

Isolation and Sequencing of a Genomic Clone Encoding Aspartic Proteinase of *Rhizopus niveus*

HIROYUKI HORIUCHI, KOJI YANAI, TAKASHI OKAZAKI, MASAMICHI TAKAGI,* AND KEIJI YANO

Department of Agricultural Chemistry, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Received 28 May 1987/Accepted 16 October 1987

A gene encoding *Rhizopus niveus* aspartic proteinase was isolated from an *R. niveus* genomic library by using oligonucleotides probes corresponding to its partial amino acid sequence, and its nucleotide sequence was determined. By comparing its deduced amino acid sequence with the amino acid sequence of rhizopuspepsin (5, 26), we concluded that the *R. niveus* aspartic proteinase gene has an intron within its coding region and that it has a preproenzyme sequence of 66 amino acids upstream of the mature enzyme of 323 amino acids.

Rhizopus niveus, a filamentous fungus, excretes a large amount of several enzymes extracellularly. Aspartic proteinase of this species (RNAP) is one of the most abundantly secreted enzymes. *Rhizopus chinensis* aspartic proteinase is one of the best characterized aspartic proteinases (9, 20, 21), and recently its primary structure was determined (5, 26). Its three-dimensional structure was also investigated at a high resolution (2, 25). The primary structure and X-ray structure of *Penicillium janthinellum* aspartic proteinase, penicillopepsin, have also been investigated (15). In addition, *Mucor pusillus* and *Mucor miehei* aspartic proteinases, *Mucor* rennins, have been characterized, and their primary structures have been determined (7, 8, 10, 27). Kurono et al. (16) have reported the purification and characterization of RNAP. Considering these data, cloning and structural analysis of a gene encoding RNAP will be interesting for the investigation of the structure-function relationships of fungal aspartic proteinases and may be useful for the construction of secretion vectors and the elucidation of the secretion mechanism of enzymes in fungi. In this paper, the isolation and characterization of the RNAP gene are described.

MATERIALS AND METHODS

Strains and media. A genomic library of *R. niveus* was constructed in *E. coli* JA221 (*recA1 leuB6 trpE5 hsdR hsdM⁺ lacY thr thi*) (4) by using pBR322 as a vector. *E. coli* JM109 [*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1 λ⁻ Δ(lac-proAB) F' traD36 proAB lacI^qZ-M15*] (30) was used as a host for plasmids pUC18 and pUC19 in nucleotides sequencing. Total DNA was isolated by a modification of the method of Hynes et al. (13) from *R. niveus* Yamazaki IFO4810 for construction of a genomic library. Potato glucose broth was purchased from Difco Laboratories, Detroit, Mich., for the cultivation of *R. niveus*. LB medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl) was used for *E. coli*.

Construction of a genomic library. Total DNA isolated from *R. niveus* was partially digested with *Sau3AI* and fractionated by sucrose density gradient centrifugation. Fractions containing fragments of 8 to 15 kilobases were pooled and ligated with *Bam*HI-digested pBR322. The ligation

mixture was used for transformation into *E. coli* JA221 by the conventional method (18).

Purification and analysis of RNAP. RNAP was purified from Gluczyme (Amano Pharmaceutical Company, Nishiharu, Aichi, Japan), a commercial digestive prepared from *R. niveus*. The purification procedure was carried out at 4°C unless indicated otherwise. Gluczyme (10 g) was dissolved in 40 ml of ice water and stirred for 1 h. The solution was centrifuged at 20,000 × *g* for 10 min to remove insoluble materials. Acetone chilled at -20°C was added to the supernatant to about 60% (vol/vol), and the mixture was stirred for 30 min at a temperature below 0°C. The precipitate was collected by centrifugation at 20,000 × *g* for 10 min and dissolved in 40 ml of ice water. The solution was centrifuged, and the supernatant was applied to a column of DEAE-Sephadex A-50 (26 by 35 cm) equilibrated with 50 mM Tris hydrochloride (pH 7.5). The column was eluted with a linear NaCl gradient from 100 to 300 mM. Fractions rich in proteinase activity were boiled in 1% sodium dodecyl sulfate-0.5% 2-mercaptoethanol for 10 min and concentrated in a Minicon apparatus (Amicon Corp., Danvers, Mass.). The solution thus obtained was applied to a high-pressure liquid chromatography column of G3000SW (Toyo Soda, Tokyo, Japan) equilibrated with 0.1% sodium dodecyl sulfate-100 mM NaCl. The eluted fraction was recovered and used as purified RNAP. It was homogeneous in sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

