

GENE 03232

Isolation of a cDNA encoding *Aspergillus oryzae* Taka-amylase A: evidence for multiple related genes

(Recombinant DNA; fungal α -amylase; multi-gene family; nucleotide sequence; intron; exon)

Norihiro Tsukagoshi, Makoto Furukawa, Hiroki Nagaba, Norie Kirita, Akio Tsuboi and Shigezo Udaka

Department of Food Science and Technology, Faculty of Agriculture, Nagoya University, Nagoya 464 (Japan)

Received by T.A. Bickle: 13 June 1989

Accepted: 3 July 1989

SUMMARY

Complementary and genomic DNAs encoding *Aspergillus oryzae* Taka-amylase A (Taa) were cloned and sequenced. The coding sequence of the cDNA comprised the signal peptide [21 amino acids (aa)] and mature Taa (478 aa). The deduced aa sequence agrees well with the published aa sequence, except for one insertion, one deletion and ten aa substitutions. These differences might be due to the difference in the strains used. Sequence comparison of the cDNA and genomic DNA indicates the presence of eight introns ranging in size from 55 to 86 bp. Southern-blot analysis showed the presence of at least two *Taa* genes, and the second gene (*Taa-G2*) was isolated. All the intron/exon junctions follow the 'GT-AG' rule, except for intron I of the first gene (*Taa-G1*).

The 5'-noncoding region was well conserved among the genomic genes and contained sequences similar to 'CAAT' and 'TATA' boxes at nucleotides -121 and -31, counted from the transcription start point, respectively. The 3'-noncoding regions, however, differed significantly from each other. *Taa-G2* contains a sequence identical to that of several independent cDNA clones, suggesting that it may be the major transcribed gene in *A. oryzae*.

INTRODUCTION

Filamentous fungi have a long history of use in fermentation. *A. oryzae* has been used in food pro-

Correspondence to: Dr. N. Tsukagoshi, Department of Food Science and Technology, Faculty of Agriculture, Nagoya University, Nagoya 464 (Japan) Tel. (052) 781-5111; Fax (052) 781-4447.

Abbreviations: aa, amino acid(s); *A.*, *Aspergillus*; Ap, ampicillin; bp, base pair(s); cDNA, DNA complementary to mRNA; DP medium, see MATERIALS AND METHODS, section a; FDA, Food and Drug Administration; GRAS, Generally Regarded as Safe; kb, kilobase(s) or 1000 bp; LB, Luria-Bertani (broth); nt, nucleotide(s); Taa, Taka-amylase A; *Taa*, gene (DNA) encoding Taa; *tsp*, transcription start point.

duction in Japan and has also been used for industrial enzyme production, such as α -amylase, the so-called Taa. *A. oryzae* and *A. niger* are both on the GRAS list of the FDA (Taylor and Richardson, 1979) and are now targets for the application of recombinant DNA technology.

α -Amylase is widely distributed in microorganisms, as well as in plants and animals (Bernfeld, 1955). Many microbial and eukaryotic α -amylase-encoding genes have been cloned and well characterized (Palva, 1982; Schibler et al., 1982; Rogers and Milliam, 1983; Young et al., 1983; Nakamura et al., 1984; Tsukagoshi et al., 1984; Yuuki et al., 1985; Boer and Hickey, 1986; Long et al., 1987).

α -Amylases of diverse origins have similar sequences in the enzymatically functional regions, probably as a consequence of convergent evolution (Ihara et al., 1985; Nakajima et al., 1986; Long et al., 1987). Among α -amylases, Taa, a glycoprotein consisting of 478 aa, has been most extensively characterized. Both the aa sequence and three-dimensional structure have been elucidated (Toda et al., 1982; Matsuura et al., 1984).

The enzymology and protein chemistry of Taa have been studied extensively, but little is yet known about the molecular biology of the mRNA and the gene(s) encoding the enzyme. Recently, the genomic DNA encoding Taa has been cloned (Christensen et al., 1988) and sequenced (Tada et al., 1989). Using the promoter sequence of the *Taa* gene, Christensen et al. (1988) demonstrated a highly effective production (3.3 mg/l) of the *Rhizomucor miehei* aspartic proteinase in *A. oryzae*.

In this paper, we describe the cloning and sequencing of the cDNA encoding Taa, as well as two genomic DNAs (*Taa-G1* and *G2*). The nt sequence of the *Taa* gene determined by Tada et al. (1989) differed in the 3'-noncoding region from those of complementary and genomic DNAs determined here. Taa appears to be encoded by a small multi-gene family.

