Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

The first structure of a bacterial diterpene cyclase: CotB2

Ronja Janke,[a]‡ Christian Görner,[b]‡ Max Hirte,[b] Thomas Brück[b]* and Bernhard Loll[a]*

[a]Institut für Chemie und Biochemie, Abteilung Strukturbiochemie, Freie Universität Berlin, Takustrasse 6, 14195 Berlin, Germany, and [b]Fachgebiet Industrielle Biokatalyse, Technische Universität München, Lichtenbergstrasse 4, 85748 Garching, Germany

‡ These authors contributed equally to this work.

Correspondence e-mail: brueck@tum.de, loll@chemie.fu-berlin.de

Received 27 January 2014
Accepted 11 March 2014

PDB references: CotB2, 4omg; F149L mutant, 4omh

Sesquiterpenes and diterpenes are a diverse class of secondary metabolites that are predominantly derived from plants and some prokaryotes. The properties of these natural products encompass antitumor, antibiotic and even insecticidal activities. Therefore, they are interesting commercial targets for the chemical and pharmaceutical industries. Owing to their structural complexity, these compounds are more efficiently accessed by metabolic engineering of microbial systems than by chemical synthesis. This work presents the first crystal structure of a bacterial diterpene cyclase, CotB2 from the soil bacterium *Streptomyces melanosporofaciens*, at 1.64 Å resolution. CotB2 is a diterpene cyclase that catalyzes the cyclization of the linear geranylgeranyl diphosphate to the tricyclic cyclooctat-9-en-7-ol. The subsequent oxidation of cyclooctat-9-en-7-ol by two cytochrome P450 monooxygenases leads to bioactive cyclooctatin. Plasticity residues that decorate the active site of CotB2 have been mutated, resulting in alternative monocyclic, dicyclic and tricyclic compounds that show bioactivity. These new compounds shed new light on diterpene cyclase reaction mechanisms. Furthermore, the product of mutant CotB2^{W288G} produced the new antibiotic compound (1*R*,3*E*,7*E*,11*S*,12*S*)-3,7,18-dolabellatriene, which acts specifically against multidrug-resistant *Staphylococcus aureus*. This opens a sustainable route for the industrial-scale production of this bioactive compound.

1. Introduction

Terpene molecules represent one of the most diverse groups of natural biomolecules (Christianson, 2006, 2008; Maimone & Baran, 2007). The initial step in terpene biosynthesis is the stepwise condensation of isopentyl diphosphate and dimethylallyl diphosphate (Poulter *et al.*, 1978) to the linear terpene cyclase substrates geranyl diphosphate (C_{10}), farnesyl diphosphate (C_{15}), geranylgeranyl diphosphate (C_{20}) and geranylfarnesyl diphosphate (C_{25}), leading to monoterpenes, sesquiterpenes and diterpenes (Christianson, 2006, 2008; Maimone & Baran, 2007; Sacchettini & Poulter, 1997; Cane, 1990). This reaction is catalyzed by isoprenyl diphosphate synthases. The subsequent cyclization reaction is performed by various terpene cyclases, using the linear isoprene precursors to build up a vast number of polycyclic hydrocarbon scaffolds. Remarkably, this complex chemical reaction, comprising changes in bonding, hybridization and the introduction of specific stereochemistry, is performed in a single step. The cyclization reaction can be subdivided into three steps: (i) initial generation of a reactive carbocation, (ii) propagation of the carbocation through the substrate, leading to hydride and/or methyl shifts, deprotonation and reprotonation and ionization of intermediates as well as the creation

of a terminal carbocation, and (iii) quenching of the carbocation by a base. Terpene cyclases can be divided into two distinct classes, which are distinguished by their substrate-activation mechanism. Class I terpene cyclases generate an allylic carbocation by the release of pyrophosphate. Members of this class contain a DD*XX*D motif (Ashby & Edwards, 1990) that binds the diphosphate function of the respective substrate with the support of Mg^{2+} ions. In contrast, class II terpene cyclases generate the carbocation by protonation of the terminal isoprene unit, involving a D*X*DD motif. Besides the differences in the activation mechanism, the two classes of terpene cyclases have an unrelated overall fold. Class I terpene cyclases establish the fold of class I terpene synthases (Wendt & Schulz, 1998) and are structurally related to isoprenyl diphosphate synthases such as farnesyl diphophate synthase. In contrast, the class II terpene cyclases have a multi-domain architecture frequently composed of two purely α-helical domains (β and γ).

Fusicoccane diterpenes feature a 5–8–5 ring motif and, depending on the functionalization of the ring system, they demonstrate a broad diversity of biological activities, incuding fungicidal and tumorstatic activities, amongst others (de Boer & de Vries-van Leeuwen, 2012; Toyomasu *et al.*, 2007; Subbarao *et al.*, 2009; Rasoamiaranjanahary *et al.*, 2003). Cyclooctatin is a member of the fusicoccane family and shows anti-inflammatory activity. In contrast to the anti-inflammatory agents derived from either aspirin or ibuprofen, which target cyclooxygenase type I or II, cyclooctatin inhibits a lysophospholipase which catalyzes the hydrolysis of the fatty-acid ester bonds of lysophospholipids (Kim *et al.*, 2009; Aoyagi *et al.*, 1992). This lysophospholipase is upregulated in eosinophilic leukocytes that occur in allergic reactions and inflammatory diseases.

The cyclooctatin gene cluster from the soil bacterium *Streptomyces melanosporofaciens* MI614-43F2 was recently characterized, comprising a GGDP synthase (CotB1), a diterpene cyclase (CotB2) and two P450 cytochromes (Kim *et al.*, 2009). CotB2 catalyzes the cyclization of GGDP to cyclooctat-9-en-7-ol, which is further functionalized by the two P450 cytochromes, CotB3 and CotB4, to cyclooctatin.

Polycyclic terpenes are chemically difficult to access owing to their complex stereochemistry. For instance, chemical synthesis of the structurally related compound cotylenol involves at least 12 steps, partly resulting in racemic intermediates and low product yields (Kato *et al.*, 1996). Organic total synthesis of novel bioactive diterpenes remains time-intensive and costly and is accompanied by the production of toxic waste streams. An elegant solution is the biotechnological application of terpene cyclase-based biocatalysts that offer a highly stereospecific, single-step route for the biosynthesis of complex macrocycles with potentially high yields and no toxic side streams. Plant-based engineering efforts are laborious and therefore bacterial terpene cyclases are of particular interest since they can be easily produced and engineered in a bacterial standard expression system such as *Escherichia coli*. Metabolic pathways consist of an upstream isoprene pathway and a heterologous downstream element (Brück *et al.*, 2014). In the upstream element the linear educts for a downstream diterpene synthase are produced. In the second conversion step of the downstream element the diterpene diphosphate is enzymatically cyclized by a terpene cyclase to a macrocyclic olefin scaffold. Successful examples of this microbial production platform have been demonstrated for taxadien-5α-ol, a precursor of taxol (paclitaxel), a highly potent anticancer drug (Ajikumar *et al.*, 2010), and for artemisinic acid, a precursor of the antimalarial drug artemisinin (Paddon *et al.*, 2013).

Currently, structural information on diterpene cyclases is limited to plant and fungal enzymes such as taxadiene synthase (TXS; Köksal, Jin *et al.*, 2011), *ent*-copalyldiphosphate synthase (Köksal, Hu *et al.*, 2011) and abietadiene synthase (Zhou *et al.*, 2012), whereas structural information is elusive for the bacterial counterparts. Here, we report the first high-resolution crystal structure of the bacterial diterpene cyclase CotB2 from the soil bacterium *S. melanosporofaciens* and demonstrate that novel olefinic macrocycles with a broad diversity and bioactivity can be generated by protein engineering of the active site.

