MOLECULAR AND CELLULAR BIOLOGY, Dec. 1987, p. 4390–4399
0270-7306/87/124390-10$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 7, No. 12

Protease B of the Lysosomelike Vacuole of the Yeast *Saccharomyces cerevisiae* Is Homologous to the Subtilisin Family of Serine Proteases

CHARLES M. MOEHLE,[1]* RICHARD TIZARD,[2] SANDRA K. LEMMON,[1] JOHN SMART,[2] AND ELIZABETH W. JONES[1]

Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213,[1] and Biogen, Cambridge, Massachusetts 02142[2]

Received 22 May 1987/Accepted 2 September 1987

The *PRB1* gene of *Saccharomyces cerevisiae* encodes the vacuolar endoprotease protease B. We have determined the DNA sequence of the *PRB1* gene and the amino acid sequence of the amino terminus of mature protease B. The deduced amino acid sequence of this serine protease shares extensive homology with those of subtilisin, proteinase K, and related proteases. The open reading frame of *PRB1* consists of 635 codons and, therefore, encodes a very large protein (molecular weight, >69,000) relative to the observed size of mature protease B (molecular weight, 33,000). Examination of the gene sequence, the determined amino-terminal sequence, and empirical molecular weight determinations suggests that the preproenzyme must be processed at both amino and carboxy termini and that asparagine-linked glycosylation occurs at an unusual tripeptide acceptor sequence.

The lysosomelike vacuole of the yeast *Saccharomyces cerevisiae* contains a number of the major hydrolases of the cell, including protease A, protease B, carboxypeptidase Y, the large aminopeptidase, repressible alkaline phosphatase, and at least one RNase (39, 42, 75). Most, if not all, of the vacuolar hydrolases are glycoproteins that are synthesized first as inactive precursors (2, 27). Like externally secreted proteins, precursors to vacuolar hydrolases pass through the endoplasmic reticulum and the Golgi body. The vacuolar hydrolase precursors are sorted from secreted proteins within the Golgi body and are transported to the vacuole (62).

Protease B, a single-subunit glycoprotein, is a serine endoprotease with a pH optimum near neutrality. Estimates of molecular weight (MW) range from 31,000 to 44,000 (8, 17, 23, 25, 31, 41, 56, 70); however, the larger size estimates may be incorrect, owing to the affinity of endogenous peptide inhibitors for protease B (25). In kinetic experiments, a 20-min pulse of spheroplasts with [^{35}S]methionine allowed the detection of a larger, 42,000-MW precursor (44). Processing of this precursor to a 33,000-MW mature enzyme requires the action of protease A, the product of the *PEP4* gene (3, 45, 77). Carbohydrate comprises about 10% of the 33,000-MW mature protease B, with mannose and amino sugars being present in a 15:2 ratio (31). The composition and size of the carbohydrate moiety are about those expected for the endoglycosidase H (endo H)-sensitive, asparagine (Asn)-linked oligosaccharide chains found on other yeast vacuolar hydrolases (44, 69). However, although the addition of tunicamycin results in the synthesis of a smaller, ca. 39,000-MW precursor to protease B, it has no effect on the activity or size of the mature enzyme (44), and treatment with endo H does not affect the size of the protease B molecule (44; data not shown; see references 2 and 27 for reviews). Since *Saccharomyces cerevisiae* does not make glycosidic side chains of the complex type found in animal cells, the resistance to endo H presumably indicates that the carbohydrate is not Asn linked and may be linked to hydroxyl groups (see reference 5 for a review).

We have recently isolated the structural gene for protease B, *PRB1*, and have described its transcription pattern (47). We report here the DNA sequence and DNA-derived amino acid sequence of the gene and gene product and our finding that protease B belongs to the subtilisin family of serine proteases. A pathway for the processing of the preproenzyme is proposed.

MATERIALS AND METHODS

Materials. Concanavalin A-Sepharose 4B, α-methylmannoside, and phenylmethylsulfonyl fluoride (PMSF) were purchased from Sigma Chemical Co., St. Louis, Mo.; DEAE and carboxymethyl cellulose were purchased from Whatman, Inc., Clifton, N.J.; Sephadex G-100 was purchased from Pharmacia, Inc., Piscataway, N.J.; YM-10 membrane and Centricon 10 filtration devices were purchased from Amicon Corp., Danvers, Mass.; bakers' yeast was purchased from Red Star Yeast, Pittsburgh, Pa.

DNA sequencing. Cloning of the *PRB1* gene has already been described (47). The DNA was sequenced across both strands by the method of Maxam and Gilbert (43).

Protein purification and sequencing. Protease B was purified from bakers' yeast by the "alternative" method of Huse et al. (25) followed by concanavalin A-Sepharose 4B affinity chromatography. Bakers' yeast (1 kg) was extracted with chloroform, and after adjustment of the pH to 7.0, the extract was incubated overnight at room temperature. After precipitation with 50% ammonium sulfate, the supernatant solution was adjusted to pH 5 and incubated overnight at 4°C. Following centrifugation, the retained supernatant solution was made 90% in ammonium sulfate. The precipitated sample was dialyzed against buffer A (0.01 M sodium phosphate [pH 7.0], 0.1 M NaCl) and applied to a DEAE-cellulose column equilibrated in buffer A. The breakthrough fraction was concentrated with 90% ammonium sulfate, dialyzed against buffer A, and batch extracted with carboxymethyl cellulose as previously described (25). The

* Corresponding author.

unbound enzyme solution was concentrated by ultrafiltration (YM-10 membrane) to 6 to 9 ml. This sample was applied to a Sephadex G-100 column (2.5 by 85 cm) and eluted with buffer A at 15 to 20 ml/h. The active fractions were pooled and loaded directly onto a column of concanavalin A-Sepharose 4B equilibrated in buffer A. The column was washed extensively with 0.01 M sodium phosphate (pH 7.0)–0.5 M NaCl until the A_{280} was <0.03. Bound protein was eluted with the same buffer containing 0.5 M α-methylmannoside. Activity was assayed by using the Azocoll assay (28), and then 1 mM PMSF was added to active fractions to prevent autoproteolysis. Portions corresponding to some of the most active fractions were pooled, concentrated to approximately 1 ml by using a Centricon-10 microconcentrator, and adjusted to 0.1% sodium dodecyl sulfate (SDS). This mixture was dialyzed extensively against 0.1% SDS–1 mM PMSF and lyophilized. Protein was electroeluted (24) from a preparative SDS-polyacrylamide gel and sequenced with a gas-liquid solid-phase sequenator (model 470A; Applied Biosystems) (22).

Computer analysis of protein sequence. The preproprotease B amino acid sequence deduced from the *PRB1* DNA sequence was compared with all of the sequences in the National Protein Information Resource by using the FASTP program of Lipman and Pearson (40). Hydropathy predictions were determined by the method of Kyte and Doolittle (37) by using the Pustell Sequence Analysis Package from International Biotechnologies, Inc. Protein secondary structure predictions were determined by the method of Chou and Fasman (11) by using a program written by William E. Brown, Carnegie Mellon University.

RESULTS

Sequence of the gene. The sequence of 2,387 bases of the DNA that encodes protease B is shown in Fig. 1. We previously reported that an uninterrupted, 2.3-kilobase mRNA is transcribed from this region (47). The 5′ and 3′ ends of the mRNA, determined by S1 nuclease mapping (7), align at the positions shown on the sequence in Fig. 1.

The open reading frame (ORF) contained within the relevant DNA sequence is 1,905 base pairs long and encodes a 635-amino-acid polypeptide. The calculated MW of the deduced peptide is 69,619. This peptide is much larger than the only reported protease B precursor of 42,000 MW, which has been identified in pulse-chase experiments (44). This precursor accumulates in a *pep4* mutant, which also is known to accumulate a larger precursor of another vacuolar hydrolase, carboxypeptidase Y (21, 45). When N-linked glycosylation is blocked by the addition of tunicamycin, the MW of the protease B precursor is reduced to 39,000 (44).

