Agric. Biol. Chem., **50** (4), 957~964, 1986

Rhizopus Raw-Starch-Degrading Glucoamylase: Its Cloning and Expression in Yeast

Toshihiko ASHIKARI, Norihisa NAKAMURA, Yoshikazu TANAKA, Naoko KIUCHI, Yuji SHIBANO, Takaharu TANAKA, Teruo AMACHI and Hajime YOSHIZUMI

Laboratory of Applied Microbiology, Research Center, Suntory Ltd., 1-1-1 Wakayamadai, Shimamoto-cho, Mishima-gun, Osaka 618, Japan

Received October 1, 1985

A glucoamylase gene has been cloned from a *Rhizopus* genomic DNA library using synthetic oligonucleotides corresponding to the amino acid sequence of the glucoamylase. Since this glucoamylase gene was not expressed in yeast cells, we have cloned a glucoamylase gene from a cDNA library prepared from *Rhizopus* mRNA. Sequence analysis of both glucoamylase genes revealed that the genomic gene contained 4 intervening sequences and the cDNA gene lacked 145 nucleotides corresponding to the N-terminal region. The glucoamylase consists of 604 amino acids including a putative signal peptide and its molecular weight was calculated to be 65,000. The glucoamylase gene to be expressed in yeast cells was constructed by recombination of both genes. The yeast cells containing this constructed glucoamylase gene secreted the glucoamylase into the culture fluid and grew at almost the normal rate on a medium containing starch as the sole carbon source.

Glucoamylases (EC 3.2.1.3) hydrolyze α(1,4)- and α(1,6)-glycosidic linkages from the nonreducing ends of starch and related maltooligosaccharides to release glucose. These enzymes have been isolated from the culture fluid of a variety of fungi and yeasts, and their properties and biological significance have been studied.[1] Generally, fungal glucoamylases are known to exist in multiple forms varying in size. A few distinct forms of *Aspergillus* glucoamylase are produced by differential RNA splicing events.[2,3] On the other hand the multiple forms of *Rhizopus* glucoamylase are thought to arise from limited proteolysis.[4,5] Recently, the genes coding for the glucoamylases of *Asp. awamori*[2] and *Asp. niger*[3] were cloned and it was discovered that their structures were the same.

Glucoamylases are especially useful in saccharifying starch in commercial alcohol production. It is unfortunate for this commercial use that *Rhizopus* produces large amounts of glucoamylase in solid culture but not in submerged culture despite extremely strong hydrolyzing activity toward raw starch compared to other glucoamylases. A strain of yeast which could produce glucoamylase would be very useful in industrial alcohol production, because it could directly convert starch into ethanol. In this report we describe the cloning and the expression of the *Rhizopus* glucoamylase gene in yeast. Several features of the glucoamylase gene are also described.

MATERIALS AND METHODS

(a) *Microorganisms.* *Rhizopus oryzae* SAM0034, a strain producing high-level glucoamylase, was selected from our culture collection. *E. coli* WA802 (*metb-1 lacY galK-2 galT-22 λ⁻ supE44 hsd-3*) was used as a bacterial host. *Saccharomyces cerevisiae* XS-30-2B (MATα *leu2 his3 ura3 trp1*), consturcted in our laboratory, was used as a yeast host and yeast cells were transformed by the method of Ito *et al.*[6]

(b) *Amino acid analysis of glucoamylase.* Glucoamylase was purified from *Rhizopus oryzae* by the method of Tsujisaka *et al.*[7] Amino acid composition and N-terminal sequence of the purified glucoamylase were determined as described previously.[8] Glucoamylase was cleaved by

BrCN and succinylated, followed by digestion with trypsin, and the peptides were purified from the digest and sequenced as described.[8] The C-terminal residue of glucoamylase was identified by carboxypeptidase A digestion.

(c) *Construction of the Rhizopus genomic DNA library.* Chromosomal DNA was isolated from fully sporulated *Rhizopus oryzae*. The spores were disrupted by a Dinomill apparatus (Willy A. Bachofen, Basel, Switzerland) with glass beads (0.25~0.5 mm) in an 0.15 M NaCl–0.05 M EDTA buffer at 4°C for 15 seconds and DNA was isolated by the method of Cryer *et al.*[9] The molecular weights of the isolated DNAs were estimated by agarose gel electrophoresis to be 1×10^6 to 10^7 daltons.

The DNAs were digested with *Hin*dIII and fractionated by sucrose density gradient centrifugation. A sample of each fraction was analyzed by both agarose gel electrophoresis and dot-hybridization to 14-mer synthetic probes. The strongest hybridizing fraction, containing approximately 4 kb DNA fragments, was used to construct the genomic gene library in *E. coli* using pBR322 as a vector.

