr His Trp Asn Val Arg Pro
 1715 1720 1725
 CTG GAA GAA GAC ACC AGT TGG AAT GCA CAA CAA CTG GAC TCC ACC GAT 5232
 Leu Glu Glu Asp Thr Ser Trp Asn Ala Gln Gln Leu Asp Ser Thr Asp
 1730 1735 1740
 CCA GAT GCT GTA GCC CAA GAT GAT CCG ATG CAC TAC AAG GTG GCT ACC 5280
 Pro Asp Ala Val Ala Gln Asp Asp Pro Met His Tyr Lys Val Ala Thr
 1745 1750 1755 1760
 TTT ATG GCG ACG TTG GAT CTG CTA ATG GCC CGT GGT GAT GCT GCT TAC 5328
 Phe Met Ala Thr Leu Asp Leu Leu Met Ala Arg Gly Asp Ala Ala Tyr
 1765 1770 1775
 CGC CAG TTA GAG CGT GAT ACG TTG GCT GAA GCT AAA ATG TGG TAT ACA 5376
 Arg Gln Leu Glu Arg Asp Thr Leu Ala Glu Ala Lys Met Trp Tyr Thr
 1780 1785 1790
 CAG GCG CTT AAT CTG TTG GGT GAT GAG CCA CAA GTG ATG CTG AGT ACG 5424
 Gln Ala Leu Asn Leu Leu Gly Asp Glu Pro Gln Val Met Leu Ser Thr
 1795 1800 1805
 ACT TGG GCT AAT CCA ACA TTG GGT AAT GCT GCT TCA AAA ACC ACA CAG 5472
 Thr Trp Ala Asn Pro Thr Leu Gly Asn Ala Ala Ser Lys Thr Thr Gln
 1810 1815 1820
 CAG GTT CGT CAG CAA GTG CTT ACC CAG TTG CGT CTC AAT AGC AGG GTA 5520
 Gln Val Arg Gln Gln Val Leu Thr Gln Leu Arg Leu Asn Ser Arg Val
 1825 1830 1835 1840
 AAA ACC CCG TTG 5532
 Lys Thr Pro Leu
 1844

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1844 amino acids
- (B) TYPE: amino acids
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53 (TcbAii):

| Features | From | To | Description |
|----------|------|--------|----------------|
| Peptide1 | 1844 | TcbAii | peptide |
| Fragment | 1 | 11 | (SEQ ID NO:1) |
| Fragment | 978 | 990 | (SEQ ID NO:23) |
| Fragment | 1387 | 1401 | (SEQ ID NO:22) |
| Fragment | 1484 | 1505 | (SEQ ID NO:24) |
| Fragment | 1527 | 1552 | (SEQ ID NO:21) |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Phe | Ile | Gln | Gly | Tyr | Ser | Asp | Leu | Phe | Gly | Asn | Arg | Ala | Asp | Asn | Tyr |
| 1 | | | | 5 | | | | | | 10 | | | | 15 | |
| Ala | Ala | Pro | Gly | Ser | Val | Ala | Ser | Met | Phe | Ser | Pro | Ala | Ala | Tyr | Leu |
| | | | 20 | | | | | 25 | | | | | | 30 | |
| Thr | Glu | Leu | Tyr | Arg | Glu | Ala | Lys | Asn | Leu | His | Asp | Ser | Ser | Ser | Ile |
| | | | 35 | | | | | 40 | | | | | | 45 | |
| Tyr | Tyr | Leu | Asp | Lys | Arg | Arg | Pro | Asp | Leu | Ala | Ser | Leu | Met | Leu | Ser |
| | | | 50 | | | | | 55 | | | | | | 60 | |
| Gln | Lys | Asn | Met | Asp | Glu | Glu | Ile | Ser | Thr | Leu | Ala | Leu | Ser | Asn | Glu |
| 65 | | | | | 70 | | | | | 75 | | | | 80 | |
| Leu | Cys | Leu | Ala | Gly | Ile | Glu | Thr | Lys | Thr | Gly | Lys | Ser | Gln | Asp | Glu |
| | | | | | 85 | | | | | 90 | | | | 95 | |
| Val | Met | Asp | Met | Leu | Ser | Thr | Tyr | Arg | Leu | Ser | Gly | Glu | Thr | Pro | Tyr |
| | | | | | 100 | | | | | 105 | | | | 110 | |
| His | His | Ala | Tyr | Glu | Thr | Val | Arg | Glu | Ile | Val | His | Glu | Arg | Asp | Pro |
| | | | | | 115 | | | | | 120 | | | | 125 | |
| Gly | Phe | Arg | His | Leu | Ser | Gln | Ala | Pro | Ile | Val | Ala | Ala | Lys | Leu | Asp |
| | | | | | 130 | | | | | 135 | | | | 140 | |
| Pro | Val | Thr | Leu | Leu | Gly | Ile | Ser | Ser | His | Ile | Ser | Pro | Glu | Leu | Tyr |
| 145 | | | | | | 150 | | | | | | | | 155 | 160 |
| Asn | Leu | Leu | Ile | Glu | Glu | Ile | Pro | Glu | Lys | Asp | Glu | Ala | Ala | Leu | Asp |
| | | | | | | 165 | | | | | | | | 170 | 175 |
| Thr | Leu | Tyr | Lys | Thr | Asn | Phe | Gly | Asp | Ile | Thr | Thr | Ala | Gln | Leu | Met |
| | | | | | | 180 | | | | | | | | 185 | 190 |
| Ser | Pro | Ser | Tyr | Leu | Ala | Arg | Tyr | Tyr | Gly | Val | Ser | Pro | Glu | Asp | Ile |
| | | | | | | 195 | | | | | | | | 200 | 205 |
| Ala | Tyr | Val | Thr | Thr | Ser | Leu | Ser | His | Val | Gly | Tyr | Ser | Ser | Asp | Ile |
| | | | | | | 210 | | | | | | | | 215 | 220 |
| Leu | Val | Ile | Pro | Leu | Val | Asp | Gly | Val | Gly | Lys | Met | Glu | Val | Val | Arg |
| 225 | | | | | | 230 | | | | | | | | 235 | 240 |
| Val | Thr | Arg | Thr | Pro | Ser | Asp | Asn | Tyr | Thr | Ser | Gln | Thr | Asn | Tyr | Ile |
| | | | | | | 245 | | | | | | | | 250 | 255 |
| Glu | Leu | Tyr | Pro | Gln | Gly | Gly | Asp | Asn | Tyr | Leu | Ile | Lys | Tyr | Asn | Leu |
| | | | | | | 260 | | | | | | | | 265 | 270 |
| Ser | Asn | Ser | Phe | Gly | Leu | Asp | Asp | Phe | Tyr | Leu | Gln | Tyr | Lys | Asp | Gly |

Ile Asp Gln Ile Gln Pro Ala Gln Ile Thr Val Asp Gly Phe Trp Glu
 645 650 655
 Glu Val Gln Thr Thr Pro Thr Ser Leu Lys Val Ile Thr Phe Ala Gln
 660 665 670
 Val Leu Ala Gln Leu Ser Leu Ile Tyr Arg Arg Ile Gly Leu Ser Glu
 675 680 685
 Thr Glu Leu Ser Leu Ile Val Thr Gln Ser Ser Leu Leu Val Ala Gly
 690 695 700
 Lys Ser Ile Leu Asp His Gly Leu Leu Thr Leu Met Ala Leu Glu Gly
 705 710 715 720
 Phe His Thr Trp Val Asn Gly Leu Gly Gln His Ala Ser Leu Ile Leu
 725 730 735
 Ala Ala Leu Lys Asp Gly Ala Leu Thr Val Thr Asp Val Ala Gln Ala
 740 745 750
 Met Asn Lys Glu Glu Ser Leu Leu Gln Met Ala Ala Asn Gln Val Glu
 755 760 765
 Lys Asp Leu Thr Lys Leu Thr Ser Trp Thr Gln Ile Asp Ala Ile Leu
 770 775 780
 Gln Trp Leu Gln Met Ser Ser Ala Leu Ala Val Ser Pro Leu Asp Leu
 785 790 795 800
 Ala Gly Met Met Ala Leu Lys Tyr Gly Ile Asp His Asn Tyr Ala Ala
 805 810 815
 Trp Gln Ala Ala Ala Ala Leu Met Ala Asp His Ala Asn Gln Ala
 820 825 830
 Gln Lys Lys Leu Asp Glu Thr Phe Ser Lys Ala Leu Cys Asn Tyr Tyr
 835 840 845
 Ile Asn Ala Val Val Asp Ser Ala Ala Gly Val Arg Asp Arg Asn Gly
 850 855 860
 Leu Tyr Thr Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Asp Val Ile
 865 870 875 880
 Thr Ser Arg Ile Ala Glu Ala Ile Ala Gly Ile Gln Leu Tyr Val Asn
 885 890 895
 Arg Ala Leu Asn Arg Asp Glu Gly Gln Leu Ala Ser Asp Val Ser Thr
 900 905 910
 Arg Gln Phe Phe Thr Asp Trp Glu Arg Tyr Asn Lys Arg Tyr Ser Thr
 915 920 925
 Trp Ala Gly Val Ser Glu Leu Val Tyr Tyr Pro Glu Asn Tyr Val Asp
 930 935 940
 Pro Thr Gln Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln
 945 950 955 960
 Ser Ile Asn Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe
 965 970 975
 Lys Thr Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile
 980 985 990
 Ser Ala Tyr His Asp Asn Val Asn Val Asp Gln Gly Leu Thr Tyr Phe

Ser Asn Lys Asp Phe Tyr Tyr Asn Leu Gln Glu Ile Glu Val Ile Ser
 1365 1370 1375

Val Thr Gly Gly Tyr Trp Ser Ser Tyr Lys Ile Ser Asn Pro Ile Asn
 1380 1385 1390

Ile Asn Thr Gly Ile Asp Ser Ala Lys Val Lys Val Thr Val Lys Ala
 1395 1400 1405

Gly Gly Asp Asp Gln Ile Phe Thr Ala Asp Asn Ser Thr Tyr Val Pro
 1410 1415 1420

Gln Gln Pro Ala Pro Ser Phe Glu Glu Met Ile Tyr Gln Phe Asn Asn
 1425 1430 1435 1440

Leu Thr Ile Asp Cys Lys Asn Leu Asn Phe Ile Asp Asn Gln Ala His
 1445 1450 1455

Ile Glu Ile Asp Phe Thr Ala Thr Ala Gln Asp Gly Arg Phe Leu Gly
 1460 1465 1470

Ala Glu Thr Phe Ile Ile Pro Val Thr Lys Lys Val Leu Gly Thr Glu
 1475 1480 1485

Asn Val Ile Ala Leu Tyr Ser Glu Asn Asn Gly Val Gln Tyr Met Gln
 1490 1495 1500

Ile Gly Ala Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Gln Gln Leu
 1505 1510 1515 1520

Val Ser Arg Ala Asn Arg Gly Ile Asp Ala Val Leu Ser Met Glu Thr
 1525 1530 1535

Gln Asn Ile Gln Glu Pro Gln Leu Gly Ala Gly Thr Tyr Val Gln Leu
 1540 1545 1550

Val Leu Asp Lys Tyr Asp Glu Ser Ile His Gly Thr Asn Lys Ser Phe
 1555 1560 1565

Ala Ile Glu Tyr Val Asp Ile Phe Lys Glu Asn Asp Ser Phe Val Ile
 1570 1575 1580

Tyr Gln Gly Glu Leu Ser Glu Thr Ser Gln Thr Val Val Lys Val Phe
 1585 1590 1595 1600

Leu Ser Tyr Phe Ile Glu Ala Thr Gly Asn Lys Asn His Leu Trp Val
 1605 1610 1615

Arg Ala Lys Tyr Gln Lys Glu Thr Thr Asp Lys Ile Leu Phe Asp Arg
 1620 1625 1630

Thr Asp Glu Lys Asp Pro His Gly Trp Phe Leu Ser Asp Asp His Lys
 1635 1640 1645

Thr Phe Ser Gly Leu Ser Ser Ala Gln Ala Leu Lys Asn Asp Ser Glu
 1650 1655 1660

Pro Met Asp Phe Ser Gly Ala Asn Ala Leu Tyr Phe Trp Glu Leu Phe
 1665 1670 1675 1680

Tyr Tyr Thr Pro Met Met Met Ala His Arg Leu Leu Gln Glu Gln Asn
 1685 1690 1695

Phe Asp Ala Ala Asn His Trp Phe Arg Tyr Val Trp Ser Pro Ser Gly
 1700 1705 1710

Tyr Ile Val Asp Gly Lys Ile Ala Ile Tyr His Trp Asn Val Arg Pro

1715 1720 1725
 Leu Glu Glu Asp Thr Ser Trp Asn Ala Gln Gln Leu Asp Ser Thr Asp
 1730 1735 1740
 Pro Asp Ala Val Ala Gln Asp Asp Pro Met His Tyr Lys Val Ala Thr
 1745 1750 1755 1760
 Phe Met Ala Thr Leu Asp Leu Leu Met Ala Arg Gly Asp Ala Ala Tyr
 1765 1770 1775
 Arg Gln Leu Glu Arg Asp Thr Leu Ala Glu Ala Lys Met Trp Tyr Thr
 1780 1785 1790
 Gln Ala Leu Asn Leu Leu Gly Asp Glu Pro Gln Val Met Leu Ser Thr
 1795 1800 1805
 Thr Trp Ala Asn Pro Thr Leu Gly Asn Ala Ala Ser Lys Thr Thr Gln
 1810 1815 1820
 Gln Val Arg Gln Gln Val Leu Thr Gln Leu Arg Leu Asn Ser Arg Val
 1825 1830 1835 1840
 Lys Thr Pro Leu
 1844

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1722 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54 (tcbAiii coding region):

CTA GGA ACA GCC AAT TCC CTG ACC GCT TTA TTC CTG CCG CAG GAA AAT 48
 Leu Gly Thr Ala Asn Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn
 1 5 10 15
 AGC AAG CTC AAA GGC TAC TGG CGG ACA CTG GCG CAG CGT ATG TTT AAT 96
 Ser Lys Leu Lys Gly Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn
 20 25 30
 TTA CGT CAT AAT CTG TCG ATT GAC GGC CAG CCG CTC TCC TTG CCG CTG 144
 Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu
 35 40 45
 TAT GCT AAA CCG GCT GAT CCA AAA GCT TTA CTG AGT GCG GCG GTT TCA 192
 Tyr Ala Lys Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser
 50 55 60
 GCT TCT CAA GGG GGA GCC GAC TTG CCG AAG GCG CCG CTG ACT ATT CAC 240
 Ala Ser Gln Gly Gly Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His
 65 70 75 80
 CGC TTC CCT CAA ATG CTA GAA GGG GCA CGG GGC TTG GTT AAC CAG CTT 288
 Arg Phe Pro Gln Met Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu
 85 90 95
 ATA CAG TTC GGT AGT TCA CTA TTG GGG TAC AGT GAG CGT CAG GAT GCG 336

Ile Gln Phe Gly Ser Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala
 100 105 110
 GAA GCT ATG AGT CAA CTA CTG CAA ACC CAA GCC AGC GAG TTA ATA CTG 384
 Glu Ala Met Ser Gln Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu
 115 120 125
 ACC AGT ATT CGT ATG CAG GAT AAC CAA TTG GCA GAG CTG GAT TCG GAA 432
 Thr Ser Ile Arg Met Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu
 130 135 140
 AAA ACC GCC TTG CAA GTC TCT TTA GCT GGA GTG CAA CAA CGG TTT GAC 480
 Lys Thr Ala Leu Gln Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp
 145 150 155 160
 AGC TAT AGC CAA CTG TAT GAG GAG AAC ATC AAC GCA GGT GAG CAG CGA 528
 Ser Tyr Ser Gln Leu Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg
 165 170 175
 GCG CTG GCG TTA CGC TCA GAA TCT GCT ATT GAG TCT CAG GGA GCG CAG 576
 Ala Leu Ala Leu Arg Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln
 180 185 190
 ATT TCC CGT ATG GCA GGC GCG GGT GTT GAT ATG GCA CCA AAT ATC TTC 624
 Ile Ser Arg Met Ala Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe
 195 200 205
 GGC CTG GCT GAT GGC GGC ATG CAT TAT GGT GCT ATT GCC TAT GCC ATC 672
 Gly Leu Ala Asp Gly Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile
 210 215 220
 GCT GAC GGT ATT GAG TTG AGT GCT TCT GCC AAG ATG GTT GAT GCG GAG 720
 Ala Asp Gly Ile Glu Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu
 225 230 235 240
 AAA GTT GCT CAG TCG GAA ATA TAT CGC CGT CGC CGT CAA GAA TGG AAA 768
 Lys Val Ala Gln Ser Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys
 245 250 255
 ATT CAG CGT GAC AAC GCA CAA GCG GAG ATT AAC CAG TTA AAC GCG CAA 816
 Ile Gln Arg Asp Asn Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln
 260 265 270
 CTG GAA TCA CTG TCT ATT CGC CGT GAA GCC GCT GAA ATG CAA AAA GAG 864
 Leu Glu Ser Leu Ser Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu
 275 280 285
 TAC CTG AAA ACC CAG CAA GCT CAG GCG CAG GCA CAA CTT ACT TTC TTA 912
 Tyr Leu Lys Thr Gln Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu
 290 295 300
 AGA AGC AAA TTC AGT AAT CAA GCG TTA TAT AGT TGG TTA CGA GGG CGT 960
 Arg Ser Lys Phe Ser Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg
 305 310 315 320
 TTG TCA GGT ATT TAT TTC CAG TTC TAT GAC TTG GCC GTA TCA CGT TGC 1008
 Leu Ser Gly Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys
 325 330 335
 CTG ATG GCA GAG CAA TCC TAT CAA TGG GAA GCT AAT GAT AAT TCC ATT 1056

Leu Met Ala Glu Gln Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile
 340 345 350
 AGC TTT GTC AAA CCG GGT GCA TGG CAA GGA ACT TAC GCC GGC TTA TTG 1104
 Ser Phe Val Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu
 355 360 365
 TGT GGA GAA GCT TTG ATA CAA AAT CTG GCA CAA ATG GAA GAG GCA TAT 1152
 Cys Gly Glu Ala Leu Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr
 370 375 380
 CTG AAA TGG GAA TCT CGC GCT TTG GAA GTA GAA CGC ACG GTT TCA TTG 1200
 Leu Lys Trp Glu Ser Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu
 385 390 395 400
 GCA GTG GTT TAT GAT TCA CTG GAA GGT AAT GAT CGT TTT AAT TTA GCG 1248
 Ala Val Val Tyr Asp Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala
 405 410 415
 GAA CAA ATA CCT GCA TTA TTG GAT AAG GGG GAG GGA ACA GCA GGA ACT 1296
 Glu Gln Ile Pro Ala Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr
 420 425 430
 AAA GAA AAT GGG TTA TCA TTG GCT AAT GCT ATC CTG TCA GCT TCG GTC 1344
 Lys Glu Asn Gly Leu Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val
 435 440 445
 AAA TTG TCC GAC TTG AAA CTG GGA ACG GAT TAT CCA GAC AGT ATC GTT 1392
 Lys Leu Ser Asp Leu Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val
 450 455 460
 GGT AGC AAC AAG GTT CGT CGT ATT AAG CAA ATC AGT GTT TCG CTA CCT 1440
 Gly Ser Asn Lys Val Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro
 465 470 475 480
 GCA TTG GTT GGG CCT TAT CAG GAT GTT CAG GCT ATG CTC AGC TAT GGT 1488
 Ala Leu Val Gly Pro Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly
 485 490 495
 GGC AGT ACT CAA TTG CCG AAA GGT TGT TCA GCG TTG GCT GTG TCT CAT 1536
 Gly Ser Thr Gln Leu Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His
 500 505 510
 GGT ACC AAT GAT AGT GGT CAG TTC CAG TTG GAT TTC AAT GAC GGC AAA 1584
 Gly Thr Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys
 515 520 525
 TAC CTG CCA TTT GAA GGT ATT GCT CTT GAT GAT CAG GGT ACA CTG AAT 1632
 Tyr Leu Pro Phe Glu Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn
 530 535 540
 CTT CAA TTT CCG AAT GCT ACC GAC AAG CAG AAA GCA ATA TTG CAA ACT 1680
 Leu Gln Phe Pro Asn Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr
 545 550 555 560
 ATG AGC GAT ATT ATT TTG CAT ATT CGT TAT ACC ATC CGT TAA 1722
 Met Ser Asp Ile Ile Leu His Ile Arg Tyr Thr Ile Arg v?
 565 570 573