MATERIALS AND METHODS

(a) Strains, phages, plasmids and media

A. oryzae JCM02239 was used for isolation of DNA and RNA, and grown at 30°C in DP medium (Iimura et al., 1987) or DP medium supplemented with 1% soluble starch. The following strains of *Escherichia coli* were used: JM103 (Messing, 1983) for DNA manipulation and growth of M13 phages; Q359 (Karn et al., 1980) for growth of recombinant λ gt10 phages; and Y1090 (Young and Davis, 1983) for growth of λ gt11 phages. pUC plasmids (Messing, 1983) were used to subclone various restriction fragments. Strains of plasmid-carrying *E. coli* were grown at 37°C in LB medium containing 50 μ g Ap/ml. Q359 was grown at 37°C in TB medium (10 g Bacto-tryptone/5 g NaCl pH 7.4) supplemented with 10 mM MgSO₄ and 0.2% maltose.

Y1090 was grown at 37°C in LB medium containing 50 μ g Ap/ml and 0.4% maltose.

(b) Isolation of DNA and RNA

Total DNA and RNA were isolated by using SDS-phenol (Palmiter, 1974) and guanidium isothiocyanate (Maniatis et al., 1982), respectively, from protoplasts which were prepared from *A. oryzae* grown at 30°C for three days in DP medium supplemented with 1% soluble starch, as described by Ballance and Turner (1985). Poly(A)⁺ RNA was obtained by chromatography on oligo(dT)-cellulose (Aviv and Leder, 1972). Phage and plasmid DNAs were isolated by standard procedures (Maniatis et al., 1982; Messing, 1983).

(c) Construction and screening of complementary and genomic DNA libraries

Double-stranded cDNA was synthesized using poly(A)⁺ RNA and the cDNA synthesis system 'plus' (Amersham, U.K.) and a cDNA library was constructed in the expression vector, λ gt11. The cDNA library was screened with an antibody to Taa raised in a rabbit, as described previously (Tsukagoshi et al., 1988).

A genomic DNA library, constructed in phage vector, λ EMBL 3, from a partial *Bcl*I digest of chromosomal DNA, was screened with ³²P-labeled cDNA fragments, M1 and T1, as probes (see Fig. 1).

RESULTS AND DISCUSSION

(a) Cloning of *Taa* complementary and genomic DNAs

cDNA libraries constructed in λ gt11 using oligo(dT)₁₂₋₁₈ and random hexanucleotides as primer were screened with antibody to Taa. Of approx. 10⁵ bacteriophages screened, ten and twelve positive clones were obtained from oligo(dT) and random hexanucleotide-primed cDNA libraries, respectively. To confirm that the clones came from the *Taa* gene, all the inserts were sequenced, since the aa sequence of the enzyme has been elucidated (Toda et al., 1982). Two clones, λ M1 (hexanucleotide-

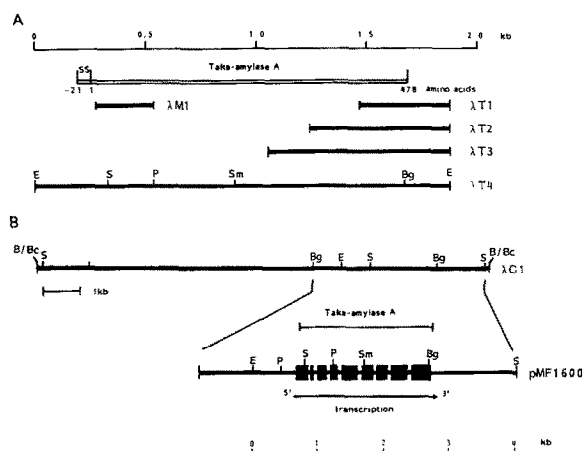


Fig. 1. Restriction maps of the complementary and genomic DNAs encoding Taa. (A) Taa and its cDNA are presented as open and blackened bars, respectively. SS represents the signal peptide of Taa. (B) The upper part indicates the 12.2-kb *Bcl*I fragment cloned in λ G1. The 4.8-kb *Bgl*II-*Sal*I fragment coding for Taa was subcloned (pMF1600, lower part). Protein-encoding (exon) regions are presented as blackened boxes. The arrow indicates the *Taa* position and direction of transcription. B, *Bam*HI; Bc, *Bcl*I; Bg, *Bgl*II; P, *Pst*I; S, *Sal*I; Sm, *Sma*I.

primed cDNA) and λ T1 [oligo(dT)-primed cDNA] shown in Fig. 1A, contained 251- and 274-bp inserts, coding for Phe-13 to Gly-95 and Val-442 to Ser-478, respectively. Two additional clones, λ T2 and λ T3, were isolated by means of the immunological procedure. All the clones, however, contained rather short inserts, up to 794 bp. Therefore, the oligo(dT)-primed cDNA library was screened with ^{32} P-labeled λ M1 and λ T1 inserts as hybridization probes. Of approx. 10^5 bacteriophages screened, one clone, λ T4, hybridizing to both probes, was isolated. λ T4 contained an approx. 1.9-kb insert, which should be sufficient to code for Taa, as judged from the published aa sequence (Fig. 1A).

