Vol. 151, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 29, 1988 Pages 25-31

NUCLEOTIDE SEQUENCE OF THE MALTOHEXAOSE-PRODUCING AMYLASE GENE FROM AN ALKALOPHILIC *Bacillus* sp. #707 AND STRUCTURAL SIMILARITY TO LIQUEFYING TYPE α-AMYLASES

Akira Tsukamoto[1,2], Kenji Kimura[1,3], Yasumasa Ishii[3], Toshiya Takano[4], and Kunio Yamane[1*]

[1] Institute of Biological Sciences, University of Tsukuba, Tsukuba, 305 Japan, [2] Central Research Laboratory, Oji-Paper Co. Ltd., Koutoh-ku, Tokyo, 135 Japan, [3] Central Laboratory, Oji-Cornstarch Co. Ltd., Ichihara, Chiba, 290 Japan and [4] National Food Research Institute, Tsukuba, 305 Japan

Received December 28, 1987

The nucleotide sequence of the gene for maltohexaose-producing amylase from an alkalophilic *Bacillus* sp. #707 was determined. Starting at an ATG initiation codon, an open reading frame was composed of 1554 bp (518 amino acids). The NH_2-terminal portion encoded a 33 amino acid-long signal peptide. The deduced amino acid sequence of the extracellular mature enzyme was more than 60 % homologous to those of the liquefying type α-amylases but not to those of the saccharifying type α-amylases. The sequence of its signal peptide was completely different from those of other α-amylases.

An alkalophilic bacterium, *Bacillus* sp. #707, produces mainly five kinds of enzymes exhibiting a starch hydrolyzing activity. Their molecular weights were estimated to be approximately 110, 95, 85, 75, and 60 kdal, respectively. Among the genes for the five enzymes, the gene for 60 kdal-amylase in the chromosomal DNA of the strain 707 was cloned into the *Escherichia coli* bacteriophage λD69 and recloned into the *E. coli* plasmid pBR322 and into the *B. subtilis* plasmid pUB110 originated from *Staphylococcus aureus*. The cloned gene was well expressed in both *E. coli* and *B. subtilis* cells. The major hydrolysis product from starch by the enzymes from those cells harboring the cloned gene was maltohexaose (1).

*To whom correspondence should be addressed.

0006-291X/88 $1.50

In order to analyze the structure of the maltohexaose-producing amylase (G6-amylase) in comparison with that of other bacterial α-amylases, the nucleotide sequence of the gene was determined.

In this paper we describe the DNA nucleotide sequence of the G6-amylase gene and the similarity of its amino acid sequence, deduced from the nucleotide sequence, to those of the liquefying type α-amylases.

MATERIALS AND METHODS

Bacterial strains and plasmids - The alkalophilic bacterium, Bacillus sp. #707, was isolated from soil. An Escherichia coli plasmid pTUE 306 was inserted a 4.4 kb DNA fragment encoding the G6-amylase gene into the plasmid pBR322. pTUE306 was maintained in E. coli HB101 (F^- hsdS20 recA13 ara-14 proA2 lacY1 galK2 repsL20 xyl-5 mlt-1 supE44 λ^-) in the presence of 50 μg/ml of ampicillin. The Bacillus subtilis plasmid pTUB812, 7.3 kb, was maintained and the G6-amylase gene was expressed in an α-amylase deficient mutant B. subtilis 207-25 (m^-_{168} hsrM recE4 amyE07 aroI906 leuA8 lys-21), a derivative of B. subtilis 168. The crude amylase preparation obtained from the culture medium (in the presence of 10 μg/ml kanamycin) of B. subtilis 207-25 harboring pTUB812 was the sole amylase component whose molecular weight was estimated to be 58,000 dalton. The same enzyme was also detected in the extracellular and periplasmic fractions of E. coli harbouring pTUB306 (1).

Preparation of plasmids and DNA fragments - The plasmids were prepared by the rapid alkaline method of Birnboim and Doly (2) and purified by CsCl equilibrium centrifugation in the presence of ethidium bromide followed by agarose gel electrophoresis. DNA fragments and plasmids in agarose gels were electroeluted into hydroxyapatite (3).

