MOLECULAR AND CELLULAR BIOLOGY, Nov. 1984, p. 2306–2315
0270-7306/84/112306-10$02.00/0

Vol. 4, No. 11

Molecular Cloning and Characterization of the Glucoamylase Gene of *Aspergillus awamori*

JACK H. NUNBERG,[1] JAMES H. MEADE,[1]* GEORGETTE COLE,[1] FRANCES C. LAWYER,[1] PETER McCABE,[1] VICKI SCHWEICKART,[1] RONY TAL,[1] VAUGHAN P. WITTMAN,[1] JEFFREY E. FLATGAARD,[2] AND MICHAEL A. INNIS[1]

Division of Molecular and Biological Research,[1] and Division of Process and Product Development,[2] Cetus Corp., Emeryville, California 94608

Received 9 April 1984/Accepted 31 July 1984

The filamentous ascomycete *Aspergillus awamori* secretes large amounts of glucoamylase upon growth in medium containing starch, glucose, or a variety of hexose sugars and sugar polymers. We examined the mechanism of this carbon source-dependent regulation of glucoamylase accumulation and found a several hundredfold increase in glucoamylase mRNA in cells grown on an inducing substrate, starch, relative to cells grown on a noninducing substrate, xylose. We postulate that induction of glucoamylase synthesis is regulated transcriptionally. Comparing total mRNA from cells grown on starch and xylose, we were able to identify an inducible 2.3-kilobase mRNA-encoding glucoamylase. The glucoamylase mRNA was purified and used to identify a molecularly cloned 3.4-kilobase *Eco*RI fragment containing the *A. awamori* glucoamylase gene. Comparison of the nucleotide sequence of the 3.4-kilobase *Eco*RI fragment with that of the glucoamylase I mRNA (as determined from molecularly cloned cDNA) revealed the existence of four intervening sequences within the glucoamylase gene. The 5′ end of the glucoamylase mRNA was mapped to several locations within a region −52 to −73 nucleotides from the translational start. Sequence and structural features of the glucoamylase gene of the filamentous ascomycete *A. awamori* were examined and compared with those reported in genes of other eucaryotes.

Glucoamylases comprise a class of extracellular enzymes, secreted by a variety of organisms, which progressively hydrolyze starch to yield glucose (17). The glucoamylase [α-(1-4),(1-6)-D-glucan glucohydrolase (EC 3.2.1.3)] of the filamentous ascomycete *Aspergillus awamori* is an inducible glycoprotein that possesses both α(1-4) and α(1-6) (debranching) activities (63, 68). Glucoamylases from *A. awamori* and other *Aspergillus* species have been widely used in commercial processes requiring the saccharification of starch (42). The glucose produced is used as a substrate for the enzymatic production of fructose syrup and as a feed source for fermentations of various organisms. Although *Aspergillus* glucoamylase production has been extensively studied in the fermentation literature, little is known at the molecular level concerning the structure and regulation of the glucoamylase gene. In this paper, we report the molecular cloning of the glucoamylase gene of *A. awamori*. The organization of the *A. awamori* glucoamylase gene is determined by comparison of the genomic DNA sequence with that of the glucoamylase mRNA as obtained from molecularly cloned glucoamylase cDNA. We describe several structural aspects of the *A. awamori* glucoamylase gene and compare structural and regulatory features with those described in gene systems of the single-celled ascomycete *Saccharomyces cerevisiae* and of higher eucaryotes.

MATERIALS AND METHODS

Strains and culture conditions. *A. awamori* (NRRL 3112) was obtained from the Northern Regional Research Laboratory, Peoria, Ill. For growth, spores were inoculated in medium containing (grams per liter): yeast extract (Difco Laboratories, Detroit, Mich.), 10; $(NH_3)_2SO_4$, 6.5; $MgSO_4 \cdot 7H_2O$, 0.2; citric acid · H_2O, 2.0; K_2HPO_4, 10; $NaNH_4HPO_4 \cdot 4H_2O$, 3.5; and various carbon sources (as described in text), 50. Cultures were agitated at 200 to 250 rpm at 30°C for 3 to 5 days.

Glucoamylase assay. Mycelia-free culture supernatants were incubated in the presence of 2% soluble starch (Difco) in 10 mM potassium succinate buffer (pH 5) for 1 to 4 h at 60°C. Glucose was determined by using a glucose analyzer (model 23, YSI Co.). One unit of activity corresponds to the formation of 10 mg of glucose per h.

Glucoamylase purification and preparation of antisera. *A. awamori* glucoamylase was purified from culture supernatants of cells grown in medium containing starch as the carbon source. *Aspergillus niger* glucoamylase was purified from a preparation (AMG 150 L) obtained from Novo Industries, Bagsvaerd, Denmark. Glucoamylase represents the major protein found in these culture supernatants. Consistent with previous results (33, 59), two molecular weight forms of glucoamylase were observed. Although molecular weight estimates of these glycoproteins vary with the analytic procedure used (59), typical values are 75,000 for glucoamylase I (GA I) and 54,000 for glucoamylase II (GA II) (32). The enzymes were purified by ion-exchange chromatography (DEAE-Sephacel; Pharmacia Fine Chemicals, Inc., Piscataway, N.J.), gel filtration chromatography (Sephacryl S-200 [Pharmacia] or Bio-Gel P-300 [Bio-Rad Laboratories, Richmond, Calif.]). On occasion, cyclohexamylose affinity chromatography (22) was used. *A. niger* GA I and GA II possess similar enzymatic activities and have been previously shown to be immunologically indistinguishable (33, 40) and to have identical N-terminal amino acid sequences (59). We have found the two forms of the *A. awamori* glucoamylase to be likewise immunologically indistinguishable. We have determined the N-terminal amino acid sequence of *A. awamori* GA I to residue 49: 1-ATLDSWLSNE ATVAR

* Corresponding author.

TAILN NIGADGAWVS GADXGIVVAX PXTDXPXYF-49 (K. Watt, unpublished data) (refer to the IUPAC-IUB Commission on Biochemical Nomenclature for an explanation of the one-letter code for amino acids). The C-terminal amino acid is arginine. The N-terminal sequence agrees with that of *A. niger* GA I published by Svenssen et al. (58). The GA I proteins of *A. niger* and *A. awamori* are also indistinguishable by immunological criteria (R. Cox, unpublished data).

Purified glucoamylase protein was used to generate antibodies in rabbits; 0.2 to 1.0 mg of protein was injected in conjunction with complete Freund adjuvant. The protein was not very immunogenic, and several injections were required to generate useful antisera.

Nucleic acid isolation. Total cellular RNA was isolated from *A. awamori* mycelia by using the guanidium thiocyanate procedure of Chirgwin et al. (14). Mycelia were wrung dry in cheesecloth and ground to a powder in liquid nitrogen with a mortar and pestle. The cell power was homogenized in guanidium thiocyanate solution containing 10 mM adenosine:vanadyl sulfate complex (7). After centrifugation to pellet cell debris, CsCl was added, and RNA was pelleted through a CsCl pad as described by Glisin et al. (20). Polyadenylated [poly(A)$^+$] RNA was isolated by two cycles of oligodeoxythymidylate cellulose chromatography (type 3; Collaborative Research, Inc., Waltham, Mass.) (2).

