

The *Saccharomyces cerevisiae* *BARI* gene encodes an exported protein with homology to pepsin

(yeast/mating factors/protease/secretion)

VIVIAN L. MACKAY*[†], S. K. WELCH*, MARGARET Y. INSLEY*, THOMAS R. MANNEY[‡], JULIE HOLLY*, GENA C. SAARI*[§], AND MICHAEL L. PARKER*[¶]

*ZymoGenetics, Inc., 2121 N. 35th Street, Seattle, WA 98103; and [‡]Department of Physics, Kansas State University, Manhattan, KS 66506

Communicated by Ira Herskowitz, September 14, 1987 (received for review October 10, 1986)

ABSTRACT *Saccharomyces cerevisiae* a cells secrete an extracellular protein, called “barrier” activity, that acts as an antagonist of α factor, the peptide mating pheromone produced by mating-type α cells. We report here the DNA sequence of *BARI*, the structural gene for barrier activity. The deduced primary translation product of 587 amino acids has a putative signal peptide, nine potential asparagine-linked glycosylation sites, and marked sequence similarity of the first two-thirds of the protein with pepsin-like proteases. Barrier activity was abolished by *in vitro* mutation of an aspartic acid predicted from this sequence homology to be in the active site. Therefore, barrier protein is probably a protease that cleaves α factor. The sequence similarity suggests that the first two-thirds of the barrier protein is organized into two distinct structural domains like those of the pepsin-like proteases. However, the *BARI* gene product has a third carboxyl-terminal domain of unknown function; deletion of at least 166 of the 191 amino acids of this region has no significant effect on barrier activity.

Saccharomyces cerevisiae a and α cells each secrete peptide mating pheromones (α factor and α factor) that act on cells of the opposite mating type to promote several physiological changes that precede cell fusion—e.g., the induction of cell-surface agglutination factors and the arrest of cells in the G₁ phase of the cell cycle (1). Mating-type α cells produce an additional extracellular activity (called “barrier” activity) that is an antagonist of α factor (2, 3). The recovery of α cells from α -factor arrest is facilitated in part by barrier activity, since α strains with mutations in the *BARI* gene lack the activity, are super-sensitive to α factor, and are slower to recover from G₁ arrest (4, 5). As *barl* mutants also mate less efficiently with α cells in mass mating mixtures, barrier activity apparently functions to establish optimal pheromone concentrations for conjugation.

It is unclear whether the extracellular barrier activity also exists in a cell-associated form (3) or if it is identical to a membrane-bound endopeptidase that cleaves α factor between leucine and lysine, the sixth and seventh amino acids (6). However, unlike barrier activity (2, 3), this endopeptidase was not detected in the culture medium. Some endopeptidase activity could be found in preparations from α and α/α cells, although barrier activity is not detected in these cell types (2, 3). Moreover, *BARI*, shown in the present paper to be the structural gene for the barrier protein, is transcribed only in α cells (ref. 7; V.L.M., unpublished).

S. cerevisiae naturally exports very few proteins to the culture medium and only four of these have been identified: α factor (8, 9), α factor (10, 11), killer toxin (12), and barrier activity (2, 3). The first three are relatively small peptides, but barrier protein has not yet been purified and characterized.

Therefore, we determined and analyzed the sequence of the *BARI* gene^{||} to learn more about the mature protein as well as features of its processing and export.

MATERIALS AND METHODS

Plasmids. YEp13 bears the *S. cerevisiae* *LEU2* gene for selection of transformants in either *S. cerevisiae* or *Schizosaccharomyces pombe* (13, 14). The promoter from the *S. cerevisiae* *ADH1* gene (15) encoding alcohol dehydrogenase I was obtained from plasmid AM5 (kindly donated by G. Ammerer, ZymoGenetics) and the transcription terminator from the *TPH1* gene (triose phosphate isomerase) (16) was isolated from plasmid M210 (also provided by G. Ammerer). Expression plasmids were constructed and transformed into *Escherichia coli* strains RR1 (17) and JM83 (18) by standard techniques.

Yeast Strains. *S. cerevisiae* strain XP635-10C (a *barl* *leu2-3,112 gal2*) was used for expression experiments. An isogenic Bar⁺ revertant XP635-10CR was constructed by cotransformation with YEp13 and a linear fragment containing the *BARI* gene (19); Leu⁺ transformants were screened for those that exhibited the Bar⁺ phenotype. The *BARI* gene replacement was confirmed by Southern blot analysis (20). *S. cerevisiae* transformations were done by the method of Beggs (21). *Schizo. pombe* strain PR118-14 (h⁻ *leu1*) was transformed as described by Russell (14).

Barrier Assay. Yeast strains were assayed for barrier activity as described by Manney *et al.* (22). Transformants to be assayed were applied to a selective (minus leucine) agar plate that had been overlaid with 0.75% agar containing RC629 cells (a *sst1-2 ade2 his6 met1 ural1 can1 cyh2 gal2; barl* and *sst1* mutations are in the same complementation group) (4) and enough α factor (Sigma) to arrest their growth (\approx 200 ng/ml). Barrier activity exported from the transformant colonies inactivates α factor surrounding the colony and thereby allows the α *barl* assay cells to grow up as a fringe of cells around the transformant colony.

DNA Sequencing. A 2750-base-pair fragment from plasmid pBAR2 was sequenced on both strands by a combination of the chemical cleavage procedure (23) and the phage M13 dideoxy procedure (24).

Computer Analysis. The deduced amino acid sequence of the *BARI* primary translation product was compared with

[†]To whom reprint requests should be addressed.

[§]Present address: Zoologisches Institut der Universität, Zurich, Winterthurststrasse 190, CH-8057 Zurich, Switzerland.

