Structural studies of biotechnologically relevant enzymes

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

Ronja Driller (née Janke)

from Ilmenau

2018

This work was carried out in the period of November 2013 to September 2018 under the supervision of Prof. Dr. Markus C. Wahl at the Institute of Chemistry and Biochemistry, Freie Universität Berlin, Germany.

First Reviewer: Prof. Dr. Markus C. Wahl

                             Structural Biochemistry

                             Institute of Chemistry and Biochemistry

                             Freie Universität Berlin

                             Takustraße 6

                             14195 Berlin, Germany

Second Reviewer: Prof. Dr. Thomas Brück

                             Werner Siemens-Chair of Synthetic Biotechnology

                             Technische Universität München

                             Lichtenbergstraße 4

                             85748 Garching, Germany

Date of defense: 3.12.2018

Contents

Contents

Abstract

Enzymes play an indispensable role in our everyday life. They are not only catalysing a vast number of reactions in our body, they are incorporated into many industrial processes. In the light of pollution and decline of fossil resources it is essential to develop new production routes to pharmaceutical and chemical target compounds. The exaggerated use of antibiotics and the resulting increase of multi-resistant bacteria is another threat of our century. Solving this problem requires deep understanding of antibiotic resistance mechanisms to enable rational drug-design.

The main focus of the thesis lies on terpene synthases, which catalyse the cyclisation of linear isoprenoid precursors to complex macrocycles. These compounds are high commercial targets of the chemical and pharmaceutical industry. Terpenes constitute the largest group of natural compounds. Especially in the class of the diterpenes many compounds possess antibacterial, anti-viral, insecticidal and anti-inflammatory properties. White biotechnology opens up new routes for the production of terpenes in bacteria in a sustainable manner. In order to optimise terpene titers in bacteria, structural knowledge of the involved terpene synthases is crucial. Here, we report the crystal structures of the diterpene synthases CotB2 and IES in complex with various substrate-analogues. Furthermore, the structure of the oleate hydratase OhyRe was elucidated. Products of this enzyme are used as surfactants and lubricants in industry.

The X-ray crystal structures of biotechnologically relevant enzymes obtained during this thesis contribute to illustrate their reaction mechanisms and binding modes.

As another part of this thesis, the structure of AlbAS, a bacterial antibiotic-binding protein, was solved. The protein structure sheds light on the binding mechanism of the antibacterial drug albicidin at a molecular level.

Zusammenfassung

Enzyme spielen eine Schlüsselrolle in unserem Leben. Sie sind nicht nur an der Katalyse von lebenswichtigen Vorgängen in unserem Körper beteiligt, sie sind ebenso unverzichtbar in vielen chemisch-industriellen Prozessen. Im Zeitalter von Klimawandel und zur Neige gehenden fossilen Rohstoffen ist es essentiell, neue Produktionswege für die Herstellung von Arzneistoffen, Verbundstoffen und Kosmetika, die bisher primär über chemische Synthese oder aus Pflanzenextraktion gewonnen werden, zu finden. Ebenso stellt auch die übertriebene Antibiotikagabe und daraus resultierend die Zunahme von Antibiotikaresistenzen die Welt vor ein Problem. Die Aufklärung von Antibiotika-Wirkmechanismen sowie bakteriellen Resistenzmechansimen ist Voraussetzung um neue antimikrobiell wirksame Substanzen zu erforschen. Im Zuge dieser Doktorarbeit wurde die Struktur von AlbAS, einem bakteriellen Antibiotika-bindenden Protein, bestimmt. Die Proteinstruktur gibt Aufschluss über das Bindeverhalten vom antibakteriellen Albicidin auf molekularer Ebene und hilft so den Resistenzmechanismus aufzuklären.

Der Hauptfokus dieser Doktorarbeit liegt auf den Terpensynthasen, einer Enzymklasse, die lineare Isoprenoide zu komplexen Makrozyklen zyklisiert. Die Produkte dieser Reaktionen sind wichtige Ausgangsstoffe und Wirkstoffe der chemischen und pharmazeutischen Industrie. Die Klasse der Terpene bildet die größte Vielfalt von Naturprodukten. Insbesondere unter den Diterpenen sind zahlreiche Vertreter mit anti-bakteriellen, anti-viralen, insektiziden und entzündungshemmenden Eigenschaften zu finden. Durch die genetische Inkorporation von Terpenzyklasen in Bakterien lassen sich kosteneffektiv und umweltschonend Terpene herstellen. Genaue strukturelle Kenntnisse der involvierten Enzyme sind dabei unabdingbar, um sie über gezielte Mutagenese zu optimieren. Im Rahmen dieser Arbeit sind die Strukturen der Diterpenzyklasen CotB2 und IES in Komplex mit verschiedenen Substratanaloga gelöst worden.

Des Weiteren wurde die Struktur der Oleat Hydratase OhyRe aufgeklärt. Produkte dieses Enzyms dienen der Industrie als Tenside und Schmiermittel.

Die Kristallstrukturen biotechnologisch relevanter Enzyme, die im Rahmen dieser Arbeit bestimmt worden sind, tragen im entscheidenden Maße zur Aufklärung der Reaktionsmechanismen und Substrat-Bindeverhalten bei.

List of Publications

I. <u>Ronja Janke</u>*, Christian Görner*, Max Hirte, Thomas Brück, Bernhard Loll (2014), The first structure of a bacterial diterpene cyclase: CotB2, *Acta Crystallographica Section D Biological Crystallography* 70:1528-1537, https://doi.org/10.1107/S1399004714005513

*These authors contributed equally to this work.

RD expressed, purified and crystallised CotB2, collected X-ray diffraction data, solved and built the structure, performed DLS experiment, prepared figures for the manuscript and wrote parts of the manuscript.

II. <u>Ronja Janke</u>, Monika Fuchs, Mathias Christmann, Thomas Brück, Bernhard Loll (2017), Herstellung von polyzyklischen Diterpenen, *BIOspektrum* 23:709-711, https://doi.org/10.1007/s12268-017-0858-5

RD wrote the manuscript.

III. Jan Lorenzen*, <u>Ronja Driller</u>*****, Ayk Waldow, Farah Qoura, Bernhard Loll, Thomas Brück (2018), *Rhodococcus erythropolis* Oleate Hydratase: a New Member in the Oleate Hydratase Family Tree - Biochemical and Structural Studies, *ChemCatChem* 10:407-414, https://doi.org/10.1002/cctc.201800008

*These authors contributed equally to this work.

RD cloned, expressed, purified and crystallised OhyRe, collected X-ray diffraction data, solved and built the structure, performed MALS experiment and wrote parts of the manuscript. RD designed the cover for the journal.

IV. Lida Rostock*, <u>Ronja Driller</u>*, Stefan Grätz, Dennis Kerwat, Leonard von Eckardstein, Daniel Petras, Claudia Alings, Franz-Josef Schmitt, Thomas Friedrich, Markus Wahl, Bernhard Loll, Andi Mainz, Roderich D. Süssmuth (2018), Molecular insights into antibiotic resistance: how a binding protein traps albicidin, *Nature Communications* 9:3095, https://doi.org/10.1038/s41467-018-05551-4

4

*These authors contributed equally to this work.

RD crystallised AlbA, collected X-ray diffraction data and built the structure, performed MALS experiments with and without albicidin and prepared figures for the manuscript.

V. <u>Ronja Driller</u>, Sophie Janke, Monika Fuchs, Evelyn Warner, Anil R. Mhashal, Dan Thomas Major, Mathias Christmann, Thomas Brück, Bernhard Loll (2018), Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis, *Nature Communications* 9:3971, https://doi.org/10.1038/s41467-018-06325-8

RD cloned, expressed, purified and crystallised CotB2 wt and mutants with and without substrate-analogues, collected X-ray diffraction data, solved and built the structures and wrote the manuscript.

Additional publications not included in this thesis:

Siamak Semsary, Ivana Crnovčić, <u>Ronja Driller</u>, Joachim Vater, Bernhard Loll, Ullrich Keller (2018), Ketonization of Proline Residues in the Peptide Chains of Actinomycins by a 4-Oxoproline Synthase, *ChemBioChem* 19:706-715

1. Introduction

Although climate change and pollution by plastic waste are more precarious than ever, industry is still relying primarily on fossil resources. The United Nations aim to break this paradigm and stated their Sustainable Development Goals (SDGs) in 2015. To implement their goals of a more sustainable future, biotechnological routes for the production of biomolecules are needed (de Lorenzo et al., 2018).

Many of the pharmaceutical active compounds were first discovered in extracts of plants, fungi and bacteria, produced by enzymes. Isolation of the compounds from their native source can be difficult and often does not provide substances at an industrial scale (Chang and Keasling, 2006). However, most of these biomolecules were then synthesised *via* organic synthesis. In the light of sustainability, this may not be the best approach, since chemical synthesis often involves toxic waste and usage of fossil resources. Furthermore, some of the sophisticated biomolecules are difficult to access by chemical synthesis or are only obtained in small yields. In recent years, biotechnological production routes have become more attractive. Due to their capability of catalysing, even chemically challenging reactions, in a very elegant and effective way, enzymes are fascinating tools in biotechnology.
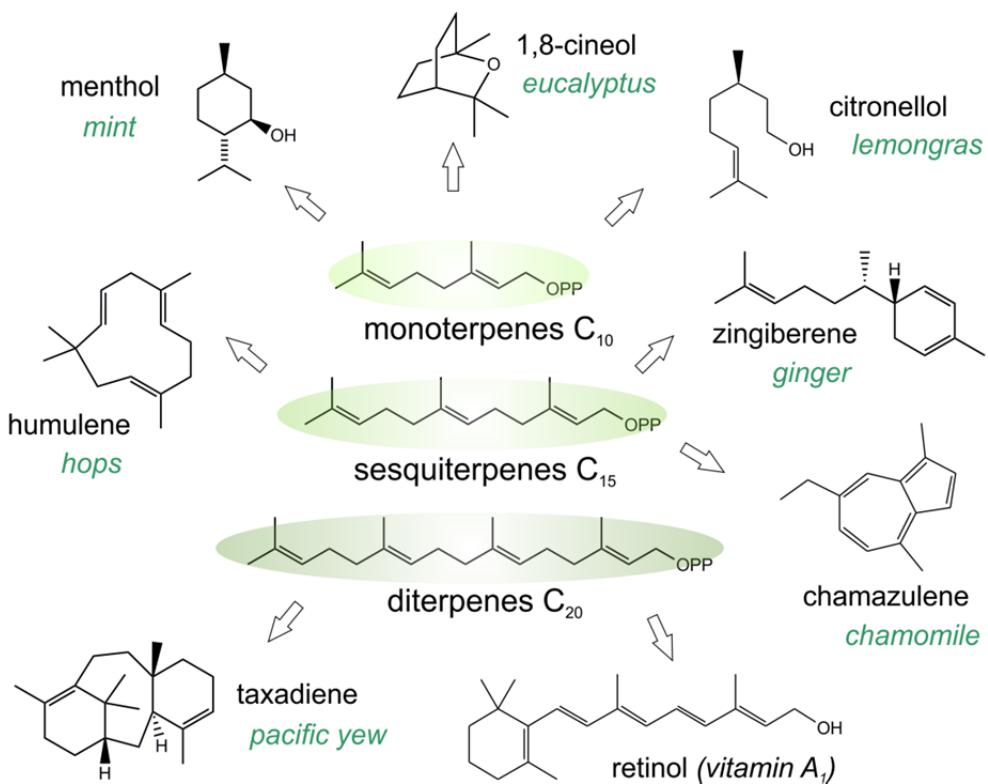
In order to systematically optimise enzymes, it is of utter importance to obtain knowledge of their overall structure and important residues lining the catalytically active site. *Via* X-ray crystallography structures of proteins and proteins in complex with their substrates can be obtained. Optimisation of enzymes can not only improve their activity under special conditions (*e.g.* temperature stable enzymes in washing powder), but can also result in new product formation.

A very specific class of biomolecules are antibiotics, which often interfere with bacterial metabolism to kill them. However, due to abuse and overdosage, especially in stock breeding, many bacteria gained resistance against common antibiotics. Therefore, new compounds are needed that are either slightly chemically modified compared to old antibiotics or target other structures of bacteria. Investigations of antibiotic-resistance mechanisms of bacteria are crucial to understand and overcome these problems.

1.1 Terpene synthases

Among the many classes of natural compounds, terpenes constitute the largest and most diverse one present in almost all species, including plants, fungi and bacteria. They are not only the source of many scents - they are also flavours in food and active substances in medicine. Because of their immense structural and bioactive diversity, terpenes are used in many modern branches of the chemical, biotechnological, agricultural and pharmaceutical industry (Aharoni et al., 2005; Bruck et al., 2014; Newman and Cragg, 2016).

Terpenes can be divided into different classes, distinguished by the number of isoprene units in their scaffold (Fig. 1). Given the vast number of different terpenes it is astonishing how simple the precursors are. All molecules belonging to a specific class of terpenes derive from the same linear precursor specific for the corresponding class. The variety of terpene molecules within the distinct classes is generated by terpene synthases (TPSs), sometimes referred to as terpene cyclases (Kirby and Keasling, 2009). These enzymes perform the cyclisation of the linear precursor to polycyclic structures with elaborate stereochemistry and with strict regioselectivity (Christianson, 2017). Further diversity of the molecules is reached by additional chemical modifications, catalysed by P450 monooxygenases for example.

Fig. 1 Selection of terpenes present in nature. Arrows indicate from which linear precursor the compound derives. Natural sources of the compounds are written in green.
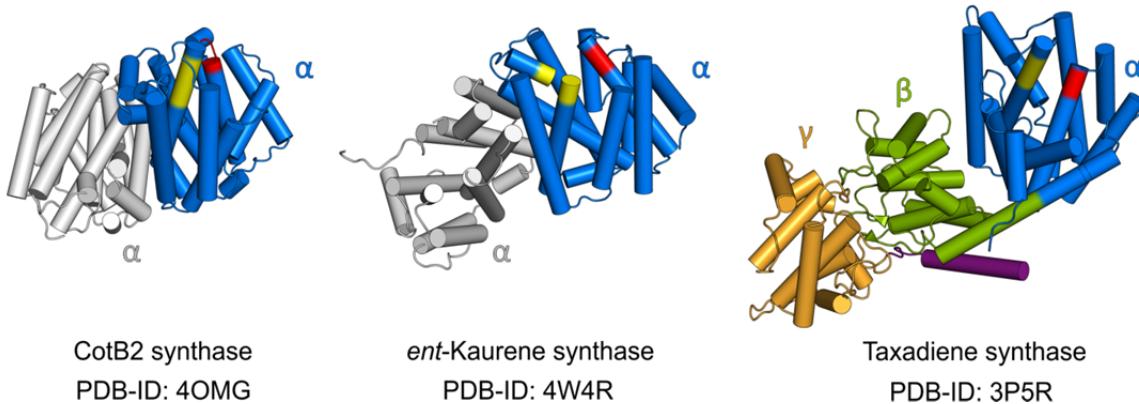
Most of the sophisticated polycyclic terpenes are difficult to access by classical chemical synthesis. Instead one can manipulate microorganisms to produce terpene scaffolds that can then be further modified by chemical synthesis (Chang and Keasling, 2006). Mutating the involved enzymes of the terpene biosynthesis pathway leads to additional, altered products, generating an immense number of potential new bioactive terpene scaffolds (Greenhagen et al., 2006).

1.1.1 Diterpene synthases

Particularly in the class of C_{20}-based diterpenes (Fig. 1) many compounds show a broad spectrum of bioactivity, such as antibiotic, insecticidal, anti-tumour, anti-viral or anti-inflammatory properties. Diterpenes are built up by four isoprene units, the linear C_{20} precursor being geranylgeranyl diphosphate (GGDP). GGDP is cyclised by diterpene synthases that start the cyclisation reaction by either removal of the diphosphate moiety (class I reaction mechanism) or protonation of the terminal isoprene unit (class II

reaction mechanism). The emerging reactive carbocation then propagates intramolecular cyclisation. Given the different cyclisation strategies, class I and class II diterpene synthases differ in their tertiary fold and motifs (Fig. 2). The common fold for class I TPSs is the α-helical bundle fold - the "α-fold", where ten to twelve long helices are connected *via* short loop segments, creating a claw-like overall structure with a large cleft forming the active site (Whittington et al., 2002). Class II TPSs consist of two different α-helical domains, termed β and γ (Oldfield and Lin, 2012).

In bacteria, diterpene synthases exclusively belong to class I TPS, whereas in fungi and plants, they can be both class I or class II TPS and even mixed class I/II. In most of the known structures, bacterial TPSs form parallel (*e.g. ent*-kaurene synthase (Liu et al., 2014)) or antiparallel (*e.g.* CotB2 synthase (Janke et al., 2014)) dimers. However, there are examples of higher oligomerisation states, *e.g.* fusicoccadiene synthase forms a hexamer (Chen et al., 2016). In plants, TPSs can adopt a more complex fold consisting of several domains. In the case of diterpene taxadiene synthase (TXS) from pacific yew tree, a catalytically active class I α-domain is present, as well as α-helical domains β and γ, known for class II TPSs (Koksal et al., 2011). However, the last two domains do not contribute to the enzyme's activity.


Fig. 2 Folds of class I diterpene synthases. Depending on the organism, TPSs adopt different folds, ranging from α-dimers to an αβγ-domain structure. CotB2 (Janke et al., 2014) and *ent*-kaurene synthase (Liu et al., 2014) are of bacterial origin, taxadiene synthase (TXS) is a plant diterpene synthase (Koksal et al., 2011). The different domains are colour-coded (α in blue, β in green and γ in gold), dimeric molecules are shown in grey. The aspartate-rich motif is coloured in red, the NSE/DTE motif in yellow. The N-terminus of TXS is shown in purple.

Members of class I TPSs usually exhibit a **DD**XX**D/E** (X=any amino acid) motif, termed aspartate-rich motif (Ashby and Edwards, 1990). The second catalytically important motif is the NSE/DTE consensus-sequence (**N,D**)DXX(**S,T**)XXX**E** (Christianson, 2006). Both motifs are crucial in class I TPSs for binding three divalent metal-ions, usually Mg^{2+}-ions, that coordinate the diphosphate moiety of the substrate GGDP. Class II TPSs feature an aspartate-rich motif only, often of the form DXDD.
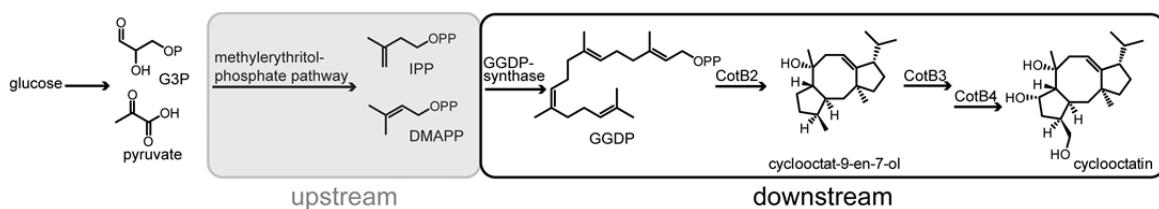
Although various structures of TPSs are available to date, X-ray crystal structures of TPSs have been exclusively solved by experimental phasing (Christianson, 2017). α-helical proteins are notoriously challenging to solve by molecular replacement, since small tilts of the helices compared to the reference model already complicate finding the correct solution. Furthermore, sequence identity for TPSs is often below 20 %. The two diterpene synthase structures presented in this thesis have been solved by single-wavelength anomalous dispersion using selenomethionine-labelled protein.

1.1.2 Terpene synthases in white biotechnology

The term white biotechnology implies saving natural resources and minimising CO_2 emission by exploiting microorganisms or enzymes to produce materials or substances that are otherwise derived from oil.

Terpenes are of special interest to the pharmaceutical industry because of their broad spectrum of activities. Since total chemical synthesis of the polycyclic compounds with many stereo-centres is cumbersome, different production strategies are needed. Using microorganisms such as *Escherichia coli* or yeast for the synthesis of terpenes not only raises sustainability by minimising waste and saving resources, but also gives the opportunity to perform protein engineering to optimise product ratios or alter product formation (Bruck et al., 2014). Furthermore, production by engineered microorganisms considerably reduces the cost compared to total chemical synthesis or extraction from natural sources, since the target compounds are produced from inexpensive sugar-based carbon sources. Prominent examples of optimised terpene production pathways in *E. coli* are taxadiene, a precursor of the anti-cancer drug taxol (Ajikumar et al., 2010), and amorpha-4,11-diene, an anti-malarial drug precursor (Anthony et al., 2009).

The terpene production chassis in bacteria involves an upstream and downstream part. In the upstream part all the necessary enzymes for the generation of building blocks from monomeric sugars are integrated. The linear terpene precursor is made up from these building blocks in the downstream part of the pipeline. The precursor is then cyclised by a TPS. Subsequently, the cyclised product can be further modified by other enzymes, *e.g.* P450 monooxygenases. An example of a terpene production pipeline is shown in Fig. 3 for the anti-inflammatory diterpene cyclooctatin.


Fig. 3 Schematic cyclooctatin production in microorganism. Monomeric sugars are used to generate the building blocks isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) in the upstream module. In the downstream element, the linear diterpene precursor geranylgeranyl diphosphate (GGDP) is generated by GGDP-synthase and cyclised by CotB2. The resulting tricyclic diterpene is then further modified by the cytochromes CotB3 and CotB4 to the anti-inflammatory drug cyclooctatin.

The downstream enzymes of the terpene production pipeline in bacteria are the ideal candidates for protein engineering. Especially the terpene cyclases are valuable targets, since they catalyse the most challenging reactions of the chassis. By site-directed mutagenesis of genes encoding these enzymes, the product spectrum can either be altered or shifted towards a specific product, if the cyclase produces a mixture of different cyclic terpenes (Yoshikuni et al., 2006).

2. Aims

X-ray crystal structures are indispensable to study an enzyme's catalytic mechanism, its precise function in a protein network or its mode of binding to ligands/ inhibitors. Furthermore, mutagenesis can be informed by knowledge of a structure. In this thesis, I aimed at solving the structures of several enzymes important to biotechnology. The following proteins were subjected to structure analysis:

I.   The bacterial diterpene synthase CotB2 and the subsequent P450 Cytochromes CotB3 and CotB4

II. The diterpene synthase IES

III. The bacterial oleate hydratase OhyRe

IV. The bacterial albicidin-binding protein AlbA

3. Results and Discussion

3.1 Diterpene synthase CotB2

This section refers to the following publications:
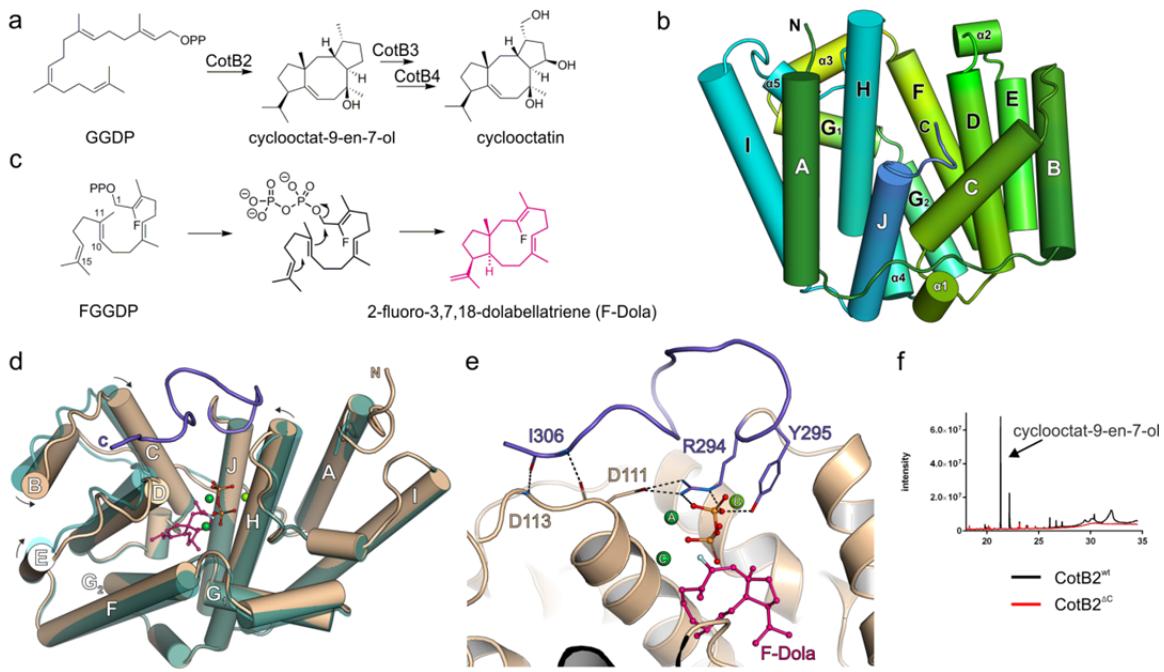
- Ronja Janke, Christian Görner, Max Hirte, Thomas Brück, Bernhard Loll (2014), The first structure of a bacterial diterpene cyclase: CotB2, *Acta Crystallographica Section D Biological Crystallography* 70:1528-1537
- Ronja Janke*,* Monika Fuchs, Mathias Christmann, Thomas Brück, Bernhard Loll (2017), Herstellung von polyzyklischen Diterpenen, *BIOspektrum* 23:709-711
- Ronja Driller, Sophie Janke, Monika Fuchs, Evelyn Warner, Anil R. Mhashal, Dan Thomas Major, Mathias Christmann, Thomas Brück, Bernhard Loll (2018), Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis, *Nature Communications* 9:3971

The bacterial diterpene synthase CotB2 (derived from *Streptomyces melanosporofaciens*), which is a member of the class I TPSs, is responsible for the cyclisation of GGDP to cyclooctat-9-en-7-ol (Fig. 4a). This fusicoccane with a 5-8-5 fused ring system is then subsequently modified by the P450 cytochromes CotB3 and CotB4 to the bioactive compound cyclooctatin (Fig. 4a). Cyclooctatin is a next-generation anti-inflammatory drug since it targets a lysophospholipase that is upregulated in eosinophilic leucocytes rather than a cyclooxygenase that is inhibited by ibuprofen or aspirin (Aoyagi et al., 1992; Kim et al., 2009).

3.1.1 Overall structure

By solving the crystal structure of CotB2 (Fig. 4b, PDB-ID 4OMG (Janke et al., 2014)), we could show that it consists mainly of α-helices connected by short loop-regions. This α-helical bundle fold is common for class I TPSs. In the crystal, two CotB2 molecules form an antiparallel dimer. This dimer represents the biological relevant unit of the enzyme which we verified by dynamic light scattering (DLS).

Fig. 4 Overall structure of the bacterial diterpene synthase CotB2 and the role of its C-terminus. a) Reaction of geranylgeranyl diphosphate (GGDP) to the bioactive compound cyclooctatin. b) Cartoon-representation of CotB2, colour gradient is from green (N-terminus) to blue (C-terminus). c) Reaction mechanism of 2-fluoro geranylgeranyl diphosphate (FGGDP) to 2-fluoro-3, 7, 18-dolabellatriene (F-Dola), catalysed by CotB2. d) Superposition of CotB2 in its open (teal) and closed (beige) state. The C-terminus of the closed structure is coloured in purple. e) The C-terminus (purple) is coordinating the diphosphate, Mg^{2+}-ions are coloured in green and F-Dola in pink. f) GC-MS spectrum of CotB2 and the CotB2 variant ΔC, monitoring product formation.

We solved the open state structure of CotB2 with no ligand or diphosphate bound. However, we observed a blob of electron density within the active site of the CotB2^{open} structure, but it was not possible to assign a molecule to fit into this density. Most likely, it is some remnant molecule of *E. coli*, that got stuck during expression of CotB2.

To get an idea of CotB2s dynamic transitions during catalysis, we aimed to determine its closed ligand-bound conformation. This task proved to be challenging since CotB2 is a very effective, fast enzyme. Initial soaking experiments with the native substrate GGDP or diphosphate failed. We then soaked CotB2 crystals with the substrate-analogue 2-fluoro geranylgeranyl diphosphate (FGGDP), a GGDP molecule with a fluorine-atom attached to the C2-position (Fig. 4c). Fluorinated substrates have been used for crystallisation of other TPSs and have been shown to stick to the active site without undergoing cyclisation, trapping the enzyme in a closed state (Koksal et al.,

2012; Koksal et al., 2011). With FGGDP we managed to solve the structure of the closed state of CotB2 (Fig. 4d, PDB-ID 6GGI). The closed conformation resembles the open one with an rmsd of 0.7 Å for 278 aligned Cα-atoms (Fig. 4d). However, there is a significant change in the C-terminal region of the enzyme. In the open state, the last 12 C-terminal residues could not be modelled due to flexibility. In the closed structure, we observed the complete C-terminus folding over the active site (Fig. 4e), thereby acting as a lid to shield the active site from bulk solvent. Interestingly, the C-terminus could not be resolved by Tomita *et al.* who co-crystallised CotB2 together with the substrate analogue geranylgeranyl thiodiphosphate (GGSDP) (PDB-ID 5GUE (Tomita et al., 2017)). In their structure they captured a pre-catalytic state of CotB2 with a single Mg^{2+}-ion bound to the active site.

In order to prove that the folded C-terminus is not a crystallographic artefact, we co-crystallised CotB2 with alendronate (AHD), a compound that mimics the diphosphate group of GGDP. The AHD-bound structure, although it crystallised in another space group, is almost identical to the intermediate-bound one (Fig. 5e, PDB-ID 6GGJ). Therefore, the ordered C-terminus represents the closed state of the enzyme. To test the importance of the C-terminus we designed a truncated construct ending with R294 and again co-crystallised with FGGDP. The resulting structure is almost identical to the open-state structure of CotB2 (rmsd of 0.4 Å for 278 Cα-atoms) and no FGGDP is bound (PDB-ID 6GGK). Activity studies performed by our collaborators (M. Fuchs, C. Görner, T. Brück, TU Munich) showed that the truncation is catalytically inactive (Fig. 4f), underlining the importance of the C-terminus.

3.1.2 Active site
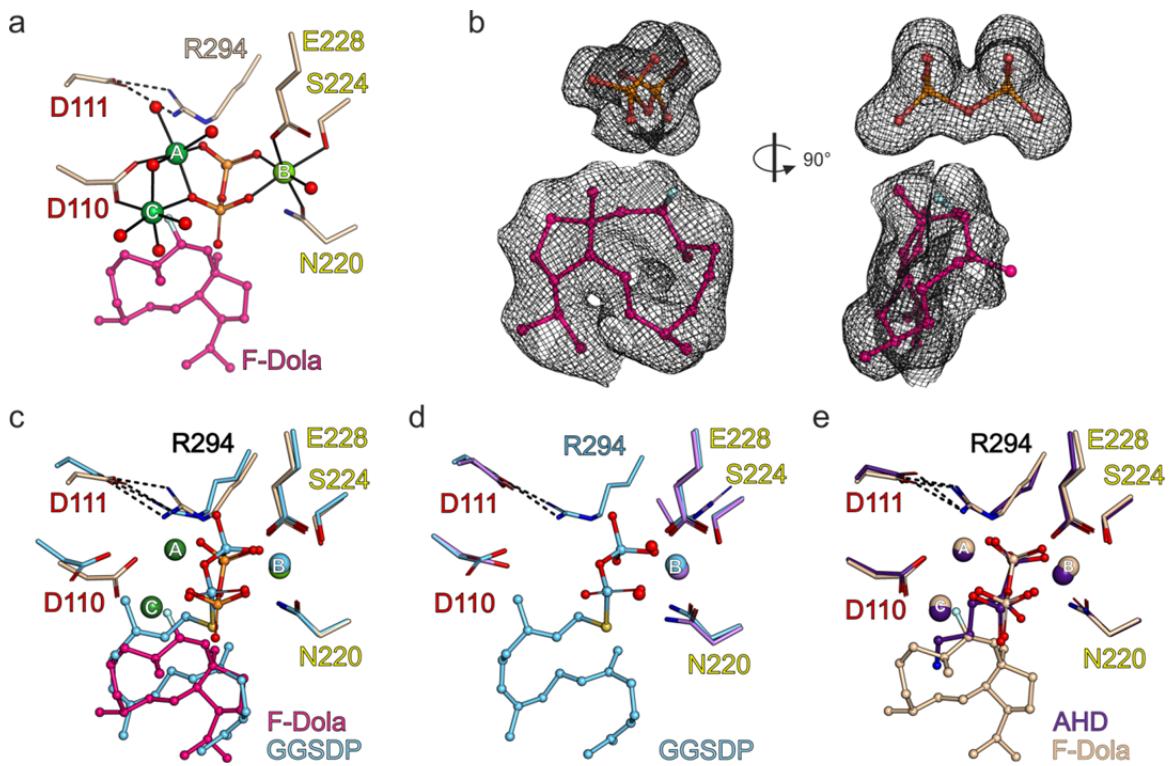
Since CotB2 belongs to the class I synthases, the two metal-binding motifs are present. The aspartate-rich motif **DD**XX**D/E** however is degenerated to [110]**DDPD**[113] in CotB2. In the literature, a couple of other TPSs are known to have altered aspartate-rich motifs as well (*e.g.* Selina-3,7(11)-diene synthase: [82]**DD**GYC**E** (Arigoni, 1975), (+)-T-Muurolol synthase: [83]**DD**EYC**D** (Hu et al., 2011)). The aspartate-rich motif of CotB2 is located at the very end of helix D with P112 introducing a kink, so that the last aspartate D113 is not facing the active site. We therefore presumed that only the first two

aspartates can play a role in catalysis. Our collaborators from TU Munich mutated the three aspartates to glutamates and performed activity assays. D110E and D113E still produce wildtype product, whereas D111E is a dead mutant. Based on the mutations it was not possible to distinguish which aspartate is coordinating the Mg^{2+}-ions. In the closed structure of CotB2, we could subsequently clearly identify the residues that are responsible for metal coordination. D110 is coordinating two Mg^{2+}-ions (Mg^{2+}_A and Mg^{2+}_C), stabilising the diphosphate moiety of the substrate (Fig. 5a). A third Mg^{2+}-ion (Mg^{2+}_B) is bound by the residues N220, S224 and E228 of the NSE/DTE motif (Fig. 5a). Since the mutation D110E does not affect catalysis, the active site is large enough to accommodate a slightly bigger amino acid residue without interfering with metal coordination. D111 is forming a salt-bridge with R294 (Fig. 5a). Interestingly, Tomita and colleagues also observe this salt-bridge in their pre-catalytic state structure (Fig. 5c and d), even though there is only one Mg^{2+}-ion bound, instead of the three that are necessary for catalysis (Tomita et al., 2017). We therefore postulate that binding of a substrate to the enzyme together with Mg^{2+}_B causes the formation of the salt-bridge, but only the full cluster of three Mg^{2+}-ions ensures the correct positioning of the substrate. Nevertheless, the salt-bridge seems to be essential for catalysis, since the mutation D111E renders the enzyme inactive. However, formation of the salt-bridge is not enough to start the cyclisation reaction. Instead, the full C-terminus is necessary as shown by the experiments with the C-terminal truncation of CotB2.

