Kor. J. Appl. Microbiol. Biotech.
Vol. 18, No. 3, 309-316 (1990)

NaCl-dependent Amylase Gene From *Bacillus circulans* F-2: Its Nucleotide Sequence

Kim, Cheorl-Ho[1*], Suk-Tae Kwon[2], Dae-Sil Lee[2], Hajime Taniguchi[3] and Yoshiharu Maruyama[1]

[1]*Laboratory of Biochemistry, Department of Agricultural Chemistry, The University of Tokyo, Tokyo 113, Japan*
[2]*Genetic Engineering Center, Korea Institute of Science and Technology (KIST), Seoul, Korea*
[3]*Laboratory of Bioconversion, National Food Research Institute (NFRI), Tsukuba, Japan.*

Bacillus circulans F-2의 NaCl 의존성 amylase 유전자의 DNA 염기배열 결정

김철호[1*] · 권석태[2] · 이대실[2] · 타니구치 하지메[3] · 마루야마 요시하루[1]

[1] 동경대학교 농예화학과 생화학교실, [2] 한국과학기술연구원 유전공학센터,
[3] 일본 농림수산성 식품종합연구소 소재화기술연구실

The sequence of a 1795 bp restriction fragment containing the *B. circulans* F-2 gene for NaCl-dependent α-amylase (Cl-amylase) is reported. The probable coding region of the gene is 1005 base pairs (335 amino acids) long. The NaCl-dependent α-amylase (*cl-amy*) sequence shows an open reading frame (ORF) with the translated molecular weight of about 38,006, which correspond to a molecular weight of about 35,000 (Mr). The gene is preceded by the sequence resembling promoter for the vegetative *B. subtilis* RNA polymerases. These are followed by the sequences resembling a *B. subtilis* ribosome binding site 5 nucleotides before the first codon of the gene. Homologous regions with other amylases were found. The N-terminal sequences of the mature proteins expressed in *E. coli* were identical to the N-terminal sequences which are analysed.

Bacillus circulans F-2 produce a variety of extracellular enzymes including amylase-pullulanase enzyme (1), NaCl-dependent α-amylase (2) and α-amylase capable of digesting potato starch granules (3, 4), and these enzymes are induced by potato starch granules and cross linked potato starch, but not with soluble starch or oligosaccharides (5). These specific characteristics are not reported in amylases of other microorganisms.

Recently we cloned the two structural genes coding the raw potato starch-digesting α-amylase (*rsda*) and NaCl-dependent α-amylase (*cl-amy*), which are the extracellular enzymes of *Bacillus circulans* F-2 (2, 6, 7). The genes, *rsda* and *cl-amy,* were stably expressed in *Escherichia coli* resulting in each active amylases which were about 93,000 and 35,000 of molecular weight, respectively and have antigenicities indistiguishable from those of authentic Cl-amylase and raw starch digesting amylase and were efficiently secreted into the periplasmic space (2, 6).

In this study we have determined the nucleotide sequence of the *cl-amy* gene and found that the gene codes for the polypeptide of 335 amino acid residues. The control regions for transcription and translation of the *cl-amy* gene have also been identified in the

Key words: NaCl-dependent amylase; N-terminal amino acid sequence; DNA sequence; Signal sequence; Promoter region; Amino acid homologous region.
*Corresponding author
Present address: Genetic Engineering Center, Korea Institute of Science and Technology, Seoul, Korea

adjacent regions. In addition we compared the amino-acid sequences of *rsda* and *cl-amy* gene product with those of amylases, which are produced by various strains of microorganisms.

Materials and Methods

Strains and plasmids

Bacillus circulans F-2 was from our collections (3). *Escherichia coli* HB101 and JM109 were used for plasmid construction. Plasmid pYKA3 is an *E. coli* plasmid made up of pYEJ001 and *B. circulans* F-2 chromosomal *cl-amylase* gene (Kim *et al.*, Reference of No.2).

