The yeast α -factor receptor: structural properties deduced from the sequence of the *STE2* gene

---

Anne C.Burkholder and Leland H.Hartwell*

Department of Genetics, SK-50, University of Washington, Seattle, WA 98195, USA

Received 29 July 1985; Revised 23 October 1985; Accepted 29 October 1985

ABSTRACT

The *STE2* gene of the yeast *Saccharomyces cerevisiae* encodes a component of the receptor for the oligopeptide pheromone α -factor (1). We have cloned and determined the nucleotide sequence of the *STE2* gene. A sequence involved in the control of cell-type expression of the *STE2* gene (2,3) was found 5' of an open reading frame that could encode a protein of 431 amino acids. The predicted *STE2* protein contains seven hydrophobic segments, suggesting that the α -factor receptor is an integral membrane protein. No extensive homology at the primary sequence level was detected between the predicted *STE2* protein and other available protein sequences.

INTRODUCTION

Peptide pheromones control cell division and conjugation in the yeast *Saccharomyces cerevisiae* (4). The oligopeptide pheromone α -factor, secreted by *MAT α* cells, elicits three responses in *MAT α* cells: division arrest in the G₁ phase of the cell cycle, induction of cell surface agglutinins, and formation of an altered morphology termed shmooing or projection formation (reviewed in 4). A specific receptor for α -factor has been detected on *MAT α* cells (1). The *STE2* locus encodes a structural component of this receptor since temperature-sensitive *STE2* mutants display thermolabile α -factor binding activity (1). This receptor mediates the cellular responses to α -factor since mutation or deletion of the *STE2* gene eliminates all three responses (5, our unpublished results).

To investigate the nature of the α -factor receptor, we cloned the *STE2* gene by complementation. We present the nucleotide sequence of 1776bp of the *STE2* gene and an analysis of the deduced amino acid sequence. Although significant homology to other proteins at the primary sequence level was not detected, a comparison of the hydropathy profiles of the predicted *STE2* protein and known membrane proteins suggests that the α -factor receptor is an integral membrane protein. The distribution of hydrophobic domains is strikingly similar to

that proposed for the STE3 gene product (D. Hagen and G. Sprague, personal communication), which may be the yeast α -factor receptor (6).

MATERIALS AND METHODS

Yeast Strains and Media

All strains are from this laboratory. Genotypes: 2047-2-4 (MATa ste2-3^{ts} ade2 leu2-3,112 trp1), 2047-9-2 (MATa ste2-3^{ts} trp1 leu2-3,112 ade2), PT2 (MATa hom3 isol can1), 4277-7 (MATa ade2-1 bar1-1 cyh2 cry1 his4-580 lys2 trp1 tyr1 SUP4-3), 4279-14 (MATa ste2-3^{ts} ade2-1 bar1-1 cyh2 cry1 his4-580 lys2 trp1 tyr1 SUP4-3), 4345-32 (MATa ade2-1 bar1-1 cyh2 cry1 his4-580 lys2 trp1 tyr1 SUP4-3), 4053-5-4 (MATa his7 leu2-3,112, trp1-289 ura3-52). Media and genetic methods were as described by Sherman et al. (7).

Yeast Transformation and Selection of Plasmids ComPLEMENTING ste2

Strain 2047-9-2 was transformed by the method of Hinnen et al. (8) with a yeast genomic library constructed by Fried and Warner (9) containing BamHI fragments ligated into plasmid YEpl3 (10). Leu^+ transformants were regenerated in 1.5% Seaplaque Agar (FMC) at 34°C until colonies grew out to the surface of the plate. Mating ability was assayed by the following replica plate mating test. Transformants were replicated to a master plate and a test plate containing a lawn of MATa tester cells (strain PT2) and incubated at 22° or 34°C for 6-20 hours. The test plate was then replicated to minimal medium to select for the growth of diploids. Transformants capable of mating at both temperatures were recovered from the master plate.

α -Factor Binding and Quantitative Mating Assays

Quantitative mating assays were as described (5). ³⁵S-labeled α -factor was obtained from D. Jenness. Cells were grown at 22°C, prepared as previously described (1) in medium containing metabolic inhibitors (10mM Na₃, 10mM KF, and 10 mM TAME), and kept on ice between manipulations. Heat-treated samples were held at 49° for 30 minutes. Treated or untreated cells (1×10^9 cells/ml) and radioactive α -factor (1×10^{-10} M) were incubated for 30 minutes at 22°C. 90 μ l aliquots (containing about 400 cpm radioactive ligand) were then removed and assayed for the amount of radioactivity bound by the procedure described previously (1).