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 amino acids

(B) TYPE: amino acids

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55 (TcbAiii):

Leu Gly Thr Ala Asn Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn
 1 5 10 15
 Ser Lys Leu Lys Gly Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn
 20 25 30
 Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu
 35 40 45
 Tyr Ala Lys Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser
 50 55 60
 Ala Ser Gln Gly Gly Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His
 65 70 75 80
 Arg Phe Pro Gln Met Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu
 85 90 95
 Ile Gln Phe Gly Ser Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala
 100 105 110
 Glu Ala Met Ser Gln Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu
 115 120 125
 Thr Ser Ile Arg Met Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu
 130 135 140
 Lys Thr Ala Leu Gln Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp
 145 150 155 160
 Ser Tyr Ser Gln Leu Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg
 165 170 175
 Ala Leu Ala Leu Arg Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln
 180 185 190
 Ile Ser Arg Met Ala Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe
 195 200 205
 Gly Leu Ala Asp Gly Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile
 210 215 220
 Ala Asp Gly Ile Glu Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu
 225 230 235 240
 Lys Val Ala Gln Ser Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys
 245 250 255
 Ile Gln Arg Asp Asn Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln
 260 265 270
 Leu Glu Ser Leu Ser Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu
 275 280 285
 Tyr Leu Lys Thr Gln Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu
 290 295 300

Arg Ser Lys Phe Ser Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg
 305 310 315 320
 Leu Ser Gly Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys
 325 330 335
 Leu Met Ala Glu Gln Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile
 340 345 350
 Ser Phe Val Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu
 355 360 365
 Cys Gly Glu Ala Leu Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr
 370 375 380
 Leu Lys Trp Glu Ser Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu
 385 390 395 400
 Ala Val Val Tyr Asp Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala
 405 410 415
 Glu Gln Ile Pro Ala Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr
 420 425 430
 Lys Glu Asn Gly Leu Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val
 435 440 445
 Lys Leu Ser Asp Leu Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val
 450 455 460
 Gly Ser Asn Lys Val Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro
 465 470 475 480
 Ala Leu Val Gly Pro Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly
 485 490 495
 Gly Ser Thr Gln Leu Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His
 500 505 510
 Gly Thr Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys
 515 520 525
 Tyr Leu Pro Phe Glu Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn
 530 535 540
 Leu Gln Phe Pro Asn Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr
 545 550 555 560
 Met Ser Asp Ile Ile Leu His Ile Arg Tyr Thr Ile Arg v ?
 565 570 573

(2) INFORMATION FOR SEQ ID NO:56

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2994 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56 (tccA)

1 ATG AAT CAA CTC GCC AGT CCC CTG ATT TCC CGC ACC GAA GAG ATC CAC 48
 1 Met Asn Gln Leu Ala Ser Pro Leu Ile Ser Arg Thr Glu Glu Ile His 16
 49 AAC TTA CCC GGT AAA TTG ACC GAT CTT GGT TAT ACC TCA GTG TTT GAT 96

17 Asn Leu Pro Gly Lys Leu Thr Asp Leu Gly Tyr Thr Ser Val Phe Asp 32
 97 GTG GTA CGT ATG CCG CGT GAG CGT TTT ATT CGT GAG CAT CGT GCT GAT 144
 33 Val Val Arg Met Pro Arg Glu Arg Phe Ile Arg Glu His Arg Ala Asp 48
 145 CTC GGG CGC AGT GCT GAA AAA ATG TAT GAC CTG GCA GTG GGC TAT GCT 192
 49 Leu Gly Arg Ser Ala Glu Lys Met Tyr Asp Leu Ala Val Gly Tyr Ala 64
 193 CAT CAG GTG TTA CAC CAT TTT CGC CGT AAT TCT CTT AGT GAA GCT GTT 240
 65 His Gln Val Leu His His Phe Arg Arg Asn Ser Leu Ser Glu Ala Val 80
 241 CAG TTT GGC TTG AGA AGT CCG TTC TCC GTA TCA GGC CCG GAT TAC GCC 288
 81 Gln Phe Gly Leu Arg Ser Pro Phe Ser Val Ser Gly Pro Asp Tyr Ala 96
 289 AAT CAG TTT CTT GAT GCA AAC ACG GGT TGG AAA GAT AAA GCA CCA AGT 336
 97 Asn Gln Phe Leu Asp Ala Asn Thr Gly Trp Lys Asp Lys Ala Pro Ser 112
 337 GGA TCA CCG GAA GCC AAT GAT GCG CCG GTA GCC TAT CTG ACT CAT ATT 384
 113 Gly Ser Pro Glu Ala Asn Asp Ala Pro Val Ala Tyr Leu Thr His Ile 128
 385 TAT CAA TTG GCC CTT GAA CAG GAA AAG AAT GGC GCC ACT ACC ATT ATG 432
 129 Tyr Gln Leu Ala Leu Glu Gln Glu Lys Asn Gly Ala Thr Thr Ile Met 144
 433 AAT ACG CTG GCG GAG CGT CGC CCC GAT CTG GGT GCT TTG TTA ATT AAT 480
 145 Asn Thr Leu Ala Glu Arg Arg Pro Asp Leu Gly Ala Leu Leu Ile Asn 160
 481 GAT AAA GCA ATC AAT GAG GTG ATA CCG CAA TTG CAG TTG GTC AAT GAA 528
 161 Asp Lys Ala Ile Asn Glu Val Ile Pro Gln Leu Gln Leu Val Asn Glu 176
 529 ATT CTG TCC AAA GCT ATT CAG AAG AAA CTG AGT TTG ACT GAT CTG GAA 576
 177 Ile Leu Ser Lys Ala Ile Gln Lys Lys Leu Ser Leu Thr Asp Leu Glu 192
 577 GCG GTA AAC GCC AGA CTT TCC ACT ACC CGT TAC CCG AAT AAT CTG CCG 624
 193 Ala Val Asn Ala Arg Leu Ser Thr Thr Arg Tyr Pro Asn Asn Leu Pro 208
 625 TAT CAT TAT GGT CAT CAG CAG ATT CAG ACA GCT CAA TCG GTA TTG GGT 672
 209 Tyr His Tyr Gly His Gln Gln Ile Gln Thr Ala Gln Ser Val Leu Gly 224
 673 ACT ACG TTG CAA GAT ATC ACT TTG CCA CAG ACG CTG GAT CTG CCG CAA 720
 225 Thr Thr Leu Gln Asp Ile Thr Leu Pro Gln Thr Leu Asp Leu Pro Gln 240
 721 AAC TTC TGG GCA ACA GCA AAA GGA AAA CTG AGC GAT ACG ACT GCC AGT 768
 241 Asn Phe Trp Ala Thr Ala Lys Gly Lys Leu Ser Asp Thr Thr Ala Ser 256
 769 GCT TTG ACC CGA CTG CAA ATC ATG GCG AGT CAG TTT TCG CCA GAG CAG 816
 257 Ala Leu Thr Arg Leu Gln Ile Met Ala Ser Gln Phe Ser Pro Glu Gln 272
 817 CAG AAA ATC ATT ACG GAG ACT GTC GGT CAG GAT TTC TAT CAG CTT AAC 864
 273 Gln Lys Ile Ile Thr Glu Thr Val Gly Gln Asp Phe Tyr Gln Leu Asn 288
 865 TAT GGT GAC AGT TCG CTT ACT GTG AAT AGT TTC AGC GAC ATG ACC ATA 912
 289 Tyr Gly Asp Ser Ser Leu Thr Val Asn Ser Phe Ser Asp Met Thr Ile 304
 913 ATG ACT GAT CGA ACA AGT TTG ACT GTA CCC CAG GTA GAA CTG ATG TTG 960
 305 Met Thr Asp Arg Thr Ser Leu Thr Val Pro Gln Val Glu Leu Met Leu 320
 961 TGT TCA ACT GTC GGA GGT TCT ACG GTT GTT AAG TCT GAT AAT GTG AGT
 1008
 321 Cys Ser Thr Val Gly Gly Ser Thr Val Val Lys Ser Asp Asn Val Ser 336
 1009 TCT GGT GAC ACG ACA GCG ACG CCA TTT GCG TAT GGC GCC CGC TTT ATT
 1056
 337 Ser Gly Asp Thr Thr Ala Thr Pro Phe Ala Tyr Gly Ala Arg Phe Ile 352
 1057 CAT GCC GGT AAG CCG GAG GCG ATT ACC CTG AGT CGC AGT GGT GCG GAG
 1104

353 His Ala Gly Lys Pro Glu Ala Ile Thr Leu Ser Arg Ser Gly Ala Glu 368
 1105 GCG CAT TTT GCT CTG ACG GTT AAC AAT CTG ACA GAT GAC AAG TTG GAC
 1152
 369 Ala His Phe Ala Leu Thr Val Asn Asn Leu Thr Asp Asp Lys Leu Asp 384
 1153 CGT ATT AAC CGC ACA GTG CGC CTG CAA AAA TGG CTG AAT CTG CCT TAT
 1200
 385 Arg Ile Asn Arg Thr Val Arg Leu Gln Lys Trp Leu Asn Leu Pro Tyr 400
 1201 GAG GAT ATT GAC CTG TTA GTG ACT TCT GCT ATG GAT GCG GAA ACA GGA
 1248
 401 Glu Asp Ile Asp Leu Leu Val Thr Ser Ala Met Asp Ala Glu Thr Gly 416
 1249 AAT ACC GCG CTG TCG ATG AAC GAC AAT ACG CTG CGT ATG TTG GGA GTG
 1296
 417 Asn Thr Ala Leu Ser Met Asn Asp Asn Thr Leu Arg Met Leu Gly Val 432
 1297 TTC AAA CAT TAT CAG GCG AAG TAT GGT GTT AGC GCT AAA CAA TTT GCT
 1344
 433 Phe Lys His Tyr Gln Ala Lys Tyr Gly Val Ser Ala Lys Gln Phe Ala 448
 1345 GGC TGG CTG CGC GTA GTG GCC CCG TTT GCC ATT ACA CCG GCA ACG CCG
 1392
 449 Gly Trp Leu Arg Val Val Ala Pro Phe Ala Ile Thr Pro Ala Thr Pro 464
 1393 TTT TTA GAC CAA GTG TTT AAC TCC GTC GGC ACC TTT GAT ACA CCG TTT
 1440
 465 Phe Leu Asp Gln Val Phe Asn Ser Val Gly Thr Phe Asp Thr Pro Phe 480
 1441 GTG ATA GAT AAT CAG GAT TTT GTC TAT ACA TTG ACC ACC GGG GGC GAT
 1488
 481 Val Ile Asp Asn Gln Asp Phe Val Tyr Thr Leu Thr Thr Gly Gly Asp 496
 1489 GGG GCG CGT GTT AAG CAT ATC AGC ACG GCA CTG GGC CTC AAT CAT CGT
 1536
 497 Gly Ala Arg Val Lys His Ile Ser Thr Ala Leu Gly Leu Asn His Arg 512
 1537 CAG TTC CTG TTA TTG GCG GAT AAT ATT GCC CGT CAA CAG GGG AAT GTC
 1584
 513 Gln Phe Leu Leu Leu Ala Asp Asn Ile Ala Arg Gln Gln Gly Asn Val 528
 1585 ACG CAA AGC ACA CTC AAC TGT AAT CTG TTT GTG GTG TCA GCT TTC TAC
 1632
 529 Thr Gln Ser Thr Leu Asn Cys Asn Leu Phe Val Val Ser Ala Phe Tyr 544
 1633 CGT CTG GCT AAT TTG GCG CGC ACA TTG GGG ATA AAT CCA GAG TCT TTC
 1680
 545 Arg Leu Ala Asn Leu Ala Arg Thr Leu Gly Ile Asn Pro Glu Ser Phe 560
 1681 TGT GCC TTG GTT GAT CGA TTA GAT GCA GGT ACA GGC ATC GTC TGG CAG
 1728
 561 Cys Ala Leu Val Asp Arg Leu Asp Ala Gly Thr Gly Ile Val Trp Gln 576
 1729 CAA TTG GCA GGG AAA CCC ACA ATC ACG GTA CCA CAA AAA GAT TCC CCG
 1776
 577 Gln Leu Ala Gly Lys Pro Thr Ile Thr Val Pro Gln Lys Asp Ser Pro 592
 1777 CTG GCG GCG GAT ATT CTG AGT TTG CTG CAA GCG CTA AGT GCG ATT GCT
 1824

593 Leu Ala Ala Asp Ile Leu Ser Leu Leu Gln Ala Leu Ser Ala Ile Ala 608
 1825 CAA TGG CAA CAA CAG CAC GAT TTA GAA TTT TCA GCA CTG CTT TTG CTG
 1872

609 Gln Trp Gln Gln Gln His Asp Leu Glu Phe Ser Ala Leu Leu Leu Leu 624
 1873 TTG AGT GAC AAC CCT ATT TCT ACC TCG CAG GGC ACT GAC GAT CAA TTG
 1920

625 Leu Ser Asp Asn Pro Ile Ser Thr Ser Gln Gly Thr Asp Asp Gln Leu 640
 1921 AAC TTT ATC CGT CAA GTG TGG CAG AAC CTA GGC AGT ACG TTT GTG GGT
 1968

641 Asn Phe Ile Arg Gln Val Trp Gln Asn Leu Gly Ser Thr Phe Val Gly 656
 1969 GCA ACA TTG TTG TCC CGC AGT GGG GCA CCA TTA GTC GAT ACC AAC GGC
 2016

657 Ala Thr Leu Leu Ser Arg Ser Gly Ala Pro Leu Val Asp Thr Asn Gly 672
 2017 CAC GCT ATT GAC TGG TTT GCT CTG CTC TCA GCA GGT AAT AGT CCG CTT
 2064

673 His Ala Ile Asp Trp Phe Ala Leu Leu Ser Ala Gly Asn Ser Pro Leu 688
 2065 ATC GAT AAG GTT GGT CTG GTG ACT GAT GCT GGC ATA CAA AGT GTT ATA
 2112

689 Ile Asp Lys Val Gly Leu Val Thr Asp Ala Gly Ile Gln Ser Val Ile 704
 2113 GCA ACG GTG GTC AAT ACA CAA AGC TTA TCT GAT GAA GAT AAG AAG CTG
 2160

705 Ala Thr Val Val Asn Thr Gln Ser Leu Ser Asp Glu Asp Lys Lys Leu 720
 2161 GCA ATC ACT ACT CTG ACT AAT ACG TTG AAT CAG GTA CAG AAA ACT CAA
 2208

721 Ala Ile Thr Thr Leu Thr Asn Thr Leu Asn Gln Val Gln Lys Thr Gln 736
 2209 CAG GGC GTG GCC GTC AGT CTG TTG GCG CAG ACT CTG AAC GTG AGT CAG
 2256

737 Gln Gly Val Ala Val Ser Leu Leu Ala Gln Thr Leu Asn Val Ser Gln 752
 2257 TCA CTG CCT GCG TTA TTG TTG CGC TGG AGT GGA CAA ACA ACC TAC CAG
 2304

753 Ser Leu Pro Ala Leu Leu Leu Arg Trp Ser Gly Gln Thr Thr Tyr Gln 768
 2305 TGG TTG AGT GCG ACT TGG GCA TTG AAG GAT GCC GTT AAG ACT GCC GCC
 2352

769 Trp Leu Ser Ala Thr Trp Ala Leu Lys Asp Ala Val Lys Thr Ala Ala 784
 2353 GAT ATT CCC GCT GAC TAT CTG CGT CAA TTA CGT GAA GTG GTA CGC CGC
 2400

785 Asp Ile Pro Ala Asp Tyr Leu Arg Gln Leu Arg Glu Val Val Arg Arg 800
 2401 TCC TTG TTG ACC CAA CAA TTC ACG CTG AGT CCT GCA ATG GTG CAA ACC
 2448

801 Ser Leu Leu Thr Gln Gln Phe Thr Leu Ser Pro Ala Met Val Gln Thr 816
 2449 TTG CTG GAC TAT CCA GCC TAT TTT GGC GCT TCC GCA GAA ACA GTG ACC
 2496

817 Leu Leu Asp Tyr Pro Ala Tyr Phe Gly Ala Ser Ala Glu Thr Val Thr 832
 2497 GAT ATC AGT TTG TGG ATG CTT TAT ACC CTG AGC TGT TAT AGC GAT TTA
 2544

833 Asp Ile Ser Leu Trp Met Leu Tyr Thr Leu Ser Cys Tyr Ser Asp Leu 848
 2545 TTG CTC CAA ATG GGT GAA GCT GGT GGT ACC GAA GAT GAT GTA CTG GCC
 2592
 849 Leu Leu Gln Met Gly Glu Ala Gly Gly Thr Glu Asp Asp Val Leu Ala 864
 2593 TAC TTA CGC ACA GCT AAT GCT ACC ACA CCG TTG AGC CAA TCT GAT GCT
 2640
 865 Tyr Leu Arg Thr Ala Asn Ala Thr Thr Pro Leu Ser Gln Ser Asp Ala 880
 2641 GCA CAG ACG TTG GCA ACG CTA TTG GGT TGG GAG GTT AAC GAG TTG CAA
 2688
 881 Ala Gln Thr Leu Ala Thr Leu Leu Gly Trp Glu Val Asn Glu Leu Gln 896
 2689 GCC GCT TGG TCG GTA TTG GGC GGG ATT GCC AAA ACC ACA CCG CAA CTG
 2736
 897 Ala Ala Trp Ser Val Leu Gly Gly Ile Ala Lys Thr Thr Pro Gln Leu 912
 2737 GAT GCG CTT CTG CGT TTG CAA CAG GCA CAG AAC CAA ACT GGT CTT GGC
 2784
 913 Asp Ala Leu Leu Arg Leu Gln Gln Ala Gln Asn Gln Thr Gly Leu Gly 928
 2785 GTT ACA CAG CAA CAG CAA GGC TAT CTC CTG AGT CGT GAC AGT GAT TAT
 2832
 929 Val Thr Gln Gln Gln Gln Gly Tyr Leu Leu Ser Arg Asp Ser Asp Tyr 944
 2833 ACC CTT TGG CAA AGC ACC GGT CAG GCG CTG GTG GCT GGC GTA TCC CAT
 2880
 945 Thr Leu Trp Gln Ser Thr Gly Gln Ala Leu Val Ala Gly Val Ser His 960
 2881 GTC AAG GGC AGT AAC TGA GCATGGCAGA GCTCACTACC TGAGTGGATT TGATTT
 2934
 961 Val Lys Gly Ser Asn End 965
 2935 TTCCGTATGG CCTAATGAGG CTATTCTAA ACCGCCATTT AAGTAAGGCA GATAATTATG
 2994