(d) *Construction of the Rhizopus cDNA library.* The total cellular RNA was isolated and purified from *R. oryzae* mycelia by the method of Chirgurin *et al.*[10] with slight modifications. Thirty grams (wet weight) of mycelia grown on YpSs agar (0.4% yeast extract, 0.1% KH_2PO_4, 0.05% $MgSO_4 \cdot 7H_2O$, 1.5% soluble starch, and 2% agar) was suspended in 50 ml guanidium thiocyanate stock solution. After addition of deionized water to a total volume of 100 ml with stirrig at room temperature, 50 ml of phenol–chloroform–isoamylalcohol (20 : 19 : 1) was added to the suspension with vigorous stirring with a Bio-mixer (Nihonseiki, Tokyo, Japan). The homogenate was centrifuged for 30 min at $1500 \times g$ and the aqueous phase containing the total RNA was pooled. The RNA was recovered from the guanidium thiocyanate homogenate by ultracentrifugation through a dense cushion of cesium chloride and mRNA was purified by oligo-(dT)-cellulose column chromatography as described by Maniatis *et al.*[11] The cDNA library was constructed by the method of Okayama and Berg[12] in *E. coli* WA802 using 15 μg of poly(A) RNA and 4.2 μg of vector-primer. Approximately 100,000 ampicillin-resistant transformants were obtained.

(e) *Other materials and methods.* Preparation of *E. coli* plasmids, transformation, analysis by restriction enzymes, nick translation of probes, and Southern hybridization were done as previously described.[11] The synthetic 14-mer probes were labeled by the methods of Donis-Keller *et al.*[13] Colony hybridization was carried out by the method of Grunstein *et al.*[14] The hybridization conditions were 18 hr at 40°C for the 14-mer mixed probes and 65°C for the probe derived from the genomic gene. The nucleotide sequence was found by the dideoxy method using M13mp10 and M13mp11.[15] *In vitro* mutagenesis was done by the method for oligonucleotide-directed site-specific mutagenesis as described by Morinaga *et al.*[16]

RESULTS

(a) *Chemical synthesis of DNA probes*

Some tryptophan containing peptides derived from *Rhizopus* glucoamylase were sequenced. One of these sequences, Thr-Trp-Ser-His-Ala-Ser-Leu-Ile-Thr-Ala-Ser, was used for designing the synthetic oligonucleotide probes. The oligonucleotides, 5′-ACNTGGTCNCAQGC-3′, were synthesized by the modified triester method (Q; T or C, N; any base).[17] One of these sequences was specifically hybridized to the glucoamylase gene in the region coding for Thr-Trp-Ser-His-Ala.

(b) *Cloning of the genomic glucoamylase gene*

The *Rhizopus* genomic gene library was screened by colony hybridization using the ^{32}P-labeled 14-mer mixed probes. Among approximately 1000 transformants, clones 39, 83, and 93, which contained an identical 4.3 kb HindIII fragment, were selected and clone 39 was used for further analysis. Figure 1 shows the restriction map of the 4.3 kb *Hin*dIII insert of pRGA39 isolated from clone 39. To elucidate the region hybridizing the 14-mer probes, pRGA39 was digested with various restriction enzymes and Southern hybridization experiments were performed. The result indicates that the region hybridized to probes was located on a 0.3 kb fragment between KpnI and DraI sites (Fig. 1). Single stranded M13 DNAs were prepared from M13mp10 and M13mp11 RF DNAs inserted by an 0.8 kb *Bgl*II-*Dra*I fragment at the *Sma*I-*Bam*HI sites. Only the M13mp11 single-stranded DNA hybridized to the 14-mer mixed probes, suggesting that the transcription of the glucoamylase gene occurs from the *Bgl*II site to the *Dra*I site (Fig. 1).

(c) *Cloning of cDNA glucoamylase gene*

The cDNA library constructed from *Rhizopus* mRNA was screened by the hybridization with a 2.0 kb DraI fragment of the

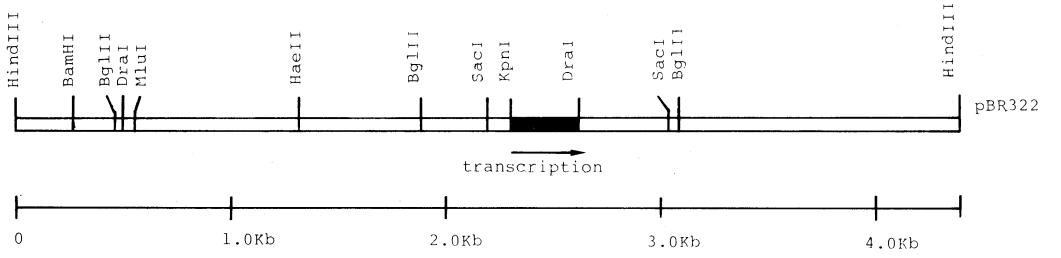

FIG. 1. The Physical Map of the Genomic Glucoamylase Gene.

The restriction map of the 4.3 kb *Hin*dIII insert of pRGA39 is presented. The closed box indicates the region hybridized to glucoamylase-specific synthetic 14-mer probes. The arrow shows the direction of transcription found by the method described in the text.