Glucoamylase mRNA was isolated from methyl mercury (MeHgOH)-agarose gels (3) after staining with ethidium bromide by the method of Sehgal (51). Low-gelling-temperature agarose (Miles Laboratories, Inc., Elkhart, Ind.) was used to form the gel, and gel slices containing glucoamylase mRNA were heated to 70°C to melt the agarose before freezing and centrifugation.

A. awamori genomic DNA was prepared from frozen mycelia as described by Murray and Thompson (41).

In vitro translation and immunoprecipitation. *A. awamori* RNA was translated in vitro by the rabbit reticulocyte lysate system (New England Nuclear Corp., Boston, Mass.). [^{35}S]methionine was used to label protein products. Immunoprecipitation, with either rabbit anti-GA I immunoglobulin G (IgG) or normal rabbit IgG, was performed essentially as described by Ivarie and Jones (24), using Formalin-fixed *Staphylococcus aureus* Cowan cells (IgGsorb, The Enzyme Centre) to precipitate immune complexes. In some cases, lysates were first treated with sodium dodecyl sulfate, and the sodium dodecyl sulfate was then removed by the formation of micelles with Triton X-100 before reaction with antibodies (69). This procedure appeared to reduce backgrounds in these experiments. Antibody specificity was demonstrated by inhibition of glucoamylase activity (Cox, unpublished data) as well as by immune electrophoresis and Ouchterlony analysis.

Molecular cloning; cDNA library preparation. cDNA was synthesized by modifications of previously published procedures (12, 43). Poly(A)$^+$ mRNA was pretreated with MeHgOH and added to the reaction mixture containing oligodeoxythymidylate as a primer and 2-mM adenosine: vanadyl sulfate as an RNase inhibitor. Before use of the cDNA as hybridization probe, remaining mRNA was degraded by treatment with sodium hydroxide. A library was prepared from cDNA essentially as described by Land et al. (28). Single-stranded cDNA was tailed with deoxycytidine and used as template for second-strand cDNA synthesis with oligodeoxyguanosine primer. Double-stranded cDNA was again tailed with deoxycytidine and annealed to plasmid pBR322 that had been linearized and tailed with deoxyguanosine at the *Pst*I site. This material was used to transform *Escherichia coli* K-12 MM294 to tetracycline resistance. The cDNA library obtained was screened by hybridization as described by Grunstein and Hogness (21).

Molecular cloning and genomic library preparation. *A. awamori* DNA was digested to completion by *Eco*RI, *Hin*dIII, *Bgl*II, and complete genomic libraries were constructed with either lambda Charon 4A (66), lambda L47.1 (34), or lambda Charon 30 (66), respectively, using modifications of previously published procedures (9, 57). The *Eco*RI library was screened as described by Benton and Davis (6), using a glucoamylase cDNA probe derived from gel-purified glucoamylase mRNA (see text). Other libraries were screened by using glucoamylase sequences derived from the *Eco*RI library.

Nucleic acid analysis. Genomic DNA was analyzed for specific glucoamylase sequences by the method of Southern (55, 65). RNA was analyzed for specific glucoamylase sequences by electrophoresis in MeHgOH-agarose gels (3) and transfer to nitrocellulose (61) as described by Alwine et al. (1). Nick-translation of DNA was performed as described by Rigby et al. (47).

DNA nucleotide sequence was determined primarily by the dideoxynucleotide method of Sanger et al. (48, 49) with the phages M13mp8 and M13mp9 (37). In some instances, the chemical method of Maxam and Gilbert (35) was used.

Restriction endonucleases and DNA polymerase I were obtained from New England Biolabs, Beverly, Mass., or Bethesda Research Laboratories, Gaithersburg, Md. Terminal deoxynucleotidyl transferase was obtained from Ratliff Biochemicals, Los Alamos, N. Mex. Reverse transcriptase (from avian myeloblastosis virus) was obtained from Life Sciences, Inc., St. Petersburg, Fla. All enzymes were used according to the instructions of the manufacturers. DNA ligase (T4) was provided by D. Gelfand; S1 nuclease was prepared by the method of Vogt (64).

Mapping of 5′ termini of glucoamylase mRNA. An oligonucleotide primer, 5′GCGAGTAGAGATCGG3′, was synthesized with a BioSearch DNA synthesizer (San Rafael, Calif.). Primer extension was performed by methods described by Proudfoot et al. (45) to determine the 5′ terminus of glucoamylase mRNA. 5′ end-labeled primer was annealed to poly(A)$^+$mRNA (92°C for 5 min; 42°C for 60 min) and extended, using avian myeloblastosis virus reverse transcriptase.

The primer was also used to prepare an internally labeled single-stranded DNA fragment complementary to the 5′ region of the glucoamylase gene for use in S1 mapping experiments. The primer was annealed to, and copied from, an *Eco*RI-*Sal*I fragment of the glucoamylase gene in phage M13mp8, as described by Hu and Messing (23), in the presence of [α-^{32}P]dATP. After digestion with *Eco*RI and alkaline denaturation the generated 213-base-pair (bp) single-stranded fragment containing the 5′ region of the glucoamylase gene was purified by electrophoresis under alkaline conditions (36). This fragment was used in S1 mapping experiments as described by Berk and Sharp (8) to determine the 5′ termini of glucoamylase mRNA.

RESULTS

Regulation of glucoamylase protein and mRNA expression. The accumulation of glucoamylase protein was examined in culture supernatants of *A. awamori* grown in medium containing different carbon sources (Fig. 1). These results confirm previous reports (46; L. L. Barton, Ph.D. thesis, University of Nebraska, 1969) that glucoamylase is secreted

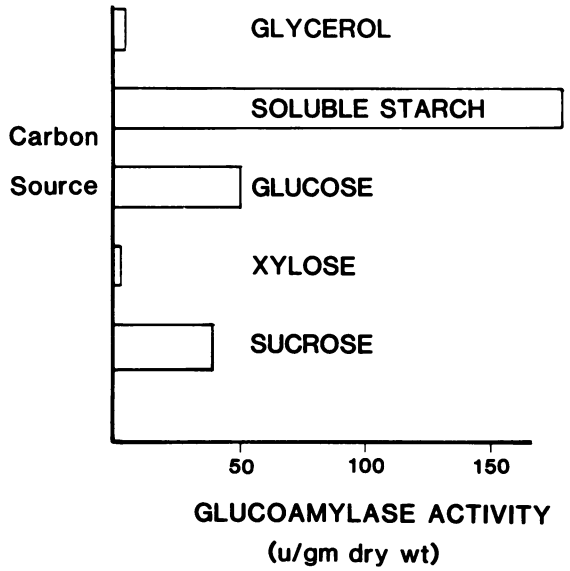

FIG. 1. Glucoamylase accumulation in culture supernatants of *A. awamori* grown to stationary phase on indicated carbon sources.

into the medium upon growth on glucose, starch, or a variety of other hexose sugars or hexose sugar polymers and that little or no glucoamylase activity is found when glycerol or xylose is used as the carbon source. The activity measurements (Fig. 1) correlate with estimates of glucoamylase protein as determined from sodium dodecyl sulfate-polyacrylamide gels (data not shown).

