Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 6038–6042, July 1991
Cell Biology

Molecular cloning and characterization of the yeast gene for squalene synthetase

(isoprenoid pathway/ergosterol synthesis/cholesterol synthesis/*Saccharomyces cerevisiae*)

SUSAN M. JENNINGS, YIM H. TSAY, TOBE M. FISCH, AND GORDON W. ROBINSON

Department of Cellular Biology, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ 08543-4000

Communicated by Joseph L. Goldstein, April 15, 1991

ABSTRACT **Squalene synthetase (farnesyl-diphosphate: farnesyl-diphosphate farnesyltransferase, EC 2.5.1.21) is a critical branch point enzyme of isoprenoid biosynthesis that is thought to regulate the flux of isoprene intermediates through the sterol pathway. The structural gene for this enzyme was cloned from the yeast *Saccharomyces cerevisiae* by functional complementation of a squalene synthetase-deficient *erg9* mutant. Identification of this *ERG9* clone was confirmed by genetic linkage analysis in yeast and expression of enzyme activity in *Escherichia coli*. The predicted squalene synthetase polypeptide of 444 amino acids (M_r, 51,753) lacks significant homology to known protein sequences, except within a region that may represent a prenyl diphosphate (substrate) binding site. The *ERG9*-encoded protein contains a PEST consensus motif (rich in proline, glutamic acid, serine, and threonine) present in many proteins with short cellular half-lives. Modeling of the protein suggests that it contains at least one, and possibly two, membrane-spanning domains. Disruption of the chromosomal squalene synthetase coding region by insertional mutagenesis indicates that *ERG9* is a single copy gene that is essential for cell growth in yeast.**

An understanding of the regulation of cholesterol synthesis is essential for assessing the role of this molecule in normal biological functions and disease states, such as atherosclerosis. Cholesterol and related sterols are produced in the polyisoprene pathway of eukaryotes from farnesyl diphosphate (FDP), an intermediate also used for the synthesis of dolichols, ubiquinones, hormones, heme A, and some isoprenylated proteins. Production of FDP from mevalonate is primarily regulated by changes in 3-hydroxy-3-methylglutaryl coenzyme A reductase enzyme levels. Other regulatory mechanisms ensure that FDP utilization by competing enzymes is proportionate to the needs of the cell for the various FDP-derived products (1). Studies with mammalian cells have shown that control of squalene synthetase (farnesyl-diphosphate:farnesyl-diphosphate farnesyltransferase, EC 2.5.1.21), the first committed enzyme of cholesterol biosynthesis, is crucial for balancing the incorporation of FDP into sterol and nonsterol isoprene synthesis (2).

Squalene synthetase is a microsomal enzyme, which, in the presence of Mg^{2+} and a reduced pyridine nucleotide, forms squalene by a two-step reductive condensation of two molecules of FDP. It has been isolated from many sources and its reaction mechanism has been investigated in some detail (3, 4). Squalene synthetase solubilized from membranes of bakers' yeast, *Saccharomyces cerevisiae*, has been described (4–8) and the purified enzyme was reported to have a molecular mass of ≈50 kDa (5–8). Yeast with defects in squalene synthetase were identified by screening a collection of mutants blocked in ergosterol biosynthesis (9). Mutants designated *erg9* were unable to convert radiolabeled mevalonate into squalene but did accumulate farnesol, a breakdown product of FDP. Because these squalene synthetase mutants were also obligate ergosterol auxotrophs, they were used in a phenotype complementation approach to clone the *ERG9* gene. In this report, we describe the structure of squalene synthetase deduced from its primary sequence and explore the consequences of genomic deletion of this gene in *S. cerevisiae*.*

MATERIALS AND METHODS

Strains. The following strains of *S. cerevisiae* were used in this study: DC67 (*MATa ERG9 leu2-4 ade1 lys2 cir°*) was from J. Hicks; W303-1A (*MATa*) and W303-1B (*MATα*) (both *ERG9 leu2-3,112 ade2-1 ura3-1 trp1-1 his3-11,15*) were from R. Rothstein; SGY336 (*MATa erg9-1*) was from F. Karst; JRY527 (*MATa ERG9 ura3-52 ade2-101 his3-d200 lys2-801 Met⁻*) was from J. Rine; SC14089 (*MATa ERG9*) was from J. Tkacz; SGY1161 an *ERG9/ERG9* diploid was from a mating of W303-1A and W303-1B; SGY969 an *ERG9/erg9* diploid was constructed from SGY336 by mating to W303-1B; SGY1011 (*MATα erg9 leu2-3,112 ade2-1 ura3-1 his3-11,15*) was a segregant from SGY969. Standard procedures were used for mating, sporulation, and tetrad analysis.

Escherichia coli strains used were DH5α [F⁻ Δ80d*lacZ* M15, *endA1, recA1, hsdR17, supE44, thi-1, gyrA, relA1,* Δ(*lacZYA-argF*)U169] and DH5αMCR (same as DH5α plus *mcrA, mcrB, mrr*) from Bethesda Research Laboratories.

