

GEN 02983

***Aspergillus oryzae* has two nearly identical Taka-amylase genes, each containing eight introns ***

(Recombinant DNA; cDNA synthesis; glycosidase; M13 phage vectors; polyadenylation; signal peptide; starch)

Maria Jesus Gines^{ab**}, Michael J. Dove^a and Verner L. Seligy^{ab}

^a The Molecular Genetics Section, Division of Biological Sciences, National Research Council of Canada, Ottawa (Canada) and

^b Department of Biology, Carleton University, Ottawa, Ontario (Canada)

Received by J.L. Slightom: 10 October 1988

Accepted: 10 December 1988

SUMMARY

cDNA and genomic DNA for two nearly identical genes, *AmyI* and *AmyII*, coding for the enzyme Taka-amylase A (TA-A) of the fungus *Aspergillus oryzae* have been cloned and characterized. These genes are apparently unlinked, differing by only 3 nucleotides (nt) out of the 2720 nt that span the coding regions. The 617-nt 5'-flanking regions differ only at nt -372 (T or A) from the putative ATG start codon and contain four sets of short, inverted repeats (IR) upstream from the putative TATAAA box at nt -100 and the transcription start point at nt -69. The coding regions consist of 499 codons disrupted by eight intervening sequences. The putative proenzymes differ by only two amino acids (aa) and consist of the 478-aa extracellular enzyme plus a 21-aa hydrophobic leader sequence. Except for the replacement site changes in codons 35 (Arg ↔ Gln) and 151 (Phe ↔ Leu), the identity of the two genes continues downstream for 58 nt past the TGA stop codon before diverging. Exon 9 codes for 94 of the 98 aa of the domain B of mature TA-A. Little conservation of TA-A exons was found when these exons were aligned with those of human amylase. The genes are flanked by at least 6 to 10 kb of unrelated chromosomal nucleotide sequence. The *Amy* genes are co-expressed, since mRNA (cDNA) specific to the 3'-UTR of both genes was recovered from mycelia grown on starch, a known inducer of TA-A biosynthesis. The 3'-UTRs of cDNAs related to *AmyI* are shorter (128 nt and 145 nt) than those of *AmyII* (179 nt and 297 nt). The *AmyI* specific 3'-UTR is characterized by the absence of IR sequences and the presence of a putative 'AATAAA' polyadenylation signal.

Correspondence to: Dr. V.L. Seligy, Molecular Genetics Section, Division of Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario, K1A 0R6 (Canada) Tel. (613)990-0825 Fax (613)952-0974.

* Dedicated to the memory of Dr. Kozo Narita, Institute for Protein Research, Osaka University, Osaka (Japan), and of Patricia Douglas, Division of Biological Sciences, National Research Council of Canada, Ottawa (Canada).

** Present address: Plant Research Centre, Experimental Farm, Agriculture Canada, K.W. Neatby Building, Ottawa, Ontario, K1C 0C6 (Canada) Tel. (613)995-3700.

Abbreviations: aa, amino acid(s); AMV, avian myeloblastosis

virus; *AmyI* and *AmyII*, genes coding for the TA-AI and TA-AII, respectively; bp, base pair(s); cDNA, DNA complementary to RNA; dNTP, any of the deoxyribonucleoside triphosphates; ds, double strand(ed); DR, direct repeat; DTT, dithiothreitol; EtdBr, ethidium bromide; ICS, internal conserved sequence; IR, inverted repeat; IVS, intervening sequence or intron; kb, 1000 bp; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame; pfu, plaque-forming unit(s); PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; SDS, sodium dodecyl sulfate; ss, single strand(ed); SSC, 0.15 M NaCl/0.015 M Na₃·citrate pH 7.6; TAE, 40 mM Tris·acetate pH 8/2 mM EDTA; TA-A, Taka-amylase A; UTR, untranslated region of 5' or 3' end of messenger RNA.

INTRODUCTION

The α -amylases (EC 3.2.1.1, α -1,4-glucan 4-glucanohydrolase) represent a major family of glycosidases for which a substantial diversity of amino acid sequences, structure and gene sequences are known (MacKay et al., 1985; Rogers, 1985; Nakajima et al., 1986). Much of our knowledge of α -amylase structure and function comes from the extensive physical and chemical studies on the commercially useful TA-A of the fungus *A. oryzae*. The inclusion of this enzyme in a patented fermentation process by Jokichi Takamine in 1894 marked the start of modern enzyme technology (Barbesgaard, 1977; Bennett, 1985). The mature, secreted TA-A is a glycoprotein comprised of 478 aa residues (Toda et al., 1982), folded into two globular domains, A (aa 1–380) and B (aa 381–478) (Matsuura et al., 1984). The aa residues involved in substrate binding and in glucosidic cleavage of starchy substrates are located in domain A. The function of the glycosylation at Asn-197 (Toda et al., 1982; Hase et al., 1982) and aa of domain B, which show very little evolutionary conservation (MacKay et al., 1985; Ito et al., 1987), remain unclear. Further testing of the TA-A model would require manipulation of the gene if it were available. We describe here the unexpected finding of two nearly identical genes, *AmyI* and *AmyII* which code for TA-A proenzymes that differ by only 2 aa. Excluding the signal peptide, *TA-AI* and *TA-AII* differ by 8 aa in comparison to the published amino acid sequence of mature, secreted TA-A (Toda et al., 1982). The consensus changes to the original amino acid sequence at aa positions 72–73 (Asp-Cys $\rightarrow$ Thr-Thr), 384–385 (Asn-Pro $\rightarrow$ Asn-Trp-Pro) and 476 (Asp is deleted) actually rule out earlier speculations on genetic variation of TA-A inferred from amino acid sequence work (Narita, 1975; Toda, 1982).

MATERIALS AND METHODS

(a) The cDNA and genomic DNA cloning

Poly(A)⁺ RNA was prepared from total RNA of *A. oryzae* (NRC401013) grown as a suspension in 2% starch medium for 6 h or 10 h as described (Erratt et al., 1984). Abundance of TA-A mRNA

was measured by in vitro translation and immunoprecipitation with anti-TA-A antibody as described earlier (Erratt et al., 1984). Blunt-ended ds cDNA synthesis and construction of cDNA libraries using *HincII*-cut M13mp9 vector were performed as described (Rutledge et al., 1988). Using *Escherichia coli* JM103 (Messing, 1983) as the host, about 3×10^5 clones (pfu) were recovered from a cDNA preparation involving 10 μ g of methyl mercury treated, 10-h poly(A)⁺ RNA, 2 μ g oligo (dT)_{12–18} and 50 units AMV reverse transcriptase (Life Sciences Inc.) in a final volume of 50 μ l.

A genomic DNA library consisting of 2.2×10^5 hybrid λ Charon 4A phage particles was constructed using 10 μ g of partially *EcoRI*-digested *A. oryzae* nuclear DNA (Seligy et al., 1980) and 6 μ g of purified, preligated, *EcoRI*-cut λ phage arms prepared as described (Slightom et al., 1980). Plaque pure clones were amplified using megaplates as described (Slightom et al., 1980). Ten randomly selected recombinant λ phage clones yielded an average insert size of 15.6 kb.