A genomic DNA library constructed in λ EMBL3 was also screened with λ M1 and λ T1 inserts as probes. Of approx. 3×10^4 bacteriophages screened, four clones hybridized to both probes and contained the same 12.2-kb insert.

One clone, λ G1, was chosen for further characterization. Southern blots of the restriction-endonuclease digestion products were sequentially probed with the λ M1 and λ T1 inserts, and a restriction map around the *Taa* gene was constructed (Fig. 1B). The gene appeared to be located in the 4-kb *Bgl*II-*Sal*I fragment, which was subcloned into pUC19 (pMF1600) for further characterization.

(b) Nucleotide sequences and characterization of complementary and genomic DNAs

All five cDNA inserts on λ M1 and λ T1-T4, and most (2935 bp) of the 3.2-kb *Eco*RI-*Sal*I fragment on pMF1600, were sequenced. Comparison of the genomic nt sequence with the cDNA sequence revealed that the gene contains nine exons (total 1497 nt) interrupted by eight introns (total 541 nt). Fig. 2 shows the nt sequence of the genomic DNA along with the deduced aa sequence.

Although the exact translation initiation site has not been elucidated, the aa sequence (Toda et al., 1982) indicates that the enzyme is synthesized as a precursor with a putative signal peptide of 21 aa residues, and that the secretory precursor is processed at a specific cleavage site between Ala-Ala residues, resulting in the formation of the mature enzyme, which consists of 478 aa. The deduced aa sequence agrees with the published sequence, except for 12 aa residues, one insertion, one deletion and ten substitutions of aa residues, as indicated in Fig. 2. These differences in the sequence might be due to differences in the strains used.

The eight introns within the *Taa* gene range in size from 55 to 86 bp, resembling other filamentous fungal introns in both their size (less than 100 nt in length) and their conserved sequences. Except for intron I, all of the other intron/exon junctions follow the 'GT-AG' rule of Breathnach and Chambon (1981). The intron/exon junction of intron I was tentatively assigned as GG-AT in comparison with the cDNA sequences of λ M1 and λ T4 inserts.

(c) Identification and characterization of the second *Taa* genomic gene

Southern blots of chromosomal DNA after cleavage with several restriction enzymes were probed at high stringency with ^{32} P-labeled cDNA fragments coding for N- and C-terminal portions of Taa (Fig. 3A). Both probes revealed prominent hybridization to one or two fragments of similar sizes, depending on the restriction enzyme used, and a few faintly hybridizing fragments. These data indicate the presence of at least two sequences related to the *Taa* gene in the *A. oryzae* genome.

Both the 7-kb and 3-kb *Eco*RI (Fig. 3) fragments were cloned and their restriction maps were con-

[illegible]

Fig. 2. Nucleotide sequence of the *Taa* genomic DNA and the deduced aa sequence of Taa. The sequences were determined as described by Sanger et al. (1977) after subcloning of appropriate restriction fragments into derivatives of bacteriophage M13. The *tsf* determined by S1 mapping analysis (Fig. 4) is indicated by +1. The nt sequence is numbered on the right. The deduced aa sequence is numbered on the left (1), from the N terminus (Ala) of the mature enzyme. The stop codon is indicated by asterisks. The cleavage site between the signal sequence and Taa is indicated by a downward arrow. Introns I–VIII, inferred by comparison with the cDNA sequence, are shown in lower-case letters. Sequences thought to be involved in the splicing process and the intron/exon junctions are underlined, except