Plasmids and Library. YEp351 (10) and YCp50 (11) are high and low copy number autonomously replicating yeast–*E. coli* shuttle vectors that confer ampicillin resistance to *E. coli* and either leucine or uracil prototrophy, respectively, to *leu2* or *ura3* yeast. pTTQ8 (12) is a bacterial *tac* promoter plasmid. pGR61 was made by subcloning the 2.4-kilobase (kb) *ERG9* insert of pSM54 into the yeast integrating vector Ylp5 (11). Plasmid-borne disruptions of the *ERG9* gene were constructed by replacement of a 1.0-kb *Eco*RI/*Sca* I fragment of pSM52 with either a 2.5-kb *Eco*RI/*Eco*RV fragment of the yeast *HIS3* gene from YEp6 (11), to create pET105, or a 2.7-kb *Eco*RI/*Sca* I fragment of the yeast *LEU2* gene from YEp13 (11), to create pET106 (see Fig. 1). The yeast genomic library used in this study was prepared by ligation of DC67 DNA partially digested with *Sau*3A into the *Bam*HI site of YEp351; it contained 14,000 clones.

Media, Growth, and Transformation Conditions. Yeast strains were grown at 28°C in either YPD (13) or YPD-E (YPD supplemented with 20 μg of ergosterol per ml) medium to obtain genomic DNA or to prepare cells for transformation. Strains harboring vectors that contained the *URA3* or *LEU2* gene were grown in YM or YM-E (ergosterol-supplemented YM) synthetic medium lacking either uracil or leucine, re-

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.

Abbreviation: FDP, farnesyl diphosphate.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M63979).

spectively (13). Yeast were grown in YPD or YM aerobically and in YPD-E or YM-E anaerobically. Transformation of yeast strains was by the spheroplast method as described (14) except for *erg9* strains, which were transformed by electroporation (15) with a Gene Pulser (Bio-Rad) using a modified procedure. Three sequential pulses of field strength 1.0 kV/cm were delivered to 10^9 yeast cells in the presence of 30 μg of library DNA in 0.5 ml of 35% PEG 4000. Leu^+ transformants were selected by embedding washed cells in YM-E regeneration agar (13) and incubating in an anaerobic chamber at 28°C for 5 days. *E. coli* cells were grown in LB broth or M9 medium (16).

Cloning of the *ERG9* Gene. A yeast genomic library constructed in the vector YEp351 was used to transform the *erg9 leu2* strain SGY1011 to leucine prototrophy. Transformants obtained after anaerobic selection were isolated from the regeneration top agar by manual disruption. Transformants were pooled, replated onto YM agar plates, and grown aerobically to screen for transformants that were ergosterol prototrophs. DNA from prototrophic colonies was used to transform *E. coli* to ampicillin resistance.

DNA Sequencing. The DNA sequence of the *ERG9* gene was determined by the dideoxynucleotide chain-termination method (17). Sequencing reactions were performed on restriction fragment subclones of a 2.4-kb *Xba* I/*Bam*HI region of pSM54 in M13, following the Sequenase (United States Biochemical) protocol. The sequences of both coding and noncoding strands were confirmed by using synthetic oligonucleotide primers derived from the initial sequence. Sequence data were assembled and analyzed with Intelligenetics and Microgenie software. NBRF/GenBank (release 63.0), PIR (release 23.0), and Swiss-Prot (release 13.0) data bases were searched for homologies to the DNA sequence and protein sequence of squalene synthetase.

DNA–DNA Hybridization Analysis. Yeast genomic DNA was prepared from wild-type or *ERG9* disruption strains and examined by DNA–DNA hybridization as described (13). Genomic DNA digested with *Xba* I was fractionated on a 0.8% agarose gel and transferred to BioTrans membrane (ICN). The blot was probed with a radiolabeled 1.1-kb *Sca* I/*Xba* I *ERG9* fragment prepared by random-primer labeling (18) and visualized by autoradiography.

Expression of the *ERG9* Gene in *E. coli* Cells. A 1.5-kb *Bgl* II PCR fragment that contained the squalene synthetase coding region and 12 base pairs (bp) of 5′ untranslated sequence was ligated into the *Bam*HI site of pTTQ8 (12) to form the expression vector pET124. The pET124 fusion gene should encode an *ERG9* protein with nine extra N-terminal amino acids. Cultures of *E. coli* DH5α transformed with either pTTQ8 or pET124 were grown at 37°C in M9 medium to an OD_{600} of 0.5 and then switched to growth in LB broth plus 0.5 mM isopropyl β-D-thiogalactopyranoside for induction. After induction, cells were suspended in breakage buffer (0.1 M NaH_2PO_4, pH 7.4/4 mM $MgCl_2$/1 mM EDTA/10 mM 2-mercaptoethanol/1 mM phenylmethylsulfonyl fluoride/2 mM benzamidine/0.1 mM EGTA/1 μg of leupeptin per ml), and lysed by three cycles of freeze–thaw. Squalene synthetase activity was measured in supernatant fractions (4000 × *g*) by a radiochemical assay (6).

