

Bacillus stearothermophilus β-Xylosidase 유전자의 염기 서열 결정 및 분석

오현주 · 최용진*

고려대학교 자연자원대학 유전공학과

Sequence Analysis of β-Xylosidase Gene from *Bacillus stearothermophilus*

Oh, Hyun-Ju and Yong-Jin Choi*

Department of Genetic Engineering, College of Agriculture, Korea University,
Seoul 136-701, Korea

Abstract — The nucleotide sequences of the *xylA* gene encoding β-xylosidase of *Bacillus stearothermophilus* and its flanking regions were determined. Three open reading frame(ORFs) were found, one of which(ORF1) appeared to code for the β-xylosidase. The 1830 base pair ORF1 encoded 609 amino acids starting from a TTG initiation codon. The molecular weight deduced from the nucleotide sequence(68 KD) was in agreement with that estimated by SDS-polyacrylamide gel electrophoresis of the purified enzyme(66 KD). The Shine-Dalgarno sequence(5'-AGGAGG-3') was found 11 bp upstream of the initiation codon. Further 15 bp upstream, there observed a potential transcription initiation signals. The putative -10 sequence(CATAAT) and -35 sequence(TTGTTA) coresponded closely to the consensus sequences for *Bacillus subtilis* RNA polymerase with major sigma factor. The guanine-plus-cytosine content of the coding region of the *xylA* gene was 56 mol%, while that of the third position of the codons was 63 mol%. Based on the comparison with the amino acid sequences of several other carbohydrate degrading enzymes, two conserved regions, possibly participating in the catalytic mechamism of β-xylosidase *xylA*, were identified in 278-298 and 329-350 regions of the translated *xylA* gene. The nucleotide sequence of the *xylA* was found to exhibit no homology to any other genes so far reproted.

Xylan은 hemicellulose의 주 구성 성분이 되고 있는 일종의 heteropolymer으로써 cellulose 다음을 차지하고 있는 자연계에서의 막대한 분포량, cellulose에 비해 비교적 용이하게 효소 분해될 수 있는 분해 특성, 최종 분해 산물인 xylose의 다양한 용도등은 물론이고 각종 hemicellulose 폐자원의 효과적인 활용으로 비롯되는 환경정화와 같은 부수적인 잇점(1)등을 고려해 볼 때 xylan 역시 cellulose 못지 않는 중요한 biomass 자원이라고 하겠다. 그러나 자연상태의 xylan은 cellulose와는 달리 매우 복잡한 구조를 지녀 β-1,4 polyxylose 주결합쇄에 acetyl, arabinosyl 및 glucuronoyl 등의 잔기가 결합된 다양한 측쇄를 가지고 있을 뿐만 아니라 phenolytic substituents를 통해 xylan 분자

상호간 또는 xylan 분자와 lignin 분자가 서로 cross-linking되어 xylan 분해계 효소작용에 대해 내성을 가지고 있다(2). 따라서 이와같은 복잡한 구조의 Natural xylan을 효율적으로 완전 가수분해하려면 상기 분자간의 cross-linking과 측쇄결합을 효과적으로 절단하는 ferulic acid esterase, acetyl xylan esterase, α-arabinofuranosidase 및 glucuronidase등의 효소와 β-1,4 주결합쇄 절단에 필수적인, 주 xylan 분해효소인 xylanase와 β-xylosidase등과 같은 여러가지 서로 다른 xylan 분해 관련 효소의 효율적인 협동작용에 의해서만 가능한 것으로 분석보고되고 있다(3).

본 연구실에서는 상기 각종 xylan 분해계 효소를 세포 내외로 다량생산하는 강력한 xylan 분해 균주인 *Bacillus stearothermophilus*를 분리하고 관계 효소의 효소적 일반 특성 내지는 관계 유전자를 분리하고 그 결과의 일부를 이미 보고한 바 있다(4, 5).

Key words: Sequencing, β-xylosidase, *Bacillus stearothermophilus*

*Corresponding author

본 연구에서는 *B. stearothermophilus*가 생산하는, 세균으로서는 아직 그 예가 보고된 바가 없는 것으로 조사되고 있는 Exo-xylanase로 분류할 수 있는 매우 특이한 β -xylosidase 생산 유전자 *xylA*(5)의 염기 서열을 결정하고 그 결과를 분석 보고한다.

재료 및 방법

사용균주 및 플라스미드

형질 전환 숙주로는 *E. coli* JM109(*recA1*, *endA1*, *gyrA96*, *thi*, *hsdR17*, *supE44*, *relA1*, (*lac-proAB*), *F*, *traD36*, *proAB*, *lacIqZM15*)를 사용하였다(6). 그리고 *B. stearothermophilus xylA* 유전자는 앞 연구에서 얻은 재조합 플라스미드인 pMG1으로부터 분리하여 사용하였으며(5) subcloning을 위한 vector로는 pUC119(7), 또 single-stranded DNA 분리를 위한 helper phage로는 M13K07을 이용하였다(8).

사용배지 및 배양조건

플라스미드 분리를 위한 *E. coli* 배양은 분리 방법에 따라 50 μ g/ml의 ampicillin을 첨가한 LB 배지(0.5% yeast extract, 1% NaCl, 1% tryptone), 또는 TB 배지(1.2% tryptone, 2.4% yeast extract, glycerol 4 ml/l, 0.17 M KH_2PO_4 , 0.72 M K_2HPO_4)를 사용하여 37°C에서 진탕배양하였다. 그리고 single-stranded DNA 분리를 위한 *E. coli* 숙주 세포의 배양은 ampicillin이 첨가된 2X YT 배지(1.6% trypton, 1% yeast extract, 0.5% NaCl)를 사용 37°C에서 진탕 배양하였다.

사용시약 및 효소

Agarose, lysozyme, TEMED, ammomium persulfate, acrylamide, N,N'-methylenebisacrylamide 등은 Sigma Co.에서, 그리고 기타 일반 시약류는 시판 일급 이상의 분석용을 사용하였다.

각종 제한 효소와 calf intestinal alkaline phosphatase(CIP), T4 DNA ligase는 한국 제철 화학(KOSCO) 또는 New England Biolab(NEB), [^{35}S]- α -dATP는 Ammersham(U.K)에서 구입하였다.

DNA sequencing kit는 United States Biochemical (USB) 회사 제품인 sequenase version 2.0을 구입, 사용하였다.

Plasmid DNA의 분리 및 형질전환

플라스미드 DNA는 alkaline lysis 처리에 의해 플라스미드 DNA를 분리한 다음 PEG(Poly ethylene glycol)를 사용하여 DNA를 침전 분리시키는 방법(9)을

이용하여 정제하였다. 형질전환은 Hanahan 방법(9)에 따라 준비한 competent cells을 이용하여 실시하였다.