The first 19 residues of the deduced polypeptide sequence (codons −280 to −262; by convention, the amino-terminal residue of the mature protease is designated codon 1) contain characteristics common to known hydrophobic leader sequences (72). This is not surprising, since protease B, like the vacuolar hydrolases carboxypeptidase Y (62) and protease A (S. Garlow and E. Jones, unpublished data), probably passes through the endoplasmic reticulum and Golgi body in transit to the vacuole. The remaining 261 amino acids between the leader and the N terminus (described below) are presumably removed during processing of the preproenzyme. Since an *Eco*RI deletion that fuses the codon for Glu −200 to the codon for Glu 332 in frame leads to no obvious phenotypic change beyond protease B deficiency (47), we infer that this long *pro* region probably has no other function. The predicted polypeptide contains five potential acceptor sites for Asn-linked glycosylation, four of which occur between the N terminus of the mature protein and the end of the ORF. Four have the sequence Asn-X-Ser at Asn −68, Asn 107, Asn 143, and Asn 181, and one has the sequence Asn-X-Thr at Asn 314.

Purification and N-terminal sequence of the protein. To enable the identification of the amino acid residues within the *PRB1*-encoded polypeptide that correspond to the N terminus of the mature enzyme, we purified protease B from bakers' yeast by the alternative method of Huse et al. (25), except that purification over a concanavalin A-Sepharose 4B affinity column proved necessary to obtain a homogeneous protein. A typical profile for the Sephadex G-100 column and the stained gel from SDS-polyacrylamide gel electrophoresis (38) of the final PMSF-inactivated protein preparation used for N-terminal sequencing are shown in Fig. 2. The MW estimated from the column was 25,000, and that estimated from SDS-polyacrylamide gel electrophoresis was 32,000. The discrepancy in these two MW estimates probably reflects ionic interaction between protease B and Sephadex, as has been observed with homologous enzymes (10; compare references 13 and 26). Sequencing of the N terminus of this protein yielded the 15-amino-acid sequence underlined in Fig. 1, starting at amino acid 1. Immunoblots of rapidly extracted *S. cerevisiae* cells prepared by using antibody raised to the 14-amino-acid N-terminal peptide revealed an antigen of about 31,000 MW, indicating that the enzyme purified by the procedure described above is probably authentic protease B (data not shown). The polypeptide encoded by the *PRB1* ORF from codon 1, corresponding to the mature N terminus, to codon 356, the stop codon, has a predicted MW of 38,140. When 2,900 kiladaltons of carbohydrate is added to that (31), mature protease B has a predicted MW of 41,000, in contrast to the observed MW of 32,000 to 33,000.

The sequenced *PRB1* gene was cloned from strain AB320 (50, 51), a laboratory strain that is related to strain X2180 and that has bakers' yeast in its pedigree (48), while the protein was purified from bakers' yeast. The 15-amino-acid sequence we determined is a perfect match of the DNA-derived amino acid sequence for amino acids 1 to 15. It has been previously reported that the amino acid compositions for the proteases B from strain X2180 and bakers' yeast are quite comparable (31; see Table 1 for these data as well). Unexpectedly, however, the amino-terminal glutamic acid that we found was different from the amino-terminal glycine reported for bakers' yeast and strain X2180 in the previous work (31).

