Published in final edited form as:

Nature. 2011 January 6; 469(7328): 116–120. doi:10.1038/nature09628.

Taxadiene Synthase Structure and Evolution of Modular Architecture in Terpene Biosynthesis

Mustafa Köksal[1], **Yinghua Jin**[2,3], **Robert M. Coates**[2], **Rodney Croteau**[4], and **David W. Christianson**[1]

[1]Roy and Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, PA 19104-6323 USA

[2]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[3]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309

[4]Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340

Abstract

With more than 55,000 members identified to date in all forms of life, the family of terpene or terpenoid natural products represents the epitome of molecular biodiversity. A particularly eminent member of this family is the polycyclic diterpenoid Taxol (paclitaxel), which promotes tubulin polymerization1 and exhibits remarkable efficacy in cancer chemotherapy2. The first committed step of Taxol biosynthesis in the Pacific yew (*Taxus brevifolia*)3 is the cyclization of the linear isoprenoid substrate geranylgeranyl diphosphate (GGPP) to form taxa-4(5), 11(12)diene4, which is catalyzed by taxadiene synthase5. The full-length form of this diterpene cyclase contains 862-residues, but an ~80-residue N-terminal transit sequence is cleaved upon maturation in plastids6. We now report the X-ray crystal structure of a truncation variant lacking the transit sequence and an additional 27 residues at the N-terminus, henceforth designated TXS. Specifically, we have determined structures of TXS complexed with 13-aza-13,14-dihydrocopalyl diphosphate (ACP, 1.82 Å resolution) and 2-fluorogeranylgeranyl diphosphate (FGP, 2.25 Å resolution). The TXS structure is the first of a diterpene cyclase and reveals a modular assembly of three α-helical domains. The C-terminal catalytic domain is a class I terpenoid cyclase, which binds and activates substrate GGPP with a three-metal ion cluster. Surprisingly, the N-terminal domain and a third "insertion" domain together adopt the fold of a vestigial class II terpenoid cyclase. A class II cyclase activates the isoprenoid substrate by protonation instead of ionization,

Correspondence and requests for materials should be addressed to D.W.C. (215-898-5714, 215-573-2201 (fax), chris@sas.upenn.edu).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions M.K. and D.W.C. performed the X-ray crystallographic studies. R.C. supplied the M79-TXS construct from which M107-TXS-CHT was ultimately prepared. Y.J. and R.M.C. synthesized 2-fluorogeranylgeranyl diphosphate. All authors contributed to the interpretation of the results and preparation of the manuscript.

Author Information The atomic coordinates and structure factors of the TXS-ACP and TXS-FGP complexes have been deposited in the Protein Data Bank with accession codes 3P5P and 3P5R, respectively.

The authors declare no competing financial interests.

and the TXS structure reveals a definitive connection between the two distinct cyclase classes in the evolution of terpenoid biosynthesis.

Although the first structures of C_{10} monoterpene7, C_{15} sesquiterpene8,9, and C_{30} triterpene10 cyclases appeared several years ago, the structure of the "missing link" in this series – a C_{20} diterpene cyclase – has remained elusive until now. Plant diterpene cyclases such as taxadiene synthase are perhaps the most intriguing because they are the largest terpenoid cyclases (800–900 residues) and they are believed to be the most closely related to the ancestral plant terpenoid synthase11,12. Triple-domain plant diterpene synthases are believed to have evolved through the fusion of single-domain and double-domain bacterial diterpene cyclases, which in turn evolved from ancient progenitors13.

The two distinct classes of terpenoid cyclases exhibit unrelated protein folds and employ different substrate activation mechanisms14–17. A class I terpenoid cyclase utilizes a trinuclear metal cluster liganded by conserved motifs **D**DXX**D** and (**N**,**D**)DXX(**S**,**T**)XXX**E** (boldface indicates typical metal ligands) to trigger the ionization of the isoprenoid substrate diphosphate group, which generates a carbocation to initiate catalysis. A class II terpenoid cyclase initiates carbocation formation by general acid catalysis, utilizing the "middle" aspartic acid in a DXDD motif to protonate an isoprenoid double bond or oxirane moiety. Taxadiene synthase lacks a DXDD motif, but contains conserved metal-binding motifs and requires Mg^{2+} for optimal catalytic activity5, indicating that it functions as a class I terpenoid cyclase.

Expression and analysis of N-terminal truncation variants of taxadiene synthase revealed that deletions of 60 or 79 residues yield catalytically active proteins, whereas deletions of 93, 113, or 126 residues yield catalytically inactive proteins18. These results implicate the N-terminal segment D^{80}DIPRLSANYHGDL^{93} in catalysis. The N-terminal truncation variant lacking 60 residues has been studied with deuterated18–21 and fluorinated22 analogues of GGPP. These studies suggest a cyclization mechanism (Figure 1) in which the diphosphate leaving group, the 14,15 π bond, and the 10,11 π bond of GGPP are optimally aligned for leaving group departure with formation of a verticillen-12-yl carbocation intermediate in the first step(s) of catalysis. Conformational inversion followed by 11α, 7α-proton transfer and transannular B/C ring closure subsequently generates the taxen-4-yl carbocation, deprotonation of which yields taxa-4(5),11(12)-diene. The intramolecular proton transfer required to initiate transannular B/C ring closure occurs without the assistance of an enzyme-bound base20. The base mediating the final deprotonation step has not yet been identified.

The successful crystallization of TXS required cocrystallization with Mg^{2+} and either ACP or FGP (molecular structures are shown in Figure S1). While not active, this truncation variant is exceptionally stable and is the only form examined that yielded satisfactory crystals. Surprisingly, TXS contains three α-helical domains and harbors the folds of *both* class I and class II terpenoid cyclases (Figure 2); this structure is representative of nearly all diterpene cyclases. The C-terminal domain (S553-V862) exhibits the class I terpenoid synthase fold first observed in farnesyl diphosphate synthase23 and subsequently observed and designated the class I terpenoid synthase fold8,14 in monoterpene and sesquiterpene

cyclases7–9. This fold is also observed in geranylgeranyl diphosphate synthase24, which generates the substrate for diterpene cyclases. Surprisingly, the N-terminal domain of TXS (M107-I135 and S349-Q552) together with the "insertion" domain25 (S136-Y348) comprise the double α-barrel class II terpenoid synthase fold first observed in the triterpene cyclase squalene-hopene cyclase10 and later observed in oxidosqualene cyclase26. TXS shares no significant overall amino acid sequence identity with these triterpene cyclases.