Proteinase activity was measured by a modification of the method of Fukumoto et al. (9). For determination of the amino acid composition, RNAP was hydrolyzed in 6 N HCl, and the hydrolysate was analyzed with a model MLC703 amino acid analyzer (Atto Co. Ltd., Tokyo, Japan). RNAP was cleaved with lysylendopeptidase (Wako Junyaku, Osaka, Japan) as follows. RNAP was dissolved in 200 μl of 0.075% sodium dodecyl sulfate-60 mM Tris hydrochloride (pH 9.0) at a concentration of 1 μg/μl. Lysylendopeptidase (5 μg) was added to the solution, and the solution was incubated for 16 h at 37°C. The resulting peptides were separated with a C₁₈ reversed-phase column (Shimpak ODS; Shimadzu Co., Tokyo, Japan). Peptides were eluted by a linear gradient of acetonitrile from 5 to 60% in 0.1% trifluoroacetic acid for 60 min at a flow rate of 0.9 ml/min.

Amino acid sequencing of RNAP and isolated peptides was performed by automated Edman degradation in a model 470A protein sequencer (Applied Biosystems).

* Corresponding author.

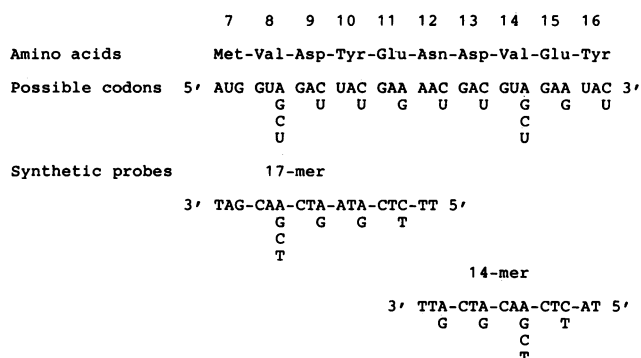


FIG. 1. Synthetic oligonucleotide probes.

Other methods. Plasmid DNA was isolated from *E. coli* by the alkaline lysis method (18). DNA sequencing was performed by the dideoxy chain termination method of Sanger et al. (23) with plasmids pUC18 and pUC19. Mixed oligonucleotides of 17-mer and 14-mer were synthesized by the automated phosphoramidite method in a model 381A DNA synthesizer (Applied Biosystems). Colony hybridization was carried out at 42°C for 48 h with ³²P-labeled probes by the method of Grunstein and Hogness (11) with some modifications. Southern blot analysis was performed by the method of Wallace et al. (29) with some modifications; the hybridization temperatures were 43°C for the 17-mer probe and 38°C for the 14-mer probe.

RESULTS

As purification by the dialysis method described by Kuro no et al. (16) was unsuccessful in our laboratory for unknown reasons, we purified RNAP from *Glucyze* by the method described above. The N-terminal amino acid sequence of RNAP obtained with our method was as follows: Ala-Ser-Gly-Ser-Val-Pro-Met-Val-Asp-Tyr-Glu-Asn-Asp-val-Glu-Tyr. RNAP purified by us has alanine at the N terminus and an optimum pH for enzyme activity of 3.0 to 3.5, while RNAP purified previously (16) was reported to have glycine at the N terminus and an optimum pH of 3.5 to 4.0. Although the amino acid compositions of these two preparations are similar, the two enzymes seem to be different. Mixed oligonucleotides of 17-mer and 14-mer (Fig. 1) were synthesized, and about 20,000 clones of the *R. niveus* genomic library were screened with the ³²P-labeled 17-mer probe. Eleven positive clones were obtained, and plasmids isolated from them were characterized by restriction endonucleases. There were four kinds of plasmids. Southern blot analysis was performed for these four kinds of plasmids with the ³²P-labeled 17-mer and 14-mer probes. One plasmid, pPRO7, had an insert of 7.4 kilobases which hybridized with both probes. The nucleotide sequence of the region of the insert hybridizable to both probes was determined by using the strategy shown in Fig. 2. The result is shown in Fig. 3. In the deduced amino acid sequence, there is a sequence of 16 amino acids (from +1 to +16; +1 is the N-terminal alanine of the mature RNAP) identical to the N-terminal amino acid sequence of RNAP shown above. So, it was strongly suggested that this region encodes RNAP.

A stop codon in-frame is present 138 bases downstream of the 14-mer probe-hybridizable sequence, indicating the presence of an intron. To confirm this, we cleaved RNAP with lysylendopeptidase, separated the resulting peptides with a C₁₈ reversed-phase column, and purified some of them. The

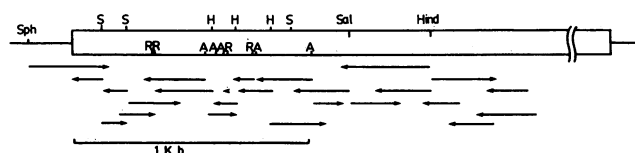


FIG. 2. Restriction map and sequencing strategy for the RNAP gene. The thin line indicates the vector pBR322 sequence, and the open box indicates the *R. niveus* DNA sequence. Abbreviations: Sph, *Sph*I; Sal, *Sal*I; Hind, *Hind*III; S, *Sau*3AI; R, *Rsa*I; A, *Alu*I; and H, *Hae*III. Kb, Kilobase.

