The extracellular acid protease gene of Yarrowia lipolytica: sequence and pH-regulated transcription

Thomas W. Young, Albert Wadeson, David J. Glover, Roger V. Quincey, Michael J. Butlin and Elizabeth A. Kamei

The gene encoding an acid extracellular protease (AXP) from Yarrowia lipolytica (Candida olea) 148 was cloned and the complete nucleotide sequence was determined. The amino acid sequence deduced from the nucleotide sequence reveals that the mature AXP consists of 353 amino acids with an M, of 37427. The gene also encodes a putative 17 amino acid hydrophobic prepeptide and a 27 amino acid propeptide containing no potential N-glycosylation sites. The mature extracellular enzyme is produced by cleavage between Phe and Ala. AXP is a member of the aspartyl family of proteases. AXP shows homology to proteases of several fungal genera and to human progastricin. The coding sequence is preceded by a potential regulatory region of 1982 bp. Transcription of both AXP and alkaline extracellular protease genes of Y. lipolytica 148 is regulated by the pH of culture.

Keywords: yeast, Yarrowia lipolytica, extracellular protease, aspartyl protease, pH regulation

INTRODUCTION

Yarrowia lipolytica (syn. Candida olea, Candida lipolytica, Saccharomycopsis lipolytica) is a heterothallic, ascospore-genous yeast which exhibits growth in budding, pseudomycelial and true mycelial forms. It has been used commercially for the production of citric acid and single-cell protein and may also be used to produce erythritol, mannitol and isopropylmalic acid. Although less well-developed than in Saccharomyces cerevisiae, the genetics of Yarrowia have been analysed and a genetic map is available (Heslot, 1990). Y. lipolytica is a natural secretor of proteins and offers advantages over Saccharomyces cerevisiae as a eukaryotic host for secretion of heterologous proteins such as bovine prochymosin (reviewed by Heslot, 1990, and Ogrydziak, 1993).

In common with many yeasts (reviewed by Ogrydziak, 1993), strains of Y. lipolytica secrete proteolytic enzymes showing optimal activity at acidic, neutral or alkaline pH values. Strain 148 secretes one acid and one alkaline protease (Nelson & Young, 1987), and the former enzyme has been considered for use in the brewing industry as a chill-proofing agent for beer (Nelson & Young, 1986). As in other strains (Ogrydziak, 1993), production of both proteolytic enzymes by strain 148 is subject to repression by nitrogen (ammonium ions, amino acids) and sulphur and may be induced by the presence of protein (Nelson, 1986).

The gene (xpr2) for the alkaline extracellular protease (AEP) from Y. lipolytica has been sequenced (Davidow et al., 1987a) and a complex potential upstream regulatory region identified. Analysis of the regulatory region (Blanchin-Roland et al., 1994) showed that a TATA box and two major upstream activation sequences were essential for promoter activity under conditions of repression or full induction. The sequences identified as important showed similarities to GCN4, TUF/RAP1 and CAR1 of Saccharomyces cerevisiae. The data of Blanchin-Roland et al. (1994) strongly suggested that the upstream activation sequence regions were targets for transcriptional factors required for assisting specific regulatory proteins.

Secretion of the AEP of Y. lipolytica CX161-1B most probably proceeds from a 55 kDa precursor produced after cleavage of a 15 amino acid signal peptide. This
precursor is then processed by a diaminopeptidase to generate a 52 kDa proprotein which is subsequently cleaved to give the mature 32 kDa secreted protein. This final step probably utilizes a KEX2-like protease (Ogrydziak, 1993). Sequence comparison between the prepro regions of Y. lipolytica AEP (Matoba et al., 1988; Davidow et al., 1987a) and the acid proteases of Saccharomyces fibuligera (Hirata et al., 1988), Candida albicans strains 10231 (Hube et al., 1991) and 10261 and Candida tropicalis (Togni et al., 1991) has led to the suggestion that there may be a common motif in processing yeast extracellular proteases (Ogrydziak, 1993). This motif is proposed to comprise a Lys-Arg (or Arg-Arg) pair of basic amino acids immediately preceding the N-terminus of the mature enzyme and two further potential common processing sites in the pro region. These sites, a Lys-Arg cleavage locus and a consensus glycosylation sequence, are in the same locations relative to the pro regions of the acid proteases. Whether or not these sites are used is unknown. In Y. lipolytica AEP, sites in the same order but with different spacings are seen. AEP is glycosylated in the pro region and it is possible that the secreted enzyme contains O-linked carbohydrate. Although the sequence of the mature protein contains two consensus N-linked glycosylation sites they are not used. The acid proteases of other strains are variously reported as being glycosylated in the secreted form.