Plasmid Construction

The 4.3kb STE2 BamHI fragment (Figure 1A) was cloned into YEpl3 (10) and YRp7 (11) to construct plasmids pAB539 and pAB503, respectively. SalI or BamHI-SalI digested plasmid pAB539 was used to generate the plasmids in Figure 1B, using the vector pRC3 and strategy of Ferguson et al. (12). These

plasmids were transformed into yeast by the lithium acetate method (13).

DNA Sequence Analysis

A nested set of sequencing endpoints was generated by the method of Henikoff (14) from BamHI, SstI doubly digested M13mp19 (15) derivatives containing the 1.6 kilobase HindIII fragment of the STE2 gene. A 0.9 kilobase Sall-BamHI fragment was cloned into M13mp18 (15) to create an endpoint in one region not covered by the nested set. Dideoxy sequencing reactions (16) of the M13 derivatives were analyzed on buffer gradient gels (17).

DNA restriction fragments, isolated from plasmids containing either the entire 4.3 kilobase STE2 BamHI fragment (Figure 1A) or a subclone of this fragment generated by Sall or Bal31 digestion, were sequenced by the method of Maxam and Gilbert (18). The sites for end-labeling were located in vector or insert sequences.

DNA sequences, and protein sequences deduced from the DNA sequence, were analyzed using the PCS program of L.M. Lagrimini and S.T. Brentano (University of Iowa, Iowa City, IA) and the program Microgenie from Beckman Instruments. The homology search to the protein sequence database (release 4.0) of the Protein Identification Resource was performed by J. Brown (Oncogen, Seattle, WA) using the strategy as described by Wilbur and Lipman (19). The program Apollo of B.R. Copeland (Utah State University, Logan, UT) was used to calculate and plot hydrophobicity.

General

Standard procedures for recombinant DNA work were as described in Maniatis et al. (20). Restriction and DNA modifying enzymes were obtained from New England Biolabs or Bethesda Research Laboratories. Radionucleotides were from New England Nuclear.

RESULTS

Cloning of the STE2 Gene

We screened plasmids from a yeast genomic library for their ability to complement the mating defect of a ste2-3^{ts} mutant. Two out of approximately 4000 Leu⁺ transformants assayed showed wild-type mating ability that was plasmid-dependent: both transformants segregated Leu⁻Ste⁻ colonies, and plasmid isolated from each would transform the ste2-3^{ts} strain 2047-9-2 to Ste⁺. A 4.3kb BamHI fragment with the restriction map shown in Figure 1A was the only yeast fragment common to both plasmids. A plasmid containing this insert alone (pAB539) would complement the mating defect of strain 2047-9-2. We mapped the segment of the BamHI insert that contains the STE2 function to a

Figure 1. Restriction map and subclones of 4.3 kb BamHI *STE2* DNA fragment. A) The cleavage sites for BamHI (B), SalI (S), MspI (M), and HindIII (H) are shown. Not all MspI sites are shown. Asterisk marks the location of the sequence homologous to the $\alpha 2$ protein binding sequence. Horizontal arrow shows the direction and extent of the open reading frame. B) Horizontal lines indicate the restriction fragments subcloned into the plasmids named at left (see Methods). Each plasmid was used to transform a *MATa ste2^{ts} trp1* strain (2047-2-4) to *Trp*⁺. Transformants were assayed for fertility by the replica plate mating test. A (-) score for complementation indicates that mating at the restrictive temperature was not above the background level of strain 2047-2-4. A (+) score indicates that mating was equivalent at the permissive and restrictive temperatures.

2.3kb SalI fragment (Figure 1B). All plasmids retaining this SalI fragment will complement the *ste2* mating defect, and the plasmid containing only this fragment (pAB520) confers *STE2* function. Complementation is neither allele-specific nor dependent on high copy number since the 2.3kb SalI fragment also complements the *ste2-8* and *ste2-9* alleles when present on the low copy-number plasmid pYE(CEN3)30 (21, data not shown).

Evidence that complementation is due to *STE2* sequences on this fragment, and not a suppressor of the *ste2-3* mutation, was provided by examining the site of chromosomal integration of plasmid pAB503 (*STE2 TRP1*) into *ste2-trp1*⁻ strains. If the plasmid contains the *STE2* gene it should integrate at the *ste2* locus. Integration at the *trp1* locus is also expected due to homology with the yeast *TRP1* gene contained in the vector sequences. The ratio of *MATa* spore types (*Trp*⁺*Ste*⁺ : *Trp*⁺*Ste*⁻ : *Trp*⁻*Ste*⁺ : *Trp*⁻*Ste*⁻) resulting from a cross of an integrated strain to a *STE2 TRP1* strain is expected to be 3:0:1:0 if integration was at the *ste2* locus, 3:1:0:0 for integration at *trp1*, and 5:1:1:1 if integration occurred at a locus unlinked to either marker. The results of this analysis (Table 1) demonstrate that the yeast sequences on plasmid pAB503 share homology to *STE2* and *TRP1*.