2. Materials and methods

2.1. Protein expression and purification

Codon-optimized wild-type CotB2 ($CotB2^{wt}$) with *Nde*I and *Xho*I restriction sites was fused to a C-terminal hexahistidine tag in pET-24a (Görner *et al.*, 2013) vector and transformed into *E. coli* Rosetta2 DE3 cells. Overexpression was performed using auto-induction medium at 37°C until an OD of 0.7 was reached, and the medium was subsequently cooled to 18°C (Studier, 2005). The cells were grown overnight and were harvested by centrifugation (6 min, 6000 rev min^{-1} at 4°C). For resuspension of the cell pellets, buffer *A* (50 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 2 m*M* $MgCl_2$, 1 m*M* DTT) was used. The cells were lysed by homogenization at 4°C and the lysate was cleared by centrifugation (1 h, 21 000 rev min^{-1} at 4°C). An Ni^{2+}–NTA (~1 ml column volume; GE Healthcare) column was equilibrated with buffer *A*. $CotB2^{wt}$ was loaded onto the column and washed with ten column volumes of buffer *A*. $CotB2^{wt}$ was eluted in a linear gradient to 50 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 2 m*M* $MgCl_2$, 1 m*M* DTT, 500 m*M* imidazole. Size-exclusion chromatography was performed with a HiLoad Superdex S200 16/60 column (GE Healthcare) equilibrated with buffer *B* (20 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl, 2 m*M* $MgCl_2$, 1 m*M* DTT). Pooled protein fractions were concentrated with Amicon Ultra 30 000. Calibration runs were performed with a high-molecular-weight standard (GE Healthcare). The $CotB2^{F149L}$ variant was purified as described for $CotB2^{wt}$. For determination of experimental phases, selenomethionine was incorporated into $CotB2^{wt}$. Transformed *E. coli* Rosetta2 DE3 cells were cultured in minimal medium containing selenomethionine (Van Duyne *et al.*, 1993) and protein expression was induced at an OD of ~0.6 by the addition of 0.5 m*M* IPTG. Purification was carried out as described above for $CotB2^{wt}$, except that all buffers contained 4 m*M* DTT.

2.2. Dynamic light scattering

Dynamic light-scattering experiments were performed with a Spectroscatter 201 instrument (RiNA) at 20°C. Prior to the experiments, the protein samples were centrifuged (13 000 rev min^{-1}, 10 min, 4°C). The protein concentration was 27.2 mg ml^{-1}. For each sample, 25 individual experiments with a time course of 3 s were recorded. Data were analyzed with the program *Spectro* supplied by the manufacturer.

2.3. Mutagenesis and product production

Saturation mutagenesis of Trp288 was performed as described previously (Görner *et al.*, 2013). The mutagenesis primers were 5′-CTGAT TTATG GCAAT TTTGT GNNNA CCACC TCCAA CAAAC GTTAT AAAAC-3′ and 5′-GTTTT ATAAC GTTTG TTGGA GGTGG TNNNC ACAAA ATTGC CATAA ATCAG-3′. In order to produce sufficient amounts for the GC-MS characterization of CotB2^{W288G}, 1 mg purified enzyme was added to 10 ml 50 m*M* Tris–HCl pH 7.5, 1 m*M* $MgCl_2$ supplemented with 45 µ*M* GGDP (Sigma–Aldrich) and incubated for 12 h at 30°C with mild agitation. The product was extracted using 5 ml ethyl acetate, dried with $MgSO_4$ and concentrated to 100 ml for GC-MS analysis.

2.4. Product identification

The diterpene products of CotB2 variants were analyzed by a Trace GC Ultra with a DSQ II (Thermo Scientific). A sample volume of 1 µl was applied by TriPlus AS onto an SGE BPX5 column (30 m, internal diameter 0.25 mm, film 0.25 µm). The initial column temperature was 50°C (maintained for 2.5 min). A temperature gradient was applied from 50 to 320°C (10°C min^{-1}) followed by 3 min maintenance at 320°C. Mass-spectrometric data were recorded at 70 eV (EI), *m*/*z* (relative intensity in %) as TIC (total ion current). The recorded *m*/*z* range was 50–650. Mass-spectrometric fragmentation patterns were compared against the NIST NIST08MS spectral database library containing a special natural product extension (Adams, 2007).

All NMR spectra were recorded in $CDCl_3$ with an Avance 500 MHz (Bruker) at 300 K. The chemical shifts are given in p.p.m. relative to $CHCl_3$ at δ = 7.26 p.p.m. (^{1}H NMR) or $CDCl_3$ at δ = 77.16 p.p.m. (^{13}C NMR); coupling constants (*J*) are given in hertz. Optical rotations were measured using a PerkinElmer 241MC polarimeter.

2.5. Crystallization and crystal cooling

For crystallization experiments, CotB2^{wt} was concentrated to 27.2 mg ml^{-1} and CotB2^{F149L} to 26.4 mg ml^{-1} as measured by the absorbance at 280 nm. Crystals were obtained by the sitting-drop vapor-diffusion method at 18 °C with a reservoir solution composed of 20%(*v*/*v*) polyethylene glycol 8000, 100 m*M* 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) pH 10.0. CotB2^{F149L} was crystallized using a reservoir solution composed of 15%(*v*/*v*) polyethylene glycol 8000, 60 m*M* CAPS pH 10.0, 2%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	SeMet CotB2^{wt}	CotB2^{wt}	CotB2^{F149L}
PDB entry	—	4omg	4omh
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Wavelength (Å)	0.9757	0.9763	0.9763
Unit-cell parameters			
a (Å)	59.3	59.1	59.5
b (Å)	101.9	100.7	98.4
c (Å)	108.9	108.7	107.8
$\alpha = \beta = \gamma$ (°)	90.0	90.0	90.0
Resolution (Å)	30.00–2.50 (2.57–2.50)	50.00–1.64 (1.68–1.64)	50.00–1.64 (1.68–1.64)
Unique reflections	43730 (3202)	79419 (5765)	77089 (5552)
Completeness	99.2 (98.0)	99.2 (98.8)	98.6 (97.5)
$\langle I/\sigma(I)\rangle$	8.1 (2.9)	17.5 (2.6)	15.8 (2.8)
R_{meas}†	0.191 (0.805)	0.068 (1.143)	0.082 (1.079)
$CC_{1/2}$	99.6 (92.8)	99.9 (76.3)	99.8 (72.5)
Multiplicity	7.7 (7.5)	6.5 (6.7)	6.6 (6.8)
Refinement			
Non-H atoms		5066	5149
R_{work}‡		0.173 (0.259)	0.183 (0.227)
R_{free}§		0.200 (0.296)	0.216 (0.264)
No. of protein chains		2	2
No. of protein residues		559	562
No. of buffer molecules		2	2
No. of water molecules		268	303
Average *B* factor (Å^2)			
Overall		24.0	23.7
Protein residues		23.5	23.0
Buffer molecules		46.7	55.10
Water molecules		30.0	30.6
R.m.s.d.¶, bond lengths (Å)		0.013	0.012
R.m.s.d.¶, bond angles (°)		1.42	1.44
Ramachandran outliers (%)		0	0
Ramachandran favored (%)		99.15	99.16

† $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of symmetry-equivalent reflections and $N(hkl)$ is the redundancy. ‡ $R_{work} = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$ (working set, no σ cutoff applied). § R_{free} is the same as R_{work} but calculated on 5% of the data that were excluded from refinement. ¶ Root-mean-square deviation from target geometries.

Crystals of both CotB2^{wt} and CotB2^{F149L} appeared after two weeks. Selenomethionine-labeled CotB2^{wt} was concentrated to 26.1 mg ml^{-1}. Crystals were obtained by the sitting-drop vapor-diffusion method at 18°C with a reservoir solution composed of 20%(*v*/*v*) polyethylene glycol 3350, 200 m*M* NaH_2PO_4. All crystals were cryoprotected by increasing the MPD concentration of the reservoir resolution to 25%(*v*/*v*) and were subsequently flash-cooled in liquid nitrogen.