효소 활성 검출

LB 고체 배지에 10 mM PNPX(p-Nitrophenyl xylopyranoside)를 200 μ l 도말, 건조시킨 다음 *E. coli* 시험균주를 37°C에서 약 20시간 획선 배양(streak culture)하여 황색의 생성여부를 관찰하여 판별하였다. 이 방법은 β -xylosidase가 PNPX에 작용하면 황색의 nitrophenol을 생성하는 촉매 반응의 원리를 이용한 활성 검출법이다(37).

Single stranded DNA의 분리

적절한 크기의 β -xylosidase 유전자 단편을 subcloning한 pUC119 재조합 플라스미드를 가지고 있는 *E. coli* JM109 균주 배양액 20 μ l를 2 ml 2X YT 배지에 접종하고 helper phage M13K07을 최종 농도가 4×10^8 pfu/ml이 되도록 첨가, 37°C에서 배양하였다.

배양 1시간 후 kanamycin 70 μ g/ml을 첨가하고 계속해서 14-18시간 배양한 다음 Sambrook 등의 소규모 분리 방법에 따라 single stranded DNA를 분리하였다(9).

염기 서열 결정

염기 서열 결정은 Sequenase version 2.0 Kit를 이용, Sanger 방법의 원리에 근거한 USB manual에 따라 실시하였다. Template로 double stranded DNA를 사용할 경우는 DNA 용액에 0.5 체적의 0.4 M NaOH와 0.4 mM EDTA 혼합용액을 첨가하여 37°C에서 30분간 처리하여 denaturation시킨 후 0.1 체적의 3 M NaOAc(pH 4.5-5.5)와 3 체적의 EtOH를 가하여 침전 분리한 DNA 3-5 μ g을 사용하였다. Single stranded DNA의 경우는 전항에 설명한 방법에 따라 분리한 DNA 1 μ g을 template로 사용하였다. primer로는 0.5 pmol의 17-mer universal primer와 0.35 pmol의 24-mer reverse primer를 사용하였다. 합성 반응이 끝난 각 반응물은 80°C에서 2분간 가열한 후 8.3 M Urea를 포함하는 polyacrylamide gel(40 \times 44 \times 0.035 cm)을 이용, 전기 영동하였다. 전기 영동이 끝난 gel은 10% methanol과 10% acetic acid 혼합용액에 15 분간 담가두어 urea를 제거한 후 건조시키고 X-ray film(Curix XP, Agfa)에 24시간 정도 노출시킨 뒤 현상하였다.

Computer analysis

염기 서열로부터 amino acid 서열의 해독, GC 함량,

codon usage 등의 분석을 위해서는 Nucleic Acid/Protein Sequence Analysis(IBM)(1989 version) 또는 DNASIS version 5.0 program을 이용하였다. 또한 *xylA* 유전자의 염기 서열과 아미노산 서열을 다른 유전자들과 비교, 분석하기 위해서는 BLAST network service를 이용하여 NCBI(National Center for Biotechnology Information)를 활용하였다.

결과 및 고찰

β -xylosidase 유전자의 염기 서열 결정

B. stearothermophilus xylA 유전자를 포함하고 있는 pMG1의 2.8 Kb 삽입 DNA 단편을 *Hind*III, *Pst*I, *Bam*HI으로 절단하여 Fig. 1에 표시되어 있는 바와 같이 600 bp의 *Hind*III-*Pst*I 단편, 200 bp의 *Bam*HI-*Bam*HI 단편, 300 bp *Bam*HI-*Pst*I 단편, 1.6 Kb의 *Pst*I-*Bam*HI 단편 등의 4개의 단편을 분리하였다. 이 중에서 200, 300, 600 bp 단편은 그 자체를 pUC119에 크로닝하여 직접 염기서열을 결정하였다. 그러나 1.6 Kb의 단편은 다시 *Sma*I, *Pst*I, 및 *Bam*HI을 동시 처리하여 약 300, 600, 700 bp의 세개의 단편을 얻었고, 또 이와 별도로 *Hinc*II, *Bam*HI, 및 *Pst*I을 동시에 처리하여 위의 세개의 단편과 부분적으로 중복되는 약 300, 350, 200, 250, 500 bp 등의 5개의 단편을 분리하고 이들 염기서열을 결정, 1.6 Kb 단편 전체 염기서열을 완결할 수 있었다. 반면 *Hind*III, *Pst*I, 및 *Bam*HI을 사용하여 처음에 분리한 600 bp, 300 bp, 200 bp 및 1.6 Kb 네 단편은 서로 중복되지 않는다. 따라서 상기 각 단편의 말단이 최소 20 bp 이상 중복될 수 있도록 Fig. 1에 표시한 A, B, C 세 primer를 이용하여 전체 2.8 Kb DNA 단편의 완전한 염기서열을 결정하였다(Fig. 2).

ORF 및 putative ribosome binding site 결정

앞 항에서 결정한 2.8 Kb DNA 단편 전체의 염기서열을 이용하여 ATG를 translation initiation codon으로 하는 ORF(open reading frame)를 검색한 결과 같은 reading frame의 크기가 서로 다른 2개의 ORFs(ORF2, ORF3)을 찾을 수 있었다(Fig. 3). ORF2(572-2361)는 596개 아미노산으로 구성, 추정 분자량이 67 KD인 polypeptide를 지시하며 ORF3(966-2361)은 465개의 아미노산, 분자량 52 KD의 polypeptide의 유전정보를 가지고 있다. 한편 재조합 프라스미드 pMG1-P(672-2811 단편 삽입)과 pMG1-HB(1-2316 단편 삽입) 어느 것도 β -xylosidase 활성을 나타내지 않는 것으로 분석되었다. 그러므로 β -xylosidase 활성을 위해서는 672번 염기의 윗쪽과 2316번 염기의