Plasmids pUC118 and 119 were used for sequencing. A 5.4 kb fragment of pUKA-1 and pUKA-2, which contained *cl-amylase* structural gene and regulatory region, was inserted into pUC118 or pUC119 from plasmid pYKA3.

Media

E. coli was grown on Luria broth (LB). Indicator plates contained 50 μg/m*l* ampincillin and 40 ug/m*l* 4'-bromo-5'-chloro-3'-indolyl β-D-galactoside (X-gal) in LB.

Isolation of DNA

Plasmid DNA was isolated from 100 m*l* cultures by the cleared lysate method (8) or from 1.5 m*l* cultures by rapid extraction method as described (9). DNA fragments were electrophoresised through either agarose or polyacrylamide and electroeluted as described (10).

Enzyme

Restriction enzyme, polynucleotide kinase, T4-induced DNA ligase, *E. coli* Exonuclease-III, Mung bean nuclease, Klenow fragment of DNA polymerase I, Calf intestinal phosphatase and bacterial alkaline phosphatase were purchased from Takara Shuzo Co. Ltd. (Kyoto, Japan), Bethesda Research Laboratories (Rockville, Md.), Boehringer Mannheim GmbH (Mannheim, West Germany). All of other enzymes used were obtained from commercial sources and used under conditions recommended by the supplier.

Chemicals

Ampicillin, X-gal (4'-bromo-5'-chloro-3'-indolyl β-D-galactoside), Pronase E, Proteinase K and IPTG (Isopropyl β-D-Thiogalactopyranoside) were purchased from Sigma Co. (USA). All other chemicals were of reagent grade.

DNA sequencing

DNA sequencing was done using the Sanger dideoxy chain termination method (11). Specific restriction fragments were cloned into the mp 18 or mp 19 M 13 vectors for dideoxy sequencing (1 & 2) and kilo sequencing by plasmid pUC118 and pUC119.

Amino acid analyis

The amino acids were identified and measured in a Hitachi 835 amino acid analyzer (Tokyo, Japan) after the enzyme preparation was hydrolyzed in 6 N HCl for 3 hr at 130°C.

Sequence analysis of NH_2-terminus of protein

The amino-terminal sequence of the native enzyme produced by *E. coli* (pHA 300) was identified in a 470A protein sequencer (Applied Biosystems Co. MD, USA).

Results

Sequence of NaCl-dependent α-Amylase (cl-amylase) Gene

(a) Sequencing strategy: The cl-amylase gene estimated to be approximately 1100 base pairs from the size of the purified cl-amylase protein and the insert DNA size of subcloned plasmid pYKA31 containing the cl-amylase gene as a 2.15 kb *Hind*III-*Eco*RV fragment.

*Sau*3AI fragments from pYKA31 containing *Bacillus circulans* F-2 sequences were isolated from acrylamide gels and subcloned into the *Bam*HI site of pUC18 or pUC19. Ligation mixes were transformed innto *E. coli* JM109 and clones were identified as white colonies on L amp plates containing Xgal. To obtain the sequences from both strands it was necessary to generate clones in which the *Sau*3AI fragments had been cloned in both orientations. The orientation of each fragment was determined, where possible, by restriction patterns with *Alu*I or *Taq*I. Plasmid DNA from clones where the orientation of the *Sau*3AI fragment could not be determined was restricted with *Hin-d*III and *Pst*I, the fragment purified, and subcloned into the *Hind*III-*Pst*I site of pUC19 (Fig. 1).

