

Three α -amylase genes of *Aspergillus oryzae* exhibit identical intron-exon organization

S. Wirsel,* A. Lachmund, G. Wildhardt and E. Rutkowski*

Institut für Biochemie, Technische Hochschule Darmstadt, Petersenstraße 22, D-6100 Darmstadt, FRG.

Summary

We have cloned three genes (*amy1*, *amy2* and *amy3*) encoding α -amylase in the filamentous fungus *Aspergillus oryzae*. The established overall sequences have a very high degree of homology, showing divergences mainly in the 3'-untranslated regions. The positions and the sequences of the eight introns were found to be absolutely identical in the three genes. The sequence analysis of the 5'-regions revealed presumptive TATA, CAAT and GC boxes. Primer extension analysis was performed to determine the transcription start. We were able to detect mRNAs from *amy1* and *amy3* but not from *amy2* with gene-specific oligonucleotide probes complementary to the 3'-noncoding regions.

Introduction

The major group of starch-hydrolysing enzymes, α -amylases (EC 3.2.1.1, α -1,4-glucan-4-glucanohydrolases), catalyse the endoamylolytic cleavage of α -1,4-glycosidic linkages in starch and related substrates releasing short oligosaccharides and α -limit dextrins. A wide range of organisms produces amylases. Alpha-amylases of microbial origin are exoenzymes and are widely used in biotechnology. Examples of the technical applications of fungal α -amylases include the production of bread, maltose syrups and asian fermentation products like soya sauce, miso and sake (Cheetham, 1985; Hesseltine, 1983).

Aspergillus oryzae is one of the most efficient α -amylase-secreting fungi. This amylase is well characterized, including the complete amino acid sequence (Toda *et al.*, 1982). From the three-dimensional structure, based on X-ray structure analysis, suggestions have arisen regarding the catalytic sites of this enzyme (Matsuura *et al.*, 1984). In contrast to the importance of this fungal enzyme and the well-established protein data there is a lack of information about the organization and the regulation of

the corresponding structural gene(s). Errat *et al.* (1984) considered an initial transcription control for the production of α -amylase in *A. oryzae*.

In higher eukaryotes, amylases often exist as different isoenzymes: for example the salivary and pancreatic types in mammals. The genetic organization of mammalian amylase genes and the regulation of their expression have been investigated (Gumucio *et al.*, 1985; 1988; Horii *et al.*, 1987; Emi *et al.*, 1988). It was shown that the human genome has a large clustered gene family encoding the different isoenzymes of α -amylase. Additionally, pseudogenes were detected in this gene family.

To date, the lower eukaryote *A. oryzae* is known to produce only one type of α -amylase. We were interested in verifying this assumption as well as in revealing the genetic organization of the structural gene(s). In this report we show the existence of several copies of α -amylase genes in the genome of *A. oryzae*. We analyse the coding regions and the distribution of introns in the genes with the intention of clarifying the existence of different isoenzymes. A further question resulting from the initial experiments was which gene copy is expressed in the fungus, with regard to the presumptive regulatory regions. A means of answering this question will be discussed with respect to the data obtained.

Results

Analysis of cDNA coding for *A. oryzae* α -amylase

Poly(A)⁺ RNA isolated from starch-grown mycelia of *A. oryzae* was used to construct a cDNA library. About 10 000 recombinant colonies were screened using a 5'-phosphorylated 63-mer derived from the published sequence of Taka amylase A (Toda *et al.*, 1982). A total of 26 positive clones were obtained. The recombinant plasmids contained cDNA inserts ranging in size from about 1100 to 1800 bp. Overlapping subclones of the longest insert (pGW8) were used to establish the nucleotide sequence of the cDNA (Fig. 1). The DNA sequence of 1822 bp comprises one open reading frame coding for 499 amino acids, a 5'-untranslated region of 20 bases, and a 3'-untranslated region of 278 bases excluding the poly(A) tail of 27 residues. Comparison with data obtained later from our genomic clones confirmed the cDNA sequence but revealed that the 5'-untranslated region of pGW8 lacks the first 42 nucleotides of the mRNA.