(2) INFORMATION FOR SEQ ID NO:57

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 965 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57 (TccA peptide)

| Features | From | To | Description | | | | | | | | | | | | | | |
|----------|------|-----|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 10 | SEQ ID NO:8 | | | | | | | | | | | | | | |
| 1 | Met | Asn | Gln | Leu | Ala | Ser | Pro | Leu | Ile | Ser | Arg | Thr | Glu | Glu | Ile | His | 16 |
| 17 | Asn | Leu | Pro | Gly | Lys | Leu | Thr | Asp | Leu | Gly | Tyr | Thr | Ser | Val | Phe | Asp | 32 |
| 33 | Val | Val | Arg | Met | Pro | Arg | Glu | Arg | Phe | Ile | Arg | Glu | His | Arg | Ala | Asp | 48 |
| 49 | Leu | Gly | Arg | Ser | Ala | Glu | Lys | Met | Tyr | Asp | Leu | Ala | Val | Gly | Tyr | Ala | 64 |
| 65 | His | Gln | Val | Leu | His | His | Phe | Arg | Arg | Asn | Ser | Leu | Ser | Glu | Ala | Val | 80 |
| 81 | Gln | Phe | Gly | Leu | Arg | Ser | Pro | Phe | Ser | Val | Ser | Gly | Pro | Asp | Tyr | Ala | 96 |
| 97 | Asn | Gln | Phe | Leu | Asp | Ala | Asn | Thr | Gly | Trp | Lys | Asp | Lys | Ala | Pro | Ser | 112 |
| 113 | Gly | Ser | Pro | Glu | Ala | Asn | Asp | Ala | Pro | Val | Ala | Tyr | Leu | Thr | His | Ile | 128 |
| 129 | Tyr | Gln | Leu | Ala | Leu | Glu | Gln | Glu | Lys | Asn | Gly | Ala | Thr | Thr | Ile | Met | 144 |

145 Asn Thr Leu Ala Glu Arg Arg Pro Asp Leu Gly Ala Leu Leu Ile Asn 160
 161 Asp Lys Ala Ile Asn Glu Val Ile Pro Gln Leu Gln Leu Val Asn Glu 176
 177 Ile Leu Ser Lys Ala Ile Gln Lys Lys Leu Ser Leu Thr Asp Leu Glu 192
 193 Ala Val Asn Ala Arg Leu Ser Thr Thr Arg Tyr Pro Asn Asn Leu Pro 208
 209 Tyr His Tyr Gly His Gln Gln Ile Gln Thr Ala Gln Ser Val Leu Gly 224
 225 Thr Thr Leu Gln Asp Ile Thr Leu Pro Gln Thr Leu Asp Leu Pro Gln 240
 241 Asn Phe Trp Ala Thr Ala Lys Gly Lys Leu Ser Asp Thr Thr Ala Ser 256
 257 Ala Leu Thr Arg Leu Gln Ile Met Ala Ser Gln Phe Ser Pro Glu Gln 272
 273 Gln Lys Ile Ile Thr Glu Thr Val Gly Gln Asp Phe Tyr Gln Leu Asn 288
 289 Tyr Gly Asp Ser Ser Leu Thr Val Asn Ser Phe Ser Asp Met Thr Ile 304
 305 Met Thr Asp Arg Thr Ser Leu Thr Val Pro Gln Val Glu Leu Met Leu 320
 321 Cys Ser Thr Val Gly Gly Ser Thr Val Val Lys Ser Asp Asn Val Ser 336
 337 Ser Gly Asp Thr Thr Ala Thr Pro Phe Ala Tyr Gly Ala Arg Phe Ile 352
 353 His Ala Gly Lys Pro Glu Ala Ile Thr Leu Ser Arg Ser Gly Ala Glu 368
 369 Ala His Phe Ala Leu Thr Val Asn Asn Leu Thr Asp Asp Lys Leu Asp 384
 385 Arg Ile Asn Arg Thr Val Arg Leu Gln Lys Trp Leu Asn Leu Pro Tyr 400
 401 Glu Asp Ile Asp Leu Leu Val Thr Ser Ala Met Asp Ala Glu Thr Gly 416
 417 Asn Thr Ala Leu Ser Met Asn Asp Asn Thr Leu Arg Met Leu Gly Val 432
 433 Phe Lys His Tyr Gln Ala Lys Tyr Gly Val Ser Ala Lys Gln Phe Ala 448
 449 Gly Trp Leu Arg Val Val Ala Pro Phe Ala Ile Thr Pro Ala Thr Pro 464
 465 Phe Leu Asp Gln Val Phe Asn Ser Val Gly Thr Phe Asp Thr Pro Phe 480
 481 Val Ile Asp Asn Gln Asp Phe Val Tyr Thr Leu Thr Thr Gly Gly Asp 496
 497 Gly Ala Arg Val Lys His Ile Ser Thr Ala Leu Gly Leu Asn His Arg 512
 513 Gln Phe Leu Leu Leu Ala Asp Asn Ile Ala Arg Gln Gln Gly Asn Val 528
 529 Thr Gln Ser Thr Leu Asn Cys Asn Leu Phe Val Val Ser Ala Phe Tyr 544
 545 Arg Leu Ala Asn Leu Ala Arg Thr Leu Gly Ile Asn Pro Glu Ser Phe 560
 561 Cys Ala Leu Val Asp Arg Leu Asp Ala Gly Thr Gly Ile Val Trp Gln 576
 577 Gln Leu Ala Gly Lys Pro Thr Ile Thr Val Pro Gln Lys Asp Ser Pro 592
 593 Leu Ala Ala Asp Ile Leu Ser Leu Leu Gln Ala Leu Ser Ala Ile Ala 608
 609 Gln Trp Gln Gln Gln His Asp Leu Glu Phe Ser Ala Leu Leu Leu Leu 624
 625 Leu Ser Asp Asn Pro Ile Ser Thr Ser Gln Gly Thr Asp Asp Gln Leu 640
 641 Asn Phe Ile Arg Gln Val Trp Gln Asn Leu Gly Ser Thr Phe Val Gly 656
 657 Ala Thr Leu Leu Ser Arg Ser Gly Ala Pro Leu Val Asp Thr Asn Gly 672
 673 His Ala Ile Asp Trp Phe Ala Leu Leu Ser Ala Gly Asn Ser Pro Leu 688
 689 Ile Asp Lys Val Gly Leu Val Thr Asp Ala Gly Ile Gln Ser Val Ile 704
 705 Ala Thr Val Val Asn Thr Gln Ser Leu Ser Asp Glu Asp Lys Lys Leu 720
 721 Ala Ile Thr Thr Leu Thr Asn Thr Leu Asn Gln Val Gln Lys Thr Gln 736
 737 Gln Gly Val Ala Val Ser Leu Leu Ala Gln Thr Leu Asn Val Ser Gln 752
 753 Ser Leu Pro Ala Leu Leu Leu Arg Trp Ser Gly Gln Thr Thr Tyr Gln 768
 769 Trp Leu Ser Ala Thr Trp Ala Leu Lys Asp Ala Val Lys Thr Ala Ala 784
 785 Asp Ile Pro Ala Asp Tyr Leu Arg Gln Leu Arg Glu Val Val Arg Arg 800
 801 Ser Leu Leu Thr Gln Gln Phe Thr Leu Ser Pro Ala Met Val Gln Thr 816
 817 Leu Leu Asp Tyr Pro Ala Tyr Phe Gly Ala Ser Ala Glu Thr Val Thr 832
 833 Asp Ile Ser Leu Trp Met Leu Tyr Thr Leu Ser Cys Tyr Ser Asp Leu 848
 849 Leu Leu Gln Met Gly Glu Ala Gly Gly Thr Glu Asp Asp Val Leu Ala 864

865 Tyr Leu Arg Thr Ala Asn Ala Thr Thr Pro Leu Ser Gln Ser Asp Ala 880
 881 Ala Gln Thr Leu Ala Thr Leu Leu Gly Trp Glu Val Asn Glu Leu Gln 896
 897 Ala Ala Trp Ser Val Leu Gly Gly Ile Ala Lys Thr Thr Pro Gln Leu 912
 913 Asp Ala Leu Leu Arg Leu Gln Gln Ala Gln Asn Gln Thr Gly Leu Gly 928
 929 Val Thr Gln Gln Gln Gln Gly Tyr Leu Leu Ser Arg Asp Ser Asp Tyr 944
 945 Thr Leu Trp Gln Ser Thr Gly Gln Ala Leu Val Ala Gly Val Ser His 960
 961 Val Lys Gly Ser Asn 965

(2) INFORMATION FOR SEQ ID NO:58

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4932 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58 (tccB)

1 ATG TTA TCG ACA ATG GAA AAA CAA CTG AAT GAA TCC CAG CGT GAT GCG 48
 1 Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu Ser Gln Arg Asp Ala 16
 49 TTG GTG ACT GGC TAT ATG AAT TTT GTG GCG CCG ACG TTG AAA GGC GTC 96
 17 Leu Val Thr Gly Tyr Met Asn Phe Val Ala Pro Thr Leu Lys Gly Val 32
 97 AGT GGT CAG CCG GTG ACG GTG GAA GAT TTA TAC GAA TAT TTG CTG ATT 144
 33 Ser Gly Gln Pro Val Thr Val Glu Asp Leu Tyr Glu Tyr Leu Leu Ile 48
 145 GAC CCG GAA GTG GCT GAT GAG GTT GAG ACG AGT CGG GTA GCA CAA GCG 192
 49 Asp Pro Glu Val Ala Asp Glu Val Glu Thr Ser Arg Val Ala Gln Ala 64
 193 ATT GCC AGC ATA CAG CAA TAT ATG ACT CGT CTG GTC AAC GGC TCT GAA 240
 65 Ile Ala Ser Ile Gln Gln Tyr Met Thr Arg Leu Val Asn Gly Ser Glu 80
 241 CCG GGG CGT CAG GCG ATG GAG CCT TCT ACA GCT AAC GAA TGG CGT GAT 288
 81 Pro Gly Arg Gln Ala Met Glu Pro Ser Thr Ala Asn Glu Trp Arg Asp 96
 289 AAT GAT AAC CAA TAT GCT ATC TGG GCT GCG GGG GCT GAG GTT CGA AAT 336
 97 Asn Asp Asn Gln Tyr Ala Ile Trp Ala Ala Gly Ala Glu Val Arg Asn 112
 337 TAC GCT GAA AAC TAT ATT TCA CCC ATC ACC CGG CAG GAA AAA AGC CAT 384
 113 Tyr Ala Glu Asn Tyr Ile Ser Pro Ile Thr Arg Gln Glu Lys Ser His 128
 385 TAT TTC TCG GAG CTG GAG ACG ACT TTA AAT CAG AAT CGA CTC GAT CCG 432
 129 Tyr Phe Ser Glu Leu Glu Thr Thr Leu Asn Gln Asn Arg Leu Asp Pro 144
 433 GAT CGT GTG CAG GAT GCT GTT TTG GCG TAT CTC AAT GAG TTT GAG GCA 480
 145 Asp Arg Val Gln Asp Ala Val Leu Ala Tyr Leu Asn Glu Phe Glu Ala 160
 481 GTG AGT AAT CTA TAT GTG CTC AGT GGT TAT ATT AAT CAG GAT AAA TTT 528
 161 Val Ser Asn Leu Tyr Val Leu Ser Gly Tyr Ile Asn Gln Asp Lys Phe 176
 529 GAC CAA GCT ATC TAC TAC TTT ATT GGT CGC ACT ACC ACT AAA CCG TAT 576
 177 Asp Gln Ala Ile Tyr Tyr Phe Ile Gly Arg Thr Thr Thr Lys Pro Tyr 192
 577 CGC TAC TAC TGG CGT CAG ATG GAT TTG AGT AAG AAC CGT CAA GAT CCG 624
 193 Arg Tyr Tyr Trp Arg Gln Met Asp Leu Ser Lys Asn Arg Gln Asp Pro 208
 625 GCA GGG AAT CCG GTG ACG CCA AAT TGC TGG AAT GAT TGG CAG GAA ATC 672
 209 Ala Gly Asn Pro Val Thr Pro Asn Cys Trp Asn Asp Trp Gln Glu Ile 224
 673 ACT TTG CCG CTG TCT GGT GAT ACG GTG CTG GAG CAT ACA GTT CGC CCG 720
 225 Thr Leu Pro Leu Ser Gly Asp Thr Val Leu Glu His Thr Val Arg Pro 240

721 GTA TTT TAT AAT GAT CGA CTA TAT GTG GCT TGG GTT GAG CGT GAC CCG 768
 241 Val Phe Tyr Asn Asp Arg Leu Tyr Val Ala Trp Val Glu Arg Asp Pro 256
 769 GCA GTA CAG AAG GAT GCT GAC GGT AAA AAC ATC GGT AAA ACC CAT GCC 816
 257 Ala Val Gln Lys Asp Ala Asp Gly Lys Asn Ile Gly Lys Thr His Ala 272
 817 TAC AAC ATA AAG TTT GGT TAT AAA CGT TAT GAT GAT ACT TGG ACA GCG 864
 273 Tyr Asn Ile Lys Phe Gly Tyr Lys Arg Tyr Asp Asp Thr Trp Thr Ala 288
 865 CCG AAT ACG ACC ACG TTA ATG ACA CAA CAA GCA GGG GAA AGT TCA GAA 912
 289 Pro Asn Thr Thr Thr Leu Met Thr Gln Gln Ala Gly Glu Ser Ser Glu 304
 913 ACA CAG CGA TCC AGC CTG CTG ATT GAT GAA TCT AGC ACC ACA TTG CGC 960
 305 Thr Gln Arg Ser Ser Leu Leu Ile Asp Glu Ser Ser Thr Thr Leu Arg 320
 961 CAA GTT AAT CTG TTG GCT ACC ACC GAT TTT AGT ATC GAT CCG ACG GAG
 1008
 321 Gln Val Asn Leu Leu Ala Thr Thr Asp Phe Ser Ile Asp Pro Thr Glu 336
 1009 GAA ACG GAC AGT AAC CCG TAT GGC CGC CTA ATG TTG GGG GTG TTT GTC
 1056
 337 Glu Thr Asp Ser Asn Pro Tyr Gly Arg Leu Met Leu Gly Val Phe Val 352
 1057 CGT CAA TTT GAA GGT GAT GGG GCC AAT AGA AAA AAT AAA CCC GTT GTT
 1104
 353 Arg Gln Phe Glu Gly Asp Gly Ala Asn Arg Lys Asn Lys Pro Val Val 368
 1105 TAT GGT TAT CTC TAT TGT GAC TCA GCT TTC AAT CGT CAT GTT CTC AGG
 1152
 369 Tyr Gly Tyr Leu Tyr Cys Asp Ser Ala Phe Asn Arg His Val Leu Arg 384
 1153 CCG TTA AGT AAG AAC TTT TTG TTC AGT ACT TAC CGT GAT GAA ACG GAT
 1200
 385 Pro Leu Ser Lys Asn Phe Leu Phe Ser Thr Tyr Arg Asp Glu Thr Asp 400
 1201 GGT CAA AAC AGC TTG CAA TTT GCG GTA TAC GAT AAA AAG TAT GTA ATT
 1248
 401 Gly Gln Asn Ser Leu Gln Phe Ala Val Tyr Asp Lys Lys Tyr Val Ile 416
 1249 ACT AAG GTT GTT ACA GGT GCA ACG GAA GAT CCC GAA AAT ACA GGA TGG
 1296
 417 Thr Lys Val Val Thr Gly Ala Thr Glu Asp Pro Glu Asn Thr Gly Trp 432
 1297 GTA AGT AAA GTT GAT GAC TTG AAA CAA GGC ACT ACT GGG GCC TAT GTG
 1344
 433 Val Ser Lys Val Asp Asp Leu Lys Gln Gly Thr Thr Gly Ala Tyr Val 448
 1345 TAT ATC GAT CAA GAT GGC CTG ACG CTT CAT ATA CAA ACC ACA ACT AAT
 1392
 449 Tyr Ile Asp Gln Asp Gly Leu Thr Leu His Ile Gln Thr Thr Thr Asn 464
 1393 GGG GAT TTT ATT AAC CGT CAT ACG TTT GGA TAT AAC GAT CTT GTA TAT
 1440
 465 Gly Asp Phe Ile Asn Arg His Thr Phe Gly Tyr Asn Asp Leu Val Tyr 480
 1441 GAT TCT AAG TCT GGT TAT GGT TTC ACG TGG TCA GGA AAT GAA GGT TTT
 1488
 481 Asp Ser Lys Ser Gly Tyr Gly Phe Thr Trp Ser Gly Asn Glu Gly Phe 496
 1489 TAT CTG GAT TAC CAT GAT GGA AAT TAT TAC ACC TTT CAT AAT GCA ATA
 1536

497 Tyr Leu Asp Tyr His Asp Gly Asn Tyr Tyr Thr Phe His Asn Ala Ile 512
 1537 ATC AAC TAC TAT CCG TCT GGA TAT GGT GGT GGA TCT GTT CCT AAT GGA
 1584

513 Ile Asn Tyr Tyr Pro Ser Gly Tyr Gly Gly Gly Ser Val Pro Asn Gly 528
 1585 ACG TGG GCG TTA GAG CAA AGG ATT AAT GAG GGA TGG GCT ATT GCT CCC
 1632

529 Thr Trp Ala Leu Glu Gln Arg Ile Asn Glu Gly Trp Ala Ile Ala Pro 544
 1633 CTG CTT GAT ACT CTC CAT ACT GTT ACT GTG AAG GGC AGT TAT ATC GCT
 1680

545 Leu Leu Asp Thr Leu His Thr Val Thr Val Lys Gly Ser Tyr Ile Ala 560
 1681 TGG GAA GGG GAA ACA CCT ACC GGT TAT AAT CTG TAT ATT CCA GAT GGT
 1728

561 Trp Glu Gly Glu Thr Pro Thr Gly Tyr Asn Leu Tyr Ile Pro Asp Gly 576
 1729 ACC GTG TTG CTA GAT TGG TTT GAT AAA ATA AAT TTT GCT ATT GGT CTT
 1776

577 Thr Val Leu Leu Asp Trp Phe Asp Lys Ile Asn Phe Ala Ile Gly Leu 592
 1777 AAT AAG CTT GAG TCT GTA TTT ACG TCG CCA GAT TGG CCA ACA CTA ACC
 1824

593 Asn Lys Leu Glu Ser Val Phe Thr Ser Pro Asp Trp Pro Thr Leu Thr 608
 1825 ACT ATC AAA AAT TTC AGT AAA ATC GCC GAT AAC CGC AAA TTC TAT CAG
 1872

609 Thr Ile Lys Asn Phe Ser Lys Ile Ala Asp Asn Arg Lys Phe Tyr Gln 624
 1873 GAA ATC AAT GCT GAG ACG GCG GAT GGA CGC AAC CTG TTT AAA CGT TAC
 1920