Yeast Microsomal Extracts and Squalene Synthetase Activity. Yeast cells harvested and then washed in distilled water by centrifugation at 8000 × *g* were disrupted by Vortex mixing vigorously with 0.45-mm glass beads in the presence of breakage buffer at 4°C. Cell debris was removed from the lysates at 4°C by sequential 10-min centrifugations at 4000 and at 10,000 rpm. The 10,000-rpm supernatant fractions were centrifuged at 4°C for 1 hr at 100,000 × *g* to produce a pellet of "microsomal membranes." Pellets were resuspended on ice in 0.2 ml of breakage buffer using a glass homogenizer and were stored frozen at −70°C. Squalene synthetase activity in microsomes was assayed at 30°C under N_2 using a gas chromatographic assay (19).

RESULTS

Cloning, Complementation Analysis, and Subcloning. DNA fragments containing the *ERG9* gene were isolated from a yeast genomic library by phenotypic complementation of a yeast *erg9* mutant. A pool of 2 × 10^5 Leu^+ transformants of the mutant strain SGY1011 was screened to identify recombinant plasmids capable of converting this strain to ergosterol prototrophy. From the prototrophs, plasmids designated pSM51, -52, and -53, which contained yeast DNA inserts ranging from 2.4 to 5.4 kb, were obtained. Restriction mapping demonstrated that the inserts shared a 2-kb region (Fig. 1). Squalene synthetase assays showed that *erg9* cells bearing either pSM51 or pSM52 possessed approximately wild-type levels of enzyme, whereas cells bearing pSM53 contained only 5% of wild-type levels (Table 1). These findings suggested that pSM51 and pSM52 carried an intact *ERG9* gene but that pSM53 contained a truncated copy. Subclones of pSM51 constructed by insertion of a 2.4-kb *Xba* I fragment in alternate orientations in the centromere-containing vector, YCp50 (11), were tested in the *erg9* mutant. Both of the resulting low copy number plasmids, pSM54 and pSM55 (Fig. 1), converted the strain to ergosterol prototrophy and expressed wild-type levels of squalene synthetase.

Genetic Confirmation of *ERG9*. To verify the identity of *ERG9* clones, the ability of insert DNA to target integration to the *ERG9* locus in the yeast genome was determined. After integration of the *ERG9*-containing plasmid pGR61 (*URA3*) into strain JRY527 (*ERG9 ura3-52*) and mating of Ura^+ integrants to strain SGY1011 (*erg9 ura3-1*), diploids were sporulated and tetrads were dissected. All 23 tetrads examined displayed $2^+/2^-$ segregation for *ERG9* and for *URA3* and cosegregation of Erg^+ and Ura^+ phenotypes, indicating that plasmid integration had occurred at the *ERG9* locus.

***ERG9* DNA Sequence.** An open reading frame of 1065 bp detected on pSM53 was extended to 1332 bp by sequencing of 5′ contiguous DNA from pSM51. Starting from the first methionine codon, the pSM51 ORF predicted a protein of 444 amino acids with a M_r of 51,753 (Fig. 2). This figure was in good agreement with the size estimated for yeast squalene synthetase by denaturing gel electrophoresis (5). The *ERG9*

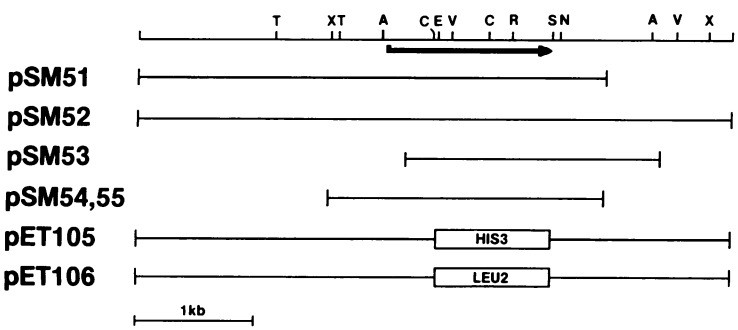

FIG. 1. Restriction map of the *ERG9* chromosomal locus and plasmid subclones. Cloned DNA fragments of 4.5, 5.4, and 2.4 kb isolated on the YEp351 plasmids pSM51, pSM52, and pSM53, respectively, are shown. Plasmids pSM54 and pSM55 are subclones of pSM51 in YCp50 (11). The location of the *ERG9* open reading frame is denoted by an arrow. An internal region of *ERG9* replaced with either a *HIS3* fragment or a *LEU2* fragment is indicated by an open box for plasmids pET105 and pET106, respectively. The ability of plasmids to complement the *erg9* mutation is indicated by + and −. A, *Acc* I; C, *Cla* I; E, *Eco*RI; N, *Nsi* I; R, *Eco*RV; S, *Sca* I; T, *Sst* I; V, *Sna*BI; X, *Xba* I.