AAACCCACAGAAGGCATTTATG ATG GTC GCG TGG TGG TCT CTA TTT CTG TAC GGC CTT CAG GTC GCG GCA CCT GCT TTG GCT
 Met Met Val Ala Trp Trp Ser Leu Phe Leu Tyr Gly Leu Gln Val Ala Ala Pro Ala Leu Ala
 -20 -10
 GCA ACG CCT GCG GAC TGG CGA TCG CAA TCC ATT TAT TTC CTT CTC ACG GAT CGA TTT GCA AGG ACG GAT GGG TCG ACG
 Ala Thr Pro Ala Asp Trp Arg Ser Gln Ser Ile Tyr Phe Leu Leu Thr Asp Arg Phe Ala Arg Thr Asp Gly Ser Thr
 1 10 20
 ACT GCG ACT TGT AAT ACT GCG GAT CAG AAA TAC TGT GGT GGA ACA TGG CAG GGC ATC ATC GAC AAG TTG GAC TAT ATC
 Thr Ala Thr Cys Asn Thr Ala Asp Gln Lys Tyr Cys Gly Gly Thr Trp Gln Gly Ile Ile Asp Lys Leu Asp Tyr Ile
 30 40 50
 CAG GGA ATG GGC TTC ACA GCC ATC TGG ATC ACC CCC GTT ACA GCC CAG CTG CCC CAG ACC ACC GCA TAT GGA GAT GCC
 Gln Gly Met Gly Phe Thr Ala Ile Trp Ile Thr Pro Val Thr Ala Gln Leu Pro Gln Thr Thr Ala Tyr Gly Asp Ala
 60 70
 TAC CAT GGC TAC TGG CAG CAG GAT ATA TAC TCT CTG AAC GAA AAC TAC GGC ACT GCA GAT GAC TTG AAG GCG CTC TCT
 Tyr His Gly Tyr Trp Gln Gln Asp Ile Tyr Ser Leu Asn Glu Asn Tyr Gly Thr Ala Asp Asp Leu Lys Ala Leu Ser
 80 90 100
 TCG GCC CTT CAT GAG AGG GGG ATG TAT CTT ATG GTC GAT GTG GTT GCT AAC CAT ATG GGC TAT GAT GGA GCG GGT AGC
 Ser Ala Leu His Glu Arg Gly Met Tyr Leu Met Val Asp Val Val Ala Asn His Met Gly Tyr Asp Gly Ala Gly Ser
 110 120 130
 TCA GTC GAT TAC AGT GTG TTT AAA CCG TTC AGT TCC CAA GAC TAC TTC CAC CCG TTC TGT TTC ATT CAA AAC TAT GAA
 Ser Val Asp Tyr Ser Val Phe Lys Pro Phe Ser Ser Gln Asp Tyr Phe His Pro Phe Cys Phe Ile Gln Asn Tyr Glu
 140 150
 GAT CAG ACT CAG GTT GAG GAT TGC TGG CTA GGA GAT AAC ACT GTC TCC TTG CCT GAT CTC GAT ACC ACC AAG GAT GTG
 Asp Gln Thr Gln Val Glu Asp Cys Trp Leu Gly Asp Asn Thr Val Ser Leu Pro Asp Leu Asp Thr Thr Lys Asp Val
 160 170 180
 GTC AAG AAT GAA TGG TAC GAC TGG GTG GGA TCA TTG GTA TCG AAC TAC TCC ATT GAC GGC CTC CGT ATC GAC ACA GTA
 Val Lys Asn Glu Trp Tyr Asp Trp Val Gly Ser Leu Val Ser Asn Tyr Ser Ile Asp Gly Leu Arg Ile Asp Thr Val
 190 200
 AAA CAC GTC CAG AAG GAC TTC TGG CCC GGG TAC AAC AAA GCC GCA GGC GTG TAC TGT ATC GGC GAG GTG CTC GAC GGT
 Lys His Val Gln Lys Asp Phe Trp Pro Gly Tyr Asn Lys Ala Ala Gly Val Tyr Cys Ile Gly Glu Val Leu Asp Gly
 210 220 230
 GAT CCG GCC TAC ACT TGT CCC TAC CAG AAC GTC ATG GAC GGC GTA CTG AAC TAT CCC ATT TAC TAT CCA CTC CTC AAC
 Asp Pro Ala Tyr Thr Cys Pro Tyr Gln Asn Val Met Asp Gly Val Leu Asn Tyr Pro Ile Tyr Tyr Pro Leu Leu Asn
 240 250 260
 GCC TTC AAG TCA ACC TCC GGC AGC ATG GAC GAC CTC TAC AAC ATG ATC AAC ACC GTC AAA TCC GAC TGT CCA GAC TCA
 Ala Phe Lys Ser Thr Ser Gly Ser Met Asp Asp Leu Tyr Asn Met Ile Asn Thr Val Lys Ser Asp Cys Pro Asp Ser
 270 280
 ACA CTC CTG GGC ACA TTC GTC GAG AAC CAC GAC AAC CCA CGG TTC GCT TCT TAC ACC AAC GAC ATA GCC CTC GCC AAG
 Thr Leu Leu Gly Thr Phe Val Glu Asn His Asp Asn Pro Arg Phe Ala Ser Tyr Thr Asn Asp Ile Ala Leu Ala Lys
 290 300 310
 AAC GTC GCA GCA TTC ATC ATC CTC AAC GAC GGA ATC CCC ATC ATC TAC GCC GGC CAA GAA CAG CAC TAC GCC GGC GGA
 Asn Val Ala Ala Phe Ile Ile Leu Asn Asp Gly Ile Pro Ile Ile Tyr Ala Gly Gln Glu Gln His Tyr Ala Gly Gly
 320 330
 AAC GAC CCC GCG AAC CGC GAA GCA ACC TGG CTC TCG GGC TAC CCG ACC GAC AGC GAG CTG TAC AAG TTA ATT GCC TCC
 Asn Asp Pro Ala Asn Arg Glu Ala Thr Trp Leu Ser Gly Tyr Pro Thr Asp Ser Glu Leu Tyr Lys Leu Ile Ala Ser
 340 350 360
 GCG AAC GCA ATC CGG AAC TAT GCC ATT AGC AAA GAT ACA GGA TTC GTG ACC TAC AAG AAC TGG CCC ATC TAC AAA GAC
 Ala Asn Ala Ile Arg Asn Tyr Ala Ile Ser Lys Asp Thr Gly Phe Val Thr Tyr Lys Asn Trp Pro Ile Tyr Lys Asp
 370 380 390
 GAC ACA ACG ATC GCC ATG CGC AAG GGC ACA GAT GGG TCG CAG ATC GTG ACT ATC TTG TCC AAC AAG GGT GCT TCG GGT
 Asp Thr Thr Ile Ala Met Arg Lys Gly Thr Asp Gly Ser Gln Ile Val Thr Ile Leu Ser Asn Lys Gly Ala Ser Gly
 400 410
 GAT TCG TAT ACC CTC TCC TTG AGT GGT GCG GGT TAC ACA GCC GGC CAG CAA TTG ACG GAG GTC ATT GGC TGC ACG ACC
 Asp Ser Tyr Thr Leu Ser Leu Ser Gly Ala Gly Tyr Thr Ala Gly Gln Gln Leu Thr Glu Val Ile Gly Cys Thr Thr
 420 430 440
 GTG ACG GTT GGT TCG GAT GGA AAT GTG CCT GTT CCT ATG GCA GGT GGG CTA CCT AGG GTA TTG TAT CCG ACT GAG AAG
 Val Thr Val Gly Ser Asp Gly Asn Val Pro Val Pro Met Ala Gly Gly Leu Pro Arg Val Leu Tyr Pro Thr Glu Lys
 450 460
 TTG GCA GGT AGC AAG ATC TGT AGT AGC TCG TGA AGGGTGAGAGTATATGATGGTACTGCTATTCAATCTGGCATTGGACAGTGAGTTTGAG
 Leu Ala Gly Ser Lys Ile Cys Ser Ser Ser stop
 470 478
 TTTGATGTACAGTTGGAGTCGTTACTGCTGTCATCCCTTATACCTCTTCGATTGTTTTTCGAACCCCTAATGCCAAGCACGCTAGTCTATTATAGGAAAGGATC
 CGGATTAATGTGTTTTTCATAACCGGTACTGTATGGTACTTCTGTATTATATACCGAAGCTCATGTATCTTACATGTATATATTATACAGACACAACCTTGG

TTACCCACC (A) 27

Fig. 1. Nucleotide sequence of cloned *A. oryzae* α -amylase cDNA and deduced amino acid sequence. 1 indicates the first amino acid of the mature protein. Underlined positions indicate differences from the published sequence of Taka amylase A (Toda *et al.*, 1982).

Identification and cloning of *A. oryzae* genomic α -amylase DNAs

Southern-blot analysis of nuclear DNA was employed to obtain initial information about the organization of the α -amylase gene(s) in *A. oryzae*. DNA was hydrolysed with the restriction enzymes *Eco*RI, *Hind*III and *Sal*I. The blotted DNA was hybridized consecutively with two nick-translated *Pst*I fragments isolated from the cDNA clone, pGW8. Figure 2 shows the corresponding fluorographs. Hybridization with the 1450bp 3'-terminal *Pst*I fragment resulted in a pattern generally consisting of three bands. Rehybridization of the membrane-bound DNA with the 370bp 5'-terminal *Pst*I fragment generated the same bands with the exception of the *Sal*I lane, in which three additional bands appeared. *Sal*I cuts the cDNA once in a region covered by the second probe. This probe should therefore overlap the appropriate genomic *Sal*I fragments. The obtained hybridization patterns indicate the existence of three amylase genes in the genome of *A. oryzae*.

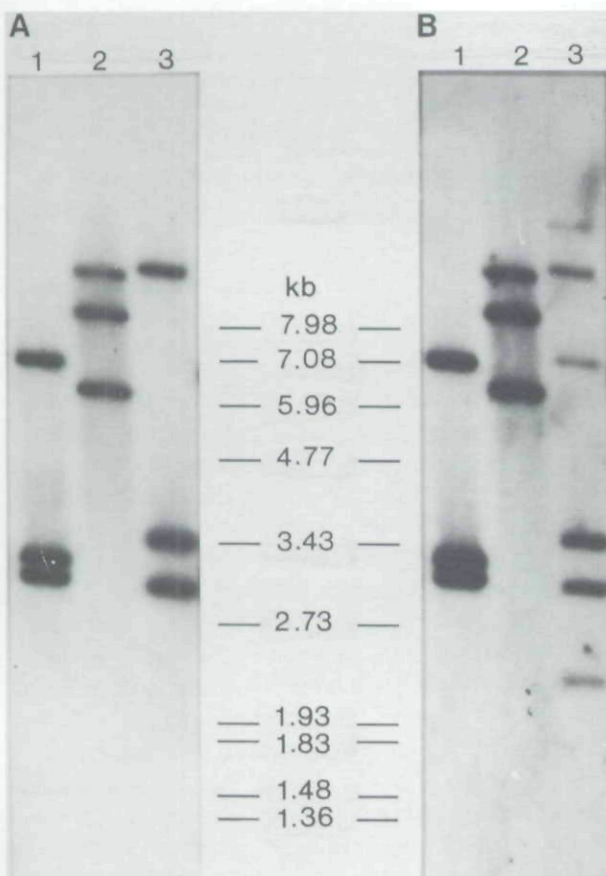


Fig. 2. Southern analysis of genomic DNA from *A. oryzae* DSM63303. DNAs hydrolysed with *Eco*RI (lane 1), *Hind*III (lane 2) and *Sal*I (lane 3) were separated on a 0.8% agarose gel. Molecular weight marker: phage SPP1 DNA hydrolysed with *Eco*RI. A. Hybridization with a nick-translated 1450bp *Pst*I fragment of pGW8. B. Hybridization with the corresponding 370bp *Pst*I probe.