NH_2-terminal amino acid sequence determination - NH_2-terminal amino acid sequence of G6-amylase was determined by the micro-sequence method using phenylthiocyanate (4).

Enzymes and chemicals - Restriction enzymes, bacterial alkaline phosphatase, T4-DNA ligase, exonuclease III and mung bean nuclease were purchased from Takara Shuzo Co. Ltd. (Kyoto, Japan), or from Bethesda Research Lab. (Maryland, USA). Each enzyme was used according to the manufacturer's specification. Ampicillin and kanamycin were from Meiji Seika Co. Ltd. (Tokyo, Japan). All other chemicals were of reagent grade.

RESULTS AND DISCUSSION

Nucleotide sequence of the G6-amylase gene - The plasmids pTUE306 and pTUB812 were stably maintained in E. coli HB101 and B. subtilis 207-25, respectively, and the G6-amylase gene was efficiently expressed in both cells. More than 60 % and 95 % of the enzyme expressed in E. coli and B. subtilis was excreted into the culture media (1). To determine the location of the G6-amylase gene, several deletion plasmids of pTUE306 were constructed and their enzyme production was assayed by measuring the starch-hydrolyzing activity. The limit of the DNA region for the expression of the activity was approximately 2.5 kb (Fig. 1).

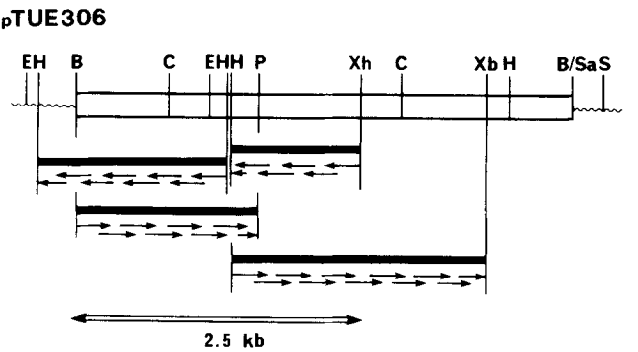
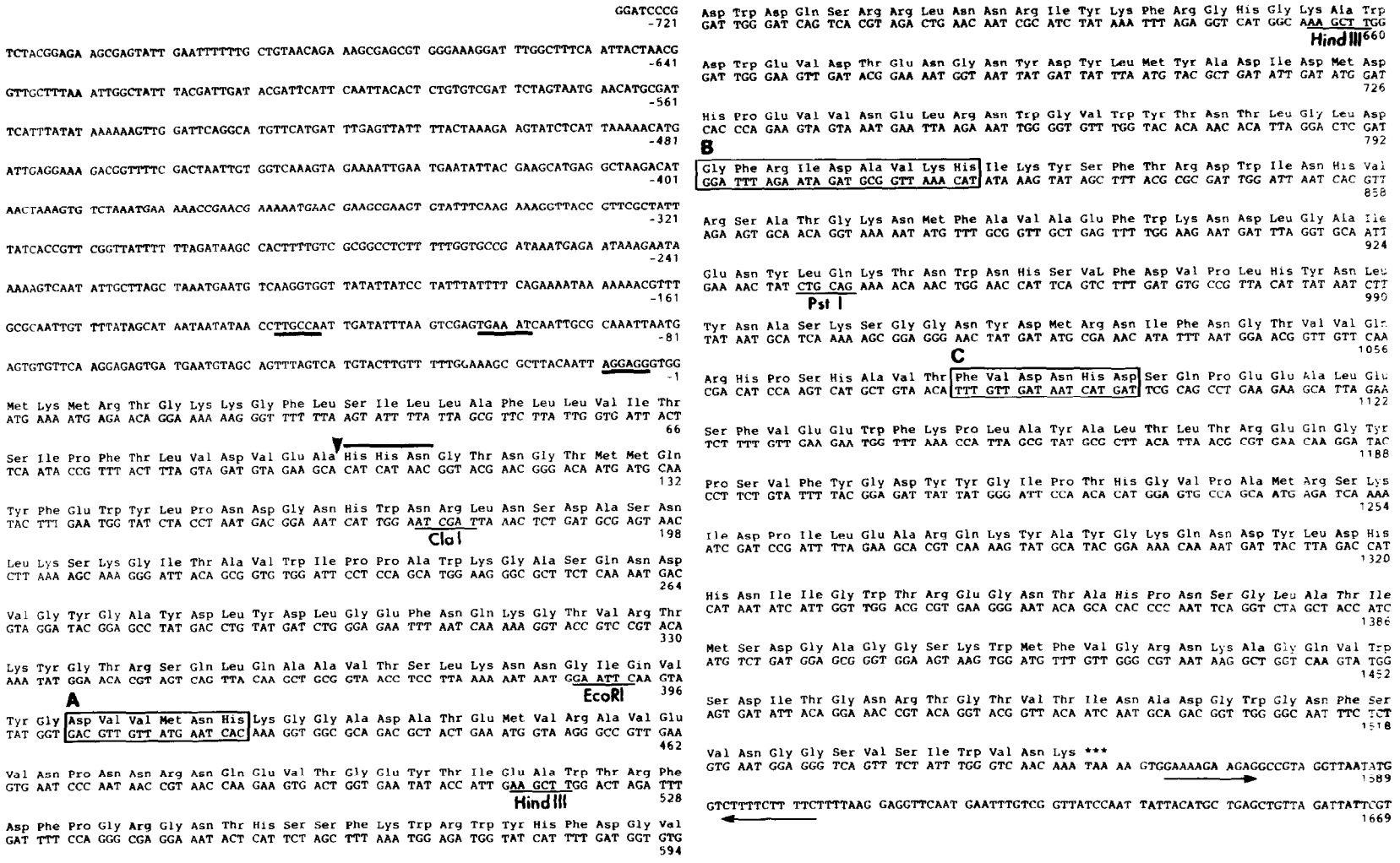