To determine the molecular basis for this carbon source-dependent accumulation of glucoamylase protein, glucoamylase mRNA levels were examined. Total cellular mRNA was isolated and used to direct the synthesis of *A. awamori* protein in a rabbit reticulocyte lysate system; glucoamylase-specific mRNA was determined by immunoprecipitation of protein products with rabbit anti-glucoamylase antibody. The results demonstrate the presence of translatable glucoamylase mRNA in RNA from starch-grown cells; in contrast, no functional glucoamylase mRNA was detected in xylose-grown cells (data not shown). This correlates with the 200-fold difference in glucoamylase protein observed in culture supernatants of these cells and suggests that the accumulation of glucoamylase protein in starch-grown cultures is controlled at the level of glucoamylase mRNA.

MeHgOH-agarose gel electrophoresis of mRNA from starch-grown cells revealed a major 2.3-kilobase (kb) mRNA, which was absent in mRNA from xylose grown cells (Fig. 2). It appeared likely that this prominent "induced" mRNA represented the mRNA of the highly expressed "induced" glucoamylase. Indeed, the 2.3-kb mRNA was found, upon elution from MeHgOH-agarose gels, to encode the synthesis of imunoprecipitable glucoamylase protein in a rabbit reticulocyte lysate system (data not shown). We will present additional evidence that further establishes the identity of the 2.3-kb mRNA as glucoamylase mRNA.

Using cDNA to the glucoamylase mRNA as a hybridization probe, we were able to isolate the glucoamylase gene from a phage library containing *A. awamori* genomic DNA. An *Eco*RI digest of genomic DNA was molecularly cloned in lambda Charon 4A, and the library produced was screened with glucoamylase cDNA. Hybridizing plaques were purified; all contained a common 3.4-kb *Eco*RI fragment that hybridized to the glucoamylase cDNA probe. This fragment was subcloned into the *Eco*RI site of pACYC184 and was designated pGAR1. Other phage libraries (*Hin*dIII, *Bgl*II) were constructed and screened with pGAR1 to isolate flanking genomic sequences. Approximately 20 kb of *A. awamori* genomic DNA surrounding the glucoamylase gene was isolated; a composite restriction map of this region is shown in Fig. 3.

The nucleotide sequence of the 3.4-kb *Eco*RI fragment containing the *A. awamori* glucoamylase gene, as isolated in pGAR1, is presented in Fig. 4. Comparison of the deduced amino acid sequence with the amino acid sequence of the GA I protein (*A. niger*, [58]; *A. awamori*, K. Watt, unpublished data) established that we had, in fact, isolated the glucoamylase gene.

Restriction analysis of *A. awamori* genomic DNA containing the glucoamylase gene was performed by the method of Southern (55). In this analysis (Fig. 5), sequences of the pGAR1 probe were shown to hybridize to genomic restriction fragments in a manner consistent with that predicted from restriction maps obtained from molecularly cloned genomic fragments (Fig. 3). No evidence was observed for additional, cross-hybridizing, glucoamylase genes. These results indicate that the glucoamylase gene exists as a single copy of the *A. awamori* genome.

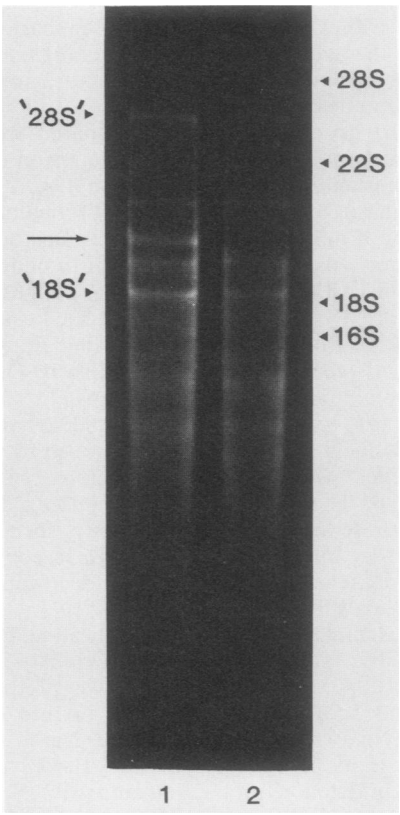

FIG. 2. Identification of glucoamylase mRNA. Poly(A)$^+$ mRNA from cells grown in medium containing starch (lane 1) or xylose (lane 2) was analyzed by MeHgOH-agarose gel electrophoresis. Human and *E. coli* rRNAs provide molecular weight markers. The *A. awamori* rRNAs are indicated by '28S' and '18S'. The major "induced" mRNA (arrow) was isolated and used to direct in vitro synthesis of glucoamylase protein.

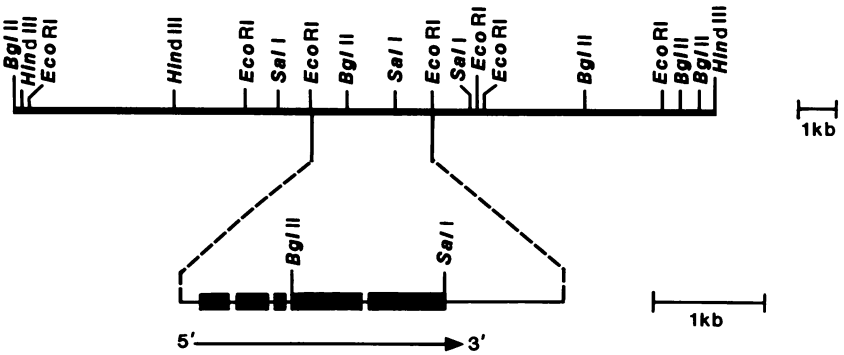

FIG. 3. Restriction endonuclease map of *A. awamori* genome surrounding the glucoamylase gene. The entire structural gene is contained within the 3.4-kb *Eco*RI fragment isolated from the Charon 4A library. The location and orientation of the glucoamylase gene are shown below; protein-encoding regions are indicated as solid boxes.

The molecularly cloned glucoamylase sequence (pGAR1) was used to confirm and extend our earlier studies in which we had shown increased levels of translatable glucoamylase mRNA in poly(A)$^+$RNA of starch-grown cells relative to cells grown on xylose. Northern analysis of these RNAs, using the method of Alwine et al. (1), demonstrates a similar increase in the level of the 2.3-kb glucoamylase mRNA (Fig. 6). Thus, the accumulation of glucoamylase protein in starch-grown cultures can be accounted for by a comparable increase in glycoamylase mRNA.

Molecular cloning of *A. awamori* glucoamylase mRNA sequences. Comparison of the coding potential of the molecularly cloned glucoamylase gene with the previously published amino acid sequence of the *A. niger* GA I protein (58) indicated the presence of several intervening sequences with the protein coding region. To confirm the identification of intervening sequences within the glucoamylase gene, double-stranded cDNA was prepared from mRNA of starch-grown *A. awamori* and a cDNA library was prepared in pBR322 as described above. Several glucoamylase cDNA-containing plasmids were identified with the pGAR1 probe; p24A2, the largest, contains 1.8 kb of sequence derived from the 3′ end of the glucoamylase mRNA. The nucleotide sequence of the glucoamylase cDNA in p24A2 was determined and found to span the genomic sequence (Fig. 4) from nucleotide 501 through the polyadenylation site at positions 2494 to 2496. (The precise polyadenylation site cannot be determined due to the presence of two adenosine residues at nucleotides 2495 and 2496.) Comparison between these sequences and with the amino acid sequence has identified four intervening sequences within the *A. awamori* glucoamylase gene. (The junctions of the first intervening sequence were deduced from our incomplete amino acid sequence data at residues 43 to 49 of *A. awamori* GA I [43-TDXPXYF-49] [see above; unpublished data].) The intervening sequences are short (ranging from 55 to 75 bp) and are all located within protein-encoding sequences.