A library of the *A. oryzae* genome was constructed in the phage λ substitution vector EMBL3 in order to isolate the three presumptive genes. From 50000 recombinant plaques screened with the 1450bp *Pst*I fragment of pGW8 we isolated 26 hybridizing clones. *Sal*I-hydrolysed phage DNAs were analysed by the same approach as that used in the genomic Southern hybridization experiments. Figure 3A demonstrates the fragment pattern of hybridization with the 1450bp probe; Fig. 3B represents the rehybridization with the 370bp 5'-terminal *Pst*I fragment. The hybridizing bands of the recombinant EMBL3 clones could be correlated to the corresponding bands of the *Sal*I-cut genomic DNA. The data show that a set of clones was isolated for each gene spanning about 20–30 kb of the corresponding genomic DNA loci. There are no indications for the clustering of the three genes within these regions. For further analysis the clones λ StW7, λ StW26 and λ StW28 were chosen, in that order, representing the genes named *amy1*, *amy2* and *amy3*, respectively.

Genetic organization of the three α -amylase genes in *A. oryzae*

Partial restriction maps were established and appropriate fragments of each gene subcloned in pUC19. The recombinant plasmids were called p19.7.4 *Eco*, p19.26.3 *Eco* and p19.28 *Sal*. Plasmid DNAs were subjected to DNA-sequencing reactions using universal primers. The sequences of the second strands were filled by priming further reactions with specific oligonucleotide probes. Figure 4 shows the major restriction sites for the established sequences and the genetic organization of the three genes. The overall DNA sequence of the genes is presented in Fig. 5. The three α -amylase genes in *A. oryzae* encode for nearly identical proteins of 478 residues. The genes *amy1* and *amy2* have identical sequences from the upstream *Eco*RI site down to the *Bgl*II site; *amy3* shows deviations from the other genes at two nucleotides which would change the amino acids at the appropriate positions. Alignment of the cDNA and the three genomic sequences demonstrates that the coding regions are disrupted by eight introns. It should be noted that there were differences neither in the localization of the introns nor in their sequences.

The transcription start was determined by primer extension (Fig. 6). An oligonucleotide which hybridizes to codons 6–10 of the mRNA could be extended to a position 63 ± 1 nucleotides upstream from the translation start. Weaker signals 66 ± 1 , 59 ± 1 and 56 ± 1 bases away from the ATG were detected beside the major band.

The DNA sequence of the 5'-flanking regions has been established to about 630bp upstream from the initiation codon. The comparison of the three genes revealed a very

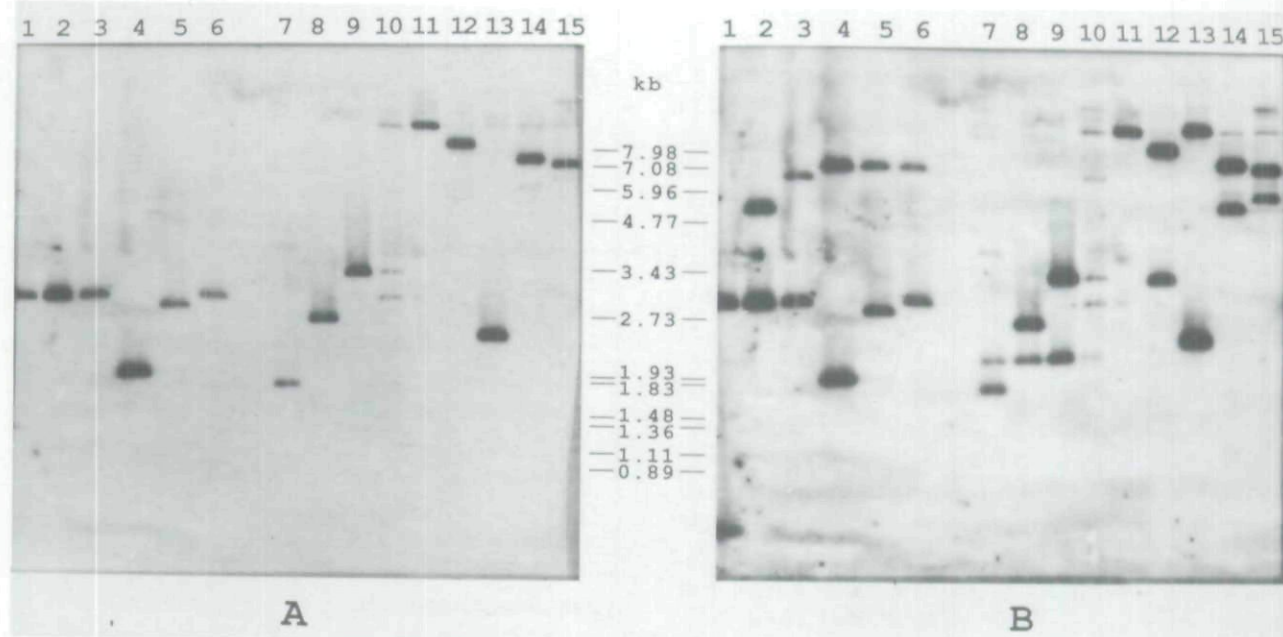


Fig. 3. Southern analysis of recombinant EMBL3 phage DNAs. Lanes 1–9 and 11–15, recombinant phage DNAs; lane 10, *A. oryzae* genomic DNA. The *Sal*I-hydrolysed DNAs were separated on a 0.8% agarose gel. Molecular weight marker: phage SPP1 DNA hydrolysed with *Eco*RI.

A. Hybridization with a nick-translated 1450 bp *Pst*I fragment of the cDNA (pGW8).

B. Hybridization with the corresponding 370 bp *Pst*I probe. In lane 6 of panel B, two fragments match exactly with a 7.1- and a 3.0 kb band of lane 10. In lanes 1–5, one of these bands is shortened, reflecting the distribution of fragments generated by the partial *Sau*3A hydrolysis during the construction of the library. Comparison with panel A revealed the order of *Sal*I fragments 5'/7.1 kb–3.0 kb/3' for this gene. The analogous evaluation of the bands in lanes 7–9 in comparison with lane 10 revealed the order 5'/2.2 kb–3.6 kb/3' for the second gene. The maximum insert size for EMBL3 is 23 kb (Frischauf *et al.*, 1983). Hence one does not expect to find recombinants with both remaining hybridizing *Sal*I fragments, which are about 20 kb and 13 kb in length. Therefore the clones in lanes 11–15 only contain parts of these. For further analysis, clones λ StW28 (lane 6), λ StW26 (lane 9) and λ StW7 (lane 14) were chosen.

high degree of homology. Only one nucleotide was changed at position 243 downstream from the *Eco*RI site in *amy3*. Typical TATA boxes are found at a distance of 32 nucleotides from the major transcription start. A CAAT motif located 125 residues upstream from the transcription start exists in all three genes. In addition, a second CAAT motif in *amy1* and *amy2* was identified just at the position of the altered nucleotide, which is about 310 bp upstream of the transcription start. *amy1* and *amy2* have the sequence CCAATT, whereas *amy3* has the sequence CCAAAT.

Analysis of the 3'-nontranslated regions of the three genes revealed striking differences. The *Bgl*II sites of the genes correspond to codons 473 to 475 of the mature protein, which is 478 residues in length. For 85 nucleotides downstream from the *Bgl*II site, the genes have exactly the same DNA sequence (Fig. 5). Beginning with the base at position 86, the three genes share no further sequence similarities. Comparison of these three genomic sequences with the appropriate region of the cDNA demonstrated that pGW8 represents the transcript of the gene *amy1*. This means that the 3' divergence of the three

structural genes is manifested in at least one transcript.

