Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 7371–7375, October 1986
Genetics

Nucleotide sequence of the yeast regulatory gene *STE7* predicts a protein homologous to protein kinases

(yeast mating type/cell-type determination/data base search/*Saccharomyces cerevisiae*)

MICHAEL A. TEAGUE*, DEBORAH T. CHALEFF†, AND BEVERLY ERREDE*‡

*Department of Chemistry, University of North Carolina, Chapel Hill, NC 27514; and †E. I. du Pont de Nemours & Co., Central Research and Development Department, Experimental Station, Wilmington, DE 19898

Communicated by Russell F. Doolittle, June 12, 1986

ABSTRACT **The nucleotide sequence of the *STE7* gene from *Saccharomyces cerevisiae* has been determined. It is one of several regulatory genes implicated in the control of cell-type-specific gene expression in yeast. The predicted amino acid sequence shows homology to several members of the protein kinase family. These results suggest protein phosphorylation may play a key role in regulation leading to cell-type specialization in yeast.**

Yeast is a model system for the study of molecular mechanisms controlling cell-type specialization in eukaryotes. There are three cell types in yeast: **a**, α, and **a**/α. Whereas **a** and α cells are haploid and are specialized for mating, **a**/α cells are diploid and are specialized for meiosis and sporulation. With the exception of the mating type locus, *MAT*, all cells contain the same genetic information. The presence of either the **a** or the α allele at *MAT* determines whether cells will express the respective **a**-specific or α-specific genes and corresponding functions. The presence of both **a** and α alleles in diploid cells heterozygous at *MAT* causes repression of haploid-specific genes. (See ref. 1 for review.)

Expression of haploid-specific genes depends upon several regulatory determinants in addition to those encoded at *MAT*. Some of these have been identified by sterile (*ste*) mutations that prevent mating in both **a** and α haploid cells. The *STE7* gene was originally identified as a mutation that produced insensitivity to cell-division arrest induced by the yeast mating hormone, α-factor. Further characterization showed that mutation of *STE7* also causes a nonmating phenotype in both **a** and α cell types (2). It appears that the *STE7* gene product is involved in transcriptional regulation of both **a**-specific and α-specific genes. On the basis of steady-state RNA analyses in *STE7*$^+$ and *ste7*$^-$ genetic backgrounds, both the α-factor and the **a**-factor hormone structural genes, the α-factor and **a**-factor receptor genes, and a gene whose product is required for **a**-factor biosynthesis are under *STE7* control (D.T.C., S. Fields, and G. Sprague, unpublished observations). It has been shown that the *STE7* gene regulates expression of the yeast transposon Ty*1* and genes under Ty*1* control (3–5). Although these genes have no apparent role in cell-type specialization, their expression in yeast is haploid-specific (4–6). To begin characterization of the *STE7* gene product, we have determined the nucleotide sequence of *STE7*. We report that the *STE7*-encoded protein is homologous to members of the protein kinase family.

MATERIALS AND METHODS

***STE7* DNA.** Cloned *STE7* DNA was obtained from plasmids pSTE7.2 (7) and pSTE7.4, shown in Fig. 1. The 2.3-kb *Hin*dIII fragment (*a–d*) from pSTE7.2 and the 0.4-kb *Cla* I–*Hin*dIII fragment (*e–f*) from pSTE7.4 were isolated by agarose gel electrophoresis (8). Plasmid DNA was prepared by the method of Norgard (9). Restriction enzymes were purchased from New England Biolabs and used under conditions recommended by the vendor.

DNA Sequencing and Data Analyses. Sequencing was performed by the chain-termination method (10) according to "shotgun cloning" and sequencing procedures described in ref. 11. Bacteriophage M13mp18 (Bethesda Research Laboratories) treated with *Sma* I (New England Biolabs) and calf intestinal phosphatase (Boehringer Mannheim) was used for shotgun cloning of random fragments from the pSTE7.2 region *a–d* (Fig. 1). The 376-bp *Cla* I–*Hin*dIII fragment (*e–f*) from pSTE7.4 was subcloned into both M13mp18 and M13mp19 vectors. Each clone was sequenced according to described procedures (11). *Escherichia coli* K-12 strain JM109 (New England Biolabs) was used for M13 cloning and phage DNA production. A 17-nucleotide oligomer, 5′ GTAAAACGACGGCCAGT 3′, synthesized using an Applied Biosystems (Foster City, CA) model 380A synthesizer, was used as a primer for M13 sequencing reactions. The Klenow fragment of *E. coli* DNA polymerase I was obtained from New England Biolabs or United States Biochemical (Cleveland, OH). The sequencing reaction mixtures contained [α-[^{35}S]thio]dATP from Dupont–NEN Products. The products were analyzed by electrophoresis in 6% polyacrylamide/7 M urea gels with a Tris borate buffer gradient (12).