Several authors have shown that environmental pH influences the activities of Yarrowia extracellular proteases (Ogrydziak, 1993). In general, high pH favours alkaline enzyme activity and low pH favours acid enzyme activity (Ogrydziak, 1993). Previous work in this laboratory showed that the accumulation of protease activity in the medium by Candida olea 148 (syn. Y. lipolytica) was influenced by environmental pH (Nelson & Young, 1987). The influence of environmental pH on enzyme synthesis and secretion is more readily measured in this strain because interference from both neutral proteases and a multiplicity of other proteases is absent. When strain 148 was grown on medium with protein as the sole source of nitrogen and sulphur, an acid (carboxyl) protease activity was produced in culture over the range pH 2-6 and an alkaline (serine) protease at pH 6-9 (Nelson & Young, 1987). However, none of the published work with this or other strains has established the biochemical basis for the pH effect. This could be facile, at the level of the known pH-dependence of enzyme activity, or involve other processes such as the secretory pathway, gene transcription or translation. The mechanism whereby environmental pH influences extracellular protease activity is important since if, for example, regulation of gene expression is involved this clearly has significance for mechanisms of cell signalling as well as providing a process for the pH-directed expression of secreted heterologous proteins.

In Aspergillus niger, PacC (a zinc-finger transcription factor) has been shown to mediate pH regulation of both acid- and alkaline-expressed genes by environmental pH (Tilburn et al., 1995). This transcription factor binds to a core hexanucleotide sequence in the promoter region of the alkaline-regulated isopenicillin N-synthetase structural gene (ipnA). The pacC gene product is activated under conditions of alkaline environmental pH by proteinase (Orejas et al., 1995). In A. niger, the pepA and pepB genes encoding acid extracellular proteases have been shown to be regulated at the level of gene transcription by the pH of the environment (Jarai & Buxton, 1994). An 18 bp sequence showing 83% similarity lies upstream of the transcription start of both these genes and may represent a locus for pH regulation of transcription. In the yeast C. albicans, environmental pH has been shown to influence the transcription of the messenger for a secreted aspartyl protease (SAP2). It is proposed that this effect, however, is dependent also on either the phase of growth or the presence of an inducing agent in the medium (Hube et al., 1994).

In this paper we report the cloning and sequencing of the gene and potential regulatory region for the secreted acid protease of the yeast Y. lipolytica strain 148, and show that the transcription of both the genes for acid and alkaline protease is regulated by the pH of the culture medium.

METHODS

Strains. Y. lipolytica 148 (syn. Saccharomyces lipolytica, C. lipolytica, C. olea; see Results) was maintained as described by Nelson & Young (1987). Genomic libraries were made in Escherichia coli DH5 (Hanahan, 1985) using the plasmid vector YCp50 (Johnson & Davis, 1984).

Maintenance and cultivation of micro-organisms. E. coli DH5 was maintained on LB broth or agar. Ampicillin (50 µg ml⁻¹) was included in the medium for maintenance of plasmids. Medium for maintenance of yeast was YM broth or agar (Difco).

For culture of yeasts to produce extracellular enzymes, a 16 l fermenter (Biolafitte) was used. Automatic monitoring and control were used to maintain an aeration rate of 300 r.p.m, temperature of 30 °C, pH of 4-5 (for acid protease production) or 6-5 (for alkaline protease production). The culture was grown with continuous aeration and the level of dissolved oxygen did not fall below 70% of saturation. The medium was 0-2% BSA, 1% (w/v) glucose, 0-143% yeast nitrogen base (YNB) without amino acids and ammonium sulphate (Gibco). Glucose solution was sterilized in situ in the fermenter. A concentrated aqueous solution (2 l) of YNB and BSA was sterilized by filtration through a 0-2 μm Sartobran-PH Capsule (Sartorius) and pumped directly into the sterile glucose solution.