N-terminal amino acid sequences of six of them were determined (Table 1). Furthermore, the amino acid composition of RNAP was analyzed (data not shown). Considering these results, it was concluded that an intron of 64 bases is present at the sequence encoding +43 alanine and that +323 asparagine is the C-terminal amino acid (Fig. 3).

The amino acid sequence of RNAP was compared with those of rhizopuspepsin, penicillopepsin, and human pepsin A. The homology between RNAP and rhizopuspepsin is 76%, that between RNAP and penicillopepsin is 42%, and that between RNAP and human pepsin A is 41% (Fig. 4). On the other hand, homology between RNAP and rat pepsin (14) is 34%, and that between RNAP and *Mucor* rennins is too low to be calculated (data not included in Fig. 4). It seems reasonable that RNAP is highly homologous to rhizopuspepsin, but it is interesting that RNAP has the same level of homology to penicillopepsin and human pepsin. The homology between RNAP and rhizopuspepsin is found throughout their structures, supporting our conclusion on the structure of the coding region of the RNAP gene (Fig. 3).

The primary translation product of this gene contains an N-terminal amino acid sequence not present in the mature RNAP (Fig. 3). A TATA box-like sequence (3) and three CAAT box-like sequences (6) are present upstream of the coding region, and a poly(A)-addition signal sequence (22) is present downstream of the coding region.

DISCUSSION

In this report, we isolated, using synthetic oligonucleotides as probes, a DNA fragment expected to encode RNAP. The amino acid sequence was deduced from the determined nucleotide sequence, and it was concluded that this fragment encodes RNAP.

Concerning the structure of the intron, TACTAAC and PuCTPuAC, which are commonly present in introns of genes in yeasts (17) and fungi (19), were not found in the intron of 64 bases in the RNAP gene. Furthermore, our recent experimental results indicate that these two sequences are not

TABLE 1. N-terminal amino acid sequences of six peptides from RNAP

Peptide no.	Amino acid sequence ^a	Residue no. in protein
1	SSTYAADGRTWSISYGDGSS	62-81
2	ASGSVPMVDYENDVE	1-15
3	LXFDXGSSXMXFXXL	31-46
4	RESSAFATDVIDGLL	106-120
5	IGGTTVSXSFXAXXD	203-217
6	QTIELA	99-104

^a All amino acids are shown in single-letter codes. X, Unidentified amino acid.

1 30
 GATCTCATTTCATCTCAAGTCACCTAA

60 90
 AAAATCCAATCATTTCCTTTTAGAGTAACTATTGTAATATTGAATTTTAAAGTCCAGTCATTGCTTACTACCTTGAT

120 150 180
 AAGAATAGGTCGCACATTTAAGATCACAGCAATTTGAATTAATGTAACAATTAATAAATTAATTTACTCTAATTGGG

210 240
 ATGAAATCACTAAAAAGAGTTAAGCAACGTTAATCCAAGAAAGGATATTGATCCCAGAAAGGACATTTATCCAGAATAC

270 300 330
 AACAGCATTTCCTTTTTTTTTTAACAACAGCAGAAAAATCCGATGCTTCCATGCCAATATCGCAGTTTATCAATCTACCA

360 390 420
 TTGTACACATATACACGTACACACATTTTTTTAAAAAATATTAATTATTAATAGGAGAAAGAAAACGTA

450 480
 ATAAAGGATCCGATATTCACCAAGTTCATTTCTCTATCAAACCATTTCTTTCTTCATTTTTTAATCAATCCTTTTCTA

510 540
 CTGAAAA ATG. AAG. TTC. ACT. TTA. ATC. TCC. TCC. TGT. GTA. GCA. CTT. GCT. GCC. ATG. ACA. CTT. GCT
 Met-Lys-Phe-Thr-Leu-Ile-Ser-Ser-Cys-Val-Ala-Leu-Ala-Ala-Met-Thr-Leu-Ala
 -60

570 600
 GTC. GAA. GCT. GCA. CCC. AAC. GGC. AAG. AAA. ATT. AAC. ATT. CCT. TTG. GCC. AAG. AAC. AAC. AGC. TAC
 Val-Glu-Ala-Ala-Pro-Asn-Gly-Lys-Lys-Ile-Asn-Ile-Pro-Leu-Ala-Lys-Asn-Asn-Ser-Tyr
 -30