All 14 *Sau*3AI fragments were thus sub-cloned in

Fig. 1. Restriction map and sequencing strategy of region surrounding the Cl-amylase gene.
The hatched box indicates the coding sequence for the Cl-amylase protein. The direction and extent of sequence determination are shown by horizontal arrows. The abbreviations of restriction endonuclease sites are A, *Acc*I; B, *Bam*HI; E, *Eco*RV, P, *Pst*I; S, *Sau*3AI.

both orientations and sequenced by the method of dideoxy sequencing (12). These fragments gave most of the sequence shown in Fig. 2. And for the complete sequence, a series of deletions was constructed using Exonuclease III, Mung bean nuclease and Klenow fragment of DNA polymerase with the same deletion procedure as described in Materials and Methods. The full sequence of 1795 base pairs is shown in Fig. 2. It was assembled using a computer programme to match sequences obtained from different runs. Most of the sequence (with the exception of short stretches at each end) has been determined for both strands (Fig. 2) and all restriction site used for end-labelling have been sequenced across.

(b) The location of NaCl-dependent α-Amylase gene: The sequence of Fig. 2 was searched for open reading frames (ORF) sufficient to code for a protein about Mr35,000-40,000 (2). There is only one such sequence of 1005 bp from base 661 to 1666. It has 335 codons equivalent to a protein of Mr 38,006. The N-terminal amino acid sequence (amino acid 1-20) resembles the signal sequences of the precursor forms of other secreted amylases known to be processed during excretion when the signal sequence is cleaved off (Fig. 3).

The N-terminal amino acid sequence of the mature enzyme begins at the number 21 of precursor enzyme

In the previous paper (2), we purified the NaCl-dependent amylase expressed in *E. coli*, extensively characterized and sequenced the N-terminal amino acid sequence from No.1 to No.7. The amino acid sequence was determined to be Ala-Ser-Lys-Asp-Val-Gly-Trp of N1-N7. To search the signal cleavage site of the enzyme. The deduced amino acid sequence of NaCl-dependent amylase from the DNA sequence was subjected to the computer programmed analysis with the above N-terminal sequence of mature enzyme. As shown in Fig. 2, the same amino acid sequences with No.1-No.7 was seen from No.21 to No.27 of the sequenced data. This result indicates that the NaCl-dependent amylase gene was expressed as a form of precursor including the signal sequence (20 amino acids) and processed by some of signal peptidase resulting in matured enzyme.

Transcription and translational signals

The consensus sequence of sigma^{55} promoters is shown in Fig. 4. The 1795 base pair DNA sequence has been scanned for sequences related to the −35 (TTGACA) and −10 (TATAAT) consensus sequence which are usually separated by 17 or 18 base pairs. There are two sequences (Fig. 4) which are in agreement with the consensus and one of these (nucleotides 593-623) is located 32 base pairs from the ATG of the NaCl-dependent amylase gene. It corresponds closely to the consensus −35 and −10 sequences which are separated by 17 base pairs.

Determination of amino acid composition and codon usage

The amino acid composition of the Cl-amylase from *E. coli* C600 carrying pYKA31 is presented in Table 2. Molar ratios obtained by amino acid analysis were closely consistent with those derived from the DNA sequence. The codon usage in the Cl-amylase gene is shown in Table 1. Of 61 codons, 59 codons have been used.

Discussion

Here we report the nucleotide sequence analysis of the NaCl-dependent α-amylase (Cl-amylase) gene of *Bacillus circulans* F-2. The cloned Cl-amylase gene is located within the 2.15 kbp *Hind*III-*Eco*RV fragment (Fig. 1).

Cl-amylase gene is contained on the plasmid, but another gene upstream and possibly a large segment of another gene down stream. The deduced protein coded for by the 1005-bp region, comprizes 335 amino acid residues, corresponding to an M_r of 38,006 (Fig. 2). The calculated molecular weight of the putative pre-amylase is thus about 3 kDa larger than that calculated