Comparison of TXS with other terpenoid cyclases reveals that cyclase architecture is modular in nature and can consist of one, two, or three domains (Figure 2). Bacterial and fungal sesquiterpene cyclases are single-domain enzymes that adopt the class I terpenoid synthase fold; the first such enzymes to yield crystal structures were pentalenene synthase8 and trichodiene synthase27, respectively. Plant monoterpene and sesquiterpene cyclases generally contain two domains: the C-terminal domain adopts the class I terpenoid synthase fold, and the N-terminal domain adopts an unrelated α-helical fold that, as first noted by Wendt and Schulz14, is homologous to the N-terminal domain of the class II triterpene cyclase, squalene-hopene cyclase10. The first plant monoterpene and sesquiterpene synthases to yield crystal structures were bornyl diphosphate synthase7 and 5-epi-aristolochene synthase9, respectively. Most plant diterpene synthases contain three domains, with the third domain being an insertion conserved in sequence and position25. Notably, Cao and colleagues correctly predicted that this domain is homologous to the insertion domain of a triterpene cyclase based on bioinformatics analysis13.

It is interesting to note that the class II triterpene cyclases squalene-hopene cyclase10 and oxidosqualene cyclase26 are monotopic membrane proteins: each penetrates, but does not completely pass through, the membranes in which they are localized. Their triterpene substrates (squalene or squalene oxide, respectively) are solubilized in the membrane and enter the active site cavity through a hydrophobic channel open to the membrane surface. A nonpolar "plateau" flanks the entrance to this channel near helix 8 in their respective insertion domains; helix 8 is quite hydrophobic in nature and likely serves as the membrane anchor (Figure 2). In contrast, TXS functions in the plastid lumen, so its insertion domain does not contain the corresponding hydrophobic components.

The active site of TXS is located in the C-terminal domain and is the exclusive binding site of the substrate analogue FGP (Figures 3a, S3a) and the bicyclic isoprenoid ACP (Figure S2; ACP does not mimic any intermediates in the TXS reaction, although it does mimic a common intermediate of many other diterpene cyclases). Metal-binding motifs that signal class I terpenoid cyclase function15,27 are conserved in TXS as D^{613}DMAD and N^{757}DTKTYQAE. The $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$ ions are coordinated by D613 and D617, and the $Mg^{2+}{}_B$ ion is chelated by N757, T761, and E765 (Figures 3b, S3b). Along with the recent observation of a trinuclear metal cluster in the active site of isoprene synthase28, the structure of the TXS-$Mg^{2+}{}_3$-FGP complex indicates that three-metal ion catalysis is conserved across the greater family of class I terpenoid synthases: C_5 hemiterpene, C_{10} monoterpene, C_{15} sesquiterpene, and C_{20} diterpene synthases.

In addition to metal coordination interactions, the diphosphate group of FGP also accepts hydrogen bonds from R754 and N757 (the latter residue also coordinates to $Mg^{2+}{}_B$) and

makes water-mediated hydrogen bonds with Y688, E691, Y835, S713, R768, and Q770. It is interesting to compare the molecular recognition of the FGP diphosphate group with that of the product diphosphate group in the plant monoterpene cyclase bornyl diphosphate synthase7 (Figure S3c). Most residues that assist the trinuclear metal cluster in binding and activating the substrate diphosphate group are conserved between these cyclases.

Class I terpenoid synthases undergo a significant structural transition from an open to a closed active site conformation upon the binding of 3 Mg^{2+} ions and the substrate diphosphate group, and this conformational transition helps to protect reactive carbocation intermediates from premature quenching by bulk solvent15,16. While the structure of the fully open conformation of TXS is unavailable, we suggest that the structural changes observed between open and closed active site conformations in plant monoterpene and sesquiterpene cyclases are representative of those that occur in the plant diterpene cyclase TXS. For example, active site closure in bornyl diphosphate synthase7 and 5-epi-aristolochene synthase9 involves conformational changes of loops flanking the mouth of the active site; additionally, the N-terminal polypeptide "caps" each active site. Specifically, the N-terminal polypeptide binds in a groove defined by the A–C and D-D1 loops on one side, and the J–K and H-H-α1 loops on the other. Tandem arginine residues in the N-terminus of bornyl diphosphate synthase make key hydrogen bond interactions in this groove. The N-terminus of 5-epi-aristolochene synthase contains only a single corresponding arginine residue, R15, that appears to serve a similar function in the structure of the closed active site conformation9. By analogy with the structures of these plant monoterpene and sesquiterpene cyclases, R84 in the missing N-terminal segment of TXS may help stabilize the fully closed, catalytically active conformation of mature taxadiene synthase. Accordingly, the closed conformations of the J–K loop and the N-terminal segment of TXS are readily modeled based on the bornyl diphosphate synthase structure to approximate the enclosed active site contour that serves as the template for GGPP cyclization (Figures 4, S4).

The active site contour of TXS encloses a larger volume in comparison with the active sites of monoterpene or sesquiterpene cyclases, which is consistent with the larger isoprenoid substrate of the diterpene cyclase. The active site cavity volumes of terpenoid synthases correlate with the hydrocarbon volume of their respective isoprenoid substrates (Table S2, Figure 4a). It has been suggested16 that the shape of the active site contour is more product-like for high-fidelity cyclases, i.e., those that generate a single cyclization product, whereas if the active site contour is less product-like, then a more promiscuous cyclase results that generates multiple cyclization products. For TXS, the active site volume is significantly larger than the volume of product taxadiene. This is consistent with the observation that TXS is a somewhat promiscuous cyclase, generating ~20% of the alternative isomer taxa-4(20),11(12)-diene21. Indeed, that TXS binds the bicyclic diterpene analogue ACP (Figure S2), which does not correspond to any intermediate in the TXS mechanism, clearly demonstrates promiscuity in ligand binding.

