Journal of Microbiology and Biotechnology
Vol. 2, No. 1, 56-65, 1992

Sequencing of the *RSDA* Gene Encoding Raw Starch-Digesting α-Amylase of *Bacillus circulans* F-2: Identification of Possible Two Domains for Raw Substrate-Adsorption and Substrate-Hydrolysis

KIM, CHEORL-HO* AND HAJIME TANIGUCHI[1]

Genome Research Program, Genetic Resource Center
Genetic Engineering Research Institute, Korea Institute of Science and Technology, P.O. Box 17, Taedok Science Town, Taejon 305-333, Korea
[1]Laboratory of Bioconversion, National Food Research Institute, Ministry of Agriculture, Forests and Fisheries, Tsukuba City, Ibaraki 305, Japan

Received 2 March 1992 / Accepted 2 April 1992

The complete nucleotide sequence of the *Bacillus circulans* F-2 *RSDA* gene, coding for raw starch digesting α-amylase (RSDA), has been determined. The *RSDA* structure gene consists of an open reading frame of 2508 bp. Six bp upstream of the translational start codon of the *RSDA* is a typical gram-positive Shine-Dalgarno sequence and the *RSDA* encodes a preprotein of 836 amino acids with an Mr of 96,727. The gene was expressed from its own regulatory region in *E. coli* and two putative consensus promoter sequences were identified upstream of a ribosome binding site and an ATG start codon. Confirmation of the nucleotide sequence was obtained and the signal peptide cleavage site was identified by comparing the predicted amino acid sequence with that derived by N-terminal analysis of the purified RSDA. The deduced N-terminal region of the RSDA conforms to the general pattern for the signal peptides of secreted prokaryotic proteins. The complete amino acid sequence was deduced and homology with other enzymes was compared. The results suggested that the Thr-Ser-rich hinge region and the non-catalytic domain are necessary for efficient adsorption onto raw substrates, and the catalytic domain (60 kDa) is necessary for the hydrolysis of substrates, as suggested in previous studies (8, 9).

Raw starch-digesting enzymes are composed of a core domain and a tail domain which contains either a Thr-Ser region or Thr-Pro-rich B-region. A distinct function has been suggested for these domains (11). Thr-Ser and Thr-Pro-rich domains act as raw substrate-binding peptides, distinct from the active site. It is believed that, at least in fungi, the peptides act in freeing the raw starch microcrystallines in such a way that the enzyme can then perform the hydrolysis step. These properties are present even in crystalline cellulose digesting cellulases (CCDC) (36). Despite extensive biochemical and ultrastructural studies of raw substrate digestion the molecular mechanisms underlying the recognition of the raw substrate and the adsorption of the enzyme onto the raw substrate remain unclear.

Previous studies of the raw starch digestibility of *B. circulans* F-2 RSDA have shown that the RSDA was proteolytically converted into two fragments, each with subtilisin (8, 9). The truncated isoenzyme, which does not bind to granular starch(BF-2A'), is generated by proteolytic modification (BF-2A). Both enzymes are active towards soluble substrates but BF-2A shows increased activity towards insoluble starches. These results lead to the proposal that BF-2A contains two separate domains: a catalytic domain containing the active site, and a raw starch adsorbing domain that confers on the enz-

*Corresponding author
Key words: Raw starch digesting amylase gene, molecular cloning, N-terminal amino acid sequence, DNA sequence, 5'-transcriptional region, amino acid homologous region, raw starch adsorbability/digestibility, C-domain.

yme the ability to be bound onto the surface of raw starch granules. BF-2A contains about 30,000 more molecular weights at the C-terminus than the truncated enzyme of BF-2A, and so this region was thought to be the adsorption domain (9).

To explain the domain structure of the *B. circulans* RSDA, we cloned the structural gene encoding the RSDA and the gene (*RSDA*). The *RSDA* was stably expressed in *E. coli*, resulting in an active enzyme which was efficiently secreted into the periplasmic space (10). In this study we determined the nucleotide sequence of the *RSDA* gene and found that the gene codes for the polypeptide of 836 amino acid residues. The control regions for transcription and translation of the *RSDA* gene have also been identified in the adjacent regions. We also compared the amino acid sequences of the *RSDA* product with those of other amylases for the hydrolytic domain containing the active site. Finally, a model mechanisim explaining raw starch-digestibility was investigated by comparing the Thr-Ser-rich sequence with those of other amylases and cellulases capable of digesting the raw substrates. Comparisons of the amino acid sequence of this RSDA with other enzymes yielded information about the similarities and differences among enzymes.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

We used *Bacillus circulans* F-2. The *E. coli* strains HB101 and JM109 were used for plasmid construction. Plasmid pYKA3 was made up of pYEJ001 and *B. circulans* F-2 chromosomal *RSDA* gene (10). Plasmids pUC 118 and 119 were used for sequencing analysis. A 5.4 Kb fragment of pUKA1 and pUKA2, which contained the *RSDA* structural gene and regulatory region, was inserted into either a pUC118 or pUC119 from plasmid pYKA3.

Recombinant DNA Techniques and Cloning Procedures

Preparation of plasmid DNA, endonuclease digestion, ligation, and bacterial transformation were carried out using standard procedures (13). Restriction endonuclease, mung bean nuclease, exonuclease III, DNA polymerase I (Klenow fragment), and T4 DNA ligase were used according to the specifications of the supplier(Pharmacia Co.).

Nucleotide Sequence Analysis and Computer Analysis of DNA Sequences

Enzymatic DNA sequence analysis (Sanger dideoxy chain termination method) was carried out on supercoiled DNA using a pUC sequencing kit purchased from Boeringer Mannheim (26). In specific cases restriction fragments were recloned into the mp18 or mp19 M13 vectors for dideoxy sequencing. All enzymatic reactions were performed at 40℃. Sequencing data were analysed and compared using the DNASIS/PROSIS software package (Hitachi Co.). Amylase proteins, which appeared to be similar to the *B. circulans* F-2 RSDA, were initially aligned using the GENETYX program (SDC, Software Developing Co. Tokyo). Protein pairs showing the greatest degree of similarity were subjected to a statistical analysis of the alignment. The purpose of the statistical analysis was to determine whether the best alignment of two proteins was significantly better than that which could be obtained from aligning random sequences generated from the chosen proteins. Free energies of RNA base pairings were calculated using the energy model of Tinico *et al.* (30).

Protein Purification, Amino Acid Analysis and Amino Acid Sequence Analysis

Raw starch-digesting amylase protein was isolated from *E. coli* C600(pYKA3) carrying the *RSDA* gene. Cell extracts were prepared by sonication (22 KHz). After centrifugation (15,000 rpm, 30 min) the supernatant was fractionated with ammonium sulfate(30 to 50% ppt) and the active protein solution was appiled to the columns of a DEAE Totopearl M650, a Phenyl-Toyopearl M650, and a TSK-Gel G3000 SWxL(Tosoh Co.). The amino acids were identified and measured with a Hitachi 835 amino acid analyzer after the purified enzyme was hydrolyzed in 6 N HCl for 3 hr at 130℃. The amino terminal sequence of the native enzyme, produced by recombinant *E. coli*, was identified with a 470A gas-phase protein sequencer(Appied Biosystems Inst. Co.).