[¶]Present address: The Pennsylvania State University, 525 Biotechnology Institute Building, University Park, PA 16802.

^{||}The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03573).

sequences in the Protein Identification Resource.**

In Vitro Mutagenesis. Site-specific mutagenesis of Asp-287 was achieved by the methods of Zoller and Smith (25) using the following sense-strand mutagenic oligonucleotides: 5' CCAGTTTTATTAGCTTCAGGAACCT 3' (aspartic acid to glutamic acid) and 5' CCAGTTTTATTAGAATCAGGAACCT 3' (aspartic acid to alanine); the antisense-strand template was the 1.3-kilobase *Sal* I–*Xho* I fragment (nucleotides 752–2070 in Fig. 2) from pBAR2 subcloned in M13mp18. Oligonucleotides were synthesized with an Applied Biosystems DNA synthesizer. Following mutagenesis, the entire template was sequenced to confirm the presence of only the desired mutation.

Tagging Barrier Protein with a Peptide Antigen. A DNA sequence encoding 14 amino acids including the 5-amino acid substance P antigen (26) and a translation stop signal was obtained from H. R. B. Pelham (Medical Research Council Laboratory of Molecular Biology, Cambridge, U.K.). This sequence was ligated to the *BARI* coding sequence at the *Eco*RI site at nucleotide 1573 (Fig. 2), corresponding to amino acid 524 of the primary translation product. The *BARI* *Eco*RI protruding end had been filled in with *E. coli* polymerase (Klenow fragment) and ligated with *Bam*HI linkers treated with polynucleotide kinase (27); the newly generated *Bam*HI site was ligated in frame with the substance P oligonucleotide. A *Sma* I site 3' to the substance P translation stop signal was ligated to an *Xba* I–*Bam*HI fragment containing the transcription terminator from the *S. cerevisiae* *TPII* gene (16); the *Xba* I end of the terminator fragment had been filled in with Klenow polymerase. A *Pvu* II–*Bgl* II fragment including nucleotides 1268–1576 of *BARI*, substance P sequence, and the *TPII* terminator was ligated with an *Sph* I–*Pvu* II fragment containing the *S. cerevisiae* *ADHI* promoter (15) and nucleotides 1–1267 of the *BARI* coding sequence and inserted into YEp13 digested with *Sph* I and *Bam*HI. The wild-type *BARI* coding sequence and the two Asp-287 mutant DNAs were used in the constructions. Yeast cells transformed with the hybrid genes were grown in selective medium and the supernatants were analyzed on electrophoretic transfer blots with anti-substance P antibody (NC1/34 HL, Accurate Chemical & Scientific, Westbury, NY) as described by Munro and Pelham (26).

Enzymes. Restriction endonucleases, DNA ligase from bacteriophage T4, polynucleotide kinase, and DNA polymerases were obtained from Bethesda Research Laboratories, New England Biolabs, and Boehringer Mannheim and were used according to suppliers' recommendations.

RESULTS AND DISCUSSION

The *BARI* gene was cloned by *in vivo* complementation of a *bar1* mutation in yeast strain XP635-10C, using a library of *S. cerevisiae* total DNA constructed in the yeast shuttle vector YEp13 (28). *Bar*⁺ transformants were first selected for growth on 0.15–0.3 unit of α factor per ml, a concentration that is inhibitory to a *bar1* cells (22), and then screened for production of barrier activity. Plasmids pBAR2 and pBAR3 were isolated from two of the *Bar*⁺ yeast transformants, transformed into *E. coli*, and shown by restriction endonuclease digestion to contain a common region. Chromosomal integration experiments using a 3.2-kilobase fragment that complements the *bar1* mutation established that the cloned gene was *BARI* (data not shown; see also ref. 7).

We initially demonstrated that *BARI* is the structural gene for the extracellular barrier protein by transforming the fission yeast *Schizo. pombe* with plasmid pBAR2 and with

the vector YEp13 and assaying the transformants for barrier activity (Fig. 1A). Although *Schizo. pombe* cells transformed with YEp13 do not secrete any barrier-like activity (rows 4 and 5), the pBAR2 transformants produce an activity that permits the growth of some a *bar1* assay cells around the *Schizo. pombe* colonies (rows 1–3). The low level of activity exported by the *Schizo. pombe* pBAR2 transformants compared to *S. cerevisiae* pBAR2 transformants (compare Fig. 1A and B) presumably reflects in part incorrect transcription starts with the *S. cerevisiae* promoter in the fission yeast (14). The identification of *BARI* as the structural gene for barrier protein was confirmed in later experiments (see below).

Subsequent subcloning, complementation tests, and integration experiments demonstrated that the functional *BARI* gene and adjacent sequences necessary for its expression and regulation were contained within 2750 base pairs extending from an *Xba* I site to an *Xho* I site (ref. 7; T.R.M. and V.L.M., unpublished). As shown in Fig. 2, DNA sequencing of this region revealed only one long open reading frame beginning at position 1 and extending for 1761 base pairs. This was demonstrated to be the *BARI* coding sequence by expression experiments in which this open reading frame was fused at the second *Xba* I site (position –16) to the *S. cerevisiae* *ADHI* promoter (15). Yeast transformed with a plasmid bearing the fusion of the *ADHI* promoter and the *BARI*