(b) Gene isolation and characterization

M13-cDNA phage was first probed by hybridization (Gillen et al., 1987) with the G + C-rich, 20-mer [³²P] oligo 5'CC(Z)CGCAT(Z)GG(Z)AC(Z)GG-(Z)AC3' (Z = A or C or G) which was back-translated from aa residues 451 to 457 of the TA-A aa sequence (Toda et al., 1982). The cDNA and genomic libraries were probed with the 500-bp cDNA insert of clone M71 labelled with [³²P]dAMP by primer extension as described by Burke (1984). Cloned DNA and total *A. oryzae* DNA, prepared by the method of Yelton et al. (1984), were subjected to restriction endonuclease site mapping using the manufacturer's protocols for the use of the various nucleases. Subcloning of DNA was carried out in pUC18 and pUC19 plasmids and M13 phage vectors (Yanisch-Perron et al., 1985). Nucleotide sequencing was carried out by the dideoxy chain-termination method described by Messing (1983).

(c) Transcript mapping

Mapping of the 5' end of the TA-A transcripts was according to Burke (1984) using 10⁶ cpm (10 ng)

of the 338-nt, *Pst*I-*Sal*I [^{32} P]DNA ss fragment synthesized by *Pol*k off a subclone of the *Amy*II gene called M13mp18.18Q. Five units of S1 nuclease (Pharmacia) were used to delete unbound portions of this DNA after it was annealed to 5 μ g of poly(A)⁺ RNA isolated from mycelia grown on starch for either 6 h or 10 h (see section a above). Size of the protected DNA fragment and its nucleotide sequence were estimated as described before (Yang et al., 1986) using as a sequencing primer the 50-nt *Pvu*I-*Sal*I DNA fragment also derived from subclone M13mp18.18Q. The 3' end(s) of the TA-A transcripts were deduced from the sequencing of several TA-A-specific cDNA clones. Computer analysis of nt sequences was carried out as previously described (MacKay et al., 1985; 1986).

RESULTS AND DISCUSSION

(a) Isolation of TA-A-specific cDNA

Previous results indicated that TA-A expression was regulated at the level of its mRNA, since TA-A mRNA, measured by in vitro translation, was 10–20 times higher when starch was used as the carbon source instead of glucose (Erratt et al., 1984). For construction of the cDNA banks we used the poly(A)⁺ RNA fraction isolated from a 10-h starch culture because this RNA yielded the maximum amount of translatable TA-A mRNA. Two TA-A-specific clones (M71 and M91) were initially isolated from 3×10^4 plaques by hybridization with the synthetic oligo probe described in MATERIALS AND

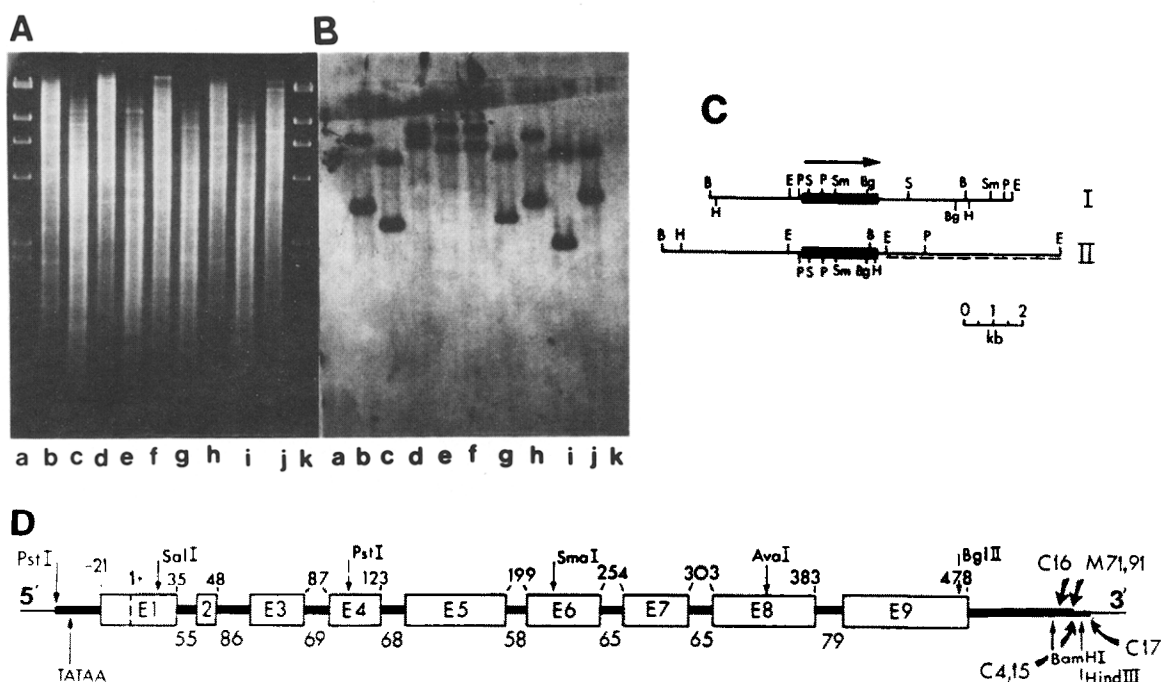


Fig. 1. Mapping of TA-A-coding chromosomal genes and flanking regions. Digests of total *A. oryzae* DNA (10 μ g per track), electrophoresed in 0.8% agarose with EtBr (panel A) were transferred to a nylon membrane and probed in $6 \times$ SSC at 68°C for 16 h with either 10^6 cpm/ml ^{32} P-labelled M71 cDNA (panel B) or the equivalent of 5.8-kb *Eco*RI fragment (hybridization result shown only by the dashed line in panel C) from the genomic clone λ Ch4A.Ao.3Cl. Hybridization of a similar DNA transfer with ^{32} P-labelled 3.35-kb *Eco*RI DNA (the *Amy*II gene) gave identical results to that obtained with probe M71. Lanes a and k are λ DNA digested with *Hind*III. Lanes b–j correspond to restriction enzyme digests: E, E + B, B, B + H, H, E + H, E, E + P, P, respectively, as defined in panel C. Panel C summarizes the mapping data showing restriction endonuclease sites and direction of transcription (arrow), where B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III; P, *Pst*I; S, *Sal*I; Sm, *Sma*I. Maps I and II correspond to loci for *Amy*I and *Amy*II defined by the 7.6-kb and 3.35-kb *Eco*RI genomic fragments, respectively. Panel D shows the refined map of the 2.8-kb *Amy* genes derived from nucleotide sequence data presented in Figs. 2 and 3. Numbers above the gene, –21, +1, 35, 48, 87, 123, 199, 254, 303, 383 and 478 refer to positions of codons defining the Met (start codon) of the secretion signal, start of the mature, secreted TA-A, the 5'-junctions of the eight introns and the last codon before the TGA stop codon, respectively. The exons are indicated by E1 to E9. Open boxes define the coding regions and numbers below the gene indicate sizes in bp of each intron. TATAA at the 5' end indicates position of putative TATA box. Curved arrows at 3' end of the gene indicate positions of the 3' end of each of the 3' UTRs determined from cDNA clones corresponding to *Amy*I (C16, M71, M91) and *Amy*II (C4, 15, 17), respectively (see Fig. 3B for details).