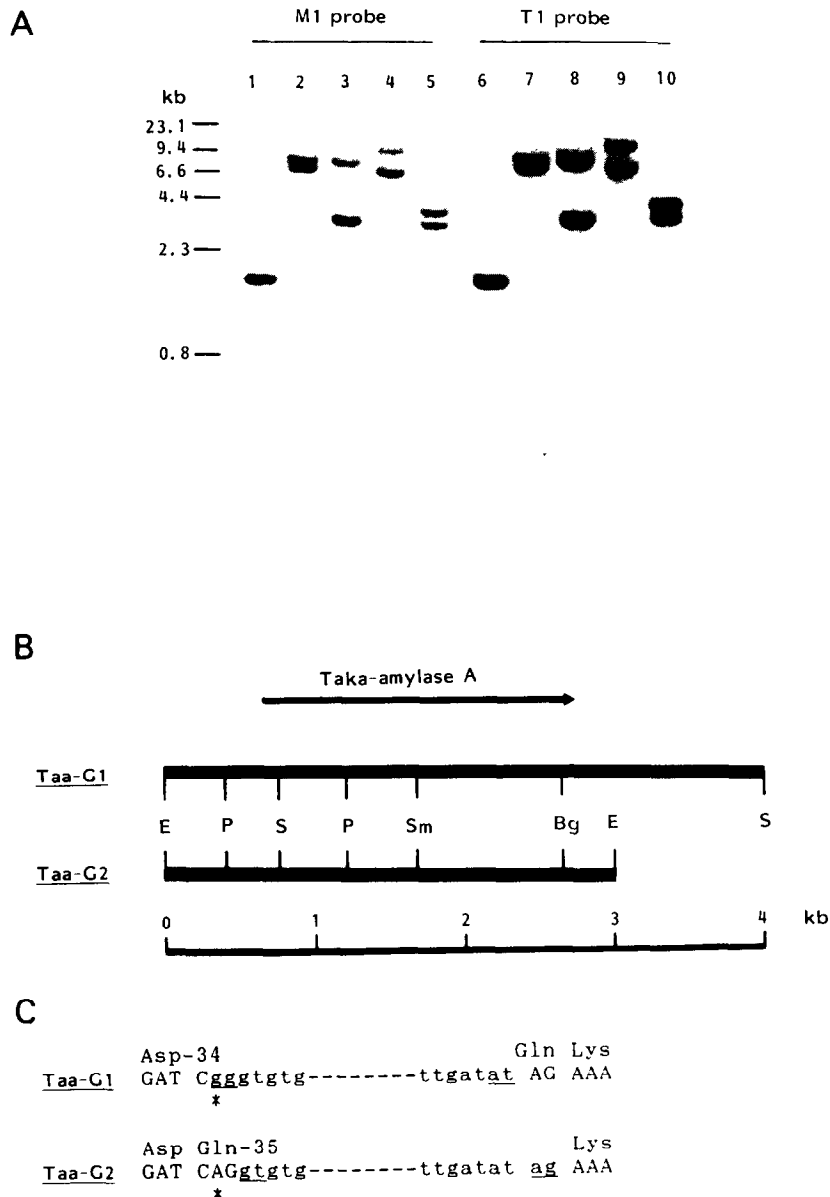


Fig. 3. Southern blot analysis of *A. oryzae* chromosomal DNA (A) and characterization of the second *Taa* gene (B, C). (A) DNA was digested with *AccI* (lanes 1 and 6), *Bam*HI (lane 2 and 7), *Eco*RI (lanes 3 and 8), *Pst*I (lanes 4 and 9) and *Sal*I (lanes 5 and 10), and processed for Southern blot hybridization with the M1 and T1 fragment isolated from λ M1 and λ T1 (Fig. 1), respectively, as probes. (B) Comparison of restriction maps of two *Taa* genomic sequences. The first and second genes are designated as *Taa-G1*, the 4-kb *Eco*RI-*Sal*I fragment, described in RESULTS AND DISCUSSION, sections a and b, and *Taa-G2*, the 3-kb *Eco*RI fragment cloned here, respectively. For abbreviations of restriction sites, see Fig. 1. (C) Nucleotide sequences around the intron I junctions of two genomic DNAs. Upper- and lower-case letters indicate the exon and intron sequences, respectively. The intron/exon junctions inferred by comparison with both the cDNA sequence and the chemically determined sequence of *Taa* (Toda et al., 1982) are underlined. The asterisks indicate the differences between two genes.

for those of intron I, which are double underlined. The palindromic sequences in the 3'-noncoding region are indicated by horizontal facing arrows. Putative CAAT and TATA boxes, the glycosylation site (Asn-196) and the polyadenylation signal are boxed. Aa residues different from the chemically determined sequence of *Taa* (Toda et al., 1982) are indicated by the following symbols above the line: ten aa substitutions (blackened circle), one inserted aa at 388 (open circle) and one deleted aa at 476 (open triangle).