Fig. 1. Physical map of plasmid pTUE306 and strategy for determining the nucleotide sequence of the G6-amylase gene of Bacillus sp. #707. □, 4.4-kb inserted DNA. ∿∿∿∿, pBR322 DNA. DNA fragments (▬) were cloned in the SmaI-site of the M13 phage vector mp11 (12), after being treated with DNA polymerase I (Klenow fragment) and dXTP. The deleted DNA fragments were then prepared by the exonuclease III digestion method of Henikoff (13) for the determination of the nucleotide sequences. ⟼, Extent and direction of sequencing. ⟺, 2.5-kb limited region for the expression of starch-hydrolyzing activity. B, BamHI site; C, ClaI site; E, EcoRI site; H. HindIII site; P, PstI site; S, SalI site; Xb, XbaI site; Xh, XhoI site; and B/Sa, joint region of BamHI and Sau3AI.

According to the strategy indicated in Fig. 1, approximately 3000 bp in the inserted DNA, into which the limited DNA region was included, was analyzed (Fig. 2). Both strands were sequenced. There was a unique open reading frame of 1,554 bp begining with the ATG codon at the nucleotide positions +1 to +3, and ending with a termination codon TAA at nucleotide 1,555 to 1,557 (518 amino acids with a molecular weight of 59,007.5 dalton).

