



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0069809
(43) 공개일자 2011년06월23일

(51) Int. Cl.

C07K 14/195 (2006.01) C12N 15/31 (2006.01)

C12S 3/04 (2006.01) C12R 1/01 (2006.01)

(21) 출원번호 10-2011-7008348

(22) 출원일자(국제출원일자) 2009년09월11일

심사청구일자 2011년04월12일

(85) 번역문제출일자 2011년04월12일

(86) 국제출원번호 PCT/KR2009/005172

(87) 국제공개번호 WO 2010/030140

국제공개일자 2010년03월18일

(30) 우선권주장

61/096,666 2008년09월12일 미국(US)

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가 1

(72) 발명자

김경현

서울특별시 서초구 서초4동 삼풍아파트 18동 703호

최인걸

서울특별시 동작구 사당1동 419-92

(뒷면에 계속)

(74) 대리인

특허법인충현

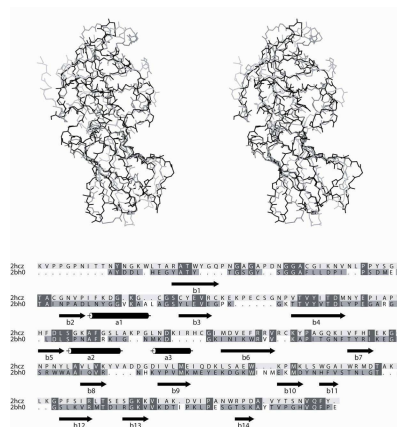
전체 청구항 수 : 총 17 항

(54) 셀룰로스 확장활성용 원핵생물 유래 익스펜신 단백질 및 이를 포함하는 셀룰로스 분해조성물

(57) 요약

셀룰로스를 확장시킬 수 있는 원핵생물 유래 익스펜신 단백질은 종래의 식물 유래 익스펜신 단백질과 동일한 기능을 수행하면서도 식물 유래 익스펜신 단백질과는 달리 대량생산이 가능하고 생산비용을 현저히 저감시킬 수 있다. 특히 목질계 바이오매스를 셀룰라제로 가수분해하여 포도당을 얻고자 할 때 셀룰라제에 의한 셀룰로스 가수분해 속도를 높여 당 수율을 현저히 향상시킬 수 있다. 이를 통해 실제 바이오에너지 생산에 있어서 원핵생물 유래 익스펜신으로 효소의 양을 줄임으로써 바이오에너지의 생산비용을 저감할 수 있다. 또한, 펄프, 면 섬유(예, 청바지) 등의 조직을 연화시키는 특성이 있어 제지산업에서의 바이오펄핑이나 섬유산업에서의 바이오스토닝과 같은 목적으로 이용될 수 있다.

대표도 - 도1



(72) 발명자

이희진

서울특별시 성북구 안암동 고려대학교

김은실

서울특별시 성북구 안암동 고려대학교

특허청구의 범위

청구항 1

셀룰로스 확장활성을 가지는, 셀룰로스 확장활성용 원핵생물 유래 익스팬신 단백질.

청구항 2

제1항에 있어서, 상기 원핵생물은 바실러스 서브틸리스, 하헬라 제주엔시스(*Hahella chejuensis*, KCTC 2396), 디티오스테리움 디스코이데움(*Dictyostelium discoideum*), 네오사르토리야 피셰리(*Neosartorya fischeri*), 아스퍼질러스 푸미가투스(*Aspergillus fumigatus*), 아스퍼질러스 클라바투스(*Aspergillus clavatus*), 아스퍼질러스 오라이제(*Aspergillus oryzae*), 아스퍼질러스 테레우스(*Aspergillus terreus*), 페니실리움 크리소제눔(*Penicillium chrysogenum*), 아스퍼질러스 나이저(*Aspergillus niger*), 에케리첼라 니둘란스(*Emericella nidulans*), 마그나포르테 그리시아(*Magnaporthe grisea*), 피레노포라 트리티시-레펜티스(*Pyrenophora tritici-repentis*), 파오스파에리아 노도룸(*Phaeosphaeria nodorum*), 스크레로티니아 스크레로티오룸(*Sclerotinia sclerotiorum*), 프란키아(*Frankia* sp.), 스트렙토마이시스 스비세우스(*Streptomyces sviveus*), 소란지움 셀룰로쭈름(*Sorangium cellulosum*), 스티그마텔라 아우렌티아카(*Stigmatella aurantiaca*), 플레시오시스티스 나시피카(*Plesiocystis pacifica*), 므소코쿠스 잔터스(*Myxococcus xanthus*), 렙토쓰릭스 콜로드니(*Leptothrix cholodnii*), 로제이플렉서스(*Roseiflexus* sp.), 로제이플렉서스 카스텐홀지(*Roseiflexus castenholzii*), 클로로플렉서스 아우란티아쿠스(*Chloroflexus aurantiacus*), 헤르페토시폰 아우란티아쿠스(*Herpetosiphon aurantiacus*), 에시도보락스 아베네(*Acidovorax avenae* subsp.), 펙토박테리움 아트록세티쿰(*Pectobacterium atrosepticum*), 바실러스 리체니포미스(*Bacillus licheniformis*), 잔토모나스 캄페스트리스(*Xanthomonas campestris* pv. *campestris*), 바실러스 푸밀루스(*Bacillus pumilus*), 잔토모나스 오라이제(*Xanthomonas oryzae* pv. *oryzae*), 랄스토니아 솔라나시아룸(*Ralstonia solanacearum*), 클라비박터 미시가넨시스(*Clavibacter michiganensis*), 잘렐라 파스티디오사(*Xylolla fastidiosa*), 나카무렐라 멀티파리타(*Nakamurella multipartite*), 마이크로모노스포라(*Micromonospora* sp.), 카테눌리스포라 애시디필라(*Catenulispora acidiphila*) 및 디키야 지(*Dickeya zeae*)로 구성된 군으로부터 선택된 원핵생물인 것을 특징으로 하는, 셀룰로스 확장활성용 원핵생물 유래 익스팬신 단백질.

청구항 3

제2항에 있어서, 상기 원핵생물은 바실러스 서브틸리스, 스티그마텔라 아우렌티아카, 잔토모나스 오라이제 또는 하헬라 제주엔시스인 것을 특징으로 하는, 셀룰로스 확장활성용 원핵생물 유래 익스팬신 단백질.

청구항 4

제1항에 있어서, 상기 익스팬신 단백질은 서열번호 1 내지 47에 기재된 군으로부터 선택된 아미노산 서열 중 어느 하나를 포함하는 것을 특징으로 하는, 셀룰로스 확장활성용 원핵생물 유래 익스팬신 단백질.

청구항 5

제4항에 있어서, 상기 익스팬신 단백질은 서열번호 20, 22, 30 또는 35에 기재된 아미노산 서열인 것을 특징으로 하는, 셀룰로스 확장활성용 원핵생물 유래 익스팬신 단백질.

청구항 6

제1항 내지 제5항 중 어느 한 항의 셀룰로스 확장활성용 원핵생물 유래 익스팬신 단백질을 포함하는 셀룰로스

분해조성물.

청구항 7

제6항에 있어서, 상기 셀룰로스 분해조성물은 셀룰로스 분해효소를 더 포함하는 것을 특징으로 하는 셀룰로스 분해조성물.

청구항 8

제7항에 있어서, 상기 셀룰로스 분해효소는 셀룰라제, 셀로비오하이드로라제, 엔도글루카나제 및 셀로비아제 으로 구성된 군으로부터 선택된 어느 하나 이상인 것을 특징으로 하는 셀룰로스 분해조성물.

청구항 9

제7항에 있어서, 상기 셀룰로스 분해조성물은 1 g의 셀룰로스에 대하여 셀룰로스 분해효소 0.01~0.05 FPU 및 원핵생물 유래 익스팬신 200~400 μ g을 포함하는 것을 특징으로 하는 셀룰로스 분해조성물.

청구항 10

셀룰로스에 제6항 내지 제9항 중 어느 한 항의 셀룰로스 분해조성물을 첨가한 후 40~60 $^{\circ}$ C에서 반응시키는 단계를 포함하는 셀룰로스 분해방법.

청구항 11

셀룰로스에 제6항 내지 제9항 중 어느 한 항의 셀룰로스 분해조성물을 첨가한 후 pH 7 이하에서 반응시키는 단계를 포함하는 셀룰로스 분해방법.

청구항 12

하헬라 제주엔시스(*Halella chejuensis*) 유래의 익스팬신(EXLX1) 아미노산 서열과 60% 이상의 상동성을 가지며 셀룰로스 확장활성을 가지는, 원핵생물 유래 익스팬신 단백질.

청구항 13

목질계 바이오매스를 제6항 내지 제9항 중 어느 한 항의 셀룰로스 분해조성물로 처리하여 상기 목질계 바이오매스에 포함된 셀룰로스를 분해하여 환원당을 생산하는 단계를 포함하는 바이오에너지 생산방법.

청구항 14

종이 또는 펄프를 제6항 내지 제9항 중 어느 한 항의 셀룰로스 분해조성물로 처리하여 상기 종이 또는 펄프를 연화시키는 단계를 포함하는 종이 또는 펄프의 생산방법.

청구항 15

섬유 또는 직물을 제6항 내지 제9항 중 어느 한 항의 셀룰로스 분해조성물로 처리하여 상기 섬유 또는 직물을 연화시키는 단계를 포함하는 섬유 또는 직물의 처리방법.

청구항 16

- a) 식물 유래 익스팬신과 구조적 유사성을 가지는 원핵생물 유래 단백질을 찾는 단계;
- b) 상기 원핵생물 유래 단백질을 클로닝하는 단계; 및
- c) 상기 클로닝된 원핵생물 유래 단백질을 균주에서 발현하는 단계를 포함하는, 원핵생물 유래 익스팬신 단백질의 대량생산방법.

청구항 17

제16항에 있어서, 상기 식물 유래 익스팬신은 서열번호 48의 아미노산 서열을 포함하는 것을 특징으로 하는, 원핵생물 유래 익스팬신 단백질의 대량생산방법.

명세서

기술분야

[0001] 본 발명은 셀룰로스 확장활성용 원핵생물 유래 익스팬신 단백질 및 이를 포함하는 셀룰로스 분해조성물에 관한 것으로서, 보다 상세하게는 셀룰로스의 확장활성을 가지는 원핵생물 유래 익스팬신 단백질 및 이를 포함하는 셀룰로스 분해조성물에 관한 것이다.

배경기술

[0002] 셀룰로스를 에탄올로 발효될 수 있는 당으로 가수분해함으로써 목질계 바이오매스를 이용하고자 하는 연구개발이 50 년 이상이 진행되고 있지만(Kim et al. 2005; Nguyen et al. 1998; Sherrad and Kressman 1945; Torget et al. 1991), 이러한 기술의 산업화는 여전히 과제로 남아 있다. 이는 헤미셀룰로스와 리그닌의 복합체에 의해 셀룰로스가 화학적 및 생화학적으로부터 단단히 보호를 받고 있어 셀룰로스를 당화하는 것이 쉽지 않은 것이 주요인으로 작용하기 때문이다. 그러므로, 셀룰로스에 바이오매스를 바이오연료나 화학소재로 전환하기 위한 전단계에서 이루어지는 전처리 및 셀룰로스 분해효소를 이용한 당화공정은 에탄올 생산에 있어 높은 비용을 요하는 공정이 된다(Gregg et al. 1998; Lynd et al. 2008; Wingren et al. 2003). 따라서, 경제적인 관점에서 볼 때, 셀룰로스의 가수분해 속도 및 정도를 개선하는 효과적인 셀룰로스 분해효소(셀룰라제)의 개발이 필수적이다.

[0003] 바이오매스의 가수분해에는 다량의 셀룰라제가 요구되므로, 리그노셀룰로스의 비용효율적인 처리에 있어 셀룰라제 효소의 비용이 가장 큰 걸림돌이 된다. 노보자임스(Novozymes) 사에 의하면, 동일한 양의 에탄올을 리그노셀룰로스로부터 생산하는 경우, 전분에 비해 40 내지 100 배 중량의 효소가 필요하다고 한다(Merino and Cherry 2007). 셀룰로스 분해효소가 성공적으로 작용하려면 몇 가지 장애를 극복해야 한다. 즉, 불용성 기질에 접근할 수 있어야 하고, 고도로 규칙적인 고분자의 정렬을 파괴시켜야 하며, 분리된 고분자 사슬을 효소의 활성부위로 가져와야 한다. 요컨대, 셀룰로스의 가수분해에는 여러 효소에 의한 협동 공격이 요구된다.

[0004] 셀로비오하이드롤라제(CBH) 및 엔도글루카나제(EG)와 같은 셀룰로스 분해효소의 상승효과는 비교적 많이 연구된 반면(Irwin et al. 1993; Medve et al. 1994; Woodward et al. 1988), 셀룰로스 분해 단백질과 비가수분해 단백질 사이의 관계는 분명히 밝혀지지 않았다. 식물 세포벽의 성장과 세포벽 내로의 침입을 촉진하기 위하여, 식물의 익스팬신(Cosgrove 2000a; Cosgrove 2000b; McQueen-Mason and Cosgrove 1994), 진균류의 스월레닌, 그리고 선충류의 익스팬신 유사단백질(Qin et al. 2004)과 같은 일부 비가수분해 단백질은 식물 세포벽과 다당류(예컨대, 셀룰로스나 헤미셀룰로스)의 정렬상태를 느슨하게 하거나 파괴할 수 있는 것으로 알려져 있다. 이러한 "파괴" 활성은 흔히 인장강도시험기(extensometer)를 사용하여 식물 세포벽 또는 거름종이의 약화 정도를 측정함으로써 측정한다(McQueen-Mason and Cosgrove 1994; Saloheimo et al. 2002). 이러한 익스팬신 및 익스팬신 유사단백질의 파괴활성은 셀룰로스가 효소에 용이하게 접근하도록 함으로써 효소에 의한 셀룰로스의 가수분해를 촉진하는 것으로 생각된다. 익스팬신과 셀룰라제에 의한 이러한 상승효과는 2001년에 처음 발견되었으나

(Cosgrove 2001), 순수한 익스팬신의 이중 과발현(heterologous overexpression)을 통해 활성 재조합 단백질을 얻고자 하는 시도는 성공하지 못하였다. 최근에, β -익스팬신과 25~40%의 서열 유사성을 갖는 초본 화분계(grass pollen group)-2/3 알레르겐인 GR2 단백질은 트리코테르마의 엔도글루카나제와 상승작용을 하는 것으로 보고되었다(Cosgrove 2007). 익스팬신과 비교할 때, GR2는 크기가 익스팬신보다 작고(~10 kDa) 대장균에서 활성 상태로 쉽게 발현된다는 장점이 있다. 최근, 옥수수 부산물에서 정제된, 아직 확인되지 않은 비가수분해 단백질(55.78 kDa)이 거름종이의 가수분해에 있어 셀룰라제와 상승작용을 하는 것으로 밝혀졌다(Han and Chen 2007). 그러나, 식물 외의 숙주 생물에서의 활성 익스팬신의 발현은 성공적이지 못하였으며, 따라서 셀룰로스의 가수분해를 촉진하기 위한 익스팬신의 이용에는 어려움이 있다.

[0005]

결국, 식물 유래 익스팬신 단백질이 셀룰로스의 확장활성을 가져 셀룰로스의 분해를 돕는 것은 이미 알려져 있으나, 이를 숙주 생물에서 발현하기 어려워 대량생산이 극히 곤란한 문제점을 가지고 있으며(미국특허 제 6326470호), 상기 익스팬신 단백질을 진균류 등의 진핵생물에서 생산하는 경우 세포벽 확장활성이 극히 미미하여 산업상 이용하기 어려운 문제가 있었다(미국공개특허 제2003-104546호).

선행기술문헌

비특허문헌

[0006]

(비특허문헌 0001) Adney B, Baker J. 1996. Measurement of cellulase activities, Chemical Analysis and Testing Task: LAP-006. National Renewable Energy Laboratory, Golden, CO, USA.

(비특허문헌 0002) Aslanidis C, Dejong PJ. 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 18(20):6069-6074.

(비특허문헌 0003) Baker JO, King MR, Adney WS, Decker SR, Vinzant TB, Lantz SE, Nieves RE, Thomas SR, Li LC, Cosgrove DJ and others. 2000. Investigation of the cell-wall loosening protein expansin as a possible additive in the enzymatic saccharification of lignocellulosic biomass. Appl. Biochem. Biotechnol. 84-6:217-223.

(비특허문헌 0004) Berlin A, Maximenko V, Bura R, Kang KY, Gilkes N, Saddler J. 2006. A rapid microassay to evaluate enzymatic hydrolysis of lignocellulosic substrates. Biotechnol. Bioeng. 93(5):880-886.

(비특허문헌 0005) Boraston AB, Kwan E, Chiu P, Warren RAJ, Kilburn DG. 2003. Recognition and hydrolysis of noncrystalline cellulose. J. Biol. Chem. 278(8):6120-6127.

(비특허문헌 0006) Carrard G, Koivula A, Soderlund H, Beguin P. 2000. Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. Proc. Natl. Acad. Sci. U.S.A 97(19):10342-10347.

(비특허문헌 0007) Cosgrove DJ. 2000a. Loosening of plant cell walls by expansins. Nature 407(6802):321-326.

(비특허문헌 0008) Cosgrove DJ. 2000b. New genes and new biological roles for expansins. Curr. Opin. Plant Biol. 3(1):73-78.

(비특허문헌 0009) Cosgrove DJ. 2001. Enhancement of accessibility of cellulose by expansins. US Patent 6326470 B1.

(비특허문헌 0010) Cosgrove DJ. 2007. Use of GR2 proteins to modify cellulosic material and to enhance enzymatic and chemical modification of cellulose. US Patent 2007/0166805 A1.

(비특허문헌 0011) Graslund S, Nordlund P, Weigelt J, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J and others. 2008. Protein production and purification. Nat. Methods 5:135-146.

(비특허문헌 0012) Gregg DJ, Boussaid A, Saddler JN. 1998. Techno-economic evaluations of a generic

wood-to-ethanol process: Effect of increased cellulose yields and enzyme recycle. *Bioresour. Technol.* 63(1):7-12.

(비특허문헌 0013) Han YJ, Chen HZ. 2007. Synergism between corn stover protein and cellulase. *Enzyme Microb. Technol.* 41(5):638-645.

(비특허문헌 0014) Holm L, Park J. 2000. DaliLite workbench for protein structure comparison. *Bioinformatics* 16(6):566-567.

(비특허문헌 0015) Irwin DC, Spezio M, Walker LP, Wilson DB. 1993. Activity studies of eight purified cellulases - specificity, synergism, and binding domain effects. *Biotechnol. Bioeng.* 42(8):1002-1013.

(비특허문헌 0016) Kende H, Bradford KJ, Brummell DA, Cho HT, Cosgrove DJ, Fleming AJ, Gehring C, Lee Y, McQueen-Mason S, Rose JKC and others. 2004. Nomenclature for members of the expansin superfamily of genes and proteins. *Plant Mol. Biol.* 55(3):311-314.

(비특허문헌 0017) Kim KH, Tucker M, Nguyen Q. 2005. Conversion of bark-rich biomass mixture into fermentable sugar by two-stage dilute acid-catalyzed hydrolysis. *Bioresour. Technol.* 96(11):1249-1255.

(비특허문헌 0018) Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller H, McMillan JM, Sheehan J and others. 2008. How biotech can transform biofuels. *Nat. Biotechnol.* 26(2):169-172.

(비특허문헌 0019) McQueen-Mason S, Cosgrove DJ. 1994. Disruption of hydrogen-bonding between plant-cell wall polymers by proteins that wall extension. *Proc. Natl. Acad. Sci. U.S.A.* 91(14):6574-6578.

(비특허문헌 0020) Medve J, Stahlberg J, Tjerneld F. 1994. Adsorption and synergism of cellobiohydrolase-I and cellobiohydrolase-II of *Trichoderma reesei* during hydrolysis of microcrystalline cellulose. *Biotechnol. Bioeng.* 44(9):1064-1073.

(비특허문헌 0021) Merino ST, Cherry J. 2007. Progress and challenges in enzyme development for Biomass utilization. *Adv. Biochem. Eng. Biotechnol.* 108: 95-120.

(비특허문헌 0022) Moser F, Irwin D, Shaolin Chen DBW. 2008. Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnol. Bioeng.* 100(6):1066-1077.

(비특허문헌 0023) Nguyen QA, Tucker MP, Boynton BL, Keller FA, Schell DJ. 1998. Dilute acid pretreatment of softwoods. *Appl. Biochem. Biotechnol.* 70-72:77-87.

(비특허문헌 0024) Petrella S, Herman R, Sauvage E, Filee P, Joris B, Charlier P. 2006. Crystal Structure of a sement derivatiove of expa from *Bacillus subtilis* at 2.5 angstrom. <http://dx.doi.org/10.2210/pdb2bh0/pdb>.

(비특허문헌 0025) Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25(13):1605-1612.

(비특허문헌 0026) Qin L, Kudla U, Roze EHA, Goverse A, Popeijus H, Nieuwland J, Overmars H, Jones JT, Schots A, Smant G and others. 2004. Plant degradation: A nematode expansin acting on plants. *Nature* 427(6969):30-30.

(비특허문헌 0027) Rudolf A, Baudel H, Zacchi G, Hahn-Hagerdal B, Liden G. 2008. Simultaneous saccharification and fermentation of steam-pretreated bagasse using *Saccharomyces cerevisiae* TMB3400 and *Pichia stipitis* CBS6054. *Biotechnol. Bioeng.* 99:783-790.

(비특허문헌 0028) Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyyssonen E, Bhatia A, Ward M, Penttila M. 2002. Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *Eur. J. Biochem.* 269(17):4202-4211.

- (비특허문헌 0029) Sampedro J, Cosgrove DJ. 2005. The expansin superfamily. *Genome Biol.* 6(12).
- (비특허문헌 0030) Sampedro J, Carey RE, Cosgrove DJ. 2006. Genome histories clarify evolution of the expansin superfamily: new insights from the poplar genome and pine ESTs. *J. Plant Res.* 119:11-21
- (비특허문헌 0031) Sherrad EC, Kressman FW. 1945. Review of processes in the United States prior to World War II. *Ind. Eng. Chem.* 37(1):5-8.
- (비특허문헌 0032) Torget R, Himmel ME, Grohmann K. 1991. Dilute sulfuric acid pretreatment of hardwood bark. *Bioresour. Technol.* 35(3):239-246.
- (비특허문헌 0033) Wingren A, Galbe M, Zacchi G. 2003. Techno-economic evaluation of producing ethanol from softwood: Comparison of SSF and SHF and identification of bottlenecks. *Biotechnol. Prog.* 19(4):1109-1117.
- (비특허문헌 0034) Woodward J, Lima M, Lee NE. 1988. The role of cellulase concentration in determining the degree of synergism in the hydrolysis of microcrystalline cellulose. *Biochem. J.* 255(3):895-899.
- (비특허문헌 0035) Xiao ZZ, Storms R, Tsang A. 2004. Microplate-based filter paper assay to measure total cellulase activity. *Biotechnol. Bioeng.* 88(7):832-837.
- (비특허문헌 0036) Yang B, Willies DM, Wyman CE. 2006. Changes in the enzymatic hydrolysis rate of Avicel cellulose with conversion. *Biotechnol. Bioeng.* 94(6):1122-1128.
- (비특허문헌 0037) Yennawar NH, Li LC, Dudzinski DM, Tabuchi A, Cosgrove DJ. 2006. Crystal structure and activities of EXPB1 (*Zea m l*), a b-expansin and group-1 pollen allergen from maize. *Proc. Natl. Acad. Sci. U.S.A.* 103(40):14664-14671.

발명의 내용

해결하려는 과제

- [0007] 본 발명은 상술한 상술한 문제점을 해결하기 위해 안출된 것으로, 본 발명의 첫번째 기술적 과제는 식물유래 셀룰로스 확장 단백질인 익스팬신이 숙주 생물에서 발현하기 어려운 문제점을 해결하기 위하여 익스팬신과 기능적, 구조적 유사성을 가지는 단백질을 원핵생물에서 찾아 이를 대량생산할 수 있는 원핵생물 유래 익스팬신 단백질을 제공하는 것이다.
- [0008] 본 발명의 두 번째 기술적 과제는 셀룰로스에 대한 셀룰로스 분해효소의 분해능이 향상된 원핵생물 유래 익스팬신 단백질을 포함하는 셀룰로스 분해조성물 및 이를 이용한 셀룰로스 분해방법을 제공하는 것이다.
- [0009] 본 발명의 세 번째 기술적 과제는 본 발명의 원핵생물 유래 익스팬신 단백질을 포함하는 셀룰로스 분해조성물의 다양한 용도를 제공하는 것이다.
- [0010] 본 발명의 네 번째 기술적 과제는, 원핵생물 유래 익스팬신 단백질의 대량생산 방법을 제공하는 것이다.

과제의 해결 수단

- [0011] 상술한 첫번째 기술적 과제를 해결하기 위하여, 본 발명은 셀룰로스 확장활성을 가지는 원핵생물 유래 익스팬신 단백질을 제공한다.
- [0012] 본 발명의 실시예에 따르면, 상기 원핵생물은 바실러스 서브틸리스, 하헬라 제주엔시스(*Hahella chejuensis*, KCTC 2396), 디티오스테리움 디스코이데움(*Dictyostelium discoideum*), 네오사르토리아 피셰리(*Neosartorya fischeri*), 아스퍼질러스 푸미가투스(*Aspergillus fumigatus*), 아스퍼질러스 클라바투스(*Aspergillus clavatus*), 아스퍼질러스 오라이제(*Aspergillus oryzae*), 아스퍼질러스 테레우스(*Aspergillus terreus*), 페니실리움 크리소제눔(*Penicillium chrysogenum*), 아스퍼질러스 나이저(*Aspergillus niger*), 에케리첼라 니둘란스(*Emericella nidulans*, 마그나포르테 그리시아(*Magnaporthe grisea*), 피레노포라 트리티시-레펜티스

(*Pyrenophora tritici-repentis*), 패오스파에리아 노도룸(*Phaeosphaeria nodorum*), 스크레로티니아 스크레로티오룸(*Sclerotinia sclerotiorum*), 프란키아(*Frankia* sp.), 스트렙토마이시스 스비세우스(*Streptomyces sviveus*), 소란지움 셀룰로쭘(*Sorangium cellulosum*), 스티그마텔라 아우렌티아카(*Stigmatella aurantiaca*), 플레시오시스티스 나시피카(*Plesiocystis pacifica*), 믹소코쿠스 잔터스(*Myxococcus xanthus*), 랩토쓰릭스 콜로드니(*Leptothrix cholodnii*), 로제이플렉서스(*Roseiflexus* sp.), 로제이플렉서스 카스텐홀지(*Roseiflexus castenholzii*), 클로로플렉서스 아우란티아쿠스(*Chloroflexus aurantiacus*), 헤르페토시폰 아우란티아쿠스(*Herpetosiphon aurantiacus*), 에시도보락스 아베네(*Acidovorax avenae* subsp), 펙토박테리움 아트로셉티쿰(*Pectobacterium atrosepticum*), 바실러스 리체니포미스(*Bacillus licheniformis*), 잔토모나스 캄페스트리스(*Xanthomonas campestris* pv. *campestris*), 바실러스 푸밀루스(*Bacillus pumilus*), 잔토모나스 오라이제(*Xanthomonas oryzae* pv. *oryzae*), 랄스토니아 솔라나시아룸(*Ralstonia solanacearum*), 클라비박터 미시가넨시스(*Clavibacter michiganensis*), 잘렐라 파스티디오사(*Xylella fastidiosa*), 나카무렐라 멀티파리티타(*Nakamurella multipartite*), 마이크로모노스포라(*Micromonospora* sp.), 카테눌리스포라 애시디필라(*Catenulispora acidiphila*) 및 디키야 지(*Dickeya zeae*)로 구성된 군으로부터 선택된 원핵생물일 수 있으며, 보다 바람직하게는 상기 원핵생물은 바실러스 서브틸리스, 스티그마텔라 아우렌티아카, 잔토모나스 오라이제 또는 하렐라 제주엔시스일 수 있다.

- [0013] 본 발명의 다른 일 실시예에 따르면, 상기 익스팬신 단백질은 서열번호 1 내지 47에 기재된 군으로부터 선택된 아미노산 서열 중 어느 하나를 포함할 수 있으며, 또는 본 발명의 단백질은 서열번호 1 내지 47의 단백질뿐 아니라 이들 단백질에 하나 이상의 치환, 결손, 역위, 전위 등과 같은 돌연변이를 통하여 본 발명의 목적인 셀룰로스 확장활성을 가지는 모든 변이 단백질을 포함한다.
- [0014] 본 발명의 또 다른 일 실시예에 따르면, 상기 익스팬신 단백질은 서열번호 20, 22, 30 또는 35에 기재된 아미노산 서열을 가질 수 있다.
- [0015] 상술한 두 번째 과제를 달성하기 위하여, 상술한 원핵생물 유래 익스팬신 단백질을 포함하는 셀룰로스 분해조성물을 제공한다.
- [0016] 본 발명의 일 실시예에 따르면, 상기 셀룰로스 분해조성물은 셀룰로스 분해효소를 더 포함할 수 있으며, 바람직하게는 상기 셀룰로스 분해효소는 셀룰라제, 셀로비오하이드로라제, 엔도글루카나제 및 셀로비아제 로 구성된 군으로부터 선택된 어느 하나 이상이 포함될 수 있다.
- [0017] 본 발명의 다른 일 실시예에 따르면, 상기 셀룰로스 분해조성물은 1 g의 셀룰로스에 대하여 셀룰로스 분해효소 0.01~0.05 FPU 및 원핵생물 유래 익스팬신 200~400 μ g을 포함할 수 있다.
- [0018] 또한, 상술한 두 번째 기술적 과제를 달성하기 위하여, 상기 셀룰로스 분해조성물을 후 40~60 $^{\circ}$ C에서 셀룰로스에 첨가하는 단계를 포함하는 셀룰로스 분해방법을 제공한다.
- [0019] 본 발명의 일 실시예에 따르면, pH 7 이하에서 상기 셀룰로스 분해조성물을 셀룰로스와 반응시킬 수 있다.
- [0020] 본 발명의 또 다른 일 실시예에 따르면, 하렐라 제주엔시스 유래의 익스팬신(EXLX1) 아미노산 서열과 60% 이상의 상동성을 가지며 셀룰로스 확장활성을 가지는 원핵생물 유래 익스팬신 단백질을 제공한다.
- [0021] 본 발명의 세 번째 기술적 과제를 달성하기 위하여, 목질계 바이오매스를 상기 셀룰로스 분해조성물로 처리하여 상기 목질계 바이오매스에 포함된 셀룰로스를 분해하여 환원당을 생산하는 단계를 포함하는 바이오에너지 생산 방법을 제공한다.
- [0022] 본 발명의 다른 일 실시예에 따르면, 종이 또는 펄프를 상기 셀룰로스 분해조성물로 처리하여 상기 종이 또는 펄프를 연화시키는 단계를 포함하는 종이 또는 펄프의 생산방법을 제공한다.
- [0023] 본 발명의 다른 일 실시예에 따르면, 상기 셀룰로스 분해조성물로 처리하여 상기 섬유 또는 직물을 연화시키는 단계를 포함하는 섬유 또는 직물의 처리방법을 제공한다.
- [0024] 본 발명의 네 번째 기술적 과제를 달성하기 위하여 식물 유래 익스팬신과 구조적 유사성을 가지는 원핵생물 유래 단백질을 찾는 단계; 상기 원핵생물 유래 단백질을 클로닝하는 단계; 및 상기 클로닝된 원핵생물 유래 단백질을 균주에서 발현하는 단계를 포함하는 원핵생물 유래 익스팬신 단백질의 대량생산방법을 제공한다.

발명의 효과

[0025] 본 발명은 셀룰로스 확장능력을 가지는 원핵생물 유래 익스팬신 단백질을 제공한다. 상기 원핵생물 유래 익스팬신 단백질은 종래의 식물 유래 익스팬신 단백질과 동일한 기능을 수행한다. 상기 원핵생물 유래 익스팬신 단백질은 식물 유래 익스팬신 단백질과 달리, 산업적 규모로 매우 낮은 비용으로 생산할 수 있다. 특히, 상기 원핵생물 유래 익스팬신 단백질과 셀룰라제를 이용하여 목질계 바이오매스를 포도당으로 가수분해하는 경우, 셀룰라제에 의한 셀룰로스의 가수분해 속도가 현저하게 증가할 수 있으며, 결과적으로 당 수율이 개선된다. 이를 통해 실제 바이오에너지 생산에 있어서 원핵생물 유래 익스팬신으로 효소의 양을 줄임으로써 바이오에너지의 생산비용을 저감할 수 있다. 또한, 상기 원핵생물 유래 익스팬신 단백질은 펄프, 면 섬유(예, 청바지) 등의 조직을 연화시키고 변화시키는 특성이 있어 바이오펄핑이나 바이오스토닝과 같은 목적으로 이용될 수 있다.

도면의 간단한 설명

[0026] 도 1은 BsEXLX1(PDB id : 2bh0)와 제아 메이스(*Zea mays*)로부터 얻은 식물 유래 익스팬신 EXPB(*Zea m 1*, PDB id : 2hcz)의 구조를 비교한 것이다. 위 : 골격구조를 중첩하여 표시한 입체도. BsEXLX1과 EXPB는 각각 검은색과 회색으로 나타내었다. 아래 : 구조에 기초한 두 단백질의 서열정렬. 두 단백질의 아미노산 잔기가 동일한 부분은 짙은 회색으로 표시하였다. EXPB 구조(2hcz)에 기초한 서열정렬에서, 2차구조는 원기둥(α -나선) 또는 화살표(β -가닥)로 표시하였다.

도 2는 재조합 대장균으로부터 얻은 단백질의 SDS-PAGE 결과이다. M : 여러 질량값의 단백질 마커; 1 : 불용성 단백질; 2 : 가용성 단백질; 3 : 정제된 BsEXLX1(24 kDa).

도 3은 100% 셀룰로스로 이루어진 거름종이를 거름종이 1 g 당 100 μ g의 BsEXLX1 또는 소 혈청 알부민(BSA) 존재 하에 시트르산염 완충액(pH 4.8) 내, 50 °C에서 0.012 FPU의 셀룰라제로 인큐베이션했을 때, BsEXLX1에 의한 셀룰로스 가수분해의 상승효과를 보여준다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 4는 BsEXLX1의 양이 셀룰로스 가수분해의 상승효과에 미치는 영향을 보여준다. 거름종이를 거름종이 1 g 당 0, 100, 200 또는 300 μ g의 BsEXLX1 또는 BSA 존재 하에 시트르산염 완충액(pH 4.8) 내, 50 °C에서 0.06 FPU의 셀룰라제로 인큐베이션하였다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 5는 StEXLX1의 양이 셀룰로스 가수분해의 상승효과에 미치는 영향을 보여준다. 거름종이를 거름종이 1 g 당 0, 100, 200 또는 300 μ g의 StEXLX1 존재 하에 시트르산염 완충액(pH 4.8) 내, 50 °C에서 0.06 FPU의 셀룰라제로 인큐베이션하였다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 6은 XoEXLX1의 양이 셀룰로스 가수분해의 상승효과에 미치는 영향을 보여준다. 거름종이를 거름종이 1 g 당 0, 100, 200, 300 또는 400 μ g의 XoEXLX1 존재 하에 시트르산염 완충액(pH 4.8) 내, 50 °C에서 0.06 FPU의 셀룰라제로 인큐베이션하였다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 7은 HcEXLX1이 셀룰로스 가수분해의 상승효과에 미치는 영향을 보여준다. 거름종이를 거름종이 1 g 당 0 및 360 μ g의 HcEXLX1 존재 하에 시트르산염 완충액(pH 4.8) 내, 50 °C에서 0.06 FPU의 셀룰라제로 인큐베이션하였다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 8은 셀룰라제의 양이 셀룰로스 가수분해의 상승효과에 미치는 영향을 보여준다. 거름종이를 거름종이 1 g 당 0, 100, 200 또는 300 μ g의 BsEXLX1 존재 하에 시트르산염 완충액(pH 4.8) 내, 50 °C에서 0.012, 0.06, 0.12 또는 0.6 FPU의 셀룰라제로 48 시간 동안 인큐베이션하였다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 9는 상등액 및 0, 6, 12 또는 24 시간 동안 셀룰라제 및/또는 BsEXLX1의 혼합물로 인큐베이션한 거름종이로부터 얻은 단백질의 SDS-PAGE 결과이다. 각 인큐베이션 시간에서 밴드 강도를 스캔하였으며, 각 단백질에 대한 결과값은 인큐베이션 시간 0에서의 값으로 표준화하였고, 75 kDa에서의 주된 단백질 밴드 강도는 셀룰라제로 확인되었다. (a) 셀룰라제만으로 처리한 거름종이로부터 얻은 상등액. (b) 셀룰라제와 BsEXLX1으로 처리한 거름종이로부터 얻은 상등액. (c) BsEXLX1만으로 처리한 거름종이로부터 얻은 상등액. (d) BSA만으로 처리한 거름종이로부터 얻은 상등액. (e) 셀룰라제만으로 처리한 거름종이. (f) 셀룰라제와 BsEXLX1으로 처리한 거름종이. (g) BsEXLX1만으로 처리한 거름종이. (h) BSA만으로 처리한 거름종이.

도 10은 반응온도가 셀룰로스 가수분해의 상승효과에 미치는 영향을 보여준다. 거름종이를 거름종이 1 g 당 300 μg 의 BsEXLX1 존재 하에 시트르산염 완충액(pH 4.8) 내, 30, 40, 50, 60 또는 70 $^{\circ}\text{C}$ 에서 0.06 FPU의 셀룰라제로 36 시간 동안 인큐베이션하였다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 11은 반응완충액의 pH가 셀룰로스 가수분해의 상승효과에 미치는 영향을 보여준다. 거름종이를 거름종이 1 g 당 300 μg 의 BsEXLX1 존재 하에 pH 3, 4, 4.8, 6 또는 7의 시트르산염 완충액 내, 50 $^{\circ}\text{C}$ 에서 0.06 FPU의 셀룰라제로 36 시간 동안 인큐베이션하였다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 12는 완충액 단독(50 mM 아세트산나트륨, pH 4.8), BsEXLX1 또는 BSA(각각 600 $\mu\text{g}/\text{mL}$)를 포함하는 완충액, 또는 8 M 요소(urea) 내에서 인큐베이션한 왓만(Whatman) 거름종이 No. 3의 인장강도를 보여준다. 실험은 3회 반복하였으며, 그 결과를 평균 $\pm$ 표준편차로 나타내었다.

도 13은 여러 용액 내에서, 30 $^{\circ}\text{C}$ 에서 1 시간 동안 인큐베이션한 왓만 거름종이 No. 3의 주사전자현미경 사진 [(a)~(d): x 300, (e)~(i): x 900]이다. (a)~(e) : BsEXLX1(600 $\mu\text{g}/\text{mL}$)을 포함하는 완충액(50 mM 아세트산나트륨, pH 4.8)으로 인큐베이션한 거름종이; (b), (f) : 8 M 요소 용액으로 인큐베이션한 거름종이; (c), (h) : BSA(600 $\mu\text{g}/\text{mL}$)를 포함하는 완충액(50 mM 아세트산나트륨, pH 4.8)으로 인큐베이션한 거름종이; (d), (i) : 완충액(50 mM 아세트산나트륨, pH 4.8)만으로 인큐베이션한 거름종이.

도 14a 내지 14n은 다양한 원핵생물의 익스펜신 유사단백질의 서열분석으로부터 얻은 서열정보를 보여준다.

도 15는 바실러스 서브틸리스(*Bacillus subtilis*)에 대한 BLAST 검색으로부터 얻은 상위 20개 상동서열에 대한 다중서열정렬의 결과를 보여준다.

도 16은 다중서열정렬을 이용한 분자다양성 트리를 보여준다.

도 17a 및 17b는 다중서열정렬을 이용한 분자다양성 트리를 보여준다.

도 18은 스티그마텔라 아우렌티아카로부터 유래한 StEXLX1의 아미노산 서열이다.

도 19는 잔토모나스 오라이제로부터 유래한 XoEXLX1의 아미노산 서열이다.

도 20은 하헬라 제주엔시스로부터 유래한 HcEXLX1의 아미노산 서열이다.

발명을 실시하기 위한 구체적인 내용

- [0027] 상술한 바와 같이, 종래의 식물 유래 익스펜신은 셀룰로스를 확장시켜 셀룰로스 분해효소의 분해능을 향상시키는 데 매우 효과적이거나, 숙주 생물에서 발현하기 어려워 대량생산이 불가능하였다. 일부 진핵세포에서 식물 유래 익스펜신과 구조적, 기능적으로 유사한 단백질이 발견되었지만, 이 경우 대량생산은 가능해도 셀룰로스 확장 활성이 현저하게 떨어지는 문제가 있었다.
- [0028] 이에 본 발명에서는 셀룰로스를 확장시켜 주는 식물 유래 익스펜신과 구조적 유사성을 갖는 단백질을 원핵생물에서 찾아 셀룰로스의 기질 분해효율을 높였다. 상기 원핵생물 유래 익스펜신 단백질은 식물 유래 익스펜신과 아미노산 서열, 구조, 그리고 기능에 있어 유사성을 가지며 셀룰로스 분해효소(특히 셀룰라제)의 분해효율을 증진시켜준다.
- [0029] 식물은 단단한 세포벽을 가지므로, 새로운 성분의 합성을 위해서는 세포벽을 이완시켜야 한다. '익스펜신'은 원래 세포벽을 확장시켜 세포벽의 합성을 촉진하는 것으로 알려진 식물 단백질이다. 익스펜신이 세포벽 분해활성을 가지는 것은 아니지만, 익스펜신은 불용성 세포벽 성분(예컨대, 셀룰로스)과 결합하여 이완활성을 제공한다. 익스펜신의 작용 메커니즘은 아직 완전히 알려지지 않았지만, 불용성 셀룰로스에 다당류들 사이의 수소결합을 깨는 것으로 생각된다. 식물 유래 익스펜신의 세포벽 이완활성을 다양한 용도에 적용한 특허는 거의 알려져 있지 않다.
- [0030] 익스펜신은 여러 식물들 사이에서 고도로 보존되어 있으며, 현재까지 서열이 밝혀진 모든 식물 게놈에서 발견된다. 진균류나 선충류와 같은 진핵생물에서는, 서열적, 기능적 상동성이 밝혀졌다. 예컨대, 트리코테르마 레세이(*T. reesei*)로부터, 식물 유래 익스펜신 단백질과 서열 유사성을 가지는, 익스펜신 유사 도메인을 갖는 스윌레닌 단백질이 확인되었다. 그러나, 식물 유래 익스펜신과 유사한 원핵생물 유래 단백질은 아직 발견되지 않았다.
- [0031] 식물 유래 익스펜신(제아 메이스 EXPB, PDB code: 2hcz)의 분자구조가 최근에 밝혀졌다. 또한 PDB에 기탁된 바실러스 서브틸리스의 *yoaJ* 단백질(YoaJ, PDB code: 1hb0) 역시 익스펜신과 높은 구조적 유사성을 보이지만, 그 기능은 아직 확인되지 않았다. 본 발명자들은 이들 구조를 기반으로 하여 서열 데이터베이스를 검색한 결과, 식

물 유래 익스팬신과 서열 유사성을 갖는 몇몇 박테리아 유래 단백질을 확인하였다. 이들은 대부분 기능이 알려지지 않은 단백질이다.

