LETTER

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] & David W. Christianson[1]

With more than 55,000 members identified so far in all forms of life, the family of terpene or terpenoid natural products represents the epitome of molecular biodiversity. A well-known and important member of this family is the polycyclic diterpenoid Taxol (paclitaxel), which promotes tubulin polymerization[1] and shows remarkable efficacy in cancer chemotherapy[2]. 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)diene[4], which is catalysed by taxadiene synthase[5]. The full-length form of this diterpene cyclase contains 862 residues, but a roughly 80-residue amino-terminal transit sequence is cleaved on maturation in plastids[6]. 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, hereafter designated TXS. Specifically, we have determined structures of TXS complexed with 13-aza-13,14-dihydrocopalyl diphosphate (1.82 Å resolution) and 2-fluorogeranylgeranyl diphosphate (2.25 Å resolution). The TXS structure reveals a modular assembly of three α-helical domains. The carboxy-terminal catalytic domain is a class I terpenoid cyclase, which binds and activates substrate GGPP with a three-metal ion cluster. 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, 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} monoterpene[7], C_{15} sesquiterpene[8,9] and C_{30} triterpene[10] cyclases appeared several years ago, the structure of the 'missing link' in this series—a C_{20} diterpene cyclase—has been unknown 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 synthase[11,12]. 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 progenitors[13].

The two distinct classes of terpenoid cyclase have unrelated protein folds and use different substrate activation mechanisms[14–17]. A class I terpenoid cyclase uses a trinuclear metal cluster liganded by conserved motifs **D**DXX**D** and (**N**,**D**)DXX(**S**,**T**)XXX**E** (bold 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, using 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 activity[5], 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 proteins[18]. 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 deuterated[18–21] and fluorinated[22] analogues of GGPP. These studies suggest a cyclization mechanism (Fig. 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 the 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, whose deprotonation 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 base[20]. The base mediating the final deprotonation step has not yet been identified.

The successful crystallization of TXS required co-crystallization with Mg^{2+} and either 13-aza-13,14-dihydrocopalyl diphosphate (ACP) or 2-fluorogeranylgeranyl diphosphate (FGP) (molecular structures are shown in Supplementary Fig. 1). Although 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 harbours the folds of both class I and class II terpenoid cyclases (Fig. 2); this structure is representative of nearly all diterpene cyclases. The C-terminal domain (S553–V862) has the class I terpenoid synthase fold that was first observed in farnesyl diphosphate synthase[23] and subsequently observed and designated the class I terpenoid synthase fold[8,14] in monoterpene and sesquiterpene cyclases[7–9]. This fold is also observed in geranylgeranyl diphosphate synthase[24], which generates the substrate for diterpene cyclases. The N-terminal domain of TXS (M107–I135 and S349–Q552) together with the 'insertion' domain[25] (S136–Y348) comprise the double α-barrel class II terpenoid synthase fold that was first observed in the triterpene cyclase squalene-hopene cyclase[10] and later observed in oxidosqualene cyclase[26]. 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 (Fig. 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 synthase[8] and trichodiene synthase[27], 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 Schulz[14], is homologous to the N-terminal domain of the class II triterpene cyclase squalene-hopene cyclase[10]. The first plant monoterpene and sesquiterpene synthases to yield crystal structures were bornyl diphosphate synthase[7] and 5-*epi*-aristolochene synthase[9], respectively. Most plant diterpene synthases contain three

[1]Roy and Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104-6323, USA. [2]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [3]Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA. [4]Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164-6340, USA.