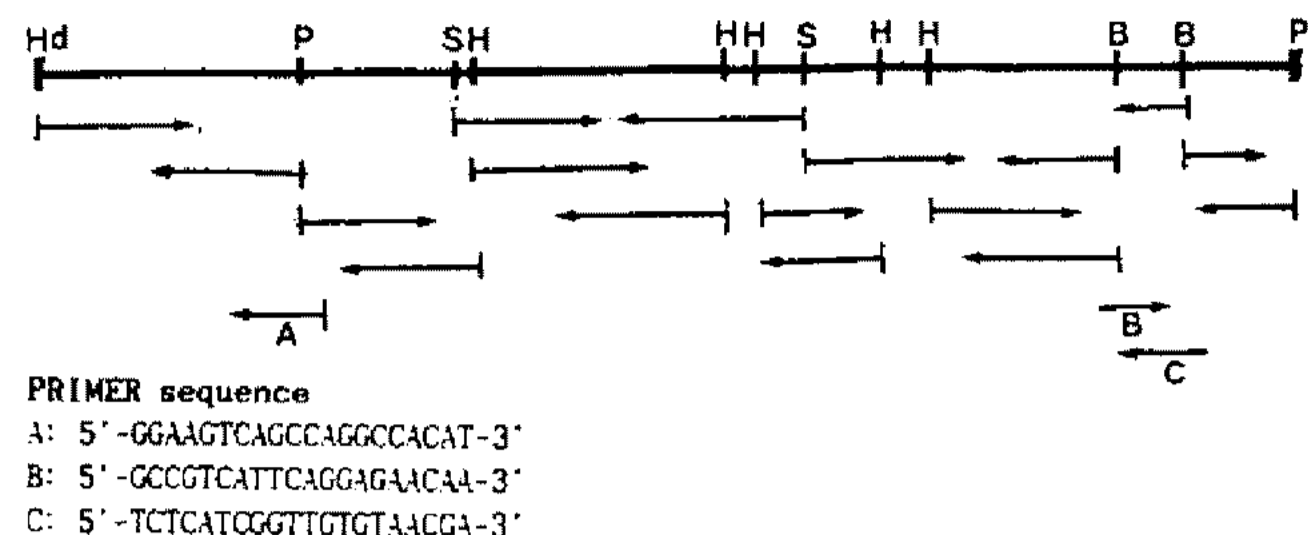


Fig. 1. Sequencing strategy of insert in pMG1.

Arrows indicate direction of sequencing. Abbreviations: Hd; *Hind*III, H; *Hinc*II, S; *Sma*I, P; *Pst*I, B; *Bam*HI.

아랫쪽 부분의 얼마간이 포함되어야 함을 알 수 있었다. 또한 정제 효소의 분자량이 66 KD로 측정되었으므로(5) ORF3은 완전한 β -xylosidase 유전자가 되지 못한다는 것을 쉽게 알 수 있었다. 반면 ORF2의 경우는 ORF3보다 위쪽으로 393 bp 더 길고 추정 분자량 역시 정제 단백질의 측정 분자량과 거의 같은 값을 보였다. 따라서 ORF2를 β -xylosidase 유전자의 완전한 coding region이라 추정하고 유전 정보의 전사와 번역의 시작과 관련된 각종 conserved sequence를 검색을 실시하였다.

그 결과 ORF2의 ATG codon으로부터 49 bp 윗쪽에 Shine-Dalgarno sequence(SD)로 추정되는 AG-GAGG 염기서열의 존재를 확인하였다. 이 부분의 염기서열을 *E. coli*와 *Bacillus subtilis*의 consensus ribosome binding site(RBS)와 비교해 본 결과(Table 1), *B. subtilis*와는 15개 중 11개가 일치하고 있었다. 그러나 *Bacillus*에서 ATG와 RBS 사이의 거리는 일반적으로 7~14 bp 정도로 보고되고 있는데 반해(10), ORF2의 경우는 49 bp를 나타내었다. 따라서 ORF2 역시 β -xylosidase 유전자의 실제 coding region이 아닐 가능성도 배재할 수 없었다. 따라서 AUG 이외에, 현재 다수의 예가 보고되고 있는 GUG, UUG 및 AUU 등을 initiation codon으로 가정하고 2.8 Kb DNA의 염기서열을 재검색해 본 결과 앞에서 설명한 SD 서열로부터 11 bp 아랫쪽에서 UUG codon을 발견하게 되었다. TTG codon을 initiation codon으로 했을 때의 ORF1(534-2363)은 ORF2와 같은 reading frame으로 609개 아미노산, 추정 분자량이 68 KD의 단백질을 생산할 수 있는 것으로 분석되었다.

현재로서는 *xylA* 유전자 산물의 N-말단 아미노산 서열을 결정하여 비교분석해 보지 않았기 때문에 TTG가 initiation codon이라고 단정할 수는 없다. 그러나 TTG와 SD sequence 사이의 거리와 *xylA* SD sequence와 *B. subtilis*의 consensus SD sequence 사이의 높은 유사성 등으로 미루어 보아 TTG가 ini-