630 660
 AAA. CCT. AGC. GCC. AAA. AAT. GCA. CTT. AAT. AAG. GCT. CTC. GCC. AAG. TAC. AAT. AGA. AGA. AAG. GTT
 Lys-Pro-Ser-Ala-Lys-Asn-Ala-Leu-Asn-Lys-Ala-Leu-Ala-Lys-Tyr-Asn-Arg-Arg-Lys-Val

690 720
 GGA. AGC. GGA. GGA. ATT. ACA. ACC. GAG. GCC. AGT. GGC. TCT. GTT. CCT. ATG. GTT. GAT. TAT. GAA. AAC
 Gly-Ser-Gly-Gly-Ile-Thr-Thr-Glu-Ala-Ser-Gly-Ser-Val-Pro-Met-Val-Asp-Tyr-Glu-Asn
 -1 1

750 780
 GAT. GTT. GAA. TAC. TAC. GGT. GAA. GTC. ACT. GTT. GGT. ACT. CCT. GGT. ATT. AAG. CTC. AAA. CTT. GAT
 Asp-Val-Glu-Tyr-Tyr-Gly-Glu-Val-Thr-Val-Gly-Thr-Pro-Gly-Ile-Lys-Leu-Lys-Leu-Asp
 30

810 840 870
 TTT. GAT. ACT. GGT. TCT. TCT. GAT. ATG. TGG. TTT. G gtaagaagttttgaaattgctgggcctttcttttattc
 Phe-Asp-Thr-Gly-Ser-Ser-Asp-Met-Trp-Phe

900 930
 actttttttttctttctttctctctag CA. TCC. ACT. TTA. TGC. TCT. TCT. TGC. AGC. AAT. TCT. CAT. ACT
 Ala-Ser-Thr-Leu-Cys-Ser-Ser-Cys-Ser-Asn-Ser-His-Thr

960 990
 AAG. TAT. GAT. CCT. AAA. AAA. TCA. AGC. ACT. TAC. GCT. GCC. GAC. GGT. CGC. ACT. TGG. TCC. ATC. TCT
 Lys-Tyr-Asp-Pro-Lys-Lys-Ser-Ser-Thr-Tyr-Ala-Ala-Asp-Gly-Arg-Thr-Trp-Ser-Ile-Ser
 60

FIG. 3—Continued on following page

60

1020 1050
 TAC.GGT.GAT.GGC.TCT.AGC.GCA.AGC.GGT.ATC.TTA.GCT.ACA.GAC.AAT.GTC.AAC.CTT.GGA.GGT
Tyr-Gly-Asp-Gly-Ser-Ser-Ala-Ser-Gly-Ile-Leu-Ala-Thr-Asp-Asn-Val-Asn-Leu-Gly-Gly
 90

1080 1100
 CTC.TTG.ATT.AAA.AAG.CAA.ACT.ATT.GAA.TTA.GCC.AAG.CGT.GAA.TCT.AGT.GCA.TTC.GCT.ACA
 Leu-Leu-Ile-Lys-Lys-Gln-Thr-Ile-Glu-Leu-Ala-Lys-Arg-Glu-Ser-Ser-Ala-Phe-Ala-Thr

1140 1170
 GAT.GTG.ATT.GAT.GGT.CTC.TTG.GGT.CTT.GGT.TTC.AAC.ACG.ATT.ACA.ACC.GTT.CGT.GGC.GTC
Asp-Val-Ile-Asp-Gly-Leu-Leu-Gly-Leu-Gly-Phe-Asn-Thr-Ile-Thr-Thr-Val-Arg-Gly-Val
 120

1200 1230
 AAG.ACT.CCA.GTC.GAC.AAT.TTG.ATC.AGT.CAA.GGT.TTG.ATC.AGC.AGA.CCC.ATC.TTC.GGT.GTT
 Lys-Thr-Pro-Val-Asp-Asn-Leu-Ile-Ser-Gln-Gly-Leu-Ile-Ser-Arg-Pro-Ile-Phe-Gly-Val
 150

1260 1290
 TAT.CTC.GGT.AAG.CAA.AGC.AAC.GGA.GGT.GGA.GGT.GAA.TAT.ATC.TTT.GGT.GGC.TAT.GAC.TCC
 Tyr-Leu-Gly-Lys-Gln-Ser-Asn-Gly-Gly-Gly-Gly-Glu-Tyr-Ile-Phe-Gly-Gly-Tyr-Asp-Ser

1320 1350
 TCC.AAG.TTC.AAG.GGT.TCC.TTA.ACT.ACT.GTC.CCT.ATC.GAT.AAC.TCA.GAA.GGC.TTC.TGG.GGT
 Ser-Lys-Phe-Lys-Gly-Ser-Leu-Thr-Thr-Val-Pro-Ile-Asp-Asn-Ser-Glu-Gly-Phe-Trp-Gly
 180