Table 1. Squalene synthetase activity in complemented *erg9* yeast

Strain	Genotype	Plasmid	Sp. act., nmol·min^{-1}·mg^{-1}	Rel. act.
JRY527	*ERG9*	None	0.45	1.0
SC14089	*ERG9*	None	0.48	1.0
SGY1011	*erg9*	None	ND	<0.01
SGY1011	*erg9*	pSM51 (*ERG9*)	1.27	2.8
SGY1011	*erg9*	pSM52 (*ERG9*)	0.59	1.3
SGY1011	*erg9*	pSM53 (*ERG9*)	0.022	0.05

Microsomal membranes prepared from disrupted yeast strains were assayed for squalene synthetase activity in duplicate assays. A specific activity (Sp. act.) of 0.46 found in wild-type (i.e., *ERG9*) strains was assigned a relative activity (Rel. act.) of 1.0. ND, activity not detectable within the limit of the assay (≈0.003 nmol of squalene per min per mg of protein).

gene of pSM53 was thus missing all 5′ upstream and the first 167 bp of its translated sequences. A 5′ *Xba* I site used for construction of pSM54 and pSM55 was located 370 bp upstream of the first *ERG9* ATG codon, indicating that efficient expression of this gene requires only these sequences. A consensus TATA transcriptional regulatory element (TATAAA) was present 100 nucleotides upstream of the start codon but no significant direct or inverted repeats, often associated with upstream activator sequence elements, were identified downstream of the *Xba* I site. A sequence similar to a canonical polyadenylylation signal (AATAAA) was identified 183 nucleotides downstream of the squalene synthetase stop codon (Fig. 2).

***ERG9* Transcript.** No TACTAAC box, associated with all intron-containing genes of *S. cerevisiae* (20), was present in the coding region, indicating that the *ERG9* transcript is unspliced. To determine the size of the *ERG9* transcript, a Northern blot analysis was performed on mRNA from a wild-type (*ERG9*) yeast strain and a single transcript of ≈1.8 kb was detected by using a radiolabeled *ERG9*-specific DNA probe (data not shown).

