A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases

Melissa Salmon[a,1], Ramesha B. Thimmappa[a,1], Robert E. Minto[b], Rachel E. Melton[a], Richard K. Hughes[a], Paul E. O'Maille[a,c], Andrew M. Hemmings[d,e], and Anne Osbourn[a,2]

[a]John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom; [b]Department of Chemistry and Chemical Biology, Indiana University–Purdue University, Indianapolis, IN 46202; [c]Food & Health Programme, Institute of Food Research, Norwich Research Park, Norwich NR4 7UH, United Kingdom; [d]School of Chemistry, University of East Anglia, Norwich NR4 7TJ, United Kingdom; and [e]School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom

Edited by Rodney B. Croteau, Washington State University, Pullman, WA, and approved June 9, 2016 (received for review April 5, 2016)

Triterpenes are structurally complex plant natural products with numerous medicinal applications. They are synthesized through an origami-like process that involves cyclization of the linear 30 carbon precursor 2,3-oxidosqualene into different triterpene scaffolds. Here, through a forward genetic screen in planta, we identify a conserved amino acid residue that determines product specificity in triterpene synthases from diverse plant species. Mutation of this residue results in a major change in triterpene cyclization, with production of tetracyclic rather than pentacyclic products. The mutated enzymes also use the more highly oxygenated substrate dioxidosqualene in preference to 2,3-oxidosqualene when expressed in yeast. Our discoveries provide new insights into triterpene cyclization, revealing hidden functional diversity within triterpene synthases. They further open up opportunities to engineer novel oxygenated triterpene scaffolds by manipulating the precursor supply.

terpenes | natural products | plant defense | cyclization | mutants

The triterpenes are one of the largest and most diverse groups of plant natural products (1). These compounds have numerous pharmaceutical, agricultural, and industrial biotechnology applications (2–5). The ability to harness this diversity to engineer known and new-to-nature triterpenes would therefore be of considerable value. Triterpenes, like sterols, are synthesized from the mevalonate pathway via the linear 30-carbon precursor 2,3-oxidosqualene (OS) (Fig. 1*A*) (2–6). The first committed step in sterol biosynthesis is cyclization of OS to cycloartenol by cycloartenol synthase. In triterpenoid biosynthesis, OS is converted to an array of alternative cyclization products by other cyclase enzymes known as triterpene synthases. The resulting scaffolds are often further elaborated by oxidation and glycosylation to triterpene glycosides (also known as saponins) (2–5). Currently, approximately 100 distinct triterpene skeletons made from OS are known from diverse plant species, the most common of which is the pentacyclic triterpene β-amyrin (1–3). Homology modeling in combination with domain swapping and site-directed mutagenesis using yeast as an expression system has yielded insights into triterpene synthase function (e.g., refs. 7–11). However, the mechanisms of triterpene cyclization are still only poorly understood.

Oats (*Avena* species) produce antifungal triterpenes known as avenacins. These β-amyrin–derived compounds are synthesized in the root tips and provide protection against attack by soil-borne pathogens (12, 13). The major avenacin A-1 is esterified with the natural fluorophore *N*-methyl anthranilate (Fig. 1*A*) and is strongly autofluorescent under ultraviolet light (13). We previously exploited this feature to identify avenacin-deficient mutants of diploid oat (*Avena strigosa*) (13). We have now characterized most of the genes and enzymes for this pathway, including the gene for the first committed step encoding β-amyrin synthase (SAD1) (14–20; Fig. 1*A*).

In our initial mutant screen, we identified a total of 10 avenacin-deficient *A. strigosa* mutants, two of which were *sad1* mutants (109 and 610) (13). These two mutants both had single nucleotide mutations resulting in premature termination of translation and mRNA degradation (14), and so did not provide information about amino acid residues important for enzyme function. The root fluorescence screen is sensitive, and we subsequently extended this screen to identify a further 82 avenacin-deficient *A. strigosa* mutants (15). Of these new mutants, 16 accumulated elevated levels of OS and were identified as candidate *sad1* mutants (21).

Here, we analyze this suite of 16 mutants and identify four with predicted amino acid changes that make stable mutant SAD1 protein. Characterization of these mutant SAD1 variants led us to identify two amino acid residues that are critical for SAD1 function. One of these amino acids is a cysteine residue within the active site that is critical for cyclization. Surprisingly, mutation at a different residue (S728F) converted SAD1 into an enzyme that makes tetracyclic (dammarane) instead of pentacyclic products. When expressed in yeast, this mutant SAD1 variant preferentially cyclizes dioxidosqualene (DOS) rather than OS, giving epoxydammaranes. Mutation of the equivalent amino acid residue of AtLUP1, an *Arabidopsis thaliana* triterpene synthase that normally cyclizes OS to pentacyclic products, similarly resulted in the generation of tetracyclic triterpenes as the major cyclization products and preferential generation of DOS-derived epoxydammaranes in yeast. This residue is therefore critical for the generation of pentacyclic rather than tetracyclic products and also appears to be a "substrate specificity switch" in both monocot and dicot triterpene synthases.

Our discoveries provide new insights into triterpene cyclization and reveal hidden functional diversity within the triterpene synthases. They open up opportunities to engineer novel oxygenated

Significance

The triterpenes are a large and highly diverse group of plant natural products. They are synthesized by cyclization of the linear isoprenoid 2,3-oxidosqualene into different triterpene scaffolds by enzymes known as triterpene synthases. This cyclization process is one of the most complex enzymatic reactions known and is only poorly understood. Here, we identify a conserved amino acid residue that is critical for both product and substrate specificity in triterpene synthases from diverse plant species. Our results shed new light on mechanisms of triterpene cyclization in plants and open up the possibility of manipulating both the nature of the precursor and product specificity, findings that can be exploited for the production of diverse and novel triterpenes.

Author contributions: M.S., R.B.T., R. E. Minto, P.E.O., A.M.H., and A.O. designed research; M.S., R.B.T., R. E. Minto, R. E. Melton, R.K.H., and A.M.H. performed research; M.S., R.B.T., and R. E. Minto contributed new reagents/analytic tools; M.S., R.B.T., R. E. Minto, R. E. Melton, A.M.H., and A.O. analyzed data; and M.S., R.B.T., R. E. Minto, A.M.H., and A.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]M.S. and R.B.T. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: anne.osbourn@jic.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1605509113/-/DCSupplemental.