RESULTS

DNA Sequencing Strategy for *RSDA* Gene

A physical map of the 5.4 kb *Hin*dIII fragment of genomic *B. circulans* F-2 DNA containing the *RSDA* is shown in Fig. 1. The nucleotide sequence of an approximately 3.25 kb region between the *Eco*RV and *Hin*dIII sites was partially determined by cloning specific restriction fragments into M13 vectord mp18 and mp19 by the method of dideoxy chain termination. Sequences near restriction sites used for cloning were resequenced away from the restriction sites to avoid potential errors related to closely spaced restriction sites by the kilo sequencing (6); a *Hin*dIII-cleaved 5.4 kb fragment of pYKA 3 was inserted into pUC118 and pUC119 resulting in new plasmids of pUKA1 and pUKA2. The new plasmids were double-digested, with the restriction endonucleases of *Sal*I(5′-protruding) and *Sst*I(3′-protruding), treated with exonuclease III, a Klenow fragment, mung bean nuclease, and then were self-ligated with T4 DNA ligase.

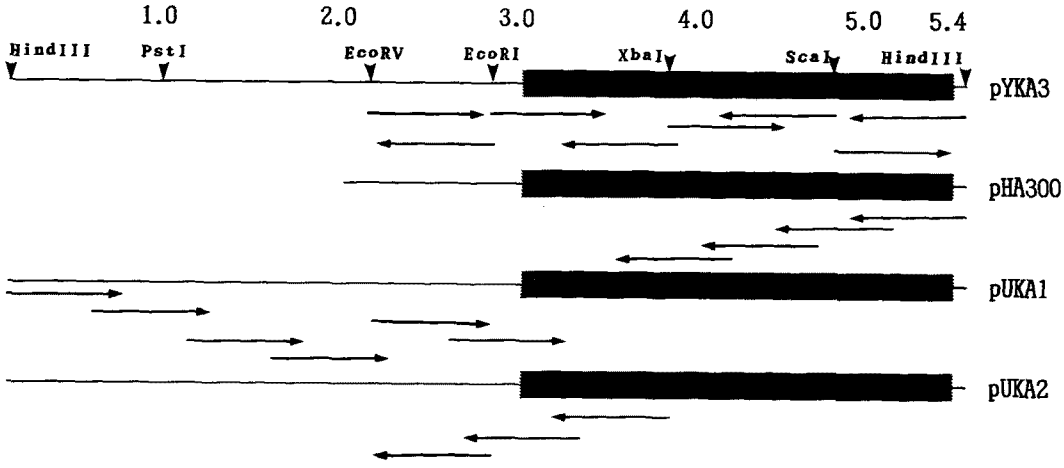

Fig. 1. Restriction map and DNA sequence strategy for the 3.2 kb fragment containing the *RSDA* gene.
The direction and length of sequence determinations are shown with arrows. The placement of the various restriction sites (exonuclease-generated subclones) used in sequencing is shown along the fragment in pUKA1 and pUKA2. The putative coding region for the *RSDA* gene is shaded.

```
-1026 CCCCCTTCCGGGAACCGTTAGCCCAACCCCGGAATCCACCGTTACGCGCGCGTTCAGGGATATCA-961
                                                                EcoRV
ATGGTGCCCGGAAGGTGAGCCGGAAGATCAACGCCTTCTAGAAGCAGAAACACAAAAAAATGCTGTATTCTTCGTATTCC-881
AGCCGTTACCAATAAAGCACGGGGAACAAGCTCCTTGACAGATATTATTTCGTGGTCGTGGTCTCGGGTACGGGAAGCAA-801
TCTCCTTGGTCCAATATTATTTCGTGGTATGGCCTTCCGGTATGTTGGTTCTTTTTAAGAGAAAAAGCAAAAAAACATGA-721
AAAATTGTGGTTCTCCTATTGCTATCTTCCCTGCCATCCTTCATTACAGTGGAGGATACAAAAAGTATAGTTTCCTTTCG-641
TTGAGCCCTACATTAGCGATGGGTAGGATTACTAATGGATCGATGTTATTATGTAAGTACTGCTTTATGTCACATCTTTT-561
TTTAGACAGACACTGTAGGCTTTCGTCGCTCTTGTTGGTTTCATACAGAAATGATTCTTCCCACCTTGACACCATTTCGT-481
CCCCACCCTCTATTTCATTAAGACGTTATCATCTACCCTTAGTTTCGGCCCGTAGGGCGATTGCCGATACCTATAGGATC-401
                                                                BamHI
CGGATGGCCATAGGCAAATCAAGGAAGTGTTCATCGTCTTATGAACAAGGTGGGCGAATGTGTCGAGTGGAGAAGATTAC-321
GTGTATGAAACCTAGACCTAGTGTCGCTATGTGCGGAGTGCTGGGACGTGCGGTGGAGTGATGACTGCTGGACGGCAGAG-241
TACTGACTGACTGAGACGGAGCTACGCGCGGCTACCAAGTTACTGGCTAATGATGGCCTAACTAAGGAACTATACTTAGG-161
AGAACCACAATTTCTATGCTTTTGCTGAGGAATTCTGATATATTAATAAATACGTTTCACGGAGCATTTCAATCTAAAAC -81
                                EcoRI
ATATATTTTTTCAAAAAAAGTGCTTGATATTTAAAAAGAAACGTGATATATTATAAAAGTGCCGTGAAACGGAGGTAGCG  -1
                      -35                -10                S D
+1
ATG.AGG.ACA.AAG.TCG.AAA.TCG.CAG.GCA.CTA.GTT.GCT.TCA.TCC.CGA.GGC.ATT.TCG.TGT.TGC   60
Met-Arg-Thr-Lys-Ser-Gln-Ala-Leu-Val-Ala-Ser-Ser-Arg-Gly-Ile-Ser-Cys-Ser-Cys-Cys   20
I C
ACG.TGC.TTC.TGG.CTC.TCG.GCT.GGA.CTA.GGC.ACC.TCG.CGT.GCT.GCT.GAG.ACT.TAC.CAT.ATG  120
Thr-Ser-Phe-Trp-Leu-Ser-Ala-Gly-Leu-Gly-Thr-Ser-Arg-Ala-[          ]-Met        40
CAT.TGG.ATA.TTA.GTG.CTA.ACC.AGT.TCC.AGG.TTC.GAA.AAC.TGG.CAC.GTT.CGA.ATT.ACA.CTG  180
His-Trp-Ile-Leu-Val-Leu-Thr-Ser-Ser-Arg-Phe-Glu-Asn-Trp-His-Val-Arg-Ile-Thr-Leu   60
                                        I
TGC.ATG.ACT.GTT.GCG.CAT.CAG.CTT.GCA.CTT|GAT.GCA.TCA.ATT.ACA.CAT|ACG.CTC.AGT.GAA  240
Cys-Met-Thr-Val-Ala-His-Gln-Leu-Ala-Leu|Asp-Ala-Ser-Ile-Thr-His|Thr-Leu-Ser-Glu   80