Taxadiene can fit in the enclosed active site contour of TXS with two alternative orientations (Figure S4). Each orientation leads to possible suggestions for active site bases that could function in the final deprotonation step of the cyclization cascade. Polar groups in the active site include S587, Q609, Y684, Y688, C719, and C830. While one of these

residues, e.g., Y688, could conceivably function as a base, taxadiene can also fit within the active site contour such that H5β of the preceding taxen-4-yl carbocation would be oriented toward the inorganic pyrophosphate product (PP_i) (Figures 4c, S4). Thus, the PP_i anion could serve as a stereospecific base, suggesting the possibility for substrate- or product-assisted catalysis.

Finally, although the N-terminal domain and the insertion domain of TXS form a double α-barrel class II terpenoid synthase fold such as that characterizing the triterpene cyclases10,26, the characteristic general acid DXDD motif and an active site cavity are absent. Regardless, the TXS structure illuminates structure-function relationships in other diterpene cyclases that contain catalytically active class II cyclase domains. For example, consider the bifunctional diterpene cyclase abietadiene synthase from the grand fir tree (*Abies grandis*). Here, the class II terpenoid cyclase domain first catalyzes the protonation-dependent cyclization of GGPP to form (+)-copalyl diphosphate, and the class I terpenoid cyclase domain then catalyzes the ionization-dependent cyclization of (+)-copalyl diphosphate to form abietadiene29. Since the structures of abietadiene synthase and TXS are expected to be homologous based on 44% amino acid sequence identity, the protonation-dependent reaction in the class II cyclase domain is presumably catalyzed in much the same manner as for a triterpene cyclase reaction. In other diterpene cyclases such as copalyl diphosphate synthase from *Arabidopsis thaliana* (related to TXS by 31% amino acid sequence identity), only the class II terpenoid cyclase domain is catalytically active; the class I terpenoid cyclase domain is vestigial and the signature metal binding motifs are absent30. Thus, biosynthetic diversity in the family of terpenoid natural products is rooted in a "mix and match" evolutionary strategy with class I and class II terpenoid cyclase folds, which can evolve together or separately as needed to generate the terpenoid product(s) required by the organism.

Methods Summary

A variety of different taxadiene synthase constructs were prepared, purified, and assessed in crystallization trials, but only one proved satisfactory for crystallization. This construct, designated TXS, was one in which 107 residues were deleted from the N-terminus and a hexa-histidine tag was added to the C-terminus to facilitate purification. TXS was expressed in *Escherichia coli* BL21 (DE3) RIL cells, purified, and cocrystallized with ACP or FGP by the sitting drop vapor diffusion method. The initial electron density map of the TXS-ACP complex was phased using single wavelength anomalous dispersion. Following map fitting, refinement converged smoothly to R/R_{free} = 0.167 / 0.205. The structure of the TXS-FGP complex was solved by molecular replacement and refined to R/R_{free} = 0.187 / 0.250. Data collection and refinement statistics are recorded in Table S1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Methods

Cloning, expression and purification of taxadiene synthase

Heterologous expression of taxadiene synthase from *Taxus brevifolia* lacking the N-terminal segment M1-V79 (M79-TXS) in *Escherichia coli* was achieved at the University of Pennsylvania using previously described procedures19. We found that this protein consistently underwent degradation at room temperature and 4 °C over a period of a few days to generate a soluble polypeptide stable for at least 4 weeks. Edman sequencing (Wistar Institute Proteomics Facility) showed that this polypeptide was missing the first 29 residues. Given its exceptional stability, this truncated polypeptide was considered a good candidate for crystallization. Accordingly, the M79-TXS gene segment corresponding to an N-terminal truncation at R107 (M107-TXS) was amplified by PCR with the following forward and reverse primers with flanking NdeI and BamHI sites, respectively: GCACATATGGAGAGTTCTACTTACCAAGAAC and GCAGGATCCTACTTGAATTGGATCAATATAAAC. A variant of the pET22b vector (Novagen, USA) (pET22bTV) was created by PCR with the following forward and reverse primers with complementary flanking restriction sites: GCAGGATCCCACCACCACCACCACC and GCACATATGTATATCTCCTTCTTAAAGTTAAAC. The M107-TXS gene and the pET22bTV vector were ligated to generate a plasmid encoding the M107-TXS polypeptide with a C-terminal hexa-histidine tag (M107-TXS-CHT), which was then used to transform *E. coli* XL1Blue cells for amplification. The resulting clones were confirmed by DNA sequencing (University of Pennsylvania School of Medicine Sequencing Facility) to have only two silent mutations and no amino acid substitutions.

The M107-TXS-CHT protein (henceforth designated “TXS”) was expressed using *E. coli* BL21 (DE3) RIL cells. Transformed cell cultures were grown in 2-L flasks containing 1 L Luria-Bertani medium with 100 mg ampicillin at 37 °C. At A_{600} = 0.6 – 0.7, cultures were equilibrated at 20 °C and expression was induced by 0.25 mM isopropyl-1-thio-β-D-galactopyranoside for 16 h. Cells were harvested by centrifugation at 6000*g* for 10 min, producing ~9 g pellet per liter of culture. The pellet was suspended in 20 mL of buffer E (50 mM K_2HPO_4 (pH 7.5), 300 mM NaCl, 10% (v/v) glycerol, 3 mM β-mercaptoethanol) containing 1 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride, then incubated at 4°C for 2 h with shaking. Cells were disrupted by sonication on ice with a large probe at medium power, 6 × (30 s on + 90 s off). Cell debris was cleared by centrifugation twice at 30,000*g* for 1 h. The clear supernatant was applied to a pre-equilibrated Talon column (Clontech Laboratories, CA) at a flow rate of 1 mL/min using an ÄKTAprime plus FPLC system (GE Healthcare Bio-Sciences AB, Sweden). The loaded column was washed 3 times with 5 column volumes of first buffer E, then buffer E plus 5 mM imidazole, then buffer E plus 10 mM imidazole. TXS was eluted with a gradient of 10–200 mM imidazole in buffer E at a flow rate of 2.5 mL/min. Selected fractions were combined, concentrated to 5 mL volume, and applied to a Superdex 200 preparatory grade 26/60 size exclusion column (GE Healthcare Bio-Sciences AB, Sweden) with buffer A (25 mM 3-(*N*-morpholino)-2-hydroxypropanesulfonic acid (MOPSO) (pH 6.8), 10% (v/v) glycerol and 1 mM dithiothreitol (DTT) containing 300 mM NaCl. Fractions from this run were combined,

concentrated to a 5-mL volume, and applied to the same column a second time with the same buffer. Fractions from the final size exclusion column were combined and concentrated to 8.6 mg/mL. The purity of TXS sample was 99% based on SDS-PAGE analysis. No hexane-extractable products were identified by gas chromatography-mass spectrometry analysis following incubation with geranylgeranyl diphosphate, indicating that this construct did not generate measurable amounts of taxadiene.