These results encouraged us to construct specific oligonucleotide probes for each gene in order to obtain information about the expression of the genes *amy2* and *amy3*. Probes were chosen which are complementary to nucleotides 76–94 downstream from the *Bgl*II sites covering the beginning of the 3' divergence. The three probes were examined in standard Southern blot analyses of the primary λ clones cut with *Sal*I to rule out possible cross-hybridization to determine the most stringent washing conditions. Each probe hybridized specifically to the expected fragment (data not shown). Poly(A)⁺ RNA isolated from starch-grown mycelia was analysed in Northern hybridization experiments (Fig. 7). As expected from the isolated cDNA, the oligonucleotide complementary to the 3'-region of *amy1* hybridized to poly(A)⁺ RNA. Surprisingly, a second signal was obtained with the specific probe derived from *amy3*. Both hybridizing mRNAs had the same length within the limits of RNA gel electrophoresis and could be correlated to the length of the cDNA. The intensity of the signals was about the same. The gene-specific probe for *amy2* gave no signal.

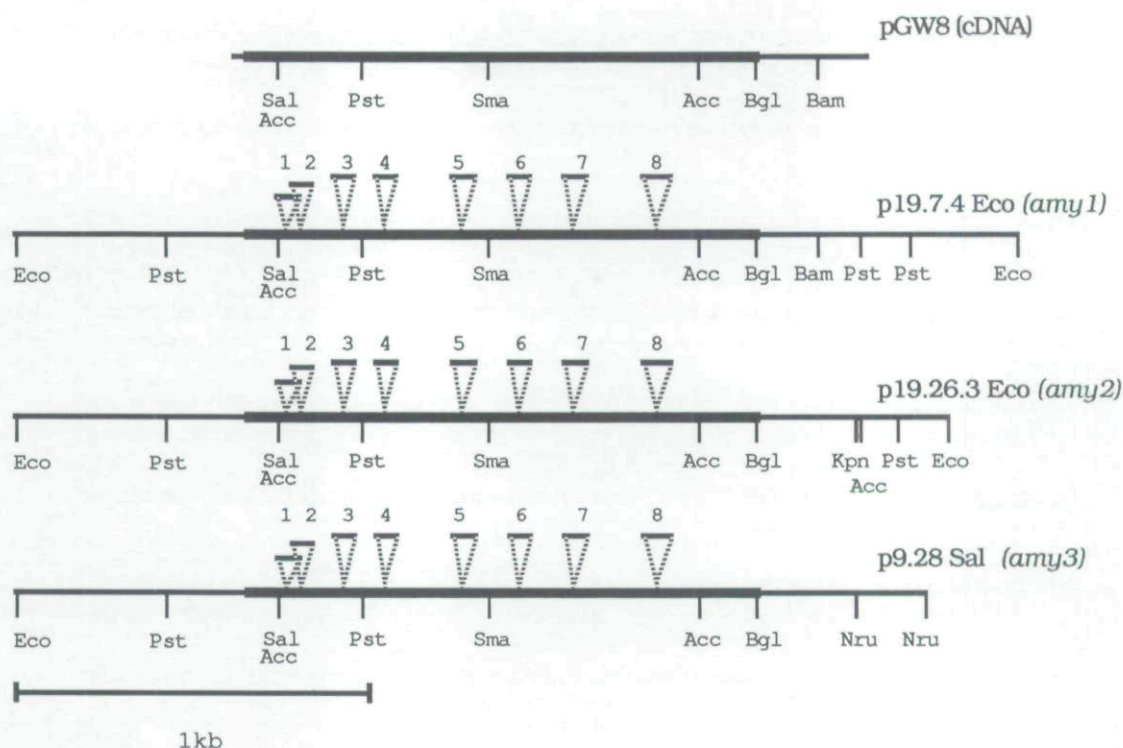


Fig. 4. Restriction maps of pGW8, p19.7.4 Eco, p19.26.3 Eco and p9.28 Sal. Schematic representation of three genomic sequences and a cDNA coding for *A. oryzae* α -amylase. The coding regions are indicated by bold lines. The eight introns and the 5'- and 3'-non-coding regions are marked by thin lines.

Discussion

In this paper we report the structural organization and the sequence analysis of the α -amylase gene family in *A. oryzae*. Although the overall gene structure in the three genes was found to be the same, we could only detect two gene-specific mRNAs in about the same amounts under the conditions investigated. We therefore analysed the 5'-sequences in order to find hints for this phenomenon. Only one nucleotide varied between the three sequenced upstream regions, but this alteration is found in *amy3*, which was shown to be transcribed. Therefore we can rule out the possibility that the upstream regions, as far as sequenced, are responsible for the lack of a detectable mRNA from *amy2*.

The presumptive promoter regions of the three *A. oryzae* α -amylase genes are identical and have a TATA box 32bp upstream from the transcription start and a CAAT motif in a distance of 125bp. This arrangement is similar to that often observed in higher eukaryotes. A second CAAT motif at -310 is only found in *amy1* and *amy2*. The minimal requirements for fungal promoters are not yet established. Only some analysed genes have the

typical core elements, i.e. TATA and CAAT boxes (Corrick *et al.*, 1987). In those cases in which they are found, the distance to the transcriptional start point seems to be more variable than in genes of higher eukaryotes. The sequence GGGCGG or its reverse complement has been determined as the core binding sites for the SP1 transcription factor in several genes of higher eukaryotes (Wingender, 1988). In a position 200bp upstream from the transcription start we found the sequence AGGGGCGGAAA and at -100 the motif TCCCCGCCCT, both of which show strong similarities to the reported consensus sequences. In a previous publication, Gwynne *et al.* (1987) found similar GC boxes in the upstream regions of the *aldA* and *alcA* genes of *Aspergillus nidulans*. We can perfectly align 12 nucleotides of our -200 motif to the region of the GC box in the *alcA* gene. Possibly these sites are involved in the initiation of transcription. A perfect palindromic structure (TTTACGTAAA) is found about 450bp upstream from the transcription start. A further feature of our sequence is a 13bp direct repeat (GGGCAACTCGCTT) separated only by 4 nucleotides. These elements of unknown function lay approximately 500bp upstream of

EcoRI 50
TTCATGGTGTGTTTATCATTTTAAATTTTATATGGCGGGTGGTGGGCAACTCGCTTCCGCGGGCAACTCGCTTACCGATTACGTTAGGGC amy1
..... amy2
..... amy3

100 150
TGATAATTACGTAAATCGTCAAGGGATGCAAGACCAAGTAGTAAACCCCGAGTCAACAGCATCCAAGCCCAAGTCTTCACGGAGA amy1
..... amy2
..... amy3

200 250
AACCCAGCGTCCACATCACGAGCGAAGGACCACCTCTAGGCATCGGACGCACCATCCAATTAGAAGCAGCAAGGAAACAGCCCAAGAA amy1
..... amy2
..... amy3

300 350
AAAGGTCGGCCCGTGGGCTTTTCTGCAACGCTGATCACGGGCAGCGATCCAACCAACACCTCCAGAGTGAAGGGCGGAAATTTAA amy1
..... amy2
..... amy3

400 450
GGGATTAAATTTCCACTCAACCAAAATCACAGTCGTCGCCCGGTATTGTCCTGCGAATGCAATTTAACTCTTCTGCGAATCGCTTGGATT amy1
..... amy2
..... amy3

500
CCCCGCCCCGCGCTAGAGCTTAAAGTATGTCCCTTGTGATGCGATGTATCACAACATATAAACTACTAGCAAGGGATGCCATGCTTGA amy1
..... amy2
..... amy3

550+ 600 Met
GGATAGCAACCGACAACATCACATCAAGCTCTCCCTTCTCTGAACAATAAACCCACAGAAGGCATTTATGATGGTCGCGTGGTGGTCTCT amy1
..... amy2
..... amy3