Small batch cultures were used to analyse the effect of extracellular pH on the synthesis of enzyme transcripts. The medium contained 2% (w/v) BSA, 1% (w/v) d-glucose, 0-143% YNB. Acid pH was maintained using 0-5 M MOPS buffer, pH 5-5, and alkaline conditions were maintained using 0-5 M MOPS buffer, pH 7-5. One hundred millilitres of medium in 250 ml baffled flasks (Sigma) were used. d-Glucose (2 ml) was sterilized by autoclaving (110 °C, 30 min) and added aseptically to 38 ml sterile distilled water. BSA (50 ml of 4%) in 1 M buffer at the appropriate pH and 10 ml 1-43% (w/v) YNB was sterilized by filtration through a 0-2 μm membrane filter (Minisart, Sartorius) and transferred aseptically to the sterilized glucose solution. Flasks were inoculated with a 10 ml preculture (48 h, 28 °C, unshaken) and incubated in a gyratory shaker (200 r.p.m.) at 28 °C. Cultures were harvested at an A₆₀₀ value of 0-8-1-5 after 16 h incubation.
Yarrowia lipolytica extracellular acid protease

Yeast classification. A computer-assisted, yeast identification system (COMPASS) formerly offered by the National Collection of Yeast Cultures, Norwich, UK, was used to classify C. olea 148. The system comprises a probabilistic matrix prepared principally from data by Kreger-van Rij (1984). Morphological and physiological tests were conducted as described in the COMPASS manual.

Purification of extracellular proteases. The AXP of strain 148 was purified following culture at pH 4.5 with BSA as the nitrogen source (Nelson & Young, 1987). The methods used were as described by Nelson & Young (1987) with the addition of two further chromatographic steps using an FPLC system (Pharmacia LKB). The first was elution from a Pharmacia Mono Q ion-exchange column (HR5/5) at a flow rate of 1 ml min⁻¹. Sample was applied to the column in 0.01 M sodium citrate pH 6.0 and eluted with 4 ml of the same buffer and then with 20 ml linear gradient from 35% (v/v) of 0.01 M sodium citrate, 1 M NaCl pH 6.0. The active material was chromatographed on a column (10 x 1 cm) of Sephadex G25 to remove low-Mᵋ components and then freeze-dried.

A single AEP was also isolated and purified from cultures grown at pH 6.5. The purification steps described by Ogrydziak & Scharf (1982) were used with the additional steps of FPLC on Mono Q and Sephadex G25. The Mono Q column was eluted at a flow rate of 2 ml min⁻¹ with 5 ml buffer A (0.01 M Tris/HCl, pH 8.4), followed by 15 ml 7% (v/v) buffer B (0.01 M Tris/HCl, 1 M NaCl, pH 8.4) and a 10 ml linear gradient from 7 to 15% of buffer B.

Protein sequencing. This was conducted by the Amino Acid Sequencing Service of the BBSRC at the Department of Biochemistry, University of Leeds, Leeds, UK.

Synthesis of oligonucleotides. A mixed oligonucleotide (see Results) to a part of the N-terminal sequence (Fig. 1) of the secreted acid protease was synthesized by Alta Bioscience (an interfaculty service of the University of Birmingham). For probing the Y. lipolytica genomic library, 0.1 µmol mixed oligonucleotide was purified by butan-1-ol extraction and electroetion from a polyacrylamide sequencing gel (Sambrook et al., 1989). The oligonucleotide was labelled using poly-nucleotide kinase (BRL) and [γ-3²P]ATP (300 Ci mmol⁻¹/11·1 TBq mmol⁻¹; Amersham). The labelled nucleic acid was purified by electrophoresis from the radioactive band excised from a polyacrylamide gel; 1 ng oligonucleotide emitted 10⁷–10⁸ c.p.m.

Construction of genomic libraries of Y. lipolytica. Genomic DNA was isolated and purified using methods based on those published by Beggs (1978) and Rigginsby et al. (1982). A partial BamHI digest of genomic DNA was cloned into the unique BamHI site of YCp50. A library of 46600 clones was made in E. coli DH5α; 32600 clones of this library were probed by colony hybridization (Sambrook et al., 1989) with the labelled mixed oligonucleotide. A second library of a complete digest of Yarrowia genomic DNA with BglII was also made in YCp50.

DNA sequencing. Restriction fragments of DNA of positive clones isolated from the first library were subcloned into pUC9. These subclones were sequenced by the dideoxy method (Sambrook et al., 1989) using the M13 cloning system (Life Technologies; Messing, 1983) and universal primers. DNA from the second library was also sequenced by the dideoxy method using M13. In this case, however, oligonucleotide primers made to both the coding and the non-coding strands were used in the sequencing reactions. All sequence information reported has been obtained from the independent analysis of both strands.