In a recent review, the presence of a "pyrophosphate sensor", an arginine 46 residues upstream the NSE/DTE motif, was discussed as a universal feature of bacterial TPSs (Dickschat, 2016). For CotB2, R175 was predicted to be this conserved residue. However, having solved the closed structure of CotB2 we can now identify R177 to be the pyrophosphate sensor that stretches out to the diphosphate moiety. R175 on the other hand is part of the dimerisation interface and is not in close vicinity to the active site.

Fig. 5 Active site of CotB2 with residues of the aspartate-rich motif labelled red and residues of the NSE/DTE motif labelled yellow, respectively. Hydrogen bonds are indicated by dashed black lines. a) Active site of CotB2 in its closed state (residues coloured in beige) with bound diphosphate and intermediate F-Dola coloured in pink, Mg^{2+}-ions presented as green spheres and water molecules as red spheres. Mg^{2+}-coordination is indicated by solid black lines. b) Polder electron density map (Liebschner et al., 2017) of F-Dola (pink) and diphosphate (orange) at a σ-level of 2.0. c) Overlay of F-Dola- and GGSDP-bound structure (cyan). d) Superposition of Mg^{2+}_B- (light purple) and GGSDP-bound structure. e) Overlay of F-Dola- and AHD-bound structure (purple).

Both the open and the closed structure of CotB2 are high resolution crystal structures (1.6 Å and 1.8 Å resolution, respectively). Surprisingly, in the closed structure, we did not find the electron density to fit the elongated tail of FGGDP. Instead, a ligand with a ring-like structure appeared to have bound deep in the active site together with a diphosphate further up the entrance (Fig. 5b). Both densities were clearly disconnected from each other. We therefore considered an initiated cyclisation reaction to have happened with the cyclised reaction intermediate being trapped in the active site. The reaction presumably halts after the first ring closure, the intermediate being 2-fluoro-3, 7, 18-dolabellatriene (F-Dola) (Fig. 4c). Since the fluorine-atom is very electronegative, the adjacent π-bond has a reduced electron density, making the ring closure at this

position very unlikely. This is the first time that a TPS has been crystallised together with an intermediate product and cleaved-off diphosphate.
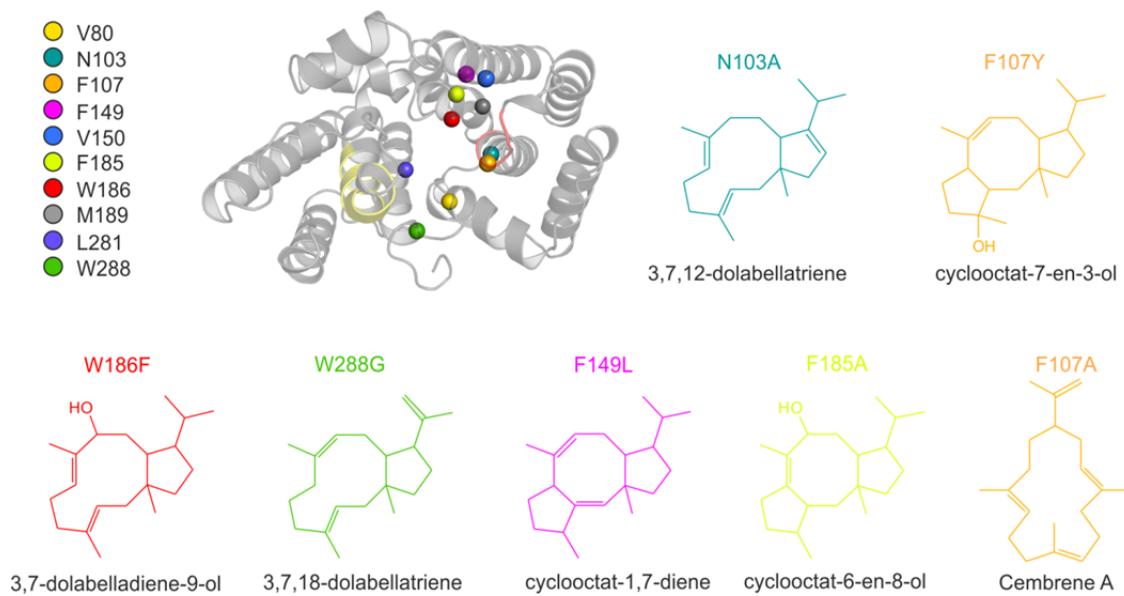
3.1.3 Mutations

It has been shown for a variety of TPSs, that single residue exchanges within the active site can have a major impact on product formation (Hyatt and Croteau, 2005; Lopez-Gallego et al., 2010; Yoshikuni et al., 2006). Not only can an exchange to another amino acid residue lead to a complete loss of enzymatic function, it can even result in a new product, which greatly differs from the native product. In case of CotB2, substitution of a single residue can generate scaffolds ranging from monocyclic to tricyclic products (Fig. 6). It is hypothesised for TPSs that the shape of the active site already determines the product. Therefore, by making the active site larger or smaller it can accommodate different macrocycles.

Our collaborators from TU Munich screened CotB2 for mutants that alter the product spectrum. They identified mutations F149L, W288G, F107Y and F107A to have a major impact on product formation (Fig. 6) (Gorner et al., 2013; Janke et al., 2014). Further, Tomita *et al.* described mutations W186F, F185A and N103A that also produce other macrocycles than CotB2^{wt} (Fig. 6) (Tomita et al., 2017). All these mutations affect residues within the active site. It seems that especially the aromatic residues have a drastic influence on the product formation. Since the products are often polycyclic compounds, the aromatic residues may help in stabilising intermediates via π -stacking.

Based on the crystal structure we proposed new residue exchanges that could be of catalytic interest. Among them is V80L, a chemically rather subtle amino acid exchange that has a massive effect on catalysis, since it renders CotB2 inactive. We can only speculate that the pocket becomes too small to accommodate the substrate.

Fig. 6 Residue exchanges in CotB2 and their effect on product formation. Single point mutations (presented as colourful spheres) within CotB2's active site (CotB2 shown in grey cartoon). Alternative products are depicted with their structural formula.

In addition to the structures described above, we determined various structures of CotB2 mutants in different catalytic states (Table 1). Structures of CotB2 mutants are the basis for quantum mechanics/ molecular mechanics (QM/MM) simulations that give detailed insights into the catalytic mechanism. The group of Prof. Dan Major is currently working with our coordinates to investigate the cyclisation mechanism of the linear GGDP.

Table 1 List of obtained structures of CotB2 mutants

Mutant	Resolution [Å]	State	Active site	PDB-ID
F149L	1.6	open	-	4OMH
F107A	1.9	open	1 Mg-ion	6GGL
F149L	1.4	closed	alendronate	unpublished
L281V	1.5	closed	alendronate	unpublished
M189C	2.2	closed	alendronate	unpublished
M189C	1.9	open	-	unpublished
W288G	2.1	open	-	unpublished
V80L	1.7	open	-	unpublished
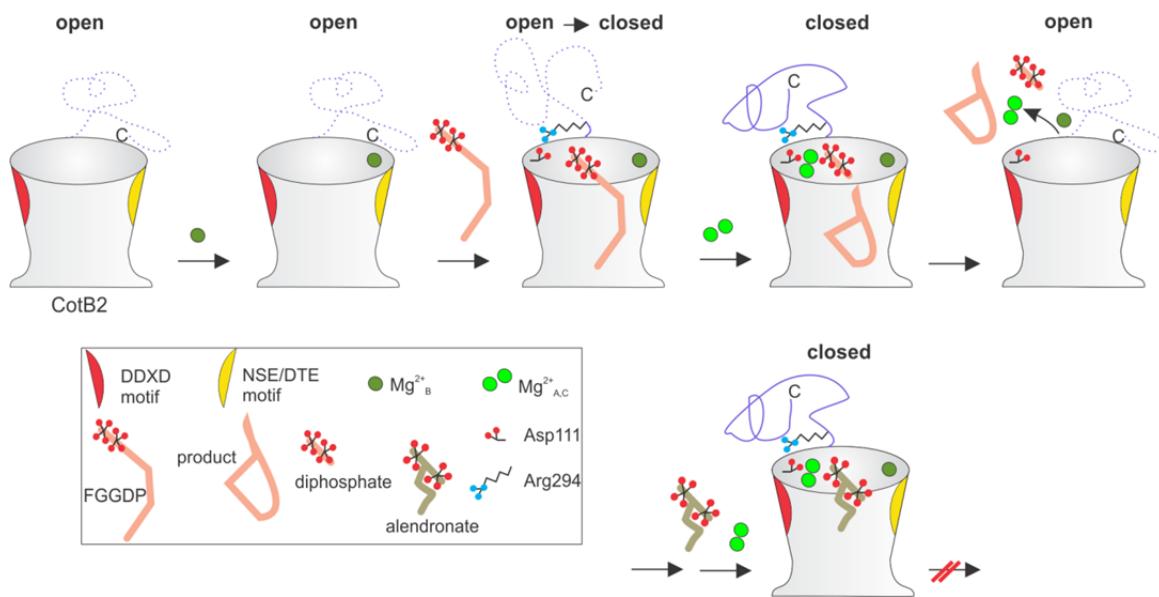
3.1.6 Biotechnological application

Residue exchanges within the active site of CotB2 are not only of interest for elucidating the catalytic mechanism - they also have an application in biotechnology. The new products are scaffolds for the chemical industry and some of them even show bioactivity. For example, 3, 7, 18-dolabellatriene, produced by the CotB2 variant W288G, is an effective antibiotic against the multi-resistant *Staphylococcus aureus* (MRSA) (Ioannou et al., 2011).

Our collaborators from TU Munich incorporate CotB2 and CotB2 variants into their bacterial terpene production pipeline. The different cyclic compounds produced by *E. coli* can be obtained from the medium by liquid-liquid extraction. Subsequently, the compounds are purified by chromatographic methods. Product titers of cyclooctatin and products of CotB2 variants can be obtained in the mg range (*e.g.* cyclooctat-1,7-diene: 3.6 mg per 6 L fermentation)(Gorner et al., 2013).

3.1.5 Mechanism

By solving the structures of CotB2 in different catalytic states, we were able to propose a mechanism for substrate binding (Fig. 7). The first step is binding of an Mg^{2+}-ion (Mg^{2+}_B). Next, the substrate binds *via* its diphosphate moiety and the salt-bridge Asp111-Arg294 is formed. Binding of Mg^{2+}_A and Mg^{2+}_C to the aspartate-rich motif ensures proper coordination of the substrate. Only after the substrate is correctly positioned within the active site, the C-terminus folds over and closes the active site. After removal of the diphosphate moiety cyclisation of the substrate is initiated.

Fig. 7 Schematic view of CotB2's reaction mechanism. CotB2 is represented by a grey cup with the aspartate-rich motif highlighted in red and the NSE/DTE motif in yellow. The reaction starts with binding of the substrate, followed by Mg^{2+}_A and Mg^{2+}_C binding. Only then the C-terminus (shown in purple) becomes folded and closes the active site.

3.1.6 CotB3/CotB4
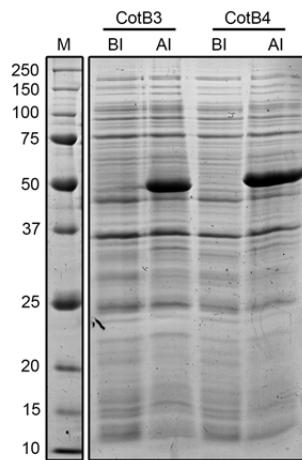
The chemical modifications necessary to obtain the bioactive compound cyclooctatin from cyclooctat-9-en-7-ol are catalysed by the two P450 cytochromes CotB3 and CotB4. Furthermore, CotB3 and CotB4 are potential candidates in the biotechnological terpene production pipeline to diversify the terpene scaffolds. Cytochromes (CYP) hydroxylate non-activated carbon atoms by utilising molecular oxygen. The majority of CYP enzymes feature a haem prosthetic group in their active site, which directs regio-specific substrate oxidation. To accomplish this redox reaction, the cofactor NAD(P)H and a corresponding electron transfer system are required. The redox partner proteins vary depending on the class of the CYP.

As CotB3 and CotB4 contain a haem group, the expression medium was supplemented with aminolevulinic acid. This compound is a precursor for the prosthetic haem group and helps *E. coli* dealing with the high level of CYP expression. Recombinant production of both CotB3 and CotB4 in *E. coli* was feasible (Fig. 8), but the proteins were insoluble. Different vectors, *E. coli* strains and medium compositions were tested, but solubility of the enzymes could not be improved. Therefore, the misfolded protein was

extracted from inclusion bodies and rapid dilution refolding was performed. Unfortunately no soluble protein was obtained by these procedures.


Fig. 8 Coomassie stained SDS-PAGE gel of CotB3 and CotB4 expression levels in *E. coli* whole cell extract (M: Bio-Rad Precision Plus Protein[TM] marker with sizes of reference proteins in kDa, BI: before induction, AI: after induction).

Expressing the cytochromes in a non-bacterial expression system does not seem to be promising as the cytochromes are of bacterial origin. However, an alternative bacterial expression host such as *Streptomyces* could be tried out to express soluble proteins. Furthermore, refolding procedures could be optimised. Apart from different refolding procedures, there are also many chemical additives to choose from that may help the refolding of proteins.
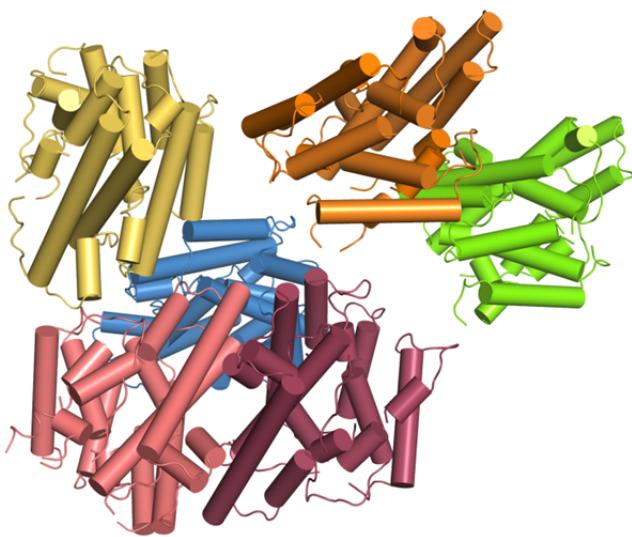
3.2 Diterpene synthase IES

A manuscript that refers to results in this section is currently in preparation. Furthermore, a patent process is ongoing, so that not all findings can be discussed in full detail in this thesis, including origin organism and residue numbering.

The diterpene synthase IES was isolated from a marine organism native to the Caribbean Sea. It produces a series of pseudopterosins, which are diterpene glycosides. These compounds have proven anti-inflammatory, anti-tumour and anti-bacterial properties (Mayer et al., 2010). Pseudopterosins seems to be very effective against multi-drug resistant *Mycobacterium tuberculosis* strains (McCulloch et al., 2012). Furthermore, they are excessively used in the cosmetic industry as anti-irritant additives or anti-aging ingredients (Onumah, 2013; Rietveld et al., 2012).
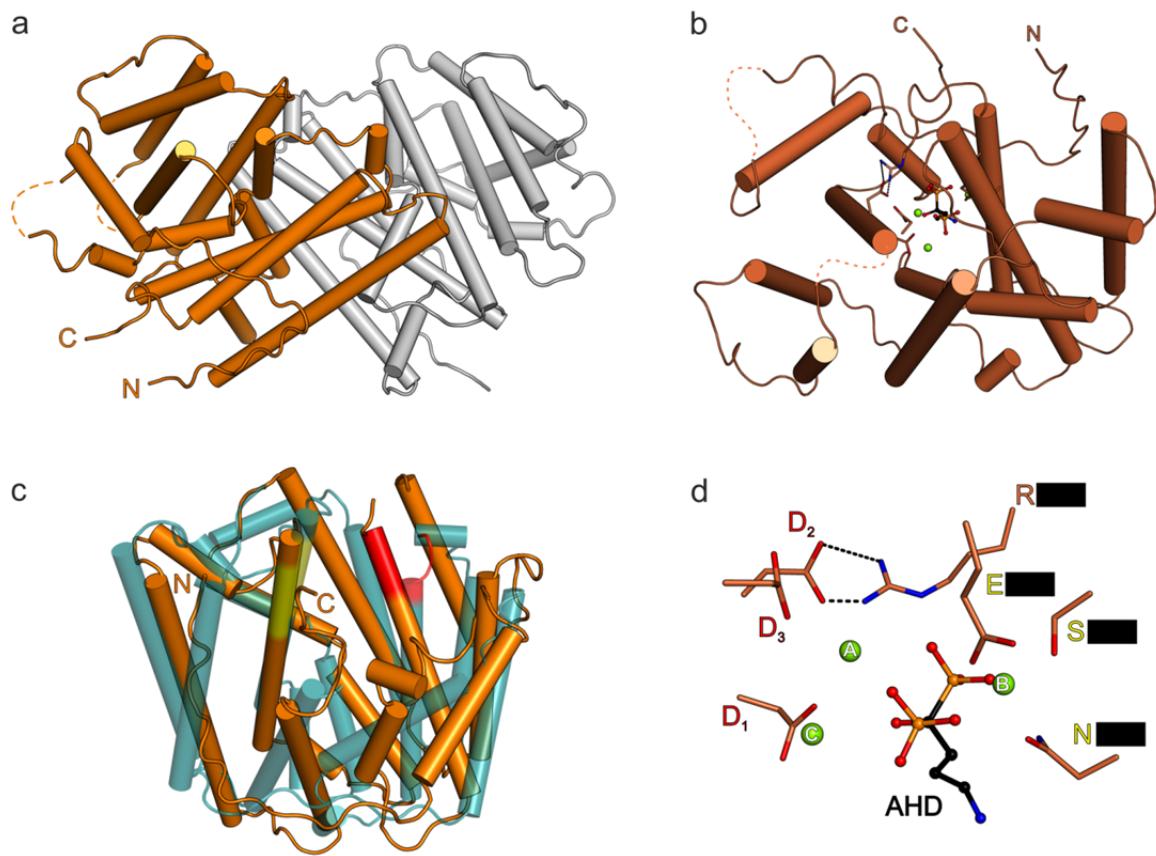
3.2.1 Overall structure

Secondary structure prediction suggested that IES belongs to class I TPSs, adopting the α-helical bundle fold. By elucidating the structure of IES, we could confirm its class I TPS α-fold. Six molecules are present in the asymmetric unit of the crystal (Fig. 9). This assembly however is not of biological relevance since IES forms a dimer in solution (Fig. 10a), as shown *via* gel filtration. Only the two pink molecules in Fig. 9 show the biological assembly, the other dimers are obtained by symmetry operations.

Fig. 9 Assembly of IES molecules in the asymmetric unit of the crystals. Six molecules of IES are present in the asymmetric unit with the two pink molecules representing the biological relevant dimer.

The IES monomer consists of ten long α-helices, connected by short loop fragments and two longer loops that could not be modelled due to flexibility (Fig. 10b). The first N-terminal helix A is not properly folded in the X-ray structure (Fig. 10b). Overlay of the IES structure with the structure of CotB2 bound to AHD resulted in a rmsd of 3.23 Å for 204 aligned Cα-atoms. Superposition of IES and CotB2^{open} resulted in a rmsd of 3.42 Å (Fig. 10c). Since IES was crystallised in the presence of AHD, it seems likely that we captured a transition state between open and closed state, in which the binding site is not fully occupied. In four out of the six molecules additional density within the active site was observed that could represent an AHD molecule and three Mg^{2+}-ions (Fig. 10d). However, since the resolution of our structure is only moderate (3.0 Å), a definite assignment was not possible. In the four molecules with additional density, up to nine residues at the C-terminus could be modelled into the electron density, compared to the two molecules with an empty active site. This observation could indicate that upon ligand binding the C-terminus of IES folds over the active site, similar to the mechanism we could show for CotB2.

Fig. 10 Overall structure of IES. a) IES dimer with residues that could not be modelled indicated by dashed orange lines. b) IES monomer (Chain B) bound to alendronate (AHD). Mg^{2+}-ions are presented as green spheres. c) Superposition of IES monomer (orange) with CotB2^{open} (teal). The aspartate-rich motifs are shown in red, the NSE/DTE motifs in yellow. d) Close-up view of the active site with residues involved in binding of AHD and Mg^{2+}-ions. Residues of the aspartate-rich motif are labelled red and residues of the NSE/DTE motif are labelled yellow, respectively. Mg^{2+}-ions are presented as green spheres and hydrogen bonds are indicated by dashed black lines.

3.2.2 Active site

IES is yet another member of class I TPSs and possesses both the aspartate-rich motif and the NSE/DTE motif. When superimposed with the structure of CotB2, the two metal-binding motifs are positioned at the same relative sites of the proteins (Fig. 10c). As in CotB2, the aspartate-rich motif is unusual as it is longer (**DD**XXX**D**) than the canonical DDXXD motif. A closer look to the active site revealed that the residues of the NSE/DTE motif coordinate one Mg^{2+}-ion (Mg^{2+}_B), similar to CotB2 (Fig. 10d). The aspartates responsible for coordinating Mg^{2+}_A and Mg^{2+}_C are only poorly resolved, as are the Mg^{2+}-ions, most likely because metal ion positions are not fully occupied. Therefore, details of the Mg^{2+}-coordination remain elusive until a better resolved structure of IES is obtained. Based on the present structure, we favour a configuration in which D_1

together with D_3 bind Mg^{2+}_A and Mg^{2+}_C. This situation would stand in sharp contrast to CotB2, where only one aspartate coordinates two Mg^{2+}-ions. The salt-bridge needed for closure of the C-terminus is formed between D_2 and the arginine residue in our structure (Fig. 10d). The presence of the salt-bridge is another hint that we captured the closed or partially closed state of the enzyme.

3.2.3 Biotechnological application

IES produces a mixture of four different compounds. One of them is of high interest for the cosmetic industry, but it only makes up 9 % of the total product mixture. Our collaborators from TU Munich were able to shift product formation towards the desired product by mutating the enzyme. Interestingly, only this mutation yielded well diffracting crystals. IES^{wt} crystals failed to diffract better than 8 Å, even after numerous optimisations of the crystallisation conditions and trying different crystallisation temperatures.

It is not possible to access the desired compound *via* total chemical synthesis. It is currently isolated from its natural source, the soft coral. By applying biotechnological production routes employing modified IES, it may be possible to satisfy the industrial demand for the specific compound in a sustainable fashion.

3.3 Oleate hydratase OhyRe

This section refers to the following publication:

- Jan Lorenzen, Ronja Driller, Ayk Waldow, Farah Qoura, Bernhard Loll, Thomas Brück (2018), *Rhodococcus erythropolis* Oleate Hydratase: a New Member in the Oleate Hydratase Family Tree - Biochemical and Structural Studies, *ChemCatChem* 10:407-414

Fatty acid hydratases catalyse hydration and dehydration reactions of fatty acids (Jin and Hanefeld, 2011). The substrate of oleate hydratases (OHs) is oleic acid which is converted into (*R*)-10-hydroxystearic acid (Fig. 11a) (Bevers et al., 2009). There is an increasing economic demand for OH products since they are used in surfactant and lubricant formulations and could even serve as biofuels (Hou, 2009; Song et al., 2013). All of the identified OHs require the cofactor FAD for catalysis.

OhyRe is a member of the OH family, deriving from the gram-positive marine bacteria *Rhodococcus erythropolis* (Goodfellow and Alderson, 1977). Up to now, eleven sub-families have been identified based on their amino acid sequence (Schmid et al., 2016). Only two structures of OHs had been reported prior to our work, namely those of LAH (belonging to family 2) (Volkov et al., 2013) and OhyA (belonging to family 11) (Engleder et al., 2015). OhyRe is the third OH structure and the first of an enzyme of the sub-family 3 of OHs.
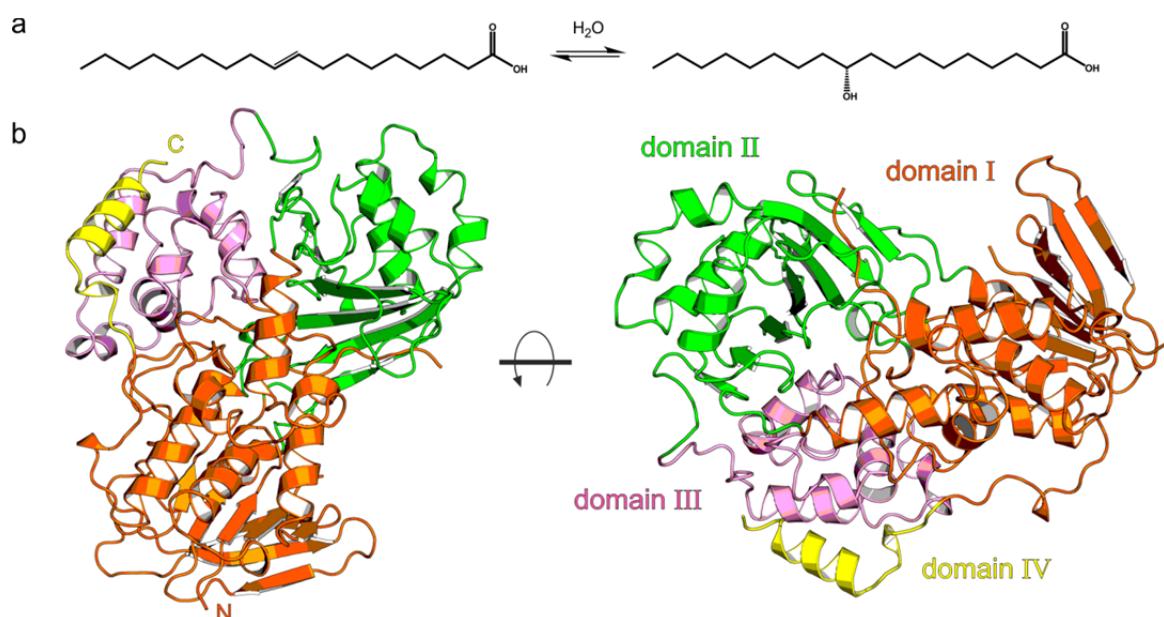
3.3.1 Overall structure

We crystallised OhyRe in space group $P6_522$, solved the structure *via* molecular replacement and refined it at a resolution of 2.64 Å (Lorenzen et al., 2018). The structure represents the open state of the enzyme, since no cofactor or substrate is bound. All attempts to co-crystallise OhyRe with fatty acid substrates failed.

OhyRe consists of three core domains (domain I - III, Fig. 11b) and a C-terminal domain IV. This last domain is only an α-helical segment in OhyRe but is 48 residues longer in OhyA and 32 residues longer in LAH. Domain I consists of two β-sheets and an α-helical part. The orientation of secondary structure elements resembles in parts the

Rossmann fold. As the Rossmann fold is a known cofactor-binding motif, it seems very likely that OhyRe binds FAD in this region. Unfortunately, we did not succeed in obtaining a crystal structure with bound FAD - neither by adding FAD during purification of OhyRe nor by co-crystallisation or soaking into the protein crystals. We therefore presume that FAD binds rather weakly to OhyRe, as was also shown for OhyA (dissociation constant K_d of $1.8 \cdot 10^6$ M). Domain II contains a large antiparallel β-sheet and an α-helical portion, whereas domain III is completely α-helical. It is not entirely clear where the substrate, oleic acid, is binding to OhyRe.


Fig. 11 The bacterial oleate hydratase OhyRe. a) Hydration/ Dehydration of oleic acid/ (*R*)-10-hydroxystearic acid catalysed by OhyRe. b) Overall structure of OhyRe, coloured by its domains (domain I in orange, domain II in green, domain III in pink and domain IV in yellow).

OhyRe is a monomer *in crystallo* as well as in solution, which we proved by MALS. This is in contrast to LAH and OhyA, which are both dimers. As OhyRe is lacking the C-terminal extension of domain IV and is also shorter at the N-terminus compared to the other two OHs, we assume that N- and C-termini play a role in dimerisation. Indeed, in LAH and OhyA the dimerisation interface mainly consists of residues of the N- and C-terminal extensions.

3.3.2 Acitivity and Mutations

Our collaborators from TU Munich biochemically characterised OhyRe and further introduced mutations to alter the enzymes activity. Interestingly, during purification of recombinantly produced protein, FAD is lost. We monitored its presence during purification at 450 nm, FAD's UV absorption maximum, and the signal is not detectable in the protein peak of the size exclusion chromatography. Instead, an additional, late-eluting peak at 450 nm, most likely represents the isolated FAD. Our collaborators found that by adding 20 µM FAD, the enzymatic activity of purified OhyRe could be recovered. OhyRe is rather temperature-tolerant, as it is active in the range from 20 °C to 45 °C with a temperature optimum at 28 °C. OhyRe works best at pH 7.2, but is tolerating pH values from 5.0 to 8.0.
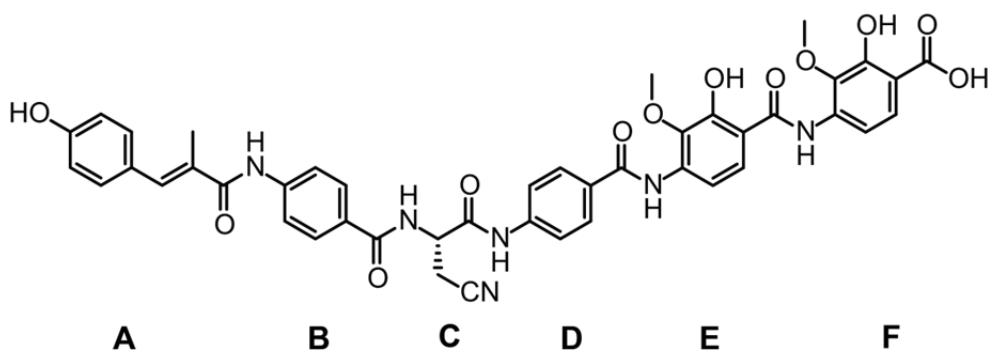
A structural comparison with OhyA revealed that residues M77 and V393 of OhyRe align with residues E122 and T436 of OhyA. The glutamate and threonin are conserved in other OHs and have been linked to enzymatic activity (Engleder et al., 2015). However, M77E and V393T variants of OhyRe resulted in a decrease of activity to 18.9 % (M77E) and 9.5 % (V393T) respectively, relative to wt OhyRe. This observation indicates that the catalytic mechanism of OhyRe differs from that of other known OHs.