- [0032] 본 발명자들은 이들 박테리아 유래 단백질 중 일부를 클로닝하여 익스팬신의 기능을 가지는지 확인하였다. 이들 박테리아 유래 단백질을 대장균 내에서 과발현시킨 후 정제하여, 본 발명자들은 이들 단백질이 식물 세포벽 성분(예컨대, 불용성 셀룰로스)에 대한 결합 및 이완활성을 가짐을 확인하였다. 본 발명자들은 이러한 활성을 셀룰라제에 의한 셀룰로스를 물질의 분해와 관련하여 시험하였다. 본 발명자들은 식물 유래 익스팬신과 서열적 또는 기능적 유사성을 가지는 이러한 원핵생물 유래 단백질을 바이오매스의 전처리 용도로 이용하고자 하였다.
- [0033] 구조에 기초한 서열분석 - 3D 구조정보를 이용한 서열 모델로부터 상동관계가 밝혀지는 경우가 간혹 있다. 본 발명자들은 익스팬신이 '바윈(Barwin) 유사 엔도글루카제'(model #) 도메인과 'PHL 화분 알레르겐'(model #) 도메인을 포함하는 것을 확인하였다. 또한, 본 발명자들은 가장 광범위한 도메인 데이터베이스 중 하나인 Pfam으로부터 이에 상응하는 도메인들(DPBB-PF03330 및 화분 알레르겐-PF01357)을 찾아내었다. 따라서, 본 발명자들은 하나의 전사해독 프레임(ORF) 내에 동시에 두 개의 서열 모델을 갖는 단백질을 익스팬신 유사체로 간주하였다.
- [0034] 또한, 본 발명자들은 BLAST의 바실러스 서브틸리스의 yoaJ 서열을 이용하여 공개된 데이터베이스들(예컨대, Genbank)을 검색하였다. P 값이 0.001보다 작은($P < 0.001$) 경우를 유사체 선정의 통계적 기준으로 하였다.
- [0035] 본 발명자들은 익스팬신과 기능적, 서열적 또는 구조적으로 유사한, 아직까지 보고된 바 없는 박테리아 유래 단백질을 확인하였다. 먼저, Protein Data Bank(PDB)에 기탁된 Zea m 1의 식물 유래 β -익스팬신(EXPB, PDB id : 2hcz)과 구조적 유사성을 가지는 박테리아 유래 익스팬신을 탐색하였다. 본 발명자들은 Charlier 등이 2006년에 구조를 밝힌(Petrella et al. 2006) 바실러스 서브틸리스의 YoaJ 단백질을 찾아내었으며, 이를 익스팬신 명명법(Kende et al. 2004)에 따라 BsEXLX1으로 명명하고, BsEXLX1의 구조를 공개하였으나, 아직 출판되지는 않았다(PDB id : 2bh0).
- [0036] 박테리아 유래 익스팬신 유사체들의 기능을 최초로 확인하기 위하여, 본 발명자들은 대장균 내에서 BsEXLX1을 포함하는 박테리아 유래 단백질을 클로닝하고 발현시켰다. 정제된 단백질의 익스팬신 유사 기능을 생체 외에서 시험하였으며 셀룰로스 가수분해에서의 역할을 조사하였다.
- [0037] 본 발명자들은 BsEXLX1과의 상동성을 기초로 하여, 스티그마텔라 아우렌티아카, 잔토모나스 오라이제 및 하헬라 제주엔시스와 같은 다른 박테리아로부터 익스팬신의 기능적 유사체를 찾아내었다. 이들 익스팬신 유사체를 각각 StEXLX1, XoEXLX1 및 HcEXLX1으로 명명하였으며, 셀룰로스 가수분해 과정에서 셀룰라제와의 상승효과를 조사하였다.
- [0038] 이어, 상기 BsEXPX1(바실러스 익스팬신 유사체)를 사용하여, UniProt DB(단백질 데이터베이스)를 검색하여 66개의 단백질 서열을 획득하였다(컷오프 기준 E 값 < 0.001).
- [0039] 이들 단백질 서열 중에서 중복되는 서열을 제거하기 위해, cd-hit이라는 프로그램을 사용하여 90%이상의 서열 동일성을 갖는 단백질을 제거하여(대표서열만 남겨두고) 총 55개의 서열을 얻었다.
- [0040] 얻어진 55개의 서열 중 같은 종에서 유래된 서열을 제거하여 총 47개의 서열을 얻었다.
- [0041] 본 발명에서 사용된 컷오프에 대한 기준은 확률에 기반하는 것이므로 절대적으로 정해진 것은 없고, 컷오프 값이 0.001(즉, 데이터 검색을 통해 얻어진 1,000개의 서열중 진짜 유사체가 아닌 경우가 하나인 경우)이면 일반적으로 유의성이 있는 것으로 간주된다. 이에 대한 자세한 내용은 <http://www.jcsg.org/psat/help/document.html>을 참고한다.) 상기 사이트의 상단을 보면 '상동 단백질 서열'를 얻기 위한 컷오프 E 값은 0.001로 상정하였다고 기재되어 있다.
- [0042] 상술한 과정을 거쳐 서열번호 1~47의 아미노산 서열을 포함하는 원핵세포 유래 익스팬신 단백질을 발견할 수 있었으며, 후술하는 실험을 통해 서열번호 1~47의 아미노산 서열을 포함하는 상기 원핵세포 유래 익스팬신 단백질이 식물 유래 익스팬신 단백질과 유사한 수준의 셀룰로스 확장활성이 있음을 확인하였다. 하헬라 제주엔시스 유래의 익스팬신(EXLX1)의 아미노산 서열(서열번호 30의 아미노산 서열)과 대략 60% 이상의 상동성을 가지며 셀룰로스 확장활성을 가지는 원핵생물 유래 단백질은 본 발명에 적용될 수 있으며, 서열번호 1 내지 서열번호 47에 기재된 군으로부터 선택된 하나의 아미노산 서열 또는 본 발명의 단백질은 서열번호 1 내지 47의 단백질뿐 아니라 이들 단백질에 하나 이상의 치환, 결손, 역위, 전위 등과 같은 돌연변이를 통하여 본 발명의 목적인 셀룰로스 분해 촉진활성을 가지는 모든 변이 단백질을 포함할 수 있다.
- [0043] 상기 원핵생물은 바실러스 서브틸리스(SEQ. ID NO : 35), 하헬라 제주엔시스(KCTC 2396, SEQ. ID NO : 30),

딕티오스텔리움 디스코이테움(SEQ. ID NO : 1, 2, 3), 네오사르토리야 피체리(SEQ. ID NO : 4), 아스퍼질러스 푸미가투스(SEQ. ID NO : 5), 아스퍼질러스 클라바투스(SEQ. ID NO : 6), 아스퍼질러스 오라이제(SEQ. ID NO : 7), 아스퍼질러스 테레우스(SEQ. ID NO : 8), 페니실리움 크리스제눔(SEQ. ID NO : 9), 아스퍼질러스 나이거(SEQ. ID NO : 10), 에케리첼라 니들란스(SEQ. ID NO : 11), 마그나포르테 그리시아(SEQ. ID NO : 12), 피레노포라 트리티시-레펜티스(SEQ. ID NO : 13), 패오스파에리아 노도룸(SEQ. ID NO : 14), 스크레로티니아 스크레로티오룸(SEQ. ID NO : 15), 프란키아(SEQ. ID NO : 16), 스트렙토마이시스 소비세우스(SEQ. ID NO : 17), 소란지움 셀룰로쭘(SEQ. ID NO : 18, 19), 스티그마텔라 아우렌티아카(스티그마텔라 아우렌티아카, SEQ. ID NO : 20, 22), 플레시오시스티스 파시피카(SEQ. ID NO : 21), 믹스코쿠스 잔터스(SEQ. ID NO : 23), 램토쓰릭스 콜로드니(SEQ. ID NO : 24), 로제이플렉서스(SEQ. ID NO : 25), 로제이플렉서스 카스텐홀지(SEQ. ID NO : 26), 클로로플렉서스 아우란티아쿠스(SEQ. ID NO : 27), 헤르페토시폰 아우란티아쿠스(SEQ. ID NO : 28, 29), 에시도보락스 아베네(SEQ. ID NO : 31), 펙토박테리움 아르토펙티쿰(SEQ. ID NO : 42), 바실러스 리체니포미스(SEQ. ID NO : 43), 잔토모나스 캠펙스트리스(SEQ. ID NO : 37), 바실러스 푸밀루스(SEQ. ID NO : 36), 잔토모나스 오라이제(SEQ. ID NO : 38), 랄스토니아 솔라나시아룸(SEQ. ID NO : 39, 40), 클라비박터 미시가넨시스(SEQ. ID NO : 32, 33, 34), 잘렐라 파스티디오사(SEQ. ID NO : 41), 나카무렐라 멀티파리타(SEQ. ID NO : 44), 마이크로모노스포라(SEQ. ID NO : 45), 카테놀리스포라 애시디필라(SEQ. ID NO : 46) 및 디키야 지(SEQ. ID NO : 47)로 구성된 군으로부터 선택된 원핵생물일 수 있으나 이에 한정되지 않으며 식물 유래 익스팬신과 구조적 유사성을 가지면서 셀룰로스를 확장시킬 수 있는 것이면 제한없이 적용될 수 있다.

[0044] 본 발명의 원핵생물 유래 익스팬신 단백질은 분리 또는 정제될 수 있다. 그러한 분리 또는 정제는 전체 세포, 세포 찌꺼기, 외래 단백질 또는 최종 조성물에서 원하지 않는 단백질을 제거하기 위하여 당업계에 인지도된 분리 기술 예를 들어 이온 교환크로마토그래피, 친화성 크로마토그래피, 소수성 분리, 투석, 단백질 분해효소 처리, 황산 암모늄 침전, 크기 비제 크로마토그래피, 여과, 겔 전기영동 또는 구배상의 분리에 의하여 성취될 수 있다.

[0045] 상술한 두 번째 과제를 달성하기 위하여, 상술한 원핵생물 유래 익스팬신 단백질을 포함하는 셀룰로스 분해조성물을 제공할 수 있다. 셀룰로스를 분해하기 위해서는 셀룰로스 분해효소로 처리하여야 한다. 이들 셀룰로스 분해효소로는 셀룰라제, 셀로비오하이드로라제, 엔도글루카나제 및 셀로비아제 등이 있는데, 이들 분해효소는 그 가격이 너무 비싸 이를 통해 제조된 제품의 가격이 지나치게 높아지는 문제가 발생하였다. 특히 목질계 바이오매스를 구성하는 다당류인 셀룰로스를 분해하여 단당류인 포도당으로 가수분해하기 위해서는 셀룰로스 분해효소(특히 셀룰라제)를 사용하여야 하는데, 상기 셀룰라제의 가격이 너무 비싸 이를 통해 제조되는 최종산물인 바이오에탄올의 가격이 지나치게 상승하는 문제가 발생하였다. 식물 유래 익스팬신 단백질은 셀룰로스를 확장시켜 셀룰라제의 접근성을 용이하게 하여 결국 셀룰라제의 사용량을 저감시킬 수 있으나, 상술한 바와 같이 대량생산이 어려워 상업화되지 못하였다. 이에 본 발명의 원핵생물 유래 익스팬신 단백질은 식물 유래 익스팬신 단백질과 거의 유사한 기능을 가지면서도 숙주 생물 내에서 발현되므로 대량생산이 가능하다.

[0046] 이를 위하여 상기 셀룰로스 분해조성물은 본 발명의 원핵생물 유래 익스팬신 단백질과 셀룰로스 분해효소를 포함할 수 있다. 이 경우 바람직하게는 셀룰로스 분해조성물은 1 g의 셀룰로스에 대하여 셀룰로스 분해효소 0.01~0.05 FPU 및 원핵생물 유래 익스팬신 200~400 μ g을 포함할 수 있다(표 1 참조).

[0047] 또한 바람직하게는, 상기 셀룰로스 분해조성물을 첨가한 후 40~60 $^{\circ}$ C, pH 7 이하에서 48 시간 이상 반응시킬 수 있다.

[0048] 본 발명의 셀룰로스 분해조성물은 본 발명의 목적하는 활성을 향상시키기 위하여 예를 들어 활성화제, 저해제, 바람직한 이온, pH 조절용 화합물 또는 다른 효소 등을 포함할 수 있다.

[0049] 본 발명은 또한 바이오에너지를 생산하는 방법을 제공한다. 구체적으로, 상기 방법은 목질계 바이오매스를 상기 셀룰로스 분해조성물로 처리하여 상기 목질계 바이오매스에 포함된 셀룰로스를 분해하여 환원당을 생산하는 단계를 포함한다.

[0050] 본 발명의 셀룰로스 분해조성물은 종이 또는 펄프를 생산하는 공정에 사용될 수 있다. 구체적으로 종이 또는 펄프를 상기 셀룰로스 분해조성물로 처리하여 상기 종이 또는 펄프를 연화시키는 단계를 포함하는, 종이 또는 펄프의 생산방법을 제공한다.

[0051] 본 발명의 셀룰로스 분해조성물은 섬유 또는 직물의 처리방법에 사용될 수 있다. 구체적으로 상기 셀룰로스 분해조성물로 처리하여 상기 섬유 또는 직물을 연화시키는 단계를 포함하는, 섬유 또는 직물의 처리방법을 제공한다.

다.

[0052] 본 발명의 셀룰로스 분해조성물에 사용되는 원핵생물 유래 익스팬신 단백질은 셀룰로스의 세포벽을 확장시켜주는 익스팬신과 구조적 유사성을 갖는다. 이러한 원핵생물 유래 익스팬신 단백질을 이용하면 셀룰로스 분해효소의 셀룰로스 기질에 대한 분해효율을 높일 수 있다. 상기 원핵생물 유래 단백질은 식물 유래 익스팬신과 아미노산 서열, 구조, 그리고 기능에 있어 유사성을 가지며 셀룰로스 분해효소의 셀룰로스 기질에 대한 분해효율을 증진시켜 줄 뿐만 아니라 산업적 규모로 생산이 가능하다.

[0053] 이하에서, 하기의 실시예를 참고로 하여 본 발명을 더욱 상세히 설명하도록 하겠다. 그러나, 하기 실시예는 설명의 목적을 위한 것으로서 본 발명의 범위를 제한하지 않는다.

[0054] <실시예 1>

[0055] 생물정보 소프트웨어

[0056] DaliLite 프로그램(Holm and Park 2000)을 사용하여 PDB에 대한 구조적 유사성 검색을 실시하였다. 박테리아 유래 익스팬신 유사체를 찾기 위하여, 본 발명자들은 모든 육상식물에서 발견되는 of β -익스팬신(PDB id : 2hcz)(Yennawar et al. 2006)의 일종인, 구조가 알려진 Zea m 1(SEQ. ID NO : 57)을 템플릿으로 사용하였다. DaliLite와 UCSF Chimera(Pettersen et al. 2004)를 사용하여 구조쌍 정렬을 수행하였으며, 역시 UCSF Chimera를 사용하여 단백질의 리본구조도를 얻었다.

[0057] <실시예 2-1>

[0058] 균주, 벡터 및 클로닝

[0059] 플라스미드 클로닝과 표적 유전자의 증폭에는 대장균 균주 DH5 α 를 사용하였고, 클로닝된 유전자는 대장균 균주 BL21(DE3)을 사용하여 발현시켰다. 균주들을 루리아-베르타니(LB) 배지(Difco, Sparks, MD, USA)를 사용하여 37 °C에서 배양한 후, 50 μ g/mL 암피실린(Sigma, St. Louis, MO, USA)을 가하여 형질전환체를 선별하였다. 바실러스 서브틸리스의 게놈 DNA를 주형으로 사용하여, PCR 증폭을 통해 표적 유전자를 얻었다. 바실러스 서브틸리스의 게놈 DNA는 버클리 구조게놈센터(BSGC, Berkeley, CA, USA)로부터 입수하였다. PCR 반응을 위한 프라이머 서열은 BsEXLX1N(5'-GAAGGAGATATAAGGATGGCATATGACGACCTGCATGAA-3', SEQ. ID NO : 49)과 BsEXLX1C(5'-ATGATGGTAATGGTGTTCAGGAACTGAACATGGCC-3', SEQ. ID NO : 50)으로, 표적 단백질의 아미노 말단에 존재할 수 있는 신호서열이 제거되도록 설계하였다. 상기 PCR 프라이머를 IDT(Coralville, IA, USA)에 의해 합성하였다. PCR 산물을 비접합 클로닝(LIC) 방법(Aslanidis and Dejong 1990)에 의해 발현벡터 내로 직접 클로닝하였다. LIC 클로닝을 위한 발현벡터로는 BSGC(Graslund et al. 2008)에서 개발한 pET21a를 약간 변형하여 사용하였다. 상기 벡터 역시 카르복실 말단에 발현된 단백질의 친화정제를 위한 6-히스티딘 태그를 가진다.

[0060] 표적 단백질의 발현 및 정제

[0061] 재조합 플라스미드를 포함하는 대장균 BL21(DE3)의 형질전환체를 600 nm에서의 흡광도가 1.0이 될 때까지 암피실린을 함유하는 LB 배지에서 배양하였다. 1 mM의 이소프로필 티오갈락토시드(IPTG)를 가한 후, 세포를 3 시간 더 배양하여 재조합 단백질이 과발현되도록 하였다. 5,000 rpm에서 30 분 동안 원심분리하여 세포를 수득한 후, 세포 펠렛을 용해완충액(20 mM 인산나트륨, 0.5 M 염화나트륨, pH 7.4)에 재분산하고 초음파처리하여 분해시켰다. 얻어진 현탁액을 4 °C에서 20 분 동안 16,000 rpm으로 원심분리하였다. 상등액은 단백질의 가용성 분획으로, 펠렛은 용해완충액에 재분산하여 불용성 분획으로 사용하였다. 두 분획을 15%(w/w) SDS-PAGE로 분석한 후 쿠마시 염색하여 발현수준을 확인하였다. 가용성 분획을 용해완충액으로 평형을 맞춘 His-트랩 컬럼(GE Healthcare, Piscataway, NJ, USA)에 로딩하였다. 컬럼을 25 mM 이미다졸을 포함하는 10 컬럼 부피의 완충액으로 세척하고, His-태그된 표적 단백질을 200 mM 이미다졸을 포함하는 완충액으로 용리하였다. 용리된 단백질을 Amicon Ultra-15 원심분리장치(Millipore, Billerica, MA, USA)를 이용하여 농축한 뒤 20 mM 인산나트륨 완충액으로 탈염하였다. 단백질을 단백질 정량키트(Pierce, USA)를 사용하여 정량하였다. 이와 비슷한 방법으로, 스티그마텔라 아우렌티아카, 잔토모나스 오라이제 및 하헬라 제주엔시스로부터 박테리아 유래 익스팬신 유사체를 대장균 BL21(DE3) 내에서 클로닝하고 발현시켰다.

[0062] 재조합 BsEXLX1의 과발현

[0063] 구조적으로 유사한 단백질이 예컨대 식물 세포벽의 이완, 셀룰로스의 파괴(약화) 또는 셀룰라제 활성의 개선과

같은, 식물 유래 익스팬신과 비슷한 분자적 기능을 가지는지 확인하기 위하여 BsEXLX1을 클로닝 및 과발현하였다. 대장균 내에서의 이중 발현이 촉진되도록 BsEXLX1의 아미노 말단에 존재할 수 있는 신호서열을 제거하였다. 재조합 단백질이 대장균 내에서 발현되었으며, 주로 가용성 분획 내에 존재하였다(도 2). 단백질을 단백질의 카르복실 말단에 위치한 6-His 태그로 정제하였다. 단백질의 수율은 배양배지 1 L 당 약 10 mg 정도였다.

[0064] 또한, 스티그마텔라 아우렌티아카, 잔토모나스 오라이제 및 하헬라 제주엔시스를 EXPB로 사용하였다.

[0065] <실시예 2-2>

[0066] 균주, 벡터 및 클로닝

[0067] 플라스미드 클로닝과 표적 유전자의 증폭에는 대장균 균주 DH5 α를 사용하였고, 클로닝된 유전자는 대장균 균주 BL21(DE3)을 사용하여 발현시켰다. 균주들을 LB 배지(Difco, Sparks, MD, USA)를 사용하여 37 °C에서 배양한 후, 50 µg/mL 암피실린(Sigma, St. Louis, MO, USA)을 가하여 형질전환체를 선별하였다. 스티그마텔라 아우렌티아카의 게놈 DNA를 주형으로 사용하여, PCR 증폭을 통해 표적 유전자를 얻었다. 스티그마텔라 아우렌티아카의 게놈 DNA는 BSGC(Berkeley, CA, USA)로부터 입수하였다. PCR 반응을 위한 프라이머 서열은 StEXLX1N과 StEXLX1C(5'-CTTTTTCAGGCTAACTGCTCAGCACCATCG-3', SEQ. ID NO : 51)로, 표적 단백질의 아미노 말단에 존재할 수 있는 발현 억제서열이 제거되도록 설계하였다. 상기 PCR 프라이머를 IDT(Coralville, IA, USA)에 의해 합성하였다. PCR 산물을 LIC 방법(Aslanidis and Dejong 1990)에 의해 발현벡터 내로 직접 클로닝하였다. LIC 클로닝을 위한 발현벡터로는 BSGC(Graslund et al. 2008)에서 개발한 pET21a를 약간 변형하여 사용하였다. 상기 벡터 역시 카르복실 말단에 발현된 단백질의 친화정제를 위한 6-히스티딘 태그를 가진다. StEXLX1이 발현되었으나 가용성은 아니었으므로, 가용성 형태를 얻기 위하여 N-말단 프라이머 서열을 다시 설계하였다. PCR 반응을 위한 프라이머 서열은 StEXLX2N (5'-ATGTCTGAACCTGACGACTCCGGCTTTACTAACGATCCTGAAGGCGAGGTTCGTGCTCTGGGTGAATTC-3', SEQ. ID NO : 52)과 StEXLX1C(위의 서열과 동일)이었다.

[0068] 표적 단백질의 발현 및 정제

[0069] 재조합 플라스미드를 포함하는 대장균 BL21(DE3)의 형질전환체를 600 nm에서의 흡광도가 1.0이 될 때까지 암피실린을 함유하는 LB 배지에서 배양하였다. 가용성 단백질이 발현되도록 하기 위하여, 최종농도 1 mM의 IPTG를 가한 후 배양온도를 16 °C로 낮추고 세포를 16 시간 더 배양하여 재조합 단백질이 과발현되도록 하였다. 5,000 rpm에서 30 분 동안 원심분리하여 세포를 수득한 후, 세포 펠렛을 용해완충액(20 mM 인산나트륨, 300 mM 염화나트륨, pH 7.4)에 재분산하고 초음파처리하여 분해시켰다. 얻어진 현탁액을 4 °C에서 30 분 동안 16,000 rpm으로 원심분리하였다. 상등액은 단백질의 가용성 분획으로, 펠렛은 용해완충액에 재분산하여 불용성 분획으로 사용하였다. 두 분획을 15%(w/w) SDS-PAGE로 분석한 후 쿠마시 염색하여 발현수준을 확인하였다. 가용성 분획을 용해완충액으로 평형을 맞춘 His-트랩 컬럼(GE Healthcare, Piscataway, NJ, USA)에 로딩하였다. 컬럼을 10 mM 이미다졸과 20 mM 인산나트륨을 포함하는 10 컬럼 부피의 완충액으로 세척하고, His-태그된 표적 단백질을 200 mM 이미다졸을 포함하는 완충액으로 용리하였다. 용리된 단백질을 Amicon Ultra-15 원심분리장치(Millipore, Billerica, MA, USA)를 이용하여 농축한 뒤 20 mM 인산나트륨 완충액으로 탈염하였다. 단백질을 단백질 정량키트(Pierce, USA)를 사용하여 정량하였다. 단백질의 수율은 배양배지 1 L 당 약 3 mg 정도였다.

[0070] <실시예 2-3>

[0071] 균주, 벡터 및 클로닝

[0072] 플라스미드 클로닝과 표적 유전자의 증폭에는 대장균 균주 DH5 α를 사용하였고, 클로닝된 유전자는 대장균 균주 BL21(DE3)을 사용하여 발현시켰다. 균주들을 LB 배지(Difco, Sparks, MD, USA)를 사용하여 37 °C에서 배양한 후, 50 µg/mL 암피실린(Sigma, St. Louis, MO, USA)을 가하여 형질전환체를 선별하였다. 잔토모나스 캄페스트리스의 게놈 DNA를 주형으로 사용하여, PCR 증폭을 통해 표적 유전자를 얻었다. PCR 반응을 위한 프라이머 서열은 XoEXLX1N(5'-ATGCAGGTCAGTACGCAAGC-3', SEQ. ID NO : 53)과 XoEXLX1C(5'-GGGAACTGTACGTGGCCG-3', SEQ. ID NO : 54)으로, 표적 단백질의 아미노 말단에 존재할 수 있는 신호서열이 제거되도록 설계하였다. 상기 PCR 프라이머를 IDT(Coralville, IA, USA)에 의해 합성하였다. PCR 산물을 LIC 방법(Aslanidis and Dejong 1990)에 의해 발현벡터 내로 직접 클로닝하였다. LIC 클로닝을 위한 발현벡터로는 BSGC(Graslund et al. 2008)에서 개발한 pET21a를 약간 변형하여 사용하였다. 상기 벡터 역시 카르복실 말단에 발현된 단백질의 친화정제를 위한 6-히스티딘 태그를 가진다.

[0073] 표적 단백질의 발현 및 정제

[0074] 재조합 플라스미드를 포함하는 대장균 BL21(DE3)의 형질전환체를 600 nm에서의 흡광도가 1.0이 될 때까지 암피실린을 함유하는 LB 배지에서 배양하였다. 가용성 단백질이 발현되도록 하기 위하여, 최종농도 1 mM의 IPTG를 가한 후 배양온도를 16 °C로 낮추고 세포를 12 시간(다른 경우에 비해 짧은 시간) 더 배양하여, 단백질 분해효소에 의한 절단 없이 재조합 단백질이 과발현되도록 하였다. 5,000 rpm에서 30 분 동안 원심분리하여 세포를 수득한 후, 세포 펠렛을 용해완충액(20 mM 인산나트륨, 300 mM 염화나트륨, pH 7.4)에 재분산하고, 1 mM IPTG 및 단백질 분해효소 억제제 카테일(Roche, USA)을 사용, 초음파처리하여 분해시켰다. 얻어진 현탁액을 4 °C에서 30 분 동안 16,000 rpm으로 원심분리하였다. 상등액은 단백질의 가용성 분획으로, 펠렛은 용해완충액에 재분산하여 불용성 분획으로 사용하였다. 두 분획을 15%(w/w) SDS-PAGE로 분석한 후 쿠마시 염색하여 발현수준을 확인하였다. 가용성 분획을 용해완충액으로 평형을 맞춘 His-트랩 컬럼(GE Healthcare, Piscataway, NJ, USA)에 로딩하였다. 컬럼을 10 mM 이미다졸과 20mM 인산나트륨을 포함하는 10 컬럼 부피의 완충액으로 세척하고, His-태그된 표적 단백질을 200 mM 이미다졸을 포함하는 완충액으로 용리하였다. 용리된 단백질을 Amicon Ultra-15 원심분리장치(Millipore, Billerica, MA, USA)를 이용하여 농축한 뒤 20 mM 인산나트륨 완충액으로 탈염하였다. 단백질을 단백질 정량키트(Pierce, USA)를 사용하여 정량하였다. 단백질의 수율은 배양배지 1 L 당 약 5 mg 정도였다.

[0075] <실시예 2-4>

[0076] 균주, 벡터 및 클로닝

[0077] 플라스미드 클로닝과 표적 유전자의 증폭에는 대장균 균주 DH5 α를 사용하였고, 클로닝된 유전자는 대장균 균주 BL21(DE3)을 사용하여 발현시켰다. 균주들을 LB 배지(Difco, Sparks, MD, USA)를 사용하여 37 °C에서 배양한 후, 50 µg/mL 암피실린(Sigma, St. Louis, MO, USA)을 가하여 형질전환체를 선별하였다. 하렐라 제주엔시스의 게놈 DNA를 주형으로 사용하여, PCR 증폭을 통해 표적 유전자를 얻었다. 하렐라 제주엔시스의 게놈 DNA는 한국생명공학연구원(KRIBB)의 J. F. Kim으로부터 입수하였다. PCR 반응을 위한 프라이머 서열은 *HcEXLX1N* (5'-GAAAATCGAGTTTCTGCGACTC-3', SEQ. ID NO : 55)과 *HcEXLX1C* (5'-TTTGTCTGCCTGATTAATAACGCC-3', SEQ. ID NO : 56)으로, 표적 단백질의 아미노 말단에 존재할 수 있는 신호서열이 제거되도록 설계하였다. 상기 PCR 프라이머를 IDT(Coralville, IA, USA)에 의해 합성하였다. PCR 산물을 LIC 방법(Aslanidis and Dejong 1990)에 의해 발현벡터 내로 직접 클로닝하였다. LIC 클로닝을 위한 발현벡터로는 BSGC(Graslund et al. 2008)에서 개발한 pET21a를 약간 변형하여 사용하였다. 상기 벡터 역시 카르복실 말단에 발현된 단백질의 친화정제를 위한 6-히스티딘 태그를 가진다.

[0078] 표적 단백질의 발현 및 정제

[0079] 재조합 플라스미드를 포함하는 대장균 BL21(DE3)의 형질전환체를 600 nm에서의 흡광도가 1.0이 될 때까지 암피실린(100 µg/L)을 함유하는 LB 배지에서 배양하였다. 가용성 단백질이 발현되도록 하기 위하여, 최종농도 1 mM의 IPTG를 가한 후 배양온도를 16 °C로 낮추고 세포를 16 시간 더 배양하여 재조합 단백질이 과발현되도록 하였다. 5,000 rpm에서 30 분 동안 원심분리하여 세포를 수득한 후, 세포 펠렛을 용해완충액(20 mM Tris-Cl 및 300 mM 염화나트륨, pH 8.0)에 재분산하고, 초음파처리하여 분해시켰다. 얻어진 현탁액을 4 °C에서 30 분 동안 16,000 rpm으로 원심분리하였다. 상등액은 단백질의 가용성 분획으로, 펠렛은 용해완충액에 재분산하여 불용성 분획으로 사용하였다. 두 분획을 15%(w/w) SDS-PAGE로 분석한 후 쿠마시 염색하여 발현수준을 확인하였다. 가용성 분획을 용해완충액으로 평형을 맞춘 His-트랩 컬럼(GE Healthcare, Piscataway, NJ, USA)에 로딩하였다. 컬럼을 10 mM 이미다졸과 20 mM Tris-Cl을 포함하는 10 컬럼 부피의 완충액으로 세척하고, His-태그된 표적 단백질을 200 mM 이미다졸을 포함하는 완충액으로 용리하였다. 용리된 단백질을 Amicon Ultra-15 원심분리장치(Millipore, Billerica, MA, USA)를 이용하여 농축한 뒤 20 mM 인산나트륨 완충액으로 탈염하였다. 단백질을 단백질 정량키트(Pierce, USA)를 사용하여 정량하였다. 단백질의 수율은 배양배지 1 L 당 약 20 mg 정도였다.

[0080] 도 17a 및 17b는 박테리아 유래 익스팬신 간의 유사성과 계통을 나타낸 뿌리 없는 계통도를 보여준다. 알려진 바와 같이, 식물 유래 익스팬신과 박테리아 유래 익스팬신은 서로 대척점에 위치하는데, 도 16에서 그 구분이 용이한 것은 BsEXLX1은 StEXLX1과 유사하기 때문이고, HcEXLX1의 경우 이들과는 달리 해양미생물에서 유래된 것(제주 마라도)으로 다른 계통에 속한다는 것이 밝혀졌다. 구체적으로 XoEXLX1의 경우 다른 익스팬신들의 크기가 25 kDa인 것과는 달리 64 kDa의 크기를 가진다. 이는 XoEXLX1이 셀룰라제 도메인을 포함하고 있기 때문이다. 또한 ProtParam을 사용하여 아미노산 서열특성을 비교한 도 18-20을 보면, StEXLX1과 HcEXLX1의 등전점(pI)은 약 산성 영역에 있고 XoEXLX1과 BsEXLX1의 등전점은 염기성 영역에 있다. 이로 인해, 정제를 하는 데 필요한 조건이 약간 달라진다. 무엇보다, 효소와의 상승효과를 얻기 위해서는 익스팬신과 효소의 최적 조건이 거의 동일해야 할 것이다. 일반적으로 염기성 익스팬신은 약산성에서도 안정적인 형태를 갖지만, 약산성 익스팬신은 침전이

발생할 수 있는 위험이 있다. 즉, 효소를 사용하는 버퍼의 pH에 따라 적절한 익스펜신의 사용이 보다 효율적이다. 또한 각각의 익스펜신의 기원이 되는 균주의 특성(가혹한 조건에서의 안정성)을 고려하여, 효소와의 상승효과가 겹치는 익스펜신을 사용하는 것이 유리할 것이다.

[0081] 식물 유래 익스펜신은 그 계통에 따라 α 형태, β 형태, $L\alpha$ 형태 및 $L\beta$ 형태로 크게 나뉘며, 이들 형태는 더 세분화된다(Sampedro, et al. 2006). 박테리아 유래 익스펜신 또한 그 계통에 따라 여러 형태로 나뉘며, 이들의 계통도 밝혀지기 시작하고 있다(*BsEXLX1*과 *HcEXLX1*이 상승효과를 나타내는 기질과 효소는 서로 다를 수 있다는 것을 예상할 수 있다. 또한, 산업적으로 이를 적용하고자 하는 바이오매스의 유형에 따라 최적의 익스펜신 역시 달라질 수 있다.).

[0082] <실시예 3>

[0083] **박테리아로부터 유래한, 옥수수 식물 유래 익스펜신과의 구조적 유사체**

[0084] 식물과 진균류에서는 다수의 익스펜신 도메인이 알려져 있지만, 박테리아에서는 익스펜신 유사 도메인이 확인되지 않았다(Kende et al. 2004; Sampedro and Cosgrove 2005). *Zea m 1* EXPB의 서열을 이용한 비중복(nr) 단백질 데이터베이스에 대한 BLAST 검색 결과, 기준(P 값 < 0.001)에 맞는 박테리아 유래 단백질은 검색되지 않았다. 따라서, 본 발명자들은 서열 불일치(예컨대, 원격 상동성)로 인해 일반적인 서열분석으로는 쉽게 검색되지 않는 구조적 유사체가 박테리아 내에 존재할 수 있다고 판단하였다. 이를 입증하기 위하여, 본 발명자들은 최근에 구조가 밝혀진 *Zea m 1* EXPB(PDB id : 2hcz)의 구조적 유사체를 PDB(Yennawar et al. 2006)에서 검색하였다. PDB 검색 결과, 바실러스 서브틸리스 유래 *BsEXLX1* 단백질(PDB id : 2bh0)의 구조가 식물 유래 익스펜신과 유사한 것으로 확인되었다. 본 발명자들은 구조적 데이터베이스 검색프로그램인 DaliLite를 이용하여, *BsEXLX1*가 *Zea m 1*과 높은 구조적 유사성을 가지는 것을 확인하였다. 유사성을 나타내는 Z 값이 20.7로, 일반적인 구조적 유사체 선별기준($Z = 2$)에 비해 통계적으로 유의하였다. 그러나, 두 단백질 사이의 서열 동일성은 21%에 불과하였다. 식물 유래 익스펜신인 *Zea m 1* EXPB(2hcz)과 *BsEXLX1*(2bh0)의 구조에 대한 평균 제곱근편차(r.m.s.d.)는 2.4 Å으로 두 구조는 거의 동일한 접힘 구조를 가지는 것으로 확인되었다. 구조비교 결과와 서열 정렬을 도 1에 제시하였다.

[0085] 도 1은 옥수수의 β 익스펜신(*Zea m1*, PDB id : 2hcz)과 *BsEXLX1*(PDB id : 2bh0)의 구조적 유사성과 아미노산 서열을 비교한 그림이다. 도 1의 아래 부분을 보면 옥수수에서 분리된 베타 익스펜신의 아미노산 서열은 *BsEXLX1*과의 유사성이 20% 미만임에도 불구하고, 위의 입체도를 보면 구조적으로 매우 유사함을 알 수 있다. 따라서 *BsEXLX1*은 옥수수 베타 익스펜신과 다른 서열을 가지지만 구조적 상동성을 가지는 익스펜신 유사체라 할 수 있다.

[0086] 염기서열 정보를 도 14a 내지 17b에 나타내었다. 도 14a~14n은 식물 유래 익스펜신과 같은 작용을 할 것으로 예측되는 도메인을 찾아 비교하기 위하여, 박테리아 유래 익스펜신들 사이에서 유사성을 보이는 특정 도메인을 다중서열정렬(ClustalW)을 이용하여 나타낸 것이다.

[0087] 도 15는 ClustalW를 이용하여 분석한 FASTA 파일을 그래픽화하여 JAVA로 표현한 것이다. 아미노산이 유사한 도메인들을 다른 색으로 표시하여 구조가 유사한 도메인을 파악하기 쉽게 하였다.

[0088] 도 16, 17a 및 17b는 파악된 구조를 비교하기 위한 다른 방법으로서, 뿌리 없는 계통도를 그린 것이다.

[0089] 이러한 방법을 이용하여 위에서 설명한, 각 익스펜신의 유사성과 차이점을 확인할 수 있다.

[0090] <실시예 4>

[0091] **효소에 의한 셀룰로스의 가수분해**

[0092] 공지된 고속 마이크로어세이(Berlin et al. 2006)를 약간 변형하여, 마이크로플레이트 상에서 셀룰로스를 가수분해하였다. 순수한 셀룰로스로 구성된 왓만 No. 1 거름종이(Whatman, Florham Park, NJ, USA)를 6 mm 직경의 디스크 형태(각 2.5 mg)로 잘라 셀룰로스 가수분해의 기질로 사용하였다. 거름종이 디스크를 96 웰 PCR 플레이트(Axygen, Union City, CA, USA)에 올려놓았다. 셀룰라제(Celluclast 1.5L, Novozymes, Bagsvaerd, Denmark)와 0.05 M 시트르산염 완충액(pH 4.8) 내 박테리아 유래 익스펜신 단백질의 최종농도는 셀룰로스 1 g을 기준으로 각각 0~0.6 FPU와 0~300 μ g이었으며, 전체 반응혼합물의 부피는 120 μ L이었다. 96 웰 플레이트를 덮고, 50 °C에서 48 시간 동안 PCR 기기(Peltier Thermal Cycler, Bio-Rad, Hercules, CA, USA) 내에서 인큐베이션하였다. 12 시간 간격으로 3 개의 샘플을 취한 후, 디니트로살리실산(DNS) 시약(Adney and Baker 1996; Xiao et al. 2004)을 사용하여 환원당의 농도를 결정하였다. 환원당과 DNS 시약에 의한 540 nm에서의 흡광도를, 포도

당 보정곡선을 이용하여 환원당 농도로 환산하였다. 얻어진 세 데이터를 평균 $\pm$ 표준편차로 나타내었다.

[0093]

BsEXLX1에 의한 셀룰로스 가수분해의 상승효과

[0094]

식물 유래 익스팬신은 가수분해 활성 없이 식물 세포벽 다당류를 파괴하는 것으로 밝혀졌는데, 이들은 다당류와 결합하여 셀룰로스 미세섬유들 사이의 수소결합을 깨는 것으로 생각된다(Cosgrove 2000a; Cosgrove 2000b; McQueen-Mason and Cosgrove 1994). 셀룰로스에 대한 파괴효과는 흔히 신장계(McQueen-Mason and Cosgrove 1994; Yennawar et al. 2006)를 이용, 셀룰로스의 신장 또는 약화 정도를 측정하여 정량화한다. 아직까지 식물 이외의 생명체에서 활성 익스팬신의 이중 발현은 보고된 바 없다. 아라비도시스 탈리아나(*Arabidopsis thaliana*)의 β -익스팬신과 유사한 선충류 단백질이 식물에서 발현된 바 있으며, 밀 초엽(coleoptile)과 오이 배축(hypocotyl)에 대해 셀룰로스 파괴활성을 가지는 것으로 확인되었다(Qin et al. 2004).

[0095]

BsEXLX1이 Zea m1 EXPB과 구조적으로 유사하다는 것은 이미 확인된 바 있으나, 박테리아 유래 익스팬신으로서의 기능과 잠재적인 용도에 대해서는 아직 연구된 바 없다. 본 발명자들은 BsEXLX1의 박테리아 유래 익스팬신으로서의 분자적 기능을 생체 외에서 시험하였다. 도 3은 셀룰라제 단독, BsEXLX1 단독 또는 셀룰라제와 BsEXLX1에 의한, 시간에 따른 거름종이 가수분해의 결과를 보여준다. 식물 유래 익스팬신은 가수분해 활성을 가지지 않는다는 기존의 보고(Cosgrove 2000a; Cosgrove 2000b; McQueen-Mason and Cosgrove 1994)와 유사하게, 거름종이를 BsEXLX1으로 인큐베이션한 경우에는 환원당이 거의 생성되지 않았다. 거름종이에 셀룰라제만 가한 경우, 시간에 따른 당 생산은 반응혼합물 내 셀룰라제의 양이 적어(예컨대, 0.012 FPU/g 셀룰로스) 낮게 나타났다. 그러나, BsEXLX1(셀룰로스 1 g 당 100 μ g)를 셀룰라제와 함께 효소 반응혼합물에 가하고 36 시간 인큐베이션한 경우, 거름종이로부터의 환원당 수율이 셀룰라제 단독(대조군)의 경우에 비해 2.6 배 증가하였다. 다른 단백질도 BsEXLX1처럼 셀룰로스 가수분해에 영향을 미칠 수 있는 지 확인하기 위하여, 음성 대조군으로서 BSA를 셀룰라제와 함께 거름종이에 가하였다. 그러나, BSA를 셀룰라제와 함께 가한 경우, 당 수율은 BSA를 가하지 않은 대조군에 비해 별다른 변화를 보이지 않았다.

[0096]

<실시예 5>

[0097]

BsEXLX1의 양이 셀룰로스 가수분해의 상승효과에 미치는 영향

[0098]

BsEXLX1의 양과 셀룰로스 가수분해의 상승효과 사이의 관계를 도 4에 도시하였다. 반응혼합물은 거름종이 1 g을 기준으로 여러 농도의 BsEXLX1(0, 100, 200 또는 300 μ g)과 0.06 FPU의 셀룰라제로 구성되었다. 셀룰로스 1 g 당 100 μ g의 BsEXLX1을 셀룰라제와 함께 가한 경우, 36 시간 후의 환원당 생성량은 BsEXLX1이 없는 경우에 비해 1.7 배 많았다. BsEXLX1의 농도가 증가함에 따라 상승효과 역시 증가하였다. 그러나, BsEXLX1의 양이 100 μ g에서 200 μ g으로, 또는 200 μ g에서 300 μ g으로 증가한 경우, 환원당 생산량의 증가는 BsEXLX1의 양이 0에서 100 μ g으로 증가한 경우에 비해 낮았다. 즉, BsEXLX1에 의한 상승효과가 낮아졌는데, BsEXLX1의 첨가량이 증가함에 따라 포화되는 것으로 보였다. 도 5, 6 및 7에서 보듯이, 다른 박테리아로부터 유래한 EXPB 계열 익스팬신의 상승효과 역시 EXLX1의 경우와 비슷하였다.

[0099]

도 5는 스티그마텔라 아우렌티아카에서 유래된 익스팬신인 StEXLX1과 거름종이 1 g 당 0.06 FPU의 Novozymes Celluclast 1.5L 셀룰라제 복합제 혼합물을 셀룰로스 기질인 거름종이에 가하여 가수분해한 후, 생성되는 환원당을 측정한 결과이다. 도 5에서 보는 바와 같이, StEXLX1의 양이 증가함에 따라(거름종이 1 g 당 100, 200, 300 μ g) 환원당의 생성량이 StEXLX1이 포함되지 않은 대조군에 비해 증가하는 것을 알 수 있다. 이러한 결과는 StEXLX1의 셀룰라제와의 상승효과를 확인시켜 준다.

[0100]

도 6은 잔토모나스 캠페스트리스에서 유래된 익스팬신인 XoEXLX1과 거름종이 1 g 당 0.06 FPU의 Novozymes Celluclast 1.5L 혼합물을 셀룰로스 기질인 거름종이에 가하여 가수분해한 후, 생성되는 환원당을 측정한 결과이다. 도 6에서 보는 바와 같이, XoEXLX1의 양이 증가함에 따라(거름종이 1 g 당 100, 200, 300 μ g) 환원당의 생성량이 XoEXLX1이 포함되지 않은 대조군에 비해 증가하는 것을 알 수 있다. 이러한 결과는 XoEXLX1의 셀룰라제와의 상승효과를 확인시켜 준다.