650 700
ATTTCTGTACGGCCTTCAGGTGCGGCGACCTGCTTTGGCTGCAACGCTGCGGACTGGCGATCGCAATCCATTATTTCTCTCACGGAT amy1
..... amy2
..... amy3

750 800
CGATTGCAAGGACGGATGGGTCGACGACTGCGACTTGTAACTACTGCGGATCAGgtgtgtgttacctactagctttcagaagaggaa amy1
..... amy2
..... amy3

850 900
taaactgacttgatatagAAATACTGTGGTGAACATGGCAGGCATCATCGACAAGttaaattgcccctttatcaaaaaaagaaggaa amy1
..... amy2
..... amy3

950 1000
aagcagaagaaaaataaaaaagaactctagtcttaaccatcacatagTTGGACTATATCCAGGGAATGGGCTTCACAGCCATCTGGA amy1
..... amy2
..... amy3

1050
TCACCCCGCTTACAGCCAGCTGCCCCAGACCACCGCATATGGAGATGCCTACCTACCTGGCTACTGGCAGCAGGATATgtaagtcgatttctt amy1
..... amy2
..... amy3

1100 1150
taaatatctacctgtcatctttacatcaatatgaactaacttgatggttttagATACTCTCTGAACGAAACTACGGCACTGCAGATGAC amy1
..... amy2
..... amy3

1200 1250
TTGAAGGCGCTCTCTTCGCCCTTCATGAGAGGGGATGTATCTTATGGTCGATGTGGTTGCTAACCATATGgttcgtggtcctttgcaac amy1
..... amy2
..... amy3

1300 1350
tgacttcgcgatgatggttcatttcagtactgacaatgagtaatatcagGGCTATGATGGAGCGGTAGCTCAGTCGATTACAGTGTGTT amy1
..... amy2
..... amy3

1400 1450
AAACCGTTCAGTTCGAAGACTACTTCCACCGTTCCTGTTTCATTCAAAATATGAAGATCAGACTCAGGTTGAGGATTGCTGGCTAGGAG amy1
..... amy2
..... amy3

1500
ATAACACTGTCTCTTGCCTGATCTCGATACCACCAAGGATGTGGTCAAGAAATGAATGGTACGACTGGTGGGATCATTTGGTATCGAACTA amy1
..... amy2
..... amy3

1550 1600
CTCCAgtaagatatttctccctcattctacaacttggtgatcgatgatacttacgaaatcagTTGACGGCCTCCGTATCGACACAGTAAA amy1
..... amy2
..... amy3

1650 1700
ACACGTCAGAGGACTTCTGGCCCGGTACAACAAAGCCGAGCGTGTACTGTATCGGCGAGGTGCTCGACGGTGATCCGGCCTACACT amy1
..... amy2
..... amy3

1750	1800	
TGTCCTACCAGAAGCTCATGGACGGCTACTGAACTATCCCATgtatggttctctccaacatgagccttcttgcaagtctcatctcctaa		amy1
.....		amy2
.....		amy3
1850	1900	
cgaacgggctaaaaccagTTACTATCCACTCCTCAACGCTTCAAGTCAACCTCCGGCAGCATGGACGACCTCTACAACATGATCAACACC		amy1
.....		amy2
.....		amy3
1950	2000	
GTCAAATCCGACTGTCCAGACTCAACACTCCTGGGCACATTGCTGAGAACACGACAAACCCACGGTTCGCTTCgttaagtcttccctttta		amy1
.....		amy2
.....		amy3
2050		
ttttccgttcccaatttccacacagaacccccacctaacaagagcaaatTTACACCAACGACATAGCCCTCGCCAAGAAGCTCGCAGCATTTC		amy1
.....		amy2
.....		amy3
2100	2150	
ATCATCTCAACGACGGAATCCCCATCATCTACGCCGGCCAAGAACGACACTACGCCGGCGGAACGACCCCGCAACCGCGAAGCAACCT		amy1
.....		amy2
.....		amy3
2200	2250	
GGCTCTCGGGCTACCCGACGACAGCGAGCTGTACAAGTTAAATGCGCTCCGCGAACGCAATCCGGAACATATGCCATTAGCAAGATACAGG		amy1
.....		amy2
.....		amy3
2300	2350	
ATTCTGTACCTACAAGgttaagcacaacctctaagcataccctaattggcctatcttcagagtatctgacacagagactaatcactggcaat		amy1
.....		amy2
.....		amy3
2400	2450	
acagAACTGGCCCATCTACAAGACGACACAACGATCGCCATGCGCAAGGGCACAGATGGGTGCGAGATCGTGACTATCTTGTCACAACAG		amy1
.....		amy2
.....		amy3
2500		
GGTGCTTCGGGTGATTCTGATACCTCTCTCTTGGTGGTGGGGTTACACAGCCGGCCAGCAATTGACGGAGGTCAATTGGCTGCACGACCG		amy1
.....		amy2
.....		amy3
2550	2600	
TGACGGTGGTTTCGGATGGAAATGTGCTGTCTTCTATGGCAGTGGGCTAOCCTAGGGTATTGTATCCGACTGAGAAGTTGGCAGGTAGCAA		amy1
.....		amy2
.....		amy3
2650	2700	
GATCTGTAGTAGCTCGTGAAGGGTGAGAGTATATGATGGTACTGCTATTCAATCTGGCATTGGACAGTGAGTTTGAGTTTGATGTACAGT		amy1
.....		amy2
.....		amy3
2750	2800	
TGGAGTCGTACTGCTGTATCCCCCTTATCTCTCGATTGTTTTTCGAACCTTAATGCCAAGCACGCTAGTCTATTATAGGAAAGGATCC		amy1
.....		amy2
.....		amy3
2850	2900	
GGATTAATGTGTTTTCATAACCGGGTACTGTATGGTACTTCTGTATTATATCACCAGAGCTCATGTATCTTACATGTATATATTATACAGA		amy1
.....		amy2
.....		amy3
2950	3000	
CACAACCTTGGTTACCCCACTATGATGTTTCTGTCAGATAATCTCTGACGATCAATCTTACCACAGGGATATGATGGCACCCCACTGGC		amy1
.....		amy2
.....		amy3
3050		
GCCTTCGCAACATCAAATTATTTGTATGATCAAGTGCATCATCTACTAATGCCAGATAGCTGAAGCTGCTAACTGCAGCACA		amy1
.....		amy2
.....		amy3
3100	3150	
CGCGCAATTCTAGGATTCCGTCAAATCTCCGGCTATTCCGCGTAGCAAAGCCCTCCAAATAAAGTGGACGCCACGAAGCATCTCAC		amy1
.....		amy2
.....		amy3
3200	3250	
CAATAATGCTGGAGGCTCTTTGTGATCATCTCTTGAAAGCCTGATCTATACCTTGACGAAATGGCCATCTTTCTATGGGATGAGTTCCA		amy1
.....		amy2
.....		amy3
3300	3350	
AAATACGCAACTACATCTAGTATACAGCGGGCTCTGTCTTCTAAAGGTTGGTCCAAAAAGGCAGCTCGGCAGAAAGCAAGGAACGGAA		amy1
.....		amy2
.....		amy3

Fig. 5. DNA sequence of *amy1*, *amy2* and *amy3*. Dots are placed in positions in which *amy2* and *amy3* have the same nucleotides as *amy1*. Small letters indicate introns; + shows the major transcription start; # marks a polyadenylation site of *amy1* as deduced from the poly(A) tail of pGW8; and the underlined sequences are described in the Discussion. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession numbers X12725 (*amy1*), X12726 (*amy2*) and X12727 (*amy3*).