1380 1410
 GTT.ACT.GTC.AAG.AGC.ACC.AAG.ATA.GGT.GGC.ACA.ACA.GTT.TCT.GCT.TCC.TTT.GAT.GCT.ATC
 Val-Thr-Val-Lys-Ser-Thr-Lys-Ile-Gly-Gly-Thr-Thr-Val-Ser-Ala-Ser-Phe-Asp-Ala-Ile
 210

1440 1470
 CTC.GAC.ACT.GGT.ACC.ACT.CTT.TTA.CTT.CTT.CCT.GAT.GAC.GTT.GCT.GCA.AAG.GTG.GCT.AGA
Leu-Asp-Thr-Gly-Thr-Thr-Leu-Leu-Leu-Leu-Pro-Asp-Asp-Val-Ala-Ala-Lys-Val-Ala-Arg

1500 1530
 TCT.TAT.GGT.GCT.TCC.GAC.AAC.GGC.GAT.GGT.ACT.TAC.AGT.ATC.ACC.TGT.GAT.ACT.TCC.AAG
 Ser-Tyr-Gly-Ala-Ser-Asp-Asn-Gly-Asp-Gly-Thr-Tyr-Ser-Ile-Thr-Cys-Asp-Thr-Ser-Lys
 240

1560 1590
 CTT.CAA.CCT.CTT.GTC.TTC.ACC.CTT.GGT.TCT.TCC.ACC.TTC.GAA.GTT.CCC.TCT.GAC.TCC.CTC
 Leu-Gln-Pro-Leu-Val-Phe-Thr-Leu-Gly-Ser-Ser-Thr-Phe-Glu-Val-Pro-Ser-Asp-Ser-Leu
 270

1620 1650
 ATC.TTT.GAA.AAG.GAT.GGT.AAC.AAA.TGT.ATT.GCT.GGT.TTT.GCT.GCT.GGT.GGT.GAT.CTT.GCC
 Ile-Phe-Glu-Lys-Asp-Gly-Asn-Lys-Cys-Ile-Ala-Gly-Phe-Ala-Ala-Gly-Gly-Asp-Leu-Ala

1680 1710
 ATC.TTG.GGT.GAT.GTC.TTT.TTG.AAG.AAC.AAC.TAT.GTT.GTC.TTT.AAC.CAA.GAA.GTC.CCT.GAA
 Ile-Leu-Gly-Asp-Val-Phe-Leu-Lys-Asn-Asn-Tyr-Val-Val-Phe-Asn-Gln-Glu-Val-Pro-Glu
 300

FIG. 3—Continued on following page

17401770

GTT.CAA.ATT.GCA.CCT.GTT.GCC.AAT.TAA. TTTATTACTTTTCCAAATTATTATTATTATTATTATTATT
 Val-Gln-Ile-Ala-Pro-Val-Ala-Asn-***

180018301860

TTGATCAATGACATTCAATAAATCTGTGTTCTTGCATTCACTGCTCATTAGTAACAGTCTTTTTACTTGCTTCAGTTA

18901920

CGCTTTAATCCATTATAATTGCATAGTACAATTGCAGATAGAGATAATACTTATGTCGAGTATATTGCTGTTAATGGG

1950

TGCTTTAGTCACATACGTA

FIG. 3. Nucleotide sequence of the RNAP gene. The deduced amino acid sequence is shown under the nucleotide sequence. The sequences CAAT, TATAA, and AATAAA are boxed, and amino acid sequences confirmed by sequencing of RNAP are underlined. Unidentified amino acids are indicated with dotted lines. Asterisks indicate the position of the stop codon; lowercase letters indicate nucleotides in the intron sequence.