METHODS, section b. Use of this probe ensured the recovery of even incomplete TA-A cDNA whose synthesis was initiated by oligo(dT) priming at the poly(A) tail of TA-A mRNA. The inserts of these clones were sequenced (see section f below). The 500-bp cDNA of M71 contains the codons for aa 360–478, the TGA stop codon and 154 nt of 3' UTR terminated with an A₉ tail. M91 is similar to M71, containing a 384-bp cDNA (codons 403–478) and exactly the same 3' UTR. M71 was used further as a hybridization probe to obtain several other cDNA and genomic clones.

(b) Cloning and mapping TA-A-coding genes

A. oryzae chromosomal DNA was first digested with the four restriction endonucleases described in Fig. 1A and probed with ³²P-labelled ss cDNA from M71 (Fig. 1B) to identify the TA-A genomic sequence. These data showed that each enzyme or pair of enzymes produces a pair of TA-A genomic fragments. Since M71 cDNA is not cut by any of the four enzymes, these results signify the possibility of two TA-A-specific chromosomal loci. The copy number and direction of transcription of each locus was further determined by using similar DNA blots probed separately with the 5' and the 3' halves of the M71 cDNA cut with *Bgl*II, located 5 codons upstream from the TGA stop codon, and with a mixed synthetic oligo probe spanning the 5' end of the TA-A-coding region (aa residues 17–25). In each case only two fragments were identified. These data are further summarized in Fig. 1C and include additional information from Fig. 2.

Screening of the genomic DNA library by plaque hybridization (see MATERIALS AND METHODS, section b) with the M71 insert, yielded five clones containing the 7.6-kb *Eco*RI fragment and six clones containing the 3.35-kb *Eco*RI fragment. Clones λCh4A.Ao.3C1 and λCh.4A.Ao.11G1, which contain the 3.35-kb and 7.60-kb *Eco*RI fragments respectively, were further compared by restriction endonuclease analysis and nucleotide sequencing to produce the map shown in Fig. 1D. In addition to the phage λ DNA arms, λCh4A.Ao.3C1 and λCh.4A.Ao.11G1 consist of three *Eco*RI fragments, adding up to 14.7 kb and 11.4 kb, respectively. Since the M71 3' UTR nucleotide sequence was the same as that of the 7.60-kb *Eco*RI fragment derived from

λCh4A.Ao.3C1 (Fig. 3), the corresponding gene was called *AmyI* and the *AmyII* gene was assigned to the 3.35-kb *Eco*RI fragment. Both *Eco*RI fragments have identical restriction sites 5' to the *Bgl*II site. Nucleotide sequences 3' of the *Bgl*II-*Hind*III fragment, including the flanking 5.8-kb *Eco*RI fragment (see dashed lines in Fig. 1C) are unique to the *AmyII* locus. Since only about 2.65 kb of the 3.35-kb and the 7.6-kb *Eco*RI DNA fragments are common in sequence, the two genes must be separated by at least 6 to 10 kb of unrelated sequences, depending on their relative orientations.

The *AmyII* gene was subcloned and sequenced first because of the readily available 3'-flanking *Bam*HI and *Hind*III sites. The use of subclones of *AmyI* and *AmyII* in M13mp18 and M13mp19 and pUC19, and additional synthetic oligo primers enabled us to obtain approx. 3.0 kb of nucleotide sequence (in both directions) for each gene, spanning the 5' *Eco*RI site to 350 nt, 3' of the TGA stop codon. The TA-A-coding region and the eight introns shown in Figs. 1D and 2, were positioned by matching the predicted coding sequence from the cDNA and genomic nucleotide sequences with that of the published TA-A amino acid sequence (Toda et al., 1982).

(c) *AmyI* and *AmyII* genes have common 5'-flanking sequences

As shown in Fig. 3A, the 617-nt sequence 5' of the putative ATG start codon of both genes is characterized by the presence of four sets of putative start codon, ATG. Only one difference in sequence was found between *AmyI* and *AmyII*, T vs. A at nt -372, respectively. For 5'-end transcript mapping, the ³²P-labelled, 338-nt long, ss *Pst*I-*Sal*I DNA from the *AmyII* was first annealed to either 6-h or 10-h RNA from starch cultures. After treatment with S1 nuclease, a single, 204-nt fragment was recovered in each case. This signifies that initiation of *Amy* gene transcriptions is uniform during growth on starch. The 5' end of this fragment corresponds in position to the base A of the sequence GGAGG, 24 nt downstream from the TATAAA sequence beginning at nt -100 and 69 nt upstream from the putative translation start codon (see section d below). This 100-nt region also has three G/C-AA-C/G sequences which are characteristic of a number of