Crystallization

TXS could not be crystallized in the absence of isoprenoid diphosphate ligands. However, excellent crystals resulted when the protein was crystallized in the presence of 13-aza-13,14-dihydrocopalyl diphosphate (ACP) or 2-fluorogeranylgeranyl diphosphate (FGP) and Mg^{2+} ions by the sitting-drop vapor-diffusion method at 4 °C (ligand synthesis is outlined in the Supplementary Information). To obtain the crystals of the TXS-ACP complex, a 1-μL drop of protein solution [5 mg/mL TXS, 25 mM MOPSO (pH 6.8), 10% glycerol, 1 mM DTT, 2.5 mM ACP, 2.5 mM $MgCl_2$] was added to a 1-μL drop of precipitant solution [100 mM Bis-Tris (pH 6.5), 25% polyethylene glycol 3350, 200 mM NaCl] and equilibrated against a 250-μL well reservoir of precipitant solution. Prism-like crystals with rounded edges appeared within 2–3 days and grew to maximal dimensions of 50 μm × 100 μm × 200 μm in 2–3 weeks. These crystals were flash-cooled after transfer to a cryoprotectant solution consisting of the mother liquor augmented with 15% ethylene glycol. For the preparation of a heavy atom derivative for phasing, crystals of the TXS-ACP complex were soaked in a cryoprotectant solution [100 mM HEPES (pH 7.5), 25% polyethylene glycol 3350, 100 mM NaCl, 100 mM $MgCl_2$, 10% glycerol] containing 2 mM methylmercury chloride for 22 h at 15°C prior to flash-cooling.

To obtain crystals of the TXS-FGP complex, a 1-μL drop of protein solution [5 mg/mL TXS, 25 mM MOPSO (pH 6.8), 10% glycerol, 1 mM DTT, 2.5 mM 2FGG, 2.5 mM $MgCl_2$] was added to a 1 μL drop of precipitant solution [100 mM HEPES (pH 7.0), 20% polyethylene glycol 3350, 200 mM $MgCl_2$] and equilibrated against a 250 μL well reservoir of precipitant solution. These crystals were flash-cooled after transfer to a cryoprotectant solution consisting of the mother liquor augmented with 10% glycerol.

X-ray diffraction data collection and processing

Crystals of the TXS-ACP complex and the TXS-FGP complex diffracted X-rays to 1.82 Å and 2.25 Å resolution, respectively, at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, beamline X-29, using incident radiation with λ = 0.945 Å and 1.008 Å, respectively. Crystals of the mercury derivatized TXS-ACP complex diffracted X-rays to 2.6 Å resolution at NSLS beamline X-25 using incident radiation with λ = 1.000 Å. All diffraction data were processed with HKL200031. Crystals of the TXS-ACP complex belonged to space group $P2_12_12_1$ with unit cell parameters a = 55.46 Å, b = 72.41 Å, c = 206.93 Å; with one molecule in the asymmetric unit, the Matthews coefficient V_M = 2.35 $Å^3$/Dalton (solvent content = 48%). Crystals of the TXS-FGP complex belonged to space group $P2_1$ with unit cell parameters a = 54.05 Å, b = 201.98 Å, c = 81.43 Å, β = 91.60°; with two molecules in the asymmetric unit, the Matthews coefficient V_M = 2.61 $Å^3$/Dalton (solvent content = 53%). Data collection and reduction statistics are recorded in Table S1.

Phasing and structure refinement

The initial electron density map of the TXS-ACP complex was phased by single wavelength anomalous dispersion (SAD) using the 2.6 Å resolution data collected from the methylmercury chloride derivative. Initially, 6 Hg^{2+} atoms were located using the program HKL2MAP32 and used for SAD phasing, search and refinement of an additional 7 Hg^{2+} sites; density modification, initial electron density map calculation, and automatic model building was performed using the AUTOSOL routine implemented in PHENIX33. This procedure built more than 50% of the protein residues into the initial electron density map, most of which were α-helices. Manual model building subsequently generated an initial model with 90% of the residues registered in the sequence. This model was used for molecular replacement calculations using the AUTOMR routine implemented in PHENIX with the 1.82 Å resolution data collected from the TXS-ACP complex. After initial rigid body refinement, iterative cycles of positional refinement, grouped and individual atomic B-factor refinement, and manual model rebuilding were performed using PHENIX and COOT34, respectively. Water molecules, Mg^{2+} ions, and the ACP molecule were included in later cycles of refinement. A total of 745 of 764 residues are present in the final model of the TXS-ACP complex; disordered segments excluded from the final model include N-terminal residues M107-S110 (M107 is the N-terminus of the construct), the C-terminal hexa-histidine tag and its associated linker residues (G863-H870), and surface loop I838-A844. An electron density map of the TXS-ACP complex is found in Figure S2.

The model of the TXS-ACP complex less ligand and solvent atoms was used as a search probe for molecular replacement calculations to solve the structure of the TXS-FGP complex at 2.25 Å resolution. Rigid body refinement, positional refinement, grouped and individual atomic B-factor refinement, and manual model rebuilding were performed using PHENIX and COOT, respectively. In the final model of the TXS-FGP complex, 746 and 736 of 764 residues are present in monomers A and B, respectively. Disordered segments excluded from the final models of monomers A and B include N-terminal residues M107-S110, the C-terminal hexa-histidine tag and its associated linker residues (G863-H870), and loop I574-R578; additionally, surface loop F837-E846 is disordered in monomer B.