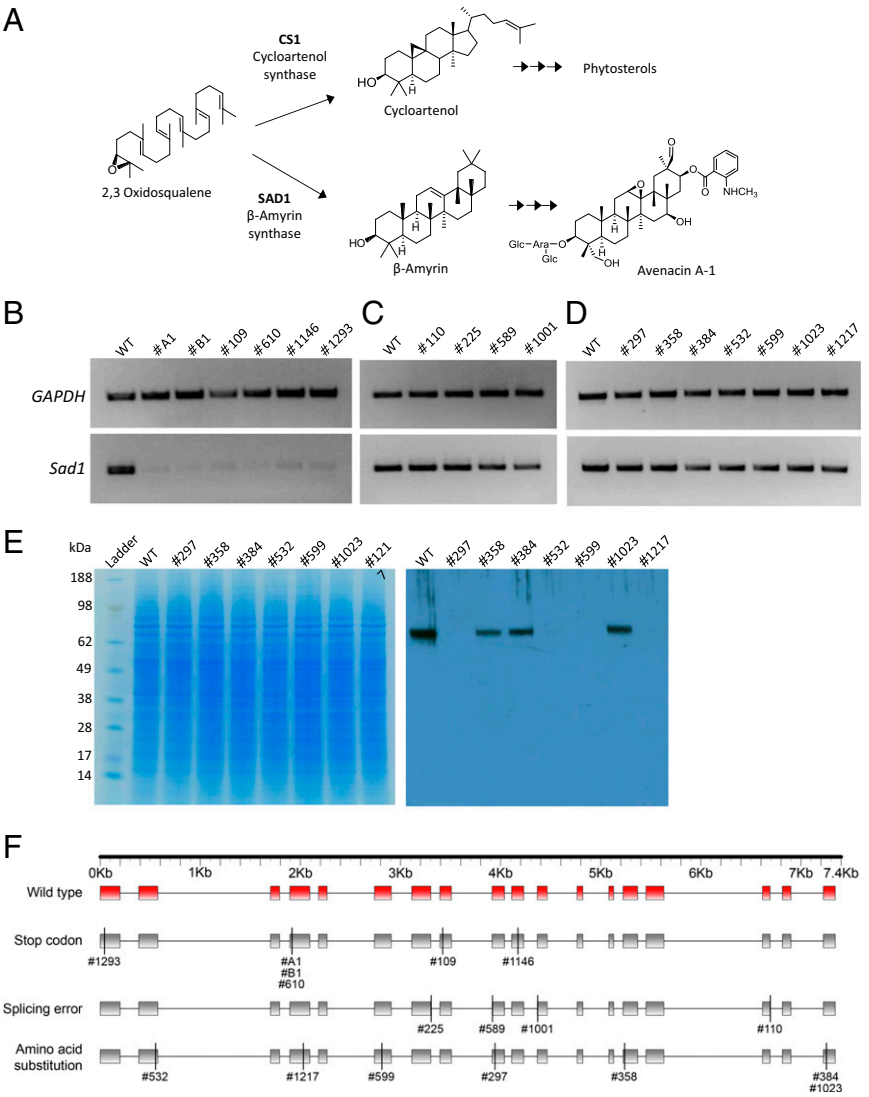


Fig. 1. Characterization of *sad1* mutants. (*A*) Biosynthesis of phytosterols and avenacin A-1 in oat. (*B–D*) RT-PCR analysis of mutant *sad1* transcript levels in mRNA extracted from the roots of wild-type (WT) oats and predicted premature termination of translation (*B*), splicing error (*C*), and amino acid substitution (*D*) mutants (Table 1). The oat glyceraldehyde-3-phosphate dehydrogenase gene (*GAPDH*) was used as a control. (*E*) Analysis of protein extracts from root tips of WT and mutant oat lines. (*Left*) Replicate gel stained with InstantBlue showing protein loading. (*Right*) Western blot analysis of extracts from predicted amino acid substitution mutants probed with antisera raised against SAD1. A single band of ~86 kDa corresponding to full-length SAD1 protein is present in the WT and mutants 358, 384, and 1023. (*F*) Locations of point mutations in *sad1* mutants. A schematic of the WT *Sad1* gene is shown at the top in red. Exons are represented by boxes and introns by lines. The location of each mutation within the gene is indicated by a vertical line.

triterpene scaffolds by manipulation of the precursor supply. Our results further illustrate the power of using a forward genetics approach to identify residues that are critical for the stability and functional diversification of triterpene synthases.

Results

Identification of Mutant SAD1 Protein Variants. The avenacin-deficient *A. strigosa* mutants were generated by using the chemical mutagen sodium azide (13), which causes single-base substitutions, usually from guanine to adenine (22, 23). DNA sequence analysis of the *Sad1* gene in each of the 16 new candidate *sad1* mutants (21) revealed single-point mutations in each case, the majority of which involved guanine-to-adenine transitions as expected. The mutants could be divided into three categories (Table 1)—those with predicted premature termination of translation mutations (as for the two original *sad1* mutants 109 and 610) (14); those with mutations at intron-exon boundaries that may give rise to splicing errors; and those with predicted amino acid substitutions.

We then assessed the *sad1* transcript levels in RNA from the root tips of these mutants by RT-PCR. The four new mutants with predicted premature termination of translation codons (Table 1), like 109 and 610 (14), had substantially reduced transcript levels (Fig. 1*B*), most likely due to nonsense-mediated mRNA decay (24). As expected, Western blot analysis using polyclonal antisera specific for SAD1 (16) failed to detect cross-reacting protein in protein preparations from the roots of these mutants (*SI Appendix*, Fig. S1*A*). Transcripts were clearly detectable in the four predicted splice site mutants (Fig. 1*C* and *SI Appendix*, Fig. S2), but further examination by RT-PCR analysis across the mutation sites revealed exon deletions (*SI Appendix*, Fig. S3). Protein that cross-reacted with the SAD1 antisera was also undetectable in these mutants (*SI Appendix*, Fig. S1*B*). These deletions may result in the formation of misfolded proteins that are targeted for degradation (25).

The transcript levels for the seven mutants with predicted amino acid substitutions were unaltered (Fig. 1*D*). Western blot analysis revealed a protein of the same molecular mass as SAD1 in root

Table 1. Sequence analysis of *sad1* mutants

Mutant	Mutation event	Predicted amino acid change
Premature termination of translation:		
A1	G1912A	Tyr-165 Stop*
B1	G1912A	Tyr-165 Stop*
109	G3417A	Tyr-380 Stop
610	G1912A	Tyr-165 Stop*
1146	G4169A	Tyr-471 Stop
1293	G39A	Tyr-13 Stop
Predicted splicing errors:		
110	G6689A	—
225	G3302A	—
589	G3914A	—
1001	G4365A	—
Predicted amino acid substitutions:		
297	G3939A	Glu-419 Lys
358	G5234A	Cys-563 Tyr
384	C7249T	Ser-728 Phe†
532	G549A	Gly-121 Glu
599	G2809A	Gly-277 Glu
1023	C7249T	Ser-728 Phe†
1217	G2025A	Gly-203 Glu

*Identical mutation ($G^{1912} \rightarrow A$). Although mutants A1, B1, and 610 all have a mutation at G1912, these mutants were isolated from different M2 families and so represent independent mutation events.
†Identical mutation ($C^{7249} \rightarrow T$). Although mutants 384 and 1023 have both undergone a cytidine to thymidine change at C7249, these mutants were isolated from different M2 families and so represent independent mutation events.

extracts from three of these mutants (358, 384, and 1023) (Fig. 1*E*). The mutations in the remaining four mutants are located in regions that are likely to be important for protein structure and presumably lead to unstable proteins that are degraded (*SI Appendix*, Fig. S4). A schematic summarizing the nature and locations of all of the *sad1* mutations is shown in Fig. 1*F*.

Conversion of S728 to F Results in the Formation of Tetracyclic Instead of Pentacyclic Triterpenes in Planta. We next examined the triterpene content of extracts from the root tips of seedlings of *A. strigosa* mutants 358, 384, and 1023. We expected to see loss of the SAD1 cyclization product β-amyrin with associated accumulation of the precursor OS. This result is indeed what we observed for the previously characterized *sad1* mutant 109, a predicted premature termination of a translation mutant that does not produce SAD1 protein; also for mutant 358, suggesting that this mutant SAD1 variant is inactive (Fig. 2*A* and *SI Appendix*, Fig. S5). Surprisingly, however, a new compound was observed in root extracts of mutants 384 and 1023 that was not present in extracts from the wild-type or *sad1* mutants 109 and 358 (Fig. 2*A* and *SI Appendix*, Fig. S5). The new compound had an elution profile and mass spectrum identical to dammaranediol-II (DM) (Fig. 2*A* and *SI Appendix*, Figs. S6 and S7). DM was not detectable in wild-type root extracts by GC-MS, although a more polar minor peak with an elution profile and mass spectrum consistent with that of epoxydammarane (epDM) was observed. Accumulation of DM in mutants of 384 and 1023 implies specificity of the downstream avenacin pathway enzymes for the β-amyrin scaffold.