-21
 ATG ATG GTC GCG TGG TGG TCT CTA TTT CTG TAC GGC CTT CAG GTC GCG GCA CCT GCT TTG GCT GCA ACG CCT GCG GAC
 met met VAL ALA TRP TRP SER LEU PHE LEU TYR GLY LEU GLN VAL ALA ALA PRO ALA LEU ALA -ALA THR PRO ALA ASP
 PwI 10 15 20 30 40 45 50 55
 TGG CGA TCG CAA TCC ATT TAT TTC CTT CTC ACG GAT CGA TTT GCA AGG ACG GAT GGG TCG ACG ACT GCG ACT TGT AAT ACT GCG GAT CCG
 TRP ARG SER GLN SER ILE TYR PHE LEU LEU THR ASP ARG PHE ALA ARG THR ASP GLY SER THR THR ALA THR CYS ASN THR ALA ASP CAG
 GTGTGTTGTACTACTAGCTTTAGAAAGAGGAATGTAAGTAACTGACTTGATATAG AAA TAC TGT GGT GGA ACA TGG CAG GGC ATC ATC GAC AAG GTAAATTGCCCC
 IVS-1 LYS TYR CYS GLY GLY THR TRP GLN GLY ILE ILE ASP LYS
 TTTATCAAAAAAAGAGGAAAAGCAGAAAAAATAAAATAAAAGAACTCTAGTCTAACCATCACATAG TTG GAC TAT ATC CAG GGA ATG GGC TTC ACA GCC
 LEU ASP TYR ILE GLN GLY MET GLY PHE THR ALA
 60 65 70 75 80 85
 ATC TGG ATC ACC CCC GTT ACA GCC CAG CTG CCC CAG ACC ACC GCA TAT GGA GAT GCC TAC CAT GGC TAC TGG CAG CAG GAT AT GTAAGTCGA
 ILE TRP ILE THR PRO VAL THR ALA GLN LEU PRO GLN THR THR ALA TYR GLY ASP ALA TYR HIS GLY TYR TRP GLN GLN ASP IL
 TTTCTTTAAATATCTACTGTCATCTTTTACATCAATGAACTAACTTGATGGTTTGTAG A TAC TCT CTG AAC GAA AAC TAC GGC ACT GCA GAT GAC TTG AAG
 E TYR SER LEU ASN GLU ASN TYR GLY THR ALA ASP ASP LEU LYS
 90 95 100
 GCG CTC TCT TCG GCC CTT CAT GAG AGG GGG ATG TAT CTT ATG GTC GAT GTG GTT GCT AAC CAT ATG GTTCGTGGTCTTTGCAACTGACTTCGGGAT
 ALA LEU SER SER ALA LEU HIS GLU ARG GLY MET TYR LEU MET VAL ASP VAL VAL ALA ASN HIS MET
 105 110 115 120
 ATGTTTCATTTCAGTACTGACAAATGAGTAATATCAG GGC TAT GAT GGA GCG GGT AGC TCA GTC GAT TAC AGT GTG TTT AAA CCG TTC AGT TCC CAA GAC
 GLY TYR ASP GLY ALA GLY SER SER VAL ASP TYR SER VAL PHE LYS PRO PHE SER SER GLN ASP
 125 130 135 140
 TAC TTC CAC CCG TTC TGT TTT ATT CAA AAC TAT GAA GAT CAG ACT CAG GTT GAG GAT TGC TGG CTA GGA GAT AAC ACT GTC TCC TTG CCT
 TYR PHE HIS PRO PHE CYS PHE ILE GLN ASN TYR GLU ASP GLN THR GLN VAL GLU ASP CYS TRP LEU GLY ASP ASN THR VAL SER LEU PRO
 145 150 155 160 165 170
 GAT CTC GAT ACC ACC AAG GAT GTG GTC AAG AAT GAA TGG TAC GAC TGG GTG GGA TCA TTG GTA TCG AAC TAC TCC A GTAAGATATTTCTCCCTC
 ASP LEU ASP THR THR LYS ASP VAL VAL LYS ASN GLU TRP TYR ASP TRP VAL GLY SER LEU VAL SER ASN TYR SER I AvaI
 175 180 185 190 195
 ATTCTCAACTTGGCTGATCGATGATACTTACGAAATCAG TT GAC GGC CTC CGT ATC GAC ACA GTA AAA CAC GTC CAG AAG GAC TTC TGG CCC GGG TAC
 LE ASP GLY LEU ARG ILE ASP THR VAL LYS HIS VAL GLN LYS ASP PHE TRP PRO GLY TYR
 200 205 210 215 220
 AAC AAA GCC GCA GGC GTG TAC TGT ATC GGC GAG GTG CTC GAC GGT GAT CCG GCC TAC ACT TGT CCC TAC CAG AAC GTC ATG GAC GGC GTA
 ASN LYS ALA ALA GLY VAL TYR CYS ILE GLY GLU VAL LEU ASP GLY ASP PRO ALA TYR THR CYS PRO TYR GLN ASN VAL MET ASP GLY VAL
 225 230 235 240 245
 CTG AAC TAT CCC AT GTATGTTCTCTCAACCATGAGCGTCTTCTGCAAGTCTCATCTCTCAACGAAACGGCTAAACCAG T TAC TAT CCA CTC CTC AAC GCC TTC
 LEU ASN TYR PRO IL E TYR TYR PRO LEU LEU ASN ALA PHE
 250 255 260
 AAG TCA ACC TCC GGC AGC ATG GAC GAC CTC TAC AAC ATG ATC AAC ACC GTC AAA TCC GAC TGT CCA GAC TCA ACA CTC CTG GGC ACA TTC
 LYS SER THR SER GLY SER MET ASP ASP LEU TYR ASN MET ILE ASN THR VAL LYS SER ASP CYS PRO SER THR LEU GLY THR PHE
 265 270 275 280 285 290
 GTC GAG AAC CAC GAC AAC CCA CGG TTC GCT TC GTAAGTCTTCCCTTTATTTTCGGTTCCTCCACACAGAACCCACCTTAACAAGAGCAAAG T TAC ACC
 VAL GLU ASN HIS ASP ASN PRO ARG PHE ALA SE R TYR THR
 295 300 305 310 315 320 325 330 335
 AAC GAC ATA GCC CTC GCC AAG AAC GTC GCA GCA TTC ATC ATC CTC AAC GAC GGA ATC CCC ATC ATC TAC GCC GGC CAA GAA CAG CAC TAC
 ASN ASP ILE ALA LEU ALA LYS ASN VAL ALA ALA PHE ILE ILE LEU ASN ASP GLY ILE PRO ILE ILE TYR ALA GLY GLN GLU GLN HIS TYR
 340 345 350 355 360 365
 GCC GGC GGA AAC GAC CCC GCG AAC CCG GAA GCA ACC TGG CTC TCG GGC TAC CCG ACC GAC AGC GAG CTG TAC AAG TTA ATT GCC TCC GCG
 ALA GLY GLY 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 ALA
 370 375 380
 AAC GCA ATC CGG AAC TAT GCC ATT AGC AAA GAT ACA GGA TTC GTG ACC TAC AAG GTAAGCACAACTCTAAGCATACCTTAATGGCTATCTTCAGAGTATC
 ASN ALA ILE ARG ASN TYR ALA ILE SER LYS ASP THR GLY PHE VAL THR TYR LYS
 385 390 395 400 405
 TGACACAAGAGACTAATCACTGGCAATACAG AAC TGG CCC ATC TAC AAA GAC GAC ACA ACG ATC GCC ATG CCG AAG GGC ACA GAT GGG TCG CAG ATC
 ASN TRP PRO ILE TYR LYS ASP ASP THR THR ILE ALA MET ARG LYS GLY THR ASP GLY SER GLN ILE
 410 415 420 425 430 435
 GTG ACT ATC TTG TCC AAC AAG GGT GCT TCG GGT GAT TCG TAT ACT CTC TCC TTG AGT GGT GCG GGT TAC ACA GCC GGC CAG CAA TTG ACG
 VAL THR ILE LEU SER ASN LYS GLY ALA SER GLY ASP SER TYR THR LEU SER LEU SER GLY ALA GLY TYR THR ALA GLY GLN GLN LEU THR
 440 445 450 455 460 465
 GAG GTC ATT GGC TGC ACG ACC GTG ACG GTT GGT TCG GAT GGA AAT GTG CCT GTT CCT ATG GCA GGT GGG CTA CCT AGG GTA TTG TAT CCG
 GLU VAL ILE GLY CYS THR THR VAL THR VAL GLY SER ASP GLY ASN VAL PRO VAL PRO MET ALA GLY GLY LEU PRO ARG VAL LEU TYR PRO
 470 475
 ACT GAG AAG TTG GCA GGT AGC AAG ATC TGT AGT AGC TCG TGA
 THR GLU LYS LEU ALA GLY SER LYS ILE CYS SER SER SER ***

Fig. 2. Nucleotide sequence of *AmyI* and *AmyII* genes and derived TA-A amino acid sequence. The 2045-nt sequence extends from the putative start codon (-21 codons upstream from the first codon corresponding to alanine; the N terminus of the mature TA-A, reported by Toda et al., 1982, is shown by the arrow) to the stop codon TGA. E1 to E9 indicate positions of the nine exons. IVS-1 to IVS-8 refer to the respective introns. Vertical bars at positions 35 and 151 show differences in nucleotide sequence found between *AmyI* and *AmyII* genes, and the derived amino acid sequence of TA-AI (Arg-35, Leu-151) and TA-AII (Gln-35, Phe-151).