DNA sequence data were assembled using DBSYSTEM programs on a VAX-11/780 computer (13). The homology search FASTP algorithm of Lipman and Pearson (14) was used to search the Protein Identification Resource protein sequence library containing 3061 sequences (March 25, 1985). The FASTP algorithm assigns initial scores for regions of local homology and optimized scores for maximum alignments obtained by allowing deletions and insertions between compared sequences. The RDF program was used to determine z values for these initial and optimized scores (14). The z value is a measure of the statistical significance of the corresponding scores: z = (similarity score − mean of random scores)/(standard deviation of random scores).

mRNA 5′-End Mapping. Strains DC38 (*MAT***a** *STE7*$^+$ *his4-519 leu2 ura3 trp1 can1-101*) and DC127 (isogenic *ste7-Δ2* derivative of DC38 constructed by gene replacement) were grown at 30°C to 1–2 × 10^7 cells per ml in YPD medium [1% (wt/vol) Bacto-yeast extract/2% (wt/vol) Bacto-peptone/2% (wt/vol) dextrose]. Total RNA was isolated according to the glass-bead lysis and extraction method (15). The preparation was enriched for poly(A)$^+$ RNA by passage over oligo(dT)-cellulose (Collaborative Research, Waltham, MA) (16). A DNA oligomer, 5′ AAGTCTTTCGTTGAAACAT 3′, complementary to the mRNA was synthesized as before. The primer-extension reactions were performed essentially according to the procedure of Reeder *et al.* (17). The reference DNA sequence was determined using the end-labeled syn-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s).

‡To whom reprint requests should be addressed.

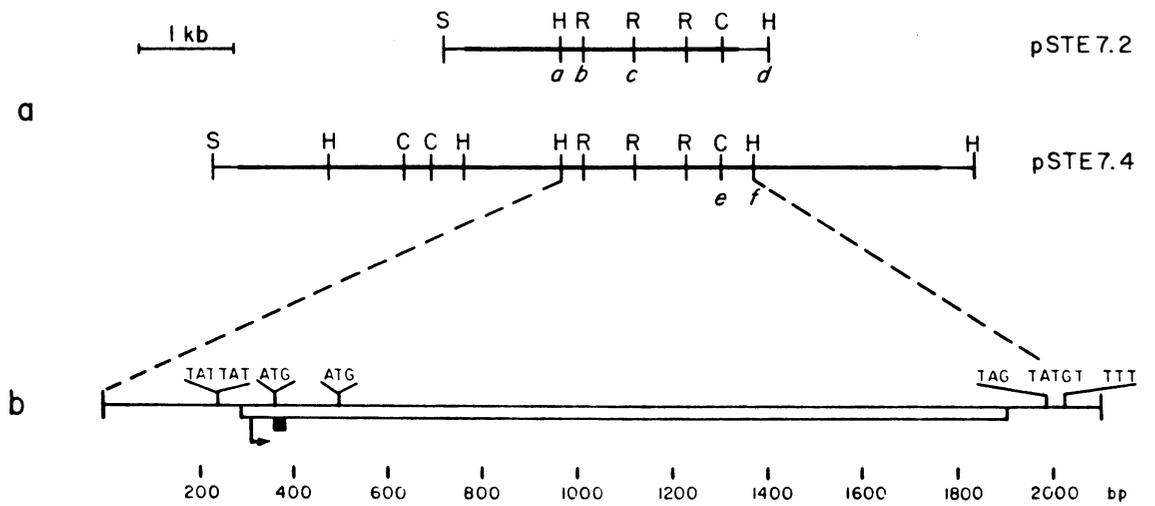

FIG. 1. (*a*) Restriction maps of the *STE7* region in the plasmids pSTE7.2 and pSTE7.4. Thick lines represent yeast DNA and thin lines represent flanking pBR322 DNA. *Cla* I (C), *Hin*dIII (H), and *Sal* I (S) restriction sites shown have been mapped in the designated plasmids. The positions for *Eco*RV (R) sites are known only for the 2.1-kilobase (kb) sequenced region. (*b*) Summary of transcriptional and translational features of the *STE7* region. The open box represents the open reading frame beginning at position 288 and ending at position 1907. The arrow indicates the major transcription initiation site at position 314. The solid box indicates the primer binding site (positions 360–379) for experiments shown in Fig. 3. See *Discussion* for description of additional features. bp, Base pairs.