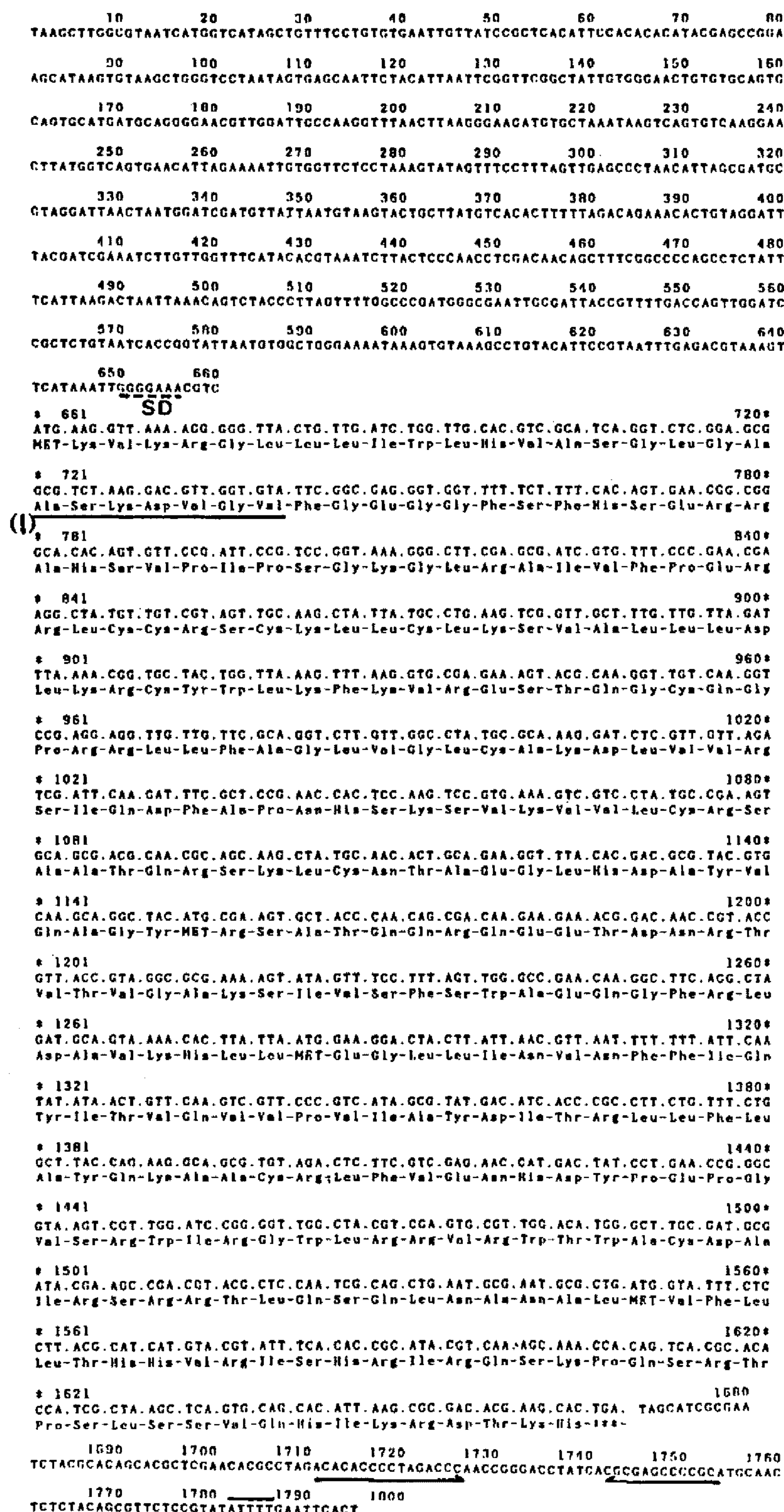

Fig. 2. DNA nucleotide and deduced amino acid sequence of the *B. circulans* F-2 Cl-amylase gene and surrounding region.

The DNA sequence of the coding strand is given from 5' to 3', numbered from nucleotide 1 at the bigining site. The Proposed ribosomal RNA binding site (SD) is underlined with a dashed line. The predicted amino acid sequence is given below the DNA sequence. The amino acids are numbered taking the NH_2-terminal amino acid of the pre-amylase as 1. Underlined amino acids have been determined by automated Edman sequencing of the purified amylase. The hairpin loop of the putative rho-independent terminator sites underlined.