[0101]

도 7은 하헬라 제주엔시스(KCTC 2396)에서 유래된 익스팬신인 HcEXLX1과 거름종이 1 g 당 0.06 FPU의 Novozymes Celluclast 1.5L 혼합물을 셀룰로스 기질인 거름종이에 가하여 가수분해한 후, 생성되는 환원당을 측정한 결과이다. 도 7에서 보는 바와 같이, 거름종이 1 g 당 360 μ g의 HcEXLX1을 가하였을 때의 환원당 생성량이 HcEXLX1이 포함되지 않은 대조군에 비해 많았다. 이러한 결과는 XoEXLX1의 셀룰라제와의 상승효과를 확인시켜 준다.

[0102] <실시예 6>

[0103] 셀룰라제의 양이 셀룰로스 가수분해의 상승효과에 미치는 영향

[0104] 셀룰라제에 의한 셀룰로스 가수분해는 전형적인 효소 반응속도를 나타낸다. 즉, 기질의 양이 고정된 경우 셀룰라제의 농도가 증가함에 따라 반응속도 역시 증가한다(Baker et al. 2000; Berlin et al. 2006). 도 8에서 보듯이, 셀룰로스의 가수분해에 의한 환원당의 생성량은 효소의 양이 거름종이 1 g 당 0.012 FPU에서 0.6 FPU로 증가함에 따라 크게 증가하였다. 이처럼 셀룰라제의 양이 0.012 또는 0.06 FPU로 작은 경우에는, 거름종이와 셀룰라제를 포함하는 반응혼합물 내에 BsEXLX1이 존재하는지 여부에 따라 환원당의 생산이 큰 차이를 보였다. 셀룰라제의 양을 거름종이 1 g 당 0.12 또는 0.6 FPU로 증가시키자, 상승효과가 낮게 나타났다. 이처럼 효소의 양이 많은 경우에는, 거름종이 1 g 당 300 μ g의 BsEXLX1과 함께 인큐베이션한 경우 BsEXLX1이 없는 대조군에 비해 약간 낮은 환원당 수율이 얻어졌다.

[0105] 실험에서 사용한 셀룰라제의 양은 셀룰로스 1 g 당 0.012~0.6 FPU로, 전처리된 리그노셀룰로스의 효소에 의한 분해정도를 시험하는 데 일반적으로 사용되는 양(일반적으로 셀룰로스 1 g 당 5 FPU 이상)(Rudolf et al. 2008; Yang et al. 2006)보다 낮았다. 거름종이 1 g 당 셀룰라제의 양이 0.012 및 0.06 FPU인 경우, 2.5 mg의 거름종이로부터 최대 150 μ g의 환원당이 생산되었으며, 이때 당 수율은 이론적 최대값의 5.5%에 해당하였다. 셀룰라제의 양이 0.12~0.6 FPU인 경우, 2.5 mg의 거름종이로부터 이론적 최대값의 21.8%에 해당하는 최대 600 μ g의 환원당이 생산되었으나, 상승효과는 크지 않았다. 셀룰라제의 양이 작을 때 상승효과가 나타나기는 하지만, 이 경우 전환율이 매우 낮으므로, 현 단계에서 BsEXLX1에 의한 상승효과를 셀룰로스 가수분해에 산업적으로 적용하기에는 무리가 있다.

[0106] 이처럼 셀룰라제의 양이 작을 때 당 수율이 낮은 경우는 탄수화물 결합모듈(CBM) 또는 셀룰로스 결합도메인(CBD)을 가지는 셀룰라제의 상승효과 연구에서 흔히 발견된다. 예컨대, 써모비피다 푸스카(*Thermobifida fusca*)의 CBM인 E7과 E8을 이 미생물의 엑소글루카나제인 Cel6B에 가하여 거름종이를 50 $^{\circ}$ C에서 168 시간 동안 가수분해한 후의 셀로비오스 수율은 각각 이론적 최대값의 5.0%와 5.1%에 불과하였다(Moser et al. 2008). 트리코테르마 레세이의 CBH I 및 II에 대한 상승효과 연구에서, 48 시간 반응 후 아비셀의 전환수율은 5~15% 범위였다(Medve et al. 1994). 바실러스 종의 Cel5A에 포함된 CBM의 역할에 관한 연구에서, 37 $^{\circ}$ C에서 48 시간 및 96 시간 동안 가수분해한 후 재생된 셀룰로스의 환원당 수율은 최대 7.4% 및 16.4%이었다(Boraston et al. 2003). 클로스트리디움 써모셀럼(*Clostridium thermocellum*)의 CBD와 CelD의 하이브리드를 사용하여 아비셀과 BMCC를 45 $^{\circ}$ C에서 48 시간 동안 가수분해하였을 때, 환원당의 최대 생산량은 각각 이론적 최대값의 8.0%와 0.2%이었다(Carrard et al. 2000).

[0107] 표 1은 셀룰라제와 셀룰로스의 다양한 조합 및 다양한 반응시간에 따른 BsEXLX1의 상승효과를 요약한 것이다.

표 1

[0108] 고정된 거름종이 양에 대하여, BsEXLX1과 셀룰라제의 조성비에 따른 BsEXLX1에 의한 셀룰로스 가수분해의 상승효과(%)

BsEXLX1 (μ g/g 셀룰로스)	셀룰라제 (FPU/g 셀룰로스)	^a 상승효과 (%)		
		반응시간(시간)		
		24	36	48
100	0.012	9.4	72.6	153.6
	0.06	55.0	60.0	66.6
	0.12	20.3	17.4	30.9
	0.6	6.8	5.9	2.5
200	0.012	37.3	129.0	205.4
	0.06	67.7	73.9	87.5
	0.12	27.7	22.0	34.7
	0.6	9.5	8.6	6.3
300	0.012	47.1	139.4	240.1
	0.06	70.7	78.9	98.3
	0.12	24.4	17.8	25.0
	0.6	2.8	6.4	-9.8

- [0109] ^a 상승효과(%) = $\left[\left\{ \frac{\text{BsEXLX1와 셀룰라제에 의해 생산된 환원당}}{\text{(셀룰라제 단독에 의해 생산된 환원당 + BsEXLX1 단독에 의해 생산된 환원당)}} \right\} - 1 \right] \times 100$
- [0110] BsEXLX1와 셀룰라제의 조성비가 상승효과에 결정적인 영향을 미치는 것으로 확인되었다. 상승효과는 BsEXLX1과 셀룰라제로 인큐베이션한 경우의 환원당 수율을 셀룰라제 단독 또는 BsEXLX1 단독으로 인큐베이션한 경우의 수율의 합으로 나누어 계산하였다. 표 1에 제시한 여러 시험조건 중에서, 셀룰로스 1 g 당 셀룰라제의 양이 0.012 FPU이고 BsEXLX1의 양이 300 μg 인 경우 반응시간이 48 시간이었을 때 상승효과가 240%로 가장 높았다. 이러한 상승효과는 셀룰라제와 BsEXLX1을 함께 가한 경우의 당 수율이 BsEXLX1 단독 및 셀룰라제 단독의 경우를 합친 수율에 비해 2.4 배 높고, 셀룰라제 단독에 의한 당 수율에 비해 5.9 배 높음을 의미한다.
- [0111] 동일한 미생물 균주로부터 얻은 셀룰라제, 예컨대 CBH와 EG의 상승효과(Medve et al. 1994; Woodward et al. 1988)와, 서로 다른 생물들로부터 얻은 셀룰라제의 상승효과(Irwin et al. 1993)는 연구된 바 있다. 그러나, 비가수분해 단백질에 의한 상승효과에 대해서는 알려진 바가 많지 않다. 식물 유래 β -익스펜신을 셀룰라제와 함께 셀룰로스에 가한 경우, 48 시간 후에 당 전환율이 익스펜신이 없는 경우에 비해 13% 증가하였다(Baker et al. 2000).
- [0112] 초본 화분에서 유래한 GR2의 재조합 단백질을 CBH I 및 EG II와 함께 거름종이를 18 시간 동안 인큐베이션하여, 효소만을 사용하여 거름종이를 인큐베이션한 경우에 비해 8.3 배의 당 수율이 얻어졌다(Cosgrove 2007). GR2와 함께 사용한 셀룰라제의 양이 특정되지 않았으나, 18 시간 후의 최종 당 수율이 이론적 최대값의 0.6%와 4.8%이므로 매우 낮았을 것으로 생각된다. 셀룰라제와 씨모비피다 푸스카에서 유래한 CBM의 상승효과에 대한 연구에서, 셀룰라제 Cel6A 및 Cel6B를 E7와 혼합한 경우의 셀로비오스 수율 증가는 셀룰라제만을 사용하여 168 시간 동안 반응한 경우에 비해 각각 12.9%와 7.5%에 불과하였다(Moser et al. 2008). 이 연구에서, Cel6B를 E7 및 E8와 혼합한 경우의 상승효과 역시 각각 이론적 최대값의 5.0%와 5.1%에 불과하였다.
- [0113] 옥수수 부산물에서 정제된 미지의 비가수분해 단백질을 셀룰로스 1 g 당 0.84 FPU의 셀룰라제와 함께 인큐베이션하였을 때, 거름종이로부터의 포도당 수율은 셀룰라제만으로 인큐베이션하였을 때에 비해 3.2 배 증가하였다(Han and Chen 2007). 이 실험의 경우 셀룰라제의 양이 적었으므로, 포도당 수율의 증가(이론적 최대값의 2.7%에서 8.2%로)는 낮은 셀룰로스 전환율에서 얻어진 것이다.
- [0114] <실시예 7>
- [0115] **BsEXLX1의 셀룰로스 결합활성 분석**
- [0116] 셀룰로스 결합활성을 측정하기 위하여, 6 mm 디스크 형태의 거름종이를 사용하여 PCR 플레이트 상에서 단백질과 거름종이를 인큐베이션하였다. 단백질 대조군으로서 15 μg 의 BSA(Sigma, St. Louis, MO, USA), 10 μg 의 BsEXLX1 또는 0.5 FPU의 셀룰라제(54 μg 의 BSA에 상응하는 양)를 포함하는 각 반응혼합물 100 μL 를 40 $^{\circ}\text{C}$ 에서 인큐베이션하였다. 0, 6, 12 및 24 시간 인큐베이션 후 샘플들(상등액 및 거름종이 디스크)을 취하였다. 거름종이를 20 μL 의 완충액으로 2 회 세척하여 얻은 거름종이 용리액과 상등액을 15% SDS-PAGE로 분석한 후 쿠마시 염색하였다.
- [0117] 표 1에서 보듯이, 예컨대 셀룰로스 1 g 당 0.6 FPU의 경우와 같이 셀룰라제의 양이 많은 경우에는, BsEXLX1의 양을 300 μg 으로 증가시킨 경우 상승효과가 다소 감소하였으며 음의(negative) 상승효과가 나타나기도 하였다. 이와 같은 음의 상승효과는 CBH I와 EG II 또는 EG I와 EG II의 상호작용에서도 보고된 바 있다(Woodward et al. 1988). 두 경우 모두 EG는 포화상태였으며 음의 상승효과는 기질의 결합부위에 대한 경쟁에 의한 것으로 생각된다. 셀룰라제와 BsEXLX1는 셀룰라제와 BsEXLX1의 농도가 높은 경우 거름종이의 결합부위를 놓고 경쟁하는 것일 수 있다. 셀룰로스 파괴활성 또는 식물 세포벽 이완활성을 갖는 익스펜신 역시 셀룰로스 또는 다른 다당류에 대한 결합활성을 가지므로(Yennawar et al. 2006), BsEXLX1의 상승효과는 셀룰로스 기질과의 결합에 의해 시작되는 것일 수 있다. 셀룰라제와 BsEXLX1이 결합부위를 공유하는 경우, BsEXLX1와 셀룰라제가 셀룰로스와 결합을 놓고 경쟁할 가능성이 있다. 결과적으로, 셀룰라제와 BsEXLX1의 양이 많은 경우에는 상승효과가 완화되는 것이다.
- [0118] 옥수수로부터 유래한 b-익스펜신이 다양한 다당류(셀룰로스, 크실란, 갈락탄 등)에 대해 결합활성을 가지므로(Yennawar et al. 2006), 본 발명자들은 BsEXLX1 또한 셀룰로스에 대한 결합활성을 가지는지 확인하였다. 도 9에서 보듯이, BsEXLX1와 셀룰라제, BSA를 독립적으로 사용하여 거름종이를 0, 6, 12 또는 24 시간 동안 인큐베이션하였다. 각 인큐베이션 혼합물의 상등액(유리된 단백질) 또는 거름종이(결합된 단백질)로부터 얻은 단백질을 SDS-PAGE로 분석하였다. 거름종이를 셀룰라제만으로 인큐베이션한 경우, 75 kDa에서의 상등액의 주된 단백질

밴드가 인큐베이션 시간이 증가함에 따라 크게 감소하였다(도 9의 (a)). 셀룰라제 밴드가 약화된 것은 셀룰라제 용액에 포함되었을 수 있는 단백질 분해효소에 의해 셀룰라제가 분해되었기 때문일 수 있다. 그러나, 도 9의 (e) 및 (f)에서 보듯이, 75 kDa에서의, 결합된 셀룰라제에 의한 밴드는 시간에 따라 확연하게 증가하였다. 따라서, 셀룰라제 용액에 포함되었을 수 있는 단백질 분해효소에 의한 단백질의 분해에 의한 영향은 미미한 것으로 생각된다. 이와 대조적으로, 거름종이로부터 얻은 셀룰라제의 밴드 강도는 인큐베이션 시간이 증가함에 따라 증가하였다(도 9의 (e)). 이러한 결과는 셀룰라제가 거름종이와 결합한다는 것을 강하게 확인시켜 준다. 대조 단백질인 BSA로 거름종이를 인큐베이션한 경우, 상등액과 거름종이 모두 셀룰라제의 밴드 강도는 크게 변하지 않았다(도 9의 (d)와 (h)). 이는 이미 보고된 바와 같이(Rudolf et al. 2008; Yang et al. 2006), BSA는 순수한 셀룰로스에 잘 결합하지 않는다는 것을 확인시켜 준다. 셀룰라제의 거름종이에 대한 결합특성은 셀룰라제를 BsEXLX1와 함께 인큐베이션한 경우에도 확인된다(도 9의 (b) 및 (f)).

[0119] BsEXLX1은 단독으로 또는 셀룰라제와 함께 인큐베이션하였을 때 거름종이에 대하여 강한 결합활성을 보였다. 상등액으로부터 얻은 BsEXLX1의 밴드 강도는 시간에 따라 약해졌으나(도 9의 (b)와 (c)), 거름종이로부터 얻은 BsEXLX1의 밴드 강도는 시간에 따라 증가하였다(도 9의 (f)와 (g)). 따라서, 위에서 언급한 바와 같이, 셀룰라제와 BsEXLX1의 농도가 일정 수준 이상으로 증가하였을 때 관찰되는 BsEXLX1에 의한 상승효과의 감소는 셀룰로스 기질의 한정된 결합부위에 대한 두 단백질의 경쟁에 기인하는 것으로 생각된다.

[0120] <실시예 8>

[0121] 온도가 셀룰로스 가수분해의 상승효과에 미치는 영향

[0122] 앞선 실험에서는 셀룰로스 가수분해의 반응온도를 제조업체에서 권장하는 셀룰라제의 적정 온도인 50 ℃로 고정하였다. 셀룰라제만을 또는 셀룰라제와 BsEXLX1를 함께 사용하여 인큐베이션한 거름종이로부터의 당 수율은 50 ℃에서 가장 높았으며(도 10), 반응온도가 더 높아지면 감소하였다. 반응온도가 70 ℃인 경우에는 반응수율이 다른 온도에 비해 매우 낮아졌는데, 이것은 효소가 열에 의해 변성되었기 때문으로 생각된다. 반응온도에 따른 BsEXLX1에 의한 셀룰라제 활성의 증가는 셀룰라제 단독의 경우와 같은 양상을 보였다.

[0123] <실시예 9>

[0124] pH가 셀룰로스 가수분해의 상승효과에 미치는 영향

[0125] 앞선 실험에서는 모두 제조업체에서 권장하는 셀룰라제의 적정 pH인 pH 4.8에서 가수분해 반응을 실시하였다. 본 발명자들은 최대 가수분해 수율을 위한 적정 pH가 BsEXLX1를 첨가하는 경우에도 유지되는지 조사하였다. pH 3-7의 범위에 대해 실험한 결과, 셀룰라제의 활성(BsEXLX1을 가하지 않은 경우) pH 4, 4.8 및 5에서 거의 일정하였는데(도 11), 이는 셀룰라제의 활성이 pH 4~6 범위에서 크게 변하지 않음을 의미한다. 그러나, BsEXLX1을 가한 경우, 셀룰라제의 활성은 pH에 더 민감하게 반응하였다. 이것은 BsEXLX1이 셀룰라제보다 pH에 민감한 것과 관련이 있을 것으로 생각된다.

[0126] <실시예 10>

[0127] BsEXLX1로 처리한 거름종이의 인장강도

[0128] BsEXLX1에 의한 거름종이의 약화 정도를 확인하기 위하여, BsEXLX1으로 처리한 거름종이의 인장강도를 인장강도 시험기(UTM; Instron, Norwood, MA, USA)로 측정하였다. 와트만 No. 3 거름종이를 2 × 5.5 cm의 띠 모양으로 잘라 UTM의 클램프 사이에 장착하였다. 클램핑된 거름종이 띠를 600 g/mL의 BsEXLX1 또는 대조용 단백질로서 BSA를 포함하는 10 mL의 아세트산나트륨 완충액(50 mM, pH 4.8) 내에서 인큐베이션하였다. 음성 대조군으로서, BsEXLX1과 같은 농도의 BSA를 포함하는 완충액 또는 완충액만을 사용하여 거름종이를 인큐베이션하였다. 양성 대조군으로서, 8 M 요소 용액을 사용하였다. 아래쪽 클램프에 0.5 mm/분의 크로스헤드 속도로 5 g의 하중을 10 분간 가하여 인큐베이션한 거름종이 띠를 잡아당겼다. 최대 힘 값(F_{max})을 측정하고, 식 $r_{max} = F_{max}/A$ (A = 단면적, F_{max} = 최대하중(kg))로부터 인장강도(r_{max})를 계산하였다. 실험은 3회 실시하였으며, 결과를 평균 ± 표준편차로 나타내었다.

[0129] BsEXLX1이 셀룰로스의 파괴에 미치는 영향을 조사하기 위하여, 정제된 BsEXLX1을 포함하는 완충액, BSA(음성 대조군), 완충액 단독(음성 대조군) 또는 8 M 요소 용액(양성 대조군) 내에서 거름종이 띠를 1 시간 동안 인큐베이션하였다. 도 12에서 보듯이, 완충액과 BSA를 포함하는 완충액에서 인큐베이션한 거름종이 띠의 인장강도는 별다른 차이가 없었다. 8 M 요소 용액을 이용한 양성 대조군 실험의 경우, 완충액 단독의 경우에 비해 인장강도가 50% 감소하였다. BsEXLX1을 포함하는 완충액에서 인큐베이션한 거름종이 띠의 인장강도는 완충액 단독의 경

우에 비해 29% 감소하였다. 이와 같이, 거름종이의 약화 정도를 확인하기 위한 인장시험으로부터 BsEXLX1 역시스월레닌(Baker et al. 2000; Saloheimo et al. 2002)이나 익스팬신(Cosgrove 2000a; Cosgrove 2000b; McQueen-Mason and Cosgrove 1994)과 유사하게 셀룰로스에 대한 파괴활성을 가짐을 알 수 있었다.

[0130] <실시예 11>

[0131] **박테리아 유래 익스팬신 BsEXLX1으로 처리한 거름종이의 SEM**

[0132] BsEXLX1으로 처리한 거름종이의 미세구조 변화를 주사전자현미경법(SEM)으로 분석하였다. 3 개의 다른 거름종이 샘플, 즉 완충액만으로 인큐베이션한 거름종이, BsEXLX1으로 인큐베이션한 거름종이, 그리고 8 M 요소 용액으로 인큐베이션한 거름종이를 거름종이의 인장강도 측정실험에서와 동일한 조건에서 준비하였다. SEM 분석에 앞서, 거름종이 샘플을 45 °C에서 1 일 동안 진공 건조오븐에서 건조한 후 금-팔라듐으로 코팅하였다. 주사전자현미경(Hitachi S-4700, Tokyo, Japan)을 사용하여 10 kV의 전압에서 샘플의 현미경사진을 촬영하였다.

[0133] SEM에 의해 촬영된 거름종이 샘플의 현미경사진을 도 13에 도시하였다. SEM 분석 결과, 다른 물질을 함유하는 완충액으로 인큐베이션한 거름종이 샘플들 간에 미세구조의 차이가 확인되었다. 양성 대조군으로서 8 M 요소 용액으로 인큐베이션한 거름종이는, 도 13의 (b)와 (f)에서 보듯이, 거름종이의 미세섬유 간의 상호연결과 중첩이 완화된 형태를 보였는데, 이는 요소의 파괴활성 때문으로 생각된다. 따라서, 도 12에서 확인된 8 M 요소 내에서의 인큐베이션에 따른 인장강도의 감소와 SEM 분석으로 확인된 구조적 변화가 잘 일치한다. 도 13의 (a)와 (e)에서 보듯이, BsEXLX1을 포함하는 완충액에서 인큐베이션한 거름종이도 이와 비슷한 구조적 변화를 보였다. 이것 역시 도 12의 BsEXLX1으로 처리한 거름종이의 인장강도 감소와 일치하는 결과이다. 한편, 완충액만으로(도 13의 (c)와 (h)) 또는 BSA를 포함하는 완충액(도 13의 (d)와 (i))으로 인큐베이션한 음성 대조군의 거름종이 샘플들은 미세섬유가 더 조밀하였으며 상호연결과 중첩 역시 강하였다. 그러므로, 완충액만으로 또는 BSA를 포함하는 완충액으로 인큐베이션한 거름종이 샘플의 SEM 분석결과 역시 도 12의 인장강도 데이터와 잘 일치한다.

[0134] 본 발명자들은 옥수수로부터 유래된 β -익스팬신과 구조적 유사성을 가지는, 바실러스 서브틸리스로부터 유래된 24 kDa 단백질인 BsEXLX1의 분자적 기능을 조사, 확인하였다. 얻어진 재조합 단백질은 식물 유래 익스팬신과 기능적 유사성을 갖는 것으로 확인되었다. BsEXLX1은 식물 유래 익스팬신과 유사하게, 거름종이에 대하여 셀룰로스 결합활성 및 셀룰로스 약화활성을 보였다. 또한, BsEXLX1은 반응완충액 내에 낮은 함량의 셀룰라제와 거름종이를 포함하는 셀룰로스 가수분해 반응혼합물에 가하였을 때 상당한 상승효과를 보였다. 셀룰라제의 양이 0.012 FPU이고 BsEXLX1의 양이 300 μ g이었을 때, BsEXLX1의 상승효과로 인해 당 수율이 셀룰라제 단독의 경우에 비해 5.7 배 증가하였다. BsEXLX1에 의한 거름종이의 약화 정도를 인장강도 측정을 통해 정량하였으며, 이를 SEM 분석에 의해 시각적으로 확인할 수 있었다.

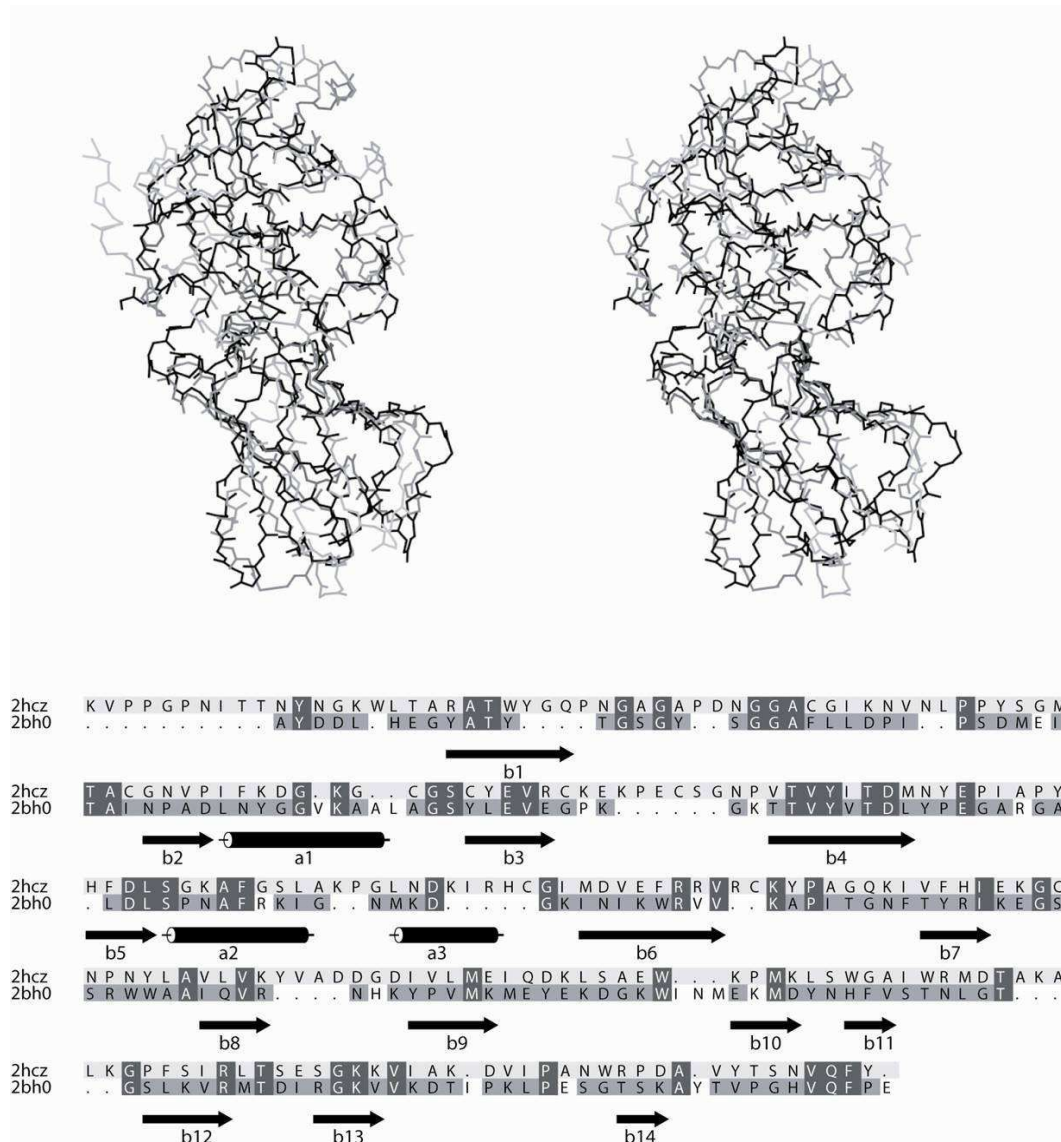
[0135] 결국 본 발명의 원핵생물 유래 익스팬신 단백질은 종래의 식물 유래 익스팬신 단백질과 동일한 기능(셀룰로스 확장활성)을 수행하면서도 식물 유래 익스팬신 단백질과는 달리 대량생산이 가능하고 생산비용을 현저히 저감시킬 수 있다. 특히 목질계 바이오매스를 셀룰라제로 가수분해하여 포도당을 얻고자 할 때 셀룰라제에 의한 셀룰로스 가수분해 속도를 높여 당 수율을 현저히 향상시킬 수 있다. 따라서, 실제 바이오에너지 생산에 있어 원핵생물 유래 익스팬신을 사용함으로써, 효소의 사용량을 줄여 바이오에너지의 생산비용을 저감할 수 있다. 또한, 펄프, 면 섬유(예, 청바지) 등의 조직을 연화시키고 질감을 개선하는 특성이 있어 바이오펄핑이나 바이오스토포닝과 같은 목적으로 이용될 수 있다.

산업상 이용가능성

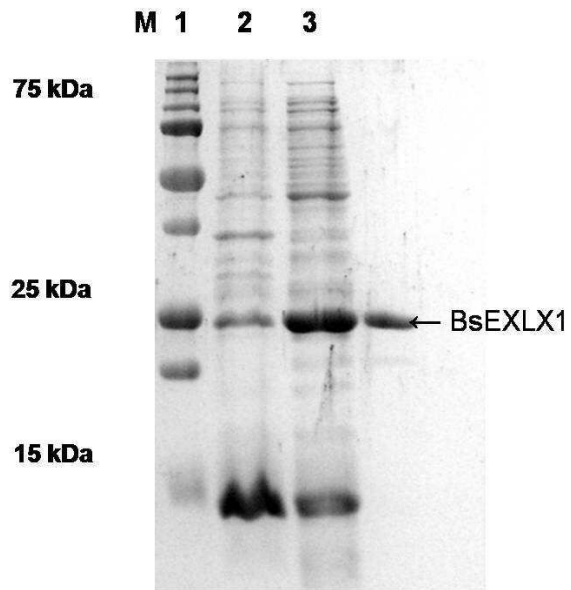
[0136] 원핵생물 유래 익스팬신을 이용하면 바이오에너지 생산에 있어 효소의 사용량을 줄여 생산비용을 저감할 수 있다. 또한, 펄프, 면 섬유(예, 청바지) 등의 조직을 연화시키고 질감을 개선하는 변화시키는 특성이 있어 제지산업에서의 바이오펄핑이나 섬유산업에서의 바이오스토포닝과 같은 목적으로 이용될 수 있다.