thetic oligonucleotide as a primer (18) and pSTE7.2 fragment *a–d* (Fig. 1) subcloned into M13mp18 as the template. Samples were analyzed by electrophoresis in buffer-gradient gels as described for sequencing (12).

```
1                                                  50                                                  100
AAGCTTGTGGTCTTAAAAGAATGTGGATGAAACAAAACTGAATCCAATGATGAGAAATAAATGTACAATCAGATAATAAACACTTTTATTGATGATTGTCTTTTCTTGCAAACCCCAAT
                  150                                                 200
GTTACAATCGCCCATTAGTTCTAAGATTGTGTTGTCCTTTTTCTTACGAAATTTTTTTCCCTCCCTTTTTTGATCACTGTCCTCGTTATAAGTGATTCGTGGGGTAACGGCGATATCATA
        250                                                 300                                                 350
CTATTATTGGCCCATCTGATGGATTAGAAGACCTACAGGTTAATGTGAAGCGACACTAATGATAAGAAAACTGCATCGGATCATATCTGTTTTTGCAGCGTGGTATATTGGTTGTTGGTC
                  400                                                 450
Met Phe Gln Arg Lys Thr Leu Gln Arg Arg Asn Leu Lys Gly Leu Asn Leu Asn Leu His Pro Asp Val Gly Asn Asn Gly Gln Leu Gln Glu Lys Thr Glu Thr His Gln Gly Gln Ser
ATGTTTCAACGAAAGACTTTACAGAGAAGGAACTTGAAAGGGCTCAATCTTAACCTGCACCCAGATGTGGGCAATAATGGCCAATTGCAGGAAAAGACAGAGACTCACCAGGGACAATCT
        500                                                 550
Arg Ile Glu Gly His Val Met Ser Asn Ile Asn Ala Ile Gln Asn Asn Ser Asn Leu Phe Leu Arg Arg Gly Ile Lys Lys Lys Leu Thr Leu Asp Ala Phe Gly Asp Asp Gln Ala Ile
CGAATAGAAGGCCACGTGATGTCTAACATTAATGCAATACAGAATAATAGCAACCTGTTTTTGCGAAGAGGCATAAAAAAAAAAACTGACGTTGGATGCGTTTGGTGATGACCAAGCTATA
                  650                                                 700
Ser Lys Pro Asn Thr Val Val Ile Gln Gln Pro Gln Asn Glu Pro Val Leu Val Leu Ser Ser Leu Ser Gln Ser Pro Cys Val Ser Ser Ser Ser Ser Leu Ser Thr Pro Cys Ile Ile
TCGAAACCAAACACTGTGGTAATACAGCAACCGCAAAATGAACCTGTTTTAGTTCTGTCTTCTCTATCACAATCCCCGTGTGTATCATCATCATCATCTTTGTCCACGCCATGCATTATA
                  750                                                 800
Asp Ala Tyr Ser Asn Asn Phe Gly Leu Ser Pro Ser Ser Thr Asn Ser Thr Pro Ser Thr Ile Gln Gly Leu Ser Asn Ile Ala Thr Pro Val Glu Asn Glu His Ser Ile Ser Leu Pro
GATGCGTACAGTAATAATTTCGGATTATCGCCATCATCCACGAATTCTACTCCCTCTACGATTCAGGGATTGTCCAATATTGCAACACCAGTTGAAAACGAACATTCGATATCACTACCA
        850                                                 900                                                 950
Pro Leu Glu Glu Ser Leu Ser Pro Ala Ala Ala Asp Leu Lys Asp Thr Leu Ser Gly Thr Ser Asn Gly Asn Tyr Ile Gln Leu Gln Asp Leu Val Gln Leu Gly Lys Ile Gly Ala Gly
CCTTTGGAGGAAAGCCTATCGCCAGCCGCAGCAGATCTGAAAGATACGTTGTCGGGAACTTCAAATGGTAATTATATACAACTCCAGGACTTGGTTCAGTTGGGGAAAATTGGTGCTGGA
                  1000                                                1050
Asn Ser Gly Thr Val Val Lys Ala Leu His Val Pro Asp Ser Lys Ile Val Ala Lys Lys Thr Ile Pro Val Glu Gln Asn Asn Ser Thr Ile Ile Asn Gln Leu Val Arg Glu Leu Ser