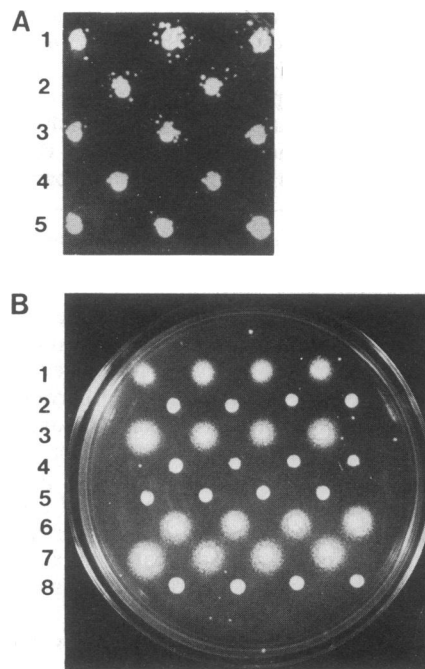


FIG. 1. Extracellular barrier activity produced by yeast transformants with the *BARI* gene or mutant forms. (A) *Schizo. pombe* strain PR118-14 transformed with pBAR2 (rows 1–3) or with the vector YEp13 (rows 4 and 5). (B) Row 1, *S. cerevisiae* strain XP635-10CR (a *BARI*) transformed with the vector YEp13. Rows 2–8, isogenic a *bar1* mutant strain XP635-10C transformed with YEp13 plasmids containing forms of the *BARI* gene as indicated: row 2, YEp13 vector control; row 3, wild-type *BARI* gene; row 4, Asp-287 mutated to glutamic acid; row 5, Asp-287 mutated to alanine; row 6, wild-type *BARI* coding sequence with the *ADHI* promoter; row 7, deletion of 166 amino acids from the *BARI* carboxyl terminus (*ADHI* promoter); row 8, deletion of 199 amino acids from the *BARI* carboxyl terminus (*ADHI* promoter). Plasmids were constructed by standard methods and contain either the *BARI* promoter or, where indicated, the *S. cerevisiae* *ADHI* promoter (15). The two carboxyl-terminal deletion genes utilize the transcription terminator from the *S. cerevisiae* triose phosphate isomerase (*TPII*) gene (16) in place of the *BARI* terminator. Yeast transformants were transformed with the expression plasmids and assayed for extracellular barrier activity (22).

**Protein Identification Resource (1985) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 6.

-680 -670
TCTAGAAGAAC AATTTGACAA

-650 -640 -630 -620 -610 -600 -590 -580 -570 -560
TGTGCTGTTG AGATACGGCA ACGAGTTGGA AGTGTGTAGC AAGGATCTAA TAACAGAGATT TGATGTGGAA GATGAAAGG ACATCTACGA GCGTAACTAT TGCAACGAAA

-540 -530 -520 -510 -500 -490 -480 -470 -460 -450
TGCTTTTGA GTTCTGTATT GAGTTAGACA GATGATGATT GATTAACGTG GATGAGTCTT TAAGAGGCGC GTTGAAGGAA AAGAGGAAAG GCGTCATAAA GGAATTCAG

-430 -420 -410 -400 -390 -380 -370 -360 -350 -340
GCTATGTTAA AAGATTGGA CTCTCTTAAA TAAGTTGTTA TTTTATGTTT ATAGATAACG GCTCTTGCCG AATTCATAGG CTGCACTCAT TCCGCTAGCT ACACATTTGT

-320 -310 -300 -290 -280 -270 -260 -250 -240 -230
TGCAITTTATT ATATTAGAGA TGCSTTGTCC CTGTTTTTCT ACCTCCGACA TCATGCTGAA ACATGGCATG TAATTACCGT AAGAGGAAAT TACATGGCGA GTGTCCACATA

-210 -200 -190 -180 -170 -160 -150 -140 -130 -120
ATAGCGACAA TAACATGTAT ACACAGCCAG CTATTCTGAA ACACACCACA TTATAGTTAT TGAATGTGTG TGTTTTTTGA TAACAGTATA AAGATGAAGA GAAAGCACGT

-100 -90 -80 -70 -60 -50 -40 -30 -20 -10
CGAGCCTTGT CATGATGAAT TCTTTAATGA TCTTCCGCTG ATTTAATCT AGTGGTCTGT ATCGCCTAAA ATCATAACAA AATAAAAAGA GTGTCTAGAA GGGTCATATA

+1 10 20 30 40 50 60 70 80 90
ATG TCT GCA AAT AAT CAT CTT TGT TTG AAA CTT ATT TTG GCG AGT TTC GCG ATT ATT AAC ACC ATT ACT GCT TTA ACA AAG GAT GGC ACT
MET Ser Ala Ile Asn His Leu Cys Leu Lys Leu Ile Leu Ala Ser Phe Ala Ile Ile Asn Thr Ile Thr Ala Leu Thr Asn Asp Gly Thr
1 10

100 110 120 130 140 150 160 170 180
GGT CAC TTA GAA TTC CTT TTA CAA CAC GAA GAG GAG ATG TAT TAC GCA ACA ACC TTA GAT ATA GGT ACA CCG TCC CAA AGT CTG ACA GTG
Gly His Leu Glu Phe Leu Leu Gln His Glu Glu Glu Met Tyr Tyr Ala Thr Thr Leu Asp Ile Gly Thr Pro Ser Gln Ser Leu Thr Val
40 50 60

190 200 210 220 230 240 250 260 270
TTG TTT GAT ACC GGA TCT GCC GAT TTT TGG GTT ATG GAT TCT AGC AAT CCC TTC TGC TTA CCA AAT TCA AAT ACG TCA TCC TAT TCA AAC
Leu Phe Asp Thr Gly Ser Ala Asp Phe Trp Val Met Asp Ser Ser Asn Pro Phe Cys Leu Pro Asn Ser Asn Thr Ser Ser Tyr Ser
70 80 90