For both structures, data reduction and refinement statistics are recorded in Table S1. Ramachandran plot statistics, calculated with PROCHECK35, were as follows. TXS-ACP complex: Allowed 93.7%; additionally allowed 5.9%; generously allowed 0.3%; disallowed 0.1%. TXS-FGP complex: Allowed 91.3%; additionally allowed 8.1%; generously allowed 0.4%; disallowed 0.1%. Simulated annealing omit maps were calculated with CNS36. Protein structure figures were prepared with the graphics program PyMol (http://www.pymol.org).

Model of TXS in the fully-closed conformation

In order to model the N-terminus and J–K loop segments of TXS in a fully-closed conformation and to calculate the active site cavity volume, the N-terminus (residues 54–81) and J–K loop (residues 574–587) segments of bornyl diphosphate synthase in its complex with 3 Mg^{2+} ions and 3-azageranyl diphosphate (PDB accession code 1N20) were "grafted" onto the structure of the TXS-Mg^{2+}_3-FGP complex and mutated to the corresponding

residues of TXS; S110 was also introduced to account for an insertion in the sequence alignment between TXS and bornyl diphosphate synthase. The conformations of the grafted segments were then subjected to 10,000 steps of gas phase conjugate gradient energy minimization using NAMD37 and the CHARMM22 force field38. During energy minimization, the grafted segments plus three adjacent residues on N- and C-terminal ends were unconstrained, while the remaining heavy atoms were fixed. Non-bonded cutoff and switch distances were set to 12 Å and 10 Å, respectively. The final structure resulting from this computation was used as the hypothetical fully-closed conformation of TXS. The meshwork representing the active site cavity of TXS was calculated with VOIDOO39 using a probe with a radius of 1.4 Å to generate a molecular surface based on solvent accessibility. To study product binding orientations in the enclosed active site, a model of taxadiene was constructed based on the coordinates of the taxane core of Taxol deposited in the Cambridge Crystallographic Data Centre with accession code TEYPAO40.

The active site cavity volume of TXS was compared with the active site cavity volumes of other terpenoid cyclases, their substrates, and their products (Table S2, Figure 4a). All volume calculations were performed with VOIDOO using a probe with a radius of 0.0 Å to generate a molecular surface based on the atomic van der Waals radii. Because the active site of the hemiterpene synthase isoprene synthase was not fully closed due to disorder of the J–K loop, the active site contour was artificially truncated by the placement of "dummy" atoms to estimate the boundary of the fully enclosed cavity. We used a similar approach to model the cavity of TXS in a chemically sensible manner.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the US National Institutes of Health for grants GM56838 (D.W.C.), GM13956 (R.M.C.), and CA55254 (R.C.) in support of this research. Y.J. thanks the University of Illinois for support through the John C. Bailar and R.C. Fuson Fellowships. Additionally, we thank Chris MacDermaid and Jeffrey Saven for advice and assistance with molecular modeling calculations. Finally, we thank the National Synchrotron Light Source at Brookhaven National Laboratory for beamline access. We are grateful to Prof. Eric Oldfield for helpful comments on the manuscript.

References

1. Schiff PB, Fant J, Horwitz SB. Promotion of microtubule assembly *in vitro* by taxol. Nature. 1979; 277:665–667. [PubMed: 423966]
2. Arbuck, SG.; Blaylock, BA. Clinical results and current issues in development. In: Suffness, M., editor. Taxol: Science and Applications. CRC Press: Boca Raton, FL; 1995. p. 379-415.
3. Wani MC, Taylor HL, Wall M, Coggon P, McPhail AT. Plant antitumor agents. VI. The isolation and structure of Taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. J Am Chem Soc. 1971; 93:2325–2327. [PubMed: 5553076]
4. Koepp AE, et al. Cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene is the committed step of Taxol biosynthesis in Pacific yew. J Biol Chem. 1995; 270:8686–8690. [PubMed: 7721772]

5. Hezari M, Lewis NG, Croteau R. Purification and characterization of taxa-4(5),11(12)-diene synthase from Pacific yew (*Taxus brevifolia*) that catalyzes the first committed step of Taxol biosynthesis. Arch Biochem Biophys. 1995; 322:437–444. [PubMed: 7574719]
6. Wildung MR, Croteau R. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of Taxol biosynthesis. J Biol Chem. 1996; 271:9201–9204. [PubMed: 8621577]
7. Whittington DA, et al. Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci USA. 2002; 99:15375–15380. [PubMed: 12432096]
8. Lesburg CA, Zhai G, Cane DE, Christianson DW. Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. Science. 1997; 277:1820–1824. [PubMed: 9295272]
9. Starks CM, Back K, Chappell J, Noel JP. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science. 1997; 277:1815–1820. [PubMed: 9295271]
10. Wendt KU, Poralla K, Schulz GE. Structure and function of a squalene cyclase. Science. 1997; 277:1811–1815. [PubMed: 9295270]
11. Trapp SC, Croteau R. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics. 2001; 158:811–832. [PubMed: 11404343]
12. Keeling CI, et al. Identification and functional characterization of monofunctional *ent*-copalyl diphosphate and *ent*-kaurene synthases in white spruce reveal different patterns for diterpene synthase evolution for primary and secondary metabolism in gymnosperms. Plant Physiol. 2010; 152:1197–1208. [PubMed: 20044448]
13. Cao R, et al. Diterpene cyclases and the nature of the isoprene fold. Proteins: Struct Funct Bioinformatics. 2010; 78:2417–2432.
14. Wendt KU, Schulz GE. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. Structure. 1998; 6:127–133. [PubMed: 9519404]
15. Christianson DW. Structural biology and chemistry of the terpenoid cyclases. Chem Rev. 2006; 106:3412–3442. [PubMed: 16895335]
16. Christianson DW. Unearthing the roots of the terpenome. Curr Opin Chem Biol. 2008; 12:141–150. [PubMed: 18249199]
17. Wendt KU, Schulz GE, Corey EJ, Liu DR. Enzyme mechanisms for polycyclic triterpene formation. Angew Chem Int Ed. 2000; 39:2812–2833.
18. Lin X, Hezari M, Koepp AE, Floss HG, Croteau R. Mechanism of taxadiene synthase, a diterpene cyclase that catalyzes the first step of Taxol biosynthesis in Pacific yew. Biochemistry. 1996; 35:2968–2977. [PubMed: 8608134]
19. Williams DC, et al. Heterologous expression and characterization of a "pseudomature" form of taxadiene synthase involved in paclitaxel (Taxol) biosynthesis and evaluation of a potential intermediate and inhibitors of the multistep diterpene cyclization reaction. Arch Biochem Biophys. 2000; 379:137–146. [PubMed: 10864451]
20. Williams DC, et al. Intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to the taxadiene precursor of taxol catalyzed by recombinant taxadiene synthase. Chem Biol. 2000; 7:969–977. [PubMed: 11137819]
21. Jin Q, Williams DC, Hezari M, Croteau R, Coates RM. Stereochemistry of the macrocyclization and elimination steps in taxadiene biosynthesis through deuterium labelling. J Org Chem. 2005; 70:4667–4675. [PubMed: 15932303]
22. Jin Y, Williams DC, Croteau R, Coates RM. Taxadiene synthase-catalyzed cyclization of 6-fluorogeranylgeranyl diphosphate to 7-fluoroverticillenes. J Am Chem Soc. 2005; 127:7834–7842. [PubMed: 15913373]
23. Tarshis LC, Yan M, Poulter CD, Sacchettini JC. Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-Å resolution. Biochemistry. 1994; 33:10871–10877. [PubMed: 8086404]
24. Chang TH, Guo RT, Ko TP, Wang AHJ, Liang PH. Crystal structure of type-III geranylgeranyl pyrophosphate synthase from *Saccharomyces cerevisiae* and the mechanism of product chain length determination. J. Biol. Chem. 2006; 281:14991–15000. [PubMed: 16554305]