AATTCTGGAACTGTGGTGAAGGCACTACATGTTCCTGATTCCAAAATAGTTGCCAAAAAAACCATTCCTGTGGAACAGAATAACAGTACAATCATCAACCAATTAGTTAGGGAATTATCT
        1100                                                1150
Ile Val Lys Asn Val Lys Pro His Glu Asn Ile Ile Thr Phe Tyr Gly Ala Tyr Tyr Asn Gln His Ile Asn Asn Glu Ile Ile Ile Leu Met Glu Tyr Ser Asp Cys Gly Ser Leu Asp
ATCGTCAAAAACGTTAAGCCCCATGAAAACATTATCACCTTCTATGGAGCTTATTATAACCAGCATATAAATAATGAAATCATAATTTTAATGGAATACTCTGATTGTGGTTCTTTAGAT
                  1250                                                1300
Lys Ile Leu Ser Val Tyr Lys Arg Phe Val Gln Arg Gly Thr Val Ser Ser Lys Lys Thr Trp Phe Asn Glu Leu Thr Ile Ser Lys Ile Ala Tyr Gly Val Leu Asn Gly Leu Asp His
AAAATACTGTCCGTTTATAAAAGGTTTGTTCAAAGAGGGACTGTTTCGAGTAAGAAAACCTGGTTCAACGAATTAACAATATCAAAAATAGCGTATGGCGTACTAAATGGCTTGGATCAT
        1350                                                1400
Leu Tyr Arg Gln Tyr Lys Ile Ile His Arg Asp Ile Lys Pro Ser Asn Val Leu Ile Asn Ser Lys Gly Gln Ile Lys Leu Cys Asp Phe Gly Val Ser Lys Lys Leu Ile Asn Ser Ile
TTGTACCGACAATATAAGATCATTCATCGTGATATCAAGCCTTCCAATGTTCTGATTAATAGTAAGGGGCAGATTAAGTTATGTGATTTTGGAGTTTCCAAAAAACTAATAAATTCTATC
        1450                                                1500                                                1550
Ala Asp Thr Phe Val Gly Thr Ser Thr Tyr Met Ser Pro Glu Arg Ile Gln Gly Asn Val Tyr Ser Ile Lys Gly Asp Val Trp Ser Leu Gly Leu Met Ile Ile Glu Leu Val Thr Gly
GCTGATACATTTGTTGGAACGTCCACTTATATGTCACCAGAGAGGATACAAGGAAACGTTTATTCTATCAAAGGGGACGTTTGGTCATTGGGCTTAATGATCATCGAGCTGGTAACTGGA
                  1600                                                1650
Glu Phe Pro Leu Gly Gly His Asn Asp Thr Pro Asp Gly Ile Leu Asp Leu Leu Gln Arg Ile Val Asn Glu Pro Ser Pro Arg Leu Pro Lys Asp Arg Ile Tyr Ser Lys Glu Met Thr
GAGTTTCCCCTAGGTGGGCATAACGATACACCTGATGGCATATTGGATTTGCTGCAACGTATTGTCAACGAGCCTTCACCAAGATTACCCAAAGACCGTATCTATTCCAAGGAAATGACA
        1700                                                1750
Asp Phe Val Asn Arg Cys Cys Ile Lys Asn Glu Arg Glu Arg Ser Ser Ile His Glu Leu Leu His His Asp Leu Ile Met Lys Tyr Val Ser Pro Ser Lys Asp Asp Lys Phe Arg His
GATTTTGTCAATAGGTGTTGTATTAAGAATGAAAGGGAAAGGTCATCGATTCATGAATTGCTACATCATGATCTTATAATGAAATACGTATCACCGTCTAAAGATGATAAATTTAGACAT
                  1850                                                1900
Trp Cys Arg Lys Ile Lys Ser Lys Ile Lys Glu Asp Lys Arg Ile Lys Arg Glu Ala Leu Asp Arg Ala Lys Leu Glu Lys Lys Gln Ser Glu Arg Ser Thr His
TGGTGTAGAAAAATAAAATCTAAAATAAAGGAAGACAAGAGAATTAAAAGAGAAGCCTTGGACCGTGCCAAGTTAGAAAAGAAACAATCGGAAAGATCAACCCATTGAATGCATGCGATA
        1950                                                2000
ACATTAGTTAGGAACATATTGCACTTTTTTTCACATCCCTTTCTCAAAGCTATATTCANATATATATAGACTCTCCTGATTATGTAAATCACAAACCAATTTTCTAATGTAACAATATTT
        2050                                                2100
TTAAAAATAATCATCAAAGTAGCACTTTATGTACAGAAAATTTCACGATTGAGAGAAGCTT
```