280 290 300 310 320 330 340 350 360
GCA ACT TAT AAT GGC GAA GAA GTT AAG CCT TCA ATT GAT TGC AGG TCT ATG AGT ACT TAT AAT GAG CAT AGA TCT TCC ACC TCA CAA TAT
Ala Thr Tyr Asn Gly Glu Glu Val Lys Pro Ser Ile Asp Cys Arg Ser Met Ser Thr Tyr Asn Glu His Arg Ser Ser Thr Tyr Gln Tyr
100 110 120

370 380 390 400 410 420 430 440 450
CTG GAA AAT GGT AAG TTT TAC ATC ACA TAT GCT GAC GGA ACA TTT GCT GAC GGT AGT TGG GGG ACG GAA ACT GTA TCA AAT AAT GGA ATT
Leu Glu Asn Gly Arg Phe Tyr Ile Thr Lys Tyr Ala Asp Gly Thr Phe Ala Asp Gly Ser Trp Gly Thr Glu Thr Val Ser Ile Asn Gly Ile
130 140 150

460 470 480 490 500 510 520 530 540
GAC ATC CCC AAT ATC CAG TTC GGA GTT GCC AAG TAT GCT ACG ACA CCC GTT AGT GGT GTT CTT GGA ATT GGG TTT CCT AGA AGA GAG TCC
Asp Ile Pro Asn Ile Ile Gln Phe Gly Val Ala Lys Tyr Ala Thr Thr Pro Val Ser Gly Val Leu Gly Ile Gly Phe Pro Arg Arg Glu Ser
160 170 180

550 560 570 580 590 600 610 620 630
GTT AAG GGC TAT GAA GGT GCT CCT AAT GAA TAT TAT CCT AAT TTT CCT CAG ATT TTA AAA AGT GAA AAA ATA ATC GAT GTG GTC GCG TAT
Val Lys Gly Tyr Glu Glu Ala Pro Asn Glu Tyr Tyr Pro Asn Phe Pro Gln Ile Leu Lys Ser Glu Lys Ile Ile Asp Val Val Ala Tyr
190 200 210

640 650 660 670 680 690 700 710 720
TGB CTG TTC TTA AAC TCA CCT GAT TCA GGT ACT GGT TCG ATT GTT TTT GGT GCC ATT GAT GAA TCA AAG TTT TCT GGT GAT TTG TTC ACT
Ser Leu Phe Leu Asn Ser Pro Asp Ser Gly Thr Gly Ser Ile Val Phe Gly Ala Ile Asp Glu Ser Lys Phe Ser Gly Asp Leu Phe Thr
220 230 240

730 740 750 760 770 780 790 800 810
TTC CCT ATG GTA AAT GAA TAT CCC ACA ATA GTC GAC GCT CCT GCA ACT TTA GCA ATG ACT ATA CAA GGA TTA GGT GCC CAA AAG AAA AGT
Phe Pro Met Val Asn Glu Tyr Pro Thr Ile Val Asp Ala Pro Ala Thr Leu Ala Met Thr Ile Gln Gly Leu Gly Ala Gln Asn Lys Ser
250 260 270

820 830 840 850 860 870 880 890 900
AGT TGT GAA CAT GAA ACG TTT ACG ACG ACC AAG TAT CCA GTT TTG TGG GAC TCA GGA ACC TCG CTA TTG AAT GCG CCC AAG GTC ATA GCA
Ser Cys Glu His Glu Thr Phe Thr Thr Thr Lys Tyr Pro Val Leu Leu Asp Ser Gly Thr Ser Leu Leu Asn Ala Pro Lys Val Ile Ala
280 290 300

910 920 930 940 950 960 970 980 990
GAT AAA ATG GCT TCT TTT GTA AAT GCG TCC TAT AGT GAA GAG GAA GGT ATA TAT ATA TTA GAC TGT CCA GTA TCT GTA GGT GAC GTG GAA
Asp Lys Met Ala Ser Phe Val Asn Ala Ser Tyr Ser Glu Glu Glu Gly Ile Tyr Ile Leu Asp Cys Pro Val Ser Val Gly Asp Val Glu
310 320 330

1000 1010 1020 1030 1040 1050 1060 1070 1080
TAC AAT TTT GAT TTC GGC GAT TTG CAA ATA AGT GGT CCA CTG TCT AGT TTG ATT TTA ACT CCC GAG ACA GGC AGC TAT TGT GGG TTT
Tyr Asn Phe Asp Phe Gly Asp Leu Gln Ile Ser Val Pro Leu Ser Ser Leu Ile Leu Ser Pro Glu Thr Glu Gly Ser Tyr Cys Gly Phe
340 350 360

1090 1100 1110 1120 1130 1140 1150 1160 1170
GCG GTC CAG CCA ACA AAC GAT TCG ATG GTT CTG GGT GAT GTG TTC CTG TCC TCT GCA TAC GTC GTA TTC GAT CTC GAT AAT TAT AAG ATA
Ala Val Gln Pro Thr Asn Asp Ser Met Val Leu Gly Asp Val Phe Leu Ser Ser Ala Tyr Val Val Phe Asp Leu Asp Asn Tyr Lys Ile
370 380 390

1180 1190 1200 1210 1220 1230 1240 1250 1260
TCT TTA GCA CAG GCA AAT TGG AAC GCA ACG GAA GTT TCG AAA AAG CTA GTA AAT ATT CAA ACA GAT GGG TCT ATT TCA GGT GCC AAA ATT
Ser Leu Ala Gln Ala Asn Trp Asn Ala Ser Glu Val Ser Lys Lys Leu Val Asn Ile Gln Thr Asp Gly Ser Ile Ser Gly Ala Lys Ile
400 410 420