도면

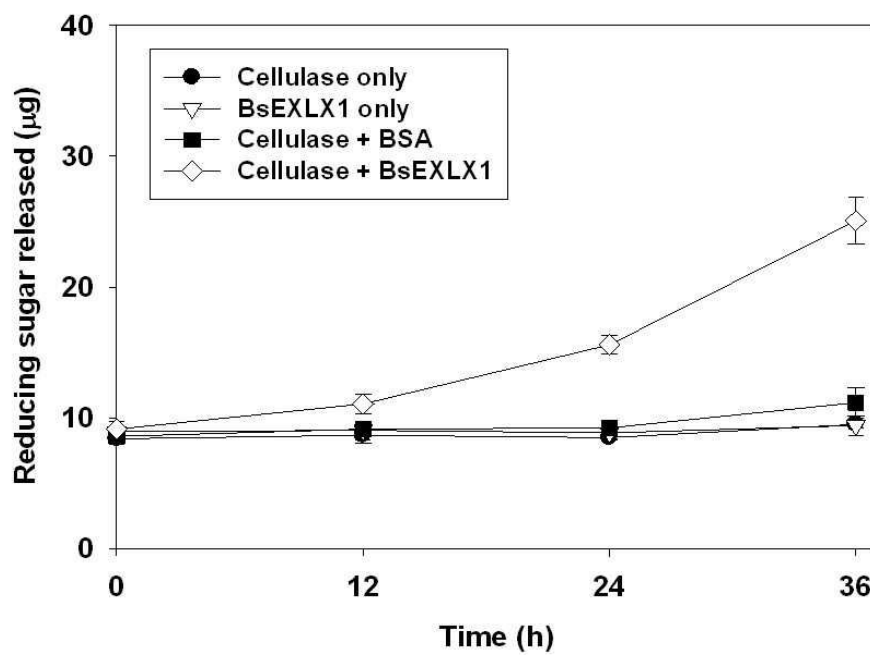
도면1



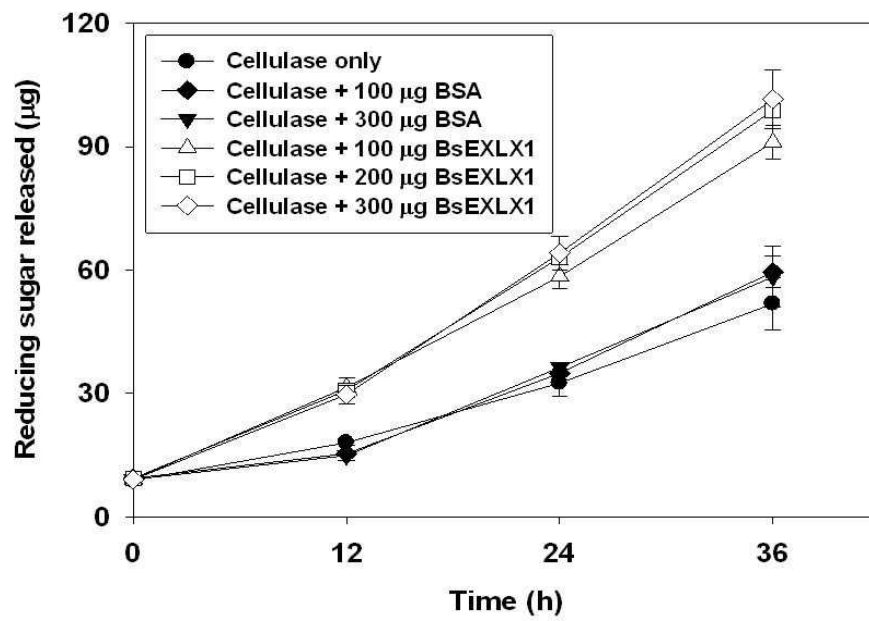
도면2



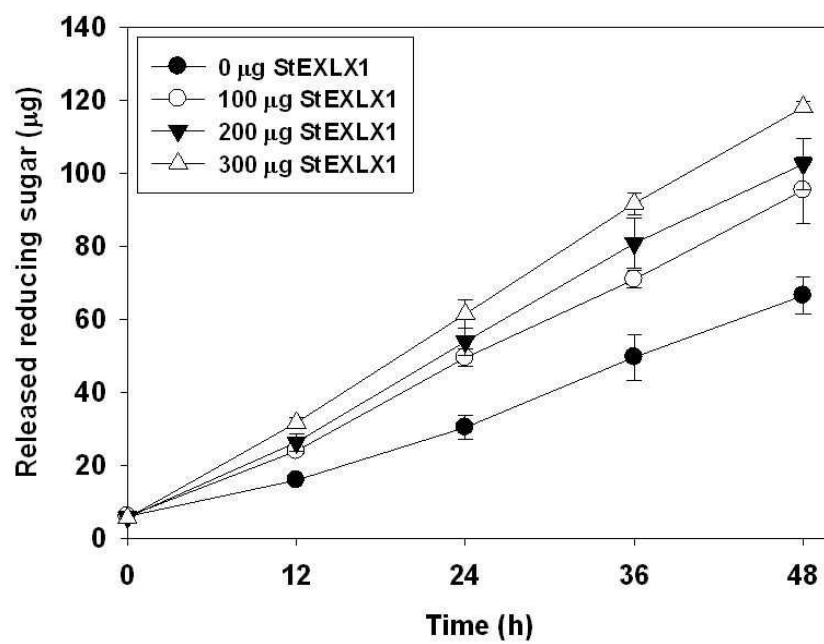
도면3



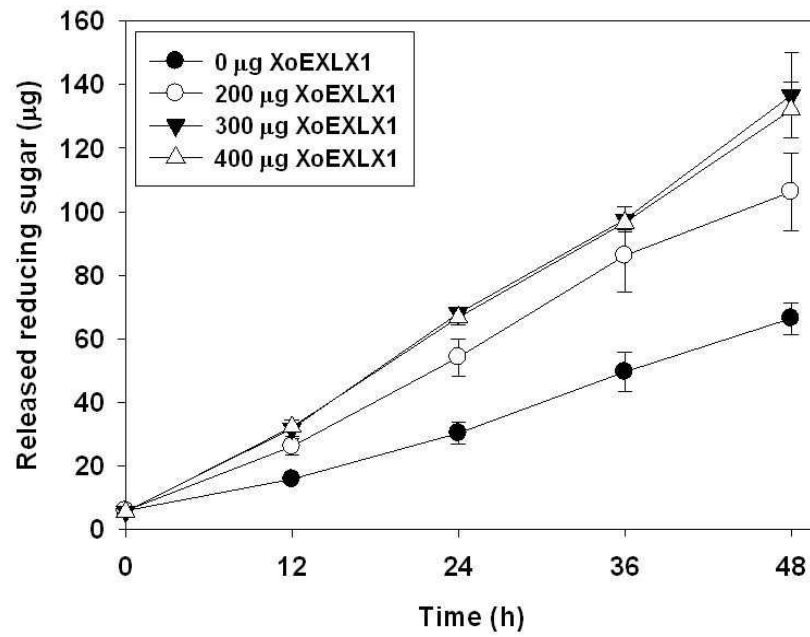
도면4



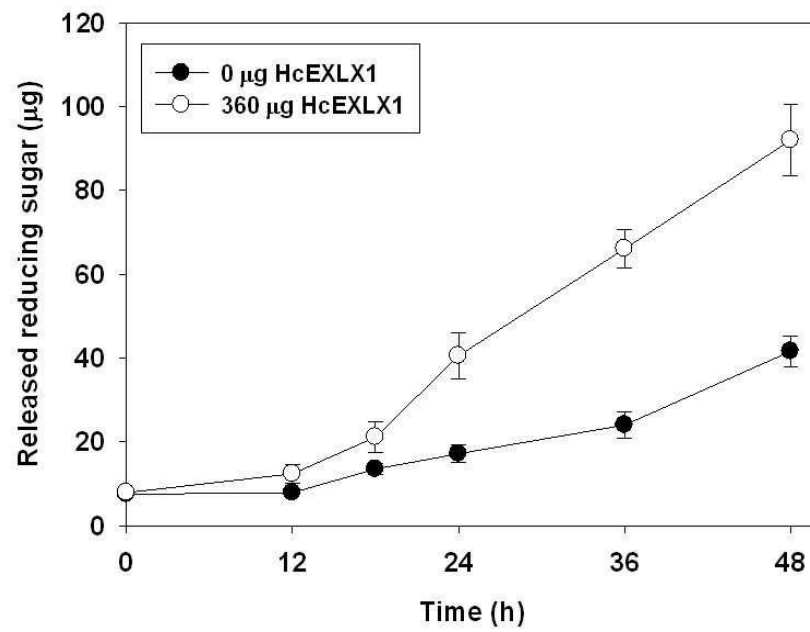
도면5



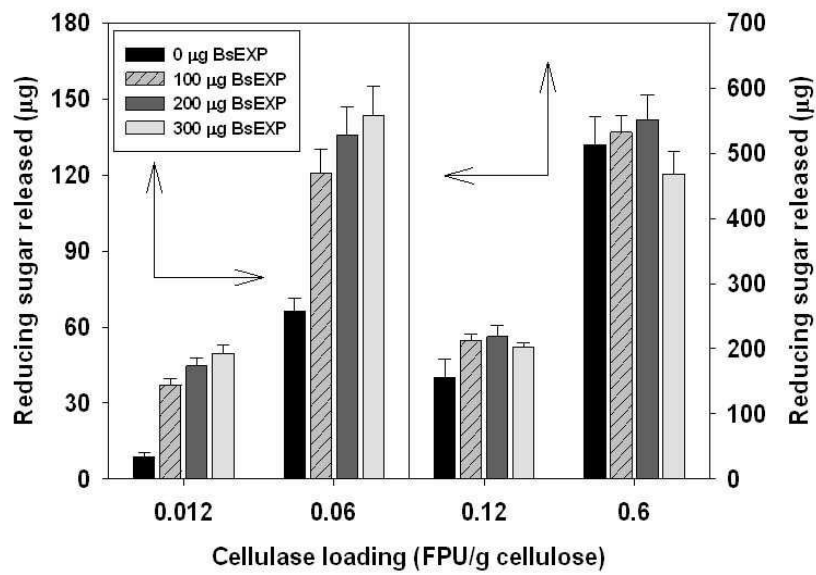
도면6



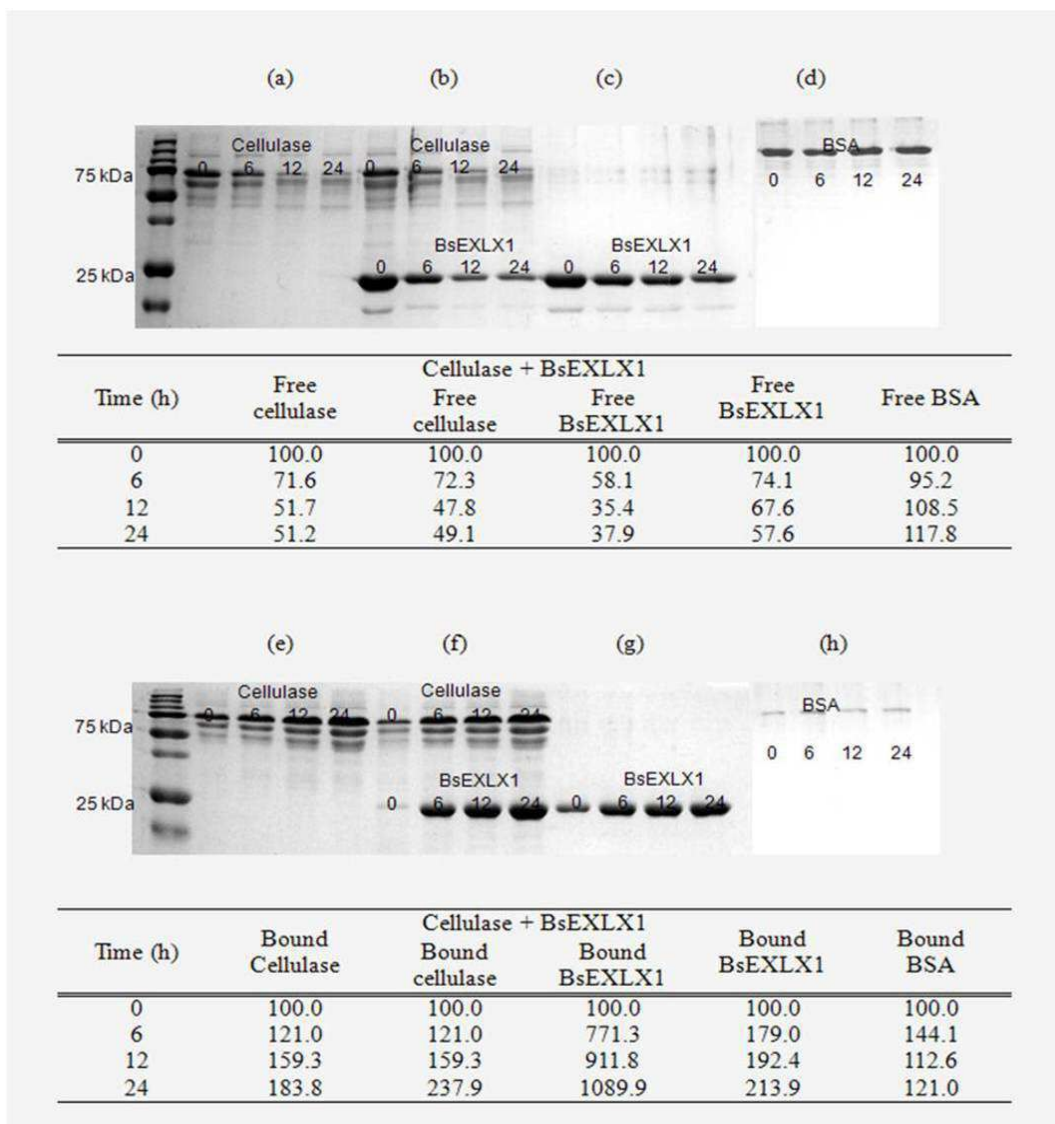
도면7



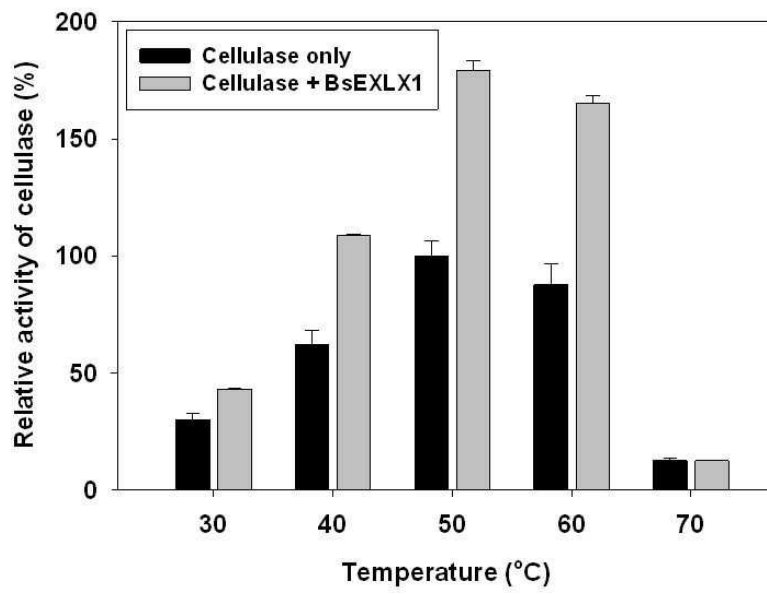
도면8



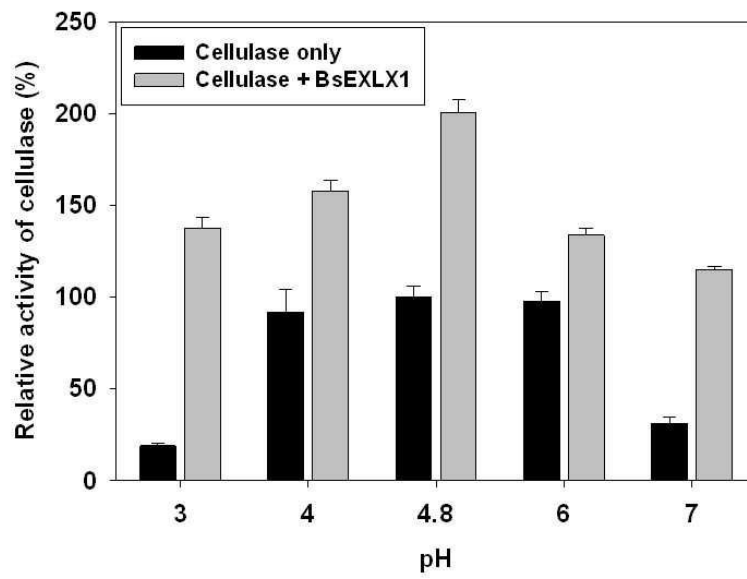
도면9



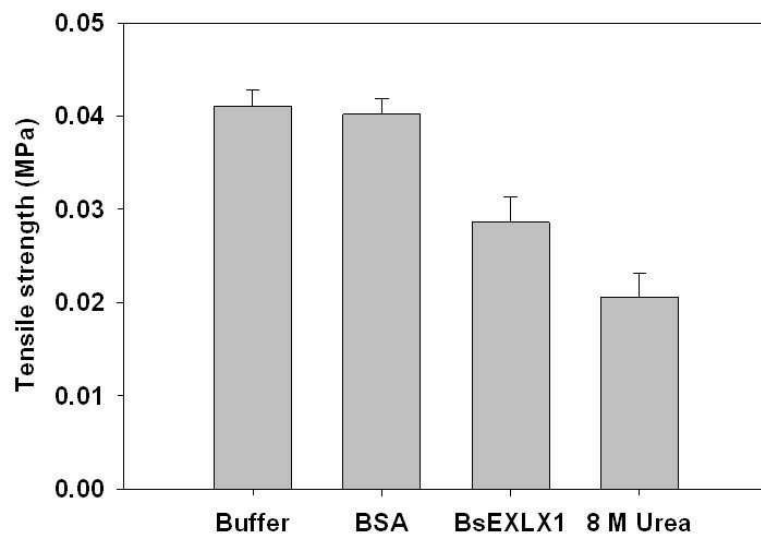
도면10



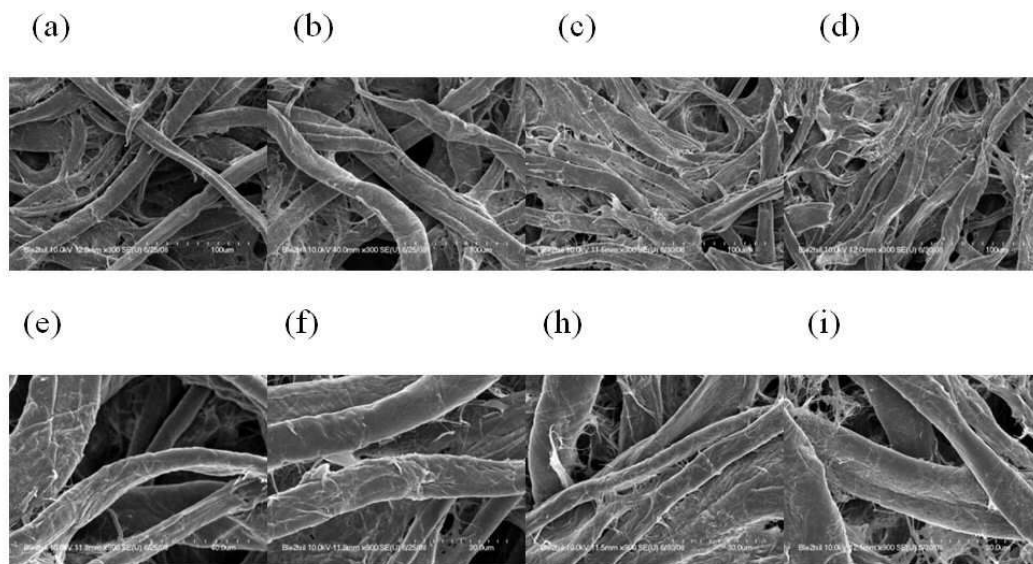
도면11



도면12



도면13



도면14a

Species	Genbank ID	E-value	aligned region
<i>Clavibacter michiganensis</i> ssp. <i>michiganensis</i> NCPPB 382	gi 148272660 ref YP_001222221.1	2.00E-11	163-239
<i>Herpetosiphon aurantiacus</i> ATCC 23779	gi 159897669 ref YP_001543916.1	1.00E-06	144-223
<i>Herpetosiphon aurantiacus</i> ATCC 23779	gi 159898342 ref YP_001544589.1	1.10E-06	144-222
<i>Roseiflexus</i> sp. RS-1	gi 148656687 ref YP_001275892.1	2.30E-10	142-220
<i>Roseiflexus castenholzii</i> DSM 13941	gi 156742279 ref YP_001432408.1	2.30E-10	143-220
<i>Chloroflexus aurantiacus</i> J-10-fl	gi 163845817 ref YP_001633861.1	1.50E-14	153-229
<i>Bacillus subtilis</i> ssp. <i>subtilis</i> 168	gi 16078923 ref NP_389744.1	1.80E-13	135-219
<i>Bacillus licheniformis</i> ATCC 14580	gi 52080641 ref YP_079432.1	2.30E-12	135-206
<i>Sorangium cellulosum</i> So ce 56	gi 162455954 ref YP_001618321.1	1.40E-06	255-327
<i>Sorangium cellulosum</i> So ce 56	gi 162453536 ref YP_001615903.1	3.20E-06	346-417
<i>Myxococcus xanthus</i> DK 1622	gi 108762346 ref YP_631763.1	6.50E-11	141-213
<i>Leptothrix cholodnii</i> SP-6	gi 171061047 ref YP_001793396.1	4.80E-12	161-235
<i>Ralstonia solanacearum</i> GM1000	gi 17545537 ref NP_518939.1	3.10E-09	149-221
<i>Acidovorax avenae</i> ssp. <i>citrulli</i> AAC00-1	gi 120612050 ref YP_971728.1	6.00E-07	159-231
<i>Hahella chejuensis</i> KCTC 2396	gi 83646399 ref YP_434834.1	7.60E-09	137-215
<i>Xylella fastidiosa</i> 9a5c	gi 15837412 ref NP_298100.1	8.30E-12	469-540
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	gi 58583659 ref YP_202675.1	1.40E-10	494-576
<i>Xanthomonas campestris</i> pv. <i>campestris</i> ATCC 33913	gi 21232964 ref NP_638881.1	4.00E-11	494-576
<i>Erwinia carotovora</i> ssp. <i>atroseptica</i> SOR11043	gi 50121146 ref YP_050313.1	7.60E-11	135-215
<i>Frankia</i> sp. EAN1pec	gi 158317309 ref YP_001509817.1	3.90E-14	201-284
<i>Bacillus licheniformis</i> ATCC 14580	gi 52786015 ref YP_091844.1	2.30E-12	135-206
<i>Xylella fastidiosa</i> M23	gi 182682471 ref YP_001830631.1	2.30E-11	488-559
<i>Xylella fastidiosa</i> M12	gi 170731100 ref YP_001776533.1	7.20E-12	488-559
<i>Xylella fastidiosa</i> Temecula1	gi 28199725 ref NP_780039.1	2.30E-11	473-544
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018	gi 84625463 ref YP_452835.1	1.40E-10	494-576
<i>Xanthomonas campestris</i> pv. <i>campestris</i> 8004	gi 66766963 ref YP_241725.1	4.00E-11	494-576
<i>Stigmatella aurantiaca</i> DW4/3-1	gi 115377675 ref ZP_01464869		

도면14b

BLAST 2.2.18+

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Sch  ffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1990), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Reference for compositional score matrix adjustment: Altschul, Stephen F., John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Sch  ffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", *FEBS J.* 272:5101-5109.

RID: 4NN4V9X01R

Database: All non-redundant GenBank CDS translations+DB+SwissProt+PIR+PRF excluding environmental samples from WGS projects

Results of PSI-Blast iteration 1

Query= gi|2619024|gb|AA084448.1| YoaJ [Bacillus subtilis]
Length=232

E-value BETTER than threshold	(Bits)	Value	Score
Sequences producing significant alignments:			
ref NP_389744.1 hypothetical protein BSU0630 [Bacillus subt...	497	4e-139	
pdb 2BBD A Chain A, Crystal Structure Of A Somet Derivative O...	422	1e-116	
ref YP_079432.1 YoaJ [Bacillus licheniformis ATCC 14880] >re...	418	2e-115	
ref YP_001902065.1 glycosidase [Xanthomonas campestris pv. c...	306	1e-81	
ref NP_638881.1 cellulase [Xanthomonas campestris pv. campe...	304	4e-81	
ref XP_02241682.1 cellulase [Xanthomonas oryzae pv. oryzae]...	295	3e-78	
ref YP_001912002.1 cellulase [Xanthomonas oryzae pv. oryzae]...	287	5e-76	
ref YP_002675.1 cellulase [Xanthomonas oryzae pv. oryzae]...	286	5e-76	
ref YP_452835.1 cellulase [Xanthomonas oryzae pv. oryzae]...	286	1e-75	
ref YP_001830631.1 Rare lipoprotein A [Xylella fastidiosa M2]...	273	1e-71	
ref XP_00681297.1 Glycoside hydrolase, family 5:Rare lipopro...	273	1e-71	
ref NP_780829.1 extracellular endoglucanase precursor [Xylel...	272	1e-71	
ref XP_00651400.1 Glycoside hydrolase, family 5:Rare lipopro...	272	1e-71	
ref NP_298100.1 extracellular endoglucanase precursor [Xylel...	269	1e-70	
ref XP_00944742.1 Hypothetical Protein RRS1_02398 [Ralstonia]...	264	6e-69	
ref YP_050513.1 putative cellulase [Klebsiella aeroterres subsp...	255	2e-66	
ref NP_518939.1 PUTATIVE ENDOGLUCANASE PROTEIN [Ralstonia so...	254	4e-66	
ref XP_01461139.1 YoaJ [Stigmatella aurantiaca DW4/3-1] >gb ...	165	3e-39	
ref XP_001275892.1 Rare lipoprotein A [Roseiflexus sp. RS-1]...	160	1e-37	
ref YP_001432408.1 Rare lipoprotein A [Roseiflexus castanhol...	157	1e-36	
ref YP_001623861.1 Rare lipoprotein A [Chloroflexus aurantia...	146	1e-33	
ref YP_001544589.1 coagulation factor S/B type domain protei...	141	5e-32	
ref XP_631763.1 putative lipoprotein Myxococcus xanthus DK...	130	1e-28	
ref YP_001793396.1 Rare lipoprotein A [Bradyrhizobium japonic...	130	1e-28	
ref YP_001509817.1 Rare lipoprotein A [Kasakia sp. KAN1pep]...	130	2e-28	
ref XP_01910738.1 putative lipoprotein [Planicystis pacific...	113	1e-23	
ref XP_001543916.1 Rare lipoprotein A [Borpetosiphon auranti...	108	4e-22	
ref YP_001705759.1 cellulase [Clavibacter michiganensis]...	108	5e-22	
ref XP_01464869.1 cellulase [Stigmatella aurantiaca DW4/3-1]...	101	6e-20	
ref YP_434834.1 Endoglucanase C-terminal domain/submit and ...	100	1e-19	
ref YP_001222221.1 putative expansin [Clavibacter michiganen...	97.8	7e-19	
ref YP_371728.1 Rare lipoprotein A [Bradyrhizobium japonic...	92.3	4e-17	
ref YP_001638321.1 hypothetical protein sce7672 [Sorangium c...	86.7	1e-15	
ref YP_001615903.1 hypothetical protein sce5260 [Sorangium c...	85.9	2e-15	
ref XP_01516438.1 YoaJ [Chloroflexus aggregans DSM 9485] >gb...	69.4	2e-10	
ref YP_001220664.1 cellulase [Clavibacter michiganensis subsp...	59.6	2e-07	
emb CAAM4467.2 Cellulase [Clavibacter michiganensis]	59.6	2e-07	
gb AAK1622.1 cellulase CelA [Clavibacter michiganensis subsp...	59.2	3e-07	
Sequences with E-value WORSE than threshold			
ref YP_001801897.1 hypothetical protein ccc_0480 [Cyanotheca]...	97.1	1.9	
ref XP_00514986.1 Rare lipoprotein A [Crocosphaera watsonii]...	36.7	1.8	
ref XP_01729178.1 rare lipoprotein A [Cyanotheca sp. CCY0110]...	35.0	5.1	
ref XP_02976533.1 rare lipoprotein A [Cyanotheca sp. PCC 742]...	34.1	9.8	
ref YP_001267828.1 methyltransferase small [Pseudomonas puti...	33.3	16	
ref XP_01089595.1 magnesium transport P-type ATPase [Synecho...	32.9	25	
ref NP_465585.1 hypothetical protein lmo2061 [Listeria monoc...	32.9	25	
ref NP_745395.1 modification methylase, HemK family [Pseudom...	32.9	25	
ref XP_212814.1 putative lipoprotein [Bacteroides fragilis N...	32.5	31	
ref YP_100649.1 hypothetical protein BF371 [Bacteroides fra...	32.5	32	
ref XP_02942337.1 Rare lipoprotein A [Cyanotheca sp. PCC 880]...	31.6	54	
ref XP_02290103.1 hypothetical protein lmonocyFSL15683 [Lis...	31.6	59	
ref XP_01667578.1 Ni/Fe-hydrogenase, b-type cytochrome subun...	31.6	64	
ref XP_01136926.1 putative sensory box histidine kinase/resp...	31.2	70	
ref XP_02066375.1 hypothetical protein BACOVA_03371 [Bactero...	31.2	82	
ref XP_747033.1 hypothetical protein Neut_0801 [Nitrosomonas]...	31.2	83	
ref XP_02942012.1 (p)ppGpp synthetase I, SpoT/RelA [Cyanotheca]...	31.2	87	
ref XP_01446236.1 phosphoribosylaminoimidazole synthetase [B...	30.8	96	
ref XP_01302790.1 TonB-dependent siderophore receptor [Sphin...	30.8	110	
ref XP_014685.1 hypothetical protein lmoF2365_2092 [Listeria]...	30.8	122	
ref XP_00989708.1 Rare lipoprotein A [Vibrio splendidus 12B0]...	30.3	123	
ref XP_984717.1 extracellular solute-binding protein, family...	30.3	126	
ref XP_01125796.1 hypothetical protein NB231_15703 [Nitrococ...	30.3	127	
ref XP_715082.1 hypothetical protein; putative nucleotide bi...	30.3	146	
ref XP_01125700.1 hypothetical protein NB231_15223 [Nitrococ...	30.3	146	
ref YP_001820117.1 glycoside hydrolase family 62 [Opuntia t...	30.3	157	
ref NP_953833.1 sensory box histidine kinase [Geobacter sulf...	30.3	157	
ref NP_219439.1 putative UvrD/REP helicase [Chlamydomonas ab...	30.3	162	
ref YP_001823109.1 hypothetical protein SGR_1597 [Streptomyces]...	29.9	182	
ref YP_001092559.1 hypothetical protein Shex_0428 [Shewanella]...	29.9	198	
ref XP_691862.1 nitrous-oxide reductase [Roseobacter denitri...	29.9	213	
ref YP_001416637.1 hypothetical protein Xaut_1735 [Xanthobac...	29.5	225	
ref YP_001250924.1 hypothetical protein LPC_1639 [Legionella]...	29.5	235	

도면14c

```

ref|SP_01718231.1| hypothetical protein ALP1_13615 [Algoriph... 29.5 240
ref|YP_608644.1| DNA methylase [Pseudomonas entomophila L48] ... 29.5 240
ref|NP_667538.1| hypothetical protein y0195 [Yersinia pestis ... 29.5 247
ref|YP_072042.1| hypothetical protein YPB3560 [Yersinia pseu... 29.5 247
ref|YP_001719226.1| conserved hypothetical protein [Yersinia ... 29.5 249
ref|NP_407126.1| hypothetical protein YP03670 [Yersinia pesti... 29.5 249
ref|SP_00366221.1| OGG419: APase involved in DNA repair [St... 29.5 249
pdb|2IDQ|A Chain A, Structure Of M98a Mutant Of Amicyanin, Cu... 29.5 251
ref|SP_01517774.1| para-aminobenzoate synthase, component I [... 29.5 263
ref|SP_02086149.1| hypothetical protein CLOBOL_03692 [Clostri... 29.5 268
ref|YP_001399396.1| hypothetical protein YpsP21758_0403 [Yer... 29.5 268
ref|SP_01039753.1| carbonic anhydrase [Krychrobacter sp. NAP1... 29.5 280
ref|SP_02466045.1| YedA C-terminal domain protein [Burkholder... 29.5 286
ref|SP_00238018.1| conserved hypothetical protein [Listeria m... 29.5 288
pdb|1LX2|A Chain A, Crystal Structure Of Selenomethionine Pro... 29.1 305
ref|YP_001899347.1| copper resistance protein CopC [Ralstonia... 29.1 307
ref|YP_001231231.1| SMC domain protein [Geobacter uranireduc... 29.1 310

ALIGNMENTS
>ref|NP_389744.1| hypothetical protein BSU16630 [Bacillus subtilis subsp. subtilis
str. 168]
gb|AA054448.1| YoaJ [Bacillus subtilis]
emb|CAB13755.1| yoaJ [Bacillus subtilis subsp. subtilis str. 168]
Length=232

Score = 497 bits (1165), Expect = 4e-139, Method: Compositional matrix adjust.
Identities = 232/232 (100%), Positives = 232/232 (100%), Gaps = 0/232 (0%)

Query 1 MGGKMSAFVGVKLLITFCFSDQASAAAYDDHMGYATTTGSGVSGGAFLLDDIPSIMKITA 60
      MGGKMSAFVGVKLLITFCFSDQASAAAYDDHMGYATTTGSGVSGGAFLLDDIPSIMKITA
Sbjct 1 MGGKMSAFVGVKLLITFCFSDQASAAAYDDHMGYATTTGSGVSGGAFLLDDIPSIMKITA 60

Query 61 INADILNYGCVKALAGSYLVKVGPKGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD 120
      INADILNYGCVKALAGSYLVKVGPKGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD
Sbjct 61 INADILNYGCVKALAGSYLVKVGPKGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD 120

Query 121 GKINIKGRVUKADITGNFTYRIKGGSSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD 180
      GKINIKGRVUKADITGNFTYRIKGGSSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD
Sbjct 121 GKINIKGRVUKADITGNFTYRIKGGSSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD 180

Query 181 YNHFVSTNLCTGSLKVMMDIRGKVVKDTIPKLESCTSKAYTVDGHVQFPE 232
      YNHFVSTNLCTGSLKVMMDIRGKVVKDTIPKLESCTSKAYTVDGHVQFPE
Sbjct 181 YNHFVSTNLCTGSLKVMMDIRGKVVKDTIPKLESCTSKAYTVDGHVQFPE 232

>pdb|2BB0|A Chain A, Crystal Structure Of A Semet Derivative Of Kapa From
Bacillus Subtilis At 2.5 Angstrom
Length=208

Score = 422 bits (988), Expect = 1e-116, Method: Compositional matrix adjust.
Identities = 200/207 (96%), Positives = 200/207 (96%), Gaps = 0/207 (0%)

Query 26 AYDDLHMGYATTTGSGVSGGAFLLDDIPSIMKITA INADILNYGCVKALAGSYLVKVG 85
      AYDDLHMGYATTTGSGVSGGAFLLDDIPSIMKITA INADILNYGCVKALAGSYLVKVG
Sbjct 2 AYDDLHMGYATTTGSGVSGGAFLLDDIPSIMKITA INADILNYGCVKALAGSYLVKVG 85

Query 86 KGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD GKINIKGRVUKADITGNFTYRIKGG 145
      KGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD GKINIKGRVUKADITGNFTYRIKGG
Sbjct 86 KGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD GKINIKGRVUKADITGNFTYRIKGG 145

Query 146 SSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD YNHFVSTNLCTGSLKVMMDIRGKV 205
      SSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD YNHFVSTNLCTGSLKVMMDIRGKV
Sbjct 146 SSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD YNHFVSTNLCTGSLKVMMDIRGKV 205

Query 206 VKDTIPKLESCTSKAYTVDGHVQFPE 232
      VKDTIPKLESCTSKAYTVDGHVQFPE
Sbjct 206 VKDTIPKLESCTSKAYTVDGHVQFPE 208

>ref|YP_079432.1| YoaJ [Bacillus licheniformis ATCC 14580]
ref|YP_091844.1| YoaJ [Bacillus licheniformis ATCC 14580]
gb|AA023794.1| YoaJ [Bacillus licheniformis ATCC 14580]
gb|AA041151.1| YoaJ [Bacillus licheniformis DSM 13]
Length=232

Score = 419 bits (979), Expect = 2e-115, Method: Compositional matrix adjust.
Identities = 194/232 (83%), Positives = 211/232 (90%), Gaps = 0/232 (0%)

Query 1 MGGKMSAFVGVKLLITFCFSDQASAAAYDDHMGYATTTGSGVSGGAFLLDDIPSIMKITA 60
      MGGKMSAFVGVKLLITFCFSDQASAAAYDDHMGYATTTGSGVSGGAFLLDDIPSIMKITA
Sbjct 1 MGGKMSAFVGVKLLITFCFSDQASAAAYDDHMGYATTTGSGVSGGAFLLDDIPSIMKITA 60

Query 61 INADILNYGCVKALAGSYLVKVGPKGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD 120
      INADILNYGCVKALAGSYLVKVGPKGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD
Sbjct 61 INADILNYGCVKALAGSYLVKVGPKGKTTVYVTDLYFGAGALDLSNPAFRKIGRMKD 120

Query 121 GKINIKGRVUKADITGNFTYRIKGGSSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD 180
      GKINIKGRVUKADITGNFTYRIKGGSSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD
Sbjct 121 GKINIKGRVUKADITGNFTYRIKGGSSRWAAIQVNNHCYPUMGMYKDGKGTINMGKMD 180

Query 181 YNHFVSTNLCTGSLKVMMDIRGKVVKDTIPKLESCTSKAYTVDGHVQFPE 232
      YNHFVSTNLCTGSLKVMMDIRGKVVKDTIPKLESCTSKAYTVDGHVQFPE
Sbjct 181 YNHFVSTNLCTGSLKVMMDIRGKVVKDTIPKLESCTSKAYTVDGHVQFPE 232

>ref|YP_001902065.1| glycosidase [Xanthomonas campestris pv. campestris]
emb|CAP49999.1| glycosidase [Xanthomonas campestris pv. campestris]
Length=666

```

도면14d

```

Score = 306 bits (715), Expect = 1e-81, Method: Compositional matrix adjust.
Identities = 138/199 (69%), Positives = 160/199 (80%), Gaps = 0/199 (0%)

Query 33 GYATYTCGSGYSGGAFLLDDIPSDIMEKITAINPADLNYGGVKAALAGSYLVKVGDPGKETTUY 92
          G AT TCGSGYSGGA LLDIPSD ITA+NP LN+GGVKAALAG+YL+V GREG TTUY
Sbjct 468 GTATYTCGSGYSGGALLDDIPSDAFITALNFPVQLNFGGVKAALAGAYLQVNGDPGKETTUY 527

Query 93 VTDLYPEGARGALDLSDNAFRKICGNMKDCKINIKGRUVKADITCNFTYRIKEGSSSRWAA 152
          VTDLYPEGA G LDLS NAF IG+M G+I I W+VV+AD+TGN YRIKEGSSSRWAA
Sbjct 528 VTDLYPEGASGGLDLSHNAFAAIGDMVQGRIPISMKVVRAPVTGNLQYRIKEGSSSRWAA 587

Query 153 IQVRHHKYPVMKMKYKCKGKWINMKMDYHHFVSTNLGTGSLKVRMTDIRGKVKVDIDPK 212
          IQVRHH YPV+K+K ++ W N+HMDYHHF+ LG L +R+TDIRKV+ DT+P+
Sbjct 588 IQVRHHAYPVVKLVKVGQSGTWKHLQMDYHHFLGQLGNQPLTALITDIRGKVLITDLR 647

Query 213 LPESGTSKAYTVPGHVQFP 231
          LPE G+ AY DGHVQFP
Sbjct 648 LPEGSGSKAYTDPGHVQFP 666

>ref|NP_638881.1| cellulase [Xanthomonas campestris pv. campestris str. ATCC 33913]
ref|YP_241725.1| cellulase [Xanthomonas campestris pv. campestris str. 8004]
gb|AA042805.1| cellulase [Xanthomonas campestris pv. campestris str. ATCC 33913]
gb|AA047705.1| cellulase [Xanthomonas campestris pv. campestris str. 8004]
Length=590

Score = 304 bits (711), Expect = 4e-81, Method: Compositional matrix adjust.
Identities = 138/199 (69%), Positives = 160/199 (80%), Gaps = 0/199 (0%)

Query 33 GYATYTCGSGYSGGAFLLDDIPSDIMEKITAINPADLNYGGVKAALAGSYLVKVGDPGKETTUY 92
          G AT TCGSGYSGGA LLDIPSD ITA+NP LN+GGVKAALAG+YL+V GREG TTUY
Sbjct 392 GTATYTCGSGYSGGALLDDIPSDAFITALNFPVQLNFGGVKAALAGAYLQVNGDPGKETTUY 451

Query 93 VTDLYPEGARGALDLSDNAFRKICGNMKDCKINIKGRUVKADITCNFTYRIKEGSSSRWAA 152
          VTDLYPEGA G LDLS NAF IG+M G+I I W+VV+AD+TGN YRIKEGSSSRWAA
Sbjct 452 VTDLYPEGASGGLDLSHNAFAAIGDMVQGRIPISMKVVRAPVTGNLQYRIKEGSSSRWAA 511

Query 153 IQVRHHKYPVMKMKYKCKGKWINMKMDYHHFVSTNLGTGSLKVRMTDIRGKVKVDIDPK 212
          IQVRHH YPV+K+K ++ W N+HMDYHHF+ LG L +R+TDIRKV+ DT+P+
Sbjct 512 IQVRHHAYPVVKLVKVGQSGTWKHLQMDYHHFLGQLGNQPLTALITDIRGKVLITDLR 571

Query 213 LPESGTSKAYTVPGHVQFP 231
          LPE G+ AY DGHVQFP
Sbjct 572 LPEGSGSKAYTDPGHVQFP 590

>ref|NP_02241682.1| cellulase [Xanthomonas oryzae pv. oryzicola BLS256]
Length=590

Score = 295 bits (688), Expect = 3e-78, Method: Compositional matrix adjust.
Identities = 141/199 (70%), Positives = 159/199 (79%), Gaps = 0/199 (0%)

Query 33 GYATYTCGSGYSGGAFLLDDIPSDIMEKITAINPADLNYGGVKAALAGSYLVKVGDPGKETTUY 92
          G AT TCGSGYSGGA LLDIPSD ITA+NP LN+GGVKAALAG+YL+V GREG TTUY
Sbjct 392 GTATYTCGSGYSGGALLDDIPSDAFITALNFPVQLNFGGVKAALAGAYLQVNGDPGKETTUY 451

Query 93 VTDLYPEGARGALDLSDNAFRKICGNMKDCKINIKGRUVKADITCNFTYRIKEGSSSRWAA 152
          VTDLYPE A G LDLS NAF IGDM DG I I W+VV+AD+TGN YRIKEGSSSRWAA
Sbjct 452 VTDLYPEAASGGLDLSYNFAAIGDMADGDIPISMKVVRADITGNVQYRIKEGSSSRWAA 511

Query 153 IQVRHHKYPVMKMKYKCKGKWINMKMDYHHFVSTNLGTGSLKVRMTDIRGKVKVDIDPK 212
          IQVRHH YPV+K+K ++ W ++ K DYHHFV TRLG L +R+TDIRKV+ D IP
Sbjct 512 IQVRHHIYPVVKLVKVGQSGTWISLQCTDYHHFVGTNLGNKPLSIRITDIRGKVLIDKIDA 571

Query 213 LPESGTSKAYTVPGHVQFP 231
          LPE G SKAY DGHVQFP
Sbjct 572 LPEYGGSKAYTDPGHVQFP 590

>ref|YP_001912002.1| cellulase [Xanthomonas oryzae pv. oryzae PKO99A]
gb|ACDS7470.1| cellulase [Xanthomonas oryzae pv. oryzae PKO99A]
Length=590

Score = 287 bits (671), Expect = 5e-76, Method: Compositional matrix adjust.
Identities = 137/199 (68%), Positives = 157/199 (78%), Gaps = 0/199 (0%)

Query 33 GYATYTCGSGYSGGAFLLDDIPSDIMEKITAINPADLNYGGVKAALAGSYLVKVGDPGKETTUY 92
          G AT TCGSGYSGGA LLDIPSD ITA+H LN+GG+KAALAG+YLEV GREGKETTUY
Sbjct 392 GTATSTCGSGYSGGALLDDIPSDAFITALNSTQNNFGGKAALAGAYLVKVTGDPGKETTUY 451

Query 93 VTDLYPEGARGALDLSDNAFRKICGNMKDCKINIKGRUVKADITCNFTYRIKEGSSSRWAA 152
          VTDLYP+GA G LDLS NAF IGH+ DG I I W+VV+AD+TGN YRIKEGSSSRWAA
Sbjct 452 VTDLYPDGASGGLDLSYNFAAIGNLADGIPISMKVVRADITGNVQYRIKEGSSSRWAA 511

Query 153 IQVRHHKYPVMKMKYKCKGKWINMKMDYHHFVSTNLGTGSLKVRMTDIRGKVKVDIDPK 212
          IQVRHH YPV+K+K ++ W ++ K YHHFV TRLG L +R+TDIRKV+ D IP
Sbjct 512 IQVRHHIYPVVKLVKVGQSGTWISLQCTDYHHFVGTNLGNKPLSIRITDIRGKVLIDKIDA 571

Query 213 LPESGTSKAYTVPGHVQFP 231
          LPE G S AY DGHVQFP
Sbjct 572 LPEYGGSAAYTDPGHVQFP 590

```

도면14e

```
>ref|YP_202675.1| cellulase [Xanthomonas oryzae pv. oryzae KACC10331]
gb|AAW77290.1| cellulase [Xanthomonas oryzae pv. oryzae KACC10331]
Length=590

Score = 286 bits (669), Expect = 8e-76, Method: Compositional matrix adjust.
Identities = 137/199 (68%), Positives = 157/199 (78%), Gaps = 0/199 (0%)

Query 33 CYATYTCGCGYSGGAFLLDPIPSIMKITAIPADLNYGGVKAALAGSYLVEGPEKKTIVY 92
      G AT TCGCGYSGGA LLDPIPSD ITA+N +H+GG+KAALAG+YLEV GPEKKTIVY
Sbjct 392 GTATSTCGCGYSGGAVLLDPIPSDAFITALNSTQMNFGGKAALAGAYLEVGTGPEKKTIVY 451

Query 93 VTDLYPEGARGALDLSNPAFRKIGNMKDKGKINIKWRUVKAPITGNFTYRIKEGSSRWAA 152
      VTDLYP+GA G LDLS NAF IGN+ DG I I W+VV+APITGN YRIKEGSSR+WAA
Sbjct 452 VTDLYPDGASGGLDLSYNFAAAGNLADGHIPISEKVVRAPIITGNVQYRIKEGSSRYWAA 511

Query 153 IQVRHHYVPMOMYKEDGKWINRMMDYHHFVSTNLCTGSLKVRMTDIRGVVVDITPE 212
      IQVR+H YPV+K+E ++ W ++ K YHHFV TNLG L +R+TDIRGV+ D IP
Sbjct 512 IQVRHHYVPMOMYKEDGKWINRMMDYHHFVSTNLCTGSLKVRMTDIRGVVVDITPE 571

Query 213 LPESGTSKAYTVDGHRVQFD 231
      LPE G S AY DGRVQFD
Sbjct 572 LPEYGSAAYPEGHRVQFD 590

>ref|YP_452835.1| cellulase [Xanthomonas oryzae pv. oryzae MAFF 311016]
dbj|BAK70561.1| cellulase [Xanthomonas oryzae pv. oryzae MAFF 311016]
Length=590

Score = 286 bits (668), Expect = 1e-75, Method: Compositional matrix adjust.
Identities = 138/206 (66%), Positives = 160/206 (77%), Gaps = 0/206 (0%)

Query 26 AYDDLHEGYATYTCGCGYSGGAFLLDPIPSIMKITAIPADLNYGGVKAALAGSYLVEGEP 85
      A++ G AT TCGCGYSGGA LLDPIPSD ITA+N +H+GG+KAALAG+YLEV GP
Sbjct 385 AWWSTFTGTATPTCGCGYSGGAVLLDPIPSDAFITALNSTQMNFGGKAALAGAYLEVGTGP 444

Query 86 KGKTTVVYVTDLYPEGARGALDLSNPAFRKIGNMKDKGKINIKWRUVKAPITGNFTYRIKEG 145
      KGKTTVVYVTDLYP+GA G LDLS NAF IGN+ DG I I W+VV+APITGN YRIKEG
Sbjct 445 KGKTTVVYVTDLYPDGASGGLDLSYNFAAAGNLADGHIPISEKVVRAPIITGNVQYRIKEG 504

Query 146 SSRWGAALQVRHHYVPMOMYKEDGKWINRMMDYHHFVSTNLCTGSLKVRMTDIRGV 205
      SSR+WGAALQVR+H YPV+K+E ++ W ++ K YHHFV TNLG L +R+TDIRGV
Sbjct 505 SSRYWGAALQVRHHYVPMOMYKEDGKWINRMMDYHHFVSTNLCTGSLKVRMTDIRGV 564

Query 206 VKDTIDPLPESGTSKAYTVDGHRVQFD 231
      + D IP LPE G S AY DGRVQFD
Sbjct 565 IADKIPALPEYGSAAYPEGHRVQFD 590

>ref|YP_001930631.1| Rare lipoprotein A [Xylella fastidiosa M23]
gb|ACB93357.1| Rare lipoprotein A [Xylella fastidiosa M23]
Length=584

Score = 273 bits (637), Expect = 1e-71, Method: Compositional matrix adjust.
Identities = 124/207 (59%), Positives = 158/207 (76%), Gaps = 0/207 (0%)

Query 26 AYDDLHEGYATYTCGCGYSGGAFLLDPIPSIMKITAIPADLNYGGVKAALAGSYLVEGEP 85
      +D G ATYT +GY GGA +LDPI S ++ITA+ND IN GG+ AA+AG+YL V+GP
Sbjct 378 GWTSESGVATYTYTCGCGGALMDPIQSHVQITALEPQTQNLGGIPAAAGAYLVVQGP 437

Query 86 KGKTTVVYVTDLYPEGARGALDLSNPAFRKIGNMKDKGKINIKWRUVKAPITGNFTYRIKEG 145
      KG TTUVYVTDLYP G+ G LDLSNPAF IGIM G+I ++W+VV AD+IGN YR+K+G
Sbjct 438 KGSTTVYVTDLYPTGSSGGLDLSNPAFASIGNMAQCRIIVQWKVVSAPVSGNLIYRVKGG 497

Query 146 SSRWGAALQVRHHYVPMOMYKEDGKWINRMMDYHHFVSTNLCTGSLKVRMTDIRGV 205
      SS WGAALQVR H+YPV+K+E +DG W+N+ F +YH+VV T LG L +R+TDIRG+
Sbjct 498 SSCWGAALQVRHHYVPMOMYKEDGKWINRMMDYHHFVSTNLCTGSLKVRMTDIRGQT 557

Query 206 VKDTIDPLPESGTSKAYTVDGHRVQFD 232
      + DT+D LP +SKAY+V G+VQF E
Sbjct 558 LIDTLALPDKASSKAYSVNGRVQFSE 584

>ref|XP_00681297.1| Glycoside hydrolase, family 5:Rare lipoprotein A [Xylella fastidiosa
Ann-1]
gb|KA033194.1| Glycoside hydrolase, family 5:Rare lipoprotein A [Xylella fastidiosa
Ann-1]
Length=584

Score = 273 bits (637), Expect = 1e-71, Method: Compositional matrix adjust.
Identities = 124/207 (59%), Positives = 158/207 (76%), Gaps = 0/207 (0%)

Query 26 AYDDLHEGYATYTCGCGYSGGAFLLDPIPSIMKITAIPADLNYGGVKAALAGSYLVEGEP 85
      +D G ATYT +GY GGA +LDPI S ++ITA+ND IN GG+ AA+AG+YL V+GP
Sbjct 378 GWTSESGVATYTYTCGCGGALMDPIQSHVQITALEPQTQNLGGIPAAAGAYLVVQGP 437

Query 86 KGKTTVVYVTDLYPEGARGALDLSNPAFRKIGNMKDKGKINIKWRUVKAPITGNFTYRIKEG 145
      KG TTUVYVTDLYP G+ G LDLSNPAF IGIM G+I ++W+VV AD+IGN YR+K+G
Sbjct 438 KGSTTVYVTDLYPTGSSGGLDLSNPAFASIGNMAQCRIIVQWKVVSAPVSGNLIYRVKGG 497

Query 146 SSRWGAALQVRHHYVPMOMYKEDGKWINRMMDYHHFVSTNLCTGSLKVRMTDIRGV 205
      SS WGAALQVR H+YPV+K+E +DG W+N+ F +YH+VV T LG L +R+TDIRG+
Sbjct 498 SSCWGAALQVRHHYVPMOMYKEDGKWINRMMDYHHFVSTNLCTGSLKVRMTDIRGQT 557

Query 206 VKDTIDPLPESGTSKAYTVDGHRVQFD 232
      + DT+D LP +SKAY+V G+VQF E
Sbjct 558 LIDTLALPDKASSKAYSVNGRVQFSE 584
```

도면14f

```
>ref|NP_780039.1| extracellular endoglucanase precursor [Xylella fastidiosa Temecula1]
gb|AA029688.1| extracellular endoglucanase precursor [Xylella fastidiosa Temecula1]
Length=569

Score = 273 bits (637), Expect = 1e-71, Method: Compositional matrix adjust.
Identities = 124/207 (59%), Positives = 158/207 (76%), Gaps = 0/207 (0%)

Query 26 AYDDLHEGYATVTCGSGGAFLLDDPISIMKITAIPADLNYGCVKAALAGSYLVEGDP 85
      +D C ATYT +GY CGA +LDPI S ++ITA+ND LN CG+ AA+AG+YL V+GP
Sbjct 363 GWDTSFSGVATYTTTCYKGGALMLDPIQSHVQITALNPTQLNLGGIDAMAGAYLRVQGP 422

Query 86 EGKTTVVVVDLYPFGARGALDLSDFNAFRKIGNMCDGKINIKRWVUKAPITGNFTYRIKGG 145
      EG TTUVVVDLYP G+ G LDLSDFNAF IGMN G+I +W+VV AP+GN YR+K+G
Sbjct 423 EGKTTVVVVDLYPFGSGGLDLSDFNAFASIGMAQGRIPVQWVUSAPVSGNLIYRVKGG 482

Query 146 SSRWGAAIQVRNRYVPMGMGYEKDGRWINGMGMQDYNHFVSTNLCTGSLKVRMTDIRGKV 205
      SS WGAAIQVR B+YPU+K+E +DG W++ K +YR+V T LG L +R+TDIRG+
Sbjct 483 SSGWGAAIQVRNRYVPLKLEICQDGTWMLPGRNRYVFCVTRIGNQPLSGRMTDIRGQT 542

Query 206 VKDTIPKLPSGCTSKAYTVPGHVQFPE 232
      + DT+D LP +SKAY+V G+VQF E
Sbjct 543 LIDTLPALPKASSKAYSUNGHVQFSE 569

>ref|XP_00651400.1| Glycoside hydrolase, family 5:Rare lipoprotein A [Xylella fastidiosa
Dixon]
>ref|XP_00681922.1| Glycoside hydrolase, family 5:Rare lipoprotein A [Xylella fastidiosa
Ana-1]
>ref|YP_001776533.1| extracellular endoglucanase precursor [Xylella fastidiosa M12]
gb|BA013637.1| Glycoside hydrolase, family 5:Rare lipoprotein A [Xylella fastidiosa
Dixon]
gb|KA032615.1| Glycoside hydrolase, family 5:Rare lipoprotein A [Xylella fastidiosa
Ana-1]
gb|JCA12903.1| extracellular endoglucanase precursor [Xylella fastidiosa M12]
Length=584

Score = 272 bits (636), Expect = 1e-71, Method: Compositional matrix adjust.
Identities = 124/207 (59%), Positives = 158/207 (76%), Gaps = 0/207 (0%)

Query 26 AYDDLHEGYATVTCGSGGAFLLDDPISIMKITAIPADLNYGCVKAALAGSYLVEGDP 85
      +D C ATYT +GY CGA +LDPI S ++ITA+ND LN CG+ AA+AG+YL V+GP
Sbjct 378 GWDTSFSGVATYTTTCYKGGALMLDPIQSHVQITALNPTQLNLGGIDAMAGAYLRVQGP 437

Query 86 EGKTTVVVVDLYPFGARGALDLSDFNAFRKIGNMCDGKINIKRWVUKAPITGNFTYRIKGG 145
      EG TTUVVVDLYP G+ G LDLSDFNAF IGMN G+I +W+VV AP+GN YR+K+G
Sbjct 438 EGKTTVVVVDLYPFGSGGLDLSDFNAFASIGMAQGRIPVQWVUSAPVSGNLIYRVKGG 497

Query 146 SSRWGAAIQVRNRYVPMGMGYEKDGRWINGMGMQDYNHFVSTNLCTGSLKVRMTDIRGKV 205
      SS WGAAIQVR B+YPU+K+E +DG W++ K +YR+V T LG L +R+TDIRG+
Sbjct 498 SSGWGAAIQVRNRYVPLKLEICQDGTWMLPGRNRYVFCVTRIGNQPLSGRMTDIRGQT 557

Query 206 VKDTIPKLPSGCTSKAYTVPGHVQFPE 232
      + DT+D LP+ +SKAY+V G+VQF E
Sbjct 558 LIDTLPALPKASSKAYSUNGHVQFSE 584

>ref|NP_298100.1| extracellular endoglucanase precursor [Xylella fastidiosa 9a5c]
gb|AA033620.1|AK033920_1| extracellular endoglucanase precursor [Xylella fastidiosa 9a5c]
Length=565

Score = 269 bits (628), Expect = 1e-70, Method: Compositional matrix adjust.
Identities = 123/200 (61%), Positives = 156/200 (78%), Gaps = 0/200 (0%)

Query 33 GYATVTCGSGGAFLLDDPISIMKITAIPADLNYGCVKAALAGSYLVEGDPGKCTTVY 92
      G ATYT +GY CGA +LDPI S ++ITA+ND LN CG+ AA+AG+YL V+GPKG TTUY
Sbjct 366 GYATVTCGSGGAFLLDDPIQSHVQITALNPTQLNLGGIDAMAGAYLRVQGPDKGCTTVY 425

Query 93 VTDLYPFGARGALDLSDFNAFRKIGNMCDGKINIKRWVUKAPITGNFTYRIKGGSSRWGA 152
      VTD YP G+ G ALDLSDFNAF IGMN G+I +W+VV AP+GN YR+K+GSS WGA
Sbjct 426 VTDYPTGSGGALDLSDFNAFASIGMAQGRIPVQWVUSAPVSGNLIYRVKGGSSRWGA 485

Query 153 IQVRNRYVPMGMGYEKDGRWINGMGMQDYNHFVSTNLCTGSLKVRMTDIRGKVVDITPK 212
      IQVR B+YPU+K+E +DG W++ K +YR+V T LG L +R+TDIRG+ + DT+D
Sbjct 486 IQVRNRYVPLKLEICQDGTWMLPGRNRYVFCVTRIGNQPLSGRMTDIRGQTLIDTLP 545

Query 213 LPESGCTSKAYTVPGHVQFPE 232
      LP+ +SKAY+V G+VQF E
Sbjct 546 LPKRASSKAYSUNGHVQFSE 565

>ref|XP_00944742.1| Hypothetical Protein RRS1_02398 [Ralstonia solanacearum UW551]
gb|KAP72760.1| Hypothetical Protein RRS1_02398 [Ralstonia solanacearum UW551]
Length=274

Score = 264 bits (615), Expect = 6e-69, Method: Compositional matrix adjust.
Identities = 127/217 (58%), Positives = 157/217 (72%), Gaps = 10/217 (4%)

Query 26 AYDDLHEGYATVTCGSGGAFLLDDPISIMKITAIPADLNYGCVKAALAGSYLVEGDP 85
      A+D C AT TCSCGSGGAFLLDPI D EITA+ND N CG+AA+AG+YL V+GP
Sbjct 58 AWDSTFTGIATATGSCGSGGAFLLDPIAKDAKITAIPVQANLGGIDAMAGAYLRVQGP 117

Query 86 EGKTTVVVVDLYPFGARGALDLSDFNAFRKIGNMCDGKINIKRWVUKAPITGNFTYRIKGG 145
      EG TTUVVVDLYP A G LDLS NAF KIG+++ G+I I+W+++ AP+GN YRIKGG
Sbjct 118 EGKTTVVVVDLYPFAASGGLDLSDFNAFSGIDGQGRIDIRKELIAPVTHGVVYRIKGG 177

Query 146 SSRWGAAIQVRNRYVPMGMGYEKDGRWINGMGMQDYNHFVSTNLCTGSLKVRMTDIRGKV 205
      S+ WGAAIQVRNRYVPU+K+E + W+++ K DYNHFV T LG L +R+TDIRG++
Sbjct 178 SSGWGAAIQVRNRYVPLKLEVFGCAWVILPADYRHFVQGLGIDVIRITDIRGKI 237

Query 206 VKDTIPKL-----PSGCTSKAYTVPGHVQFPE 232
      + D +D L S SK Y V G+VQF E
Sbjct 238 LVNKLDPILKDKCTPKSANQSPCKPYFVQGNVQFPE 274
```