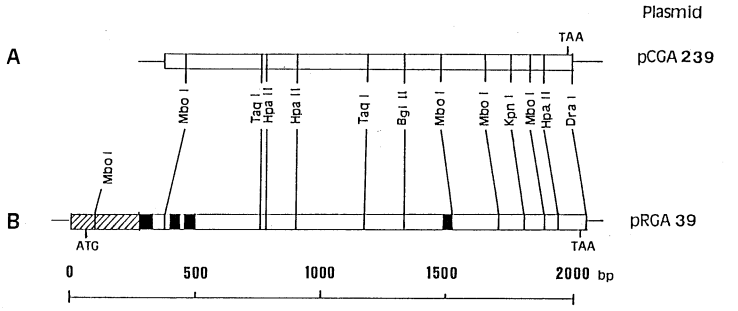

FIG. 2. The Restriction Maps of Glucoamylase Genes.

The restriction sites are shown for the cDNA (A) in pCGA239, and the genomic gene (B) in pRGA39. The shaded boxes represent the region lacking in the cDNA region and closed box represents the intervening sequences. ATG and TAA indicate the initiation and termination codons of the glucoamylase gene, respectively.

Rhizopus genomic glucoamylase gene. Several clones containing glucoamylase cDNA were obtained. One of these clones contained pCGA239, which had the largest cDNA insert (2.0 kb), covering almost the entire coding sequence of the glucoamylase. The restriction map of the glucoamylase cDNA in pCGA239 is shown in Fig. 2 with that of the genomic glucoamylase gene in pRGA39. Comparison of these two maps revealed that both genes had almost the same restriction sites although the gene sizes were different. These results suggested that the glucoamylase gene had at least one intervening sequence.

(d) *The nucleotide sequences of glucoamylase genes*

The nucleotide sequences of the glucoamylase genes derived from genomic DNA and cDNA are shown in Fig. 3. Comparison of the sequences reveals that the genomic gene contains 4 intervening sequences and the cDNA gene lacks 145 nucleotides coresponding to the N-terminal amino acid sequence. A large open reading frame which started from ATG at position 115, continuing to TAA at position 2160 and interrupting by four intervening sequences was the best candidate for the glucoamylase.