TAT.TAT.CAG.TGT.GCA.TAT.TAT.GCG.TAT.ACT.GGA.CAA.GAT.TCG.GAT.AGG.AAG.ATC.ACC.CGC.  300
Tyr-Tyr-Gln-Cys-Ala-Tyr-Tyr-Ala-Tyr-Thr-Gly-Gln-Asp-Ser-Asp-Arg-Lys-Ile-Thr-Arg  100
CTT.CTG.TTT.CTG.GCT.TAC.GAG.AAG.GCA.GCG.TCT.TGC.CTC.TTC.TTC.CGG.CAG.CAT.GAC.TAT  360
```

```
Leu-Leu-Phe-Leu-Ala-Tyr-Glu-Lys-Ala-Ala-Ser-Cys-Leu-Phe-Phe-Arg-Gln-His-Asp-Tyr  120
CCT.GAA.CCG.GGC.GTA.AGT.CGT.TGG.ATC.CGG.GGT.TGG.CTA.CGT.CGA.GTG.CGT.TGG.ACA.TGG  420
Pro-Glu-Pro-Lys-Val-Ser-Arg-Trp-Ile-Arg-Gly-Trp-Leu-Arg-Arg-Val-Arg-Trp-Thr-Trp  140
GCT.TGC.GAT.GCG.ATA.CGA.ATC.CGA.CGT.ACG.CTC.CAA.TCG.CAG.CTG.AAT.GCG.AAT.GCG.CTG  480
Ala-Cys-Asp-Ala-Ile-Arg-Ser-Arg-Arg-Thr-Leu-Gln-Thr-Gln-Leu-Asn-Ala-Asn-Ala-Leu  160
ATG.GTA.TTT.CTC.CTT.ACG.CAT.CAT.GTA.CGT.ATT.TCA.CAC.CGC.ATA.CGT.CAA.AGC.AAC.CAT  540
Met-Val-Phe-Leu-Leu-Thr-His-His-Val-Arg-Ile-Ser-His-Arg-Ile-Arg-Gln-Ser-Asn-His  180
AGT.CAC.GCA.CAC.CAT.CGC.CAA.GCT.CAG.TGC.AGC.ACA.TTA.AGC.GCG.ACA.CGA.AGC.ACT.GAT  600
Ser-His-Ala-His-His-Arg-Gln-Ala-Gln-Cys-Ser-Thr-Leu-Ser-Ala-Thr-Arg-Ser-Thr-Asp  200

                                                                    XbaI
AGC.ATC.GCG.AAT.CTA.CGC.ACA.GCA.CGC.TCG.AAC.ACG.CCT.AGA.CAC.ACC.TCT.AGA.CCC.AAC  660
Ser-Ile-Ala-asn-Leu-Arg-Thr-Ala-Arg-Ser-Asn-Thr-Pro-Arg-His-Thr-Ser-Arg-Pro-Asn  220
CGG.GAC.CTA.TCA.CGC.GAG.CCC.CGC.ATG.CAA.CTC.TCT.ACA.GCG.TTC.TCC.GTA.TAC.CCC.GAA  720
Arg-Asp-Leu-Ser-Arg-Glu-Pro-Arg-Met-Gln-Leu-Ser-Thr-Ala-Phe-Ser-Val-Tyr-Pro-Glu  240
TTC.ACT.GAC.CAT.AAG.TTC.CTT.GAC.ACT.GAC.TTA.TTT.AGC.ACA.TCT.TCC.CTT.AAG.TTA.AAC  780
Phe-Thr-Asp-His-Lys-Phe-Leu-Asp-Thr-Asp-Leu-Phe-Ser-Thr-Ser-Ser-Leu-Lys-Leu-Asn  260
CTG.GCA.ATC.CAA.CGT.TCC.CCT.GCA.TCA.TGC.ACT.GCA.CTG.CAC.ACA.GTT.CCC.ACA.ATA.GCC  840
Leu-Ala-Ile-Gln-Arg-Ser-Pro-Ala-Ser-Cys-Thr-Ala-Leu-His-Thr-Val-Pro-Thr-Ile-Ala  280
GAA.CCG.AAT.TAC.TGT.AGA.ATT.GCT.CAC.TAT.CAG.GAC.CCA.GCT.TAC.ACT.TAT.GCT.TCC.GGC  900
Glu-Pro-Asn-Tyr-Cys-Arg-Ile-Ala-His-Tyr-Gln-Asp-Pro-Ala-Tyr-Thr-Tyr-Ala-Ser-Gly  300

                    II
TCG.TAT.GTG.TGT|GGA.ATG.CGA.GCG.GAT.AAC.AAT.TCA.CAC|AGG.AAA.CAG.CTA.TGT.CCA.CGA  960
Ser-Tyr-Val-Cys|Gly-Met-Arg-Ala-Asp-Asn-Asn-Ser-His|Arg-Lys-Gln-Leu-Cys-Pro-Arg  320

TTA.CGC.CAA.GCT.CAA.AGG.AGA.AAA.ACC.GCT.TTG.TTC.GCG.ATT.GTA.GCT.GGC.GGC.TTT.CGT 1020
Leu-Arg-Gln-Ala-Gln-Arg-Arg-Lys-Thr-Ala-Leu-Phe-Ala-Ile-Val-Ala-Gly-Gly-Phe-Arg  340
AGA.CCT.AGT.GTC.GCT.ATG.TGC.GGA.GTG.CTG.GGA.CGT.GCG.GTG.GAG.TCA.AGA.CTG.CTG.GAC 1080
Arg-Pro-Ser-Val-Ala-Met-Cys-Gly-Val-Leu-Gly-Arg-Ala-Val-Glu-Ser-Arg-Leu-Leu-Asp  360
GGC.AGA.GTA.CTG.ACT.GAC.TGC.GAC.GGA.GCT.ACG.CGC.GGC.TAC.CAA.GGT.TGG.GGC.CAA.AGA 1140
Gly-Arg-Val-Leu-Thr-Asp-Cys-Asp-Gly-Ala-Thr-Arg-Gly-Tyr-Gln-Gly-Trp-Gly-Gln-Arg  380
GTT.GAG.TCA.ATT.AAT.AAA.TTG.CGT.TTG.GTA.TGC.GTT.TCA.GTG.GGA.AAC.CTG.TGG.CAG.TGG 1200
Val-Glu-Ser-Ile-Asn-Lys-Leu-Arg-Leu-Val-Cys-Val-Ser-Val-Gly-Asn-Leu-Trp-Gln-Trp  400

                                    III
GCA.TTA.ATG.AAT.GGC.AGG.GGG.AAG.GGG.TTG.GTA.TTG.GGC|AGG.TGG.TTC.TTA|CAT.AGA.GGA 1260
Ala-Leu-Met-Asn-Gly-Arg-Gly-Lys-Gly-Leu-Val-Leu-Gly|Arg-Trp-Phe-Leu|His-Arg-Gly  420

AGT.GAT.GCT.ACT.GCT.AAA.AGA.GCG.TCA.GTG.TGC.AGG.ATC.TTG.AGG.TGC.ATG.ATC.TTA.ATA 1320
Ser-Asp-Ala-Thr-Ala-Lys-Arg-Glu-Ser-Val-Cys-Arg-Ile-Leu-Arg-Cys-Met-Ile-Leu-Ile  440
AGT.AGC.GTA.GTA.TTT.CAT.GAG.CTT.AAG.GGT.CAG.AGG.AAC.TTC.GTG.CTG.CTC.AAT.TTG.GTG 1380
Ser-Ser-Val-Val-Phe-His-Glu-Leu-Lys-Gly-Gln-Arg-Asn-Leu-Val-Leu-Leu-Asn-Leu-Val  460
TCA.TAT.GGC.TCA.CCA.TGC.CAG.CGA.ATA.GTT.ATC.GGC.AGG.TTA.AGC.AGA.TGT.CGT.TAC.AGA 1440
Ser-Tyr-Gly-Ser-Pro-Cys-Gln-Arg-Ile-Val-Ile-Gly-Arg-Leu-Arg-Arg-Cys-Arg-Tyr-Arg  480