도면14g

```
>ref|YP_050313.1| putative cellulase [Erwinia carotovora subsp. atroseptica SCRI1043]
emb|CAG7512.1.1| putative cellulase [Erwinia carotovora subsp. atroseptica SCRI1043]
Length=229

Score = 255 bits (596), Expect = 2e-66, Method: Compositional matrix adjust.
Identities = 126/235 (53%), Positives = 168/235 (71%), Gaps = 11/235 (4%)

Query 1  MGGKMSAFVGMVLLTIFCFSPQASAAY-----DDLHEGYATYTCGSGYSGGAFLLDPIPSD 55
          M K I S      + L+ C D + A+      DD+ GYAT TCGSY GGA LLDPIP +
Sbjct 1  MNRKTS-----LALSALCVIPLINTAHAGWELDDICYGYATATGSGYSGGALLDPIPN 55

Query 56  MEITAINPADINYGCVKAALAGSYLEVEGPKCKTIVYVTDLYPEGARGALDLSNPAFRKI 115
          MEITA+N + L+Y CVKA+LAG+YL+V GPKG T VYVTDLYPEG ALDLS NAF KI
Sbjct 56  MEITAINRSQLDYRCGVKASLAGAYLVNGPKGSTVYVYVTDLYPEGGDCALDLSNPAFEKI 115

Query 116  GNMKDCCKINIKWRVVKADITCNFTYRIKEGSSRWAAIQVRNHKYPVMRMEYEKDKGWIN 175
          C++DGRINI W +++AD+ GN YRIKEGS+ +WAA+Q RN KYPV++M+Y +D +W+
          GILDGCKINILDWLIEAFVKGHVYIYRIKEGSPYAAVGFHVHYFVIEGKYNHNGWA 175

Query 176  MERKMDYNHFVSTNLCTGSLKVRMTDIRKVVKDTIPKLESGTSKAYTVPGHVQF 230
          +K DYNHF+ ++G + + TD+G V+ DT+D + +S SS AY + G+VQ
Sbjct 176  AGCTDYNHFIVEHVMNDIDIEFTDVKGNVLSDTLPMSQS-TSSAYLITGNVQL 229

>ref|NP_518939.1| PUTATIVE ENDOGLUCANASE PROTEIN [Ralstonia solanacearum GMI1000]
emb|CAD14520.1| putative endoglucanase c-terminal domain harboring protein [Ralstonia
solanacearum]
Length=256

Score = 254 bits (593), Expect = 4e-66, Method: Compositional matrix adjust.
Identities = 124/210 (59%), Positives = 153/210 (72%), Gaps = 10/210 (4%)

Query 33  GYATYTCGSGYSGGAFLLDPIPSDMEITAINPADINYGCVKAALAGSYLEVEGPKCKTIVY 92
          G AT TCGSGYSGGAFLLDPI D EITA+NPA N GG+ AA+AG+YL V+GPKG TTVY
Sbjct 47  GATATCGSGYSGGAFLLDPIPTKDRREITAINPAQANLGGIPAAHAGAYLRVQGPKGCTTVY 106

Query 93  VTDLYPEGARGALDLSNPAFRKIGNMKDCCKINIKWRVVKADITCNFTYRIKEGSSRWAA 152
          VTDLYPE A G LDLS NAF KIC++ G+I ++WR++ P+TCN YRIKEGS+ WAA
Sbjct 107  VTDLYPEAASGLDLSYNPAFAKIGDLGQGRIPVQWRLIKGPFVTHVYRIKEGSPMWAA 166

Query 153  IQVRNHKYPVMRMEYEKDKGWINMERKMDYNHFVSTNLCTGSLKVRMTDIRKVVKDTIPK 212
          IQVRNH +PVK+E + W++ K DYNHF T LG L ++R+TDING++ D +D
Sbjct 167  IQVRNHTEFPVVKLEVFQCKAWVSLPKADYNHFVCTQLGDKPLVIRITDIRCRILVDKLP 226

Query 213  L-----DE-----SGTSKAYTVPGHVQFPE 232
          L P+ S SK Y V G VQF E
Sbjct 227  LKDCCTPQKAGEASPCSKPYFVQGVQFSE 256

>ref|ZP_01461139.1| YoaJ [Stigmatella aurantiaca DW4/3-1]
gb|EA068056.1| YoaJ [Stigmatella aurantiaca DW4/3-1]
Length=243

Score = 165 bits (382), Expect = 3e-39, Method: Compositional matrix adjust.
Identities = 87/224 (38%), Positives = 129/224 (57%), Gaps = 26/224 (11%)

Query 9  VGMVLLTIFCFSPQASA-----AYDDLHEGYATY---TCGSGYSGGAFLLDPI 53
          VGM+ LT C S + A + +G AT+ TC G+ G D P
Sbjct 15  VGMGLGTA-CSSSEPDSSGFTNDPEGEVRALGEFRQGIATWYNATGEGHCG----YDASP 69

Query 54  SDMEITAINPADINYGCVKAALAGSYLEVEGPKCKTIVYVTDLYPEGARGALDLSNPAFR 113
          DM+ A+N +A+ GS EVEGPKG V + D PE G LDLS AF
Sbjct 70  RDMVAAMNAPQF----ANSVCGSCAEVEGPKCTVVRVRIVDSCEPCGPGHLDLSEGAFA 125

Query 114  KIGNMKDCCKINIKWRVVKADITCNFTYRIKEGSSRWAAIQVRNHKYPVMRMEYEKDKGW 172
          KI + DG++ +WR V ++G YR KEGSS+WW AIQVRNH+ D+ K+E++++ G
Sbjct 126  KLAADVADGRVGTWRPFVTCVSGPVYRFKEGSSQWWTATQVRNHRLPIQGLEWQRENGA 185

Query 173  WINMERKMDYNHFV-STNLCTGSLKVRMTDIRKVVKDTIPKLE 215
          W++H++ DYN+H+ S + T + K+R+T G++H+DT+D++ E
Sbjct 186  WVMQRQDYNVFLASPGVDTATKLRVTASDCQVLEDTLPVKE 229

>ref|YP_001275892.1| Rare lipoprotein A [Roseiflexus sp. RS-1]
gb|ABQ89942.1| Rare lipoprotein A [Roseiflexus sp. RS-1]
Length=238

Score = 160 bits (370), Expect = 1e-37, Method: Compositional matrix adjust.
Identities = 82/197 (41%), Positives = 119/197 (60%), Gaps = 13/197 (6%)

Query 37  YTCGSGYSGGAFLLDPIPSDMEITAINPADINYGCVKAALAGSYLEVEGPKCKTIVYVTDL 96
          Y G +F DP D D+ +A+N AD + AAL G++EV GPKG+ V + D
Sbjct 50  YNADGTGNCSE--DPSPELDMVAAMNBADYD-----NAALCCTFIEVICPKRVVVRIVDR 103

Query 97  YDEARGALDLSNPAFRKIGNMKDCCKINIKWRVVKADITCNFTYRIKEGSSRWAAIQVR 156
          PE ARG +D+SP AF +I ++ G++ I+WR+V +TC YR KEGSS+WW A+Q+R
Sbjct 104  CDEARGCDVDMSPQAFARIADLSAGRVDIRWRIVSDALTCPLAYRKEGSSQWWTAVQIR 163

Query 157  NIKYPVMRMEYEKDKGWINMERKMDYNHFV-STNLCTGSLKVRMTDIRKVVKDTIPKLE 214
          NH+ PV + EY + DG+W ++ N+EV S+ +G C R+TD+ G V+ DT L
Sbjct 164  NHRHPVARFEYLRSDGQWQTVPRVMDHYFVASSGMCGPCPTTFRTVDVYGHVLTDTGVPIL 223

Query 215  ESGTSKAYTVPGHVQFP 231
          E C + G QFP
Sbjct 224  EGG-----VINGAAQFP 235
```


도면14h

```
>>ref|YP_001422408.1| Rare lipoprotein A [Roseiflexus castaneholisii DSM 13941]
gb|ABU55390.1| Rare lipoprotein A [Roseiflexus castaneholisii DSM 13941]
Length=227

Score = 157 bits (262), Expect = 1e-36, Method: Compositional matrix adjust.
Identities = 95/246 (38%), Positives = 136/246 (55%), Gaps = 26/246 (10%)

Query 1 MKKMSAFUGMVLITICFSPQASAAVDDLHEGY-----ATY-TGSG--YSG---GAF 47
      M + AF G+ LL + P A +D H Y A Y TG G V+ G
Sbjct 1 MPILAILAFAGCLL-LAGLIDVVRAQ-RDTFTVPLPLVVRDAQYKGGSTVVAADGTNC 59

Query 40 LLDIPSPDMEITAINPADLNYGGVKAALAGSYLEVEGPEKGTTVYVTDLYPEGARGALDL 107
      + DP D D+ + A+N D + AAL G++EV DNG TV + D PE ARG +D+
Sbjct 59 MFDSPHDIAVAAAMNIDVD---NAALCGAFIEVIGPKGSUTVRIVDRCEPCARGDVIM 114

Query 108 SPNAFRKIGNMKDGGKINIKRWAVFKAPITGNFTYVIRKEGSSRWAAIQVRNHKYPVMMEY 167
      SP AF +I ++ G++ I+RW+V + G YR KEGS+HW A+Q+RNH+ PV + EY
Sbjct 115 SPQAFERIAIDLAGRVPIRWLVSPNLAGPIAYRKEGSSQWNTAVQIRNHNVPVARFEY 174

Query 168 EK-DGKWINMEKMDYNHFV-STNLGTSLKVRMTDIRKVVKDTIPKLPSGTSKAYTVP 225
      + DG+M + + YN+EV S+ +G G R+TD+ G V+ DT E G +
Sbjct 175 LRADGQWCTVPRGMVNYFVASSGMSPGPYTFRVTDMYGNVLTDTDIPFVEGG-----VIN 229

Query 226 GHVQFP 231
      G QFP
Sbjct 220 GAAQFP 228

>>ref|YP_001628961.1| Rare lipoprotein A [Chloroflexus aurantiacus J-10-E1]
ref|XP_02985499.1| Rare lipoprotein A [Chloroflexus sp. Y-400-E1]
gb|ABY33472.1| Rare lipoprotein A [Chloroflexus aurantiacus J-10-E1]
gb|EDU23459.1| Rare lipoprotein A [Chloroflexus sp. Y-400-E1]
Length=258

Score = 146 bits (329), Expect = 1e-22, Method: Compositional matrix adjust.
Identities = 80/217 (36%), Positives = 117/217 (53%), Gaps = 14/217 (6%)

Query 25 AAYDDLHEGYATYTGSGYSGGAFLLDPIPSDMEITAINPADLNYGGVKAALAGSYLEVEG 84
      A + L G AT+ G+ L DPID D + AI+ LNYG A G+Y+EV G
Sbjct 43 PAQNPPLSGEATFYLEDGSGSCLEDFPIGDKHVAIS--YLNYG--NADYCGAYVEVG 95

Query 85 PKGTTTUVYVTDLYPEGA---RGALDSPNAFRKIGNMKDGGKINIKRWAVFKAPITGNFTYR 141
      P+G V + D+ P+ + LDLSF AF +I G++ I WRV+ P+G+ +
Sbjct 89 PQGTVVVRIVDMCPDNPGCCGNHLDSPFAEDRIAPRWGRUPITRWVISPGLSGPVQFH 158

Query 142 IKEGSSRWAAIQVRNHKYPVMMEY-EKDGEWINMEKMDYNHFV---STNLGTSLKV 196
      +K+GS+ HW A QVR+H+ P+ +MEY G+H+ + + YN+P+ N G +
Sbjct 159 LKDGSSPWLAIQVRNHNPIARMEYRTPQGQWVQLNTTNYTIRQCQCANGELGPTL 218

Query 197 RMTDIRKVVKDTIP--KLPSGTSKAYTVPGRVQFP 231
      R+TDI G V+ +T+ L S S V G QFP
Sbjct 199 RITDIYGNVLTETVGTFTLSRSSNSPELVSGTQFP 255

>>ref|YP_001544559.1| coagulation factor 5/8 type domain protein [Herpetosiphon aurantiac
ATCC 29779]
gb|ABX04461.1| coagulation factor 5/8 type domain protein [Herpetosiphon aurantiac
ATCC 29779]
Length=477

Score = 141 bits (326), Expect = 5e-22, Method: Compositional matrix adjust.
Identities = 91/231 (39%), Positives = 124/231 (53%), Gaps = 22/231 (9%)

Query 12 VLLITICFSPQASA---AYDD---LHEGYATYTGSGYSGGAFLLDPIPSDMEITAINPA 64
      +LLE F Q SA D L G AT+ G L P+PDM + A+N
Sbjct 17 LLLTSALFVAGSARHMQTLADPWALRSGRATFYDPTVGMGNCSL-PVPSIMLLAAMTT 75

Query 65 DLNYGGVKAALAGSYLEVEGPEKGTTVYVTDLYPEGARGALDLSPNATFRKIGNMKDGGKIN 124
      D YG A G+Y+ V GP+G TV + D P G +DLSF AF +I ++ G +
Sbjct 76 D--YG--LADYCGAYVTVNGPRGVTVEIIRACPGCVVGGIDLSPQAFERIAALEAGNVP 131

Query 125 IKRWVVKAP-ITGNFTYVIRKEGSSRWAAIQVRNHKYPVMMEY-EKDGEWINMEKMDYN 182
      I W++ AP ++GN Y KEGS+HWA +QVRNH+ + K EY G + N+++ YN
Sbjct 132 ITWGLTAPNVSGVLYVYKESGQWAGVQVRNHNATATFTYRNPGQITQWYRVQVW 191

Query 193 RF-VSTNLGTSLKVRMTDIRKVVKDTIPKLPSGTSKAYTVPGRVQFP 231
      +F V +GGG R+TD+ G V+ D TDP E +V G+ QFP
Sbjct 192 YFLVPGMGTPFTFRITDVVGNVETDNTIPLASEG-----VAGNQQFP 236

>>ref|YP_621762.1| putative lipoprotein [Myxococcus xanthus DK 1622]
gb|ABF91411.1| putative lipoprotein [Myxococcus xanthus DK 1622]
Length=238

Score = 120 bits (300), Expect = 1e-28, Method: Compositional matrix adjust.
Identities = 68/199 (34%), Positives = 108/199 (54%), Gaps = 14/199 (7%)

Query 29 ABAAYDDLHEGYATY---TGSGYSGGAFLLDPIPSDMEITAINPADLNYGGVKAALAGSY 79
      A + +G AT+ TGSG +P D+ + A+N +A G
Sbjct 24 GGLPLGEGQGLATFYNATGSGNCH---VEPT-GDLAVAMNTPQY----ANSAACQC 84

Query 80 LEVEGPEKGTTVYVTDLYPEGARGALDLSPNATFRKIGNMKDGGKINIKRWVVKAPITGNFT 139
      ++ GREG V + D PE G LLS AF +I M+ G+NI W V + GN
Sbjct 85 VDITGPKGSVRVRIYDNCPECEGHLDSREAFARIAEQGQGVNITWTFVSCDVAGNIE 144

Query 140 YRIKEGSSRWAAIQVRNHKYPVMMEYEED-GEWINMEKMDYNHFV-STNLGTSLKVR 197
      Y K GS+ HW AIQVRNH+ P+ K+E+ + G W ++ + YN+EV+ + +G G VR
Sbjct 145 YRIKNGSNPWTATQVRNHLPICQLEWRRTGGWQDVPRSYNYFVNLSGMGSGPFSVR 204

Query 198 RMTDIRKVVKDTIPKLPS 216
      +T + G+ +DT+ ++ ++
Sbjct 205 VTAVDGQQLDRLRVIDR 223
```

도면14i

```
>ref|YP_001793396.1| Rare lipoprotein A [Leptothrix cholodnii SP-6]
gb|ACB36631.1| Rare lipoprotein A [Leptothrix cholodnii SP-6]
Length=254

Score = 130 bits (300), Expect = 1e-28, Method: Compositional matrix adjust.
Identities = 78/203 (38%), Positives = 111/203 (54%), Gaps = 14/203 (6%)

Query 31 HEGYATTTGSGYSGGAFLLDDIPSDMEITAINPADLNYGCVKAAAGSYLVEGPKKTT 90
+G TY + GA D +D + A+N D Y G +A G + + V GP G T
Sbjct 62 RQCECTYYAA-TGACACSYDA-SADRMVAAMNHTD--YAG--SAACGEHVUTGDLCTVT 115

Query 91 VVVTDLVPEGARGALDLSDNAFRIKIGNMKDKGKINIKWVUVKADITCNFTYRIKESSRWW 150
V + D DE A G +DLS AF +I G+ I WVV + +G YR KES+R+W
Sbjct 116 VRIVDECEPCAGCQVLLSAEAFARIAEPVAGVITWGVTCGAVSGPVQYVYKESSTYVW 175

Query 151 AATQVRNHKYPVMKMEYEK--GKWINMEKMDYNHV--STNLCTGSLKVRMTDIRGKVVK 207
ATQVRNH P+ +E + D G WI + + DYN+V +G G+L+VR+T G +
Sbjct 176 TATQVRNHVPDIARLEIQDDCTGNIAVPRTDYNIVVHPVAIGACALQVRVTSSTGVALI 235

Query 208 DTIPKLDESSTSKAYTVGCHVQF 230
D +D+ P+ G V G QF
Sbjct 236 DRLPQ-PQG-----LVDGCTGQF 253

>ref|YP_001509817.1| Rare lipoprotein A [Frankia sp. EAN1pec]
gb|ABW14911.1| Rare lipoprotein A [Frankia sp. EAN1pec]
Length=293

Score = 130 bits (299), Expect = 2e-28, Method: Compositional matrix adjust.
Identities = 73/189 (38%), Positives = 109/189 (57%), Gaps = 7/189 (3%)

Query 30 LHEGYATTTGSGYSGGAFLLDDIP-SDMEITAINPADLNYGCVKAAAGSYLVEGPKGK 88
+ G AT+ G+ GG + D + M + A+N +L+Y +A G+Y+V GP G
Sbjct 99 VEGCPATHYCAD-GGNCMFDRLLTDFAMFVVAHN--ELDYETARAC--GAYIEVTDGGGT 153

Query 89 TTVVVTDLVPEGARGALDLSDNAFRIKIGNMKDKGKINIKWVUVKADITCNFTYRIKESSR 148
T V VTD DE G LDIS AF +I G + + + WR+V G + +R+KESG
Sbjct 154 TVVGVTRDCDECGDGLDLDSQAFAIACGVEGLVDVTWRLVSPADIGAVQFVWKEGSSA 213

Query 149 WWAATQVRNHKYPVMKMEYEKDGKWINMEKMDYNHVST-NLCTGSLKVRMTDIRGKVVK 207
+W ATQVRNH+ DV+ +E +G W + + +N+V+ LG G VR+TD+ G+ +
Sbjct 214 YWLAATQVRNHRNPVSLVRVNGAWTALOREWNNYFVADPQCLGPGGPTVRITDVYGEQLV 273

Query 208 DTIPKLDES 216
+T+ D S
Sbjct 274 ETVNLAPAS 282

>ref|XP_01910738.1| putative lipoprotein [Plesiocystis pacifica SIR-1]
gb|EM74334.1| putative lipoprotein [Plesiocystis pacifica SIR-1]
Length=298

Score = 113 bits (260), Expect = 1e-23, Method: Compositional matrix adjust.
Identities = 63/170 (37%), Positives = 86/170 (50%), Gaps = 16/170 (9%)

Query 71 VKAAAGSYLVEGPKKCTTVVVTDLVPEGAR-----GALDLSDNAFRIKIGNMKDKGKINI 125
AA+ G+ V GPKG + + D D G +DL+P AF +IG + DG I I
Sbjct 133 ADAAAGCAGCARVTGPKGSVDIGIVDQCIATNDPICYEGHIDLMPATEGICEIDGIDPI 192

Query 126 KRWVUVKADITCNFTYRIKESSRWWAATQVRNHKYPVMKMEYEK--DGKWINMEKMDYNHV 194
W + + G Y KESR+WW + VRNH P+ K EY+ G + + + YH F
Sbjct 193 TWELISDDDDCTIAYSFKESGSGWTCVLVRNHNDIAKFEYDTGGNFKAVDPEGYNFF 252

Query 185 VSTN-LCTGSLKVRMTDIRGKVVDK-ID-KLDESSTSKAYTVGCHVQF 231
V + GTC R+TDI G V+ D ID +L + V G QF
Sbjct 253 VDDGPTGPTFRVTDIRGHVITVDIDPLQLGDP-----VVGSGQF 295

>ref|YP_001543916.1| Rare lipoprotein A [Haerpetosiphon aurantiacus ATCC 23779]
gb|AK03798.1| Rare lipoprotein A [Haerpetosiphon aurantiacus ATCC 23779]
Length=273

Score = 108 bits (248), Expect = 4e-22, Method: Compositional matrix adjust.
Identities = 64/183 (34%), Positives = 99/183 (54%), Gaps = 11/183 (6%)

Query 30 LHEGYATV-TGSGYSGGAFLLDDIPSDMEITAINPADLNYGCVKAAAGSYLVEGPKK 88
+G ATY T +G + D+SD+ A+N D A G+ + + + GP G
Sbjct 44 IRGCGATYYTATGLGNCI-----DVPSDLLFGAMNNDY----ATADYCGAFVKTGPLGS 96

Query 89 TTVVVTDLVPEGARGALDLSDNAFRIKIGNMKDKGKINIKWVUVKAD-ITCNFTYRIKESG 147
TV +TD DE G +DLS AF +I N G + I W+ + + D TG +Y R+GSS
Sbjct 97 VTQVTDGDCGCGCHLDLSQAFAIANDVTCIVDITWQLISNDATTEKVSVEPKGSS 156

Query 148 RWAATQVRNHKYPVMKMEYEK--DGKWINMEKMDYNHV--STNLCTGSLKVRMTDIRGKV 205
+WW A+Q RNH+ + K EY + + +N+E+ + +GTC R TD+ G +
Sbjct 157 QWFTAVQVRNHRNAIAKFEYRMCQAAYKVAPEFIFNYIAEAGMTCGPYSRTTDVYGNM 216

Query 206 VKD 208
+ D
Sbjct 217 IVD 219

>ref|YP_001705759.1| cellulase CelA [Clavibacter michiganensis subsp. sepedonicus]
emb|CAQ03244.1| cellulase CelA [Clavibacter michiganensis subsp. sepedonicus]
Length=684

Score = 108 bits (248), Expect = 5e-22, Method: Compositional matrix adjust.
Identities = 72/208 (34%), Positives = 105/208 (50%), Gaps = 13/208 (6%)

Query 20 SPQASAAAYDDLEGYATV-----TGSGYSGGAFLLDDIPSDMEITAINPADLNYGCVKAAAL 75
QA+ G AT+ TG+ + G + +D+D A++ + YGG AA
Sbjct 474 DSGATPDSQATTHGRATHYSLCTGNTIANGCNCMPAUVADRMVAVSSPE--YGC--AAA 529

Query 76 AGSYLVEGPKKCTTVVVTDLVPEGARGALDLSDNAFRIKIGNMKDKGKINIKWVUVKADIT 135
GS+L V GPKG V + D E G LDIS AFR IG+ G I I + V+ D
Sbjct 530 CDSHLLATGPKCTVAVQIVDQCECEGCHLDLSRAFAICGDFDAGIDIDYTTVDNDPSV 589

Query 136 GNTTYRIKESSRWWAATQVRNHKYPVMKMEYEKDGKWINMEKMDYNHVSTN-LCTGSL 194
R+KESSRWWA +Q+ N + +E G+W+ + + DY +V+ + + G +
Sbjct 590 DAVAVRVKESSRWWAGLQILNAGNRIDRVEVRAGGQWLPDSRTDYCYWVTPSDPIDSPH 649

Query 195 KVRMTDIRGK-VVKDTIPKLDESSTSKA 221
VR+TD G+ VV I R S A
Sbjct 650 TVVVTQYGRVVLDRMAGCTGSA 677
```

도면14j

```
>ref|XP_01464869.1| cellulase [Stigmatella aurantiaca D64/3-1]
gb|EAG64870.1| cellulase [Stigmatella aurantiaca D64/3-1]
Length=248

Score = 101 bits (231), Expect = 6e-20, Method: Compositional matrix adjust.
Identities = 80/139 (58%), Positives = 79/139 (56%), Gaps = 2/139 (2%)

Query 73 AALAGSYLVESGPRGRTTVVTDLYPEGARGALDLSNPAFKIGNMKDKGINKRWVKA 122
      AA+ G+ EV GP GK V D+ A + LS A+ K+ + G++ + W++V
Sbjct 91 AAVCGACAEVSGPGRKVVARVVDICDSCAPDQMLLSQAYSKVAS--GGGVKMSRLVAC 148

Query 123 PITGNTFYRDKGSSRW-WAALQVRNKKYPMKMEYERKDKRWINMERMDYNEFVSTNLT 191
      + G Y TR+GS + + AIGVRNKK D+ +E + W + + +YNT+V 126
Sbjct 149 SVGGPIRDKKDGSGPYFAIGVRNKKYPIKSLVVMGSSNMTLTERKRYFVGNALGT 208

Query 192 GSLKVRMTDIRKVVKDTI 210
      I+VR+T G++ +DT+
Sbjct 209 PPLQVRVTGTEGVLEDTL 227

>ref|YP_434884.1| Endoglucanase C-terminal domain/subunit and related protein [Bacillus
chejuensis KCTC 2296]
gb|ABCS0409.1| Endoglucanase C-terminal domain/subunit and related protein [Bacillus
chejuensis KCTC 2296]
Length=248

Score = 100 bits (229), Expect = 1e-19, Method: Compositional matrix adjust.
Identities = 76/216 (35%), Positives = 105/216 (48%), Gaps = 22/216 (10%)

Query 22 QAAAYVDLREGVATYTGSGVSGGAFLLDPIPSDEITAINPADLNYGVAALAGSYLVESGPR 81
      SAA LREG TY Y+GG P+P+ M A+N D Y G +A G ++
Sbjct 31 RTSA--LREGSTYYE--YNGSGHCVVPVA-METAAHNTD--YNGSGAC--GGCVK 80

Query 62 VEGPK-GRFTVY-VTDLYPEGARGALDLSNPAFKIGNMKDKGINKRWVKA PITGNT 139
      V GK+ V V D P G +DL+ AF +I ++ G+I I W V
Sbjct 81 VEMHKKSVVARVDSGSGHSGSDVLTDAAFAGISPLEAGRIPIWDVPCDVPVLL 140

Query 140 YLRGSSRWAAALQVRNKKYPMKMEY--EKDGKWINMERMDYNEFV-STNLTGSGPR 195
      Y EGSS+W A+QVR +YFY + Y G + + DRYFY + +GG
Sbjct 141 Y-FMGSSGQWFAVQVRQRYPVSSLAYRESGSTRSGYQETAREDYNYFVERSGAGTGPFD 199

Query 196 VRMTDIRKVVKDTIPKLESOTSKATVPGRVQFP 231
      R+TDI G V++ L +SG + QFP
Sbjct 200 FRTDIRGVHLGAGNTL-QSGVP---INTQVQFP 230

>ref|YP_001222221.1| putative expansin [Clavibacter michiganensis subsp. michiganensis
NCPPB 352]
emb|CAN01526.1| putative expansin [Clavibacter michiganensis subsp. michiganensis
NCPPB 352]
Length=258

Score = 97.8 bits (223), Expect = 7e-19, Method: Compositional matrix adjust.
Identities = 64/177 (36%), Positives = 87/177 (49%), Gaps = 9/177 (5%)

Query 33 GYATYTGSGVSG---GAFLLDPIPSDEITAINPADLNYGVAALAGSYLVESGPR 88
      GYAT+ G G G L IP D A+ P DL GG A GSY +V GP G
Sbjct 59 GYTHYILGPGRTTHSGHSLPILKDLVAVGP-DLYAG--ACGSGTDTGPHGT 114

Query 89 TVVVTDLYPEGARGALDLSNPAFKIGNMKDKGINKRWVKA PITGNTFYRDKGSSRW 148
      V V D E G LGS ATR LG+ G I + V A ++R K+GS
Sbjct 115 VREVAVDSHCCHSLDLSEAFRAIGDYDAGITPSYVVAASFVPLSFNKDSSA 174

Query 149 WAAALQVRNKKYPMKMEYERKDKWINMERMDYNEFVSTN-LGTGSLKVRMTDIRK 204
      +WAA+QV + + +E +W+ + DY +++ +G G VR+TD G+
Sbjct 175 WAAALQVLGAVRLRVELAVGARGVFLSLDIXYVLA PGYVAGSPFTRVTDITGR 231

>ref|YP_971728.1| RaxA lipoprotein A [Acidovorax avenae subsp. citrulli AAC00-1]
gb|AAM2954.1| RaxA lipoprotein A [Acidovorax avenae subsp. citrulli AAC00-1]
Length=351

Score = 92.3 bits (210), Expect = 4e-17, Method: Compositional matrix adjust.
Identities = 62/182 (34%), Positives = 95/182 (52%), Gaps = 15/182 (8%)

Query 36 TVTSGS---YSGAFLDPIPSDEITAINPADLNYGVAALAGSYLVESGPR--KGR 89
      TTTS G Y GG P D P + A+N D +X G +A G+YLEV P R
Sbjct 51 TVTSGSTFYGVDSGNCLEPESVLEVAIN--DSYVSGAC--GAYLEVLAATSKCV 106

Query 90 TVVVTDLYPEGARGALDLSNPAFKIGNMKDKGINKRWVKA PITGNTFYRDKGSSRW 149
      V V + P+ LGL+ A +I + G +++WR V P T + REGSS
Sbjct 107 VVRVDRKCPDCDPHGLDLAIPALAGIAPIDAGIVLSRWRYVSGPDTFA-SVVERGSSAS 165

Query 150 WAAALQVRNKKYPMKMEY--EKDGKWINMERMDYNEFVSTN-LGTGSLKVRMTDIRK 208
      W+A+QVRN + V + Y G ++ +E+ YN+R++ +G G +++TD+ G+V
Sbjct 166 WHALQVRQNAVATLAVRSGSGSTVPLSGQVNTLAPGSGSPEDTAKTNTGQV 225

Query 206 VR 207
      ++
Sbjct 226 LE 227

>ref|YP_001618321.1| hypothetical protein sce7672 [Sorangium cellulosum 'So ce 56']
emb|CAN97041.1| hypothetical protein [Sorangium cellulosum 'So ce 56']
Length=346

Score = 86.7 bits (197), Expect = 1e-15, Method: Compositional matrix adjust.
Identities = 57/175 (32%), Positives = 82/175 (46%), Gaps = 24/175 (13%)

Query 51 PIPSDMEITAINPADLNYGVAALAGSYLVESGPRGRTTVVTDLYPEGAR----- 102
      P A+N AD N AA+ G+ +EV +G K K V D P
Sbjct 156 PFSSGQYAMPTADYN-----AAMCGACVEVSDGGR-KVGLAVDQCPLATNPCKA 210

Query 103 GALLDSNPAFKIGNMKDKG-----NKRWVKA PITGNTFYRDKGSSRWAAIT 153
      G +DLG NAF +IG +G + I WR + P E +R+K+ S++W I
Sbjct 211 GHLDSNNAFLGIGEREGYLGTTNGGAAGKIRYIPCTEAVSERLKDANKYVNEI 270

Query 154 QVRNKKYPMKMEYERKDKWINMERMDYNEFV-STNLTGSLKVRMTDIRKVV 206
      V R +P+ K+E E +G W + YN+R N+G VR TDI G V
Sbjct 271 LVGSHAPTEKLEVEINOTGATATGQYNYFTVGDGNNAPTRVRATDINGSTV 325
```

도면14k

```
>ref|YP_001615903.1| hypothetical protein ace5260 [Sorangium cellulosum 'So ce 56']
emb|CAN95423.1| hypothetical protein [Sorangium cellulosum 'So ce 56']
Length=438

Score = 85.9 bits (195), Expect = 2e-15, Method: Compositional matrix adjust.
Identities = 59/185 (31%), Positives = 91/185 (49%), Gaps = 22/185 (11%)

Query 58 ITAINPADINYGKKAALAGSYLEV---EGPKGKTTV---YVTDLYPEGARGALDLSFNA 111
      A+N AD N AA G+ +EV +G K TV T P+ G +DLS A
Sbjct 256 EALNTADYN---TAAACGACVEVSRRDGRKVATVVDQCPTASNPKCKAGHIDLSKEA 311

Query 112 FRKIGNMKDG-----KINIKWRVVKAPITGNFTYRIKEGSSRWAAIQVRNHKYPV 162
      F +IG + +G + I W+ V P+ N + +KE + WW + +V+ H+ PV
Sbjct 312 EAQIGALDEGYLGTGNGGVRGKISWKYVPCFVEENVSVLKEPGNAWNLRLVQGHRTFV 371

Query 163 MKMEYEKGKWINMEKMDYNHF--VSTNLGTGSLKVRMTDIRGKVVKDTI-PKLPESGTS 219
      K+E G+W + YN++ S N+G G +VR+TD+ GK ++ +I E TS
Sbjct 372 RKLEVYVGGRWTEATRQFPYNYWRVSGNGMGQFWRVRVTDVLGKTETSINTTTAEQLTS 431

Query 220 KAYTV 224
      + V
Sbjct 432 SQFPV 436

>ref|EP_01516438.1| YoaJ [Chloroflexus aggregans DSM 9485]
gb|EAV0929.1| YoaJ [Chloroflexus aggregans DSM 9485]
Length=104

Score = 69.4 bits (156), Expect = 2e-10, Method: Compositional matrix adjust.
Identities = 34/98 (34%), Positives = 53/98 (54%), Gaps = 7/98 (7%)

Query 141 RIKEGSSRWAAIQVRNHKYPVMEYEK-DGKWINMEKMDYNHFVSTNLGTGSLK---- 195
      +K+GS+ WW A QVR+H+ P+ K+EY G+W+ + + YN+F+ G +
Sbjct 4 HLKDGSNPWLAFQVRHHRNPLAKLEYRTEQGQWVQLNRTFYNFYIRECQCAGGERGPFV 63

Query 196 VRMTDIRGKVVKDTIP--KLPESGTSKATTVPGHVQFP 231
      +R+TDI G + +T+ L G S V G QFP
Sbjct 64 LRVTDIYGNLTETVQFTLLSSGGLSPGELVSGSQFP 101

>ref|YP_001220654.1| cellulase [Clavibacter michiganensis subsp. michiganensis NCPPB
382]
emb|CAM98473.1| cellulase [Clavibacter michiganensis subsp. michiganensis NCPPB
382]
Length=746

Score = 59.6 bits (133), Expect = 2e-07, Method: Composition-based stats.
Identities = 41/106 (38%), Positives = 50/106 (47%), Gaps = 38/106 (35%)

Query 68 YGGVKAALAGSYLEVEGPKGKTTVVYVTDLYPEGAR-----GALDLSFNAFRKI 115
      Y G AA G++L+V GPKG TV R G LDLS AFR +
Sbjct 586 YSG--AAACGTFLDVTGPKG--TV-----RVQVADQCHGCEVGHLDLSEAFRAL 631

Query 116 GNMKGKINIKWRVVKAP-----ITGNFTYRIKEGSSRWAAIQV 155
      G+ G I I + V+ P I R+KEGSSRWAA +QV
Sbjct 632 GDFNAGIIPISYVTVRDPAGETVAI-----RVKEGSSRWAGLQV 671

>emb|CAA44467.2| Cellulase [Clavibacter michiganensis]
Length=745

Score = 59.6 bits (133), Expect = 2e-07, Method: Composition-based stats.
Identities = 41/106 (38%), Positives = 50/106 (47%), Gaps = 38/106 (35%)

Query 68 YGGVKAALAGSYLEVEGPKGKTTVVYVTDLYPEGAR-----GALDLSFNAFRKI 115
      Y G AA G++L+V GPKG TV R G LDLS AFR +
Sbjct 586 YSG--AAACGTFLDVTGPKG--TV-----RVQVADQCHGCEVGHLDLSEAFRAL 631

Query 116 GNMKGKINIKWRVVKAP-----ITGNFTYRIKEGSSRWAAIQV 155
      G+ G I I + V+ P I R+KEGSSRWAA +QV
Sbjct 632 GDFNAGIIPISYVTVRDPAGETVAI-----RVKEGSSRWAGLQV 671

>gb|AAK16222.1| cellulase CelA [Clavibacter michiganensis subsp. sepedonicus]
Length=727

Score = 59.2 bits (132), Expect = 3e-07, Method: Composition-based stats.
Identities = 44/103 (42%), Positives = 50/103 (48%), Gaps = 32/103 (31%)

Query 68 YGGVKAALAGSYLEVEGPKGKTTVVY--TDLYPEGARGALDLSFNAFRKIGNMKDGKINI 125
      YGG AA GS+L V GPKG TV V D E G LDLS AFR IG+ G I
Sbjct 567 YGG--AAACGSHLLVTGPKG--TVRVQIVDQCHECEIGHLDLSEAFRAIGDFDAGII-- 620

Query 126 KWRVVKAPITGNFTY-----RIKEGSSRWAAIQV 155
      PI+ Y R+KEGSSRWAA +Q+
Sbjct 621 -----PIS-----YTVRDPSPFAVAVRVKEGSSRWAGLQI 652
```


도면141

```
>ref|YP_001615903.1| hypothetical protein sce5260 [Sorangium cellulosum 'So ce 56']
emb|CAN95423.1| hypothetical protein [Sorangium cellulosum 'So ce 56']
Length=438

Score = 85.9 bits (195), Expect = 2e-15, Method: Compositional matrix adjust.
Identities = 59/185 (31%), Positives = 91/185 (49%), Gaps = 22/185 (11%)

Query 58 ITAINPADLNYGGVKAALAGSYLEV---EGPKGKTTV---YVTDLYPEGARGALDLS PNA 111
A+N AD N AA G+ +EV +G K TV T P+ G +DLS A
Sbjct 256 FAALTADYN---TAACGACVEVSRDDGRKVATVVDQCPATASNPKCKAGHIDLSKEA 311

Query 112 FRKIGNMKDG-----KINIKWRVVKAPITGNFTYRIKEGSSRWAAIQVRNHKYPV 162
F +IG + +G + I W+ V P+ N + +KE + WW ++V+ H+ PV
Sbjct 312 FAQIGALDEGYLGTGGGGVRGKISWYVPCPVEENVSVLKEPGNAWNNELRVQGHRTFV 371

Query 163 MMEYEKDGKWINMEKMDYNHF--VSTNLGTGSLKVRMTDIRGKVVKDTI--EKLPESGTS 219
K+E G+W + YN++ S N+G G +VR+TD+ GK ++ +I E TS
Sbjct 372 RKLEVVYGGRTATRQPPYNYWRVGSNGMGQPRVRVTDVLGKTIETSIINTTAEQLTS 431

Query 220 KATTV 224
+ V
Sbjct 432 SQPFV 436

>ref|EP_01516438.1| YoaJ [Chloroflexus aggregans DSM 9485]
gb|EAV09929.1| YoaJ [Chloroflexus aggregans DSM 9485]
Length=104

Score = 69.4 bits (156), Expect = 2e-10, Method: Compositional matrix adjust.
Identities = 34/98 (34%), Positives = 53/98 (54%), Gaps = 7/98 (7%)

Query 141 RIKEGSSRWAAIQVRNHKYPVMKMEYEK-DGKWINMEKMDYNHFVSTNLGTGSLK---- 195
+K+GS+ WW A QVR+H+ P+ K+EY G+W+ + + YN+F+ G +
Sbjct 4 HLKDGSNFWWLAQVRHHRNPIAKLEYRTFQGGWVQLNRTFNYFIRECQAGGERGPFV 63

Query 196 VRTDIRGKVVKDTIF--KLPESGTSKATVPGHVQFP 231
+R+TDI G + +t+ L G S V G QFP
Sbjct 64 LRVTDIYGNLTETVQFTLLSSGGLSPGELVSGSQFP 101

>ref|YP_001220664.1| cellulase [Clavibacter michiganensis subsp. michiganensis NCPFB
382]
emb|CAM98473.1| cellulase [Clavibacter michiganensis subsp. michiganensis NCPFB
382]
Length=746

Score = 59.6 bits (133), Expect = 2e-07, Method: Composition-based stats.
Identities = 41/106 (38%), Positives = 50/106 (47%), Gaps = 38/106 (35%)

Query 68 YGGVKAALAGSYLEVGGPKGKTTVYVTDLYPEGAR-----GALDLSPNAFRKI 115
Y G AA G+LAV GPKG TV R G LDLS AFR +
Sbjct 586 YSG--AAACGTFLDVTGPKG--TV-----RVQVADQCHGCEVGHLDLSEAFRAL 631

Query 116 GNMKGKINIKWRVVKAP-----ITGNFTYRIKEGSSRWAAIQV 155
G+ G I I + V+ P I R+KEGSSRWAA +QV
Sbjct 632 GDNAGIIPISYVTVRDPAGFTVAI-----RVKEGSSRWAGLQV 671

>emb|CAA44467.2| Cellulase [Clavibacter michiganensis]
Length=745

Score = 59.6 bits (133), Expect = 2e-07, Method: Composition-based stats.
Identities = 41/106 (38%), Positives = 50/106 (47%), Gaps = 38/106 (35%)

Query 68 YGGVKAALAGSYLEVGGPKGKTTVYVTDLYPEGAR-----GALDLSPNAFRKI 115
Y G AA G+LAV GPKG TV R G LDLS AFR +
Sbjct 586 YSG--AAACGTFLDVTGPKG--TV-----RVQVADQCHGCEVGHLDLSEAFRAL 631

Query 116 GNMKGKINIKWRVVKAP-----ITGNFTYRIKEGSSRWAAIQV 155
G+ G I I + V+ P I R+KEGSSRWAA +QV
Sbjct 632 GDNAGIIPISYVTVRDPAGFTVAI-----RVKEGSSRWAGLQV 671

>gb|AAK16222.1| cellulase CelA [Clavibacter michiganensis subsp. sepedonicus]
Length=727

Score = 59.2 bits (132), Expect = 3e-07, Method: Composition-based stats.
Identities = 44/103 (42%), Positives = 50/103 (48%), Gaps = 32/103 (31%)

Query 68 YGGVKAALAGSYLEVGGPKGKTTVV---TDLYPEGARGALDLSPNAFRKIGNMKDGKINI 125
YGG AA GS+L V GPKG TV V D E G LDLS AFR IG+ G I
Sbjct 567 YGG--AAACGSHLIVGPKG--TVRVQIVDQCECEIGHLDLSEAFRAIGDFDAGII-- 620

Query 126 KWRVVKAPITGNFTY-----RIKEGSSRWAAIQV 155
PI+ Y R+KEGSSRWAA +Q+
Sbjct 621 -----PIS-----YITVRDPSVFAVAVRVKEGSSRWAGLQI 652
```