1 AAGCTTGGCGAATGACCTGGCCAGCGACCCGTCGGGGTGTGGAGCGGGTTCTGCCA 60
 61 CGGCACAATAACTCGGAAGACGGGCTCATCGACGTTAAATGCAGGCAATAAAGTCA 120
 121 TCGGCTCCGTCGGAGCGCCGGCTTTCTGTCACATCCTCAGGGCCAGGAGCAGCAAAT 180
 181 CGCCGGGCTGCTGCACATAACGTTTCCACCAGCGGTCATCACCTGACTGCCTGCCAACA 240
 241 GCAAGTTCACCTCGAACAACGGGTGGCTGTGACGTGGATAAGGACACTCCGGGTAACC 300
 301 GTCCTGCAAAATGCTGTGGTAAAGAGTTTAAACGTGAATCATATTGGCAGCAATTTGGTA 360
 361 ACCTGATGCTCTCTGGCCACTCTCATCTCCGCTCTTCTAGATGATCCCGTCTCACC 420
 421 GCTGTGAACGATTGGATAAATCGATTGGCGGTTTGGACATGGCGCCCCCAAGCCT 480
 481 TGTTATAATCAAAGCTAGCATAATCGCTGTCAAGTGAACAGGAGGTAATCAACTTGGGT539
M G 2
 540 AAA AAT ATA TTT TTT AAC GCG CAT CAC TCT CCG ATG GGA TGT TCG 584
 3 K N I F F N A H H S P M G C S 17
 585 CCA GTA TTT ACG TTA GGT TAT CCC GGC CAA AGC GGC GGC TTC GAC 629
 18 P V F T L G Y P G Q S G G F D 32
 630 CTT GAG CTC GGC AGC CCG CGG AAC CAA AAC ATC TAC ATT GGC CTG 674
 33 L E L G S P P N Q N I Y I G L 47
 675 CAG GAC GAT GGG CAG GAG CGA TAC CGG GCT TTC CCT TTC TTC GGC 719
 48 Q D D G Q E R Y R A F P F F G 62
 720 GAA GGT CTG GAC GAG CGT GAT CGT TAC ACA ACC GAT GAG AGC GAT 764
 63 E G L D E R D R Y T T D E S D 77
 765 TCA GGC TCC AGC GAG CGA CAT TCC GTA CAG CAG GAA GAA AGC GGT 809
 78 S G S S E R H S N Q Q E E S G 92
 810 CTC ATC ATT CCG TAC AAC CGG GAC GAA ATC GTC CGC CGA TTT GGG 854
 93 L I I P Y N R D E I V R R F G 107
 855 GCA GCG ATC GAC GAA TGG CAA GCC GGA GAT TTG ACG TTC CGC CTG 899
 108 A A I D E W Q A G D L T F R L 122
 900 ATC TCC CCG TTT GAA GGC GTG CCC GAT CCC GAG ACA GCC ACG GAA 944
 123 I S P F E G V P D P E T A T E 137
 945 GAG CAG CTG CGG TTC GCG CTG ATG CCT GCT GTG CTT GCA GAG ATC 989
 138 E Q L R F A L M P A V L A E I 152
 990 ACG GTG GAC AAT ACC CGG GGA ACG TCA ACG CGC AAA GGC TTC TTT 1034
 153 T V D N T R G T S T R K A F F 167
 1035 GGC CTG CAA AAC AAC GAT CCT GTC AGC GCC ATC CGC CGT TTG GAA 1079
 168 G L Q N N D P V S A I R R L E 182
 1080 GAC GTA ACG GGC GGC CGG ATT TGC GGC ATC GGC CAA GGT CGC CAT 1124
 183 D V T G G R I C G I G Q G R H 197
 1125 ACC GCC ATT GCG CGA GCG ATC CGC GAT TAC CAG CGC CGG ATT TTT 1169
 198 T A I A R A I R D Y Q R R I F 212
 1170 TAC GAT GGA ATC GAT TCT CAA GCC AAA GGT TCC GGA AAA CCT CCG 1214
 213 Y D G I D S Q A K G S G K P P 227
 1215 GTT TGG CCT TGG TCC GGT AGG GCG GTG CTT ATG GAT ACC CCG GCC 1259
 228 V W P W S G R A V L M D T P A 242
 1260 GGC GAG AAA GGC ACT TAC CGC TTC GGC CTC TGC TTC TAC CGC CCG 1304
 243 G E K A T Y R F G L C F Y R P 257
 1305 GTT CGT CAC AAC CGG AAT CGA TGC TCC TAC TAC TAT TCG CGC TAT 1349
 258 V R H N R N R C S Y Y Y S R Y 272
 1350 TTC CCG AAC ATC GAA GAA GTA CCT GAG TAT GCG TTA GCG CAA TTT 1394
 273 F A N I E E V A E Y A L A Q F 287
 1395 AAC CGC AAG CTG GAG GTC GCC CGC CAA GCC GAC GCG CTT GTG AAC 1439
 288 N R K L E V A R Q A D A L V N 302
 1440 GAT GCC AAG CTC AGC GAG GAT CAG AAA TTT ATC GTC GCG CAT GCC 1484
 303 D A K L S E D Q K F I V A H A 317
 1485 ATC CGC AGC TAT TAC GGT TCC ACG GAG CTG CTG GAG CAT GAC GGC 1529
 318 I R S Y Y G S T E L L E H D G 332
 1530 AAG CCG CTC TGG ATT GTT AAC GAA GGC GAA TAC CGG ATG ATG AAT 1574
 333 K P L W I V N E G E Y R M M N 347
 1575 ACG TTT GAT TTG CTG GTT GAC CAG TTG TTT TAC GAG CTC AAA ATG 1619
 348 T F D L L V D Q L F Y E L K M 362
 1620 AAT CCC TGG ACG GTC AAA AAT GAG CTT GAC CTT TAC ACC AGC CGC 1664
 363 N P W T V K N E L D L Y T S R 377
 1665 TAT AGC TAT CGA GAT ACG GTT GAT TTC CCG GGA GAT GCA ACC GAA 1709
 378 Y S Y R D T V D F P G D A T E 392
 1710 TAT CCT GGC GGC ATC AGC TTT ACG CAT GAC ATG GGC GTG GCG AAC 1754
 393 Y P G G I S F T H D M G V A N 407
 1755 TGG TTT TCT CTT CCC GGC TAT TCG TCT TAT GAA CTG CAC GCC ATC 1799
 408 S F S L P G Y S S Y E L H A I 422
 1800 GAT GAC TGC TTC TCT CAT ATG ACG CAT GAG CAG CTC GTC AAC TGG 1844
 423 D D C F S H M T H E Q L V N W 437
 1845 GTG CTA AGC GCC ACC GTT TAC GTG CAT CAA ACC AAC GAC CGC GAA 1889
 438 V L S A T V Y V A Q T N D R E 452
 1890 TGG CTG GAA CCG CAA CCT GCC GAT TCT GGA GCA AGC CTG GAT AGC 1934
 453 W L E P Q P A D S G A S L D S 467
 1935 ATG GTC AAC CGC GAC CAT CCT GAC CCG TCC CAG CGT ACC GGC ATG 1979
 468 M V N R D H P D P S Q R T G M 482
 1980 ATG GGA CTC GAC AGC TCC CGT ACG ATG GGC GGA CGG AAA TCA CCA 2024
 483 M G L D S S R T M G G R K S P 497
 2025 CCT ATA CAG CCT GGA CGT CTC GTT GGG GCA AGC GCG GAA CAA CAT 2069
 498 P I Q P G R L V G A S A E Q H 512
 2070 CTA TTT GCC GGC AAG AGC TGG GCG GCC TAT GTA GCG ATG GAG AAG 2114
 513 L F A G K S W A A Y V A M E K 527
 2115 CTG TTT CGT GAC AAC GGC CGC AAA TCG GCG GCT GCA CTG GCC GGA 2159
 528 L F R D N G R K S A A A L A G 542
 2160 CGA CAA CCG GAG CTT GGC GCC GGC ACG ATT ACT TCC CAT CTG CTG 2204
 543 R Q P E L G A G T I T S H L L 557
 2205 CCG GGC GGT ATA TTC CAG CCG TCA TTC AGG AGA ACA ATG ACT CCA 2249
 558 P G G I F Q P S F R R T M T P 572
 2250 AGA TCA TCC CTG CGA TCG AGG GTC TTA TCT TCC CGC TAT ACA CCG 2294
 573 R S S L R S R V L S S R Y T P 587
 2295 GAT GCG AAG AAG CGC TGG ATC CAA ACG GCC GAT TCG GTG CAT ATA 2339
 588 D A K K R W I Q T A D S V H I 602
 2340 TTG CCG CGC TGT CCA CGC ATT TGA ACACGGTCTTCAGCCGGCATCTGCC 2390
 603 L R R C P R I *
 2391 TGTTGAGGACGGCGGCTGGAAGCTGTCTCAACGACCAACAACCTTTGGCTCAGCAAAA 2450
 2451 TTTATCTCTGCCAGTTTCATTGCCCGTGGATCCTTGGCTTGTCTGGGACGAACGGGGCC 2510
 2511 AAGCCGCCGATGCGGCATGTGGCCTGGCTGACTTCCGAAGCCGAACCGCTTACGGAGCTGG 2570
 2571 AGCGACCAGATGATGGCCGGCTGATTTGGCGTAGCCGGTATTATCCGTGGCTGTAACG 2630
 2631 GCGATTTTGTGGCTGGATGAGAAATAAACCCGGCCGCAACAATTAAGGAGCCCTTAGG 2690
 2691 GCGCGGCTTTCCGTTATACGTGCGTGGGGGGCGCTGATTGCGCCCGGCATCGTTCCG 2750
 2751 TTGATAACAAGCCCTCATCCTGGATTGAACAGCTCCTGCAACCGGATATCCGAGCCTGCA 2810
 2811 G

