The EMBO Journal vol.3 no.7 pp.1581–1585, 1984

Two different types of intervening sequences in the glucoamylase gene from *Aspergillus niger*

E.Boel, M.T.Hansen, I.Hjort, I.Høegh and N.P.Fiil

Laboratory of Genetics, Novo Research Institute, DK-2880 Bagsvaerd, Copenhagen, Denmark

Communicated by N.P.Fiil

One single glucoamylase gene could be identified in the chromosomal DNA of *Aspergillus niger* by Southern blot analysis. This glucoamylase gene was isolated from a genomic library of *A. niger* DNA. The glucoamylase gene is situated on a 2.5-kb *Eco*RI-*Eco*RV fragment and contains five intervening sequences in the coding region. One 169-bp intron is involved in differential mRNA processing leading to the two different glucoamylase enzymes G1 and G2; the other four introns are all very small ranging from 55 to 75 bp in length. One intron has a significant homology to the coding region which immediately follows, and it contains the internal conserved sequence TACTAAC, which is also found in yeast chromosomal gene introns, and is thought to participate in mRNA splicing. Two transcription initiation sites and a typical eukaryotic promoter region with TATAAT and CAAT boxes are located upstream from the gene.

Key words: *A. niger* genomic library/internal conserved sequence/glucoamylase promoter/transcription initiation/mRNA splicing

Introduction

Glucoamylases (1,4-α-D-glucan glucohydrolase EC 3.2.1.3) are extracellular glycoproteins synthesized by several filamentous fungi. The enzymes, like the *Aspergillus niger* glucoamylases, are industrially important because of their capacity to release D-glucose from the non-reducing ends of starch and related oligo- and polysaccharides (Pazur and Ando, 1959; Reilly, 1979). Most fungal glucoamylases are found in multiple forms varying in size (Pazur *et al.*, 1971; Ueda, 1981). Two glucoamylases from *A. niger*, G1 and G2, sharing a common NH_2-terminal polypeptide chain, but differing in the COOH-terminal region, have been characterized (Svensson *et al.*, 1982). The total amino acid sequence of the *A. niger* glucoamylase G1 has been determined (Svensson *et al.*, 1983). Recently, we described the molecular cloning of cDNA synthesized from *A. niger* glucoamylase mRNA, and suggested that differential mRNA splicing could explain the existence of the two forms of the enzyme in this organism (Boel *et al.*, 1984). Since differential splicing of primary transcripts leading to different protein products is an uncommon form of gene expression, we wanted to examine further the gene structure leading to this phenomenon in a lower eukaryote. We have therefore isolated genomic clones that contain glucoamylase-specific regions from an *A. niger* chromosomal DNA library. We present evidence that only one glucoamylase gene exists in the genome.

Results

Isolation of glucoamylase-specific genomic recombinants

The *A. niger* genomic library, constructed as described in Materials and methods, was screened with a single-stranded cDNA synthesized from *A. niger* poly(A)$^+$ RNA with a glucoamylase-specific tetradecamer mixture as primer (Boel *et al.*, 1984). Ten positive colonies were identified among 25 000 and, after a re-screening to purify the positive colonies, their DNA inserts were characterized with restriction endonucleases. Two types of glucoamylase-specific recombinants were identified. One type, pCAMG91 had an insert of 4.7 kb, and carried a glucoamylase gene on a 2.5-kb *Eco*RI-*Eco*RV fragment (Figure 1), while the other type, pCAMG101 had an insert of 4.8 kb and carried the 3′ part of the same glucoamylase gene on a 1.8-kb *Nde*I-*Eco*RV fragment.

Structure of the A. niger glucoamylase gene

The nucleotide sequence of the 2.5-kb *Eco*RI-*Eco*RV fragment from pCAMG91 was determined, and it contained the whole glucoamylase gene as shown in Figures 1 and 2. The 1.8-kb *Nde*I-*Eco*RV fragment of pCAMG101 represented the 3′ part of the glucoamylase gene starting at the *Nde*I restriction endonuclease site in intron B, and had the exact same DNA sequence as the gene from pCAMG91. The chromosomal nucleotide sequence was in total agreement with the sequence of the glucoamylase G1 cDNA (Boel *et al.*, 1984) except for the presence in the glucoamylase gene of four intervening sequences: A, B, C and D (Figures 1 and 2) in addition to the 169-bp intervening sequence found to be responsible for the production of the two glucoamylase enzyme variants G1 and G2 (Boel *et al.*, 1984). A typical eukaryotic promoter region was found at the 5′ end of the gene (Figure 2), and at its 3′ end the last 10 nucleotides just before the poly(A) addition site have five bases in common with the model sequence TTTTCACTGC of Benoist *et al.* (1980) for 3′-terminal regions of eukaryotic genes. Two out of three elements in the consensus yeast transcription termination sequence of Zaret and Sherman (1982), namely TAG....TAGT are located 56 bases upstream from the polyadenylation site, while the third element, TTT, is absent. As in many yeast mRNAs (Zaret and Sherman, 1982), the poly(A) addition signal AATAAA found in most mRNAs from higher eukaryotes is also absent.