3.4 Antibiotic-binding AlbAS

This section refers to the following publication:

- Lida Rostock, Ronja Driller, Stefan Grätz, Dennis Kerwat, Leonard von Eckardstein, Daniel Petras, Claudia Alings, Franz-Josef Schmitt, Thomas Friedrich, Markus Wahl, Bernhard Loll, Andi Mainz, Roderich D. Süssmuth (2018), Molecular insights into antibiotic resistance: how a binding protein traps albicidin, *Nature Communications* 9:3095

The antibiotic albicidin (Fig. 12) inhibits DNA gyrase and is therefore toxic to most Gram-positive and Gram-negative bacteria, such as *S. aureus* and *E. coli* (Hashimi et al., 2007; Kretz et al., 2015). However, some bacteria developed resistance strategies against albicidin. AlbA from *Klebsiella oxytoca* is an albicidin-binding protein that belongs to the MerR (Mercury Resistance (Miller, 1999)) transcriptional regulator family (Walker et al., 1988). To reveal the molecular basis of albicidin sensing AlbA, we analysed a truncated construct, termed AlbAS, which is missing the DNA-binding domain.
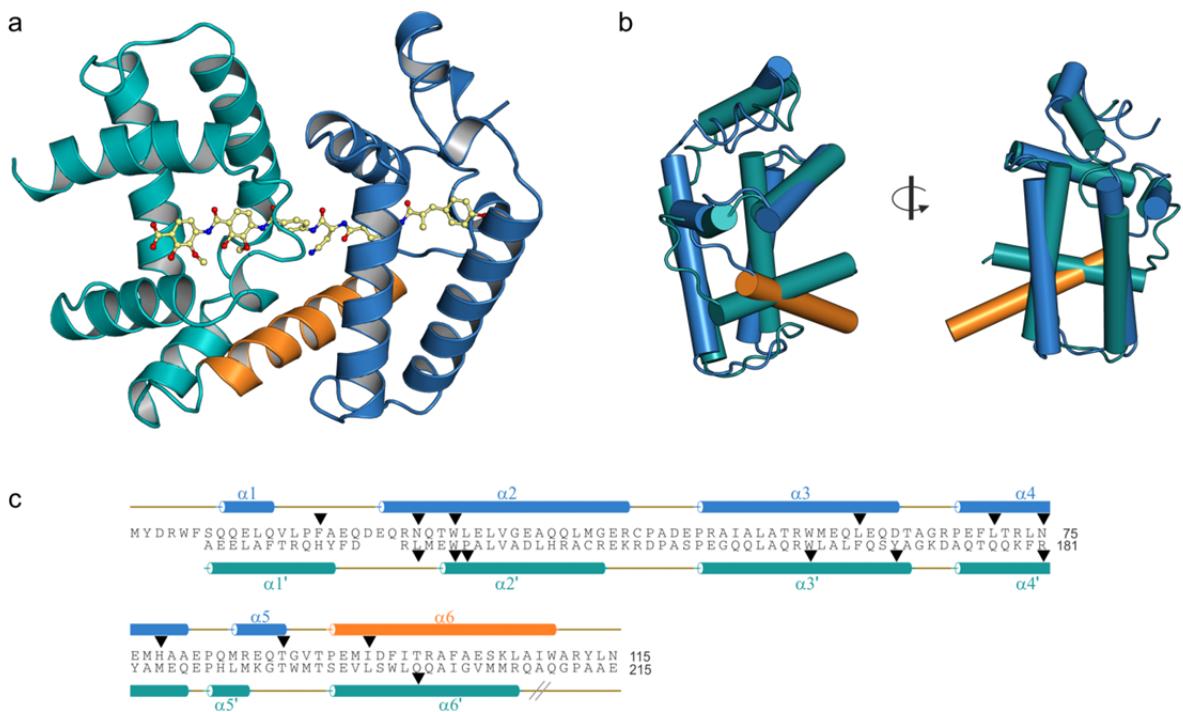

Fig. 12 Chemical structure of albicidin. It can be divided into six distinct building blocks (A-F).

3.4.1 Overall structure

We determined the structure of AlbAS in complex with albicidin by SAD phasing of selenomethionine-labelled protein crystals at a resolution of 1.7 Å (Rostock et al., 2018). AlbAS is a purely α-helical protein, consisting of ten helices. It can be divided into two parts, the N-terminal domain (NTD) and the C-terminal domain (CTD), which are

connected by a long linker helix α6 (Fig. 13a). The structures of NTD and CTD superimpose well as reflected in a rather low rmsd of 1.7 Å (Fig. 13c) for 84 aligned Cα-atoms. Merely the linker helix (grouped with the NTD in the superposition) adopts a slightly different orientation compared to the last helix of the CTD (Fig. 13b). Therefore, it is possible that NTD and CTD arose through gene duplication. Depletion of either one of the domains results in loss of albicidin-binding-capability.


Fig. 13 Overall structure of AlbAS. a) Structure of AlbAS in cartoon representation; NTD is coloured in blue, CTD in cyan and the linker helix in orange. Albicidin is shown as ball-and-stick model coloured in yellow with oxygen atoms coloured in red and nitrogen atoms coloured in blue. b) Superposition of NTD and CTD, helices are shown as cylinders, same colour-code as in a). c) Secondary structure alignment of NTD and CTD. Helices are indicated as solid bars.
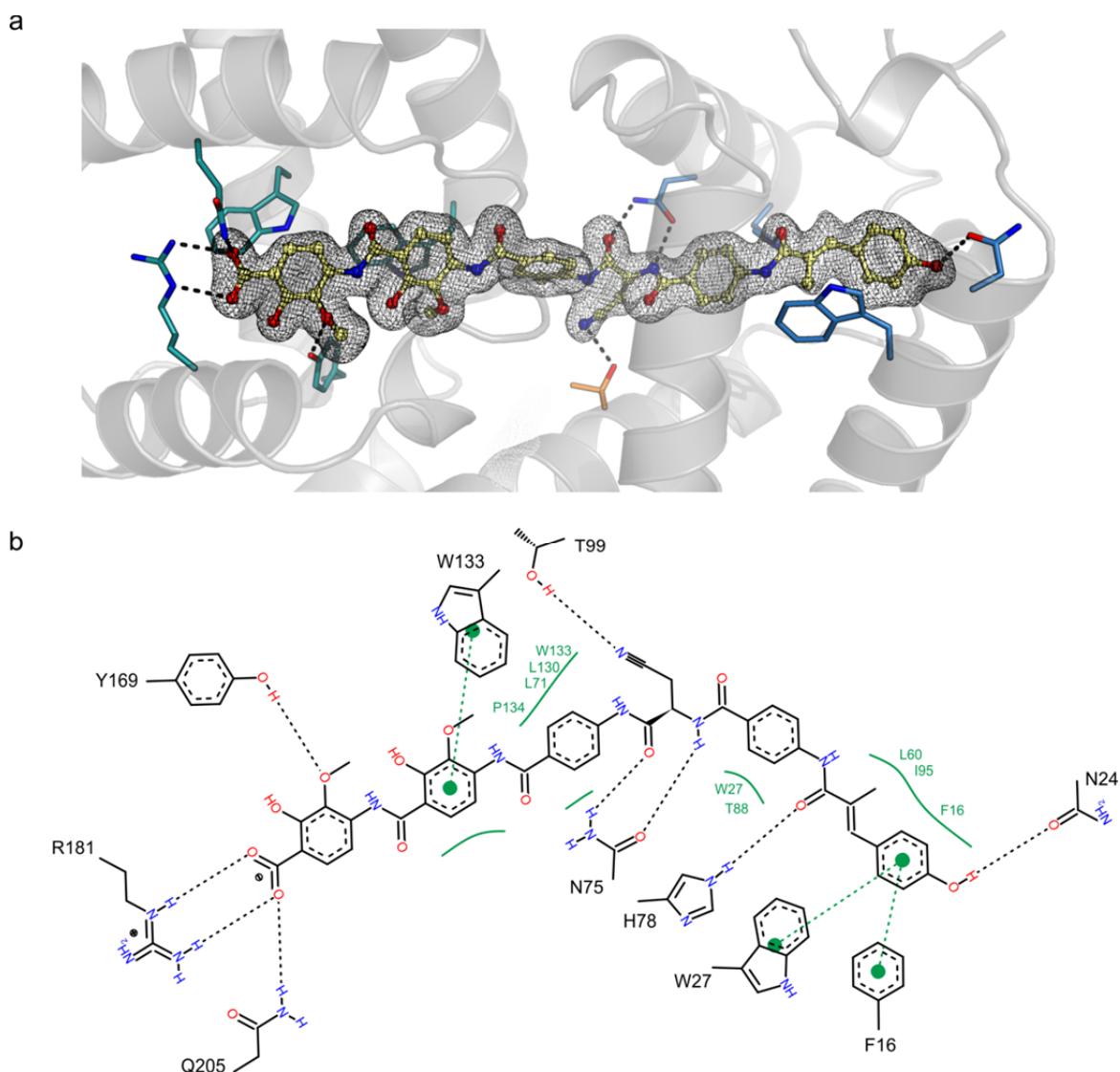
3.4.2 Binding of albicidin

The albicidin molecule binds tightly to AlbAS (K_d of 6 nM), bridging the NTD and CTD (Fig. 13a). It is bound by residues of the NTD, CTD and linker helix, primarily through hydrogen bonding and π-stacking of neighbouring aromatic residues (Fig. 14a). Due to the large number of aromatic residues, the tunnel to which albicidin binds, is rather hydrophobic. Residues F16, W27 and W133 directly contribute to binding *via* π-stacking to albicidin's aromatic groups. Additionally, hydrogen-bonds are formed between albicidin and residues N24, N75, H78, T99, Y169 and Q205 (Fig. 14b). In the CTD of

AlbAS, a prominent salt-bridge between R181 and the carboxylate moiety of albicidin is formed. This salt-bridge might help in ensuring that albicidin is bound to AlbAS in the correct orientation.

The electron density is well defined for each atom of albicidin (Fig. 14a), suggesting a rigid binding to AlbAS.
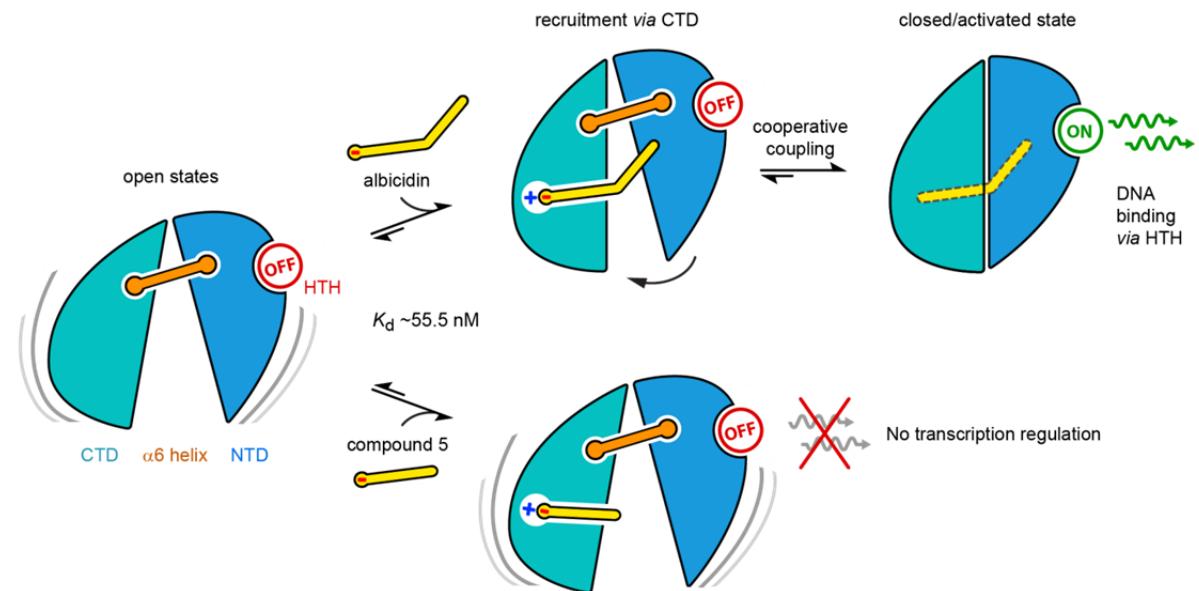
Fig. 14 Binding mode of albicidin to ALbAS. a) Polder map (Liebschner et al., 2017) of albicidins electron density shown as black mesh contoured at 2 σ. Albicidin-interacting residues of AlbAS are shown in stick representation and coloured by the domain they are residing in. c) PoseView (Stierand and Rarey, 2007) ligand interaction network of albicidin. Hydrogen bonds are indicated by black dashed lines and π-stacking by green dashed lines.

To examine whether AlbAS undergoes any conformational change when binding to albicidin, we determined the hydrodynamic radii, R_h, of free and albicidin-bound AlbAS *via* MALS measurements. As albicidin is a rather large ligand compared with the molecular size of AlbAS, one might expect that considerable protein conformational changes are necessary to properly bind it. However, the values of R_h for apo AlbAS (2.0 ± 0.08 nm) and albicidin-bound AlbAS (2.1 ± 0.08 nm) were almost identical. These findings do not exclude conformational changes during binding, but suggest that AlbAS has a large, maybe preformed, substrate-binding pocket.

3.4.3 Mechanism

DNA binding by AlbA is mediated by an N-terminal HTH domain, which is absent in the truncated AlbAS. In its apo state, AlbA is not capable of binding DNA. Upon binding of albicidin, the NTD and CTD of AlbA are rigidified and the protein is activated for DNA binding, thus allowing transcriptional regulation of antibiotic-resistance genes (Fig. 15). Albicidin-fragments lacking the C-terminal part of the compound which are still strongly binding to AlbAS, do not induce DNA binding by AlbA.


Fig. 15 Proposed Mechanism of albicidin-binding in AlbA. The two domains of AlbA (NTD in blue and CTD in cyan) are shown linked by helix α6 coloured in orange.

4. Outlook

A future that does not rely on fossil resources requires exploiting the power of enzymes. Not only can enzymes be used to obtain chemically diverse compounds and antibiotics as in the case of TPSs, they are also necessary for the production of fuels and are essential components in wastewater treatment plants or can be used for the degradation of biodegradable plastics (Roohi et al., 2017). Chen and Yu *et al.* recently reported on enzymes that are able to degrade estrogen (Chen et al., 2017). The increasing amount of estrogen in our water has become a health hazard especially for marine animals.

The concept of active site closure in TPSs is largely accepted, but until now there is only weak experimental evidence. With our data of the closed CotB2 structure, we could demonstrate the importance of the C-terminus for active site closure. These results might also be applicable to other TPSs, as we have evidence that for the diterpene synthase IES the C-terminus again facilitates active site closure. Our structures of CotB2 and its variants can be exploited for extensive computational modeling, whose outcome may predict catalytic details, which in turn may guide further mutagenesis. Since CotB2's product spectrum can be altered by mutagenesis, the CotB2 terpene production chassis by our collaborators from TU Munich generates many different small molecules. These compounds likely possess bioactivity or can act as enzyme inhibitors. It may thus be advantageous to include them in substance collections that are used for high-throughput inhibitor/modulator screening. Furthermore, CotB2 molecules could be coupled to resin to simplify terpene production in an *in vitro* approach. Very recently, Kirschning and co-workers evaluated the substrate flexibility of sesquiterpene synthases (Oberhauser et al., 2018). Interestingly, the synthases were able to convert unnatural farnesyl diphosphate derivatives to heteroatom-modified macrocyclic terpenes. A similar approach could be tried out for diterpene synthases as well. Including unnatural terpene precursors in the terpene production chassis would further enlarge the product spectrum.

TPSs and their variants are a promising toolbox for the chemical and pharmaceutical industry, as they diversify the product spectrum and create access to molecules not available through total chemical synthesis. There are most likely many more TPSs producing pharmacologically interesting products that are not discovered yet. A cumbersome task for the next years will be to identify them in the genomes of plants and bacteria to exploit them for the biotechnological sector. X-ray crystal structures of TPSs can help to characterise them, since it is not always possible to assign the correct product of a novel TPS. Poulter and coworkers presented a computational chemistry approach that uses X-ray crystal structures to predict the function of a TPS (Chow et al., 2015).

To unveil the complete mechanism of AlbA binding to DNA on a structural level, it is necessary to crystallise full-length protein including the DNA-binding domain. Maybe a co-crystallisation of AlbA together with DNA is possible. Furthermore, truncated variants of albicidin or derivatives can be used for co-crystallisation to expand the knowledge of albicidin's binding mode and its release.

Intensive studies on enzymes are necessary to improve catalytic activities and to discover their potential for new applications in a sustainable way of life. Crystal structures are of tremendous help to understand an enzyme's mechanism and to generate novel tailor-made enzymes.

References

Aharoni, A., M.A. Jongsma, and H.J. Bouwmeester. 2005. Volatile science? Metabolic engineering of terpenoids in plants. *Trends Plant Sci*. 10:594-602.

Ajikumar, P.K., W.H. Xiao, K.E. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T.H. Phon, B. Pfeifer, and G. Stephanopoulos. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science*. 330:70-74.

Anthony, J.R., L.C. Anthony, F. Nowroozi, G. Kwon, J.D. Newman, and J.D. Keasling. 2009. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metab Eng*. 11:13-19.

Aoyagi, T., T. Aoyama, F. Kojima, S. Hattori, Y. Honma, M. Hamada, and T. Takeuchi. 1992. Cyclooctatin, a new inhibitor of lysophospholipase, produced by Streptomyces melanosporofaciens MI614-43F2. Taxonomy, production, isolation, physico-chemical properties and biological activities. *J Antibiot (Tokyo)*. 45:1587-1591.

Arigoni, D. 1975. Stereochemical Aspects of Sesquiterpene Biosynthesis. *Pure Appl Chem*. 41:219-245.

Ashby, M.N., and P.A. Edwards. 1990. Elucidation of the deficiency in two yeast coenzyme Q mutants. Characterization of the structural gene encoding hexaprenyl pyrophosphate synthetase. *J Biol Chem*. 265:13157-13164.

Bevers, L.E., M.W. Pinkse, P.D. Verhaert, and W.R. Hagen. 2009. Oleate hydratase catalyzes the hydration of a nonactivated carbon-carbon bond. *J Bacteriol*. 191:5010-5012.

Bruck, T., R. Kourist, and B. Loll. 2014. Production of Macrocyclic Sesqui- and Diterpenes in Heterologous Microbial Hosts: A Systems Approach to Harness Nature's Molecular Diversity. *Chemcatchem*. 6:1142-1165.

Chang, M.C., and J.D. Keasling. 2006. Production of isoprenoid pharmaceuticals by engineered microbes. *Nat Chem Biol*. 2:674-681.

Chen, M., W.K. Chou, T. Toyomasu, D.E. Cane, and D.W. Christianson. 2016. Structure and Function of Fusicoccadiene Synthase, a Hexameric Bifunctional Diterpene Synthase. *ACS chemical biology*. 11:889-899.

References

Chen, Y.L., C.P. Yu, T.H. Lee, K.S. Goh, K.H. Chu, P.H. Wang, W. Ismail, C.J. Shih, and Y.R. Chiang. 2017. Biochemical Mechanisms and Catabolic Enzymes Involved in Bacterial Estrogen Degradation Pathways. *Cell Chem Biol*. 24:712-724 e717.

Chow, J.Y., B.X. Tian, G. Ramamoorthy, B.S. Hillerich, R.D. Seidel, S.C. Almo, M.P. Jacobson, and C.D. Poulter. 2015. Computational-guided discovery and characterization of a sesquiterpene synthase from Streptomyces clavuligerus. *Proc Natl Acad Sci U S A*. 112:5661-5666.

Christianson, D.W. 2006. Structural biology and chemistry of the terpenoid cyclases. *Chemical reviews*. 106:3412-3442.

Christianson, D.W. 2017. Structural and Chemical Biology of Terpenoid Cyclases. *Chem Rev*. 117:11570-11648.

de Lorenzo, V., K.L.J. Prather, G.Q. Chen, E. O'Day, C. von Kameke, D.A. Oyarzun, L. Hosta-Rigau, H. Alsafar, C. Cao, W.Z. Ji, H. Okano, R.J. Roberts, M. Ronaghi, K. Yeung, F. Zhang, and S.Y. Lee. 2018. The power of synthetic biology for bioproduction, remediation and pollution control: The UN's Sustainable Development Goals will inevitably require the application of molecular biology and biotechnology on a global scale. *Embo Rep*. 19.

Dickschat, J.S. 2016. Bacterial terpene cyclases. *Nat Prod Rep*. 33:87-110.

Engleder, M., T. Pavkov-Keller, A. Emmerstorfer, A. Hromic, S. Schrempf, G. Steinkellner, T. Wriessnegger, E. Leitner, G.A. Strohmeier, I. Kaluzna, D. Mink, M. Schurmann, S. Wallner, P. Macheroux, K. Gruber, and H. Pichler. 2015. Structure-Based Mechanism of Oleate Hydratase from Elizabethkingia meningoseptica. *Chembiochem*. 16:1730-1734.

Goodfellow, M., and G. Alderson. 1977. The actinomycete-genus Rhodococcus: a home for the "rhodochrous" complex. *J Gen Microbiol*. 100:99-122.

Gorner, C., I. Hauslein, P. Schrepfer, W. Eisenreich, and T. Bruck. 2013. Targeted Engineering of Cyclooctat-9-en-7-ol Synthase: A Stereospecific Access to Two New Non-natural Fusicoccane-Type Diterpenes. *Chemcatchem*. 5:3289-3298.

Greenhagen, B.T., P.E. O'Maille, J.P. Noel, and J. Chappell. 2006. Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proc Natl Acad Sci U S A*. 103:9826-9831.

Hashimi, S.M., M.K. Wall, A.B. Smith, A. Maxwell, and R.G. Birch. 2007. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob Agents Chemother*. 51:181-187.

Hou, C.T. 2009. Biotechnology for fats and oils: new oxygenated fatty acids. *N Biotechnol*. 26:2-10.

Hu, Y., W.K. Chou, R. Hopson, and D.E. Cane. 2011. Genome mining in Streptomyces clavuligerus: expression and biochemical characterization of two new cryptic sesquiterpene synthases. *Chem Biol*. 18:32-37.

Hyatt, D.C., and R. Croteau. 2005. Mutational analysis of a monoterpene synthase reaction: altered catalysis through directed mutagenesis of (-)-pinene synthase from Abies grandis. *Arch Biochem Biophys*. 439:222-233.

Ioannou, E., A. Quesada, M.M. Rahman, S. Gibbons, C. Vagias, and V. Roussis. 2011. Dolabellanes with antibacterial activity from the brown alga Dilophus spiralis. *J Nat Prod*. 74:213-222.

Janke, R., C. Gorner, M. Hirte, T. Bruck, and B. Loll. 2014. The first structure of a bacterial diterpene cyclase: CotB2. *Acta Crystallogr D Biol Crystallogr*. 70:1528-1537.

Jin, J., and U. Hanefeld. 2011. The selective addition of water to C=C bonds; enzymes are the best chemists. *Chem Commun (Camb)*. 47:2502-2510.

Kim, S.Y., P. Zhao, M. Igarashi, R. Sawa, T. Tomita, M. Nishiyama, and T. Kuzuyama. 2009. Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two p450 hydroxylases. *Chem Biol*. 16:736-743.

Kirby, J., and J.D. Keasling. 2009. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu Rev Plant Biol*. 60:335-355.

Koksal, M., W.K. Chou, D.E. Cane, and D.W. Christianson. 2012. Structure of 2-methylisoborneol synthase from Streptomyces coelicolor and implications for the cyclization of a noncanonical C-methylated monoterpenoid substrate. *Biochemistry*. 51:3011-3020.

Koksal, M., Y. Jin, R.M. Coates, R. Croteau, and D.W. Christianson. 2011. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature*. 469:116-120.

Kretz, J., D. Kerwat, V. Schubert, S. Gratz, A. Pesic, S. Semsary, S. Cociancich, M. Royer, and R.D. Sussmuth. 2015. Total synthesis of albicidin: a lead structure from Xanthomonas albilineans for potent antibacterial gyrase inhibitors. *Angew Chem Int Ed Engl*. 54:1969-1973.

Liebschner, D., P.V. Afonine, N.W. Moriarty, B.K. Poon, O.V. Sobolev, T.C. Terwilliger, and P.D. Adams. 2017. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol*. 73:148-157.

References

Liu, W., X. Feng, Y. Zheng, C.H. Huang, C. Nakano, T. Hoshino, S. Bogue, T.P. Ko, C.C. Chen, Y. Cui, J. Li, I. Wang, S.T. Hsu, E. Oldfield, and R.T. Guo. 2014. Structure, function and inhibition of ent-kaurene synthase from Bradyrhizobium japonicum. *Sci Rep*. 4:6214.

Lopez-Gallego, F., G.T. Wawrzyn, and C. Schmidt-Dannert. 2010. Selectivity of fungal sesquiterpene synthases: role of the active site's H-1 alpha loop in catalysis. *Appl Environ Microbiol*. 76:7723-7733.

Lorenzen, J., R. Driller, A. Waldow, F. Qoura, B. Loll, and T. Bruck. 2018. Rhodococcus erythropolis Oleate Hydratase: a New Member in the Oleate Hydratase Family TreeBiochemical and Structural Studies. *Chemcatchem*. 10:407-414.

Mayer, A.M., K.B. Glaser, C. Cuevas, R.S. Jacobs, W. Kem, R.D. Little, J.M. McIntosh, D.J. Newman, B.C. Potts, and D.E. Shuster. 2010. The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends Pharmacol Sci*. 31:255-265.

McCulloch, M.W., B. Haltli, D.H. Marchbank, and R.G. Kerr. 2012. Evaluation of pseudopteroxazole and pseudopterosin derivatives against Mycobacterium tuberculosis and other pathogens. *Mar Drugs*. 10:1711-1728.

Miller, S.M. 1999. Bacterial detoxification of Hg(II) and organomercurials. *Essays Biochem*. 34:17-30.

Newman, D.J., and G.M. Cragg. 2016. Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod*. 79:629-661.

Oberhauser, C., V. Harms, K. Seidel, B. Schroder, K. Ekramzadeh, S. Beutel, S. Winkler, L. Lauterbach, J.S. Dickschat, and A. Kirschning. 2018. Exploiting the Synthetic Potential of Sesquiterpene Cyclases for Generating Unnatural Terpenoids. *Angew Chem Int Ed Engl*. 57:11802-11806.

Oldfield, E., and F.Y. Lin. 2012. Terpene biosynthesis: modularity rules. *Angew Chem Int Ed Engl*. 51:1124-1137.

Onumah, N. 2013. A novel anti-inflammatory in treatment of acne vulgaris: the pseudopterosins. *J Drugs Dermatol*. 12:1177-1179.

Rietveld, M., D. Janson, R. Siamari, J. Vicanova, M.T. Andersen, and A. El Ghalbzouri. 2012. Marine-derived nutrient improves epidermal and dermal structure and prolongs the life span of reconstructed human skin equivalents. *J Cosmet Dermatol*. 11:213-222.

Roohi, K. Bano, M. Kuddus, M.R. Zaheer, Q. Zia, M.F. Khan, G.M. Ashraf, A. Gupta, and G. Aliev. 2017. Microbial Enzymatic Degradation of Biodegradable Plastics. *Curr Pharm Biotechnol*. 18:429-440.

Rostock, L., R. Driller, S. Gratz, D. Kerwat, L. von Eckardstein, D. Petras, M. Kunert, C. Alings, F.J. Schmitt, T. Friedrich, M.C. Wahl, B. Loll, A. Mainz, and R.D. Sussmuth. 2018. Molecular insights into antibiotic resistance - how a binding protein traps albicidin. *Nat Commun*. 9:3095.

Schmid, J., L. Steiner, S. Fademrecht, J. Pleiss, K.B. Otte, and B. Hauer. 2016. Biocatalytic study of novel oleate hydratases. *Journal of Molecular Catalysis B: Enzymatic*. 133:S243-S249.

Song, J.W., E.Y. Jeon, D.H. Song, H.Y. Jang, U.T. Bornscheuer, D.K. Oh, and J.B. Park. 2013. Multistep enzymatic synthesis of long-chain alpha,omega-dicarboxylic and omega-hydroxycarboxylic acids from renewable fatty acids and plant oils. *Angew Chem Int Ed Engl*. 52:2534-2537.

Stierand, K., and M. Rarey. 2007. From modeling to medicinal chemistry: Automatic generation of two-dimensional complex diagrams. *Chemmedchem*. 2:853-860.

Tomita, T., S.Y. Kim, K. Teramoto, A. Meguro, T. Ozaki, A. Yoshida, Y. Motoyoshi, N. Mori, K. Ishigami, H. Watanabe, M. Nishiyama, and T. Kuzuyama. 2017. Structural Insights into the CotB2-Catalyzed Cyclization of Geranylgeranyl Diphosphate to the Diterpene Cyclooctat-9-en-7-ol. *ACS chemical biology*. 12:1621-1628.

Volkov, A., S. Khoshnevis, P. Neumann, C. Herrfurth, D. Wohlwend, R. Ficner, and I. Feussner. 2013. Crystal structure analysis of a fatty acid double-bond hydratase from Lactobacillus acidophilus. *Acta Crystallogr D Biol Crystallogr*. 69:648-657.

Walker, M.J., R.G. Birch, and J.M. Pemberton. 1988. Cloning and characterization of an albicidin resistance gene from Klebsiella oxytoca. *Mol Microbiol*. 2:443-454.

Whittington, D.A., M.L. Wise, M. Urbansky, R.M. Coates, R.B. Croteau, and D.W. Christianson. 2002. Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc Natl Acad Sci U S A*. 99:15375-15380.

Yoshikuni, Y., T.E. Ferrin, and J.D. Keasling. 2006. Designed divergent evolution of enzyme function. *Nature*. 440:1078-1082.

Abbreviations

°C	degree celsius
Å	ångström
AHD	alendronate
CTD	C-terminal domain
CYP	cytochrome
DLS	dynamic light scattering
DMAPP	dimethylallyl diphosphate
E. coli	*Escherichia coli*
FAD	flavin adenine dinucleotide
FGGDP	2-fluoro geranylgeranyl diphosphate
F-Dola	2-fluoro-3, 7, 18-dolabellatriene
GC-MS	Gas chromatography-mass spectrometry
GGDP	geranylgeranyl diphosphate
GGSDP	geranylgeranyl thiodiphosphate
IPP	isopentyl diphosphate
K_d	dissociation constant
kDa	kilodalton
MALS	multi-angle light scattering
NTD	N-terminal domain
OH	oleate hydratase
PDB	RCSB Protein Data Bank
QM/MM	quantum mechanics/ molecular mechanics
R_h	hydrodynamic radius
rmsd	root-mean-square deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamid gel electrophoresis
S. aureus	*Staphylococcus aureus*
TPS	terpene synthase
wt	wild type

Table of Figures

Acknowledgement

First of all, I would like to thank Dr. Bernhard Loll for his superb supervision. You never failed to support and encourage me, to motivate me to go on even if things were not going smoothly. Thank you for sharing all your knowledge about crystallography with me. I appreciate your trust in me early on in my Ph.D. to let me present our results.

Thank you to Prof. Dr. Markus Wahl for giving me the opportunity to work in his lab and for supporting me throughout my Ph.D. Huge thanks to Prof. Dr. Thomas Brück for letting me work on his exciting projects.

Further, I would like to thank the whole Wahl lab for the fun time, fruitful discussions and constant support. During my time in the lab I met so many wonderful people with whom it has been a pleasure to work with. Special thanks to Claudia Alings, Clemens Langner, Karin Hesse, Nicole Holton, Dr. Eva Absmeier and Jan Driller for your help and motivation. Thank you to Jan and Eva for reviewing my thesis. Thanks also go to my students, especially to Ayk Waldow and Bobby Draegert. And thank you Francesca De Bortoli for sharing the yummiest treats with me.

Thank you to my collaborators of the projects presented in this thesis: Lida Rostock, Christian Görner, Jan Lorenzen, Monika Fuchs, Sophie Janke, Prof. Dr. Mathias Christmann and Prof. Dr. Dan Thomas Major. Thanks also go to my collaborators and mentors of projects not related to this thesis, namely Prof. Dr. Florian Heyd, Dr. Christoph Böttcher, Dr. Christoph Weise and Dr. Benno Kuropka.
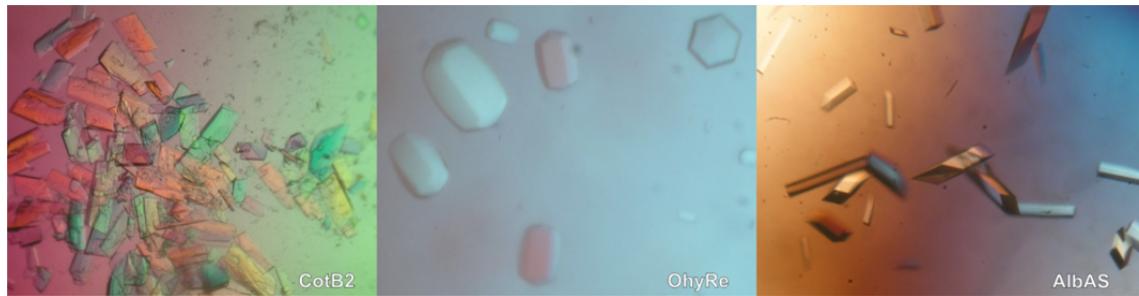
A big thank you to the beamline scientists for help and access to the MX-beamlines of Bessy II storage ring (Berlin) and beamlines P11 and P14 of Petra III at DESY (Hamburg), special thanks to Dr. Manfred Weiss for career advice.

I am very grateful to the Christiane Nüsslein-Volhard foundation and Elsa-Neumann stipend (NaFöG) for funding me throughout my Ph.D.

Last but not least, I would like to thank my family and friends. You made it possible for me to arrange family and career. I am a very lucky person to have you in my life.

And thank you to the protein samples that eventually crystallised. I would not have been able to complete my Ph.D. without you. ;)

Publication I

The first structure of a bacterial diterpene cyclase: CotB2
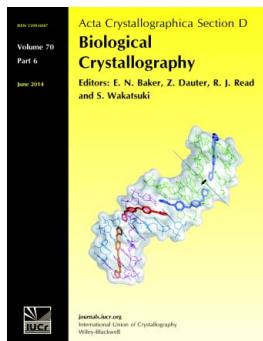
Ronja Janke, Christian Görner, Max Hirte, Thomas Brück and Bernhard Loll

Acta Crystallographica Section D: Biological Crystallography welcomes the submission of papers covering any aspect of structural biology, with a particular emphasis on the structures of biological macromolecules and the methods used to determine them. Reports on new protein structures are particularly encouraged, as are structure–function papers that could include crystallographic binding studies, or structural analysis of mutants or other modified forms of a known protein structure. The key criterion is that such papers should present new insights into biology, chemistry or structure. Papers on crystallographic methods should be oriented towards biological crystallography, and may include new approaches to any aspect of structure determination or analysis. Papers on the crystallization of biological molecules will be accepted providing that these focus on new methods or other features that are of general importance or applicability.