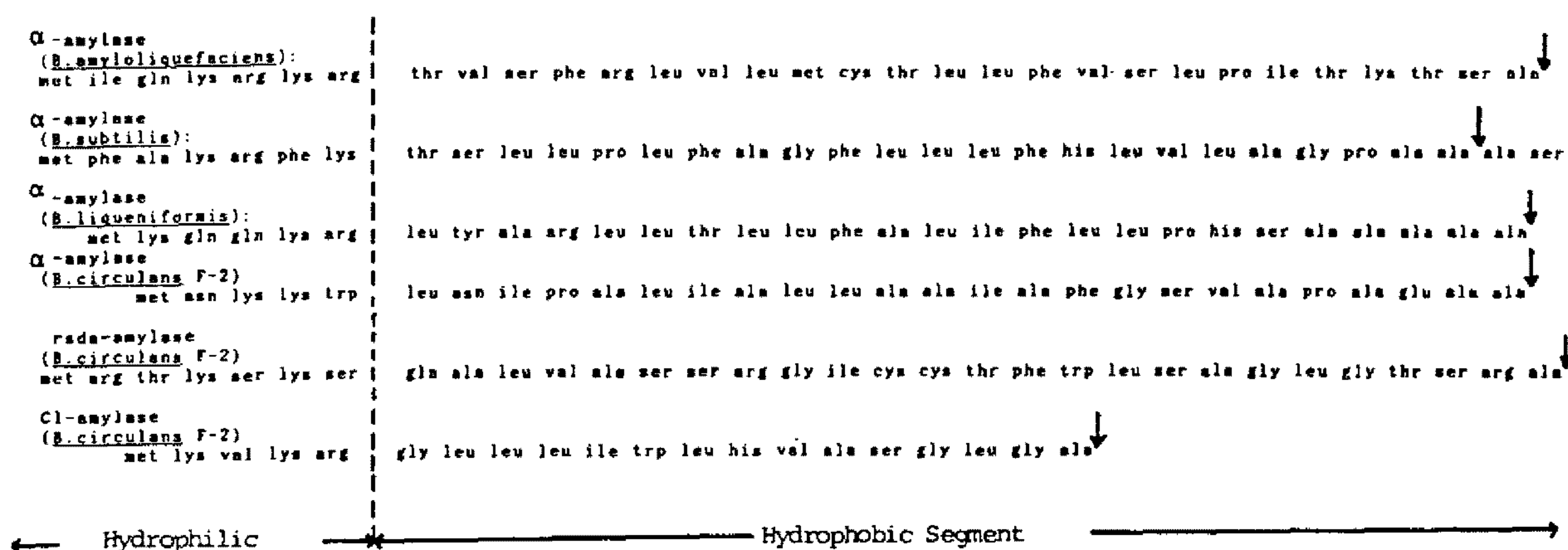

Fig. 3. NH_2-terminal amino acid sequences of precursors to extracellular amylases of *Bacilli*.
The amino acid sequences of the NH_2 terminal of six precursors to extracellular amylases of *Bacilli* are shown and compared to the sequence of Cl-amylase deduced from the DNA sequence. Published sequence are for α-amylases from *B. amylofaciens* (22), *B. subtilis* (23, 24), *B. licheniformis* (25) and *B. circulans* F-2 (7, 19). The arrows show known or postulated cleavage sites.

NUCLEOTIDES	TTGACA	TATAAT
546 - 552	ATTACCGTTTTGACCAGTTGGATCCGCTCTGTAATCACCGGTAT	
593 - 623	GGAAAATAATTGAAAAAGCCTGTACATTCCGTACATTCCGTAAT	

Fig. 4. Regions showing homology to *B. subtilis* RNA initiation sites.
On the top line is shown the consensus for the "-35' and "-10" regions of sigma[55] promoters. Sequence 5' to the Cl-amylase gene are shown. The nucleotides refer to the numbering given in Fig. 1. Nucleotides that match the consensus sequence are underlined.