25. Bohlmann J, Meyer-Gauen G, Croteau R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc Natl Acad Sci USA. 1998; 95:4126–4133. [PubMed: 9539701]
26. Thoma R, et al. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. Nature. 2004; 432:118–122. [PubMed: 15525992]
27. Rynkiewicz MJ, Cane DE, Christianson DW. Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. Proc Natl Acad Sci USA. 2001; 98:13543–13548. [PubMed: 11698643]
28. Köksal M, Zimmer I, Schnitzler J-P, Christianson DW. Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission. J Mol Biol. 2010; 402:363–373. [PubMed: 20624401]
29. Peters RJ, Ravn MM, Coates RM, Croteau R. Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. J Am Chem Soc. 2001; 123:8974–8978. [PubMed: 11552804]
30. Prisic S, Peters RJ. Synergistic substrate inhibition of *ent*-copalyl diphosphate synthase: a potential feed-forward inhibition mechanism limiting gibberellin metabolism. Plant Physiol. 2007; 144:445–454. [PubMed: 17384166]

References for Methods

31. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.
32. Pape T, Schneider TR. HKL2MAP: a graphical user interface for phasing with SHELX programs. J Appl Cryst. 2004; 37:843–844.
33. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. 2010; D66:213–221.
34. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr. 2010; D66:486–501.
35. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - a program to check the stereochemical quality of protein structures. J Appl Cryst. 1993; 26:283–291.
36. Brünger AT, et al. Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallogr. 1998; D54:905–921.
37. Phillips JC, et al. Scalable molecular dynamics with NAMD. J Computational Chem. 2005; 26:1781–1802.
38. MacKerell AD Jr, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem. 1998; B102:3586–3616.
39. Kleywegt GJ, Jones TA. Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr. 1994; D50:178–185.
40. Gao Q, Chen SH. An unprecedented side chain conformation of paclitaxel (Taxol®): crystal structure of 7-mesylpaclitaxel. Tetrahedron Lett. 1996; 37:3425–3428.

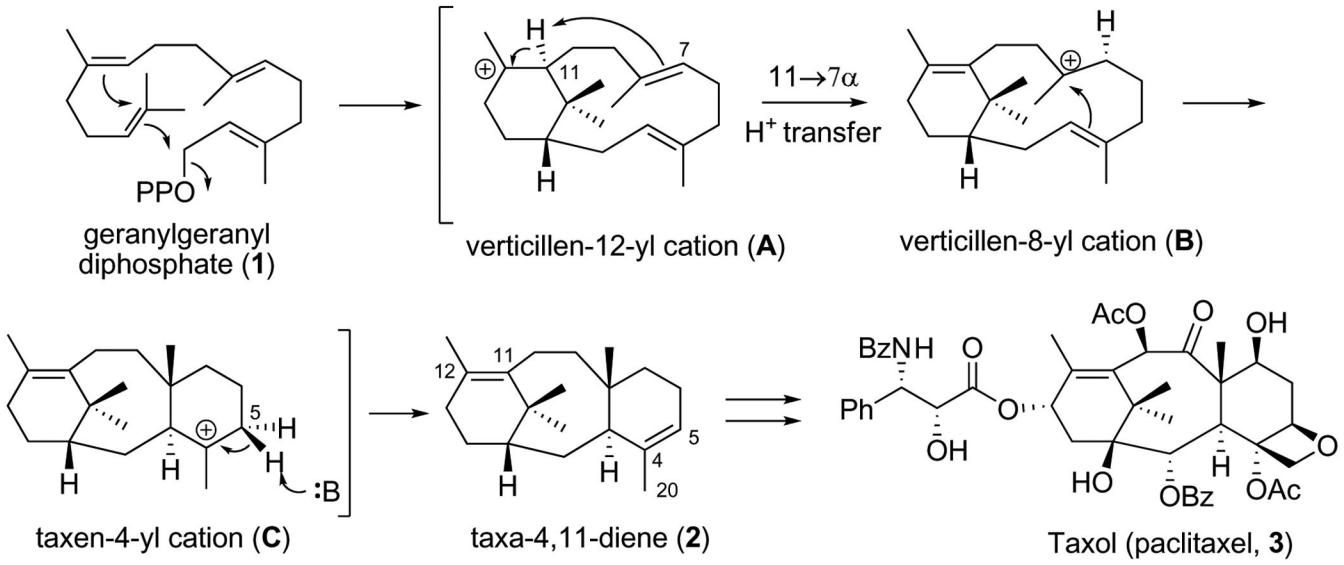

Figure 1. Proposed catalytic mechanism of taxadiene synthase

The cyclization of GGPP to form taxadiene is the first committed step of Taxol (paclitaxel) biosynthesis in yew species (OPP = diphosphate, Ph = phenyl, Ac = acetyl, Bz = benzyl). Taxadiene is converted to Taxol through a lengthy series of oxidation and acylation steps.