To examine the extent of complementation, fertility and receptor stability were assayed in the *MATa ste2-3^{ts}* strains containing plasmid pAB503 or the

TABLE 1. Analysis of Plasmid Integration Sites

Strain	Integrated Plasmid		MATa Spore Phenotype ⁽¹⁾			
			Trp ⁺ Ste ⁺	Trp ⁺ Ste ⁻	Trp ⁻ Ste ⁺	Trp ⁻ Ste ⁻
			<u>INTEGRATION AT ste2</u>			
B3-1(2)	pAB503	observed	59	0	18	1 ⁽³⁾
		expected	58.5	0	19.5	0
B3-12-3(4)	pAB503	observed	30	0	6	0
		expected	27	0	9	0
<u>INTEGRATION AT trp1</u>						
B3-3-4(4)	pAB503	observed	33	11	0	0
		expected	33	11	0	0
YR-4-3(4)	YRp7	observed	23	29	0	0
		expected	26	26	0	0

(1) The indicated strain was crossed to a MAT α STE2 TRP1 strain (PT2 or 4053-5-4) and the resultant diploids sporulated and dissected. Data are from tetrads with four viable spores. The analysis is limited to MATa spores since ste2 mutations do not affect MAT α cells. The Ste⁻ phenotype was temperature-sensitive. (2) Plasmid was integrated into strain 2047-2-4. (3) In this spore the plasmid was probably excised and subsequently lost, resulting in loss of both TRP1 and STE2. (4) Plasmid was integrated into strain 4277-14.

vector YRp7 integrated into the genome (Table 2). Integrants were chosen for this analysis to ensure that each cell in the assay contained a plasmid. The quantitative test of mating ability confirms the results from replica-plate mating tests that sequences on the 4.3kb BamHI fragment confer wild-type levels of fertility to a ste2 mutant. Mutant receptors in ste2-3^{ts} cells grown at permissive temperature can be distinguished from those in wild-type

Table 2. Complementation of Ste2 Mutant Phenotype

Strain and Relevant Genotype	Integrated Plasmid ⁽¹⁾	³⁵ S α -Factor Bound (cpm bound) ⁽²⁾		Mating Ability ⁽³⁾ (# diploids)
		untreated	heat-treated	
4277-7 MATa STE2	none	248 $\pm$ 17	258 $\pm$ 5	2.4 x 10 ⁵
B3-12-3 MATa ste2-3 ^{ts}	pAB503	337 $\pm$ 15	309 $\pm$ 5	2.5 x 10 ⁵
B3-3-4 MATa ste2-3 ^{ts}	pAB503	317 $\pm$ 9	329 $\pm$ 5	2.1 x 10 ⁵
YR-4-3 MATa ste2-3 ^{ts}	YRp7	255 $\pm$ 7	48 $\pm$ 2	1.3 x 10 ³
4345-32 MAT α STE2	none	16 $\pm$ 15	25 $\pm$ 1	not done

(1) Integrated at site shown in Table 1.

(2) Mean and standard deviation of triplicate assays.

(3) Cells were pre-grown and mated at the restrictive temperature (34°).

Figure 2. Sequencing strategy. Arrows denote the direction and extent of sequence reactions. A) The cleavage sites for BamHI (B), SalI (S), MspI (M), and HindIII (H) are shown. The position of the proposed initiation codon and the termination codon of the open reading frame are indicated. B) Sequences determined by the dideoxy chain-termination method. C) Sequences determined by the chemical cleavage method. Arrow with circle indicates sequence was read on both strands. Other arrows denote sequences read from a 3' end-labeled strand.

cells by their loss of α -factor binding activity after the cells are heated for 30 minutes at 49° (1). Heat-treatment reduces α -factor binding activity more than 80% in *ste2-3^{ts}* cells containing only vector sequences, but the cells with the plasmid pAB503 integrated show no loss of binding. These results provide further evidence that the 4.3kb BamHI fragment contains the *STE2* gene.