Computer analysis of DNA sequences. Most analysis was conducted using versions 7 and 8 of the University of Wisconsin Genetics Computer Group (UWGCG) software on a VAX computer at the University of Birmingham. The same software and current databases were also used at Sequet, Daresbury, UK, via the Joint Academic Network.

Isolation of total RNA. The methods used were based on those described by Quincey & Arnold (1984) modified from those reported by Berger (1987a, b) and Kohrer & Domdey (1991). All glassware and spatulas were baked overnight at 160 °C prior to use. Solutions were treated with 0.1% diethyl pyrocarbonate (DEPC) overnight and then autoclaved for 15 min. Solutions of heat-labile materials were made up in DEPC-treated water and all chemicals were handled with baked spatulas.

Cultures were grown in shake flasks and 20 ml culture (or a volume equivalent to 0.5 g centrifuged weight of cells) were poured through crushed ice into a sterile, polypyrrole 50 ml centrifuge tube. The tube was kept at 4 °C. The pellet was washed by resuspension in 20 ml cold DEPC-treated water and collected as above. To the pellet, 2 g baked, acid-washed, 0.45 mm glass beads (Sigma) and 2 ml AC I (sterile-filtered 50 mM sodium acetate, pH 5.2, 10 mM EDTA pH 8.0, 0.2% heparin containing 10 mM ribonucleoside vanadyl complexes; Sigma) were added. The whole was mixed vigorously on a vortex mixer for 8 x 30 s periods with intermittent 30 s periods of cooling on ice. Two millilitres ACE II (50 mM sodium acetate, pH 5.2, 10 mM EDTA pH 8.0, 5% SDS, 0.1% 2-mercaptoethanol, 10 nM ribonucleoside vanadyl complexes) and 4 ml phenol equilibrated to pH 4.5 (Sigma) were added. The whole was mixed vigorously on a vortex mixer for 4 x 30 s periods with intermittent 30 s periods of cooling on ice. The phases were separated by centrifugation, the organic phase was removed and the aqueous and inter-phases were re-extracted with an equal volume of phenol/chloroform by vigorous mixing and cooling as before. The combined aqueous phases were re-extracted with phenol/chloroform. This process was repeated until no white interphase remained and then one further phenol/chloroform extraction was done followed by a final chloroform extraction. Finally, the aqueous layer was transferred to a fresh tube and the nucleic acid was precipitated by adding 0.1 vol. of 3 M sodium acetate, pH 5.2, and 2.5 vols cold absolute ethanol, mixing and leaving overnight at -20 °C.

Fractionation of RNA through gels containing formaldehyde. The steps outlined above for the elimination of ribonuclease activity from materials and equipment were observed. In addition electrophoresis tanks, gel trays and combs were treated with DEPC overnight and rinsed with DEPC-treated water prior to use.

A volume of resuspended, ethanol-precipitated RNA (containing about 10 µg nucleic acid) was centrifuged at 4 °C and the pellet was washed by suspension and re-centrifugation in cold 70% (v/v) ethanol and finally dried under vacuum. The pellet was dissolved in 3.5 µl water and 3.5 µl 5 M lithium chloride were added. The whole was left either on ice for at least 1 h or overnight at -20 °C to precipitate RNA. The purified nucleic acid was collected by centrifugation at 4 °C, washed by suspension and centrifugation in cold 70% ethanol and dried under vacuum. The pellet was dissolved in 3 µl water and 15 µl deionized formamide buffer (500 µl deionized formamide, 100 µl 1X MOPS buffer, 150 µl 38% formaldehyde; this buffer was stored at -20 °C). MOPS buffer (1X) was made from a 5X stock of 0.1 M MOPS (pH 7.0), 40 mM sodium acetate and 5 mM EDTA (pH 8.0). Deionized formamide was made by stirring 1 g AG 501-X8 mixed bed resin (Bio-Rad) with 20 ml formamide. Deionized formamide was stored at -20 °C in airtight cryovials.
The nucleic acid was denatured by incubation at 60 °C for 15 min, cooled on ice and mixed with 3 μl 6× RNA loading buffer: 5% (v/v) glycerol, 1 mM EDTA (pH 8.0), 0.25% bromophenol blue, 0.25% xylene cyanol FF. Samples were electrophoresed in 15× agarose-formaldehyde gels (Sambrook et al., 1989). Tanks with large buffer reservoirs (Pharmacia GNA 280) were used and electrophoresis was at 34 V cm⁻¹. Half way through the run, electrophoresis was stopped and the electrophoresis buffer (1× MOPS) was removed from the anode and cathode reservoirs, mixed, returned and the electrophoresis was re-started. The lane containing markers was cut from the gel and stained in 0.2 μg ethidium bromide ml⁻¹ for 25 min. The gel was washed by three consecutive 10 min washes with water to remove formaldehyde, and photographed using transmitted UV illumination. The remainder of the gel was neutralized by soaking in 10% (w/v) glycine for 25 min and processed for Northern blotting.