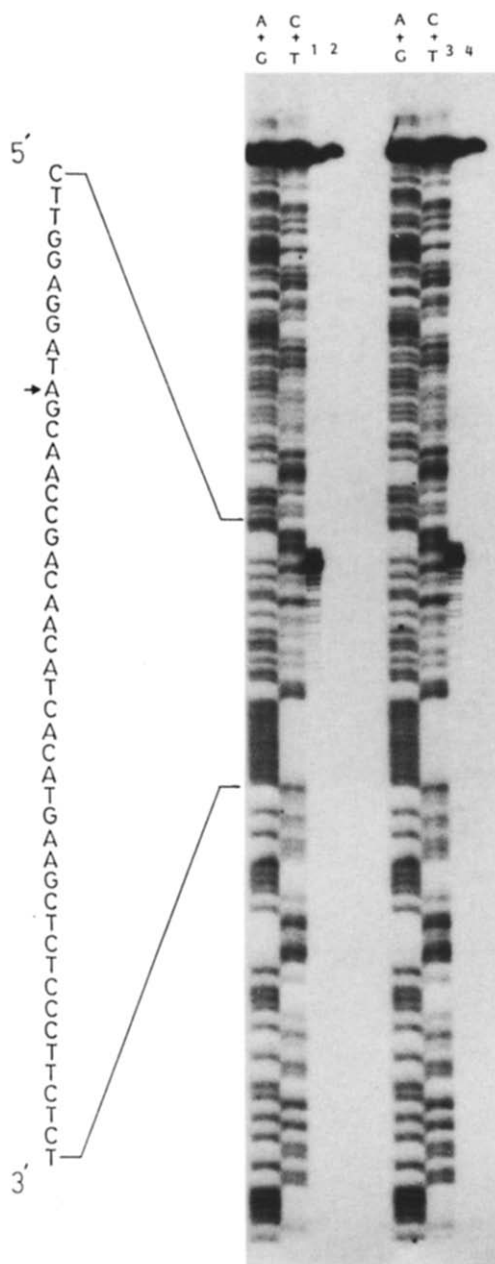


Fig. 4. S1 mapping analysis of the *Taa* transcript. Total and poly(A)⁺ RNAs were prepared as described in MATERIALS AND METHODS, section b. The *Eco*RI-*Sal*I fragment (734 bp) of pMF1600 (Fig. 1) was isolated and labeled with ³²P at the 5' end of the sense strand, followed by digestion with *Pst*I. The resultant *Pst*I-*Sal*I fragment (339 bp), which contains 200 bp of the 5'-noncoding sequence and 139 bp of the 5'-coding sequence, was used as a probe DNA. The probe DNA (5×10^4 cpm, 100 ng) was incubated in the presence of 80% formamide at 50°C for 3 h with 50 µg of total RNA (lane 1) or 1 µg of poly(A)⁺ RNA (lane 3) or without RNA (lanes 2 and 4). The hybridized molecules were treated with 75 units of S1 nuclease (Bethesda Research Laboratories, Inc.) in 0.3 ml of reaction mixture at 37°C for 30 min. S1-treated DNAs were subjected to

structed (Fig. 3B). The 7-kb *Eco*RI fragment showed exactly the same restriction map as that of the *Taa* gene described above and designated as *Taa-G1*. The 3-kb *Eco*RI fragment, on the other hand, had a restriction map different in the 3'-noncoding region from that of *Taa-G1* and was designated *Taa-G2*. Both genes revealed almost the same restriction map in the structural gene region (Fig. 3B). A partial nt sequence of *Taa-G2* was determined and compared to that of *Taa-G1* (Fig. 5). Both genes contained nearly identical sequences in the 5'-noncoding regions, but differed significantly from each other in the 3'-noncoding regions. A striking feature of *Taa-G2* from the standpoint of the intron/exon junction, is that the intron I also follows the 'GT-AG' rule (Fig. 3C). G at nt 231 of *Taa-G1* was substituted for A in *Taa-G2*. This is also the case with the third gene (*Taa-G3*) isolated from another strain by Tada et al. (1989).

(d) The 5'-noncoding region

To define the site of initiation of transcription, S1 mapping analysis was performed using *Taa-G1*, total RNA and poly(A)⁺ RNA (Fig. 4). The transcription start point (*tsp*) was localized to the A, 64 nt upstream from the ATG start codon (Fig. 2). The 5'-noncoding region of all three genomic DNAs had essentially the same sequence (Fig. 5A) and contained the characteristic sequences noted in other eukaryotic genes. The sequence, TATAAAT, is located at nt -31. It resembles the consensus sequence for the 'TATA' box, 5'-TATA(A/T)A-(A/T)-3', which is normally found 26-36 bp upstream from the position corresponding to the 5' end of the mRNA (Gannon et al., 1979). Other sequences, ATGCAATTGA or ATGCAATTTA, which are located at nt -121 have homology to both the 'CCAAT' box (Benoist et al., 1980) and the homeoprotein-binding consensus sequence (octamer motif) 'ATGCAAATNA' (Thali et al., 1988). It will be of interest to determine whether any

electrophoresis on an 8% polyacrylamide gel containing 8 M urea. The ³²P-labeled probe DNA was simultaneously processed for nt sequence determination (A + G and C + T) as described by Maxam and Gilbert (1980). The nt sequence of the sense strand around the position of the S1-protected fragment (*tsp*) indicated by an arrow is shown on the left.

