Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 11628–11632, December 1993
Chemistry

Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen

(sterol biosynthesis/2,3-epoxysqualene/secondary metabolite/gene cloning)

E. J. COREY*, SEIICHI P. T. MATSUDA*, AND BONNIE BARTEL†

*Department of Chemistry, Harvard University, Cambridge, MA 02138; and †Whitehead Institute for Biomedical Research, Cambridge, MA 02142

Contributed by E. J. Corey, September 2, 1993

ABSTRACT **Whereas vertebrates and fungi synthesize sterols from epoxysqualene through the intermediate lanosterol, plants cyclize epoxysqualene to cycloartenol as the initial sterol. We report the cloning and characterization of *CAS1*, an *Arabidopsis thaliana* gene encoding cycloartenol synthase [(*S*)-2,3-epoxysqualene mutase (cyclizing, cycloartenol forming), EC 5.4.99.8]. A yeast mutant lacking lanosterol synthase [(*S*)-2,3-epoxysqualene mutase (cyclizing, lanosterol forming), EC 5.4.99.7] was transformed with an *A. thaliana* cDNA yeast expression library, and colonies were assayed for epoxysqualene mutase activity by thin-layer chromatography. One out of ≈10,000 transformants produced a homogenate that cyclized 2,3-epoxysqualene to the plant sterol cycloartenol. This activity was shown to be plasmid dependent. The plasmid insert contains a 2277-bp open reading frame capable of encoding an 86-kDa protein with significant homology to lanosterol synthase from *Candida albicans* and squalene-hopene cyclase (EC 5.4.99.-) from *Bacillus acidocalcarius*. The method used to clone this gene should be generally applicable to genes responsible for secondary metabolite biosynthesis.**

The epoxysqualene mutases compose a family of exceptionally sophisticated catalysts that convert (*S*)-2,3-epoxysqualene, through a remarkable sequence of cation-π cyclizations and cationic 1,2 rearrangements of carbon and hydrogen, to polycyclic triterpenes with dramatic structural differences from the starting material. Cycloartenol synthase (1, 2) [(*S*)-2,3-epoxysqualene mutase (cyclizing, cycloartenol forming), EC 5.4.99.8] must break 11 bonds and form 11 new ones to transform the acarbocyclic epoxysqualene (Fig. 1, compound **1**) to the plant sterol precursor cycloartenol (Fig. 1, compound **2**), a pentacyclic triterpene that contains nine chiral centers. The mammalian (3, 4) and fungal (5) lanosterol synthases [(*S*)-2,3-epoxysqualene mutase (cyclizing, lanosterol forming), EC 5.4.99.7] produce lanosterol (Fig. 1, compound **3**), the common precursor of cholesterol in vertebrates and ergosterol in fungi. Lanosterol synthase catalyzes a slightly less complex reaction than do the plant terpene synthases, but the former are more commonly studied, in part, because they are more readily accessible (6–8). Mammalian liver and *Saccharomyces cerevisiae* (bakers' yeast) homogenates contain sufficient lanosterol synthase activity to be used directly in enzymatic reactions. Moreover, these sources produce only one epoxysqualene mutase; epoxysqualene introduced into hog liver or yeast homogenates yield lanosterol as the sole product. The plant cycloartenol synthase presents considerably more experimental difficulty. Plants generally cyclize epoxysqualene to one or more triterpenes in addition to cycloartenol, which precludes using crude homogenates as a clean source of a single catalytic activity. In addition, plant homogenates often contain too little enzymatic activity to produce useful amounts of products. A cloned and heterologously expressed cycloartenol synthase would circumvent the experimental problems associated with crude plant homogenates. The molecular details of cycloartenol synthase could then be scrutinized by using synthetic substrate analogs, as we have done with lanosterol synthase (9–11).