The deduced amino acid sequence corresponds completely to the known amino acid sequence of the peptides from glucoamylase (Fig. 3). For instance, the N-terminal amino acid sequence of the mature glucoamylase, Ala-Ser-Ile-Pro-, was found at residues 26 ~ 29, suggesting that the glucoamylase contains a signal peptide. The putative signal peptide contains 16 hydrophobic amino acids including N-terminus methionine. The basic amino acid lysine is located at the 9th position from

```
        10          20          30          40          50          60          70          80          90          100         110
GATCTCAATT TGTGTTGTGA TATATTCAGA TTTAAAATTT CAAACATATA TAAGACGCGT TTATTTCCTC GTTTTTCAAA AATCATCACT TGTCTTCAAA TTGATCTTTC TCTA ATG
                                                                                                                          MET
120                         150                         180                         210
CAA  CTG TTC AAT TTG CCA TTG AAA GTT TCA TTC TTT CTC GTC CTC TCT TAC TTT TCT TTG CTC GTT TCT GCT GCA AGC ATT CCT AGT AGT  GCT TCT
GLN  LEU PHE ASN LEU PRO LEU LYS VAL SER PHE PHE LEU VAL LEU SER TYR PHE SER LEU LEU VAL SER ALA Ala Ser Ile Pro Ser Ser  Ala Ser

                            240                         277                 297                         317
GTC CAG CTT GAT TCA TAC AAT TAC GAT GGC TCT ACT TTT TCA GGA AAA ATT TAT gtaggttgca tataattcaa acattaaaaa aatattaatg ttcatgctgt ac
Val Gln Leu Asp Ser Tyr Asn Tyr Asp Gly Ser Thr Phe Ser Gly Lys Ile Tyr

                        341                         371                         401
gctgtag GTC  AAG AAC ATT GCT TAC TCC AAG AAG GTT ACT GTA ATT TAC GCC GAT GGC TCT GAC AAC TGG AAT AAT AAT GGA AAC ACC ATT  GCT GCT
        Val  Lys Asn Ile Ala Tyr Ser Lys Lys Val Thr Val Ile Tyr Ala Asp Gly Ser Asp Asn Trp Asn Asn Asn Gly Asn Thr Ile  Ala Ala

                    431                         461                         491
TCT TAC TCT GCT CCT ATT TCT GGA TCA AAT TAC GAA TAC TGG ACA TTC TCT GCC TCC ATT AAT GGT ATC AAG GAG TTC TAC ATT AAG gtaacttatt tt
Ser Tyr Ser Ala Pro Ile Ser Gly Ser Asn Tyr Glu Tyr Trp Thr Phe Ser Ala Ser Ile Asn Gly Ile Lys Glu Phe Tyr Ile Lys

    523                 543                 563                 581                         611
tactttat gatatttgcc cttatactta attaactaac ccttttttct ctatag TAT GAG GTC  AGT GGA AAA ACA TAC TAT GAT AAC AAC AAT TCT GCC  AAT TAC
                                              Tyr Glu Val  Ser Gly Lys Thr Tyr Tyr Asp Asn Asn Asn Ser Ala  Asn Tyr

                636                 656                 680                         710
CAA Ggt aaagataata atatacaaat gcctacaact attcacattt ttatagTA TCT ACA TCC AAG CCT ACT ACT ACT ACT GCT ACT GCT ACT ACT ACT  ACC GCT
Gln Val                                                 Ser Thr Ser Lys Pro Thr Thr Thr.Thr Ala Thr Ala Thr Thr Thr  Thr Ala

            740                 770                 800
CCT TCC  ACT TCA ACC ACG ACT CCC CCC TCA AGC TCT GAG CCA GCT ACT TTC CCA ACT GGT AAC TCT ACA ATC TCC TCA TGG ATT AAG AAG  CAA GAA
Pro Ser  Thr Ser Thr Thr Thr Pro Pro Ser Ser Ser Glu Pro Ala Thr Phe Pro Thr Gly Asn Ser Thr Ile Ser Ser Trp Ile Lys Lys  Gln Glu

        830                 860                 890
GGT ATC  AGC CGC TTT GCT ATG CTT CGA AAC ATC AAT CCT CCT GGA AGC GCT ACC GGT TTC ATT GCT GCC TCA CTC TCT ACC GCT GGT CCC  GAT TAC
Gly Ile  Ser Arg Phe Ala Met Leu Arg Asn Ile Asn Pro Pro Gly Ser Ala Thr Gly Phe Ile Ala Ala Ser Leu Ser Thr Ala Gly Pro  Asp Tyr

920                         950                         980                         1010
TAC TAT  GCT TGG ACT CGT GAT GCT GCA TTA ACC TCC AAT GTA ATT GTT TAC GAA TAC AAC ACT ACT TTG TCC GGT AAT AAG ACT ATC CTC  AAC GTC
Tyr Tyr  Ala Trp Thr Arg Asp Ala Ala Leu Thr Ser Asn Val Ile Val Tyr Glu Tyr Asn Thr Thr Leu Ser Gly Asn Lys Thr Ile Leu  Asn Val

                    1040                        1070                        1100
CTC AAG  GAC TAT GTT ACA TTC TCA GTC AAG ACC CAA TCA ACT TCT ACC GTC TGT AAC TGC CTT GGT GAG CCT AAG TTC AAT CCT GAT GCT  TCT GGC