Fig. 6. Determination of the transcription start by primer extension. About 25 fmol [^{32} P]-labelled oligonucleotide complementary to codons 6–10 of α -amylase transcripts was hybridized to different amounts of poly(A) $^{+}$ RNA from starch-induced mycelia and extended with reverse transcriptase. The resulting cDNAs were fractionated on a 6% sequencing gel. The cDNA was transcribed from 1600 ng (lane 1) and 800 ng poly(A) $^{+}$ RNA (lane 2). Lane 3: control (no RNA). An *EcoRI/SalI* subclone of p19.7.4 *Eco* (Fig. 4) was sequenced using the same primer. The sequencing reactions A, C, G and T served as precise markers to determine the distances of the labelled fragments in lanes 1 and 2 from the TATA box. These are shown on the right.

the transcription start. At the moment we can only speculate about the regulatory nature of the motifs described. Future experiments may elucidate possible *in vivo* functions. It would be very interesting to identify the upstream regulatory elements in the α -amylase genes of *A. oryzae* and to compare them with the upstream activating sites in yeast and the enhancers in vertebrates. Recent results indicate that common principles of transcriptional activation have been retained in the evolution of eukaryotes (Guarente, 1988).

The sequence analysis of the coding regions revealed deviations from the previously described protein sequence of Taka amylase A (Toda *et al.*, 1982) which are listed in Table 1. The data of our genomic clones confirmed the cDNA sequence presented in Fig. 1. Firstly, the deviations could be caused by different isoforms of α -amylase in different strains investigated. Beyond the genes of a wild-type *A. oryzae* described here we cloned an α -amylase gene of a production strain (to be published

elsewhere). The established DNA sequence of the latter supports the observed amino acid changes outlined in Table 1. Secondly, the published protein sequence could be incorrect. In this case the consequences for the established structure of the protein have to be examined.

Comparing our three genomic sequences we determined that *amy3*, which was shown to be transcribed, encodes a protein with variations at two positions. Codon 35 is changed from CAG to CGG (Gln to Arg) and codon 151 from TTC to CTC (Phe to Leu). These differences are not identical to those shown in Table 1.

It has been proposed that codon selection has a regulatory function in the expression of genes in *Saccharomyces cerevisiae* (Bennetzen and Hall, 1982). *A. oryzae* uses 60 codons out of the possible 61 codons for the synthesis of α -amylase. Generally, there is only a weak bias in codon preference. In those cases in which it is observed, it is quite similar to that of the glucoamylase of *Aspergillus niger* (Boel *et al.*, 1984). Using the published protein sequence and the first ATG of our cDNA sequence as the translational startpoint we found a leader peptide of 21 residues followed by 478 amino acids of the mature enzyme. A hydrophilicity plot revealed the hydrophobic character of the leader sequence (data not shown). Leader peptides cover three regions designated as n-, h- and c-regions (von Heijne, 1985). The c-region follows the '(-3, -1)'-rule for defining the cleavage site of the leader (von Heijne, 1984). We find alanine residues in these positions as expected. The positive charge of the N-terminal regions has a statistical mean value of +1.7, which is somewhat higher than the +1.0 of the α -amylase leader.

The nucleotide sequence ACCAUGG has been proposed as a consensus for the region of the translation start in higher eukaryotes (Kozak, 1986) and the sequence A(Y)AA(U)AUGUCU has been proposed for the corresponding site in yeast (Cigan and Donahue, 1987). The

Table 1. Differences between the deduced protein sequence of pGW8 and the published sequence of Taka amylase A.

aa position	pGW8	Taka amylase A
72	Thr	Asp
73	Thr	Cys
80	His	Thr
85	Gln	Thr
324	Ile	Leu
385	Trp	—
387	Ile	Tyr
388	Tyr	Ile
427	Gly	Ser
476	Ser	Ser
—	—	Asp
477	Ser	Ser

The numbering of the amino acid positions is applied to the deduced protein sequence of pGW8 shown in Fig. 1.

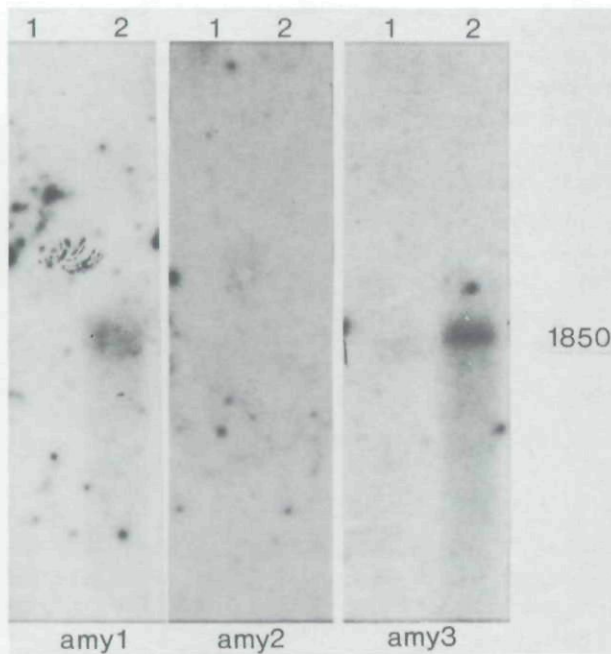


Fig. 7. Northern analysis of *A. oryzae* RNA. RNAs from starch-induced mycelia were fractionated on oligo(dT) cellulose, glyoxylated and separated on a 1.2% agarose gel. Lanes 1, poly(A)⁻ fraction; lanes 2, poly(A)⁺ fraction. Each lane contained about 8 μ g RNA. *amy1*, *amy2* and *amy3*: hybridization with [³²P]-labelled oligonucleotides complementary to 3'-untranslated regions of the corresponding genes (Fig. 5).

sequence around the translation start in the *A. oryzae* α -amylase genes shows no significant similarity to one of the two.

The comparison of the intron distribution revealed further striking homologies. Not only the positions but also the sequences of the eight introns are absolutely identical in the three amylase genes of *A. oryzae*. The features of the eight introns are summarized in Table 2. The small size of the introns is typical for filamentous fungi. All introns obey the GT/AG rule for 5' and 3' splice sites (Mount, 1982), but the previously proposed consensus sequences (5' GTRNGT-----YAG 3') for fungal intron boundaries (Corrick *et al.*, 1987) are not found exactly in all introns. The

amylase genes of *A. oryzae* have a higher variability in the intron junctions. An internal conserved sequence located near the 3'-boundary of introns has been suggested as a putative splicing signal. In yeast, it is a strictly conserved sequence (TACTAAC), and in filamentous fungi, the weakened consensus RCTRAC has been proposed (Corrick *et al.*, 1987). In each intron we found a derivative of the latter sequence in a distance of 8 to 18 bp from the 3' splicing sites.

The strong homology of the three genes only extends to a point 70 bp downstream from the stop codon. With gene-specific oligonucleotide probes derived from the 3'-untranslated sequences we showed that two of the three α -amylase genes in *A. oryzae* are transcribed under the conditions investigated. We can imagine some reasons for the lack of a signal from *amy2*. Firstly, if the end of the message of *amy2* terminates before the region from which the oligonucleotide was derived we can never prove the existence of this specific mRNA. Secondly, this gene could represent a silent copy in general or a copy which is only transcribed in specific environmental or cellular conditions. Thirdly, the effect could be caused by reduced mRNA stability.

To date, it is uncertain whether the signal AATAAA commonly found in polyadenylated mRNAs of higher eukaryotes functions in the same manner in fungi. Comparison of the 3'-terminal sequences from *amy1* with the end of the cDNA revealed that no canonical AATAAA signal precedes the polyadenylation site. Nevertheless, it is found in the truncated form ATAA about 90 nucleotides upstream from the site of polyadenylation in pGW8. On the other hand, in *amy3*, which has also been shown to be transcribed, there exists the sequence AATAAA about 140 nucleotides downstream from the stop codon. Its relationship to the actual polyadenylation site has yet to be demonstrated. Previously reported results indicate that further signals should be involved in the generation of mature 3'-termini of polyadenylated RNAs. McLauchlan *et al.* (1985) found an 8 bp consensus sequence (YGTGTTY) and Renan (1987) identified a distinct 12 bp consensus

Table 2. Analysis of 5'-, 3'-intron boundaries and putative internal splice signals.