Northern blotting. The methods described by Sambrook et al. (1989) were used. Probes were made by PCR (30 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 60 s) using 5' and 3' primers to the published gene sequences of Y. lipolytica AEP (Davidow et al., 1987a). The primers were made to bases 2356–2379 (5') and 2590–2613 (3'). A probe for the AEP from Y. lipolytica (this paper) was made from a 255 bp BamH1–NcoI fragment obtained from a 500 bp BamH1–SalI fragment of clone 13.1 subcloned in pUC9. The probe extends from the 5' BamH1 site to the first NcoI site. Probes were purified by gel electrophoresis and electroelution.

**RESULTS AND DISCUSSION**

Classification of *C. olea* 148

The identity score (maximum value 1.000000) was 0.9999722 with *Y. lipolytica*. No other yeast in the database gave scores close to this value. Taken together with the sequence data for the alkaline protease and the codon usage (see below), it is clear that *C. olea* 148 is indeed a strain of *Y. lipolytica*.

**Purification of extracellular proteases and N-terminal sequences**

A single acid protease with an estimated *M*ₚ by SDS-PAGE of 39000 was purified to greater than 98% purity as judged by SDS-PAGE. The enzyme eluted from Mono Q in a single peak of activity 19 ml into the salt gradient, and 30 μg was used for obtaining the N-terminal sequence. The alkaline protease eluted from the Mono Q column after 15 ml of salt gradient and was greater than 95% pure as judged by SDS-PAGE, and 120 μg was used for N-terminal sequencing. The N-terminal sequences of both enzymes are shown in Fig. 1.

For the alkaline protease, the N-terminal amino acid and residues at positions 4, 18, 19 and 48 were not unequivocally characterized. However, data (not shown) indicated that the most probable residue at both 4 and 48 is Thr. With these assignments, 49 of the 52 residues are identical to those reported by Davidow et al. (1987a) for the AEP from *Y. lipolytica*. In the latter enzyme, positions 18 and 19 are occupied by Lys. Given the fact that difficulties are frequently experienced in sequencing Lys, it is reasonable to assume that both 18 and 19 are Lys residues in the *Y. lipolytica* 148 sequence, and it seems likely that this organism secretes the same AEP as other strains of *Y. lipolytica*.

The sequence obtained from the acid protease was unique when compared to proteins in the SWISS-PROT database.
Fig. 3. Complete sequence of \textit{Y. lipolytica} cloned DNA containing the gene for the AXP. Regions of repeated sequence are underlined. The probable transcription start site is shown by the arrow. The coding region for the AXP gene begins at 1983 and ends at 3176; the first, second and third stop codons are shown in bold and underlined. The extreme 5' translation is a putative open reading frame. No other ORFs of significant size were detected. The translated sequence is shown as single-letter code. Underlined amino acids: Ala$_1$, and Ala$_{16}$, predicted site of signal cleavage; Phe$_{44}$ and Ala$_{45}$ mature sequence cleavage site.

using the UWGGCG \textsc{tfasta} program. It did however show 44.4\% identity in a 36 amino acid overlap with an aspartyl protease, endothiapepsin, from the chestnut blight fungus, \textit{Cryptococcus parasitica} (\textit{Endothia parasitica}) (Razanamparany \textit{et al.}, 1992).

**Gene cloning**

A mixed 18-mer oligonucleotide was synthesized corresponding to amino acids Trp$_3$ to Asp$_{45}$ of the purified acid protease (Fig. 1). The mixture comprised 16 oligonucleotides (only the forms with T at the 3' end were made). This was labelled and used to probe the \textit{Bam}HI library made in \textit{E. coli} DH5, and from the first screen, 15 presumptive positive clones were isolated. These were re-screened and four positives were obtained. These four were sequenced and only one (designated 13.1) contained sequence identical to both the probe and the N-terminal sequence of the secreted AXP. The 5' end of the cloned DNA corresponded to residue Gly$_3$ of the N-terminal protein sequence.