FIG. 2. Nucleotide sequence of a 2.1-kb *Hin*dIII fragment containing the *STE7* coding region. The translated sequence is shown above the DNA sequence beginning at the first ATG of the open reading frame. Landmark restriction endonuclease recognition sequences (underlined) are *Cla* I (ATCGAT) at position 1724; *Eco*RV (GATATC) at positions 231, 827, and 1350; and *Hin*dIII (AAGCTT) at positions 1 and 2095. (See Fig. 1.)

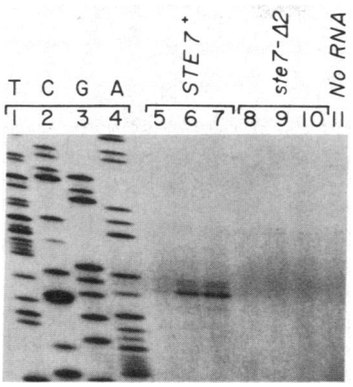

FIG. 3. Autoradiogram of *STE7* RNA primer-extension products analyzed on DNA sequencing gels. Lanes 1–4: DNA sequence ladder of the transcription start region from *STE7*. Lanes 5–7: products of primer-extension reactions using 1, 10, and 25 μg of poly(A)$^+$ RNA from *STE7*$^+$ strain DC38. Lanes 8–10: products of control reactions using the corresponding amounts of poly(A)$^+$ RNA from *ste7-Δ2* strain DC127. Lane 11: mock reaction with no RNA. The *ste7-Δ2* allele is a deletion of *Eco*RV fragment *b–c* (Fig. 1), which encompasses the primer binding site.

RESULTS

***STE7* DNA Sequence.** Previous complementation analysis of temperature-sensitive *ste7* mutants with clones from a yeast genomic DNA library showed that the functional portion of *STE7* is contained on the 2.3-kb *Hin*dIII fragment of pSTE7.2 (Fig. 1*a* and ref. 7). This fragment, which includes 1.9 kb of yeast DNA and 0.4 kb of pBR322, was isolated and sequenced. The DNA sequence revealed 1897 bp of yeast DNA containing an open reading frame of 537 codons with no termination codon. Subsequent cloning of a 376-bp *Cla* I–*Hin*dIII fragment (*e–f*) from pSTE7.4 (Fig. 1*a*), which overlaps the 3′ end of the determined sequence by 174 bp, allowed the completion of the open reading frame and determination of 3′ flanking sequence. The 2.1-kb sequence shown in Fig. 2 has been determined on both strands except for nucleotides 2083–2100.

The 2100-bp DNA sequence was translated in all six frames. The longest open reading frame extends from position 288 to 1907, where it ends with a TGA translation termination codon (Fig. 1*b*). The amino acid sequence of the predicted *STE7* gene product is shown in Fig. 2. The sequence begins with a methionine at the first ATG in the open reading frame.

***STE7* RNA.** The polarity of the *STE7* transcript as predicted from the DNA sequence was confirmed by blot hybridization analysis of electrophoretically fractionated RNA, using strand-specific M13 hybridization probes (results not shown). The 5′ end of the *STE7* mRNA was mapped to its cognate DNA by the primer-extension method (17). The results are shown in Fig. 3. One major extension product was produced in proportion to RNA concentration in the reaction with *STE7*$^+$ RNA from yeast strain DC38. The 5′ map position of the major band is located at nucleotide 314. Minor bands were visible at positions 311, 312, and 318 or 319. No bands were observed for extension reactions with control RNA from the isogenic *ste7-Δ2* deletion strain DC127 or with no RNA.