2.6. X-ray data collection, structure determination and refinement

Synchrotron diffraction data were collected on beamline 14.2 of the MX Joint Berlin laboratory at BESSY (Berlin, Germany) or beamline P14 of PETRA III (Deutsches Elektronen Synchrotron, Hamburg, Germany). X-ray data collection was performed at 100 K. Diffraction data were processed with *XDS* (Kabsch, 2010; Table 1). Experimental phases were determined by single anomalous dispersion with *SHARP* (Vonrhein *et al.*, 2007) using the selenomethionine-labeled

CotB2^{wt} data set (Supplementary Table S1). The structure of CotB2^{F149L} was determined by isomorphous replacement. For calculation of the free *R* factor, a randomly generated set of 5% of the reflections from the diffraction data set was used and was excluded from the refinement. An initial model of CotB2^{wt} was built with *ARP*/*wARP* (Langer *et al.*, 2008). The structure was initially refined by applying a simulated-annealing protocol and in later refinement cycles by maximum-likelihood restrained refinement using *PHENIX* (Adams *et al.*, 2010; Afonine *et al.*, 2012) including TLS refinement with each polypeptide chain as an individual TLS group (Winn *et al.*, 2001). In initial rounds of refinement NCS restraints were applied, which were completely released in the final stages of refinement. Model building and water picking was performed with *Coot* (Emsley *et al.*, 2010). Model quality was evaluated with *MolProbity* (Chen *et al.*, 2010) and *PROCHECK* (Laskowski *et al.*, 1993). Secondary-structure elements were assigned with *DSSP* (Kabsch & Sander, 1983), and *ALSCRIPT* (Barton, 1993) was used for secondary-structure-based sequence alignments. Figures were prepared using *PyMOL* (DeLano, 2002).

3. Results

3.1. Overall structure

To unravel the architecture of CotB2 from *S. melanosporofaciens*, which catalyzes the cyclization of linear GGDP to cyclooctat-9-en-7-ol (Fig. 1*a*), we overexpressed CotB2 in *E. coli* and subsequently purified and crystallized the protein. Initially, we attempted to solve the structure by molecular replacement using the coordinates of the α-domain of TXS (Köksal, Jin *et al.*, 2011), since primary-sequence comparison showed that the α-domain of TXS is the closest relative to CotB2 for which coordinates are available, with a sequence identity of 9.5%. Since we could not obtain a molecular-replacement solution, we altered our search model by deletion of the flexible parts and truncation to a polyalanine model, but this approach was not successful either. Next, we used different structures of sesquiterpene synthases, such as trichodiene synthase (Rynkiewicz *et al.*, 2001), epi-isozizaene synthase (Aaron *et al.*, 2010), pentalene synthase (Lesburg *et al.*, 1997) and aristolochene synthase (Shishova *et al.*, 2007), as search models, but were unsuccessful. The failure of the molecular-replacement approach is most likely to be owing to modulation of the secondary-structure elements, leading to differences in packing/tilting of the core α-helices and making molecular-replacement attempts difficult. Notably, most of the terpene synthase structures deposited in the PDB have been solved by experimental phasing. Additional burdens for the molecular-replacement approach are conformational changes upon the binding of metal ions and substrate that lead to a closure of the active site accompanied by changes in the tilting of flanking α-helices. For instance, the unliganded and ligand-bound form of fungal aristolochene synthase differ by a root-mean-square deviation (r.m.s.d.) of 2.11 Å (Shishova *et al.*, 2007). Since all molecular-replacement attempts failed, we prepared selenomethionine-labeled CotB2^{wt} and solved the high-resolution structure by single anomalous dispersion to 1.64 Å resolution in space group $P2_12_12_1$ (Table 1 and Supplementary Table S1). The asymmetric unit contains two CotB2 polypeptide chains that are practically indistinguishable, with an r.m.s.d. of 0.14 Å for 274 C^{α} pairs. The electron density is of excellent quality, which not only allows us to trace the polypeptide chain but also to assign one sodium ion and one buffer molecule per CotB2 monomer. CotB2 adopts the classical α-helical bundle fold of class I terpene synthases (Wendt & Schulz, 1998; Fig. 1*b*). Our model is complete except for the 15 N-terminal and ten C-terminal residues. CotB2 consists of ten core α-helices (A–J) that are connected by short loop segments and an additional five short α-helices (α1–α5; Fig. 1*b*). The core α-helices surround a large, deep cleft: the active site (Fig. 1*b*). Since we have crystallized CotB2 in the absence of substrate or substrate analog, we describe here the open state of the enzyme.

The two polypeptide chains in the asymmetric unit of CotB2 are arranged as homodimers (Supplementary Fig. S1), with the dimer interface being established by residues of α-helices F,

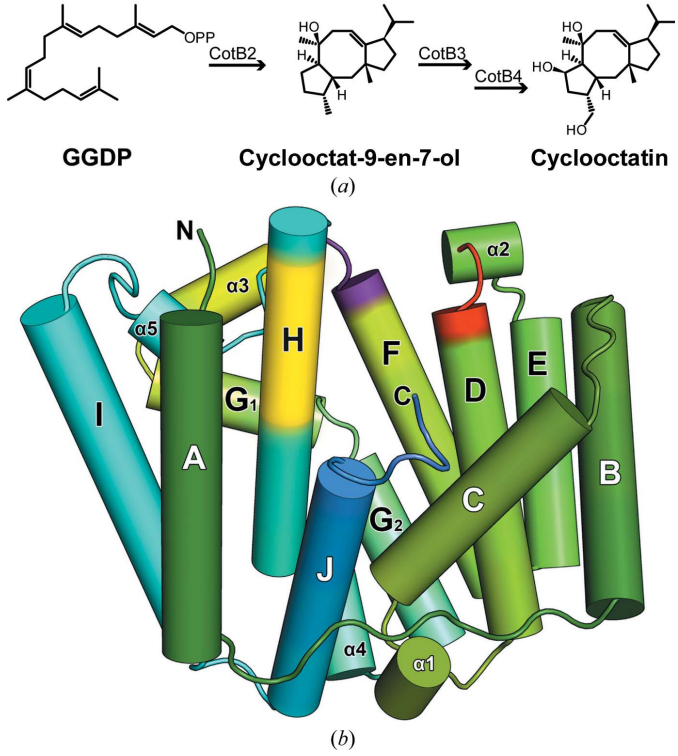


Figure 1
Cyclooctatin biosynthesis and structure of CotB2. (*a*) Chemical structures are given for the educt and product of CotB2. Cyclooctat-9-en-7-ol is further functionalized by CotB3 and CotB4, two cytochrome P450 enzymes, to yield cyclooctatin. (*b*) One monomer of CotB2 is shown colored in a gradient from green (N-terminus) to blue (C-terminus). α-Helices are drawn as cylinders. The location of the aspartate-rich motif is indicated in red and the NSE/DTE motif is marked in yellow. The double conformation at the C-terminal end of α-helix F is indicated in purple.

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: RR5067).

α3, G_1, G_2, α4, H and I. About 1477 Å^2 of solvent-accessible surface area is buried upon dimerization, and the *PISA* server (Krissinel & Henrick, 2007) suggests the presence of a stable dimer in solution. To further elaborate whether the dimer exists in solution as well, we performed size-exclusion chromatography and dynamic light scattering. SEC revealed a molecular weight of 72 kDa and DLS a molecular weight of 86 kDa (Supplementary Fig. S2) indicative of the presence of dimeric CotB2 in solution. The two active sites of CotB2 are arranged in an antiparallel fashion, resembling the arrangement initially observed for the monoterpene (+)-bornyl diphosphate synthase (Whittington *et al.*, 2002) and the sesquiterpene trichodiene synthase (Rynkiewicz *et al.*, 2001), but is in contrast to the parallel dimer described for farnesyl diphosphate synthase (Tarshis *et al.*, 1994). Using a *DALI* search (Holm & Rosenström, 2010) with the coordinates of CotB2, we tentatively expected to find the known plant structures of diterpene cyclases, but surprisingly the structurally closest neighbors are monoterpene and sesquiterpene cyclases. The structurally closest neighbors to be found were the monoterpene synthase 4*S*-limonene (PDB entry 2ong; Hyatt *et al.*, 2007) as well as the sesquiterpene cyclases trichodiene (PDB entry 2ps5; Vedula *et al.*, 2008), aristolochene (PDB entry 2oa6; Shishova *et al.*, 2007) and epi-isoziziaene (PDB entry 3lg5; Aaron *et al.*, 2010), all with a comparable r.m.s.d. of 3.3–3.6 Å.