	1	10	20	30	40	50	60	70	80	
RNAP:	A SCSVPMVDYEN DVEYYGQVTVGTPGKRLKLDFTGSSDMPFAST LCSS--CSNSHTKYDPKSSST AADGRTWSISYGDGSSASGI									
RPEP:	ACVGTVPMTDYEN DVEYYGQVTVGTPGKFNLDFTGSSDLWIAST LCTN--CGSRQTKYDPKSSST YQADGRTWSISYGDGSSASGI									
PPEP:	AA SGVATNTPTAN DREYITFTVIG GTTLNLDFTGSDLWVPSLPAAS--QCSHSVYNP SATGKELGRTWSISYGDGSSASGI									
HPEP:	VDEQLENYLMEYFTFIGIGTFAQDTTVVFDTGSSMLWVPSV YCSLACTN ENRNFEDSST YCSTSEVMSITYGTS SMTGI									
	90	100	110	120	130	140	150	160	170	180
	LATDNVNLGGLLIKQTIELAKRE SSAPATD VIDGLLGLGFNFITTVRGVKTPV DNLISQGLISRPIFGVYLGKQSNNGGGGEYIFGGYDSSKFKGSI									
	LATDNVNLGGLLIKQTIELAKRE AASFANG ENDGLLGLGFNFITTVRGVKTPV DNLISQGLISRPIFGVYLGKASNGGGGEYIFGGYDSSKFKGSI									
	VFTDSMVGCVTARCAVQAQQI SAQRCQDTMNDGLLGLAFSSINTVPPQSQT FFDTVKSSLAQPLFAVALKHE--QGVYDFGIDSSKYTGSI									
	LGYDITVVGGISDTNIFGUSENPGSFLYYA PFDGTLGLAYPEI--SSSGATPFDNIWNQGLVSCDLEFVYL SADDQSGSVMIFGGIDSSVYTGSI									
	190	200	210	220	230	240	250	260	270	
	TTVPIDNSEGFVGVTVKIS IKIGGTVVSA SFDAILDTGTTLLLPDDV--AAKVARSYGASDNGDGTYSITIC DTSKIQPLVFTLGSSTFEVPSDSLII									
	TTVPIDNSRGLWGITVDR ATVGTSTVAS SFDGILDGTTLLILENNV--AASVARAYGASDNGDGTYSITIC DTSRFLPLVFTLGSSTFEVPSDSLIV									
	ITGVDNSGPFVSPVDS--YTAGSQSGD GFSGIADTGTTLLLEDVSVSQIYSVSGAQDSNAGQVFDG--SINLPDFSVSISG YDQATVH--GSLIN									
	NWVF--VTVEGTVKTTVDSITMNEAIAQNEGCCAIVDTGTSLITGFTSP--IANNISDIGASENSDCDMVVSASATSSPPDIYFTINGVQYVPVPSAYII									
	280	290	300	310	320					
	--FEKCGNKCIAG--F-----AAG GDLAILGDVFLKNNVVFVFNQGVPEVQIAPVAN									
	--FBEYQGCACIAG--F-----GYCNFDFAITGDFLKNNVVFVFNQGVPEVQIAPVAQ									
	YGPSSGSGSTCLGGIQS-----NSG IGFSIFGDIIFLKSQVVFVSDGPGGFAFQA									
	--LQSEF--SCISG--FQGMNLPTES--GELMLLGDVFTIRQYFTVFD RANNQVGLIAPVA									

FIG. 4. Comparison of the amino acid sequence of RNAP with those of rhizopuspepsin (RPEP) (5, 26), penicillopepsin (PPEP) (15), and human pepsin A (HPEP) (24). The numbers above the amino acid sequence are based on RNAP. Identical amino acids in two or more enzymes are boxed. Dashes indicate gaps introduced to obtain maximal homology.

	10	20		
RNAP	MKFTLISSCVLAAMTLAVEAA			
	* * * * *	* * * * *		
MPR	MLFSKISSAILLTAASFALTSR			

FIG. 5. Homology of the putative pre-enzyme sequence of RNAP with that of *M. pusillus* rennin (MPR) (27). Asterisks indicate conserved amino acids.

present in the four introns in the RNase Rh gene of *R. niveus* [H. Horiuchi, K. Yanai, M. Takagi, K. Yano, E. Wakabayashi, A. Sanda, S. Mine, K. Ohgi, and M. Irie, J. Biochem. (Tokyo), in press]. A similar finding was also reported for the four introns in the glucoamylase gene isolated from *R. oryzae* (1). It is known that there are eight introns in the genes of human pepsinogen A (24) and in the calf prochymosin gene (12), but their positions are different from those in the RNAP gene.

Although there is another methionine codon at -52, we speculate that a methionine codon at -66 may be the translation start point of RNAP for the following two reasons. Firstly, a sequence of about 20 amino acids between

TABLE 2. Codon usage of the RNAP gene in comparison with those of the rhizopuspepsin (5), *R. oryzae* glucoamylase (1), *M. pusillus* rennin (27), rat pepsinogen (14), and human pepsinogen A (24) genes.