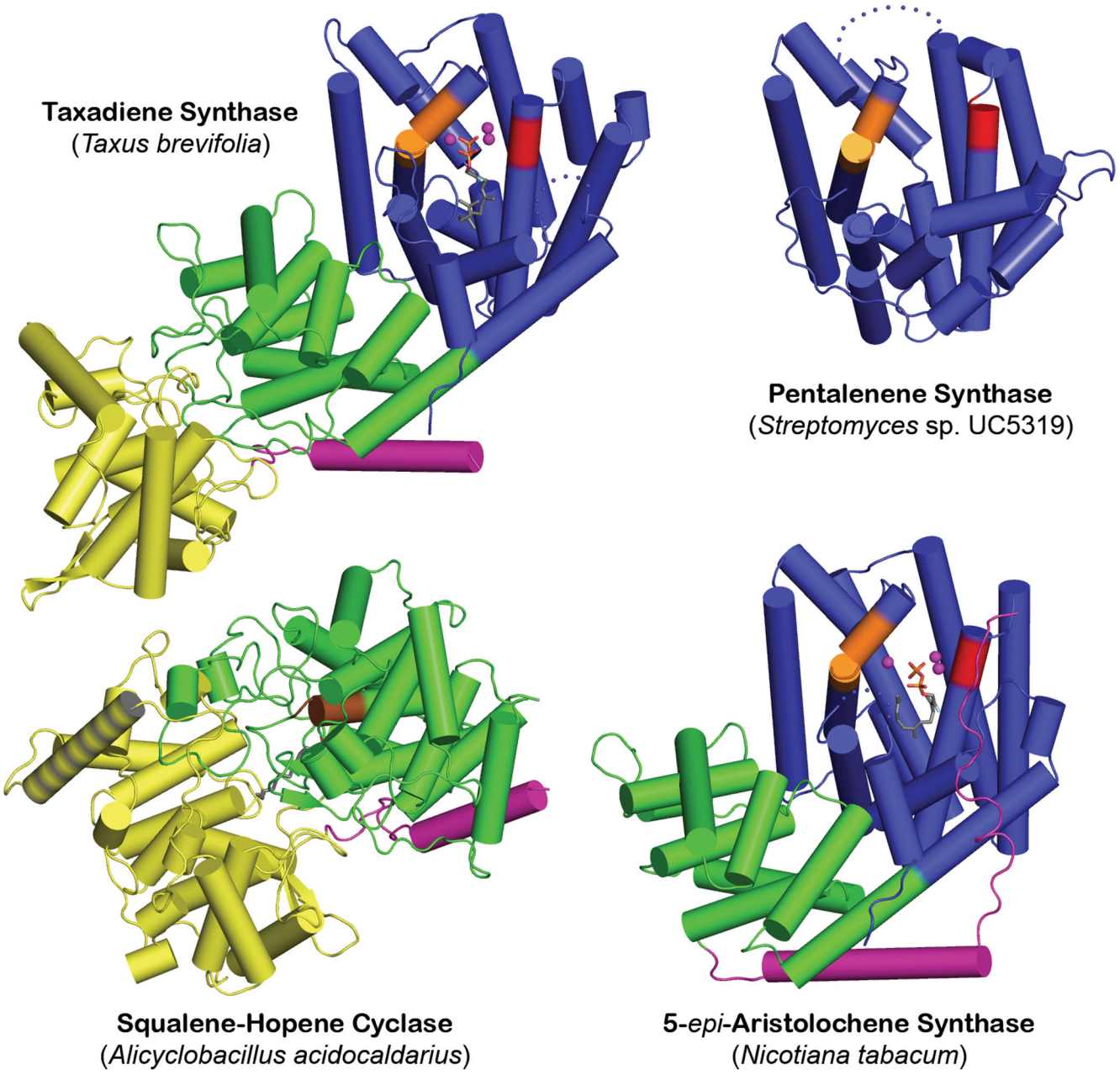

Figure 2. Structural relationships among terpenoid cyclases

The class I terpenoid cyclase fold of pentalenene synthase8 (PDB 1PS1) (blue, "α-domain"13) contains metal-binding motifs DDXXD and (N,D)DXX(S,T)XXXE (red and orange, respectively); in 5-epi-aristolochene synthase9 (PDB 1LZ9), this domain is linked to a smaller vestigial domain (green, "β-domain"13). A related domain is found in the class II terpenoid cyclase fold of squalene-hopene cyclase10 (PDB 1SQC), where it contains the general acid motif DXDD (brown) and a second domain (yellow, "γ-domain"13) inserted between the first and second helices; a hydrophobic plateau flanking helix 8 (gray stripes) enables membrane insertion. Taxadiene synthase (PDB 3P5R) contains both class I and class II terpenoid cyclase folds, but only the class I domain is catalytically active. The role of N-

termini (purple) in class I plant cyclases is to "cap" the active site, as shown for 5-epi-aristolochene synthase.