Leu Lys  Asp Tyr Val Thr Phe Ser Val Lys Thr Gln Ser Thr Ser Thr Val Cys Asn Cys Leu Gly Glu Pro Lys Phe Asn Pro Asp Gly  Ser Gly

            1130                        1160                        1190
TAT ACT  GGT GCT TGG GGA AGA CCT CAA AAT GAT GGA CCT GCT GAA CGT GCT ACT ACC TTC ATT TTG TTT GCT GAC AGT TAT CTT ACT CAA  ACA AAG
Tyr Thr  Gly Ala Trp Gly Arg Pro Gln Asn Asp Gly Pro Ala Glu Arg Ala Thr Thr Phe Ile Leu Phe Ala Asp Ser Tyr Leu Thr Gln  Thr Lys

        1220                        1250                        1280
GAT GCT  TCC TAT GTC ACT GGT ACA CTC AAG CCT GCT ATC TTC AAG GAC TTG GAC TAT GTC GTC AAT GTC TGG TCT AAT GGC TGT TTC GAT  TTA TGG
Asp Ala  Ser Tyr Val Thr Gly Thr Leu Lys Pro Ala Ile Phe Lys Asp Leu Asp Tyr Val Val Asn Val Trp Ser Asn Gly Cys Phe Asp  Leu Trp

    1310                        1340                        1370
GAA GAA  GTC AAC GGT GTT CAC TTC TAT ACT TTA ATG GTT ATG CGT AAG GGT TTG CTT CTT GGT GCA GAT TTC GCT AAA CGT AAC GGT GAC  TCT ACT
Glu Glu  Val Asn Gly Val His Phe Tyr Thr Leu Met Val Met Arg Lys Gly Leu Leu Leu Gly Ala Asp Phe Ala Lys Arg Asn Gly Asp  Ser Thr

                1430                        1460                        1490
CGT GCA  TCT ACC TAT AGC AGC ACT GCA TCC ACT ATT GCA AAC AAG ATC TCT AGC TTC TGG GTT TCT TCT AAT AAC TGG ATT CAA GTC AGT  CAA AGC
Arg Ala  Ser Thr Tyr Ser Ser Thr Ala Ser Thr Ile Ala Asn Lys Ile Ser Ser Phe Trp Val Ser Ser Asn Asn Trp Ile Gln Val Ser  Gln Ser

            1520                        1550                        1580
GTT ACT  GGT GGT GTC AGT AAA AAG GGT TTG GAT GTC TCC ACA TTG TTG GCT GCT AAC CTT GGT AGT GTT GAT GAT GGA TTC TTC ACT CCT  GGC TCT
Val Thr  Gly Gly Val Ser Lys Lys Gly Leu Asp Val Ser Thr Leu Leu Ala Ala Asn Leu Gly Ser Val Asp Asp Gly Phe Phe Thr Pro  Gly Ser

        1605                1625                1645                        1680
GAA AAG gtaaattaac atgcataaaa gaataagact ataacattac taattcacat tttctactca g ATC CTT  GCC ACT GCT GTT GCT GTT GAA GAC TCC  TTC GCT
Glu Lys                                                                   Ile Leu  Ala Thr Ala Val Ala Val Glu Asp Ser  Phe Ala

            1710                        1740                        1770
TCC TTG  TAT CCT ATC AAC AAA AAC CTT CCA TCT TAC CTT GGT AAC TCT ATT GGT AGA TAT CCT GAA GAC ACT TAC AAT GGT AAC GGA AAC  TCT CAA
Ser Leu  Tyr Pro Ile Asn Lys Asn Leu Pro Ser Tyr Leu Gly Asn Ser Ile Gly Arg Tyr Pro Glu Asp Thr Tyr Asn Gly Asn Gly Asn  Ser Gln

    1800                        1830                        1860
GGA AAC  TCT TGG TTC TTG GCT GTA ACT GGT TAC GCT GAG CTC TAT TAC CGT GCC ATC AAG GAA TGG ATC GGC AAC GGT GGT GTC ACT GTC  AGC AGC
Gly Asn  Ser Trp Phe Leu Ala Val Thr Gly Tyr Ala Glu Leu Tyr Tyr Arg Ala Ile Lys Glu Trp Ile Gly Asn Gly Gly Val Thr Val  Ser Ser

                1920                        1950                        1980
ATA AGT  TAA CCC TTC TTC AAG AAG TTT GAT TCA TCT GCT ACA TCT GGA AAG AAG TAC ACT GTT GGT ACC TCC GAC TTT AAC AAC CTT GCT  CAA AAT
Ile Ser  Leu Pro Phe Phe Lys Lys Phe Asp Ser Ser Ala Thr Ser Gly Lys Lys Tyr Thr Val Gly Thr Ser Asp Phe Asn Asn Leu Ala  Gln Asn

            2010                        2040                        2070
ATT GCA  CTC GCT GCT GAC CGT TTC TTG TCC ACT GTC CAG CTC CAT GCT CAC AAC AAT GGA TCT CTT GCT GAA GAG TTT GAC CGC ACC ACT  GGT TTA
Ile Ala  Leu Ala Ala Asp Arg Phe Leu Ser Thr Val Gln Leu His Ala His Asn Asn Gly Ser Leu Ala Glu Glu Phe Asp Arg Thr Thr  Gly Leu

        2100                        2130                        2160
TCC ACC GGT GCT AGA GAC TTG ACC TGG TCT CAC GCT TCT TTA ATC ACC GCT TCT TAC GCT AAG GCT GGT GCA CCT GCC GCT TAAGCTGTAA ATTTAAATGC
Ser Thr Gly Ala Arg Asp Leu Thr Trp Ser His Ala Ser Leu Ile Thr Ala Ser Tyr Ala Lys Ala Gly Ala Pro Ala Ala

    2190                        2230                        2270
AAAGCATTAC AGCTTATTTT CTTTTTCAAA TAAAAACATA TTGATATGTT CATAACATTT TCTTGTTTGT TGTACTGTAA TATGGGTAAC CACATAAGCA TAAACAGCAA
```

