JOURNAL OF BACTERIOLOGY, Sept. 1987, p. 4399–4402
0021-9193/87/094399-04$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 169, No. 9

Nucleotide Sequence of the β-Cyclodextrin Glucanotransferase Gene of Alkalophilic *Bacillus* sp. Strain 1011 and Similarity of Its Amino Acid Sequence to Those of α-Amylases

KENJI KIMURA,[1,2] SHINSUKE KATAOKA,[1] YASUMASA ISHII,[2] TOSHIYA TAKANO,[3] AND KUNIO YAMANE[1]*

Institute of Biological Sciences, University of Tsukuba, Sakura,[1] and National Food Research Institute, Yatabe,[3] Ibaraki-ken 305, and Central Laboratory, Oji-Cornstarch Co. Ltd., Ichihara, Chiba-ken 290,[2] Japan

Received 19 February 1987/Accepted 1 June 1987

The nucleotide sequence of the gene for cyclodextrin glucanotransferase of alkalophilic *Bacillus* sp. strain 1011 was determined. The deduced amino acid sequence at the NH_2-terminal side of the enzyme showed a high homology with the sequences of α-amylase in the three regions which constitute the active centers of α-amylases.

α-Amylases and cyclodextrin glucanotransferases (CGTases) are able to cleave α-1,4-glucosidic bonds within the molecules of amylose and starch. The hydrolysis products from the substrates by α-amylases include glucose, maltose, and maltooligosaccharides. In contrast, CGTases degrade the substrates mainly to cyclodextrins, in which six to eight glucose units are joined by means of α-1,4-glucosidic bonds (2, 6, 8). Thus, α-1,4-glucosidic bonds can be reformed through the activity of CGTase, in addition to being cleaved by its amylase activity.

β-Cyclodextrin glucanotransferase (β-CGTase) from an alkalophilic bacterium, *Bacillus* sp. strain 1011, is an important enzyme for the food and pharmaceutical industries. The enzyme is capable of degrading starch to β-cyclodextrin, in which seven glucose units are joined by α-1,4-glucosidic bonds. To study the structure and mechanism of action of β-CGTase, we cloned the β-CGTase gene from the chromosomal DNA of *Bacillus* sp. strain 1011 and analyzed its nucleotide sequence. In this paper we describe the nucleotide sequence of the β-CGTase gene and the similarity of its amino acid sequence to the three common amino acid sequences, found in various α-amylases, which constitute the active centers of the enzymes.

The alkalophilic bacterium *Bacillus* sp. strain 1011, which was isolated from soil, produces the extracellular β-CGTase. The β-CGTase gene from the chromosomal DNA of strain 1011 was cloned in the *Escherichia coli* bacteriophage λD69 and recloned in the *E. coli* plasmid pBR322. The constructed plasmid, in which a 5.3-kilobase-pair (kb) DNA fragment had been inserted, was designated pTUE217 (Fig. 1) (5a). The plasmid was extracted by the rapid alkaline method of Birnboim and Doly (1) and purified by CsCl-ethidium bromide equilibrium centrifugation followed by agarose gel electrophoresis. The plasmid DNA and DNA fragments in the agarose gels were electroeluted onto hydroxyapatite (12). The culture medium and culture conditions were as described previously (19). Restriction enzymes, bacterial alkaline phosphatase, DNA polymerase I (Klenow fragment), and T4 DNA ligase were purchased from Takara Shuzo Co., Ltd., Kyoto, Japan, or Bethesda Research Laboratories, Gaithersburg, Md. Each enzyme was used according to the manufacturer's specifications.

Plasmid pTUE217 was stably maintained in *E. coli* HB101 cells, and the β-CGTase gene was efficiently expressed in the cells. More than 70% of the enzyme expressed was excreted into the culture medium. The major hydrolysis product from starch by degradation with the extracellular β-CGTase of *E. coli*(pTUE217) was β-cyclodextrin, as in the case of the enzyme from the parental *Bacillus* strain 1011 (5a). To determine the location of the β-CGTase gene in pTUE217, several deletion plasmids were constructed and their enzyme production was assayed by means of starch-hydrolyzing activity (13). The limit of the DNA region for the expression of the activity was approximately 2.5 kb (Fig. 1).