Predicted STE7 Protein. The FASTP algorithm of Lipman and Pearson was used to compare the predicted STE7 protein sequence to the Protein Identification Resource database (14). The 24 highest-scoring matches were all kinase or kinase-related proteins. Representatives from each kinase subfamily among the first 24 matches and the highest-scoring unrelated protein identified by this search are shown in Fig. 4*a*. An estimate of the statistical significance of these matches was made by comparing the predicted STE7 protein sequence to randomly permuted versions of sequences for the potentially related proteins. The significance-of-similarity values (*z* values) determined by the RDF program are given in parentheses in Fig. 4*a* (see ref. 14). The *z* values we find for the higher scoring matches are statistically significant according to the guidelines of Lipman and Pearson. In addition, no matches outside of the kinase family were found that met these criteria. We conclude from this comparison that the predicted *STE7* gene product is related to the protein kinase family of polypeptides.

DISCUSSION

A summary of *STE7* sequence features is presented in Fig. 1*b*. A sequence resembling the consensus hexamer of the transcription control sequence TATAAA is located at position 241 (TATTAT), 74 bp upstream from the mapped initiation site. Although the "TATA box" for eukaryotic genes is usually found 20–30 bp from the transcription initiation site, recent studies in yeast indicate that the TATA box can function 90 or more bp upstream from the site of transcription initiation (19, 20). The tripartite transcription termination control consensus sequence (TAG/TATGT/TTT) identified by Zaret and Sherman (21) is present in the *STE7* sequence between nucleotides 1986 and 2021. Transcription termination has been found to occur 10–21 nucleotides from this termination sequence in four cases where 3′ map positions have been determined (21). Based on the 5′ map position and the inferred 3′ end, the predicted *STE7* transcription unit is close to the observed 1.8-kb size for the *STE7* RNA (7). The longest open reading frame in the 2.1-kb sequence extends from position 288 to 1907, where it ends with a UGA translation termination codon. The transcription initiation site resides within this region and is 46 bases upstream from the first methionine codon at position 360. Because eukaryotic translation usually initiates at the first AUG (22), the predicted *STE7* gene product is a 515 amino acid polypeptide. Another possible translation initiation codon is found at position 499. Neither ATG is flanked by nucleotides that strictly conform to the consensus for preferred translation initiation sites. However, both are preceded by a purine at position −3, a feature that is believed important for initiation (22).

The predicted primary sequence of the *STE7* product was compared to a large data base of known and predicted protein sequences by the FASTP algorithm (14). The results of the data base search indicate that the predicted *STE7* product is related to the protein kinase family of polypeptides. The most compelling evidence for kinase relatedness of the predicted STE7 protein arises from an examination of the homologies that are illustrated in Fig. 4*b*. All protein kinases are closely related to one another over a region of about 260 amino acids that is believed to constitute the catalytic domain. Of particular note are four sequences in the predicted STE7 protein that are characteristic of the kinase family (23, 24). One is the sequence Gly-Xaa-Gly-Xaa-Xaa-Gly, which is thought to be involved in ribose ring interactions with ATP (25). The second conserved sequence, Ala-Xaa-Lys, is found another 15–20 amino acids toward the COOH terminus in all protein kinases. Because the conserved lysine residue in pp60$^{v\text{-}src}$ and in cAMP-dependent protein kinase is modified by the ATP analogue 5′-*p*-fluorosulfonylbenzoyladenosine, it has been suggested that this lysine may be involved in a phosphate group interaction at the catalytic center (25, 26). Approximately 130 amino acids further toward the COOH terminus, the third conserved sequence, Asp-Phe-Gly, is also found in STE7. The fourth conserved and catalytically important region in protein kinases consists of an Ala-Pro-Glu sequence (27). At the homologous position in STE7 we