625 Glu Ile Asn Ala Glu Thr Ala Asp Gly Arg Asn Leu Phe Lys Arg Tyr 640
 1921 AGT ACT CAA ACT TTC GGA CTT ACC AGC GGT GCG ACT TAT TCT ACA ACT
 1968

641 Ser Thr Gln Thr Phe Gly Leu Thr Ser Gly Ala Thr Tyr Ser Thr Thr 656
 1969 TAT ACT TTG TCT GAG GCG GAT TTC TCC ACT GAT CCG GAC AAA AAC TAC
 2016

657 Tyr Thr Leu Ser Glu Ala Asp Phe Ser Thr Asp Pro Asp Lys Asn Tyr 672
 2017 CTA CAG GTT TGT TTG AAT GTC GTG TGG GAT CAT TAT GAC CGC CCG TCA
 2064

673 Leu Gln Val Cys Leu Asn Val Val Trp Asp His Tyr Asp Arg Pro Ser 688
 2065 GGG AAA AAA GGG GCT TAT TCT TGG GTC AGT AAG TGG TTT AAC GTC TAT
 2112

689 Gly Lys Lys Gly Ala Tyr Ser Trp Val Ser Lys Trp Phe Asn Val Tyr 704
 2113 GTT GCG TTG CAA GAT AGC AAA GCT CCG GAT GCC ATT CCT CGA TTA GTT
 2160

705 Val Ala Leu Gln Asp Ser Lys Ala Pro Asp Ala Ile Pro Arg Leu Val 720
 2161 TCC CGT TAC GAT AGT AAA CGT GGT CTG GTG CAA TAT CTG GAC TTC TGG
 2208

721 Ser Arg Tyr Asp Ser Lys Arg Gly Leu Val Gln Tyr Leu Asp Phe Trp 736
 2209 ACC TCA TCA TTA CCC GCG AAA ACC CGT CTT AAC ACC ACC TTT GTG CGT
 2256

737 Thr Ser Ser Leu Pro Ala Lys Thr Arg Leu Asn Thr Thr Phe Val Arg 752
 2257 ACT TTG ATT GAG AAG GCT AAT CTG GGG CTG GAT AGT TTG CTG GAT TAC
 2304

753 Thr Leu Ile Glu Lys Ala Asn Leu Gly Leu Asp Ser Leu Leu Asp Tyr 768
 2305 ACC TTG CAG GCA GAT CCT TCT CTG GAA GCA GAT TTA GTG ACT GAC GGC
 2352

769 Thr Leu Gln Ala Asp Pro Ser Leu Glu Ala Asp Leu Val Thr Asp Gly 784
 2353 AAA AGC GAA CCA ATG GAC TTT AAT GGT TCA AAC GGT CTC TAT TTC TGG
 2400

785 Lys Ser Glu Pro Met Asp Phe Asn Gly Ser Asn Gly Leu Tyr Phe Trp 800
 2401 GAA TTG TTC TTT CAC CTG CCG TTT TTG GTT GCT ACA CGC TTT GCC AAC
 2448

801 Glu Leu Phe Phe His Leu Pro Phe Leu Val Ala Thr Arg Phe Ala Asn 816
 2449 GAA CAG CAA TTT TCG CCG GCA CAA AAG AGT TTG CAT TAC ATC TTT GAC
 2496

817 Glu Gln Gln Phe Ser Pro Ala Gln Lys Ser Leu His Tyr Ile Phe Asp 832
 2497 CCG GCG ATG AAA AAC AAG CCA CAC AAT GCC CCG GCT TAT TGG AAT GTA
 2544

833 Pro Ala Met Lys Asn Lys Pro His Asn Ala Pro Ala Tyr Trp Asn Val 848
 2545 CGT CCG TTG GTT GAA GGA AAC AGC GAT TTG TCA CGT CAT TTG GAC GAT
 2592

849 Arg Pro Leu Val Glu Gly Asn Ser Asp Leu Ser Arg His Leu Asp Asp 864
 2593 TCT ATA GAC CCA GAT ACT CAA GCT TAT GCT CAT CCG GTG ATA TAC CAG
 2640

865 Ser Ile Asp Pro Asp Thr Gln Ala Tyr Ala His Pro Val Ile Tyr Gln 880
 2641 AAA GCG GTG TTT ATT GCC TAT GTC AGT AAC CTG ATT GCT CAG GGA GAT
 2688

881 Lys Ala Val Phe Ile Ala Tyr Val Ser Asn Leu Ile Ala Gln Gly Asp 896
 2689 ATG TGG TAT CGC CAA TTG ACT CGT GAC GGT CTG ACT CAG GCC CGT GTC
 2736

897 Met Trp Tyr Arg Gln Leu Thr Arg Asp Gly Leu Thr Gln Ala Arg Val 912
 2737 TAT TAC AAT CTG GCC GCT GAA TTG CTA GGG CCT CGT CCG GAT GTA TCG
 2784

913 Tyr Tyr Asn Leu Ala Ala Glu Leu Leu Gly Pro Arg Pro Asp Val Ser 928
 2785 CTG AGT AGC ATT TGG ACG CCG CAA ACC CTG GAT ACC TTA GCA GCC GGG
 2832

929 Leu Ser Ser Ile Trp Thr Pro Gln Thr Leu Asp Thr Leu Ala Ala Gly 944
 2833 CAA AAA GCG GTT TTA CGT GAT TTT GAG CAC CAG TTG GCT AAT AGT GAT
 2880

945 Gln Lys Ala Val Leu Arg Asp Phe Glu His Gln Leu Ala Asn Ser Asp 960
 2881 ACC GCT TTA CCC GCA TTG CCG GGC CGC AAT GTC AGC TAC TTG AAA CTG
 2928

961 Thr Ala Leu Pro Ala Leu Pro Gly Arg Asn Val Ser Tyr Leu Lys Leu 976
 2929 GCA GAT AAT GGC TAC TTT AAT GAA CCG CTC AAT GTT CTG ATG TTG TCT
 2976

977 Ala Asp Asn Gly Tyr Phe Asn Glu Pro Leu Asn Val Leu Met Leu Ser 992
 2977 CAC TGG GAT ACG TTG GAT GCA CGG TTA TAC AAT CTG CGT CAT AAC CTG
 3024
 993 His Trp Asp Thr Leu Asp Ala Arg Leu Tyr Asn Leu Arg His Asn Leu
 1008
 3025 ACC GTT GAT GGC AAG CCG CTT TCG CTG CCG CTG TAT GCT GCG CCT GTT
 3072
 1009 Thr Val Asp Gly Lys Pro Leu Ser Leu Pro Leu Tyr Ala Ala Pro Val
 1024
 3073 GAT CCG GTA GCG TTG TTG GCT CAG CGT GCT CAG TCC GGC ACG TTG ACG
 3120
 1025 Asp Pro Val Ala Leu Leu Ala Gln Arg Ala Gln Ser Gly Thr Leu Thr
 1040
 3121 AAT GGC GTC AGT GGC GCC ATG TTG ACG GTG CCG CCA TAC CGT TTC AGC
 3168
 1041 Asn Gly Val Ser Gly Ala Met Leu Thr Val Pro Pro Tyr Arg Phe Ser
 1056
 3169 GCT ATG TTG CCG CGA GCT TAC AGC GCC GTG GGT ACG TTG ACC AGT TTT
 3216
 1057 Ala Met Leu Pro Arg Ala Tyr Ser Ala Val Gly Thr Leu Thr Ser Phe
 1072
 3217 GGT CAG AAC CTG CTT AGT TTG TTG GAA CGT AGC GAA CGA GCC TGT CAA
 3264
 1073 Gly Gln Asn Leu Leu Ser Leu Leu Glu Arg Ser Glu Arg Ala Cys Gln
 1088
 3265 GAA GAG TTG GCG CAA CAG CAA CTG TTG GAT ATG TCC AGC TAT GCC ATC
 3312
 1089 Glu Glu Leu Ala Gln Gln Gln Leu Leu Asp Met Ser Ser Tyr Ala Ile
 1104
 3313 ACG TTG CAA CAA CAG GCG CTG GAT GGA TTG GCG GCA GAT CGT CTG GCG
 3360
 1105 Thr Leu Gln Gln Gln Ala Leu Asp Gly Leu Ala Ala Asp Arg Leu Ala
 1120
 3361 CTG CTA GCT AGT CAG GCT ACG GCA CAA CAG CGT CAT GAC CAT TAT TAC
 3408
 1121 Leu Leu Ala Ser Gln Ala Thr Ala Gln Gln Arg His Asp His Tyr Tyr
 1136
 3409 ACT CTG TAT CAG AAC AAC ATC TCC AGT GCG GAA CAA CTG GTG ATG GAC
 3456
 1137 Thr Leu Tyr Gln Asn Asn Ile Ser Ser Ala Glu Gln Leu Val Met Asp
 1152
 3457 ACC CAA ACG TCA GCA CAA TCC CTG ATT TCT TCT TCC ACT GGT GTA CAA
 3504
 1153 Thr Gln Thr Ser Ala Gln Ser Leu Ile Ser Ser Ser Thr Gly Val Gln
 1168

3505 ACT GCC AGT GGG GCA CTG AAA GTG ATC CCG AAT ATC TTT GGT TTG GCT
3552
1169 Thr Ala Ser Gly Ala Leu Lys Val Ile Pro Asn Ile Phe Gly Leu Ala
1184
3553 GAT GGC GGC TCG CGC TAT GAA GGA GTA ACG GAA GCG ATT GCC ATC GGG
3600
1185 Asp Gly Gly Ser Arg Tyr Glu Gly Val Thr Glu Ala Ile Ala Ile Gly
1200
3601 TTA ATG GCT GCC GGA CAA GCC ACC AGC GTG GTG GCC GAG CGT CTG GCA
3648
1201 Leu Met Ala Ala Gly Gln Ala Thr Ser Val Val Ala Glu Arg Leu Ala
1216
3649 ACC ACG GAG AAT TAC CGC CGC CGC CGT GAA GAG TGG CAA ATC CAA TAC
3696
1217 Thr Thr Glu Asn Tyr Arg Arg Arg Arg Glu Glu Trp Gln Ile Gln Tyr
1232
3697 CAG CAG GCA CAG TCT GAG GTC GAC GCA TTA CAG AAA CAG TTG GAT GCG
3744
1233 Gln Gln Ala Gln Ser Glu Val Asp Ala Leu Gln Lys Gln Leu Asp Ala
1248
3745 CTG GCA GTG CGC GAG AAA GCA GCT CAA ACT TCC CTG CAA CAG GCG AAG
3792
1249 Leu Ala Val Arg Glu Lys Ala Ala Gln Thr Ser Leu Gln Gln Ala Lys
1264
3793 GCA CAG CAG GTA CAA ATT CGG ACC ATG CTG ACT TAC TTA ACT ACT CGT
3840
1265 Ala Gln Gln Val Gln Ile Arg Thr Met Leu Thr Tyr Leu Thr Thr Arg
1280
3841 TTC ACC CAG GCG ACT CTG TAC CAG TGG CTG AGT GGT CAA TTA TCC GCG
3888
1281 Phe Thr Gln Ala Thr Leu Tyr Gln Trp Leu Ser Gly Gln Leu Ser Ala
1296
3889 TTG TAT TAT CAA GCG TAT GAT GCC GTG GTT GCT CTC TGC CTC TCC GCC
3936
1297 Leu Tyr Tyr Gln Ala Tyr Asp Ala Val Val Ala Leu Cys Leu Ser Ala
1312
3937 CAA GCT TGC TGG CAG TAT GAA TTG GGT GAT TAC GCT ACC ACT TTT ATC
3984
1313 Gln Ala Cys Trp Gln Tyr Glu Leu Gly Asp Tyr Ala Thr Thr Phe Ile
1328
3985 CAG ACC GGT ACC TGG AAC GAC CAT TAC CGT GGT TTG CAA GTG GGG GAG
4032
1329 Gln Thr Gly Thr Trp Asn Asp His Tyr Arg Gly Leu Gln Val Gly Glu
1344
4033 ACA CTG CAA CTC AAT TTG CAT CAG ATG GAA GCG GCC TAT TTA GTT CGT

4080
 1345 Thr Leu Gln Leu Asn Leu His Gln Met Glu Ala Ala Tyr Leu Val Arg
 1360
 4081 CAC GAA CGC CGT CTT AAT GTG ATC CGT ACT GTG TCG CTC AAA AGC CTA
 4128
 1361 His Glu Arg Arg Leu Asn Val Ile Arg Thr Val Ser Leu Lys Ser Leu
 1376
 4129 TTG GGT GAT GAT GGT TTT GGT AAG TTA AAA ACC GAA GGC AAA GTC GAC
 4176
 1377 Leu Gly Asp Asp Gly Phe Gly Lys Leu Lys Thr Glu Gly Lys Val Asp
 1392
 4177 TTT CCA TTA AGC GAA AAG CTG TTT GAC AAC GAC TAT CCG GGG CAC TAT
 4224
 1393 Phe Pro Leu Ser Glu Lys Leu Phe Asp Asn Asp Tyr Pro Gly His Tyr
 1408
 4225 TTG CGC CAG ATT AAA ACT GTG TCA GTG ACG TTG CCG ACG TTA GTC GGG
 4272
 1409 Leu Arg Gln Ile Lys Thr Val Ser Val Thr Leu Pro Thr Leu Val Gly
 1424
 4273 CCG TAT CAA AAC GTG AAG GCA ACG CTC ACT CAG ACC AGC AGC AGT ATA
 4320
 1425 Pro Tyr Gln Asn Val Lys Ala Thr Leu Thr Gln Thr Ser Ser Ser Ile
 1440
 4321 TTG TTA GCA GCA GAT ATC AAT GGT GTT AAA CGT CTC AAT GAT CCG ACA
 4368
 1441 Leu Leu Ala Ala Asp Ile Asn Gly Val Lys Arg Leu Asn Asp Pro Thr
 1456
 4369 GGT AAA GAG GGT GAT GCG ACG CAT ATT GTC ACC AAT CTG CGT GCC AGC
 4416
 1457 Gly Lys Glu Gly Asp Ala Thr His Ile Val Thr Asn Leu Arg Ala Ser
 1472
 4417 CAG CAG GTG GCG CTC TCT TCT GGC ATT AAT GAT GCC GGT AGC TTT GAG
 4464
 1473 Gln Gln Val Ala Leu Ser Ser Gly Ile Asn Asp Ala Gly Ser Phe Glu
 1488
 4465 TTG CGT TTG GAA GAT GAG CGC TAT CTA TCA TTT GAG GGG ACT GGA GCT
 4512
 1489 Leu Arg Leu Glu Asp Glu Arg Tyr Leu Ser Phe Glu Gly Thr Gly Ala
 1504
 4513 GTT TCC AAA TGG ACT CTT AAC TTC CCG CGT TCT GTG GAT GAG CAT ATT
 4560
 1505 Val Ser Lys Trp Thr Leu Asn Phe Pro Arg Ser Val Asp Glu His Ile
 1520
 4561 GAC GAT AAG ACA TTG AAA GCG GAT GAG ATG CAG GCC GCA CTG TTG GCG
 4608

1521 Asp Asp Lys Thr Leu Lys Ala Asp Glu Met Gln Ala Ala Leu Leu Ala
 1536
 4609 AAT ATG GAT GAT GTG CTG GTG CAG GTG CAT TAT ACC GCC TGC GAC GGC
 4656
 1537 Asn Met Asp Asp Val Leu Val Gln Val His Tyr Thr Ala Cys Asp Gly
 1552
 4657 GGC GCC AGT TTC GCA AAC CAG GTC AAG AAA ACA CTC TCT TAA CATTAAC TTT 4708
 1553 Gly Ala Ser Phe Ala Asn Gln Val Lys Lys Thr Leu Ser End 1565
 4709 TAACTAATCC CTCCACTCT GTTCGCCAGA GTGGGAGAAG GTTTGCATA TCTAAAATCA 4768
 4770 ATCTTGCGAT CTTTCTCCAT TTCATTGGAA GGGAAGCTGT AAAACAAATA AGGAATATGA 4828
 4829 TATG 4932

(2) INFORMATION FOR SEQ ID NO:59

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1565 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59 (TccB peptide)

| Features | From | To | Description |
|----------|------|----|-------------|
|----------|------|----|-------------|

| | | | |
|-----|---|--|--|
| 1 | Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu Ser Gln Arg Asp Ala | | |
| 16 | | | |
| 17 | Leu Val Thr Gly Tyr Met Asn Phe Val Ala Pro Thr Leu Lys Gly Val | | |
| 32 | | | |
| 33 | Ser Gly Gln Pro Val Thr Val Glu Asp Leu Tyr Glu Tyr Leu Leu Ile | | |
| 48 | | | |
| 49 | Asp Pro Glu Val Ala Asp Glu Val Glu Thr Ser Arg Val Ala Gln Ala | | |
| 64 | | | |
| 65 | Ile Ala Ser Ile Gln Gln Tyr Met Thr Arg Leu Val Asn Gly Ser Glu | | |
| 80 | | | |
| 81 | Pro Gly Arg Gln Ala Met Glu Pro Ser Thr Ala Asn Glu Trp Arg Asp | | |
| 96 | | | |
| 97 | Asn Asp Asn Gln Tyr Ala Ile Trp Ala Ala Gly Ala Glu Val Arg Asn | | |
| 112 | | | |
| 113 | Tyr Ala Glu Asn Tyr Ile Ser Pro Ile Thr Arg Gln Glu Lys Ser His | | |
| 128 | | | |
| 129 | Tyr Phe Ser Glu Leu Glu Thr Thr Leu Asn Gln Asn Arg Leu Asp Pro | | |
| 144 | | | |
| 145 | Asp Arg Val Gln Asp Ala Val Leu Ala Tyr Leu Asn Glu Phe Glu Ala | | |
| 160 | | | |
| 161 | Val Ser Asn Leu Tyr Val Leu Ser Gly Tyr Ile Asn Gln Asp Lys Phe | | |
| 176 | | | |
| 177 | Asp Gln Ala Ile Tyr Tyr Phe Ile Gly Arg Thr Thr Thr Lys Pro Tyr | | |
| 192 | | | |
| 193 | Arg Tyr Tyr Trp Arg Gln Met Asp Leu Ser Lys Asn Arg Gln Asp Pro | | |