ACT.GGA.CTC.AGA.TGT.TAT.GAA.CGC.GAA.GTC.GTA.TTT.GCA.TTA.GCG.CTT.AGC.GCT.AAG.CGC 1500
Thr-Gly-Leu-Arg-Cys-Tyr-Glu-Arg-Glu-Val-Val-Phe-Ala-Leu-Ala-Leu-Ser-Ala-Lys-Arg  500

                                        IV
ACA.GCG.GTG.TGC.TGT.AGT.AGA.CGA.TTC.ACA.CGT.GGA|TTT.GTT.TCT.AAC.CAT.GAC|GAG.AAT 1560
Thr-Ala-Val-Cys-Cys-Ser-Arg-Arg-Phe-Thr-Arg-Gly|Phe-Val-Ser-Asn-His-Asp|Glu-Asn  520
```

```
GAT.TCC.TAT.ACA.CTC.ATG.CCT.CGC.TGC.TTA.GTA.TGT.GTA.GTC.CGA.AAG.CAA.CGA.ATA.TTA 1620
Asp-Ser-Tyr-Thr-Leu-Met-Pro-Arg-Cys-Leu-Val-Cys-Val-Val-Arg-Lys-Gln-Arg-Ile-Leu  540
CGA.GAT.GCT.CAA.ACA.CAA.GAA.CTT.TTA.GAA.GAA.GAG.ATA.GTA.GAT.AAA.ATC.AGA.TAT.AGT 1680
Arg-Asp-Ala-Gln-Thr-Gln-Glu-Leu-Leu-Glu-Glu-Glu-Ile-Val-Asp-Lys-Ile-Arg-Tyr-Ser  560
GAT.ATA.CGA.ATG.TCC.GAG.CTT.TTA.TTA.GAA.TAT.AAG.TGT.TAT.CGC.ACA.CTG.AGC.AAT.GCC 1740
Asp-Ile-Arg-Met-Ser-Glu-Leu-Leu-Leu-Glu-Tyr-Lys-Cys-Tyr-Arg-Thr-Leu-Ser-Asn-Ala  580
GAT.ACA.GAG.ATA.CGG.AGA.TAT.ATA.ATG.AAC.AAA.TAC.CAC.ACG.CGA.GCA.GAA.AAG.TGC.GTT 1800
Asp-Thr-Glu-Ile-Arg-Arg-Tyr-Ile-Met-Asn-Lys-Tyr-His-Thr-Arg-Ala-Glu-Lys-Cys-Val  600
CAA.CCA.GTG.AAA.CAA.ATA.GAA.GAA.AAG.CGA.TCG.CGG.TTA.TTA.AAA.TTG.CTA.AAA.GAA.AAT 1860
Gln-Pro-Val-Lys-Gln-Ile-Glu-Glu-Lys-Arg-Ser-Arg-Leu-Leu-Lys-Leu-Leu-Lys-Glu-Asn  620
AAA.TCT.TCC.CGT.CTT.AGA.TTA.GAC.CGA.GAT.AAT.AAA.TCT.AAA.GTG.AAG.GCA.CTT.TGG.CGA 1920
Lys-Ser-Ser-Arg-Leu-Arg-Leu-Asp-Arg-Asp-Asn-Lys-Ser-Lys-Val-Lys-Ala-Leu-Trp-Arg  640
GTG.AGC.CAA.CAC.GTA.AAA.GTT.CTA.GTG.AAA.CAG.GAG.GGA.GCG.ATG.GCA.ACA.AAG.TGC.CGT 1980
Val-Ser-Gln-Gln-Val-Lys-Val-Leu-Val-Lys-Gln-Glu-Gly-Ala-Met-Ala-Thr-Lys-Cys-Arg  660
GAA.AAT.ATT.ATA.GAG.TGC.GAA.AAT.TTA.TCG.TTC.GTG.AAA.AAA.CTT.TTT.TAT.ATA.CAA.AAT 2040
Glu-Asn-Ile-Ile-Glu-Cys-Glu-Asn-Leu-Ser-Phe-Val-Lys-Lys-Leu-Phe-Tyr-Ile-Gln-Asn  680
CTA.ACT.TTC.ACG.AGC.AGT.TTG.CAT.AAA.GAA.TTA.TAT.AGT.CTT.AGA.GTC.AGG.TTT.CGT.ATG 2100
Leu-Thr-Phe-Thr-Ser-Ser-Leu-His-Lys-Glu-Leu-Tyr-Ser-Leu-Arg-Val-Arg-Phe-Arg-Met  700
                                                    ScaI
TTT.AAC.ACC.AAG.AGG.ATT.AGC.TAT.TGG.ATC.GCA.TGT.TAT.TAT.AGA.AGT.ACT.GCT.TTA.TTG 2160
Phe-Asn-Thr-Lys-Arg-Ile-Ser-Tyr-Trp-Ile-Ala-Cys-Tyr-Tyr-Arg-Ser-Thr-Ala-Leu-Leu  720
TCA.GCA.TCT.TTT.TTT.AGA.ACA.GAA.ACA.GCT.GAG.GCT.TTC.GAT.CGA.AAT.CTT.GTT.GGT.TTC 2220
Ser-Ala-Ser-Phe-Phe-Arg-Thr-Glu-Thr-Ala-Glu-Ala-Phe-Asp-Arg-Asn-Leu-Val-Gly-Phe  740
ATA.CAA.GTA.AAT.CTT.ACT.CCC.ACC.TCG.GAC.ACA.TTT.CGT.CCC.CAA.CCT.CTA.TTT.CAT.AAG 2280
Ile-Gln-Val-Asn-Leu-Thr-Pro-Thr-Ser-Asp-Thr-Phe-Arg-Pro-Gln-Pro-Leu-Phe-His-Lys  760
ACG.ATT.AAC.ATC.TAC.CCT.AAG.TTT.TGG.CCC.GAT.GGC.CGC.AAT.TCC.GCA.TTA.CCC.TCT.CTC 2340
Thr-Ile-Asn-Ile-Tyr-Pro-Lys-Phe-Trp-Pro-Asp-Gly-Arg-Asn-Ser-Ala-Leu-Pro-Ser-Leu  780
CAT.GTT.GGA.TCC.GTC.TCT.TTT.AAT.GAG.GCG.GTT.TCG.CTT.TTT.CCT.CTC.GCC.GTG.TGG.TAT 2400
His-Val-Gly-Ser-Val-Ser-Phe-Asn-Glu-Ala-Val-Ser-Leu-Phe-Pro-Leu-Ala-Val-Trp-Tyr  800
TTG.TTC.ATT.ATA.TTA.TCA.CCG.TAT.CTT.CTG.TAT.CCG.GCA.TTT.CTC.CAG.TGT.GCG.AGA.AAC 2460
Leu-Phe-Ile-Ile-Leu-Ser-Pro-Tyr-Leu-Leu-Tyr-Pro-Ala-Phe-Leu-Gln-Cys-Ala-Arg-Asn  820
                                                        2508
ACC.TTA.TAT.TCT.ATA.AAA.GCC.TCG.GAT.CAT.CGC.ATA.TCT.AGA.TAT.TCC.TGA-TTTTATCATGA 2520
Thr-Leu-Tyr-Ser-Ile-Lys-Ala-Ser-Asp-His-Arg-Ile-Ser-Arg-Tyr-Ser-***              836
                                                          T C
ATCATGATCTATTTGCTTCTAAACTTCTTGTGTTTAGGCACTCTGTAATAATTCGTTGCTTTCGGATACACAATGAATG 2599
CAGCCAGAGGCAAGTGTGTATAGGAATCATTTCTCATAATATCTCCTTCACGCAAGAGCGTTATTCAATTCTTACGATT 2678
ACAAATAAGAGTAAA 2793
```