Intron	Length	5'-Boundary	Internal signal	Distance to 3'-boundary	3'-Boundary
1	55bp	G T G T G T	A C T G A C	9bp	T A G
2	85bp	G T A A A T	C C T A A C	10bp	T A G
3	69bp	G T A A G T	A C T A A C	13bp	T A G
4	68bp	G T T C G T	A C T G A C	15bp	C A G
5	58bp	G T A A G A	A C T T A C	8bp	C A G
6	66bp	G T A T G G	C C T A A C	18bp	C A G
			G C T A A A	6bp	
7	65bp	G T A A G T	C C T A A C	10bp	A A G
8	79bp	G T A A G C	A C T A A T	14bp	C A G
			A C T G G C	7bp	

Introns 6 and 8 show two alternative internal consensus sequences.

sequence (TTRNNNTTTTTT) located 5–20 bp downstream from the poly(A) sites in many genes of higher eukaryotes. Our sequence data did not match the expected requirements concerning their positions and the length of the message. Further analysis of the 3'-regions of the three genes revealed an interesting element consisting of an overlapping direct repeat 15 nucleotides in length located immediately before the 3' divergence.

In the investigations described in this publication we found no hints that *A. oryzae* possesses several α -amylase isoenzymes in a strict sense. Nevertheless, the fungal genome contains a gene family encoding this enzyme. The reasons for the presence of multiple genes for α -amylase in *A. oryzae* are unknown. On the one hand it could be a quantitative effect to increase the synthesis of enzyme. On the other hand, the question remains as to why *amy2* seems not to be transcribed.

The overall DNA sequences of the three genes analysed in this paper show a very high degree of homology, indicating that the genes have a common ancestor. We suggest two possible reasons for the observed homology. Firstly, there could have been a strong evolutionary tendency not to change the acquired organization of the genes. Secondly, the increase of the gene copy number could be the result of very recent gene duplications. In a previous analysis of the human amylase genes, Gumucio *et al.* (1988) showed a recent reorganization of this gene family. The authors assumed less than 1 million years for the evolution of the three salivary type genes and about 7 million years for the two pancreatic type genes. The suggestion that recent events led, likewise, to the present gene arrangement in *A. oryzae* is supported by the finding that the intron-exon organization is found to be identical in the three gene copies, even at nucleotide level.

Interestingly, the phenomenon of a gene family for α -amylase is a common feature of higher eukaryotes, e.g. mammals and plants (Gumucio, *et al.*, 1985; 1988; Baulcombe *et al.*, 1987). One might even speculate about an evolutionary reason to explain the presence of several copies of α -amylase genes in the genomes of all eukaryotes.

To our knowledge there have only been two analyses reported for fungal gene families. The three genes for the elongation factor EF-1 α in *Mucor racemosus* are transcribed, but transcription has different levels (Sundstrom *et al.*, 1987). The EF-1 α genes also show only two amino acid changes but they have different 5'-sequences as well as different 3'-untranslated regions. A further feature concerned the distribution of introns. An intron near the 3'-end exists in all three genes. A second intron near the 5'-end is only present in two of them. Two of the genes encoding alcohol dehydrogenase in *A. nidulans* have been isolated and sequenced (McKnight *et al.*, 1985; Gwynne *et al.*, 1987). Comparison reveals that the gene products are real

isoenzymes, showing 85 differences out of 348 amino acids. In these examples of fungal gene families, the observed differences are more marked than those of the α -amylase genes in *A. oryzae*.

In future we will investigate whether it is possible to find conditions in which *amy2* is transcribed, perhaps in a pattern of differential expression. Our further intention is to identify the *cis*-acting elements of the upstream regions in the α -amylase genes in *A. oryzae* and to compare them with previously analysed examples from yeast and higher eukaryotes.

Experimental procedures

Strains and culture conditions

The source of fungal RNA and genomic DNA throughout this investigation was *A. oryzae* DSM63303 (German type strain collection, Braunschweig, FRG). The mycelia were grown as described earlier (Errat *et al.*, 1984). *Escherichia coli* strains DH2 and DH5 α (Hanahan, 1985) were the host strains for the plasmids pUC9 and pUC19 (Yanisch-Perron *et al.*, 1985). *E. coli* strains NM538 and NM539, respectively, were host strains for the propagation of the phage lambda EMBL3 and its derivatives (Frischauf *et al.*, 1983). Bacterial cells and phage were grown in LM media under the appropriate selective conditions (Maniatis *et al.*, 1982).

Northern and Southern analyses

DNAs and glyoxal-denatured RNAs were separated on agarose gels, transferred to Hybond N membranes (Amersham) and hybridized with [³²P]-labelled probes, according to the manufacturer's suggestions.

Isolation of poly(A)⁺ RNA and cDNA cloning

Fungal total RNA was prepared from starch-induced mycelia (Errat *et al.*, 1984) with the guanidinium isothiocyanate procedure as described by Chirgwin *et al.* (1979). Poly(A)⁺ RNA was isolated by chromatography on oligo(dT) cellulose (Taylor, 1979). Double-stranded cDNA was synthesized according to the procedure of Gubler and Hoffman (1982) using M-MLV-reverse transcriptase (BRL). 40 ng of ds cDNA was incubated with 1 U of Mung bean nuclease (Pharmacia) for 30 min at 37°C, followed by fractionation on a Bio-Gel A-50 column (Huynh *et al.*, 1985). Fractions containing RNA of about 2.7 to 0.9 kb were pooled. The dC-tailing of the cDNAs was carried out using established methods (Maniatis *et al.*, 1982). 10 ng of dG-tailed pUC9 (Pharmacia) was annealed with approximately 1 ng of the dC-tailed cDNAs and transformed into *E. coli* DH5 α (Hanahan, 1985). The colonies were transferred to a Hybond N membrane and screened with a [³²P]-labelled oligonucleotide probe derived from the published *A. oryzae* α -amylase protein sequence (Toda *et al.*, 1982). This probe, a 63-mer complementary to codons 375–395 was synthesized using the phosphoramidite method (Beaucage *et al.*, 1981). The sequence of the probe was selected using the criteria of Agarwal *et al.* (1981).

Isolation of nuclear DNA and cloning of the genomic DNAs

The preparation of fungal nuclei was carried out essentially as described by Gealt *et al.* (1976). DNA was isolated from the nuclear pellet fraction using standard procedures (Maniatis *et al.*, 1982). The DNA was partially digested with *Sau*3A and 15- to 20-kbp fragments were isolated from a NaCl gradient (Kaiser and Murray, 1985). DNA from bacteriophage EMBL3 was digested with *Bam*HI and *Eco*RI, followed by the removal of the small linker fragments (Frischauf *et al.*, 1983). Then about 80 ng of *Aspergillus* DNA fragments was ligated to 500 ng of vector arms and the products were packaged *in vitro* using the method of Hohn (1979). An aliquot of the packaging extract was plated on an *E. coli* NM539 lawn. Plaque hybridization was performed on a Hybond N membrane with a nick-translated cDNA probe (see Results). Phage DNAs were isolated from plaque-purified positive clones according to Helms *et al.* (1985).

DNA sequencing and computer analysis

DNAs in double-stranded forms were sequenced using the dideoxy method with the published modifications (Sanger *et al.*, 1977; Biggin *et al.*, 1983; Tabor and Richardson, 1987). The data were processed on an Apple Macintosh SE with a software package from Texco (West Lebanon, N.H.).

Primer extension

The hybridization and extension reactions with starch-induced poly(A)⁺ RNA were carried out according to Williams and Mason (1985). The synthetic oligonucleotide primer (5'-AATGGATTGC-GATCGCCA-3') complementary to codons 6-10 of the cDNA sequence of pGW8 (Results) was extended with M-MLV-reverse transcriptase (BRL).