Fig. 2. Nucleotide sequence and deduced amino acid sequence for the 2.8-kbp insert in pMG1. Putative Shine-Dalgarno (S.D.) regions and initiation codon, and putative catalytic residues are underlined. The translational termination is indicated by an asterisk (*).

tiation codon일 가능성이 매우 높으며 따라서 ORF1이 실제 coding region이라고 판단된다.

실제로 McLaughlin 등(11)과 Moran 등(10)은 gram-positive 세균의 경우 translation initiation 효율과 관련해서 SD sequence의 상보성의 중요성을 강조하고 있으며 특히 McLaughlin등은 *S. aureus* β -lactamase 유전자의 UUG codon의 높은 효율은 codon과 anticodon 사이의 약한 base pairing을 16S rRNA와 SD sequence 사이의 강한 상호작용이 보충해 주기 때문이라고 주장하고 있어 본 *xylA* 유전자의 TTG initiation codon을 뒷받침 해 주고 있다.

Putative promoter sequence와 transcription terminator

추정 ORF1 SD sequence의 윗쪽 부분의 염기 서열을 *B. subtilis* RNA polymerase major sigma factor consensus promoter 서열과 비교해 본 결과 -35 element는 TTGTTA로 6개 염기 중 4개가 일치했고 -10 element는 CATAAT로 6개 염기 중 5개 염기가 일치되고 있다(Table 2). 그러나 현재 보고되고 있는 같은 균종의 *nprT*의 promoter와는 큰 차이를 나타

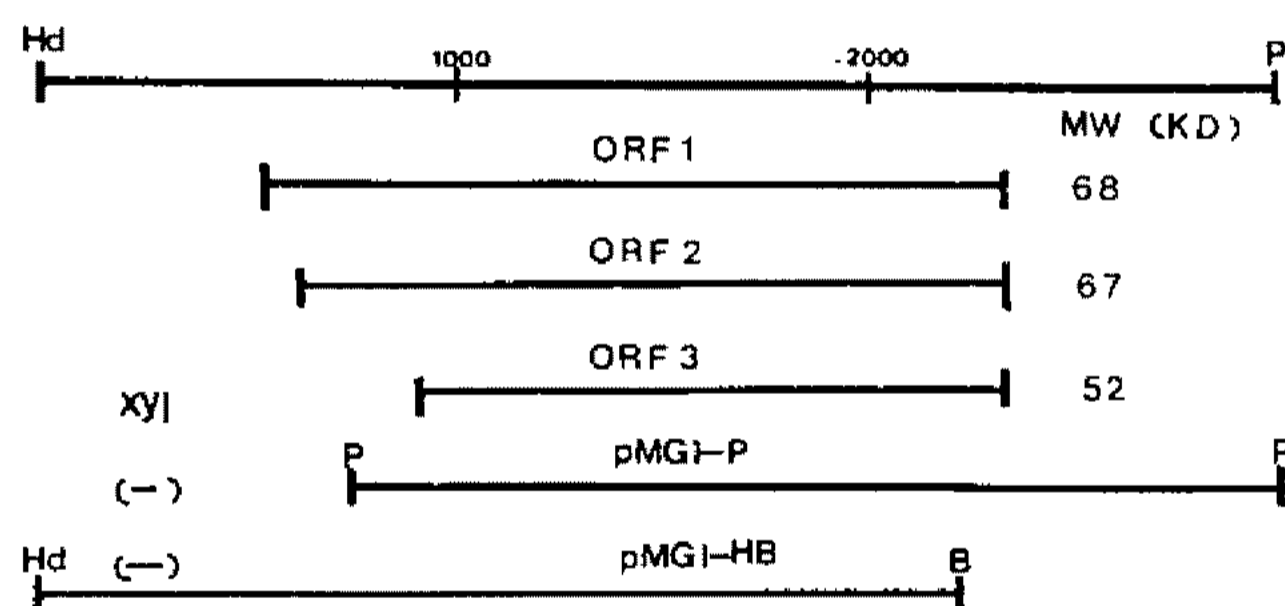


Fig. 3. Diagram showing ORFs and selected subclones with a qualitative evaluation of xylosidase by chromogenic indicator plates.

내고 있고 또 *B. subtilis*의 경우 -35 region과 -10 region 사이의 거리는 일반적으로 17~19 bp(13)로 보고되고 있으나 본 *xylA* 유전자의 경우 14 bp로 매우 드문 promoter 구조인 것으로 생각된다. 물론 그 예는 드물지만 *B. pumilus* β -xylosidase 유전자 promoter도 -35 region과 -10 region 사이의 거리가 14 bp라고 보고되고 있다(12).

또한 ORF1의 3'-말단 부분도 분석해 보았으나 ρ -independent transcription terminator로 인정할 만한 stem-loop 구조(13)는 발견할 수 없었다.

G+C 함량의 비교

ORF1의 G+C 함량은 56%, codon의 세 번째, 즉 Wobble position 염기의 G+C 함량은 63%로 산출되었다. 일반적으로 세균 genome의 GC 함량, 특히 Wobble position의 GC 함량과 세균의 최적 생육 온도와는 비례적인 상관관계를 나타내고 있다는 통계 보고가 있다. 50°C 전후의 최적 생육온도를 지닌 본 *B. stearothermophilus*의 *xylA* 유전자의 GC 함량 역시 Table 3에 표시되어 있는 바와 같이 moderate thermophile 균종의 GC 함량 범위내의 값을 나타내고 있어 GC 함량과 최적 생육 온도와의 관계를 입증해 주고 있다고 하겠다.