Fig. 2. Nucleotide sequence of the *RSDA* gene and its flanking region, and the amino acid sequence deduced from the *RSDA* gene.

The amino acid sequence, corresponding to the open reading frame beginning with the first ATG codon, is identified below the sequence. The 5′ initiation codon and 3′ termination codon are indicated with "I.C" and "T.C"(and ***), respectively. The putative −10 and −35 regions of the 5′-flanking promoter regions are also indicated by underline. "S.D." indicates the putative Shine-Dalgarno sequence. The N-terminal end of the mature protein is shaded. The N-terminal amino acid sequence(shaded) is preceded by a 34-amino acid signal sequence. Homologous regions I, II, III and IV, responsible for the enzyme activity, are blocked.The palindromin sequence of a possible transcription terminator is underlined with facing arrows. The numbers refer to the nucleotide (amino acid) position, with numbering starting at the first residue of the coding sequence. The last digit of each number is aligned with the relevant residue.

Finally hundreds of deleted plasmids, covering various lengths of the insert DNA, were isolated and kilo-sequenced.

Structure of the *RSDA* Gene

Fig. 2 illustrates the nucleotide sequence of the *RSDA* gene and its flanking regions. There was only one long

open reading frame (ORF) starting from an ATG codon (nucleotide +1 to +3). The ORF was composed of 2508 bp (836 amino acids). For the initiation of transcription, putative '−35' was observed at −56 (TTGATA) and '−10' was observed at −34 (TATATTA). The distance between these two regions (17 bp) correspond to the distances from the initiation sites and Pribnow boxes in other transcriptional sites of the proteins in gram-positive *Bacillus sublitis* (17). Furthermore, the TATATTA sequence is similar to the consensus sequence for the Pribnow box TATAAT (17). Upstream of these potential transcriptional initiation sites there were several inverted repeat structures in nucleotide sequences of −250 to −10 which were almost entirely A and T. Downstream of termination codons, TGA(+2509 to +2611) was detected. This termination codon TGA, seems to be the terminating signal for the *RSDA*. Downstream from '−10, as far as −250 bp, we observed three ATG codons (at +1, −17, and −146) in the beginning of the ORF. ATG at +1 was assumed to be the starting codon of the *RSDA* because a possible Shine-Dalarno sequence (near nucleotide −6 of GGAGG) (17) was observed immediately upstream from this ATG, and it was followed by a cluster of codons encoding basic and hydrophobic amino acids for secretional signaling. Following the initial methionine is a sequence of amino acids that appears similar to other identified signal sequence peptides. Beginning at amino acid +35, a potential cleavage site is located with the sequence Ala-Ala-Glu. Since this enzyme is a secretory protein it should possess a signal peptide which commonly has a cluster of hydrophobic amino acids. Neither the ATG at −146 nor the ATG at −135 has these neighboring characterisitics. If our assumption is correct, the enzyme precursor consists of 836 amino acid residues, corresponding to a molecular weight of 96,000 and a 5'-noncoding region of about 1 kb. These results are in good agreement with our previous results (10) which showed transformed *E. coli* (pYKA3) can make functional raw starch-digesting amylase in periplasmic space.

The deduced translated molecular weight of a 2,508 bp open reading frame (ORF), beginning with the first ATG(+1), agrees with the published size estimated of a raw starch-digesting enzyme protein (about 93,000 Da), if the predicted leader peptide is removed (29).

Amino Acid Composition of the RSDA Enzyme Produced in *E. coli* and Codon Usage

To check the amino acid composition of the pre-enzyme, the 93,000 Da active enzyme was purified from the periplasmic fraction of the *E. coli* cells carrying the *RSDA* gene, and the amino acid composition was analyzed. The result is shown in Table 1 together with the amino acid composition deduced from the DNA sequences. The amino acid composition of the purified enzyme was in good agreement with that of the putative mature form of the *RSDA* gene product deduced from the DNA sequence, if it is cleaved at Alanine at position +35.

Table 1. Amino acid composition of Raw Starch Digesting Amylase from *B. circulans* F-2

Amino acid (one-letter)	Amino acid analysis	From DNA sequence[a]
Ala(A)	60(7.1)	61(7.3%)
Cys(C)	29(3.6)	29(3.5)
His(H)	27(3.2)	25(3.0)
Met(M)	14(1.6)	14(1.7)
Thr(T)	50(6.0)	51(6.1)
Asn(N)	30(3.6)	33(3.9)
Asp(D)	29(3.4)	27(3.2)
Glu(E)	33(4.1)	34(4.1)
Gln(Q)	30(3.6)	33(3.9)
Ile(I)	43(4.9)	40(4.8)
Phe(F)	30(3.6)	92(11.0)
Trp(W)	N.D	15(1.8)
Pro(P)	25(2.9)	24(2.9)
Tyr(Y)	33(4.0)	35(4.2)
Arg(R)	96(12.0)	93(11.1)
Gly(G)	34(4.0)	32(3.8)
Lys(K)	35(4.3)	37(4.4)
Ser(S)	84(9.9)	78(9.3)
Val(V)	49(5.9)	50(6.0)

[a]The values in the parentheses are the number of individual amino acid residues charged when the first 33 amino acid residues from the N-terminus are omitted. The remaining residues did not change.
N.D., not determined.