Genomic DNA blot hybridization

Chromosomal *A. niger* DNA was digested with the restriction endonuclease enzymes *Eco*RI, *Eco*RV, *Nde*I, *Bgl*II and *Bss*HII, electrophoresed and blotted onto nitrocellulose. After hybridization of the blot with the glucoamylase-specific 1100 bases long cDNA synthesized with *A. niger* mRNA as template and a tetradecamer mixture as primer (Boel *et al.*,

© IRL Press Limited, Oxford, England.

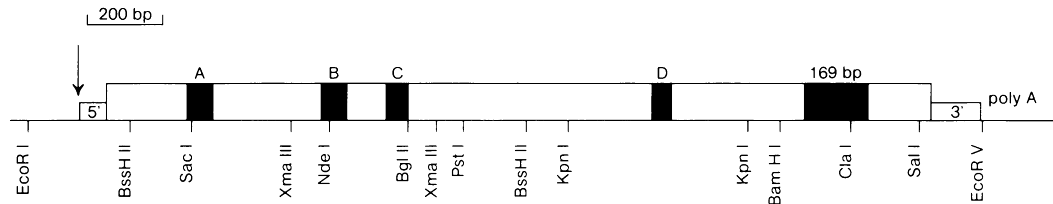

Fig. 1. Structure and restriction endonuclease map of the glucoamylase gene from *A. niger*. The 5'- and 3'-untranslated ends of the mRNA coding regions are shown by half thickness. Vertical arrow shows initiation of transcription. Black boxes are intervening sequences. The 169-bp intron is only spliced out to generate the glucoamylase G2 mRNA, while it is not removed in the G1 mRNA (Boel *et al.*, 1984).

```
                                                                                                190
                                                                                 TTCGTCGCCTAATGTCTCGTCCGTTCACAAACTG
                  150                         120                            90                        60
AAGAGCTTGAAGTGGCGAGATGTCTCTGCAGGAATTCAAGCTAGATGCTAAGCGATATTGCAT [GGCAATAT] GTGTTGATGCATGTGCTTCTTCCTTCAGCTTCCCCTCGTGCGAGTGA
                                EcoRI
         30                               1
GGTTTGGC [TATAAAT] TGAAGTGGTTGGTCGGGGTTCCGTGAGGGGCTGAAGTGCTTCCTCCCTTTTAGGCGCAACTGAGAGCCTGAGCTTCATCCCCAGCATCATTACACCTCAGCA

                                     10                                     20                                          30
ATG TCG TTC CGA TCT CTA CTC GCC CTG AGC GGC CTC GTC TGC ACA GGG TTG GCA AAT GTG ATT TCC AAG CGC GCG ACC TTG GAT TCA TGG
Met Ser Phe Arg Ser Leu Leu Ala Leu Ser Gly Leu Val Cys Thr Gly Leu Ala Asn Val Ile Ser Lys Arg Ala Thr Leu Asp Ser Trp
                                     40                                     50                                          60
TTG AGC AAC GAA GCG ACC GTG GCT CGT ACT GCC ATC CTG AAT AAC ATC GGG GCG GAC GGT GCT TGG GTG TCG GGC GCG GAC TCT GGC ATT
Leu Ser Asn Glu Ala Thr Val Ala Arg Thr Ala Ile Leu Asn Asn Ile Gly Ala Asp Gly Ala Trp Val Ser Gly Ala Asp Ser Gly Ile
                                     70
GTC GTT GCT AGT CCC AGC ACG GAT AAC CCG GAC T GTATGTTTCGAGCTCAGATTTAGTATGAGTGTGTCATTGATTGATTGATGCTGACTGGCGTGTCGTTTGTTGTAG
Val Val Ala Ser Pro Ser Thr Asp Asn Pro Asp                                      IVS A
                                 80                                 90                                          100
AC TTC TAC ACC TGG ACT CGC GAC TCT GGT CTC GTC CTC AAG ACC CTC GTC GAT CTC TTC CGA AAT GGA GAT ACC AGT CTC CTC TCC ACC
Tyr Phe Tyr Thr Trp Thr Arg Asp Ser Gly Leu Val Leu Lys Thr Leu Val Asp Leu Phe Arg Asn Gly Asp Thr Ser Leu Leu Ser Thr
                                 110                                120                                         130
ATT GAG AAC TAC ATC TCC GCC CAG GCA ATT GTC CAG GGT ATC AGT AAC CCC TCT GGT GAT CTG TCC AGC GGC GCT GGT CTC GGT GAA CCC
Ile Glu Asn Tyr Ile Ser Ala Gln Ala Ile Val Gln Gly Ile Ser Asn Pro Ser Gly Asp Leu Ser Ser Gly Ala Gly Leu Gly Glu Pro