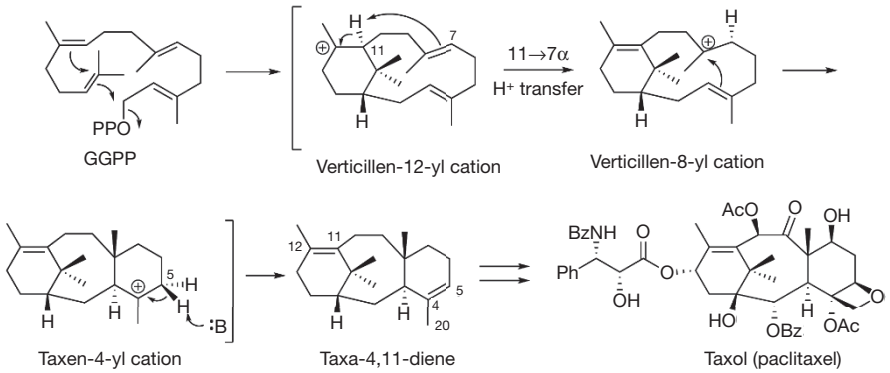

Figure 1 | Proposed catalytic mechanism of taxadiene synthase. The cyclization of GGPP to form taxadiene is the first committed step of Taxol (paclitaxel) biosynthesis in yew species. OPP, diphosphate; Ph, phenyl; Ac, acetyl; Bz, benzyl. Taxadiene is converted to Taxol through a lengthy series of oxidation and acylation steps.

domains, the third being an insertion conserved in sequence and position[25]. It was correctly predicted that this domain is homologous to the insertion domain of a triterpene cyclase on the basis of bioinformatics analysis[13].

It is interesting to note that the class II triterpene cyclases squalene-hopene cyclase[10] and oxidosqualene cyclase[26] are monotopic membrane proteins: each penetrates, but does not completely pass through, the membranes in which they are localized. Their triterpene substrates (squalene and 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 probably serves as the membrane anchor (Fig. 2). In contrast, TXS functions in the plastid lumen, so its insertion domain does not contain the corresponding hydrophobic components.