Of the 64 codons, 62 codons were used in the *RSDA* gene product deduced from the DNA sequence of 59 codons in the preenzymes of *B. circulans* (21), *B. subtilis* (34), *B. amyloliquefaciencs* (34) and *A. awamorifrom* (5) (data not shown). These five proteins showed quite similar codon usage, except that the usage of UUA(Leu), CUU(Leu), AUA(Ile) and AGA(Arg) in *B. circulans* F-2 pre-enzyme were much higher than in the other proteins.

Identification of the N-terminus of the Pre-enzyme by Correlation of the DNA and Partial Amino Acid Sequences

To pinpoint the N-terminus of the pre-enzyme, the purified enzyme(purified to homogeneity from *E. coli* carrying plasmid pYKA3, specific activity 52 uints/mg protein) was subjected to sequential Edman degradation and five amino acid residues from the N-terminus were determined. As shown in Fig. 2, the five amino acid

residues from the N-terminus of the RSDA enzyme produced in *E. coli* were Ala-Glu-Thr-Tyr-His, which were identical to those deduced for this ORF (from No. 35 to 39). Thus, the precursor of the enzyme must be initiated at position +1 to +3, corresponding to the methionine at position +1. Therefore, we calculated that the signal sequences of the enzyme from *B. circulans* F-2 is 34 residues long.

Ribosome-binding Sites (RBS, Shine-Dalgarno), Transcriptional Region and Termination Region of the *RSDA* Gene

Fig. 3 details the sequence of the areas preceding and following the coding regions. The initial methionine is preceded by a possible ribosome binding site (RBS, Shine-Dalgarno) sequence (near nucleotide −6) that is similar to other such sequences seen in *B. subtilis* (15, 17). There is another possible ribosome binding site sequence (near nucleotide −36) which has homology with *B. subtilis* 16s RNA. If this sequence is used for RBS it will be followed closely by GTG, which can be used as an initiation codon in *B. subtilis* (17) and a termination codon (TGA) one codon later. These features are shown in detail in Fig. 3. These regions are of interest because a somewhat similar sequence of multiple ribosome binding site sequences, followed by potential initiation codons, has been postulated to play a role in the efficient translation of mRNA (12, 35). The calculated free energies for the interaction of RBS with the 3′-end of *B. subtilis* 16S rRNA are −13.7 to −21.3 Kcal/mole, which is within the range of the more stringent RBS normally seen for Gram-positive genes (18).

The region adjacent to the area of the 3′-end of the *RSDA* gene was analyzed by the SDC computer system. Following the termination codon of the enzyme (nucleotides 2509-2511) there are four inverted structures which could play a role in mRNA transcription termination. These potential terminator sequences composed of an AT-rich inverted repeat, are found in the region between nucleotides 50 and 200 downstream from the termination codon. The second of these sequences begins at nucleotide 57, is 16 nucleotides long, and has a free energy of 16.3 Kcal/mol (30). This sequence is followed by a T-rich region which is similar to termination structures seen in *E. coli* (25). Such sequences are also found in several genes from Gram-positive organism *B. subtilis* (20, 22, 32), but the importance of these sequences in transcription termination in Gram-postive bacteria has not yet been eatablished.

Homologous Regions of Active Site-related Amino acid Sequences of Some Bacterial and Fungal Amylases

To compare the amino acid sequences of the various different enzymes produced by some bacterial and fungal strains, some homologous regions of consensus amino acid sequences I, II, III and IV, which are considered to be the regions responsible for enzyme active sites, were analyzed by the Gene Bank computer data base. No notable similarities were seen in the coding regions for the structural genes. However, when we compared the deduced sequence of our enzyme with those of other amylases published, we were able to find four homologous regions. In these regions active site-related amino acids were conserved in all sequences (Fig. 4). One interesting feature in these homologous regions of

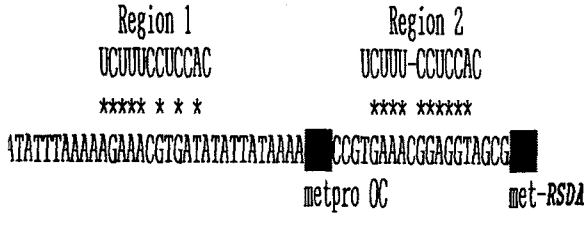

Fig. 3. Possibble ribosome binding sites preceding the *RSDA* gene.

The regions immediately preceding the coding sequence of the *RSDA* gene, that showing complemetarity to the 3′-end of *Bacillus subtilis* 16S rRNA, are shown. The possible pairing with the 3′-end of 16S rRNA is indicated above each sequence(RBS) by "****" indicating matching bases (3). Below the DNA sequence the possible translation start sites and the stop codons of the *RSDA* gene are shown. The calculated free energies of interaction(ΔG) for region 1 and 2 are −13.7kcal and −21.3 kcal/mole, respectively, using the rules of Tinoco *et al.* (1973) (30).

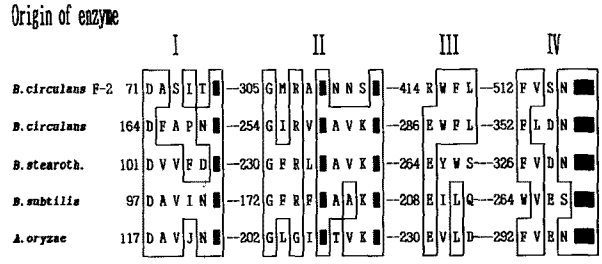

Fig. 4. Homologous regions of the amino acid sequence of some bacterial and fungal amylases.

Amino acid residues are shown as one-letter abbreviations. Numbering of the amino acids starts from the amino terminal of the mature enzymes, except for *B. circulans* and *B. circulans* F-2, which starts from the amino terminal methionine of the putative precursor. The residues located at active centers of Taka-amylase A are shaded. Roman numerals show homologous regions suggested by Nakajima *et al.* (1986) (19).

Origin of Enzyme	T-S rich region	No. T & S residues/25a.a	%
B. circulans F-2	727 [redacted]	6	24
S. diasiticus	481 [redacted]	7	28
A. oryzae	500 [redacted]	17	68
A. awamori	446 [redacted]	18	72

Fig. 5. Threonine(T) and Serine(S)-rich regions of amino acid sequences of some bacterial and fungal raw starch-digesting enzymes.
Threonine and Serine residues are shaded.

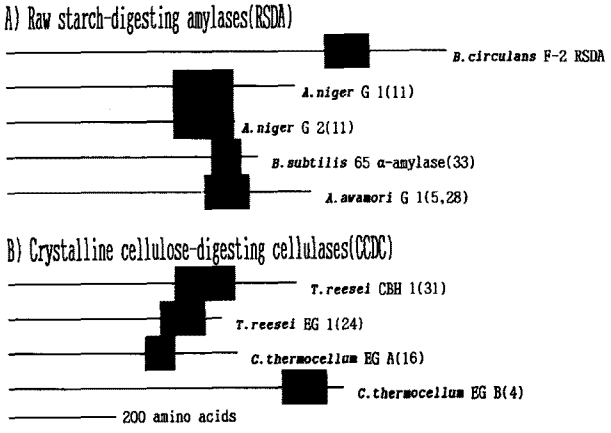

Fig. 6. Schematic structures of Ser-Thr-rich regions from different raw starch-digesting amylases and crystalline cellulose-digesting cellulases.
Each enzyme consists of a catalytic domain probably linked to the C-domain that is most likely responsible for disrupting the structure of the semi- crystalline substrate for cellulase, and the raw starch substrate for amylase. The drawing is only roughly to scale (11). The bar on the left side indicates the hydrolytic domain; the shaded bar indicates the raw substrate specific domain (by O-glycosylation); G 1 or 2, glucoamylase 1 and 2; CBH 1, cellobiohydrolase 1; EG 1, endoglucanasae 1; EG A and B, endoglucanase A and B.