1270 1280 1290 1300 1310 1320 1330 1340 1350
GCT ACA GCT GAA CCC TGG TCC ACC AAT GAA CCA TTT ACA GTC ACC TCT GAC ATT TAT TCA TCT ACA GGC TGC AAG AGT AAG CCT TTT CTT
Ala Thr Ala Glu Pro Trp Ser Gly Thr Asn Glu Pro Phe Thr Val Thr Ser Asp Ile Tyr Ser Ser Thr Gly Cys Lys Ser Arg Pro Phe Leu
430 440 450

1360 1370 1380 1390 1400 1410 1420 1430 1440
CAA TCA TCG ACA GCC TCT TCG CTT ATT GCA GAA ACA AAC GTA CAA AGT CCG AAC TGC TCT ACG AAG ATG CCA GGC ACT AGA TCA ACT ACT
Gln Ser Ser Thr Ala Ser Ser Leu Ile Ala Glu Thr Asn Val Gln Ser Arg Asn Cys Ser Thr Lys Met Pro Gly Thr Arg Ser Thr Thr
460 470 480

1450 1460 1470 1480 1490 1500 1510 1520 1530
GTC TTA AGT AAG CCT ACT CAA AAT AGT GCT ATG CAT CAA AGT ACA GGT GCT GTC ACA AAC TCA AAT GAA ACT AAA TTA GAA TTA TCC
Val Leu Ser Lys Pro Thr Gln Asn Ser Ala Met His Gln Ser Thr Gly Ala Val Thr ACT Thr Ser Asn Glu Thr Lys Leu Glu Leu Ser
490 500 510

1540 1550 1560 1570 1580 1590 1600 1610 1620
TGB ACT ATG GCA AAT TCG GGC AGT GTC TCG CTT CCC ACT TCG AAT TCA ATA GAC AAA GAG TTC GAA CAT TCG AAA TCT CAA ACT ACC AGC
Ser Thr Met Ala Asn Ser Gly Ser Val Ser Leu Pro Thr Ser Asn Ser Ile Asp Lys Glu Phe Glu His Ser Lys Ser Gln Thr Thr Ser
520 530 540

1630 1640 1650 1660 1670 1680 1690 1700 1710
GAT CCA AGT GTA GCG GAG CAT TCT ACG TTT AAC CAA ACG TTT GTA CAT GAA ACT AAA TAT CCG CCT ACT CAA AAG ACA GTC ATA ACA GAA
Asp Pro Ser Val Ala Glu His Ser Thr Phe Asn Gln Thr Phe Val His Glu Thr Lys Tyr Arg Pro Thr His Lys Thr Val Ile Thr Glu
550 560 570

1720 1730 1740 1750 1760 1770 1780 1790 1800
ACT GTC ACG AAG TAT TCT ACA GTC TTA ATA AAT GTC TGT AAA CCA ACA TAT TAAGAAAT CTGAGTACA ATTTCTTTAT AGCATATAAA TATCAAAAT
Thr Val Thr Lys Tyr Ser Thr Val Leu Ile Asn Val Cys Lys Pro Thr Tyr
580

1820 1830 1840 1850 1860 1870 1880 1890 1900 1910
ATAGTCATTTT TTATACATG GAAAGCATAA TAAAAAACA AGGGGAGTIT TACTGATATC ATTTGGATAT AATAAACAAA ATAGTAAAT TATGACCCA TCACAAATTT

1930 1940 1950 1960 1970 1980 1990 2000 2010 2020
GAACAGTAG CACATTTATC ATTGAATAAA TGCTAAAAAA ATCTCCCCGA CGGGGAATG AACCCCGATC TGCCACGCGA CAAGCGCCCA TTCTGACCAT TAAACTATCA

2040 2050 2060
CGAATATTA GATGTGATAC TGTGTATTA CCGGCTCGAG

Fig. 2. Nucleotide sequence of 2750 base pairs encompassing the *BAR1* coding and regulatory sequences and the deduced amino acid sequence of the primary translation product. The *MAT α 2* binding site is underlined, the putative signal peptide cleavage site after amino acid 24 is indicated with an arrowhead, the potential asparagine-linked glycosylation sites are starred, and the three proposed domains (by analogy with pepsin-like proteases) are delimited with brackets.

Protein	Precursor/ proenzyme	Active form	Sequence
BAR	45-71	21-47	YATTLDTGTPSQSLTVLFDTGSAADFVW
PEHU	76-102	14-40	Y T* IGTP*Q *TV*FDTGS* *WV
PEPO	59-85	18-44	Y T* IGTP*Q *TV*FDTGS* *WV
PEPG	58-84	14-40	Y T* IGTP*Q *TV*FDTGS* *WV
PECH	59-85	17-43	Y T* IGTP Q **V*FDTGS* *WV
REMSS	83-109	20-46	Y * IGTP Q** V*FDTGSA *WV
KHPGD		15-41	Y * IGTP Q *TV*FDTGS* *WV
CMBO	74-100	16-42	Y * *GTP Q *TVLFDTGSA*DFVW
CMUMF		20-46	YA * IGTP Q * *LFDTGSA*DW
PEPLBJ		17-41	Y T * IG * L FDTGSAD*WV
BAR	284-310	260-286	VLLDSGTSLLNAPKVIADKMASFVNAS
PEHU	274-300	212-238	**D*GTSLL P * S * AS
PEPG	256-282	212-238	**D*GTSLL P * S * AS
PECH	257-283	215-241	**D*GTSLL P ** * S
REMSS	283-309	220-246	V**D*C*S** AP * * A
KHPGD		221-247	**D*GTS* P * **
CMBO	271-297	213-239	*LD*GTS L P * * A*
CMUMF		226-252	* *D*GT ** P A K*
PEPLBJ		210-236	* D*GT*LL * SV *
BAR	370-393	346-369	VLDGVFLSSAYVVFLLDNYKISLA
PEHU	362-385	300-323	*LGDVF* * VFD N * LA
PEPG	344-367	300-323	*LGDVF* Y VFD N K* LA
PECH	341-364	299-322	*LGDVF* YV*FD N K* L*
REMSS	375-398	312-335	VLG F* Y FD N *I *A
KHPGD		313-336	*LGDVF* Y VFD D ** LA
CMBO	355-378	297-320	*LGDVF* Y VFD N * LA
CMUMF		320-342	* G *FL V*D* N *I *A
PEPLBJ		297-320	**GD*FL S YVFD D * *A