Sequence of the *STE2* Gene

The strategy outlined in Figure 2 was used to determine the sequence of the *STE2* gene. The sequence of 1776 nucleotides of the *STE2* gene is presented in Figure 3. It was not necessary to sequence the entire 2.3kb SalI fragment containing the *STE2* complementing activity (Figure 1), since transposon Tn5 insertions into this fragment that map to the region we have sequenced abolish *STE2* function, whereas insertions that map outside this region do not affect *STE2* activity (P. Talbert, A. Burkholder, and L. Hartwell, unpublished results). Our determination of the sequence of nucleotides -276 to -178, the only region for which we have data from one strand only, agrees with that reported for this region by Miller et al. (3).

The *STE2* gene contains a single long open reading frame and no matches for the consensus RNA splicing site in yeast, GGTATGT...TACTAACA...AG (22). The first two in-frame ATG codons (at nucleotides 1 and 160) have two characteristics thought to be important for efficient translation initiation, an A at position -3 and a T at position +6 relative to the A of the ATG codon (23,24). Although there is no direct evidence to indicate which of these ATG codons initiates protein synthesis, the position of the proposed $\alpha 2$ protein binding site suggests that the ATG at position 1 is more likely to be the

correct translation initiation codon. In MAT α cells, the $\alpha 2$ protein binds to the promoters of STE2, STE6, and other genes that specify MAT α cell type, to negatively regulate their expression at the level of mRNA production (2). An $\alpha 2$ protein binding region is located in the STE2 gene at nucleotides -199 through -231. This region is homologous to the $\alpha 2$ binding site of the STE6 gene (2), matching at 27 out of 33 nucleotides, although it is in the opposite orientation with respect to the direction of transcription as that found in the STE6 gene (K. Wilson and I. Herskowitz, personal communication). In the STE2 gene, the distance from the center of the proposed $\alpha 2$ binding sequence to the first and second ATG codons of the open reading frame (at nucleotides 1 and 160) is 215bp and 373bp, respectively. The distance from the center of the $\alpha 2$ binding region to the predicted start codon in four other MAT α specific genes is 195 to 250bp (2). Only the distance to the first in-frame ATG codon in STE2 falls within this interval, suggesting that it may be the site of translation initiation.

There are two TATA boxes 53bp and 59bp upstream of the ATG codon at nucleotide 1. If transcription is initiated 30-35bp downstream of TATA (nucleotides -19 to -30), a typical distance in eukaryotic genes (25), translation from the first AUG codon in the resultant messenger RNA would begin at position 1. However, since transcription in yeast sometimes begins over 100bp from TATA (26), and since there is another TATA box at nucleotide 49, it is possible that transcription, and therefore translation, initiates further downstream.

There are several stop codons in all three reading frames downstream of the predicted translation termination codon at nucleotide 1294. The sequence AATAAA, postulated to be important for polyadenylation of eukaryotic messenger RNAs (27), is at an appropriate position 147 bases downstream of the termination codon. Partial matches are found for the consensus sequence TAG..TAGT or TATGT..(AT-rich)...TTT that occurs in the 3' region of many yeast genes and may function in transcription termination (28). For example, TAGT..(13bp)...TTT appears at nucleotide 1458. The sequence TTTTATA, suggested to be important for transcription termination in yeast (29), does not appear, although there are several AT-rich regions. These features support the conclusion that we have correctly identified the translation termination codon.

Characteristics of the Predicted Amino Acid Sequence

Translation from the initiation codon at nucleotide 1 to the stop codon at nucleotide 1294 would result in a 431 amino acid protein with a predicted