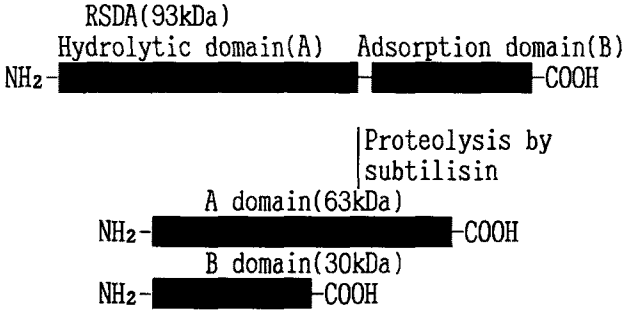

Fig. 7. Domain structure of the *B. circulans* F-2 RSDA and proteolytical separation of each domain.
The RSDA protein can be converted to the two domains as previously reported (8).

each amylase is that the four homologous regions are located only in the N-terminal region (two of third) (Fig. 2).

Possible Hydrolytic and Adsorption Domain of the Enzyme

In previous studies (8, 9) we treated the RSDA with proteolytic enzymes to investigate raw starch digestibility. When this enzyme was digested with subtilisin it was converted into two components of the N-domain and the C-domain. The N-domain was a portion which contained the active site to hydrolyze soluble substrates, but not raw materials. However, the C-domain showed only an adsorption ability onto the raw substrate without any hydrolytic activity. Therefore, a bifunctional organisation of the RSDA was proposed. One domain, corresponding to the carboxyl-terminal of the enzyme, was implicated in the adsorption process. Whereas the other contained the hydrolytic function. As shown in Fig. 5, the *B. circulans* F-2 RSDA contains a Ser-Thr-rich region, strongly suggesting that the RSDA has two domains in its structure. This result supports the model system previously proposed to explain the relationship between the raw starch-digestibility and the molecular structure of the enzyme (Fig. 6 and 7). Furthermore, the Ser-Thr-rich region was located on the C-terminus region of the protein.

DISCUSSION

We found only one open reading frame, which begain at the ATG codon and terminated at the TGA codon at 2509 to 2611. This coding sequence is unambigously the *RSDA* structural gene because: (a); the canonical Shine-Dalgarno(SD) ribosome binding sequence GGAGG, which is commonly observed in Gram-positive ribosome binding sites (3, 15, 17), precedes 4 nucleotides upstream from the initiation codon: (b); the calculated molecular weight of the predicted polypeptide (Mr=96, 727) is consistent with that of the raw starch-digesting amylase (about 93,000) purified from the donor strain *B. circulans* F-2, as reported previously (10, 29): (c); the amino acid residue Ala-Glu-Thr-Tyr-His, from the amino terminus of the enzyme produced in *E. coli*, was identical to that deduced for this ORF.

The putative promoter sequence of the *RSDA* gene has a strong homology with the consensus sequence transcribed by 55 RNA polymerase(TTGATA) for "−35", and the TATATTA sequence for "−10" (17) (Fig. 3). Other candidates for the −35 and −10 region were ATTCTGA (positions −139 through −133) and ATAAATA (positions −115 through −109), which corresponded to the σ^{55} RNA polymerase recognition sequence and were identical to the consensus sequence.

However, this is too long for the transcriptional start site of ATG. The distance between the −35 and −10 region of these two putative promoters is 17 bp, which is the prepared spacing of σ^{55} and σ^{43} RNA polymerases (17). As this σ^{55}-type promoter sequence is also homologous to the consensus sequence of *E. coli*, we anticipated that this promoter also functioned in *E. coli* and, thus, hampered our attempt to subclone the *RSDA* gene in *E. coli*, as described previously (10).

Even though the two sequences, GGAGG and GTGATA, with free energies of −21.3 and −13.7 Kcal /mole, respectively, existed on the sequence, the sequence GGAGG was more highly complementary to the 3′-end of *B. subtilis* 16S rRNA than GTGATA (18). Although the precise translation initiation site has not been determined, the enzyme must be translated from the ATG codon located 4 bp downstream of this putative RBS(GGAGG). The *RSDA* gene of *B. circulans* F-2 is, at least, followed by a typical prokaryotic transcription terminator sequence (25). Of the four putative terminators, the calculated free energy for the corresponding RNA hairpin structure (30) is −18.6 Kcal/mole.

The putative signal sequence was 34 amino acid residues long, with three basic amino acids near the N-terminal end followed by a hydrophobic region and Alanine at the C-terminus. These observations are consistent with the signal peptides from other Gram-positive bacteria (22, 32, 28).

Although the plasmid pHA300-encoded RSDA enzyme had the same properties as raw starch-digesting amylase from *B. circulans* F-2 (10), its Mr, deduced from the sequence data, was slightly higher (93,000 vs 96, 000) than that of the donor strain. This result suggested that the difference might be due to processing of the protein in the extraenzyme system of *B. circulans* F-2.

Several years ago Nishizawa *et al.*(21) compared the amino acid sequences of different α-amylases with that of *B. circulans* and reported the presence of six homologous regions; A, B, I, II, III and IV. Of these, four highly homologous regions(named as consensus regions I, II, III and IV) were the same as those reported by Nakajima *et al.* (19). A comparison of the nucleotide sequence and the deduced protein sequence of the *B. circulans* F-2 *RSDA* and the amino terminal portion of other *Bacillus* species amylases showed no obvious regions of homology. But, when we compared the deduced sequence of our RSDA enzyme with the consensus regions of α-amylase (19, 21) we were able to find the consensus homologous regions I to IV in the *B. circulans* F-2 amylase (Fig. 4). However, we could not find regions A and B. The His and Asp residues in regions II and IV may be in the active site of Taka-amylase A (14), suggested by low angle X-ray diffraction studies. They are also found in *B. circulans* F-2 raw starch-digesting amylase. This result, taken together with the results of the ORF determination, suggests that the ORF is likely to code an amylase of the α-amylase type.