The 5′ end of the glucoamylase mRNA was determined by primer extension and S1 mapping techniques. A synthetic oligonucleotide 5′GCGAGTAGAGATCGG3′, complementary to sequences within the signal peptide-encoding region near the 5′ end of the glucoamylase mRNA, was used to prime reverse transcriptase synthesis from the mRNA template. Several primer extension products are synthesized from total poly(A)$^+$mRNA from starch-grown cells (Fig. 7A). To examine the possible effects of RNA secondary structure on this pattern, primer extension was performed at both 42 and 50°C; the pattern of primer extension was unchanged, supporting the conclusion that the observed extension products represent distinct 5′ mRNA termini.

To corroborate these results, S1 mapping experiments were performed with an internally labeled single-stranded DNA fragment generated by the use of the synthetic oligonucleotide to prime reverse transcription of 5′ genomic glucoamylase sequences in phage M13mp8 (see above). After hybridization to total poly(A)$^+$mRNA from starch-grown cells and digestion with S1 nuclease (8), protected DNA was examined by electrophoresis under denatured conditions (49). The results (Fig. 7B) demonstrate a pattern coincident with that observed in primer extension experiments. (One site predicted from primer extension results is absent in the S1 experiment.) Taken together, these results indicate that several distinct 5′ termini exist within the population of glucoamylase mRNA. The termini are localized within a region −73 to −52 from the site of translation initiation (Fig. 4). The coincidence of bands generated by primer extension and S1 mapping demonstrates the absence of additional intervening sequences within this region of the glucoamylase gene.

DISCUSSION

We report the molecular cloning and structural characterization of the glucoamylase gene of *A. awamori*. The results presented here examine, on a molecular level, the regulation and organization of glucoamylase gene expression in the filamentous ascomycete *A. awamori*.

Regulation of glucoamylase gene expression. Growth of *A. awamori* in medium containing glucose, starch, or a variety of other hexose sugars or hexose sugar polymers resulted in the accumulation of high levels of glucoamylase in culture supernatants. Little or no glucoamylase was secreted into culture supernatants of cells grown in medium containing glycerol or xylose. We have shown that the increase in glucoamylase accumulation in cultures grown on starch resulted from a comparable increase in the level of glucoamylase mRNA found in these cells.

The 2.3-kb glucoamylase mRNA from starch-grown cells constituted 1 to 5% of cellular poly(A)$^+$ RNA, as judged from its appearance on MeHgOH-agarose gels. We estimate that this represents a several hundredfold increase in glucoamylase mRNA over the level found in xylose-grown cells. We speculate that the increased levels of glucoamylase mRNA accumulation in starch-grown cells resulted from increased levels of transcription of the glucoamylase gene.

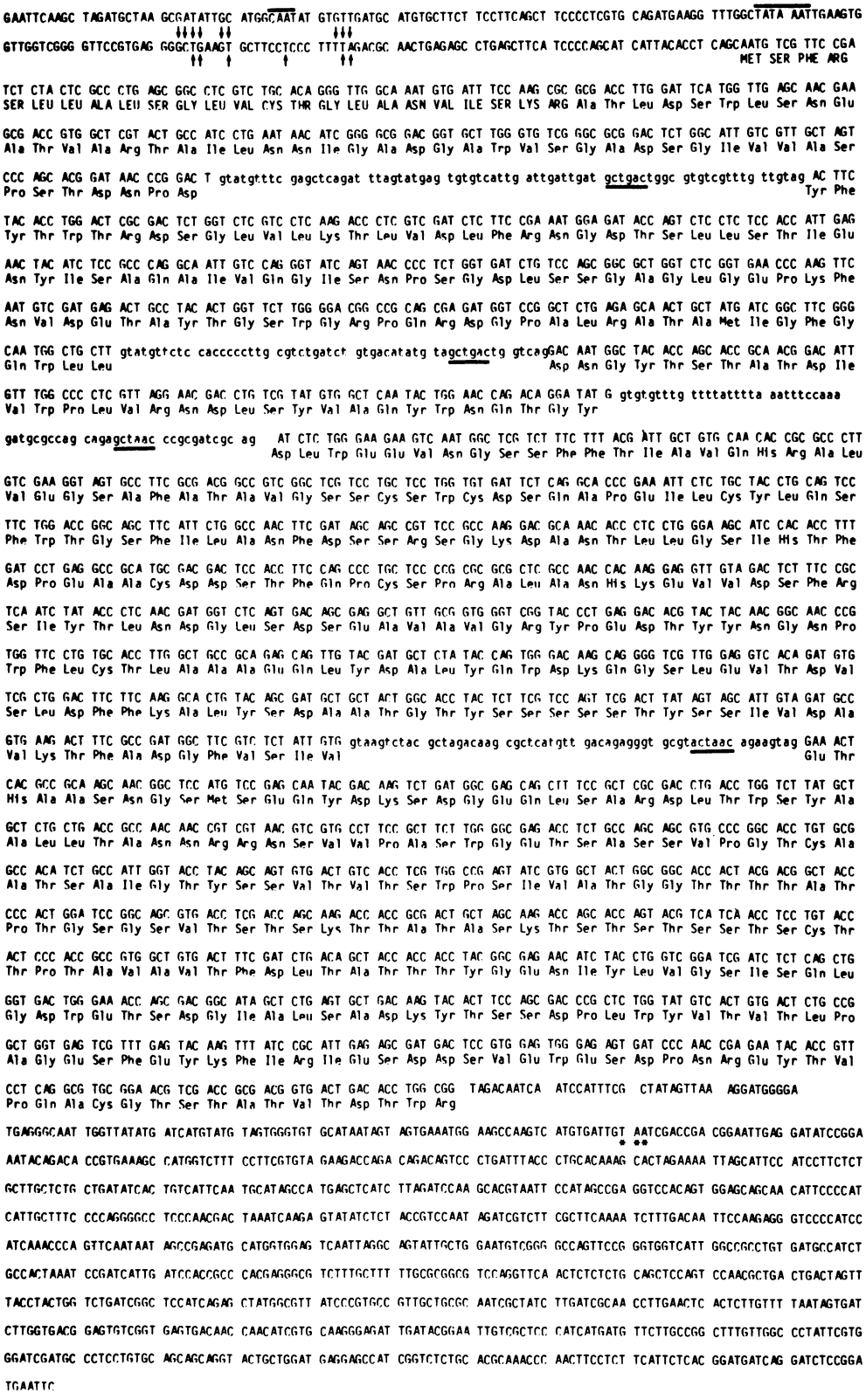

FIG. 4. Nucleotide sequence of 3.4-kb *Eco*RI fragment containing the *A. awamori* glucoamylase gene. Intervening sequences are in lowercase letters. The deduced protein sequence is also presented; amino acid residues −24 to −1, in uppercase letters, represent the presumed signal peptide sequence. Regions of transcription initiation, as determined from primer extension and S1 mapping products (↑ and ↓, respectively), and the polyadenylation site (***) of glucoamylase mRNA are indicated. (The polyadenylation site cannot be precisely determined due to the presence of A's within the genomic sequence at the site.) The TATA and CCAAT sequences, upstream of transcription initiation, are overlined; the consensus sequence PuCTPuAC found near the 3′ terminus of intervening sequences of ascomycetes is underlined.