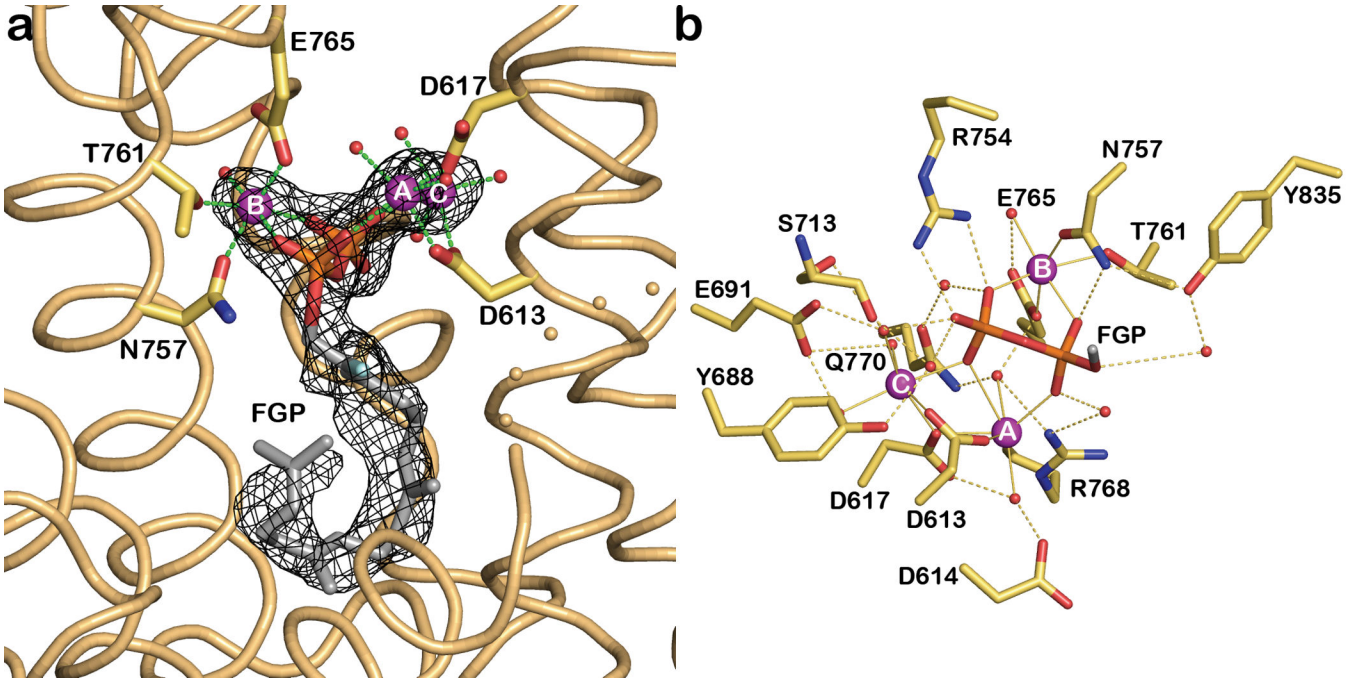

Figure 3. Substrate analogue binding to TXS

a, Simulated annealing $|F_o|-|F_c|$ omit map in which FGP and 3 Mg^{2+} ions are omitted from the structure factor calculation (contoured at 3.0σ); the side chains of metal ligands are indicated. **b**, Molecular recognition of the substrate diphosphate group in the TXS active site. For clarity, the isoprenoid moiety of FGP is truncated to one carbon (gray). Metal coordination and hydrogen bond interactions are indicated by green and black dashed lines, respectively. Atoms are color coded as follows: carbon = yellow, nitrogen = blue, oxygen = red, phosphorus = orange; Mg^{2+} ions and water molecules appear as purple or red spheres, respectively. A corresponding stereofigure is found in Figure S3.

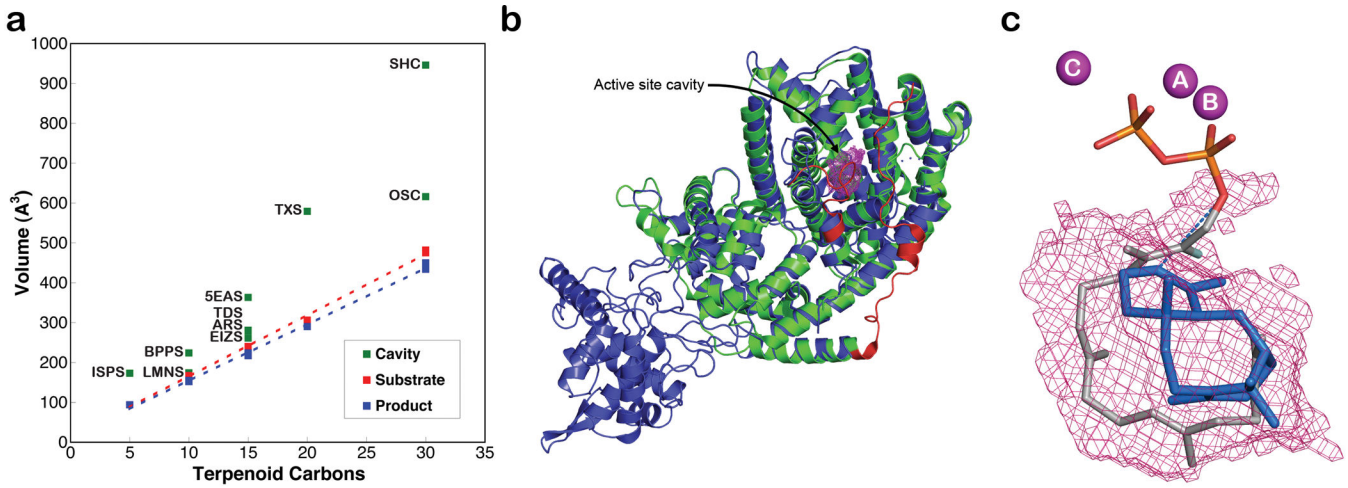

Figure 4. Active site cavities of terpenoid synthases

a, Active site volumes are generally slightly larger than corresponding substrate and product volumes, perhaps to accommodate structural changes better during the cyclization cascade. Abbreviations are outlined in Table S2. **b**, Superposition of TXS (blue) and bornyl diphosphate synthase (green) guides the modeling of the J–K loop and the N-terminal segment of TXS (red) to define the enclosed active site cavity (solvent-accessible surface, magenta meshwork). **c**, One orientation of taxadiene (blue) fits in the active site cavity such that the H5β atom of the preceding taxen-4-yl cation would be oriented toward the diphosphate leaving group, suggesting that the PP_i anion could serve as the stereospecific base that terminates the cyclization cascade. Three Mg^{2+} ions and FGP are shown for reference; all protein atoms are omitted for clarity. A corresponding stereofigure is found in Figure S4.