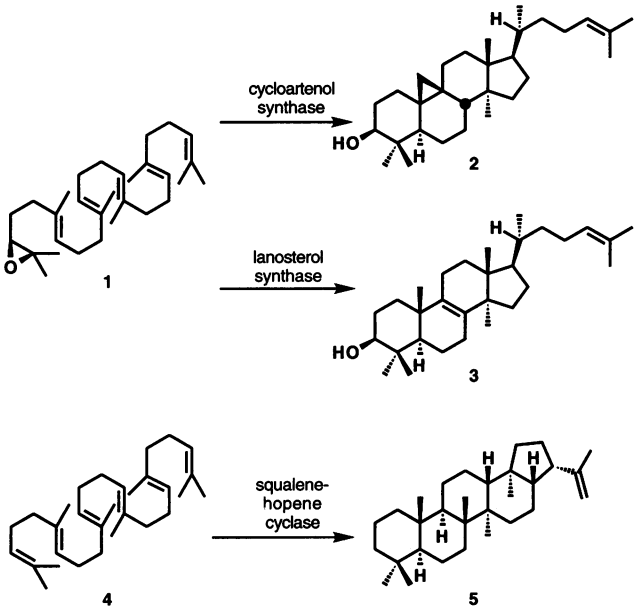

FIG. 1. Reactions catalyzed by terpene synthases. Epoxysqualene (**1**) is cyclized by lanosterol synthase to lanosterol (**3**), and by cycloartenol synthase to cycloartenol (**2**). Squalene-hopene cyclase cyclizes squalene (**4**) to hopene (**5**).

As with many enzymes involved in secondary metabolism, cycloartenol synthase could not be cloned by complementing a mutation in a metabolic pathway because the organisms most amenable to this technique (*Escherichia coli, S. cerevisiae*) do not produce this enzyme. We present here a simple, rapid procedure to clone epoxysqualene mutases that requires neither pure protein nor a metabolic mutant in that gene. Yeast transformed with a plant cDNA library under a constitutive yeast promoter were assayed for cycloartenol synthase activity by TLC. A mutant yeast strain (SMY1 lacking *erg7*) was used as the host to abolish background epoxysqualene mutase activity in the assay. The crucifer *Arabidopsis thaliana* was chosen as the source of genetic material. *A. thaliana* has a small genome, is easily propagated and transformed, and provides an excellent genetic system in which to generate recombinant plants engineered with altered abilities to synthesize cycloartenol.

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.

This technique should be broadly applicable to other genes that cannot be cloned by complementation, such as those involved in secondary metabolism.‡

EXPERIMENTAL PROCEDURES

Strains and Plasmids. The following yeast strains were used: GL7 (12) (*MATa gal2 hem3-6 erg7*), YPH252 (13) (*MATα ura3-52 lys2-801 ade2-101 trp1-Δ1 his3-Δ200 leu2-Δ1*) and SMY1 [*gal2 hem3-6 erg7 ura3-52*, obtained by crossing strain GL7 to strain YPH252, sporulating the resultant diploid, and selecting a segregant with the desired phenotype (14)]. The plasmid vector pBluescript II (Stratagene) was used for subcloning and was propagated in *E. coli* strain DH5α (15). The *A. thaliana* cDNA library (16) was provided by F. Karst (Université de Poitiers). pSM60 is a plasmid from this library containing the *CAS1* gene under the control of the *PGK1* promoter and terminator sequences.

Media, Growth, and Transformation Conditions. Strains GL7 and SMY1 were propagated according to published procedures for strain GL7, which requires media supplemented with ergosterol (20 μg/ml), heme (13 μg/ml), and Tween 80 (5 mg/ml) (12). Yeast were transformed according to published procedures (17). Transformants were selected on synthetic complete medium plates without uracil (14) and supplemented with ergosterol (20 μg/ml), heme (13 μg/ml), and Tween 80 (5 mg/ml), and propagated in the corresponding liquid medium. *E. coli* were transformed, selected, and propagated according to published procedures (15).

Cloning. The yeast strain SMY1 was transformed with an *A. thaliana* cDNA expression library (16). Nine hundred and fifty pools of ≈10 transformants each were grown to saturation in 2.5 ml of liquid medium at 30°C. Cells from 2 ml of each pool were collected by centrifugation, washed once with water, resuspended in 50 μl of 100 mM sodium phosphate buffer at pH 7.0 containing 0.1% Triton X-100, 1 mM dithiothreitol, and epoxysqualene at 1 mg/ml. The cells were permeabilized by first freezing at −78°C in dry ice followed by thawing at 30°C (18). After 3 hr at room temperature, 5 μl of each reaction mixture was spotted directly onto a TLC plate (0.25-mm Kieselgel 60; Merck) alongside a cycloartenol standard, and the water was removed at 30 torr (1 torr = 133.3 Pa) for 1 hr. The plate was then developed with hexane/ethyl ether, 1:1, and visualized with *p*-anisaldehyde stain (19). In this system, the retardation factors (R_f) are 0.82 for epoxysqualene, 0.49 for cycloartenol and lanosterol, and 0.31 for ergosterol.