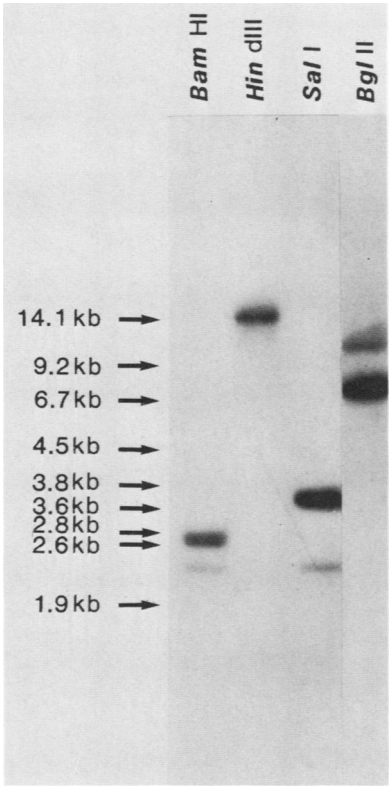

FIG. 5. Analysis of *A. awamori* DNA by using pGAR1 glucoamylase probe. Genomic DNA was digested to completion with the indicated restriction endonucleases, and the fragments were separated by agarose gel electrophoresis and transferred to nitrocellulose. Hybridization, with nick-translated pGAR1, was as described by Wahl et al. (65).

At present, we have no direct data to support this hypothesis.

Multiple forms of glucoamylase protein. Two forms of glucoamylase protein of differing molecular weights were found in culture supernatants of *A. awamori* cultures growing on starch. These forms were enzymatically and immunologically related and had identical N-terminal amino acid sequences. The peptide molecular weight of the glucoamylase protein deduced from the DNA sequence (69,000) agrees with that observed for the major in vitro translation product of glucoamylase mRNA (data not shown) and is consistent with that estimated for an unglycosylated form of the larger of the two forms of glucoamylase, GA I. The amino acid composition (unpublished data) and C terminus of the deduced GA I protein are also consistent with this identification.

The two forms of glucoamylase were not derived from two related glucoamylase genes, as we have demonstrated the glucoamylase gene to exist as a single copy gene in the *A. awamori* genome. Southern analysis of genomic DNA, as well as the analysis of molecularly cloned genomic sequences, showed no evidence for multiple genes. Possible mechanisms for the generation of multiple forms of protein from a single gene include post-translational processing and differential mRNA splicing. During preparation of this manuscript, Boel et al. (10) have reported the isolation of an additional form of glucoamylase cDNA from *A. niger*; this cDNA differs from the larger glucoamylase cDNA isolated by both groups in that it appears to result from an additional mRNA splicing event within sequences encoding the C-terminal portion of GA I. The authors propose that this smaller mRNA may encode the smaller GA II protein. In our in vitro translation studies, we saw a minor immunoprecipitable product which may correspond to GA II, but we have not isolated a cDNA copy of the presumptive GA II mRNA.

Mature GA I and GA II contain an identical N-terminal amino acid sequence and initiate at the Ala residue shown in Fig. 4. Consistent with the observed secretion of glucoamylase protein, a presumptive signal peptide sequence can be deduced from the molecularly cloned DNA sequence. This 24-amino-acid region possesses the characteristic features of eucaryotic signal peptides (56), containing a long hydrophobic region preceded by a basic amino acid. The presence of a Lys-Arg dipeptide preceding the cleavage site raises the possibility of a second, trypsin-like processing step in the formation of the mature protein.

Characterization of the glucoamylase gene. Comparison of the nucleotide sequence of the molecularly cloned glucoamylase gene with that of GA I mRNA, as determined from molecularly cloned cDNA, and with the GA I amino acid sequence, has revealed the presence of four intervening sequences within the *A. awamori* glucoamylase gene. The intervening sequences were short (ranging from 55 to 75 bp) and were all located within protein-encoding sequences. Short intervening sequences have also been found in genes of other filamentous ascomycetes, namely, the cellobiohydrolase gene of *Trichoderma reesei* (54) and the glutamate dehydrogenase gene (26) and histone H3 and H4 genes (67) of *Neurospora crassa*. In contrast, the *trp-1* gene of *N. crassa* has been shown to lack intervening sequences (50). Although virtually ubiquitous within genes of higher eucaryotes (11), intervening sequences appear to be limited to a

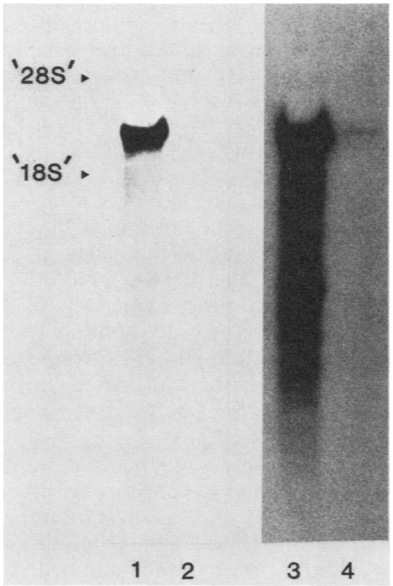

FIG. 6. Analysis of *A. awamori* mRNA by using pGAR1 glucoamylase probe. mRNA from cells grown in medium containing starch (lanes 1 and 3) or xylose (lanes 2 and 4) was separated by MeHgOH-agarose gel electrophoresis, transferred to nitrocellulose, and hybridized to nick-translated pGAR1 DNA as described by Thomas (61). Lanes 3 and 4 represent a long exposure to demonstrate the presence of low levels of glucoamylase mRNA in xylose-grown cells. '28S' and '18S' markers refer to *A. awamori* rRNAs.