                                 140                                150                                         160
AAG TTC AAT GTC GAT GAG ACT GCC TAC ACT GGT TCT TGG GGA CGG CCG CAG CGA GAT GGT CCG GCT CTG AGA GCA ACT GCT ATG ATC GGC
Lys Phe Asn Val Asp Glu Thr Ala Tyr Thr Gly Ser Trp Gly Arg Pro Gln Arg Asp Gly Pro Ala Leu Arg Ala Thr Ala Met Ile Gly
                                                                                            170
TTC GGG CAG TGG CTG CTT GTATGTTCTCCACCCCCTTGCGTCTGATCTGTGACATATGTAGCTGACTGGTCAG GAC AAT GGC TAC ACC AGC ACC GCA ACG GAC
Phe Gly Gln Trp Leu Leu                     IVS B                                Asp Asn Gly Tyr Thr Ser Thr Ala Thr Asp
        180                                     190
ATT GTT TGG CCC CTC GTT AGG AAC GAC CTG TCG TAT GTG GCT CAA TAC TGG AAC CAG ACA GGA TAT G GTGTGTTTGTTTTATTTTAAATTTCCAAAGA
Ile Val Trp Pro Leu Val Arg Asn Asp Leu Ser Tyr Val Ala Gln Tyr Trp Asn Gln Thr Gly Tyr                         IVS C
                                  200                                210                                        220
TGCGCCAGCAGAGCTAACCCGCGATCGCAG AT CTC TGG GAA GAA GTC AAT GGC TCG TCT TTC TTT ACG ATT GCT GTG CAA CAC CGC GCC CTT GTC
                               Asp Leu Trp Glu Glu Val Asn Gly Ser Ser Phe Phe Thr Ile Ala Val Gln His Arg Ala Leu Val
                                 230                                240                                         250
GAA GGT AGT GCC TTC GCG ACG GCC GTC GGC TCG TCC TGC TCC TGG TGT GAT TCT CAG GCA CCC GAA ATT CTC TGC TAC CTG CAG TCC TTC
Glu Gly Ser Ala Phe Ala Thr Ala Val Gly Ser Ser Cys Ser Trp Cys Asp Ser Gln Ala Pro Glu Ile Leu Cys Tyr Leu Gln Ser Phe
                                 260                                270                                         280
TGG ACC GGC AGC TTC ATT CTG GCC AAC TTC GAT AGC AGC CGT TCC GGC AAG GAC GCA AAC ACC CTC CTG GGA AGC ATC CAC ACC TTT GAT
Trp Thr Gly Ser Phe Ile Leu Ala Asn Phe Asp Ser Ser Arg Ser Gly Lys Asp Ala Asn Thr Leu Leu Gly Ser Ile His Thr Phe Asp
                                 290                                300                                         310
CCT GAG GCC GCA TGC GAC GAC TCC ACC TTC CAG CCC TGC TCC CCG CGC GCG CTC GCC AAC CAC AAG GAG GTT GTA GAC TCT TTC CGC TCA
Pro Glu Ala Ala Cys Asp Asp Ser Thr Phe Gln Pro Cys Ser Pro Arg Ala Leu Ala Asn His Lys Glu Val Val Asp Ser Phe Arg Ser
                                 320                                330                                         340
ATC TAT ACC CTC AAC GAT GGT CTC AGT GAC AGC GAG GCT GTT GCG GTG GGT CGG TAC CCT GAG GAC ACG TAC TAC AAC GGC AAC CCG TGG
Ile Tyr Thr Leu Asn Asn Gly Leu Ser Asp Ser Glu Ala Val Ala Val Gly Arg Tyr Pro Glu Asp Thr Tyr Tyr Asn Gly Asn Pro Trp
                                 350                                360                                         370
TTC CTG TGC ACC TTG GCT GCC GCA GAG CAG TTG TAC GAT GCT CTA TAC CAG TGG GAC AAG CAG GGG TCG TTG GAG GTC ACA GAT GTG TCG
Phe Leu Cys Thr Leu Ala Ala Ala Glu Gln Leu Tyr Asp Ala Leu Tyr Gln Trp Asp Lys Gln Gly Ser Leu Glu Val Thr Asp Val Ser
                                 380                                390                                         400
CTG GAC TTC TTC AAG GCA CTG TAC AGC GAT GCT GCT ACT GGC ACC TAC TCT TCG TCC AGT TCG ACT TAT AGT AGC ATT GTA GAT GCC GTG