3.2. Active site

The 'aspartate-rich' motif is located at the C-terminal end of α-helix D and is supposed to bind the diphosphate moiety of GGDP *via* Mg^{2+} ions (Figs. 1*b* and Supplementary Fig. S3). Strikingly, instead of the classical DD*XX*D motif of class I terpene synthases, CotB2 contains an altered ^{110}DD*X*D^{113} motif characteristic of neither class I nor class II terpene synthases. Whereas in all other class I terpene cyclases the aspartate-rich motif resides at the C-terminal end of α-helix D, in CotB2 α-helix D is shorter and a proline residue adjacent to the third aspartate of the aspartate-rich motif introduces a kink. Consequently, Asp113 points to the opposing side away from the active site. The 'NSE/DTE' motif (Cane & Kang, 2000) ^{218}IVNDFYSYDRE^{228} resides in α-helix H (Supplementary Fig. S4). Both motifs are considered to be important for the ionization-dependent step in the cyclization reaction and for the binding of Mg^{2+} ions. Presumably, the motifs trigger ionic cleavage of the C—O bond in the GGDP substrate. The floor of the active site is decorated by hydrophobic residues (Val80, Phe107, Trp109, Phe149, Val150, Ile181, Phe185, Met189, Trp186, Leu281 and Trp288), conferring the overall shape to the cavity that serves as a template for the binding of GGDP and the cyclization reaction. At the upper end of the pocket the first two aspartate residues of the aspartate-rich motif and residues of the NSE/DTE motif point into the active site. Upon binding of GGDP the active site closes to prevent the emerging reactive carbocations from being quenched by the bulk solvent. GGDP is stabilized by Mg^{2+} ions and by positively charged amino-acid residues. In the structure of CotB2, the basic residues Arg177 and Arg227 protrude into the pocket, most likely providing charge compensation for the diphosphate group of GGDP. The finite C-terminus is unstructured and not resolved in the electron density. Residue Arg294 is located in this unstructured C-terminal end in our modeled CotB2 structure. We suggest that upon binding of the catalytically obligatory Mg^{2+} ions and the GGDP substrate, the active site closes and could bring the guanidinium function of Arg294 into close proximity to the diphosphate function of GGDP. In all of our structures we observe a double conformation at the C-terminal end of helix F (^{155}MFRD^{158}), reflected by a double conformation of the protein backbone (Fig. 1*b*) as well as the amino-acid side-chain rotamers. The two backbone conformations are distinguished by the resulting length of the α-helix, which is longer or shorter by one amino-acid residue. It is interesting to note that the secondary-structure elements close to the active site display such plasticity and this might explain the observation that the so-called secondary layer residues in the vicinity of the active site can influence the dynamic properties of the active-site residues (Greenhagen *et al.*, 2006).

Table 2
Single amino-acid exchange of CotB2 altering the product of the cyclization reaction.

Mutation to proline, arginine, aspartate, glutamate, threonine, valine or isoleucine at position 107 yielded no detectable product formation.

Residue	Variant	Product	Reference
CotB2^{wt}	—	Cyclooctat-9-en-7-ol	Kim *et al.* (2009)
Asp110	Glutamate	Cyclooctat-9-en-7-ol	This study
Asp111	Glutamate	No product	This study
Asp113	Glutamate	Cyclooctat-9-en-7-ol	This study
Phe107	Glycine	*R*-Cembrene A	This study
	Alanine	*R*-Cembrene A	This study
	Leucine	Cyclooctat-9-en-7-ol isomers	Görner *et al.* (2013)
	Tyrosine	Cyclooctat-1,7-diene	Görner *et al.* (2013)
Phe149	Histidine	Cyclooctat-7-en-3-ol	Görner *et al.* (2013)
	Leucine	Cyclooctat-7-en-3-ol	Görner *et al.* (2013)
	Valine	Cyclooctat-7-en-3-ol	Görner *et al.* (2013)
	Glycine	Cyclooctat-7-en-3-ol	Görner *et al.* (2013)
	Tyrosine	No product	Görner *et al.* (2013)
Trp288	Glycine	Dolabellatriene	This study

3.3. Plasticity of the residues in the active site of CotB2

In previous work, we constructed a homology model of CotB2 (Görner *et al.*, 2013) that was based on the structure of the α-domain of taxadiene synthase (TXS; Köksal, Jin *et al.*, 2011). This homology model enabled us to identify single amino-acid residues that are potentially important for carbocation migration during the cyclization reaction. Mutational analysis of these residues revealed alternative products, as summarized in Table 2. The substitution of Phe149 by leucine led to cyclooctat-7-en-3-ol (Fig. 2*a*), with migration of the double bond from C9/C10 to C7/C8 and the hydroxyl moving from position C7 to C3 (Görner *et al.*, 2013). Interestingly, the Phe149 position tolerates other substitutions as well, all of which lead to the formation of cyclooctat-7-en-3-ol (Table 2).

To study the structural integrity of CotB2 variants in comparison to CotB2^{wt} and to analyze whether our mutations

would have an effect on the protein surroundings of the active site, we crystallized the variant CotB2^{F149L}. Phases for CotB2^{F149L} were obtained by Fourier synthesis with the refined structure of CotB2^{wt}. The structure was refined to 1.64 Å resolution (Table 1). The variant CotB2^{F149L} superimposes with CotB2^{wt} with an r.m.s.d. of 0.3 Å for 276 C$^{\alpha}$ pairs. Major differences are observed for both termini that are better resolved in the electron density in the case of CotB2^{F149L}. Residue 149 is located in α-helix F (Supplementary Fig. S3). The side chains of Phe149 and Leu149 occupy identical positions. Differences are observed at different locations in the active-site pocket (Supplementary Fig. S5). Whereas in CotB2^{wt} the hydroxyl function of Ser152 occupies different rotamers, we observe only a single rotamer conformation in CotB2^{F149L}. Consequently, the number of waters in and the position of the surrounding water network are altered. Asn220 of the NSE/DTE motif, which is likely to be involved in Mg^{2+} coordination, adopts two side-chain conformations in CotB2^{F149L} but only one in CotB2^{wt}. It is also interesting to note that in CotB2^{F149L} the aromatic rings of Tyr77, Phe107 and Trp288 are slightly rotated towards the active site and hence reduce the active-site volume. This further extends to the N-terminal portion of α-helix C, which is translated by 0.5 Å towards α-helix D. In general, the differences are more pronounced in one of the monomers.