208

209 Ala Gly Asn Pro Val Thr Pro Asn Cys Trp Asn Asp Trp Gln Glu Ile

224

225 Thr Leu Pro Leu Ser Gly Asp Thr Val Leu Glu His Thr Val Arg Pro

240

241 Val Phe Tyr Asn Asp Arg Leu Tyr Val Ala Trp Val Glu Arg Asp Pro

256

257 Ala Val Gln Lys Asp Ala Asp Gly Lys Asn Ile Gly Lys Thr His Ala

272

273 Tyr Asn Ile Lys Phe Gly Tyr Lys Arg Tyr Asp Asp Thr Trp Thr Ala

288

289 Pro Asn Thr Thr Thr Leu Met Thr Gln Gln Ala Gly Glu Ser Ser Glu

304

305 Thr Gln Arg Ser Ser Leu Leu Ile Asp Glu Ser Ser Thr Thr Leu Arg

320

321 Gln Val Asn Leu Leu Ala Thr Thr Asp Phe Ser Ile Asp Pro Thr Glu

336

337 Glu Thr Asp Ser Asn Pro Tyr Gly Arg Leu Met Leu Gly Val Phe Val

352

353 Arg Gln Phe Glu Gly Asp Gly Ala Asn Arg Lys Asn Lys Pro Val Val

368

369 Tyr Gly Tyr Leu Tyr Cys Asp Ser Ala Phe Asn Arg His Val Leu Arg

384

385 Pro Leu Ser Lys Asn Phe Leu Phe Ser Thr Tyr Arg Asp Glu Thr Asp

400

401 Gly Gln Asn Ser Leu Gln Phe Ala Val Tyr Asp Lys Lys Tyr Val Ile

416

417 Thr Lys Val Val Thr Gly Ala Thr Glu Asp Pro Glu Asn Thr Gly Trp

432

433 Val Ser Lys Val Asp Asp Leu Lys Gln Gly Thr Thr Gly Ala Tyr Val

448

449 Tyr Ile Asp Gln Asp Gly Leu Thr Leu His Ile Gln Thr Thr Thr Asn

464

465 Gly Asp Phe Ile Asn Arg His Thr Phe Gly Tyr Asn Asp Leu Val Tyr

480

481 Asp Ser Lys Ser Gly Tyr Gly Phe Thr Trp Ser Gly Asn Glu Gly Phe

496

497 Tyr Leu Asp Tyr His Asp Gly Asn Tyr Tyr Thr Phe His Asn Ala Ile

512

513 Ile Asn Tyr Tyr Pro Ser Gly Tyr Gly Gly Gly Ser Val Pro Asn Gly

528

529 Thr Trp Ala Leu Glu Gln Arg Ile Asn Glu Gly Trp Ala Ile Ala Pro

544

545 Leu Leu Asp Thr Leu His Thr Val Thr Val Lys Gly Ser Tyr Ile Ala

560

561 Trp Glu Gly Glu Thr Pro Thr Gly Tyr Asn Leu Tyr Ile Pro Asp Gly
576
577 Thr Val Leu Leu Asp Trp Phe Asp Lys Ile Asn Phe Ala Ile Gly Leu
592
593 Asn Lys Leu Glu Ser Val Phe Thr Ser Pro Asp Trp Pro Thr Leu Thr
608
609 Thr Ile Lys Asn Phe Ser Lys Ile Ala Asp Asn Arg Lys Phe Tyr Gln
624
625 Glu Ile Asn Ala Glu Thr Ala Asp Gly Arg Asn Leu Phe Lys Arg Tyr
640
641 Ser Thr Gln Thr Phe Gly Leu Thr Ser Gly Ala Thr Tyr Ser Thr Thr
656
657 Tyr Thr Leu Ser Glu Ala Asp Phe Ser Thr Asp Pro Asp Lys Asn Tyr
672
673 Leu Gln Val Cys Leu Asn Val Val Trp Asp His Tyr Asp Arg Pro Ser
688
689 Gly Lys Lys Gly Ala Tyr Ser Trp Val Ser Lys Trp Phe Asn Val Tyr
704
705 Val Ala Leu Gln Asp Ser Lys Ala Pro Asp Ala Ile Pro Arg Leu Val
720
721 Ser Arg Tyr Asp Ser Lys Arg Gly Leu Val Gln Tyr Leu Asp Phe Trp
736
737 Thr Ser Ser Leu Pro Ala Lys Thr Arg Leu Asn Thr Thr Phe Val Arg
752
753 Thr Leu Ile Glu Lys Ala Asn Leu Gly Leu Asp Ser Leu Leu Asp Tyr
768
769 Thr Leu Gln Ala Asp Pro Ser Leu Glu Ala Asp Leu Val Thr Asp Gly
784
785 Lys Ser Glu Pro Met Asp Phe Asn Gly Ser Asn Gly Leu Tyr Phe Trp
800
801 Glu Leu Phe Phe His Leu Pro Phe Leu Val Ala Thr Arg Phe Ala Asn
816
817 Glu Gln Gln Phe Ser Pro Ala Gln Lys Ser Leu His Tyr Ile Phe Asp
832
833 Pro Ala Met Lys Asn Lys Pro His Asn Ala Pro Ala Tyr Trp Asn Val
848
849 Arg Pro Leu Val Glu Gly Asn Ser Asp Leu Ser Arg His Leu Asp Asp
864
865 Ser Ile Asp Pro Asp Thr Gln Ala Tyr Ala His Pro Val Ile Tyr Gln
880
881 Lys Ala Val Phe Ile Ala Tyr Val Ser Asn Leu Ile Ala Gln Gly Asp
896
897 Met Trp Tyr Arg Gln Leu Thr Arg Asp Gly Leu Thr Gln Ala Arg Val
912
913 Tyr Tyr Asn Leu Ala Ala Glu Leu Leu Gly Pro Arg Pro Asp Val Ser

928

929 Leu Ser Ser Ile Trp Thr Pro Gln Thr Leu Asp Thr Leu Ala Ala Gly

944

945 Gln Lys Ala Val Leu Arg Asp Phe Glu His Gln Leu Ala Asn Ser Asp

960

961 Thr Ala Leu Pro Ala Leu Pro Gly Arg Asn Val Ser Tyr Leu Lys Leu

976

977 Ala Asp Asn Gly Tyr Phe Asn Glu Pro Leu Asn Val Leu Met Leu Ser

992

993 His Trp Asp Thr Leu Asp Ala Arg Leu Tyr Asn Leu Arg His Asn Leu

1008

1009 Thr Val Asp Gly Lys Pro Leu Ser Leu Pro Leu Tyr Ala Ala Pro Val

1024

1025 Asp Pro Val Ala Leu Leu Ala Gln Arg Ala Gln Ser Gly Thr Leu Thr

1040

1041 Asn Gly Val Ser Gly Ala Met Leu Thr Val Pro Pro Tyr Arg Phe Ser

1056

1057 Ala Met Leu Pro Arg Ala Tyr Ser Ala Val Gly Thr Leu Thr Ser Phe

1072

1073 Gly Gln Asn Leu Leu Ser Leu Leu Glu Arg Ser Glu Arg Ala Cys Gln

1088

1089 Glu Glu Leu Ala Gln Gln Gln Leu Leu Asp Met Ser Ser Tyr Ala Ile

1104

1105 Thr Leu Gln Gln Gln Ala Leu Asp Gly Leu Ala Ala Asp Arg Leu Ala

1120

1121 Leu Leu Ala Ser Gln Ala Thr Ala Gln Gln Arg His Asp His Tyr Tyr

1136

1137 Thr Leu Tyr Gln Asn Asn Ile Ser Ser Ala Glu Gln Leu Val Met Asp

1152

1153 Thr Gln Thr Ser Ala Gln Ser Leu Ile Ser Ser Ser Thr Gly Val Gln

1168

1169 Thr Ala Ser Gly Ala Leu Lys Val Ile Pro Asn Ile Phe Gly Leu Ala

1184

1185 Asp Gly Gly Ser Arg Tyr Glu Gly Val Thr Glu Ala Ile Ala Ile Gly

1200

1201 Leu Met Ala Ala Gly Gln Ala Thr Ser Val Val Ala Glu Arg Leu Ala

1216

1217 Thr Thr Glu Asn Tyr Arg Arg Arg Arg Glu Glu Trp Gln Ile Gln Tyr

1232

1233 Gln Gln Ala Gln Ser Glu Val Asp Ala Leu Gln Lys Gln Leu Asp Ala

1248

1249 Leu Ala Val Arg Glu Lys Ala Ala Gln Thr Ser Leu Gln Gln Ala Lys

1264

1265 Ala Gln Gln Val Gln Ile Arg Thr Met Leu Thr Tyr Leu Thr Thr Arg

1280

1281 Phe Thr Gln Ala Thr Leu Tyr Gln Trp Leu Ser Gly Gln Leu Ser Ala
1296
1297 Leu Tyr Tyr Gln Ala Tyr Asp Ala Val Val Ala Leu Cys Leu Ser Ala
1312
1313 Gln Ala Cys Trp Gln Tyr Glu Leu Gly Asp Tyr Ala Thr Thr Phe Ile
1328
1329 Gln Thr Gly Thr Trp Asn Asp His Tyr Arg Gly Leu Gln Val Gly Glu
1344
1345 Thr Leu Gln Leu Asn Leu His Gln Met Glu Ala Ala Tyr Leu Val Arg
1360
1361 His Glu Arg Arg Leu Asn Val Ile Arg Thr Val Ser Leu Lys Ser Leu
1376
1377 Leu Gly Asp Asp Gly Phe Gly Lys Leu Lys Thr Glu Gly Lys Val Asp
1392
1393 Phe Pro Leu Ser Glu Lys Leu Phe Asp Asn Asp Tyr Pro Gly His Tyr
1408
1409 Leu Arg Gln Ile Lys Thr Val Ser Val Thr Leu Pro Thr Leu Val Gly
1424
1425 Pro Tyr Gln Asn Val Lys Ala Thr Leu Thr Gln Thr Ser Ser Ser Ile
1440
1441 Leu Leu Ala Ala Asp Ile Asn Gly Val Lys Arg Leu Asn Asp Pro Thr
1456
1457 Gly Lys Glu Gly Asp Ala Thr His Ile Val Thr Asn Leu Arg Ala Ser
1472
1473 Gln Gln Val Ala Leu Ser Ser Gly Ile Asn Asp Ala Gly Ser Phe Glu
1488
1489 Leu Arg Leu Glu Asp Glu Arg Tyr Leu Ser Phe Glu Gly Thr Gly Ala
1504
1505 Val Ser Lys Trp Thr Leu Asn Phe Pro Arg Ser Val Asp Glu His Ile
1520
1521 Asp Asp Lys Thr Leu Lys Ala Asp Glu Met Gln Ala Ala Leu Leu Ala
1536
1537 Asn Met Asp Asp Val Leu Val Gln Val His Tyr Thr Ala Cys Asp Gly
1552
1553 Gly Ala Ser Phe Ala Asn Gln Val Lys Lys Thr Leu Ser 1565

(2) INFORMATION FOR SEQ ID NO:60

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3132 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60 (tccc)

1 ATG AGT CCG TCT GAG ACT ACT CTT TAT ACT CAA ACC CCA ACA GTC AGC 48
1 Met Ser Pro Ser Glu Thr Thr Leu Tyr Thr Gln Thr Pro Thr Val Ser 16

49 GTG TTA GAT AAT CGC GGT CTG TCC ATT CGT GAT ATT GGT TTT CAC CGT 96
 17 Val Leu Asp Asn Arg Gly Leu Ser Ile Arg Asp Ile Gly Phe His Arg 32
 97 ATT GTA ATC GGG GGG GAT ACT GAC ACC CGC GTC ACC CGT CAC CAG TAT
 144
 33 Ile Val Ile Gly Gly Asp Thr Asp Thr Arg Val Thr Arg His Gln Tyr 48
 145 GAT GCC CGT GGA CAC CTG AAC TAC AGT ATT GAC CCA CGC TTG TAT GAT
 192
 49 Asp Ala Arg Gly His Leu Asn Tyr Ser Ile Asp Pro Arg Leu Tyr Asp 64
 193 GCA AAG CAG GCT GAT AAC TCA GTA AAG CCT AAT TTT GTC TGG CAG CAT
 240
 65 Ala Lys Gln Ala Asp Asn Ser Val Lys Pro Asn Phe Val Trp Gln His 80
 241 GAT CTG GCC GGT CAT GCC CTG CGG ACA GAG AGT GTC GAT GCT GGT CGT
 288
 81 Asp Leu Ala Gly His Ala Leu Arg Thr Glu Ser Val Asp Ala Gly Arg 96
 289 ACT GTT GCA TTG AAT GAT ATT GAA GGT CGT TCG GTA ATG ACA ATG AAT
 336
 97 Thr Val Ala Leu Asn Asp Ile Glu Gly Arg Ser Val Met Thr Met Asn
 112
 337 GCG ACC GGT GTT CGT CAG ACC CGT CGC TAT GAA GGC AAC ACC TTG CCC
 384
 113 Ala Thr Gly Val Arg Gln Thr Arg Arg Tyr Glu Gly Asn Thr Leu Pro
 128
 385 GGT CGC TTG TTA TCT GTG AGC GAG CAA GTT TTC AAC CAA GAG AGT GCT
 432
 129 Gly Arg Leu Leu Ser Val Ser Glu Gln Val Phe Asn Gln Glu Ser Ala
 144
 433 AAA GTG ACA GAG CGC TTT ATC TGG GCT GGG AAT ACA ACC TCG GAG AAA
 480
 145 Lys Val Thr Glu Arg Phe Ile Trp Ala Gly Asn Thr Thr Ser Glu Lys
 160
 481 GAG TAT AAC CTC TCC GGT CTG TGT ATA CGC CAC TAC GAC ACA GCG GGA
 528
 161 Glu Tyr Asn Leu Ser Gly Leu Cys Ile Arg His Tyr Asp Thr Ala Gly
 176
 529 GTG ACC CGG TTG ATG AGT CAG TCA CTG GCG GGC GCC ATG CTA TCC CAA
 576
 177 Val Thr Arg Leu Met Ser Gln Ser Leu Ala Gly Ala Met Leu Ser Gln
 192
 577 TCT CAC CAA TTG CTG GCG GAA GGG CAG GAG GCT AAC TGG AGC GGT GAC
 624
 193 Ser His Gln Leu Leu Ala Glu Gly Gln Glu Ala Asn Trp Ser Gly Asp
 208
 625 GAC GAA ACT GTC TGG CAG GGA ATG CTG GCA AGT GAG GTC TAT ACG ACA
 672
 209 Asp Glu Thr Val Trp Gln Gly Met Leu Ala Ser Glu Val Tyr Thr Thr

224
 673 CAA AGT ACC ACT AAT GCC ATC GGG GCT TTA CTG ACC CAA ACC GAT GCG
 720
 225 Gln Ser Thr Thr Asn Ala Ile Gly Ala Leu Leu Thr Gln Thr Asp Ala
 240
 721 AAA GGC AAT ATT CAG CGT CTG GCT TAT GAC ATT GCC GGT CAG TTA AAA
 768
 241 Lys Gly Asn Ile Gln Arg Leu Ala Tyr Asp Ile Ala Gly Gln Leu Lys
 256
 769 GGG AGT TGG TTG ACG GTG AAA GGC CAG AGT GAA CAG GTG ATT GTT AAG
 816
 257 Gly Ser Trp Leu Thr Val Lys Gly Gln Ser Glu Gln Val Ile Val Lys
 272
 817 TCC CTG AGC TGG TCA GCC GCA GGT CAT AAA TTG CGT GAA GAG CAC GGT
 864
 273 Ser Leu Ser Trp Ser Ala Ala Gly His Lys Leu Arg Glu Glu His Gly
 288
 865 AAC GGC GTG GTT ACG GAG TAC AGT TAT GAG CCG GAA ACT CAA CGT CTG
 912
 289 Asn Gly Val Val Thr Glu Tyr Ser Tyr Glu Pro Glu Thr Gln Arg Leu
 304
 913 ATA GGT ATC ACC ACC CGG CGT GCC GAA GGG AGT CAA TCA GGA GCC AGA
 960
 305 Ile Gly Ile Thr Thr Arg Arg Ala Glu Gly Ser Gln Ser Gly Ala Arg
 320
 961 GTA TTG CAG GAT CTA CGC TAT AAG TAT GAT CCG GTG GGG AAT GTT ATC
 1008
 321 Val Leu Gln Asp Leu Arg Tyr Lys Tyr Asp Pro Val Gly Asn Val Ile 336
 1009 AGT ATC CAT AAT GAT GCC GAA GCT ACC CGC TTT TGG CGT AAT CAG AAA
 1056
 337 Ser Ile His Asn Asp Ala Glu Ala Thr Arg Phe Trp Arg Asn Gln Lys 352
 1057 GTG GAG CCG GAG AAT CGC TAT GTT TAT GAT TCT CTG TAT CAG CTT ATG
 1104
 353 Val Glu Pro Glu Asn Arg Tyr Val Tyr Asp Ser Leu Tyr Gln Leu Met 368
 1105 AGT GCG ACA GGG CGT GAA ATG GCT AAT ATC GGT CAG CAA AGC AAC CAA
 1152
 369 Ser Ala Thr Gly Arg Glu Met Ala Asn Ile Gly Gln Gln Ser Asn Gln
 384
 1153 CTT CCC TCA CCC GTT ATA CCT GTT CCT ACT GAC GAC AGC ACT TAT ACC
 1200
 385 Leu Pro Ser Pro Val Ile Pro Val Pro Thr Asp Asp Ser Thr Tyr Thr 400
 1201 AAT TAC CTT CGT ACC TAT ACT TAT GAC CGT GGC GGT AAT TTG GTT CAA
 1248
 401 Asn Tyr Leu Arg Thr Tyr Thr Tyr Asp Arg Gly Gly Asn Leu Val Gln 416
 1249 ATC CGA CAC AGT TCA CCC GCG ACT CAA AAT AGT TAC ACC ACA GAT ATC

1296
 417 Ile Arg His Ser Ser Pro Ala Thr Gln Asn Ser Tyr Thr Thr Asp Ile 432
 1297 ACC GTT TCA AGC CGC AGT AAC CGG GCG GTA TTG AGT ACA TTA ACG ACA
 1344
 433 Thr Val Ser Ser Arg Ser Asn Arg Ala Val Leu Ser Thr Leu Thr Thr 448
 1345 GAT CCA ACC CGA GTG GAT GCG CTA TTT GAT TCC GGC GGT CAT CAG AAG
 1392
 449 Asp Pro Thr Arg Val Asp Ala Leu Phe Asp Ser Gly Gly His Gln Lys 464
 1393 ATG TTA ATA CCG GGG CAA AAT CTG GAT TGG AAT ATT CGG GGT GAA TTG
 1440
 465 Met Leu Ile Pro Gly Gln Asn Leu Asp Trp Asn Ile Arg Gly Glu Leu 480
 1441 CAA CGA GTC ACA CCG GTG AGC CGT GAA AAT AGC AGT GAC AGT GAA TGG
 1488
 481 Gln Arg Val Thr Pro Val Ser Arg Glu Asn Ser Ser Asp Ser Glu Trp 496
 1489 TAT CGC TAT AGC AGT GAT GGC ATG CGG CTG CTA AAA GTG AGT GAA CAG
 1536
 497 Tyr Arg Tyr Ser Ser Asp Gly Met Arg Leu Leu Lys Val Ser Glu Gln 512
 1537 CAG ACG GGC AAC AGT ACT CAA GTA CAA CGG GTG ACT TAT CTG CCG GGA
 1584
 513 Gln Thr Gly Asn Ser Thr Gln Val Gln Arg Val Thr Tyr Leu Pro Gly 528
 1585 TTA GAG CTA CGG ACA ACT GGG GTT GCA GAT AAA ACA ACC GAA GAT TTG
 1632
 529 Leu Glu Leu Arg Thr Thr Gly Val Ala Asp Lys Thr Thr Glu Asp Leu 544
 1633 CAG GTG ATT ACG GTA GGT GAA GCG GGT CGC GCA CAG GTA AGG GTA TTG
 1680
 545 Gln Val Ile Thr Val Gly Glu Ala Gly Arg Ala Gln Val Arg Val Leu 560
 1681 CAC TGG GAA AGT GGT AAG CCG ACA GAT ATT GAC AAC AAT CAG GTG CGC
 1728
 561 His Trp Glu Ser Gly Lys Pro Thr Asp Ile Asp Asn Asn Gln Val Arg 576
 1729 TAC AGC TAC GAT AAT CTG CTT GGC TCC AGC CAG CTT GAA CTG GAT AGC
 1776
 577 Tyr Ser Tyr Asp Asn Leu Leu Gly Ser Ser Gln Leu Glu Leu Asp Ser 592
 1777 GAA GGG CAG ATT CTC AGT CAG GAA GAG TAT TAT CCG TAT GGC GGT ACG
 1824
 593 Glu Gly Gln Ile Leu Ser Gln Glu Glu Tyr Tyr Pro Tyr Gly Gly Thr 608
 1825 GCG ATA TGG GCG GCG AGA AAT CAG ACA GAA GCC AGC TAC AAA TTT ATT
 1872
 609 Ala Ile Trp Ala Ala Arg Asn Gln Thr Glu Ala Ser Tyr Lys Phe Ile 624
 1873 CGT TAC TCC GGT AAA GAG CGG GAT GCC ACT GGA TTG TAT TAT TAC GGC
 1920
 625 Arg Tyr Ser Gly Lys Glu Arg Asp Ala Thr Gly Leu Tyr Tyr Tyr Gly 640
 1921 TAC CGT TAT TAT CAA CCT TGG GTG GGT CGA TGG TTG AGT GCT GAT CCG
 1968
 641 Tyr Arg Tyr Tyr Gln Pro Trp Val Gly Arg Trp Leu Ser Ala Asp Pro 656
 1969 GCG GGA ACC GTG GAT GGG CTG AAT TTG TAC CGA ATG GTG AGG AAT AAC