Origin	I	II	III	IV
α-amylase B.cir.	---164DFAPNH---	254GIRVDAVKH---	286EWFL---	352FLDNHD---
rsda-amylase B.cir.	----71DASINH---	305GMRADNNKH---	414RWFL---	512FVSNHI---
Cl-amylase B.cir.	---124DFAPNH---	197GFRLDAVKH---	237LLFL---	250FVDNHD---
α-amylase B.sub.	----97DAVINH---	172GFRFDAAKH---	208EILQ---	264WVESHD---
α-amylase B.stear.	---101DVVFDH---	230GFRLDAVKH---	264EYWS---	326FVDNHD---

Fig. 5. Homologous regions of amino acid sequences of *B. circulans* F-2 amylases and some bacterial α-amylases.
(B. cir, *B. circulans* F-2; B. stear., *B. stearothermophilus*; B.sub., *B. subtilis*). Amino acids are shown as one-letter abbreviations. Numbering of the amino acids starts from amino terminal of the mature enzymes except *B. circulans* F-2 amylases, which start from the amino-terminal methionine of putative precursors. The residues located at active centers of Taka-amylase A are underlined. Roman numbers show homologous regions suggested by Nakajima *et al.* (20).

Table 1. Codon usage of the Cl-amylase gene of *B. circulans* F-2

Amino acid	Codon	Number of codons	Amino acid	Codon	Number of codons	Amino acid	Codon	Number of codons
Phe	UUU	9	Gln	CAA	12	Gly	GGU	10
	UUC	5		CAG	5		GGC	6
Leu	UUA	8	Asn	AAU	3		GGA	2
	UUG	6		AAC	5		GGG	2
	CUU	5	Lys	AAA	7	Ser	UCU	2
	CUC	5		AAG	12		UCC	4
	CUA	9	Asp	GAU	5		UCA	4
	CUG	6		GAC	6		UCG	3
ILe	AUU	6	Glu	GAA	9	Pro	CCU	1
	AUC	4		GAG	2		CCC	2
	AUA	5	Cys	UGU	4		CCA	2
Met	AUG	4		UGC	7		CCG	10
Val	GUU	12	Term	UGA	1	Thr	ACU	2
	GUC	6	Trp	UGG	7		ACC	4
	GUA	6	Arg	CGU	8		ACA	2
	GUG	6		CGC	5	Ala	GCU	5
Tyr	UAU	3		CGA	9		GCC	1
	UAC	4		CGG	4		GCA	9
Term	UAA	0	Ser	AGU	9		GCG	11
	UAG	0		AGC	4			
His	CAU	3	Arg	AGA	2			
	CAC	9		AGG	5			

[a] A total of 335 amino acids and 59 sense codons were used in preamylase of *B. circulans* F-2 (Cl-amy).

for the exoenzyme (2). Compared with the amino acid sequence of mature Cl-amylase, an additional twenty amino acids were observed at the N-terminal of the cloned Cl-amylase gene The sequence terminates with alanine, which is generally observed at the cleavage site of the sigma peptide. The putative Cl-amylase signal sequence has similar properties to the other *Bacillus* signal sequences with a hydrophilic segment of five amino acids ending at either at arginine followed by a hydrophobic segment which may end at valine 14. Of the 20 amino acids from leucine 7 to alanine 20, 10 have hydrophathy indices greater than 1. Moreever the Cl-amylase sequence shows homology with the other signal sequences as shown in Fig. 3. It is therefore suggested that a signal peptide of about 20 aa residues is processed during enzyme secretion, as has been shown for many other secreted proteins.