Leu Asp Phe Phe Lys Ala Leu Tyr Ser Asp Ala Ala Thr Gly Thr Tyr Ser Ser Ser Ser Ser Thr Tyr Ser Ser Ile Val Asp Ala Val
                                 410
AAG ACT TTC GCC GAT GGC TTC GTC TCT ATT GTG GTAAGTCTACGCTAGACAAGCGCTCATGTTGACAGAGGGTGCGTACTAACAGAAGTAG GAA ACT CAC GCC
Lys Thr Phe Ala Asp Gly Phe Val Ser Ile Val                       IVS D                                  Glu Thr His Ala
            420                                 430                                         440
GCA AGC AAC GGC TCC ATG TCC GAG CAA TAC GAC AAG TCT GAT GGC GAG CAG CTT TCC GCT CGC GAC CTG ACC TGG TCT TAT GCT GCT CTG
Ala Ser Asn Gly Ser Met Ser Glu Gln Tyr Asp Lys Ser Asp Gly Glu Gln Leu Ser Ala Arg Asp Leu Thr Trp Ser Tyr Ala Ala Leu
            450                                 460                                         470
CTG ACC GCC AAC AAC CGT CGT AAC TCC GTC GTG CCT GCT TCT TGG GGC GAG ACC TCT GCC AGC AGC GTG CCC GGC ACC TGT GCG GCC ACA
Leu Thr Ala Asn Asn Arg Arg Asn Ser Val Val Pro Ala Ser Trp Gly Glu Thr Ser Ala Ser Ser Val Pro Gly Thr Cys Ala Ala Thr
            480                                 490                                         500
TCT GCC ATT GGT ACC TAC AGC AGT GTG ACT GTC ACC TCG TGG CCG AGT ATC GTG GCT ACT GGC GGC ACC ACT ACG ACG GCT ACC CCC ACT
Ser Ala Ile Gly Thr Tyr Ser Ser Val Thr Val Thr Ser Trp Pro Ser Ile Val Ala Thr Gly Gly Thr Thr Thr Thr Ala Thr Pro Thr
            510                                 520                                         530
GGA TCC GGC AGC GTG ACC TCG ACC AGC AAG ACC ACC GCG ACT GCT AGC AAG ACC AGC ACC A|GT ACG TCA→TCA ACC TCC TGT ACC ACT CCC
Gly Ser Gly Ser Val Thr Ser Thr Ser Lys Thr Thr Ala Thr Ala Ser Lys Thr Ser Thr Ser Thr Ser Ser Thr Ser Cys Thr Thr Pro
            540                                 550                                         560
ACC GCC GTG GCT GTG ACT TTC GAT CTG ACA GCT ACC ACC ACC TAC GGC GAG AAC ATC TAC CTG GTC GGA TCG ATC TCT CAG CTG GGT GAC
Thr Ala Val Ala Val Thr Phe Asp Leu Thr Ala Thr Thr Thr Tyr Gly Glu Asn Ile Tyr Leu Val Gly Ser Ile Ser Gln Leu Gly Asp
            570                                 580                                         590
TGG GAA ACC AGC GAC GGC ATA GCT CTG AGT GCT GAC AAG TAC←ACT TCC AG|C GAC CCG CTC TGG TAT GTC ACT GTG ACT CTG CCG GCT GGT
Trp Glu Thr Ser Asp Gly Ile Ala Leu Ser Ala Asp Lys Tyr Thr Ser Ser Asp Pro Leu Trp Tyr Val Thr Val Thr Leu Pro Ala Gly
            600                                 610                                         620
GAG TCG TTT GAG TAC AAG TTT ATC CGC ATT GAG AGC GAT GAC TCC GTG GAG TGG GAG AGT GAT CCC AAC CGA GAA TAC ACC GTT CCT CAG
Glu Ser Phe Glu Tyr Lys Phe Ile Arg Ile Glu Ser Asp Asp Ser Val Glu Trp Glu Ser Asp Pro Asn Arg Glu Tyr Thr Val Pro Gln
            630                                 640
GCG TGC GGA ACG TCG ACC GCG ACG GTG ACT GAC ACC TGG CGG TGA CAATCAATCCATTTCGCTATAGTTAAAGGATGGGGATGAGGGCAATTGGTTATATGATCATG
Ala Cys Gly Thr Ser Thr Ala Thr Val Thr Asp Thr Trp Arg Stop
                                                          ↓
TATGTAGTGGGTGTGCATAATAGTAGTGAAATGGAAGCCAAGTCATGTGATTGTAATCGACCGACGGAATTGAGGATATCCGGAAATACAGACACCGGG
                                                                              EcoRV
```