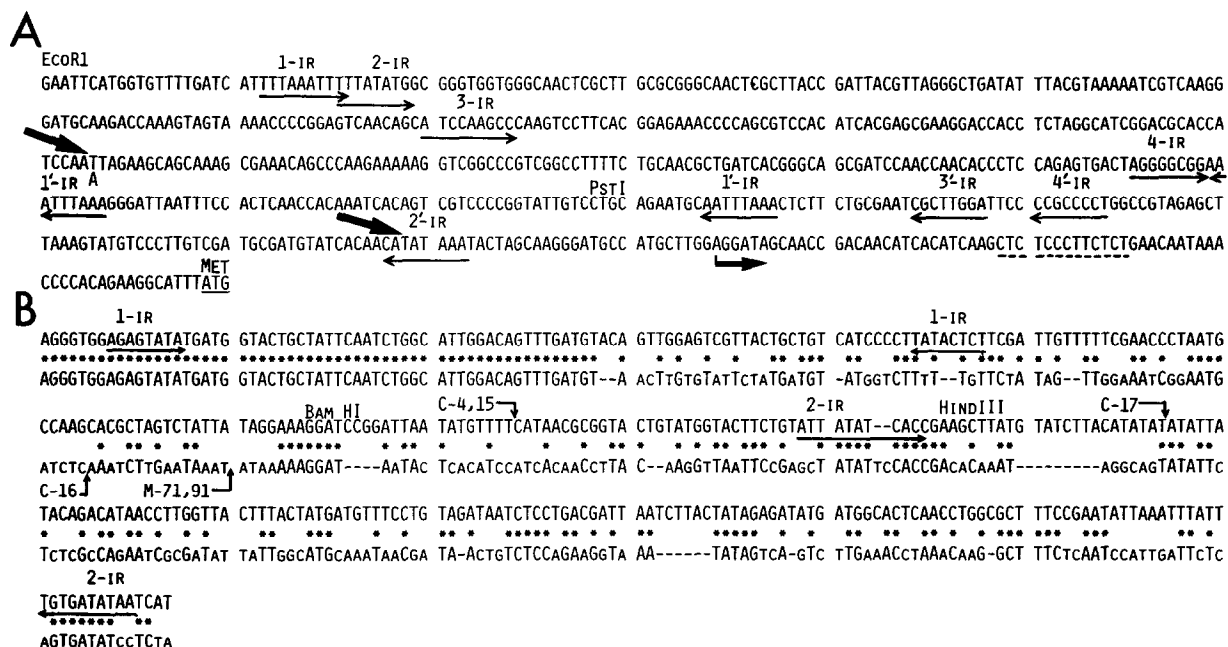


Fig. 3. Sequence of noncoding 5' and 3' regions of *AmyI* and *AmyII*. (A) 5' UTR and flanking upstream nucleotide sequence. Heavy arrows indicate positions of the only nt difference found between *AmyI* (T) and *AmyII* (A) at nt -372, the putative TATA box starting at nt -100 and the putative transcription start point at nt -69, relative to the first in-frame methionine (Met) start codon at +1. The four pairs of indirect repeat sequences (IR-1,1'; IR-2,2'; IR-3,3' and IR-4,4') are indicated by thin horizontal arrows. (B) Nucleotide sequence of the 3' UTR and downstream flanking regions of *AmyII* (upper line) and *AmyI* (lower line) and their related cDNA clones C-4, C-15, C-17 and C-16, M-71, M-91, respectively. The bent arrows after each cDNA clone indicate the last nt before putative poly(A) addition to the respective mRNA. Asterisks indicate regions of nt identity and dashed lines indicate gaps required to align nt sequences. Sequences in lower-case letters in *AmyI* without asterisks emphasize no identity with *AmyII* sequences. The two pairs of IR sequences (IR-1,1'; IR-2,2') found only in *AmyII* are shown by straight horizontal arrows.

highly transcribed genes of lower eukaryotes (Nevins, 1983). The third one, at nt -44, is just upstream from a pyrimidine track CTCTCCCTT-CTCT which has a sequence similar to that found in close association with presumed positions of transcription initiation of the genes for TrpC of *Aspergillus nidulans* and glucoamylase of *Aspergillus awamori* (Mullaney et al., 1986).

(d) Coding region of *AmyI* and *AmyII* genes

The 2045-nt sequence spanning the coding region is shown in Fig. 2. The first ORF, beginning with two methionines is 63 nt (21 codons) upstream from the N-terminal alanine (Ala-1) codon of the mature TA-A (Toda et al., 1982) and 617 nt downstream from the *EcoRI* site. Two other putative start codons are found at 140 nt and 145 nt upstream from Ala-1 but not in the same reading frame.

The coding region consists of 499 codons disrupted by eight introns and terminated at the TGA

stop codon. Except for the two replacement site substitutions in codons 35 (*AmyI* has CGG/Arg and *AmyII* has CAG/Gln) and 151 (*AmyI* has CTC/Leu and *AmyII* has TTC/Phe), the nucleotide sequences of the two genes are identical. The change in codon 35 is 1 nt before the 5' end of the first intron (IVS-1). Aside from the 21-aa leader sequence, there are only a few differences between the nucleotide-sequence-derived amino acid sequence and the determined amino acid sequence of TA-A (Table I and Fig. 2). The properties of either Arg or Gln at position 35 would be different than that of Glu, believed to be involved in amylose binding (Matsuura et al., 1984). However, Phe-151 of TA-AII does agree with the earlier work of Toda and colleagues (1982). Correction of the sequence of aa residues 72-73, Asp-Cys → Thr-Thr has been reported recently by Toda (cited in Onishi et al., 1986, pp. 234-235). The four differences located between aa residues 384 and 476 were found in TA-AI and TA-AII and amino acid sequences deduced from eight independent

cDNA clones. These all included the changes at residues 386–387 and 426. Insertion of Trp between codons Asn-384 and Pro-385 (the original published positions) and deletion of Asp-476 maintains the total number of residues at 478. The 2 aa differences