도면14m

```
>xf|YE_001801897.1| hypothetical protein ccs_0480 [Cyanobacterium sp. ATCC 5142]
gb|ACB48831.1| hypothetical protein ccs_0480 [Cyanobacterium sp. ATCC 5142]
Length=108

Score = 27.1 bits (80), Expect = 1.2, Method: Compositional matrix adjust.
Identities = 32/126 (26%), Positives = 59/126 (46%), Gaps = 22/126 (18%)

Query 1 MKKIMATVGMVLLTFECFSGQASAAVDDLHEGVATYTGSGVSGGAFLLDPIPSDMEITA 60
      MKK+ S + M+ L Q + Y ++G T +G ++ A +
Sbjct 1 MKKV-SLSILMIALAAMPFAQAQTATYVASHYQGQKTASGVRENNSAMM-----A 48

Query 61 INFADINYGCVKAAALAGSYLEVEGPKGKTTVVVTDLYPEGARGALDLSFNATRKIGNMKD 120
      +E+ L +G + +V R + +V V + + R +LDLS +AFR IG+R G+I +K +V+
Sbjct 49 AHPD-LFTG-----TKVVTNERTGKSVIVRIV--DRCRSLDLQGAFTTGSALGRIIVL 97

Query 121 GRINIK 126
      G+I ++
Sbjct 98 GRIPVR 103

>xf|XE_00518884.1| Raxa lipoprotein A [Coccomphasma watsonii WH 8501]
gb|EAM51921.1| Raxa lipoprotein A [Coccomphasma watsonii WH 8501]
Length=108

Score = 36.7 bits (79), Expect = 1.8, Method: Compositional matrix adjust.
Identities = 20/55 (36%), Positives = 34/55 (61%), Gaps = 2/55 (3%)

Query 77 GSYLEVEGPKGKTTVVVTDLYPEGARGALDLSFNATRKIGNMKDGRINIKRVRVVK 131
      G+ +V R + +V V + + R +LDLS +AFR IG+R G+I +K +V+
Sbjct 56 GPKVVTNERTGKSVIVRIV--DRCRSLDLQGAFTTGSALGRIIVL 108

>xf|XE_01729178.1| raxa lipoprotein A [Cyanobacterium sp. CCY0110]
gb|EAM591381.1| raxa lipoprotein A [Cyanobacterium sp. CCY0110]
Length=99

Score = 35.0 bits (75), Expect = 5.1, Method: Compositional matrix adjust.
Identities = 29/116 (25%), Positives = 52/116 (45%), Gaps = 22/116 (19%)

Query 11 MVLLTFECFSGQASAAVDDLHEGVATYTGSGVSGGAFLLDPIPSDMEITAINADNADLYGG 70
      M+ L Q + Y ++G T +G ++ A + A +E+ L +G
Sbjct 1 MIALAAMPFAQAQTATYVASHYQGQKTASGVRENNSAMM-----AHPD-LFTG- 47

Query 71 VKAALAGSYLEVEGPKGKTTVVVTDLYPEGARGALDLSFNATRKIGNMKDGRINIK 126
      + +V R + +V V + + R +LDLS +AFR IG+R G+I +K +V+
Sbjct 48 -----TKVVTNERTGKSVIVRIV--DRCRSLDLQGAFTTGSALGRIIVL 94

>xf|XE_02976533.1| raxa lipoprotein A [Cyanobacterium sp. FCC 7424]
gb|EDU13976.1| raxa lipoprotein A [Cyanobacterium sp. FCC 7424]
Length=99

Score = 34.1 bits (73), Expect = 9.8, Method: Compositional matrix adjust.
Identities = 30/99 (30%), Positives = 49/99 (49%), Gaps = 15/99 (15%)

Query 35 ATYTCGSGVSGGAFLLDPIPSDMEITAINADNADLYGG-VKAAALAGSYLEVEGPKGKTTV- 92
      AT+ SGV G S + A +E+ L +G VR + GK+ V
Sbjct 14 ATYTCGSGVSGGAFLLDPIPSDMEITAINADNADLYGG-VKAAALAGSYLEVEGPKGKTTV- 64

Query 93 VTDLYPEGARGALDLSFNATRKIGNMKDGRINIKRVRVVK 131
      + D R +LDLS +AFR IG+R G+I +K +V+
Sbjct 65 VYDC-----RCRSLDLQGAFTTGSALGRIIVL 99

>xf|YE_001267828.1| methyltransferase small [Pseudomonas putida F1]
gb|ADQ75644.1| methyltransferase small [Pseudomonas putida F1]
Length=217

Score = 32.2 bits (71), Expect = 16, Method: Compositional matrix adjust.
Identities = 28/97 (29%), Positives = 42/97 (43%), Gaps = 4/97 (4%)

Query 34 VATTGSGVSGGAFLLDPIPSDMEITA--INADNADLYGGVKAALAG-SYLEVEGPKGKTT 90
      +A G G GA L+ ++A INE L Y + AALAG S L VE
Sbjct 140 HAVDIGCGSGVGGALLIARAAGHAQVAVDINFLALRYTAINAALAGVSNLSVEPDLDDG 199

Query 91 VYVT-DLYPEGARGALDLSFNATRKIGNMKDGRINIK 126
      + T DL LD + +R G + +++
Sbjct 200 ISGTFILVANEFFVMDNARYRSGGVIGADLIR 236

>xf|XE_01084595.1| magnesium transport P-type atpase [Synecococcus sp. WH 5701]
gb|EAQ75926.1| magnesium transport P-type atpase [Synecococcus sp. WH 5701]
Length=401

Score = 32.9 bits (70), Expect = 25, Method: Composition-based stats.
Identities = 10/11 (90%), Positives = 10/11 (90%), Gaps = 0/11 (0%)

Query 98 PEGARGALDLS 108
      PEGARG LDLS
Sbjct 391 PEGARGELDLS 401

>xf|NR_465585.1| hypothetical protein lmo2061 [Listeria monocytogenes ECD-9]
xf|XE_01940935.1| conserved hypothetical protein [Listeria monocytogenes FSL N3-165]
xf|XE_01940935.1| conserved hypothetical protein [Listeria monocytogenes 104033]
emb|CAD00129.1| lmo2061 [Listeria monocytogenes]
gb|EAD14321.1| conserved hypothetical protein [Listeria monocytogenes 104033]
gb|EAD14321.1| conserved hypothetical protein [Listeria monocytogenes FSL N3-165]
Length=205

Score = 32.9 bits (70), Expect = 25, Method: Compositional matrix adjust.
Identities = 19/32 (59%), Positives = 22/32 (69%), Gaps = 7/32 (22%)

Query 1 MKKIMATVGMVLLTFECFSGQASAAVDDLH-----EGVATYT 38
      MKKIM+ V VLL +E QASA L + + TTT
Sbjct 1 MKKIMSVIV-VLLAVTETFGASAHVAVLPSSTVDSWETTT 41
```

도면14n

```
>=cF1NF_746295.11 modification methylase, HsaK Family [Pseudomonas putida ET2440]
gb|AA060899.1|AE016510_5 modification methylase, HsaK Family [Pseudomonas putida ET2440]
Length=236

Score = 32.9 bits (70), Expect = 25, Method: Compositional matrix adjust.
Identities = 21/52 (40%), Positives = 27/52 (52%), Gaps = 5/52 (9%)

Query 24 YATTTGGGYSGGAFLLDFIESDMEITA--INFADLNYGGVKAALAG--YLEVE 03
+ A G G GA L+ ++A INF L Y + AALAG S L VE
Sbjct 135 HAVDIGCGSGVGCALLIARAAGHAQVSADVINFALRYTAINAALAGVSNLSVE 191

>=cF1VF_212314.11 putative lipoprotein [Bacteroides fragilis NCTC 9342]
emb|CA000896.1| putative lipoprotein [Bacteroides fragilis NCTC 9342]
Length=626

Score = 32.5 bits (69), Expect = 31, Method: Compositional matrix adjust.
Identities = 20/67 (29%), Positives = 32/67 (47%), Gaps = 7/67 (10%)

Query 165 MEYERDGRW----INME--KMDYNHVS-TNLGTGSLVVRVTDIGRVVVDITPELDESG 217
+EVYKDGKW +N E + Y + + T L+R TD R + + I
Sbjct 427 VEYRDKRWLPLVKINEETTVGYKRLRFEFVSTDKLRIRTDARACLCINNIEAYAGE 486

Query 218 TSKAYTV 224
T+ +TV
Sbjct 487 TADFTTV 492

>=cF1VF_100649.11 hypothetical protein BF2271 [Bacteroides fragilis YCH46]
gb|BAD00115.1| conserved hypothetical protein [Bacteroides fragilis YCH46]
Length=626

Score = 32.5 bits (69), Expect = 32, Method: Compositional matrix adjust.
Identities = 20/67 (29%), Positives = 32/67 (47%), Gaps = 7/67 (10%)

Query 165 MEYERDGRW----INME--KMDYNHVS-TNLGTGSLVVRVTDIGRVVVDITPELDESG 217
+EVYKDGKW +N E + Y + + T L+R TD R + + I
Sbjct 427 VEYRDKRWLPLVKINEETTVGYKRLRFEFVSTDKLRIRTDARACLCINNIEAYAGE 486

Query 218 TSKAYTV 224
T+ +TV
Sbjct 487 TADFTTV 492

>=cF1ZF_02942337.11 Rare lipoprotein A [Cyanobacteria sp. PCC 8801]
gb|EDT59682.1| Rare lipoprotein A [Cyanobacteria sp. PCC 8801]
Length=108

Score = 31.6 bits (67), Expect = 54, Method: Compositional matrix adjust.
Identities = 27/101 (26%), Positives = 46/101 (45%), Gaps = 11/101 (10%)

Query 31 HEGYATYTCGGYSGGAFLLDFIESDMEITAINFADLNYGGVKAALAGSYLEVEGPRKKT 90
ATY S Y G S+ + A +P+ L GS ++V +
Sbjct 19 QAQTATYYSZYQCGKRTASGVRENSQPMANPS-----LPL-GAEVKEVTNENNGKS 69

Query 91 VVVFDLYPEGARGALDLSFNATRKIGNMKDGEINIKRWVVK 131
V V + R +DDL3 AFR+IG+ G+I + +V++
Sbjct 70 VIVIV--DRCRCSTDL3QAATRGCGSLRGRIVSIEVL 109

>=cF1ZF_02290103.11 hypothetical protein Immunoglobulin 15682 [Listeria monocytogenes
FSL J1-208]
Length=199

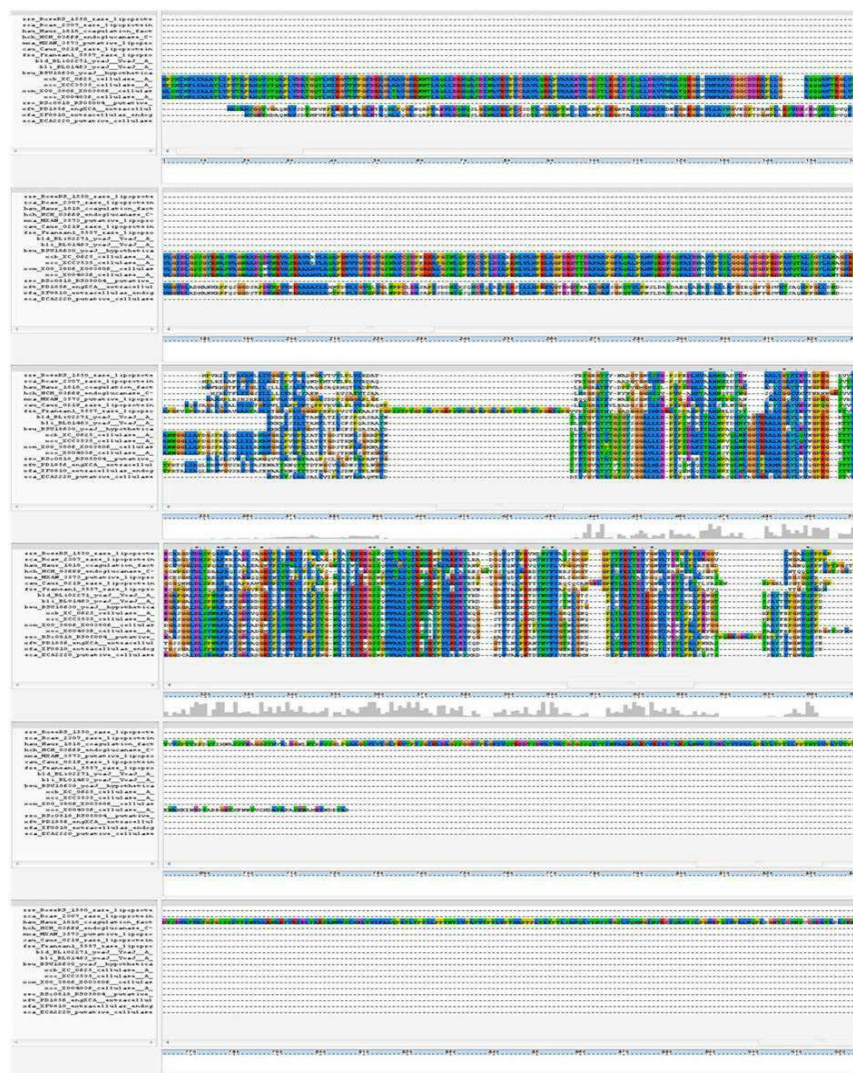
Score = 31.6 bits (67), Expect = 59, Method: Compositional matrix adjust.
Identities = 19/32 (59%), Positives = 23/32 (72%), Gaps = 7/32 (22%)

Query 1 MKKIMSAYFGNVLLTIFCTCFQASAAAYDDLH-----EGYATYT 38
MKKI+3+ V VLL IF QASA L + + TYT
Sbjct 1 MKKISSIV--VLLAIFITFFQASAHVSVLESESTVD3WETTT 41

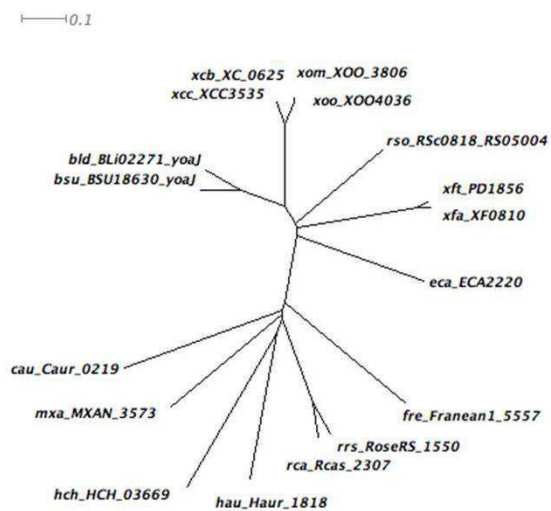
Database: All non-redundant GenBank CDS translations+EDB+SwissProt+PIR+DRE excluding
environmental samples from WGS projects
Posted date: Jan 6, 2008 9:57 AM
Number of letters in database: 1,058,578,688
Number of sequences in database: 3,390,049

Lambda K 0.280 H 1.79
Gapped
Lambda K 0.294 H 0.610
Matrix: DAM20
Gap Penalties: Existence: 9, Extension: 1
Number of Sequences: 3390049
Number of Hits to DB: 137167288
Number of extensions: 14999639
Number of successful extensions: 155957
Number of sequences better than 2e+06: 124728
Number of HSP's better than 2e+06 without gapping: 0
Number of HSP's gapped: 155549
Number of HSP's successfully gapped: 155435
Length of query: 232
Length of database: 1058578688
Length adjustment: 32
Effective length of query: 200
Effective length of database: 950097120
Effective search space: 190019424000
Effective search space used: 190019424000
T: 13
A: 40
K: 15 (7.4 bits)
N: 35 (14.0 bits)
R: 58 (24.6 bits)
S: 32 (16.0 bits)
D: 32 (16.0 bits)
```

도면15



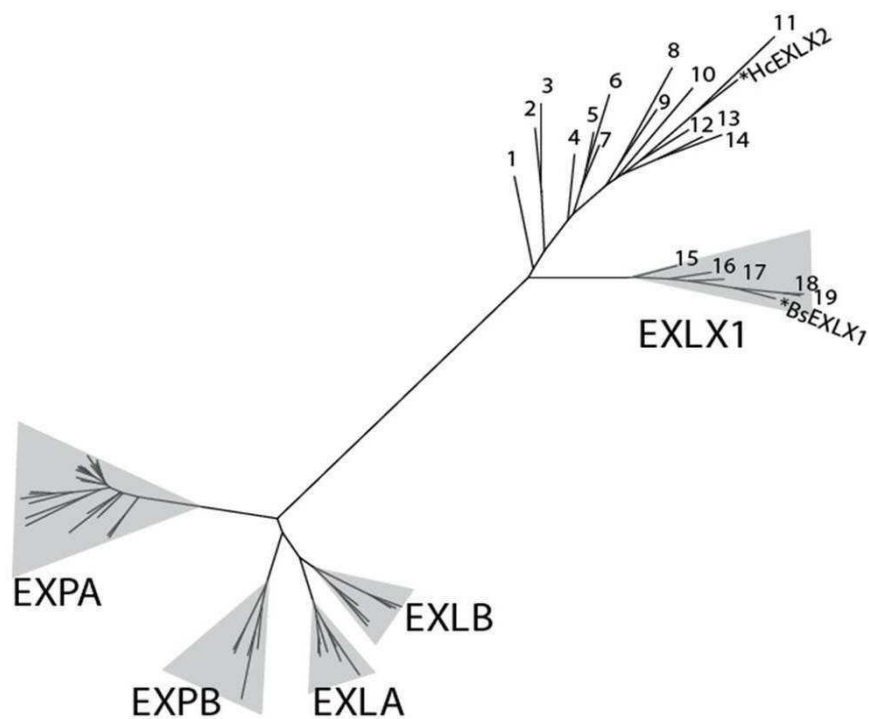
도면16



도면17a

No.	Name	Gene ID
1	<i>Leptothrix cholodnii</i>	gi 171061047
2	<i>Clavibacter michiganensi</i>	gi 148272660
3	<i>Nakamurella multipartita</i>	gi 229224499
4	<i>Frankia</i> sp.	gi 158317309
5	<i>Micromonospora</i> sp.	gi 238061663
6	<i>Streptomyces svceus</i>	gi 254404534
7	<i>Catenulispora acidiphila</i>	gi 229246366
8	<i>Plesiocystis</i>	gi 149922302
10	<i>Herpetosiphon aurantiacus</i>	gi 159898342
11	<i>Chloroflexus aurantiacus</i>	gi 163845817
12	<i>Acidovorax avenae</i>	gi 120612050
*	<i>Hahella chejuensis</i>	gi 83646399
13	<i>Roseiflexus</i> sp.	gi 148655687
14	<i>Myxococcus</i>	gi 108762346
15	<i>Stigmatella aurantiaca</i>	gi 115373847
16	<i>Xanthomonas campestris</i>	gi 188990055
17	<i>Ralstonia solanacearum</i>	gi 83747707
18	<i>Xylella fastidiosa</i> Dixon	gi 71275113
19	<i>Pectobacterium atrosepticum</i>	gi 50121146
*	<i>Bacillus subtilis</i>	gi 16078923

도면17b



도면18

SEPDDSGFTNDPEGEVRALGEFRQGIATWYNATGEGHCGYDASPKDMD
VAAMNAPQFANSVCGSCAEVEGPKGTVRVRIVDSCPECGPGLDLSE
QAFAKIAAVADGRVQTRWRPVTCVSGPVRYRFKEGSSQWWTAIQVR
NHRLPIQKLEWQRENGAWVDMQRQDYNFLASPGVDTATTKLRVTAS
DGQVLEDTLPRVEEKKVFDGAEQFSLKK.

도면19

MLSHIHFLIALATLVPVTAPAMQVSTQAPLV DATGQTLHIGVTWPGFDR
AGLTAVGMRNNTLAQLLDRMQASDINAVRVPVCAAVLQRAPVAAAEV
AGDSTLRGLDSLQLLDAVVHAATQRGMHVMFAFADGGCDDRAPLLGA
QQQAWTRGLVTLARRYGGNANVLGIDLGSSGYRNASWAGNAADQNW
NRVASRAAAMVLAQAPRWVVGVEGVGSNAVCSDPGRKALGSNLQPF
CVPLDIAPRHLVLMPLKLAGPDRDTTDAFAAPGFAQALPAMWQRDFGQF
AIDHTVVPVSLGGGLGDGDPDPAWQTALSGYLANTGMRS AFLGSWET
SNANNGGLAPDGSPRADKLLILRHAWGMLPVMPIATATGDSTEPAIG
KNAWNSTFTGTATPTGSGYSGGAVLLDPIPSDAFITALNSTQMNFGGIK
AALAGAYLEVTGPKGKTTVYVTDLYPDGASGGLDLSYNAFAAIGNLAD
GHIPISWKVVRAPITGNVQYRIKEGSSRYWAAIQVRHHIYPVVKLEVQK
GSTWTS LPKTPYNHFVGTNLGNKPLSIRITDIRGKVIADKIPALPEYGGSA
AYFEPGHVQFP.

도면20

MSKLTFLIIGSILLLLCVRASYAENRVSATHTSAALHEGEGTYFFYNGGG
HCSVPVPAMFTAAMNQTDYNGSQACGGCVKVTNRNNGKS VVARVDDS
CPGCNPGDVDLTDAFAAQISPLEAGRIPISWDYVPCDYPSVLLYFMEGSS
QWWTAVQVREQRYPVSSLAYRESGSTGSYQEIAREDYNYFVERSGMGT
GPFDFRITDIYGHVLEAGNITLQSGVPINTQQQFPSMGTS GVINQADK.

서 열 목 록

- <110> KOREA UNIVERSITY RESEARCH AND BUSINESS FOUNDATION
- <120> PROKARYOTIC EXPANSIN PROTEIN FOR ACTVATION CELLULOSE EXPANSION
AND CELLULOSE-DEGRADING COMPOSING THE SAME
- <150> US61/096,666
- <151> 2008-09-12
- <160> 57
- <170> Kopatent In 1.71
- <210> 1
- <211> 238
- <212> PRT
- <213> Q86AV4: (Expansin-like protein 5 from Dictyostelium discoideum)
- <400> 1

Met Arg Ile Asn Phe Lys Leu Ile Leu Ile Ile Leu Thr Ser Phe Tyr
1 5 10 15

Gly Ile Ile Asn Cys Gln Ser Thr Cys Pro Tyr Ser Lys Thr Val Ile
20 25 30

Asn Gly Ala Ser Ala Thr Phe Tyr Ser Ala Met Asp Asn Gly Asn Cys
35 40 45

Gly Phe Gly Lys Leu Thr Gly Pro Thr Gly Pro Gly Asn Tyr Met Ile
50 55 60

Ala Ala Leu Gly Thr Lys Leu Tyr Gln Asn Gly Ala Gln Cys Gly Gln
65 70 75 80

Cys Phe Lys Ile Ser Asn Ser Lys Asn Ala Ser Val Thr Val Met Ala
85 90 95

Thr Asp Ser Cys Asn Asp Ala Gly Tyr Cys Gln Arg Asp Asn His Phe
100 105 110

Asp Leu Ser Pro Thr Ala Phe Ser Ile Leu Gly Ala Gln Ser Gln Gly
115 120 125

Val Leu Asp Gly Leu Ser Tyr Val Lys Val Pro Cys Arg Val Ser Gly
130 135 140

Asn Val Lys Val Met Leu Lys Asp Gly Ser Asn Ala Tyr Trp Thr Ser
145 150 155 160

Phe Leu Val Phe Asn Asn Ala Ile Asp Val Lys Gln Val Ser Ile Lys
165 170 175

Leu Ser Gly Ser Ser Thr Tyr Val Pro Leu Thr Gln Thr Thr Tyr Asn
180 185 190

Tyr Trp Pro Ser Ser Ile Thr Ala Gly Ser Phe Gln Val Arg Ile Glu
195 200 205

Ser Ile Gly Gly Glu Phe Ile Tyr Val Thr Ile Pro Ser Val Val Ser
210 215 220

Ser Lys Ile Tyr Asp Thr Gly Ser Gln Phe Ser Ser Ser Cys
225 230 235

<210> 2

<211> 335
 <212> PRT
 <213> Q7KWS2: (Expansin-like protein 3 from Dictyostelium discoideum)
 <400> 2

```

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

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

      115            120            125
Asp Gly Leu Ser Tyr Val Lys Val Pro Cys Glu Val Ser Gly Asn Val
      130            135            140
Lys Ile Met Met Lys Asp Gly Ser Asn Asp Phe Trp Thr Ser Phe Phe
      145            150            155            160
Val Phe Asn Ser Lys Val Ile Ile Lys Gln Val Ser Ile Lys Leu Ser
      165            170            175
Asn Ser Asn Gln Phe Val Pro Leu Ser Gln Thr Thr Tyr Asn Tyr Trp
      180            185            190

Pro Thr Ser Ile Thr Gly Gly Gln Phe His Val Arg Ile Glu Ser Ile
      195            200            205
Gly Gly Glu Phe Ile Tyr Val Thr Ile Pro Lys Val Glu Ser Arg Lys
      210            215            220
    
```

Val Tyr Glu Thr Ser Gly Gln Phe Ser Thr Ser Cys Ser Asn Leu Asn
225 230 235 240

Glu Asn Asn Pro Ile Asn Tyr Lys Pro Gln Thr Phe Asn Ser Gln Gln
245 250 255

Thr Ser Asn Asn Gln Asn Ser Asn Thr Gln Thr Pro Thr Lys Gln Pro

260 265 270
Ser Pro Asn Ser Gln Asn Phe Ile Pro Ser Tyr Cys Gln Gln Tyr Ile
275 280 285

Gln Lys Pro Asn Tyr Ile Phe Ala Lys Glu Ser Lys Glu Met Leu Val
290 295 300

Leu Asn Glu Asn Glu Asn Ile Glu Ser Asn Ser Leu Lys Leu Leu Pro
305 310 315 320

Asn Phe Leu Leu Leu Ile Leu Ile Ile Leu Leu Asn Ile Asn Phe
325 330 335

<210> 3

<211> 286

<212> PRT

<213> Q55G31: (Expansin-like protein 1 from Dictyostelium discoideum)

<400> 3

Met Lys Thr Phe Val Leu Phe Val Ile Leu Leu Cys Leu Thr Phe Leu
1 5 10 15

Ser Ile Ser Lys Ser Glu Thr Cys Pro Phe Ser Gln Ser Leu Val Ser
20 25 30

Gly Ala Ser Ala Thr Tyr Tyr Thr Asp Pro Asn Ala Gly Asn Cys Gly
35 40 45

Tyr Glu Asn Leu Met Gly Pro Leu Gly Pro Gly Asn Leu Phe Ile Ala

50 55 60
Ala Leu Gly Pro Asn Leu Tyr Asn Asn Gly Lys Asn Cys Gly Gln Cys
65 70 75 80

Phe Asn Ile Ser Ser Pro Tyr Thr Asn Arg Ser Val Val Ile Met Ala
85 90 95

Thr Asp Ser Cys Pro Asp Ser Gly Tyr Cys Gln Arg Ser Ser His Phe

100 105 110
 Asp Leu Ser Thr Gln Ala Phe Asp Val Leu Gly Ala Gln Ser Ile Gly
 115 120 125

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

195 200 205
 Glu Ser Val Gly Gly Glu Phe Ile Tyr Ile Thr Ile Pro Lys Val Glu
 210 215 220
 Ser Arg Lys Gln Tyr Glu Thr Ser Ser Gln Phe Ser Val Asp Gly Cys
 225 230 235 240
 Val Gly Thr Pro Ser Gly Pro Ser Gly Gly Leu Gly Ser Pro Ser Thr
 245 250 255
 Gly Ala Ser Ile Gly Thr Pro Ser Asp Ala Ser Ser Leu Thr Leu Tyr
 260 265 270

Ala Leu Phe Ser Leu Thr Ile Leu Phe Leu Val Met Leu Asn
 275 280 285

<210> 4

<211> 352

<212> PRT

<213> A1DF29: (Extracellular cellulase CelA, putative from Neosartorya fischeri)

<400> 4

Met Leu Tyr Gln Arg Leu Thr Thr Leu Gly Val Ala Ala Leu Val Ala
 1 5 10 15
 Ala Ser Ser Val Ser Ala Ser Pro Met Val Gly Gly Ala Lys Ser Arg
 20 25 30

Cys Arg Ala Gly Tyr Asn Lys Ala Val Ser His Val Pro Thr Thr His

35 40 45

Glu Ala Thr Pro Ser Ile Asn Pro Pro Val Glu Leu Ser Glu Ala Pro

50 55 60

Ser Gln Ser Pro Trp Pro Thr Val Asp Ser Pro Ser Ala Ala Ser Ile

65 70 75 80

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

85 90 95

Glu Asp Pro Asp Ala Ser Ser Ser Ser Ser Ser Ser Thr Thr Ser

100 105 110

Ala Leu Pro Ser Ser Pro Ala Thr Gln Gln Gln Asp Thr Val Pro Ala

115 120 125

Ala Pro Ala Glu Pro Ala Thr Ala Ala Ser Gln Thr Asn Asn Lys Ala

130 135 140

Ala Ala Ala Ala Ser Ser Ser Ser Ser Thr Thr His Ser Gly Lys Ala

145 150 155 160

Thr Phe Tyr Gly Gly Asn Ile Ser Gly Gly Thr Cys Ser Phe Thr Gly

165 170 175

Tyr Thr Leu Pro Ser Gly Leu Phe Gly Thr Ala Tyr Ser Gly Ala Ala

180 185 190

Trp Asn Asn Ala Ala Glu Cys Gly Ala Cys Val Ser Val Lys Gly Pro

195 200 205

Asn Gly Lys Thr Ile Lys Ala Met Ile Val Asp Gln Cys Pro Glu Cys

210 215 220

Glu Gln Asp His Leu Asp Leu Phe Gln Asp Ala Phe Thr Gln Leu Ala

225 230 235 240

Asp Val Ser Lys Gly Ile Ile Pro Ile Thr Trp Ser Phe Val Pro Cys

245 250 255

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

260 265 270

Tyr Trp Phe Ser Met Gln Val Met Asn Ala Asn Glu Pro Val Ala Lys

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

 325 330 335
 Asn Val Gly Thr Ser Ser Gly Ser Ser Val Thr Ala Lys Ser Asn Leu
 340 345 350
 <210> 5
 <211> 344
 <212> PRT
 <213> Q4WUD1: (Extracellular cellulase CelA/allergen Asp F7-like from Aspergillus fumigatus)
 <400> 5
 Met Leu Tyr Gln Arg Leu Thr Ala Leu Gly Val Ala Ala Leu Val Ala
 1 5 10 15
 Ala Ser Ser Val Ser Ala Ser Pro Met Ala Arg Gly Val Lys Gly Arg

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

 Ser Ser Ser Leu Ser Thr Thr Ser Ala Leu Pro Ser Ser Pro Ala Thr
 100 105 110
 Gln Glu Gln Asp Thr Val Ser Ala Ala Pro Ala Glu Pro Thr Thr Ala
 115 120 125
 Ala Pro Glu Thr Val Asn Lys Ala Ala Thr Thr Ala Ser Ser Ser Ser
 130 135 140

Ser Thr Thr His Ser Gly Lys Ala Thr Phe Tyr Gly Gly Asn Val Ser
 145 150 155 160
 Gly Gly Thr Cys Ser Phe Thr Gly Tyr Thr Leu Pro Ser Gly Leu Phe

165 170 175
 Gly Thr Ala Tyr Ser Gly Ala Ala Trp Asn Asn Ala Ala Glu Cys Gly
 180 185 190
 Ala Cys Val Ser Val Thr Gly Pro Asn Gly Lys Thr Ile Lys Ala Met
 195 200 205
 Ile Val Asp Gln Cys Pro Glu Cys Glu Gln Asp His Leu Asp Leu Phe
 210 215 220
 Gln Asn Ala Phe Thr Gln Leu Ala Asp Val Ser Lys Gly Ile Ile Pro
 225 230 235 240

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

305 310 315 320
 Ser Gly Ala Thr Leu Thr Val Thr Asn Val Gly Thr Ser Ser Gly Ser
 325 330 335
 Ser Val Thr Ala Lys Ser Asn Leu
 340

<210> 6

<211> 428

<212> PRT

<213> A1CAQ3: (Cellulase, putative from Aspergillus clavatus)

<400> 6

Met Arg Tyr Gln Arg Leu Thr Thr Leu Gly Val Ala Ala Leu Gly Ala

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

Ala Gly Leu Phe Gly Thr Ala Leu Ser Lys Ala Arg Trp Ser Asp Ala
260 265 270

Ala Glu Cys Gly Ala Cys Val Ser Val Thr Gly Pro Asn Gly Asn Ser
275 280 285

Ile Lys Ala Met Ile Val Asp Gln Cys Pro Glu Cys Glu Ser Asn His
290 295 300

Leu Asp Leu Phe Gln Asp Ala Phe Ala Glu Leu Ala Asp Ile Ser Lys
305 310 315 320

Gly Ile Ile Gly Ile Asp Trp Ser Tyr Val Pro Cys Glu Ile Asp Ser
325 330 335

Pro Leu Val Leu Lys Asn Lys Glu Gly Thr Ser Arg Tyr Trp Phe Ser
340 345 350

Met Gln Val Val Asn Ala Asn Glu Pro Val Ala Ser Leu Glu Val Ser
355 360 365

Thr Asn Gly Gly Ser Thr Trp Gln Pro Thr Thr Arg Thr Tyr Tyr Asn
370 375 380

Phe Phe Glu Asn Ala Ser Gly Phe Gly Ala Asp Thr Val Asp Val Arg
385 390 395 400

Val Thr Gly Val Ser Gly Lys Ser Leu Thr Val Lys Asn Val Gly Val
405 410 415

Gly Ser Ser Ser Ser Val Thr Ala Ser Ser Asn Leu
420 425

<210> 7

<211> 420

<212> PRT

<213> Q2U7S7: (Predicted protein from Aspergillus oryzae)

<400> 7

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

Thr Gly Ser Val Ser Ala Ser Pro Leu Ile Arg His Glu Gly Glu Ser
20 25 30

Val Cys Pro Ser Gly Tyr Thr Gln Ser Val Tyr Tyr Val Thr Val Thr

35 40 45
 Ala Ser Ser Thr Pro Ala Ser Thr Ser Ser Val Glu Pro Thr Thr Thr
 50 55 60
 Ile Glu Ser Thr Ser Thr Val Thr Glu Thr Thr Val Ile Thr Pro Glu

 65 70 75 80
 Ile Pro Ala Gln Ser Pro Thr Ser Thr Pro Val Glu Ile Pro Ala Pro
 85 90 95
 Val Glu Thr Pro Ala Pro Val Glu Thr Pro Ser Pro Ala Glu Thr Pro
 100 105 110
 Ala Pro Val Glu Thr Asn Thr Pro Val Glu Pro Thr Thr Ser Ser Ser
 115 120 125
 Thr Thr Glu Thr Pro Val Val Ala Pro Thr Ile Ala Thr Pro Ser Thr
 130 135 140

 Ala Asp Val Gln Pro Thr Glu Val Val Ala Glu Pro Ser Thr Ser Ser
 145 150 155 160
 Ser Ser Thr Glu Glu Pro Thr Ala Thr Pro Ile Ala Ala Glu Thr Pro
 165 170 175
 Ser Thr Thr Val Asp Ala Gln Pro Thr Thr Ala Ala Ala Ala Pro Thr
 180 185 190
 Thr Lys Gln Leu Lys Leu Ser Thr Thr Ser Thr Ala Ala Pro Ala Ala
 195 200 205
 Ser Val Thr Ser Ser Ser Thr Gly Ser Ser Ser Ser Ser Ser Ser Ser

 210 215 220
 Ser Glu Ser Asn Ser Gly Glu Ala Thr Phe Tyr Gly Gly Asn Leu Ser
 225 230 235 240
 Gly Gly Ala Cys Ser Phe Thr Gly Tyr Thr Leu Pro Ser Asn Leu Phe
 245 250 255
 Gly Thr Ala Leu Gly Ser Pro Arg Trp Asp Asn Ala Ala Glu Cys Gly
 260 265 270
 Ala Cys Val Ala Val Thr Gly Pro Asn Gly Asn Thr Ile Lys Ala Met
 275 280 285

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

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

Glu Val Thr Ala
 420

<210> 8
 <211> 341
 <212> PRT

<213

> Q0CDB1: (Putative uncharacterized protein from Aspergillus terreus)

<400> 8

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

50 55 60
 Ala Ile Glu Thr Ser Val Ser Val Gly Leu Pro Pro Ser Ser Ser Thr

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

325 330 335

Ala Ser Thr Asn Phe

340

<210> 9

<211> 378

<212> PRT

<213> B6HQC7: (Pc22g15890 protein from *Penicillium chrysogenum*)

<400> 9

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

1 5 10 15

Thr Val Ser Ala Gly Pro Leu Gly Ala Arg Glu His Asp Gly Tyr Cys

20 25 30

Pro Lys Gly Tyr Thr Met Ser Val Tyr Tyr Lys Thr Ile Thr Val Glu

35 40 45

Ser Tyr Pro Ser Thr Thr Ser Val Glu Ser Thr Pro Ala Ile Val Glu

50 55 60

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

65 70 75 80

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

85 90 95

Ser Ser Ser Ala Pro Ala Ala Glu Thr Glu Val Ala Glu Thr Ala Ala

100 105 110

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

115 120 125

Thr Ala Ala Val Ile Glu Thr Ile Ala Ala Pro Ala Val Lys Thr Pro

130 135 140

Val Val Gln Pro Ser Thr Thr Glu Ala Pro Ala Glu Val Ala Ser Thr

145 150 155 160

Ser Thr Lys Ser Ser Thr Ser Thr Ser Ser Thr Lys Ser Ala Ser Thr

165 170 175

Gly Ser Ser Asn Gly Thr Pro Gly Lys Ala Thr Phe Tyr Gly Gly Asn

180 185 190
 Val Gly Gly Gly Thr Cys Ser Phe Ser Gly Tyr Thr Leu Pro Ser His
 195 200 205
 Leu Phe Gly Thr Ala Leu Ser Leu Gln Arg Trp Asp Asp Ala Ala Asn
 210 215 220
 Cys Gly Ala Cys Val Ser Val Thr Gly Pro Lys Gly Asn Ser Ile Lys
 225 230 235 240
 Ala Met Ile Val Asp Gln Cys Pro Glu Cys Glu Ser Asn His Leu Asp
 245 250 255
 Leu Phe Gln Glu Ala Phe Ala Glu Leu Ser Asp Ile Ser Ala Gly Ile
 260 265 270
 Ile Gln Thr Thr Trp Ser Tyr Val Pro Cys Asp Leu Asp Gly Pro Leu
 275 280 285
 Lys Leu Lys Asn Lys Glu Gly Thr Ser Ala Tyr Trp Phe Ser Met Gln
 290 295 300
 Val Val Asn Ala Asn Glu Ala Val Thr Ala Leu Glu Val Ser Thr Asp
 305 310 315 320
 Gly Gly Ser Ser Trp Gln Ser Thr Thr Arg Thr Tyr Tyr Asn Tyr Phe
 325 330 335
 Glu Asn Thr Ala Gly Phe Gly Thr Ser Thr Val Asp Val Arg Ile Thr
 340 345 350
 Gly Ala Ser Gly Ser Thr Val Val Val Lys Asp Val Gly Val Ser Ser
 355 360 365
 Gly Ser Glu Val Thr Ala Gly Ser Asn Leu
 370 375
 <210> 10
 <211> 524
 <212> PRT
 <213> A2QGR1: Xanthomonas campestris from Aspergillus niger)
 <400> 10
 Met Lys Tyr Arg Cys Leu Ala Ser Leu Gly Ile Ala Ala Leu Gly Ala

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

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

 145 150 155 160
 Ala Glu Ala Ser Pro Ala Ala Glu Thr Ser Thr Ala Val Glu Ala Thr
 165 170 175
 Ser Ser Ser Ser Ser Ser Val Val Glu Val Thr Thr Asn Ala Ala Ala
 180 185 190
 Asn Ala Ala Ala Thr Thr Ala Ala Ala Gln Pro Thr Pro Glu Ala Val
 195 200 205
 Thr Thr Glu Val Ala Thr Thr Glu Ala Ala Ala Thr Thr Glu Ala Ala
 210 215 220

 Ala Gln Pro Ser Pro Glu Ala Ala Thr Thr Glu Ala Val Thr Thr Gln
 225 230 235 240
 Ala Ala Ala Ala Ala Ala Thr Thr Glu Ala Ala Thr Thr Gln Glu Ala
 245 250 255
 Ala Ala Ala Ala Thr Thr Glu Ala Val Val Gln Ser Thr Ala Glu Ala

260 265 270
 Ala Thr Thr Glu Ala Ala Thr Thr Glu Ala Ala Ala Ala Thr Gln Ala
 275 280 285
 Ala Ala Thr Thr Val Glu Ser Lys Ser Thr Ser Thr Ser Ser Thr Ala

 290 295 300
 Ala Ser Ala Ala Thr Ser Ser Ala Ser Ser Ser Ser Ser Ser Thr Ser
 305 310 315 320
 Ser Ala Leu Ser Ser Glu Tyr Ser Gly Glu Ala Thr Phe Tyr Gly Gly
 325 330 335
 Asn Val Ser Gly Gly Thr Cys Ser Phe Thr Glu Tyr Thr Ile Pro Ser
 340 345 350
 Gly Leu Phe Gly Thr Ala Leu Ser Ser Gln Arg Trp Asp Asn Ala Ala
 355 360 365

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

 435 440 445
 Gln Val Val Asn Ser Asn Leu Pro Val Thr Lys Leu Glu Val Ser Thr
 450 455 460
 Asp Ser Gly Ser Thr Trp Gln Ser Thr Thr Arg Ser Ser Tyr Asn Tyr
 465 470 475 480
 Phe Glu Asn Gln Ser Gly Phe Gly Thr Asp Thr Val Asp Val Arg Val
 485 490 495
 Thr Ala Glu Gly Gly Ser Gln Ile Thr Val Lys Asn Val Ser Val Ser
 500 505 510