Expression of the *ERG9* Gene in *E. coli* Cells. To establish that *ERG9* encodes squalene synthetase, expression of the gene was evaluated in an organism (*E. coli*) that lacks this enzyme. Under growth conditions known to induce expression from the bacterial *tac* promoter (LB broth + isopropyl β-D-thiogalactopyranoside), *E. coli* cells containing the

```
TCTAGAGACCCTGCGAGCGTGTCCCGGTGGGTTCTGGGAGCTCTAACTCCGCAGGAACTACAAACCTTGCTTACACAGA   79
 XbaI
GTGAACCTGCTGCCTGGCGTGCTCTGACTCAGTACATTTCATAGCCCATCTTCAACAACAATACCGACTTACCATCCTA  158

TTTGCTTTGCCCTTTTTCTTTTCCACTGCACTTTGCATCGGAAGGCGTTATCGGTTTTGGGTTTAGTGCCTAAACGAGC  237

AGCGAGAACACGACCACGGGCTATATAAATGGAAAGTTAGGACAGGGGCAAAGAATAAGAGCACAGAAGAAGAGAAAAG  316

ACGAAGAGCAGAAGCGGAAAACGTATACACGTCACATATCACACACACACAATGGGAAAGCTATTACAATTGGCATTGC  395
                      AccI                    M  G  K  L  L  Q  L  A  L              9

ATCCGGTCGAGATGAAGGCAGCTTTGAAGCTGAAGTTTTGCAGAACACCGCTATTCTCCATCTATGATCAGTCCACGTC  474
H  P  V  E  M  K  A  A  L  K  L  K  F  C  R  T  P  L  F  S  I  Y  D  Q  S  T  S   36

TCCATATCTCTTGCACTGTTTCGAACTGTTGAACTTGACCTCCAGATCGTTTGCTGCTGTGATCAGAGAGCTGCATCCA  553
  P  Y  L  L  H  C  F  E  L  L [N] L  T  S  R  S  F  A  A  V  I  R  E  L  H  P   62

GAATTGAGAAACTGTGTTACTCTCTTTTATTTGATTTTAAGGGCTTTGGATACCATCGAAGACGATATGTCCATCGAAC  632
  E  L  R  N  C  V  T  L  F  Y  L  I  L  R  A  L  D  T  I  E  D  D  M  S  I  E   88

ACGATTTGAAAATTGACTTGTTGCGTCACTTCCACGAGAAATTGTTGTTAACTAAATGGAGTTTCGACGGAAATGCCCC  711
H  D  L  K  I  D  L  L  R  H  F  H  E  K  L  L  L  T  K  W  S  F  D  G  N  A  P  115
                                                  EcoRI
CGATGTGAAGGACAGAGCCGTTTTGACAGATTTCGAATCGATTCTTATTGAATTCCACAAATTGAAACCAGAATATCAA  790
  D  V  K  D  R  A  V  L  T  D  F  E  S  I  L  I  E  F  H  K  L  K  P  E  Y  Q  141

GAAGTCATCAAGGAGATCACCGAGAAAATGGGTAATGGTATGGCCGACTACATCTTGGATGAAAATTACAACTTGAATG  869
  E  V  I  K  E  I  T  E  K  M  G  N  G  M  A  D  Y  I  L  D  E  N  Y  N  L  N  167

GGTTGCAAACCGTCCACGACTACGACGTGTACTGTCACTACGTAGCTGGTTTGGTCGGTGATGGTTTGACCCGTTTGAT  948
G  L  Q  T  V  H  D  Y  D  V  Y  C  H  Y  V  A  G  L  V  G  D  G  L  T  R  L  I  194

TGTCATTGCCAAGTTTGCCAACGAATCTTTGTATTCTAATGAGCAATTGTATGAAAGCATGGGTCTTTTCCTACAAAAA 1027
  V  I  A  K  F  A [N] E  S  L  Y  S  N  E  Q  L  Y  E  S  M  G  L  F  L  Q  K  220

ACCAACATCATCAGAGACTACAATGAAGATTTGGTCGATGGTAGATCCTTCTGGCCCAAGGAAATCTGGTCACAATACG 1106
  T  N  I  I  R  D  Y  N  E  D  L  V  D  G  R  S  F  W  P  K  E  I  W  S  Q  Y  246

CTCCTCAGTTGAAGGACTTCATGAAACCTGAAAACGAACAACTGGGGTTGGACTGTATAAACCACCTCGTCTTAAACGC 1185
A  P  Q  L  K  D  F  M  K  P  E  N  E  Q  L  G  L  D  C  I  N  H  L  V  L  N  A  273

ATTGAGTCATGTTATCGATGTGTTGACTTATTTGGCCAGTATCCACGAGCAATCCACTTTCCAATTTTGTGCCATTCCC 1264
  L  S  H  V  I  D  V  L  T  Y  L  A  S  I  H  E  Q  S  T  F  Q  F  C  A  I  P  299

CAAGTTATGGCCATTGCAACCTTGGCTTTGGTATTCAACAACCGTGAAGTGCTACATGGCAATGTAAAGATTCGTAAGG 1343
  Q  V  M  A  I  A  T  L  A  L  V  F  N  N  R  E  V  L  H  G  N  V  K  I  R  K  325

GTACTACCTGCTATTTAATTTTGAAATCAAGGACTTTGCGTGGCTGTGTCGAGATTTTTGACTATTACTTACGTGATAT 1422
G  T  T  C  Y  L  I  L  K  S  R  T  L  R  G  C  V  E  I  F  D  Y  Y  L  R  D  I  352

CAAATCTAAATTGGCTGTGCAAGATCCAAATTTCTTAAAATTGAACATTCAAATCTCCAAGATCGAACAATTCATGGAA 1501
  K  S  K  L  A  V  Q  D  P  N  F  L  K  L  N  I  Q  I  S  K  I  E  Q  F  M  E  378

GAAATGTACCAGGATAAATTACCTCCTAACGTGAAGCCAAATGAAACTCCAATTTTCTTGAAAGTTAAAGAAAGATCCA 1580
  E  M  Y  Q  D  K  L  P  P  N  V  K  P  N  E  T  P  I  F  L  K  V  K  E  R  S  404

GATACGATGATGAATTGGTCCCAACCCAACAAGAAGAAGAGTACAAGTTCAATATGGTTTTATCTATCATCTTGTCCGT 1659
R  Y  D  D  E  L  V  P  T  Q  Q  E  E  E  Y  K  F  N  M  V  L  S  I  I  L  S  V  431

TCTTCTTGGGTTTTATTATATATACACTTTACACAGAGCGTGAAGTCTGCGCCAAATAACATAAACAAACAACTCCGAA 1738
  L  L  G  F  Y  Y  I  Y  T  L  H  R  A  *                                      444

CAATAACTAAGTACTTACATAATAGGTAGAGGCCTATCCTTAAAGATAACCTTATATTTCATTACATCAACTAATTCGA 1817
          ScaI
CCTTATTATCTTTCGAATTGAAATGCATTATACCCATCGGTACGTCTAGCTTTGTCACCTTCCCCAGTAAACGTTGTTT 1896
                 NsiI
CTTGCCGACAAACAATGTGGCCCTCTCTCCGTCAATCTGTAACGACCCAAATCGTATTAAAGTTTCGCCGTCCTGTTCA 1975

CTGAACCTTCCCTCATTTGGAGAATCTCTCCTCGCCAGCGACGCAAAGTCCTTAGGCAACTCTAGTTCACCTTGAATCT 2054
```

FIG. 2. *ERG9* nucleotide sequence and deduced squalene synthetase amino acid sequence. The translated sequence is shown in single-letter code below the *ERG9* nucleotide sequence. Potential membrane-spanning domains are underlined and a single PEST sequence is double-underlined. Potential N-linked glycosylation sites are bracketed and a putative prenyl substrate motif is underlined with a dashed line. In *ERG9* 5′ and 3′ untranslated regions, a TATA transcriptional regulatory element and a possible polyadenylylation signal are boxed. Restriction sites are labeled and underlined.

Table 2. Squalene synthetase activity in *E. coli* transformants

Plasmid	IPTG induction, hr	Specific activity, nmol·min^{-1}·mg^{-1}
pTTQ8	0	ND
	3	ND
	6	ND
pET124	0	0.006
	3	0.195
	6	0.446

Supernatant fractions (4000 × *g*) prepared from *E. coli* transformants were assayed for squalene synthetase activity in duplicate. ND, activity not detectable within the limit of the assay (≈0.003 nmol of squalene per min per mg of protein).

ERG9 overexpression plasmid pET124 had squalene synthetase activity levels comparable to those of wild-type yeast (Table 2). Cells bearing the parent plasmid pTTQ8 failed to produce measurable enzyme.