A

Taa-G3 CATATA
Taa-G1 AATGCAATTGAACTCTTCTGCGAATCGCTTATCCCGCCCTTGGCCGTAGAGCTTAAAGTATGTCCTTGTGATGGGATGTATCACAACCAATA -167
Taa-G2 AATGCAATTGAACTCTTCTGCGAATCGCTTATCCCGCCCTTGGCCGTAGAGCTTAAAGTATGTCCTTGTGATGGGATGTATCACAACCAATA
Taa-cDNA ttcggccatataaaatggttcattgttcattactctataatgctaattgttagattagcacaactatgactgggcaaatccgcccggccata
 +1
Taa-G3 AATAGTAGCAAGGGATGCCATGCTTGGAGGATAGCAACCGACAACATCACATCAAGCTCTCCCTTCTCTGAACAATAAACCCACAGAAGGCATTATG
Taa-G1 AATAGTAGCAAGGGATGCCATGCTTGGAGGATAGCAACCGACAACATCACATCAAGCTCTCCCTTCTCTGAACAATAAACCCACAGAAGGCATTATG 67
Taa-G2 AATAGTAGCAAGGGATGCCATGCTTGGAGGATAGCAACCGACAACATCACATCAAGCTCTCCCTTCTCTGAACAATAAACCCACAGAAGGCATTATG
Taa-cDNA gatagatcatctcctctcggacgttgcggaagcaacgacaacatcacatcagctctcctctctctgaacataaacccacagaaggcatttatg Met

B

Taa-G3 AGCTCGTGAAGGGTGGAGACTATATGATGGTACTGCTATTCAATCTGGCATTGGACAGTGAGTTGAGTTTGATGTACAGTTGGAGTCGTTACTGCTGT
Taa-G1 AGCTCGTGAAGGGTGGAGACTATATGATGGTACTGCTATTCAATCTGGCATTGGACAGTGAGTTGAGTTTGATGTACAGTTGGAGTCGTTACTGCTGT 2195
Taa-G2 AGCTCGTGAAGGGTGGAGACTATATGATGGTACTGCTATTCAATCTGGCATTGGACAGTGAGTTGAGTTTGATGTACAGTTGGAGTCGTTACTGCTGT
Taa-cDNA agctcgtgaagggtggagactatgatggtactgctattcaatctggcattggacagtggatttgagtttgatgtacagtataaatctagtgtaacttt SerSerSer
Taa-G3 CATCCCTTATACCTCTTCGATTGTTTTTGAACCCATAATGCCAAGCACGCTAGTCTATTATAGGAAAGGATCCCCGGG
Taa-G1 GGTCTTTTGTTCATAGTTGGAAATCGGAATGATCTCAATCTTGAATAATATAAAAGGATAAATACTCACATCCATCACAACCTTACAAGGTTAAT 2295
Taa-G2 GCACCCACCAACCAATGAACGGGACCGGCCCGCTCTGAGAGCCCGTCTCGAATCCCTGTTGGTCATCTTCCATCGCTTCGCTCCAGAGGCGAGGA
*Taa-cDNA** gcaaccacacacaaatgaacggcaccggcccgactgagagcccgctctcgaaaccccaaaaaaaaaaaaaaaaaaaaaa
Taa-cDNA gcaaccacacacaaatgaacggcaccggcccgactgagagcccgctctcgaaaccccaaaaaaaaaaaaaaaaaaaaaa

Fig. 5. Comparison of the 5'- and 3'-noncoding regions of cDNA and genomic DNA. cDNA and genomic nt sequences are indicated by lower- and upper-case letters, respectively. *Taa-G1*, *G2* and cDNA were cloned in this work. *Taa-G3* was sequenced by Tada et al. (1989). The sequence of *Taa-G3* upstream from nt -117 was not determined. The nt sequence is numbered on the right. The asterisks indicate the nt identical to cDNA. (A) Comparison of the 5'-noncoding regions. Putative CAAT and TATA boxes are boxed. The *tsp* is indicated by +1. (B) Comparison of the 3'-noncoding regions. The polyadenylation signals are boxed. *Taa-cDNA**, the insert on λ T1, contained the sequence 3 nt longer than other cDNA just upstream of poly(A) tail. The three blackened circles mark the stop codon.

transcriptional activators participate in the expression of the gene, since Taa is an inducible enzyme.