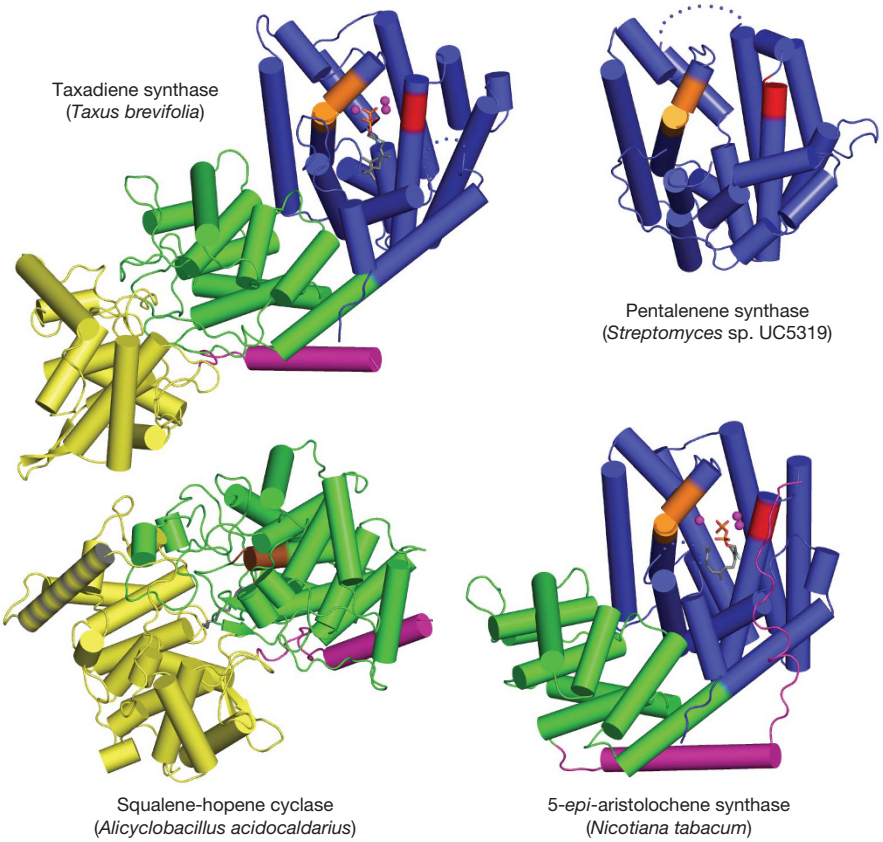

Figure 2 | Structural relationships among terpenoid cyclases. The class I terpenoid cyclase fold of pentalenene synthase[8] (PDB accession code 1PS1) (blue; 'α domain'[13]) contains metal-binding motifs DDXXD and (N,D)DXX(S,T)XXXE (red and orange, respectively); in 5-*epi*-aristolochene synthase[9] (PDB accession code 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 cyclase[10] (PDB accession code 1SQC), where it contains the general acid motif DXDD (brown) and a second domain (yellow; 'γ domain'[13]) inserted between the first (purple) and second helices; a hydrophobic plateau flanking helix 8 (grey stripes) enables membrane insertion. Taxadiene synthase (PDB accession code 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.

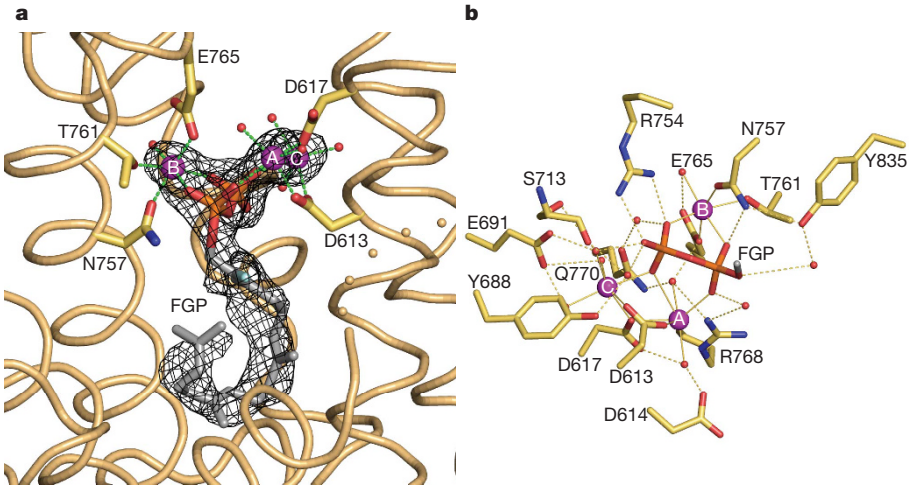

Figure 3 | Binding of substrate analogue to TXS. **a**, Simulated annealing $|F_o| - |F_c|$ omit map in which FGP and three Mg^{2+} ions are omitted from the structure factor calculation (contoured at 3.0σ); the side chains of metal ligands are indicated. **b**, Molecular recognition of the substrate diphosphate group in the TXS active site. For clarity, the isoprenoid moiety of FGP is truncated to one carbon (grey). Metal coordination and hydrogen bond interactions are indicated by thin solid and dashed lines, respectively. Atoms are colour-coded as follows: yellow, carbon; blue, nitrogen; red, oxygen; orange, phosphorus. Mg^{2+} ions (A, B and C) and water molecules are shown as purple and red spheres, respectively. A corresponding stereo figure is shown in Supplementary Fig. 3.

The active site of TXS is located in the C-terminal domain and is the exclusive binding site of the substrate analogue FGP (Fig. 3a and Supplementary Fig. 3a) and the bicyclic isoprenoid ACP (Supplementary Fig. 2; 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 function[15,27] 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 (Fig. 3b and Supplementary Fig. 3b). Along with the recent observation of a trinuclear metal cluster in the active site of isoprene synthase[28], 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 synthase[7] (Supplementary Fig. 3c). 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 after the binding of three Mg^{2+} ions and the substrate diphosphate group, and this conformational transition helps to protect reactive carbocation intermediates from premature quenching by bulk solvent[15,16]. Although 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 synthase[7] and 5-*epi*-aristolochene synthase[9] involves conformational changes of loops flanking the mouth of the active site; in addition, 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