DNA and RNA Methods. Plasmid DNA was isolated from yeast as described (20). Standard DNA analysis and cloning techniques were done as described (15). DNA fragments were purified from agarose gels by using Qiaex (Qiagen, Chatsworth, CA). DNA was sequenced by the dideoxynucleotide chain-termination method with the Applied Biosystems automated system. The reported sequence was sequenced at least once on each strand.

Hybridization probes were labeled by using [α-^{32}P]dCTP and random hexamer primers (International Biotechnologies). DNA and RNA were purified from 16-day-old rosette leaves of *A. thaliana* var. Columbia (15). DNA was digested with restriction endonucleases, electrophoresed through 1% agarose, and transferred to GeneScreen (DuPont) in 0.4 M NaOH. Total RNA was electrophoresed through 1% agarose-formaldehyde (15) and transferred to GeneScreen in 1.5 M NaCl/0.15 M sodium citrate, pH 7.0. Blots were hybridized overnight at 65°C, as described (21), and were washed in 30 mM NaCl/3 mM sodium citrate/0.1% SDS, pH 7.0 at 25°C.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. U02555).

Enzymatic Synthesis and Identification of Cycloartenol. SMY1 cells transformed with pSM60 (10 g) were suspended in 50 ml of 100 mM sodium phosphate buffer at pH 7.0 and lysed in a French Press at 20,000 psi (1 psi = 6.9 kPa), and the homogenate was centrifuged at 10,000 × *g* for 20 min. One milliliter of (±)-2,3-epoxysqualene solution (10 mg/ml) in 20% Triton X-100 (22, 23) was added to the supernatant. After 8 hr at 25°C, 100 ml of tetrahydrofuran was added to terminate the reaction. The precipitate (denatured protein and salt) was removed by filtration, and the filtrate was concentrated under vacuum. The remaining residue was extracted three times with three volumes of ethyl ether, and the combined organic extracts were dried over sodium sulfate, concentrated under vacuum, and purified by silica gel chromatography (5% ethyl ether in hexane) to yield 3.8 mg [79% based on (*S*)-epoxysqualene] of a colorless solid. This material was identical to an authentic sample of cycloartenol by ^{1}H and ^{13}C NMR, IR, and MS analyses, which gave spectra consistent with available literature data (24–26). Retardation factor (R_f) = 0.49 (silica gel, hexane/ethyl ether, 1:1). Fourier transform IR spectroscopy (thin film) 3309.2, 2962.0, 1449.4, 1375.2, 1096.6, 1047.5, 1004.6, 994.1 cm^{-1}. ^{1}H NMR (500 MHz, $CDCl_3$) δ 0.33 (d, J = 4.2 Hz, 1H), 0.55 (d, J = 4.2 Hz, 1H), 0.81 (s, 3H), 0.89 (s, 6H), 0.96 (s, 3H), 0.97 (s, 3H), 0.99–1.14 (m, 2H), 1.21–1.46 (m, 9H), 1.49 (dd, J = 4.9, 12.4, 1H), 1.54–1.64 (m, 7H), 1.59 (s, 3H), 1.68 (s, 3H), 1.73–1.79 (m, 1H), 1.82–1.93 (m, 2H), 1.95–2.08 (m, 2H), 3.24–3.32 (m, 1H), 5.10 (t, J = 7.1, 1H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 14.0, 17.6, 18.0, 18.3, 19.3, 20.1, 21.1, 25.0, 25.4, 25.7, 26.0, 26.2, 26.5, 28.1, 29.9, 30.4, 32.0, 33.0, 35.6, 35.9, 36.4, 40.6, 45.3, 47.1, 48.0, 48.8, 52.3, 78.9, 125.3, 130.9. MS (fast atom bombardment): m/z 449 $[M+Na]^+$, 426 $[M]^+$, 409, 393. High-resolution mass spectrum parent ion (fast atom bombardment, $[M+Na]^+$) m/z calculated for $[C_{30}H_{50}ONa]^+$: 449.3759, found: 449.3748.

RESULTS

We have cloned the *A. thaliana* cycloartenol synthase gene while demonstrating the feasibility of a chromatographic screen for heterologously expressing transformants. Strain SMY1 was transformed with a yeast expression library containing *A. thaliana* cDNAs (16) under the control of the constitutive yeast *PGK1* promoter (27), and the resultant transformants were divided into pools and assayed for epoxysqualene mutase activity by TLC (see *Experimental Procedures*). The single positive pool was replated for single colonies, and homogenates of 3 of 20 colonies metabolized epoxysqualene in the same assay system. Plasmid DNA was isolated from these strains and used to transform *E. coli*. Plasmids from the three resultant *E. coli* lines were shown to be identical to one another by restriction enzyme mapping, and the plasmid was named pSM60. A homogenate of SMY1 transformed with pSM60 cyclized epoxysqualene to the plant sterol cycloartenol (see *Experimental Procedures*). Because *S. cerevisiae* cannot cyclize epoxysqualene to cycloartenol and yeast transformed with pSM60 acquires the ability to catalyze this reaction, the cDNA in this plasmid is the structural gene for cycloartenol synthase. The gene encoding this cDNA was therefore named *CAS1* (*c*yclo*a*rtenol *s*ynthase).