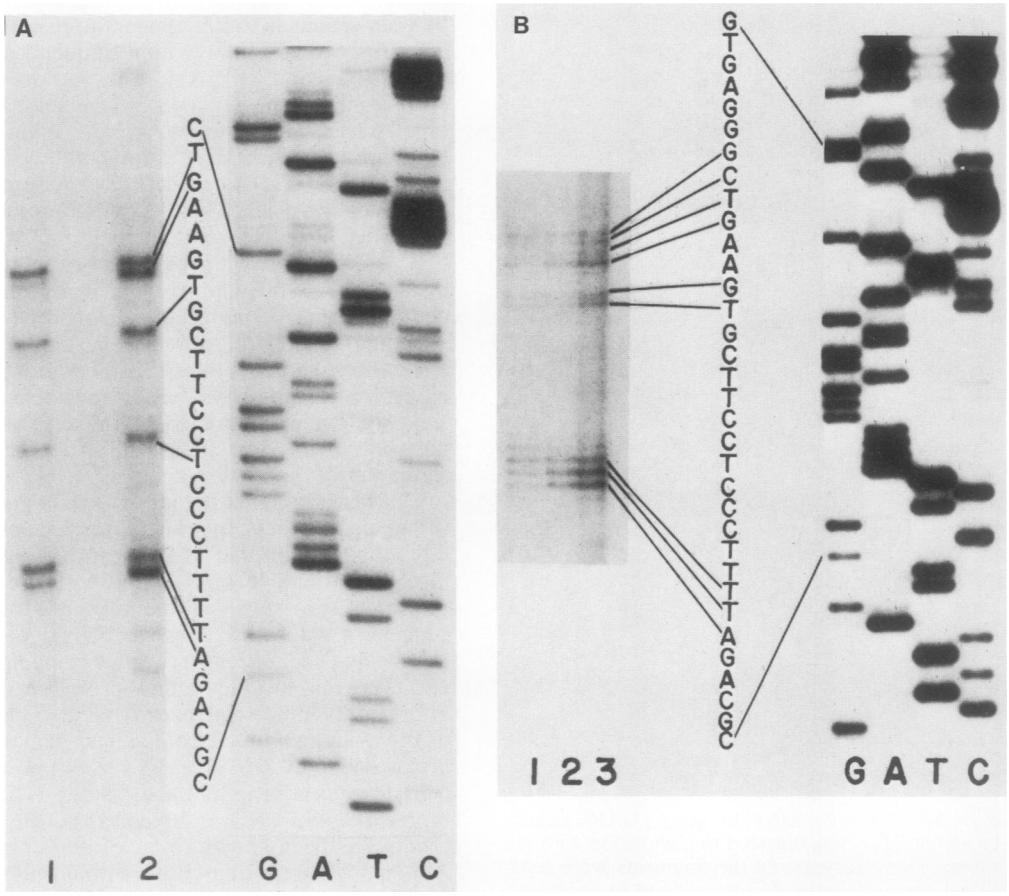

FIG. 7. Mapping of glucoamylase mRNA 5′ termini. (A) Primer extension. The 15-mer, 5′-^{32}P-GCGAGTAGAGATCGG3′, complementary to a region near the 5′ terminus of glucoamylase mRNA, was hybridized to total cellular mRNA and used to prime reverse transcriptase synthesis. The products of primer extension at 42°C (lane 1) and 50°C (lane 2) are displayed on a sequencing gel (49) in parallel with m13-dideoxynucleotide sequencing reactions, utilizing the identical 15-mer primer. The sequence presented represents the glucoamylase mRNA sequence and is complementary to that read from the sequencing reactions shown. (B) S1 nuclease mapping. An internally labeled single-stranded 213-bp DNA fragment corresponding to the 5′ end of the glucoamylase gene was hybridized to total poly(A)$^+$ mRNA from starch-grown cells, and S1 nuclease-resistant products were analyzed by gel electrophoresis under denatured conditions (49). Products resulting from increasing times of S1 nuclease digestion (15, 30, and 45 min) are displayed in lanes 1, 2, and 3, respectively. Sequencing reactions and the sequence presented are as described above.

small number of specific genes of the single-celled ascomycete *S. cerevisiae*. Among nuclear genes in *S. cerevisiae*, only the actin gene (18), the *MATa1* gene (38), and several genes encoding ribosomal proteins (rp51 [60], S10 [31], and cyh2 [25]) have been shown to contain intervening sequences. Analysis of additional genes from ascomycetes is necessary to establish whether differences in the frequency of intervening sequences exist between the single-celled and filamentous ascomycetes.

Sequences at intervening sequence junctions of the glucoamylase gene were found to conform closely to consensus splice junction sequences of *S. cerevisiae* in particular (30) and from eucaryotes in general (39). An additional consensus sequence, PuCTPuAC, was found within and near the 3′ terminus of all intervening sequences of the glucoamylase gene. An identical sequence is similarly located within intervening sequences of other filamentous ascomycetes (*T. reesei* [54] and *N. crassa* [26, 67]) and is related to the consensus sequence TACTAAC found within intervening sequences of *S. cerevisiae* genes (30). This latter sequence has been postulated to be required for RNA splicing in *S. cerevisiae*.

Using the methods of primer extension and S1 mapping, we determined several distinct 5′ termini within the population of glucoamylase mRNA. The multiple transcription initiation sites mapped to a region −73 to −52 nucleotides from the translational start site. DNA sequences 5′ of the region of transcription initiation contained sequences homologous to consensus sequences previously shown to be involved in transcription initiation by RNA polymerase II. A sequence related to the Goldberg-Hogness TATA sequence (19; M. L. Goldberg, Ph.D. thesis, Stanford University, Stanford, Calif., 1979) (TATAAAT) was found 31 bp upstream of the first transcription initiation site, and a sequence related to the canonical CCAAT box (15) (CAAT) was found 62 bp upstream of the TATA box.

A virtually universal feature of the 3′ untranslated region of mRNAs in higher eucaryotes is the presence of the sequence AATAAA preceding the site of polyadenylation (44). Although this sequence has been shown to be required for polyadenylation of simian virus 40 late mRNAs (16), isolated examples exist in which the presence of this sequence is neither necessary (52, 53) nor sufficient (62) for polyadenylation. In contrast to the case in higher eucary-

TABLE 1. Codon utilization in the glucoamylase gene *A. awamori*

Amino acid	Codon	No. of times used	Amino acid	Codon	No. of times used
Phe	UUU	4	Ter	UAA	0
	UUC	18		UAG	1
	UUA	0	His	CAU	0
	UUG	6		CAC	4
Leu	CUU	3			
	CUC	17	Gln	CAA	4
	CUA	2		CAG	13
	CUG	20			
			Asn	AAU	6
Ile	AUU	12		AAC	19
	AUC	11			
	AUA	1	Lys	AAA	0
				AAG	13
Met	AUG	3			
			Asp	GAU	21
Val	GUU	6		GAC	23
	GUC	15			
	GUA	2	Glu	GAA	9
	GUG	19		GAG	17
Ser	UCU	16	Cys	UGU	3
	UCC	19		UGC	7
	UCA	4			
	UCG	14	Ter	UGA	0
Pro	CCU	4	Trp	UGG	19
	CCC	10			
	CCA	0	Arg	CGU	4
	CCG	8		CGC	7
				CGA	4
Thr	ACU	20		CGG	3
	ACC	39			
	ACA	5	Ser	AGU	12
	ACG	10		AGC	23
Ala	GCU	25	Arg	AGA	1
	GCC	19		AGG	1
	GCA	10			
	GCG	11	Gly	GGU	14
				GGC	22
Tyr	UAU	6		GGA	7
	UAC	21		GGG	4

otes, the 3′ untranslated region of mRNAs in *S. cerevisiae* may or may not contain this sequence or related sequences (70). Zaret and Sherman presented a consensus sequence preceding the polyadenylation site of several genes of *S. cerevisiae*; this consensus included the sequences TAGT or TATGT, which the authors postulated to be involved in polyadenylation, transcription termination, or both (70). The 3′ untranslated region of the *A. awamori* glucoamylase mRNA did not contain the sequence AATAAA. (The truncated sequence ATAA appeared 35 bp preceding the polyadenylation site and may be significant.) We did, however, find the sequences TAGT and TATGT preceding the site of polyadenylation of glucoamylase mRNA. The functional significance of specific sequences in the 3′ untranslated region of mRNA remains unclear.