	-270	-260	-250	-240	-230	-220	-210	-200	-190	
	TACTGCTTAGGACCTGTCCTGGCAAGTCGCAGATTGAAGTTTTTTTCAACCATGTAATTTTCCTAATTGGGTAAGTACATGATGAACACATA									
	-180	-170	-160	-150	-140	-130	-120	-110	-100	
	TGAAGAAAAAGCTTTCTCTACATATTCAAGATTTTTTTCTGTGGGTGGAATACTATTTAAGGAGCTGCTATTAGTATCTTTATTGACTTCAAAG									
	-90	-80	-70	-60	-50	-40	-30	-20	-10	
	CAATACGATACCTTTTTCTTTTCACTGCTCTGGCTATAATTATAATTGGTTACTTAAAAATGCACCGTTAAGAACCATATCCAAGAATCAAAA									
1	10	20	30	40	50	60				
ATG TCT GAT GCG GCT CCT TCA TTG AGC AAT CTA TTT TAT GAT CCA ACG TAT AAT CCT GGT CAA AGC ACC										23
Met Ser Asp Ala Ala Pro Ser Leu Ser Asn Leu Phe Tyr Asp Pro Thr Tyr Asn Pro Gly Gln Ser Thr										
70	80	90	100	110	120	130				
ATT AAC TAC ACT TCC ATA TAT GGG AAT GGA TCT ACC ATC ACT TTC GAT GAG TTG CAA GGT TTA GTT AAC										46
Ile Asn Tyr Thr Ser Ile Tyr Gly Asn Gly Ser Thr Ile Thr Phe Asp Glu Leu Gln Gly Leu Val Asn										
140	150	160	170	180	190	200				
AGT ACT GTT ACT CAG GCC ATT ATG TTT GGT GTC AGA TGT GGT GCA GCT GCT TTG ACT TTG ATT GTC ATG										69
Ser Thr Val Thr Gln Ala Ile Met Phe Gly Val Arg Cys Gly Ala Ala Ala Leu Thr Leu Ile Val Met										
210	220	230	240	250	260	270				
TGG ATG ACA TCG AGA AGC AGA AAA ACG CCG ATT TTC ATT ATC AAC CAA GTT TCA TTG TTT TTA ATC ATT										92
Trp Met Thr Ser Arg Ser Arg Lys Thr Pro Ile Phe Ile Ile Asn Gln Val Ser Leu Phe Leu Ile Ile										
280	290	300	310	320	330	340				
TTG CAT TCT GCA CTC TAT TTT AAA TAT TTA CTG TCT AAT TAC TCT TCA GTG ACT TAC GCT CTC ACC GGA										115
Leu His Ser Ala Leu Tyr Phe Lys Tyr Leu Leu Ser Asn Tyr Ser Ser Val Thr Tyr Ala Leu Thr Gly										
350	360	370	380	390	400	410				
TTT CCT CAG TTC ATC AGT AGA GGT GAC GTT CAT GTT TAT GGT GCT ACA AAT ATA ATT CAA GTC CTT CTT										138
Phe Pro Gln Phe Ile Ser Arg Gly Asn Val His Val Tyr Gly Ala Thr Asn Ile Ile Gln Val Leu Leu										
420	430	440	450	460	470	480				
GTG GCT TCT ATT GAG ACT TCA CTG GTG TTT CAG ATA AAA GTT ATT TTC ACA GGC GAC AAC TTC AAA AGG										161
Val Ala Ser Ile Glu Thr Ser Leu Val Phe Gln Ile Lys Val Ile Phe Thr Gly Asp Asn Phe Lys Arg										
490	500	510	520	530	540	550				
ATA GGT TTG ATG CTG ACG TCG ATA TCT TTC ACT TTA GGG ATT GCT ACA GTT ACC ATG TAT TTT GTA AGC										184
Ile Gly Leu Met Leu Thr Ser Ile Ser Phe Thr Leu Gly Ile Ala Thr Val Thr Met Tyr Phe Val Ser										
560	570	580	590	600	610	620				
GCT GTT AAA GGT ATG ATT GTG ACT TAT AAT GAT GTT AGT GCC ACC CAA GAT AAA TAC TTC AAT GCA TCC										207
Ala Val Lys Gly Met Ile Val Thr Tyr Asn Asp Val Ser Ala Thr Gln Asp Lys Tyr Phe Asn Ala Ser										
630	640	650	660	670	680	690				
ACA ATT TTA CTT GCA TCC TCA ATA AAC TTT ATG TCA TTT GTG CTC GTA GTT AAA TTG ATT TTA GCT ATT										230
Thr Ile Leu Leu Ala Ser Ser Ile Asn Phe Met Ser Phe Val Leu Val Val Lys Leu Ile Leu Ala Ile										
700	710	720	730	740	750					
AGA TCA AGA AGA TTC CTT GGT CTC AAG CAG TTC GAT AGT TTC CAT ATT TTA CTC ATA ATG TCA TGT CAA										253
Arg Ser Arg Arg Phe Leu Gly Leu Lys Gln Phe Asp Ser Phe His Ile Leu Leu Ile Met Ser Cys Gln										
760	770	780	790	800	810	820				
TCT TTG TTG GTT CCA TCG ATA ATA TTC ATC CTC GCA TAC AGT TTG AAA CCA AAC CAG GGA CAA GAT GTC										276
Ser Leu Leu Val Pro Ser Ile Ile Phe Ile Leu Ala Tyr Ser Leu Lys Pro Asn Gln Gly Thr Asp Val										
830	840	850	860	870	880	890				
TTG ACT ACT GTT GCA ACA TTA CTT GCT GTA TTG TCT TTA CCA TTA TCA TCA ATG TGG GCC ACG GCT GCT										299
Leu Thr Thr Val Ala Thr Leu Leu Ala Val Leu Ser Leu Pro Leu Ser Ser Met Trp Ala Thr Ala Ala										
900	910	920	930	940	950	960				
AAT AAT GCA TCC AAA ACA AAC ACA ATT ACT TCA GAC TTT ACA ACA TCC ACA GAT AGG TTT TAT CCA GGC										322
Asn Asn Ala Ser Lys Thr Asn Thr Ile Thr Ser Asp Phe Thr Thr Ser Thr Asp Arg Phe Tyr Pro Gly										
970	980	990	10							