Crystallography Journals Online is available from **journals.iucr.org**

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

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 activ-
ities. 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 resolu-
tion. 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 dime-
thylallyl diphosphate (Poulter *et al.*, 1978) to the linear terpene
cyclase substrates geranyl diphosphate (C_{10}), farnesyl dipho-
sphate (C_{15}), geranylgeranyl diphosphate (C_{20}) and geranyl-
farnesyl 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 precur-
sors to build up a vast number of polycyclic hydrocarbon
scaffolds. Remarkably, this complex chemical reaction,
comprising changes in bonding, hybridization and the intro-
duction 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 reproto-
nation 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$X$$X$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 DXDD 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 mM Tris–HCl pH 7.5, 500 mM NaCl, 2 mM MgCl$_2$, 1 mM 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 ($\sim$1 ml column volume; GE Healthcare) column was equilibrated with buffer A. CotB2^{wt} was loaded onto the column and washed with ten column volumes of buffer A. CotB2^{wt} was eluted in a linear gradient to 50 mM Tris–HCl pH 7.5, 500 mM NaCl, 2 mM MgCl$_2$, 1 mM DTT, 500 mM imidazole. Size-exclusion chromatography was performed with a HiLoad Superdex S200 16/60 column (GE Healthcare) equilibrated with buffer B (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$, 1 mM 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 $\sim$0.6 by the addition of 0.5 mM IPTG. Purification was carried out as described above for CotB2^{wt}, except that all buffers contained 4 mM 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 mM Tris–HCl pH 7.5, 1 mM MgCl$_2$ supplemented with 45 µM 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 $\delta = 7.26$ p.p.m. (¹H NMR) or CDCl$_3$ at $\delta = 77.16$ p.p.m. (¹³C 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 mM 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 mM 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	50.00–1.64	50.00–1.64
	(2.57–2.50)	(1.68–1.64)	(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 (Å²)			
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 mM NaH$_2$PO$_4$. 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).
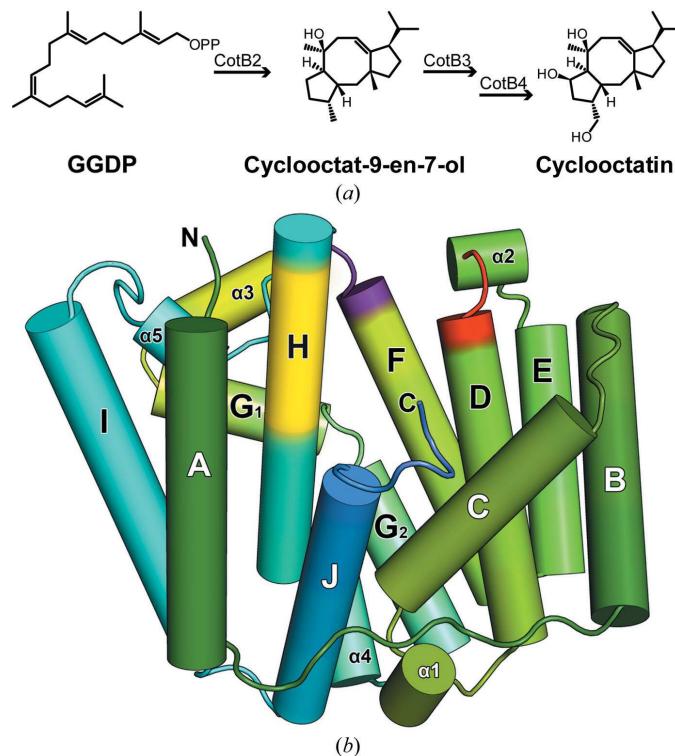
3. Results

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$^\alpha$ 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

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

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).

3.3. Plasticity of the residues in the active site of CotB2
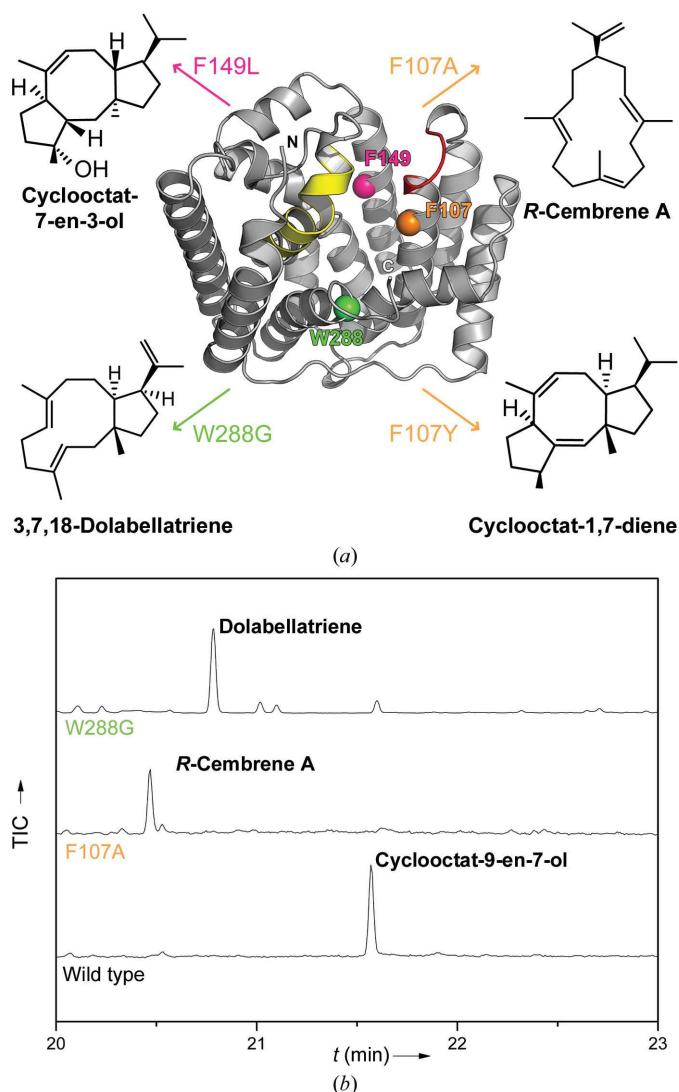
In previous work, we constructed a homology model of CotB2 (Görner *et al.*, 2013) that was based on the structure of the α-domain of taxadiene synthase (TXS; Köksal, Jin *et al.*, 2011). This homology model enabled us to identify single amino-acid residues that are potentially important for carbocation 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


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.

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.

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. 2a). 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. 2b). 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, ^3J(H,H) = 7.6 Hz, 1H], 5.00 [t, ^3J(H,H) = 6.2 Hz, 1H] and 5.08 [t, ^3J(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. 3a and Supplementary Fig. S3). The overall arrangement of the core α-helices of both proteins is comparable (Fig. 3a

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. 3a). 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}DDXXD^{617}) and ^{110}DDXD^{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. 3b). 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.
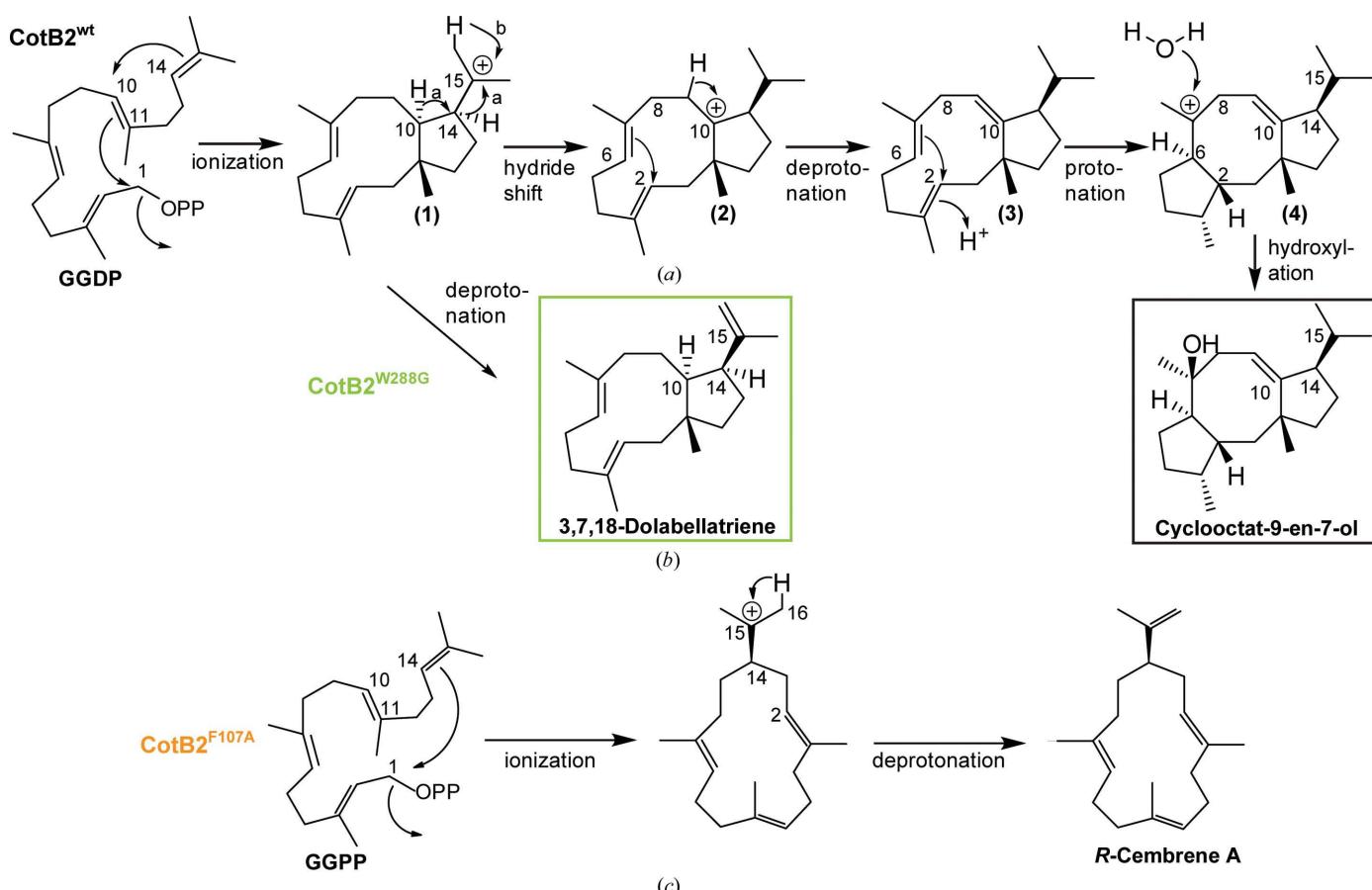
sphate function of GGDP. The guanidinium function of Arg754 of TXS and Arg177 of CotB2 superimpose (Fig. 3b), even though they reside on different core α-helices. Notably, the recognition and the mechanism of diphosphate activation seem to be conserved between plant and bacterial diterpene and monoterpene 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 steric clashes of CotB2 amino-acid side chains with 2-fluorogeranylgeranyl diphosphate. These clashes are observed at the bottom of the active site (Figs. 3b and 3c) 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 ^1H and ^{13}C spectra diverged by less than 0.2 p.p.m. (^{13}C) and 0.02 p.p.m. (^1H) 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 (1R,3E,7E,11S,12S)-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 cyclooct-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 cyclooct-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 cyclooct-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 cyclooct-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 pentalene synthase from *Streptomyces* UC5319 reveals a superposition of catalytically important residues. Residues Phe107 and Phe149 of CotB2 superimpose with Phe77 and Tyr146 of pentalene 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 pentalene (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.

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.

Supplemental information to:

Crystal structure of the first bacterial diterpene cyclase: CotB2

Ronja Janke[1,4], Christian Görner[2,4], Thomas Brück[2,*], and Bernhard Loll[1,*]

[1] Institut für Chemie und Biochemie, Abteilung Strukturbiochemie, Freie Universität Berlin, Takusstr. 6, 14195 Berlin, Germany

[2] Fachgebiet Industrielle Biokatalyse, Technische Universität München, Garching, Germany

[4] These authors contributed equally to this work

* To whom correspondence should be addressed:

Thomas Brück

Fachgebiet für Industrielle Biokatalyse

Zentrum für Weiße Biotechnologie

Department für Chemie

Technische Universität München

Lichtenberg Str. 4, 85748 Garching, Germany

E-mail: brueck@tum.de

Bernhard Loll

Institut für Chemie und Biochemie

Abteilung Strukturbiochemie

Freie Universität Berlin

Takusstr. 6,

14195 Berlin, Germany

Email: loll@chemie.fu-berlin.de
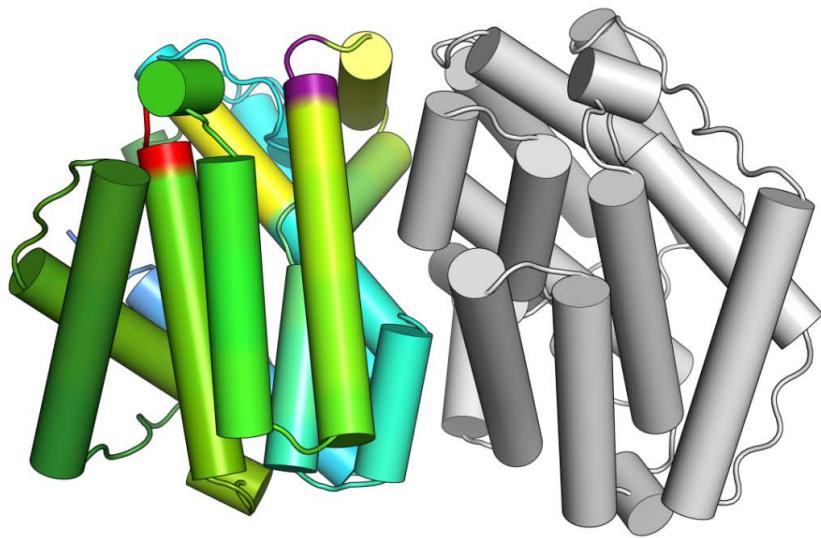
Table S1 Phasing statistics

Phasing statistics	
resolution [Å] [a]	29.66 - 2.50
	(2.56 - 2.50)
R_{cullis} (acentric) [b]	0.864 (0.964)
phasing power (acentric) [c]	1.042 (0.698)
no. of selenium sites	16
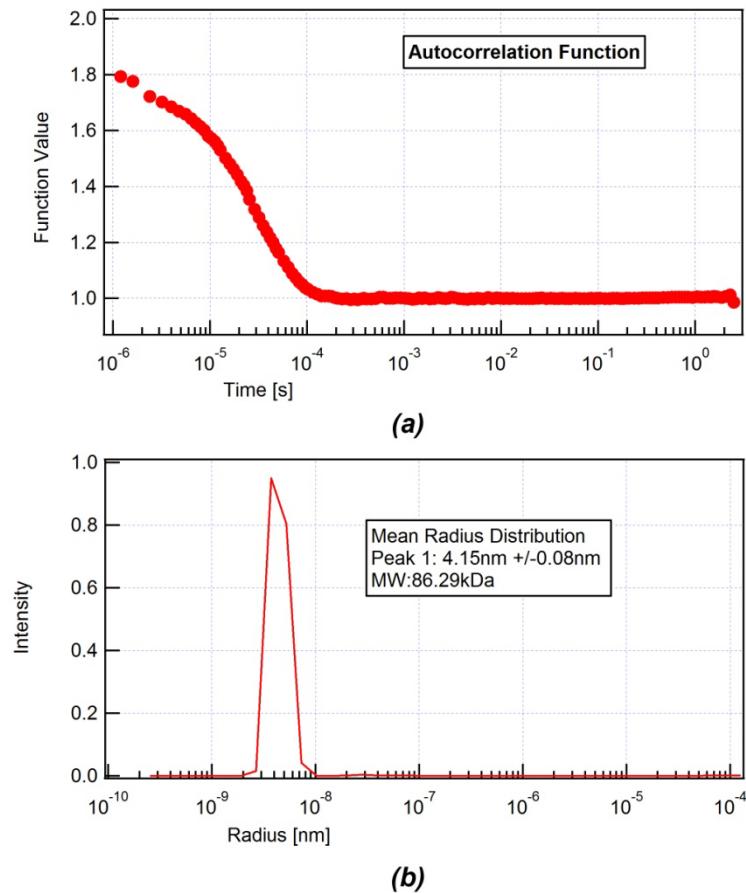
[a] values in parentheses refer to the highest resolution shell.

[b] $R_{cullis} = \langle\text{phase-integrated lack of closure}\rangle \; / \; \langle|F_{PH} - F_P|\rangle$
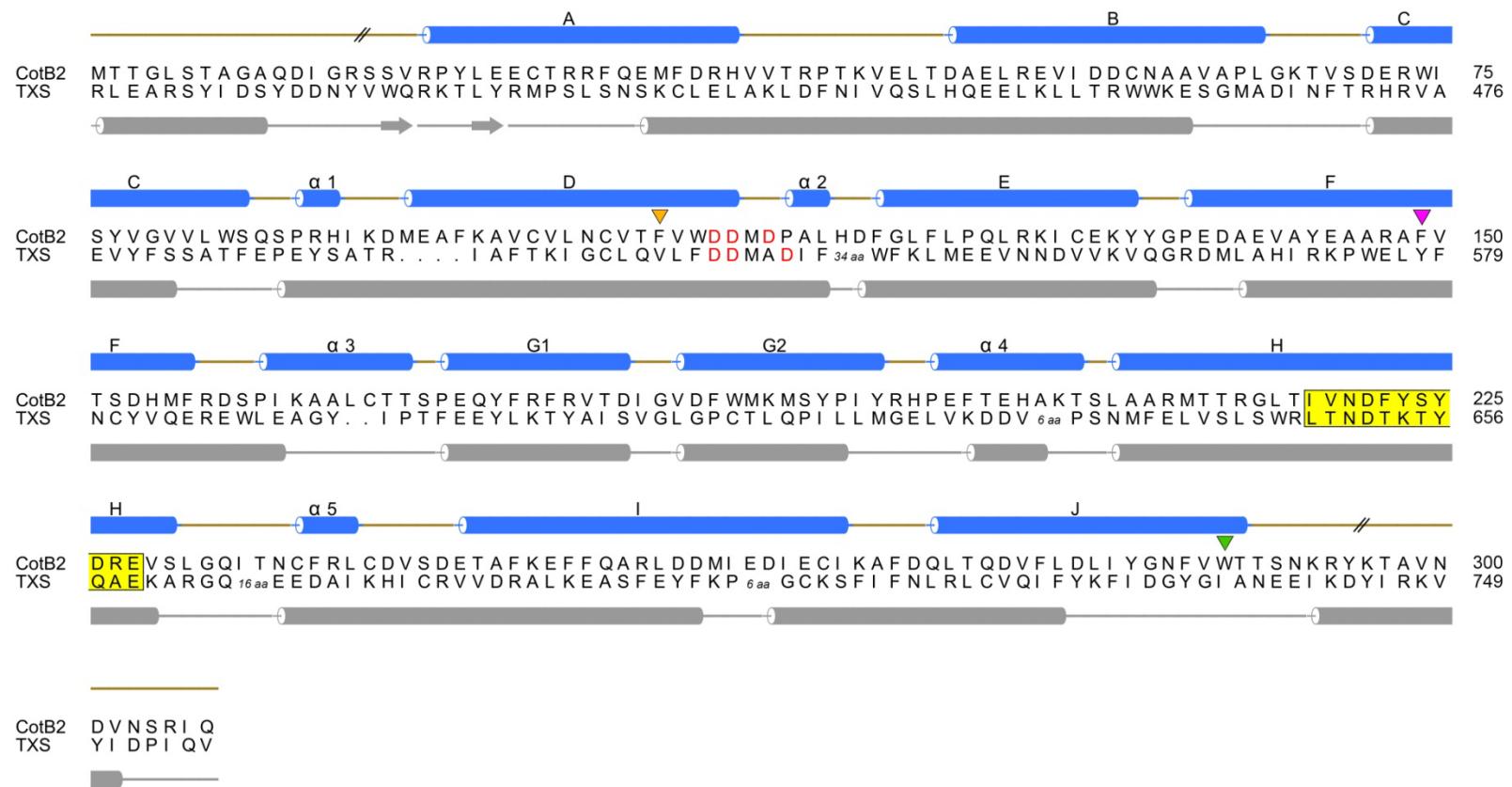
[c] Phasing power $= \langle[|F_H(\text{calc})|/\text{phase-integrated lack of closure}]\rangle$
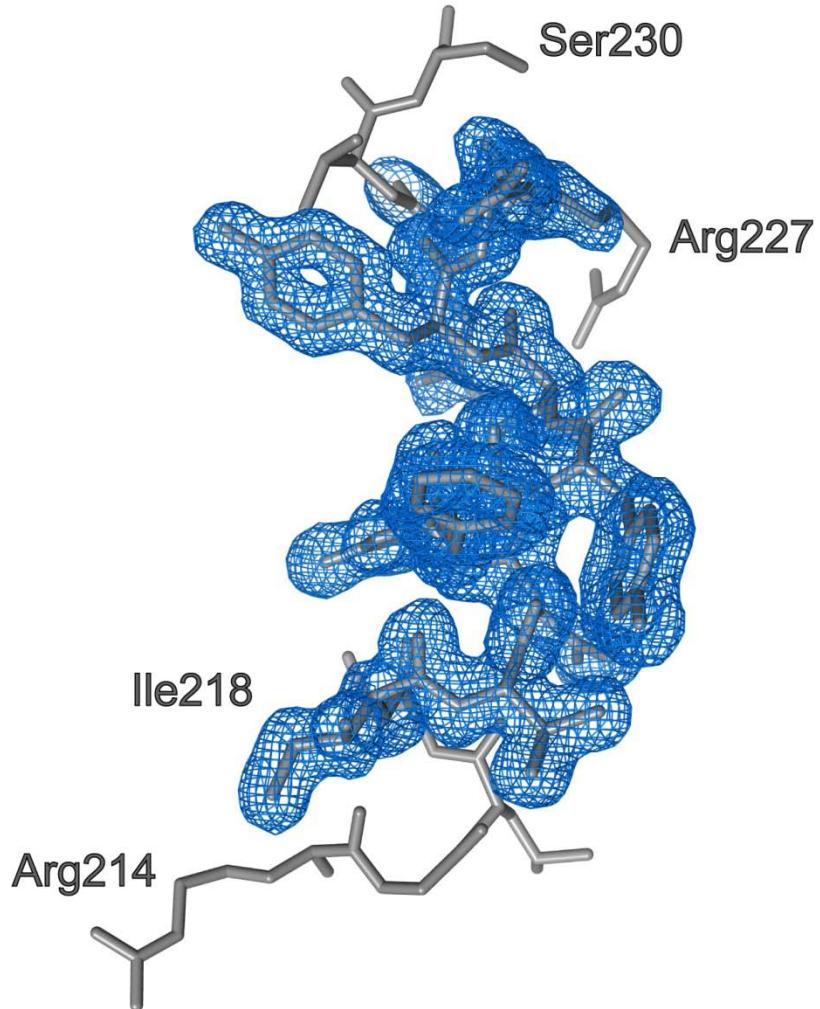
Figure S1 Dimeric arrangement of CotB2. In one monomer the protein is coloured in a gradient from green (N-terminus) to blue (C-terminus). α-helices are drawn as cylinders. The second monomer is drawn in gray. The location of the aspartate-rich motif is indicated in red and the NSE/DTE-motif is marked in orange. The double conformation at the C-terminal end of helix F is indicated in purple.
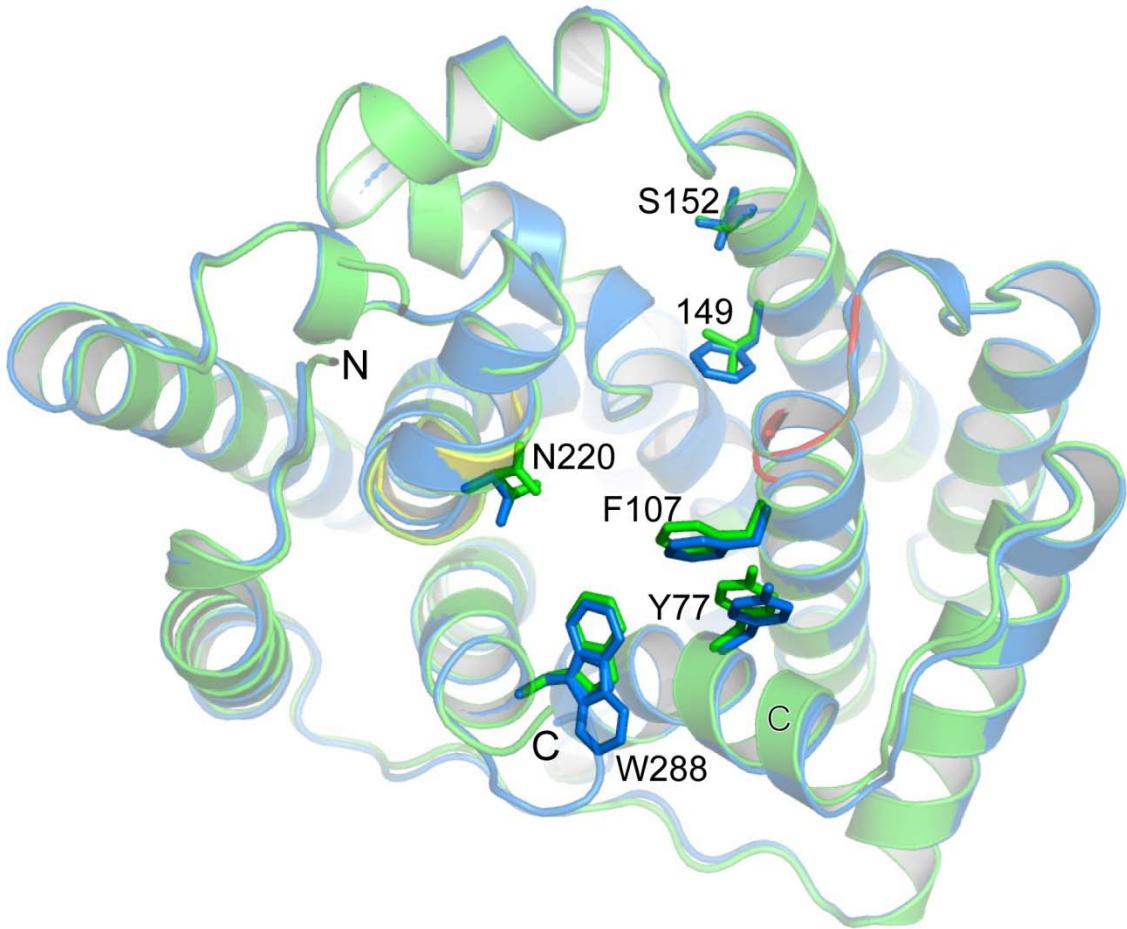
Figure S2 Dynamic light scattering experiments with CotB2. (*a*) The autocorrelation function of experimental data (red circles). The time of the experiments is logarithmically plotted against the amplitude. (*b*) The logarithmic size distribution is plotted versus the intensity. The experimental molecular weight was computed for CotB2 to 86 kDa (theoretical Mw 35.5 kDa per monomer).

CotB2 MTTGLSTAGAQDIGRSSVRPYLEECTRRFQEMFDRHVVTRPTKVELTDAELREVIDDCNAAVAPLGKTVSDERWI 75
TXS   RLEARSYIDSYDDNYVVWQRKTLYRMPSLSNSKCLELAKLDFNIVQSLHQEELKLLTRWWKESGMADINFTRHRVA 476

CotB2 SYVGVVLWSQSPRHIKDMEAFKAVCVLNCVTFVWDDMDPALHDFGLFLPQLRKICEKYYGPEDAEVAYEAARAFV 150
TXS   EVYFSSATFEPEYSATR....IAFTKIGCLQVLFDDMADIF 34aa WFKLMEEVNNDVVKVQGRDMLAHIRKPWELYF 579

CotB2 TSDHMFRDSPIKAALCTTSPEQYFRFRVTDIGVDFWMKMSYPIYRHPEFTEHAKTSLAARMTTRGLTIVNDFYSY 225
TXS   NCYVQEREWLEAGY..IPTFEEYLKTYAISVGLGPCTLQPILLMGELVKDDV 6aa PSNMFELVSLSWRLTNDTKTY 656

CotB2 DREVSLGQITNCFRLCDVSDETAFKEFFQARLDDMIEDIECIKAFDQLTQDVFLDLIYGNFVWTTSNKRYKTAVN 300
TXS   QAEKARGQ 16aa EEDAIKHICRVVDRALKEASFEYFKP 6aa GCKSFIFNLRLCVQIFYKFIDGYGIANEEIKDYIRKV 749

CotB2 DVNSRIQ
TXS   YIDPIQV

Figure S3 Structure-based alignment of CotB2 and the α-domain of TXS. On top the amino acid sequence of CotB2, its secondary structure elements are depicted. Below the amino acid sequence of TXS as well as the secondary structure elements of TXS. The aspartate-rich motif is highlighted in red and the NSE/DTE-motif in yellow. Residues subjected to mutagenesis are indicated by triangles: F105 (orange), F149 (magenta) and W288 (green).

4

Figure S4 NSE/DTE-motif. The NSE/DTE ([218]IVNDFYSYDRE[228]) motif of CotB2 resides on α-helix H. Amino acids are depicted in stick representation. Final m*2Fo*-D*Fc* electron density map is shown for the residues [218]IVNDFYSYDRE[228] contoured at σ = 1.0.

Figure S5 Superposition of CotB2^wt and CotB2^F149L. View into the active site of CotB2. CotB2^wt is drawn in blue and CotB2^F149L in green. Active site residues undergoing conformational changes upon mutation of phenylalanine to leucine at position 149 are shown in stick representation. The aspartate-rich motif is highlighted in red and the NSE/DTE-motif in yellow.

Figure S6 Mass spectrometry of the diterpene product dolabellatriene of CotB2[W288G] (retention time 20.71 min), recorded on a Trace GC Ultra with DSQII (Thermo Scientific), *m/z* was analyzed from [50-650].

Publication II

Ronja Janke, Monika Fuchs, Mathias Christmann, Thomas Brück, Bernhard Loll
(2017), Herstellung von polyzyklischen Diterpenen, *BIOspektrum* 23:709-711

Herstellung von polyzyklischen Diterpenen mittels weißer Biotechnologie

Ronja Janke[1], Monika Fuchs[2], Mathias Christmann[3], Thomas Brück[2], Bernhard Loll[1, 4]

[1] Institut für Chemie und Biochemie, Strukturbiochemie, Freie Universität Berlin

[2] Fachgebiet für Industrielle Biokatalyse, Zentrum für Weiße Biotechnologie, Fakultät für Chemie Technische Universität München

[3] Institut für Chemie und Biochemie, Organische Chemie, Freie Universität Berlin

[4] moloX GmbH

Abstract:

Diterpenes constitute a versatile class of natural biomolecules predominantly derived from plants, fungi and prokaryotes. Properties of these natural products include anti-tumor, anti-inflammatory, antibiotic, and insecticidal activities, which makes these compounds high value commercial targets for the chemical and pharmaceutical industry. Since chemical synthesis of terpenes can be difficult, production can be alternatively performed in engineered microorganisms on natural occurring routes.

Terpene bilden die größte Gruppe von Naturstoffen mit derzeit über 60.000 bekannten Verbindungen [3]. Aufgrund ihrer vielfältigen biologischen Eigenschaften finden sie Einsatz in nahezu allen Bereichen der Industrie, beispielsweise als Duftstoffe, Arzneimittel, Insektizide oder auch als Biokraftstoff. So reicht die strukturelle Diversität dieser Stoffklasse von einfachen azyklischen Duftstoffen wie Citronellol über Monozyklen, wie zum Beispiel Menthol (**Abb. 1**), zu komplizierten polyzyklischen Verbindungen wie dem Anti-Tumor-Wirkstoff Taxol. Grundbausteine aller Terpene sind C5-basierte Isopreneinheiten, deren Anzahl im Terpengrundgerüst die Gruppenzugehörigkeit und strukturelle Komplexität bestimmen (**Abb. 1**). Der erste Schritt in der Terpenbiosynthese ist die Produktion linearer, aliphatischer Polyprenyldiphosphate. In der Natur können diese durch Transferasen zu strukturell einfachen Carotinoiden verlängert werden. Die strukturelle Vielfalt der resultierenden C10-C30 Polyprenyle entsteht durch Zyklisierungen, welche durch die Enzymfamilie der Terpensynthasen katalysiert wird. So können ausgehend von nur wenigen Vorläufermolekülen eine immense Anzahl von zyklischen Grundgerüsten generiert werden, die dann von weiteren Enzymen, wie z.B. P450 Monooxygenasen, zusätzlich funktionalisiert werden.