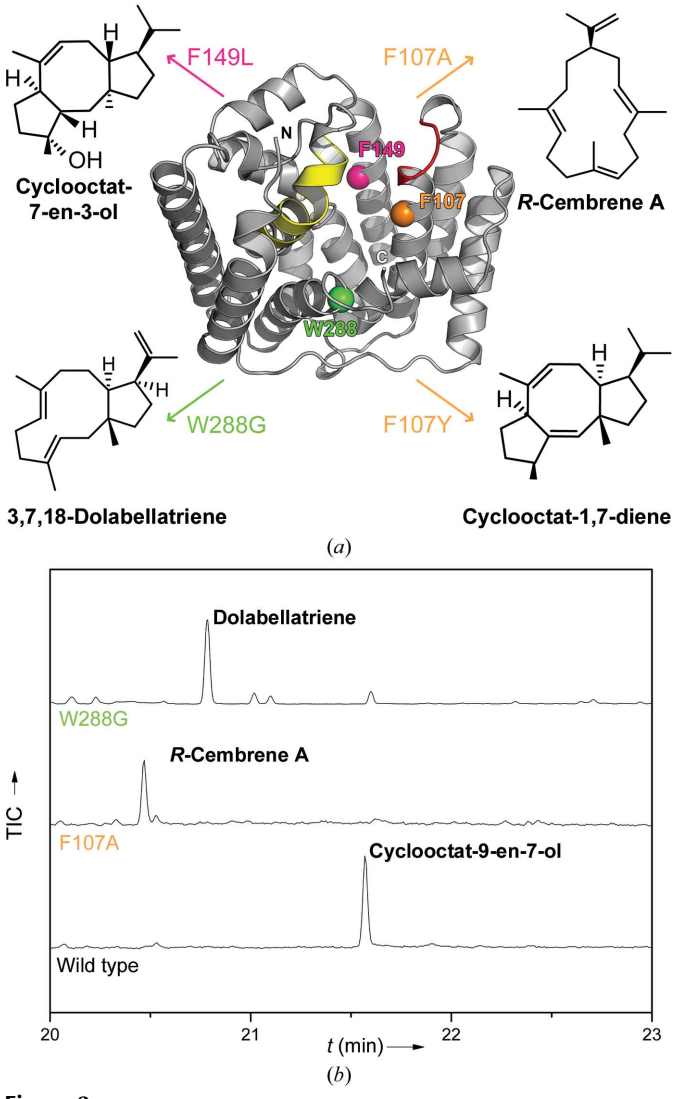


Figure 2
Overview of CotB2 variants and product outcome. (*a*) The structure of CotB2 is shown in cartoon representation. The metal-binding motifs are highlighted in red and yellow. Mutants of CotB2 are indicated by colored spheres: Phe149 in pink, Phe107 in orange and Trp288 in green. The different products generated by the respective mutants are depicted by color-coded arrows. Chemical structures are shown for the different products. (*b*) GC-MS analysis of different CotB2 products. The retention time for CotB2^{wt} is 21.60 min (cyclooctat-7-en-3-ol), that for CotB2^{F107A} is 20.47 min (*R*-cembrene A) and that for CotB2^{W288G} is 20.78 min (dolabellatriene). Mass-spectrometric analysis of the dolabellatriene product of CotB2^{W288G} is shown in Supplementary Fig. S6.

Mutational studies at position Phe107, located in α-helix F, revealed less tolerance to mutagenesis in contrast to position 149 (Table 1). Replacement by an aromatic tyrosine residue retained the 5–8–5 ring structure but led to the formation of a diene and elimination of the hydroxyl function (Table 2 and Fig. 2*a*). Replacement by alanine led to the formation of a cembrene-like diterpene (Görner *et al.*, 2013). Until now, we could not identify it using our diterpene production system. Using an *in vitro* conversion of GGDP by CotB2^{F107A}, we were now able to identify the structure by mass spectrometry (Supplementary Fig. S6), ^{1}H NMR and the specific rotation (Vanderah *et al.*, 1978; Schwabe *et al.*, 1988) as *R*-cembrene A (Fig. 2*b*). The proton chemical shifts revealed the characteristic olefinic proton signals at δ = 4.74 (s, 1H) and 4.68 (s, 1H) as singlets as well as triplet signals at δ = 5.22 [t, ^{3}J(H,H) = 7.6 Hz, 1H], 5.00 [t, ^{3}J(H,H) = 6.2 Hz, 1H] and 5.08 [t, ^{3}J(H,H) = 6.2 Hz, 1H]. Comparison of the remaining proton signals shows less than 0.02 p.p.m. divergence from the literature values (Bai & Jain, 2008). The CotB2^{F107A} variant results in a change in the shape and the volume of the active-site pocket. Consequently, the GGDP can fold along an altered pathway leading to *R*-cembrene A. Cembrene molecules are natural compounds with cytotoxic, pheromonal and antimicrobial activities (Bonnard *et al.*, 2010; Casanova *et al.*, 2006; Tietze *et al.*, 2008). Hence, compounds with a cembrene scaffold or a cembrene scaffold that is hydroxylated by downstream cytochrome P450 monooxygenases could potentially lead to the development of novel antibiotics.

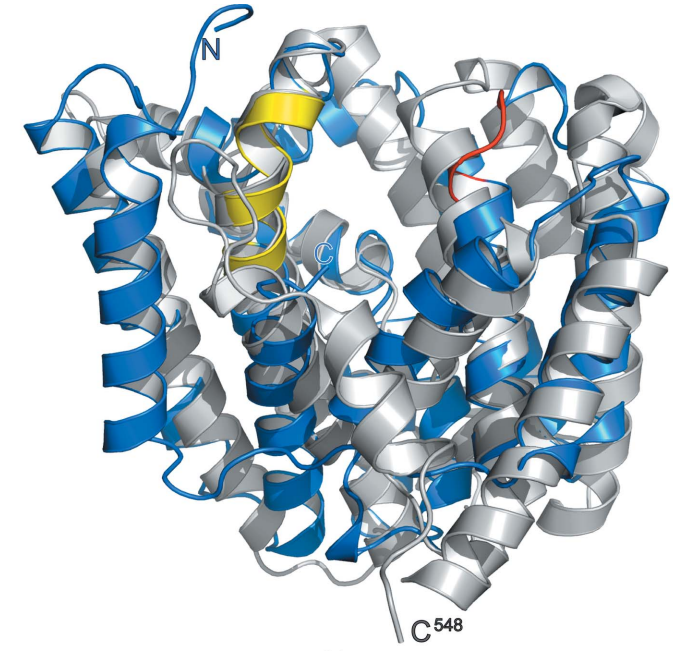
3.4. Comparison with taxadiene cyclase

To date, structural information is only available for three diterpine cyclases (Zhou *et al.*, 2012; Köksal, Hu *et al.*, 2011; Köksal, Jin *et al.*, 2011). TXS is the most attractive candidate for comparison since it has only a single class I terpene synthase activity. Furthermore, TXS and CotB2 both catalyze the cyclization of GGDP to tricyclic products: taxa-4,11-diene with a 6–8–6 ring motif and cyclooctat-9-en-7-ol with a 5–8–5 ring scaffold, respectively. The α-domains of TXS (Köksal, Jin *et al.*, 2011) and CotB2 share a sequence identity of 9.5% and superimpose with an r.m.s.d. of 2.9 Å for 199 pairs of C$^{\alpha}$ atoms (Fig. 3*a* and Supplementary Fig. S3). The overall arrangement of the core α-helices of both proteins is comparable (Fig. 3*a*

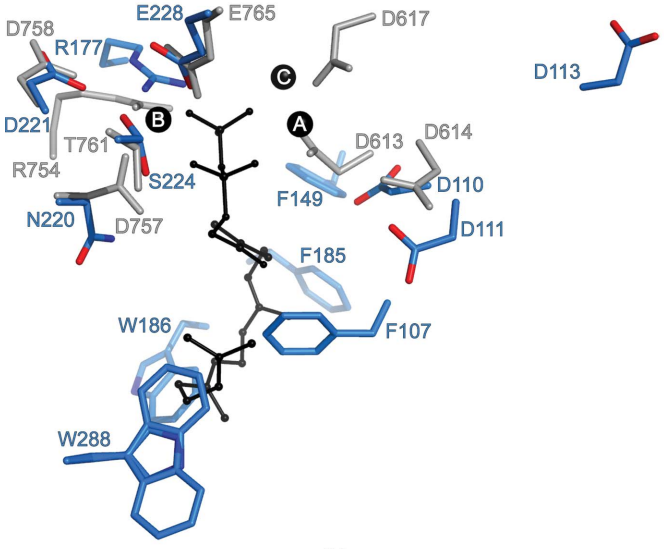
and Supplementary Fig. S3). Differences are restricted to the lengths of the loop regions. Major differences are observed at both the N-terminus and the C-terminus. In comparison to TXS, CotB2 has an additional N-terminal α-helix which is absent in TXS (Fig. 3*a*). As a result, the N-terminus of CotB2 is in close proximity to the active site, and the missing 15 N-terminal residues in our crystal structure might fold upon substrate binding and thereby close and shield the active site.