According to the strategy indicated in Fig. 1, approximately 3,000 base pairs (bp) in the inserted DNA, in which the limited DNA region was included, were determined. Both strands were sequenced. There was a unique open

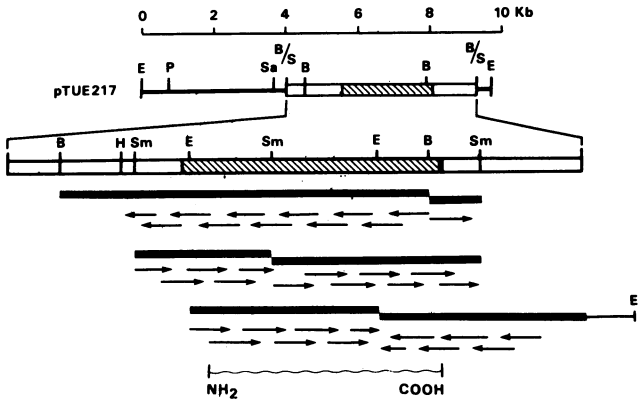

FIG. 1. Physical map of plasmid pTUE217 and strategy for determining the nucleotide sequence of the β-CGTase gene of *Bacillus* sp. strain 1011. ▭▨▭, 5.3-kb inserted DNA; ▨, 2.5-kb limited region for expression of starch-hydrolyzing activity; ———, pBR322 DNA. ▬, DNA fragments were cloned in the *Sma*I-site of the M13 phage vector mp10 (9) after being treated with DNA polymerase I (Klenow fragment) and dXTP if necessary. Their deleted DNA fragments were then prepared by the exonuclease III digestion method of Henikoff (3) for determination of the nucleotide sequences. →, Extent and direction of sequencing. Abbreviations: B, *Bam*HI site; E, *Eco*RI site; H, *Hin*dIII site; Sa, *Sal*I site; Sm, *Sma*I site; B/S, joint regions of *Bam*HI and *Sau*3AI sites.

* Corresponding author.