FIG. 3. Amino acid homologies between barrier activity and pepsin-like proteases. Only identical amino acids are shown; conservative changes are denoted with an *. Proteins: BAR, barrier activity; PEHU, human pepsinogen; PEPO, bovine pepsinogen A; PEPG, porcine pepsinogen A; PECH, chicken pepsinogen; REMSS, mouse renin; KHPGD, porcine cathepsin D; CMBO, bovine prochymosin; CMUMF, mucor carboxyl protease; PEPLBJ, penicillopepsin. Amino acid numbering is shown for the enzymes in their proenzyme or precursor forms (primary translation product for BAR) and their active forms (BAR minus a 24-amino acid signal peptide). The three conserved regions are aligned with each other at the aspartic acid residues. Amino acids are shown in the one-letter code: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

coding sequence export as much or more barrier activity than yeast transformed with the wild-type *BAR1* gene (Fig. 1B, row 6 vs. row 3). The 5' noncoding sequence contains the *MATa2* binding site (nucleotides -235 to -264) (29, 30), which is responsible for repression of *BAR1* transcription in α and a/α cells and other elements necessary for *BAR1* expression (7). The 587-amino acid primary translation product includes a putative signal peptide with a predicted cleavage site after amino acid 24 (31) and nine potential asparagine-linked glycosylation sites, features that are consistent with the secretion and export of the polypeptide.

The deduced amino acid sequence of the *BAR1* primary translation product was compared with sequences in the Protein Identification Resource.** Strong similarities were found between short regions of the *BAR1* product and conserved sequences of members of the "acid protease" family or pepsin-like proteases (Fig. 3). It has been reported that the yeast vacuolar protease A and a number of retroviral proteases (including the protease encoded by the acquired immunodeficiency syndrome human immunodeficiency virus) are also homologous to pepsin (32-34). Additional homology was demonstrated by computer-generated alignment between porcine pepsinogen (35-38) and the first 396 amino acids of the *BAR1* product (Fig. 4). Their similarity extends over the entire length of pepsinogen (PEPG) and results from a high percentage of identities and an accumulation of many conservative replacements. However, relative to pepsinogen, the *BAR1* product has three large insertions between PEPG amino acids 86 and 87, 182 and 183, and 253 and 254. Comparison of the *BAR1* sequence with the three-di-

PEPG:	LVKVPLVRKSLQNLKDGKLDKFLKTHKHPASKYFPEAAALCDEPLENYLDTEYFGT (60)
BAR:	-----MSAINHLCKLIILASFVIINTI TALINDGCTHLEFLIQHEEMYATT (48)
	:*:
PEPG:	IGIGTPAQDFTVLFDTGSAADFVWVPSV-----YCSSLACSDHN (97)
BAR:	LDIGTPSQSLTVLFDTGSAADFVWVMDSSNPFCPLNSNTSSYNATYNGEEVKPSIDCRMS (108)
	:*****:
PEPG:	FPNPDSSSTFEA-TSQELSIITYGTGSM-TGLIYDVTQVQVCGISDTMQIFCLSHTEPFGSFL (155)
BAR:	TYNHRSSYQYLENPSFYIITYADGTFADGCSWGTETVSIINGIDIPHIQFAVAKYAT---- (164)
	::*:
PEPG:	YYAPFDGILGLAYPSISASGATPVDFN-----LMDQLVSVQDLFVYVSSNDSDGS (206)
BAR:	--TPVSGVLGIGFPPRRRESVKGYEGAPNEYYPNPFQILKTEKIIDVVAYSLPLNSPDSGTG (222)
	*:
PEPG:	VVLLGGIDSSYYTGLNWPVSVVEGYQITLDSITMDCTIACSGCC----- (253)
BAR:	SIVFGAIDESKFSGLFTFPVNE-----YPTIVDAPATLAMIQGLGAQNKSSCEHET (276)
	::*:
PEPG:	-----QAVDVTGTSLLTCTSAIANIQSDIGASENSDCEMVISCSSIDSLPDI VFTING (307)
BAR:	FTTTKYPVLLDSGTSLLNAPKVIADKMASFVNASYTEEEAYIILDSFV-SVGDVYMFDF (335)
	::*:
PEPG:	VQFYLSPSAYILQDDSDCTSCFEGMDVPTSSGELHILGDFVIRQYTVFDRANKVGLAP (367)
BAR:	GLAQISVFLSSLLSPETQCSYCGFAVQPTNDSM-VLDGVFLSSAYVVFLLDNYKISLAQ (394)
	:*:
PEPG:	VA (369)
BAR:	AN (396)