2016
 657 Ala Gly Thr Val Asp Gly Leu Asn Leu Tyr Arg Met Val Arg Asn Asn 672
 2017 CCC ATC ACA TTG ACT GAC CAT GAC GGA TTA GCA CCG TCT CCA AAT AGA
 2064
 673 Pro Ile Thr Leu Thr Asp His Asp Gly Leu Ala Pro Ser Pro Asn Arg 688
 2065 AAT CGA AAT ACA TTT TGG TTT GCT TCA TTT TTG TTT CGT AAA CCT GAT
 2112
 689 Asn Arg Asn Thr Phe Trp Phe Ala Ser Phe Leu Phe Arg Lys Pro Asp 704
 2113 GAG GGA ATG TCC GCG TCA ATG AGA CGG GGA CAA AAA ATT GGC AGA GCC
 2160
 705 Glu Gly Met Ser Ala Ser Met Arg Arg Gly Gln Lys Ile Gly Arg Ala 720
 2161 ATT GCC GGC GGG ATT GCG ATT GGC GGT CTT GCG GCT ACC ATT GCC GCT
 2208
 721 Ile Ala Gly Gly Ile Ala Ile Gly Gly Leu Ala Ala Thr Ile Ala Ala 736
 2209 ACG GCT GGC GCG GCT ATC CCC GTC ATT CTG GGG GTT GCG GCC GTA GGC
 2256
 737 Thr Ala Gly Ala Ala Ile Pro Val Ile Leu Gly Val Ala Ala Val Gly 752
 2257 GCG GGG ATT GGC GCG TTG ATG GGA TAT AAC GTC GGT AGC CTG CTG GAA
 2304
 753 Ala Gly Ile Gly Ala Leu Met Gly Tyr Asn Val Gly Ser Leu Leu Glu 768
 2305 AAA GGC GGG GCA TTA CTT GCT CGA CTC GTA CAG GGG AAA TCG ACG TTA
 2352
 769 Lys Gly Gly Ala Leu Leu Ala Arg Leu Val Gln Gly Lys Ser Thr Leu 784
 2353 GTA CAG TCG GCG GCT GGC GCG GCT GCC GGA GCG AGT TCA GCC GCG GCT
 2400
 785 Val Gln Ser Ala Ala Gly Ala Ala Ala Gly Ala Ser Ser Ala Ala Ala 800
 2401 TAT GGC GCA CGG GCA CAA GGT GTC GGT GTT GCA TCA GCC GCC GGG GCG
 2448
 801 Tyr Gly Ala Arg Ala Gln Gly Val Gly Val Ala Ser Ala Ala Gly Ala 816
 2449 GTA ACA GGG GCT GTG GGA TCA TGG ATA AAT AAT GCT GAT CGG GGG ATT
 2496
 817 Val Thr Gly Ala Val Gly Ser Trp Ile Asn Asn Ala Asp Arg Gly Ile 832
 2497 GGC GGC GCT ATT GGG GCC GGG AGT GCG GTA GGC ACC ATT GAT ACT ATG
 2544
 833 Gly Gly Ala Ile Gly Ala Gly Ser Ala Val Gly Thr Ile Asp Thr Met 848
 2545 TTA GGG ACT GCC TCT ACC CTT ACC CAT GAA GTC GGG GCA GCG GCG GGT
 2592
 849 Leu Gly Thr Ala Ser Thr Leu Thr His Glu Val Gly Ala Ala Ala Gly 864
 2593 GGG GCG GCG GGT GGG ATG ATC ACC GGT ACG CAA GGG AGT ACT CGG GCA
 2640
 865 Gly Ala Ala Gly Gly Met Ile Thr Gly Thr Gln Gly Ser Thr Arg Ala 880
 2641 GGT ATC CAT GCC GGT ATT GGC ACC TAT TAT GGC TCC TGG ATT GGT TTT
 2688
 881 Gly Ile His Ala Gly Ile Gly Thr Tyr Tyr Gly Ser Trp Ile Gly Phe 896
 2689 GGT TTA GAT GTC GCT AGT AAC CCC GCC GGA CAT TTA GCG AAT TAC GCA

2736

897 Gly Leu Asp Val Ala Ser Asn Pro Ala Gly His Leu Ala Asn Tyr Ala 912

2737 GTG GGT TAT GCC GCT GGT TTG GGT GCT GAA ATG GCT GTC AAC AGA ATA

2784

913 Val Gly Tyr Ala Ala Gly Leu Gly Ala Glu Met Ala Val Asn Arg Ile 928

2785 ATG GGT GGT GGA TTT TTG AGT AGG CTC TTA GGC CGG GTT GTC AGC CCA

2832

929 Met Gly Gly Gly Phe Leu Ser Arg Leu Leu Gly Arg Val Val Ser Pro 944

2833 TAT GCC GCC GGT TTA GCC AGA CAA TTA GTA CAT TTC AGT GTC GCC AGA

2880

945 Tyr Ala Ala Gly Leu Ala Arg Gln Leu Val His Phe Ser Val Ala Arg 960

2881 CCT GTC TTT GAG CCG ATA TTT AGT GTT CTC GGC GGG CTT GTC GGT GGT

2928

961 Pro Val Phe Glu Pro Ile Phe Ser Val Leu Gly Gly Leu Val Gly Gly 976

2929 ATT GGA ACT GGC CTG CAC AGA GTG ATG GGA AGA GAG AGT TGG ATT TCC

2976

977 Ile Gly Thr Gly Leu His Arg Val Met Gly Arg Glu Ser Trp Ile Ser 992

2977 AGA GCG TTA AGT GCT GCC GGT AGT GGT ATA GAT CAT GTC GCT GGC ATG

3024

993 Arg Ala Leu Ser Ala Ala Gly Ser Gly Ile Asp His Val Ala Gly Met

1008

3025 ATT GGT AAT CAG ATC AGA GGC AGG GTC TTG ACC ACA ACC GGG ATC GCT

3072

1009 Ile Gly Asn Gln Ile Arg Gly Arg Val Leu Thr Thr Thr Gly Ile Ala

1024

3073 AAT GCG ATA GAC TAT GGC ACC AGT GCT GTG GGA GCC GCA CGA CGA GTT

3120

1025 Asn Ala Ile Asp Tyr Gly Thr Ser Ala Val Gly Ala Ala Arg Arg Val

1040

3121 TTT TCT TTG TAA 3132

1041 Phe Ser Leu End 1043

(2) INFORMATION FOR SEQ ID NO:61

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1043 amino acids

(B) TYPE: amino acid

(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61 (TccC peptide)

1 Met Ser Pro Ser Glu Thr Thr Leu Tyr Thr Gln Thr Pro Thr Val Ser 16
 17 Val Leu Asp Asn Arg Gly Leu Ser Ile Arg Asp Ile Gly Phe His Arg 32
 33 Ile Val Ile Gly Gly Asp Thr Asp Thr Arg Val Thr Arg His Gln Tyr 48
 49 Asp Ala Arg Gly His Leu Asn Tyr Ser Ile Asp Pro Arg Leu Tyr Asp 64
 65 Ala Lys Gln Ala Asp Asn Ser Val Lys Pro Asn Phe Val Trp Gln His 80
 81 Asp Leu Ala Gly His Ala Leu Arg Thr Glu Ser Val Asp Ala Gly Arg 96
 97 Thr Val Ala Leu Asn Asp Ile Glu Gly Arg Ser Val Met Thr Met Asn 112

113 Ala Thr Gly Val Arg Gln Thr Arg Arg Tyr Glu Gly Asn Thr Leu Pro 128
 129 Gly Arg Leu Leu Ser Val Ser Glu Gln Val Phe Asn Gln Glu Ser Ala 144
 145 Lys Val Thr Glu Arg Phe Ile Trp Ala Gly Asn Thr Thr Ser Glu Lys 160
 161 Glu Tyr Asn Leu Ser Gly Leu Cys Ile Arg His Tyr Asp Thr Ala Gly 176
 177 Val Thr Arg Leu Met Ser Gln Ser Leu Ala Gly Ala Met Leu Ser Gln 192
 193 Ser His Gln Leu Leu Ala Glu Gly Gln Glu Ala Asn Trp Ser Gly Asp 208
 209 Asp Glu Thr Val Trp Gln Gly Met Leu Ala Ser Glu Val Tyr Thr Thr 224
 225 Gln Ser Thr Thr Asn Ala Ile Gly Ala Leu Leu Thr Gln Thr Asp Ala 240
 241 Lys Gly Asn Ile Gln Arg Leu Ala Tyr Asp Ile Ala Gly Gln Leu Lys 256
 257 Gly Ser Trp Leu Thr Val Lys Gly Gln Ser Glu Gln Val Ile Val Lys 272
 273 Ser Leu Ser Trp Ser Ala Ala Gly His Lys Leu Arg Glu Glu His Gly 288
 289 Asn Gly Val Val Thr Glu Tyr Ser Tyr Glu Pro Glu Thr Gln Arg Leu 304
 305 Ile Gly Ile Thr Thr Arg Arg Ala Glu Gly Ser Gln Ser Gly Ala Arg 320
 321 Val Leu Gln Asp Leu Arg Tyr Lys Tyr Asp Pro Val Gly Asn Val Ile 336
 337 Ser Ile His Asn Asp Ala Glu Ala Thr Arg Phe Trp Arg Asn Gln Lys 352
 353 Val Glu Pro Glu Asn Arg Tyr Val Tyr Asp Ser Leu Tyr Gln Leu Met 368
 369 Ser Ala Thr Gly Arg Glu Met Ala Asn Ile Gly Gln Gln Ser Asn Gln 384
 385 Leu Pro Ser Pro Val Ile Pro Val Pro Thr Asp Asp Ser Thr Tyr Thr 400
 401 Asn Tyr Leu Arg Thr Tyr Thr Tyr Asp Arg Gly Gly Asn Leu Val Gln 416
 417 Ile Arg His Ser Ser Pro Ala Thr Gln Asn Ser Tyr Thr Thr Asp Ile 432
 433 Thr Val Ser Ser Arg Ser Asn Arg Ala Val Leu Ser Thr Leu Thr Thr 448
 449 Asp Pro Thr Arg Val Asp Ala Leu Phe Asp Ser Gly Gly His Gln Lys 464
 465 Met Leu Ile Pro Gly Gln Asn Leu Asp Trp Asn Ile Arg Gly Glu Leu 480
 481 Gln Arg Val Thr Pro Val Ser Arg Glu Asn Ser Ser Asp Ser Glu Trp 496
 497 Tyr Arg Tyr Ser Ser Asp Gly Met Arg Leu Leu Lys Val Ser Glu Gln 512
 513 Gln Thr Gly Asn Ser Thr Gln Val Gln Arg Val Thr Tyr Leu Pro Gly 528
 529 Leu Glu Leu Arg Thr Thr Gly Val Ala Asp Lys Thr Thr Glu Asp Leu 544
 545 Gln Val Ile Thr Val Gly Glu Ala Gly Arg Ala Gln Val Arg Val Leu 560
 561 His Trp Glu Ser Gly Lys Pro Thr Asp Ile Asp Asn Asn Gln Val Arg 576
 577 Tyr Ser Tyr Asp Asn Leu Leu Gly Ser Ser Gln Leu Glu Leu Asp Ser 592
 593 Glu Gly Gln Ile Leu Ser Gln Glu Glu Tyr Tyr Pro Tyr Gly Gly Thr 608
 609 Ala Ile Trp Ala Ala Arg Asn Gln Thr Glu Ala Ser Tyr Lys Phe Ile 624
 625 Arg Tyr Ser Gly Lys Glu Arg Asp Ala Thr Gly Leu Tyr Tyr Tyr Gly 640
 641 Tyr Arg Tyr Tyr Gln Pro Trp Val Gly Arg Trp Leu Ser Ala Asp Pro 656
 657 Ala Gly Thr Val Asp Gly Leu Asn Leu Tyr Arg Met Val Arg Asn Asn 672
 673 Pro Ile Thr Leu Thr Asp His Asp Gly Leu Ala Pro Ser Pro Asn Arg 688
 689 Asn Arg Asn Thr Phe Trp Phe Ala Ser Phe Leu Phe Arg Lys Pro Asp 704
 705 Glu Gly Met Ser Ala Ser Met Arg Arg Gly Gln Lys Ile Gly Arg Ala 720
 721 Ile Ala Gly Gly Ile Ala Ile Gly Gly Leu Ala Ala Thr Ile Ala Ala 736
 737 Thr Ala Gly Ala Ala Ile Pro Val Ile Leu Gly Val Ala Ala Val Gly 752
 753 Ala Gly Ile Gly Ala Leu Met Gly Tyr Asn Val Gly Ser Leu Leu Glu 768
 769 Lys Gly Gly Ala Leu Leu Ala Arg Leu Val Gln Gly Lys Ser Thr Leu 784
 785 Val Gln Ser Ala Ala Gly Ala Ala Ala Gly Ala Ser Ser Ala Ala Ala 800
 801 Tyr Gly Ala Arg Ala Gln Gly Val Gly Val Ala Ser Ala Ala Gly Ala 816
 817 Val Thr Gly Ala Val Gly Ser Trp Ile Asn Asn Ala Asp Arg Gly Ile 832

833 Gly Gly Ala Ile Gly Ala Gly Ser Ala Val Gly Thr Ile Asp Thr Met 848
 849 Leu Gly Thr Ala Ser Thr Leu Thr His Glu Val Gly Ala Ala Ala Gly 864
 865 Gly Ala Ala Gly Gly Met Ile Thr Gly Thr Gln Gly Ser Thr Arg Ala 880
 881 Gly Ile His Ala Gly Ile Gly Thr Tyr Tyr Gly Ser Trp Ile Gly Phe 896
 897 Gly Leu Asp Val Ala Ser Asn Pro Ala Gly His Leu Ala Asn Tyr Ala 912
 913 Val Gly Tyr Ala Ala Gly Leu Gly Ala Glu Met Ala Val Asn Arg Ile 928
 929 Met Gly Gly Gly Phe Leu Ser Arg Leu Leu Gly Arg Val Val Ser Pro 944
 945 Tyr Ala Ala Gly Leu Ala Arg Gln Leu Val His Phe Ser Val Ala Arg 960
 961 Pro Val Phe Glu Pro Ile Phe Ser Val Leu Gly Gly Leu Val Gly Gly 976
 977 Ile Gly Thr Gly Leu His Arg Val Met Gly Arg Glu Ser Trp Ile Ser 992
 993 Arg Ala Leu Ser Ala Ala Gly Ser Gly Ile Asp His Val Ala Gly Met 1008
 1009 Ile Gly Asn Gln Ile Arg Gly Arg Val Leu Thr Thr Thr Gly Ile Ala 1024
 1025 Asn Ala Ile Asp Tyr Gly Thr Ser Ala Val Gly Ala Ala Arg Arg Val 1040
 1041 Phe Ser Leu 1043

(2) INFORMATION FOR SEQ ID NO:62: TcaAiv

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62: TcaAiv

Asn Ile Gly Gly Asp

1 5

(2) INFORMATION FOR SEQ ID NO:63: TcaAii-syn

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63: TcaAii-syn

Cys Leu Arg Gly Asn Ser Pro Thr Asn Pro Asp Lys Asp Gly Ile

1 5 10 15

Phe Ala Gln Val Ala

20

(2) INFORMATION FOR SEQ ID NO:64: TcaAiii-syn

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67: TcaC-syn

Cys Tyr Lys Ala Pro Gln Arg Gln Glu Asp Gly Asp Ser Asn Ala
 1 5 10 15
 Val Thr Tyr Asp Lys
 20

(2) INFORMATION FOR SEQ ID NO:68: TcbAii-syn

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68: TcbAii-syn

Cys Tyr Asn Glu Asn Pro Ser Ser Glu Asp Lys Lys Trp Tyr Phe
 1 5 10 15
 Ser Ser Lys Asp Asp
 20

(2) INFORMATION FOR SEQ ID NO:69: TcbAiii-syn

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69: TcbAiii-syn

Cys Phe Asp Ser Tyr Ser Gln Leu Tyr Glu Glu Asn Ile Asn Ala
 1 5 10 15
 Gly Glu Gln Arg Ala
 20

(2) INFORMATION FOR SEQ ID NO:70: TcdAii-syn

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70: TcdAii-syn

Cys Asn Pro Asn Asn Ser Ser Asn Lys Leu Met Phe Tyr Pro Val

1 5 10 15
 Tyr Gln Tyr Ser Gly Asn Thr
 20

(2) INFORMATION FOR SEQ ID NO:71: TcdAiii-syn

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71: TcdAiii-syn

Val Ser Gln Gly Ser Gly Ser Ala Gly Ser Gly Asn Asn Asn Leu
 1 5 10 15
 Ala Phe Gly Ala Gly
 20

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72: 160 kDa - Hb

Met Gln Asp Ser Pro Glu Val Ala Ile Thr Thr Leu
 1 5 10

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73: 170 kDa - WIR

Met Gln Arg Ser Ser Glu Val Ser
 1 5

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74: 180 kDa - H9

Met Gln Asp Ile Pro Glu Val Gln Leu Asn

1 5 10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75: 170 kDa - Hm(2)

INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

Met Gln Asp Ser Pro Glu Val Ser Val Thr Gln Asn

1 5 10

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76: 74 kDa - H9

Ser Glu Ser Leu Phe Thr Gln Ser Leu Lys Glu Ala Arg Arg Asp

1 5 10 15

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77: 71 kDa - Hb

Met Asn Leu Ile Glu Ala Lys Leu Gln Glu Asn Arg Asp Ala

1 5 10

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78: 170 kDa - H9

Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu Ser Gln Arg Asp
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79: 109 kDa - Hm

Met Leu Asp Ile Met Glu Lys Gln Leu Asn Glu Ser Glu Arg Asp
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80: 170 kDa - WX-1

Met Gln Asp Ser Arg Glu Val Ser
 1 5

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81: 69 kDa - H9

Leu Arg Ser Ala Xxx Ser Ala Leu Thr Thr Leu Leu
 1 5 10

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82: 64 kDa - HP88

Leu Lys Leu Ala Asp Asn Gly Tyr Phe Asn Glu Pro Leu Asn Val
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83: 70 kDa - NC-1

Leu Lys Leu Ala Asp Asn Ser Tyr Phe Asn Glu Pro Leu Asn
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84: 60 kDa - WIR

Ser Lys Asp Glu Ser Lys Ala Asp Ser Gln Leu Val Tyr His Thr
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85: 58 kDa - NC-1

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

1 5 10

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86: 60 kDa - WX-12

Met Leu Asn Pro Ile Val Arg Lys Phe Glu Tyr Gly Glu His Thr

1 5 10 15

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87: 60 kDa - Hm

Ala Glu Ile Tyr Asn Lys Asp Gly Asn Lys Leu Asp Leu Tyr Gly

1 5 10 15

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: protein

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88: 140 kDa - Hm

Asn Leu Ile Glu Ala Thr Leu Glu Gln Asn Leu Arg Asp Ala

1 5 10 15

(57) 청구의 범위

청구항 1

곤충에 대하여 작용 활성을 갖는 포토랍두스 (Photorhabdus) 단백질 독소의 유효량을 포함하는 조성물.