To obtain the regions 5' to clone 13.1, restriction digests of \textit{Yarrowia} genomic DNA were probed with a 329 bp \textit{Bam}HI–\textit{Bgl}II fragment from the 5' end of the clone (see Fig. 2). This screen identified an approximately 2.5 kb \textit{Bgl}II fragment which would contain the N-terminus of the enzyme and upstream sequences. This fragment was isolated from a \textit{Bgl}II library by colony hybridization using the 329 bp \textit{Bam}HI–\textit{Bgl}II fragment of clone 13.1 as a probe. The isolated clone was designated 5.2.

**Gene sequencing**

The sequencing strategies for clones 13.1 and 5.2 are shown in Fig. 2. The 5' end of 5.2 overlaps the 5' end of 13.1. Both strands were sequenced in both directions and all segments overlap. The complete sequence of 3478 bp of \textit{Yarrowia} DNA including the coding sequence and its translation are presented in Fig. 3. Computer analysis...
shows a large open reading frame extending from 1983 to 3173. Within this open reading frame lies the N-terminus of the mature secreted AXP beginning with Ala at 2115 bp. The first stop codon is at 3174 bp. The mature enzyme is therefore 353 amino acids in length.

**Codon usage**

Values calculated for codon usage in the AXP are presented in Table 1 together with combined values for three highly expressed *Y. lipolytica* genes, XPR2 (Davidow et al., 1987a), LEU2 (Davidow et al., 1987b) and PYK1 (Strick et al., 1992). There is very good agreement in terms of preferred codon usage throughout. CGT (Arg) and ATA (Ile) remain unused and TGA was not found as the first stop codon.

**Comparison with other proteins**

Comparison of the sequence of mature AXP with proteins in the database using the programFASTA gave a best score with the candidapepsin 1 precursor of *C. albicans*. Table 2 shows the results of the analysis for the 22 best scores. With the exception of progastricin precursors from human and Macaque monkey and a cathepsin E precursor from rabbit, all other matches are to fungal enzymes. The majority are from the yeast genus Candida; *C. albicans* secreted aspartyl proteases (candidapepsins precursors 1–7) and enzymes from *Candida parapsilosis* and *C. tropicalis*. The precursor of the acid protease from the yeast *Saccharomyces fibuliger* (taxonomically the most closely related yeast) is eleventh in the list. Related enzyme precursors are also found in the fungal genera.
Table 2. Proteins in the SWISS-PROT database showing similarity to the secreted AXP of Y. lipolytica 148