Ser Gly Ser Ser Val Thr Ala Ser Ser Asn Phe Glu

515 520

<210> 11

<211> 366

<212> PRT

<213> Q5AVE5: (Putative uncharacterized protein from Emericella nidulans)

<400> 11

Met Lys Tyr Gln His Phe Ser Ser Ile Ala Leu Ala Ala Leu Ser Ala

1 5 10 15

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

20 25 30

Cys Pro Ala Gly Tyr Ser Pro Ser Val Tyr Tyr Ile Thr Val Thr Ala

35 40 45

Glu Pro Ser Ser Thr Val Arg Pro Thr Ser Ser Ala Pro Ile Ser Ser

50 55 60

Thr Pro Thr Ser Thr Ser Thr Ser Thr Ser Glu Thr Ser Leu

65 70 75 80

Thr Thr Leu Ala Ser Thr Ser Thr Gly Asp Val Thr Val Thr Ser Ser

85 90 95

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

100 105 110

Ala Thr Ser Thr Thr Ser Glu Ser Ala Thr Ser Thr Ser Ala Leu Ser

115 120 125

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

130 135 140

Thr Thr Ala Ala Glu Lys Thr Ser Ser Thr Ser Thr Ser Ser Ser

145 150 155 160

Ser Ser Thr Ser Thr Asn Ser Gly Ala Thr Thr Gly Glu Ala Thr Phe

165 170 175

Tyr Gly Gly Asn Leu Ser Gly Gly Thr Cys Ser Phe Thr Asp Tyr Thr

180 185 190

Leu Pro Ser His Leu Ser Gly Val Ala Phe Ser Gly Gln Ala Trp Asp

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

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

325 330 335
 Val Arg Val Thr Gly Val Asp Gly Gly Val Ile Thr Val Lys Asn Val
 340 345 350
 Ser Val Ala Ser Gly Glu Ser Val Thr Ala Asp Gly Asn Phe
 355 360 365

<210> 12
 <211> 223
 <212> PRT
 <213> A4RA67: (Putative uncharacterized protein from Magnaporthe grisea)
 <400> 12

Met Ala Ser His Ser Ala Ile Ile Thr Arg Leu Leu Val Leu Ile Leu
 1 5 10 15

Ala Ile Thr Val Arg Gln Thr Ala Ala Val Asn Gly Lys Ala Ser His
 20 25 30
 Tyr Gly Gly Asn Leu Ser Gly Gly Asn Cys Met Leu Thr Ser Tyr Thr
 35 40 45

Ile Pro Ala Gly Thr Phe Gly Thr Ala Ile Ala Ile Ser Asn Tyr Asp
50 55 60
Asn Ser Asn Met Cys Gly Val Cys Leu Asn Val Lys Gly Pro Ser Gly
65 70 75 80
Ser Met Lys Val Met Val Val Asp Ser Cys Pro Asp Cys Pro Pro Asn
85 90 95
Lys Leu Asp Leu Phe Glu Asn Gly Phe Ser Asn Ile Gly Asp Phe Asn
100 105 110
Ala Gly Ile Val Asp Val Asp Trp Glu Val Thr Pro Cys Asp Ile Thr
115 120 125
Glu Pro Leu Gln Val Arg Asn Lys Asp Gly Thr Ser Lys Tyr Trp Phe
130 135 140
Ser Met Gln Val Leu Asn Ser Asn Gln Gln Val Gln Ser Leu Glu Val
145 150 155 160

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

<210> 13

<211> 397

<212> PRT

<213> B2VVM5: (Putative uncharacterized protein from *Pyrenophora tritici-repentis*)

<400> 13

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

35 40 45
 Val Ala Asp Ala Thr Thr Ser Pro Gly Ser Cys Gly Pro Ser Ile Val

 50 55 60
 Ser Val Thr Met Thr Glu Lys Val Tyr Val Thr Val Thr Tyr Thr Ser
 65 70 75 80
 Gly Thr Tyr Val Thr Pro Asp Ala Val Lys Thr Ile Asp Val Thr Ser
 85 90 95
 Thr Ser Thr Leu Asp Ile Tyr Thr Thr Val Thr Leu His Pro Ser Gly
 100 105 110
 Ser Val His Pro Ser Gly Gly Tyr Phe Ser Asn Ser Thr Thr Val Ala
 115 120 125

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

 195 200 205
 Gly Asn Leu Ser Gly Gly Thr Cys Ser Phe Val Gly Tyr Thr Ile Pro
 210 215 220
 Ala Gly Ile Tyr Gly Thr Ala Ile Ser Asp Phe Asn Trp Asp Thr Ala
 225 230 235 240
 Gly Ala Cys Gly Thr Cys Val Ser Val Thr Gly Pro Lys Gly Asn Thr
 245 250 255
 Val Lys Ala Met Val Val Asp Gln Cys Pro Gly Cys Gly Pro Asn His
 260 265 270

 Leu Asp Leu Phe Pro Asp Gly Phe Ala Ala Leu Ala Ser Pro Asn Ala
 275 280 285

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

340 345 350
 Phe Phe Glu Asn Pro Ala Gly Phe Gly Thr Asp Ala Val Asp Val Arg
 355 360 365
 Val Thr Ser Thr Asp Gly Lys Ser Ile Val Val Lys Asn Val Gly Ile
 370 375 380
 Lys Ser Glu Ser Lys Thr Thr Ala Gly Gly Asn Phe Ala
 385 390 395

<210> 14

<211> 397

<212> PRT

<213> Q0UGV6: (Putative uncharacterized protein from Phaeosphaeria nodorum)

<400> 14

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

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

Asn Phe Phe Glu Gln Pro Ser Gly Phe Gly Thr Asp Ser Val Asp Ile

355 360 365
Lys Ile Thr Ser Ile Asp Gly Lys Thr Val Val Val Lys Gly Val Ser

370 375 380
Val Thr Pro Asn Ala Leu Lys Thr Ala Gly Gly Asn Phe

385 390 395

<210> 15

<211> 304

<212> PRT

<213> A7EFJ4: (Putative uncharacterized protein from *Sclerotinia sclerotiorum*)

<400> 15

Met Val Phe Ser Ile Lys Asn Val Ala Ile Ala Ser Ala Phe Leu Ser

1 5 10 15

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

20 25 30

Thr Ser Tyr Gly Gly Asn Thr Asn Gly Gly Ala Cys Ser Phe Ser Thr

35 40 45

Tyr Thr Leu Pro Ser Asn Ile Leu Gly Thr Ala Leu Ser Asp Ser Asn

50 55 60

Trp Ala Thr Ala Ala Asn Cys Gly Arg Cys Val Ser Val Thr Gly Pro

65 70 75 80

Ser Gly Asn Asn Ile Thr Ala Met Ile Thr Asp Glu Cys Pro Gly Cys

85 90 95

Gly Thr Asn His Leu Asp Leu Tyr Gln Asn Ala Phe Thr Lys Leu Ala

100 105 110

Ala Leu Ser Val Gly Ile Ile Asp Ile Thr Trp Asp Tyr Val Pro Cys

115 120 125

Pro Ile Thr Thr Pro Leu Gln Val His Leu Lys Glu Gly Val Ser Ala

130 135 140

Tyr Trp Phe Ser Met Gln Val Val Asn Ala Ala Glu Gly Val Ser Lys

145 150 155 160

Leu Glu Val Ser Thr Asp Gly Ser Lys Thr Trp Lys Ser Thr Thr Arg
165 170 175
Thr Thr Tyr Asn Phe Phe Glu Asn Ser Ser Gly Phe Gly Thr Thr Ser
180 185 190
Val Asp Val Arg Val Thr Gly Leu Ser Gly Ser Thr Val Ile Ile Lys
195 200 205
Asn Val Ala Val Thr Pro Gly Leu Val Val Thr Ala Ser Gly Asn Leu
210 215 220
Lys Ser Gly Ser Thr Gly Ser Ala Lys Thr Val Val Pro Ala Thr Val
225 230 235 240
Ser Ser Thr Ser Ser Glu Val Asp Thr Val Glu Pro Val Val Ala Ala
245 250 255
Asp Ala Val Ala Thr Pro Ser Ala Ser Pro Val Ala Ala Val Ser Thr
260 265 270
Pro Asp Val Val Ala Thr Pro Ala Ser Ser Pro Ser Pro Ser Val Ala
275 280 285
Glu Glu Pro Thr Pro Glu Ser Glu Glu Ser Asp Glu Cys Asp Glu Ser
290 295 300

<210> 16
<211> 293
<212> PRT
<213> A8L5V4: (Rare lipoprotein A from Frankia sp.)
<400> 16

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

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

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

 195 200 205
 Glu Gly Ser Ser Ala Tyr Trp Leu Ala Ile Gln Val Arg Asn His Arg
 210 215 220
 Asn Pro Val Val Ser Leu Glu Val Arg Val Asn Gly Ala Trp Thr Ala
 225 230 235 240
 Leu Pro Arg Glu Met Trp Asn Tyr Phe Val Ala Pro Gln Gly Leu Gly
 245 250 255
 Pro Gly Pro Phe Thr Val Arg Ile Thr Asp Val Tyr Gly Glu Gln Leu
 260 265 270

 Val Glu Thr Val Asn Leu Ala Pro Ala Ser Val Gln Thr Thr Gly Ser
 275 280 285
 Gln Phe Ala Arg His
 290
 <210> 17
 <211> 302
 <212> PRT

<213> B5HYX2: (Rare lipoprotein A from Streptomyces sviveus)

<400> 17

Met Val Gly Ser Thr Ala Gly Leu Val Val Thr Gly Leu Leu Val Cys

1 5 10 15

Leu Ala Ile Ala Met Gln Pro Asp Arg Lys Asp Asp Ala Gly Arg Ala

20 25 30

Ala Ala Thr Ala Val Ala Asp Ser Gln Glu Thr Gly Pro Ala Ser Ser

35 40 45

Pro Thr Gly Ser Lys Ser Pro Thr Pro Ser Pro Ser Ala Val Ser Ala

50 55 60

Ser Leu Lys Ala Lys Ala Lys Lys Pro Lys Ala Ser Ala Ser Pro Ser

65 70 75 80

Ala Thr Lys Pro Ser Ser Ala Ser Pro Arg Lys Ala Thr Thr Thr Ser

85 90 95

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

100 105 110

Thr Val Tyr Lys Ala Gly Val Gly Asp Gly Ala Cys Ser Tyr Gly Pro

115 120 125

Ser Ser Asp Met Met Ile Ala Ala Met Asn Thr Thr Asp Tyr Glu Thr

130 135 140

Ser Lys Ala Cys Gly Ala Tyr Val Phe Val Arg Ala Ala Asn Gly Asn

145 150 155 160

Ser Val Thr Val Arg Ile Thr Asn Glu Cys Pro Leu Pro Cys Ala Pro

165 170 175

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

180 185 190

Ser Thr Gly Arg Leu Ala Val Thr Trp Ser Leu Val Ser Pro Ser Ser

195 200 205

Val Gly Thr Ile Ser Ile Arg Tyr Lys Ile Gly Ser Ser Pro Tyr Trp

210 215 220

Cys Gly Ile Gln Ala Ile Gly His Arg Asn Pro Leu Ala Arg Leu Glu

225 230 235 240

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

245 250 255
Tyr Phe Leu Ser Pro Asp Gly Ser Gly Cys Gly Lys Ala Ile Arg Leu

260 265 270
Thr Asp Ile Tyr Gly Glu Gln Leu Thr Val Asn Gly Val Ala Leu Arg

275 280 285
Pro Asp Val Val Gln Ser Thr Gly Leu Gln Phe Ala Arg His

290 295 300
<210> 18

<211> 346

<212> PRT

<213> A9F5W0: (Putative uncharacterized protein from Sorangium cellulosum)

<400> 18

Met Lys Leu Gly Ser Met Pro Arg Gly Ile Ser Trp Pro Leu Val Ser

1 5 10 15
Leu Ile Ala Leu Gly Val Ala Gly Cys Ser Asp Gly Gly Gly Gly Asp

20 25 30
Ala Gly Ser Ala Glu Ser Ala Ser Ser Ala Ala Gly Gly Pro Gly Gly

35 40 45
Ala Gly Ala Val Gly Ala Ser Ala Ser Ala Gly Thr Gly Ala Gly Gly

50 55 60
Gly Asp Gly Val Thr Thr Gly Ala Gly Ala Gly Pro Gly Ser Ser Thr

65 70 75 80
Thr Ala Gly Ser Thr Ser Ser Thr Ser Gly Gly Gly Gly Asp Pro Ser

85 90 95
Ala Thr Ser Thr Gly Ala Ser Thr Ser Thr Gly Thr Gly Ser Gln Pro

100 105 110
Val Ser Cys Asp Tyr Pro Ala Glu Tyr Ala Asn Gly Ser Ile Thr Phe

115 120 125
Tyr Thr Leu Asp Met Gly Ser Thr Glu Val Asn Cys Ser Tyr Pro Ile

130 135 140

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

210 215 220
 Glu Glu Arg Glu Gly Tyr Leu Gly Thr Thr Asn Gly Gly Ala Ala Gly
 225 230 235 240
 Lys Ile Ser Trp Arg Tyr Ile Pro Cys Pro Thr Thr Ser Ala Val Ser
 245 250 255
 Phe Arg Leu Lys Asp Ala Ser Asn Lys Tyr Trp Asn Glu Ile Leu Val
 260 265 270
 Glu Gly His Ala His Pro Ile Glu Lys Leu Glu Val Glu Ile Asn Gly
 275 280 285

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

<210> 19
 <211> 438
 <212> PRT
 <213> A9FSQ7: (Putative uncharacterized protein from Sorangium cellulosum)

<400> 19
 Met Arg Leu Gly Ala Tyr Arg Ser Ala Val Ala Leu Ser Phe Leu Gly

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

 65 70 75 80
 Ile Asn Leu Asn Gln Ala Ala Val Ser Ser Trp Trp Asn Ala Ala Ser
 85 90 95
 Ser Gln Ser Gly Ser Thr Leu Thr Val Ser Pro Arg Ser Gly Ser Pro
 100 105 110
 Ser Ile Pro Val Gly Ser Ser Val Ser Phe Gly Phe Cys Ala Asn Ala
 115 120 125
 Thr Gly Ser Asn Tyr Arg Pro Thr Leu Val Ser Val Thr Lys Thr Gly
 130 135 140

 Gly Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
 145 150 155 160
 Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ala Gly Ser Thr Gly Ser
 165 170 175
 Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Ser Gly Gly Ser
 180 185 190
 Gly Gly Ser Thr Gly Ala Gly Gly Ser Ser Ala Thr Cys Ser Ile Pro
 195 200 205
 Leu Pro Ser Tyr Thr Ser Gly Asn Gly Ser Ala Thr Trp Tyr Ser Leu

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

260 265 270
Cys Val Glu Val Ser Arg Asp Asp Gly Arg Lys Val Val Ala Thr Val
275 280 285

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

355 360 365
Thr Pro Val Arg Lys Leu Glu Val Tyr Val Gly Gly Arg Trp Thr Glu
370 375 380
Ala Thr Arg Gln Pro Tyr Asn Tyr Trp Arg Val Gly Ser Gly Asn Met
385 390 395 400
Gly Gln Gly Pro Trp Arg Val Arg Val Thr Asp Val Leu Gly Lys Thr
405 410 415
Ile Glu Thr Ser Ile Asn Thr Thr Thr Ala Glu Gln Leu Thr Ser Ser
420 425 430

Gln Phe Pro Val Cys Gln

435

<210> 20

<211> 243

<212> PRT

<213> Q08V48: (Cellulase from *Stigmatella aurantiaca*)

<400> 20

Met Arg Phe Gly Lys Ala Phe Ser Ser His Ala Leu Leu Val Ala Gly

1 5 10 15

Val Leu Ala Leu Thr Ala Cys Ser Asp Asn Gly Asp Gly Asp Gly Asp

20 25 30

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Asp Val Gly

35 40 45

Leu Gly Lys Asp Pro Lys Asp Gly Arg Val Asn Ala Tyr Glu Thr Glu

50 55 60

Gly Asp Gly Ser Cys Ser Phe Gly Pro Pro Val Gly Glu Lys Phe Val

65 70 75 80

Ala Ala Leu Ser Ser Thr Glu Phe Leu Gly Ala Ala Val Cys Gly Ala

85 90 95

Cys Ala Glu Val Ser Gly Pro Asn Gly Lys Val Val Ala Arg Val Val

100 105 110

Asp Ile Cys Asp Ser Cys Ala Pro Asp Gln Met Leu Leu Ser Gln Glu

115 120 125

Ala Tyr Ser Lys Val Ala Ser Gly Gly Glu Val Lys Met Ser Trp Lys

130 135 140

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

145 150 155 160

Gly Ser Gly Pro Tyr Ser Tyr Phe Ala Ile Gln Val Arg Asn His Lys

165 170 175

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

180 185 190

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

195 200 205

Pro Pro Leu Gln Val Arg Val Thr Gly Thr Thr Gly Glu Val Leu Glu

210 215 220

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

225 230 235 240

Gln Phe Lys

<210> 21

<211> 298

<212> PRT

<213> A6GCR8: (Putative lipoprotein from Plesiocystis pacifica SIR-1)

<400> 21

Met Ser Pro Ala Arg Ser Ile Ala Leu Leu Ser Leu Thr Leu Val Ser

1 5 10 15

Leu Phe Gly Cys Ala Glu Glu Pro Ser Glu Glu Gly Ser Thr Gly Gly

20 25 30

Ala Thr Leu Gly Phe Pro Asp Thr Gly Asp Glu Ala Asp Thr Thr Thr

35 40 45

Ala Gly Glu Ser Ser Glu Gly Glu Ala Ser Ser Gly Glu Ser Gly Ser

50 55 60

Gly Glu Ser Ser Gly Glu Thr Ser Gly Glu Ala Ser Ser Gly Asp Thr

65 70 75 80

Ala Ser Glu Thr Ser Gly Gly Gln Gly Ser Cys Gly Pro Ala Gln Val

85 90 95

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

100 105 110

Cys Ser Tyr Pro Thr Gly Thr Leu Pro Gln Phe Tyr Gly Ala Leu Asn

115 120 125

Thr Ala Gln Tyr Ala Asp Ala Ala Met Cys Gly Ala Cys Ala Arg Val

130 135 140

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

145 150 155 160

Ile Ala Thr Asn Pro Ile Cys Tyr Glu Gly His Ile Asp Leu Asn Pro

165 170 175

Pro Ala Phe Glu Gln Ile Gly Glu Ile Ile Asp Gly Ile Ile Pro Ile

180 185 190

Thr Trp Glu Leu Ile Ser Cys Asp Asp Pro Gly Thr Ile Ala Tyr Ser

195 200 205

Phe Lys Glu Gly Ser Ser Gln Trp Trp Thr Gly Val Leu Val Arg Asn

210 215 220

His Arg Asn Pro Ile Ala Lys Phe Glu Tyr Asp Thr Gly Gly Gly Asn

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

 Val Gly Gly Ser Gln Phe Pro Ala Cys Asp
 290 295
 <210> 22
 <211> 243
 <212> PRT
 <213> Q098B5: (YoaJ from *Stigmatella aurantiaca* DW4/3-1)
 <400> 22
 Met Arg Ala Leu Pro Leu Val Ser Leu Arg His Leu Thr Ala Val Gly
 1 5 10 15
 Met Leu Gly Leu Thr Ala Cys Ser Ser Ser Glu Pro Asp Asp Ser Gly
 20 25 30
 Phe Thr Asn Asp Pro Glu Gly Glu Val Arg Ala Leu Gly Glu Phe Arg

 35 40 45
 Gln Gly Ile Ala Thr Trp Tyr Asn Ala Thr Gly Glu Gly His Cys Gly
 50 55 60
 Tyr Asp Ala Ser Pro Lys Asp Met Asp Val Ala Ala Met Asn Ala Pro
 65 70 75 80
 Gln Phe Ala Asn Ser Ala Val Cys Gly Ser Cys Ala Glu Val Glu Gly
 85 90 95
 Pro Lys Gly Thr Val Arg Val Arg Ile Val Asp Ser Cys Pro Glu Cys
 100 105 110

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

Ala Val Ser Gly Pro Val Arg Tyr Arg Phe Lys Glu Gly Ser Ser Gln
 145 150 155 160
 Trp Trp Thr Ala Ile Gln Val Arg Asn His Arg Leu Pro Ile Gln Lys
 165 170 175
 Leu Glu Trp Gln Arg Glu Asn Gly Ala Trp Val Asp Met Gln Arg Gln
 180 185 190
 Asp Tyr Asn Tyr Phe Leu Ala Ser Pro Gly Val Asp Thr Ala Thr Thr
 195 200 205
 Lys Leu Arg Val Thr Ala Ser Asp Gly Gln Val Leu Glu Asp Thr Leu
 210 215 220
 Pro Arg Val Glu Glu Lys Lys Val Phe Asp Gly Ala Glu Gln Phe Ser
 225 230 235 240
 Leu Lys Lys

<210> 23
 <211> 233
 <212> PRT
 <213> Q1D6G0: (Putative lipoprotein from Myxococcus xanthus)

<400> 23
 Met Ala Met His Pro Leu Arg Ser Lys Ser Arg Arg Phe Leu Leu Pro
 1 5 10 15
 Val Ala Ala Val Phe Val Ala Cys Gly Gly Cys Ser Asp Asp Pro Ser
 20 25 30
 Gly Gly Gly Ala Pro Leu Gly Glu Glu Gln Gln Gly Ile Ala Thr Phe
 35 40 45
 Tyr Asn Ala Thr Gly Ser Gly Asn Cys Ser Tyr Glu Pro Thr Gly Asp
 50 55 60
 Leu Met Val Ala Ala Met Asn Thr Pro Gln Tyr Ala Asn Ser Ala Ala
 65 70 75 80
 Cys Gly Gln Cys Val Asp Ile Thr Gly Pro Lys Gly Ser Val Arg Val
 85 90 95
 Arg Ile Val Asp Arg Cys Pro Glu Cys Glu Ser Gly His Leu Asp Leu

100 105 110
 Ser Arg Glu Ala Phe Ala Arg Ile Ala Glu Met Gln Gln Gly Arg Val
 115 120 125
 Asn Ile Thr Trp Thr Pro Val Ser Cys Asp Val Ala Gly Asn Ile Glu
 130 135 140

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

 210 215 220
 Ser Ala Glu Gly Ser Gly Gln Phe Arg
 225 230
 <210> 24
 <211> 254
 <212> PRT
 <213> B1Y0F2: (Rare lipoprotein A from *Leptothrix cholodnii*)
 <400> 24
 Met Arg Arg His Pro Arg Arg Phe Pro His Arg Met Leu Asp Leu Glu
 1 5 10 15
 Ser Ala Leu Lys Arg Leu Ala Cys Leu Cys Thr Leu Ala Gly Leu Ala
 20 25 30
 Ala Cys Gly Gly Gly Ser Pro Leu Ala Ala Ala Ser Pro Thr Val Asp

 35 40 45
 Thr Asp Gly Gly Gly Thr Gly Ala Thr Leu Gly Ala Thr Arg Gln Gly
 50 55 60
 Glu Gly Thr Tyr Tyr Ala Ala Thr Gly Ala Gly Ala Cys Ser Tyr Asp
 65 70 75 80

Ala Ser Ala Asp Arg Met Val Ala Ala Met Asn His Thr Asp Tyr Ala

85 90 95

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

100 105 110

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

115 120 125

Asp Val Asp Leu Ser Ala Glu Ala Phe Ala Arg Ile Ala Glu Pro Val

130 135 140

Ala Gly Arg Val Pro Ile Thr Trp Gln Val Val Thr Gly Ala Val Ser

145 150 155 160

Gly Pro Val Gln Tyr Arg Tyr Lys Glu Gly Ser Thr Arg Tyr Trp Thr

165 170 175

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

180 185 190

Gln Pro Asp Gly Thr Gly Asn Trp Ile Ala Val Pro Arg Thr Asp Tyr

195 200 205

Asn Tyr Phe Val His Pro Val Ala Ile Gly Ala Gly Ala Leu Gln Val

210 215 220

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

225 230 235 240

Pro Gln Gly Gly Leu Val Val Asp Gly Thr Gly Gln Phe Gln

245 250

<210> 25

<211> 1512

<212> PRT

<213> A5UTY9: (Putative uncharacterized protein from Roseiflexus sp.)

<400> 25

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

1 5 10 15

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

20 25 30

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

35 40 45
 Pro Val Gly Pro Gln Pro Gly Asp Val Phe Arg Gln Phe Gln Asp Asp

 50 55 60
 Leu Asn Ser Arg Gln Thr Ser Val Cys Asn Ser Ser Pro Ala Asn Cys
 65 70 75 80
 Gly Thr Ile Pro Ser Glu Gln Gln Arg Gln Ala Arg Asn Val Asp Leu
 85 90 95
 Ala Asn Ala Val Arg Ala Glu Ala Val Phe Glu Tyr Trp Gly Gly His
 100 105 110
 Leu Ala Ser Thr Gln Gln Glu Phe Asn Val Asn Gly Gly Pro Trp Thr
 115 120 125

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

 195 200 205
 Thr Arg Ile Tyr Tyr Asn Ser Thr Ile Pro Arg Pro Asp Gly Ala Ile
 210 215 220
 Thr Ser Pro Thr Asp Gly Ala Thr Ile Thr Glu Asn Arg Ser Ala Asn
 225 230 235 240
 Gln Phe Pro Arg Ile Thr Val Arg Ala Gln Ala Ser Thr Gly Arg Ser
 245 250 255
 Ile Ala Arg Val Asp Leu Ile Gly Asn Tyr Leu Asp Tyr Asp Trp Asp
 260 265 270

 Gly Asp Gly Ile Leu Arg Glu Trp His Tyr Ala Leu Arg His Gly Val
 275 280 285

Trp Glu Arg Thr Ile Gly Thr Ser Thr Ser Pro Ser Arg Gly Thr Ala
 290 295 300
 Thr Ser Gly Glu Tyr Asp Phe Thr Trp Asn Thr Glu Trp Ile Pro Asp
 305 310 315 320
 Gln Asp Gln Pro Val Gln Leu Met Ala Arg Ile Thr Asp Asn Thr Gly
 325 330 335
 Val Ser Tyr Met Thr Glu Ala Val Thr Val Thr Phe Asn Arg Val Gly
 340 345 350
 Arg Ser Val Arg Met Tyr Thr Thr Trp Ser Asp Tyr Tyr Gly Gln Arg
 355 360 365
 Val Phe Gly Pro Pro Arg Asn Phe Gly Phe Asn Thr Tyr Ser Gln Pro
 370 375 380
 Arg Ile Gly Lys Thr Ala Phe Ile Ser Ile Pro Asp Pro Leu Thr Asn
 385 390 395 400
 Ala Thr Arg Ala Arg Val Ile Leu Tyr Ser Trp Ser Gly Arg Ser Ser
 405 410 415
 Thr Ala Leu Ser Thr Thr Val Ala Ile Asn Ala Phe Ser Val Trp Ser
 420 425 430
 Gly Asn Ser Phe Gly Ile Phe His Asp Tyr Asn Phe Asp Thr Arg Asp
 435 440 445
 Leu Asn Asn Pro Arg Asp Phe Leu Asn Gln Gly Asp Asn Ile Leu Arg
 450 455 460
 Ile Trp Ser Asn Arg Thr Asp His Ala Leu Glu Val Leu Trp Pro Gly
 465 470 475 480
 Pro Ala Ile Phe Val Glu Tyr Ser Ala Pro Pro Glu Thr Pro Val Leu
 485 490 495
 Arg Ala Ile Asp Thr Ser Phe Ser Thr Pro Glu Asp Gln Ser Leu Ser
 500 505 510
 Phe Thr Leu Pro Val Ser Asn Gln Ala Arg Ala Ala Leu Gln Tyr Thr
 515 520 525
 Ile Leu Thr Pro Pro Ala Leu Gly Thr Leu Arg Gly Ile Ala Pro Asn
 530 535 540

Leu Thr Tyr Ile Pro Asn Pro Asp Val Asn Gly Ser Asp Ser Phe Val
545 550 555 560

Phe Arg Val Asp Asp Gln Tyr Gly Arg Ser Ala Thr Ala Thr Val Ala
565 570 575

Ile Asn Ile Gln Ala Val Asp Asp Pro Pro Glu Ile Thr Leu Gly Gly
580 585 590

Gly Ala Leu Leu Asn Pro Asp Asp Val Leu Asp Arg Ser Gly Arg Val
595 600 605

Arg Asp Pro Asp Ser Ser Ser Leu Thr Ala Thr Val Asp Tyr Gly Asp
610 615 620

Gly Arg Gly Glu Gln Pro Leu Ser Leu Gly Ala Asp Gly Ser Phe Arg

625 630 635 640

Leu Ile Tyr Arg Tyr Arg Asn Gly Gly Arg Phe Pro Leu Thr Val Arg
645 650 655

Val Ser Asp Gly Ala Gln Thr Ser Val Ala Thr Val Leu Val Thr Val
660 665 670

Ile Thr Pro Gly Thr Pro Pro Pro Pro Gly Asp Trp Ser Leu Thr Gly
675 680 685

Trp Arg Tyr Arg Val Pro Leu Glu Ala Arg Ser Gly Ala Tyr Pro Arg
690 695 700

Val Asp His Val Ala Glu Phe Asp Val Asn Phe Thr Thr Ala Leu Ser
705 710 715 720

Gln Ala Gly Gly Ser Gly Ala Leu Leu Arg Glu Ser Ile Arg Val Val
725 730 735

Glu Val Asp Ala Gly Gly Asn Val Ile Asp Pro Asn Val Val Tyr Gln
740 745 750

Phe Asp Pro Val Ser Ala Tyr Asn Pro Thr Thr Asn Ala Arg Gly Thr
755 760 765

Leu Leu Val Leu Leu Arg Gly Arg Thr Glu Ala Asn Thr Thr Arg Arg

770 775 780

Phe Tyr Val Tyr Phe Glu Thr Gln Gly Arg Phe Ala Pro Pro Thr Phe

785 790 795 800
 Thr Pro Leu Thr Val Val Val Thr Pro Thr Leu Ile Tyr Gln Arg Asp
 805 810 815
 Leu Ala Thr Val Ile Glu Thr Arg Thr Ser Asp Gly Gln Arg Asn Ala
 820 825 830
 Thr Phe Tyr Tyr Ser Leu Asn Gly Gly Ala Leu Ile Ser Met Phe Asp
 835 840 845

 Arg Ser Asn Ala Asp Trp Ile Ser Tyr Asn Pro Ala Ala Asp Ser Ala
 850 855 860
 Gly Ala Tyr Arg Gly Ile Pro Asn Val Gly Pro Val Phe His Pro Gly
 865 870 875 880
 Tyr Tyr Asn Val Phe Pro Ile Pro Ser Ala Ser Gly Pro Leu Gln Gly
 885 890 895
 Asn Glu Pro Gly Thr Asn Gln Arg Ser Ser Thr Thr Ile Glu Arg Glu
 900 905 910
 Gly Thr Leu Arg Thr Ile Val Arg Ser Val Ser Ser Asn Gly Gln Trp

 915 920 925
 Glu Ala Val Trp Glu Phe Phe Pro Asn Ser Val Arg Met Leu Val Leu
 930 935 940
 Pro Arg Glu Pro Arg Gly Ser Asn Pro Phe Leu Tyr Trp Trp Leu Tyr
 945 950 955 960
 Glu Gly Thr Pro Gly Gly Thr Leu Asn Glu Gly Asp Arg Trp Val Arg
 965 970 975
 Gly Asp Gly Leu Ser Gly Ile Ala Val Asp Pro Ser Phe Ser Phe Asn
 980 985 990

 Glu Arg Phe His Ala Asp Ile Asn Gly Pro Glu Trp Val Tyr Phe Arg
 995 1000 1005
 Ala Gly Glu Pro Ser Gln Gly Gly Arg Ala Leu Trp Leu Ala His His
 1010 1015 1020
 Asp Asp Thr Asp Thr Thr Ile Glu Met Tyr Arg Tyr Gln Asn Asp Ser
 1025 1030 1035 1040
 Ser Gly Trp Gln Asn Gly Met Thr Val Phe Gly Phe Gly Arg Arg Glu

1045 1050 1055
Trp Asp Gly Ser Ser Gln Thr Leu Arg Pro Glu Leu Arg Asn Gln Arg

1060 1065 1070
Arg Phe Ser Leu Gly Phe Ile Asp Asn Ala Pro Leu Thr Phe Asp Ser

1075 1080 1085
Ala Val Gly Arg Glu Ile Glu Gly Arg Tyr Arg Asp Leu Ala Ala Thr

1090 1095 1100
Val Gly Arg Gly Glu Arg Leu Ala Pro Asn Leu Ala Pro Trp Ala Pro

1105 1110 1115 1120
Gly Glu Thr Tyr Ser Val Arg Pro Gly Gln Thr Leu Asn Val Pro Ala

1125 1130 1135

Pro Gly Val Arg Ala Asn Asp Ser Asp Gly Asp Gly Asp Pro Phe Gln

1140 1145 1150

Val Thr Leu Thr Gly Glu Pro Gln Asn Gly Thr Leu Thr Phe Asn Ser

1155 1160 1165

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

1170 1175 1180

Val Phe Thr Tyr Thr Ala Ser Asp Gly Arg Leu Ser Ala Pro Pro Thr

1185 1190 1195 1200

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

1205 1210 1215

Ser Ser Ala Thr Thr Thr Asn Leu Arg Pro Leu Thr Val Pro Ala Pro

1220 1225 1230

Gly Val Leu Glu Gly Ala Ser Asp Pro Asp Gly Asp Pro Leu Thr Ala

1235 1240 1245

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

1250 1255 1260

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

1265 1270 1275 1280

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

1285 1290 1295

Val Arg Ile Thr Val Arg Lys Glu Asn Arg Pro Pro Thr Ala Arg Gly
1300 1305 1310

Ser Ser Ala Thr Thr Thr Asn Leu Arg Pro Leu Thr Val Pro Ala Pro
1315 1320 1325

Gly Val Leu Glu Gly Ala Ser Asp Pro Asp Gly Asp Pro Leu Thr Ala
1330 1335 1340

Val Leu Ile Ser Pro Pro Ala Gly Gly Thr Leu Ser Leu Asn Pro Asp
1345 1350 1355 1360

Gly Ser Tyr Thr Tyr Thr Pro Arg Ser Asp Phe Ile Gly Glu Asp Val
1365 1370 1375

Phe Gln Ile Ala Ala Ser Asp Gly Glu Ala Gln Ser Ala Pro Val Glu
1380 1385 1390

Val Arg Ile Thr Val Arg Lys Glu Asn Arg Pro Pro Thr Ala Arg Gly
1395 1400 1405

Ser Ser Ala Thr Thr Thr Asn Leu Arg Pro Leu Thr Val Pro Ala Pro
1410 1415 1420

Gly Val Leu Glu Gly Ala Ser Asp Pro Asp Gly Asp Pro Leu Thr Ala
1425 1430 1435 1440

Val Leu Ile Ser Ser Pro Ala Gly Gly Thr Leu Ser Leu Asn Pro Asp
1445 1450 1455

Gly Ser Tyr Thr Tyr Thr Pro Arg Ser Asp Phe Ile Gly Glu Asp Val
1460 1465 1470

Phe Gln Ile Ala Ala Ser Asp Gly Glu Ala Gln Ser Ala Pro Val Glu
1475 1480 1485

Val Arg Ile Thr Val Thr Arg Glu Val Ala Ser Thr Pro Glu Tyr Arg
1490 1495 1500

Val Phe Leu Pro Leu Leu Val Arg
1505 1510

<210> 26
<211> 237
<212> PRT
<213> A7NLJ9: (Rare lipoprotein A from Roseiflexus castenholzii)

<400> 26

Met Pro Leu Arg Ile Leu Ala Phe Ala Gly Val Cys Leu Leu Leu Ala

1 5 10 15

Gly Ile Ile Pro Val Val Arg Ala Gln Asn Asp Thr His Thr Val Tyr

20 25 30

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

35 40 45

Tyr Tyr Ala Ala Asp Gly Thr Gly Asn Cys Met Phe Asp Pro Ser Pro

50 55 60

His Asp Leu Met Val Ala Ala Met Asn His Ile Asp Tyr Asp Asn Ala

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

Ile Ala Tyr Arg Phe Lys Glu Gly Ser Asn Gln Trp Trp Thr Ala Val

145 150 155 160

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

165 170 175

Ala Asp Gly Gln Trp Gln Thr Val Pro Arg Gln Met Tyr Asn Tyr Phe

180 185 190

Val Ala Ser Ser Gly Met Gly Pro Gly Pro Tyr Thr Phe Arg Val Thr

195 200 205

Asp Met Tyr Gly Asn Val Leu Thr Asp Thr Asp Ile Pro Phe Val Glu

210 215 220

Gly Gly Val Ile Asn Gly Ala Ala Gln Phe Pro Pro Pro

225 230 235

<210> 27

<211> 258
 <212> PRT
 <213> A9WC79: (Rare lipoprotein A from *Chloroflexus aurantiacus*)

<400> 27

Met Gln Trp Ser Ala Gly Arg Leu Gln Ile Ala Leu Val Ile Cys Leu
 1 5 10 15

Thr Gly Phe Phe Trp Leu Ile Ala Asn Pro Val Phe Thr Gln Pro Pro
 20 25 30

Tyr Arg Val His Leu Pro Leu Val Val Arg Pro Ala Gln Asn Pro Leu
 35 40 45

Leu Ser Gly Glu Ala Thr Phe Tyr Leu Glu Ala Asp Gly Ser Gly Ser
 50 55 60

Cys Leu Phe Asp Pro Ile Pro Gly Asp Lys Met Val Ala Ala Ile Ser

65 70 75 80

Tyr Leu Asn Tyr Gly Asn Ala Asp Tyr Cys Gly Ala Tyr Val Glu Val
 85 90 95

Phe Gly Pro Gln Gly Ser Val Val Val Arg Ile Val Asp Met Cys Pro
 100 105 110

Asp Asn Pro Gly Cys Gly Gln Asn His Leu Asp Leu Ser Pro Glu Ala
 115 120 125

Phe Asp Arg Ile Ala Pro Arg Ala Trp Gly Arg Val Pro Ile Thr Trp
 130 135 140

Arg Val Ile Ser Pro Pro Leu Ser Gly Pro Val Gln Phe His Leu Lys
 145 150 155 160

Asp Gly Ser Asn Pro Trp Trp Leu Ala Phe Gln Val Arg His His Arg
 165 170 175

Asn Pro Ile Ala Arg Met Glu Tyr Arg Thr Pro Gln Gly Gln Trp Val
 180 185 190

Gln Leu Asn Arg Thr Thr Tyr Asn Tyr Phe Ile Arg Gln Cys Gln Cys
 195 200 205

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

210 215 220
 Gly Asn Val Leu Thr Glu Thr Val Gln Phe Thr Thr Leu Ser Arg Ser
 225 230 235 240
 Ser Asn Ser Pro Gly Glu Leu Val Ser Gly Thr Gly Gln Phe Pro Tyr
 245 250 255
 Gly Pro

<210> 28
 <211> 477
 <212> PRT
 <213> A9B7T3: (Coagulation factor 5/8 type domain protein from Herpetosiphon aurantiacus)
 <400> 28
 Met Ala Trp Arg Lys Gly Thr Phe Phe Ser Phe Gly Leu Ile Leu Ser

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

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

145 150 155 160

Val Gln Val Arg Asn His Arg Asn Ala Ile Ala Lys Phe Glu Tyr Arg
165 170 175

Asn Pro Gln Gly Ile Phe Gln Asn Val Asp Arg Val Gln Tyr Asn Tyr
180 185 190

Phe Leu Val Pro Gly Gly Met Gly Thr Gly Pro Phe Thr Phe Arg Leu
195 200 205

Thr Asp Val Tyr Gly Asn Val Phe Thr Asp Asn Thr Ile Pro Leu Arg
210 215 220

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

Pro Gly Ser Thr Pro Thr Ala Ser Pro Thr Arg Thr Ala Ser Pro Thr
245 250 255

Val Val Ser Pro Thr Thr Arg Pro Asp Gln Thr Ile Met Asn Ala Ser
260 265 270

Ser Thr Glu Ala Gly Gly Ser Thr His Tyr Ala Leu Asp Gly Asn Leu
275 280 285

Asn Thr Arg Trp Ser Ser Gly Leu Pro Gln Ala Ala Gly Gln Trp Ile
290 295 300

Tyr Ile Gly Leu Pro Arg Val Thr Pro Ile Ser Gly Ile Lys Leu Asp
305 310 315 320

Ala Gly Ser Ser Gly Gly Asp Tyr Pro Ala Gly Phe Ile Val Gln Thr
325 330 335

Arg Asp Asp Thr Ser Asp Trp Ala Thr Val Ala Thr Gly Ser Gly Ser
340 345 350

Ser Gln Ile Thr Thr Ile Asn Phe Ala Ala Arg Asn Ala Arg Tyr Val
355 360 365

Arg Ile Glu Leu Thr Ala Arg Ser Ala Asn Trp Trp Ser Ile His Glu
370 375 380

Leu Thr Val Ile Phe Ala Ala Gln Pro Ala Thr Leu Thr Pro Thr Ile
385 390 395 400

Leu Pro Pro Thr Asn Thr Ile Val Pro Ala Thr Val Thr Pro Thr Leu
405 410 415

Val Pro Thr Met Ala Pro Pro Pro Leu Thr Ala Thr Pro Thr Leu Ser
420 425 430

Ala Trp Gln Ala Tyr Thr Asn Tyr Ser Val Gly Ser Leu Val Gln His

435 440 445
Asn Gly Ile Asn Tyr Arg Cys Ile Gln Ala His Thr Ser Leu Pro Gly

450 455 460
Trp Glu Pro Gln Ile Val Pro Ala Leu Trp Gln Pro Leu

465 470 475
<210> 29

<211> 273

<212> PRT

<213> A9B0R8: (Rare lipoprotein A from Herpetosiphon aurantiacus)

<400> 29

Met Leu Val Arg Thr His Ala Val Arg Trp Leu Leu Phe Val Ser Leu
1 5 10 15

Phe Met Val Ile Phe Gly Ser Tyr Ala Trp Gln Gln Gln Ser Phe Ala
20 25 30

Gln Asn Pro Ala Thr Ile Thr Asn Pro Trp Ala Ile Arg Gln Gly Gln
35 40 45

Ala Thr Tyr Tyr Thr Ala Thr Gly Leu Gly Asn Cys Ser Ile Pro Val
50 55 60

Pro Ser Asp Leu Leu Phe Gly Ala Met Asn Asn Pro Asp Tyr Ala Thr
65 70 75 80

Ala Asp Tyr Cys Gly Ala Phe Val Lys Ile Thr Gly Pro Leu Gly Ser

85 90 95
Val Thr Val Gln Ile Thr Asp Arg Cys Pro Glu Cys Gln Thr Gly His

100 105 110
Ile Asp Leu Ser Pro Gln Ala Phe Asp Arg Ile Ala Asn Arg Val Thr

115 120 125
Gly Ile Val Pro Ile Thr Trp Gln Leu Ile Ser Asn Pro Ala Thr Thr

130 135 140
Gly Lys Val Ser Tyr His Phe Lys Asp Gly Ser Ser Gln Trp Trp Thr

145 150 155 160

Ala Val Gln Pro Arg Asn His Arg Asn Ala Ile Ala Lys Phe Glu Tyr

165 170 175

Arg Met Gly Gln Ala Ala Tyr Lys Val Ala Pro Arg Phe Ile Phe Asn

180 185 190

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

195 200 205

Thr Thr Asp Val Tyr Gly Asn Met Ile Val Asp Glu Asn Ile Val Leu

210 215 220

Gly Asp Asp Val Ile Arg Thr Gly Thr Gln Gln Phe Pro Tyr Met Ala

225 230 235 240

Pro Pro Gly Ser Ala Thr Pro Thr Pro Ser Pro Thr Pro Arg Pro Thr

245 250 255

Ile Pro Pro Glu Ser Leu Thr Glu His Val Trp Leu Pro Tyr Thr Arg

260 265 270

Lys

<210> 30

<211> 243

<212> PRT

<213> HcEXLX=Q2SG15: (Endoglucanase C-terminal domain/subunit from *Hahella chejuensis*)