Homology between protease B and subtilisins. When we compared the *PRB1* coding sequence with all of the sequences in the National Protein Information Resource protein sequence database, we discovered that mature protease B shares significant homology with subtilisins (Fig. 3). The homology to the fungal member of this protease family, proteinase K (26), is even more striking. Within the region of overlap of protease B with proteinase K, 118 of 277 or 43% of the residues are identical. Corresponding values for the homology of protease B to thermitase and the secreted subtilisins are 33 to 35% and to the intracellular subtilisin 30% (30). The active-site residues (55, 78) aspartate (Asp 45 in protease B), histidine (His 77), and serine (Ser 239) are embedded in long stretches of homology. Like proteinase K (26) and thermitase (46), protease B has a cysteine residue (Cys 81) in close proximity to the active-site histidine residue. The most striking difference between the protease B

```
AACACACCCG CGATAAAGAG CGCGATGAAT ATAAAAGGGG GCCAATGTTA CGTCCCGTTA TATTGGAGTT CTTCCCATAC AAACTTAAGA GTCCAATTAG CTTCATCGCC
AATAAAAAAA CAAACTAAAC CTAATTCTAA CAAGCAAAG

ATG AAG TTA GAA AAT ACT CTA TTT ACA CTC GGT GCC CTA GGG AGC ATC TCT GCT GCT TTG GTC ATC CCA AAT CTT GAA AAT GCC GCC GAC
MET Lys Leu Glu Asn Thr Leu Phe Thr Leu Gly Ala Leu Gly Ser Ile Ser Ala Ala Leu Val Ile Pro Asn Leu Glu Asn Ala Ala Asp  -251

CAC CAC GAA CTG ATT AAC AAG GAA GAT CAC CAC GAG AGA CCC AGA AAA GTG GAA TTC ACT AAG GAC GAT GAT GAG GAG CCA TCT GAC TCT
His His Glu Leu Ile Asn Lys Glu Asp His His Glu Arg Pro Arg Lys Val Glu Phe Thr Lys Asp Asp Asp Glu Glu Pro Ser Asp Ser  -221

GAA GAT AAA GAA CAT GGA AAG TTC CAT AAG AAG GGC CGC AAG GGC CAA GAC AAG GAG TCT CCG GAA TTC AAC GGT AAA CGT GCA AGT GGC
Glu Asp Lys Glu His Gly Lys Phe His Lys Lys Gly Arg Lys Gly Gln Asp Lys Glu Ser Pro Glu Phe Asn Gly Lys Arg Ala Ser Gly  -191

TCT CAT GGG AGC GCC CAC GAG GGA GGA AAG GGC ATG AAG CCT AAG CAT GAA AGT TCC AAT GAT GAT GAT AAT GAT GAT AAG AAG AAG AAG
Ser His Gly Ser Ala His Glu Gly Gly Lys Gly Met Lys Pro Lys His Glu Ser Ser Asn Asp Asp Asp Asn Asp Asp Lys Lys Lys Lys  -161

CCT CAC CAT AAG GGT GGC TGC CAC GAA AAT AAG GTG GAG GAG AAG AAG ATG AAA GGT AAG AAA GTC AAG GGC AAG AAG CAC CAC GAA AAG
Pro His His Lys Gly Gly Cys His Glu Asn Lys Val Glu Glu Lys Lys Met Lys Gly Lys Lys Val Lys Gly Lys Lys His His Glu Lys  -131

ACG TTG GAG AAA GGG AGG CAC CAC AAC AGG CTG GCT CCT CTC GTG TCC ACT GCA CAA TTC AAC CCA GAC GCG ATC TCC AAG ATC ATC CCC
Thr Leu Glu Lys Gly Arg His His Asn Arg Leu Ala Pro Leu Val Ser Thr Ala Gln Phe Asn Pro Asp Ala Ile Ser Lys Ile Ile Pro  -101

AAC CGC TAC ATT ATA GTC TTC AAG AGA GGT GCC CCT CAA GAA GAG ATC GAT TTC CAC AAG GAA AAC GTC CAG CAG GCA CAA CTT CAA TCC
Asn Arg Tyr Ile Ile Val Phe Lys Arg Gly Ala Pro Gln Glu Glu Ile Asp Phe His Lys Glu Asn Val Gln Gln Ala Gln Leu Gln Ser   -71

GTA GAG AAC TTA TCT GCC GAA GAC GCT TTC TTC ATT TCT ACT AAA GAC ACC TCC TTG TCC ACC TCT GAG GCT GGC GGT ATC CAG GAC TCA
Val Glu Asn Leu Ser Ala Glu Asp Ala Phe Phe Ile Ser Thr Lys Asp Thr Ser Leu Ser Thr Ser Glu Ala Gly Gly Ile Gln Asp Ser   -41

TTC AAC ATC GAT AAT CTT TTC TCC GGT TAC ATC GGT TAC TTC ACC CAA GAG ATT GTC GAC TTG ATA CGT CAA AAC CCA TTA GTA GAC TTT
Phe Asn Ile Asp Asn Leu Phe Ser Gly Tyr Ile Gly Tyr Phe Thr Gln Glu Ile Val Asp Leu Ile Arg Gln Asn Pro Leu Val Asp Phe   -11

GTT GAG AGA GAC TCT ATT GTC GAG GCT ACA|GAA TTT GAC ACT CAA AAT AGC GCC CCA TGG GGG TTG GCC CGT ATT TCC CAC AGA GAG CGC
Val Glu Arg Asp Ser Ile Val Glu Ala Thr|Glu Phe Asp Thr Gln Asn Ser Ala Pro Trp Gly Leu Ala Arg Ile Ser His Arg Glu Arg    20

CTC AAC CTG GGG TCC TTC AAC AAG TAT CTC TAC GAT GAT GAT GCC GGT CGC GGT GTC ACG TCC TAT GTT ATT GAC ACG GGT GTC AAC ATC
Leu Asn Leu Gly Ser Phe Asn Lys Tyr Leu Tyr Asp Asp Asp Ala Gly Arg Gly Val Thr Ser Tyr Val Ile Asp Thr Gly Val Asn Ile    50

AAC CAC AAG GAC TTC GAA AAG AGA GCC ATT TGG GGG AAA ACC ATC CCA CTT AAC GAC GAA GAT CTC GAC GGT AAC GGC CAC GGT ACC CAC
Asn His Lys Asp Phe Glu Lys Arg Ala Ile Trp Gly Lys Thr Ile Pro Leu Asn Asp Glu Asp Leu Asp Gly Asn Gly His Gly Thr His    80

TGT GCC GGT ACT ATC GCT TCC AAA CAC TAC GGT GTC GCT AAA AAT GCC AAC GTT GTT GCG GTG AAA GTC TTG AGA TCA AAC GGG TCT GGT
Cys Ala Gly Thr Ile Ala Ser Lys His Tyr Gly Val Ala Lys Asn Ala Asn Val Val Ala Val Lys Val Leu Arg Ser Asn Gly Ser Gly   110

ACC ATG TCT GAT GTC GTC AAA GGT GTC GAA TAT GCC GCA AAG GCG CAC CAA AAA GAA GCC CAA GAA AAG AAA AAG GGG TTC AAA GGT TCC
Thr Met Ser Asp Val Val Lys Gly Val Glu Tyr Ala Ala Lys Ala His Gln Lys Glu Ala Gln Glu Lys Lys Lys Gly Phe Lys Gly Ser   140

ACA GCC AAT ATG TCG CTT GGT GGT GGC AAG TCC CCA GCT TTG GAC TTG GCT GTT AAT GCA GCC GTT GAA GTC GGT ATT CAC TTT GCC GTG
Thr Ala Asn Met Ser Leu Gly Gly Gly Lys Ser Pro Ala Leu Asp Leu Ala Val Asn Ala Ala Val Glu Val Gly Ile His Phe Ala Val   170

GCT GCT GGT AAC GAA AAC CAA GAC GCT TGT AAT ACC TCC CCA GCT TCT GCT GAC AAG GCC ATC ACC GTC GGA GCT TCC ACG TTG AGC GAT
Ala Ala Gly Asn Glu Asn Gln Asp Ala Cys Asn Thr Ser Pro Ala Ser Ala Asp Lys Ala Ile Thr Val Gly Ala Ser Thr Leu Ser Asp   200

GAC AGA GCC TAC TTT TCC AAC TGG GGT AAG TGT GTC GAC GTT TTC GCC CCA GGT TTA AAC ATT TTG TCC ACC TAC ATT GGA AGT GAT GAC
Asp Arg Ala Tyr Phe Ser Asn Trp Gly Lys Cys Val Asp Val Phe Ala Pro Gly Leu Asn Ile Leu Ser Thr Tyr Ile Gly Ser Asp Asp   230

GCC ACC GCC ACT TTA TCG GGT ACC TCA ATG GCT TCC CCT CAC GTT GCT GGT TTG TTG ACC TAC TTT TTG TCT TTG CAA CCA GGT TCT GAT
Ala Thr Ala Thr Leu Ser Gly Thr Ser Met Ala Ser Pro His Val Ala Gly Leu Leu Thr Tyr Phe Leu Ser Leu Gln Pro Gly Ser Asp   260

AGT GAA TTT TTC GAA TTG GGC CAG GAT TCT TTG ACT CCT CAG CAA TTG AAA AAG AAG CTA ATT CAT TAC AGT ACA AAG GAC ATT TTG TTC
Ser Glu Phe Phe Glu Leu Gly Gln Asp Ser Leu Thr Pro Gln Gln Leu Lys Lys Lys Leu Ile His Tyr Ser Thr Lys Asp Ile Leu Phe   290

GAT ATC CCT GAA GAC ACT CCA AAT GTT TTA ATC TAC AAC GGT GGT GGT CAA GAT TTG TCC GCT TTC TGG AAT GAT ACC AAA AAG TCC CAT
Asp Ile Pro Glu Asp Thr Pro Asn Val Leu Ile Tyr Asn Gly Gly Gly Gln Asp Leu Ser Ala Phe Trp Asn Asp Thr Lys Lys Ser His   320

TCA TCT GGT TTC AAG CAA GAA TTG AAC ATG GAT GAA TTC ATT GGA TCA AAG ACC GAT TTA ATC TTT GAT CAA GTG AGA GAT ATT CTT GAT
Ser Ser Gly Phe Lys Gln Glu Leu Asn Met Asp Glu Phe Ile Gly Ser Lys Thr Asp Leu Ile Phe Asp Gln Val Arg Asp Ile Leu Asp   350

AAA TTG AAT ATT ATT TAA
Lys Leu Asn Ile Ile

TTCTTCATTT AGAAAAATTT CAGCTGCTTT TTTTTTTCTT TTTCTTTCCT TAGGCGTCTC GAGGTTACAA GTCGGAGTCC CTCTTCACTA TCGTTTGTCC ACTTTTTTTA
TATCCCCATT ATTTTCAATC TGAATTTCAT TTTTTTTTTT TAATTCATGA AATTTATATG TCCCACGTAT TACTACATAT TTGCGTTTTT AATTAAATAA ATAACTGTTA
CTTTTATTAT ATCTTATTTG CAGATCACTT ATCTGATCAA ATGTTTTCGT TTTCGTGTGT GGTGACGATG TATTAGGTAC GCGAAATAAA CAAAACAAAC AAACAAGGCC
```

FIG. 1. Sequence of the *PRB1* gene and its flanking regions from strain AB320 (50, 51). The DNA sequence is shown together with the predicted amino acid sequence of the *PRB1*-encoded protein. (It is assumed that the first AUG codon of the transcript is used as the start of translation.) The vertical arrow marks the amino terminus of the mature enzyme. The underlined sequence of 15 amino acids corresponds to the N-terminal sequence found by analysis of protease B purified from bakers' yeast. Asterisks identify the aspartic acid, histidine, and serine residues corresponding to active-site residues in the homologous subtilisins. Wavy underlining identifies cysteine residues. Diamonds identify the positions of potential signals for asparagine-linked glycosylation (Asn-X-Thr or Asn-X-Ser) (63). Horizontal arrows identify the beginning and end of the *PRB1* mRNA, as determined by S1 nuclease mapping (7).

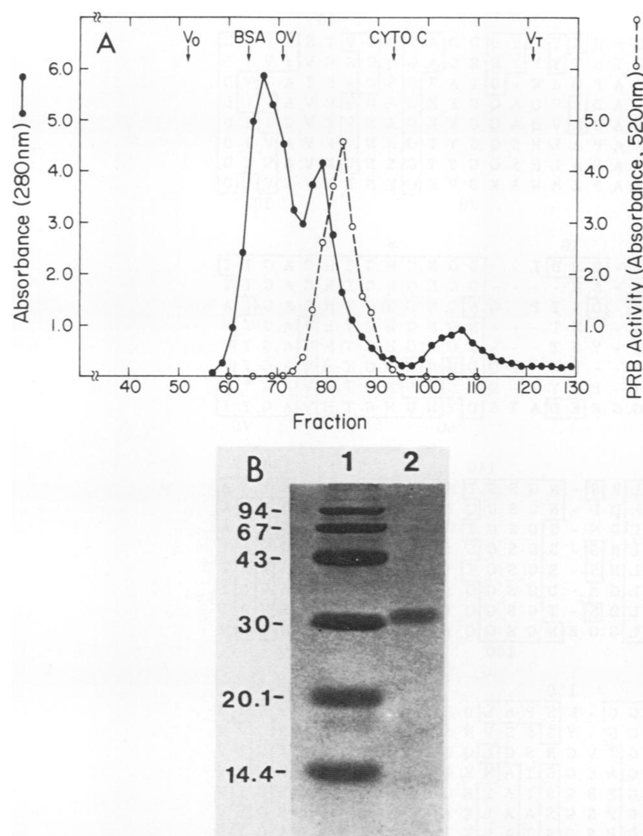

FIG. 2. Size estimates for protease B (PRB). Protease B was purified as described in Materials and Methods. (A) Typical elution profile from the Sephadex G-100 column. MW standards were ferritin (V_0), excluded; bovine serum albumin (BSA), 67,000; ovalbumin (OV), 43,000; and cytochrome *c* (CYTO C), 14,400. The total volume (V_T) was calculated by the formula $V_T = \pi r^2 h$. (B) Coomassie blue-stained gel from SDS-polyacrylamide gel electrophoresis (16% acrylamide) of the final PMSF-inactivated protein preparation (lane 2) used for N-terminal sequence analysis. Lane 1, MW markers: phosphorylase *b*, 94,000; BSA, 67,000; OV, 43,000; carbonic anhydrase, 30,000; soybean trypsin inhibitor, 20,100; and CYTO C, 14,400.