Squalene Synthetase Protein Sequence. The IFIND and FASTDB search algorithms did not reveal significant homologies between the predicted squalene synthetase amino acid sequence and any entry in the PIR or the Swiss-Prot protein sequence data bases. Several proteins that use prenyl substrates similar to those of squalene synthetase contain aspartate-rich consensus sequences proposed to be binding sites for the charged head group of these substrates (Fig. 3) (21). No perfect match to these sequence motifs was found in squalene synthetase, but a related sequence that may serve a similar role was identified extending from residue 214 to 236 (Figs. 2 and 3). No consensus dinucleotide (NADPH) binding motif (22) or endoplasmic reticulum membrane-retention motif (23) was found in the *ERG9* peptide sequence.

Squalene synthetase is an integral membrane protein, so its primary structure was examined for possible membrane transit peptides or transmembrane domains. No region within the N-terminal 100 amino acids was found that resembled a typical signal sequence (24). A hydrophobicity plot (25) revealed four very hydrophobic regions in the protein (Figs. 2 and 4) that met the criteria of Eisenberg *et al.* (26) for the prediction of membrane-associated helices. However, a hydrophobic moment plot classified only the third as a transmembrane sequence and the others as membrane-proximal sequences. By contrast, the algorithm of Rao and Argos (27) identified both the third and fourth domains as transmembrane sequences. A candidate PEST sequence (28), YDDELVPTQQEEEY (Fig. 2, residues 406–419), was identified just proximal to the fourth hydrophobic domain.

Deletion/Disruption of the *ERG9* Gene. To produce *ERG9* gene disruptions, a segment of the coding region on pSM52 was replaced with sequences that contained either the *HIS3* gene or the *LEU2* gene. The gene constructs designated pET105 and pET106 (Fig. 1) were then used to replace chromosomal alleles in diploid yeast by the one-step gene disruption method of Rothstein (29). Transformants of the homozygous strain SGY1161 (*ERG9/ERG9*) or the heterozygous strain SGY969 (*ERG9/erg9*) were selected as histidine or leucine prototrophs and screened for those bearing *ERG9* disruptions by probing chromosomal DNA with a radiolabeled *ERG9* restriction fragment. Five transformants yielded the pattern of *Xba* I restriction fragments expected to arise from integration of either the *HIS3* or the *LEU2* gene into an *ERG9* allele. Hybridization to chromosomal DNA of an undisrupted diploid gave only a single band, indicating that *ERG9* is present at just a single location in the *S. cerevisiae* genome (data not shown).

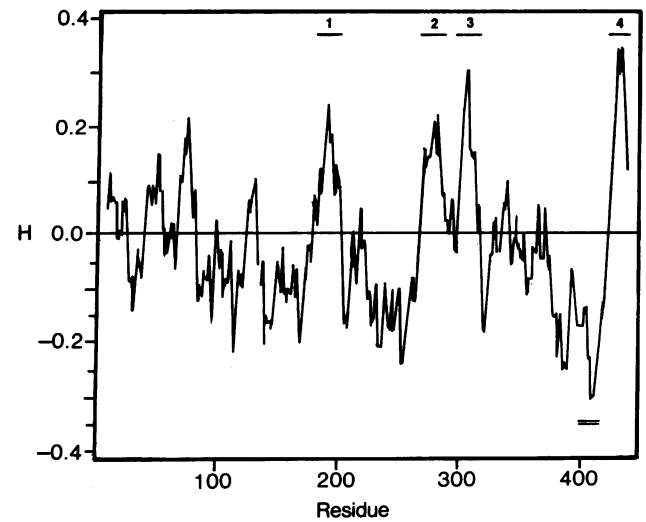

FIG. 4. Hydrophobicity plot of the amino acid sequence of yeast squalene synthetase. The hydrophobicity (H) of each residue, calculated with the SOAP program of Microgenie using a Kyte and Doolittle algorithm (25) and a window size of 15 amino acids, is plotted as a function of residue number. Hydrophobic regions that are predicted to be membrane associated (26) are labeled 1–4 and denote the sequences underlined in Fig. 2. A hydrophilic region containing a PEST sequence is double-underlined.

To examine the consequences of *ERG9* gene disruption, His$^+$ and Leu$^+$ transformants of diploid strain SGY1161 were sporulated to obtain haploids bearing these disruptions. Tetrad dissection on YPD medium yielded only two viable spores from each four-spored ascus. However, if dissection was done on YPD-E medium and tetrads were incubated anaerobically to promote sterol uptake (30), the disrupted diploids yielded predominantly tetrads with four viable spores. For 11 tetrads derived from a diploid bearing a *HIS3* disruption, the 2 His$^+$/2 His$^-$ pattern expected for segregation of a heterozygous marker was seen. Significantly, ergosterol auxotrophy segregated 2$^+$/2$^-$ and all the Erg$^-$ spores were His$^+$, showing that the Erg$^-$ phenotype was linked to the *ERG9*::*HIS3* disruption. Analogous results were obtained with 22 tetrads from a diploid bearing a *LEU2* disruption. As a control, 22 tetrads dissected from the wild-type parent diploid all gave four spores that were prototrophic for ergosterol but auxotrophic for both leucine