Viele der von Pflanzen, Bakterien und Pilzen produzierten Terpene besitzen Bioaktivität. Insbesondere in der Gruppe der C20 basierten Diterpene, deren Grundgerüst aus vier C5-Isopreneinheiten besteht, existieren viele Verbindungen mit antiviralen, entzündungshemmenden, antitumorösen oder antimikrobiellen Eigenschaften. Diese Eigenschaften machen die Klasse der Diterpene besonders interessant für die pharmazeutische Industrie. Allerdings sind die hochkomplexen Polyzyklen mit mehreren Stereozentren nur aufwändig durch klassische organische Synthese zugänglich. Eine Alternative ist die Herstellung von Terpenen in genetisch modifizierten Mikroorgansimen. Dabei werden alle zur Synthese benötigten Gene in Mikroorganismen wie *Escherichia coli* oder Hefen exprimiert. So können z.B. mit Zucker als Ausgangsstoff komplexe Terpene hergestellt werden

(**Abb. 2**). Durch die Verwendung von modularen plasmid-basierten Produktionssystemen können einzelne Gene der „upstream" und „downstream" Syntheselinie ausgetauscht werden um das Produkt oder die Produktivität zu verändern [8]. Um danach stabile Produktionsorganismen zu erhalten kann man die Produktionskaskade in das Genom des Wirts integrieren. Diese biologischen Pipelines stellen eine attraktive Alternative im Vergleich zur organischen Totalsynthese dar, da strukturell hochkomplexe Makrozyklen innerhalb einer Zelle unter nachhaltigen Produktionsbedingungen generiert werden können. Die hier produzierten Makrozyklen können in Folge als Ausgangsstoffe von Chemikern weiter modifiziert werden.

Um die Ausbeute und Bandbreite der Produkte bei diesem biotechnologischen Ansatz zu optimieren, ist es unerlässlich über genaue strukturelle Informationen der Terpensynthasen zu verfügen. Denn über Mutationen im aktiven Zentrum dieser Enzyme lassen sich die Zyklisierungsreaktionen gezielt steuern, was die Produktvielfalt erweitert. Neue Produkte bedeuten aber auch immer potenziell veränderte möglicherweise neue biologische Aktivitäten.

Diterpensynthasen

Diterpene, zu denen wichtige Wirkstoffe wie Taxol und Artimisinin gehören, sind von besonderem biotechnologischen Interesse [2]. Diterpensynthasen nutzen das Vorläufermolekül Geranylgeranyldiphosphat (GGDP) als Ausgangsstoff für ihre Zyklisierungsreaktion (**Abb. 2**). Anhand ihres Reaktionsmechanismus werden die Synthasen in zwei verschiedene Klassen unterteilt [3]. Wir beziehen uns in diesem Aufsatz ausschließlich auf die häufiger vorkommende Klasse I. Bei dieser Klasse wird die Pyrophosphatgruppe des GGDP über drei Mg^{2+} Kationen koordiniert, wobei der hydrophobe Schwanz/Endgruppe in eine hydrophobe Tasche im Inneren des Enzyms zeigt. Durch Abspaltung der Pyrophosphatgruppe wird ein hochreaktives Carbokation generiert und die Zyklisierungsreaktion initiiert. Die erste aufgeklärte Proteinstruktur einer pflanzlichen Diterpensynthase, ist die der Taxadiensynthase aus *Taxus brevifolia* [9], welche ein Intermediat in der Biosynthese von Taxol produziert. Hier sei angemerkt, dass die chemische Synthese von Taxol extrem komplex ist und über 51 Syntheseschritte beinhaltet mit einer Ausbeute von 0.4% [10]. Die erste Struktur einer bakteriellen Diterpensynthase ist die von CotB2 aus dem Bakterium *Streptomyces melanosporofaciens* (**Abb. 4**) [7].

Kombination von chemischen und biotechnologischen Methoden

Der große klinische Erfolg von Taxol und Taxol-Derivaten in Kombination mit der geringen Verfügbarkeit aus pflanzlichen Quellen hat dazu geführt, dass ein reges Interesse an einem effizienten Zugang zu dieser Substanzklasse besteht. Pfeiffer und Stephanopoulos [1] ist es gelungen, die Fermentation um den Faktor 15.000 zu optimieren, dass Taxadien in einer Ausbeute von 1 g/L erhalten werden konnte. Auch die erste Oxidation konnte eine P450-vermittelte Oxidationsreaktion erreicht werden. Einen völlig anderen Ansatz verfolgten Baran und Mitarbeiter[11]. Zunächst wurde ein oxidiertes Taxadienon in 8 Stufen aus Tetramethylethen synthetisiert. Dieses wurde dann durch geschickte chemische C-H-Oxidationen[4] in 18 Stufen in das Taxabaccatin III überführt (**Abb. 3**). Auch wenn dieses Molekül in seiner Komplexität vom Taxol noch etwas entfernt ist, stellt dieses

2

Verfahren einen großen Fortschritt dar. Ein zukünftiges Forschungsgebiet wird sein, die Vorteile aus den biotechnologischen Cyclase-Prozessen mit der chemischen Modifikation unfunktionalisierter Terpene zu verbinden.

Strukturelle Charakterisierung der Diterpensynthase CotB2
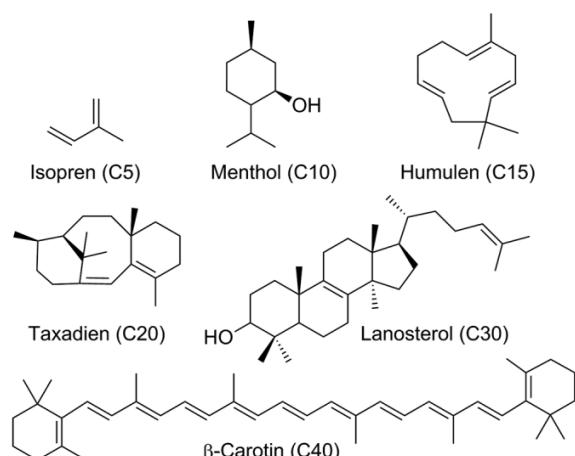
CotB2 generiert Cyclooctat-9-en-7-ol, welches vom Bakterium weiter zu Cyclooctatin modifiziert wird (**Abb. 2**). Cyclooctatin ist ein entzündungshemmender Wirkstoff, der im Gegensatz zu Ibuprofen und Aspirin eine andere Enzymklasse inhibiert. Wir konnten erfolgreich das Gen von CotB2 in *E. coli* einführen (**Abb. 2**) [5, 7]. Die Ausbeute an hochreinem Cyclooctat-9-en-7-ol aus 400 ml *E. coli* Kultur beträgt 25 mg. Mittels Röntgenstrukturanalyse klärten wir die Kristallstruktur von CotB2 auf [7]. Strukturell nimmt CotB2 die klassische Klasse I Faltung an, bestehend aus einem Bündel von α-Helices (**Abb. 4**). Die beiden Magnesium-bindenden Motive, das Aspartat-reiche Motiv und das NSE/DTE Motiv, befinden sich im aktiven Zentrum und sorgen für die stabile Koordination der Pyrophosphatgruppe. Basierend auf der Struktur wurden durch gezielte Mutagenese einzelne Aminosäuren im aktiven Zentrum von CotB2 ausgetauscht. In unserer Produktionspipeline wurde dann lediglich das native Enzym durch die neu erzeugten Varianten ersetzt. Anstelle des trizyklischen Wildtyp-Produktes konnten nun auch mono- und dizyklische Produkte erzeugt werden. Mit anderen Worten wurde durch den Austausch einzelner Aminosäuren die Reihenfolge und der Ort der Bindungsknüpfung soweit verändert, dass es zu anderen Ringschlussreaktionen kommt. Dieses Prinzip der Produktmodifizierung wurde exemplarisch an CotB2 studiert, lässt sich prinzipiell aber auch auf andere Terpensynthasen übertragen.

Aufgrund ihrer geringen Aminosäuresequenzidentität ist es schwer Terpensynthasen in Genomen zu identifizieren. Wir können davon ausgehen, dass noch viele hunderte Terpensynthasen auf ihre Entdeckung warten. Bei den P450 Monooxygenasen, die die anschließende Terpenfunktionalisierung katalysieren, verhält es sich ganz ähnlich. Auch hier ist ein Aufspüren durch die niedrige Sequenzidentität und insbesondere die der zughörigen Redoxsysteme sehr schwer [6]. Es kann erwartet werden, das Organische Synthese und Biokatalyse in Zukunft erfolgreich bei der Erschließung des Chemischen Raumes und bei der Synthese neuer bioaktiver Strukturen zusammenarbeiten.
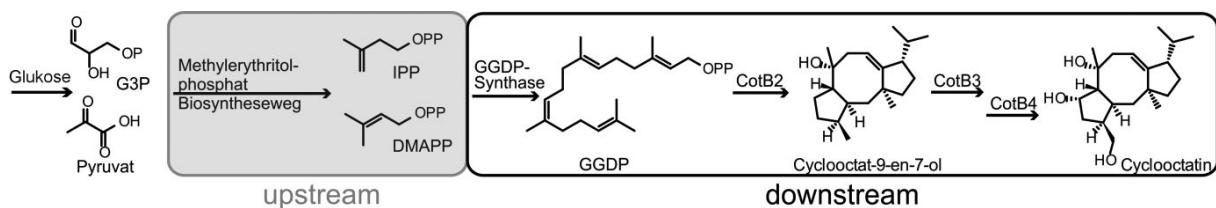
Literatur

1. Boghigian BA, Salas D, Ajikumar PK, Stephanopoulos G, Pfeifer BA (2012) Analysis of heterologous taxadiene production in K- and B-derived Escherichia coli. Appl Microbiol Biotechnol 93:1651-1661
2. Brück T, Kourist R, Loll B (2014) Production of macrocyclic sesqui- and diterpenes in heterologous microbial hosts: A systems approach to harness nature's molecular diversity. ChemCatChem DOI: 101002/cctc201300733
3. Christianson DW (2008) Unearthing the roots of the terpenome. Curr Opin Chem Biol 12:141-150
4. Christmann M (2008) Selective oxidation of aliphatic C-H bonds in the synthesis of complex molecules. Angew Chem Int Ed Engl 47:2740-2742
5. Görner C, Häuslein I, Schrepfer P, Eisenreich W, Brück T (2013) Targeted engineering of cyclooctat-9-en-7-ol synthase: A stereospecific access to two new, non-natural fusicoccane-type diterpenes. ChemCatChem:10.1002/cctc.201300285
6. Görner C, Schrepfer P, Redai V, Wallrapp F, Loll B, Eisenreich W, Haslbeck M, Bruck T (2016) Identification, characterization and molecular adaptation of class I redox systems for the production of hydroxylated diterpenoids. Microb Cell Fact 15:86
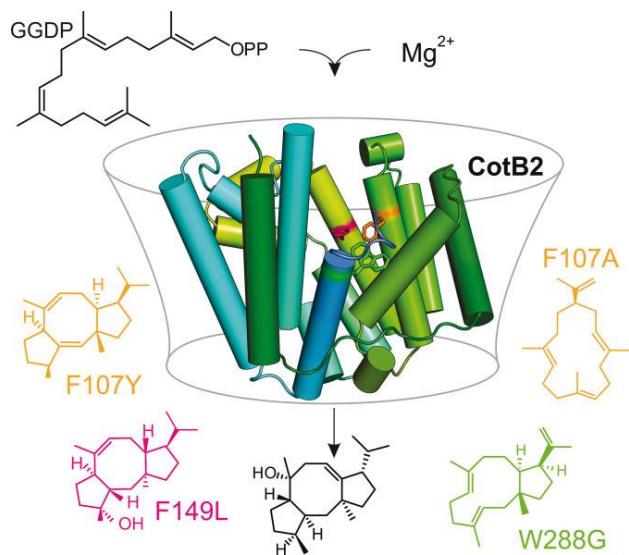
7. Janke R, Görner C, Hirte M, Brück T, Loll B (2014) The first structure of a bacterial diterpene cyclase: CotB2. Acta Crystallogr D Biol Crystallogr 70:1528-1537
8. Kemper K, Hirte M, Reinbold M, Fuchs M, Bruck T (2017) Opportunities and challenges for the sustainable production of structurally complex diterpenoids in recombinant microbial systems. Beilstein J Org Chem 13:845-854
9. Köksal M, Jin Y, Coates RM, Croteau R, Christianson DW (2011) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. Nature 469:116-120
10. Nicolaou KC, Yang Z, Liu JJ, Ueno H, Nantermet PG, Guy RK, Claiborne CF, Renaud J, Couladouros EA, Paulvannan K, et al. (1994) Total synthesis of taxol. Nature 367:630-634
11. Yuan C, Jin Y, Wilde NC, Baran PS (2016) Short, Enantioselective Total Synthesis of Highly Oxidized Taxanes. Angew Chem Int Ed Engl 55:8280-8284
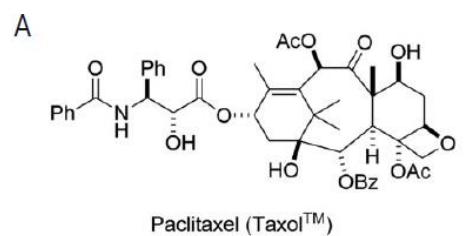
Abb. 1: Übersicht der verschiedenen Terpen-Klassen. Abhängig von der Anzahl der Skelett-Kohlenstoffatome werden Terpene in Hemiterpene (C5), Monoterpene (C10), Sesquiterpene (C15), Diterpene (C20), Triterpene (C30) und Carotinoide (C40) eingeteilt. Die Zyklisierung der linearen Vorläufermoleküle wird durch Terpensynthasen katalysiert.
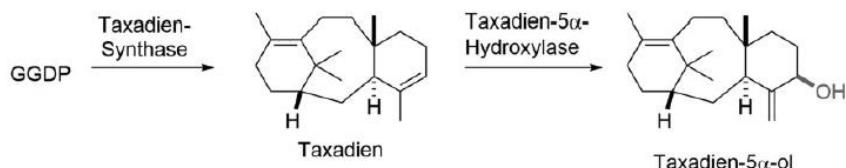
Abb. 2: Schematische Darstellung der Cyclooctatin Produktion in Bakterien. Ausgehend von Glukose werden im „upstream"-Bereich die Bausteine für das Diterpen-Vorläufermolekül GGDP geschaffen, welches dann im „downstream"-Bereich durch CotB2 zyklisiert und mit Hilfe der P450 Monooxygenasen CotB3 und CotB4 weiter modifiziert wird.
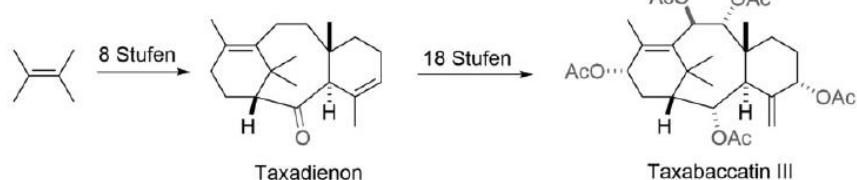
Abb. 3: Ansätze zur Herstellung von Paclitaxel (Taxol®): a) chemische Struktur; b) biotechnologische Herstellung und Oxidation von Taxadien; c) chemische Synthese von Taxadienon und Oxidation an vier weiteren Positionen im Molekül.

Abb. 4: Kristallstruktur von CotB2. Gezeigt ist ein Monomer, dessen Farbverlauf vom grünen N-Terminus bis zum blauen C-Terminus reicht. Die α-Helices sind als Zylinder präsentiert. Über Punktmutationen im aktiven Zentrum ist es gelungen, neue Produkte zu erzeugen. Die Aminosäuren, die durch Punktmutationen verändert worden sind, sowie die durch die Mutation resultierenden Produkte, sind farbig hervorgehoben: F149 in violett, F107 in orange sowie W288 in grün.

Publication III

Jan Lorenzen, Ronja Driller, Ayk Waldow, Farah Qoura, Bernhard Loll, Thomas Brück (2018), Rhodococcus erythropolis Oleate Hydratase: a New Member in the Oleate Hydratase Family Tree - Biochemical and Structural Studies, *ChemCatChem* 10:407-414

Rhodococcus erythropolis Oleate Hydratase: a New Member in the Oleate Hydratase Family Tree - Biochemical and Structural Studies

Jan Lorenzen[a,d], Ronja Driller[b,d], Ayk Waldow[b], Farah Qoura[a], Bernhard Loll[b,c,e] and Thomas Brück[a,e]

a Professorship for Industrial Biocatalysis, Technical University Munich, Lichtenberg Str. 4, 85748   Garching, Germany

b Institute of Chemistry and Biochemistry, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany

c moloX GmbH, Takustr. 6, 14195 Berlin, Germany

d These authors contributed equally to this work

e To whom correspondence should be addressed:

Prof. Dr. Thomas Brück
Professorship for Industrial Biocatalysis
Department of Chemistry
Technical University Munich
Lichtenberg Str. 4, 85748 Garching, Germany
Tel.: +49-(0) 89 289-13253
E-mail: brueck@tum.de

Dr. Bernhard Loll
Institute of Chemistry and Biochemistry
Structural Biochemistry
Freie Universität Berlin
Takustr. 6,
14195 Berlin, Germany
Phone: +49 (0) 30 838-57348
E-mail: loll@chemie.fu-berlin.de
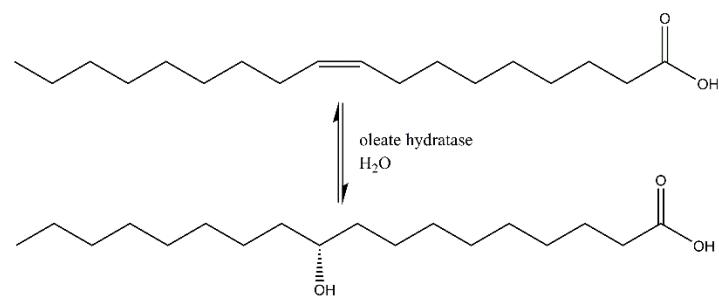
Abstract

Recently, the enzyme family of oleate hydratases (OHs: EC 4.2.1.53) have gained a rising scientific and economic interest, as these FAD-binding bacterial enzymes do not require cofactor recycling and feature high thermal and pH stability. Their products, hydroxy fatty acids, are used in specialty chemical applications including surfactant and lubricant formulations. The "oleate hydratase engineering database", established by Schmid et al. (2017), divides all OHs into 11 families (HFam1 to 11). To date, only two crystal structures of homodimeric OHs from the families HFam2 and HFam11 have been reported. In this study, we biophysically characterized an OH belonging to the HFam3 family, originating from the marine bacterium *Rhodococcus erythropolis*, for the first time. The crystal structure revealed that this new OH (OhyRe) surprisingly is a monomer in its active form. This particular feature provides new avenues for enzyme engineering and recycling via immobilization.

Introduction

Unsaturated fatty acids are toxic to many bacteria due to their surfactant properties, destabilization of membranes and due to the unsaturated fatty acids inhibitory effect on the fatty acid biosynthesis.[1] Hence, mechanisms have evolved to counteract these effects when organisms are exposed to unsaturated fatty acids, such as linoleic acid or oleic acid. In consequence, unsaturated fatty acids are saturated by hydrogenation. However, an intermediate step in this hydration process is catalyzed by fatty acid hydratases.[2] Hydration of unsaturated fatty acids has so far only been identified in Gram-negative and Gram-positive bacteria [3] and is mediated by hydratases.

Hydratases are a group of lyases that catalyze hydration and dehydration of a substrate[4]. Even though many hydratases are known, only few OHs have been described.[2] OHs belong to the group of fatty acid hydratases (EC 4.2.1.53) and convert oleic acid into (R)-10-hydroxystearic acid (10-HSA).[5] To date all characterized OHs bind an FAD cofactor,[2] albeit the role of FAD

Figure 1: Oleate hydratase (OH) reaction scheme.

has remained enigmatic.

Merely two structures of OHs have been reported. One of the OH from *Lactobacillus acidophilus* (LAH), crystalized in the apo state as well as bound to the preferred substrate linoleic acid.[6] The second structure of an OH (OhyA) has been determined from *Elizabethkingia meningoseptica*, bound to its cofactor FAD.[7] Even though LAH and OhyA convert different

substrates, both utilize FAD as a cofactor. Interestingly the bound FAD was oxidized in both hydratases, but reduced FAD led to significantly higher activity in OhyA.[7]

Most recently, Schmid et al. assembled a "hydratase engineering database" including a total of 2046 protein sequences.[8] Sequence comparison revealed 11 distinct families (HFam1 to 11) with a sequence identity of 62 % within the respective family. The only two structurally characterized OHs, LAH and OhyA fall into two different families, HFam2 and HFam11. For all other families structural information has remained elusive. Remarkably, in regions attributed to decorate the active site, amino acid sequences are not strictly conserved within these families.

Hydratases are in particular interesting for commercial applications since their products, hydroxyl fatty acids, have a wide range of industrial and pharmaceutical applications. Hydroxyl fatty acids can be utilized in applications for the food industry as precursors of lactones,[9] as well as an emollient for cosmetics. Further applications are the employment as a surfactant or as additives in lubricant formulation and as a green material in polymer science.[10]

In this study, we aimed at obtaining novel insights into the reaction mechanism of the OhyRe enzyme of *R. erythropolis*, using a combined structural, biophysical and biochemical approach. The OhyRe enzyme was biochemically characterized in terms of pH and temperature optimum as well as substrate specificity. The kinetic constants were determined for the conversion reaction of oleic acid to 10-HSA. Furthermore, we were able to solve the crystal structure of OhyRe at 2.64 Å resolution. The structure of OhyRe provides structural information for the OH HFam3 family of OHs for the first time. We have compared the overall architecture as well as structural features of the OhyRe active site to a set of reported OHs. As we identified two key residues, which differed in the OhyRe structure, these were replaced by site directed mutagenesis to residues conserved in the other family members. Most interestingly, these conservative amino acid replacements resulted in a strong decrease of hydration activity, indicating that OhyRe may follow a distinct reaction mechanism.

Results & Discussion
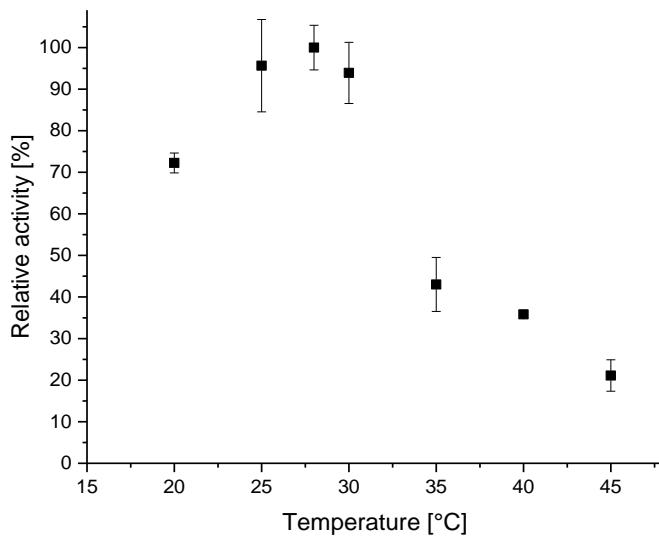
Isolation & characterization of OhyRe derived from *R. erythropolis*

R. erythropolis is a Gram-positive, non-motile representative of the nocardiaceae family.[11] *R. erythropolis* is a wide spread soil bacterium which is mainly found in marine habitats, from coastal to deep sea sediments [12] and exhibits a broad range of biochemically interesting enzymes.[13] As the genome of *R. erythropolis* has recently been deciphered,[14] we have performed sequence specific genome mining, targeting alternative OH activities in this extremophile organism.

To this end we have identified the 1.68 kB sequence related to known OHs albeit at a very low sequence identity level of max. 38 %. The gene for the putative OH was termed *ohyRe*, according to the nomenclature used by Bevers et al.[5a] for the *Elizabethkingia meningoseptica* OH ohyA. The corresponding OhyRe protein is composed of 559 amino acids with a calculated molecular weight of 66.3 kDa. The N-terminally His$_6$-tagged protein was successfully expressed in *Escherichia coli BL21(DE3)* cells. Initial activity tests with the crude extract resulted in high
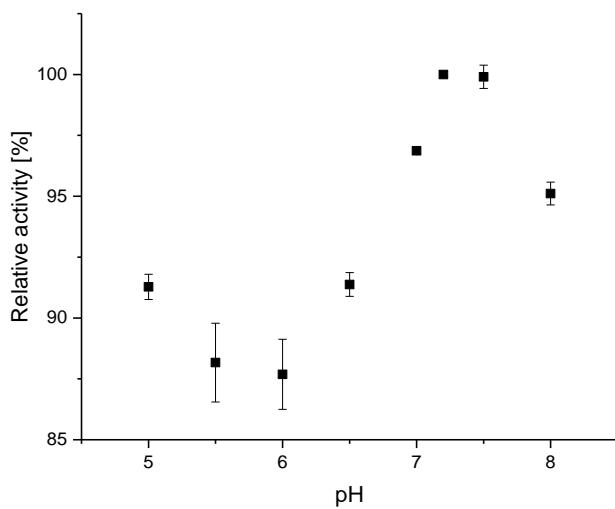
conversion rates (approx. 85%) from oleic acid to 10-HSA, confirming that OhyRe indeed displayed OH activity. To further characterize the enzyme, we have devised a Ni^{2+}-affinity purification protocol. After final His-Trap HP affinity chromatography, the purified enzyme was colorless and lost its hydration activity (Fig. S1). This observation suggested a loss of the FAD cofactor, which has been uniformly reported to be essential for the hydration activity in all known OHs. Moreover, the only reported loss of FAD during purification was also associated with a strong decrease in hydration activity.[6] Therefore, the addition of various FAD concentrations (5 µM to 100 µM) to the purified enzyme preparation has been evaluated to reconstitute the activity of purified OhyRe. The addition of FAD to a final concentration of 20 µM lead to the recovery of the enzymatic activity of OhyRe and was applied for all following experiments. After establishing an efficient purification protocol, OhyRe was biochemically characterized.

Initial tests demonstrated that purified OhyRe has a temperature optimum for the hydration activity at 28 °C (Fig 2). Interestingly, OhyRe exhibits a relatively wide temperature tolerance with high activity in the low temperature region (72 % of max. at 20 °C) and still moderate activity in the high temperature area (21 % of max. at 45 °C). This wide temperature tolerance might reflect the diversity of the occurrence of *R. erythropolis* from costal to deep sea habitats.

Figure 2: Determination of the temperature optimum for the oleate hydratase (OH) OhyRe from *R. erythropolis* in 20 mM Tris-HCl buffer at pH 7.2. Relative activity is given to the temperature optimum at 28 °C (100 %). All experiments were carried out in biological duplicates and technical triplicates.

The determination of the pH optimum (Fig. 3) showed a high pH tolerance in a range from pH 5.0 to 8.0 for OhyRe, the highest activity was reached at pH of 7.2.

Figure 3: Determination of the pH optimum for the oleate hydratase (OH) OhyRe from *R. erythropolis* at 28°C. Relative activity is given to the optimum at pH 7.2 (100 %). All experiments were carried out in biological duplicates and technical triplicates.

These results correlate with the recommended cultivation conditions for *R. erythropolis* published by the German Collection of Microorganisms and Cell Culture (Leibniz Institute DMSZ).

The substrate spectrum of OhyRe was experimentally determined in reactions with the purified enzyme and different free fatty acids (FFAs) (Table 1). All experiments were carried out in biological duplicates and technical triplicates. The enzyme showed neither detectable activity for any saturated fatty acid tested (C14:0 to C18:0), nor for myristoleic acid, vaccenic acid, eicosatrienoic acid and eicosatetraenoic acid, respectively. Low conversion rates could be observed for linoleic acid, α- and γ-linolenic acid. The highest affinity, under the given conditions, was detected for oleic acid, followed by palmitoleic acid. Complex lipids, like the tested triolein or microalgae derived oils were not converted by the enzyme. These results correspond to published data for other known OHs [6] and confirm the preference of OHs for the 9-Z double bond in the fatty acid carbon chain. The kinetic parameters were experimentally defined for OhyRe, with oleic acid as the preferred substrate, and resulted in a K_M value of 0.49 ± 0.1 mM and a k_{cat} value of 34 ± 5 min^{-1}.
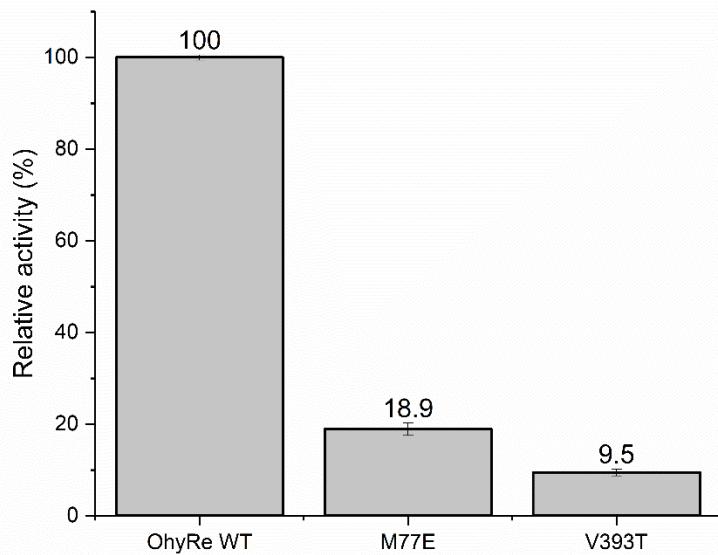
Table 1: Tested substrates, identified products and corresponding specific activity of conversion reaction with OhyRe. All experiments were carried out in biological duplicates and technical triplicates.

substrate		product	specific activity [nmol mg-enzyme^{-1} min^{-1}]
C16:1 ω7	palmitoleic Acid	10-hydroxypalmitic acid	1205 ± 70.6
C18:1 ω9	oleic Acid	10-hydroxystearic acid	1266 ± 30.3
C18:2 ω6	linoleic Acid	10-hydroxy-12(Z)-octadecenoic acid	156 ± 3.3

5

| C18:3 ω3 | α-linolenic Acid | 10-hydroxy-12,15(Z,Z)-octadecadienoic acid | 150 ± 1.7 |
| C18:3 ω6 | γ-linolenic Acid | 10-hydroxy-6,12(Z,Z)-octadecadienoic acid | 615 ± 36.0 |

OhyRe mutants

Engleder et al. [7] demonstrated by multiple sequence alignment of 10 OHs that the catalytically crucial residues are highly conserved within the OH family. The alignment of *OhyRe* with 14 different OHs (Fig. S2) predominantly confirmed this hypothesis, but also depicted two exceptions. Interestingly, two amino acid residues that are highly conserved in all other OHs differ in the sequence of *OhyRe*, namely the residues M77 and V393. Further structural comparison of OhyRe and OhyA (Fig. S3) showed that these residues correlate with the residues E122 and T436 from OhyA, which are responsible for the activation of the water molecule or the binding of the carboxylate, [7] respectively. A conservative substitution of the two identified residues to the familial conserved residues (M77E, V393T) was executed by site directed mutagenesis and resulted in the effect depicted in Fig. 4.

Figure 4: Relative conversion activity from oleic acid to 10-HSA from OhyRe wild type enzyme (OhyRe WT), set as 100 %, and its respective mutants (M77E, V393T). All experiments were carried out in biological duplicates and technical triplicates.
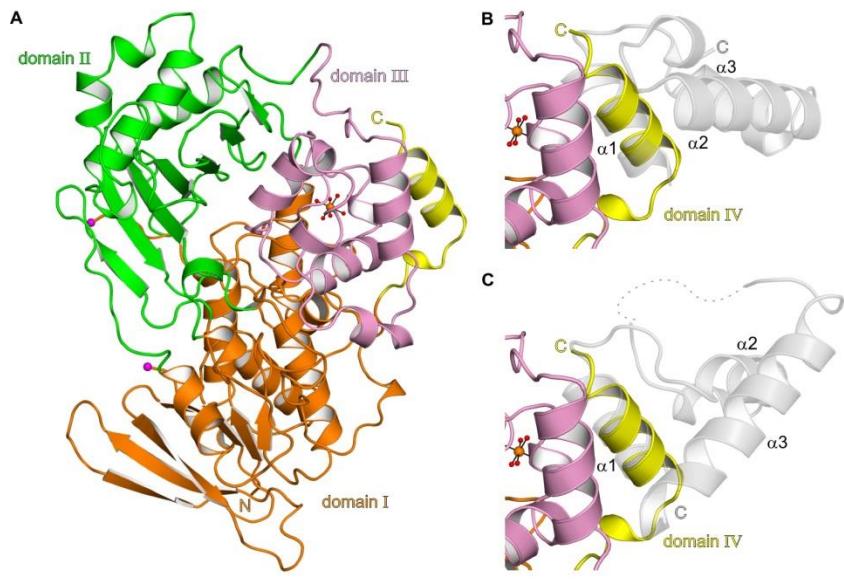
Fig. 4 shows the relative hydration activity of the OhyRe wild type enzyme compared to the two constructed mutants M77E and V393T with oleic acid as substrate. The conservative substitution of both residues resulted in a strong decrease in the conversion efficiency from oleic acid to 10-HSA of 81 % and 90 %, respectively. This strong decrease in hydration activity

supports the hypothesis that OHs from the HFam3 family might follow a distinct catalytic mechanism.

In comparison to the two other structurally characterized OHs LAH [6] and OhyA [7], OhyRe shows a sequence identity of 35 % and 34 %, respectively. Beside the low sequence identity, both the LAH and OhyA have large N- and C-terminal amino acid sequence extensions (Fig. S4 and 6A,B), which are absent in OhyRe. To understand whether those extensions are common for all other OHs, we performed a sequence alignment (Fig. S2). Interestingly, OH sequences can be roughly divided in sequences with amino acid extensions at the N- and the C-termini present or absent. Consequently, this raised the question, whether OhyRe might be structurally different compared to the members of the HFam2 and HFam11 families. Hence we tried to crystalize the protein for structure determination by X-ray crystallography.