TABLE 1

Relationship of exons to functional and conserved aa residues in TA-A

Residue number ^a	Exon	Conserved regions ^b
Changes to original published amino acid sequence		
Glu-35 to Arg-35 (TA-AI) or Gln-35 (TA-AII)	E1	
Asp-Cys-73 to Thr-Thr-73	E3	A
Thr-80 to His-80	E3	A
Thr-85 to Gln-85	E3	A
Phe-151 to Leu-151 (TA-AI)	E4	B
Asn-Pro-385 to Asn-Trp-Pro-386 (inserted)	E8	
Tyr-Ile-387 to Ile-Tyr-388	E8	
Ser-426 to Gly-427	E9	
Asp-476 (deleted)	E9	
Catalytic residues		
Glu-230	E6	C
Asp-297	E7	D
Disulfide,-S-S- bonds		
30–38	E1, E2	A (only in E2)
150–164	E5	B
227 (unpaired)	E6	C
240–283	E6, E7	C, D
439–474 (corrected position is 440–475)	E9	
Carbohydrate attachment		
Asn-197	E5	B
Substrate binding sites		
Glu-35 change to Arg-35 or Gln-35	E1	
Tyr-74, Tyr-79, Trp-83	E3	A
His-122	E4	A
Tyr-155, Glu-156, Asp-168	E5	B
Asp-206, Lys-209, His-210	E6	B
Val-231, Leu-232, Asp-233	E6	C
His-296	E7	D
Asp-340, Arg-344	E7	

^a Functional aa described by Matsuura et al. (1984).

^b The four conserved regions A to D predicted by α -amylase aa sequence comparisons (MacKay et al., 1985). Regions A to D span aa residues 56–124, 154–210, 228–234 and 286–300, located in exons E3–E4, E5–E6, E6 and E7, respectively.

between TA-AI and TA-AII cannot account for the earlier reported TA-A heterogeneity (Narita, 1975; Toda et al., 1982; Hase et al., 1982; Ohnishi et al., 1986). The C-terminal TA-AI and TA-AII sequence differs from that reported by Toda et al. (1982) but agrees with the earlier version reported by Narita (1975) as Lys-Ile-Cys-Ser-Ser-Ser.

(e) Introns

The most striking feature of the two genes is the presence of the eight introns which splits the coding region into nine exons, indicated as E1 to E9 in Figs. 1D and 2. The sequence of the 3' UTR from the cDNAs and the 5' UTR mapping established that there are no other introns present. All 5' exon-intron and 3' intron-exon junctions of these genes conform to the consensus sequence GTPuAGT and T/CAG described earlier (Mount, 1982; Guthrie et al., 1986). The GT (positions 1 and 2) and G (position 5) of the 5'-junction consensus sequence are invariant in all except the G-5 of IVS-2. The third position also shows high conservation and to a lesser extent the fourth and sixth positions. The trinucleotide, 3'-junction sequence is highly conserved, except in IVS-7 which is AAG.

A sequence within each intron resembling the PuCTPuAC, ICS of filamentous ascomycetes (Guthrie et al., 1986) is also present, but not the TACTAAC sequence required for *S. cerevisiae* mRNA precursor splicing (Langford et al., 1984). Only introns 1, 3 and 4 contain ACTPuACTT which is similar to caGCTGACTg found in *Trichoderma reesei* (Shoemaker et al., 1983; Van Arsdell et al., 1987) and *A. niger* (Boel et al., 1984) glycosidases. The introns IVS-4 and IVS-8 also contain a second ICS. These sequences are identical in IVS-4 and differ by 1 nt in the two found in IVS-8. The relative distances of these sequences from the 3'-junction are 15 nt and 14 nt and 45 nt and 27 nt, respectively.

Four introns interrupt codons: Ile-87 (IVS-3), Ile-200 (IVS-5), Ile-254 (IVS-6) and Ser-303 (IVS-7). The remaining introns are located between codons. Translations through the eight introns would not be possible because of the presence of either in-frame stop codons (IVS-1, IVS-2, IVS-3 and IVS-7) or frame-shift sequences (IVS-4, IVS-5, IVS-6 and IVS-8). A summary of exon distribution in relation to conserved, known functional regions of

TA-A is given in Table I. Two of the larger conserved amino acid sequence blocks (A and B) are disrupted by introns. Alignment of exons of the *Amy* genes with those reported for human amylase (Nishide et al., 1986) and plant (Baulcombe et al., 1987) was done by establishing amino acid sequence identity (MacKay et al., 1985; Nakajima et al., 1986) before locating the position of known introns. Homology between TA-A and plant was very low and no predictions were possible. The analysis of TA-A and human amylase was more encouraging but revealed nothing as precise as that reported for the triose phosphate isomerase gene family (Gilbert et al., 1986). For example, four introns, IVS-1 (or IVS-2), IVS-3, IVS-6 and IVS-7 of the *Amy* genes and IVS-2, IVS-3, IVS-5 and IVS-7 of the human amylase gene could be placed within 7, 9, 15 and 21 codons from each other, respectively. It is interesting that only three codons separated IVS-8 of the *Amy* genes and IVS-9 of the human even though very little amino acid sequence identity occurs in the flanking exons. Further, exon 9 defines 94 of the 98 aa proposed to be in domain B of the TA-A structure (Matsuura et al., 1984).

(f) Evidence for co-transcription of *AmyI* and *AmyII*

A total of 15 additional cDNA clones containing partial *AmyI* or *AmyII* sequences, in reverse orientation (i.e., the cDNA insert of the M13 viral DNA template is anti-sense) were obtained using the *AmyII* (3.35-kb *EcoRI* fragment) as a hybridization probe. The terminations of the 3' UTR of six of these cDNA clones are shown in Fig. 3B along with the corresponding *Amy* gene sequences. Substantial differences in nucleotide sequence of the two genes are found after the first 58-nt downstream from the stop codon. The cDNA clones M71, M91 and C-16 have identical sequences, differing from each other only in location of the predicted terminal oligo(A) tail. C-16 is 17 nt shorter than M-71 and has a 128-nt 3' UTR with an 8-mer oligo(A) tail. All three cDNAs match with the 3' UTR sequence of *AmyI*. The terminal oligo(A) track of C-16 is located 8 nt upstream from an AATAAA sequence which serves as a polyadenylation signal in higher animal mRNA (Nevins, 1983). However, the terminal oligo(A) track of M71 and M91 ends only 3 nt downstream from this puta-

tive signal rather than the usual 11 to 19 nt for most gene transcripts (Nevins, 1983). This sequence is not present in either *AmyII* or its related cDNAs, C-4, C-15 and C-17 (Fig. 3B). The 3' UTRs of C-4 and C-15 clones end 179 nt from the TGA stop codon at A₁₅ and A₉ tracks, respectively. C-17 contains a 240-nt 3' UTR, 63 nt more than C-4, terminating in an A₁₃ tail 22 nt beyond the *HindIII* site. These data further confirm that two genes with different 3'-flanking sequences occur in this fungus. The data also indicate that these genes are co-transcribed in mycelia cultured for 10 h in starch medium and that at least two types of transcripts with different sites for polyadenylation are produced.