Ginseng and other medicinal plants accumulate biologically active epoxydammarane saponins at levels as high as 5%, but the biosynthetic origin of the oxacyclic triterpenoid scaffold is not known (26). Stereoisomers of 2-epoxydammarane have been identified as cyclization products generated by the *A. thaliana* mixed product triterpene synthase AtLUP1 in yeast when fed with exogenous DOS (26). Although small peaks with similar elution times to epDM were observed in the mutant extracts, we were unable to detect epDM in these lines by GC-MS. Collectively, these data indicate that the S728F mutation results in a change in product specificity, converting SAD1 into an enzyme that yields primarily tetracyclic rather than pentacyclic cyclization products.

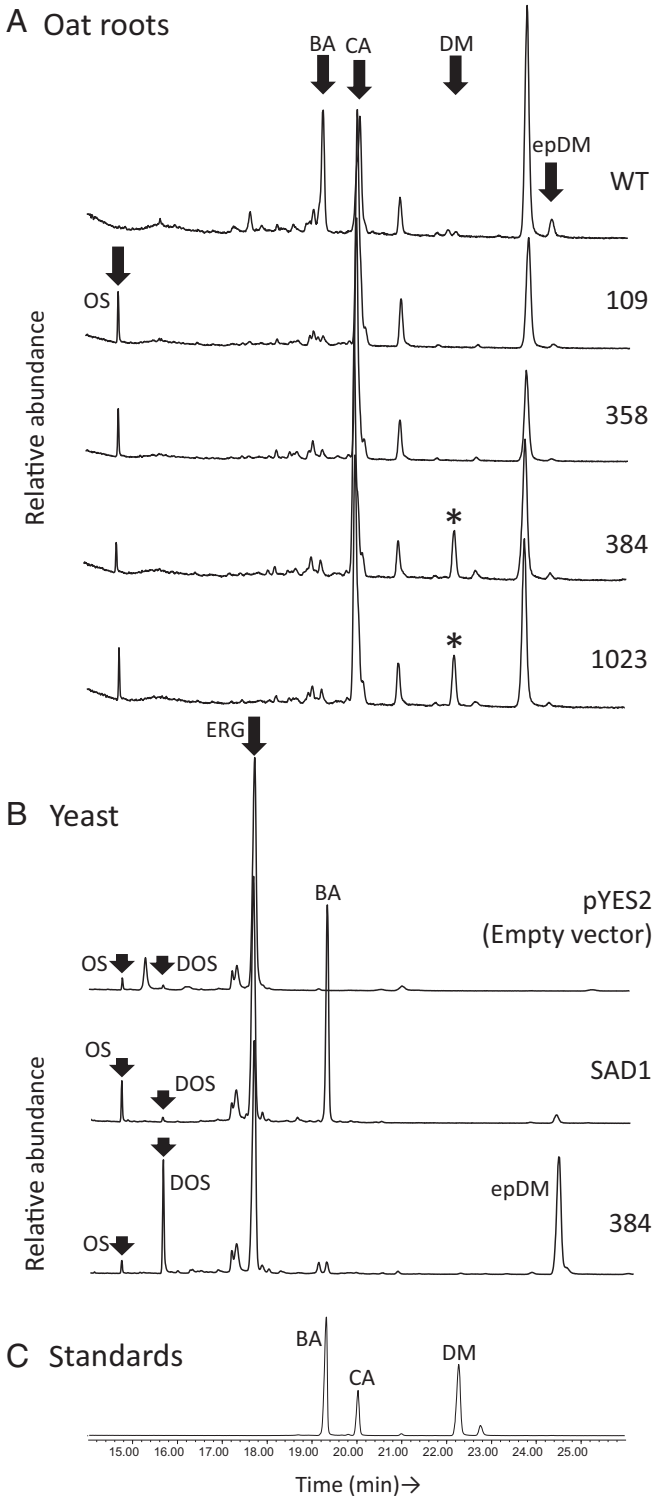


Fig. 2. Effects of mutations on cyclization. Analysis was carried out by GC-MS. Total ion chromatograms (TIC) are shown. (*A*) Analysis of oat root tip extracts from the WT and *sad1* mutants 109, 358, 384, and 1023. (*B*) Analysis of extracts from yeast expressing the WT SAD1 protein or the S728F SAD1 mutant variant. (*C*) β-amyrin, cycloartenol, and dammarenediol-II standards. BA, β-amyrin; CA, cycloartenol; ERG, ergosterol.

Heterologous Expression of the S728F SAD1 Mutant Variant in Yeast. We then expressed the S728F SAD1 variant in yeast. cDNAs encoding the wild-type SAD1 protein and the mutant SAD1 variant were cloned into the yeast expression vector pYES2 under the control of a galactose-inducible promoter, expressed in the yeast strain GIL77 (*gal2 hem3-6 erg7 ura3-167*) (27) (*SI Appendix*, Fig. S8), and yeast extracts were analyzed by GC-MS (Fig. 2*B*). β-Amyrin was the major triterpene product detected when the wild-type SAD1 protein was expressed in yeast. epDM was also detected as a minor product (Fig. 2*B*). The S728F SAD1 mutant variant produced small amounts of β-amyrin by comparison. Unexpectedly, however, this variant generated a major peak that appeared to correspond to epDM, with only trace amounts of DM (Fig. 2*B* and *SI Appendix*, Fig. S9). DOS was also clearly detectable in extracts from yeast expressing the SAD1 mutant variant but not in those from the empty vector control or expressing the wild-type SAD1 protein. Thus, in yeast, the SAD1 mutant enzyme appears to preferentially cyclize DOS rather than OS, yielding predominantly epDM rather than DM (Fig. 3*B*). Accumulation of DOS in yeast expressing the SAD1 mutant variant may be an equilibrium effect due to pull through by DOS cyclization and could be suggestive of metabolome formation.

To confirm the identity of the putative epDM cyclization product, we grew a large-scale (1 L) culture of the yeast strain expressing the S728F SAD1 mutant variant and purified ∼2 mg of this compound (*Methods*). The purified triterpene was examined by ^{1}H-NMR