Fig. 2. Nucleotide sequence of the glucoamylase gene from *A. niger* as determined from the insert of a genomic recombinant pCAMG91 and confirmed from pCAMG101. IVS designates intervening sequences not found in the glucoamylase mRNAs, while the 169-bp intron between Ser (527) and Ser (583) is maintained in the G1 mRNA and spliced out in the G2 mRNA (Boel *et al.*, 1984). CAAT and TATAA regions in the promoter are boxed, and the site of polyadenylation is shown with an arrow. The two transcription initiation sites are indicated by open horizontal arrows, and bases upstream from the first site of transcription start have been assigned negative numbers. Sequences in IVS A, B, C and the 169 bases IVS which show homology to conserved sequences from *T. reesei* (Shoemaker *et al.*, 1983) have been underlined. The sequence in IVS D which is homologous to the internal conserved sequence from yeast (Langford and Gallwitz, 1983) has also been underlined.

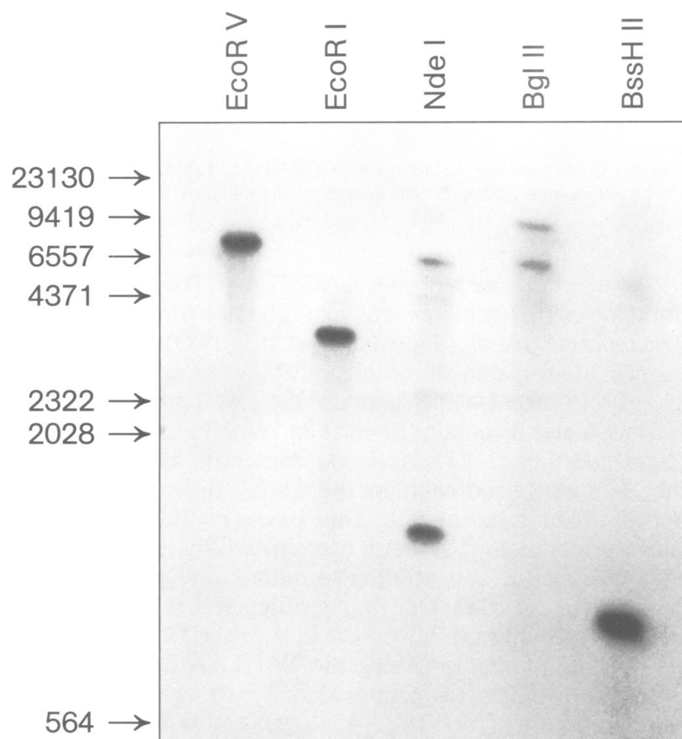

Fig. 3. Southern blot analysis of *A. niger* chromosomal DNA. DNA was digested with restriction enzymes as indicated, run on a 0.7% agarose gel, blotted onto nitrocellulose and gene-containing fragments were visualized by autoradiography. λ DNA cut with *Hind*III was used as size marker.

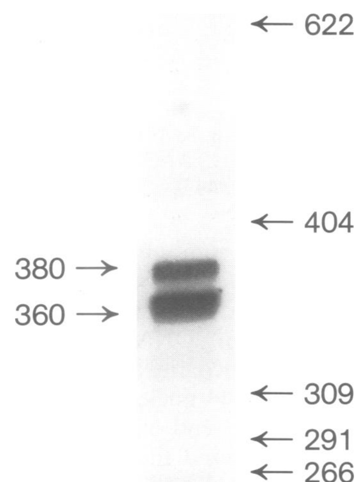

Fig. 4. Autoradiograph of a 5% polyacrylamide/urea gel of d(GTAGTTCTCAATGGT) primed glucoamylase cDNA (Boel *et al.*, 1984). The cDNA was transcribed from 5 μg of *A. niger* total mRNA, using 5′ end-labeled primer and unlabeled dNTPs. The two reverse transcripts (360 and 380 bases) were sequenced according to Maxam and Gilbert (1980) to give the sequence from the point of priming and up to the 5′ end of the glucoamylase mRNA. The nucleotide length of molecular markers are indicated.

1984) it was autoradiographed (Figure 3). There are no *Eco*RI and *Eco*RV restriction sites in the glucoamylase gene, and only one band was found to hybridize when the DNA was cut with either of these enzymes. *Nde*I and *Bgl*II each cut the glucoamylase gene once, and two hybridizing bands were detected. *Bss*HII cuts the glucoamylase gene twice to give an internal fragment of 1035 bp and two flanking fragments. The 1100 bases long cDNA probe used for hybridization covers the 1035-bp *Bss*HII fragment, while each of the two flanking fragments have only a short overlap to the probe. The observed strong hybridization to the internal fragment is therefore to be expected, while the flanking fragments can only be seen after prolonged exposure. These data imply that the *A. niger* genome contains only one glucoamylase gene, i.e. the one isolated in the genomic recombinant pCAMG91.