The 5'-flanking region of the cDNA was identical to the genomic DNAs up to 64 nt upstream from the translation initiation site (Fig. 5A), confirming the *tsp* determined by the S1 nuclease assay. The additional cDNA fragment of 125 nt does not reveal any significant homology to any of the genomic nt sequences determined so far. At present, it is not known whether the additional sequence has any role at the transcriptional or the translational level.

(e) The 3'-noncoding region

All the genomic DNAs had exactly the same sequence up to 68 or 72 nt downstream from the stop codon and diverged thereafter (Fig. 5B). Only *Taa-G1* contains the canonical polyadenylation sequence, 'AATAAA' (Proudfoot and Brownlee, 1976), 136 nt downstream from the stop codon.

The 3'-flanking sequences determined for the cDNAs of four clones, λ T1-T4, showed a micro-

heterogeneity just upstream from the poly(A) tail, but were identical to all the genomic nt sequences up to 68 or 72 nt downstream from the stop codon. After this point, all of the cDNA sequences were identical to *Taa-G2* (Fig. 5B), suggesting that *Taa-G2* may be the main gene active in *A. oryzae*. In common with some other fungal genes (Ballance, 1986; Tien and Tu, 1987), the typical polyadenylation signal is not present within the 3'-flanking sequences of the cDNAs. The sequence, CAATG, located 41 nt upstream from the poly(A) tail is, however, quite homologous to the sequence, CAYTG, implicated as a polyadenylation signal (Berget, 1984).

(f) Conclusions

(1) The cDNA-encoding Taa consisted of 1855 nt and contained an additional fragment of 125 nt upstream from the *tsp*. The coding sequence of the cDNA comprised the signal peptide of 21 aa and mature Taa of 478 aa.

(2) Two genomic DNAs (*Taa-G1* and *Taa-G2*)

were isolated and sequenced. Both genes contained eight introns ranging in size from 55 to 86 nt. All the intron/exon junctions followed the 'GT-AG' rule, except intron I of *Taa-G1*.

(3) The 5'-noncoding region was well conserved among genomic DNAs and contained sequences similar to 'CAAT' and 'TATA' boxes at nt -121 and -31 from the *tsp*, respectively.

(4) The 3'-noncoding regions differed significantly from each other. *Taa-G2* contained a sequence identical to the cDNA and was suggested to be the major transcribed gene in *A. oryzae*.