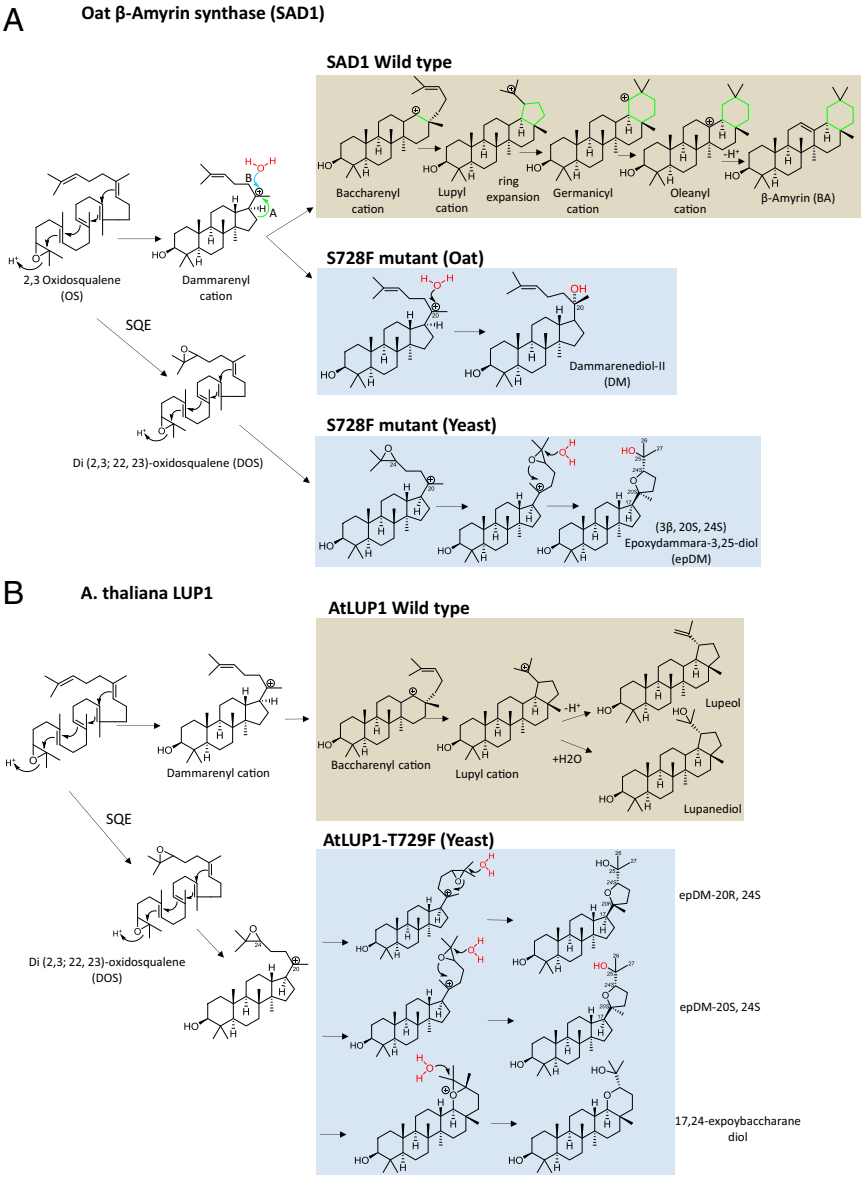
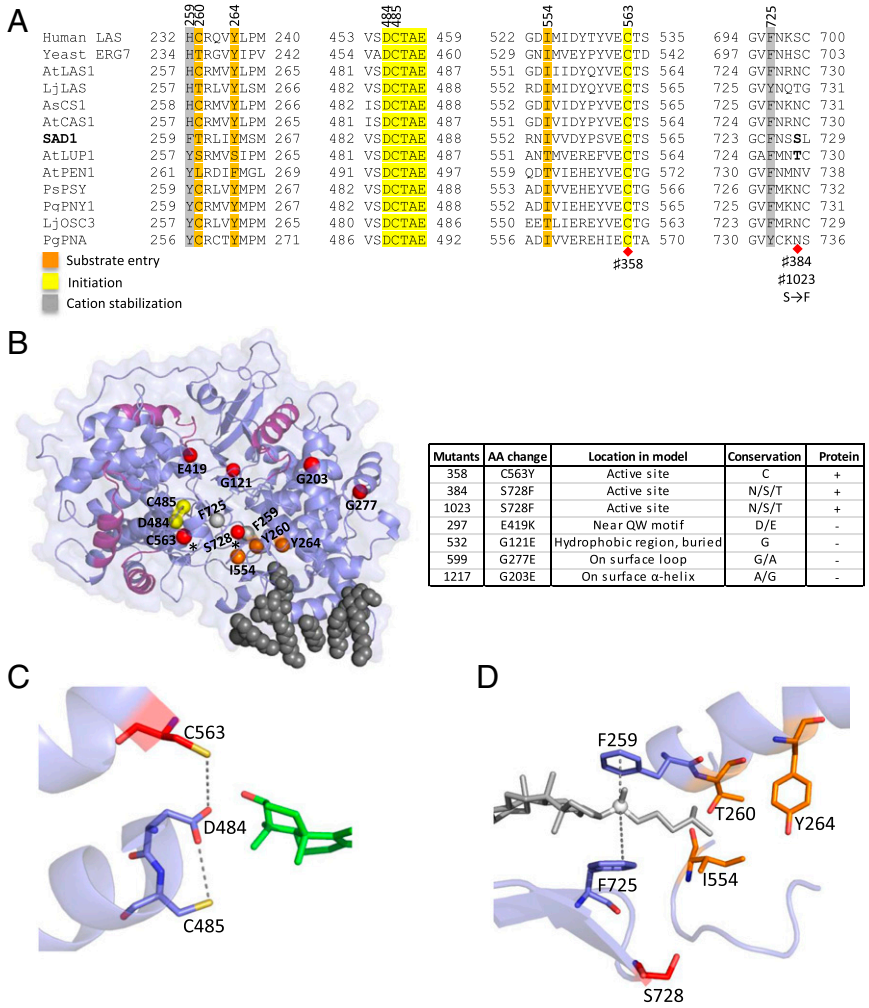


Fig. 3. Cyclization reactions carried out by the WT and mutant triterpene synthases from oat and *A. thaliana*. (*A*) The WT SAD1 protein first converts OS to the tetracyclic dammarenyl cation and then traverses through a series of cations to give the oleanyl cation. The final step is the deprotonation of the oleanyl cation and the release of β-amyrin (BA). In contrast, the S728F SAD1 variant catalyses an alternative cyclization reaction along the path indicated to give tetracyclic products. This mutant cyclase can accept both OS and DOS as substrates, cyclizing them to DM and epDM, respectively. Accumulation of OS can lead to reacceptance of OS as a substrate by endogenous squalene epoxidase, resulting in transformation of OS into DOS. The WT SAD1 enzyme is also able to accept DOS and cyclize it to epDM, but to a much lesser extent than the S728F SAD1 variant (∼6% of total cyclic products in oat and ∼3% of total cyclic products in yeast). (*B*) The cyclization reactions catalyzed by WT ATLUP1 and the ATLUP1-T729F variant are shown. WT AtLUP1 is also able to accept DOS, cyclizing it to 2-epDM epimers as opposed to the single epimer observed for SAD1. For AtLUP1–T729F, cyclization is predominantly DOS-mediated.

spectroscopy at 400 MHz in $CDCl_3$ solution (*SI Appendix*, Figs. S10 and S11). C-24*S* or C-24*R* epimers of the epoxydammaranes can be distinguished by comparing the ^{1}H-NMR chemical shifts of H-24, Me-26, and Me-27 positions. Between C-20 and C-24, four different combinations of configurations are possible. Molecules with 20*R*, 24*R* configuration are not known in nature. Three possible configuration pairs (20*S*, 24*S*; 20*R*, 24*S*; 20*S*, 24*R*) are known to occur in natural products; chemical shifts of assignable resonances are listed in *SI Appendix*, Table S1. Because epDM, which is made together with DM, has 20*S* configuration, it was anticipated to maintain the *S* configuration at C-20. Chemical shifts at the diagnostic protons (H-24, Me-26, Me-27) confirmed epDM has 20*S*, 24*S* configuration. Chemical shifts and coupling constants at H-24 vary dramatically for molecules with locally diastereomeric configurations at C-20 and C-24. For example, epDM with 20*R*, 20*S*, at H-24 δ 3.73 (dd, J = 7.7, 6.9) was similar to the skeleton with 20*S*, 24*R*, at H-24 δ 3.73 (dd, J = 7.5, 7.5). For an arrangement with both stereocenters *S*, H-24 appeared at δ 3.639 (dd, J = 10.1, 5.3), a signal similar to that of the yeast-derived epDM [H-24 δ 3.64 (dd, J = 9.9, 5.4)]. Based on this shift and coupling data, together with other spectral attributes, the epDM product generated by the S728F mutant variant of SAD1 was assigned the configuration 20*S*, 24*S* and designated as (3*S*, 20*S*, 24*S*)-20,24-epoxydammarane-3,25-diol (Fig. 3*A*).