The G6-amylase of B. subtilis 207-25 harboring pTUB812 was purified and the amino acid sequence of its NH_2-terminal end was determined to be His-His-Asn. These amino acids corresponded to the nucleotide positions +100 to +108 in the analyzed sequence. These results suggested that the first 33 amino acids, from the initiator Met to Ala, constituted a signal peptide involved in the secretion of the exported proteins. The deduced amino acid sequence of this signal peptide is a typical bacterial signal peptide (several positively charged amino acids, followed by a hydrophobic amino acid core and a COOH-terminal Ala residue). Thus the extracellular G6-amylase would be composed of 485 amino acids giving molecular weight of 55,372 dalton. Upstream of the ATG initiation codon, the sequence for the ribosome binding site, AGGAGG, was found at nucleotide positions -11 to -5. The sequence, TTGCCAATTGATATTTAAGTCGAGTGAAAT


GGATCCCG -721
TCTACGGAGA AGCGAGTATT GAATTTTTTG CTGTAACAGA AAGCGAGCGT GGGAAAGGAT TTGGCTTTCA ATTACTAACG -641
GTTGCTTTAA ATTGGCTATT TACGATTGAT ACGATTCATT CAATTACACT CTGTGTCGAT TCTAGTAATG AACATGCGAT -561
TCATTTATAT AAAAAAGTTG GATTCAGGCA TGTTCATGAT TTGAGTTATT TTACTAAAGA AGTATCTCAT TAAAAACATG -481
ATTGAGGAAA GACGGTTTTC GACTAATTGT GGTCAAAGTA GAAAATTGAA TGAATATTAC GAAGCATGAG GCTAAGACAT -401
AACTAAAGTG TCTAAATGAA AAACCGAACG AAAAATGAAC GAAGCGAAGT GTATTTCAAG AAAGGTTACC GTTCGCTATT -321
TATCACCGTT CGGTTATTTT TTAGATAAGC CACTTTTGTC GCGGCCTCTT TTTGGTGCCG ATAAATGAGA ATAAAGAATA -241
AAAAGTCAAT ATTGCTTAGC TAAATGAATG TCAAGGTGGT TATATTATCC TATTTATTTT CAGAAAATAA AAAAACGTTT -161
GCGCAATTGT TTTATAGCAT AATAATATAA CCTTGCCAAT TGATATTTAA GTCGAGTGAA ATCAATTGCG CAAATTAATG -81
AGTGTGTTCA AGGAGAGTGA TGAATGTAGC AGTTTAGTCA TGTACTTGTT TTTGGAAAGC GCTTACAATT AGGAGGGTGG -1
Met Lys Met Arg Thr Gly Lys Lys Gly Phe Leu Ser Ile Leu Leu Ala Phe Leu Leu Val Ile Thr
ATG AAA ATG AGA ACA GGA AAA AAG GGT TTT TTA AGT ATT TTA TTA GCG TTC TTA TTG GTG ATT ACT 66
Ser Ile Pro Phe Thr Leu Val Asp Val Glu Ala His His Asn Gly Thr Asn Gly Thr Met Met Gln
TCA ATA CCG TTT ACT TTA GTA GAT GTA GAA GCA CAT CAT AAC GGT ACG AAC GGG ACA ATG ATG CAA 132
Tyr Phe Glu Trp Tyr Leu Pro Asn Asp Gly Asn His Trp Asn Arg Leu Asn Ser Asp Ala Ser Asn
TAC TTT GAA TGG TAT CTA CCT AAT GAC GGA AAT CAT TGG AAT CGA TTA AAC TCT GAT GCG AGT AAC 198
ClaI
Leu Lys Ser Lys Gly Ile Thr Ala Val Trp Ile Pro Pro Ala Trp Lys Gly Ala Ser Gln Asn Asp
CTT AAA AGC AAA GGG ATT ACA GCG GTG TGG ATT CCT CCA GCA TGG AAG GGC GCT TCT CAA AAT GAC 264
Val Gly Tyr Gly Ala Tyr Asp Leu Tyr Asp Leu Gly Glu Phe Asn Gln Lys Gly Thr Val Arg Thr
GTA GGA TAC GGA GCC TAT GAC CTG TAT GAT CTG GGA GAA TTT AAT CAA AAA GGT ACC GTC CGT ACA 330
Lys Tyr Gly Thr Arg Ser Gln Leu Gln Ala Ala Val Thr Ser Leu Lys Asn Asn Gly Ile Gln Val
AAA TAT GGA ACA CGT AGT CAG TTA CAA GCT GCG GTA ACC TCC TTA AAA AAT AAT GGA ATT CAA GTA 396
A
EcoRI
Tyr Gly Asp Val Val Met Asn His Lys Gly Gly Ala Asp Ala Thr Glu Met Val Arg Ala Val Glu
TAT GGT GAC GTT GTT ATG AAT CAC AAA GGT GGC GCA GAC GCT ACT GAA ATG GTA AGG GCC GTT GAA 462
Val Asn Pro Asn Asn Arg Asn Gln Glu Val Thr Gly Glu Tyr Thr Ile Glu Ala Trp Thr Arg Phe