Overall structure

We obtained crystals of OhyRe that belong to space group $P6_522$. The diffraction of the OhyRe crystals was anisotropic, resulting in complete data set to a resolution of 2.64 Å. The structure of OhyRe was determined by molecular replacement with two protein chains in the asymmetric unit corresponding to a very high solvent content of about 75%. The structure was refined with good stereochemistry to good R factors (Table S1, supporting information). Both polypeptide chains are practically identical with a root mean square deviation (rmsd) of 0.1 Å for 522 pairs of Cα-atoms. The structure is complete, except for the 5 N-terminal residues and a loop region ranging from residue 42 to 67 due to missing electron density caused by its high flexibility (Fig. 5A). No redox cofactor could be identified in the electron density maps. Hence our structure represents the apo-state of OhyRe from *R. erythropolis*.
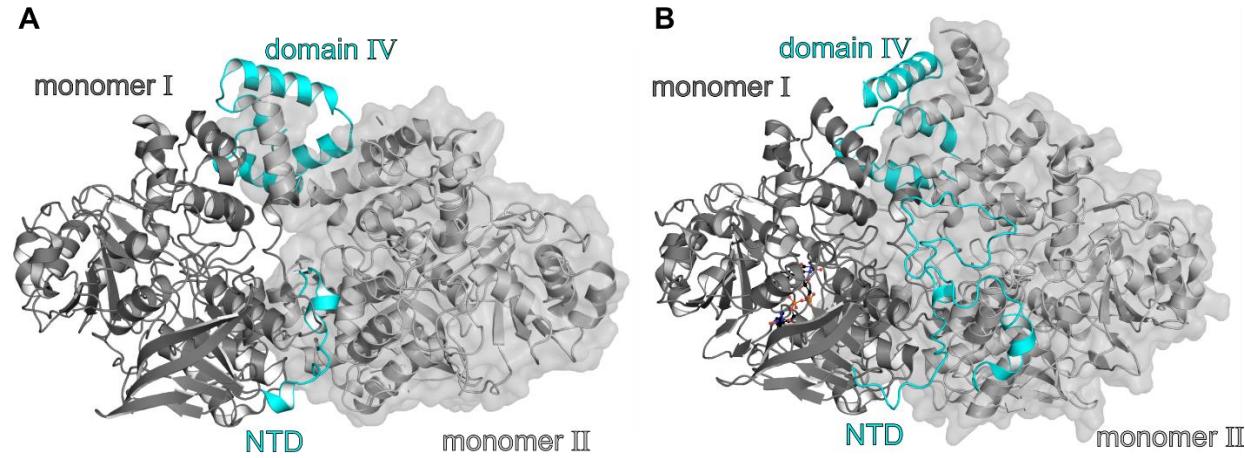
Figure 5: Overall structure of OhyRe in cartoon representation. **(A)** Domain organization of OhyRe: Domain I in orange, domain II in green, domain III in pink and domain IV in yellow. The bound Mg^{2+} and coordinating waters are shown as orange and red spheres, respectively. Magenta spheres indicate the boarders of a flexible linker, ranging from residue 61 to 86, which could not be modelled due to its flexibility. **(B)** Zoom onto domain IV with a superposition of OhyRe and LAH, domain IV of LAH is shown in gray. **(C)** Superposition of OhyRe and OhyA, domain IV of OhyA is shown in gray. Dashed line indicates an un-modelled loop region.

The closest related structures, as analyzed by the DALI software,[15] are the OHs LAH (PDB-id 4ia6) and OhyA (PDB-id 4uir). OhyRe superimposes with a rmsd of 1.8 Å for 500 pairs of Cα-atoms and 1.6 Å for 501 pairs of Cα-atoms, respectively. In analogy to the structures of LAH and OhyA, OhyRe is composed of three core domains (residues 1–540; Fig. 5A and 6) that are related to other FAD-dependent enzymes. Domain I (residues 6–115, 223–300, 322–344, and 478–540) is a mixed α/β domain composed of a parallel five-stranded β-sheet packed between two α-helices on one side and a three-stranded antiparallel β-sheet on the other side (Fig. 5A), resembling a variant of the Rossmann fold. Domain II (residues 116–131, 301–321, and 345–477) consists of an antiparallel β-sheet (Fig. 5A) flanked by three α-helices defining the cofactor- and substrate-binding site in conjunction with domain I. Domain III (residues 132-222) is exclusively α-helical (Fig. 5A) and its fold is structurally related to monoamine oxidases.[16] Domain IV (residues 541-559) of OhyRe is composed of merely one single α-helix. In contrast to the structures of LAH and OhyA the C-terminal domain is truncated by 32 or 48 residues, respectively. As a consequence two α-helices are absent in domain IV of OhyRe (Fig. 5B and C). The structure of LAH has been determined in apo and product-bound state. Upon substrate recognition and binding the two C-terminal α-helices of domain IV undergo a large conformational change and thereby allow the substrate to enter a cavity that ends at a cleft between domain I and III.[6]

For both the LAH and OhyA a homodimeric quaternary structure in solution as well as *in crystallo* has been reported.[6-7] For other OHs such as the one from *L. fusiformis*,[17] *S. maltophilia* [18] and *M. caseolyticus* [19] dimers in solution were reported as well. As analyzed by the PISA server,[20] OhyRe is a monomer *in crystallo*. To analyze the situation in solution and to

8

confirm our structural analysis, we performed multi-angle light scattering (MALS) experiments. Our experiments clearly revealed that OhyRe is monomeric in solution (Fig. S5). The latter finding is in agreement to our observations for OhyRe assembly in the crystal and strengthens our hypothesis that OHs, lacking N- and C-terminal extensions, are monomers. Additionally we examined whether the extensions might have functional implications. As discussed above, in the structure of LAH, domain IV reflects the C-terminal extension, suggested to be involved in substrate gating. Since the two reported and our crystal structures cover nearly the complete amino acid sequence of the respective protein, we superimposed the structures and inspected in particular the monomer-monomer interfaces of dimeric LAH and OhyA. Most of the dimer interface is established by residues that reside within the N- or C-terminal extension, including domain IV of LAH (Fig. 6A) and OhyA (Fig. 6B). In summary, we propose that the presence or absence of N- and C-terminal extensions within the amino acid sequence directly influences the oligomeric state of OHs. In case of OhyRe the extensions are absent and hence the protein is monomeric in solution.

Notably, the crystal structure revealed a coordinated Mg^{2+} ion close to the putative active site of OhyRe (section S1). However, experimental data showed that the Mg^{2+} ion has no influence on the hydration activity of OhyRe.
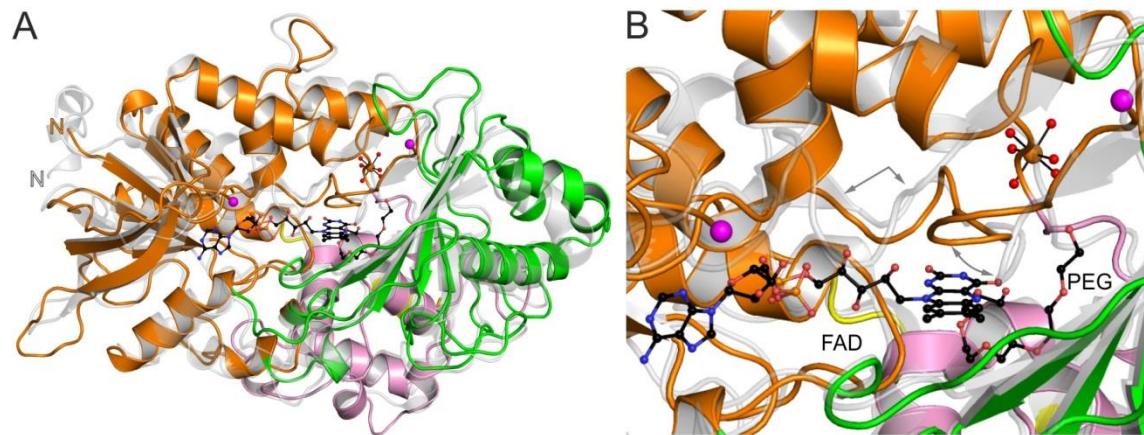
Figure 6: Oligomeric state of oleate hydratases (OHs). **(B)** LAH monomer I is shown in dark gray. Monomer II in surface and cartoon representation (light gray). The N-terminal domain (NTD) and domain IV in cyan. **(C)** The NTD OhyA of each monomer meanders into the other monomer.

FAD binding

All characterized OHs utilize FAD as cofactor.[2] Different roles of the FAD cofactor have been proposed either being directly involved in catalysis or indirectly having merely a structural role to stabilize the enzyme. In the crystal structure of LAH, no FAD could be located in its binding site.[6] In contrast, in the crystal structure of OhyA, FAD is bound, but only to one of the two polypeptide chains.[7] Furthermore a K_d of 1.8×10^{-6} M was determined by isothermal calorimetry.[7] All indicate a rather low binding affinity of FAD to studied OHs. For our structural studies, we attempted to add FAD prior to the final purification step or to incubate OhyRe with FAD prior to the crystallization experiments. Moreover, we tried to soak FAD in our apo OhyRe

crystals. All our attempts to gain structural information of OhyRe bound to FAD remained unsuccessful so far. A primary sequence analysis reveals the presence of a characteristic recognition sequence $GXGXXGX_{21}E/D$ for nucleotide binding, also reflected by the Rossmann fold of the protein. A superposition of the structures of OhyA and OhyRe (Fig. 7A and B) clearly reveals that the cofactor binding site is large enough to accommodate a FAD molecule. Furthermore in the crystal structure of OhyA a PEG molecule was identified that might mimic a bound oleate molecule. For latter substrate enough space would be in the active site of our OhyRe structure (Fig. 7A and B).
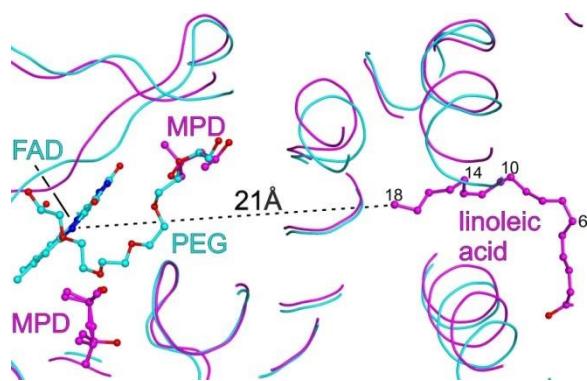
**Figure 7**: Superposition of OhyRe and OhyA. Same color coding as in Figure 5A. **(A)** OhyRe is colored as in Figure 1 and OhyA in gray. View on the active site of OhyA. Magenta spheres indicate the boarders of a flexible linker in OhyRe, ranging from residue 42 to 67, which could not be modelled due its flexibility. In contrast upon FAD binding in OhyA this loop region is getting structured, indicated by two gray arrows. Moreover a short loop fragment rearranges upon cofactor/substrate binding (gray arrow-headed arc). In general, there would be enough space in OhyRe to accommodate the cofactor as well as the substrate. Solvent exposed on the back of the active site, the Mg^{2+} binding site located. **(B)** Same view in (A) zoomed on the FAD cofactor binding site of OhyA.

Putative active site with catalytically important residues

As mentioned above OHs can be, based on their amino acid sequences, sub-divided in 11 families.[8] Based on the structure of OhyA, that belongs to the family HFam1,[8] E122 was proposed to be important for cofactor binding and being involved in catalysis by activating a water molecule for the attack on the partially charged double bond of the substrate.[7] Latter residue resides within a conserved loop region [118]RGGREM[123], almost strictly conserved within all OHs. In the sequence of the OhyRe, that falls into the family HFam3, the ultimate and penultimate amino acids are different, but characteristic for members of the largest HFam1 as well HFam3 family.[8] In the respective loop region OhyRe has the sequence [73]RGGRML[78]. The exchange of OhyA E122 to OhyRe M77, will drastically alter the chemistry and can certainly not activate a water molecule. To investigate the influence of amino acid at position 77 in OhyRe, we generated a point mutation M77E. This single mutation led to a drastic decrease in hydration activity of OhyRe. OhyA Y241 was proposed to be as well important for catalysis, by protonation of the double bond [7] and is conserved with OhyRe Y205. Hence, it is tempting to propose that OhyRe might have a different catalytic mechanism as proposed for OhyA.

Interestingly, the fatty acid double-bond hydratase from *Lactobacillus plantarum*, LPH,[21] also exhibits the HFam3 family specific motive RGGREM. In contrast to LPH, which solely converts linoleic acid, OhyRe exhibits hydration activity for both, C16 and C18 unsaturated fatty acids. In addition, LPH was shown to be a homotrimer in its active form, whereas OhyRe is a monomer.

The exact binding site of the substrates of OHs could not yet be resolved. In the crystal structure of LAH, a cavity was observed running from the surface of the protein towards the interface formed by domain I and III. This site is known to be a substrate binding site in other related enzyme families. Volkov *et al.* identified a linoleic acid (LA) molecule bound in this cavity with its carboxylate oriented towards the protein surface and the hydrophobic tail pointing to the protein interior (Fig. 8). In contrast, a PEG molecule has been identified close to the FAD cofactor in the structure of OhyA, possibly mimicking a bound substrate molecule (Fig. 8). Strinkly this binding site is fundamental different to the "interdomain"-binding site as described for LAH (Fig. 7B and Fig. 8). To date, there is no structural information on an OH with a real substrate bound to the enzyme. Therefore it is very speculative which amino acid residues are crucial for catalysis. A definitive answer could only provide a structure with bound substrate. This would allow the identification of key amino acids important for catalysis. These residues could be then targeted by structure-based mutagenesis. Moreover, the given resolution of 2.64 Å of our OhyA structure, does not allow the identification of water molecules within the active site. But, One can imagine that the precise position of water molecules might be crucial for catalysis. Based on the structure of OhyA, the T436A mutant had been generated, showing a dramatic decrease in activity. Latter observation was interpreted as the hydroxyl of T436 being important for recognition and binding of the carboxylate function of the substrate. Taken together, both structures not only display distinct substrate binding sites, but also differ in the orientation of the carboxylate. In order to increase the enzymatic activity of OhyRe, we speculated that a mutation V393T (equivalent to T436 in OhyA) could potentially improve recognition and binding of the substrate. But unexpectedly, the activity was drastically reduced (Fig 4).

Figure 8: View on the potential nucleotide and substrate binding site. Superposition of LAH (magenta) and OhyA (cyan) as ribbon. FAD and PEG bound to LAH are shown in stick representation. The linoleic acid substrate of LAH and 2-methyl-2,4-pentanediol (MPD) molecules bound to LAH are shown in stick representation. Notably a 2-methyl-2,4-pentanediol (MPD) molecule in the LAH structure superimposed partially with the bound PEG molecule in the OhyA structure. The terminal carbon atom of linoleic acid is in distance of 21 Å to FAD, indicated by a dashed black line.

Conclusion

The enzyme family of OHs offers the sustainable conversion and valorization of biogenic fatty acids in high value lubricant additives. In contrast to heme dependent P450 hydroxylases, the conversion of oleic acid to 10-hdroxysteric acid by OHs can be accomplished without the need for co-factors. In this study a new OH from *R. erythropolis* could be identified by targeted genome mining. The biochemical characterization showed that the enzyme, OhyRe, exhibits a high pH-tolerance as well as high hydration activity over a wide temperature range. The crystal structure of OhyRe was solved to a resolution of 2.64 Å, surprisingly revealing a monomeric structure for the enzyme, which contrasts all other reported members of the OH family. These results were ultimately confirmed by MALS experiments of the protein in solution. To our knowledge, OhyRe is the first strictly monomeric OH and the first member of the HFam3 family of OHs that was characterized and crystallized so far. The HFam3 family of OHs exhibits high potential for further biocatalytical applications, as the monomeric enzyme provides a variety of options for immobilization strategies. The free N- and C-terminal ends are not involved in dimer association and could e.g. be feasible targets for the fusion of anchor-proteins. Furthermore, two mutants of OhyRe, M77E and V393T, were generated to clarify the importance of these two residues in the catalytic mechanism of OhyRe. These conservative substitutions, to residues being highly conserved within in the OH family, resulted in a drastic decrease of hydration activity of 80 to 90 %. This suggests that OhyRe might follow a distinct reaction mechanism.

Materials and Methods

Chemicals

All chemicals utilized in this work were purchased from Sigma-Aldrich (Munich, Germany) and Carl Roth (Karlsruhe, Germany), at the highest purity grade available.

Bacterial strains & plasmids

The *Escherichia coli* strains XL-1 Blue and DH5α were used for cloning whereas the strains BL21 (DE3) and Rosetta2 were employed for protein expression. All plasmids and *E. coli* strains were obtained from Novagene/Merck Millipore.

Cloning for characterization experiments

The myosin-cross-reactive antigen coding gene O5Y_00450 from *Rhodococcus erythropolis CCM2595*, was taken as a template for a codon-optimized gene-synthesis (Life Technologies, Regensburg), for an *E. coli* host strain. The obtained synthetic gene was sub-cloned in a pET28a expression plasmid and transformed into chemically competent *F* cells.

Cloning for crystallization experiments

For protein production in *E. coli*, the *OhyRe* ORF was amplified with the forward primer 5'-TATACCATGGGAATGAGCAGCAATCTGAGCC-3' and reverse primer 5'-TATAAAGCTTTTAACGAAACATGGTAACTGCTGC-3' containing *Hind*III and *Nco*I recognition

sites, respectively) and cloned into the pETM11 vector, resulting in a construct with N-terminal hexa-histidine-tag that could be cleaved off by TEV protease.

OhyRe mutant construction

The mutants *OhyRe-M77E* and *OhyRe-V393T* were constructed by overlap-extension-PCR using primer sets listed in Table S2 (supporting information).

Protein expression and protein purification for characterization experiments

The expression of the native OhyRe or respective mutants thereof was carried out in *E. coli BL21DE3* cells, grown in Laure Broth (LB) medium. Pre-cultures were inoculated from a cryo-stock and grown over night in 100 ml LB in a 500 ml baffled flask at 120 rpm at 37°C. Main cultures were grown up to an optical density, measured at 600 nm (OD_{600}), of 0.6 – 0.8, before the expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. After 16 h of incubation at 16°C the cells were harvested and the resuspended (20 mM Tris-HCl buffer and 25 mM imidazole pH 7.2) cell pellets were disrupted by high pressure homogenization (EmulsiFlex-B15, AVESTIN). A subsequent centrifugation step at 20.000 g for 40 min at 4°C (Beckmann coulter J-20 XP) was applied for the separation of the cell debris from the soluble protein fraction. The soluble protein fraction was utilized for affinity chromatography via a Ni^{2+}-NTA His-trap column (HisTrap FF, GE Healthcare; flow rate 1 ml/min). The purified protein solution was desalted using HiPrep 26/10 desalting column (GE Healthcare). Protein amounts were quantified using 2-D quant kit (GE Healthcare) according to manufacturer's instructions.

Protein expression and protein purification for crystallization experiments

OhyRe fused to an N-terminal hexa-histidine-tag in pET-M11 vector was transformed in *E. coli* Rosetta2. Protein was expressed in auto-induction media at 37 °C until an OD_{600} ~ 0.8 was reached and subsequently cooled down to 20 °C.[22] Cells grew over night and were harvested by centrifugation (6 min, 6'000 rpm at 4 °C). For resuspension of the cell pellet, buffer A was used (50 mM Tris/HCl pH 7.5, 500 mM NaCl, 30 mM imidazole and 1 mM DTT). Cells were lysed by homogenization at 4 °C and the lysate was cleared by centrifugation (1 h, 21'000 rpm at 4 °C). Ni^{2+}-NTA beads (cv ~ 1 ml; GE Healthcare) were equilibrated with buffer A. OhyRe was loaded on the column and washed with 10 cv of buffer A. OhyRe was eluted in a linear gradient to buffer A supplemented to 500 mM imidazole. Pooled fractions were dialyzed over night against buffer B (20 mM Tris/HCl pH 7.5, 150 mM NaCl and 1 mM DTT) supplemented with TEV protease to cleave off the N-terminal His-tag. Uncleaved protein was separated from cleaved protein via an additional Ni^{2+}-NTA affinity chromatography. Size exclusion chromatography was performed with a HighLoad Superdex S200 16/60 column (GE Healthcare), equilibrated with buffer B. Pooled protein fractions were concentrated with Amicon-Ultra 30'000 to 19.2 mg/ml as measured by the absorbance at 280 nm.

Enzyme characterization

The determination of the enzymatic properties of the OH from *Rhodococcus erythropolis CM2595* was (unless otherwise stated) executed under the following standard reaction conditions. The tests were carried out in a reaction volume of 200 µl in 20 mM Tris-HCl buffer (pH 7.2), containing the purified enzyme (final conc. of 5 µM), FAD (final conc. 20 µM) and 720 µM OA as substrate at 28°C for 15 min. **Thermo-stability** was tested in a temperature range from 20 – 45°C. **pH tolerance** was monitored in a range from pH 5 – 8. To determine the **substrate specificity** of the myosin-cross-reactive antigen from *Rhodococcus erythropolis CM2595*, the purified enzyme was tested on myristic acid (C14:0), myristoleic acid (C14:1), palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), vaccenic acid (C18:1 *trans*-11) oleic acid (C18:1, *cis*-9), linoleic acid (C18:2), linolenic acid (C18:3) and triolein as potential substrates. For the determination of the **kinetic parameters** of OhyRe towards oleic acid, different substrate concentrations (90 µM to 1.44 mM) and reaction times (1 to 15 min) were tested under standard conditions. The reactions were stopped by the addition of an equal volume of ethyl acetate (EtOAc) and instant, intensive vortexing. The resulting solvent phase was separated from the water phase and applied for silylation and subsequent analysis.

Lipid analysis

The preparation of the extracted lipid fractions for GC measurements was performed according to Volkov et al.[23] Hydroxylated fatty acids were identified by GC-MS by injecting 1 µl into a Thermo Scientific™ TRACE™ Ultra Gas Chromatograph coupled to a Thermo DSQ™ II mass spectrometer and the Triplus™ Autosampler injector. Column: Stabilwax ® fused silica capillary (30 m × 0.25 mm, film thickness 0.25 µm) Restek. (Program: initial column temperature 150 °C, increasing (4 °C/min) up to a final temperature of 250 °C. Carrier gas: hydrogen (flow rate 3.5 ml/min). Peaks were identified by comparison to fatty acid standards or by specific molecular masses detected. Extracts resulting from kinetic experiments were analyzed with the Shimadzu™ GC-2025 system equipped with a flame ionization detector. Column: Zebron ZB-WAX (30 m x 0.32 mm, film thickness 0,25 µm) Phenomenex. Carrier gas: hydrogen (3.00 ml/min). Program: initial column temperature 150 °C for 1 min; increasing 5°C/min to 240°C, hold for 6 min. Peaks were identified by comparison to the respective standards.

Crystallization

Crystals were obtained by the sitting-drop vapor-diffusion method at 18 °C with a reservoir solution composed of 100 mM Bis-Tris/HCl pH 5.5 to pH 6.0, 255 mM to 300 mM magnesium formate and 5 % (v/v) glycerol. Crystals were cryo-protected with 25% (v/v) MPD supplemented to the reservoir resolution and subsequently flash-cooled in liquid nitrogen.

Diffraction data collection, structure determination and refinement

Synchrotron diffraction data were collected at the beamline 14.1 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 [24] (Table S1). The structure was solved via molecular replacement in PHASER [25] by using the structure of the hydratase from *Lactobacillus*

acidophilus (PDB 4ia5) [6] as a search model. Crystals of OhyRe belong to the space group *P*6$_5$22, with two molecules in the asymmetric unit. For the refinement 2.8% of the reflections were set aside for the calculation of the R$_{free}$. Model building and water picking was performed with COOT.[26] The structure was initially refined by applying a simulated annealing protocol and in later refinement cycles by maximum-likelihood restrained refinement using PHENIX.refine.[27] Model quality was evaluated with MolProbity [28] and the JCSG validation server.[29] Secondary structure elements were assigned with DSSP [30] and ALSCRIPT [31] was used for secondary structure based sequence alignments. Figures were prepared using PyMOL.[32]

Multi-angle light scattering (MALS)

MALS experiment was performed at 18°C. OhyRe was loaded onto a Superdex 200 increase 10/300 column (GE Healthcare) that was coupled to a miniDAWN TREOS three-angle light scattering detector (Wyatt Technology) in combination with a RefractoMax520 refractive index detector. For calculation of the molecular mass, protein concentrations were determined from the differential refractive index with a specific refractive index increment (*dn*/*dc*) of 0.185 ml/g. Data were analyzed with the ASTRA 6.1.4.25 software (Wyatt Technology).

Accession numbers

The atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under the accession code 5ODO.

Acknowledgement

References

[1]   a) D. L. Greenway, K. G. Dyke, *Journal of General Microbiology* **1979**, *115*, 233-245;
        b) C. J. Zheng, J. S. Yoo, T. G. Lee, H. Y. Cho, Y. H. Kim, W. G. Kim, *FEBS Lett* **2005**, *579*, 5157-5162.
[2]   A. Hiseni, I. W. C. E. Arends, L. G. Otten, *ChemCatChem* **2015**, *7*, 29-37.
[3]   D. L. Greenway, K. G. Dyke, *J Gen Microbiol* **1979**, *115*, 233-245.
[4]   J. Jin, U. Hanefeld, *Chem Commun (Camb)* **2011**, *47*, 2502-2510.
[5]   a) L. E. Bevers, M. W. Pinkse, P. D. Verhaert, W. R. Hagen, *J Bacteriol* **2009**, *191*, 5010-5012;
        b) E.-Y. Jeon, J.-H. Lee, K.-M. Yang, Y.-C. Joo, D.-K. Oh, J.-B. Park, *Process Biochemistry* **2012**, *47*, 941-947.
[6]   A. Volkov, S. Khoshnevis, P. Neumann, C. Herrfurth, D. Wohlwend, R. Ficner, I. Feussner, *Acta Crystallogr D Biol Crystallogr* **2013**, *69*, 648-657.
[7]   M. Engleder, T. Pavkov-Keller, A. Emmerstorfer, A. Hromic, S. Schrempf, G. Steinkellner, T. Wriessnegger, E. Leitner, G. A. Strohmeier, I. Kaluzna, D. Mink, M. Schurmann, S. Wallner, P. Macheroux, K. Gruber, H. Pichler, *Chembiochem* **2015**, *16*, 1730-1734.
[8]   J. Schmid, L. Steiner, S. Fademrecht, J. Pleiss, K. B. Otte, B. Hauer, *Journal of Molecular Catalysis B: Enzymatic* **2017**.
[9]   Y. Wache, M. Aguedo, A. Choquet, I. L. Gatfield, J. M. Nicaud, J. M. Belin, *Appl Environ Microbiol* **2001**, *67*, 5700-5704.
[10]   a) C. T. Hou, *N Biotechnol* **2009**, *26*, 2-10;
        b) J. W. Song, E. Y. Jeon, D. H. Song, H. Y. Jang, U. T. Bornscheuer, D. K. Oh, J. B. Park, *Angew Chem Int Ed Engl* **2013**, *52*, 2534-2537.
[11]   M. Goodfellow, G. Alderson, *Journal of General MIcrobiology* **1977**, *100*, 25.
[12]   a) B. R. Langdhal, P. Bisp, K. Ingvorsen, *Microbiology* **1996**, *142*;
        b) S. C. Heald, P. F. B. Brandao, R. Hardicre, A. T. Bull, *Antonie van Leeuwenhoek* **2001**, *80*.
[13]   C. C. de Carvalho, M. M. da Fonseca, *Appl Microbiol Biotechnol* **2005**, *67*, 715-726.
[14]   H. Strnad, M. Patek, J. Fousek, J. Szokol, P. Ulbrich, J. Nesvera, V. Paces, C. Vlcek, *Genome Announc* **2014**, *2*.
[15]   L. Holm, P. Rosenstrom, *Nucleic Acids Res* **2010**, *38*, W545-549.
[16]   L. De Colibus, M. Li, C. Binda, A. Lustig, D. E. Edmondson, A. Mattevi, *Proc Natl Acad Sci U S A* **2005**, *102*, 12684-12689.
[17]   B. N. Kim, Y. C. Joo, Y. S. Kim, K. R. Kim, D. K. Oh, *Appl Microbiol Biotechnol* **2012**, *95*, 929-937.
[18]   Y. C. Joo, E. S. Seo, Y. S. Kim, K. R. Kim, J. B. Park, D. K. Oh, *J Biotechnol* **2012**, *158*, 17-23.
[19]   Y. C. Joo, K. W. Jeong, S. J. Yeom, Y. S. Kim, Y. Kim, D. K. Oh, *Biochimie* **2012**, *94*, 907-915.
[20]   E. Krissinel, K. Henrick, *J Mol Biol* **2007**, *372*, 774-797.
[21]   J. Ortega-Anaya, A. Hernandez-Santoyo, *Biochim Biophys Acta* **2015**, *1848*, 3166-3174.
[22]   F. W. Studier, *Protein Expr Purif* **2005**, *41*, 207-234.
[23]   A. Volkov, A. Liavonchanka, O. Kamneva, T. Fiedler, C. Goebel, B. Kreikemeyer, I. Feussner, *J Biol Chem* **2010**, *285*, 10353-10361.
[24]   W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125-132.
[25]   A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *Journal of applied crystallography* **2007**, *40*, 658-674.
[26]   P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486-501.
[27]   a) P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D.

C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213-221;
b) P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta Crystallogr D Biol Crystallogr* **2012**, *68*, 352-367.

[28]   V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 12-21.

[29]   H. Yang, V. Guranovic, S. Dutta, Z. Feng, H. M. Berman, J. D. Westbrook, *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 1833-1839.

[30]   W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577-2637.

[31]   G. J. Barton, *Protein Eng* **1993**, *6*, 37-40.

[32]   W. DeLano, *The PyMOL Molecular Graphics System* http://www.pymol.org . **2002**.

Supporting Information
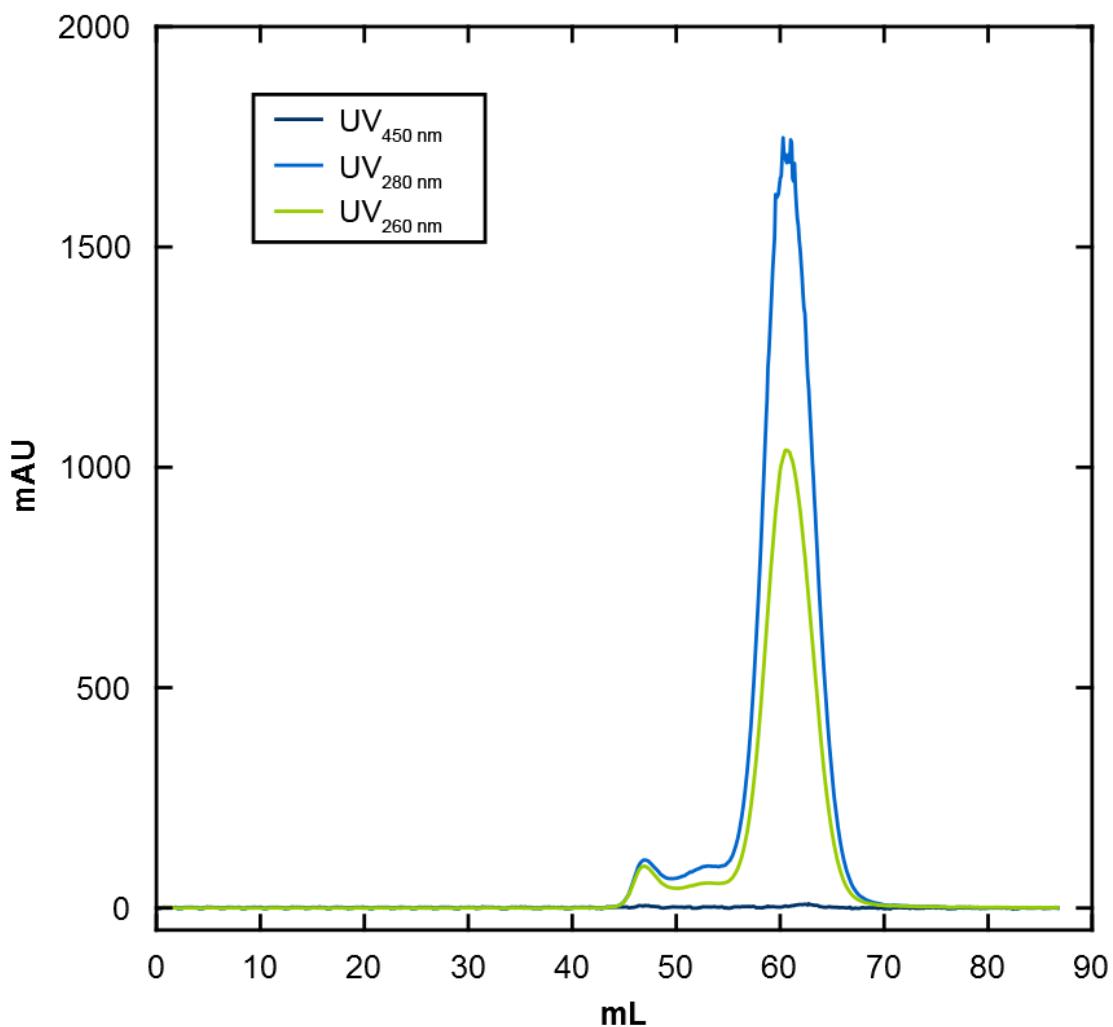
Rhodococcus erythropolis Oleate Hydratase: a New Member in the Oleate Hydratase Family Tree— Biochemical and Structural Studies

Jan Lorenzen+,[a] Ronja Driller+,[b] Ayk Waldow,[b] Farah Qoura,[a] Bernhard Loll,*[b, c] and Thomas Brück*[a]

cctc_201701350_sm_miscellaneous_information.pdf

Table of content:

1 Supplementary Figures and Sections

Figure S1: Adsorption spectra of final gel filtration chromatography of OhyRe at wave lengths of 260 nm, 280 nm and 450 nm.