Beside these differences, a number of residues within the active site of the α-domain of TXS and CotB2 superimpose very well. Hence, the structure of TXS is a very valuable template to postulate catalytically important residues of CotB2. The first two aspartate residues of the aspartate-rich motif of TXS (^{613}DD*XX*D^{617}) and ^{110}DD*X*D^{113} of CotB2 superimpose. Therefore, it seems likely that Asp110 of CotB2 directly coordinates Mg_A^{2+} as well as Mg_C^{2+} and Asp111 of CotB2 possibly coordinates Mg_C^{2+} indirectly *via* a water molecule. To further elaborate the function of the residues within the aspartate-rich motif, we performed site-directed mutagenesis. Mutation of the two flanking aspartate residues did not interfere with product formation, whereas enzymatic activity was lost for the CotB2^{D111E} mutation (Table 2). Asp111 of CotB2 might not only be involved in Mg^{2+} coordination, since it is hydrogen bonded with its carboxylate function to the hydroxyl of Tyr77 (2.7 Å) in α-helix C and seems to be rather important for protein stabilization.

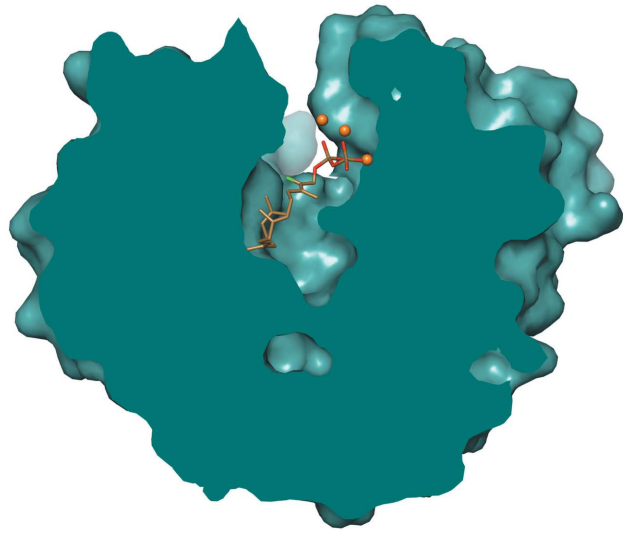
A major difference is observed at the C-terminal end of α-helix D harboring the aspartate-rich motif, which is shorter in CotB2; consequently, the third aspartates of the aspartate-rich motifs do not superimpose (Fig. 3 and Supplementary Fig. S3), with Asp113 of CotB2^{D113} being solvent-exposed. Hence, it seems unlikely that Asp113 is involved in metal binding. In the bacterial sesquiterpene synthase epi-isozizaene, the third aspartate of the aspartate-rich motif is also not involved in metal binding (Aaron *et al.*, 2010). Strikingly, for epi-isozizaene synthase it has been shown that the first aspartate residue is crucial for activity (Lin & Cane, 2009). Residues Asn757, Asp758, Thr761 and Glu765 of the NSE/DTE motif in TXS, which are involved in the direct coordination of Mg_B^{2+}, superimpose with Asn220, Asp221, Ser224 and Glu229 of CotB2 (Fig. 3*b*). Beside the metal coordination, we can also derive information on possible interactions with the dipho-



(*a*)



(*b*)



(*c*)

Figure 3
Superposition of CotB2^{wt} and the α-domain of TXS (PDB entry 3p5r). (*a*) CotB2^{wt} is drawn in blue, with its aspartate-rich motif indicated in red and the NSE/DTE motif in yellow. The α-domain of TXS (residues 548–900) is depicted in grey. (*b*) Residues of the aspartate-rich and the NSE/DTE motif as well as residues involved in the stabilization of the diphosphate function of the substrate are shown. Amino-acid side chains of CotB2^{wt} are depicted in blue and those of TXS are depicted in grey. For TXS the bound 2-fluorogeranylgeranyl diphosphate is drawn in stick representation and the three bound Mg^{2+} ions are shown as black spheres. (*c*) Surface representation of CotB2 (dark teal). The view is vertically rotated 180° compared with that in (*a*). Section through the active site. The 2-fluorogeranylgeranyl diphosphate is shown in stick representation and Mg^{2+} ions are shown as orange spheres as observed in the crystal structure of TXS. Its position was obtained by a superposition of CotB2 and TXS.

sphate function of GGDP. The guanidinium function of Arg754 of TXS and Arg177 of CotB2 superimpose (Fig. 3*b*), even though they reside on different core α-helices. Notably, the recognition and the mechanism of diphosphate activation seem to be conserved between plant and bacterial diterpene and monoterpene cyclases.

Superposition of the bound 2-fluorogeranylgeranyl diphosphate and Mg^{2+} ions, as modeled in the structure of TXS, on the structure of CotB2 reveals sterical clashes of CotB2 amino-acid side chains with 2-fluorogeranylgeranyl diphosphate. These clashes are observed at the bottom of the active site (Figs. 3*b* and 3*c*) and are caused by the aromatic side chain of Phe107 and the indole of Trp288 located at the very C-terminal end of α-helix J of CotB2. Consequently, the volume of the active site of CotB2 is decreased compared with TXS.

3.5. Importance of CotB2^{W288G} for carbocation migration

Based on our structural observations, we investigated using saturation mutagenesis whether Trp288 is essential for the cyclization process. Interestingly, the variant CotB2^{W288G} showed the formation of a new, unexpected product. Our *E. coli* diterpene production system was able to produce 1.7 mg of this compound per litre (Görner *et al.*, 2013). Analysis by GC-MS (Supplementary Fig. S6) followed by detailed NMR analysis revealed the formation of a bicyclic dolabellane carbon skeleton (Fig. 2). The NMR data of the ^{1}H and ^{13}C spectra diverged by less than 0.2 p.p.m. (^{13}C) and 0.02 p.p.m. (^{1}H) from the literature values (Ioannou *et al.*, 2011). Furthermore, NOESY NMR correlations and the specific rotation were in very good agreement with the literature (Ioannou *et al.*, 2011). The data are consistent with the biocatalytic formation of (1*R*,3*E*,7*E*,11*S*,12*S*)-3,7,18-dolabellatriene (Fig. 4). The formation of a dolabellane can be explained by an early proton elimination within the proposed reaction mechanism for cyclooctat-9-en-7-ol (Kim *et al.*, 2009; Fig. 4). After ionization of GGDP, a two-ring system with a carbocation at C15 (**1**) is formed by bonding between C1/C11 and C10/C14. During the biocatalytic formation of dolabellane the cyclization mechanism is stopped after deprotonation at C16. In the biocatalytic formation of cyclooctat-9-en-7-ol the carbocation migrates to C10 by two consecutive 1,2-hydride shifts (C10 to C14 and C14 to C15), providing structure (**2**). Deprotonation of (**2**) yields (**3**), which subsequently undergoes closure of the third ring by bond formation between C6 and C2 facilitated by protonation of (**3**) at C3. Eventually, water is added to (**3**), which results in the formation of cyclooctat-9-en-7-ol. Since the absolute configuration of the CotB2-derived dolabellane product has been determined in