GTG AAT CCC AAT AAC CGT AAC CAA GAA GTG ACT GGT GAA TAT ACC ATT GAA GCT TGG ACT AGA TTT 528
HindIII
Asp Phe Pro Gly Arg Gly Asn Thr His Ser Ser Phe Lys Trp Arg Trp Tyr His Phe Asp Gly Val
GAT TTT CCA GGG CGA GGA AAT ACT CAT TCT AGC TTT AAA TGG AGA TGG TAT CAT TTT GAT GGT GTG 594
Asp Trp Asp Gln Ser Arg Arg Leu Asn Asn Arg Ile Tyr Lys Phe Arg Gly His Gly Lys Ala Trp
GAT TGG GAT CAG TCA CGT AGA CTG AAC AAT CGC ATC TAT AAA TTT AGA GGT CAT GGC AAA GCT TGG
HindIII 660
Asp Trp Glu Val Asp Thr Glu Asn Gly Asn Tyr Asp Tyr Leu Met Tyr Ala Asp Ile Asp Met Asp
GAT TGG GAA GTT GAT ACG GAA AAT GGT AAT TAT GAT TAT TTA ATG TAC GCT GAT ATT GAT ATG GAT 726
His Pro Glu Val Val Asn Glu Leu Arg Asn Trp Gly Val Trp Tyr Thr Asn Thr Leu Gly Leu Asp
CAC CCA GAA GTA GTA AAT GAA TTA AGA AAT TGG GGT GTT TGG TAC ACA AAC ACA TTA GGA CTC GAT 792
B
Gly Phe Arg Ile Asp Ala Val Lys His Ile Lys Tyr Ser Phe Thr Arg Asp Trp Ile Asn His Val
GGA TTT AGA ATA GAT GCG GTT AAA CAT ATA AAG TAT AGC TTT ACG CGC GAT TGG ATT AAT CAC GTT 858
Arg Ser Ala Thr Gly Lys Asn Met Phe Ala Val Ala Glu Phe Trp Lys Asn Asp Leu Gly Ala Ile
AGA AGT GCA ACA GGT AAA AAT ATG TTT GCG GTT GCT GAG TTT TGG AAG AAT GAT TTA GGT GCA ATT 924
Glu Asn Tyr Leu Gln Lys Thr Asn Trp Asn His Ser Val Phe Asp Val Pro Leu His Tyr Asn Leu
GAA AAC TAT CTG CAG AAA ACA AAC TGG AAC CAT TCA GTC TTT GAT GTG CCG TTA CAT TAT AAT CTT 990
PstI
Tyr Asn Ala Ser Lys Ser Gly Gly Asn Tyr Asp Met Arg Asn Ile Phe Asn Gly Thr Val Val Gln
TAT AAT GCA TCA AAA AGC GGA GGG AAC TAT GAT ATG CGA AAC ATA TTT AAT GGA ACG GTT GTT CAA 1056
C
Arg His Pro Ser His Ala Val Thr Phe Val Asp Asn His Asp Ser Gln Pro Glu Glu Ala Leu Glu
CGA CAT CCA AGT CAT GCT GTA ACA TTT GTT GAT AAT CAT GAT TCG CAG CCT GAA GAA GCA TTA GAA 1122
Ser Phe Val Glu Glu Trp Phe Lys Pro Leu Ala Tyr Ala Leu Thr Leu Thr Arg Glu Gln Gly Tyr
TCT TTT GTT GAA GAA TGG TTT AAA CCA TTA GCG TAT GCG CTT ACA TTA ACG CGT GAA CAA GGA TAC 1188
Pro Ser Val Phe Tyr Gly Asp Tyr Tyr Gly Ile Pro Thr His Gly Val Pro Ala Met Arg Ser Lys
CCT TCT GTA TTT TAC GGA GAT TAT TAT GGG ATT CCA ACA CAT GGA GTG CCA GCA ATG AGA TCA AAA 1254
Ile Asp Pro Ile Leu Glu Ala Arg Gln Lys Tyr Ala Tyr Gly Lys Gln Asn Asp Tyr Leu Asp His
ATC GAT CCG ATT TTA GAA GCA CGT CAA AAG TAT GCA TAC GGA AAA CAA AAT GAT TAC TTA GAC CAT 1320
His Asn Ile Ile Gly Trp Thr Arg Glu Gly Asn Thr Ala His Pro Asn Ser Gly Leu Ala Thr Ile
CAT AAT ATC ATT GGT TGG ACG CGT GAA GGG AAT ACA GCA CAC CCC AAT TCA GGT CTA GCT ACC ATC 1386
Met Ser Asp Gly Ala Gly Gly Ser Lys Trp Met Phe Val Gly Arg Asn Lys Ala Gly Gln Val Trp
ATG TCT GAT GGA GCG GGT GGA AGT AAG TGG ATG TTT GTT GGG CGT AAT AAG GCT GGT CAA GTA TGG 1452
Ser Asp Ile Thr Gly Asn Arg Thr Gly Thr Val Thr Ile Asn Ala Asp Gly Trp Gly Asn Phe Ser
AGT GAT ATT ACA GGA AAC CGT ACA GGT ACG GTT ACA ATC AAT GCA GAC GGT TGG GGC AAT TTC TCT 1518
Val Asn Gly Gly Ser Val Ser Ile Trp Val Asn Lys ***
GTG AAT GGA GGG TCA GTT TCT ATT TGG GTC AAC AAA TAA AA GTGGAAAAGA AGAGGCCGTA GGTTAATATG 1589
GTCTTTTCTT TTCTTTTAAG GAGGTTCAAT GAATTTGTCG GTTATCCAAT TATTACATGC TGAGCTGTTA GATTATTCGT 1669