REFERENCES

- Aviv, H. and Leder, P.: Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA* 69 (1972) 1408-1412.
- Ballance, D.J.: Sequences important in gene expression in filamentous fungi. *Yeast* 2 (1986) 224-236.
- Ballance, D.J. and Turner, G.: Development of a high-frequency transforming vector for *Aspergillus nidulans*. *Gene* 36 (1985) 321-331.
- Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P.: The ovalbumin gene-sequence of putative control regions. *Nucleic Acids Res.* 8 (1980) 127-142.
- Berget, S.M.: Are U4 small nuclear ribonucleoproteins involved in polyadenylation? *Nature* 309 (1984) 179-182.
- Bernfeld, P.: Amylases, α and β . *Methods Enzymol.* 1 (1955) 149-158.
- Boer, P.H. and Hickey, D.A.: The α -amylase gene in *Drosophila melanogaster*: nucleotide sequence, gene structure and expression motifs. *Nucleic Acids Res.* 14 (1986) 8399-8411.
- Breathnach, R. and Chambon, P.: Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50 (1981) 349-383.
- Christensen, T., Woeldike, H., Boel, E., Mortensen, S.B., Hjortshøj, K., Thim, L. and Hansen, M.T.: High level expression of recombinant genes in *Aspergillus oryzae*. *Bio/Technology* 6 (1988) 1419-1422.
- Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Gerapin, A., Cami, B. and Chambon, P.: Organization and sequences at the 5' end of a cloned complete ovalbumin gene. *Nature* 278 (1979) 428-434.
- Ihara, H., Sasaki, T., Tsuboi, A., Yamagata, H., Tsukagoshi, N. and Udaka, S.: Complete nucleotide sequence of a thermophilic α -amylase gene: homology between prokaryotic and eukaryotic α -amylases at the active sites. *J. Biochem.* 98 (1985) 95-103.
- Iimura, Y., Gomi, K., Uzu, H. and Hara, S.: Transformation of *Aspergillus oryzae* through plasmid-mediated complementation of methionine-auxotrophic mutation. *Agric. Biol. Chem.* 51 (1987) 323-328.
- Karn, J., Brenner, S., Barnett, L. and Cesareni, G.: Novel bacteriophage λ cloning vector. *Proc. Natl. Acad. Sci. USA* 77 (1980) 5172-5176.
- Long, C.M., Virolle, M.-J., Chang, S.-Y., Chang, S. and Bibb, M.J.: α -Amylase gene of *Streptomyces limosus*: nucleotide sequence, expression motifs and amino acid sequence homology to mammalian and invertebrate α -amylase. *J. Bacteriol.* 169 (1987) 5745-5754.
- Maniatis, T., Fritsch, E.F. and Sambrook, J.: *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.
- Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M.: Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.* 95 (1984) 697-702.
- Maxam, A.M. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65 (1980) 499-560.
- Messing, J.: New M13 vectors for cloning. *Methods Enzymol.* 101 (1983) 20-78.
- Nakajima, R., Imanaka, T. and Aiba, S.: Comparison of amino acid sequences of eleven different α -amylases. *Appl. Microbiol. Biotechnol.* 23 (1986) 355-360.
- Nakamura, Y., Ogawa, M., Nishide, T., Emi, M., Kosaki, G., Himeno, S. and Matsubara, K.: Sequences for cDNAs for human salivary and pancreatic α -amylases. *Gene* 28 (1984) 263-270.
- Palmiter, R.D.: Magnesium precipitation of ribonucleoprotein complexes. Expedient techniques for the isolation of undegraded polysomes and messenger ribonucleic acid. *Biochemistry* 13 (1974) 3606-3615.
- Palva, I.: Molecular cloning of α -amylase gene from *Bacillus amyloliquefaciens* and its expression in *B. subtilis*. *Gene* 19 (1982) 81-87.
- Proudfoot, N.J. and Brownlee, G.G.: 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 263 (1976) 211-214.
- Rogers, J.C. and Milliam, C.: Isolation and sequence analysis of a barley α -amylase cDNA clone. *J. Biol. Chem.* 258 (1983) 8169-8174.
- Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74 (1977) 5463-5467.
- Schibler, U., Pittet, A., Young, R.A., Hagenbuehle, O., Tosi, M., Gellman, S. and Wellauer, P.K.: The mouse α -amylase multi-gene family: sequence organization of members expressed in the pancreas, salivary gland and liver. *J. Mol. Biol.* 155 (1982) 247-266.
- Tada, S., Iimura, Y., Gomi, K., Takahashi, K., Hara, S. and Yoshizawa, K.: Cloning and nucleotide sequence of the genomic Taka-amylase A gene of *Aspergillus oryzae*. *Agric. Biol. Chem.* 53 (1989) 593-599.
- Taylor, M.J. and Richardson, T.: Applications of microbial enzymes in food systems and in biotechnology. *Adv. Appl. Microbiol.* 25 (1979) 7-35.
- Thali, M., Muller, M.M., DeLorenzi, M., Matthias, P. and Bienz, M.: *Drosophila* homoeotic genes encode transcriptional activators similar to mammalian OTF-2. *Nature* 336 (1988) 598-601.

- Tien, M. and Tu, C.P.D.: Cloning and sequencing of a cDNA for a ligninase from *Phanerochaete chrysosporium*. *Nature* 326 (1987) 520.
- Toda, H., Kondo, K. and Narita, K.: The complete amino acid sequence of Taka-amylase A. *Proc. Jpn. Acad.* 58B (1982) 208–212.
- Tsukagoshi, N., Ihara, H., Yamagata, H. and Udaka, S.: Cloning and expression of thermophilic α -amylase gene from *Bacillus stearothermophilus* in *Escherichia coli*. *Mol. Gen. Genet.* 193 (1984) 58–63.
- Tsukagoshi, N., Ando, Y., Tomita, Y., Uchida, R., Takemura, T., Sasaki, T., Yamagata, H., Udaka, S., Ichihara, Y. and Takahashi, K.: Nucleotide sequence and expression in *Escherichia coli* of cDNA of swine pepsinogen: involvement of the amino-terminal portion of the activation peptide segment in restoration of the functional protein. *Gene* 65 (1988) 285–292.
- Young, M., Galizzi, A. and Henner, D.: Nucleotide sequence of the amylase gene from *Bacillus subtilis*. *Nucleic Acids Res.* 11 (1983) 237–249.
- Young, R.A. and Davis, R.W.: Yeast RNA polymerase II genes: isolation with antibody probes. *Science* 222 (1983) 827–842.
- Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N. and Udaka, S.: Complete nucleotide sequence of a gene coding for heat- and pH-stable α -amylase of *Bacillus licheniformis*: comparison of the amino acid sequences of three bacterial liquefying α -amylases deduced from the DNA sequences. *J. Biochem.* 98 (1985) 1147–1156.