The nucleotide and predicted amino acid sequences of the CAS1 cDNA are shown in Fig. 2. The 2277-bp open reading frame is capable of encoding an 86-kDa protein. Gel-blot analysis of RNA prepared from *A. thaliana* leaves revealed a ≈2.3-kb message that hybridized to the CAS1 probe (data not shown), suggesting that the cloned cDNA is full length. Although *A. thaliana* cycloartenol synthase has not been purified, the epoxysqualene cyclizing enzymes from two other plants have been purified [*Pisum sativum*, 55 kDa (28); *Rabdosia japonica*, 54 kDa (29)]; they are significantly smaller than the predicted CAS1 protein and may be proteo-

```
                                      tcaaactcattttccgatcaaacggcatatttttccctctcgagagagtgata    53
ATGTGGAAACTGAAGATCGCGGAAGGAGGTAGTCCATGGCTTAGAACCACCAATAATCACGTCGGAAGACAGTTTTGGGAGTTCGATCCG   143
 M  W  K  L  K  I  A  E  G  G  S  P  W  L  R  T  T  N  N  H  V  G  R  Q  F  W  E  F  D  P
AATCTCGGTACTCCTGAGGATCTCGCCGCCGTCGAAGAAGCTAGGAAGTCTTTTCCGGATAATCGATTCGTGCAGAAACATAGCGCCGAT   233
 N  L  G  T  P  E  D  L  A  A  V  E  E  A  R  K  S  F  P  D  N  R  F  V  Q  K  H  S  A  D
CTGCTTATGCGCCTTCAGTTTTCAAGAGAAAATTTGATTAGCCCAGTTTTACCTCAAGTCAAAATCGAAGACACTGATGATGTTACAGAG   323
 L  L  M  R  L  Q  F  S  R  E  N  L  I  S  P  V  L  P  Q  V  K  I  E  D  T  D  D  V  T  E
GAGATGGTGGAAACCACGTTAAAGAGGGGTCTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGGTGATTATGGTGGT   413
 E  M  V  E  T  T  L  K  R  G  L  D  F  Y  S  T  I  Q  A  H  D  G  H  W  P  G  D  Y  G  G
CCTATGTTTCTTCTCCCAGGACTGATAATTACACTCTCCATAACTGGAGCACTGAATACAGTATTGTCGGAACAACATAAACAAGAAATG   503
 P  M  F  L  L  P  G  L  I  I  T  L  S  I  T  G  A  L  N  T  V  L  S  E  Q  H  K  Q  E  M
CGCCGTTATCTCTATAATCACCAGAATGAGGACGGAGGTTGGGGTTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAAC   593
 R  R  Y  L  Y  N  H  Q  N  E  D  G  G  W  G  L  H  I  E  G  P  S  T  M  F  G  S  V  L  N
TATGTTACTCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATGGAGAAAGGACGAGACTGGATACTAAAATATGGTGGT   683
 Y  V  T  L  R  L  L  G  E  G  P  N  D  G  D  G  D  M  E  K  G  R  D  W  I  L  K  Y  G  G
GCTACCAATATTACATCTTGGGGGAAAATGTGGCTATCGGTACTTGGAGCTTTTGAATGGTCCGGAAATAACCCACTGCCACCTGAGATA   773
 A  T  N  I  T  S  W  G  K  M  W  L  S  V  L  G  A  F  E  W  S  G  N  N  P  L  P  P  E  I
TGGCTTCTCCCATATTTCCTGCCAATACATCCAGGAAGGATGTGGTGCCATTGTCGAATGGTGTACTTGCCGATGTCGTATTTGTATGGA   863
 W  L  L  P  Y  F  L  P  I  H  P  G  R  M  W  C  H  C  R  M  V  Y  L  P  M  S  Y  L  Y  G
AAAAGGTTTGTGGGTCCCATAACGTCCACTGTTTTATCACTGAGAAAGGAGCTTTTCACAGTACCATATCATGAAGTCAACTGGAATGAA   953
 K  R  F  V  G  P  I  T  S  T  V  L  S  L  R  K  E  L  F  T  V  P  Y  H  E  V  N  W  N  E
GCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCCACTTGTGCAAGATATTCTTTGGGCATCACTTCATAAGATTGTTGAG  1043
 A  R  N  L  C  A  K  E  D  L  Y  Y  P  H  P  L  V  Q  D  I  L  W  A  S  L  H  K  I  V  E
CCTGTTCTGATGCGATGGCCTGGTGCAAATTTGAGAGAAAAGGCTATAAGAACCGCAATAGAACATATTCATTATGAAGATGAGAATACT  1133