(positions -128 to -99), seemed to be the most probable promoter for transcription. At 3' end, a short distance from the termination codon of the G6-amylase gene, there was a palindromic sequence.

Three amino acid sequences, which are considered to constitute the active centers of various α-amylases from bacteria to human (5,6) and cyclodextrin glucanotransferase (6), were also found in the predicted amino acid sequence of the G6-amylase (A, B, and C regions in Fig. 2).

Comparison of G6-amylase with other bacterial α-amylases - The predicted amino acid sequence of the G6-amylase originated from Bacillus sp. #707 was compared with those of α-amylases of Bacillus origin. The sequence showed almost no homology with that of B. subtilis α-amylase (7), which belongs to the saccharifying type α-amylase. In contrast, it was quite similar to those of B. amyloliquefaciens (8), B. licheniformis (9) and B. stearothermophilus α-amylases (5,10,11), which belong to the liquefying type α-amylases (Fig. 3). The homologous amino acid sequences in the four amylases start at position 6, Glu residue of the G6-amylase, and continued to the COOH-terminals, except for the B. stearothermophilus α-amylase. The sequence of the extracellular mature G6-amylase was 65.5, 65.9 and 66.3 % homologous to those of B. amyloliquefaciens, B. licheniformis, and B. stearothermophilus α-amylases, respectively. 53.6 % of the amino acid sequence of the G6-amylase was identical with those of the other three α-amylases. The position of the A, B, and C-regions in the four enzymes was also the same. In addition, more than 60 % of the DNA nucleotide sequences of the coding regions for the extracellular enzymes in the four amylases