Homology Modeling. To further investigate the likely impact of the amino acid substitutions that we had observed on SAD1 stability and function, we generated a homology model of SAD1 by using the human lanosterol synthase crystal structure (28) as a template. The locations of the seven predicted SAD1 amino acid substitutions are shown in Fig. 4*B*. The C563Y mutation in line 358 results in stable but inactive protein (Figs. 1 and 2). This mutation affects



Mutants	AA change	Location in model	Conservation	Protein
358	C563Y	Active site	C	+
384	S728F	Active site	N/S/T	+
1023	S728F	Active site	N/S/T	+
297	E419K	Near QW motif	D/E	-
532	G121E	Hydrophobic region, buried	G	-
599	G277E	On surface loop	G/A	-
1217	G203E	On surface α-helix	A/G	-

Fig. 4. Effects of mutations on protein structure and function. (*A*) Selected regions of 13 functionally characterized oxidosqualene cyclases from diverse organisms were aligned and annotated according to known protein structure–function relationships. Conserved residues involved in substrate entry, reaction initiation, and cation stabilization are shown in orange, yellow and gray, respectively. Red diamonds indicate the positions of the mutations in *A. strigosa* mutants 358, 384, and 1023. (*B*) Location of the amino acid substitutions mapped onto a homology model of SAD1. Mutations are shown as red spheres; mutations that still yield full-length SAD1 protein (C563Y, S728F) are indicated by asterisks. Amino acid residues involved in initiation of cyclization (yellow), substrate entry (orange), and stabilization of cation intermediates (light gray) are indicated. The QW motifs involved in protein stability are shown as purple helices, and detergent molecules that suggest how the enzyme is orientated in the membrane are shown as dark gray spheres. The table shows the amino acid change in each of the *sad1* mutants, their location in the tertiary structure, and the degree of amino acid conservation at each of these positions across diverse oxidosqualene cyclases. (*C*) *sad1* mutant 358 has a mutation at Cys563 (red), a residue that is hydrogen bonded to the catalytic aspartate (D484). Hydrogen bonds are shown by dashed lines. The substrate is shown in green. (*D*) *sad1* mutants 384 and 1023 both have a mutation at Ser728 (red), which is in close proximity to residues involved in substrate access (orange) and Phe725 (purple), involved in stabilization of the tetracyclic C-20 cationic intermediate. Cation–π interactions are shown by gray dashed lines.

a residue close to the catalytic aspartate D484 of the conserved DCTAE motif (Fig. 4*A*). This conserved aspartate has been implicated in oxidosqualene cyclase function in *Euphorbia tirucalli* β-amyrin synthase (7). C485 (Fig. 4*C*) is known to be required for the function of the oxidosqualene cyclase lanosterol synthase in *Saccharomyces cerevisiae* (29). C563 has been proposed to have a role in initiation of cyclization based on its proximity to D484, following the determination of the human lanosterol synthase crystal structure (28), but its function has not been tested. The C563Y substitution introduces a bulky tyrosine residue that is likely to interfere with the hydrogen bonding interaction critical for lowering the pK_a of D484 to facilitate protonation of the epoxide group (Fig. 3*C*) (30). We predict that this substitution will inhibit reaction initiation, which would inactivate cyclization.

The wild-type SAD1 protein converts OS to the pentacyclic cyclization product β-amyrin through a series of cationic intermediates (Fig. 3*A*). Protein modeling and docking analysis predict that in the S728F variant, the aromatic side chain of the newly introduced phenylalanine together with that of F725 sandwich the carbocationic center of the dammarenyl cation (Fig. 4*D*). As a result, ring expansion of the dammarenyl cation is likely to be compromised, giving rise to formation of a truncated tetracyclic OS cyclization product (Fig. 3*A*). This product is presumably converted to the diol DM by interception of the C-20 cation by a water molecule. The wild-type SAD1 protein is able to generate low levels of epDM in oat and in yeast (Fig. 2 *A* and *B*), suggesting that although OS is its preferred substrate, it is capable of using DOS. In contrast, the S728F mutant SAD1 variant preferentially accepts DOS as a substrate when expressed in yeast and cyclizes this to epDM (Fig. 3*B*). As speculated in Shan et al. (26), side-chain rotation of the distal epoxide to the carbocation at C-20 may enable movement of a positive charge from carbon to oxygen and attack of a nearby water molecule to give an alcohol at C-25. The epoxide-containing side chain of the dammarenyl cation is therefore free to rotate within the active site to participate in cyclization. S728 is located close to a loop containing I554, T260, and Y264. This loop is believed to block the opening to the substrate access channel and undergo a conformational change to allow substrate entry to the active site (Fig. 4*D*). The S728F amino acid substitution may alter the hydrogen-bonding network of this loop and control access of the two substrates (OS and DOS) into the active site. However, further experiments are required to investigate this mechanism in detail. The addition of water in triterpene cyclization is known for the triterpene synthases PgPNA, AtLUP1, and AtPEN1 (ARAB) (31–35). Unlike AtLUP1, which makes two stereoisomers of epDM (20*R*, 24*S* and 20*S*, 24*S*), S728F accumulates only 20*S*, 24*S*-epDM in a highly stereospecific manner. Cyclization of the side chain of DOS occurs in a manner consistent with cation quenching by a positioned water molecule, resulting in a reaction that is more stereospecific in S728F than in the AtLUP1 enzyme.

We next investigated the effects of amino acid substitution at the corresponding position in the *A. thaliana* triterpene synthase AtLUP1 (Thr729; Fig. 4*A* and *SI Appendix*, Fig. S12). Wild-type AtLUP1 and a mutated version in which Thr729 had been converted to phenylalanine (T729F) were expressed in yeast. Wild-type AtLUP1 produced the OS-derived pentacyclic triterpenes lupeol and lupanediol as the major cyclization products, and also the epoxydammarendiols epDM (20R, 24S) and 17,24-expoybaccharane diol from DOS as minor products (Figs. 3*B* and 5). Interestingly, the AtLUP1-T729F mutant yielded primarily epoxydammarendiol products (Figs. 3*B* and 5 and *SI Appendix*, Figs. S13 and S14). Thus, the T729F mutation in AtLUP1 leads to a change in product specificity from pentacyclic to tetracyclic triterpenes, and also to preferential cyclization of DOS instead of OS in yeast as seen for the SAD1 S728F mutant variant. As seen for SAD1 S728F, yeast strains expressing the mutant form of AtLUP1 also accumulated elevated levels of DOS (*SI Appendix*, Fig. S15). The S728F SAD1 mutant enzyme shows high stereoselective cyclization of DOS to the epDM-20S, 24S epimer, whereas AtLUP1 is less selective, generating two epimers and other additional minor products. Elucidation of the crystal structures of the wild-type and mutant forms of these enzymes may in the future enable the exact nature of this control to be understood in more detail.

Discussion

In tritepenoid biosynthesis, cyclization of OS is the key step that determines the nature of the triterpene scaffold. The enzymes that catalyze this process—triterpene synthases—belong to multigene families in plant genomes. Here, we have shown that a single-point mutation that causes an amino acid substitution close to the active site dramatically alters product specificity in both SAD1 and LUP1, so uncovering hidden functional diversity in these triterpene synthases. It is conceivable that nature explores alternate modes of cyclization through such single mutational steps, as has been suggested for diterpene synthases (36–43). A dedicated cyclase that makes epDM as its major product has not been reported before our work to our knowledge. Triterpene glycosides based on the DM and epDM skeletons have important pharmaceutical and antibacterial properties (44–46). Our results open up the possibility of manipulating both the nature of the precursor and the product specificity of the cyclization process for the production of diverse and novel triterpenes. They further demonstrate the power of forward genetics screens in plants for elucidating the enzymatic mechanisms of this versatile and fascinating group of enzymes.