1040	1050	1060	1070	1080	1090	1100	
GAC CTA TAT CCT AGA AGG AAG GAA ACA ACA TCG GAT AAA CAT TCG GAA AGA ACT TTT GTT TCT GAG ACT							
Asp Leu Tyr Pro Arg Arg Lys Glu Thr Thr Ser Asp Lys His Ser Glu Arg Thr Phe Val Ser Glu Thr							368
1110	1120	1130	1140	1150	1160	1170	
GCA GAT GAT ATA GAG AAA AAT CAG TTT TAT CAG TTG CCC ACA CCT ACG AGT TCA AAA AAT ACT AGG ATA							
Ala Asp Asp Ile Glu Lys Asn Gln Phe Tyr Gln Leu Pro Thr Pro Thr Ser Lys Asn Thr Arg Ile							391
1180	1190	1200	1210	1220	1230	1240	
GGA CCG TTT GCT GAT GCA AGT TAC AAA GAG GGA GAA GTT GAA CCC GTC GAC ATG TAC ACT CCC GAT ACG							
Gly Pro Phe Ala Asp Ala Ser Tyr Lys Glu Gly Glu Val Glu Pro Val Asp Met Tyr Thr Pro Asp Thr							414
1250	1260	1270	1280	1290	1300	1310	
GCA GCT GAT GAG GAA GCC AGA AAG TTC TGG ACT GAA GAT AAT AAT AAT TTA TGA TCAAAATTACGGCTTTGA							
Ala Ala Asp Glu Glu Ala Arg Lys Phe Trp Thr Glu Asp Asn Asn Asn Leu ---							431
1320	1330	1340	1350	1360	1370	1380	1390
AAAAATAATTCGTGACCTTCGGTATAAGGTTACTACTAGATTGAGTCTCATCAGATGCACACATTCTCTATAAAAAAATGGTATC							
1410	1420	1430	1440	1450	1460	1470	1480
TTTCTTATTTGATAATTTAACTCCTTTACATAATAACATCTCGTAAGTAGTGGTAGAAACCACTTTGCTTTTACGAGTTCAAGCTT							

Figure 3. Nucleotide and deduced amino acid sequence of STE2. The nucleotide sequence is numbered with position 1 at the first base of the proposed translation initiation codon. The first digit of the numeral above the sequence denotes the numbered base. The underlined nucleotides in the sequence from -199 to -231 are homologous to the $\alpha 2$ protein binding site (2). In the 5' untranslated region two possible "TATA" sequences are underlined. The sequence AATAAA, in the 3' untranslated region and a possible signal for mRNA polyadenylation, is doubly underlined. Amino acids are numbered at right. Asparagine residues at potential N-linked glycosylation sites are marked by an asterisk. Cysteine residues are underlined. Location of restriction sites noted in Figure 2: HindIII, -177 to -172 and 1492 to 1497; SalI, 1219 to 1224; MspI, 341 to 345.

molecular weight of 47,794. No significant homologies were detected in a comparison of the predicted STE2 amino acid sequence (Figure 3) to a protein sequence database (see methods). A comparison of the STE2 protein to the amino acid sequences recently obtained for the insulin receptor (30) and for the putative yeast α -factor receptor, deduced from the sequence of the STE3 gene (D. Hagen and G. Sprague, personal communication), again revealed no extensive homologies.

On the basis of the hydropathy of the deduced amino acid sequence (Figure 4), the STE2 protein can be divided into three regions. The amino-terminal region (residues 1 to approximately 50) is neither highly hydrophobic or hydrophilic. The amino terminus does not have a long stretch of hydrophobic residues flanked by charged amino acids, characteristic of a signal sequence (31). This region contains potential glycosylation sites (Asn-X-Thr or Asn-X-Ser)(32) at positions 25, 32, and 46.

The central region (approximately residues 51 to 301) is defined by the seven hydrophobic peaks. The calculated hydrophobicity (values above 1.36) of

Figure 4. Hydropathy plot of predicted STE2 protein. Hydropathy values are those of Kyte and Doolittle (33) converted to a scale where the value for Arg=0 and Ile=2.0. Window size is 19. Hydrophobicity above 1.36 indicates a potential membrane-spanning sequence (33).

the amino acid tracts corresponding to these peaks strongly suggest that there are seven transmembrane segments in this region of the STE2 protein (33). The three potential glycosylation sites in this region (positions 105, 205, and 301) are located in the relatively hydrophilic sequences between or adjacent to the hydrophobic segments. The only cysteine residues in the amino acid sequence are located in this region, at positions 59 and 252.