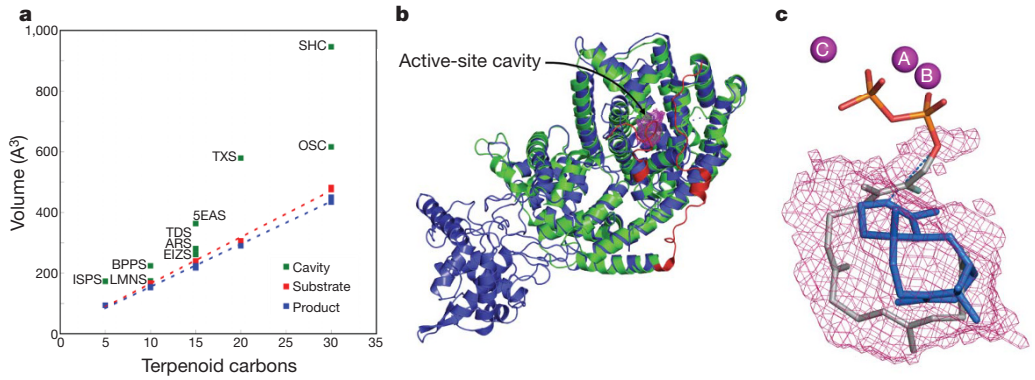

Figure 4 | Active-site cavities of terpenoid synthases. **a**, Active-site volumes are generally slightly larger than corresponding substrate and product volumes, perhaps to accommodate structural changes better during the cyclization cascade. Abbreviations are defined in Supplementary Table 2. **b**, Superposition of TXS (blue) and bornyl diphosphate synthase (green) guides the modelling of the J–K loop and the N-terminal segment of TXS (red) to define the enclosed active-site cavity (the magenta meshwork indicates the solvent-accessible surface). **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 towards 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 (A, B and C) and FGP are shown for reference; all protein atoms are omitted for clarity. A corresponding stereo figure is shown in Supplementary Fig. 4.

terminus of 5-*epi*-aristolochene synthase contains only a single corresponding arginine residue, R15, that seems to serve a similar function in the structure of the closed active-site conformation[9]. By analogy with the structures of these plant monoterpene and sesquiterpene cyclases, R84 in the missing N-terminal segment of TXS may help to 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 modelled on the basis of the bornyl diphosphate synthase structure to approximate the enclosed active-site contour that serves as the template for GGPP cyclization (Fig. 4 and Supplementary Fig. 4).

The active-site contour of TXS encloses a larger volume than 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 (Fig. 4a and Supplementary Table 2). It has been suggested[16] that the shape of the active-site contour is more product-like for high-fidelity cyclases—that is, those that generate a single cyclization product—whereas if the active-site contour is less product-like, a more promiscuous cyclase results that generates multiple cyclization products. For TXS, the active-site volume is significantly larger than the volume of the product taxadiene. This is consistent with the observation that TXS is a somewhat promiscuous cyclase, generating about 20% of the alternative isomer taxa-4(20),11(12)-diene[21]. Indeed, the fact that TXS binds the bicyclic diterpene analogue ACP (Supplementary Fig. 2), 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 (Supplementary Fig. 4). 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. Although one of these residues, for example 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 towards the inorganic pyrophosphate (PP_i) product (Fig. 4c and Supplementary Fig. 4). Thus, the PP_i anion could serve as a stereospecific base, suggesting the possibility for substrate-assisted 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 cyclases[10,26], the characteristic general acid DXDD motif and an active-site cavity are absent. Nevertheless, 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 catalyses the protonation-dependent cyclization of GGPP to form (+)-copalyl diphosphate, and the class I terpenoid cyclase domain then catalyses the ionization-dependent cyclization of (+)-copalyl diphosphate to form abietadiene[29]. Because the structures of abietadiene synthase and TXS are expected to be homologous, on the basis of 44% amino-acid sequence identity, the protonation-dependent reaction in the class II cyclase domain is presumably catalysed 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 absent[30]. 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 hexahistidine tag was added to the C terminus to facilitate purification. TXS was expressed in *Escherichia coli* BL21 (DE3) RIL cells, purified, and co-crystallized with ACP or FGP by the sitting-drop vapour-diffusion method. The initial electron density map of the TXS–ACP complex was phased by using single-wavelength anomalous dispersion. After 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 shown in Supplementary Table 1.