Codon selection of isoacceptor tRNAs has been postulated to serve a regulatory function in the expression of specific genes in *S. cerevisiae* and *E. coli*. Extreme codon bias toward the major isoacceptor tRNA species has been observed in several highly expressed genes of *S. cerevisiae* (5). An opposite bias has been observed in several *E. coli* genes expressed at very low levels (27). Species differences in codon utilization within well-expressed genes may reflect species differences in the distribution of isoacceptor tRNA species. We examined codon utilization in the glucoamylase gene of *A. awamori* (Table 1). Clear preferences in codon utilization were seen for many of the amino acids present in this glucoamylase. The pattern of codon preferences was quite consistent among the three filamentous ascomycetes examined to date (*A. awamori*, *T. reesei*, and *N. crassa*) and adhered to the preferences previously noted by Schechtman and Yanofsky (50). This pattern differed strikingly, in several instances, from that observed in a compilation of genes from the single-celled ascomycete *S. cerevisiae* (5); differences in codon preference were most evident in Glu, Gln, Cys, Pro, and Arg codons. We speculate that these differences may reflect evolutionary distance between the yeasts and the filamentous ascomycetes (13).

We have described several structural aspects of the *A. awamori* glucoamylase gene and have compared structural and regulatory features with those described in other eucaryotic gene systems. Differences exist in features thought to be involved in gene expression in *S. cerevisiae* and higher eucaryotes. This is perhaps best demonstrated by the inability of *S. cerevisiae* to process heterologous intervening sequences (4, 29). Species specificity, and evolution, of control elements is not unexpected. Ascomycetes other than *S. cerevisiae* have been studied to a limited extent; examination of structural differences between genes of the single-celled ascomycete *S. cerevisiae* and the filamentous ascomycetes, e.g., *A. awamori*, may be informative in defining and studying the evolution of eucaryotic gene control elements.

ACKNOWLEDGMENTS

We thank our colleagues at Cetus who were involved in aspects of this work: D. Gelfand for many helpful discussions throughout the work; S. Shoemaker, R. Cox, J. Thomson, and K. Watt for glucoamylase purification, protein sequencing, and antibody production; K. Myambo for assistance in DNA sequencing; and D. Mark and S. D. Yu for preparation of the cDNA library. We thank T. White for support and helpful discussions. The oligonucleotide was provided by the DNA synthesis group. We also thank E. McCallan, E. Ladner, and L. Pearce for assistance in preparation of the manuscript.

This work was supported by National Distillers and Chemical Corp.

LITERATURE CITED

1. **Alwine, J. C., D. J. Kemp, and G. R. Stark.** 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U.S.A. **74**:5350–5354.
2. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid cellulose. Proc. Natl. Acad. Sci. U.S.A. **69**:1408–1412.
3. **Bailey, J. M., and N. Davidson.** 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Anal. Biochem. **70**:75–85.
4. **Beggs, J. D., J. van den Berg, A. van Ocyer, and C. Weissman.** 1980. Abnormal expression of chromosomal rabbit β-globin gene in *Saccharomyces cerevisiae*. Nature (London) **283**:835–840.
5. **Bennetzen, J. L., and B. D. Hall.** 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.
6. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
7. **Berger, S. L., and C. S. Birkenmeier.** 1979. Inhibition of

intractable nucleases with ribonucleoside-vanadyl complexes: isolation of messenger ribonucleic acid from resting lymphocytes. Biochemistry **18**:5143–5149.
8. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
9. **Blattner, F. R., A. E. Blechl, K. Denniston-Thompson, H. E. Faber, J. E. Richards, J. L. Slightom, P. W. Tucker, and O. Smithies.** 1978. Cloning human fetal λ globin and mouse α-type globin DNA: preparation and screening of shotgun collections. Science **202**:1279–1283.
10. **Boel, E., I. Hjort, B. Svensson, F. Norris, and K. E. Norris.** 1984. Glucoamylases G1 and G2 from *Aspergillus niger* are synthesized from two different but closely related mRNAs. EMBO J. **3**:1097–1102.
11. **Breatnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
12. **Chang, A. C. Y., J. H. Nunberg, R. J. Kaufman, H. A. Erlich, R. T. Schimke, and S. N. Cohen.** 1978. Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. Nature (London) **275**:617–624.
13. **Chen, M.-W., A. Jozef, G. Volckaert, E. Huysmans, A. Vandenberghe, and R. De Wachter.** 1984. The nucleotide sequences of the 5 S rRNAs of seven molds and a yeast and their use in studying ascomycete phylogeny. Nucleic Acids Res. **12**:4881–4892.
14. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
15. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, A. E. Blechl, J. L. Slightom, O. Smithies, F. E. Baralle, C. C. Shoulder, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
16. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24**:251–260.
17. **Fogarty, W. M., and C. T. Kelley.** 1979. Starch degrading enzymes of microbial origin, p. 87–150. *In* M. J. Bull (ed.), Progress in industrial microbiology, vol. 15. Elsevier Scientific Publishing Co., Amsterdam.
18. **Gallwitz, D., and I. Sures.** 1980. Structure of a split yeast gene: complete nucleotide sequence of the actin gene, in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. **77**:2546–2550.
19. **Gannon, F., K. O'Hare, F. Perrin, J. P. LePennec, C. Benoist, M. Cochet, R. Breathnach, A. Royal, A. Garapin, B. Cami, and P. Chambon.** 1979. Organization and sequences at the 5′ end of a cloned complete ovalbumin gene. Nature (London) **278**:428–434.
20. **Glisin, V., R. Crkvenjakow, and C. Byns.** 1974. Ribonucleic acid isolation by cesium chloride centrifugation. Biochemistry **13**:2633–2637.
21. **Grunstein, M., and D. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961–3965.
22. **Hoschke, A., E. Laszlo, and J. Hollo.** 1976. Application of cycloamylase ligand affinity chromatography for analysis of amylolytic enzymes. Starch **12**:426–432.
23. **Hu, N., and J. Messing.** 1982. The making of strand-specific M13 probes. Gene **17**:271–277.
24. **Ivarie, R. D., and P. P. Jones.** 1979. A rapid sensitive assay for specific protein synthesis in cells and in cell free translation: use of *Staphylococcus aureus* as an adsorbent for immune complexes. Anal. Biochem. **97**:24–35.
25. **Kaufer, N. F., H. M. Fried, W. F. Schwindinger, M. Jasin, and J. R. Warner.** 1983. Cycloheximide resistance in yeast: the gene and its protein. Nucleic Acids Res. **11**:3123–3135.
26. **Kinnaird, J. H., M. A. Keighren, J. A. Kinsey, M. Eaton, and J. R. S. Fincham.** 1982. Cloning of the *am* (glutamate dehydrogenase) gene of *Neurospora crassa* through the use of a synthetic DNA probe. Gene **20**:387–396.
27. **Konigsberg, W., and G. N. Godson.** 1983. Evidence for use of rare codons in the dnaG gene and other regulatory genes for *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **80**:687–691.
28. **Land, H., M. Grez, H. Hauser, W. Lindenmaier, and G. Schultz.** 1981. 5′ terminal sequences of eucaryotic mRNA can be cloned with high efficiency. Nucleic Acids Res. **9**:2251–2266.
29. **Langford, C., W. Nellen, J. Niessing, and D. Gallwitz.** 1983. Yeast is unable to excise foreign intervening sequences from hybrid gene transcripts. Proc. Natl. Acad. Sci. U.S.A. **80**:1496–1500.
30. **Langford, C. J., and D. Gallwitz.** 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell **33**:519–527.
31. **Leer, R. J., M. M. C. van Raamsdonk-Duin, M. T. Molenaar, L. H. Cohen, W. H. Mager, and R. J. Planta.** 1982. The structure of the gene coding for the phosphorylated ribosomal protein S10 in yeast. Nucleic Acids Res. **10**:5869–5878.
32. **Lineback, D. R., L. A. Aira, and R. L. Horner.** 1972. Structural characterization of the two forms of glucoamylase from *Aspergillus niger*. Cereal Chem. **49**:283–298.
33. **Lineback, D. R., I. J. Russell, and C. Rasmussen.** 1969. Two forms of the glucoamylase of *Aspergillus niger*. Arch. Biochem. Biophys. **134**:539–553.
34. **Loenen, W. A., and W. J. Brammar.** 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene **10**:249–259.
35. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
36. **McDonell, M. W., M. N. Simon, and F. W. Studier.** 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. **110**:119–146.
37. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
38. **Miller, A. M.** 1984. The yeast MATa1 gene contains two introns. EMBO J. **3**:1061–1065.
39. **Mount, S. M.** 1982. A catalog of splice junction sequences. Nucleic Acids Res. **10**:459–472.
40. **Munjunath, P., and M. R. Raghavendra Rao.** 1980. Immunochemical relationship between glucoamylases I and II of *Aspergillus niger*. J. Biosci. **2**:163–169.
41. **Murray, M. G., and W. F. Thompson.** 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. **8**:4321–4325.
42. **Norman, B. E.** 1979. The application of polysaccharide degrading enzymes in the starch industry, p. 339–376. *In* R. C. W. Berkeley, G. W. Gooday, and D. C. Ellwood (ed.), Microbial polysaccharides and polysaccharases. Academic Press, Inc., New York.
43. **Payvar, F., and R. T. Schimke.** 1979. Methylmercury hydroxide enhancement of translation and transcription of ovalbumin and conalbumin mRNA's. J. Biol. Chem. **254**:617–624.
44. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
45. **Proudfoot, N. J., M. H. M. Shander, J. L. Manley, M. L. Gefter, and T. Maniatis.** 1980. Structure and *in vitro* transcription of human globin genes. Science **209**:1329–1336.
46. **Ramachandran, N., K. R. Sreekantiah, and V. Sreenivasa Murthy.** 1979. Influence of media composition on the production of alpha-amylase and amyloglucosidase by a strain of *Aspergillus niger*. Starch **31**:134–138.
47. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling DNA to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
48. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
49. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci.