Methods

Plant Material. Wild-type and mutant *A. strigosa* lines were grown as described (13). Transcript, protein, and triterpene analysis were carried out by using root tips (terminal 0.5 cm) from 3-d-old seedlings, and genomic DNA was extracted from 6-d-old seedlings.

DNA, RNA, and Protein Analysis. Genomic DNA was isolated from 6-d-old seedlings of *A. strigosa* by using the DNeasy Plant Mini kit (Qiagen). The wild-type and mutant forms of the *Sad1* gene were amplified in four segments. Purified DNA segments from each of the *sad1* mutants were sequenced by using sets of primers along the length of the *Sad1* gene, and each mutant was sequenced to at least twofold coverage. Primers used for *Sad1* amplification and sequencing are shown in *SI Appendix*, Table S2. Genomic DNA samples from each mutant were sent for Diversity Array Technology (DArT) analysis (Diversity Arrays Technology Pty Ltd) to confirm that each mutant line was independent (47).

For analysis of transcripts in oats, total RNA was extracted from 0.5-cm root tips of 3-d-old oat seedlings by using TRI-REAGENT (Sigma catalog no. T9424) and the extract was treated with DNase I (Roche). For RT-PCR, cDNA was synthesized by using 1 μg of DNase-treated RNA. First-strand cDNA synthesis was carried out by using SuperScript II Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions, and cDNA was amplified by PCR. For Northern blot analysis, 10 μg of total RNA was used. RNA was separated on a 1.2% (wt/vol) agarose/0.25 M formaldehyde gel and transferred to a Hybond-N^+ nylon membrane (Amersham) overnight. cDNA probes were labeled with ^{32}P-dCTP by using the Rediprime II Random Prime Labeling Kit (Amersham). Hybridizations were carried out overnight at 65 °C in 10 mL of Church Buffer containing 0.1 mg/mL salmon sperm DNA (Sigma) and 50 μL of ^{32}P-dCTP labeled probe. The membrane was exposed to a BAS-IIIS imaging plate (Fuji) overnight and imaged by using a Typhoon 9200 Variable Mode Imager (Amersham).

For protein and immunoblot analysis of oat root protein, total protein was extracted from 0.5-cm root tips of 3-d-old oat seedlings. Root tips were ground in protein extraction buffer [50 mM Tris·HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 10% (vol/vol) glycerol, 1% (wt/vol) PVPP, 1% (vol/vol) Triton X-100 (Boehringer Mannheim), 1× Complete protease inhibitor (Roche)] for 1 min with a plastic pestle followed by incubation at 4 °C for 2 h. Proteins were denatured, separated on NuPAGE gels (4–12% acrylamide gradient) (Invitrogen), and blotted onto nitrocellulose membranes (Bio-Rad) by using the manufacturer's protocol. Membranes were probed with anti-SAD1 antisera (1:10,000 dilution)

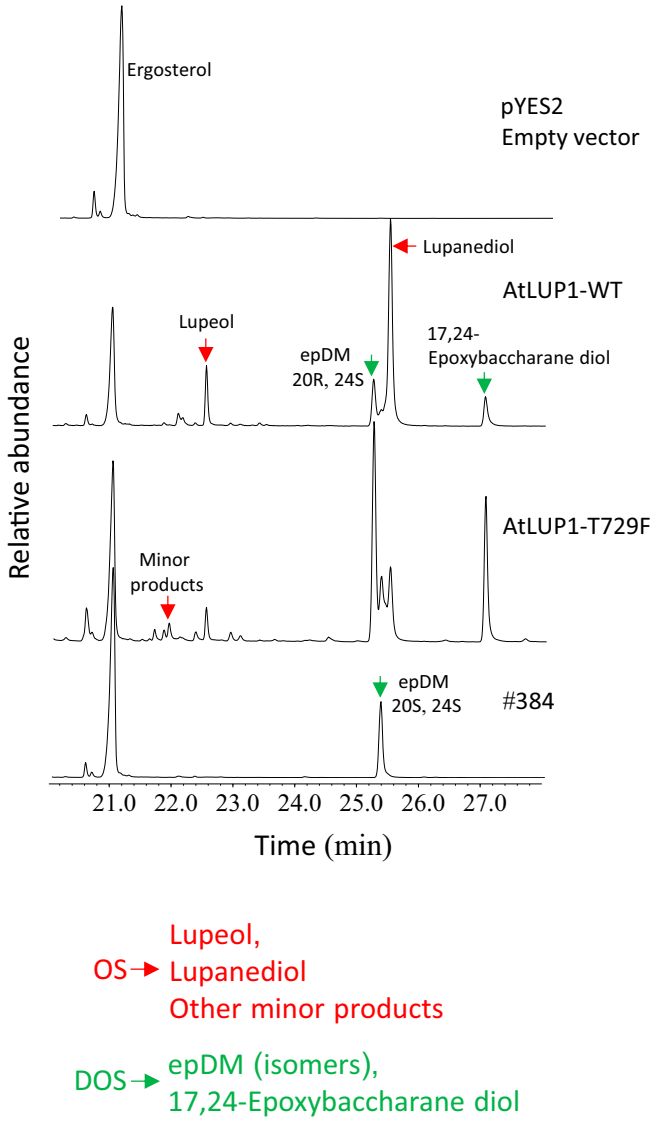


Fig. 5. Total ion chromatograms of yeast extracts expressing the WT and mutant triterpene synthases. In AtLUP1, lupeol and lupanediol are derived from OS cyclization, whereas epDM isomers (20S, 24R and 20S, 24S) and 17, 24 epoxybaccharane diol are derived from DOS cyclization. For the SAD1 mutant variant S728F (384), the epDM isomer-20S, 24S is derived from DOS. Peaks with red arrows indicate OS-derived cyclization products, and ones with green indicate DOS-derived cyclization products.

(16) followed by detection with a goat anti-rat IgG horseradish peroxidase-labeled secondary antibody (Sigma-Aldrich) according to the manufacturer's protocol.

Extraction and Analysis of Triterpenes from Oat Roots. The triterpene content of oat root tips was analyzed by TLC and GC-MS. Root tips (~50 per line) were ground, mixed with 0.5 mL of saponification reagent [20% (wt/vol) KOH in 50% (vol/vol) ethanol], and incubated at 65 °C for 2 h before extraction with an equal volume of hexane. The extraction step was repeated twice more to maximize triterpene recovery. The extract was then dried down, and the residue dissolved in 500 μL of hexane. For rapid qualitative analysis, extracts were run on TLC plates (Silica gel on Al foil, 10 cm × 5 cm, FLU.K.A, catalog no. 70644) by using a hexane:ethyl acetate (6:1) solvent system. Compounds were visualized by spraying the plates with acetic acid: H_2SO_4: p-anisaldehyde (48:1:1 vol/vol) and heating to 120 °C for 5 min on a TLC plate heater. For GC-MS analysis, 100-μL aliquots of hexane extract were dried down and the residues were resuspended in 100 μL of Tri-Sil Z reagent (Sigma, catalog no. 92718) before incubating at 65 °C for 30 min in a dry heat bath. Samples were run on either a HP-5MS column (30 m × 0.25 mm i.d., 0.25-μm film) (Agilent) or a ZB-5HT column (35 m × 0.25 mm i.d., 0.10-μm film) (Phenomenex) by using an Agilent 7890B GC machine. The injector port, source, and transfer line temperatures were set at 250 °C; an oven temperature program from 80 °C (2 min) to 290 °C (30 min) at 20 °C/min was used. The carrier gas was helium; the flow rate was 1.2 mL/min. Samples were injected in splitless mode with either a 1-μL or a 3-μL sample volume. The output was used to search the NISTv8 library to assign identity to peaks in the GC-MS traces. Product abundance was calculated as the percentage of total cyclic products using integrated peak areas. All experiments were repeated to confirm reproducibility of the triterpene profiles of the wild-type and mutant samples.