a

Identifier	Protein	Score I	Score O
OKBO2C	Bovine cyclic AMP-dependent kinase (CAPK)	97 (10.30)	178 (12.20)
TVBY8	Yeast cell division control protein 28 (CDC28)	93 (10.98)	141 (11.02)
TVRTM	Rat kinase-related transforming protein (mos)	85 (10.56)	171 (12.53)
TVFFS	Kinase-related transforming protein (src)	82 (9.38)	153 (20.52)
GNFVG9	Avian sarcoma virus gag-yes polyprotein p90	78 (6.34)	130 (6.33)
TVYUH	Avian erythroblastosis virus kinase-related transforming protein (erbB)	75 (7.86)	152 (11.91)
GNFVFP	Avian sarcoma virus gag-fps polyprotein	74 (8.46)	126 (6.90)
GNMVCS	Feline sarcoma virus gag-fes polyprotein	73 (7.56)	114 (9.54)
GNVWGM	Abelson murine leukemia virus gag-abl polyprotein	72 (6.61)	134 (9.23)
GNMVRR	Feline sarcoma virus gag-fgr polyprotein	71 (7.73)	136 (8.48)
GNUWMM	Avian retrovirus gag-mht (gag-mil) polyprotein	70 (6.74)	97 (6.25)
GNMVF6	Murine sarcoma virus gag-raf polyprotein	69 (5.80)	171 (11.14)
FESC	Scenedesmus quadricauda ferredoxin	58 (6.76)	58 (4.94)

b

```
YIQLQDLVQLGKIGAGNSGTVVKALHV...PDSKIVAKKTIPVEQNNSTIINQ.LVRELSIVKNVKPHENIITFYGAYYN  STE7

TAHLDQFERIKTLGTGSFGRVMLVKHM...ETGNHYAMKILDKQKVVKLKQIEHTLNEKRILQAVNFPF....LVKLEFS  CAPK
SGELANYKRLEKVGEGTYGVVYKALDLRPGQGQRVVALKKIRLESEDEGVPSTAIR.EISLLKELKDD...NIVRLYDIV  CDC28
FIDWGQVCLLHRLGSGGFGSVYKATY.....HGVPVAIKQVNKCTRTLRASQRNFWAELNIARLHHDNIIRVVAASTRTP  MOS
KMEASEVMLSTRIGSGSFGTVYKGKWH......GDVAVKILKVVDPTPEQLQAFRNEVAVLRKTR.............  RAF

QHINNEIIILMEYSDCGSLDKILSVYKRFVQRGTVSSKKTWFNELTISKIAYGVLNGLDHLYRQYKIIHRDIKPSNVLIN  STE7

FKDNSNLYMVMEYVPGGEMFSHLRRIGRFSEPHARFYAAQIVLTFEY.............LHSLDLIYRDLKPENLLID  CAPK
HSDAHKLYLVFEFLDLDLKRYMEGIPKDQPLGADIVKKFMMQLCKGI............AYCHSHRILHRDLKPQNLLIN  CDC28
EGSNSLGTIIMEFGGNVTLHQVIYGATRSPEPLSCREQLSLGKCLKY....SLDIVNGLLFLHSQSILHLDLKPANILIS  MOS
...HVNILLFMGYMTKDNLAIVTQWCEGSSLYKHLHVQETKFQMFQLIDIARQTAQGMDYLHA.KNIIHRDMKSNNIFLH  RAF

SKGQIKLCDFGVSKKLINSIADTFV....GTSTYMSPERIQ...GNVYSIKGDVWSLGLMIIELVTGEFPLGGHNDTPDG  STE7

QQGYIQVTDFGFAKRVKGRTWTL.....CGTPEYLAPEIIL...SKGYNKAVDWWALGVLIYEMAAGYPPFFADQPIQIY  CAPK
KDGNLKLGDFGLARAFGVPLRAY..THEIVTLWYRAPEVL..LGGKQYSTGVDTWSIGCIFAEMCNRKPIFSGDSEIDQI  CDC28
EKDVCKISDFGCSQKLQDLRCRPSLHHIGGTYTHQAPELLK...GEIATPKADIYSFGITLWQMTTREVPYSGEPQYVQY  MOS
EGLTVKIGDFGLATVKSRWSGSQQVEQPTGSVLWMAPEVIRMQDDNPFSFQSDVYSYGIVLYELMAGELPYAHINNRDQI  RAF
```

FIG. 4. (*a*) Protein sequences similar to the hypothetical STE7 protein. Identifiers and protein descriptive names are those used by the National Biomedical Research Foundation, Washington, DC. Initial scores (I) and optimized scores (O) were determined by the FASTP algorithm, using ktup = 1 (14). Values given in parentheses are *z* values determined by the RDF algorithm (14). The RDF routine was run at ktup = 1, with 50 random shuffles for each comparison. (*b*) Sequence alignment of four high-scoring proteins with the putative *STE7* gene product reveals a high degree of homology within conserved domains common to the kinase family. Each sequence is compared directly to STE7, with identities indicated by + and similarities by −. Dots indicate gaps inserted to maximize homology. Regions of conservation that are discussed in the text are indicated by brackets. Standard one-letter amino acid abbreviations are used.

find the closely related Ser-Pro-Glu sequence. These and other identities and similarities illustrated by the comparison in Fig. 4*b* suggest that the putative *STE7* gene product conforms to the conserved catalytic domain of protein kinases.