<table>
<thead>
<tr>
<th>SWISS-PROT</th>
<th>Accession no.</th>
<th>Description</th>
<th>Organism</th>
<th>Identity (%)</th>
<th>Overlap (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sw:car1_canal</td>
<td>P28872</td>
<td>Candida pep81 1 precursor (EC 3.4.23.24) (ACP)</td>
<td>Candida utilis</td>
<td>26:6</td>
<td>353</td>
</tr>
<tr>
<td>sw:car2_canal</td>
<td>P28871</td>
<td>Candida pep81 2 precursor (EC 3.4.23.24) (ACP)</td>
<td>Candida utilis</td>
<td>26:2</td>
<td>370</td>
</tr>
<tr>
<td>sw:yap3_yeast</td>
<td>P32329</td>
<td>Aspartic proteinase 3 precursor (EC 3.4.23 ...)</td>
<td>Saccharomyces cerevisiae</td>
<td>22:8</td>
<td>395</td>
</tr>
<tr>
<td>sw:car4_canal</td>
<td>P43093</td>
<td>Candida pep81 4 precursor (EC 3.4.23.24)</td>
<td>Candida utilis</td>
<td>28:6</td>
<td>357</td>
</tr>
<tr>
<td>sw:car7_canal</td>
<td>P43096</td>
<td>Candida pep81 7 precursor (EC 3.4.23.24)</td>
<td>Candida utilis</td>
<td>23:7</td>
<td>211</td>
</tr>
<tr>
<td>sw:pep_c_hum</td>
<td>P20142</td>
<td>Progastricin precursor (EC 3.4.23.3) (pepsinogen C)</td>
<td>Homo sapiens</td>
<td>22:4</td>
<td>330</td>
</tr>
<tr>
<td>sw:pep_maccu</td>
<td>P03955</td>
<td>Progastricin precursor (EC 3.4.23.3) (pepsinogen C) (fragment)</td>
<td>Macaca fascicularis</td>
<td>22:7</td>
<td>330</td>
</tr>
<tr>
<td>sw:car3_rhini</td>
<td>Q03699</td>
<td>Rhizopus pep8in 3 precursor (EC 3.4.23.21)</td>
<td>Rhizopus niveus</td>
<td>25:2</td>
<td>337</td>
</tr>
<tr>
<td>sw:car5_canal</td>
<td>P43094</td>
<td>Candida pep81 5 precursor (EC 3.4.23.24)</td>
<td>Candida utilis</td>
<td>26:8</td>
<td>355</td>
</tr>
<tr>
<td>sw:car1_rhini</td>
<td>P10602</td>
<td>Rhizopus pep8in 1 precursor (EC 3.4.23.21)</td>
<td>Candida utilis</td>
<td>24:0</td>
<td>371</td>
</tr>
<tr>
<td>sw:car_pacni</td>
<td>P22929</td>
<td>Acid protease precursor (EC 3.4.23 ...)</td>
<td>Saccharomyces fibuligera</td>
<td>26:7</td>
<td>345</td>
</tr>
<tr>
<td>sw:car6_canal</td>
<td>P43095</td>
<td>Candida pep81 6 precursor (EC 3.4.23.24)</td>
<td>Candida utilis</td>
<td>27:8</td>
<td>353</td>
</tr>
<tr>
<td>sw:car3_canal</td>
<td>P43092</td>
<td>Candida pep81 3 precursor (EC 3.4.23.24)</td>
<td>Candida utilis</td>
<td>26:5</td>
<td>351</td>
</tr>
<tr>
<td>sw:carv_canal</td>
<td>P10977</td>
<td>Candida pep81 1 precursor (EC 3.4.23.24) (ACP)</td>
<td>Candida utilis</td>
<td>24:3</td>
<td>342</td>
</tr>
<tr>
<td>sw:carv_rhich</td>
<td>P06026</td>
<td>Rhizopus pep8in 1 precursor (EC 3.4.23.21) (fragment)</td>
<td>Rhizopus chinensis</td>
<td>23:5</td>
<td>362</td>
</tr>
<tr>
<td>sw:car1_canpa</td>
<td>P32951</td>
<td>Candida pep81 1 precursor (EC 3.4.23.24) (ACP)</td>
<td>Candida parapsilosis</td>
<td>25:0</td>
<td>360</td>
</tr>
<tr>
<td>sw:carp_cantr</td>
<td>Q00663</td>
<td>Candida pep8in 1 precursor (EC 3.4.23.24) (ACP)</td>
<td>Candida tropicalis</td>
<td>23:9</td>
<td>368</td>
</tr>
<tr>
<td>sw:carp_rhich</td>
<td>P09177</td>
<td>Mucoropepsin precursor (EC 3.4.23.23) (Mucor rennin)</td>
<td>Rhizomucor pusillus</td>
<td>21:6</td>
<td>338</td>
</tr>
<tr>
<td>sw:cate_rabid</td>
<td>P43139</td>
<td>Cathepsin E precursor (EC 3.4.23.34)</td>
<td>Orceinococcus candidatus</td>
<td>24:1</td>
<td>336</td>
</tr>
<tr>
<td>sw:ssa1_tichi</td>
<td>P32834</td>
<td>Aspartic protease Sxal precursor (EC 3.4.23 ...)</td>
<td>Schizosaccharomyces pombe</td>
<td>17:7</td>
<td>317</td>
</tr>
<tr>
<td>sw:carp_rhimi</td>
<td>P00799</td>
<td>Mucoropepsin precursor (EC 3.4.23.23) (Mucor rennin)</td>
<td>Rhizomucor melilii</td>
<td>24:2</td>
<td>219</td>
</tr>
<tr>
<td>sw:carp_crypa</td>
<td>P11838</td>
<td>Endothiapepsin precursor (EC 3.4.23.22)</td>
<td>Cryptococcus parastisati</td>
<td>28:4</td>
<td>317</td>
</tr>
</tbody>
</table>

Rhizopus and Cryptococci. Fig. 4 shows a comparison, using the program clustal v (Higgins et al., 1991) of the amino acid sequence of prepro AXP with the sequences of six of the proteins of Table 2. In common with these other proteases, AXP contains two active site Asp/Thr/Gly motifs characteristic of aspartyl proteases. The active site Asp residues are marked by arrows in Fig. 4. No homology is detected between the prepro region of AXP and the other proteins.