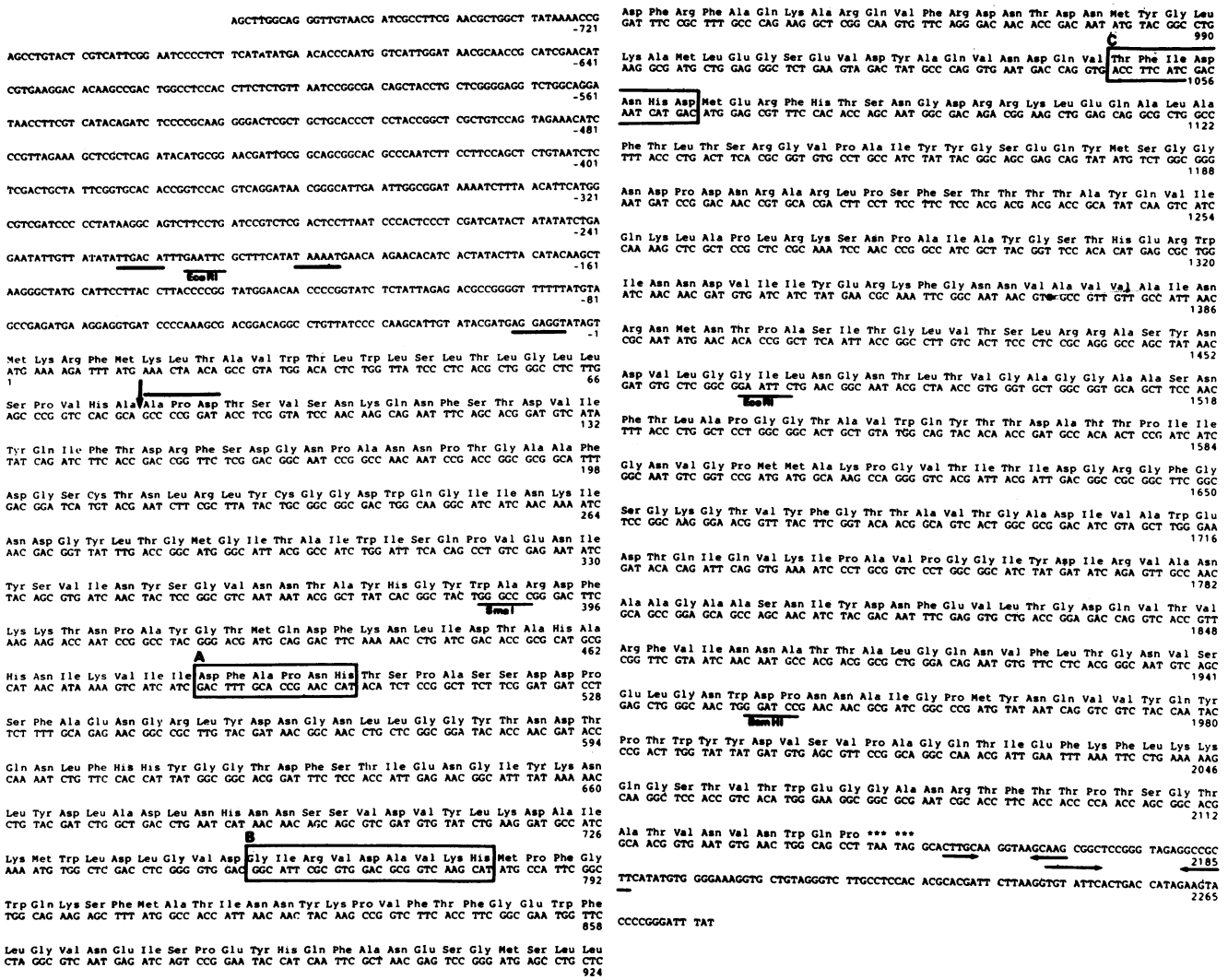

FIG. 2. Nucleotide sequence of the β-CGTase gene of the alkalophilic *Bacillus* sp. strain 1011. The nucleotides were determined by the dideoxy chain termination method of Sanger et al. (11) after the fragments to be sequenced were cloned and deleted according to the strategy shown in Fig. 1. Numbering of the nucleotide sequence begins at the most probable initiation codon, ATG (codon position 6 in the open reading frame). The noncoding DNA strand from 5′ to 3′ is shown with its corresponding amino acid sequence. The three amino acids at the NH_2-terminal region of the extracellular β-CGTase from strain 1011 are shown by a solid line above the sequence. The cleavage site between the signal peptide and extracellular mature β-CGTase is indicated by vertical arrow. The most probable RNA polymerase-binding site (TTGACA), the potential Pribnow box (TAAAAT), and a ribosome-binding site (AGGAGGT) are underlined. The sequences containing inverted repeat structures are designated by horizontal arrows. The three regions (A, B, and C) in which the deduced amino acid sequence of β-CGTase has a high homology with those of the α-amylases (Fig. 3) are boxed.

reading frame of 2,154 bp beginning with the ATG initiation codon and ending with two termination codons, TAA and TGA. Within this sequence, there were three potential initiator ATG codons (codon positions 1, 6, and 10, where position 1 corresponds to the first methionine of the open reading frame). Among them, the ATG codon at position 6 seemed to be the true initiator. There was a typical sequence for the ribosome-binding site (AGGAGGT) at 6 to 12 bp upstream from the ATG codon, and the amino acid sequence at the beginning of the ATG codon resembled a typical signal peptide (several positively charged amino acids followed by a run of hydrophobic amino acid core and a COOH-terminal alanine residue) (5). Furthermore, when the DNA fragment after the ATG codon at position 6 was fused downstream of the *tac* promoter in the *E. coli* expression vector pDR540 by using *Bam*HI linker DNA, the gene was well expressed (K. Kimura, T. Takano, and K. Yamane, unpublished data). Therefore, the ATG codon at position 6 was placed at nucleotides 1 to 3. The nucleotide sequence determined and the amino acid sequence deduced from the nucleotide sequence are illustrated in Fig. 2. The structural gene for the β-CGTase consisted of 2,139 bp (713 amino acids with a total molecular weight of 78,339).