Residues 302 to 431 form a hydrophilic region. The carboxy-terminal 23 amino acids of this region are not essential for receptor function, since plasmid pAB520, and low copy-number plasmids containing the same *Sal*I fragment, have STE2 complementing activity yet do not contain the sequences 3' of the *Sal*I site (nucleotides 1219 to 1224) that encode residues 409 to 431 of the STE2 protein (Figures 1 and 3). The 3 or 16 residues substituted by translation into the adjacent vector sequences in these constructs (the number depends on the orientation of the fragment) do not resemble the missing amino acids (data not shown).

DISCUSSION

The hydropathic structure of the STE2 gene product strongly suggests that it is an integral membrane protein, and is consistent with its known role as a component of the α -factor receptor (1). However, the STE2 protein differs in several respects from mammalian peptide hormone receptors. The receptors for

insulin and epidermal growth factor are characterized by an amino-terminal signal sequence, an extracellular ligand-binding domain, a single transmembrane segment and a carboxy-terminal cytoplasmic domain containing kinase activity (30,34). The presence of several potential membrane-spanning regions of the STE2 protein suggests that the α -factor receptor may be folded with as many as seven transmembrane sequences. There does not appear to be an amino-terminal signal sequence in the STE2 protein. These differences, combined with the lack of significant homology at the primary level, suggest that the yeast α -factor receptor is both structurally and functionally different than the peptide hormone receptors involved in growth control in higher systems.

The transmembrane structure predicted for the α -factor receptor resembles that of bacteriorhodopsin and the acetylcholine receptor (35,36). This similarity raises the possibility that the α -factor receptor functions as a ligand-activated membrane channel. However, the presence of several transmembrane domains does not necessarily specify a function as a channel, since bovine rhodopsin has seven membrane-spanning segments and is thought to function by interacting with a cytoplasmic protein to trigger an enzymatic cascade (35,37). The predicted structure of the α -factor receptor resembles that of bovine rhodopsin in the number of transmembrane segments and the nature of the amino terminus, which is glycosylated and lacks a signal sequence (35,38).

MAT α cells recover from the arrest of cell division by α -factor after prolonged exposure to the pheromone, a result that indicates the presence of an adaptation process (39). Adaptation by covalent modification of the cytoplasmic domain of a receptor is a strategy employed in both bacterial and mammalian systems (40,41). The potential glycosylation sites in the amino terminus, and the potential for phosphorylation at the several serine, threonine, and tyrosine residues in the carboxy terminus of the STE2 protein allow for the possibility that covalent modification of the α -factor receptor plays a role in recovery. Adaptation could also occur by endocytosis of receptors, but the sequence provides no information on this point since no structural feature is known to be responsible for receptor endocytosis in coated pits (42).

Mutations in the STE3 gene only affect the fertility of MAT α cells, suggesting that it may encode the yeast α -factor receptor (6). The predicted STE3 protein, deduced from the sequence of the STE3 gene, has seven hydrophobic regions followed by a hydrophilic carboxy terminus (D. Hagen and

G. Sprague, personal communication). The similarity of the hydropathy plots of the predicted STE2 and STE3 proteins suggests that the two mating factor receptors have a similar structure, and may signal division arrest and other mating responses by a similar mechanism. However, the lack of significant sequence homology between the two suggests that they do not interact with a common protein.

ACKNOWLEDGEMENTS

We thank C. Furlong and R. Garber for advice on DNA sequencing, J. Brown and S. Eisenberg for assistance with the computer analysis, H. Fried for contributing the yeast genomic clone bank, D. Jenness for his gift of radioactive α -factor, and D. Hagen, G. Sprague, A. Johnson, K. Wilson, and I. Herskowitz for communication of results prior to publication. We also thank D. Burke, D. Koshland, and J. Konopka for their comments on the manuscript. This investigation was supported by NIH research grant number GM17709 from the Institute of General Medical Sciences and a grant from the American Business Research Foundation. A.C.B. was supported by Public Health Service Training Grant 5T32GM07735.