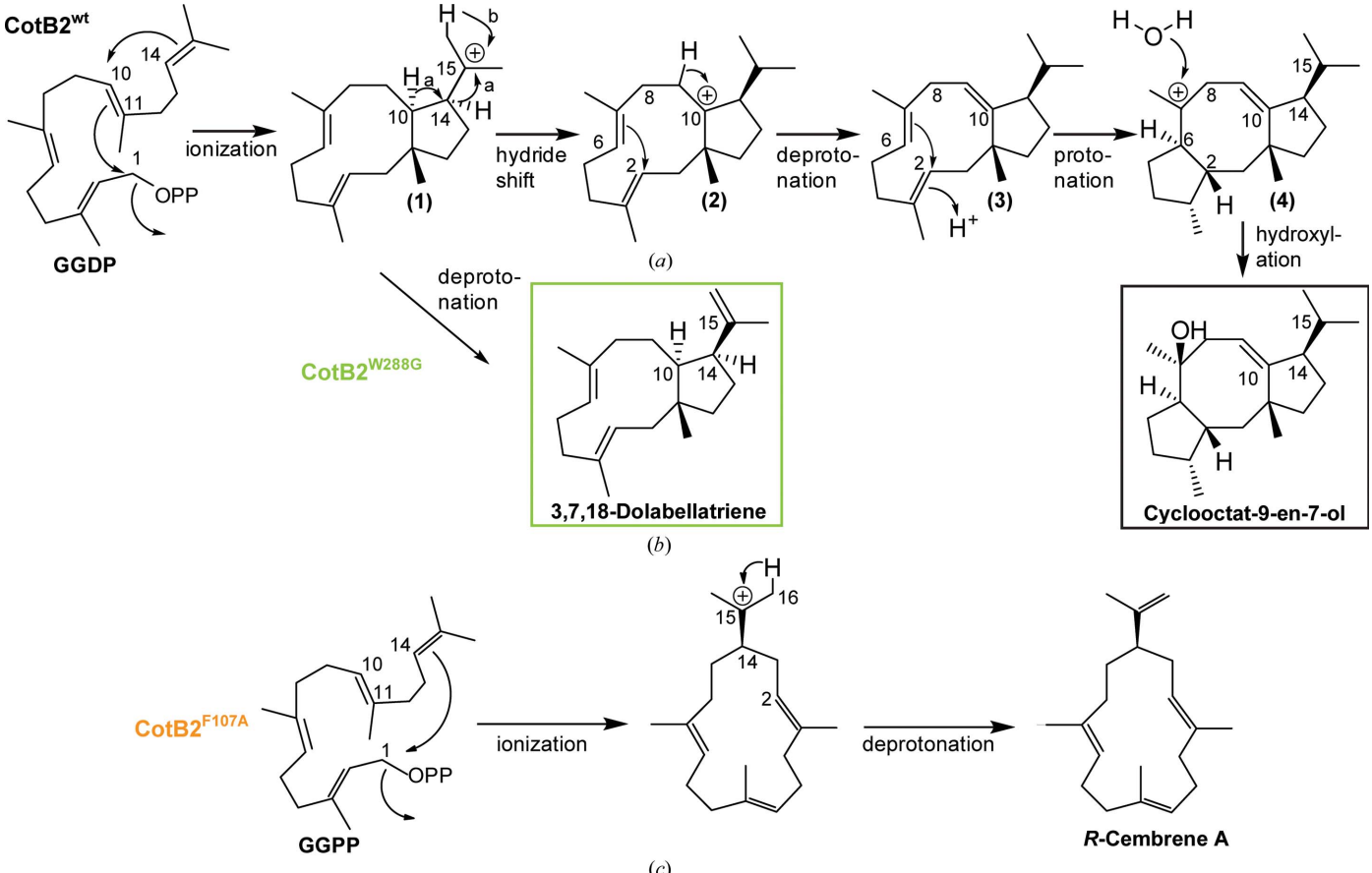


Figure 4
Proposed reaction mechanism. The final products are indicated in boxes. (*a*) The formation of cyclooctat-9-en-7-ol by CotB2^{wt}, (*b*) the formation of 3,7,18-dolabellatriene by CotB2^{W288G} and (*c*) the formation of *R*-cembrene A by CotB2^{F107A}.

this report, the cumulative data indicate that the formation of (**2**) is facilitated by an energetically rather unfavorable tandem *cis*-hydride transfer. This mechanistic feature has also been reported in the biocatalysis of fusiccocin (Stoessl *et al.*, 1988; Banerji *et al.*, 1978), and it is a likely option in the formation of dolabellane and cyclooctat-9-en-7-ol. Therefore, a consecutive *cis*-hydride transfer may be a particular feature of the biosynthesis of fusicoccane-type macrocycles. Consequently, CotB2^{W288G} may play an important role in the cyclization process by facilitating the potential *cis* tandem 1,2-hydride shifts and promoting the migration of the carbocation. Based on the experimentally determined absolute configuration of dolabellatriene, we can superimpose this information on cyclooctat-9-en-7-ol and suggest the (2*R*, 3*R*, 6*S*, 7*S*, 11*R* and 14*R*) configuration (Fig. 4). The formation of (1*R*,3*E*,7*E*,11*S*,12*S*)-3,7,18-dolabellatriene has significant implications as this structure has been shown to have potential antibiotic activity against multidrug-resistant *Staphylococcus aureus* (MRSA; Ioannou *et al.*, 2011). To our knowledge, this is the first diterpene produced in a recombinant microbial production system which demonstrates significant bioactivity in the absence of any additional functional decoration. Interestingly, our recombinant *E. coli* system was capable of producing milligram quantities of dolabellatriene without toxicity effects. The reason for this may be linked to the fact that *E. coli* is Gram-negative while *Staphylococcus* is a Gram-positive organism. Therefore, pharmaceutical production of the promising new antibiotic dolabellatriene in *E. coli* is an efficient route to secure an industrial-scale supply.

4. Conclusion

In this study, we determined the structure of CotB2 (Fig. 1*b*), the first structure of a diterpene cyclase of bacterial origin. CotB2 catalyzes the cyclization of GGDP to cyclooctat-9-en-7-ol, the precursor of the anti-inflammatory drug cyclooctatin. The latter compound belongs to the family of fusicoccanes that are produced in bacteria and fungi. Owing to their broad biological activity (de Boer & de Vries-van Leeuwen, 2012), they represent a potential lead scaffold for chemical and pharmaceutical products. Moreover, cyclooctatin is of particular interest since it is a next-generation anti-inflammatory drug that lacks severe side effects compared with existing anti-inflammatory agents derived from either aspirin or ibuprofen. It would be a new lead compound for the pharmaceutical industry and various derivates of cyclooctatin could be created through combinatorial biotechnology.

Bacterial CotB2 belongs to the class I terpene synthases but contains an altered aspartate-rich (^{110}DD*X*D^{113}) motif for Mg^{2+} binding compared with the classical DD*XX*D motif. Hence, the Mg^{2+} coordination must be substantially different compared with the classical class I terpene cyclases.

Comparison of the aspartate-rich motif of the cyclase domains of CotB2 and PaFS (Toyomasu *et al.*, 2007) from the plant-pathogenic fungi *Phomopsis amygdale* reveals that fungi have the classical aspartate-rich motif. Consequently, the degenerated aspartate-rich motif of the model bacterial CotB2 seems to be an exception. In contrast to CotB2, the fungal PaFS is a bifunctional fusion protein with prenyltransferase and cyclase domains. This suggests convergent evolution of fusicoccane-centred cyclases in bacteria and fungi. In further studies, it would be of interest to determine the structurally related catalytic differences between bacterial and fungal derived fusicoccanes.

Structural comparison of CotB2 and pentalenene synthase from *Streptomyces* UC5319 reveals a superposition of catalytically important residues. Residues Phe107 and Phe149 of CotB2 superimpose with Phe77 and Tyr146 of pentalenene synthase. Hence, we propose that Phe107 and Phe149 of CotB2 are optimally located to stabilize the reactive carbocation intermediates through quadropole-charge (Burley & Petsko, 1988) and dipole interactions, as suggested for the sesquiterpene cyclase pentalenene (Lesburg *et al.*, 1997). The formation of *R*-cembrene A is not part of the mechanism for the formation of cyclooctat-9-en-7-ol owing to the connection of C1 and C14 (Fig. 4) to form a monomeric ring structure. It can be assumed that Phe107 is essential for the correct orientation of the GGDP substrate and migration of the positive charge through the carbon skeleton. Aromatic residues such as Phe107, Phe149 and Trp288 may stabilize carbocation cyclization intermediates through electrostatic quadrupole-charge interactions and in addition by polar residues that are able to participate in electrostatic stabilization.