 P  V  L  M  R  W  P  G  A  N  L  R  E  K  A  I  R  T  A  I  E  H  I  H  Y  E  D  E  N  T
AGGTACATCTGCATAGGTCCCGTGAACAAGGTATTAAATATGCTTTGCTGTTGGGTAGAAGACCCAAACTCAGAGGCTTTCAAGTTGCAC  1223
 R  Y  I  C  I  G  P  V  N  K  V  L  N  M  L  C  C  W  V  E  D  P  N  S  E  A  F  K  L  H
CTACCAAGAATCCATGACTTTCTCTGGTTAGCTGAAGATGGAATGAAGATGCAGGGTTATAACGGAAGCCAGCTATGGGATACAGGTTTT  1313
 L  P  R  I  H  D  F  L  W  L  A  E  D  G  M  K  M  Q  G  Y  N  G  S  Q  L  W  D  T  G  F
GCTATTCAAGCGATTTTGGCAACTAACCTCGTCGAAGAATATGGGCCCGTTTTGGAAAAAGCACATTCATTTGTCAAGAATTCCCAGGTG  1403
 A  I  Q  A  I  L  A  T  N  L  V  E  E  Y  G  P  V  L  E  K  A  H  S  F  V  K  N  S  Q  V
TTAGAAGACTGCCCTGGAGATCTGAATTACTGGTATCGCCACATTTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCC  1493
 L  E  D  C  P  G  D  L  N  Y  W  Y  R  H  I  S  K  G  A  W  P  F  S  T  A  D  H  G  W  P
ATCTCTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGCTATCCAAAGTTCCCAAGGAGATTGTTGGTGAACCAATAGATGCAAAA  1583
 I  S  D  C  T  A  E  G  L  K  A  A  L  L  L  S  K  V  P  K  E  I  V  G  E  P  I  D  A  K
CGGTTATATGAAGCTGTTAATGTTATCATTTCTTTACAGAATGCAGATGGAGGCCTCGCAACATATGAGCTCACCAGGTCATACCCTTGG  1673
 R  L  Y  E  A  V  N  V  I  I  S  L  Q  N  A  D  G  G  L  A  T  Y  E  L  T  R  S  Y  P  W
TTAGAGCTAATCAACCCAGCAGAAACCTTTGGCGATATTGTTATTGATTATCCTTACGTGGAATGTACATCAGCTGCTATCCAAGCTTTG  1763
 L  E  L  I  N  P  A  E  T  F  G  D  I  V  I  D  Y  P  Y  V  E  C  T  S  A  A  I  Q  A  L
ATATCATTTCGAAAGCTGTATCCTGGTCATCGAAAGAAGGAAGTAGATGAGTGCATTGAGAAGGCGGTTAAGTTCATTGAATCCATTCAA  1853
 I  S  F  R  K  L  Y  P  G  H  R  K  K  E  V  D  E  C  I  E  K  A  V  K  F  I  E  S  I  Q
GCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGCTTCACGTATGGTACGTGGTTTGGAGTGAAAGGGCTGGTAGCTGTTGGAAAA  1943
 A  A  D  G  S  W  Y  G  S  W  A  V  C  F  T  Y  G  T  W  F  G  V  K  G  L  V  A  V  G  K
ACATTGAAAAACTCTCCACATGTTGCTAAAGCTTGTGAATTTCTATTGTCGAAACAACAACCTTCGGGCGGCTGGGGAGAAAGCTATCTT  2033
 T  L  K  N  S  P  H  V  A  K  A  C  E  F  L  L  S  K  Q  Q  P  S  G  G  W  G  E  S  Y  L
TCATGTCAAGACAAGGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACAGCATGGGCTATGCTCGCACTCATTGGTGCT  2123
 S  C  Q  D  K  V  Y  S  N  L  D  G  N  R  S  H  V  V  N  T  A  W  A  M  L  A  L  I  G  A
GGGCAAGCTGAGGTAGACCGGAAACCACTACACCGGGCTGCAAGATACTTGATTAATGCTCAAATGGAGAATGGTGATTTTCCACAACAG  2213
 G  Q  A  E  V  D  R  K  P  L  H  R  A  A  R  Y  L  I  N  A  Q  M  E  N  G  D  F  P  Q  Q
GAAATAATGGGAGTCTTCAATAGGAACTGCATGATAACATATGCCGCGTATCGAAACATTTTTCCGATATGGGCGTTGGGGGAGTACCGT  2303
 E  I  M  G  V  F  N  R  N  C  M  I  T  Y  A  A  Y  R  N  I  F  P  I  W  A  L  G  E  Y  R
TGTCAGGTATTATTGCAACAAGGAGAATGAattaagttttcaaattggttatttctctagggaaggaactgtgtagcttccatgctttac  2393
 C  Q  V  L  L  Q  Q  G  E ter
tctttccaaagttgtgttatccgtaatggatctgctctgatggttacgtgaaccaatctttgtgtaattgtttcaaactcaaatcttggt  2483
ggcagattttgntggnttctcttgaatacccgattttaggataatgtcttgttttgngaaaacagacattttcttgt              2560
```