Acknowledgements

This work was supported by the Bundesminister für Forschung und Technologie, Förderungskennzeichen PTB 038634. A.L. is a recipient of a Dechema grant. We thank Professor H.G. Gassen for his suggestions and his continuous interest in our work. The authors wish to thank Jürgen Stollwerk for the synthesis of the oligonucleotides used in many experiments. We also thank Rainer Fuchs and Martin Weber for assistance in computing. Dr Sabine Bertram and Marion Weber are thanked for critical reading of the manuscript.

References

Agarwal, K.L., Brunstedt, J., and Noyes, B.E. (1981) A general method for detection and characterization of an mRNA using an oligonucleotide probe. *J Biol Chem* **256**: 1023-1028.

Baulcombe, D.C., Huttly, A.K., Martensen, R.A., Barker, R.F., and Jarvis, M.G. (1987) A novel wheat α -amylase gene (α -amy3). *Mol Gen Genet* **209**: 33-40.

Beaucage, C.M., and Caruthers, M.H. (1981) Deoxynucleoside phosphoramidites. A new class of key intermediates for deoxy-polynucleotide synthesis. *Tetrahedron Lett* **22**: 1859-1862.

Bennetzen, J.L., and Hall, B.D. (1982) The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase I. *J Biol Chem* **257**: 3026-3031.

Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc Natl Acad Sci USA* **80**: 3963-3965.

Boel, E., Hjort, I., Svensson, B., Norris, F., Norris, K.E., and Fiil, N.P. (1984) Glucoamylases G1 and G2 from *Aspergillus niger* are synthesized from two different but closely related mRNAs. *EMBO J* **3**: 1097-1102.

Cheetham, P.S.J. (1985) The applications of enzymes in industry. In *Enzyme Biotechnology*. Wiseman, A. (ed.). Chichester: Ellis Horwood Ltd, pp. 329, 343-345.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**: 5294-5299.

Cigan, A.M., and Donahue, T.F. (1987) Sequence and structural features associated with translational initiator regions in yeast — a review. *Gene* **59**: 1-18.

Corrick, C.M., Twomey, A.P., and Hynes, M.J. (1987) The nucleotide sequence of the *amdS* gene of *Aspergillus nidulans*. *Gene* **53**: 63-71.

Emi, M., Horii, A., Tomita, N., Nishide, T., Ogawa, M., Mori, T., and Matsubara, K. (1988) Overlapping two genes in human DNA: a salivary amylase gene overlaps with a gamma-actin pseudogene that carries an integrated human endogenous retroviral DNA. *Gene* **62**: 229-235.

Erratt, J.A., Douglas, P.E., Moranelli, F., and Seligy, V.L. (1984) The induction of α -amylase by starch in *Aspergillus oryzae*: evidence for controlled mRNA expression. *Can J Biochem Cell Biol* **62**: 678-690.

Frischauf, A.-M., Lehrach, H., Poustka, A., and Murray, N. (1983) Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* **170**: 827-842.

Gealt, M.A., Sheir-Neiss, G., and Morris, N.R. (1976) The isolation of nuclei from the filamentous fungus *Aspergillus nidulans*. *J Gen Microbiol* **94**: 204-210.

Guarente, L. (1988) UASs and enhancers: common mechanisms of transcriptional activation in yeast and mammals. *Cell* **52**: 303-305.

Gubler, U., and Hoffman, B.J. (1982) A simple and very efficient method for generating cDNA libraries. *Gene* **25**: 263-269.

Gumucio, D.L., Wiebauer, K., Dranginis, A., Samuelson, L.C., Treisman, L., Caldwell, R.M., Antonucci, T.K., and Meisler, M.H. (1985) Evolution of the amylase multigene family. *J Biol Chem* **260**: 13483-13489.

Gumucio, D.L., Wiebauer, K., Caldwell, R.M., Samuelson, L.C., and Meisler, M.H. (1988) Concerted evolution of human amylase genes. *Mol Cell Biol* **8**: 1197-1205.

Gwynne, D.I., Buxton, F.P., Sibley, S., Davies, R.W., Lockington, R.A., Scazzocchio, C., and Sealy-Lewis, H.M. (1987) Comparison of the *cis*-acting control regions of two coordinately controlled genes involved in ethanol utilization in *Aspergillus nidulans*. *Gene* **51**: 205-216.

Hanahan, D. (1985) Techniques for transformation of *E. coli*. In *DNA Cloning*. Vol. 1. Glover, D.M. (ed.). Oxford: IRL Press, pp. 109-135.

Helms, C., Graham, M., Dutchik, J., and Olson, M. (1985) A new method for purifying λ DNA from phage lysates. *DNA* **4**: 39-49.

Hesseltine, C.W. (1983) Microbiology of oriental fermented foods. *Annu Rev Microbiol* **37**: 575-601.

Hohn, B. (1979) *In vitro* packaging of λ and cosmid DNA. *Meth Enzymol* **68**: 299-309.

- Horii, A., Emi, M., Tomita, N., Nishide, T., Ogawa, M., Mori, T., and Matsubara, K. (1987) Primary structure of human pancreatic α -amylase gene: its comparison with human salivary α -amylase gene. *Gene* **60**: 57–64.
- Huynh, T.V., Young, R.A., and Davis, R.W. (1985) Construction and screening cDNA libraries in λ gt10 and λ gt11. In *DNA Cloning*. Vol. 1. Glover, D.M. (ed.). Oxford: IRL Press, pp. 64–66.
- Kaiser, K., and Murray, N.E. (1985) The use of phage lambda replacement vectors in the construction of representative genomic libraries. In *DNA Cloning*. Vol. 1. Glover, D.M. (ed.). Oxford: IRL Press, pp. 41–42.
- Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eucaryotic ribosomes. *Cell* **44**: 283–292.
- Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Matsuura, Y., Masami, K., Harada, W., and Kakudo, M. (1984) Structure and possible catalytic residues of Taka-Amylase A. *J Biochem* **95**: 697–702.
- McKnight, G.L., Kato, H., Upshall, A., Parker, M.D., Saari, G., and O'Hara, P.J.O. (1985) Identification and molecular analysis of a third *Aspergillus nidulans* alcohol dehydrogenase gene. *EMBO J* **4**: 2093–2099.
- McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B. (1985) The consensus YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucl Acids Res* **13**: 1347–1368.
- Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucl Acids Res* **10**: 459–472.
- Renan, M.J. (1987) A conserved 12-bp element downstream from mRNA polyadenylation sites. *Gene* **60**: 245–254.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463–5467.
- Sundstrom, P., Lira, L.M., Choi, D., Linz, J.E., and Sypherd, P.S. (1987) Sequence analysis of the EF-1 α gene family of *Mucor racemosus*. *Nucl Acids Res* **15**: 9997–10006.
- Tabor, S., and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci USA* **84**: 4767–4771.
- Taylor, J.M. (1979) The isolation of eukaryotic messenger RNA. *Annu Rev Biochem* **48**: 681–717.
- Toda, H., Kondo, K., and Narita, K. (1982) The complete amino acid sequence of Taka amylase A. *Proc Jpn Acad* **58**: 208–212.
- Williams, J.G., and Mason, P.J. (1985) Hybridisation in the analysis of RNA. In *Nucleic Acid Hybridisation*. Hames, B.D., and Higgins, S.J. (eds). Oxford: IRL Press, pp. 152–159.
- Wingender, E. (1988) Compilation of transcription regulating proteins. *Nucl Acids Res* **16**: 1879–1902.
- von Heijne, G. (1984) Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. *EMBO J* **3**: 2315–2318.
- von Heijne, G. (1985) Signal sequences. The limits of variation. *J Mol Biol* **184**: 93–105.
- Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103–119.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.