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

Received 6 August; accepted 29 October 2010.
Published online 15 December 2010.

1. Schiff, P. B., Fant, J. & Horwitz, S. B. Promotion of microtubule assembly *in vitro* by taxol. *Nature* **277,** 665–667 (1979).
2. Arbuck, S. G. & Blaylock, B. A. in *Taxol: Science and Applications* (ed. Suffness, M.) 379–415 (CRC Press, 1995).
3. Wani, M. C., Taylor, H. L., Wall, M., Coggon, P. & McPhail, A. T. Plant antitumor agents. VI. The isolation and structure of Taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **93,** 2325–2327 (1971).
4. Koepp, A. E. *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.* **270,** 8686–8690 (1995).
5. Hezari, M., Lewis, N. G. & 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.* **322,** 437–444 (1995).
6. Wildung, M. R. & Croteau, R. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of Taxol biosynthesis. *J. Biol. Chem.* **271,** 9201–9204 (1996).
7. Whittington, D. A. *et al.* Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl Acad. Sci. USA* **99,** 15375–15380 (2002).
8. Lesburg, C. A., Zhai, G., Cane, D. E. & Christianson, D. W. Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. *Science* **277,** 1820–1824 (1997).
9. Starks, C. M., Back, K., Chappell, J. & Noel, J. P. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **277,** 1815–1820 (1997).
10. Wendt, K. U., Poralla, K. & Schulz, G. E. Structure and function of a squalene cyclase. *Science* **277,** 1811–1815 (1997).
11. Trapp, S. C. & Croteau, R. Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* **158,** 811–832 (2001).
12. Keeling, C. I. *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.* **152,** 1197–1208 (2010).
13. Cao, R. *et al.* Diterpene cyclases and the nature of the isoprene fold. *Proteins Struct. Funct. Bioinf.* **78,** 2417–2432 (2010).
14. Wendt, K. U. & Schulz, G. E. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure* **6,** 127–133 (1998).
15. Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106,** 3412–3442 (2006).
16. Christianson, D. W. Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* **12,** 141–150 (2008).
17. Wendt, K. U., Schulz, G. E., Corey, E. J. & Liu, D. R. Enzyme mechanisms for polycyclic triterpene formation. *Angew. Chem. Int. Ed.* **39,** 2812–2833 (2000).
18. Lin, X., Hezari, M., Koepp, A. E., Floss, H. G. & Croteau, R. Mechanism of taxadiene synthase, a diterpene cyclase that catalyzes the first step of Taxol biosynthesis in Pacific yew. *Biochemistry* **35,** 2968–2977 (1996).
19. Williams, D. C. *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.* **379,** 137–146 (2000).
20. Williams, D. C. *et al.* Intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to the taxadiene precursor of taxol catalyzed by recombinant taxadiene synthase. *Chem. Biol.* **7,** 969–977 (2000).
21. Jin, Q., Williams, D. C., Hezari, M., Croteau, R. & Coates, R. M. Stereochemistry of the macrocyclization and elimination steps in taxadiene biosynthesis through deuterium labelling. *J. Org. Chem.* **70,** 4667–4675 (2005).
22. Jin, Y., Williams, D. C., Croteau, R. & Coates, R. M. Taxadiene synthase-catalyzed cyclization of 6-fluorogeranylgeranyl diphosphate to 7-fluoroverticillenes. *J. Am. Chem. Soc.* **127,** 7834–7842 (2005).
23. Tarshis, L. C., Yan, M., Poulter, C. D. & Sacchettini, J. C. Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-Å resolution. *Biochemistry* **33,** 10871–10877 (1994).