FIG. 2. Nucleotide and deduced amino acid sequences of the CAS1 cDNA. The DNA sequence of the *CAS1* open reading frame is shown in uppercase letters, and the 5′ and 3′ untranslated regions are shown in lowercase letters. The sequence was followed by 58 adenoside residues beginning at residue 2560.

lytic fragments produced as posttranslational modifications or during the purification process.

Gel-blot analysis of genomic *A. thaliana* DNA hybridized with the CAS1 probe at high stringency revealed a strongly hybridizing band and several weakly hybridizing bands (data not shown). The most predominant band probably represents the *CAS1* gene, and the other cross-hybridizing bands may represent additional terpene synthases.

The CAS1 protein is 34% identical to the *Candida albicans* lanosterol synthase (30, 31) and 18% identical to the *Bacillus acidocalcarius* squalene-hopene cyclase (EC 5.4.99.-) (32). The homology (Fig. 3) of these proteins is not unexpected, considering that the three enzymes catalyze mechanistically related reactions of squalene or its 2,3 oxide (Fig. 1).

DISCUSSION

The most difficult and time-consuming aspect of gene isolation is often obtaining either pure protein for partial sequence determination or antibody production, or a mutant for complementation experiments. Strategies that depend on pure protein or mutants are not always practicable; protein purification is often nontrivial, and many enzymes, such as most of those involved in secondary metabolism, are not represented in organisms commonly used in complementation experiments. When neither a mutant in the desired gene nor pure protein product of that gene is readily available, an alternate approach is to use an *in vitro* assay to identify library transformants that heterologously express a desired enzymatic activity. When a simple assay can be devised, screening directly for enzymatic activity present in transformants is a generally applicable, expeditious alternative to the traditional methods of cloning genes. The only limitation of this method is that the gene must encode an enzyme that can be expressed at detectable levels in an easily transformed organism such as *E. coli* or *S. cerevisiae*. If the host cell has an enzyme related to the target enzyme, using a host with a mutation in that enzyme simplifies the assay by eliminating background activity.

Yeast was chosen as the host to clone the *A. thaliana* cycloartenol synthase for several reasons. (*i*) Yeast is easily transformed, yielding up to 10^6 transformants per μg of DNA (17); (*ii*) a eukaryotic host such as yeast is more likely to process or modify correctly eukaryotic proteins than is a prokaryotic host such as *E. coli*; and (*iii*) strong yeast promoters are known (33) that should produce readily detectable activity in positive transformants.