---

Fig. 2. Nucleotide sequence of the G6-amylase gene of the alkalophilic Bacillus sp. #707. The nucleotides were determined by the dideoxy chain termination method of Sanger et al (14) 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. The noncoding DNA strand from 5' to 3' is shown with its corresponding amino acid sequence. The three amino acids in the NH_2-terminal region of the extracellular G6-amylase from B. subtilis 207-25 harbouring pTUB812 are shown by a solid line above the sequence. The cleavage site between the signal peptide and extracellular G6-amylase is indicated by a vertical arrow. The most probable RNA polymerase binding site (TTGCCA), the potential Pribnow box (TGAAAT), and a ribosome-binding site (AGGAGG) are underlined. The sequences containing an inverted repeat structure are indicated by horizontal arrows. The three regions (A, B, and C), in which the deduced amino acid sequence of G6-amylase has a high homology with the regions forming the active center of α-amylases, are boxed.

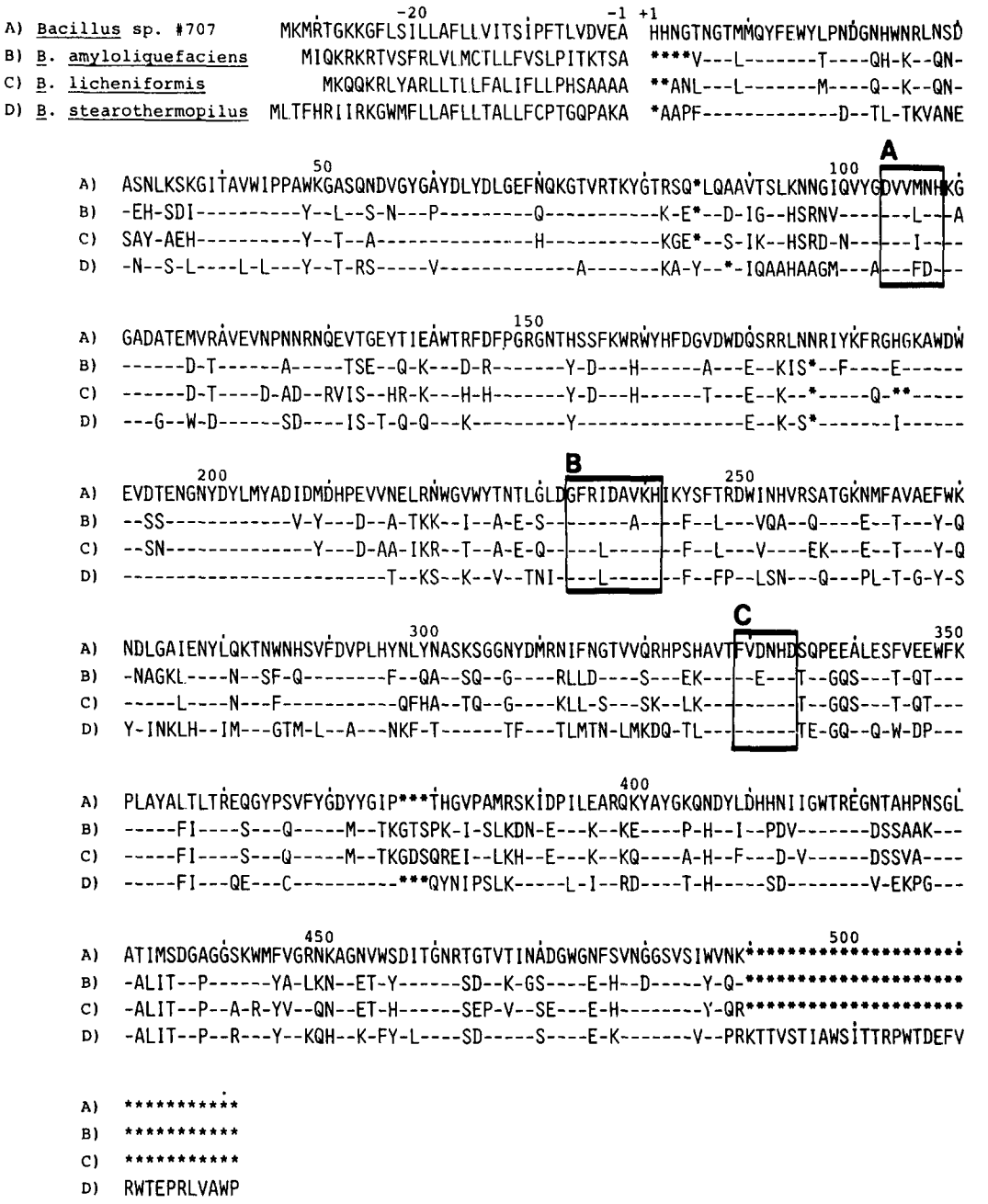


Fig. 3. Comparison of the deduced amino acid sequences of G6-amylase of Bacillus sp. #707 (A) and α-amylases of B. amyloliquefaciens (B), B. licheniformis (C) and B. stearothermophilus (D). -, The same amino acid residues with those of G6-amylase. *, Gaps introduced into the sequences to maximize homology.

genes were homologous to each other. Thus the DNA region encoding the mature G6-amylase seemed to be derived from a common ancestor gene with the genes for the three α-amylases.

In contrast, the predicted amino acid sequences of the signal peptides and NH_2-terminal amino acids in the four enzymes were completely different from each other. Moreover, there was no homology among the DNA nucleotide sequences in the upstream regions of the translation initiation sites. Therefore, it is possible to consider that the four amylases which were produced by the insertion of the same DNA encoding an enzyme with starch-hydrolyzing activity into different gene(s) in each Bacillus strain, resulted from mutations, and that the difference in their enzymatic properties, such as thermostability and substrate specificity, might be acquired.