The adsorption spectra of the gel filtration chromatography of OhyRe depicted in Fig. S1 show a clear peak for the enzyme at 260 nm and 280 nm. At 450 nm, the characteristic wave length for the detection of FAD, no corresponding peak could be detected. These data clearly show the loss of the FAD cofactor during the protein purification.

Figure S2: Sequence alignment of 14 different OHs from *Rhodococcus erythropulus* (Rre; HFam3), *Staphylococcus aureus* (Sau; HFam2), *Lysinibacillus fusiformis* (Lfu; HFam2), *Macrococcus caseolyticus* (Mca; HFam2), *Lactobacillus acidophilus* (Lac; HFam2), *Ochrobactrum anthropi* (Oan; HFam2), *Bifidobacterium breve* (Bbr; HFam2), *Streptococcus pyogenes* (Spy; HFam2), *Elizabethkingia meningoseptica* (Eme; HFam11), *Myroides odoratus* (Mod; HFam11), *Cellulophaga algicola* (Cal; HFam11), *Stenotrophomonas maltophilia* (Sma; HFam2), *Corynebacterium kroppenstedtii* (Ckr; HFam10) and *Chryseobacterium gleum* (Cgl; HFam11).
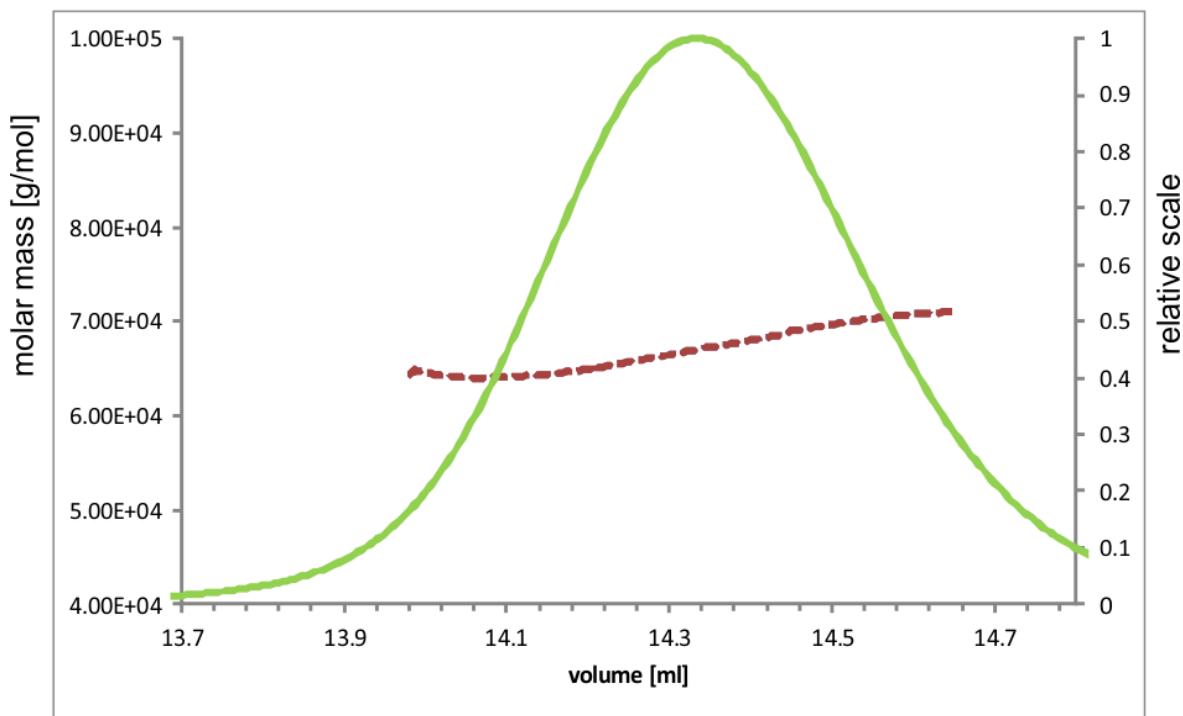
Figure S3: Superposition of putative active sites of OhyRe and OhyA. Highlighted in magenta are the residues M77 and V393 subjected to mutagenesis. (A) Residues of OhyRe residing in domain I are shown as orange sticks and residues of OhyA in gray (B) Residues of OhyRe residing in domain II are shown as green sticks and residues of OhyA in gray.

The superposition of the putative active sites of OhyRe (colored) and OhyA (gray), showed that the residues M77 and V393 from OhyRe are located in the exact same position in the three dimensional structure, like the residues E122 and T435 from OhyA.

OhyRe   MSSNLSHKAYMI GAGI GNL 19
OhyA MNPI TSKFDKVLNASSEYGHVNHEPDSSKEQQRNTPQKSMPFSDQI GNYQRNKGI PVQSYDNSKI YI I GSGI AGM 75

OhyRe SAAVYLI RDGEWNGEDI TI MGLD- MHGANDGESAATFQHQYGHRELGNDAGFI NRGGRMLNEETYENLWDVLSAV 93
OhyA   SAAYYFI RDGHVPAKNI TFLEQLHI DGGSL- - - - - - - - - - - - DGAGNPTDGYI I RGGREM- DMTYENLWDMFQDI 137

OhyRe PSLDNP- GKSVTDDI LDFDHAHPTHDVARLI DRRDGI RNKGENDYKHMQFDNKDRYLLTKLMTMPESDEAKLDDI 166
OhyA   PALEMPAPYSVLDEYRLI NDNDSNYSKARLI NNKGE- - - - I K- DFSKFGLNKMDQLAI I RLLLKN- - K- EELDDL 204

OhyRe SI EQWFEETPHFFTTNFWYMWETTFAFKRVSSAMELRRYMNRMI LEFSRI QTLAGVTRSPYNQYESI I LPMRTFL 241
OhyA   TI EDYF- - SESFLKSNFWTFWRTMFAFENWHSLLELKLYMHRFLHAI DGLNDLSSLVFPKYNQYDTFVTPLRKFL 277

OhyRe EGKGVKFVNELKI TEFVFKDTPLRDEI I VTGLDYENVRTGEKGRI DVAEGDFVFDTNGSI TDSSSI GDLDTPI VE 316
OhyA   QEKGVNI HLNTLVKDLDI HI N- - TEGKVVEGI I TEQ- - DGKEVKI PVGKNDYVI VTTGSMTEDTFYGNNKTAPI I 348

OhyRe DMR- - - - - YAPSALLWKQATEHFYDLGNPDKFFGDRAQSEWTSFTVTTSSHELI NEI SRI TKQLP- - - - - - GNAL 380
OhyA   GI DNSTSGQSAGWKLWKNLAAKSEI FGKPEKFCSNI EKSAWESATLTCKPSALI DKLKEYSVNDPYSGKTVTGGI 423

OhyRe NTFVDSNVLLSI VVHHQPHYHAQKENEGVFWGYCLFPRKDGDYVKKPFI EMTGREMLEETLGHLEALDESGTLAA 455
OhyA   I TI TDSNWLMSFTCNRQPHFPEQPDDVLVLWVYALFMDKEGNYI KKTMLECTGDEI LAELCYHLG- - - - - - I ED 491

OhyRe RRQEI MDSVVNSI PSHMPYASALFNRRAVGDRPLVVPKHSKNLAFI SQFAELPFDMVFTEQYSVRCAQVAVYKFL 530
OhyA   QLENVQKN- TI VRTAFMPYI TSMFMPRAKGDRPRVVPEGCKNLGLVGQFVETNNDVVFTMESSVRTARI AVYKLL 565

OhyR   GI PEDKLTKMHHYEKDPKVLAKAAVTMFR 559
OhyA   NLNKQVP- DI NPLQYDI RHLLKAAKTLNDDKPFVGEGLLRKVLKGTYFEHVLPAGAAEEEEHESFI AEHVNKFRE 639

OhyA   WVKGI RG 646

Figure S2: Structure-based sequence alignment of OhyRe and OhyA. Secondary structure assignment was performed with DSSP based on the respective crystal structure. Secondary structure elements of OhyRe are depicted above the primary sequence and below for OhyA. The sequence of OhyRe is colored according to the domain architecture. Slanted lines indicated sections of OhyRe and OhyA that are not included in the crystal structures. Gray triangles indicate residues of OhyRe that are involved in hydrogen bonding contact to Mg^{2+} coordinating water molecules. Black triangle shows F548 in π-stacking interaction with the flavine moiety of FAD. Other residues in hydrogen bonding distance to FAD are labeled with a black circle. Golden circles show OhyA E122 and Y241 that were proposed to be of catalytic importance.
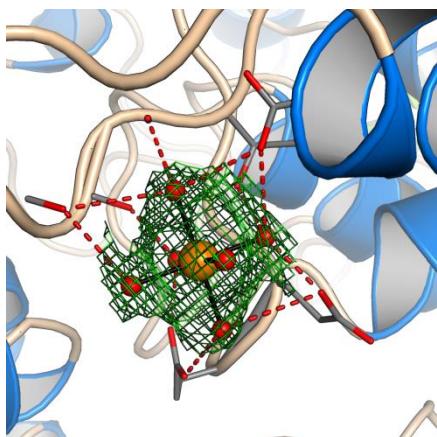
Figure S3: SEC/MALS analysis of OhyRe. Refractive index signal is shown as green curve. Solid, green curve represents the refractive index trace and the red curve the molecular mass at the corresponding elution volumes.

Section S1: Mg²⁺ binding site

In each monomer, we detected one Mg^{2+} binding site, being solvent exposed on the backside of the putative active site of OhyRe (Fig. 5A and S6). The binding site is located at the intersection of domain I, domain II and domain III. The Mg^{2+}-ion is octahedral coordinated by six water molecules (Fig. S6 and S7) that are bound to the carboxylates of Glu80, Glu81, Asp111 and the hydroxyl of Thr223 as well as to the backbone carbonyls of Ala220, Gly221, and Val222. Notably amino acid residues in the second coordination sphere of Mg^{2+} are not conserved within other OHs. For the OhyA of *E. meningoseptica* a bound Ca^{2+}-ion has been detected by plasma optical emission spectroscopy analysis, but no influence on the activity could be associated (Bevers et al. 2009). Furthermore, in its crystal structure no Ca^{2+} could be identified (Engleder et al. 2015). Given the close proximity to the putative active site, we questioned whether the Mg^{2+} could have an influence on the catalytic activity of the OhyRe enzyme. Hence we tested the enzymatic activity under standard conditions and added EDTA to a final concentration of 5 mM to the reaction mixture. The addition of EDTA had no detectable effect on the conversion activity of OhyRe, suggesting that the Mg^{2+} doesn't play a crucial role in the hydration reaction, but might have an effect on the protein stability.

Figure S4: Octahedral Mg^{2+} coordination. Same color coding as in Figure 5A. The bound Mg^{2+} and coordinating waters are shown as orange and red spheres, respectively. Hydrogen bonding interactions, with a distance cut-off < 3.1 Å, to amino acid residues are shown as red dashed lines.

Figure S5: Octahedral Mg^{2+} coordination to water molecules is indicated by black lines. Hydrogen bonding (< 3.1 Å) interactions to amino acid residues are shown as red dashed lines. 2Fo-Fc electron density contoured at 1.0 σ depicted as green mesh.

Table S1: Crystallographic data collection and model refinement statistics.

Data Collection	
Wavelength [Å]	0.98141
Temperature [K]	100
Space group	$P6_522$
Unit Cell Parameters a=b, c [Å]	261.75, 127.30
Resolution range [Å][a]	50.00 - 2.64 (2.80 - 2.64)
Reflections [a] Unique Completeness [%] Multiplicity	 75230 (11910) 99.8 (99.4) 22.2 (22.3)
Data quality [a] Intensity [$I/\sigma(I)$] R_{meas} [%] $CC_{1/2}$[b] Wilson B value [Å^2]	 13.66 (1.80) 31.0 (217.4) 99.6 (78.8) 46.2
Refinement	
Resolution range [Å][a]	49.47 - 2.64 (2.73 - 2.64)
Reflections [a] Number Test Set (2.8 %)	 75186 (7348) 2097 (204)
R_{work}	0.190 (0.306)
R_{free}	0.221 (0.343)
Contents of the Asymmetric Unit Protein, Molecules, Residues, Atoms	 1, 2, 1057, 8603

Ligands (Molecules)	Formate (14), glycerin (2),Magnesium (2)
Water, Molecules	417
Mean Temperature factors [Å^2]^b	
All Atoms	36.5
Macromolecules	36.6
Ligands	44.2
Water Oxygens	34.0
RMSD from Target Geometry	
Bond Lengths [Å]	0.011
Bond Angles [°]	0.84
Validation Statistics [c]	
Ramachandran Plot	
Residues in Allowed Regions [%]	3.0
Residues in Favored Regions [%]	97.0
MOLPROBITY Clashscore [d]	4.81

[a] data for the highest resolution shell in parenthesis
[b] calculated with PHENIX (Adams et al, 2010)
[c] calculated with MOLPROBITY (Chen et al, 2010)
[d] Clashscore is the number of serious steric overlaps (> 0.4) per 1,000 atoms.

Table S2: Primers used for OhyRe mutant construction

name	sequence	feature
OhyRe-fw (for overlap)	5'-GGA GAT ATA CCA TGG GCA GCA-3'	*Nco*I cutting site (red)
OhyRe-rv (for overlap)	5'-GTG GTG CTC GAG TTA ACG AAA-3'	*Xho*I cutting site (red)
M77E fw	5'-TTT ATC AAT CGT GGT GGT CGT GAA CTG AAT GAA G-3'	Methionine to glutamic acid (red)
M77E rv	5'-CAC GAT TGA TAA AAC CTG CAT CAT TAC CCA GT-3'	
V393T fw	5'-C AAT GTG CTG CTG AGC ATT ACC GTT CAT CAT C-3'	Valine to threonine (red)
V393T rv	5'-CAG CAG CAC ATT GCT ATC AAC AAA GGT ATT CA-3'	

Publication IV

Lida Rostock, Ronja Driller, Stefan Grätz, Dennis Kerwat, Leonard von Eckardstein, Daniel Petras, Claudia Alings, Franz-Josef Schmitt, Thomas Friedrich, Markus Wahl, Bernhard Loll, Andi Mainz, Roderich D. Süssmuth (2018), Molecular insights into antibiotic resistance: how a binding protein traps albicidin, *Nature Communications* 9:3095

Molecular insights into antibiotic resistance - how a binding protein traps albicidin

Lida Rostock[1], Ronja Driller [2], Stefan Grätz[1], Dennis Kerwat[1], Leonard von Eckardstein[1], Daniel Petras [1,3], Maria Kunert[1], Claudia Alings[2], Franz-Josef Schmitt[4], Thomas Friedrich [4], Markus C. Wahl [2,5], Bernhard Loll [2], Andi Mainz [1] & Roderich D. Süssmuth[1]

The worldwide emergence of antibiotic resistance poses a serious threat to human health. A molecular understanding of resistance strategies employed by bacteria is obligatory to generate less-susceptible antibiotics. Albicidin is a highly potent antibacterial compound synthesized by the plant-pathogenic bacterium *Xanthomonas albilineans*. The drug-binding protein AlbA confers albicidin resistance to *Klebsiella oxytoca*. Here we show that AlbA binds albicidin with low nanomolar affinity resulting in full inhibition of its antibacterial activity. We report on the crystal structure of the drug-binding domain of AlbA (AlbAS) in complex with albicidin. Both α-helical repeat domains of AlbAS are required to cooperatively clamp albicidin, which is unusual for drug-binding proteins of the MerR family. Structure-guided NMR binding studies employing synthetic albicidin derivatives give valuable information about ligand promiscuity of AlbAS. Our findings thus expand the general understanding of antibiotic resistance mechanisms and support current drug-design efforts directed at more effective albicidin analogs.

[1] Institut für Chemie, Biologische Chemie, Technische Universität Berlin, Straße des 17. Juni 124, 10623 Berlin, Germany. [2] Institut für Chemie und Biochemie, Strukturbiochemie, Freie Universität Berlin, Takustraße 6, 14195 Berlin, Germany. [3] Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0751, USA. [4] Institut für Chemie, Biophysikalische Chemie, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany. [5] Helmholtz Zentrum Berlin für Materialien und Energie, Macromolecular Crystallography, Albert-Einstein-Straße 15, D-12489 Berlin, Germany. These authors contributed equally: Lida Rostock, Ronja Driller. Correspondence and requests for materials should be addressed to R.D.S. (email: suessmuth@chem.tu-berlin.de)

The increasing number of antibiotic-resistant bacteria worldwide has become a major public health threat. Discovery and study of different bacterial resistance strategies as well as the development of new antimicrobial compounds, which are also able to overcome the protecting outer-membrane barrier of Gram-negative bacteria, are of great urgency[1]. Albicidin is a phytotoxic small molecule synthesized by the Gram-negative plant pathogen *Xanthomonas albilineans* that causes leaf scald disease in sugarcane plants[2]. Besides its phytotoxicity, albicidin is bactericidal in the nanomolar range against Gram-positive and in particular against Gram-negative bacteria with low minimal inhibitory concentrations (MIC), e.g., against *Staphylococcus aureus* (4.0 µg mL^{-1}), *Salmonella enteritidis* (0.5 µg mL^{-1}), *Pseudomonas aeruginosa* DSM 117 (1.0 µg mL^{-1}), and *Escherichia coli* (0.063 µg mL^{-1})[3,4]. The molecular target of albicidin is DNA gyrase (topoisomerase II), an enzyme essential for bacterial DNA replication. Albicidin has been reported to block the ATP-dependent DNA cleavage-religation step at the gyrase A subunit[3]. The biosynthesis of albicidin is based on a hybrid polyketide synthase/nonribosomal peptide synthetase[5,6]. The structure of albicidin consists of six building blocks (Fig. 1a): methyl *p*-coumaric acid (MCA-1) at the N terminus; two *p*-aminobenzoic acids (*p*ABA-2 and *p*ABA-4); the unusual non-proteinogenic α-amino acid β-cyano-L-alanine (L-Cya-3) centered in the structure; and two 4-amino-2-hydroxy-3-methoxybenzoic acids (*p*MBA-5 and *p*MBA-6) at the C-terminal end. The structure elucidation[5,7] of albicidin and a total synthesis route[4] enabled initial structure-activity relationship (SAR) studies[8–10].

Previously, a number of resistance mechanisms against albicidin have been described: examples include the nucleoside transporter Tsx, for which mutations have been reported that compromise its abilities to import albicidin[11,12], or the endopeptidase AlbD from *Pantoea dispersa*[13], which cleaves albicidin into two inactive fragments[14]. Another strategy that counteracts the antibacterial effect of albicidin is drug binding, as it is also known for the tetracycline-binding protein (TetR family)[15] or the thiostrepton-binding protein (MerR family)[16]. Examples of such albicidin-binding proteins are AlbA from *Klebsiella oxytoca*[17] and AlbB from *Alcaligenes denitrificans*[18], which exert protective effects for survival of the host strains. Previous studies on AlbA/AlbB have demonstrated non-covalent binding of albicidin[17,18]. Furthermore, far-ultraviolet (UV) circular dichroism (CD) spectroscopy has indicated a mostly α-helical structure for AlbA[19]. Specific amino-acid residues such as His125 have been suggested to play a role in albicidin binding[19,20]. Due to high similarities to the DNA-binding domains of transcriptional regulator proteins NifA and NtrC (nitrogen regulatory protein) a classification of AlbA/AlbB as transcription activator proteins has been suggested[19].

Transcriptional regulator proteins are widespread in nature. For instance, the MerR (Mercury Resistance) transcriptional regulator family is known to activate several multi-drug resistance (MDR) systems in response to environmental stress[21]. Members of this family are characterized by a highly conserved N-terminal winged helix-turn-helix (HTH) DNA-binding motif. The HTH motif is connected by a long coiled-coil linker to a C-terminal effector-binding domain that differs in structure and function among the family members[21]. Activation of these regulators occurs in response to oxidative stress and heavy metals or small molecules, e.g., antibiotics, binding to the effector domain[22]. Well-known members of the MerR family are the transcriptional regulators BmrR from *Bacillus subtilis*[22,23], and SoxR from *Escherichia coli*[24]. Another prominent and well-described MerR system is the thiopeptide-binding protein TipA (thiostrepton-induced protein A) from *Streptomyces lividans*[25,26]. The *tipA* gene encodes for two alternate in-frame translation products: the long form TipAL consisting of the N-terminal HTH DNA-binding domain (TipAN), and the shorter C-terminal thiopeptide drug-binding domain (TipAS)[27]. Drug binding to TipAL leads to an upregulation of its own expression and confers resistance against several thiopeptide antibiotics, e.g., thiostrepton, nosiheptide, and promothiocin A[16]. Previously, Grzesiek and co-workers have solved the solution structure of TipAS bound to various thiopeptides[16,27], which for the first time gave insights into binding determinants and domain dynamics of the protein.
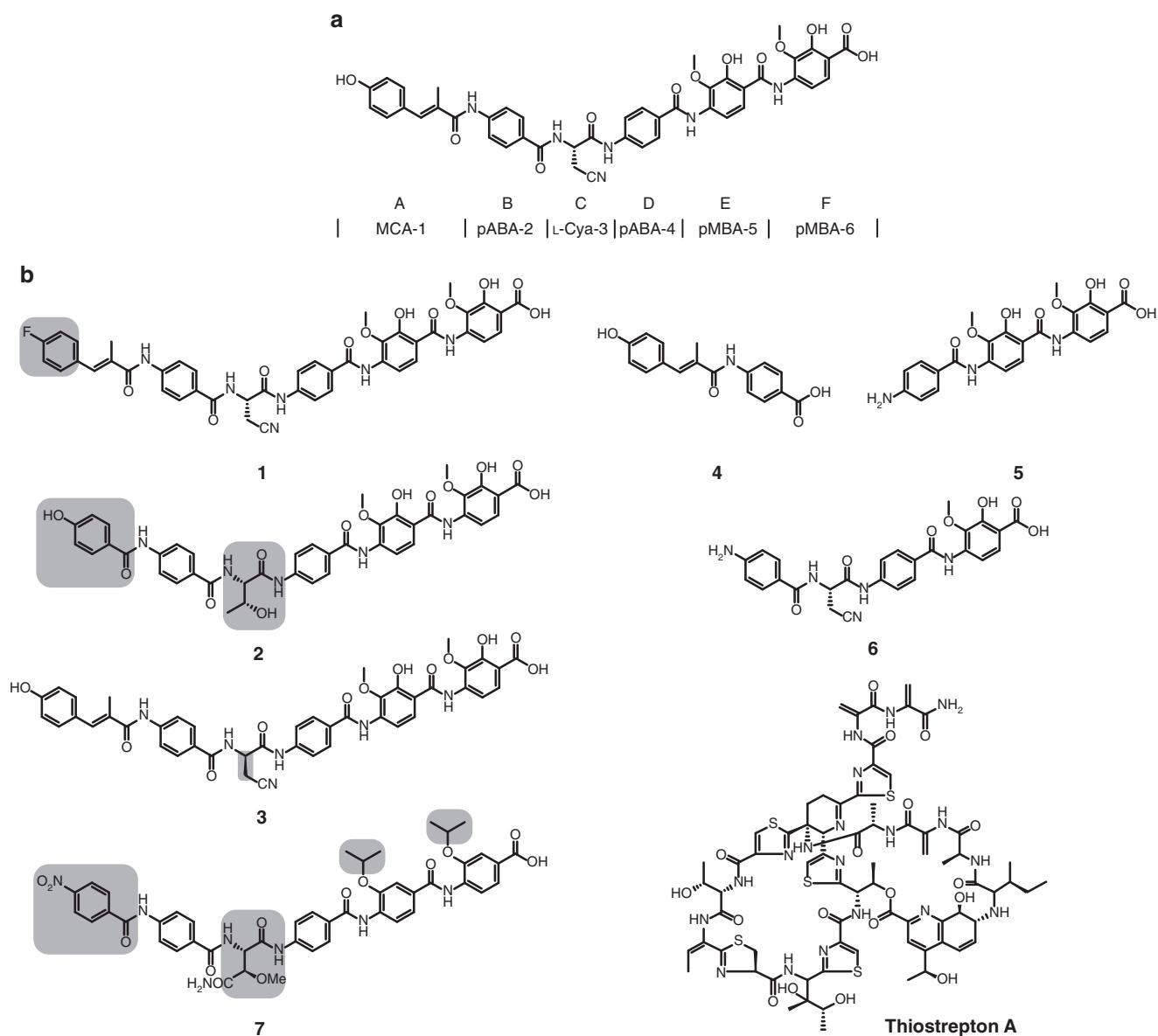
In this work, we characterized the MerR-like drug-binding protein AlbA that displays a high binding affinity to albicidin. We elucidated the crystal structure of the complex between albicidin and the drug-binding domain AlbAS, demonstrating that this domain is structurally distinct from other drug-binding domains of MerR family members, such as BmrR[23] or TipAS[27]. The unusual topology of AlbAS comprises two tandem domains that are required to wrap around and fully enclose the drug in the core of the protein. Nuclear magnetic resonance (NMR), CD, and fluorescence spectroscopic measurements provided detailed insights into the binding mechanism and the structural elements required for drug binding of AlbA. Consistent with our structural data, binding studies on synthetic albicidin derivatives revealed ligand promiscuity of AlbAS for aromatic acylpentapeptides as well as further supported our model of the binding mechanism. Our findings represent valuable knowledge about antibiotic capture as an important resistance mechanism and how to possibly circumvent it in the case of albicidin.

Results

AlbA is a MerR-like transcriptional regulator. A sequence similarity search using Basic Local Alignment Search Tool (BLAST) for AlbA (accession number (AC): Q8KRS7)[17] and AlbB (AC: Q44567)[18] suggests that they belong to the MerR family of transcriptional regulators (*E*-values of 10^{-160} (AlbA) and 10^{-37} (AlbB)). Moreover, AlbA shows high structural homology to the thiopeptide-binding protein TipAS (Supplementary Figure 1). Further homologous proteins were identified in various Gram-negative bacteria belonging to the family of Enterobacteriaceae like *Raoultella ornithinolytica*, *Leclercia adecarboxylata*, or *Enterobacter mori* (Supplementary Figure 2A and Supplementary Figure 3). Bioinformatic comparison of AlbA and AlbB showed that while being distantly related in primary structure, both proteins most probably share high secondary structure similarity (Supplementary Figure 2B). A striking difference between both proteins is the isoelectric point (AlbA: pI = 4.99, AlbB: pI = 10.15) due to the reversed content of acidic (15% for AlbA versus 8% for AlbB) and basic amino-acid residues (10% for AlbA versus 15% for AlbB).

Bioinformatic analysis of the *K. oxytoca* and *A. denitrificans* genomes revealed a putative HTH motif in AlbA and AlbB that might be involved in DNA binding. This domain is conserved in several MerR-transcription activators[21]. In analogy to the TipA system[27], we assumed that the *albA*/*albB* genes each encode for two proteins: a full-length protein and a truncated protein as a result of two in-frame translation products. Based on the TipA nomenclature[27], we termed the full-length versions AlbAL/AlbBL (AlbAL 40 kDa, 348 amino acids, AC: WP_016808841). Hence, AlbAL/AlbBL consist of the MerR-characteristic N-terminal HTH DNA-binding domain and the C-terminal drug-binding domain AlbAS/AlbBS. Sequence alignments with TipAL and secondary structure prediction illustrated that AlbAS (25.8 kDa, 221 amino acids) may comprise an unusual tandem arrangement of two TipAS-like drug-binding domains (Supplementary Figure 1). Interested in the capture mechanism toward albicidin, we subsequently focused on AlbAS and validated the results by employing AlbAL in key experiments (Supplementary Figure 4A).

Fig. 1 Structure of albicidin. **a** Albicidin is composed of six building blocks (denoted A–F): *p*-coumaric acid (MCA-1); *p*-aminobenzoic acid (*p*ABA-2 and *p*ABA-4); non-proteinogenic α-amino acid L-Cya-3; and two 4-amino-2-hydroxy-3-methoxybenzoic acids (*p*MBA-5 and *p*MBA-6). **b** Overview of tested compounds. Structural variations with respect to albicidin are highlighted in gray
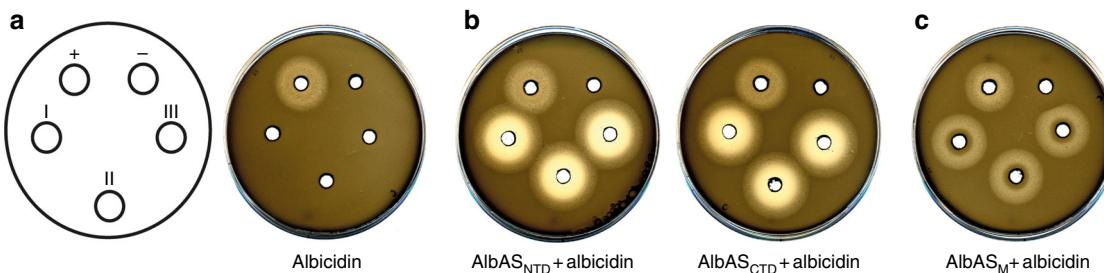
AlbA captures albicidin with low nanomolar binding affinity. Initial in vitro agar diffusion assays with *E. coli* and a pre-incubated AlbA-albicidin mixture clearly demonstrated the albicidin-neutralizing effect of both AlbAS (Fig. 2a) and AlbAL (Supplementary Figure 5A).

MIC values for albicidin were determined against four *K. oxytoca* and seven *Klebsiella pneumoniae* strains as well as several ESKAPE organisms such as *P. aeruginosa*, *R. ornithinolytica*, *Acinetobacter baumanii*, and *Enterobacter cloacae*. All strains were resistant against albicidin with MIC values of ≥64 μg mL^{-1}. By bioinformatic analysis, we identified AlbA homologs in many of these pathogens (Supplementary Figure 3) indicating that AlbA may function as a widespread resistance mechanism.

Far-UV CD spectroscopy of AlbAS and AlbAL in the absence and presence of albicidin showed signal minima at wavelengths of 209 and 222 nm, which are characteristic of α-helical secondary structures (Fig. 3a and Supplementary Figure 5B). No significant changes in secondary structure were induced by albicidin binding

to both proteins. However, assessing the thermodynamic protein stabilities showed that the melting temperature, T_m, of AlbAS increased significantly upon albicidin binding from 66 to 81 °C (Fig. 3b). Binding of albicidin to AlbAL caused a less-pronounced increase in T_m from 71 to 74 °C (Supplementary Figure 5C).

We further inspected the structural rearrangement of AlbAS upon albicidin binding via two-dimensional (2D) NMR experiments with ^{15}N-labeled AlbAS. The ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectrum of AlbAS was characterized by considerable line-broadening and signal overlap (Fig. 3e). Peak picking yielded only 95 of the expected 209 cross-peaks. By contrast, addition of albicidin not only caused chemical shift changes but also the appearance of new cross-peaks (approximately 175 cross-peaks in total). For example, the well-resolved indole region of tryptophans showed three versus nine cross-peaks in the absence and presence of the ligand, respectively. A tentative assignment of these side-chain signals using Trp-to-Phe mutagenesis revealed that at least W27, W56, W133, and W195

Fig. 2 Agar diffusion assays. **a–c** AlbAS, AlbAS$_{NTD}$, AlbAS$_{CTD}$, and AlbAS$_M$ (40 μM) with 40 μM albicidin (in triplicates I–III). The first plate illustrates the sample arrangement on the assay plates. Positive control with only 40 μM albicidin (+) and negative control with 40 μM protein in 5% DMSO and buffer (−) are shown on top. Representative results are shown for three independent experiments

became detectable upon albicidin binding (Supplementary Figure 6). Interestingly, the observation of doubled resonances for the side chains of W56 and W133 indicated that these residues adopt two different and slowly interconverting conformations in the albicidin–AlbAS complex (Supplementary Figure 6). The NMR data suggest that the extensive line-broadening for the unbound AlbAS is caused by internal dynamics on a microsecond to millisecond timescale, which are frozen in a more rigid albicidin-bound state.

Size-exclusion chromatography with multi-angle light scattering (SEC-MALS) measurements yielded molecular masses, M, of 27.1 and 27.2 kDa for AlbAS in the absence and presence of albicidin, respectively (Supplementary Figure 7A). In contrast to monomeric AlbAS, the full-length protein AlbAL formed a dimer in solution in both the unbound state (M of 82.2 kDa) and albicidin-bound state (M of 85.5 kDa) (Supplementary Figure 7B). Dimerization has been shown for other MerR proteins as well, enabling DNA promoter binding[28,21]. In agreement with a dimeric state, ^{2}H,^{15}N-labeled AlbAL yielded rather broad resonance lines in ^{1}H-^{15}N transverse relaxation-optimized spectroscopy (TROSY) experiments. More importantly, albicidin binding to AlbAL caused many of the characteristic chemical shift perturbations observed for AlbAS–albicidin indicating a very similar binding mode (Supplementary Figure 5D and E).