청구항 2

제1항에 있어서, 포토랍두스 독소가 포토랍두스의 정제된 배양물, 유전자전환 식물, 바쿨로바이러스, 또는 이중 미생물 숙주에 의해 생산되는 조성물.

청구항 3

제2항에 있어서, 포토랍두스 독소가 포토랍두스 루미네센스(*P. luminescens*)의 정제된 배양물에 의해 생산되는 조성물.

청구항 4

제2항에 있어서, 독소가 ATCC 55397로 지정된 포토랍두스 루미네센스 균주의 정제된 배양물로부터 생산되는 조성물.

청구항 5

제2항에 있어서, 독소가 W-14로 지정된 포토랍두스 루미네센스 균주의 정제된 배양물에 의해 생산되는 조성물.

청구항 6

제1항에 있어서, 독소가 WX-1, WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-9, WX-10, WX-11, WX-12, WX-14, WX-15, H9, Hb, Hm, HP88, NC-1, W30, WIR, B2, ATCC# 43948, ATCC# 43949, ATCC# 43950, ATCC# 43951, ATCC# 43952, DEP1, DEP2, DEP3, 피. 지알란드리카, 피. 해피알루스, HB-Arg, HB 오스웨고, HB 레위스톤, K-122, HMGD, 인디커스, GD, PWH-5, 메기디스, HF-85, 에이. 카우스, MP1, MP2, MP3, MP4, MP5, GL98, GL101, GL138, GL55, GL217 또는 GL257로 지정된 포토랍두스 균주의 정제된 배양물에 의해 생산되는 조성물.

청구항 7

제2항에 있어서, 독소가 WX-1, WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-9, WX-10, WX-11, WX-12, WX-14, WX-15, H9, Hb, Hm, HP88, NC-1, W30, WIR, B2, ATCC# 43948, ATCC# 43949, ATCC# 43950, ATCC# 43951, ATCC# 43952, DEP1, DEP2, DEP3, 피. 지알란드리카, 피.해피알루스, HB-Arg, HB 오스웨고, HB 레위스톤, K-122, HMGD, 인디커스, GD, PWH-5, 메기디스, HF-85, 에이. 카우스, MP1, MP2, MP3, MP4, MP5, GL98, GL101, GL138, GL55, GL217 또는 GL257로 지정된 포토랍두스 루미네센스 균주의 정제된 배양물로부터 생산되는 조성물.

청구항 8

제1항에 있어서, 독소가 서열 12의 아미노산 서열로 표시되는 조성물.

청구항 9

제6항에 있어서, 조성물이 포토랍두스의 정제된 배양물들로부터 생산된 하나 이상의 독소의 혼합물인 조성물.

청구항 10

제1항 또는 제6항에 있어서, 곤충이 나비목 (*Lepidoptera*), 딱정벌레목 (*Coleoptera*), 벌목 (*Hymenoptera*), 파리목 (*Diptera*), 방시목 (*Dictyoptera*), 응애목 (*Acarina*) 또는 동시목 (*Homoptera*)인 조성물.

청구항 11

제1항 또는 제6항에 있어서, 곤충종이 딱정벌레목에 속하며, 써던 옥수수 뿌리벌레, 웨스턴 옥수수 뿌리벌레, 콜로라도 감자 딱정벌레, 밀웬, 볼 바구미 또는 잔디 땅벌레인 조성물.

청구항 12

제1항 또는 제6항에 있어서, 곤충종이 나비목에 속하며, 사탕무 행혈구더기, 블랙 야도충, 양배추 자벌레, 코들링 나방, 옥수수 귀벌레, 유럽 옥수수 천공충, 박각시나방 또는 회색담배나방인 조성물.

청구항 13

제1항 또는 제6항에 있어서, 독소가 분무성 살충제로서 제형화되는 조성물.

청구항 14

제1항 또는 제6항에 있어서, 독소가 미끼 매트릭스로서 제형화되고 지상 또는 지하 미끼 스테이션에 전달되는 조성물.

청구항 15

포토랍두스 속의 정제된 박테리아 배양물에 의해 생산된, 곤충에 대해 작용 활성을 갖는 단백질 독소의 유효량을 곤충에 경구적으로 전달하는 것을 포함하는 곤충 방제 방법.

청구항 16

제15항에 있어서, 박테리아가 포토랍두스 루미네센스의 정제된 배양물인 방법.

청구항 17

제15항에 있어서, 독소가 ATCC 55397로 지정된 *포토라프두스 루미네센스* 균주의 정제된 배양물로부터 생산되는 방법.

청구항 18

제16항에 있어서, 독소가 W-14로 지정된 *포토라프두스 루미네센스* 균주의 정제된 배양물로부터 생산되는 방법.

청구항 19

제15항에 있어서, 독소가 WX-1, WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-9, WX-10, WX-11, WX-12, WX-14, WX-15, H9, Hb, Hm, HP88, NC-1, W30, WIR, B2, ATCC# 43948, ATCC# 43949, ATCC# 43950, ATCC# 43951, ATCC# 43952, DEP1, DEP2, DEP3, 피. 지알란드리카, 피.해피알루스, HB-Arg, HB 오스웨고, HB 레위스톤, K-122, HMGD, 인디커스, GD, PWH-5, 메기디스, HF-85, 에이. 카우스, MP1, MP2, MP3, MP4, MP5, GL98, GL101, GL138, GL55, GL217 또는 GL257로 지정된 *포토라프두스* 균주의 정제된 배양물로부터 생산되는 방법.

청구항 20

제15항에 있어서, 독소가 WX-1, WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-9, WX-10, WX-11, WX-12, WX-14, WX-15, H9, Hb, Hm, HP88, NC-1, W30, WIR, B2, ATCC# 43948, ATCC# 43949, ATCC# 43950, ATCC# 43951, ATCC# 43952, DEP1, DEP2, DEP3, 피. 지알란드리카, 피.해피알루스, HB-Arg, HB 오스웨고, HB 레위스톤, K-122, HMGD, 인디커스, GD, PWH-5, 메기디스, HF-85, 에이. 카우스, MP1, MP2, MP3, MP4, MP5, GL98, GL101, GL138, GL155, GL217 또는 GL257로 지정된 *포토라프두스 루미네센스* 균주의 정제된 배양물로부터 생산되는 방법.

청구항 21

제19항에 있어서, 하나 이상의 독소의 혼합물이 *포토라프두스*의 정제된 배양물로부터 생산되고 상기 독소가 곤충에 경구적으로 전달되는 방법.

청구항 22

제15항에 있어서, 독소가 독소를 코딩하는 유전자로 형질전환된 원핵세포 숙주에 의해 생산되는 방법.

청구항 23

제15항에 있어서, 독소가 독소를 코딩하는 유전자로 형질전환된 진핵세포 숙주에 의해 생산되는 방법.

청구항 24

제23항에 있어서, 진핵세포 숙주가 바쿨로바이러스인 방법.

청구항 25

제15항 또는 제19항에 있어서, 곤충이 나비목, 딱정벌레목, 벌목, 파리목, 방시목, 응애목 또는 동시목인 조성물.

청구항 26

제15항 또는 제19항에 있어서, 곤충종이 딱정벌레목에 속하며, 셔던 옥수수 뿌리벌레, 웨스턴 옥수수 뿌리벌레, 콜로라도 감자 딱정벌레, 밀웜, 볼 바구미 또는 잔디 땅벌레인 방법.

청구항 27

제15항 또는 제19항에 있어서, 곤충종이 나비목에 속하며, 사탕무 행렬구더기, 블랙 야도충, 양배추 자벌레, 코틀링 나방, 옥수수 귀벌레, 유럽 옥수수 천공충, 박각시나방 또는 회색담배나방인 방법.

청구항 28

제15항 또는 제19항에 있어서, 독소가 분무성 살충제로서 제형화되는 방법.

청구항 29

제15항 또는 제19항에 있어서, 독소가 미기 매트릭스로서 제형화되고 지상 또는 지하 미기 스테이션에 전달되는 방법.

청구항 30

포토라프두스 또는 *포토라프두스 루미네센스*로부터 유전 물질을 분리하기 위해 사용되는, 서열 1, 서열 2, 서열 3, 서열 4, 서열 5, 서열 6, 서열 7, 서열 8, 서열 9, 서열 10, 서열 13, 서열 14, 서열 15, 서열 16, 서열 17, 서열 18, 서열 19, 서열 20, 서열 21, 서열 22, 서열 23, 서열 24, 서열 36, 서열 37, 서열 38, 서열 39, 서열 40, 서열 41, 서열 42, 서열 43, 서열 62, 서열 72, 서열 73, 서열 74, 서열 75, 서열 76, 서열 77, 서열 78, 서열 79, 서열 80, 서열 81, 서열 82, 서열 83, 서열 84, 서열 85, 서열 86, 서열 87 및 서열 88로 이루어진 군 중에서 선택된 아미노산

서열의 DNA 코딩 영역의 적어도 일부에 상응하는 적어도 하나의 RNA 또는 DNA 올리고뉴클레오티드 분자를 제작하는 단계를 포함하는 단백질 서브유닛을 코딩하는 유전자의 단리 방법.

청구항 31

5' 에서 3' 프로모터, 단백질을 코딩하는 DNA 서열 및 전사 터미네이터를 갖는 키메라 DNA 구조물을 제작한 다음 키메라 DNA 구조물을 숙주내로 전달하는 것을 포함하는, *포토라프투스* 속의 정제된 박테리아 배양물에 의해 생산되는 단백질을 원핵세포 또는 진핵세포 숙주내에서 단백질이 곤충에 대해 작용 활성을 가질 수 있는 유효량으로 발현시키는 방법.

청구항 32

제31항에 있어서, 단백질이 딱정벌레목, 나비목, 파리목, 동시목, 벌목, 방시목 및 응애목으로 이루어진 군 중에서 선택된 곤충에 대해 작용 활성을 갖는 방법.

청구항 33

제31항에 있어서, DNA 서열에 의해 코딩되는 단백질이 서열 1, 서열 2, 서열 3, 서열 4, 서열 5, 서열 6, 서열 7, 서열 8, 서열 9, 서열 10, 서열 13, 서열 14, 서열 15, 서열 16, 서열 17, 서열 18, 서열 19, 서열 20, 서열 21, 서열 22, 서열 23, 서열 24, 서열 36, 서열 37, 서열 38, 서열 39, 서열 40, 서열 41, 서열 42, 서열 43, 서열 62, 서열 72, 서열 73, 서열 74, 서열 75, 서열 76, 서열 77, 서열 78, 서열 79, 서열 80, 서열 81, 서열 82, 서열 83, 서열 84, 서열 85, 서열 86, 서열 87 및 서열 88로 이루어진 군 중에서 선택된 N-말단 아미노산 서열을 갖는 방법.

청구항 34

제31항에 있어서, DNA 서열에 의해 코딩되는 단백질이 서열 12, 서열 26, 서열 28, 서열 30, 서열 32, 서열 34, 서열 35, 서열 47, 서열 49, 서열 51, 서열 53, 서열 55, 서열 57, 서열 59 및 서열 61로 이루어진 군 중에서 선택된 아미노산 서열을 포함하는 방법.

청구항 35

숙주내에서 활성인 5' 에서 3' 전사 프로모터; 곤충에 대해 작용 활성을 갖는 포토라프투스 단백질을 코딩하는 DNA 서열; 및 전사 터미네이터를 포함하는, 원핵세포 또는 진핵세포 숙주내 발현에 적합한 키메라 DNA 구조물.

청구항 36

제35항에 있어서, DNA 서열에 의해 코딩되는 단백질이 서열 1, 서열 2, 서열 3, 서열 4, 서열 5, 서열 6, 서열 7, 서열 8, 서열 9, 서열 10, 서열 13, 서열 14, 서열 15, 서열 16, 서열 17, 서열 18, 서열 19, 서열 20, 서열 21, 서열 22, 서열 23, 서열 24, 서열 36, 서열 37, 서열 38, 서열 39, 서열 40, 서열 41, 서열 42, 서열 43, 서열 62, 서열 72, 서열 73, 서열 74, 서열 75, 서열 76, 서열 77, 서열 78, 서열 79, 서열 80, 서열 81, 서열 82, 서열 83, 서열 84, 서열 85, 서열 86, 서열 87 및 서열 88로 이루어진 군 중에서 선택된 N-말단 아미노산 서열을 갖는 키메라 DNA 구조물.

청구항 37

제35항에 있어서, DNA 서열에 의해 코딩되는 단백질이 서열 12, 서열 26, 서열 28, 서열 30, 서열 32, 서열 34, 서열 35, 서열 47, 서열 49, 서열 51, 서열 53, 서열 55, 서열 57, 서열 59 및 서열 61로 이루어진 군 중에서 선택된 아미노산 서열을 갖는 키메라 DNA 구조물.

청구항 38

제35항에 있어서, *포토라프투스 루미네센스* 단백질을 코딩하는 DNA 서열이 서열 11, 서열 25, 서열 27, 서열 29, 서열 31, 서열 33, 서열 46, 서열 48, 서열 50, 서열 52, 서열 54, 서열 56, 서열 58 및 서열 60으로 이루어진 군 중에서 선택되는 키메라 DNA 구조물.

청구항 39

제35항에 있어서, 숙주가 바콜로바이러스 또는 식물 세포인 키메라 DNA 구조물.

청구항 40

포토라프투스 속의 박테리아에 의해 생산되고 곤충에 대해 작용 활성을 갖는 단백질의 유효량을 코딩할 수 있는 DNA 분자를 포함하는, 단리되고 실질적으로 정제된 제제.

청구항 41

제40항에 있어서, 박테리아가 *포토라프투스 루미네센스*인 제제.

청구항 42

서열 1, 서열 2, 서열 3, 서열 4, 서열 5, 서열 6, 서열 7, 서열 8, 서열 9, 서열 10, 서열 13, 서열 14, 서열 15, 서열 16, 서열 17, 서열 18, 서열 19, 서열 20, 서열 21, 서열 22, 서열 23, 서열 24, 서열 36, 서열 37, 서열 38, 서열 39, 서열 40, 서열 41, 서열 42, 서열 43, 서열 62, 서열 72, 서열 73, 서열 74, 서열 75, 서열 76, 서열 77, 서열 78, 서열 79, 서열 80, 서열 81, 서열 82, 서열 83, 서열 84, 서열 85, 서열 86, 서열 87 및 서열 88로 이루어진 군 중에서 선택된 N-말단 아미노산 서열을 갖는, *포토라프투스* 또는 *포토라프투스 루미네센스*에 의해 생산된 단백질을

포함하는 정제된 제제.

청구항 43

서열 1, 서열 2, 서열 3, 서열 4, 서열 5, 서열 6, 서열 7, 서열 8, 서열 9, 서열 10, 서열 13, 서열 14, 서열 15, 서열 16, 서열 17, 서열 18, 서열 19, 서열 20, 서열 21, 서열 22, 서열 23, 서열 24, 서열 38, 서열 39, 서열 40, 서열 41, 서열 42, 서열 43, 서열 62, 서열 72, 서열 73, 서열 74, 서열 75, 서열 76, 서열 77, 서열 78, 서열 79, 서열 80, 서열 81, 서열 82, 서열 83, 서열 84, 서열 85, 서열 86, 서열 87 및 서열 88로 이루어진 군 중에서 선택된 N-말단 아미노산 서열을 갖는 단백질을 포함하는 정제된 단백질 제제.

청구항 44

서열 12, 서열 26, 서열 28, 서열 30, 서열 32, 서열 34, 서열 35, 서열 47, 서열 49, 서열 51, 서열 53, 서열 55, 서열 57, 서열 59 및 서열 61로 이루어진 군 중에서 선택된 단백질을 포함하는 정제된 단백질 제제.

청구항 45

천연 숙주로부터 단리된, 서열 11, 서열 25, 서열 27, 서열 29, 서열 31, 서열 33, 서열 46, 서열 48, 서열 50, 서열 52, 서열 54, 서열 56, 서열 58 및 서열 60으로 이루어진 군 중에서 선택된 DNA 서열을 포함하는 정제된 DNA 제제.

청구항 46

18 kDa 내지 약 230 kDa; 약 160 kDa 내지 약 230 kDa; 100 kDa 내지 약 160 kDa; 약 80 kDa 내지 약 100 kDa; 또는 약 50 kDa 내지 약 80 kDa의 대략적인 분자량을 갖는 적어도 하나의 서브유닛을 갖는 *포토라프두스 루미네센스* 단백질을 포함하는 정제된 단백질 제제.

청구항 47

약 280 kDa의 대략적인 분자량을 갖는 적어도 하나의 서브유닛을 갖는 *포토라프두스 루미네센스* 단백질을 포함하는 정제된 단백질 제제.

청구항 48

ATCC 55397을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 49

제48항에 있어서, 배양물이 곤충에 대해 작용 활성을 갖는 단백질 독소를 생산하는 ATCC 55397의 변종인 배양물.

청구항 50

곤충에 대해 작용 활성을 갖는 포토라프두스 단백질의 유효량을 발현하는 능력을 식물에 부여하는 키메라 인공 유전자 구조물을 그의 게놈내에 포함하는 유전자전환 식물.

청구항 51

제50항에 있어서, 식물이 미립자 상에 피복된 유전 물질의 세포내로의 직접적인 도입촉진, 아그로 박테리아 (*Agrobacteria*) 기법, 휘스커스 기법 또는 일렉트로포레이션 기법을 사용하여 형질전환 되는 유전자전환 식물.

청구항 52

제50항에 있어서, 선택 마커가 카나마이신, 네오마이신, 글리포세이트, 히그로마이신, 메토크세이트, 포스포노트리신(비알로포스), 클로로술푸론, 브로목시닐, 달라폰 등으로 이루어진 군 중에서 선택되는 유전자전환 식물.

청구항 53

제50항에 있어서, 프로모터가 옥토펜 신타제, 노팔린 신타제, 만노핀 신타제, 35S, 19S, 35T, 리볼로스-1,6-비스포스페이트(RUBP) 카르복실라제의 작은 서브유닛(ssu), 베타-콘글리시닌, 파제올린, 알코올 데하이드로게나제(ADH), 열 속크, 유비퀴틴, 제인, 올레오신, 나핀 또는 아실 캐리어 단백질(ACP)로 이루어진 군 중에서 선택되는 유전자전환 식물.