FIG. 3. The Nucleotide Sequence for the *Rhizopus* Glucoamylase Gene.

The intervening sequences are shown in small letters. The predicted amino acid sequence is shown under the nucleotide sequence. (The putative signal peptide is shown by capital letters.) The amino acid sequences found for peptides derived from the glucoamylase are underlined. The nucleotide sequences hybridized with glucoamylase-specific synthetic 14-mer probes are overlined.

the N-terminus of the signal peptide and the signal peptide is cleaved between Ala-Ala at the 25～26th position. These features are consistent with those of other signal peptides.[18]

The C-terminal sequence, Ala-Ala, is located just before the stop codon TAA. The 3′-noncoding region is A-T rich and contains the nucleotide sequence AATAAA that has been

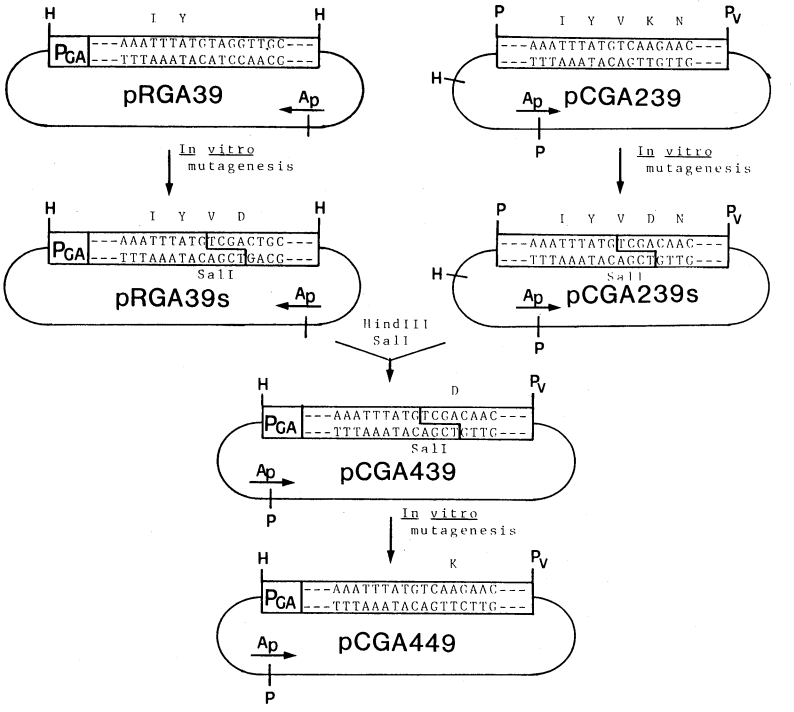

FIG. 4. The Construction of the Glucoamylase Gene to be Expressed in Yeast.

Boxes represented the fragments containing the glucoamylase genes. I, Y, K, N, and D are one letter descriptions of amino acids. *Sal*I sites were created in pRGA39 and pCGA239 by *in vitro* mutagenesis (see MATERIALS AND METHODS) resulting in pRGA39s and pCGA239s, respectively. pCGA439 was constructed by recombination of pRGA39s and pCGA239s at *Hin*dIII and *sal*I sites. The altered amino acid (D) was restored to the original amino acid (K) by *in vitro* mutagenesis.

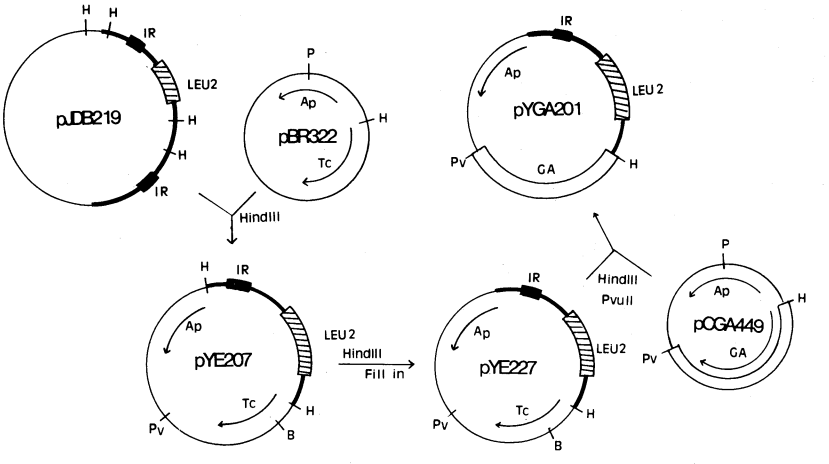

FIG. 5. The Construction of a Shuttle Vector Containing the Glucoamylase Gene.

The restriction sites used are shown. The thin line represents the pBR322 moiety and the thick line represents the yeast DNA moiety. IR indicates the inverted repeat sequence of 2 μM DNA. The restriction sites P, H, B, and Pv indicates *Pst*I, *Hin*dIII, *Bam*HI and *Pvu*II, respectively.

proposed to be necessary for polyadenylation.[19] The glucoamylase gene encodes 604 amino acids and the mature glucoamylase consists of 579 amino acids. The molecular weight of the mature glucoamylase is calculated to be 62,197 daltons which correlates well with the molecular size reported by Takahashi *et al.*[4]

The four intervening sequences identified within the *Rhizopus* genomic glucoamylase gene are short (from 48 to 66 bp) and A-T rich (71 to 77%). These features are similar to those of *Aspergillus*[2,3] and *Trichoderma*[20] secreted enzymes.

(e) *Expression of the glucoamylase gene in yeast*

The genomic glucoamylase gene was not expressed in yeast, probably due to the existence of the intervening sequences. The cDNA glucoamylase gene also is not expressed due to the lack of the N-terminal coding sequence. Therefore we constructed a glucoamylase gene which could be expressed in yeast cells (Fig. 4). The resulting recombinant plasmid pCGA449 contains both the 5′- and the 3′-noncoding sequences and the entire glucoamylase coding sequence without the intervening sequences.