sequence and those of the other members of this family is the 62-amino-acid extension of the *PRB1* ORF C terminus relative to all the other C termini save that of intracellular subtilisin.

The DNA sequence for the protease B precursor contains four cysteine codons corresponding to amino acids −154, 81, 180, and 211, the last three of which reside within the mature protease B polypeptide. Kominami et al. (31) reported titration by 5,5′-dithiobis (2-nitrobenzoic acid) of 0.85 mol of cysteine per mol of purified protease B, raising the possibility that two may be in a disulfide bond. Since Kominami et al. (31) also reported that *p*-hydroxymercuribenzoate eliminated the reaction of the active-site serine with diisopropylfluorophosphate, the most likely candidate for the free cysteine is Cys 81, located near the active-site His 77. Proteinase K contains five cysteine residues, one free at the position corresponding to Cys 81 of protease B and the remaining four in two disulfide bonds. None of the other cysteine residues of protease B and proteinase K are found in homologous positions (26). The single cysteine in thermitase is homologous to Cys 81 (46).

The glycine residue at position 166 (165 in our numbering in Fig. 3) is invariant in extracellular subtilisins and forms an important part of the binding cleft. Varying the residue at this position by in vitro mutagenesis had profound effects on substrate specificity (14). Interestingly, this portion of the primary sequence corresponds to a two-amino-acid gap in the fungal and actinomycete sequences.

The protease B primary sequence shows no homology to those of the trypsin-chymotrypsin family of serine proteases. Not surprisingly then, the pattern of cleavage of the oxidized β chain of insulin catalyzed by protease B (32) closely resembles those of the subtilisins (34, 54) and proteinase K (34) but differs from those of chymotrypsin (57) and trypsin (73). The DNA-derived amino acid sequences of two cloned genes for subtilisins reveal that they are synthesized as preproenzymes (61, 71, 74, 76). Although the two preproenzyme sequences are virtually identical for the subtilisins, they show no primary sequence homology to the corresponding region of the protease B precursor (data not shown).

The primary, secondary, and tertiary structures of the subtilisins have been studied extensively. Computer-generated predictions of the hydropathy (37) and the secondary structure (11) of the *PRB1*-encoded protein are shown in Fig. 4. Not unexpectedly, since protease B is a vacuolar hydrolase and presumably follows the route through the endoplasmic reticulum and Golgi body to the vacuole like other vacuolar hydrolases (62), there is an initial hydrophobic leader. It is followed, however, by a very hydrophilic stretch of amino acids. Between residues −250 and −121, there are 29 negatively and 51 positively charged residues. Whether or how the cell accomplishes transport of this sequence through the membrane into the lumen of the endoplasmic reticulum awaits resolution. Within most of the *pro* sequence, as well as the leader sequence, extended regions of α helix are predicted, including one that ends near residue 1, the N-terminal glutamic acid of the mature polypeptide. Little secondary structure is predicted within the polypeptide sequence of the mature enzyme, and most of this predicted secondary structure in short stretches. However, the α helices predicted for Met 112-Lys 134, Ser 151-Ile 166, and Ala 241-Leu 253 are found in positions that correspond to those of α helices found in the crystal structure of subtilisin BPN′ from *Bacillus amyloliquefaciens* (D, Glu 103-Asn 117; E, Ser 132-Ser 145; and F, Thr 242-Asn 252, respectively) (78). Likewise, the predicted β-pleated sheet segments Ser 41-Ile 44, Val 98-Arg 105, and Val 164-Val 170 appear to be counterparts of Val 28-Asp 32, Ala 89-Lys 94, and Val 148-Ala 152, respectively, in the subtilisin BPN′ crystal structure (78). Interestingly, the F α helix presents the serine (239 for protease B) to the active site, and the Ser 41-Ile 44 β-pleated sheet is adjacent to the active-site Asp 45. Conspicuously absent from the prediction for protease B is the C α helix in subtilisin BPN′ (His 64-Ala 73) that runs from the active-site His 64 (His 77 in protease B). A cysteine residue (Cys 81) is present in the protease B sequence within this otherwise quite homologous segment.

DISCUSSION

Maturation of protease B. Mature protease B is a glycoprotein with an MW of approximately 32,000 to 33,000 (8, 17, 23, 25, 31, 41, 56, 70; this study), of which 8 to 9% is estimated to be tunicamycin- and endo H-insensitive carbohydrate (31), leaving a polypeptide MW of about 29,000 to 30,000. Kinetic studies have detected a larger, 42,000-MW precursor to protease B that contains a tunicamycin-