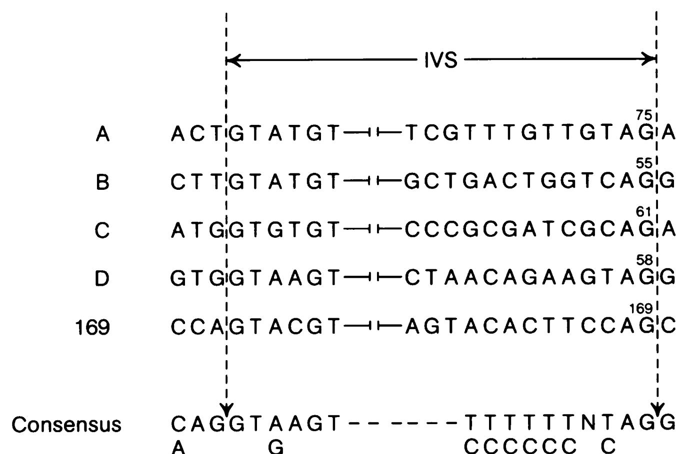

Fig. 5. Intron/exon junctions from the five intervening sequences in the *A. niger* glucoamylase gene. The junctions are aligned with the corresponding consensus sequences for eukaryotic splice sites (Mount, 1982). The length of the intervening sequences is indicated.

Discussion

The glucoamylase gene promoter

To localize the initiation point of transcription, cDNA was synthesized on *A. niger* mRNA with a 5′ end-labeled glucoamylase-specific pentadecamer as primer (Boel *et al.*, 1984). The primer is complementary to the glucoamylase mRNA in the region coding for Thr-101 to Tyr-105. Resolution of the cDNA products on an acrylamide/urea gel (Figure 4) revealed two major bands corresponding to 360 and 380 bases, respectively. The nucleotide sequences of these two cDNAs were determined from the point of priming up to the 5′ end of the mRNA. We have identified the initiation points of transcription, by comparing the mRNA sequence with the corresponding gene sequence. From this experiment we conclude that the glucoamylase gene has two initiation points as shown in Figure 2. The long reverse transcript (380 bases) pointed to the second A in the sequence GAAG as the initiation point, while the shorter reverse transcript (360 bases) identified the second A in the sequence CAAC as the initiation point (Figure 2). Multiple initiation points in genes from lower eukaryotes have been demonstrated in, for example, *Saccharomyces cerevisiae* alcohol dehydrogenase I (ADHI) (Bennetzen and Hall, 1982), iso-1-cytochrome (CYC1) (Faye *et al.*, 1981) and the tryptophan synthase gene (TRP5) (Zalkin and Yanofsky, 1982), and in *Schizosaccharomyces pombe* cytochrome C (Russell, 1983). Transcription from the ADH I gene of *S. cerevisiae* starts at two different sites, in both cases at the second A in the sequence CAAG. In the 3-phosphoglycerate kinase gene from the same organism, transcription starts at the second A in the sequence GAAG (Hitzeman *et*

Fig. 6. Sequence homology between the 58 bases of intron D and the start of the following region encoding the amino acids Glu(413) to Lys(428). Homolc is indicated with asterisks. The TACTAAC internal conserved sequence in the intron is boxed. Amino acids are numbered from the initiating methionine. Horizontal arrows mark the start and the stop of the D intron.

al., 1982). Thus there are a number of examples of highly transcribed genes in the lower eukaryotes, on which transcription starts at the second A in the sequence C/G-A-A-C/G. Judged from the amount of cDNA synthesized in the above experiment, 70% of the glucoamylase mRNAs would have a 5′-untranslated region of 44 nucleotides, while 30% of the transcripts would have a 5′ end of 68 nucleotides. A typical TATAA box is found around the −35 region of the promoter (Figure 2) and in the −100 region there is a typical CAAT box (Breathnach and Chambon, 1981). In that respect the promoter is very similar to other eukaryotic promoters.

Arrangement and structure of the intervening sequences

Five intervening sequences have been localized in the glucoamylase gene (Figures 1 and 2). The 169 bp long intron has been described recently and was found to be responsible for the expression of two glucoamylase enzymes G1 and G2 from only one gene by differential mRNA processing (Boel *et al.*, 1984). The other four introns are all very short, ranging from 55 to 75 nucleotides, and none of them has been found to persist in a translatable mRNA. Since data concerning the active site or other functional domains in the glucoamylase enzyme are not presently available, we cannot yet ascribe any domain-dividing function, if any, to the introns in this gene.

The exon/intron junctions have been compared (Figure 5) with the consensus for eukaryotic splice junctions compiled by Mount (1982). All the introns begin with GTNNGT as has also been noticed in yeast 5′ splice junctions (Pikielny *et al.*, 1983). The C- and T-rich region found at the 3′ end of introns in other eukaryotes is less pronounced in this *A. niger* gene, but all five introns end in T/CAG.