The extracellular β-CGTase of *Bacillus* sp. strain 1011 was purified, and the amino acid sequence of its NH_2-terminal end was determined to be Ala-Pro-Asp by the microsequence method using phenylisothiocyanate (15). These amino acids corresponded to nucleotide positions 82 to 90 in the analyzed DNA sequence. These results suggested that the first 27 amino acids, from the initiator methionine to alanine, constitute a signal peptide involved in the secretion of proteins. Thus, the extracellular β-CGTase would be

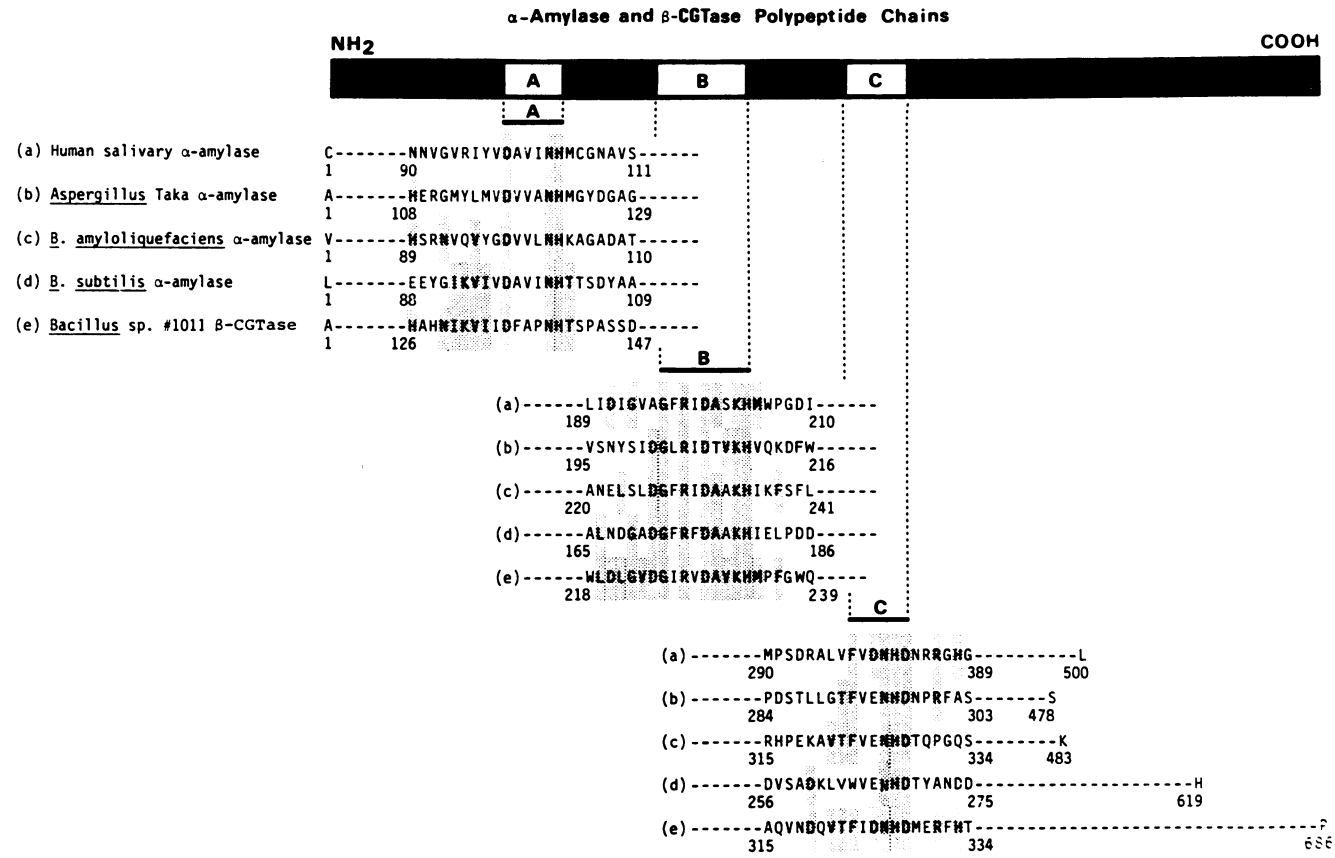

FIG. 3. Amino acid homology between β-CGTase of *Bacillus* sp. strain 1011 and α-amylases from the human salivary gland (10), *Aspergillus oryzae* (16), *Bacillus amyloliquefaciens* (14), and *B. subtilis* (20) at the three common regions which are considered to constitute the active centers of α-amylases from various origins. The amino acid residues identical with those of β-CGTase are shaded. The amino acid residues at the NH_2 termini, at and near the three common regions, and at the COOH termini of the enzymes are indicated. Numbering of the amino acid sequences of the enzymes starts at the NH_2-terminal amino acid of each mature enzyme.

composed of 686 amino acids with a total molecular weight of 75,225. The molecular weight of the purified β-CGTase was estimated to be 66,000 by sodium dodecyl sulfate disc gel electrophoresis. Upstream of the ATG initiation codon, the sequence for the ribosome-binding site was found at nucleotide positions −12 to −6 as described above. The sequence TTGACATTTGAATTCGCTTTCATATAAAAT (positions −225 to −196) seemed to be the most probable promoter for transcription. At the 3′ end, a short distance from the termination codon of the β-CGTase gene, there were palindromic sequences.

Amino acid sequences of many α-amylases from various sources (bacterial to human) have been determined or have been predicted from the DNA nucleotide sequences. Three amino acid sequences (A, B, and C regions in Fig. 3), which constitute the active centers of the enzymes, have been found in various α-amylases (4, 7, 16). The deduced amino acid