Although AXP apparently differs in M_r, inhibition by diazocetyl-norleucine methyl ester and 1,2-epoxy-3(p-nitrophenoxy)-propane (Nelson & Young, 1986) from the enzymes from Y. lipolytica CX161-1B (Yamada & Ogrydziak, 1983), sequence from clone 13.1 has been used to successfully disrupt (by integrative transformation) acid protease production in Y. lipolytica (C. Gaillardin, personal communication). It seems highly likely therefore that the enzyme is produced by other Yarrowia strains. It is not clear at present, however, which if any of the three AXP's reported for strain CX161-1B correspond to AXP from strain 148.

The preproenzyme, glycosylation and processing

The preproenzyme comprises 397 residues and has an M_r of 42081. The first 18 residues of the amino terminus are all significantly hydrophobic. Using the weight matrix for eukaryotic signal peptide cleavage sites of von Heijne (1986), the predicted site is between Ala_14 and Ala_18. This site shows a P value of 114. Alternative but much less likely sites are given at Ala_13/Phe_14, Ala_18/Pro_19 and Ala_20/Asn_21, with P values of 1-8, 0-6 and 2-3, respectively. There are no N-glycosylation sites in the pro region of the protein and the mature enzyme is produced by cleavage between Phe_14 and Ala_18, and comprises 353 amino acids and has a calculated M_r of 37427 and P I of 4-05 which agrees well with the biochemical estimates (39000 and 4-1, respectively). Comparison of the pro region of AXP with those of other secreted yeast proteases (acid proteases from C. albicans strains 10231 and 10261, C. tropicalis, Saccharomyces fibuligera and the AEP of Y. lipolytica) reveals striking differences. It is much shorter, it contains no N-linked glycosylation sequences and lacks a Lys-Arg cleavage site immediately preceding the N-terminus of the
secreted protein. It appears therefore that the secretory motif of *Y. lipolytica* acid protease is distinct from the common motif for yeast extracellular protease processing suggested by Ogrydziak (1993). Of particular note is the lack of the Lys-Arg (or Arg-Arg) processing site in the propeptide.

**Sequence upstream and downstream of the AXP gene**

Computer analysis indicates ORFs in each of the three reading frames from the 5' end (the largest is from 3 to 111, Fig. 3). None of the translated sequences show convincing homology to other sequences. Accordingly, the AXP gene potentially contains a 1982 bp upstream regulatory region. Analysis of mRNA by 5' extension (data not shown) indicates that the transcription start is at position 1957.

No sequence matching the consensus for pH regulation reported for *A. niger* (Jarai & Buxton, 1994) is present. However, in the AXP sequence, two hexanucleotide sequences corresponding to the core recognition sequence of the ipmA promoter (Tilburn et al., 1995) are found. The core sequence 5' GCCARG 3' is found at positions 1100 and 1121 on the top strand.

Downstream of the coding region, second and third stop codons are found at positions 3189 and 3231, respectively (Fig. 3). No consensus polyadenylation site has been detected. The estimated size of the transcript is thus at least 1217 bases.

**Effect of environmental pH on transcription of AEP and AXP**

Total RNA, isolated from strain 148 grown at pH 5.5 and 7.5 in buffered batch culture, was separated by electrophoresis under denaturing conditions and probed with labelled probes specific to AEP and AXP sequences. Transcript sizes were estimated as 1350 bases for AXP and 1550 for AEP (Fig. 5). The results presented in Fig. 5 also show that under acid conditions only AXP transcript is detected and under alkaline conditions only AEP. That the observations with regard to AXP transcript result from pH and not the influence of buffering ions was shown by growing the cells in culture at pH 4.5 with pH control (by automated acid or base addition in response to ipnA promoter) (Tilburn et al., 1995) are found. The core sequence 5' GCCARG 3' is found at positions 1100 and 1121 on the top strand.

Downstream of the coding region, second and third stop codons are found at positions 3189 and 3231, respectively (Fig. 3). No consensus polyadenylation site has been detected. The estimated size of the transcript is thus at least 1217 bases.
is that environmental pH regulates the transcription apparatus. Precisely how this is achieved remains to be determined; it may, as in *Aspergillus*, involve a PacC-type gene product. However if it does, the regulatory events in *Yarrowia* must be more complex since the AEP promoter does not contain a pacC core recognition sequence. Alternatively an as-yet-unidentified transcription factor or factors with different specificities may be involved.

REFERENCES


Received 9 April 1996; revised 28 May 1996; accepted 31 May 1996.