The nucleotide sequence of intron D displays two special features. First it is very homologous (Figure 6) to the subsequent coding region beginning with Glu (413). To the extent that the sequence of the intron diverges from the following coding region, the divergence occurs preferentially within a specific region. This region coincides with the second special feature of intron D, namely the occurrence of the internal conserved sequence TACTAAC which is also found in yeast chromosomal gene introns (Pikielny *et al.*, 1983; Langford and Gallwitz, 1983). As proposed by these authors for the yeast genes, the TACTAAC sequence could function in a self-splicing mechanism by substituting for the U1 snRNA in aligning the junction regions for proper mRNA processing. However, only intron D has the TACTAAC sequence and, therefore, might be spliced by this proposed *cis*-acting mechanism; introns A, B, C and the 169-bp intron would presumably have to be processed by a mechanism involving the *trans*-acting U1 snRNAs, as has been proposed for the intervening sequences of higher eukaryotes. snRNAs have recently been described in yeast (Wise *et al.*, 1983; Tollervey *et al.*, 1983), but none of them have been identified as belonging to the mammalian U1 type, and all *S. cerevisiae* chromosomal gene introns characterized so far have the TACTAAC sequence.

The internal base sequence CAGCTGACTG was foun both introns of the gene for exo-cellobiohydrolase I fr *Trichoderma reesei* (Shoemaker *et al.*, 1983). Similar quences are found in all the introns of the glucoamylase ge except for intron D which contains the TACTAAC sequer Introns A and B contain the eight and nine 3′-terminal ba respectively, of the *Trichoderma* consensus sequence w the 169 base intron contains the whole consensus seque but in a form interrupted by four bases (Figure 2). Introı shows less homology in that it contains a 12-base interrupt of the sequence and at three positions one pyrimidin changed for the other. Our findings suggest that two differ splicing mechanisms are working in *A. niger*; one would t self-splicing mechanism using the TACTAAC sequence junction alignment, as proposed for yeast, while the ot could make use of a U1 snRNA (Rogers and Wall, 1980). glucoamylase gene from *A. niger* has thus been demonstra to contain two different types of intervening sequences: is tion of more chromosomal genes from the filamentous fu and related lower eukaryotes might help to clarify the r tionship between the two types of introns.

Materials and methods

Enzymes, reagents, mRNA extraction and cDNA synthesis have been des ed (Boel *et al.*, 1984).

Preparation of A. niger DNA

A 500 ml culture of *A. niger* strain BU-1 grown out on YPD (Sherman e 1981) was concentrated by centrifugation, washed and resuspended in 5 of SED (1 M sorbitol, 25 mM Na_2EDTA, pH 8.0, 6.7 mg/ml DTT) cubated at 30°C for 15 min, recentrifuged and suspended in 50 ml of (1 M sorbitol, 0.1 M Na-citrate, 10 mM Na_2EDTA, pH 5.8). Cell degrading enzyme, Novozyme 234 (50 mg), was added and incubation tinued at 30°C for 30 min. After centrifugation and resuspension in 50 n TE (10 mM Tris, pH 8.0, 1 mM EDTA), glass beads (0.5 mm diameter) added, and the mixture shaken by hand for 3 x 1 min. 20 ml of 10% SDS 50 ml of phenol was added and the mixture again shaken. The water p was separated by centrifugation, re-extracted once with phenol and twice PCI (phenol, chloroform, isoamylalcohol, 50:50:1) before ethanol preci tion. The resuspended nucleic acids were treated with RNase A (20 μg/r 4°C overnight) followed by banding of the DNA on a CsCl equilibrium d ty gradient. After dialysis, the isolated DNA was again phenol and PC tracted, ethanol precipitated and resuspended in TE. The final yield of D was ~2 mg and of a mol. wt. >30 kb as judged from agarose gel trophoresis.

Construction of a genomic library

Samples of the purified *A. niger* DNA were digested with either *Mbo*I or to find the amounts of enzyme needed to produce fragments predominant the range 4−12 kb. Half this amount of each enzyme was used i preparative digest used for the library. The digest was electrophoresed 0.7% agarose gel and fragments in the range 4−12 kb isolated on DE membranes (NA45, Schleicher and Schull). The pooled fragments ligated onto the isolated 2.3-kb *Bam*HI-*Nde*I fragment containing th lactamase gene and the origin of replication from pBR322. The ligation ı ture was used for transformation into an *E. coli* MC1000 (Casabadan Cohen, 1980) derivative (r^-m^+,leuB6), selecting for ampicillin resistance

Colony screening

The genomic library constructed in pBR322 was screened for glucoamy specific recombinants by the methods of Hanahan and Meselson (1983) glucoamylase-specific cDNA (1100 bases long) synthesized with *A. n*

mRNA as template and the tetradecamer mixture d(T-T-A/G-T-C-C-C-A-T/C-T-G-A/G-T-A) as primer (Boel *et al.*, 1984) was used as hybridization probe.