Other sequences believed to play a role in transcription termination and/or polyadenylation in some genes of yeast and filamentous fungi (Zaret and Sherman, 1982) are found in *AmyI* and *II* and their respective cDNAs. The *AmyII* and corresponding cDNA clones, C-4, C-15 and C-17 contain TAGT, TATG and TATGTTTT, similar to yeast sequences. The yeast TAGT sequence is also found in the 3' UTR of the *AmyI* gene and cDNA clones C-16, M71 and M91 but not CAA/TTG/C which is proposed to play a role in polyadenylation in *Aspergillus* sp. (Mullaney et al., 1986).

(g) Conclusions

This is the first characterization of the TA-A amino acid sequence based on the nucleotide sequence determined from two, very closely related (99.6%) genes and their mRNAs (cDNA). Comparison of the predicted amino acid sequence with that of the published TA-A sequence revealed only a few minor differences. A typical, hydrophobic signal peptide accounts for the 21 aa difference in their sizes (499 vs. 478). The remaining differences involve changes to three single and two paired aa and a single insertion and a single deletion of an aa. These changes are located outside the highly conserved regions (MacKay et al., 1985; Nakajima et al., 1986) and do not involve any specific functions. We found that all of these aa changes increased the aa identity (from 51% to 53%) calculated between the α -amylase (*ALP1*) of *Saccharomyces fibuliger* (Ito et al., 1987) and the TA-A amino acid sequence of Toda et al. (1982). The absence of Glu at position 35, Asp at positions 72 and 426 and

a Cys at position 73 suggests that they are not necessary for function.

The finding of two nearly identical genes which appear to be unlinked by at least 6 to 10 kb of unrelated flanking sequence and are both transcriptionally active, at least during maximum TA-A mRNA production (Erratt et al., 1984), cannot readily explain the earlier reports of TA-A heterogeneity. The high nucleotide sequence identity of *AmyI* and *AmyII* suggests that these genes have diverged very little from each other in comparison to human or plant amylase genes (Nishide et al., 1986; Baulcombe et al., 1987). The 617-nt or more of upstream sequence common to these two genes may be important in the regulation of their transcription. The co-ordinate expression of these genes could explain why this organism is a high producer of TA-A, however, we do not know if other strains of *A. oryzae* have the same number of *Amy* gene copies. Glucoamylase production from *A. niger* apparently involves only one gene (Boel et al., 1984) which is regulated by glucose (Innis et al., 1985). The α -amylase (*ALP1*) of *Sa. fibuligera* is also expressed by a single gene (Ito et al., 1987). Our preliminary experiments indicate that the *Amy* genes are repressed by glucose in the manner predicted in earlier work (Erratt et al., 1984). What is not clear is why more than one type of mRNA is produced with slightly different sizes of 3' UTR and what role the introns might play in gene expression. The 3' UTR length of the *AmyI* and *AmyII* cDNAs are two to six times larger than those of vertebrate and invertebrate amylases (Nishide et al., 1986; Boer and Hickey, 1986). However, they are not unusual in size compared to those reported for other fungal glycosidase genes (Shoemaker et al., 1983; Van Arsdel et al., 1987; Boel et al., 1984; Innis et al., 1985).

The finding of eight small introns (range 55 to 86 nt) in fungal ascomycete genes is the highest number recorded so far. However, the ligninase isozyme H8 from the basidiomyceteous fungus *Phanerochaete chrysosporium* has recently been reported to have eight introns 49 to 69 nt in size (Smith et al., 1988). The number of introns in fungi and yeasts are usually 2 to 5 (Gilbert et al., 1986). It is interesting that the recently sequenced 494 aa residue amylase (*ALP1*) of the yeast *Sa. fibuligera* (Ito et al., 1987) does not contain any intervening sequences yet it is 53% identical with the amino acid sequence pre-

dicted from the *Amy* gene sequences. The *ALP1* gene can be expressed directly in *Saccharomyces cerevisiae* which has not been found to be the case for other glycosidase genes of fungi (Innis et al., 1985; Van Arsdel et al., 1987), including the *Amy* genes (V.L.S. and M.J.D., unpublished).

With the exception of fruit fly (Boer et al., 1986) and yeast (Ito et al., 1987), the presence of introns in eukaryotic amylase genes appears to be the rule. The number of introns within the coding region of the *A. oryzae* α -amylase genes is one less than that found for the human salivary gland amylase gene which contains ten introns but only nine divide the coding region into ten exons (Nishide et al., 1986). The relationship of the exon distribution within the coding region to functional domains of the amylases encoded by these two genes was analyzed but the results were not too striking. In this case the conservation of exons in relation to amino acid sequence identity blocks is much less than that found for triose phosphate isomerase (Gilbert et al., 1986). However, it is interesting that exon 9 of the *Amy* genes of *A. oryzae* accurately defines the structural domain B of TA-A (Matsuura et al., 1984) and that the equivalent exon in human salivary amylase, based on our aa sequence alignments, would be exon 10 (also exon 11).

The presence of two virtually identical genes (exons and introns) in *A. oryzae* tends to rule out the possibility that introns function to prevent recombination between the *Amy* genes as has been suggested for related cellulase genes (Van Arsdel et al., 1987). The role of the various introns and the 5' and 3' UTR sequences in coordinated expression of the two *Amy* genes, including mRNA stability and control in translation of the TA-A will be studied further.

ACKNOWLEDGEMENTS

We thank J. Michniewicz for synthesis of oligos, L. Bramall for help with computer analysis and Dr. M. Yaguchi for review of this manuscript. M.J.G. was supported by a Government of Canada Award to Foreign Nationals and by NSERC funds to V.L.S. This is NRCC No. 29872.