청구항 54

제50항에 있어서, 배아 조직, 칼루스 조직 유형 I 또는 II, 배축조직, 분열조직 또는 탈분화중의 식물 조직이 유전자전환 식물을 제조하는데 사용되는 유전자전환 식물.

청구항 55

제50항에 있어서, 키메라 유전자가 곤충에 대해 작용 활성을 갖는 포토라프두스 단백질을 코딩하는 DNA 서열이고, 유전자의 적어도 하나의 코돈이 식물에 바람직한 코돈으로 변형된 유전자전환 식물.

청구항 56

곤충이 섭취하는 유전자전환 식물에 의해 생산된 단백질 독소의 유효량을 곤충에게 경구적으로 전달하는 것을 포함하는 곤충 방제 방법.

청구항 57

포토람두스 속의 정제된 박테리아 배양물로부터의 정제된 DNA 서열을 포함하는 물질의 조성물.

청구항 58

H9를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 59

Hb를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 60

Hm을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 61

HP88을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 62

NC-1를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 63

W30을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 64

WIR을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 65

B2를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 66

*피. 지알란드리카*를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 67

*피. 헤피알루스*를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 68

HB-Arg를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 69

HB 오스웨고를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 70

HB 레위스톤을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 71

K-122을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 72

HMGD을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 73

인디커스를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 74

GD을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 75

PWH-5을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 76

메기디스를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 77

HF-85을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 78

에이. 카우스를 포함하는 실질적으로 순수한 미생물 배양물.

청구항 79

MP1을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 80

MP2을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 81

MP3을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 82

MP4을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 83

MP5을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 84

GL98을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 85

GL155을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 86

GL101을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 87

GL138을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 88

GL217을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 89

GL257을 포함하는 실질적으로 순수한 미생물 배양물.

청구항 90

- a) 6개 이상의 아미노산인 단백질의 단편을 단리하고,
- b) 상기 단백질 단편으로 포유류 종을 면역화시키고,
- c) 상기 포유류 종의 비장으로부터 작용 활성을 갖는 단백질 단편에 대한 항체 또는 그를 함유하는 혈청을 회수하는 것을 포함하는, 엔테로박테라카에과의 박테리아에 의해 제조된 작용 활성을 갖는 단백질의 일부인 단백질 단편에 대한 항체를 제조하는 방법.

청구항 91

제90항에 있어서, 상기 단백질 단편이 서열 63, 서열 64, 서열 65, 서열 66, 서열 67, 서열 68, 서열 69, 서열 70 및 서열 71로 이루어진 군 중에서 선택되는 방법.

청구항 92

제90항에 있어서, 상기 박테리아가 *포토람두스* 속인 방법.

청구항 93

제90항에 있어서, 상기 박테리아가 *포토람두스 루미네센스*인 방법.

청구항 94

- a) 30개 이상의 뉴클레오티드를 갖는 DNA 서열의 단편을 단리하고,
- b) 방사성 또는 화학제제로 상기 DNA 단편을 표지하고,
- c) 엔테로박테라카에 cDNA 또는 엔테로박테라카에 게놈 라이브러리인 DNA 라이브러리에 상기 DNA 단편을 하이브리드시키고,
- d) 작용 활성을 갖는 단백질을 코딩하는 라이브러리의 DNA에 하이브리드된 단편을 선별하는 것을 포함하는 엔테로박테라카에과의 박테리아에 의해 생산된 작용 활성을 갖는 단백질의 일부를 코딩

하는 DNA 단편의 선별 방법.

청구항 95

제94항에 있어서, 상기 박테리아가 *포토람두스* 속인 방법.

청구항 96

제95항에 있어서, 상기 박테리아가 *포토람두스 루미네센스*인 방법.

청구항 97

- a) 12개 이상의 뉴클레오티드를 갖는 DNA 단편인 둘 이상의 프라이머를 단리하고,
- b) 단계 a)의 프라이머를 사용하여, 폴리머라제 연쇄 증합반응 기법으로 엔테로박테라카에 유래의 DNA 단편을 증폭하여, 상기 DNA 단편을 정제하고,
- c) 상기 정제된 DNA 단편을 방사성 또는 화학 제제로 표지하고,
- d) 엔테로박테라카에 cDNA 또는 엔테로박테라카에 게놈 라이브러리인 DNA 라이브러리에 상기 정제된 DNA 단편을 하이브리드시키고,
- e) 상기 라이브러리로 부터 정제된 DNA 단편과 동일하거나 더 큰 크기의 DNA 단편을 선별하는 것을 포함하며, 여기서 선별된 DNA 단편 또는 그의 일부는 작용 활성을 갖는 단백질을 코딩하는 것인 엔테로박테라카에 과의 박테리아에 의해 생산된 작용 활성을 갖는 단백질의 일부를 코딩하는 DNA 단편을 선별하는 방법.

청구항 98

제97항에 있어서, 상기 박테리아가 *포토람두스* 속인 방법.

청구항 99

제98항에 있어서, 상기 박테리아가 *포토람두스 루미네센스*인 방법.

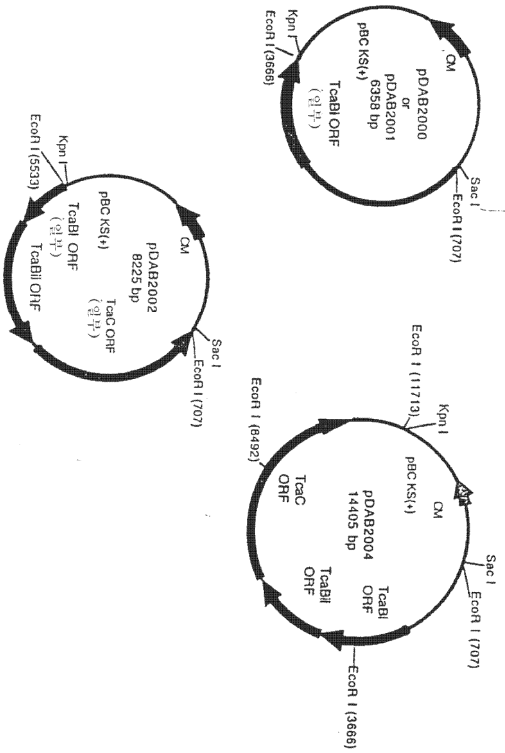
도면

도면1

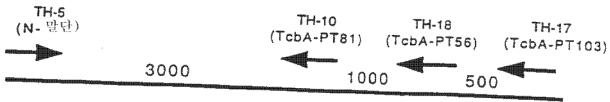
```

1  ATG  CAG  GAT  TGT  CCG  GAA  GTA  TCG  ATT  ACA  ACG  CAG  TCA  CTT  CCC  AAA  GGT  GGC  GGT
   TAC  GTC  CTA  ACA  GCG  CTT  CAT  AGC  TAA  TGT  TCC  GAC  AGT  GAA  GCG  TTT  CCA  CCG  CCA
19  Het  Gln  Asp  Cys  Pro  Glu  Val  Ser  Ile  Thr  Thr  Leu  Ser  Leu  Pro  Lys  Gly  Gly  Gly
   P235H
58  GCT  ATC  AAT  GGC  ATG  GCA  GAA  GCA  CAG  MAT  GCT  GGC  GGC  CCT  GAT  GGA  ATG  GGC  TCC
   CGA  TAG  TTA  CCG  TAC  CTT  CTT  CGT  GAC  TTA  CCA  CCG  CGG  GGA  CTA  CCT  TAC  CCG  ACG
208  Ala  Ile  Asn  Gly  Het  Gly  Glu  Ala  Leu  Leu  Asn  Ala  Ala  Gly  Pro  Asp  Gly  Het  Ala  Ser
115  GTA  TCT  CTG  CCA  TTA  CCG  CTT  TCG  ACG  GGC  AGA  GAG  ACG  GCT  CCT  GGA  TTA  TCG  CTG
398  Leu  Ser  Leu  Pro  Leu  Pro  Leu  Ser  Thr  Gly  Arg  Gly  Thr  Ala  Pro  Gly  Leu  Ser  Leu
172  ATT  TAC  AGC  AAC  AGT  GCA  GGT  AAT  GGG  CGT  TTC  GGC  ATC  GTC  TGG  CAA  TGC  GGT  GTT
   TAA  ATG  TCG  TTG  TCA  CCG  CCA  TTA  CCG  GGA  AAG  CCG  TAG  GCG  ACC  GTT  ACG  CCA  CAA
588  Ile  Tyr  Ser  Asn  Ser  Ala  Gly  Asn  Gly  Pro  Phe  Gly  Ile  Gly  Trp  Gln  Cys  Gly  Val
229  ATG  TCC  ATT  AGC  CCA  CCG  ACC  CAA  CAT  GAG  CTT  CAA  CAT  TCA  CCA  CTT
   TAC  ACG  TAA  TCG  GCT  GCG  TGG  GTT  GAA  CCG  GAA  GTT  GAA  ACT  GCT  CCA
719  Het  Ser  Ile  Ser  Arg  Arg  Thr  Gln  His  Gly  Leu  Gln  His  ...  Arg  Arg
   P235H
    
```

도면2



도면3



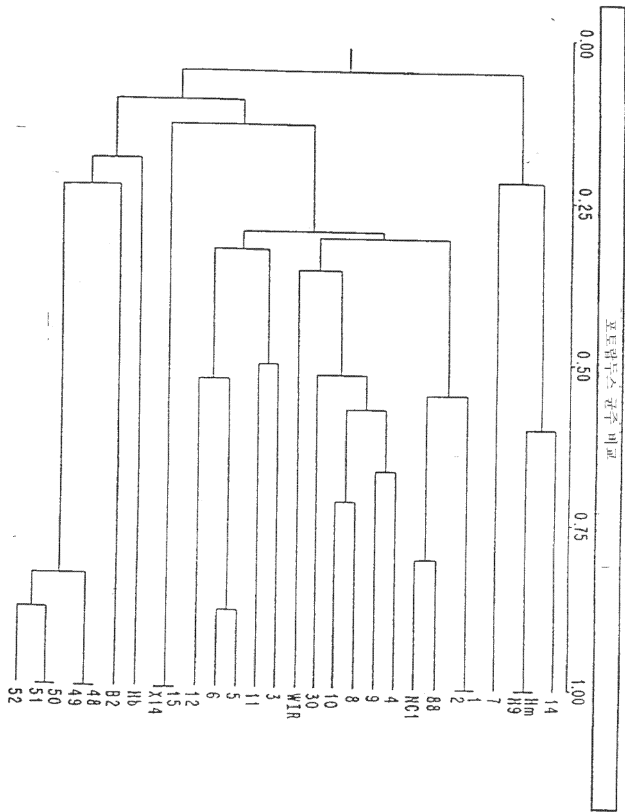
도면4a

| | | | | | |
|--------|------------|------------|------------|------------|-------------|
| TcbA | 1740 | 1750 | 1760 | 1770 | 1780 |
| | SSAQALKNDS | EPMDFSGANA | LYFWELFYIT | PMMMAHRLIQ | EQNFDAANHW |
| TcaBi | gS | nPvDFSGpyg | iYlWEiFfhi | PfIvtvRmqt | EQryedAdtW> |
| | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| TcbA | 1790 | 1800 | 1810 | 1820 | 1830 |
| | FRYVWSPSGY | IVDGKLIAYH | WNVRPLEEDT | SWNAQQLDST | DPDAVAQDDP |
| TcaBi | rdangql | | | | |
| | 490 | 510 | 520 | 530 | |
| | ykYifrsaGY | ImDGskprY- | WNVmPLqLDT | aWdttQpatT | DPDviAmaDP> |
| | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| TcbA | 1840 | 1850 | 1860 | 1870 | 1880 |
| | MHYKVAIFMA | TLDLLMARGD | AAVRQLERDT | LAEAKMWTQ | ALNLLGDEPO |
| TcaBi | 540 | 550 | 560 | 570 | 580 |
| | MHYKlaiFlh | TLDLLIARGD | sAYRQLERDT | LWEAKQyYiQ | AqqLLQprPd> |
| | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| TcbA | 1890 | 1900 | 1910 | 1920 | 1930 |
| | VMLSTTWANP | TLGNAASKIT | QQVRQQVLTQ | LRLNSRVKTP | LLGTANSLTA |
| TcaBi | ihttntWpNP | TLsk> | | | |
| | ~~~~~ | ~~~~~ | | | |
| TcbA | 1940 | 1950 | 1960 | 1970 | 1980 |
| | LFLPQENSKL | KGYWRILAQR | MFNLRHNLSI | DGQPLSLPLY | AKPADPKALL |
| TcaBii | 20 | 30 | 40 | 50 | 60 |
| | _FLPpyNgvL | lGYWdkLeLR | lyNLRHNLSI | DGQPLnLPLY | AtPvDPKtLq> |
| | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| TcbA | 1990 | 2000 | 2010 | 2020 | 2030 |
| | SAAVSASQGG | ADLPKAPLTI | HRFPQMLEGA | RGLVNQLIQF | GSSLLGYSER |
| TcaBii | 70 | 80 | 90 | 100 | 110 |
| | rqqaggdgtG | sspaggqgsv | qRyPllvErA | RsaVslLtQF | GnSLqtLEh> |
| | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| TcbA | 2040 | 2050 | 2060 | 2070 | 2080 |
| | QDAEAMSQLL | QTQASELILT | SIRMQDNQLA | ELDSEKTALQ | VSLAGVQQRF |
| TcaBii | 120 | 130 | 140 | 150 | 160 |
| | QDnEkmtiLL | QTQgeailkh | ghdiQqNnLk | gLqhsLTALQ | aSrdGdtLRq> |
| | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |

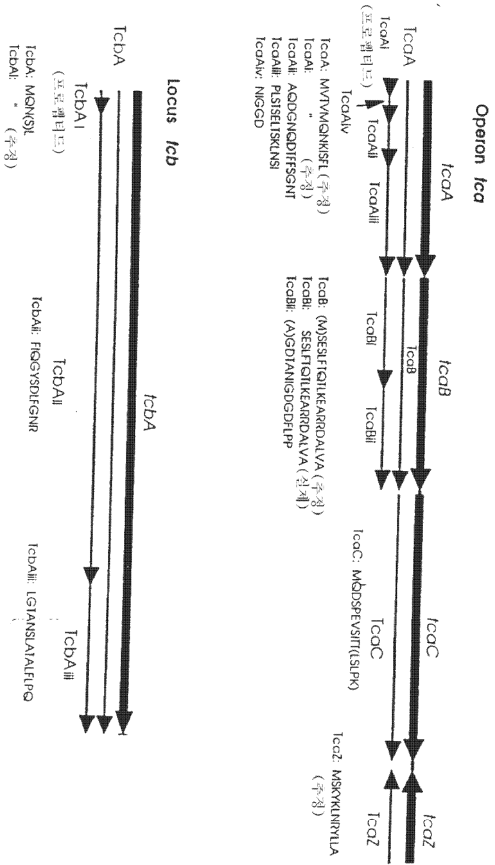
도면4b

| | | | | | |
|--------|------------|------------|-------------|------------|-------------|
| TcbA | 2090 | 2100 | 2110 | 2120 | 2130 |
| | DSYSQLYEEN | INAGEQRALA | LRSESAIESQ | GAQISRMAQA | GVIDMAPNIFG |
| TcaBii | 170 | 180 | 190 | 200 | 210 |
| | khYSdLingg | lsAaEiagLt | LRStamI-tn | GvatglliaG | GinavPNvFG> |
| TcbA | 2140 | 2150 | 2160 | 2170 | 2180 |
| | LADGGMHYGA | IAYAIADGIE | LSASAKMVA | EKVAQSEIYR | RRRQEWKIQR |
| TcaBii | 220 | 230 | 240 | 250 | 260 |
| | LANGGsewGA | pligsggatq | vgAgiqdgsA | gisevtagYq | RRqeEWalQR> |
| TcbA | 2190 | 2200 | 2210 | 2220 | 2230 |
| | DNAQAEINQL | NAQLESLSIR | REAAEMQKEY | LKTQQAQAQA | QLTFLRSKFS |
| TcaBii | 270 | 280 | 290 | 300 | 310 |
| | DiAdnEItQL | dAQiqSLqeq | itmAqkQitl | seTeQAnAQA | iydlqttrFt> |
| TcbA | 2240 | 2250 | 2260 | 2270 | 2280 |
| | NQALYSWLRG | RLSGTYFQFY | DLAVSRCLMA | EQSYQWEAND | NSISFVKFGA |
| TcaBii | 320 | 330 | 340 | 360 | |
| | gQALYnWmaG | RLSalYyQmY | DstlpiCLqP | kaalvqEgek | esdSifqvvp> |
| TcbA | 2290 | 2300 | 2310 | 2320 | 2330 |
| | WQGYAGLLC | GEALIQNLAQ | MEEAYLKWES | RALEVERTVS | LAVVYDSLEG |
| TcaBii | 370 | 380 | 390 | 400 | 410 |
| | WndlwgLLa | GEGLsseLqk | ldaiwLargg | igLEaiRTVS | Ldtlfgt--G> |
| TcbA | 2340 | 2350 | 2360 | 2370 | 2380 |
| | NDRFNLAEQI | PALLDRKGST | AGTKKNGLSL | ANAILSASVK | LSDLKLGTDY |
| TcaBii | 420 | 430 | 440 | 450 | |
| | ----tLsEnI | nkvLn-GETv | spsggvtLaL | tgdlfqAtld | LSqLgLdnsY> |
| TcbA | 2390 | 2400 | 2410 | 2420 | 2430 |
| | PDSIVGSNKV | RRIKQISVSL | PALVGPYQDV | QAMLSYGGST | QLPKGCSALA |
| TcaBii | 460 | 470 | 480 | 490 | 500 |
| | -n--lGneKk | RRIKrIaVtL | PtLLGFPYQDI | eATLvmGaea | aLshGvndgg> |
| TcbA | 2440 | 2450 | 2460 | 2470 | |
| | VSHGTNDSGQ | FQLDFNDGKY | LPFEGIALDD | QGTLNLQFPN | |
| TcaBii | 510 | 520 | 530 | | |
| | rIvtdfndsR | F-LpF-eGrd | attgtleLn> | | |

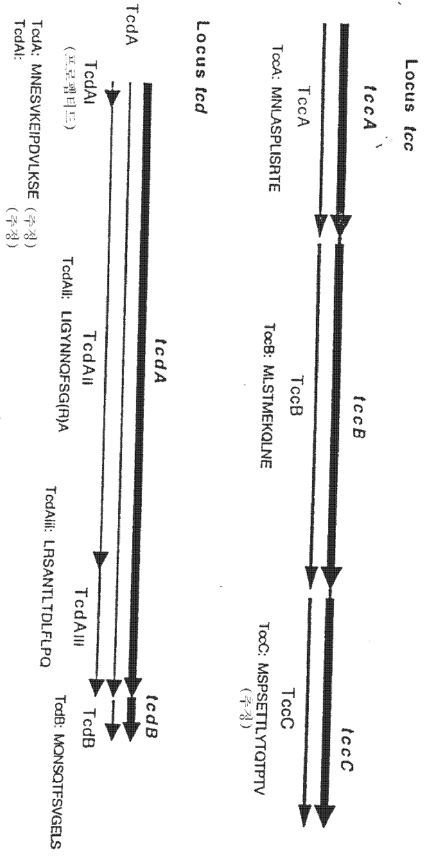
도면5



도면6a



9면도



도면7