A TATAAT sequence (13, 14), which seems to function as a TATA box, was found in the upstream region of the cloned Cl-amylase gene. The putative promoter is followed by a sequence (GGGAAAA) which is complementary to the sequence CC(U) CCUUUU found to the 3' terminus of the 16S ribosomal RNA of *B. substilis.* A complex between these two sequence would have a binding energy G = -13.6 kcals per mole by Tinoco *et al.* (18), which is in the range observed for other *Bacillus* ribosome binding site sequences (15-16). This sequence is followed by an ATG five nucleotides away, which is typical for *Bacillus* genes. Although the precise translation initiation site has not been determined, amylase must be translated from ATG codons located 5 bp downstream of this putative ribosome-binding site. The Cl-amylase gene is followed by a typical prokaryotic rho-independent transcription terminator sequence (G = -19.2) (18).

Recently, Nakajima *et al.* (1986) compared the amino acid sequence of 11 different α-amylases and reported the presence of four highly homologous

Table 2. Amino acid composition of NaCl-dependent - amylase

Amino acid	Amino acid analysis Mol %	From DNA sequence[a] No.	Mol %
Ala	7.39	26	7.76
Cys	3.29	11	3.28
His	3.77	12	3.58
Met	1.20	4	1.19
Thr	4.11	14	4.18
Asn	2.43	8	2.39
Asp	3.32	11	3.28
Glu	3.28	11	3.28
Gln	4.98	17	5.07
Ile	4.54	15	4.48
Leu	10.54	39	11.64
Phe	4.32	14	4.18
Trp	2.01	7	2.09
Pro	3.04	10	2.99
Tyr	1.98	7	2.09
Arg	9.88	33	9.85
Gly	6.34	20	5.97
Lys	5.71	19	5.67
Ser	8.11	27	8.06
Val	8.98	30	8.96

[a] The values in the parentheses are the number of individual amino acid residues charged when the first 20 amino acid residues from the NH_2-terminus are omitted. The remaining residues did not change.

regions (20). When we compared the deduced protein sequence of the *B. circulans* F-2 Cl-amylase with other two amylases (rsda-amylase and α-amylase) of this bacterium, which are previously cloned and sequenced by Kim *et al.* (6, 7) and Nishizawa *et al.* (19), and the amino terminal portion of other *Bacillus* species's amylase shows no obvious regions of homology, with the exception of the signal sequence regions of those genes (Fig. 3). However, we were able to find all the homologous regions as shown in Fig. 5. His and Asp residues in regions 2 and 4 are suggested to be in the active site of take-amylase A (21), and they are also found in *B. circulans* F-2 Cl-amylase. These observations, taken together with the results of the ORF determination, suggests that the ORF is likely to code an amylase of α-amylase and indicate that the *B. circulans* F-2 amylase system is multiple, and thus those genes have independent structures.

요　약

Bacillus circulans F-2의 생산하는 NaCl 의존성 amylase(NaCl-dependent amylase) 유전자를 함유하는 1795 bp의 DNA 염기배열을 결정하였다. 본 유전자의 ORF는 총염기수 1005 bp(335 아미노산)로 구성되며, 분자량 38,006의 amylase 단백질을 code하였다. 이는 공여균의 분비 amylase의 분자량 약 35,000과 일치하였다. 본 유전자의 상류영역(upstream region)에는 고초균(*Bacillus subtiis*)의 전형적인 전사발현영역(transcriptional region)과 상보적인 DNA 영역이 존재하였다. 성숙단백질의 N-말단측 아미노산 배열은 Ala-Ser-Lys-Val-Gly이며, 분비에 필요한 20개의 signal 아미노산 배열을 갖는 전형적인 분비 단백질임이 확인 되었다. 한편 다른 amylase들과의 비교결과, smylase 활성발현과 밀접히 관련되 있는 4개 부위의 상보성영역(homologous region)을 가지고 있었다.