Triterpene Standards. Dammarenediol-II (catalog no. CFN99476, 98% HPLC pure) was purchased from Wuhan ChemFaces Biochemical Co. Ltd, China, and β-Amyrin and cycloartenol from Extrasynthese. The standards were dissolved and diluted to 0.5 mg/mL in hexane before derivatization and GC-MS.

AtLUP1 Cloning and Site-Directed Mutagenesis. AtLUP1 (AT1G78970)-ORF was amplified from total cDNA of young *A. thaliana* (Col-0) seedlings by PCR using Gateway primers (*SI Appendix*, Table S2) and cloned into pDONOR207. The T729F mutation was created by site-directed mutagenesis using pDONOR207:AtLUP1 plasmid as the template. The oligonucleotide design strategy and conditions for PCR amplification followed those described (48). The oligonucleotides used for site-directed mutagenesis are listed in *SI Appendix*, Table S2.

Yeast Cloning and Expression. All cloning and expression analysis was carried out in the yeast strain GIL77 (*gal2 hem3-6 erg7 ura3-167*) (32). Expression vectors were constructed by using in vivo homologous recombination in yeast. The ORFs of the wild-type *Sad1* gene, *AtLUP1* (AT1G78970), and mutant variants were amplified from pDONOR207 entry vectors by using the oligonucleotides for yeast cloning shown in *SI Appendix*, Table S2. Each primer contained a region that overlapped with the pYES2 vector sequences (the 5′ end of the forward primer overlapped with the GAL1 promoter sequence, and the 5′ of the reverse primer with the CYC1 terminator sequence). The 3′ ends of the primers matched the beginning and end of the *Sad1* ORF. The ORFs of the wild-type and mutant lines were amplified by using these primers, and the PCR fragments obtained were cotransformed into GIL77 strain along with XbaI/HindIII-linearized pYES2 vector. Yeast transformation was performed by using standard protocols (Yeastmaker Yeast transformation system 2, Clontech Laboratories). This resulted in vivo recombination between the pYES2 vector and the *Sad1* ORFs. Plasmids were recovered from yeast, transformed into *E. coli*, and checked by sequencing.

For expression analysis, yeast strains were grown at 28 °C in 5-mL cultures in selective medium [SD-URA+ 2% (wt/vol) glucose + supplements] until saturation (~2 d). The supplements used were as follows: ergosterol (Fluka), 20 μg/mL; hemin (Sigma-Aldrich), 13 μg/mL; and Tween-80 (Sigma-Aldrich), 5 mg/mL. Cells were then pelleted, washed in ddH_2O, transferred to induction medium [SD-URA + 2% (wt/vol) galactose], and incubated for a further 2 d to allow accumulation of triterpenes. They were then pelleted and washed once with ddH_2O before triterpene extraction as described for oat roots.

For protein analysis, yeast cells were resuspended in protein extraction buffer [50 mM Tris·HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 10% (vol/vol) glycerol, 1% (wt/vol) PVPP, 1% (vol/vol) Triton X-100 (Boehringer Mannheim), 1× Complete protease inhibitor (Roche)] and lysed by using a French press (with two passes at 1125 p.s.i., 4 °C). The preparations were then incubated on ice for 2 h and then centrifuged at 21,130 × *g* for 20 min. The supernatants were used for protein and Western blot analysis as described previously (16).

Purification and Structural Elucidation of (3S, 20S, 24S)-20,24-epoxydammarane-3, 25-diol. epDM was extracted from a 1-L culture of a yeast transformant expressing the SAD1 S728F variant. Cells were pelleted, round, and then extracted by using the methods described above for oat roots. The organic residue was loaded onto a silica gel column 10 cm long and 0.5 cm in diameter (LC60A35-70 μm; Fluorochem) in a Pasteur pipette that had been preequilibrated with an ethyl acetate:hexane (1:9) solvent system. The column was washed with 5–6 column volumes of 1:9 ethyl acetate:hexane to remove oxidosqualene, dioxidosqualene, and other nonpolar yeast components. Next, the solvent was switched in a step gradient to 1:6 and then 1:4 ethyl acetate:hexane, and 0.5-mL fractions were collected and analyzed by TLC. Fractions containing epDM were combined and dried in a rotary evaporator. The purity of the compound was assessed by using

GC-MS. Through this process, we obtained ~2 mg of the compound. To assign configuration to epDM, we recorded ^{1}H-NMR of epDM in $CDCl_3$ at 400 MHz (Bruker Avance III).

Homology Modeling and Sequence Alignments. For homology modeling of SAD1, human lanosterol synthase was used as a template (PDB ID code; 1W6K) to generate a model using Modeler (49). The models obtained were subjected to stereochemical validation by using Prosa II (50), Prove (51), and Procheck (52). Models were visualized by using PyMOL (53). Protein sequences were aligned by using Clustal W, and sequence features were viewed and annotated manually using functional information available for human lanosterol synthase (28).

The orientation and position of SAD1 relative to a virtual membrane were predicted by using the PPM server (54). This approach allows the calculation of the rotational and translational positions of transmembrane and peripheral proteins in membranes using their 3D structure as input. Hydrophobicity was calculated by using the TopPred II server (55).

ACKNOWLEDGMENTS. This work was supported by European Union Grant KBBE-2013-7 (TriForC), the Biotechnology and Biological Sciences Research Council Institute Strategic Programme Grant Understanding and Exploiting Plant and Microbial Metabolism BB/J004561/1, the John Innes Foundation (A.O., R. E. Melton, R.K.H., and P.E.O.), and a Norwich Research Park studentship award (to M.S.). R. E. Minto is grateful for sabbatical leave provided by Indiana University–Purdue University, Indianapolis.