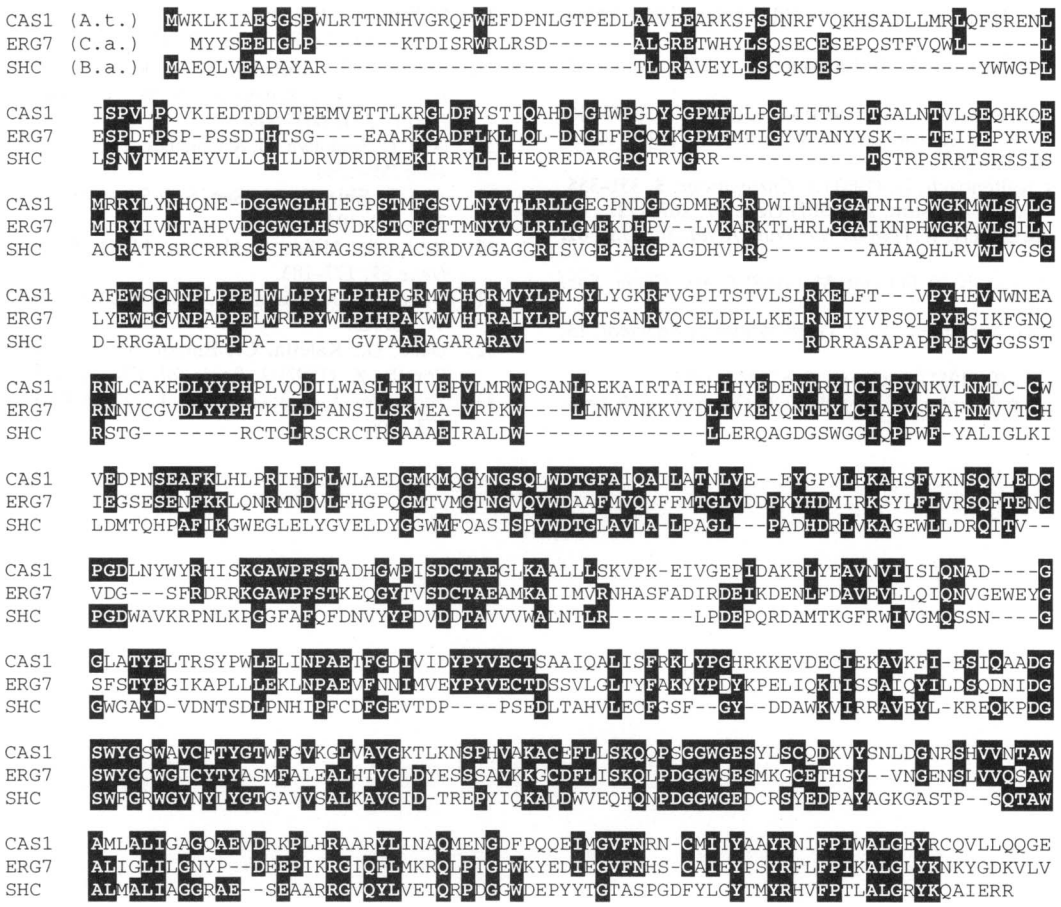

FIG. 3. Alignment of terpene synthase amino acid sequences. The sequences shown are *A. thaliana* CAS1 [CAS1 (A. t.)], *C. albicans* ERG7 [ERG7 (C. a.) (31)], and *B. acidocalcarius* squalene-hopene cyclase [SHC (B. a.) (32)]. Amino acid positions identical in at least two of the three sequences are boxed; hyphens indicate gaps introduced to maximize alignment.

TLC is an ideal assay method for this purpose; it is very sensitive for identifying small organic molecules (10 pmol of lanosterol is visible with *p*-anisaldehyde stain) and is a simple technique that can be performed rapidly with inexpensive equipment. Unlike quantitative analytical methods such as GC-MS or HPLC, TLC can accommodate crude homogenate assays without workup, which would be a major portion of the labor with a GC-MS- or HPLC-based screen. The sensitivity of TLC permits assaying pools of transformants rather than assaying individual colonies, which also minimizes labor.

Cycloartenol and lanosterol are nearly indistinguishable by chromatography. Although a TLC-based *in vitro* assay can provide qualitative data about the conversion of epoxysqualene to cyclic products, it allows neither quantitation nor identification of the newly synthesized triterpene. Therefore, TLC cannot be used to distinguish between wild-type yeast producing normal levels of yeast lanosterol synthase and recombinant yeast producing *A. thaliana* cycloartenol synthase in addition to native yeast lanosterol synthase. To circumvent this problem, we used the yeast strain SMY1 (*erg7*), which lacks lanosterol synthase and so cannot convert epoxysqualene to lanosterol, a compound that comigrates with cycloartenol on TLC.

This cloned and heterologously expressed *A. thaliana* cycloartenol synthase will facilitate purification of the enzyme and allow detailed analysis of the reaction mechanism by using substrate analogs. The regions of identity between the known terpene synthases shown in Fig. 3 will simplify cloning other epoxysqualene mutases.

We are indebted to the DNA sequencing facility of Pfizer Inc. (Groton, CT) and, in particular, Drs. Michael Fogliano and Arthur E. Franke for the sequence shown in Fig. 2, to Prof. Francis Karst for the *A. thaliana* cDNA library, to Prof. Charles Dahl for the GL7 strain, and to Prof. William N. Lipscomb, Dr. John Celenza, Prof. Per O. Ljungdahl, and Prof. Jennifer Normanly for comments on the manuscript. This research was assisted financially by the National Institutes of Health.