References

1. Sata, H., M. Umeda, C.H. Kim, H. Taniguchi and Y. Maruyama: *Biochem. Biophys. Acta.* **991**, 388-394 (1989).
2. Kim, C.H., H. Taniguchi and Y. Maruyama: *Arch. Biochem. and Biophys.* In Press (1990).
3. Taniguchi, H., M.J. Jae, N. Yoshigi and Y. Maruyama: *Agric. Biol. Chem.* **47**, 511-519 (1983).
4. Taniguchi, H., F. Odashima, M. Igarashi, Y. Maruyama and M. Nakamura: *Agric. Biol. Chem.* **46**, 2017-2115 (1982).
5. Sata, H., H. Taniguchi and Y. Maruyama: *Agric. Biol. Chem.* **50**, 2803-2809 (1987).
6. Kim, C.H., H. Sata, H. Taniguchi and Y. Maruyama: *Biochim. Biophys. Acta.* In Press (1990).
7. Kim, C.H., H. Taniguchi and Y. Maruyama: *Biochemie.* In Press (1990).
8. Birnbiom, H.C. and J. Doly: *Nucleic Acids Res.* **7**, 1513-1523 (1979).
9. Davis, R.W., M. Thomas, J. Cameron. T.P. St. John, S. Schever and R.A. Oadgett: *Methods in Enzymols.* **65**, 404-411 (1980).
10. Maniatis, T., E.F. Fritsch and J. Sambrook: In *Molecular Cloning:* A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. p.164 and p.178 (1982).
11. Sager, F., S. Nicklen and A.R. Coulson: *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467 (1977).

12. Messing, J.: In Third Clevelend Symposium on Macromolecules: Recombinant DNA, Ed. A. Walton, Elsevier, Amsterdam, 143-153 (1981).
13. Moran, C.P., Jr., N. Lang, S.F.J. Legrice, G. Lee, M. Stephens, A.L. Sonenshein, J. Pero and R. Losick: *Mol. Gen. Genet.* **186**, 339-345 (1982).
14. Murrary, C.L. and J.C. Rabinowitz: *J. Biol. Chem.* **257**, 1053-1062 (1982).
15. McLaughlin, J.R., Murray, C.L. and J.C. Rabinowitz: *J. Biol. Chem.* **47**, 1087-1097 (1981).
16. Palva, I., R.F. Petterssen, N. Kalkkinen, P. Lehtovaara, M. Sarvas and H. Solderlund: *Gene* **15**, 43-51 (1981).
17. Tinoco, I., P.N. Borer, B. Dengler, M.D. Levine, O.C. Uhlenbeck, D.M. Crothers and J. Gralla: *Nature New Biol.* **246**, 40-41 (1973).
18. Rosenber, M. and D. Court: *Ann. Rev. of Genetics* **13**, 319-353 (1979).
19. Nishizawa, M., F. Ozawa and F. Hishinuma: *DNA* **6**, 255-265 (1987).
20. Nakajima, R., T. Imanaka and S. Aiba: *Appl. Microbiol. Biotech.* **23**, 355-360 (1985).
21. Matsuura, Y., M. Kusunoki, W. Harada and M. Kakudo: *J. Biochem.* **95**, 697-702 (1984).
22. Takkine, K., R.F. Pettersson, N.M. Kalkkinen, I. Palva, H. Soderlund and L. Kaarianinen: *J. Biol. Chem.* **258**, 1007-1013 (1983).
23. Yang, M., A. Glaizzi and D. Henner: *Nucleic Acids Res.* **11**, 237-249 (1983).
24. Yamazaki, H., K. Ohmura, A. Nakayama, Y.M. Takeichi, K. Otozai, M. Yamasaki, G. Tamura and K. Yamane: *J. Bacteriol.* **156**, 327-337 (1983).
25. Stephens, M., S.A. Ortlepp, J.F. Ollington and D.J. McConnell, *J. Bacteriol.* **158**, 369-372 (1984).

(Received May 16, 1990)