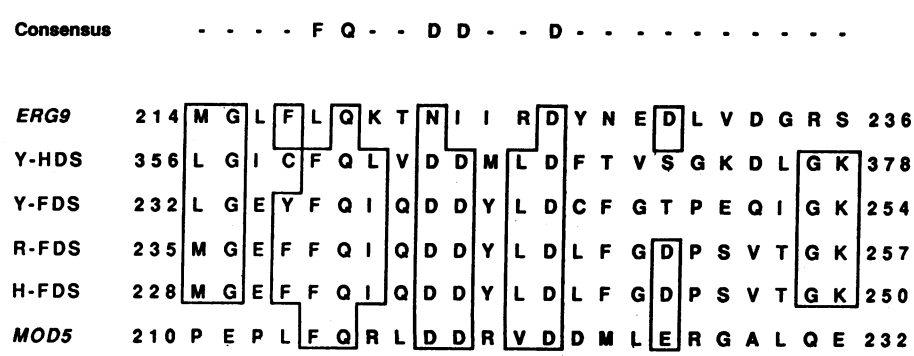

FIG. 3. Alignment of a peptide domain in squalene synthetase (*ERG9*) to a proposed allylic prenyl diphosphate binding motif (21). A consensus motif derived from the sequences of yeast hexaprenyl diphosphate synthetase (Y-HDS); farnesyl diphosphate synthetases from yeast (Y-FDS), rat (R-FDS), and human (H-FDS); and yeast tRNA isopentenyltransferase (*MOD5*) is shown. Numbers indicate residue positions in the primary sequences (cited in ref. 21).

and histidine. DNA–DNA blot hybridization using tetrad DNA derived from a diploid bearing *ERG9*::*HIS3* verified that disrupted and wild-type *ERG9* alleles segregated 2:2 (data not shown). Biochemical assay of cell-free extracts prepared from two sets of tetrads also showed that only the His^+ spores lacked squalene synthetase activity. Finally, RNA blot analysis of mRNA from a His^+ haploid showed that a normal *ERG9* transcript was missing in this strain (data not shown).

DISCUSSION

Comparison of the deduced amino acid sequence of squalene synthetase and the sequences of other cloned enzymes known to bind FDP or related substrates reveals no overall sequence conservation. However, weak homology was noted between a domain of squalene synthetase and a highly conserved sequence that is thought to comprise an allylic prenyl diphosphate binding site of polyprenyl synthetases (21). If this domain of squalene synthetase functions in prenyl substrate (FDP) binding, its dissimilarity may reflect a unique feature of the enzyme—namely, the catalysis of a head-to-head (1′–1) condensation reaction rather than the head-to-tail (1′–4) isoprene condensations catalyzed by many other prenyltransferases. Mevalonate pulse-labeling experiments with mammalian cells have suggested that squalene synthetase also has a lower affinity for FDP than do some of the other FDP-utilizing enzymes (1, 2). It is therefore possible that a divergent prenyl substrate binding site in squalene synthetase could have evolved, in part, to assist cells in maintaining steady ubiquinone and dolichol synthesis in the face of a variable requirement for intracellular sterol synthesis.

Modeling of the squalene synthetase secondary structure shows that this protein has four hydrophobic α-helical domains, one (the third) or two (the third and fourth) of which are predicted to constitute transmembrane domains. Interestingly, the most C-terminal domain (the fourth) has all the characteristics of a "simple" eukaryotic transmembrane domain (24), present in many proteins with a single membrane-spanning sequence. Such sequences are located within 40 residues of the C terminus; lack asparagine, glutamine, histidine, tryptophan, proline, and charged amino acids; and are immediately followed by multiple positively charged residues. Based on this, we speculate that only the fourth domain of squalene synthetase may be used as a transmembrane tether and that the rest of the protein projects from the surface of the endoplasmic reticulum membrane. In this topological model, the other hydrophobic domains could serve as pockets for binding the isoprene moieties of two FDP molecules.

Of potential significance for the regulation of squalene synthetase is the presence of a strong PEST sequence located just proximal to the fourth hydrophobic domain of the protein. PEST sequences are present in many proteins that experience rapid intracellular turnover and typically contain potential sites for serine/threonine phosphorylation, which may be used to regulate specific degradation by proteases such as calpains (28). Two consensus sites for casein kinase II phosphorylation are located near the *ERG9* PEST sequence (Ser-404 and Thr-413), as well as at other positions within the protein. 3-Hydroxy-3-methylglutaryl CoA reductase has PEST sequences within the linker region between its membrane-spanning domains and its catalytic domain (31). If the proposed topological model for squalene synthetase is correct, the *ERG9* PEST sequence may also be located within such a linker region, suggesting a similar architecture and turnover mechanism for both enzymes.

The deletion/disruption results demonstrate that squalene synthetase is a single-copy gene that is essential for vegetative growth of haploid yeast. Deletion of the structural genes for either of two enzymes that follow squalene synthetase in the sterol pathway, squalene epoxidase (*ERG1*) or lanosterol demethylase (*ERG16*), leads to similar ergosterol auxotrophy in yeast (J. Tkacz and M. Kurtz, personal communication; ref. 32). Because squalene synthetase is a branchpoint and likely regulatory enzyme in polyisoprene biosynthesis, the availability of *ERG9* null mutants and the cloned gene should prove valuable for future investigations of FDP and sterol regulation.