FIG. 4. Alignment of the amino acid sequence of porcine pepsinogen and the first 396 residues of the *BAR1* primary translation product. Porcine pepsinogen is cleaved to pepsin after amino acid 43 (36). The nine-gap alignment was formed and evaluated statistically by the Sankoff algorithm (39). Nonparametric analysis of incremental scores demonstrated that nine was the maximum number of gaps that was allowed on a statistical basis. Comparison of the score of the alignment shown here with the scores of 100 pairs of randomized sequences of the same lengths and amino acid compositions (39) demonstrated that the similarity between the *BAR1* product and pepsinogen is statistically highly significant. The alignment of the authentic sequences generated a score that was 16 standard deviations greater than the mean of the scores of the randomized sequences. Under the assumption of a normal distribution, this deviation corresponds to a probability of 10^{-57} . Asterisks (*) indicate identities; colons (:) indicate conservative replacements—i.e., amino acid pairs having greater-than-average frequency in the scheme of McLachlan (40). The two active site aspartic acid residues are indicated with closed circles (●). Amino acids are shown in the one-letter code: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

mensional structure of pepsinogen (41) indicates that these may be surface loops and that the hydrophobic core of pepsin is conserved in the *BAR1* protein. The pronounced sequence similarity between the two proteins strongly suggests that the part of the *BAR1* product shown in Fig. 4 is homologous (i.e., coancestral) with pepsin and that it functions as a proteolytic enzyme.

Two of the regions with strong homology between *BAR1* and pepsin-like proteases contain the two reactive site aspartic acid residues (corresponding to positions 63 and 287 in the *BAR1* precursor) that are characteristic of pepsin-like proteases (38, 42-46). We mutated one of these putative active site residues to test whether it is necessary for barrier activity. Asp-287 was mutated *in vitro* (25) to glutamic acid and to alanine and the mutant *BAR1* genes were transformed into yeast. As shown in Fig. 1B (rows 4 and 5), both mutations abolished detectable extracellular barrier activity.

To determine if the mutant proteins were still exported to the medium, we constructed hybrid genes in which DNA encoding the carboxyl-terminal 63 amino acids of the *BAR1* primary translation product was replaced with a synthetic DNA that codes for a 14-amino acid sequence containing a part of the neuropeptide substance P that can be detected with a monoclonal antibody (26). Yeast transformed with the

hybrid derived from the wild-type *BARI* gene exported approximately as much active barrier protein as cells transformed with the full-length *BARI* gene (data not shown), indicating that the carboxyl-terminal 63 amino acids are not required for barrier activity or export. As expected, yeast transformed with substance P hybrid genes derived from the Asp-287 mutant *BARI* genes did not produce active barrier protein. Electrophoretic transfer blot analysis of culture medium from cells transformed with the wild-type or mutant *BARI*-substance P hybrids indicated, however, that the mutant hybrid proteins were exported to the culture medium as efficiently as the wild-type hybrid protein (Fig. 5). Fig. 5 also demonstrates that the exported hybrid protein is heavily glycosylated, migrating as a heterogeneous band near the top of the gel with an apparent $M_r > 200,000$.

Pepsin-like proteases are organized into two distinct structural domains, each of which contains an active site aspartic acid. The two domains are related by a 2-fold axis of approximate rotational symmetry. Since the first two-thirds of the barrier sequence aligns well with the entire pepsinogen sequence, one can conclude tentatively that this region of barrier protein is also organized as two domains (indicated in Fig. 2). The role of the last third of the barrier protein is unclear, as a similar domain is not found in the other members of the pepsin family. This region should therefore not be required for the barrier protein to function as a protease. Accordingly, deletion of 166 amino acids from the carboxyl terminus of the *BARI* product does not significantly affect barrier activity (Fig. 1B, row 7). However, a deletion of 199 amino acids, which extends into the second domain, abolishes activity (Fig. 1B, row 8).

Conservation of other amino acids near the conserved aspartic acid residues, not only in pepsinogen but also in other members of the family, argues that other positions are essential for protease activity and/or structure. The ease with

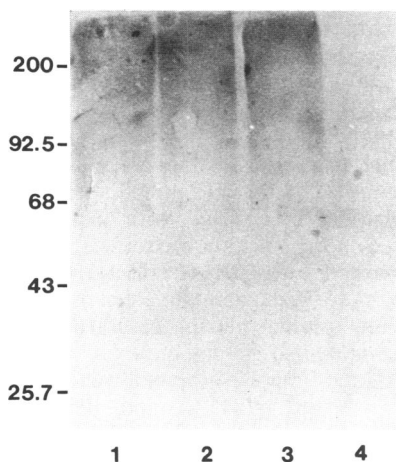


FIG. 5. Secretion of mutant barrier proteins. *S. cerevisiae* strain XP635-10C was transformed with plasmids bearing hybrid genes of the *ADHI* promoter-*BARI*-substance P oligonucleotide-*TPII* terminator. Transformants were grown in selective medium lacking leucine for 31 hr. Barrier protein was precipitated from 12.5 ml of the supernatant by the addition of an equal volume of cold 95% ethanol, and the dried pellet was resuspended in 200 μ l of sample buffer for electrophoresis (after denaturation) on a 10% polyacrylamide gel containing sodium lauryl sulfate. Following electrophoresis, the proteins were transferred to nitrocellulose and immunoblotted as described by Munro and Pelham (26). Lane 1, Asp-287 mutated to alanine; lane 2, Asp-287 mutated to glutamic acid; lane 3, wild-type *BARI*; lane 4, YEp13 vector. Molecular weights are shown as $M_r \times 10^{-3}$.

which specific or localized random mutations can be generated in the *BARI* gene and then reintroduced into yeast for assay suggests that this may be a useful system for analyzing the functions of active site and other conserved residues of pepsin-like proteases.