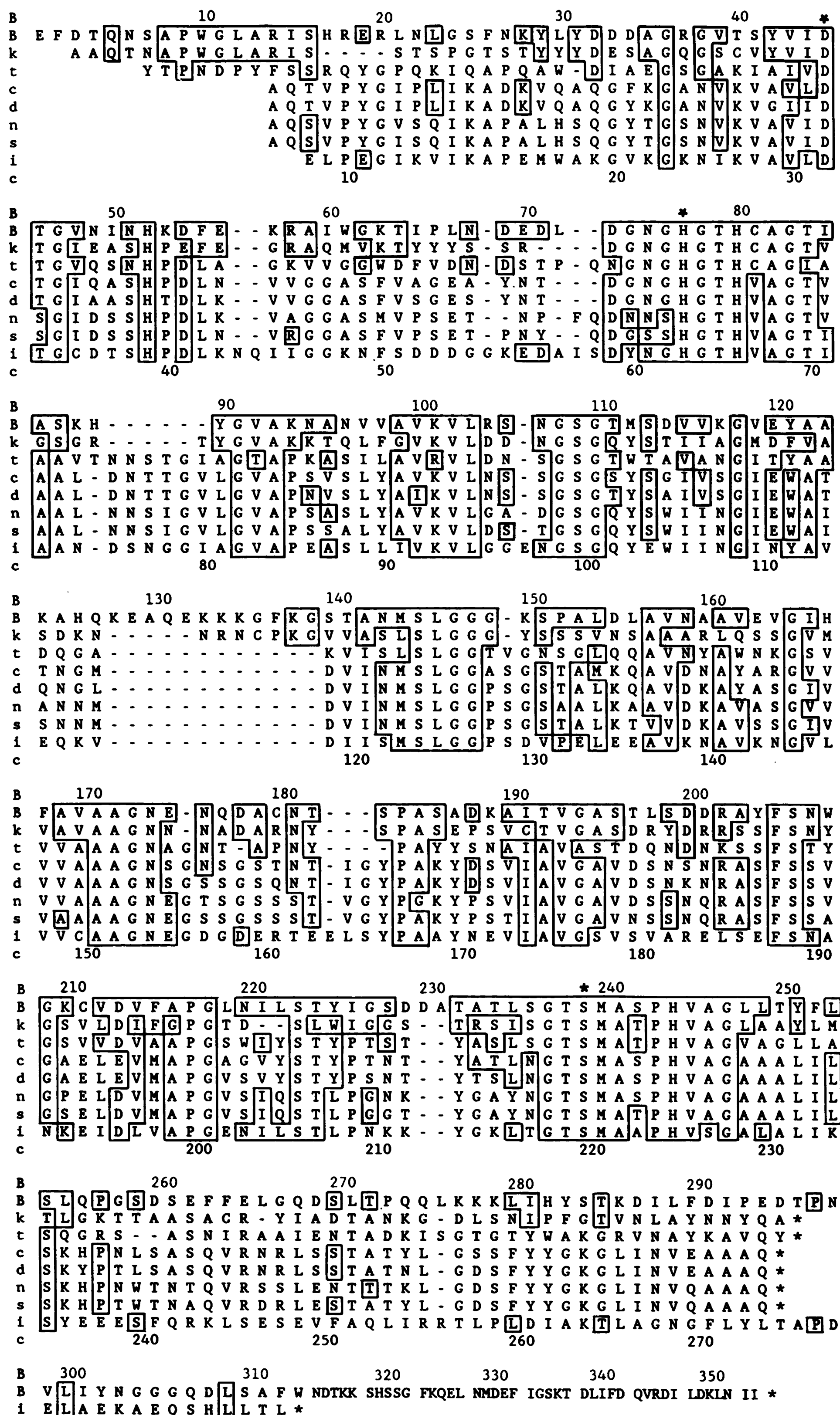

FIG. 3. Homology of protease B with other subtilisin family proteases. B, Protease B of *S. cerevisiae*; k, *Tritirachium album* Limber proteinase K (26); t, *Thermoactinomyces vulgaris* thermitase (46); c, *Bacillus subtilis* Carlsberg subtilisin (60); d, *B. subtilis* DY subtilisin (52); n, *B. amyloliquefaciens* subtilisin (also called BPN') (74); s, *B. subtilis* subsp. *amylosacchariticus* subtilisin (35); i, *B. subtilis* intracellular subtilisin (30). Boxes enclose residues identical to the corresponding protease B residue. The active-site aspartic acid, histidine, and serine residues are marked by asterisks. The sequences presented begin with the known N-terminal residues and proceed to the end of the protein or ORF.

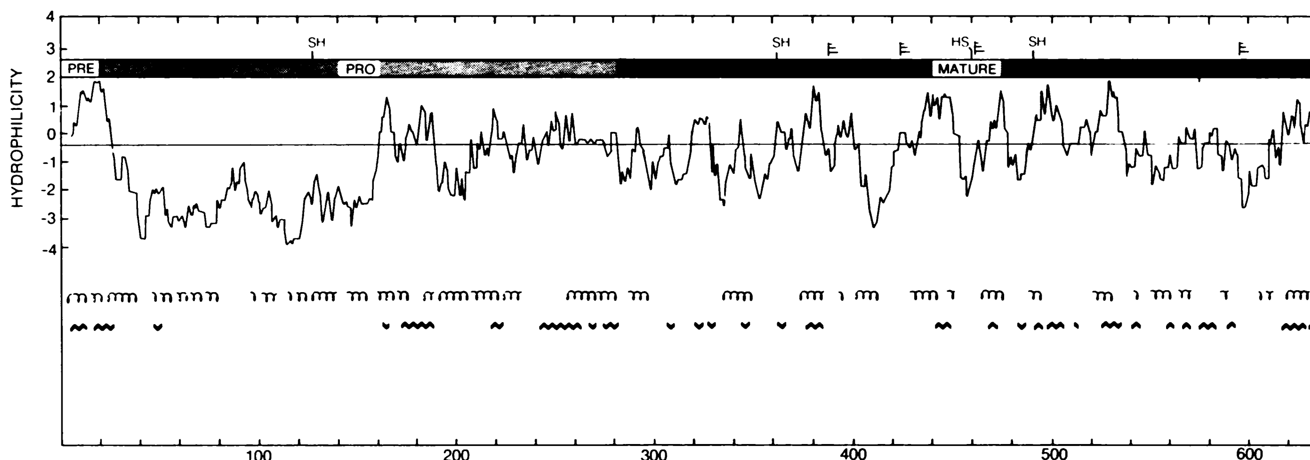

FIG. 4. Computer-aided secondary structure analysis of the predicted *PRB1* polypeptide. Hydrophilicity is represented by the solid line (determined by the method of Kyte and Doolittle [37] by using a segment length of nine). The α-helix (loops) and β-sheet (wavy lines) conformations predicted by the method of Chou and Fasman (11) are also shown. The locations of the four cysteine residues are shown. Potential Asn-linked glycosylation acceptor sequences are indicated ( ). The position of the N-terminal glutamic acid is located at the junction between the blocks designated "PRO" and "MATURE." The inferred end of the signal sequence is located at the junction between the blocks designated "PRE" and "PRO." A mark below the block designated MATURE corresponds to the position of the C termini of the homologous subtilisins.

sensitive glycosidic side chain (44). The maturation step that converts the 42,000-MW precursor to the 33,000-MW enzyme requires the action of protease A, the product of the *PEP4* gene (3, 77), and is presumed to occur by removal of a polypeptide. In *pep4* mutants, a 42,000-MW precursor accumulates (45).

The ORF that includes sequence information for protease B predicts a 635-amino-acid polypeptide with an MW of 69,619, much larger than the known 42,000-MW precursor (44). The estimated MW for the 355-amino-acid polypeptide extending from the N terminus to the end of the ORF is 38,140 (Fig. 5). With the addition of two types of carbohydrate moieties, one (probably distributed over more than one linkage) with an MW of about 2,900 for the 15 mannose, 2 to 3 hexosamine, tunicamycin-insensitive component corresponding to the carbohydrate found in the mature enzyme (31) and one with an MW of about 3,000 for the 16 mannose, 2 acetylglucosamine, tunicamycin-sensitive component that was reported to be part of the 42,000-MW precursor but not the mature enzyme (44), an MW of around 44,000 would be obtained. This MW is tantalizingly close to the observed precursor MW of 42,000. Moreover, cleavage of the

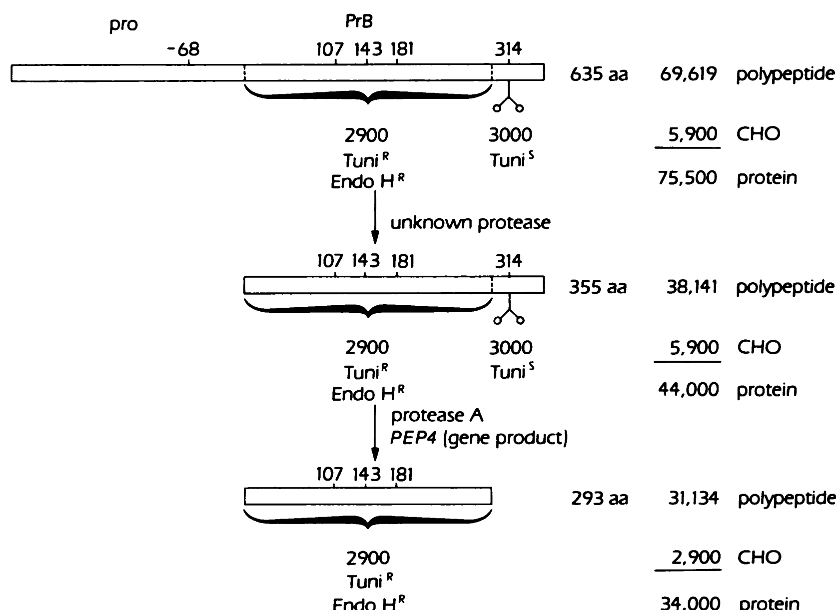

FIG. 5. Proposed processing pathway of preproprotease B. The open bars correspond to the primary sequence of the polypeptide forms. Dotted lines indicate positions of cleavages to follow. The *pro* designation includes all amino acids (aa) between the initial methionine and the N-terminal glutamic acid of the mature protein. Two classes of carbohydrate side chains (CHO) are indicated, one resistant to tunicamycin ($Tuni^R$) and endo H (Endo H^R) and the other sensitive to tunicamycin ($Tuni^S$). See the text for other details.

42,000-MW precursor at position 293, corresponding to the C termini of the homologous subtilisins, proteinase K, and thermitase, would result in the removal of 62 amino acids (MW, 7,007) as well as a tunicamycin-sensitive, Asn-linked glycosidic side chain on Asn 314 (MW, 3,000). This would leave a 293-amino-acid polypeptide with an MW of 31,134 containing tunicamycin- and endo H-insensitive carbohydrate with a total MW of 2,900, giving a protein with an MW of 34,034, very close to the observed MW of 32,000 to 33,000 (Fig. 5) (see Table 1).