To examine the expression of the constructed glucoamylase gene, the shuttle vector pYGA201 was constructed (Fig. 5). pYGA201 consists of the glucoamylase gene from pCGA449, a replication origin derived from yeast 2 μm DNA, and selectable markers (Ap^R for *E. coli* and *LEU2* for yeasts). This plasmid was introduced into *Saccharomyces cerevisiae* XS-30-2B (MATα *trp*1 *leu*2 *his*3 *ura*3). Transformants grew at nearly the normal rate in a medium containing starch as the sole carbon source and secreted the glucoamylase into the culture fluid. These results have already been reported in another paper.[21]

DISCUSSION

The *Rhizopus* genomic glucoamylase gene contains four intervening sequences. The consensus splice junction sequence, GT-AG,[22] is found in these intervening sequences. Other consensus sequences, TACTAAC[23] and PuCTPuAC[2] in the intervening sequences of yeast and fungi, are not found in the *Rhizopus* glucoamylase gene. There were no more than 6 nucleotides in common between four intervening sequences of the *Rhizopus* glucoamylase gene. The mRNA transcribed from the genomic glucoamylase gene in yeast was not accurately spliced due to the difference of splicing systems between *Saccharomyces* and *Rhizopus*, resulting in a failure of the gene to be expressed in *Saccharomyces* cells.

The bias of codon selection for the *Rhizopus* glucoamylase gene was compared with those of filamentous ascomycetes and *Saccharomyces cerevisiae* (Table I). Preference in codon usage for *Rhizopus* glucoamylase gene is not consistent with those of *Aspergillus awamori*,[2] *Trichoderma reesei*[20] or *Neurospora crassa*.[24] These filamentous ascomycetes preferred G and C at the third position of codons. The pattern of codon selection of *Rhizopus* glucoamylase is rather similar to that of *Saccharomyces cerevisiae* codon selection.[25] In particular, acid phosphatase,[26] one of the secretion enzymes of yeast, is in good agreement with *Rhizopus* glucoamylase.

Rhizopus glucoamylase exists in multiple forms, Gluc1, Gluc2, and Gluc3, differing in molecular size.[3] It is thought that Gluc2 and Gluc3 are produced by limited proteolysis of Gluc1 accompanied by the loss of N-terminal peptide fragments. These peptide fragments were isolated as fragment H and L.[4] These results suggest that there is a single gene coding for *Rhizopus* glucoamylase, and this is confirmed by the hybridization experiment. Only one band was detected in the *Hin*dIII fragments of *Rhizopus* DNA when the DNA fragment derived from the genomic glucoamylase gene was used as a probe (data not shown). The N-termini of Gluc1, Gluc2, and Gluc3 were alanine, glutamic acid, and lysine, respectively. Judging from both carbohydrate contents and amino acid compositions of Gluc2 and Gluc3,[4,5] the N-termini can be assigned to glutamic acid at the 134th residue for Gluc2 and lysine at the 85th or the 91st

TABLE I. THE CODON USAGE FOR VARIOUS GLUCOAMYLASE GENES AND YEAST PHO5 GENE

The table lists the codon usages in following secretion enzymes: rh, *Rhizopus* glucoamylase gene; as *Aspergillus oryzae* glucoamylase gene (2); sa, *Saccharomyces diastaticus* glucoamylase gene (27) and *Saccharomyces cerevisiae* PHO5 gene (16).

	Glucoamylase			PHO5		Glucoamylase			PHO5		Glucoamylase			PHO5		Glucoamylase			PHO5
	rh	as	sa	sa		rh	as	sa	sa		rh	as	sa	sa		rh	as	sa	sa
UUU	8	4	17	8	UCU	34	16	29	9	UAU	13	6	12	7	UGU	2	3	4	10
UUC	21	18	14	19	UCC	15	19	16	5	UAC	22	21	14	25	UGC	1	7	5	0
UUA	6	0	6	9	UCA	11	4	19	6	UAA	—	—	—	—	UGA	—	—	—	—
UUG	1	6	19	20	UCG	0	14	9	2	UAG	—	—	—	—	UGG	12	19	11	7
CUU	12	3	5	0	CCU	15	4	15	5	CAU	1	0	6	5	CGU	7	4	4	5
CUC	10	17	6	0	CCC	4	10	5	1	CAC	4	4	7	3	CGC	2	7	1	0
CUA	0	2	6	4	CCA	4	0	21	7	CAA	9	4	13	12	CGA	1	4	2	0
CUG	0	20	14	2	CCG	0	8	3	0	CAG	2	13	9	2	CGG	0	3	4	0
AUU	16	12	22	13	ACU	36	20	48	16	AAU	20	6	14	9	AGU	8	12	7	11
AUC	12	11	13	7	ACC	15	39	27	20	AAC	25	19	27	22	AGC	10	23	10	2
AUA	1	1	8	1	ACA	9	5	32	2	AAA	6	0	11	8	AGA	3	1	7	7
AUG	4	3	8	7	ACG	1	10	19	0	AAG	25	13	11	16	AGG	0	1	3	1
GUU	13	6	19	9	GCU	42	25	16	13	GAU	15	21	16	20	GGU	28	14	13	22
GUC	18	15	19	16	GCC	7	19	14	14	GAC	12	23	19	21	GGC	6	22	19	4
GUA	4	2	11	2	GCA	8	10	7	2	GAA	11	9	20	23	GGA	12	7	11	1
GUG	0	19	7	1	GCG	0	11	7	0	GAG	6	17	11	2	GGG	0	4	6	2

residue for Gluc3. All three forms of glucoamylase hydrolyze gelatinized starch at similar rates, but only the largest one (Gluc1) is able to adsorb to raw starch and degrade this substrate. It seems likely that the N-terminal region of the *Rhizopus* glucoamylase must be involved in raw starch adsorption and degradation.