Plasmid DNA isolation and DNA sequence analysis

DNA was isolated from *Escherichia coli* as described (Birnboim and Doly, 1979). The nucleotide sequence of DNA fragments were determined by the chemical cleavage procedure (Maxam and Gilbert, 1980).

Southern genomic blot analysis

A. niger DNA was digested with *Eco*RV, *Eco*RI, *Nde*I, *Bgl*II and *Bss*HII, electrophoresed on a 0.7% agarose gel and blotted onto nitrocellulose paper according to Southern (1975). The blot was performed in a Bio Rad trans-blot cell run at 10 V in 5 mM NaOH, 250 mM ammonium acetate. The immobilized DNA was hybridized at 68°C for 48 h with the primer-extended urea gel purified cDNA described under Colony screening. The hybridization buffer was 2 x SSC (1 x SSC is 0.150 M NaCl and 0.015 M sodium citrate, pH 7), 0.2% bovine serum albumin, 0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.1% SDS and 50 μg/ml sonicated, denatured salmon sperm DNA. Four washings were carried out at 42°C, each for 15 min in 1 x SSC. The dried filter was autoradiographed for 5 days using an intensifier screen.

Acknowledgements

The authors thank Ms. Ulla Rinus, Ms. Else Jørgensen and Ms. Jytte Ulrich for excellent secretarial assistance.

References

Bennetzen,J.L. and Hall,B.D. (1982) *J. Biol. Chem.* **257**, 3018-3025.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.,* **8**, 127-142.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.,* **7**, 1513-1523.
Boel,E., Hjort,I., Svensson,B., Norris,F., Norris,K.E. and Fiil,N.P. (1984) *EMBO J.,* **3**, in press.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.,* **50**, 349-383.
Casabadan,M.J. and Cohen,S.N. (1980) *J. Mol. Biol.,* **138**, 179-207.
Faye,G., Leung,D.W., Tatchell,K., Hall,B.D. and Smith,M. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 2258-2262.
Hanahan,D. and Meselson,M. (1983) *Methods Enzymol.,* **100**, 333-342.
Hitzeman,R.A., Hagie,F.E., Hayflick,J.S., Chen,C.Y., Seeburg,P.H. and Derynck,R. (1982) *Nucleic Acids Res.,* **23**, 7791-7808.
Langford,C.J. and Gallwitz,D. (1983) *Cell,* **33**, 519-527.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 499-560.
Mount,S.M. (1982) *Nucleic Acids Res.,* **10**, 459-472.
Pazur,J.H. and Ando,T. (1959) *J. Biol. Chem.,* **234**, 1966-1970.
Pazur,J.H., Knull,H.R. and Cepure,A. (1971) *Carbohydr. Res.,* **20**, 83-96.
Pikielny,C.W., Teen,J.L. and Rosbash,M. (1983) *Cell,* **34**, 395-403.
Reilly,P.J. (1979) *Appl. Biochem. Bioeng.,* **2**, 185-206.
Rogers,J. and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 1877-1879.
Russell,P. (1983) *Nature,* **301**, 167-169.
Sherman,F., Fink,G.R. and Hicks,J.B. (1981) *Methods in Yeast Genetics,* published by Cold Spring Harbor Laboratory Press, NY.
Shoemaker,S., Schweickart,V., Ladner,M., Gelfand,D., Kwok,S., Myambo, K. and Innis,M. (1983) *BIO/TECHNOLOGY,* **1**, 691-696.
Southern,E.M. (1975) *J. Mol. Biol.,* **98**, 503-517.
Svensson,B., Pedersen,T.G., Svendsen,I., Sakai,T. and Ottesen,M. (1982) *Carlsberg Res. Commun.,* **47**, 55-69.
Svensson,B., Larsen,K., Svendsen,I. and Boel,E. (1983) *Carlsberg Res. Commun.,* **48**, 529-544.
Tollervey,D., Wise,J.A. and Guthrie,C. (1983) *Cell,* **35**, 753-762.
Ueda,S. (1981) *Trends Biochem. Sci.,* **6**, 89-90.
Wise,J.A., Tollervey,D., Maloney,D., Swerdlow,H., Dunn,E.J. and Guthrie, C. (1983) **35**, 743-751.
Zalkin,H. and Yanofsky,C. (1982) *J. Biol. Chem.,* **257**, 1491-1500.
Zaret,K.S. and Sherman,F. (1982) *Cell,* **28**, 563-573.

Received on 19 March 1984; revised on 19 April 1984