24. Chang, T. H., Guo, R. T., Ko, T. P., Wang, A. H. J. & Liang, P. H. Crystal structure of type-III geranylgeranyl pyrophosphate synthase from *Saccharomyces cerevisiae* and the mechanism of product chain length determination. *J. Biol. Chem.* **281,** 14991–15000 (2006).
25. Bohlmann, J., Meyer-Gauen, G. & Croteau, R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl Acad. Sci. USA* **95,** 4126–4133 (1998).
26. Thoma, R. *et al.* Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* **432,** 118–122 (2004).
27. Rynkiewicz, M. J., Cane, D. E. & Christianson, D. W. Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl Acad. Sci. USA* **98,** 13543–13548 (2001).
28. Köksal, M., Zimmer, I., Schnitzler, J.-P. & Christianson, D. W. Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission. *J. Mol. Biol.* **402,** 363–373 (2010).
29. Peters, R. J., Ravn, M. M., Coates, R. M. & Croteau, R. Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J. Am. Chem. Soc.* **123,** 8974–8978 (2001).
30. Prisic, S. & Peters, R. J. Synergistic substrate inhibition of *ent*-copalyl diphosphate synthase: a potential feed-forward inhibition mechanism limiting gibberellin metabolism. *Plant Physiol.* **144,** 445–454 (2007).

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

Acknowledgements We thank C. MacDermaid and J. Saven for advice and assistance with molecular modelling calculations, and E. Oldfield for helpful comments on the manuscript. We thank the National Synchrotron Light Source at Brookhaven National Laboratory for beamline access. The US National Institutes of Health provided 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.

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 are deposited in the Protein Data Bank with accession codes 3P5P and 3P5R, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.W.C. (chris@sas.upenn.edu).

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 procedures described previously[19]. We found that this protein consistently underwent degradation at 20 °C 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 lacked 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 *Nde*I and *Bam*HI sites, respectively: 5′-GCACATATGGAGAGTTCTACT TACCAAGAAC-3′ and 5′-GCAGGATCCTACTTGAATTGGATCAATATAA AC-3′. A variant of the pET22b vector (pET22bTV; Novagen) was created by PCR with the following forward and reverse primers with complementary flanking restriction sites: 5′-GCAGGATCCCACCACCACCACCACC-3′ and 5′-GCACA TATGTATATCTCCTTCTTAAAGTTAAAC-3′. The gene encoding M107-TXS and the pET22bTV vector were ligated to generate a plasmid encoding the M107-TXS polypeptide with a C-terminal hexahistidine 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 in *E. coli* BL21 (DE3) RIL cells. Transformed cell cultures were grown in 2-l flasks containing 1 l of Luria–Bertani medium with 100 mg of ampicillin at 37 °C. At an attenuance (D_{600}) of 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 6,000*g* for 10 min, producing about 9 g of pellet per litre 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 2-mercaptoethanol) containing 1 mg ml^{-1} lysozyme and 1 mM phenylmethylsulfonyl fluoride, then incubated at 4 °C for 2 h with shaking. Cells were disrupted by sonication on ice six times (30 s on and 90 s off) with a large probe at medium power. 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) at a flow rate of 1 ml min^{-1} with an ÄKTAprime plus fast performance liquid chromatography system (GE Healthcare Bio-Sciences AB). The loaded column was washed three times with 5 column volumes of 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^{-1}. Selected fractions were combined, concentrated to a volume of 5 ml, and applied to a Superdex 200 preparatory-grade 26/60 size-exclusion column (GE Healthcare Bio-Sciences AB) with buffer A (25 mM 3-(*N*-morpholino)-2-hydroxypropanesulfonic acid (MOPSO) pH 6.8, 10% (v/v) glycerol, 1 mM dithiothreitol (DTT)) containing 300 mM NaCl. Fractions from this run were combined, concentrated to a volume of 5 ml 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^{-1}. The purity of the TXS sample was 99% by SDS–PAGE analysis. No hexane-extractable products were identified by gas chromatography–mass spectrometry analysis after incubation with GGPP, 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 ACP or FGP and Mg^{2+} ions by the sitting-drop vapour-diffusion method at 4 °C (ligand synthesis is outlined in Supplementary Information). To obtain the crystals of the TXS–ACP complex, a 1-μl drop of protein solution (5 mg ml^{-1} 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 before flash-cooling.