The identification of residues with plasticity in the active site of CotB2 allowed us to design single-point mutations of aromatic residues potentially important for the migration of the emerging carbocation. The mutation of Phe107, Phe149 and Trp288 altered the product outcome (Fig. 4 and Table 2). We were able to identify two alternative products. For wild-type CotB2 the natural product is cyclooctat-9-en-7-ol, which features a tricyclic 5–8–5 ring system characteristic of fusicoccanes. In contrast, the single-point mutation CotB2^{F107A} led to the monocyclic cembrene A with a single 14-membered ring system. Mutation to CotB2^{W288G} led to a bicyclic dolabellatriene product with a 5–11-membered ring system. Dolabellatriene shows antibacterial activity (Ioannou *et al.*, 2011), but to date it is not accessible by organic synthesis and hence its production depends on isolation from its natural source. Once such a microbial production platform has been successfully established for a wild-type terpene cyclase, it could be very easily adapted to variants of terpene cyclase. The incorporation of the CotB2^{W288G} variant into a bacterial downstream element of a microbial production platform offers an attractive route to access this bioactive compound. The efficient production of dolabellatriene in recombinant *E. coli* opens a new route for efficient industrial-scale supply of this new and highly promising drug candidate.

We are grateful to C. Alings for excellent technical support and C. Weise for the mass-spectrometry service. We are grateful to M. Wahl for continuous encouragement and support. RJ is supported by the Heinrich Böll foundation. RJ and BL are grateful for Deutsche Forschungsgemeinschaft

Grant SFB958/A6. We acknowledge beamtime and support at beamline P14-2 of PETRA III (Deutsches Elektronen Synchrotron, Hamburg, Germany). We accessed beamlines of the BESSY II (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung II) storage ring (Berlin, Germany) *via* the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück Centrum and the Leibniz-Institut für Molekulare Pharmakologie.

References

Aaron, J. A., Lin, X., Cane, D. E. & Christianson, D. W. (2010). *Biochemistry*, **49**, 1787–1797.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Adams, R. P. (2007). *Identification of Essential Oil Components By Gas Chromatography/Mass Spectrometry*. Carol Stream: Allured Publishing Group.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B. & Stephanopoulos, G. (2010). *Science*, **330**, 70–74.
Aoyagi, T., Aoyama, T., Kojima, F., Hattori, S., Honma, Y., Hamada, M. & Takeuchi, T. (1992). *J. Antibiot.* **45**, 1587–1591.
Ashby, M. N. & Edwards, P. A. (1990). *J. Biol. Chem.* **265**, 13157–13164.
Bai, S. & Jain, M. (2008). *Magn. Reson. Chem.* **46**, 791–793.
Banerji, A., Hunter, R., Mellows, G., Sim, K. & Barton, D. H. R. (1978). *J. Chem. Soc. Chem. Commun.*, pp. 843–845.
Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.
Boer, A. H. de & de Vries-van Leeuwen, I. J. (2012). *Trends Plant Sci.* **17**, 360–368.
Bonnard, I., Jhaumeer-Laulloo, S. B., Bontemps, N., Banaigs, B. & Aknin, M. (2010). *Mar. Drugs*, **8**, 359–372.
Brück, T., Kourist, R. & Loll, B. (2014). *ChemCatChem.*, doi:10.1002/cctc.201300733.
Burley, S. K. & Petsko, G. A. (1988). *Adv. Protein Chem.* **39**, 125–189.
Cane, D. E. (1990). *Chem. Rev.* **90**, 1089–1103.
Cane, D. E. & Kang, I. (2000). *Arch. Biochem. Biophys.* **376**, 354–364.
Casanova, C., Hamilton, J. G., Trigo, J. R. & Costa, A. I. (2006). *Mem. Inst. Oswaldo Cruz*, **101**, 113–115.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Christianson, D. W. (2006). *Chem. Rev.* **106**, 3412–3442.
Christianson, D. W. (2008). *Curr. Opin. Chem. Biol.* **12**, 141–150.
DeLano, W. (2002). *PyMOL.* http://www.pymol.org.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Görner, C., Häuslein, I., Schrepfer, P., Eisenreich, W. & Brück, T. (2013). *ChemCatChem*, **11**, 3289–3298.
Greenhagen, B. T., O'Maille, P. E., Noel, J. P. & Chappell, J. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 9826–9831.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Hyatt, D. C., Youn, B., Zhao, Y., Santhamma, B., Coates, R. M., Croteau, R. B. & Kang, C. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 5360–5365.
Ioannou, E., Quesada, A., Rahman, M. M., Gibbons, S., Vagias, C. & Roussis, V. (2011). *J. Nat. Prod.* **74**, 213–222.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kato, N., Okamoto, H. & Takeshita, H. (1996). *Tetrahedron*, **52**, 3921–3932.
Kim, S.-Y., Zhao, P., Igarashi, M., Sawa, R., Tomita, T., Nishiyama, M. & Kuzuyama, T. (2009). *Chem. Biol.* **16**, 736–743.
Köksal, M., Hu, H., Coates, R. M., Peters, R. J. & Christianson, D. W. (2011). *Nature Chem. Biol.* **7**, 431–433.
Köksal, M., Jin, Y., Coates, R. M., Croteau, R. & Christianson, D. W. (2011). *Nature (London)*, **469**, 116–120.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lesburg, C. A., Zhai, G., Cane, D. E. & Christianson, D. W. (1997). *Science*, **277**, 1820–1824.
Lin, X. & Cane, D. E. (2009). *J. Am. Chem. Soc.* **131**, 6332–6333.
Maimone, T. J. & Baran, P. S. (2007). *Nature Chem. Biol.* **3**, 396–407.
Paddon, C. J. *et al.* (2013). *Nature (London)*, **496**, 528–532.
Poulter, C. D., Argyle, J. C. & Mash, E. A. (1978). *J. Biol. Chem.* **253**, 7227–7233.
Rasoamiaranjanahary, L., Marston, A., Guilet, D., Schenk, K., Randimbivololona, F. & Hostettmann, K. (2003). *Phytochemistry*, **62**, 333–337.
Rynkiewicz, M. J., Cane, D. E. & Christianson, D. W. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 13543–13548.
Sacchettini, J. C. & Poulter, C. D. (1997). *Science*, **277**, 1788–1789.
Schwabe, R., Farkas, I. & Pfander, H. (1988). *Helv. Chim. Acta*, **71**, 292–297.
Shishova, E. Y., Di Costanzo, L., Cane, D. E. & Christianson, D. W. (2007). *Biochemistry*, **46**, 1941–1951.
Stoessl, A., Rock, G. L., Stothers, J. B. & Zimmer, R. C. (1988). *Can. J. Chem.* **88**, 1084–1090.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Subbarao, G. V. *et al.* (2009). *Proc. Natl Acad. Sci. USA*, **106**, 17302–17307.
Tarshis, L. C., Yan, M., Poulter, C. D. & Sacchettini, J. C. (1994). *Biochemistry*, **33**, 10871–10877.
Tietze, L. F., Brazel, C. C., Hölsken, S., Magull, J. & Ringe, A. (2008). *Angew. Chem. Int. Ed. Engl.* **47**, 5246–5249.
Toyomasu, T., Tsukahara, M., Kaneko, A., Niida, R., Mitsuhashi, W., Dairi, T., Kato, N. & Sassa, T. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 3084–3088.
Vanderah, D. J., Rutledge, N., Schmitz, F. J. & Ciereszko, L. S. (1978). *J. Org. Chem.* **43**, 1614–1616.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
Vedula, L. S., Jiang, J., Zakharian, T., Cane, D. E. & Christianson, D. W. (2008). *Arch. Biochem. Biophys.* **469**, 184–194.
Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
Wendt, K. U. & Schulz, G. E. (1998). *Structure*, **6**, 127–133.
Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B. & Christianson, D. W. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 15375–15380.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Zhou, K., Gao, Y., Hoy, J. A., Mann, F. M., Honzatko, R. B. & Peters, R. J. (2012). *J. Biol. Chem.* **287**, 6840–6850.