We are grateful to Hugh Pelham for providing the substance P oligonucleotides and to Ken Walsh, Patrick O'Hara, Gerald Reeck, Kirk Clark, and Rattan Nath for their assistance in the computer analysis. This research was supported in part by grants from the National Science Foundation (V.L.M.) and the National Institutes of Health (T.R.M.).

1. Sprague, G. F., Jr., Blair, L. C. & Thorner, J. (1983) *Annu. Rev. Microbiol.* **37**, 623-660.
2. Hicks, J. B. & Herskowitz, I. (1976) *Nature (London)* **260**, 246-248.
3. Manney, T. R. (1983) *J. Bacteriol.* **155**, 291-301.
4. Chan, R. K. & Otte, C. A. (1982) *Mol. Cell. Biol.* **2**, 11-20.
5. Sprague, G. F., Jr., & Herskowitz, I. (1981) *J. Mol. Biol.* **153**, 305-321.
6. Ciejak, E. & Thorner, J. (1979) *Cell* **18**, 623-635.
7. Kronstad, J. W., Holly, J. A. & MacKay, V. L. (1987) *Cell* **50**, 369-377.
8. Betz, R., MacKay, V. L. & Duntze, W. (1977) *J. Bacteriol.* **132**, 462-472.
9. Strazdis, J. R. & MacKay, V. L. (1982) *J. Bacteriol.* **151**, 1153-1161.
10. Duntze, W., MacKay, V. L. & Manney, T. R. (1970) *Science* **168**, 1472-1473.
11. Duntze, W., Stötzler, D., Bücking-Throm, E. & Kalbitzer, S. (1973) *Eur. J. Biochem.* **35**, 357-365.
12. Palfree, R. & Bussey, H. (1979) *Eur. J. Biochem.* **93**, 487-493.
13. Broach, J. R., Strathern, J. N. & Hicks, J. B. (1979) *Gene* **8**, 121-133.
14. Russell, P. R. (1983) *Nature (London)* **301**, 167-169.
15. Ammerer, G. (1983) *Methods Enzymol.* **101**, 192-201.
16. Alber, T. & Kawasaki, G. (1982) *J. Mol. Appl. Genet.* **1**, 419-434.
17. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heynecker, H. L., Boyer, M. W., Crosa, J. H. & Falkow, S. (1977) *Gene* **2**, 95-113.
18. Messing, J. (1983) *Methods Enzymol.* **101**, 20-78.
19. Rothstein, R. J. (1983) *Methods Enzymol.* **101**, 202-211.
20. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503-517.
21. Beggs, J. D. (1978) *Nature (London)* **275**, 104-109.
22. Manney, T. R., Jackson, P. & Meade, J. (1983) *J. Cell Biol.* **96**, 1592-1600.
23. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560-564.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
25. Zoller, M. J. & Smith, M. (1984) *DNA* **3**, 479-488.
26. Munro, S. & Pelham, H. R. B. (1984) *EMBO J.* **3**, 3087-3093.
27. Maniatis, T., Fritsch, E. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
28. Nasmyth, K. A. & Reed, S. I. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2119-2123.
29. Johnson, A. D. & Herskowitz, I. (1985) *Cell* **42**, 237-247.
30. Miller, A. M., MacKay, V. L. & Nasmyth, K. A. (1985) *Nature (London)* **314**, 598-603.
31. von Heinje, G. (1985) *J. Mol. Biol.* **184**, 99-105.
32. Ammerer, G., Hunter, C. P., Rothman, J. H., Saari, G. C., Valls, L. A. & Stevens, T. H. (1986) *Mol. Cell. Biol.* **6**, 2490-2499.
33. Woolford, C. A., Daniels, L. B., Park, F. J., Jones, E. W., Van Arsdell, J. N. & Innis, M. A. (1986) *Mol. Cell. Biol.* **6**, 2500-2510.
34. Toh, H., Kikuno, R., Hayashida, H., Miyata, T., Kugimiya, W., Inouye, S., Yuki, S. & Saigo, K. (1985) *EMBO J.* **4**, 1267-1272.
35. Ong, E. B. & Perlmann, G. E. (1968) *J. Biol. Chem.* **243**, 6104-6109.
36. Stepanov, V. M., Baratova, L. A., Pugacheva, I. B., Belyanova, L. P., Revina, L. P. & Timokhina, E. A. (1973) *Biochem. Biophys. Res. Commun.* **54**, 1164-1170.
37. Moravsek, L. & Kostka, V. (1974) *FEBS Lett.* **43**, 207-211.
38. Sepulveda, P., Marciniuszyn, J., Jr., Liu, D. & Tang, J. (1975) *J. Biol. Chem.* **250**, 5082-5088.
39. Reeck, G. R. & Teller, D. C. (1985) in *Progress in Nonhistone Protein Research*, ed. Bekhor, I. (CRC, Boca Raton, FL), Vol. 2, pp. 1-21.
40. McLachlan, A. D. (1971) *J. Mol. Biol.* **61**, 409-424.
41. James, M. N. G. & Sielecki, A. R. (1986) *Nature (London)* **319**, 33-38.
42. Baudys, M. & Kostka, V. (1983) *Eur. J. Biochem.* **136**, 89-99.
43. Misono, K. S., Chang, J.-J. & Inagami, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4858-4862.
44. Sodek, J. & Hofmann, T. (1970) *Can. J. Biochem.* **48**, 1014-1016.
45. Bayliss, R. S., Knowles, J. R. & Wybrandt, G. B. (1969) *Biochem. J.* **113**, 377-386.
46. Chang, W.-J. & Takahashi, K. (1974) *J. Biochem.* **76**, 467-474.