The constructed *Rhizopus* glucoamylase gene containing the 5′-flanking sequence and the entire coding sequence without the intervening sequences was expressed in the yeast cells and glucoamylase was secreted into the culture medium as shown previously. These results indicate that the 5′-flanking region of *Rhizopus* glucoamylase gene promoted gene expression and the signal peptide functioned properly in yeast cells. However the expression level of glucoamylase in these yeast cells is not sufficiently high enough to allow the practical production of glucoamylase and ethanol from raw starch. High level expression of glucoamylase in yeast and efficient ethanol production from starch materials are now in progress and will be reported elsewhere.

Acknowledgments. We thank Dr. Hiroshi Matsubara and Dr. Sadao Wakabayashi, Osaka University, for sequencing of the glucoamylase. We also thank Dr. Takehiro Oshima for helpful discussions.

REFERENCES

1) S. Ueda, *Trends Biochem. Sci.*, **6**, 89 (1981).
2) J. H. Nunberg, J. H. Meade, G. Cole, F. C. Lawyer, P. MacCabe, V. Schwerckart, R. Tal, V. P. Wittman, J. E. Flatgaard and M. A. Innis, *Molec. Cell. Biol.*, **4**, 2306 (1984).
3) E. Boel, I. Hjort, B. Svensson, F. Norris, K. E. Norris and N. P. Fiil, *EMBO J.*, **3**, 1097 (1984).
4) T. Takahashi, Y. Tsuchida and M. Irie, *J. Biochem.*, **84**, 1183 (1978).
5) T. Takahashi, Y. Tsuchida and M. Irie, *J. Biochem.*, **92**, 1623 (1982).
6) H. Ito, Y. Fukuda, K. Murata and A. Kimura, *J. Bacteriol.*, **153**, 163 (1983).
7) Y. Tsujisaka, N. Hamada and S. Takenishi,

Proceedings of the Symposium on Amylase (in Japanese), 10th meeting of amylase kenkyu kai, Osaka, Japan, October, 1975, p. 61.
8) Y. Tanaka, Y. Fukumori and T. Yamanaka, *Biochim. Biophys. Acta*, **707**, 14 (1982).
9) D. R. Cryer, R. Ecdeshall and J. Marmur, "Methods in Cell Biology," Vol. 12, ed. by D. M. Prescott, Academic Press Inc., New York, 1975, p. 39.
10) J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald and W. J. Rutter, *Biochmistry*, **18**, 5294 (1979).
11) T. Maniatis, E. F. Fritsch and J. Sambrook, "Molecular Cloning," Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982, pp. 188～209.
12) H. Okayama and P. Berg, *Molec. Cell. Biol.*, **2**, 161 (1982).
13) H. Donis-Keller, A. M. Maxam and W. Gilbert, *Nucl. Acids. Res.*, **4**, 2527 (1977).
14) M. Crunstein and J. Walls, "Methods in Enzymology," Vol. 68, Academic Press Inc., New York, 1979, p. 379.
15) J. Messing and J. Vieira, *Gene*, **19**, 269 (1982).
16) Y. Morinaga, T. Franceshini, S. Inouye and M. Inouye, *Bio/Technology*, **2**, 636 (1984).
17) K. Miyoshi, R. Arentzen, T. Huang and K. Itakura, *Nucl. Acids Res.*, **8**, 5509 (1980).
18) M. E. E. Watson, *Nucl. Acids. Res.*, **12**, 5145 (1984).
19) M. Fitzgerald and T. Shenk, *Cell*, **24**, 251 (1981).
20) S. Shoemaker, V. Schweickart, M. Ladner, D. Gelfand, S. Kwok, K. Myambo and M. Innis, *Bio/Technology*, **1**, 691 (1983).
21) T. Ashikari, N. Nakamura, Y. Tanaka, N. Kiuchi, Y. Shibano, T. Tanaka, T. Amachi and H. Yoshizumi, *Agric. Biol. Chem.*, **49**, 2521 (1985).
22) M. R. Lerner, J. A. Boyle, S. M. Mount, S. L. Wolin and J. A. Stertz, *Nature*, **283**, 220 (1980).
23) C. J. Langford and D. Gallwitz, *Cell*, **33**, 519 (1983).
24) M. G. Schechman and C. Yanofsky, *J. Mol. Appl. Genet.*, **2**, 83 (1983).
25) J. L. Bennetzen and B. D. Hall, *J. Biol. Chem.*, **257**, 3036 (1982).
26) K. Arima, T. Oshima, I. Kubota, N. Nakamura, T. Mizunaga and A. Toh-e, *Nucl. Acids Res.*, **11**, 1657 (1983).
27) I. Yamashita, K. Suzuki and S. Fukui, *J. Bacteriol.*, **161**, 567 (1985).