Kominami et al. (31) have reported the amino acid compositions of protease B purified from bakers' yeast and from a standard laboratory strain, X2180. In Table 1 we have compared the reported compositions of these two preparations to the DNA-derived compositions of the 355- and 293-amino-acid polypeptides discussed above. It is clear that the reported compositions agree more closely with the composition of the 293-amino-acid polypeptide. The simplest hypothesis to reconcile the size estimates for the product of the ORF, the known precursor, and the mature enzyme is that translation and glycosylation result in an as-yet-undetected precursor with an MW of 75,500 (69,619-MW polypeptide plus 5,900-MW carbohydrate) that is processed to yield the 42,000-MW precursor reported by Mechler et al. (44). We propose that the 42,000-MW precursor possesses the same N terminus as mature protease B. This 42,000-MW protein is then further processed by protease A at or near amino acid 293 to yield the 293-amino-acid protease B with the mature C terminus. This hypothesis would account for the finding that the 42,000-MW precursor accumulates in the *pep4-3* mutant (45).

The consensus tripeptide sequences for Asn-linked glycosylation are Asn-X-Thr/Ser(Cys) (63). The postulated processing pathway assumes that the Asn-X-Thr sequence at Asn 314 is an acceptor for Asn-linked (and therefore tunicamycin- and endo H-sensitive) glycosylation but that the Asn-X-Ser sequences at Asn −68, Asn 107, Asn 143, and Asn 181 are not. In Table 2 we summarize data on acceptor sites for Asn-linked glycosylation for glycoproteins in *S. cerevisiae*. The preponderance of the tripeptide acceptors are of the Asn-X-Thr type. For vacuolar hydrolases, excluding protease B, all known sites are of the Asn-X-Thr type, and all are used. For secreted proteins, both types of tripeptide acceptor are present. Comparison of the known number of glycosidic side chains with the number of potential tripeptide acceptor sites of each type requires that Asn-X-Thr be an acceptor for glycosylation on secreted proteins. Where Asn-X-Thr and Asn-X-Ser tripeptides are both present, the number of chains found is fewer than the number of tripeptide acceptors present. There is only one instance, that of killer toxin, in which Asn-X-Ser is known to be used as an acceptor (19). It is possible then that the four Asn-X-Ser sequences at Asn −68, Asn 107, Asn 143, and Asn 181 in the protease B polypeptide are not used and that the tunicamycin- and endo H-insensitive glycosidic residues are hydroxyl linked. Certainly in in vitro studies with calf liver microsomal fractions, Asn-X-Thr is a much better acceptor than is either Asn-X-Ser or Asn-X-Cys (6). The only observation unaccounted for by the hypothesis is the reported presence of hexosamine in mature protease B (31), for all hexosamine is normally present in Asn-linked chitobiose residues in *S. cerevisiae* (5). This hypothesis does not address the question of glycosylation at Asn −68.

An alternative, albeit less tidy, hypothesis would place a glycosylated Asn −68 in the 42,000-MW precursor, with protease A-mediated maturation occurring at the N terminus of mature protease B. This alternative hypothesis is less likely for three reasons. First, according to this model, the N terminus of the precursor that accumulates in the *pep4-3* mutant must be found at or before Asn −68. Therefore, the predicted MW of this precursor would be 53,000 or greater, as opposed to the observed MW of 42,000. Second, it follows from the first reason that mature protease B would have a predicted MW of 44,000, as opposed to the observed MW of 32,000 to 33,000. Third, this alternative model would require glycosylation at the less commonly used tripeptide acceptor sequence (Asn-X-Ser) rather than the more commonly used one (Asn-X-Thr) (Table 2).

TABLE 1. Amino acid composition of protease B

Amino acid	No. of residues/mol of enzyme in:			
	Bakers' yeast[a]	X2180[a]	AB320[b]	
			Amino acids 1 to 355	Amino acids 1 to 293
Aspartic acid[c]	44.1	44.3	52	39
Threonine	21.1	20.1	21	18
Serine	23.7	23.2	31	26
Glutamic acid[c]	18.2	19.5	25	19
Proline	10.2	10.4	10	9
Glycine	27.0	27.3	34	29
Alanine	28.0	29.4	33	32
Half-cystine	1.3	ND[d]	3	3
Valine	17.5	19.3	21	19
Methionine	1.8	ND	4	3
Isoleucine	11.4	12.8	19	13
Leucine	21.1	22.0	29	23
Tyrosine	8.0	7.8	10	9
Phenylalanine	9.1	9.6	15	11
Lysine	20.4	17.3	26	21
Histidine	8.4	7.5	10	9
Arginine	6.0	5.7	8	7
Tryptophan	2.3	ND	4	3

[a] Data are from Kominami et al. (31) and are based on an MW of 33,000.
[b] The composition is derived from the DNA sequence. AB320 is related to X2180 (48).
[c] Aspartic acid plus asparagine and glutamic acid plus glutamine, respectively.
[d] ND, Not determined.

Several of the *S. cerevisiae* tripeptide acceptor sequences in Table 2, including the Asn-Asp-Thr sequence at Asn 314 in the protease B precursor sequence, have aspartic acid as the middle amino acid. On statistical grounds, Struck and Lennarz (63) have inferred that aspartic acid may be unacceptable in the center position. Because all of the *S. cerevisiae* proteins in Table 2 are known to be glycosylated and hence to pass through the compartments in which glycosylation occurs, a direct determination of whether Asn-Asp-Thr or Asn-Asp-Ser serves as a glycosylation acceptor is now possible.

Protease homology. Protease B of *S. cerevisiae* is a 33,000-MW glycoprotein whose primary sequence shows striking homology to those of the subtilisins, proteinase K, and thermitase. Like proteinase K and thermitase, but unlike the subtilisins, protease B is a serine protease that contains a free cysteine residue that presumably is near the active site. It is very likely that other eucaryotes produce related serine proteases. Protease B of *Candida albicans*, the alkaline proteases of *Aspergillus* species, and thermomycolin from *Malbranchea pulchella* seem likely candidates for homologous enzymes. Protease B of *C. albicans* is sensitive to diisopropylfluorophosphate and mercurial compounds

TABLE 2. Acceptor sites for Asn-linked glycosylation in yeast glycoproteins

Glycoprotein	No.[a] of:			Reference[b]
	Asn-X-Thr sites	Asn-X-Ser sites	Carbohydrate side chains	
Vacuolar hydrolases				
Carboxypeptidase Y	4	0	4	PS (65); C (20, 67)
Protease A	2	0	2	PS (12); DS (3, 77); C (12)
Alkaline phosphatase	2	0	2	DS (Y. Oshima, personal communication); C (53)
Protease B	1 (1)	4	1 or 2[c]	DS (this work); C (31, 44)
Secreted proteins				
Mating factor α	3	0	3	DS (36); C (29)
Killer toxin	2	1	3	DS (9, 59); C (19)
Invertase	10 (2)	3 (1)	9[c]	DS (66); C (68)
Acid phosphatase	9 (2)	3	8[c]	DS (4); C (33)
Lysine-arginine-cleaving endopeptidase	5[d]	1[e]	3 or 4[c]	DS (R. Fuller, and J. Thorner, personal communication); C (R. Fuller and J. Thorner, personal communication)
α-Galactosidase	7 (1)	1	8[f]	DS (64); C (64)

[a] Numbers in parentheses represent the numbers of tripeptide acceptors with aspartic acid as X.
[b] PS, Protein sequence data; DS, DNA sequence data; C, number of glycosidic chains found.
[c] When the number of acceptor asparagines used is fewer than the number of acceptors available, it is unknown which asparagines are used.
[d] One of the five is beyond the C-terminal transmembrane domain.
[e] X is proline, which is expected to be unacceptable (63).
[f] Data are inconclusive.

and catalyzes a subtilisinlike cleavage of the oxidized β chain of insulin (15). Thermomycolin (18) produces a subtilisinlike cleavage pattern of insulin, is retarded in gel filtration like protease B and proteinase K (10, 13, 26, 55), has an N-terminal sequence homologous to those of proteinase K and protease B (18), and has an active-site sequence for the serine residue corresponding to that of the subtilisin family rather than to that of the trypsin-chymotrypsin family (18). The alkaline serine proteases of *Aspergillus* species likewise have the substrate specificity (18) and active-site serine sequences (49) of the subtilisin family. Interestingly, some of these contain cysteine residues, and others do not (49). The sulfhydryl reagent-sensitive serine proteases of *Phycomyces* species (16) and *Neurospora crassa* (1, 58) are other possible homologous enzymes.