Note Added in Proof. The putative substrate-binding domain of *ERG9* is highly homologous to ones present in CrtB proteins from *R. sphaeroides* and *E. herbicola*. CrtB protein, like squalene synthetase, catalyzes a head-to-head condensation of two identical prenyl substrates but uses geranylgeranyl (C_{20}) instead of farnesyl (C_{15}) diphosphates (33).

We thank F. Karst for providing the *erg9* strain and J. Tkacz and A. M. Queenan for their help in characterization of the *erg9* mutation. We acknowledge C. Ciosek and D. Slusarchyk for the measurement of microsomal squalene synthetase activity. We thank C. Ciosek, J. Gorman, R. Gregg, M. Ondetti, and R. Tanaka for their comments on the manuscript. We thank the reviewers for their assistance in aligning the prenyl substrate homology region. We thank J. Alcantara for his contributions in the preparation of the illustrations. We acknowledge the Bristol-Myers Squibb Pharmaceutical Research Institute for supporting basic aspects of research in sterol metabolism.

1. Brown, M. S. & Goldstein, J. L. (1980) *J. Lipid Res.* **21**, 505–517.
2. Faust, J. R., Goldstein, J. L. & Brown, M. S. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5018–5022.
3. Poulter, D. C. & Rilling, H. C. (1981) in *Biosynthesis of Isoprenoid Compounds*, eds. Porter, J. W. & Spurgeon, S. L. (Wiley, New York), Vol. 1, pp. 414–441.
4. Popjack, G. & Agnew, W. S. (1979) *Mol. Cell. Biochem.* **27**, 97–116.
5. Kuswik-Rabiega, G. & Rilling, H. C. (1987) *J. Biol. Chem.* **262**, 1505–1509.
6. Agnew, W. S. (1985) *Methods Enzymol.* **110**, 359–373.
7. Sasiak, K. & Rilling, H. C. (1988) *Arch. Biochem. Biophys.* **260**, 622–627.
8. Agnew, W. S. & Popjack, G. (1978) *J. Biol. Chem.* **253**, 4574–4583.
9. Karst, F. & Lacroute, F. (1977) *Mol. Gen. Genet.* **154**, 269–277.
10. Hill, J. E., Myers, A. M., Koerner, T. J. & Tzagoloff, A. (1986) *Yeast* **2**, 163–167.
11. Parent, S. A., Fenimore, C. M. & Bostian, K. A. (1985) *Yeast* **1**, 83–138.
12. Stark, M. J. R. (1987) *Gene* **51**, 255–267.
13. Naumovski, L. & Friedberg, E. C. (1984) *Mol. Cell. Biol.* **4**, 290–295.
14. Hinen, A., Hicks, J. B. & Fink, G. R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1929–1933.
15. Hashimoto, H., Morikawa, H., Yamada, Y. & Kimura, A. (1985) *Appl. Microbiol. Biotechnol.* **21**, 336–339.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 440–441.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
19. Biller, S. A., Forster, C., Gordon, E. M., Harrity, T., Scott, W. A. & Ciosek, C. P., Jr. (1988) *J. Med. Chem.* **31**, 1869–1871.
20. Langford, C. J., Klinz, F. J., Donath, C. & Gallwitz, D. (1984) *Cell* **36**, 645–653.
21. Ashby, M. N., Spear, D. H. & Edwards, P. A. (1990) in *Molecular Biology of Atherosclerosis*, ed. Attie, A. D. (Elsevier, Amsterdam), pp. 27–34.
22. Moller, W. & Amons, R. (1985) *FEBS Lett.* **186**, 1–7.
23. Jackson, M. R., Nilsson, T. & Peterson, P. A. (1990) *EMBO J.* **9**, 3153–3162.
24. Von Heijne, G. (1985) *Curr. Top. Membr. Transport* **24**, 151–179.
25. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
26. Eisenberg, D., Schwartz, E., Komaromy, M. & Wall, R. (1984) *J. Mol. Biol.* **179**, 125–142.
27. Rao, J. K. M. & Argos, P. (1986) *Biochim. Biophys. Acta* **869**, 197–214.
28. Rogers, S., Wells, R. & Rechsteiner, M. (1986) *Science* **234**, 364–368.
29. Rothstein, R. J. (1983) *Methods Enzymol.* **101**, 202–211.
30. Andreasen, A. A. & Stier, T. J. B. (1953) *J. Cell. Comp. Physiol.* **41**, 23–35.
31. Basson, M. E., Thorsness, M., Finer-Moore, J., Stroud, R. M. & Rine, J. (1988) *Mol. Cell. Biol.* **8**, 3797–3808.
32. Kalb, V. F., Woods, C. W., Dey, C. R., Sutter, T. R., Turi, T. G. & Loper, J. C. (1987) *DNA* **6**, 529–537.
33. Armstrong, G. A. (1989) Ph.D. dissertation (University of California at Berkeley).