As repored previously (8, 29) the *B. circulans* F-2 RSDA showed efficient adsorbability and digestibility of raw potato starch granules, which were most resistant to enzymatic degradation. This *B. circulans* F-2 enzyme was shown to be the best source to digest raw potato starch. Some enzymes were shown to digest raw materials as substrates; cellulases of *T. reesei* (36), *C.thermocellum* (1, 7) and *A. neger* (23) and glucoamylases of *A. niger* (2, 27), *A. awamori* (5, 28) and our RSDA enzyme are examples. The homologous terminal domains of *T. reesei* cellulase and *B. subtilis* 65 amylase, including the Serine and Threonine-rich sequence, could be removed by limited proteolysis (33, 36). Seemingly, our RSDA was converted to an N-domain and a C-domain. Of these two the N-domain, a truncated enzyme form, showed the same hydrolytic activity on soluble substrates as the original enzyme. The activity of these truncated enzymes towards microcrystalline cellulose and raw starch are impaired, but the activity towards small soluble substrates are not affected. Thus, It is suggested that the hydrolytic active sites must be localized. Consensus regions I through IV, which are in the N-terminal region as shown in Fig. 4, are proposed as the active site for the enzyme). The C-terminal domain regions seem to have a role in substrate binding and solubilization. It is of interest to note that fungal glucoamylases and bacterial α-amylases (including our amylase which is capable of digesting raw substrates) have, at least superficially, a similar structure to cellulases. A C-terminal domain (Serine and Threonine-rich region) seems to be responsible for binding to raw starch, and this is also linked to the hydrolytic domain by a region rich in Serine and Threonine (2, 27). The presence of extensive glycosylation (O-glycosylation in Serine and Threonine) in most fungal and many bacterial cellulases seems to suggest a functional rule for raw carbohydrate. The *T. reesei* (36) cellobiohydrolases, *A. awamori* var. *kawachi* (5) glucoamylases and α-amylase of *B. subtilis* 65 have been shown to be localized to the conserved C-terminal regions. Our cloned *RSDA* gene also carried the Ser-Thr-rich area in the C-terminal region of the gene product. Even the amount of the Serine and Threonine residues is less than in other cellulases and glucoamylases (Fig. 5 and 6). This structural basis indicates that the raw substrate-adsorbing ability(or affinity) of *B. circulans* F-2 amylase may result from the Ser-Thr-rich region.

Acknowledgement

Authors are highly indebted to the late Prof. Y. Maruyama, Laboratory of Biological Chemistry, Dept of Agricultural Chemistry, University of Tokyo, Japan for his kind direction on this study.

REFERENCES

1. **Beguin, P., Cornet, P. and Aubert, J-P.** 1985. *J. Bacteriol.* **162**: 102-105.
2. **Boel, E., Hjort, I., Svensson, B., Norris, F., Norris, K.E. and Fui, N.P.** 1984. *EMBO. J.* **3**: 1097-1102.
3. **Duggleby, C. and Jones, S.A.** 1983. *Nucleic Acids Res.* **11**: 3965-3076.
4. **Grepinet, O. and Beguin, P.** 1986. *Nucleic Acids Res.* **14**: 1791-1799.
5. **Hayashida, S., Kuroda, K., Ohta, K., Kuhar, S., Fukuda, K. and Sakaki, Y.** 1989. *Agric. Biol. Chem.* **53**: 923-929.
6. **Henikoff, S.** 1984. *Gene.* **28**: 351-359.
7. **Joliff, P., Cornet, P. and Aubert, J-P.** 1985. *J. Bacteriol.* **162**: 102-105.
8. **Kim, C-H., Zhang, G-W., Y. Maruyama and H. Taniguchi.** 1990. *Agric. Biol. Chem.* **54**: 2767-2768.
9. **Kim, C-H. and H. Taniguchi.** 1992. *Biochim. Biophys. Acta.* Submitted.
10. **Kim, C-H., H. Sata, H. Taniguchi and Y. Maruyama.** 1990. *Biochem. Biophys. Acta.* **1048**: 223-228.
11. **Knowles, J., Lehtovaara, P. and Teeri, T.** 1987. *Trends in Biotechnol.* **5**: 255-261.
12. **Losick, R. and Perr, J.** 1981. *Cell.* **25**: 582-584.
13. **Maniatis, T., E.F. Fritsch and J. Sambrook.** 1982. in Laboratory Mannual for Molecular Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 164 and 178.
14. **Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M.** 1984. *J. Biochem.* **95**: 697-702.
15. **Mclaughlin, J.R., Murray, C.L. and Rabinowitz, J.C.** 1981. *J. Biol. Chem.* **256**: 11283-11291.
16. **Millet, J.** 1987. *FEMS. Microbiol. Lett.* **29**: 145-154.
17. **Moran, C.P., Jr. N. Lang, S.F.J. Legrice, G. Lee, M. Stephens, A.L. Sonenshein, J. Pero, and R. Losick.** 1982. *Mol. Gen. Genet.* **186**: 339-346.
18. **Murray, C.L. and Rabinowitz, J.C.** 1982. *J. Biol. Chem.* **257**: 1053-1062.
19. **Nakajima, R., Imanaka, T. and Aiba, S.** 1986. *Appl. Microbiol. Biotechnol.* **23**: 355-360.
20. **Neugebauer, K., Sprengel, R. and Schaller, H.** 1981. *Nucleic Acids Res.* **9**: 2577-2589.
21. **Nishizawa, M., Ozawa, F. and Hishinuma, F.** 1987. *DNA.* **6**: 255-265.
22. **Palva, I., Pettessen, R.F., Kalkkinen, N., Lehtovaara, P., Sarvas, M. and Soderlund, H.** 1981. *Gene.* **15**: 43-51.
23. **Penttila, M.E.** 1984. *Mol. Gen. Genet.* **194**: 494-499.
24. **Penttila, M., Lehtovaara, P., Nevalainen, H., Bhikhakhai, R. and Knowles, J.** 1986. *Gene.* **30**: 253-263.
25. **Rosenberg, M. and Court, D.** 1979. *Ann. Rev. Genet.* **13**: 319-353.
26. **Sanger, F., S. Nicklen and A.R. Coulson.** 1977. *Proc. Natl. Acad. Sci. USA.* **74**: 5463-5467.
27. **Svensson, B., Larsen, K., Svendson, I. and Boel, E.** 1983. *Carlsberg Res. Commun.* **48**: 1097-1102.
28. **Tada, S., Iimura, Y., Gomi, K., Takahashi, K., Hara, S. and Yoshizawa, K.** 1989. *Agric. Biol. Chem.* **53**: 593-599.
29. **Taniguchi, H., M.J. Chung, N. Yoshigi and Y. Maruyama.** 1983. *Agric. Biol. Chem.* **47**: 511-519.
30. **Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J.** 1973. *Nature New Biol.* **246**: 40-41.
31. **Teeri, T., Salovouri, I. and Knowles, J.** 1983. *Bio/Technology,* **1**: 696-699.
32. **Tomoyuki, S. and N. Tsuchida.** 1983. *Nucleic Acids Res.* **11**: 7679-7693.
33. **Teramoto, Y., Kira, I. and Hayashida, S.** 1989. *Agric. Biol. Chem.* **53**: 601-605.
34. **Yamasaki, H., Ohmura, K., Nakayama, A., Takaichi, Y., Otozai, K., Yamasaki, M., Tamura, G. and Yamane, K.** 1983. *J. Bacteriol.* **156**: 327-337.
35. **Yang, M., Galizzi, A. and Henner, D.** 1983. *Nucleic Acids Res.* **11**: 237-249.
36. **Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhi, R. and Petterson, G.** 1986. *FEBS Lett.* **204**: 223-227.