U.S.A. **74**:5463–5467.

50. **Schechtman, M. G., and C. Yanofsky.** 1983. Structure of the trifunctional *trp*-1 gene from *Neurospora crassa* and its aberrant expression in *Escherichia coli*. J. Mol. Appl. Genet. **2**:83–99.
51. **Sehgal, P. B.** 1981. Procedures to estimate site of interferon mRNA and UV irradiation to estimate the site of its primary transcript. Methods Enzymol. **79**:111–124.
52. **Setzer, D. R., M. McGrogan, J. H. Nunberg, and R. T. Schimke.** 1980. Size heterogeneity in the 3′ end of dihydrofolate reductase messenger RNAs in mouse cells. Cell **22**:361–370.
53. **Setzer, D. R., M. McGrogan, and R. T. Schimke.** 1982. Nucleotide sequence surrounding multiple polyadenylation sites in the mouse dihydrofolate reductase gene. J. Biol. Chem. **257**:5143–5147.
54. **Shoemaker, S., V. Schweickart, M. Ladner, D. Gelfand, S. Kwok, K. Myambo, and M. Innis.** 1983. Molecular cloning of exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27. Biotechnology **1**:691–696.
55. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
56. **Steiner, D. F., P. S. Quinn, C. Patzelt, S. J. Chan, J. Marsh, and H. S. Tager.** 1980. Proteolytic cleavage in the posttranslational processing of proteins. p. 175–202. *In* D. M. Prescott and L. Goldstein (ed.), Cell biology, vol. 4. Academic Press, Inc., New York.
57. **Sternberg, N., D. Tiemeier, and L. Enquist.** 1977. *In vitro* packaging of a λ dam vector containing *Eco*RI DNA fragments of *Escherichia coli* and phage P1. Gene **1**:255–280.
58. **Svensson, B., K. Larsen, I. Svendsen, and E. Boel.** 1983. The complete amino acid sequence of the glycoprotein, glucoamylase G1, from *Aspergillus niger*. Carlsberg Res. Commun. **48**:529–544.
59. **Svensson, B., T. G. Pedersen, I. Svendsen, T. Sakai, and M. Ottesen.** 1982. Characterization of two forms of glucoamylase from *Aspergillus niger*. Carlsberg Res. Commun. **47**:55–69.
60. **Teem, J. L., and R. Rosbash.** 1983. The expression of a beta-galactosidase gene containing the ribosomal protein 51 intron is sensitive to the RNA 2 mutation of yeast. Proc. Natl. Acad. Sci. U.S.A. **80**:4403–4407.
61. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. **77**:5201–5205.
62. **Tosi, M., R. A. Young, O. Hagenbuchle, and U. Schibler.** 1981. Multiple polyadenylation sites in a mouse α-amylase gene. Nucleic Acids Res. **9**:2313–2323.
63. **Ueda, S., R. Ohba, and S. Kano.** 1974. Fractionation of glucoamylase system from black Koji mold and the effects of adding isoamylase and alpha-amylase on amylolysis by the glucoamylase fractions. Starch **26**:374–378.
64. **Vogt, V. M.** 1973. Purification and further properties of single-strand-specific nuclease from *Aspergillus oryzae*. Eur. J. Biochem. **33**:192–200.
65. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to DBM-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. **76**:3683–3687.
66. **Williams, B. G. and F. R. Blattner.** 1980. Bacteriophage lambda vectors for DNA cloning, p. 201–281. *In* J. K. Setlow and A. Hollaender (ed.), Genetic engineering, vol. 2. Plenum Publishing Corp., New York.
67. **Woudt, L. P., A. Pastink, A. E. Kempers-Veenstra, A. E. M. Jansen, W. H. Mager, and R. J. Planta.** 1983. The genes coding for histone H3 and H4 in *Neurospora crassa* are unique and contain intervening sequences. Nucleic Acids Res. **11**:5347–5360.
68. **Yamasaki, Y., Y. Suzuki, and J. Ozawa.** 1977. Three forms of α-glucosidase and glucoamylase from *Aspergillus awamori*. Agric. Biol. Chem. **41**:2149–2161.
69. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. U.S.A. **80**:1194–1198.
70. **Zaret, K. S., and F. Sherman.** 1982. DNA sequence required for efficient transcription termination in yeast. Cell **28**:563–573.