REFERENCES

- Barbesgaard, P.: Industrial enzymes produced by members of the genus *Aspergillus*. In Smith, J.E. and Pateman, J.A. (Eds.), *Genetics and Physiology of Aspergillus*. The British Mycological Society Symposium Series, Vol. 1, Academic Press, London, 1977, pp. 382–404.
- Baulcombe, D., Huttly, A.K., Martienssen, R.A., Barker, R.F. and Jarvis, M.G.: A novel wheat α -amylase gene (α -Amy3). *Mol. Gen. Genet.* 209 (1987) 33–40.
- Bennet, J.W.: Molds, manufacturing and molecular genetics. In Timberlake, W.E. (Ed.), *Molecular Genetics of Filamentous Fungi*. UCLA Symp. Mol. Cell. Biol. Vol. 34, Liss, New-York, 1985, pp. 345–366.
- Boel, E., Hansen, M.T., Hjort, I., Hoegh, I. and Fiil, N.P.: Two different types of intervening sequences in the glucoamylase gene from *Aspergillus niger*. *EMBO J.* 3 (1984) 1581–1585.
- Boer, H. and Hickey, D.A.: The alpha-amylase gene in *Drosophila melanogaster*: nucleotide sequence, gene structure and expression motifs. *Nucleic Acids Res.* 14 (1986) 8399–8411.
- Burke, J.F.: High-sensitivity S1 mapping with single-stranded [^{32} P]DNA probes synthesized from bacteriophage M13mp templates. *Gene* 30 (1984) 63–65.
- Erratt, J.A., Douglas, P., Morannelli, F. and Seligy, V.L.: The induction of alpha-amylase by starch in *Aspergillus oryzae*: evidence for controlled mRNA expression. *Can. J. Biochem. Cell Biol.* 62 (1984) 678–690.
- Gilbert, W., Marchionni, M. and McKnight, G.: On the antiquity of introns. *Cell* 46 (1986) 151–154.
- Gillen, M.F., Banville, D., Rutledge, R.G., Narang, S., Seligy, V.L., Whitfield, J. and MacManus, P.: The synthesis of a complete complementary DNA for the oncodevelopmental-binding protein, oncomodulin. *J. Biol. Chem.* 262 (1987) 5308–5312.
- Guthrie, C., Riedel, N., Parker, R., Swerdlow, H. and Paterson, B.: Genetics of snRNAs and RNA processing in yeast. In Hicks, J. (Ed.), *Yeast Cell Biology*, UCLA Symp. Mol. Cell. Biol. Vol. 33, Liss, New-York, 1986, pp. 301–321.
- Hase, S., Fujimura, K., Kanoh, M. and Ikenaka, T.: Studies on the heterogeneity of Taka-amylase A: isolation of an amylase having one *N*-acetylglucosamine residue as the sugar chain. *J. Biochem.* 92 (1982) 265–270.
- Innis, M.A., Holland, M.J., McCabe, P.C., Cole, G.E., Wittman, V.P., Tal, R., Watt, W.K., Gelfand, D.H., Holland, J.P. and Meade, J.H.: Expression, glycosylation and secretion of an *Aspergillus* glucoamylase by *Saccharomyces cerevisiae*. *Science* 228 (1985) 21–26.
- Ito, T., Yamashita, I. and Fukui, S.: Nucleotide sequence of the α -amylase gene (*ALP1*) in the yeast *Saccharomycopsis fibuligera*. *FEBS Lett.* 219 (1987) 339–342.
- Langford, C.J., Klinz, F.J., Donath, C. and Gallwitz, D.: Point mutations identify the conserved, intron-contained TAC-TAAC box as an essential splicing signal sequence in yeast. *Cell* 36 (1984) 645–653.
- MacKay, R.M., Baird, S., Dove, M.J., Erratt, J.A., Gines, M.J., Moranelli, F., Nasim, A., Willick, G.E., Yaguchi, M. and Seligy, V.L.: Glucanase gene diversity in prokaryotic and eukaryotic organisms. *Biosystems* 18 (1985) 279–292.
- MacKay, R.M., Lo, A., Willick, G.E., Zucker, M., Baird, S., Dove, M., Moranelli, F. and Seligy, V.L.: Structure of a *Bacillus subtilis* endo- β -1,4-glucanase. *Nucleic Acids Res.* 14 (1986) 9159–9170.
- Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M.: Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.* 95 (1984) 697–702.
- Messing, J.: New M13 vectors for cloning. *Methods Enzymol.* 101 (1983) 20–78.
- Mount, S.M.: A catalogue of splice junction sequences. *Nucleic Acids Res.* 10 (1982) 459–472.
- Mullaney, E.J., Hamer, J.E., Roberti, K.A., Yelton, M. and Timberlake, W.E.: Primary structure of the *trpC* gene from *Aspergillus nidulans*. *Mol. Gen. Genet.* 199 (1986) 37–45.
- Nakajima, R., Imanaka, T. and Aiba, S.: Comparison of amino acid sequences of eleven different α -amylases. *Appl. Microbiol. Biotechnol.* 23 (1986) 355–360.
- Narita, K.: Amino acid sequence of the C-terminal sixty-six residues of Taka-amylase A. *Proc. Japan. Acad.* 51 (1975) 285–290.
- Nevins, J.R.: The pathway of eukaryotic mRNA formation. *Annu. Rev. Biochem.* 52 (1983) 441–466.
- Nishide, T., Nakamura, Y., Emi, M., Yamamoto, P., Ogawa, M., Mori, T. and Matsubara, K.: Primary structure of human salivary α -amylase gene. *Gene* 41 (1986) 299–304.
- Ohnishi, M., Sakano, Y. and Taniguchi, H. (Eds.): *Amylases*. Japanese Scientific Societies Press, Tokyo, 1986 (Japanese original translated by M. Yaguchi).
- Rogers, J.C.: Conserved amino acid sequence domains in alpha-amylases from plants, animals and bacteria. *Biochem. Biophys. Res. Commun.* 128 (1985) 470–476.
- Rutledge, R.G., Neelin, J.M. and Seligy, V.L.: Isolation and expression of cDNA clones coding for two sequence variants of *Xenopus laevis* histone H5. *Gene* 70 (1988) 117–126.
- Seligy, V.L., Thomas, D.Y. and Miki, B.L.A.: *Saccharomyces cerevisiae*, Sep or 2 μ m: intracellular distribution, stability and nucleosomal-like packaging. *Nucleic Acids Res.* 8 (1980) 3371–3381.
- Shoemaker, S., Schweikart, V., Ladner, M., Gelfand, D., Kwok, S., Myambo, K. and Innis, M.: Molecular cloning of exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27. *Bio/Technology* 1 (1983) 691–696.
- Slightom, J.L., Blechl, A.E. and Smithies, O.: Human fetal γ - and A-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21 (1980) 627–638.
- Smith, T.L., Schalch, H., Gaskell, J., Covert, S. and Cullen, D.: Nucleotide sequence of a ligninase gene from *Phanerochaete chrysosporium*. *Nucleic Acids Res.* 16 (1988) 1219.
- Taylor, G.R.: Southern transfer of native DNA using nylon membranes. *Anal. Biochem.* 148 (1985) 524–526.
- Toda, H., Kondo, K. and Narita, K.: The complete amino acid sequence of Taka-amylase A. *Proc. Japan. Acad.* 58 (Ser.B) (1982) 208–212.
- Van Arsdell, J.N., Kwok, S., Scheickart, V.L., Ladner, M.B., Gelfand, D.H. and Innis, M.A.: Cloning, characterization, and expression in *Saccharomyces cerevisiae* of endoglucanase I *Trichoderma reesei*. *Bio/Technology* 5 (1987) 60–64.

- Yanisch-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33 (1985) 103–119.
- Yang, R.C.A., Dove, M., Seligy, V.L., Lemieux, C., Turmel, M. and Narang, S.A.: Complete nucleotide sequence and mRNA mapping of the large subunit gene of ribulose-1-5-bisphosphate carboxylase/oxygenase (Rubisco) from *Chlamydomonas moewusii*. *Gene* 50 (1986) 259–270.
- Yelton, M.M., Hamer, J.E. and Timberlake, W.E.: Transformation of *Aspergillus nidulans* by using a *trpC* plasmid. *Proc. Natl. Acad. Sci. USA* 81 (1984) 1470–1474.
- Zaret, K.S. and Sherman, F.: DNA sequence required for efficient transcription termination in yeast. *Cell* 28 (1982) 563–573.