Codon usage

xylA 유전자 발현에 이용되고 있는 각 codon의 이용 빈도를 분석하여 *E. coli* 경우와 비교해 보았다. Table 4와 같이 본 *xylA* 유전자의 높은 GC 함량(56%)과 관계되는 특징으로 codon의 세 번째 염기가 G 혹은 C인 codon을 선호하는 경향이 *E. coli*의 경우와 특히 다른 점이라고 평가되었으며 기타 codon의 사용 빈도는 *E. coli*의 경우와 크게 다르지 않았다. 그러나

Table 1. Nucleotide sequences of putative ribosome binding sites

Organism	Nucleotide sequence	Reference
<i>E. coli</i>	CTAGTGGAGGAAT	10
<i>B. subtilis</i>	AGAAAGGAGGTGATC	10
<i>B. stearothermophilus xylA</i>	GAACAGGAGGTAATC	This study
<i>B. subtilis</i> 16S rRNA	3'UCUUUCCUCCACUAG 5'	10

Table 2. Promoter sequences in *Bacillus* sp.

Gene	-35 Region	-10 Region	spacer length	Reference
<i>B. subtilis</i> major σ factor consensus sequence	TTGACA	TATAAT	17-19	10
<i>B. stearothermophilus nprT</i>	TTTTCC	TATTTT	18	14
<i>B. stearothermophilus xylA</i>	TTGTTA	CATAAT	13	This study

Table 3. Comparison of GC content of the *B. stearothermophilus xylA* gene with those of other genes of thermophile and methophile

Strains	Gene	% G+C	% G+C the third base of codon	Reference
Moderate thermophile	<i>B. stearothermophilus xylA</i>	56	63	this study
	<i>B. stearothermophilus</i> neural protease	58	72	14
	α -Amylase	50	57	15
	Tyrosyl-tRNA synthetase	54	70	16
Extream thermophile	<i>T. thermophilus Isopropylmalate dehydrogenase</i>	79	89	17
Methophile	<i>B. subtilis</i> Neutral protease	44	42	18
	Subtilisin	46	41	19
	<i>B. amyloliquefaciens</i> Neutral protease	46	49	20
	Alkaline protease	49	46	20
	average of 64 <i>E. coli</i> genes	53	55	21

Table 4. Comparison of codon usage frequency

CODON	%AGE	CODON	%AGE	CODON	%AGE	CODON	%AGE
TTT-Phe	2.5% (1.3)	TCT-Ser	1.1% (1.3)	TAT-Tyr	2.0% (1.0)	TGT-Cys	0.3% (0.4)
TTC-Phe	2.3% (2.2)	TCC-Ser	2.0% (1.5)	TAC-Tyr	2.5% (1.5)	TGC-Cys	0.7% (0.5)
TTA-Leu	0.5% (0.7)	TCA-Ser	0.8% (0.4)	TAA-***	0.0%	TGA-***	0.2%
TTG-Leu	1.0% (0.9)	TCG-Ser	1.1% (0.6)	TAG-***	0.0%	TGG-Trp	1.5% (0.7)
CTT-Leu	1.3% (0.8)	CCT-Pro	1.6% (0.5)	CAT-His	2.1% (0.7)	CGT-Arg	1.3% (3.1)
CTC-Leu	1.5% (0.8)	CCC-Pro	1.0% (0.3)	CAC-His	0.5% (1.2)	CGC-Arg	3.4% (2.0)
CTA-Leu	0.3% (0.2)	CCA-Pro	0.7% (0.7)	CAA-Gln	2.1% (1.0)	CGA-Arg	1.3% (0.2)
CTG-Leu	3.0% (6.8)	CCG-Pro	2.5% (2.5)	CAG-Gln	2.0% (3.2)	CGG-Arg	1.6% (0.2)
ATT-Ile	1.3% (2.2)	ACT-Thr	0.5% (1.1)	AAT-Asn	1.0% (1.0)	AGT-Ser	0.0% (0.3)
ATC-Ile	2.6% (3.7)	ACC-Thr	1.5% (2.4)	AAC-Asn	2.6% (2.8)	AGC-Ser	2.8% (1.4)
ATA-Ile	0.7% (0.2)	ACA-Thr	0.7% (0.3)	AAA-Lys	1.6% (4.1)	AGA-Arg	0.3% (0.1)
ATG-Met	2.3% (2.8)	ACG-Thr	2.6% (0.8)	AAG-Lys	1.1% (1.3)	AGG-Arg	0.5% (0.1)
GTT-Val	1.1% (2.9)	GCT-Ala	0.7% (2.6)	GAT-Asp	3.6% (2.5)	GGT-Gly	1.5% (3.8)
GTC-Val	1.3% (1.2)	GCC-Ala	3.3% (2.2)	GAC-Asp	3.0% (3.0)	GGC-Gly	4.6% (3.1)
GTA-Val	0.8% (1.8)	GCA-Ala	1.0% (2.3)	GAA-Glu	2.8% (4.9)	GGA-Gly	1.8% (0.4)
GTG-Val	1.5% (2.2)	GCG-Ala	2.8% (3.2)	GAG-Glu	3.1% (1.8)	GGG-Gly	0.5% (0.6)

The number in parenthesis represents the condon usage frequency of *E. coli*(36). Three asterisks represent translation termination codon.

His의 경우는 CAC codon보다 CAT codon이 더 높은 빈도로 이용되고 있는 특징을 보이고 있다.

Putative catalytic site의 amino acid sequences의 비교

현재 hen egg white lysozyme(HEWL)(22)과 taka-amylase는 많은 연구 결과 그 촉매 기작이 잘 확립되어 있어 다른 탄수화물 분해 효소(carbohydrate hy-

drolases)의 작용기작연구의 대표적인 model로 이용되고 있다(22). 즉, HEWL 경우 Glu-35와 Asp-52 잔기가 촉매 반응에 직접 관계하는 일반적인 산-염기 촉매반응에 의해 가수 분해 반응을 촉매하고 있는 것으로 보고되고 있다(22). 실제로 지금까지 보고되고 있는 많은 각종 cellulase와 amylases의 촉매부위에 상기 산성 아미노산 잔기가 잘 보존되어 있으며(23). 또, Morosoli 등은 *Cryptococcus albidus*의 xylanase