ACKNOWLEDGMENTS

We thank Kathleen Conley for purifying protease B, John R. Thompson for extensive instruction in using IBM personal computers and software, William E. Brown for the computer program for analyzing protein secondary structure and instruction in its use, and Klaus Breddam for helpful discussions.

This work was supported by Public Health Service research grants AM18090 and GM29713 from the National Institutes of Health (NIH) (to E.W.J.), NIH postdoctoral fellowship F32AI06884 (to S.K.L.), NIH training grant T32GM08067 (to C.M.M.), and Biogen.

LITERATURE CITED

1. **Abbott, R., and G. Marzluf.** 1984. Major extracellular protease of *Neurospora crassa*. J. Bacteriol. **159:**505–510.
2. **Achstetter, T., and D. Wolf.** 1985. Proteinases, proteolysis and biological control in the yeast *Saccharomyces cerevisiae*. Yeast **1:**139–157.
3. **Ammerer, G., C. Hunter, J. Rothman, G. Saan, L. Valls, and T. Stevens.** 1986. *PEP4* gene of *Saccharomyces cerevisiae* encodes proteinase A, a vacuolar enzyme required for processing of vacuolar precursors. Mol. Cell. Biol. **6:**2490–2499.
4. **Bajwa, W., B. Meyhack, H. Rudolph, A.-M. Schweingruber, and A. Hinnen.** 1984. Structural analysis of the two tandemly repeated acid phosphatase genes in yeast. Nucleic Acids Res. **12:**7721–7739.
5. **Ballou, C.** 1982. Yeast cell wall and cell surface, p. 335–360. *In* J. Strathern, E. Jones, and J. Broach (ed.), Molecular biology of the yeast *Saccharomyces*, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
6. **Bause, E., and G. Legler.** 1981. The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis. Biochem. J. **195:**639–644.
7. **Berk, A., and P. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12:**721–732.
8. **Betz, H., H. Hinze, and H. Holzer.** 1974. Isolation and properties of two inhibitors of proteinase B from yeast. J. Biol. Chem. **249:**4515–4521.
9. **Bostian, K., Q. Elliot, H. Bussey, V. Burn, A. Smith, and D. Tipper.** 1984. Sequence of the preprotoxin dsRNA gene of type I killer yeast: multiple processing events produce a two-component toxin. Cell **36:**741–751.
10. **Chestukhina, G., A. Epremyan, A. Gaida, A. Osterman, O. Khodova, and V. Stepanov.** 1982. Thiol dependent serine proteinases. II. An intracellular serine proteinase of *Bacillus cereus*. Bioorg. Khim. **8:**1649–1658.
11. **Chou, P., and G. Fasman.** 1974. Prediction of protein conformation. Biochemistry **13:**222–245.
12. **Dreyer, T., B. Halkier, I. Svendsen, and M. Ottesen.** 1986. Primary structure of the aspartic proteinase A from *Saccharomyces cerevisiae*. Carlsberg Res. Commun. **51:**27–41.
13. **Ebeling, W., N. Hennrich, M. Klockow, H. Metz, H. Orth, and H. Lang.** 1974. Proteinase K from *Tritirachium album* Limber. Eur. J. Biochem. **47:**91–97.
14. **Estell, D., T. Graycar, J. Miller, D. Powers, J. Burnier, P. Ng, and J. Wells.** 1986. Probing steric and hydrophobic effects on enzyme-substrate interactions by protein engineering. Science **233:**659–663.
15. **Farley, P., M. Shepherd, and P. Sullivan.** 1986. The purification and properties of yeast proteinase B from *Candida albicans*. Biochem. J. **236:**177–184.