Amino acid	Codon	No. of each codon in ^a :					
		RNAP	RPEP	RGAM	MPR	RAPP	HPEP
Leu	UUA	5	3	6	0	0	0
	UUG	6	7	1	12	4	1
	CUU	10	4	12	10	8	2
	CUC	6	8	10	6	16	7
	CUA	0	0	0	1	0	0
	CUG	0	0	0	2	15	25
Arg	CGU	2	6	7	2	1	1
	CGC	1	2	2	4	1	3
	CGA	0	0	1	0	0	0
	CGG	0	0	0	0	0	0
	AGA	2	1	3	0	0	2
	AGG	0	0	0	0	3	2
Pro	CCU	8	7	15	9	5	4
	CCC	2	4	4	6	10	14
	CCA	1	1	4	4	1	3
	CCG	1	0	0	0	1	0
Gln	CAA	6	11	9	7	3	0
	CAG	0	1	2	10	23	17
Lys	AAA	5	1	6	7	3	0
	AAG	14	12	25	15	8	8
Ala	GCU	11	12	42	13	2	4
	GCC	5	10	7	9	13	14
	GCA	5	0	8	7	1	1
	GCG	0	0	0	1	0	1
Val	GUU	13	9	13	13	2	4
	GUC	10	13	18	14	14	15
	GUA	0	0	4	3	4	1
	GUG	2	0	0	2	6	9
Gly	GGU	29	33	28	14	8	4
	GGC	7	10	6	23	27	21
	GGA	3	2	12	8	6	8
	GGG	0	0	0	0	8	3
Ser	UCU	13	11	34	14	11	9
	UCC	10	8	15	14	11	12
	UCA	2	0	11	0	2	3
	UCG	0	0	0	3	0	0
	AGU	4	2	8	2	8	5
	AGC	7	8	10	9	9	15
Thr	ACU	16	15	36	11	8	4
	ACC	6	11	15	19	16	20
	ACA	5	1	9	2	2	3
	ACG	1	2	1	0	0	1
Ile	AUU	7	5	16	7	7	4
	AUC	11	16	12	12	9	23
	AUA	1	0	1	0	1	0
Asn	AAU	4	5	20	0	3	2
	AAC	10	13	25	20	10	16
Phe	UUU	9	7	8	7	7	4
	UUC	7	10	21	22	15	13
Tyr	UAU	7	7	13	2	6	2
	UAC	5	7	22	17	16	16

Continued

TABLE 2—Continued

Amino acid	Codon	No. of each codon in ^a :					
		RNAP	RPEP	RGAM	MPR	RAPP	HPEP
Glu	GAA	11	6	11	5	4	3
	GAG	0	1	6	9	16	14
Cys	UGU	2	4	2	0	0	0
	UGC	2	0	1	4	7	7
His	CAU	1	0	1	0	0	0
	CAC	0	0	4	3	3	2
Asp	GAU	17	12	15	18	3	8
	GAC	8	11	12	16	13	14
Trp	UGG	3	4	12	2	6	7
Met	AUG	2	2	4	5	11	7

^a RAPP, *Rhizopuspepsin*; RGAM, *R. oryzae* glucoamylase; MPR, *M. pusillus* rennin; RPEP, rat pepsinogen; HPEP, human pepsinogen A.

–66 and –44 has the typical characteristics of signal peptides or preenzyme sequences (28); secondly, 14 amino acids between –66 and –52 are 50% homologous with amino acids in the presumptive signal peptide of *M. pusillus* rennin (Fig. 5). Thus, RNAP may have a preproenzyme sequence of 66 amino acids upstream of the mature sequence. A signal sequence or a preenzyme sequence may be about 20 amino acids long, and a proenzyme sequence may consist of about 46 amino acids. This proenzyme sequence, however, has low homology with those of human pepsinogen A, rat pepsinogen, calf prochymosin, *M. pusillus* rennin, and *M. miehei* rennin.

The codon usage of the RNAP gene was compared with those of the rhizopuspepsin gene, the *R. oryzae* glucoamylase gene, the *M. pusillus* rennin gene, the rat pepsinogen gene, and the human pepsinogen A gene (Table 2). Codon usage was rather similar among fungal genes but different between fungal genes and mammalian genes.

ACKNOWLEDGMENTS

We thank K. Nakaya and his colleagues of Showa University, Tokyo, Japan, for their help and advice in the determination of the amino acid composition and sequence and Amano Pharmaceutical Company for supplying Gluczyme.

LITERATURE CITED

- Ashikari, T., N. Nakamura, Y. Tanaka, N. Kiuchi, Y. Shibano, T. Tanaka, T. Amachi, and H. Yoshizumi. 1986. *Rhizopus* raw-starch-degrading glucoamylase: its cloning and expression in yeast. *Agric. Biol. Chem.* 50:957–964.
- Bott, R., E. Subramanian, and D. R. Davies. 1982. Three-dimensional structure of the complex of the *Rhizopus chinensis* carboxyl proteinase and pepstatin at 2.5-Å resolution. *Biochemistry* 21:6956–6962.
- Breathnach, R., and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349–383.
- Clarke, L., and J. Carbon. 1978. Functional expression of cloned yeast DNA in *Escherichia coli*: specific complementation of argininosuccinate lyase (*argH*) mutations. *J. Mol. Biol.* 120:517–532.
- Delaney, R., R. N. S. Wong, G. Meng, N. Wu, and J. Tang. 1987. Amino acid sequence of rhizopuspepsin isozyme p15. *J. Biol. Chem.* 262:1461–1467.
- Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C.