Table 5. Amino acid sequence alignment of conserved regions

Organism or protein ^a	sequence			Reference
HEWL	35	44	52	25
	F E S N F N T Q A T . N R N T D G S . . . T D Y			
A.n.(glu)	331	338	349	26
	P E D T Y . Y N G . . N P W F L C T F A A A E Q			
B.f.xylB	342	350	362	27
	S E D F Y S L T D . . N T G F L R L K L R P E A			
B.f.end1	320	328	337	28
	G E T S A T N R N . . N T A E R V K W A . . D Y			
B.f.xynA	355	362	369	29
	N E K P L I W S . . . N I G C A K P A Y . . D E			
B.p.xynB	325	335	343	12
	I E C T R L A Q L N W N T C S M Q F V . . . E E			
B.t.xylA	278	288	298	This study
	E <u>E</u> <u>V</u> A E Y A L A Q F <u>N</u> R <u>K</u> . L E V A R Q A D A			
	329	339	350	
	L <u>E</u> H D G K P L W I V <u>N</u> E G E Y R M M N T F D L			
C.t.celA	458	466	475	30
	Q <u>E</u> <u>V</u> V N Y C I D . . N L <u>K</u> Y L R C M R . . D F			
C.t.xynZ	418	427	436	31
	G E A L L R A D V . . N R S G K V D S . . T D Y			
C.t.celB	130	138	149	32
	T E G G H P L L D L . N L K Y L F C M R . . D F			
C.t.celD	376	384	395	33
	D E E Y L R D F E . . N R A A Q F S K K E A D F			
C.s.xynB	408	418	427	34
	R E V F V E R I D E Y N A N P K R V W L . . E M			
T.r.CBH II	244	254	263	23
	L E C I N Y A V T Q L N L P N V A M Y L . . D A			
K.f.	586	596	605	
β -glucosidase				35
	G E W E T E G Y D R E N M D L P K R T N . . E L			
S.c.EG I	33	42	50	25
	N E S C A E F G N Q . N I P G V K N . . . T D Y			

^aAbbreviations: A.n., *Aspergillus niger*; glu, glucoamylase; B.f., *B. fibrisolvans* endoglucanase I, xylosidase; B.p., *Bacillus pumilus* xylosidase; B.t., *B. stearothermophilus*; C.t., *Clostridium thermocellum* endoglucanases; C.s., *Caldocellum saccharolyticum* xylosidase; T.R., *Trichoderma reesei* cellobiohydrolase II; K.f., *Kluyveromyces fragilis*; S.c., *Schizophyllum commune* endoglucanase I.

역시 동일한 동일한 촉매 기작에 의해 xylan을 분해하고 있다고 보고하고 있다(24). 따라서 본 *xylA* 유전자 산물의 아미노산 서열을 상기 HEWL을 비롯한 몇가지 탄수화물 분해 효소와 비교 분석해 보았다. Table 5와 같이 *xylA* 산물도 두 부분에서 conserved amino acid sequence를 발견할 수 있었으며 특히 Glu-278과 Asp-298 영역은 *Clostridium thermocellum* endoglucanase와는 22개 아미노산 중 8개가 동일하여 본 β -xylosidase의 촉매 부위일 가능성이 큰 부분이라 생각된다.

다른 유전자들과의 유사성

B. stearothermophilus xylA 유전자의 염기 서열을 1993년 10월 현재까지의 각종 data bank에 수록된 모든 DNA 염기 서열들과 비교해 본 결과 유사성 가진 유전자를 발견할 수 없었다. 또한 유전자 산물의 아미노산 서열의 경우도 앞항에서 설명한 catalytic region으로 추정되는 매우 짧은 서열이 보존되어 있을 뿐 전체적으로 볼 때 유사성이 인정되는 단백질은 없는 것으로 분석됨으로써 세균으로서는 아직 보고된 바 없는 극히 드문 본 유전자의 산물의 특성을 잘

설명해 주고 있다고 하겠다.

요 약

강력한 xylan 분해 균주인 *Bacillus stearothermophilus*의 β -xylosidase 유전자의 염기 서열을 결정, 분석하였다. 본 *xylA* 유전자는 1830 bp의 ORF을 가져 609 아미노산을 지시하며 TTG를 translation initiation codon으로 하고 있는 것으로 분석되었다. 염기 서열로부터 추정된 유전자 산물의 분자량은 68 KD로 SDS-polyacrylamide gel 전기 영동 결과로부터 산출한 분자량(66 KD)과 잘 일치되었다. Initiation codon 11 bp 윗쪽에 Shine-Dalgarno sequence로 추정되는 5'-AGGAGG-3'가 확인되었고 31 bp 윗쪽에는 promoter로 추정되는 염기 서열이 존재하며 -10 sequenc (CATAAT)와 -35 sequence(TTGTA)는 *Bacillus subtilis* RNA polymerase major sigma factor의 consensus sequence와 높은 유사성을 나타내고 있다. 한편 *xylA* 유전자의 GC 함량은 56%이며 codon의 세번째 염기의 GC 함량은 63%로 고온성 세균의 GC 함량의 일반적인 경향을 나타내었다.

또한 본 *xylA* 유전자 염기 서열에서 추정된 아미노산 서열을 다른 탄수화물 분해 효소와 비교 분석해 본 결과 촉매활성부위로 보고되고 있는 Glu-Asp/Glu 영역을 두 부분에서 발견하였으나 그 중 Glu-278과 Asp-298 보존 영역은 지금까지 보고되고 있는 다른 효소의 아미노산서열과 높은 유사성을 보였다. 최종적으로 *xylA* 유전자의 염기 서열을 기존에 발표된 다른 유전자와 비교해 본 결과 유사성을 인정할 수 있는 어떤 유전자도 발견할 수 없었다.

감사의 말

본 연구는 1993년도 교육부의 유전공학 연구비 지원에 의해 이루어진 것입니다.

참고문헌

1. Dekker, R.F.H. 1985. Biodegradation of hemicelluloses. Pp. 505-533. In T. Higuchi, *Biosynthesis and biodegradation of wood components*. Academic Press, Inc., Orlando, Fla.
2. Chanda, S.K., E.L. Hirst, J.K.N. Jones, and E.G.V. Percival. 1950. Constitution of Xylan from Esparto Grass. *J. Chem. Soc.* **12**: 39.
3. Ken K.Y. Wong., Larry U.L. and John N. Saddler. 1988. Multiplicity of β -1,4-xylanase in microorga-