16. **Fischer, E.-P.** 1979. Limited proteolysis in *Phycomyces*, p. 139–144. *In* G. Cohen and H. Holzer (ed.), Limited proteolysis of microorganisms. U.S. Department of Health, Education, and Welfare publication no. (NIH) 79-1591. National Institutes of Health, Bethesda, Md.
17. **Fujishiro, K., Y. Sanada, H. Tanaka, and N. Katunuma.** 1980. Purification and characterization of protease B. J. Biochem. (Tokyo) **87:**1321–1326.
18. **Gaucher, G., and K. Stevenson.** 1976. Thermomycolin. Methods Enzymol. **45:**415–433.
19. **Hanes, S., V. Burn, S. Sturley, D. Tipper, and K. Bostian.** 1986. Expression of a cDNA derived from the yeast killer preprotoxin gene: implications for processing and immunity. Proc. Natl. Acad. Sci. USA **83:**1675–1679.
20. **Hasilik, A., and W. Tanner.** 1978. Biosynthesis of the vacuolar yeast glycoprotein carboxypeptidase Y. Conversion of precursor into enzyme. Eur. J. Biochem. **91:**567–575.
21. **Hemmings, B., G. Zubenko, A. Hasilik, and E. Jones.** 1981. Mutant defective in processing of an enzyme located in the lysosome-like vacuole of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **78:**435–439.
22. **Hewick, R., M. Hunkapiller, L. Hood, and W. Dreyer.** 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. **256:**7990–7997.
23. **Hinze, H., H. Betz, T. Saheki, and H. Holzer.** 1975. Formation of a complex between yeast proteinases A and B. Hoppe-Seyler's Z. Physiol. Chem. **356:**1259–1264.
24. **Hunkapiller, M., E. Lujan, F. Ostrander, and L. Hood.** 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. Methods Enzymol. **91:**227–236.
25. **Huse, G., G. Kopperschlaeger, and E. Hofmann.** 1982. A new procedure for the purification of proteinase B from baker's yeast and interaction of the purified enzyme with a specific inhibitor. Acta Biol. Med. Ger. **41:**991–1002.
26. **Jany, K.-D., G. Lederer, and B. Mayer.** 1986. Amino acid sequence of proteinase K from the mold *Tritirachium album* Limber. Proteinase K—a subtilisin-related enzyme with disulfide bonds. FEBS Lett. **199:**139–144.
27. **Jones, E.** 1984. The synthesis and function of proteases in *Saccharomyces*: genetic approaches. Annu. Rev. Genet. **18:** 233–270.
28. **Jones, E., G. Zubenko, and R. Parker.** 1982. *PEP4* gene function is required for expression of several vacuolar hydrolases in *Saccharomyces cerevisiae*. Genetics **102:**665–677.
29. **Julius, D., R. Schekman, and J. Thorner.** 1984. Glycosylation and processing of prepro-α-factor through the yeast secretory pathway. Cell **36:**309–318.
30. **Koide, T., A. Nakamura, T. Uozumi, and T. Beppu.** 1986. Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. J. Bacteriol. **167:**110–116.
31. **Kominami, E., H. Hoffschulte, and H. Holzer.** 1981. Purification and properties of proteinase B from yeast. Biochim. Biophys. Acta **661:**124–135.
32. **Kominami, E., H. Hoffschulte, L. Lenschel, K. Maier, and H. Holzer.** 1981. The substrate specificity of proteinase B from baker's yeast. Biochim. Biophys. Acta **661:**136–141.
33. **Kozulic, B., S. Barbaric, B. Ries, and P. Mildner.** 1984. Study of the carbohydrate part of yeast acid phosphatase. Biochem. Biophys. Res. Commun. **122:**1083–1090.
34. **Kraus, E., H. Kiltz, and U. Femfert.** 1976. The specificity of proteinase K against oxidized insulin B chain. Hoppe-Seyler's Z. Physiol. Chem. **357:**233–237.
35. **Kurihara, M., F. Markland, and E. Smith.** 1972. Subtilisin Amylosaccharíticus. III. Isolation and sequence of the chymotryptic peptides and the complete amino acid sequence. J. Biol. Chem. **247:**5619–5631.
36. **Kurjan, J., and I. Herskowitz.** 1982. Structure of a yeast pheromone gene (MFα): a putative α-factor precursor contains four tandem copies of mature α-factor. Cell **30:**933–943.
37. **Kyte, J., and R. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
38. **Laemmli, U.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
39. **Lenney, J., P. Matile, A. Wiemken, M. Schellenberg, and J. Meyer.** 1974. Activities and cellular localization of yeast proteases and their inhibitors. Biochem. Biophys. Res. Commun. **60:**1378–1383.
40. **Lipman, D., and W. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227:**1435–1441.
41. **Looze, Y., L. Gillet, M. Deconinck, B. Couteaux, and E. Polastro.** 1979. Protease B from *Saccharomyces cerevisiae*. Int. J. Pept. Protein Res. **13:**253–259.
42. **Matern, H., H. Betz, and H. Holzer.** 1974. Compartmentation of inhibitors of proteinases A and B and carboxypeptidase Y in yeast. Biochem. Biophys. Res. Commun. **60:**1051–1057.
43. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
44. **Mechler, B., M. Müller, H. Müller, F. Muessdoerffer, and D. Wolf.** 1982. *In vivo* biosynthesis of the vacuolar proteinases A and B in the yeast *Saccharomyces cerevisiae*. J. Biol. Chem. **257:**11203–11206.
45. **Mechler, B., M. Müller, H. Müller, and D. Wolf.** 1982. *In vivo* biosynthesis of vacuolar proteinases in proteinase mutants of *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun. **107:**770–778.
46. **Meloun, B., M. Baudys, V. Kostka, G. Hausdorf, C. Frommel, and W. Hohne.** 1985. Complete primary structure of thermitase from *Thermoactinomyces vulgaris* and its structural features related to the subtilisin-type proteinases. FEBS Lett. **183:** 195–200.
47. **Moehle, C., M. Aynardi, M. Kolodny, F. Park, and E. Jones.** 1987. Protease B of *Saccharomyces cerevisiae*: isolation and regulation of the *PRB1* structural gene. Genetics **115:**255–263.
48. **Mortimer, R., and J. Johnston.** 1986. Genealogy of principal strains of the Yeast Genetic Stock Center. Genetics **113:** 35–43.
49. **Nakagawa, Y.** 1970. Alkaline proteinases from Aspergillus. Methods Enzymol. **19:**581–591.
50. **Nasmyth, K., and S. Reed.** 1980. Isolation of genes by complementation in yeast. Proc. Natl. Acad. Sci. USA **77:** 2119–2123.
51. **Nasmyth, K., and K. Tatchell.** 1980. The structure of transposable yeast mating type loci. Cell **19:**753–764.
52. **Nedkov, P., W. Oberthur, and G. Braunitzer.** 1983. Die Primarstruktur von Subtilisin DY. Hoppe-Seyler's Z. Physiol. Chem. **364:**1537–1540.
53. **Onishi, H., J. Tkacz, and J. Lampen.** 1979. Glycoprotein nature of yeast alkaline phosphatase. J. Biol. Chem. **254:**11943–11952.
54. **Ottesen, M., and I. Svendsen.** 1970. The subtilisins. Methods Enzymol. **19:**199–215.
55. **Pahler, A., A. Banerjee, J. Dattagupta, T. Fujiwara, K. Lindner, G. Pal, D. Suck, G. Weber, and W. Saenger.** 1984. Three-dimensional structure of fungal proteinase K reveals similarity to bacterial subtilisin. EMBO J. **3:**1311–1314.
56. **Sanada, Y., K. Fujishiro, H. Tanaka, and H. Katanuma.** 1975. Isolation and characterization of yeast protease B. Biochem. Biophys. Res. Commun. **86:**815–821.
57. **Sanger, F., and H. Tuppy.** 1951. The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. Biochem. J. **49:**481–490.
58. **Siepen, D., P.-H. Yu, and M.-R. Kula.** 1975. Proteolytic enzymes of *Neurospora crassa*: purification and some properties of five intracellular proteinases. Eur. J. Biochem. **56:**271–281.
59. **Skipper, N., D. Thomas, and P. Lau.** 1984. Cloning and sequencing of the preprotoxin-coding region of the yeast M1 double-stranded RNA. EMBO J. **3:**107–111.
60. **Smith, E., R. DeLange, W. Evans, M. Landon, and F. Markland.** 1968. Subtilisin Carlsberg. V. The complete sequence; comparison with subtilisin BPN'; evolutionary relationships. J. Biol. Chem. **243:**2184–2191.
61. **Stahl, M., and E. Ferrari.** 1984. Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro-derived dele-

tion mutation. J. Bacteriol. **158**:411–418.

62. **Stevens, T., B. Esmon, and R. Schekman.** 1982. Early stages in the yeast secretory pathway are required for transport of carboxypeptidase Y to the vacuole. Cell **30**:439–448.
63. **Struck, D., and W. Lennarz.** 1980. The function of saccharide-lipids in synthesis of glycoproteins, p. 35–83. *In* W. Lennarz (ed.), The biochemistry of glycoproteins and proteoglycans. Plenum Publishing Corp., New York.
64. **Sumner-Smith, M., R. Bozzato, N. Skipper, R. Davies, and J. Hopper.** 1985. Analysis of the inducible *MEL1* gene of *Saccharomyces carlsbergensis* and its secreted product, alpha-galactosidase (melibiase). Gene **36**:333–340.
65. **Svendsen, I., B. Martin, T. Viswanatha, and J. Johansen.** 1982. Amino acid sequence of carboxypeptidase Y. II. Peptides from enzymatic cleavages. Carlsberg Res. Commun. **47**:15–27.
66. **Taussig, R., and M. Carlson.** 1983. Nucleotide sequence of the yeast *SUC2* gene for invertase. Nucleic Acids Res. **11**:1943–1954.
67. **Trimble, R., and F. Maley.** 1977. The use of endo-β-N-acetylglucosaminidase H in characterizing the structure and function of glycoproteins. Biochem. Biophys. Res. Commun. **78**:935–944.
68. **Trimble, R., and F. Maley.** 1977. Subunit structure of external invertase from *Saccharomyces cerevisiae*. J. Biol. Chem. **252**: 4409–4412.
69. **Trimble, R., F. Maley, and F. Chu.** 1983. Glycoprotein biosynthesis in yeast: protein conformation affects processing of high mannose oligosaccharides on carboxypeptidase Y and invertase. J. Biol. Chem. **258**:2562–2567.
70. **Ulane, R., and E. Cabib.** 1976. The activating system of chitin synthetase from *Saccharomyces cerevisiae*. Purification and properties of the activating factor. J. Biol. Chem. **251**:3367–3374.
71. **Vasantha, N., L. Thompson, C. Rhodes, C. Banner, J. Nagle, and D. Filpula.** 1984. Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bacteriol. **159**:811–819.
72. **von Heijne, G.** 1985. Signal sequences: the limits of variation. J. Mol. Biol. **184**:99–105.
73. **Wang, S.-S., F. Carpenter.** 1967. Kinetics of the tryptic hydrolysis of the oxidized B chain of bovine insulin. Biochemistry **6**:215–224.
74. **Wells, J., E. Ferrari, D. Henner, D. Estell, and E. Chen.** 1983. Cloning, sequencing and secretion of *Bacillus amyloliquefacians* subtilisin in *Bacillus subtilis*. Nucleic Acids Res. **11**: 7911–7925.
75. **Wiemken, A., M. Schellenberg, and K. Urech.** 1979. Vacuoles: the sole compartments of digestive enzymes in yeast (*Saccharomyces cerevisiae*)? Arch. Microbiol. **123**:23–35.
76. **Wong, S.-L., C. Price, D. Goldfarb, and R. Doi.** 1984. The subtilisin E gene of *Bacillus subtilis* is transcribed from a σ^{37} promoter *in vivo*. Proc. Natl. Acad. Sci. USA **81**:1184–1188.
77. **Woolford, C., L. Daniels, F. Park, E. Jones, J. Van Arsdell, and M. Innis.** 1986. The *PEP4* gene encodes an aspartyl protease implicated in the posttranslational regulation of *Saccharomyces cerevisiae* vacuolar hydrolases. Mol. Cell. Biol. **6**:2500–2510.
78. **Wright, C., R. Alden, and J. Kraut.** 1969. Structure of subtilisin BPN′ at 2.5 Angstrom resolution. Nature (London) **221**: 235–242.