The STE7 protein has not yet been shown to be an active protein kinase. However, the prospects for this demonstration are promising. The relationship of the predicted STE7 protein sequence to members of the kinase family is similar to that found for the yeast *CDC28* gene product, for which protein kinase activity has been demonstrated (28, 29).

Protein kinases, which include several known oncogenes, are intimately involved in the control of normal growth in eukaryotic cells. In this role, they appear to be components of regulatory cascades and to function as mediators for a variety of effector molecules. In some cases it appears these proteins may modulate the activity of the cell's transcriptional machinery (30). The implication that follows is that the *STE7* gene product may similarly participate in a cascade of regulatory events resulting in cell-type specialization in yeast. The genetic and molecular methods available with yeast make this an excellent system in which to test for and characterize components of the postulated network.

We thank Drs. D. Chelsky, C. Falco, and C. Hutchison III for critical reading of the manuscript. This research was supported by U.S. Public Health Service Research Grant GM30619.

1. Herskowitz, I. & Oshima, Y. (1981) in *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, eds.

Strathern, J. N., Jones, E. W. & Broach, J. R. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 181–209.
2. Hartwell, L. H. (1980) *J. Cell Biol.* **85,** 811–822.
3. Dubois, E., Jacobs, E. & Jauniaux, J. C. (1982) *EMBO J.* **1,** 1133–1139.
4. Errede, B., Cardillo, T. S., Sherman, F., Dubois, E., Deschamps, J. & Wiame, J. M. (1980) *Cell* **22,** 427–436.
5. Taguchi, A. K. W., Ciriacy, M. & Young, E. T. (1983) *Mol. Cell. Biol.* **4,** 61–68.
6. Elder, R. T., St. John, T. P., Stinchcomb, D. T. & Davis, R. W. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45,** 581–584.
7. Chaleff, D. T. & Tatchell, K. (1985) *Mol. Cell. Biol.* **5,** 1878–1886.
8. Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112,** 295–298.
9. Norgard, M. V. (1981) *Anal. Biochem.* **113,** 34–42.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
11. Bankier, A. T. & Barrell, B. G. (1983) in *Techniques in Nucleic Acid Biochemistry*, ed. Flavell, R. A. (Elsevier, Ireland), pp. 1–34.
12. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.
13. Staden, R. (1982) *Nucleic Acids Res.* **10,** 4731–4751.
14. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
15. Broach, J. R., Atkins, J. F., McGill, C. & Chow, L. (1979) *Cell* **16,** 827–839.
16. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
17. Reeder, R. H., Roan, J. G. & Dunaway, M. (1983) *Cell* **35,** 449–456.
18. McGraw, R. A., III (1984) *Anal. Biochem.* **143,** 298–303.
19. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 369–375.
20. Chen, W. & Struhl, K. (1985) *EMBO J.* **4,** 3273–3280.
21. Zaret, K. S. & Sherman, F. (1982) *Cell* **28,** 563–573.
22. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
23. Hunter, T. & Cooper, J. A. (1985) *Annu. Rev. Biochem.* **54,** 897–930.
24. Barker, W. C. & Dayhoff, M. O. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2836–2839.
25. Kamps, M. P., Taylor, S. S. & Sefton, B. M. (1984) *Nature (London)* **310,** 589–591.
26. Zoller, M. J., Nelson, N. C. & Taylor, S. S. (1981) *J. Biol. Chem.* **256,** 10837–10842.
27. Bryant, D. L. & Parsons, J. T. (1984) *Mol. Cell. Biol.* **4,** 862–866.
28. Lörincz, A. T. & Reed, S. I. (1984) *Nature (London)* **307,** 183–185.
29. Reed, S. I., Hadwiger, J. A. & Lörincz, A. T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4055–4059.
30. Weinberg, R. A. (1985) *Science* **230,** 770–776.