- nisms: functions and applications. *Microbial. Rev.* **52**: 305-317.
4. Oh, S.W., S.S. Park, Y.I. Park, and Y.J. Choi. 1992. Molecular cloning and expression of *Bacillus stearothermophilus* β -D-xylosidase gene in *E. coli*. *Kor. J. Appl. Microbiol. Bioeng.* **20**: 136-142.
5. Mun, A.R. and Y.J. Choi. 1992. Purification and characterization of exo-xylanase from *Escherichia coli* cells harboring the recombinant plasmid pMG1. *Kor. J. Appl. Microbiol. Biotechnol.* **20**: 574-582.
6. Bolivar, F. and K. Backman. 1979. Plasmids of *Escherichia coli* as cloning vectors. *Meth. Enzymol.* **68**: 245.
7. Vieira, J. and J. Messing. 1982. The pUC plasmids, and M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**: 259.
8. Vieira, J. and J. Messing. 1987. Production of single-stranded plasmid DNA. *Meths Enzymol.* **153**: 3.
9. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1989. *Molecular Clonig, a laboratory manual*. Cold Spring Harbor Laboratory, New York.
10. C.P. Moran Jr., N. Lang, S.F.J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. **186**: 339-346.
11. J.R. McLaughlin, C.L. Murray, and J.C. Rabinowitz. 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β -latamase gene. *J. Biol. Chem.* **256**: 11283-11291.
12. W.Z. XU, Y. Shima, S. Megoro, and I. Urabe. 1991. Sequence and properties of β -xylosidase from *Bacillus pumilus* IPO. *Eur. J. Biochem.* **202**: 1197-1203.
13. B. Lewin. 1990. *GENES IV*, Cell Press, Cambridge.
14. M. Takagi, T. Imanaka, and S. Aiba. 1985. Nucleotide sequence and promoter region for the neutral protease gene from *Bacillus stearothermophilus*. *J. Bacteriol.* **163**: 824-831.
15. Nakajima, R. T. Imanaka, and S. Aiba 1985. Nucleotide sequence of a *Bacillus stearothermophilus* α -amylase gene. *J. Bacteriol.* **163**: 404-409.
16. Winter, G., G.L.E. Koch, B. Hartley, and D. Barker. 1983. The amino acid sequence of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*. *Eur. J. Biochem.* **132**: 383-387.
17. Kagaqa, Y., H. Nojima, M. Ishizuka, T. Makajima, T. Yasuhara, T. Tanaka, and T. Oishima. 1984. Highguanine plus cytosine content in the third letter of codons of an extreme thermophile.

- J. Biol. Chem.* **259**: 2956-2960.
18. Yang, M.Y., E. Ferrari, and D.J. Henner. 1984. Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. *J. Bacteriol.* **160**: 15-21.
 19. Stahl, M.L. and E. Rerrari. 1984. Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro-derived deletion mutation. *J. Bacteriol.* **158**: 411-418.
 20. Vasantha, M.L.D. Thompson, C. Rhodes, C. Banner, J. Nagle, and D. Filpula. 1984. Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain an large open reading frame between the regions coding for signal sequence and mature protein. *J. Bacteriol.* **159**: 811-819.
 21. Grosjean, H. and Fiers, W. 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective usage in efficiently expressed genes. *Gene* **18**: 199-209.
 22. Teeri, T.P. Lehtovaara, S. Kauppinen, I. Salovouri, and J. Knowles. 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzyme: gene sequence and expression of cellobiohydrolase. *Gene* **51**: 43-52.
 23. Rouvinen, J.T. Bergfors, T. Teeri, J.K. Knowles, and T.A. Jones. 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* **249**: 380-386.
 24. Morosoli, R., C. Roy, and M. Yaguchi. 1986. Isolation and partial primary sequence of a xylanase from the yeast *Cryptococcus albidus*. *Biochem. Biophys. Acta.* **870**: 473-478.
 25. Yaguchi, M.C. Toy, C.F. Rollin, M.G. Paice, and L. Jurasek. 1983. A fungal cellulase shows sequence homology with the active site of hen egg-white lysozyme. *Biochem. Biophys. Res. Commun.* **226**: 408-411.
 26. Boel, E., M.T. Hansen, I. Hjort, and N.P. Fiil. 1984. Two different types of intervening sequence in the glucoamylase gene from *Aspergillus niger*. *EMBO J.* **3**: 1581-1585.
 27. Utt, E.A., C.K. Eddy, K.F. Keshav, and L.O. Ingram. 1991. Sequencing and expression of the *Butyrivibrio fibrisolvens* xylB gene encoding a novel bifunctional protein with β -D-xylosidase and α -L-arabinofuranosidase activities. *Appl. Environ. Microbiol.* **57**: 1227-1234.
 28. Berger, E., W.A. Jones, D.T. Jones, and D.R. Woods. 1989. Cloning and sequencing of an endoglucanase(end1) gene from *Butyrivibrio fibrisolvens* H17c. *Mol. Gen. Genet.* **219**: 193-198.
 29. Mannarelli, B.M., S. Evans, and D. Lee. 1990. Cloning, sequencing, and expression of a xylanase gene from the anaerobic rumen bacterium *Butyrivibrio fibrisolvens*. *J. Bacteriol.* **172**: 4247-4254.
 30. Benguin, P., P. Cornet, and J.P. Aubert. 1985. Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. *J. Bacteriol.* **162**: 102-105.
 31. Grepinet, O., M.C. Chebrou, and P. Beguin. 1988. Purification of *Clostridium thermocellum* xylanase Z expressed in *Escherichia coli* and identification of the corresponding product in the culture medium of *C. thermocellum*. *J. Bacteriol.* **170**: 4576-4581.
 32. Fukusaki, E., W. Panbangred, A. Shinmyo, and H. Lkada. 1984. The complete nucleotide sequence of the xylanase gene(xylA) of *Bacillus pumilus*. *FEBS Lett.* **171**: 197-201.
 33. Joliff, G., P. Benguin, and J.P. Aubert. 1986. Nucleotide sequence of the cellulase gene celD encoding endoglucanase D of *Clostridium thermocellum*. *Nucleic Acids Res.* **14**: 8605-8613.
 34. Luthi, E.D.F. Love, J. McAnulty, C.Wallace, P.A. Caughey, D. Saul, and P. Bergquist. 1990. Cloning, sequence analysis, and expression of genes encoding xylan-degrading enzymes from the thermophilus *Caldocellum saccharolyticum*. *Appl. Environ. Microbiol.* **56**: 1017-1024.
 35. Raynal, A.C. Gerbaud, M.C. Francingurd, and M. Guerineau. 1987. Sequence and transcription of the β -glucosidase gene of *Kluyveromyces fragilis* cloned in *Saccharomyces cerevisiae*. *Curr. Genet.* **12**: 175-184.
 36. Allf-Steinberger, C. 1984. Evidence for coding pattern on the non-coding strand of the *Escherichia coli* genome. *Nucleic Acids Res.* **12**: 235-2241.
 37. Yu, J.H. Chung, K.S. Chung, and D.H. Oh. 1989. Molecular cloning and expression of a xylanase gene from thermophilic alkalophilic *Bacillus* sp. K-17 in *E. coli*. *Kor. J. Appl. Microbiol. Bioeng.* **17**: 178.

(Received December 21, 1993)