- O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot. 1980. The structure and evolution of the human beta-globin gene family. *Cell* 21:653-668.
7. Etoh, Y., H. Shoun, K. Arima, and T. Beppu. 1982. Photo-oxidation of a histidyl residue of milk-clotting acid protease, Mucor rennin. *J. Biochem. (Tokyo)* 91:747-753.
 8. Etoh, Y., H. Shoun, T. Ogino, S. Fujiwara, K. Arima, and T. Beppu. 1982. Proton magnetic resonance spectroscopy of an essential histidyl residue in a milk-clotting acid protease, Mucor rennin. *J. Biochem. (Tokyo)* 91:2039-2046.
 9. Fukumoto, J., D. Tsuru, and T. Yamamoto. 1967. Studies on mold protease. Part I. Purification, crystallization and some enzymatic properties of acid protease of *Rhizopus chinensis*. *Agric. Biol. Chem.* 31:710-717.
 10. Gray, G. L., K. Hayenga, D. Cullen, L. J. Wilson, and S. Norton. 1986. Primary structure of *Mucor miehei* aspartic protease: evidence for a zymogen intermediate. *Gene* 48:41-53.
 11. Grunstein, M., and D. S. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* 72:3961-3965.
 12. Hidaka, M., K. Sasaki, T. Uozumi, and T. Beppu. 1986. Cloning and structural analysis of the calf prothymosin gene. *Gene* 43:197-203.
 13. Hynes, M. J., C. M. Corrick, and J. A. King. 1983. Isolation of genomic clones containing the *amdS* gene of *Aspergillus nidulans* and their use in the analysis of structural and regulatory mutations. *Mol. Cell. Biol.* 3:1430-1439.
 14. Ichihara, Y., K. Sogawa, K. Morohashi, Y. Fujii-Kuriyama, and K. Takahashi. 1986. Nucleotide sequence of a nearly full-length cDNA coding for pepsinogen of rat gastric mucosa. *Eur. J. Biochem.* 161:7-12.
 15. James, M. N. G., and A. R. Sielecki. 1983. Structure and refinement of penicillopepsin at 1.8Å resolution. *J. Mol. Biol.* 163:299-361.
 16. Kurono, Y., M. Chidimatsu, K. Horikoshi, and Y. Ikeda. 1971. Isolation of a protease from a *Rhizopus* product. *Agric. Biol. Chem.* 35:1668-1675.
 17. Langford, C. J., and D. Gallwitz. 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell* 33:519-527.
 18. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 19. Nunberg, J. H., J. H. Meade, G. Cole, F. C. Lawyer, P. McCabe, V. Schweickart, R. Tal, V. P. Wittman, J. E. Flatgaard, and M. A. Innis. 1984. Molecular cloning and characterization of the glucoamylase gene of *Aspergillus awamori*. *Mol. Cell. Biol.* 4:2306-2315.
 20. Oka, T., and K. Morihara. 1973. Comparative specificity of microbial acid proteinases for synthetic peptides. I. Primary specificity. *Arch. Biochem. Biophys.* 156:543-551.
 21. Oka, T., and K. Morihara. 1974. Comparative specificity of microbial acid proteinases for synthetic peptides. Primary specificity with Z-tetrapeptides. *Arch. Biochem. Biophys.* 165:65-71.
 22. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature (London)* 263:211-214.
 23. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
 24. Sogawa, K., Y. Fujii-Kuriyama, Y. Mizukami, Y. Ichihara, and K. Takahashi. 1983. Primary structure of human pepsinogen gene. *J. Biol. Chem.* 258:5306-5311.
 25. Subramanian, E., I. D. A. Swan, M. Liu, D. R. Davies, J. A. Jenkins, I. J. Tickle, and T. L. Blundell. 1977. Homology among acid proteases: comparison of crystal structures at 3Å resolution of acid proteases from *Rhizopus chinensis* and *Endothia parasitica*. *Proc. Natl. Acad. Sci. USA* 74:556-559.
 26. Takahashi, K. 1987. The amino acid sequence of rhizopuspepsin, an aspartic proteinase from *Rhizopus chinensis*. *J. Biol. Chem.* 262:1468-1478.
 27. Tonouchi, N., H. Shoun, T. Uozumi, and T. Beppu. 1986. Cloning and sequencing of a gene for Mucor rennin, an aspartate protease from *Mucor pusillus*. *Nucleic Acids Res.* 14:7557-7569.
 28. von Heijne, G. 1985. Signal sequences. The limits of variation. *J. Mol. Biol.* 184:99-105.
 29. Wallace, R. B., M. J. Johnson, T. Hirose, T. Miyake, E. H. Kawashima, and K. Itakura. 1981. The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit beta-globin DNA. *Nucleic Acids Res.* 9:879-894.
 30. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33:103-119.