<400> 30

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

1 5 10 15

Cys Val Arg Ala Ser Tyr Ala Glu Asn Arg Val Ser Ala Thr His Thr

20 25 30

Ser Ala Ala Leu His Glu Gly Glu Gly Thr Tyr Tyr Phe Tyr Asn Gly

35 40 45

Gly Gly His Cys Ser Val Pro Val Pro Ala Met Phe Thr Ala Ala Met

50 55 60

Asn Gln Thr Asp Tyr Asn Gly Ser Gln Ala Cys Gly Gly Cys Val Lys

65 70 75 80

Val Thr Asn Arg Asn Asn Gly Lys Ser Val Val Ala Arg Val Asp Asp

85 90 95

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

100 105 110

Phe Ala Gln Ile Ser Pro Leu Glu Ala Gly Arg Ile Pro Ile Ser Trp

115 120 125

Asp Tyr Val Pro Cys Asp Tyr Pro Ser Val Leu Leu Tyr Phe Met Glu

130 135 140

Gly Ser Ser Gln Trp Trp Thr Ala Val Gln Val Arg Glu Gln Arg Tyr

145 150 155 160

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

165 170 175

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

180 185 190

Met Gly Thr Gly Pro Phe Asp Phe Arg Ile Thr Asp Ile Tyr Gly His

195 200 205

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

210 215 220

Thr Gln Gln Gln Phe Pro Ser Met Gly Thr Ser Gly Val Ile Asn Gln

225 230 235 240

Ala Asp Lys

<210> 31

<211> 351

<212> PRT

<213> A1TSL6: (Rare lipoprotein A from *Acidovorax avenae* subsp)

<400> 31

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

1 5 10 15

Lys Ser Leu Leu Ser Val Phe Gly Arg Arg Cys Ala Leu Ala Trp Gly

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

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

Gln Val Gly Asn Glu Leu Ser Val Gln Gly Ala Ala Trp Asn Arg Thr
 325 330 335
 Leu Gln Pro Gly Ala Ser Ala Gln Phe Gly Phe Cys Ala Asn Arg
 340 345 350

<210> 32

<211> 727

<212> PRT

<213> Q9AF65: (Cellulase CelA from *Clavibacter michiganensis* subsp. *sepedonicus*)

<400> 32

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

Phe Phe Ala Leu Ala Val Gly Ala Thr Phe Gly Ala Phe Ala Ala Pro
 20 25 30

Ala Leu Ala Ala Thr Ala Ala Gly Thr Gly Ala Val Ala Ser Pro Pro
 35 40 45

Gly Trp Leu His Thr Ala Gly Gly Lys Ile Val Thr Ala Ser Gly Ala
 50 55 60

Pro Tyr Thr Ile Arg Gly Ile Ala Trp Phe Gly Met Glu Thr Ser Ser
 65 70 75 80

Cys Ala Pro His Gly Leu Asp Thr Ile Thr Leu Ala Gly Gly Met Gln
 85 90 95

His Ile Lys Gln Leu Gly Phe Thr Thr Val Arg Leu Pro Phe Ser Asn
 100 105 110

Gln Cys Leu Ala Ala Ser Gly Val Thr Gly Val Asp Ala Asp Pro Ser
 115 120 125

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

130 135 140
 Ala Lys Ala Ala Gly Leu Asp Val Ile Leu Asp Gln His Arg Pro Asp
 145 150 155 160

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

 225 230 235 240
 Val Glu Gly Ile Asp His Gln Ala Asp Gly Thr Gly Thr Trp Trp Gly
 245 250 255
 Gly Ala Leu Asp Ser Ala Ala Thr Ala Ser Val Arg Leu Thr Val Ala
 260 265 270
 Asn Arg Val Val Tyr Ser Pro His Asp Tyr Pro Ser Thr Ile Tyr Gly
 275 280 285
 Gln Pro Trp Phe Ser Ala Ser Asn Tyr Pro Thr Asn Leu Pro Gly Ile
 290 295 300

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

 370 375 380

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

Lys Thr Thr Ser Val Val Asn Ser Trp Gly Met Arg Cys Ser Val Ala
450 455 460
Ser Asn Thr Val Thr Cys Thr Ser Thr Asp Trp Ala Ser Lys Leu Ala
465 470 475 480
Ala Gly Gln Thr Val Arg Val Gly Val Gln Leu Ala Gly Gly Pro Ala
485 490 495
Pro Ala Ser Pro Arg Ile Ser Val Thr Ala Ala Gly Thr Pro Pro Ser
500 505 510
Gln Ala Thr Pro Pro Ser Gln Ala Thr Pro Pro Ser Gln Ala Thr Thr

515 520 525
His Gly Arg Ala Thr His Tyr Ser Leu Gly Thr Gly Asn Thr Ile Ala
530 535 540
Asn Gly Asn Cys Ser Met Pro Ala Val Pro Ala Asp Arg Met Tyr Val
545 550 555 560
Ala Val Ser Ser Pro Glu Tyr Gly Gly Ala Ala Ala Cys Gly Ser His
565 570 575
Leu Leu Val Thr Gly Pro Lys Gly Thr Val Arg Val Gln Ile Val Asp
580 585 590

Gln Cys His Glu Cys Glu Ile Gly His Leu Asp Leu Ser Glu Glu Ala
595 600 605
Phe Arg Ala Ile Gly Asp Phe Asp Ala Gly Ile Ile Pro Ile Ser Tyr
610 615 620
Thr Thr Val Arg Asp Pro Ser Val Pro Ala Val Ala Val Arg Val Lys
625 630 635 640

Glu Gly Ser Ser Arg Trp Trp Ala Gly Leu Gln Ile Leu Asn Ala Gly
645 650 655
Asn Arg Ile Asp Arg Val Glu Val Arg Ala Gly Gly Gln Trp Leu Pro
660 665 670
Leu Ser Arg Thr Asp Tyr Gly Tyr Trp Val Thr Pro Ser Pro Ile Pro
675 680 685
Asp Gly Pro Met Thr Val Arg Val Thr Asp Gln Tyr Gly Arg Ser Val
690 695 700
Val Leu Pro Gly Ile Arg Ile Ala Pro Gly Glu Thr Gln Ser Thr Ala
705 710 715 720
Arg Arg Ile Tyr Gln Met His
725

<210> 33

<211> 746

<212> PRT

<213> A5CLK3: (Cellulase from *Clavibacter michiganensis* subsp. *michiganensis*)

<400> 33

Met Arg Lys Ala Ser Val Glu Val Val Phe Ser Ala Ala Gly Trp Leu
1 5 10 15
Pro Tyr Asp Pro Tyr Met Thr Phe Arg Gln Val Arg Ala Ser Leu Val
20 25 30
Leu Arg Leu Val Leu Leu Leu Ala Leu Val Val Gly Thr Thr Ser Ala
35 40 45
Ala Phe Ala Ala Pro Val Ser Ala Ala Thr Val Ala Gly Pro Val Ala
50 55 60
Ala Ala Ser Ser Pro Gly Trp Leu His Thr Ala Gly Gly Lys Ile Val
65 70 75 80
Thr Ala Ser Gly Ala Pro Tyr Thr Ile Arg Gly Ile Ala Trp Phe Gly
85 90 95
Met Glu Thr Ser Ser Cys Ala Pro His Gly Leu Asp Thr Ile Thr Leu
100 105 110
Ala Gly Gly Met Gln His Ile Lys Gln Met Gly Phe Thr Thr Val Arg

115 120 125
 Leu Pro Phe Ser Asn Gln Cys Leu Ala Ala Ser Gly Val Thr Gly Val
 130 135 140
 Ser Ala Asp Pro Ser Leu Ala Gly Leu Thr Pro Leu Gln Val Met Asp
 145 150 155 160
 His Val Val Ala Ser Ala Lys Ser Ala Gly Leu Asp Val Ile Leu Asp
 165 170 175
 Gln His Arg Pro Asp Ser Gly Gly Gln Ser Glu Leu Trp Tyr Thr Ser
 180 185 190
 Gln Tyr Pro Glu Ser Arg Trp Ile Ser Asp Trp Arg Met Leu Ala Lys
 195 200 205
 Arg Tyr Ala Ala Glu Pro Thr Val Ile Gly Val Asp Leu His Asn Glu
 210 215 220
 Pro His Gly Ala Ala Thr Trp Gly Thr Gly Ala Ala Thr Thr Asp Trp
 225 230 235 240
 Arg Ala Ala Ala Glu Arg Gly Gly Asn Ala Val Leu Ala Glu Asn Pro
 245 250 255
 Asn Leu Leu Val Leu Val Glu Gly Ile Asp His Glu Ala Asp Gly Ser
 260 265 270
 Gly Thr Trp Trp Gly Gly Ala Leu Gly Leu Val Gly Asn Ala Pro Val
 275 280 285
 Arg Leu Ser Val Ala Asn Arg Val Val Tyr Ser Pro His Asp Tyr Pro
 290 295 300
 Ser Thr Ile Tyr Gly Gln Ser Trp Phe Ser Ala Ser Asn Tyr Pro Ala
 305 310 315 320
 Asn Leu Pro Gly Ile Trp Asp Ala His Trp Gly Tyr Leu Ala Lys Lys
 325 330 335
 Asp Ile Ala Pro Val Leu Val Gly Glu Phe Gly Thr Lys Phe Glu Thr
 340 345 350
 Thr Ser Asp Lys Gln Trp Leu Asn Thr Leu Val Gly Tyr Leu Ser Ser
 355 360 365

Thr Gly Ile Ser Ser Ser Phe Trp Ala Phe Asn Pro Asn Ser Gly Asp
370 375 380

Thr Gly Gly Ile Val Lys Ser Asp Trp Val Thr Pro Glu Gln Ala Lys
385 390 395 400

Leu Asp Ala Leu Ala Pro Ile Leu His Pro Ser Pro Gly Ser Gly Pro
405 410 415

Gly Ser Gly Gly Ser Gly Ser Gln Pro Ala Pro Gln Pro Asp Pro Ala
420 425 430

Asn Pro Gly Ala Val Ser Ala Lys Trp Gln Pro Gly Ser Ser Trp Ala
435 440 445

Ser Gly Tyr Val Ala Asn Ile Asp Val Thr Ala Thr Ala Ala Val Thr
450 455 460

Gly Trp Thr Val Ser Trp Ala Ser Pro Gly Thr Thr Arg Val Val Asn
465 470 475 480

Ser Trp Gly Met Arg Cys Ser Val Ala Ser Gly Thr Val Ser Cys Thr
485 490 495

Gly Thr Asp Trp Ala Ser Lys Leu Ala Ala Gly Gln Thr Val His Val
500 505 510

Gly Leu Gln Ala Ser Gly Gly Pro Ala Pro Ser Ser Pro Arg Leu Thr
515 520 525

Ala Thr Ala Ala Ala Val Pro Pro Ala Gln Pro Thr Pro Pro Ala Arg
530 535 540

Pro Thr Thr His Gly Arg Ala Thr His Tyr Ser Leu Gly Gln Gly Asn
545 550 555 560

Thr Ile Ala Asn Gly Asn Cys Ser Met Pro Ala Val Pro Ala Asp Arg
565 570 575

Met Tyr Val Ala Val Ser Ser Pro Glu Tyr Ser Gly Ala Ala Ala Cys
580 585 590

Gly Thr Phe Leu Asp Val Thr Gly Pro Lys Gly Thr Val Arg Val Gln
595 600 605

Val Ala Asp Gln Cys His Gly Cys Glu Val Gly His Leu Asp Leu Ser
610 615 620

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

625 630 635 640

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

645 650 655

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

660 665 670

Asn Ala Gly Asn Arg Ile Asp Arg Val Glu Ile Gln Ala Gly Arg Gln

675 680 685

Trp Leu Pro Leu Thr Arg Thr Asp Tyr Gly Tyr Trp Val Thr Pro Ser

690 695 700

Pro Ile Gln Asp Gly Pro Leu Thr Val Lys Val Thr Asp Gln Tyr Gly

705 710 715 720

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

725 730 735

Arg Thr Ala Ser Arg Phe Tyr Pro Val His

740 745

<210> 34

<211> 253

<212> PRT

<213> A5CR21: (Putative expansin from *Clavibacter michiganensis* subsp. *michiganensis*)

<400> 34

Met Thr Arg Ala Ser Ala Leu Pro Gly Arg Pro Arg Ala Pro Arg Arg

1 5 10 15

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

20 25 30

Leu Val Ala Val Ala Gly Pro Ala Met Ala Ala Ser Ala Ala Ser Ala

35 40 45

Ala Pro Ala Ala Gly Pro Ala Arg Val Ser Gly Tyr Ala Thr His Tyr

50 55 60

Ser Leu Gly Pro Asp Gly Arg Thr Thr Asn Gly Asn Cys Ser Leu Pro

65 70 75 80

Ala Ile Pro Lys Asp Arg Leu Tyr Val Ala Val Gly Pro Asp Leu Tyr
85 90 95

Ala Gly Gly Ala Gly Cys Gly Ser Tyr Phe Asp Val Thr Gly Pro His
100 105 110

Gly Thr Val Arg Val Glu Val Ala Asp Ser Cys His Glu Cys Val His
115 120 125

Gly His Leu Asp Leu Ser Glu Glu Ala Phe Arg Ala Ile Gly Asp Tyr
130 135 140

Asp Ala Gly Ile Ile Thr Thr Ser Tyr Val Pro Val Ala Ala Ser Thr

145 150 155 160

Val Pro Pro Leu Ser Phe Arg Phe Lys Asp Gly Ser Ser Ala Tyr Trp
165 170 175

Ala Ala Leu Gln Val Leu Asp Ala Gly Val Arg Leu Arg Ser Val Glu
180 185 190

Leu Trp Val Gly Ala Arg Trp Val Pro Leu Ser Leu Thr Asp Tyr Gly
195 200 205

Tyr Trp Leu Ala Pro Gly Tyr Val Gly Ala Gly Pro Phe Thr Val Arg
210 215 220

Val Thr Asp Thr Thr Gly Arg Thr Ala Thr Val Gln Gly Ile Val Leu
225 230 235 240

Asp Pro Met Arg Leu Gln His Thr Ala Ser Arg Leu Arg
245 250

<210> 35

<211> 232

<212> PRT

<213> 034918: (BsEXLX1=Expansin-yoaJ from Bacillus subtilis)

<400> 35

Met Lys Lys Ile Met Ser Ala Phe Val Gly Met Val Leu Leu Thr Ile
1 5 10 15

Phe Cys Phe Ser Pro Gln Ala Ser Ala Ala Tyr Asp Asp Leu His Glu

20 25 30

Gly Tyr Ala Thr Tyr Thr Gly Ser Gly Tyr Ser Gly Gly Ala Phe Leu

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

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

 165 170 175
 Glu Lys Met Asp Tyr Asn His Phe Val Ser Thr Asn Leu Gly Thr Gly
 180 185 190
 Ser Leu Lys Val Arg Met Thr Asp Ile Arg Gly Lys Val Val Lys Asp
 195 200 205
 Thr Ile Pro Lys Leu Pro Glu Ser Gly Thr Ser Lys Ala Tyr Thr Val
 210 215 220
 Pro Gly His Val Gln Phe Pro Glu
 225 230
 <210> 36
 <211> 220

 <212> PRT
 <213> B4ALC5: (YoaJ from Bacillus pumilus)
 <400> 36
 Met Leu Ser Phe Ala Val Pro Asn Met Lys Ala Ala Ser Val Leu Asp
 1 5 10 15

Glu Val Arg Glu Ser Tyr Ala Thr Tyr Thr Gly Ser Gly Tyr Ser Gly
 20 25 30
 Gly Ala Ala Leu Leu Asp Pro Ile Pro Ser Asp Met Lys Ile Thr Ala
 35 40 45
 Leu Asn Pro His Asp Tyr Asn Tyr Lys Gly Ile Asn Ala Ala Leu Ala
 50 55 60
 Gly Ala Tyr Leu Glu Val Lys Gly Pro Lys Gly Lys Thr Val Val Tyr
 65 70 75 80
 Val Thr Asp Lys Tyr Pro Glu Gly Ala Lys Gly Ala Leu Asp Leu Ser
 85 90 95
 His Asn Ala Phe Ala Lys Ile Gly Asn Met Ala Asp Gly Lys Ile Pro
 100 105 110
 Ile Ser Trp Lys Ile Val Lys Ala Pro Ile Ser Gly Asn Val Val Tyr
 115 120 125
 Arg Ile Lys Glu Gly Ser Ser Lys Trp Trp Ala Ala Ile Gln Val Arg
 130 135 140
 Asn His Lys Tyr Pro Ile Met Lys Met Glu Tyr Tyr Lys Asp Gly Gln
 145 150 155 160
 Trp Val Asn Met Glu Lys Thr Asp Tyr Asn His Phe Leu Gly Phe Asn
 165 170 175
 Met Gly Ser Lys Ser Leu Pro Val Arg Ile Thr Asp Ile Arg Gly Val
 180 185 190
 Ile Val Lys Asp Lys Leu Pro Ala Leu Pro Ser Thr Ala Ala Ser Gln
 195 200 205
 Ala Tyr Thr Val Lys Gly Asn Val Gln Leu Pro Asn
 210 215 220
 <210> 37
 <211> 666
 <212> PRT
 <213> BORNG4: (Glycosidase from Xanthomonas campestris)
 <400> 37
 Met Thr Ser Glu Gln Met Pro Ser Leu Ala Glu Arg Ala Leu His Gly

1 5 10 15
 Thr Gly Ser Val Ala Pro Arg Gln Ala Ala Gln Arg Pro Leu Arg Ala
 20 25 30

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

 100 105 110
 Thr Leu His Ile Arg Gly Val Thr Trp Pro Gly Phe Asp Arg Ala Gly
 115 120 125
 Leu Ala Ala Val Gly Met Arg Asn Asn Thr Leu Ala Gln Leu Leu Asp
 130 135 140
 Arg Met Gln Ala Ser Asp Ile Asn Ala Val Arg Val Pro Val Cys Ala
 145 150 155 160
 Ala Val Leu Gln Arg Ala Pro Val Ala Ala Ala Glu Val Ala Gly Asp
 165 170 175

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

 245 250 255

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

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

 385 390 395 400
 Ser Ala Phe Leu Gly Ser Trp Glu Thr Gly Asn Ala Asn Asn Gly Gly
 405 410 415
 Leu Leu Ala Pro Asp Gly Ser Pro Arg Ala Asp Lys Leu Leu Ile Leu
 420 425 430
 Arg His Ala Trp Gly Arg Leu Pro Val Met Pro Ala Ile Ala Thr Ala
 435 440 445
 Thr Gly Asp Ser Thr Lys Asn Ala Ser Gly Lys Lys Pro Trp Asn Ser
 450 455 460

 Thr Phe Thr Gly Thr Ala Thr Val Thr Gly Ser Gly Tyr Ser Gly Gly
 465 470 475 480
 Ala Leu Leu Leu Asp Pro Ile Pro Ser Asp Ala Phe Ile Thr Ala Leu
 485 490 495
 Asn Pro Val Gln Leu Asn Phe Gly Gly Val Lys Ala Ala Leu Ala Gly
 500 505 510

Ala Tyr Leu Gln Val Asn Gly Pro Lys Gly Thr Thr Thr Val Tyr Val
515 520 525
Thr Asp Leu Tyr Pro Glu Gly Ala Ser Gly Gly Leu Asp Leu Ser His
530 535 540
Asn Ala Phe Ala Ala Ile Gly Asp Met Val Gln Gly Arg Ile Pro Ile
545 550 555 560
Ser Trp Lys Val Val Arg Ala Pro Val Thr Gly Asn Leu Gln Tyr Arg
565 570 575
Ile Lys Glu Gly Ser Ser Arg Trp Trp Ala Ala Ile Gln Val Arg Asn
580 585 590
His Ala Tyr Pro Val Val Lys Leu Glu Val Lys Gln Gly Ser Thr Trp
595 600 605
Lys Asn Leu Gln Lys Met Asp Tyr Asn His Phe Leu Gly Glu Gln Leu
610 615 620
Gly Asn Gln Pro Leu Thr Leu Arg Ile Thr Asp Ile Arg Gly Lys Val
625 630 635 640
Leu Thr Asp Thr Leu Pro Arg Leu Pro Glu Asp Gly Ser Lys Pro Ala
645 650 655
Tyr Phe Glu Pro Gly His Val Gln Phe Pro
660 665
<210> 38
<211> 590
<212> PRT
<213> Q2NYR6: (Cellulase from Xanthomonas oryzae)
<400> 38
Met Leu Ser His Ile His Phe Leu Ile Ala Leu Ala Thr Leu Val Pro
1 5 10 15
Val Thr Ala Pro Ala Met Gln Val Ser Thr Gln Ala Pro Leu Val Asp
20 25 30
Ala Thr Gly Gln Thr Leu His Ile Arg Gly Val Thr Trp Pro Gly Phe
35 40 45
Asp Arg Ala Gly Leu Thr Ala Val Gly Met Arg Asn Asn Thr Leu Ala

50 55 60
 Gln Leu Leu Asp Arg Met Gln Ala Ser Asp Ile Asn Ala Val Arg Val

 65 70 75 80
 Pro Val Cys Ala Ala Val Leu Gln Arg Ala Pro Val Ala Ala Ala Glu

 85 90 95
 Val Ala Gly Asp Ser Thr Leu Arg Gly Leu Asp Ser Leu Gln Leu Leu

 100 105 110
 Asp Ala Val Val His Ala Ala Thr Gln Arg Gly Met His Val Met Phe

 115 120 125
 Ala Phe Ala Asp Gly Gly Cys Asp Asp Arg Ala Pro Leu Leu Gly Ala

 130 135 140

 Gln Gln Gln Ala Trp Thr Arg Gly Leu Val Thr Leu Ala Arg Arg Tyr
 145 150 155 160
 Gly Gly Asn Ala Asn Val Leu Gly Ile Asp Leu Gly Ser Ser Gly Tyr

 165 170 175
 Arg Asn Ala Ser Trp Ala Gly Asn Ala Ala Asp Gln Asn Trp Asn Arg

 180 185 190
 Val Ala Ser Arg Ala Ala Ala Met Val Leu Ala Gln Ala Pro Arg Trp

 195 200 205
 Val Val Gly Val Glu Gly Val Gly Ser Asn Ala Val Cys Ser Asp Pro

 210 215 220
 Gly Arg Lys Ala Leu Gly Ser Asn Leu Gln Pro Phe Ala Cys Val Pro
 225 230 235 240
 Leu Asp Ile Ala Pro Arg His Leu Val Leu Met Pro Lys Leu Ala Gly

 245 250 255
 Pro Asp Arg Asp Thr Thr Asp Ala Phe Ala Ala Pro Gly Phe Ala Gln

 260 265 270
 Ala Leu Pro Ala Met Trp Gln Arg Asp Phe Gly Gln Phe Ala Ile Asp

 275 280 285

 His Thr Val Val Pro Val Ser Leu Gly Gly Gly Leu Gly Asp Gly Asp
 290 295 300

Pro Arg Asp Pro Ala Trp Gln Thr Ala Leu Ser Gly Tyr Leu Ala Asn
305 310 315 320
Thr Gly Met Arg Ser Ala Phe Leu Gly Ser Trp Glu Thr Ser Asn Ala
325 330 335
Asn Asn Gly Gly Leu Leu Ala Pro Asp Gly Ser Pro Arg Ala Asp Lys
340 345 350
Leu Leu Ile Leu Arg His Ala Trp Gly Met Leu Pro Val Met Pro Ala
355 360 365
Ile Ala Thr Ala Thr Gly Asp Ser Thr Glu Pro Ala Ile Gly Lys Asn
370 375 380
Ala Trp Asn Ser Thr Phe Thr Gly Thr Ala Thr Pro Thr Gly Ser Gly
385 390 395 400
Tyr Ser Gly Gly Ala Val Leu Leu Asp Pro Ile Pro Ser Asp Ala Phe
405 410 415
Ile Thr Ala Leu Asn Ser Thr Gln Met Asn Phe Gly Gly Ile Lys Ala
420 425 430
Ala Leu Ala Gly Ala Tyr Leu Glu Val Thr Gly Pro Lys Gly Lys Thr
435 440 445
Thr Val Tyr Val Thr Asp Leu Tyr Pro Asp Gly Ala Ser Gly Gly Leu
450 455 460
Asp Leu Ser Tyr Asn Ala Phe Ala Ala Ile Gly Asn Leu Ala Asp Gly
465 470 475 480
His Ile Pro Ile Ser Trp Lys Val Val Arg Ala Pro Ile Thr Gly Asn
485 490 495
Val Gln Tyr Arg Ile Lys Glu Gly Ser Ser Arg Tyr Trp Ala Ala Ile
500 505 510
Gln Val Arg His His Ile Tyr Pro Val Val Lys Leu Glu Val Lys Gln
515 520 525
Gly Ser Thr Trp Thr Ser Leu Pro Lys Thr Pro Tyr Asn His Phe Val
530 535 540
Gly Thr Asn Leu Gly Asn Lys Pro Leu Ser Ile Arg Ile Thr Asp Ile
545 550 555 560

Arg Gly Lys Val Ile Ala Asp Lys Ile Pro Ala Leu Pro Glu Tyr Gly
565 570 575

Gly Ser Ala Ala Tyr Phe Glu Pro Gly His Val Gln Phe Pro
580 585 590

<210> 39

<211> 274

<212> PRT

<213> B4SGD8: (Endoglucanase c-terminal domain harboring protein from Ralstonia solanacearum)

<400> 39

Met Ser Cys Phe Lys Lys Trp Gln Trp Ile Ala Ile Met Leu Leu Met
1 5 10 15

Leu Pro Val Leu Ser Ala Arg Ala Asn Gly Gln Thr Thr Ala Ala Gly
20 25 30

Asn Val Ala Ala Ala Gln Gly Ser Cys Leu Gly Thr Gly Thr Ala Ala
35 40 45

Gly Asn Gly Ala Thr Thr Pro Ala Asn Ala Trp Asp Ser Thr Phe Thr
50 55 60

Gly Ile Ala Thr Ala Thr Gly Ser Gly Tyr Ser Gly Gly Ala Phe Leu
65 70 75 80

Leu Asp Pro Ile Ala Lys Asp Ala Glu Ile Thr Ala Leu Asn Pro Val
85 90 95

Gln Ala Asn Leu Gly Gly Ile Arg Ala Ala Met Ala Gly Ala Tyr Leu

100 105 110

Arg Val Gln Gly Pro Lys Gly Cys Thr Thr Val Tyr Val Thr Asp Leu
115 120 125

Tyr Pro Glu Ala Ala Ser Gly Gly Leu Asp Leu Ser Tyr Asn Ala Phe
130 135 140

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

Leu Ile Ala Ala Pro Val Thr Gly Asn Val Val Tyr Arg Ile Lys Glu
165 170 175

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

245 250 255
Pro Ser Pro Cys Ser Lys Pro Tyr Phe Val Gln Gly Asn Val Gln Phe
260 265 270
Pro Glu

<210> 40
<211> 256
<212> PRT
<213> Q8Y172:endoglucanase c-terminal domain harboring protein from *Ralstonia solanacearum*)
<400> 40

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

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

165 170 175
 Val Lys Leu Glu Val Phe Gln Gly Lys Ala Trp Val Ser Leu Pro Lys
 180 185 190
 Ala Asp Tyr Asn His Phe Val Gly Thr Gln Leu Gly Asp Lys Pro Leu
 195 200 205
 Val Ile Arg Ile Thr Asp Ile Arg Gly Arg Ile Leu Val Asp Lys Leu
 210 215 220
 Pro Pro Leu Leu Lys Asp Cys Thr Pro Gln Lys Ala Gly Glu Ala Ser
 225 230 235 240

Pro Cys Ser Lys Pro Tyr Phe Val Gln Gly Lys Val Gln Phe Ser Glu
 245 250 255

<210> 41
 <211> 584
 <212> PRT
 <213> B0U4Y6: (Extracellular endoglucanase from Xylella fastidiosa)
 <400> 41

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

35 40 45
 Glu Thr Ala Gln His Leu Pro Gln Gly Leu Asp Gln Arg Pro Trp Arg
 50 55 60

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

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

 180 185 190
 Pro Asp Trp Asn Arg Val Val Glu Arg Ala Ala Ala Ala Ile Leu Ala
 195 200 205
 Met Asn Pro Asp Trp Leu Ile Gly Val Gln Pro Val Gly Leu Asn Pro
 210 215 220
 Pro Cys Leu Asp Ala Ser Ala Pro Ile Ser Asp Asp Asn Ile Lys Ser
 225 230 235 240
 Gln His Cys Val Gln Leu Arg Ile Pro Ala Arg Asn Leu Leu Leu Met
 245 250 255

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

Gly Leu Leu Asp Asn Asp Gly Arg Thr Pro Arg Thr Ala Leu Ile Ala

325 330 335

Gln Leu His Arg Trp Trp Gly Val Ser Arg Val Asp Val Ala Ser Glu

340 345 350

Asn Ala Ala Thr Lys Asn Gln Thr Thr Thr Asp Thr Asn Gly Cys Val

355 360 365

Thr Gly Asp Ser Ser Val Pro Leu Asn Gly Trp Asp Thr Ser Phe Ser

370 375 380

Gly Val Ala Thr Tyr Thr Tyr Thr Gly Tyr Lys Gly Gly Ala Leu Met

385 390 395 400

Leu Asp Pro Ile Gln Ser His Val Gln Ile Thr Ala Leu Asn Pro Thr

405 410 415

Gln Leu Asn Leu Gly Gly Ile Pro Ala Ala Met Ala Gly Ala Tyr Leu

420 425 430

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

435 440 445

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

450 455 460

Ala Ser Ile Gly Asn Met Ala Gln Gly Arg Ile Pro Val Gln Trp Lys

465 470 475 480

Val Val Ser Ala Pro Val Ser Gly Asn Leu Ile Tyr Arg Val Lys Lys

485 490 495

Gly Ser Ser Gly Trp Trp Ala Ala Ile Gln Val Arg Glu His Arg Tyr

500 505 510

Pro Val Leu Lys Leu Glu Ile Cys Gln Asp Gly Thr Trp Leu Asn Leu

515 520 525

Pro Lys Arg Asn Tyr Asn Tyr Phe Val Gly Thr Arg Leu Gly Asn Gln

530 535 540

Pro Leu Ser Met Arg Met Thr Asp Ile Arg Gly Gln Thr Leu Ile Asp

545 550 555 560

Thr Leu Pro Ala Leu Pro Lys Lys Ala Ser Ser Lys Ala Tyr Ser Val

565 570 575
 Asn Gly Asn Val Gln Phe Ser Glu
 580
 <210> 42
 <211> 229
 <212> PRT
 <213> *Pectobacterium atrosepticum*
 <400> 42
 Met Asn Lys Ile Thr Ser Leu Ala Leu Ser Ala Leu Cys Val Ile Pro
 1 5 10 15

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

 85 90 95
 Tyr Pro Glu Gly Gly Asp Cys Ala Leu Asp Leu Ser Phe Asn Ala Phe
 100 105 110
 Glu Lys Ile Gly Asp Leu Arg Asp Gly Lys Ile Asn Ile Asp Trp Thr
 115 120 125
 Leu Ile Glu Ala Pro Val Asn Gly Asn Val Ile Tyr Arg Ile Lys Glu
 130 135 140
 Gly Ser Asn Pro Tyr Trp Ala Ala Val Gln Phe Arg Asn Val Lys Tyr
 145 150 155 160

 Pro Val Ile Glu Met Lys Tyr Met Arg Asp Asn Gln Trp Val Ala Ala
 165 170 175
 Gln Lys Thr Asp Tyr Asn His Phe Ile Val Glu His Val Gly Met Asn
 180 185 190

Asp Ile Pro Ile Glu Phe Thr Asp Val Lys Gly Asn Val Leu Ser Asp
 195 200 205
 Thr Leu Pro Pro Met Ser Gln Ser Thr Ser Ser Ala Tyr Leu Ile Thr
 210 215 220
 Gly Asn Val Gln Leu
 225

<210> 43
 <211> 232
 <212> PRT
 <213> Bacillus licheniformis
 <400> 43

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

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

165 170 175
 Glu Lys Thr Asp Tyr Asn His Phe Val Ser Thr Asn Leu Gly Thr Ser
 180 185 190
 Pro Leu Lys Val Arg Ile Thr Asp Ile Arg Gly Lys Val Val Lys Asp

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

225 230
 <210> 44
 <211> 251
 <212> PRT
 <213> Nakamurella multipartita
 <400> 44

Met Ala Thr Gly Val Glu Pro Ile Asn Pro Glu Phe Ser Leu Leu Thr
 1 5 10 15
 Glu Ser Thr Thr Pro Leu Ala Gly Ala Arg Arg Pro Arg Arg Arg Thr

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

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

Gly Ile Ile Pro Val Thr Tyr Ser Thr Val Lys Asn Pro Thr Val Pro
 145 150 155 160
 Ala Leu Asn Phe Arg Phe Lys Asn Gly Ser Ser Arg Trp Trp Phe Ala

 165 170 175
 Leu Gln Val Leu Asn His Gly Asn Arg Leu Gln Ser Val Ser Val Gln
 180 185 190
 Val Asn Gly Lys Trp Val Ala Ala Thr Leu Ala Asp Tyr Asn Tyr Trp
 195 200 205
 Ile Tyr Gln Pro Gly Ala Gly Pro Gly Pro Tyr Thr Leu Leu Val Lys
 210 215 220
 Asp Ile Tyr Gly Gln Gln Ala Ile Val Ser Gly Val Thr Met Ser Pro
 225 230 235 240

Thr Leu Val Gln Thr Thr Thr Ala Arg Leu Lys
 245 250

<210> 45
 <211> 303
 <212> PRT
 <213> Micromonospora sp.
 <400> 45

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

Ala Ala Gly Thr Ser Thr Pro Ala Val Gly Lys Ser Gly Pro Gly Phe
 50 55 60
 Pro Thr Ala Ala Ala Asp Ala Thr Thr Pro Ala Thr Thr Pro Thr Thr
 65 70 75 80
 Ala Arg Pro Ser Ala Thr Gly Ser Pro Ser Pro Lys Ser Ser Gly Val
 85 90 95
 Ala Pro Leu Ala Gly Arg Ile Arg Pro Asn Val Thr Tyr Arg Gly Lys

100 105 110
Ala Thr Phe Tyr Asp Ala Gly Asp Gly Gly Gly Ala Cys Leu Phe Gly

115 120 125
Pro Ala Ser Asp Leu Met Ile Gly Ala Met Asn Gln Thr Asp Tyr Glu

130 135 140
Ser Ala Lys Ala Cys Gly Ala Tyr Val Leu Val Lys Ala Ala Asn Gly

145 150 155 160
Asn Ser Val Thr Val Arg Ile Thr Asn Leu Cys Pro Leu Pro Cys Ala

165 170 175
Pro Gly Gln Ile Asp Leu Ser Pro Gln Ala Phe Ala Lys Leu Ala Asn

180 185 190
Arg Ser Leu Gly Glu Val Pro Ile Thr Trp Lys Leu Leu Ser Pro Ser

195 200 205
Met Ser Asp Thr Ile Ser Ile Arg Tyr Lys Val Gly Ser Ser Gln Trp

210 215 220
Trp Cys Gly Ile Gln Ala Ile Gly His Arg Asn Pro Val Ala Leu Leu

225 230 235 240
Glu Val Arg Thr Ser Ser Gly Trp Arg Gln Leu Pro Arg Pro Glu Tyr

245 250 255
Asn Tyr Phe Ile Ser Ala Asn Gly Ser Gly Cys Gly Gly Pro Ile Arg

260 265 270
Ile Thr Asp Ile Tyr Gly Glu Lys Leu Thr Ile Ser Gly Ile Lys Leu

275 280 285
Lys Pro Asp Val Val Gln Pro Thr Arg Val Gln Phe Ala Arg His

290 295 300
<210> 46

<211> 326
<212> PRT

<213> Catenulispora acidiphila
<400> 46

Met Lys Ala Thr His Pro Thr Thr His Arg Ser Arg Trp Leu Pro Ile

1 5 10 15

Ser Thr Ala Ala Gly Leu Leu Val Val Ala Gly Leu Ile Ala Gly Leu

20 25 30

Ala Met Gly Ser Ser Ser Arg His Thr Ala Asp Pro Pro Pro Ala Ala

35 40 45

Thr Gly Phe Ala Thr Thr Gly Ser Gly Met Thr Ala Thr Pro Asn Ser

50 55 60

Ala Pro Thr Ser Ala Glu Pro Ser Ala Pro Ala Val Ala Ser Gly Thr

65 70 75 80

Ala Ser Ser Ala Ser Ser Pro Thr Val Ala Ala Ala His Lys Thr Pro

85 90 95

Gly Lys Thr Ser Thr Thr Ser Gln Thr Ser Ala Thr Pro Thr His Gly

100 105 110

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

115 120 125

Gly Val Thr Tyr Gln Gly Val Ala Thr Glu Tyr Ser Ala Ala Asp Gly

130 135 140

Asp Gly Ala Cys Leu Phe Gly Pro Ala Ala Asp Met Met Ile Ala Ala

145 150 155 160

Met Asn Glu Leu Asp Tyr Gln Asn Ser Glu Ala Cys Gly Ala His Val

165 170 175

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

180 185 190

Glu Cys Pro Tyr Pro Cys Ala Pro Gly Gln Leu Asp Leu Ser Gln Gln

195 200 205

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

210 215 220

Trp Gln Leu Val Ser Pro Ala Ala Ala Gly Asn Val Ser Ile Arg Tyr

225 230 235 240

Lys Val Gly Ser Ser Gln Tyr Trp Cys Gly Ile Gln Val Ile Gly Glu

245 250 255

Arg Asn Pro Val Ala Arg Leu Glu Val Ala Ala Gly Ser Gly Trp Gln

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

 305 310 315 320
 Val Gln Phe Ser Arg His
 325
 <210> 47
 <211> 230
 <212> PRT
 <213> Dickeya zeae
 <400> 47
 Met Met Lys Leu Lys Ser Val Leu Leu Ala Ser Ser Val Leu Phe Leu
 1 5 10 15
 Pro Leu Thr Gln Gln Ala Gln Ala Ala Trp Glu Leu Gly Asp Ile Cys
 20 25 30
 Tyr Gly Tyr Ala Thr Ala Thr Gly Ser Gly Tyr Ser Gly Gly Ala Leu

 35 40 45
 Leu Leu Asp Pro Ile Pro Ser Asn Met Glu Ile Thr Ala Leu Asn Arg
 50 55 60
 Ala Gln Leu Asp Tyr Lys Gly Val Lys Ala Ala Leu Ala Gly Ala Tyr
 65 70 75 80
 Leu Lys Val Thr Gly Pro Lys Gly Ser Thr Ile Val Tyr Val Thr Asp
 85 90 95
 Leu Tyr Pro Glu Gly Gly Asp Cys Ala Leu Asp Leu Ser Phe Asn Ala
 100 105 110

 Phe Glu Lys Ile Gly Asn Leu Gln Asp Gly Lys Ile Asn Ile Gln Trp
 115 120 125
 Glu Leu Val Lys Ala Pro Val Ser Gly Asn Val Val Tyr Arg Val Lys
 130 135 140

Glu Gly Ser Asn Pro Tyr Trp Ala Ala Val Gln Phe Arg Asn Val Lys
 145 150 155 160
 Tyr Pro Ile Ile Glu Met Lys Tyr Leu Arg Gly Thr Gln Trp Val Ser
 165 170 175
 Ala Pro Lys Thr Asp Tyr Asn His Phe Ile Leu Glu Phe Val Gly Lys
 180 185 190
 Asn Asp Ile Pro Ile Glu Phe Thr Asp Ile Lys Gly Asn Ile Leu Ser
 195 200 205
 Asp Thr Leu Pro Pro Met Ser Asp Ser Thr Ser Ser Ala Tyr Leu Ile
 210 215 220
 Thr Gly Lys Val Gln Leu
 225 230
 <210> 48
 <211> 267
 <212> PRT
 <213> ExpB1, Q40638 from *Oryza sativa* subsp. *japonica* (Rice)
 <400> 48
 Met Ala Ser Ser Ser Leu Leu Leu Ala Cys Val Val Val Ala Ala Met

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

 Gly Asn Asp Pro Ile Phe Lys Asp Gly Lys Gly Cys Gly Ser Cys Phe
 85 90 95
 Glu Ile Lys Cys Ser Lys Pro Glu Ala Cys Ser Asp Lys Pro Ala Leu
 100 105 110
 Ile His Val Thr Asp Met Asn Asp Glu Pro Ile Ala Ala Tyr His Phe

115 120 125
Asp Leu Ser Gly Leu Ala Phe Gly Ala Met Ala Lys Asp Gly Lys Asp

130 135 140
Glu Glu Leu Arg Lys Ala Gly Ile Ile Asp Thr Gln Phe Arg Arg Val

145 150 155 160
Lys Cys Lys Tyr Pro Ala Asp Thr Lys Ile Thr Phe His Ile Glu Lys

165 170 175
Ala Ser Asn Pro Asn Tyr Leu Ala Leu Leu Val Lys Tyr Val Ala Gly

180 185 190
Asp Gly Asp Val Val Glu Val Glu Ile Lys Glu Lys Gly Ser Glu Glu

195 200 205
Trp Lys Ala Leu Lys Glu Ser Trp Gly Ala Ile Trp Arg Ile Asp Thr

210 215 220

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

225 230 235 240

Gly Glu Lys Ile Ile Ala Glu Asp Ala Ile Pro Asp Gly Trp Lys Ala

245 250 255

Asp Ser Val Tyr Lys Ser Asn Val Gln Ala Lys

260 265

<210> 49

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 49

gaaggagata taaggatggc atatgacgac ctgcatgaa

39

<210> 50

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 50

atgatggtaa tgggtgttcag gaaactgaac atggcc 36

<210> 51

<211> 31

<212> PRT

<213> Artificial Sequence

<220><223> primer

<400> 51

Cys Thr Thr Thr Thr Thr Cys Ala Gly Gly Cys Thr Ala Ala Ala Cys

1 5 10 15

Thr Gly Cys Thr Cys Ala Gly Cys Ala Cys Cys Ala Thr Cys Gly

20 25 30

<210> 52

<211> 70

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 52

atgtctgaac ctgacgactc cggctttact aacgatactg aaggcgaggt tcgtgctctg 60

ggtgaatttc 70

<210> 53

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 53

atgcaggtca gtacgcaagc 20

<210> 54

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> primer

<400> 54

gggaaactgt acgtggccg 19

<210> 55
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 55
 gaaaatcgag ttctgacgac tc 22
 <210> 56
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> primer
 <400> 56
 ttgtctgcc tgattaataa cgcc 24
 <210> 57
 <211>
 > 242
 <212> PRT
 <213> zea m1 from Zea mays(Zea mays EXP B4, PBD code:2hcz)
 <400> 57
 Lys Val Pro Pro Gly Pro Asn Ile Thr Thr Asn Tyr Asn Gly Lys Trp
 1 5 10 15
 Leu Thr Ala Arg Ala Thr Trp Tyr Gly Gln Pro Asn Gly Ala Gly Ala
 20 25 30
 Pro Asp Asn Gly Gly Ala Cys Gly Ile Lys Asn Val Asn Leu Pro Pro
 35 40 45
 Tyr Ser Gly Met Thr Ala Cys Gly Asn Val Pro Ile Phe Lys Asp Gly
 50 55 60
 Lys Gly Cys Gly Ser Cys Tyr Glu Val Arg Cys Lys Glu Lys Pro Glu
 65 70 75 80
 Cys Ser Gly Asn Pro Val Thr Val Tyr Ile Thr Asp Met Asn Tyr Glu
 85 90 95
 Pro Ile Ala Pro Tyr His Phe Asp Leu Ser Gly Lys Ala Phe Gly Ser
 100 105 110

Leu Ala Lys Pro Gly Leu Asn Asp Lys Ile Arg His Cys Gly Ile Met
 115 120 125

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

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