sequence of β-CGTase was compared with these regions of representative α-amylases (Fig. 3). The three common regions, which are also recognized in the amino acid sequence of β-CGTase, are boxed in Fig. 2. The amino acid homology in the A regions of β-CGTase and the α-amylases was slightly low due to the presence of alanine and proline residues in this region of β-CGTase, whereas high homology was observed in the B and C regions. The arrangement of the A, B, and C regions from the NH_2 terminus of the β-CGTase polypeptide was the same as that in the α-amylases. Furthermore, the distances from the A to the B region and from the B to the C region of β-CGTase are also quite similar to those of human salivary gland and *Aspergillus* Taka α-amylases (10, 16). Other regions of β-CGTase did not show such a high homology with the sequences of α-amylases.

In contrast, the COOH-terminal region of the β-CGTase was completely different from those of the α-amylases. Many α-amylases are composed of approximately 500 amino acids, and an extracellular α-amylase of *Bacillus natto* IAM 1212 consists of only 436 amino acids (17). Therefore, the COOH-terminal region of β-CGTase would contain an extra 200 to 250 amino acids in addition to the polypeptide exhibiting the amylase activity. The sequence in the extra region showed a low similarity to that in the COOH-terminal side of the *B. subtilis* α-amylase, which also contains an extra COOH-terminal region (17, 18, 20). These findings suggest that β-CGTase may consist of two protein domains; the one in the NH_2-terminal side cleaves the α-1,4-glucosidic bond in starch, and the other in the COOH-terminal side catalyzes other activities, including the reconstitution of an α-1,4-glucosidic linkage for cycling the maltooligosaccharide produced. However, it is possible that the active sites used in breaking the α-1,4-glucosidic bond will participate in cyclization also.

This work was supported in part by a grant from the Ministry of Education, Science and Culture and in part by a grant from the

Biomass Conversion Project, Ministry of Agriculture, Forestry and Fisheries of Japan.