1. Xu R, Fazio GC, Matsuda SPT (2004) On the origins of triterpenoid skeletal diversity. *Phytochemistry* 65(3):261–291.
2. Osbourn A, Goss RJM, Field RA (2011) The saponins: Polar isoprenoids with important and diverse biological activities. *Nat Prod Rep* 28(7):1261–1268.
3. Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A (2014) Triterpene biosynthesis in plants. *Annu Rev Plant Biol* 65:225–257.
4. Moses T, Papadopoulou KK, Osbourn A (2014) Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit Rev Biochem Mol Biol* 49(6):439–462.
5. Augustin JM, Kuzina V, Andersen SB, Bak S (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* 72(6):435–457.
6. Chappell J (2002) The genetics and molecular genetics of terpene and sterol origami. *Curr Opin Plant Biol* 5(2):151–157.
7. Ito R, Masukawa Y, Hoshino T (2013) Purification, kinetics, inhibitors and CD for recombinant β-amyrin synthase from *Euphorbia tirucalli* L and functional analysis of the DCTA motif, which is highly conserved among oxidosqualene cyclases. *FEBS J* 280(5):1267–1280.
8. Segura MJR, Jackson BE, Matsuda SPT (2003) Mutagenesis approaches to deduce structure-function relationships in terpene synthases. *Nat Prod Rep* 20(3):304–317.
9. Kushiro T, Shibuya M, Masuda K, Ebizuka Y (2000) Mutational studies on triterpene syntheses: Engineering lupeol synthase into β-amyrin synthase. *J Am Chem Soc* 122(29):6816–6824.
10. Chang CH, et al. (2013) Protein engineering of oxidosqualene-lanosterol cyclase into triterpene monocyclase. *Org Biomol Chem* 11(25):4214–4219.
11. Racolta S, Juhl PB, Sirim D, Pleiss J (2012) The triterpene cyclase protein family: A systematic analysis. *Proteins* 80(8):2009–2019.
12. Turner EM (1960) The nature of resistance of oats to the take-all fungus. III. Distribution of the inhibitor in oat seedlings. *J Exp Bot* 11:403–412.
13. Papadopoulou K, Melton RE, Leggett M, Daniels MJ, Osbourn AE (1999) Compromised disease resistance in saponin-deficient plants. *Proc Natl Acad Sci USA* 96(22):12923–12928.
14. Haralampidis K, et al. (2001) A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proc Natl Acad Sci USA* 98(23):13431–13436.
15. Qi X, et al. (2006) A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. *Proc Natl Acad Sci USA* 103(49):18848–18853.
16. Geisler K, et al. (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc Natl Acad Sci USA* 110(35):E3360–E3367.
17. Mugford ST, et al. (2009) A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell* 21(8):2473–2484.
18. Mugford ST, et al. (2013) Modularity of plant metabolic gene clusters: A trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell* 25(3):1078–1092.
19. Owatworakit A, et al. (2013) Glycosyltransferases from oat (*Avena*) implicated in the acylation of avenacins. *J Biol Chem* 288(6):3696–3704.
20. Qi X, et al. (2004) A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. *Proc Natl Acad Sci USA* 101(21):8233–8238.
21. Qin B, et al. (2010) High throughput screening of mutants of oat that are defective in triterpene synthesis. *Phytochemistry* 71(11-12):1245–1252.
22. Rines HW (1985) Sodium-azide mutagenesis in diploid and hexaploid oats and comparison with ethyl methanesulfonate treatments. *Environ Exp Bot* 25(1):7–16.
23. Al-Qurainy F, Khan S (2009) Mutagenic effects of sodium azide and its application in crop improvement. *World Appl Sci J* 6(12):1589–1601.
24. Chiba Y, Green PJ (2009) mRNA degradation machinery in plants. *J Plant Biol* 52(2): 114–124.
25. Ellgaard L, Helenius A (2003) Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* 4(3):181–191.
26. Shan H, Segura MJR, Wilson WK, Lodeiro S, Matsuda SPT (2005) Enzymatic cyclization of dioxidosqualene to heterocyclic triterpenes. *J Am Chem Soc* 127(51):18008–18009.
27. Kushiro T, Shibuya M, Ebizuka Y (1998) β-amyrin synthase–cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. *Eur J Biochem* 256(1):238–244.
28. Thoma R, et al. (2004) Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* 432(7013):118–122.
29. Oliaro-Bosso S, Schulz-Gasch T, Balliano G, Viola F (2005) Access of the substrate to the active site of yeast oxidosqualene cyclase: An inhibition and site-directed mutagenesis approach. *ChemBioChem* 6(12):2221–2228.
30. Gandour RD (1981) On the importance of orientation in general base catalysis by carboxylate. *Bioorg Chem* 10(2):169–176.
31. Tansakul P, Shibuya M, Kushiro T, Ebizuka Y (2006) Dammarenediol-II synthase, the first dedicated enzyme for ginsenoside biosynthesis, in *Panax ginseng*. *FEBS Lett* 580(22):5143–5149.
32. Kushiro T, et al. (2006) Stereochemical course in water addition during LUP1-catalyzed triterpene cyclization. *Org Lett* 8(24):5589–5592.
33. Segura MJR, Meyer MM, Matsuda SPT (2000) *Arabidopsis thaliana* LUP1 converts oxidosqualene to multiple triterpene alcohols and a triterpene diol. *Org Lett* 2(15): 2257–2259.
34. Xiang T, et al. (2006) A new triterpene synthase from *Arabidopsis thaliana* produces a tricyclic triterpene with two hydroxyl groups. *Org Lett* 8(13):2835–2838.
35. Kolesnikova MD, et al. (2007) Stereochemistry of water addition in triterpene synthesis: The structure of arabidiol. *Org Lett* 9(11):2183–2186.
36. Keeling CI, Weisshaar S, Lin RPC, Bohlmann J (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc Natl Acad Sci USA* 105(3): 1085–1090.
37. Criswell J, Potter K, Shephard F, Beale MH, Peters RJ (2012) A single residue change leads to a hydroxylated product from the class II diterpene cyclization catalyzed by abietadiene synthase. *Org Lett* 14(23):5828–5831.
38. Zerbe P, Chiang A, Bohlmann J (2012) Mutational analysis of white spruce (*Picea glauca*) ent-kaurene synthase (PgKS) reveals common and distinct mechanisms of conifer diterpene synthases of general and specialized metabolism. *Phytochemistry* 74:30–39.
39. Potter K, Criswell J, Zi J, Stubbs A, Peters RJ (2014) Novel product chemistry from mechanistic analysis of ent-copalyl diphosphate synthases from plant hormone biosynthesis. *Angew Chem Int Ed Engl* 53(28):7198–7202.
40. Irmisch S, et al. (2015) One amino acid makes the difference: The formation of ent-kaurene and 16α-hydroxy-ent-kaurane by diterpene synthases in poplar. *BMC Plant Biol* 15:262.
41. Mafu S, et al. (2015) Efficient heterocyclisation by (di)terpene synthases. *Chem Commun (Camb)* 51(70):13485–13487.
42. Potter KC, Jia M, Hong YJ, Tantillo D, Peters RJ (2016) Product rearrangement from altering a single residue in the rice syn-copalyl diphosphate synthase. *Org Lett* 18(5): 1060–1063.
43. Potter KC, et al. (2016) Blocking deprotonation with retention of aromaticity in plant ent-copalyl diphosphate synthase leads to product rearrangement. *Angew Chem Int Ed Engl* 55(2):634–638.
44. Wu G, et al. (2013) Pseudoginsenoside F11, a novel partial PPAR γ agonist, promotes adiponectin oligomerization and secretion in 3T3-L1 adipocytes. *PPAR Res* 2013: 701017.
45. Wang X, et al. (2014) Pseudoginsenoside-F11 (PF11) exerts anti-neuroinflammatory effects on LPS-activated microglial cells by inhibiting TLR4-mediated TAK1/IKK/NF-κB, MAPKs and Akt signaling pathways. *Neuropharmacology* 79:642–656.
46. Zhou Z, et al. (2013) Synthesis and biological evaluation of novel ocotillol-type triterpenoid derivatives as antibacterial agents. *Eur J Med Chem* 68:444–453.
47. Wenzl P, et al. (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci USA* 101(26):9915–9920.
48. Erijman A, Dantes A, Bernheim R, Shifman JM, Peleg Y (2011) Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *J Struct Biol* 175(2):171–177.
49. Eswar N, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Prot Bioinformatics* 5:5.6.
50. Wiederstein M, Sippl MJ (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35(Web Server issue):W407-10.
51. Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol* 264(1):121–136.
52. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* 26(2):283–291.
53. The PyMOL Molecular graphics system (Schrödinger, LLC), Version 1.7.4.
54. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res* 40(Database issue):D370–D376.
55. Claros MG, von Heijne G (1994) TopPred II: An improved software for membrane protein structure predictions. *Comput Appl Biosci* 10(6):685–686.