To obtain crystals of the TXS–FGP complex, a 1-μl drop of protein solution (5 mg ml^{-1} TXS, 25 mM MOPSO pH 6.8, 10% glycerol, 1 mM DTT, 2.5 mM FGP, 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.

Collection and processing of X-ray diffraction data. Crystals of the TXS–ACP and TXS–FGP complexes 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 $\lambda = 0.945$ 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 $\lambda = 1.000$ Å. All diffraction data were processed with HKL2000 (ref. 31). 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 was 2.35 Å^3 Da^{-1} (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, $\beta = 91.60°$, with two molecules in the asymmetric unit; $V_M = 2.61$ Å^3 Da^{-1} (solvent content 53%). Data collection and reduction statistics are shown in Supplementary Table 1.

Phasing and structure refinement. The initial electron density map of the TXS–ACP complex was phased by single-wavelength anomalous dispersion (SAD) with the 2.6-Å resolution data collected from the methylmercury chloride derivative. Initially, six Hg^{2+} atoms were located by using the program HKL2MAP[32] and used for SAD phasing, search and refinement of an additional seven Hg^{2+} sites; density modification, initial electron density map calculation and automatic model building were performed with the AUTOSOL routine implemented in PHENIX[33]. 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 with the AUTOMR routine implemented in PHENIX with the 1.82-Å resolution data collected from the TXS–ACP complex. Initial rigid-body refinement, iterative cycles of positional refinement, and grouped and individual atomic *B*-factor refinement were performed with PHENIX. Manual model rebuilding was performed with COOT[34]. Water molecules, Mg^{2+} ions and the ACP molecule were included in later cycles of refinement. A total of 745 out 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 hexahistidine tag and its associated linker residues (G863–H870) and surface loop I838–A844. An electron density map of the TXS–ACP complex is shown in Supplementary Fig. 2.

The model of the TXS–ACP complex without its 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 and grouped and individual atomic *B*-factor refinement were performed with PHENIX. Manual model rebuilding was performed with COOT. In the final model of the TXS–FGP complex, 746 and 736 out of 764 residues were present in monomers A and B, respectively. Disordered segments excluded from the final models of monomers A and B included N-terminal residues M107–S110, the C-terminal hexahistidine tag and its associated linker residues (G863–H870) and loop I574–R578; in addition, surface loop F837–E846 was disordered in monomer B.

For both structures, data reduction and refinement statistics are shown in Supplementary Table 1. Ramachandran plot statistics, calculated with PROCHECK[35], 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 CNS[36]. Protein structure figures were prepared with the graphics program PyMol (http://www.pymol.org/).

Model of TXS in the fully closed conformation. 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 three 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 NAMD[37] and the CHARMM22 force field[38]. During energy minimization the grafted segments plus three adjacent residues on the N-terminal 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 VOIDOO[39], 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 on the basis of the coordinates of the taxane core of Taxol deposited in the Cambridge Crystallographic Data Centre with accession code TEYPAO[40].

The active-site cavity volume of TXS was compared with the active-site cavity volumes of other terpenoid cyclases, their substrates and their products (Fig. 4a and Supplementary Table 2). 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 as a result of 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.

31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
32. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.* **37,** 843–844 (2004).
33. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* **D66,** 213–221 (2010).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr.* **D66,** 486–501 (2010).
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26,** 283–291 (1993).
36. Brünger, A. T. *et al.* Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54,** 905–921 (1998).
37. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
38. MacKerell, A. D. Jr *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).
39. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr.* **D50,** 178–185 (1994).
40. Gao, Q. & Chen, S. H. An unprecedented side chain conformation of paclitaxel (Taxol®): crystal structure of 7-mesylpaclitaxel. *Tetrahedr. Lett.* **37,** 3425–3428 (1996).