1. Rees, H. H., Goad, L. J. & Goodwin, T. S. (1968) *Tetrahedron Lett.* **9,** 723–725.
2. Rees, H. H., Goad, L. J. & Goodwin, T. W. (1969) *Biochim. Biophys. Acta* **176,** 892–894.
3. Corey, E. J., Russey, W. E. & Ortiz de Montellano, P. R. (1966) *J. Am. Chem. Soc.* **88,** 4750–4851.
4. Dean, P. D. G., Ortiz de Montellano, P. R., Bloch, K. & Corey, E. J. (1967) *J. Biol. Chem.* **242,** 3014–3015.
5. Barton, D. H. R., Gosden, A. F., Mellows, G. & Widdowson, D. A. (1968) *Chem. Commun.*, 1067–1068.
6. Harrison, D. M. (1985) *Nat. Prod. Rep.* **2,** 525–560.
7. Harrison, D. M. (1988) *Nat. Prod. Rep.* **5,** 387–415.
8. Harrison, D. M. (1990) *Nat. Prod. Rep.* **7,** 459–484.
9. Corey, E. J. & Virgil, S. C. (1991) *J. Am. Chem. Soc.* **113,** 4025–4026.
10. Corey, E. J., Virgil, S. C. & Sarshar, S. (1991) *J. Am. Chem. Soc.* **113,** 8171–8172.
11. Corey, E. J., Virgil, S. C., Liu, D. R. & Sarshar, S. (1992) *J. Am. Chem. Soc.* **114,** 1524–1525.
12. Gollub, E. G., Liu, K., Dayan, J., Adlersberg, M. & Sprinson, D. B. (1977) *J. Biol. Chem.* **252,** 2846–2854.
13. Sikorski, R. S. & Hieter, P. (1989) *Genetics* **122,** 19–27.
14. Guthrie, C., Fink, G. R., eds. (1991) *Guide to Yeast Genetics and Molecular Biology* (Academic, New York).
15. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K., eds. (1993) *Current*

Protocols in Molecular Biology (Wiley-Interscience, New York).
16. Minet, M., Dufour, M.-E. & Lacroute, F. (1992) *Plant J.* **2**, 417–422.
17. Schiestl, R. H. & Gietz, R. D. (1989) *Curr. Genet.* **16**, 339–346.
18. Miozzari, G. F., Niederberger, P. & Hütter, R. (1978) *Anal. Biochem.* **90**, 220–233.
19. Stahl, E. & Kaltenbach, U. (1961) *J. Chromatogr.* **5**, 351–355.
20. Hoffman, C. S. & Winston, F. (1987) *Gene* **57**, 267–272.
21. Church, G. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
22. Corey, E. J., Lin, K. & Jautelat, M. (1968) *J. Am. Chem. Soc.* **90**, 2724–2726.
23. Nadeau, R. G. & Hanzlik, R. P. (1968) *Methods Enzymol.* **15**, 346–351.
24. Barton, D. H. R. (1951) *J. Chem. Soc.* 1444–1451.
25. Aplin, R. T. & Hornby, G. M. (1966) *J. Chem. Soc. B* **1966**, 1078–1079.
26. Radics, L., Kajtar-Peredy, M., Corsano, S. & Standoli, L. (1975) *Tetrahedron Lett.* **17**, 4287–4290.
27. Tuite, M. F., Dobson, M. J., Roberts, N. A., King, R. M., Burke, D. C., Kingsman, S. M. & Kingsman, A. J. (1982) *EMBO J.* **1**, 603–608.
28. Abe, I., Sankawa, U. & Ebizuka, Y. (1992) *Chem. Pharm. Bull.* **40**, 1755–1760.
29. Abe, I., Ebizuka, Y., Seo, S. & Sankawa, U. (1989) *FEBS Lett.* **249**, 100–104.
30. Kelly, R., Miller, S. M., Lai, M. H. & Kirsch, D. R. (1990) *Gene* **87**, 177–183.
31. Buntel, C. J. & Griffin, J. H. (1992) *J. Am. Chem. Soc.* **114**, 9711–9713.
32. Ochs, D., Kaletta, C., Entian, K.-D., Beck-Sickinger, A. & Poralla, K. (1992) *J. Bacteriol.* **174**, 298–302.
33. Schneider, J. C. & Guarante, L. (1991) *Methods Enzymol.* **194**, 373–388.