*To whom correspondence should be addressed

REFERENCES

1. Jenness, D.D., Burkholder, A.C., and Hartwell, L.H. (1984) *Cell* **35**, 521-529.
2. Johnson, A.D., and Herskowitz, I. (1985) *Cell*, **42**, 237-247.
3. Miller, A.M., MacKay, V.L., and Nasmyth, K.A. (1985) *Nature* **314**, 598-603.
4. Thorner, J. (1980) in *The Molecular Biology of the Yeast Saccharomyces cerevisiae: Life Cycle and Inheritance*, Strathern, J.N., Jones, E.W., and Broach J.R., Eds. pp 143-180, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
5. Hartwell, L.H. (1980) *J. Cell Biol.* **85**, 811-822.
6. MacKay, V., and Manney, R. (1974) *Genetics* **76**, 273-288.
7. Sherman, F., Fink, G., and Lawrence, C.W. (1979) *Methods in Yeast Genetics*. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
8. Hinnen, A., Hicks, J., and Fink, G.R. (1978). *Proc. Natl. Acad. Sci. USA* **75**, 1929-1933.
9. Fried, H.M., and Warner, J.R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 238-242.
10. Beggs, J.D. (1978) *Nature* **275**, 104-109.
11. Stinchcomb, D.T., Struhl, K., and Davis, R.W. (1979) *Nature* **282**, 39-43.
12. Ferguson, J., Groppe, J.C., Reed, S.I. (1981) *Gene* **16**, 191-197.
13. Ito, H., Fukada, Y., Murata, K. and Kimura, A. (1983). *J. Bacteriol.* **153**, 163-168.
14. Henikoff, S. (1984) *Gene* **28**, 351-359.
15. Norrander, J., Kempe, T., and Messing, J. (1983) *Gene* **26**, 101-106.

16. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
17. Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA **80**, 3963-3965.
18. Maxam, A.M., and Gilbert, W. (1980) Meth. Enzymol. **65**, 499-560.
19. Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA **80**, 726-730.
20. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
21. Fitzgerald-Hayes, M., Clarke, L., and Carbon, J. (1982) Cell **29**, 235-244.
22. Langford, C.J., and Gallwitz, D. (1983) Cell **33**, 519-527.
23. Kozak, M. (1984) Nucl. Acids Res. **12**, 857-872.
24. Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J. Kingsman, S.M., Perkins, R.E., Conroy, S.C., Dunbar, B., and Fothergill, L.A. (1982) Nucl. Acids Res. **10**, 2625-2637.
25. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349-383.
26. Sentenac, A. and Hall, B. (1982) in The Molecular Biology of the Yeast *Saccharomyces cerevisiae*: Metabolism and Gene Expression, Strathern J.N., Jones E.W., and Broach, J.R., Eds. pp. 561-606, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
27. Fitzgerald, M. and Shenk, T. (1981) Cell **24**, 251-260.
28. Zaret, K.S., and Sherman, F. (1982) Cell **28**, 563-573.
29. Henikoff, S., Kelly, J.D., Cohen, E.H. (1983) Cell **33**, 607-614.
30. Ullrich, A., Bell, J.R., Chen, E.Y., Herrera, R., Petruzzelli, L.M., Dull, T.J., Gray, A., Coussens, L., Liao, Y.-C., Tsubokawa, M., Mason, A., Seeburg, P.H., Grunfeld, C., Rosen, O.M. and Ramachandran, J. (1985) Nature **313**, 756-761.
31. Kreil, G. (1981) Annu. Rev. Bioch. **50**, 317-348.
32. Struck, D.K., Lennarz, W.J., and Brew, K. (1978) J. Biol. Chem. **253**, 5786-5794.
33. Kyte, J., and Doolittle, R.F. (1982) J. Mol. Biol. **157**, 105-132.
34. Ullrich, A., Coussens, L., Hayflick, J.S., Dull, T.J., Gray, A., Tam, A.W., Lee, J., Yarden, Y., Libermann, T.A., Schlessinger, J., Downward, J., Mayes, E.L.V., Whittle, N., Waterfield, M.D., and Seeburg, P.H. (1984) Nature **309**, 418-425.
35. Ovchinnikov, Y.A. (1982) FEBS Lett. **148** 179-191.
36. Finer-Moore, J., and Stroud, R.M. (1984) Proc. Natl. Acad. Sci. USA **81**, 155-159.
37. Stryer, L. (1984) Nature **312**, 498-499.
38. Nathans, J., and Hogness, D.S. (1983) Cell **34**, 807-814.
39. Moore, S.A. (1984) J. Biol. Chem. **259**, 1004-1010.
40. Koshland, D.E., Jr. (1981) Annu. Rev. Biochem. **50**, 765-782.
41. Miller, J.L., and Dratz, E.A. (1983) Biophys. J. **41**, 343a.
42. Yamamoto, T., Davis, C.J., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L., and Russel, D.W. (1984) Cell **39**, 27-38.