Since the DNA nucleotide and amino acid sequence of the G6-amylase gene was quite similar to those of the genes for the liquefying type of α-amylases, it is suggested that the G6-amylase may be one kind of liquefying type α-amylase.

ACKNOWLEDGEMENTS

This work was supported in part by a grant from the Ministry of Education, Science and Culture and in part from the Biomass Conversion Project, Ministry of Agriculture, Forestry and Fisheries of Japan.

REFERENCES

1. Kimura, K., Tsukamoto, A., Ishii, Y., Takano, T., and Yamane, K. (1987) Appl. Microbiol. Biotechnol. in press.
2. Birnboim, H.C., and Doly, J. (1979) Nucleic Acids Res. 7:1513-1523.
3. Tabak, H.F., and Flavell, R.A. (1978) Nucleic Acids Res. 5:2321-2332.
4. Tarr, G.E. (1977) Methods Enzymol. 47:335-357.
5. Ihara, H., Sasaki, T., Tsuboi, A., Yamagata, H., Tsukagoshi, N., and Udaka, S. (1985) J. Biochem. 98:95-103.
6. Kimura, K., Kataoka, S., Ishii, Y., Takano, T., and Yamane, K. (1987) J. Bacteriol. 169:4399-4402.
7. Yamazaki, H., Ohmura, K., Nakayama, A., Takeichi, Y., Otozai, K., Yamasaki, M., Tamura, G., and Yamane, K. (1983) J. Bacteriol. 156: 327-337.
8. Takkinen, K., Petterson, R.F., Kalkkinen, N., Palva, I., Söderlund, H., and Kääriäinen, L. (1983) J. Biol. Chem. 258:1007-1013.
9. Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N., and Udaka, S. (1985) J. Biochem. 98:1147-1156.
10. Nakajima, R., Imanaka, T., and Aiba, S. (1985) J. Bacteriol. 163: 401-406.
11. Sohma, A., Fujita, T., and Yamane, K. (1987) J. Gen. Microbiol. 133:3271-3277.
12. Messing, J., and Vieira, J. (1982) Gene 19:269-276.
13. Henikoff, S. (1984) Gene 28:351-359.
14. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA. 74:5463-5467.