LITERATURE CITED

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
2. **French, D.** 1957. The Schardinger dextrins. Adv. Carbohydr. Chem. **12**:189–260.
3. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351–359.
4. **Ihara, H., T. Sasaki, A. Tsuboi, H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1985. Complete nucleotide sequence of a thermostable α-amylase gene: homology between prokaryotic and eukaryotic amylases at the active sites. J. Biochem. **98**:95–103.
5. **Inouye, M., and S. Halegoua.** 1980. Secretion and membrane localization of proteins in *Escherichia coli*. Crit. Rev. Biochem. **7**:339–371.

5a. **Kimura, K., T. Takano, and K. Yamane.** 1987. Molecular cloning of the β-cyclodextrin synthetase gene from an alkalophilic *Bacullus* and its expression in *Escherichia coli* and *Bacillus subtilis*. Appl. Microbiol. Biotechnol. **26**:149–153.

6. **Kobayashi, S., K. Kainuma, and S. Suzuki.** 1978. Purification and some properties of *Bacillus macerans* cycloamylose (cyclodextrin) glucanotransferase. Carbohydr. Res. **61**:229–238.
7. **Matsuura, Y., M. Kusunoki, W. Harada, and M. Kakudo.** 1984. Structure and possible catalytic residues of Taka-amylase A. J. Biochem. **95**:697–702.
8. **Matsuzaki, H., K. Yamane, K. Yamaguchi, Y. Nagata, and B. Maruo.** 1974. Hybrid α-amylases produced by transformants of *Bacillus subtilis*. I. Purification and characterization of extracellular α-amylases produced by the parental strains and transformants. Biochim. Biophys. Acta **363**:235–247.
9. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
10. **Nakamura, Y., M. Ogawa, T. Nishide, M. Emi, G. Kosaki, S. Himeno, and K. Matsubara.** 1984. Sequences of cDNAs from human salivary and pancreatic α-amylases. Gene **28**:263–270.
11. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
12. **Tabak, H. F., and R. A. Flavell.** 1978. A method for the recovery of DNA from agarose gels. Nucleic Acids Res. **5**:2321–2332.
13. **Takano, T., M. Fukuda, M. Monma, S. Kobayashi, K. Kainuma, and K. Yamane.** 1986. Molecular cloning, DNA nucleotide sequencing, and expression in *Bacillus subtilis* cells of the *Bacillus macerans* cyclodextrin glucanotransferase gene. J. Bacteriol. **166**:1118–1122.
14. **Takkinen, K., R. F. Petterson, N. Kalkkinen, I. Palva, H. Söderlund, and L. Kääriäinen.** 1983. Amino acid sequence of α-amylase from *Bacillus amyloliquefaciens* deduced from the nucleotide sequence of the cloned gene. J. Biol. Chem. **258**:1007–1013.
15. **Tarr, G. E.** 1977. Improved manual sequencing methods. Methods Enzymol. **47**:335–357.
16. **Toda, H., K. Kondo, and K. Narita.** 1982. The complete amino acid sequence of Taka-amylase A. Proc. Jpn. Acad. **58**:208–212.
17. **Yamane, K., Y. Hirata, T. Furusato, H. Yamazaki, and A. Nakayama.** 1984. Changes in the properties and molecular weights of *Bacillus subtilis* M-type and N-type α-amylases resulting from a spontaneous deletion. J. Biochem. **96**: 1849–1858.
18. **Yamane, K., T. Shiroza, T. Furusato, K. Nakamura, K. Nakazawa, K. Yanagi, M. Yamasaki, and G. Tamura.** 1985. *Bacillus subtilis* secretion vectors for proteins and oligopeptides constructed from *B. subtilis* α-amylase genes, p. 117–123. *In* J. A. Hoch and P. Setlow (ed.), Molecular biology of microbial differentiation. American Society for Microbiology, Washington, D.C.
19. **Yamane, K., Y. Takeichi, T. Masuda, F. Kawamura, and H. Saito.** 1982. Construction and physical map of a *Bacillus subtilis* specialized transducing phage ρ11 containing *B. subtilis LYS*$^+$ gene. J. Gen. Appl. Microbiol. **28**:417–428.
20. **Yang, M., A. Galizzi, and D. Henner.** 1983. Nucleotide sequence of the amylase gene from *Bacillus subtilis*. Nucleic Acids Res. **11**:237–249.