In order to estimate the binding strength between AlbAS and albicidin, we monitored characteristic signals in a well-resolved reporter region of the ^{1}H-^{15}N HSQC spectrum during a titration series (Supplementary Figure 8A). The simultaneous observation of signals arising from both AlbAS and its albicidin-bound state, as well as and the mutual change in populations of these two states indicated a slow-exchange regime in agreement with strong ligand binding in the nM–pM range (Supplementary Figure 8B and Supplementary Figure 8C)[29]. Owing to this very strong interaction and the high protein concentrations required in the NMR experiments, we employed fluorescence spectroscopy to avoid ligand depletion effects and to ultimately determine the dissociation constant, K_d, of albicidin binding to AlbAS and AlbAL. The experimental read-out was based on quenching of fluorescence emission of AlbAS/AlbAL upon binding of albicidin (Fig. 4a). Analysis of the observed binding response for AlbAS yielded a K_d of 5.6 ± 0.2 nM and indicated positive cooperativity with a Hill coefficient, n, of 3.0 ± 0.5 (Fig. 4b). The full-length protein AlbAL showed a similar binding affinity (K_d of 7.4 ± 0.9 nM) and cooperativity (n of 2.0 ± 0.2) (Supplementary Figure 5F). These results demonstrated that the drug-binding domain of AlbA acts as a high-affinity capture system for albicidin.
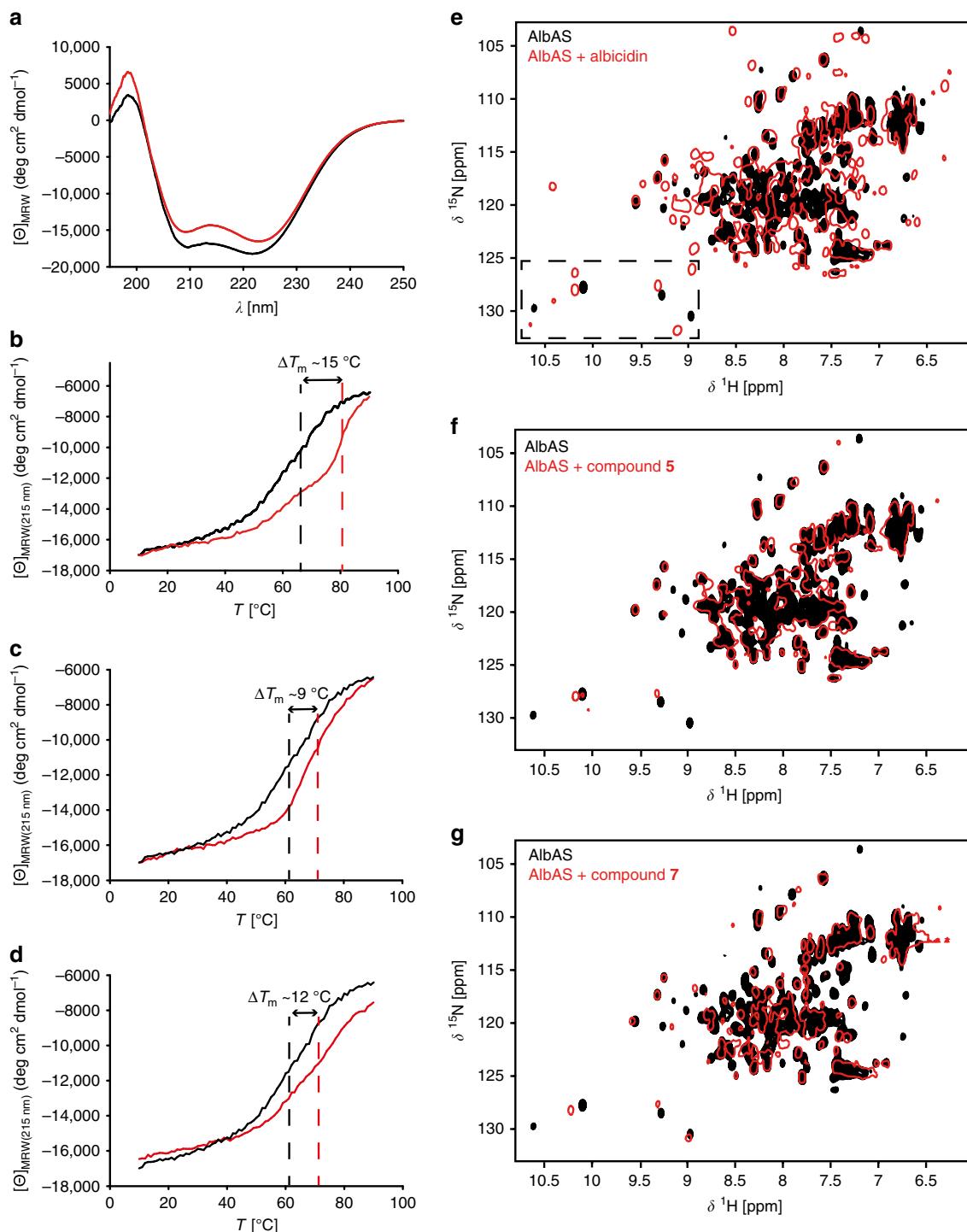
It has been shown previously that albicidin exhibits a half maximal inhibitory concentration (IC$_{50}$) of 40 nM for its molecular target DNA gyrase[3]. To investigate whether AlbAS is capable of protecting DNA gyrase against albicidin, we performed an NMR-based in vitro competition experiment employing the

gyrase construct GyrBA59 from *E. coli* (Supplementary Figure 9A). Purified and enzymatically active GyrBA59 (Supplementary Figure 9B and C) was incubated with MgCl$_2$, ATP, supercoiled pBR322 plasmid DNA, and equimolar amounts of AlbAS. An equimolar amount of albicidin was then added to the mixture to allow for direct competition. The overlay of ^{1}H-^{15}N correlation spectra showed that AlbAS fully populated the albicidin-bound state with no free AlbAS being detectable (Supplementary Figure 9D and E). This demonstrated that AlbAS can efficiently sequester albicidin before the inhibitor targets DNA gyrase.

The structure of the AlbAS–albicidin complex. We obtained well-diffracting crystals for the AlbAS–albicidin complex and solved its structure by single-wavelength anomalous dispersion with selenomethionine-labeled AlbAS (Fig. 5a). Model refinement (1.7 Å resolution) converged with excellent statistics and geometry (Supplementary Table 1 and Supplementary Figure 10A and B). The structure of AlbAS could be fully modeled except for the C-terminal residues 215–221. An asymmetric unit contains two copies of the protein that are virtually identical with a root mean square deviation (rmsd) of 0.6 Å for all Cα atoms. AlbAS adopts an all-α-helical fold in agreement with CD spectroscopy (Fig. 3a). As postulated from our bioinformatic analysis (Supplementary Figure 11A and Supplementary Figure 1), the structure of AlbAS is organized in two repeat units, i.e., the N-terminal domain (NTD, residues M1-R112) and the C-terminal domain (CTD, residues Y113-E221). A 24-residue α-helix (P92-R112) connects the NTD with the CTD. To reflect the tandem arrangement in AlbAS, we designated helices of the NTD as α1-α6 and the equivalent helices of the CTD as α1′-α6′ (Supplementary Figure 11A). The NTD and CTD are arranged relative to each other via 2$_1$ screw pseudosymmetry and can be superimposed (using boundaries D20-L114 and D128-A215) with an rmsd of 1.7 Å (Fig. 5b and Supplementary Figure 11A).

A DALI[30] search for structurally related folds resulted in very low Z-scores. The most related structure with a Z-score of 5.8 is the drug-binding domain TipAS (PDB 2MBZ, in complex with promothiocin A)[27] with a sequence identity of only 13%. TipAS (G111-P253) superimposes with both the NTD (M1-R112) and CTD (Y113-Q214) of AlbAS with Cα rmsd values of 2.2 and 5.0 Å, respectively. The deviations mainly arise from structural differences at the N-terminal regions, whereas helices α2-α6 of the NTD (D20-L114) and helices α2′-α6′ of the CTD (Y128-A215) can be well aligned to helices α9-α13 of TipAS (Y155-P253) yielding rmsd values of 2.5 and 2.8 Å, respectively.
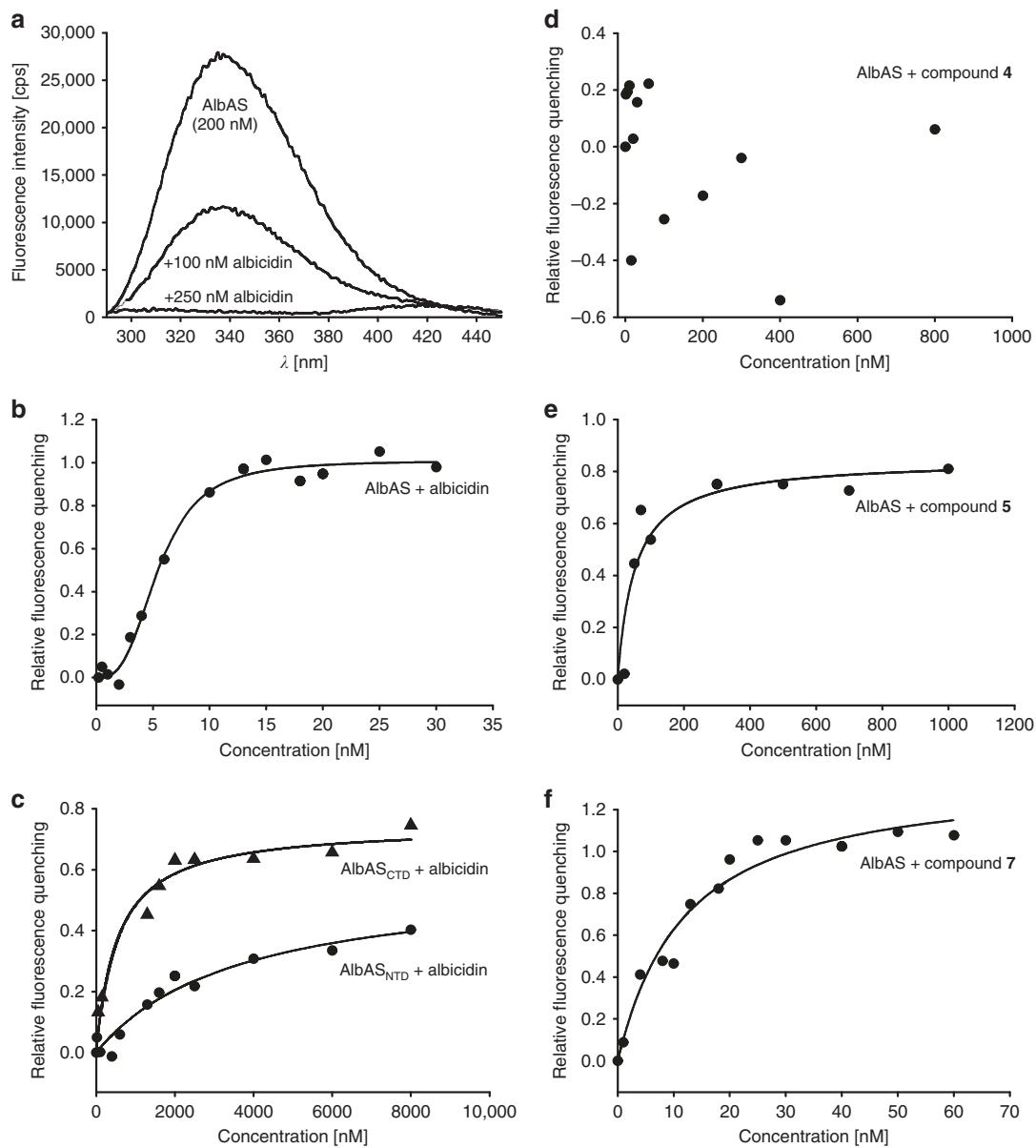
To test the albicidin-binding potential of the individual domains, we prepared constructs comprising either the NTD (AlbAS$_{NTD}$) or the CTD (AlbAS$_{CTD}$) (Supplementary Figure 4B). Both constructs adopted folded states in solution as judged by

Fig. 3 Impact of albicidin binding on structure and stability of AlbAS. **a** Far-UV CD spectra of AlbAS in the absence (black) or presence (red) of albicidin (molar ratio 1.5:1 albicidin:AlbAS) at 21 °C. **b–d** CD-detected thermal unfolding of AlbAS in the absence (black) or presence (red) of albicidin, compound **5**, or compound **7**. T_m values are indicated by dashed lines. For the overlay, the $[\Theta]_{MRW(215\,nm)}$ signals of AlbAS–albicidin and AlbAS–compound **5** were scaled to that of AlbAS alone using a factor of 1.157 and 1.091, respectively. **e–g** Overlay of ^1H-^{15}N HSQC NMR spectra of AlbAS in the absence (black) or presence (red) of albicidin, compound **5**, or compound **7**, respectively. CD experiments were performed once. NMR experiments with the AlbAS–albicidin complex were performed five times and a representative spectrum is shown in **e**. NMR experiments using albicidin derivatives were conducted once

^1H-^{15}N HSQC spectra (Supplementary Figure 12A and B). However, in agar diffusion assays they failed to suppress the antibacterial effects of albicidin (Fig. 2b). This observation can be ascribed to diminished affinities toward albicidin, as fluorescence spectroscopy yielded K_d values of $3.4 \pm 0.2\,\mu M$ (AlbAS$_{NTD}$) and $0.6 \pm 0.1\,\mu M$ (AlbAS$_{CTD}$), respectively (Fig. 4c).
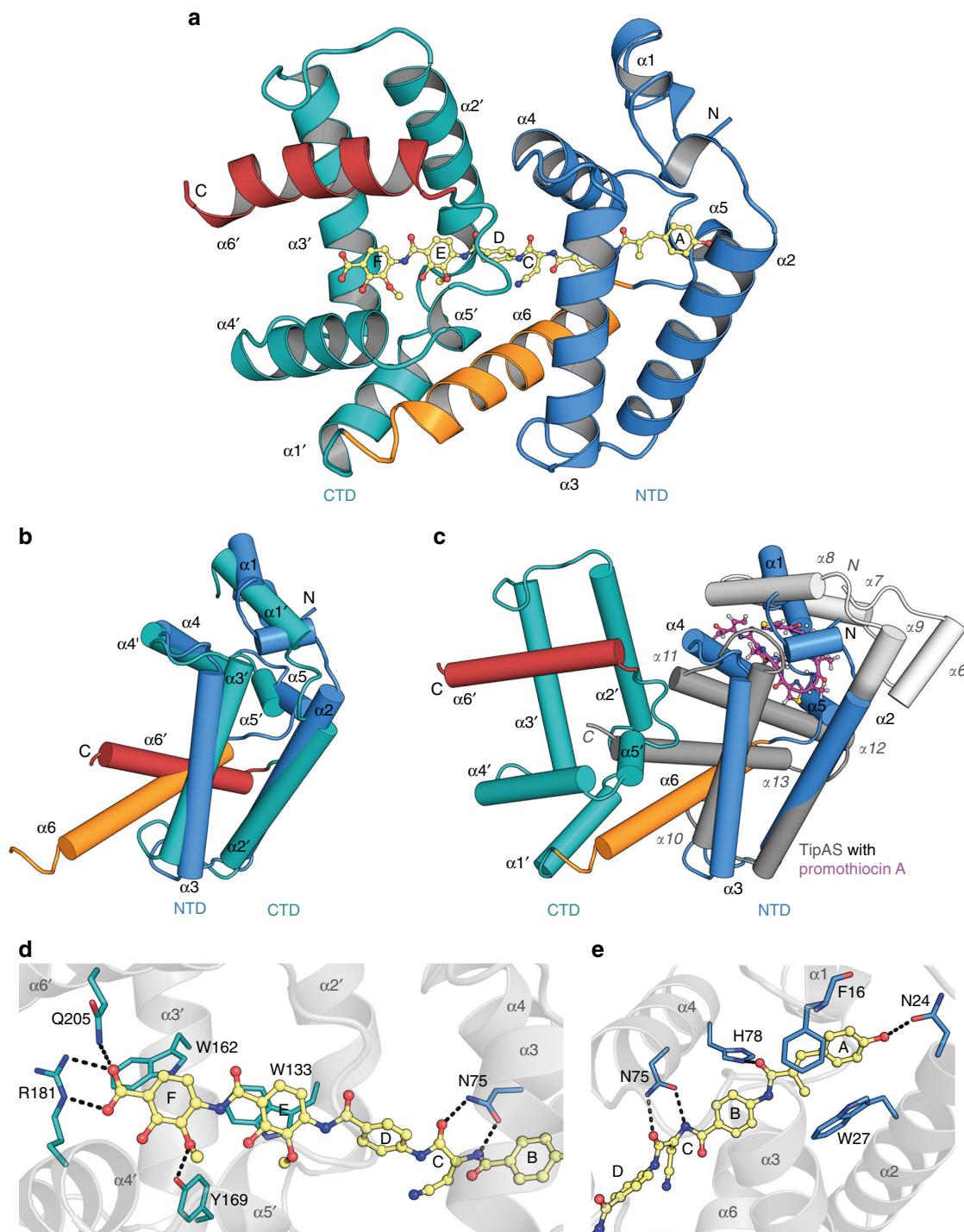
In contrast to TipAS, which binds its macrocyclic ligands in an N-terminal pocket (Fig. 5c)[27], the elongated and rather rigid albicidin molecule is accommodated in a central tunnel of AlbAS thereby pervading both, its NTD and CTD. Interestingly, both repeat units utilize an equivalent pocket to harbor the different segments of albicidin, albeit with opposite entrance sites

Fig. 4 Determination of binding affinities by monitoring quenching of fluorescence emission of AlbAS. **a** Fluorescence quenching of AlbAS (200 nM; $\lambda_{exc} =$ 280 nm) with increasing albicidin concentrations. **b** Nonlinear regression (Eq. (2), see Methods section) of fluorescence quenching data yielding a K_d of 5.6 ± 0.2 nM and a Hill coefficient, n, of 3.0 ± 0.5 for the interaction of albicidin with AlbAS (20 nM). **c** Nonlinear regression of the quenching data yielded K_d values of 3.4 ± 0.2 and 0.6 ± 0.1 µM for the interaction of albicidin with the dissected domains AlbAS$_{NTD}$ (100 nM) and AlbAS$_{CTD}$ (100 nM), respectively. **d** AlbAS (20 nM) titrated with compound **4** did not lead to significant quenching of the fluorescence emission signal. **e** Titration of AlbAS (20 nM) with compound **5** yielded a K_d of 55.5 ± 3.6 nM (Eq. (2), see Methods section). **f** Titration of AlbAS (20 nM) with compound **7** yielded a K_d of 14.0 ± 2.9 nM (Eq. (3), see Methods section). Representative binding curves are shown for two independent experiments each. Errors are given as s.d

(Supplementary Figure 11B). The binding tunnel is decorated by residues of all helices, except for the N-terminal helix α1 (Fig. 5a), and exhibits a predominantly hydrophobic inner surface. The building blocks A, B, and C of albicidin are surrounded by the NTD of AlbAS, whereas building blocks D, E, and F are enclosed by the CTD. The high-salt crystallization conditions reflect the hydrophobic character of the interaction, which is mediated by several hydrophobic residues pointing into the tunnel, e.g., F16, W27, W56, L60, L71, T88, I95, L130, W133, P134, W162, and W195 (an overview of all interacting residues is given in Supplementary Figure 13). Five out of eight Trp side chains (W27, W56, W133, W162, and W195) line the inner surface of the tunnel. Residues F16, W27, W133, and W162 are π-stacked to

the aromatic building blocks A, E, and F of albicidin. Besides strong hydrophobic interactions, a salt bridge between the carboxylate moiety of albicidin and R181 of the CTD might play an important role for the orientation and fixation of the ligand in the tunnel (Fig. 5d). Furthermore, several hydrogen bonds contribute to the binding of albicidin. For example, pMBA-6 (building block F) establishes two hydrogen bonds to Y169 and Q205 of the CTD (Fig. 5d), while N24 and H78 of the NTD hydrogen-bond to the hydroxyl and carbonyl moiety of MCA-1 (building block A), respectively (Fig. 5e). Owing to its rigid scaffold with a single α-amino acid, albicidin adopts a V-shaped conformation with the Cα atom of L-Cya-3 at the center and the building blocks A-B-C and D-E-F forming the two arms that span

Fig. 5 Structure of the AlbAS–albicidin complex and key residues of AlbAS for albicidin binding. **a** AlbAS is shown in cartoon representation with the NTD (α1-α6) in blue and the CTD (α1′-α6′) in cyan. The elongated helix α6 connecting both domains is highlighted in orange. Albicidin is shown in ball-and-stick representation (yellow) and pervades both domains of AlbAS. **b** Superposition of the NTD and CTD of AlbAS. α-helical elements are labeled (a prime indicates assignment to the CTD). The protruding helix α1 of the NTD is predicted to be part of a longer α-helical linker region of AlbAL, which connects the HTH motif and the drug-binding region of the protein. **c** Superposition of the AlbAS–albicidin complex (ligand not shown) and TipAS (PDB 2MBZ)[27] bound to promothiocin A (ball-and-stick representation, carbon magenta). A region that is unfolded in isolated TipAS and forms an α-helical lid (α6-α9) after drug binding is colored in white. Helices present in the unbound form of TipAS are drawn in dark gray. **d** Interactions of the AlbAS CTD with albicidin portion D-E-F. **e** Interactions of the AlbAS NTD with albicidin portion A-B-C. Residues of the NTD, the CTD, and albicidin are colored in blue, cyan, and yellow, respectively, with oxygens in red and nitrogens in blue. Hydrogen bonds and ionic interactions are indicated by dashed lines

an angle of about 130°. Importantly, the side chain of N75 of AlbAS forms two hydrogen bonds with the amide proton and the carbonyl oxygen of the central L-Cya-3 (Φ and Ψ values of 66.0° and −128.7°), thereby stabilizing this specific conformer of albicidin. The cyanomethylene side chain of L-Cya-3 is oriented outwards the binding tunnel approaching helices α6 and α1′. Interestingly, the amide proton of pABA-4 is pointing at the triple bond of the nitrile moiety with a distance of about 1.4 Å.

Besides the mainly hydrophobic character of the interaction between AlbAS and albicidin, we propose that N75, R181, and Q205 are important residues for recognition of albicidin. Based on the crystal structure we generated the triple variant AlbAS$_M$ with amino-acid substitutions N75A, R181A, and Q205A (Supplementary Figure 4A). Employing the agar diffusion assay as a first line of evidence, AlbAS$_M$ was indeed not capable of efficiently suppressing the antibacterial activity of albicidin (Fig. 2c), albeit binding was not fully impeded (Supplementary Figure 14).

AlbAS tolerates structural variations of albicidin. The above-described interaction between AlbAS and albicidin raised further questions on the tolerance of AlbAS toward structural variations of albicidin and thus the possible design of albicidin analogs, which could escape arrest by AlbAS. Based on previous SAR studies, we were equipped with a number of synthetic albicidin derivatives and substructures thereof[8,9]. We employed the agar diffusion assay with *E. coli* and ^{1}H-^{15}N HSQC experiments in order to monitor the ability of AlbAS to bind albicidin analogs (Supplementary Figure 14 and 15).

From our library of albicidin-derived compounds we chose seven derivatives (compounds **1–7**) as well as the TipAS-specific ligand thiostrepton A as a control substance (Fig. 1b). These albicidin derivatives comprised variations of building blocks A and C, which otherwise stabilize the AlbAS–albicidin complex through hydrogen bonds to N24 and H78 (building block A), N75 (L-Cya-3, building block C), as well as π–π stacking and hydrophobic interactions (building block A) to W27 of AlbAS. In compound **1**, a fluorine atom replaced the phenolic hydroxyl group of the coumaric acid (building block A), thus eliminating hydrogen bond capabilities to N24. The agar diffusion assay showed that the presence of AlbAS abolished the antibacterial activity of compound **1** (Supplementary Figure 15). Moreover, the HSQC signal pattern clearly indicated strong binding by AlbAS similar to that of albicidin (Supplementary Figure 16).

Compound **2** bore two structural modifications: (i) it was shorter than albicidin due to a substitution of coumaric acid by a para-hydroxy benzoic acid moiety; and (ii) L-Cya-3 was substituted by the sterically more demanding L-Thr with H-bond donor and acceptor capabilities. Despite the expectedly missing hydrogen bond to residue N24 and the variation in the side chain of L-Cya-3 (building block C), the agar diffusion assay indeed confirmed binding of the strongly antibacterial compound **2** (Supplementary Figure 15). Accordingly, we observed the characteristic HSQC signal pattern as in the case of albicidin binding to AlbAS (Supplementary Figure 14 and 16). Apparently, both structural modifications (building blocks A and C) had no significant influence on binding by AlbAS.

In order to assess the importance of the only stereocenter in albicidin for binding to AlbAS, we also tested the antibacterially active enantiomer of albicidin[4] bearing D-Cya-3 (compound **3**). Remarkably, AlbAS was still able to bind compound **3** (Supplementary Figure 14) and thus to prohibit its antibacterial effects (Supplementary Figure 15).

Finally, we tested truncated albicidin derivatives (compounds **4**, **5**, and **6**) to evaluate binding contributions from the NTD and CTD (Fig. 1b). Compound **4** only consisted of the building blocks A and B, and therefore addressed binding contributions from the NTD of AlbAS. On the contrary, compound **5** comprised only the C-terminal building blocks D, E, and F, and was thought to potentially accommodate in the CTD of AlbAS. Ultimately, compound **6** was devoid of building blocks A and F and was thus hypothesized to miss essential interactions with residues N24, W27, W162, Y169, R181, and Q205 in the NTD and CTD. The truncated albicidin variants **4**, **5**, and **6** did not display antibacterial activity (Supplementary Figure 15), however, the change in the NMR signal pattern clearly demonstrated interaction between AlbAS and the respective derivatives (Supplementary Figure 14). Notably, while compounds **5** and **6** again caused spectra in the slow-exchange regime (strong binding), compound **4** showed a gradual chemical shift change typical for fast-exchange systems (weak binding, Fig. 4d). Fluorescence quenching experiments with AlbAS and compound **5** resulted in a K_d of 55.5 ± 3.6 nM (Fig. 4e). Compound **4** had no consistent effect on fluorescence emission spectra of AlbAS suggesting a very low binding affinity (Fig. 4d). Taken together, these results underline the importance of the CTD as a fixation point for albicidin.

Facing the high structural similarities between albicidin and the anti-Gram-negative compounds cystobactamid[31] and coralmy-cin[32], we tested binding capabilities of an albicidin derivative (compound **7**)[33] with cystobactamid-like features (Fig. 1b). The pattern of NMR reporter signals of AlbAS clearly pointed to a strong binding of compound **7** in the nanomolar K_d range. Our fluorescence quenching experiments revealed a K_d of 14.0 ± 2.9 nM for compound **7** (Fig. 4f), which is similar to the binding affinity of AlbAS toward albicidin (5.6 ± 0.2 nM), however, no cooperative behavior was observed. Analysis of the agar diffusion assay showed that AlbAS is able to neutralize the antibacterial effects of compound **7**, albeit with decreased efficiency (Supplementary Figure 15). Finally, thiostrepton A was devoid of any binding affinity toward AlbAS (Supplementary Figure 15 and Supplementary Figure 14).

Discussion

AlbA has been reported as a binding protein conferring resistance against the potent antibiotic albicidin[17,19,20]. Previous studies already mentioned a single high-affinity binding site for albicidin in AlbAS[20,34], and an α-helical structure[19], but they lacked both the knowledge of the chemical structure of albicidin and the high-resolution structure of AlbAS. Our antibacterial tests using an in vitro agar diffusion assay combined with biophysical binding studies via CD, NMR, and fluorescence spectroscopy underscore the high binding affinity of AlbAS and AlbAL toward albicidin (K_d of 5.6 ± 0.2 and 7.4 ± 0.9 nM, respectively) and other derivatives. The sensitivity of fluorescence quenching experiments was essential for the determination of such a low K_d value. However, as we were restricted to minimum protein concentrations of 20 nM, the determined K_d values could still represent upper boundaries. Therefore, the previously reported K_d of 64 nM[34] is at least one order of magnitude to high. Surprisingly, quenching of AlbAS emission by albicidin proved to be highly efficient (100% quenching) possibly reflecting the network of tryptophans in the binding tunnel, in particular in that of the CTD (Fig. 5c).

Competition experiments with the DNA gyrase construct GyrBA59 and albicidin (Supplementary Figure 9D and E) demonstrated the strong binding affinity of AlbAS and its ability to protect the molecular target from inhibition. In light of previously determined IC$_{50}$ values of 40 nM for DNA gyrase, we suggest that AlbA may either be more abundant in the cell than

DNA gyrase and/or exhibit superior association rates k_{on} to efficiently capture albicidin after cell entry.

The crystal structure of the drug-binding domain AlbAS revealed a tandem architecture with the NTD and CTD representing structural repeats of each other that are connected by the extended helix α6 of the NTD (Fig. 5a). Intriguingly, nature has adapted to the elongated and fairly rigid scaffold of albicidin by combining two TipAS-type folding units to fully enclose the highly potent toxin. Albicidin is thus clamped between the NTD and CTD and stabilized predominantly by hydrophobic and π–π stacking interactions (Fig. 5d, e, Supplementary Figure 11A and B). Based on our structural data, we can exclude the previously suggested direct involvement of residue H125 in albicidin binding[20], since this residue is located in helix α1′ and oriented toward the protein exterior (Supplementary Figure 17). Several hydrogen bonds, and in particular the salt bridge between the carboxylic acid of albicidin and R181 of AlbAS, most likely confer ligand selectivity. We assume that R181 is relevant to orient albicidin in the binding cleft by anchoring the carboxylic acid of pMBA-6 (building block F) in the CTD.

Attempts to crystallize AlbAS in the absence of albicidin did not succeed. According to this, the increase in observable signals in ^{1}H-^{15}N HSQC spectra of AlbAS upon albicidin binding suggests that free AlbAS displays inherent protein dynamics on the micro-to-millisecond timescale. Based on the repeat structure of AlbAS with few direct contacts between the repeat units even in the albicidin-bound state, we hypothesize that albicidin may fix the otherwise flexible relative orientation of the NTD and CTD. In addition, dynamics within the binding tunnel in the absence of albicidin may also account for some of these observations, as only three out of eight expected HSQC signals for Trp side chains were detected with good signal-to-noise ratio (Supplementary Figure 6 and 16). Ligand-induced stabilization has also been observed for the structurally related drug-binding protein TipAS[16], for which binding of thiopeptides induces an otherwise non-observable α-helical lid to pack onto the ligand (Fig. 5c)[27]. In contrast, however, the comparison of AlbAS/AlbAL and the AlbAS/AlbAL–albicidin complexes by CD spectroscopy showed that the α-helical content was virtually independent of ligand binding (Fig. 3a and Supplementary Figure 5B). The dynamic nature of the vacant protein suggests a possible binding mechanism: AlbAS exploits its inherent structural plasticity in order to clamp the elongated and rather rigid albicidin molecule in a buried tunnel of about 33 Å in length that is built from both repeat units (Supplementary Figure 18). The observed positive cooperativity with Hill coefficients n of 2.0–3.0 is in agreement with a binding process that includes recruitment of albicidin to one domain favoring subsequent association to the neighboring domain. Such a cooperative binding behavior has been also observed for the MerR protein BmrR[35]. The strong binding of albicidin consequently leads to a well-defined NTD-CTD arrangement in AlbAS causing an increase in thermal stability, ΔT_m, of about 15 °C.

Our concept of ligand-mediated domain-domain bridging is supported by binding effects caused by albicidin fragments: despite strong binding to AlbAS (K_d of 55.5 ± 3.6 nM), the C-terminal segment of albicidin (compound 5) did not cause the appearance of additional resonances in the HSQC spectrum of AlbAS, and neither did compounds 4 and 6 (Supplementary Figure 14). This observation demonstrates that the fragments were not capable of efficiently stabilizing the NTD-CTD arrangement (Supplementary Figure 18), while all tested full-length analogs of albicidin did (Supplementary Figure 14). Intriguingly, the full-length cystobactamid-type analog (compound 7) was not able to stabilize AlbAS to a similar extent as albicidin (Fig. 3d, g). Despite binding with low nanomolar affinity (K_d of 14.0 ± 2.9 nM) similar to albicidin, compound 7 showed no

positive cooperativity (Fig. 4f). Moreover, compound 7 provoked chemical shift patterns in HSQC spectra of AlbAS reminiscent of those of the C-terminal albicidin fragment (compound 5). We could not achieve crystallization of the AlbAS–compound 7 complex either. Taken together, these observations clearly support that cystobactamid is mainly bound to the CTD of AlbAS without properly bridging the NTD and CTD. As AlbAS can tolerate a shortened building block A (para-hydroxy benzoic acid in compound 2, see Supplementary Figure 14), the steric demand of β-methoxy-L-Asn and its anticipated effect on the backbone conformation in building block C appears to hinder the stabilization of both domains. In a previous study, we identified various naturally occurring derivatives of albicidin, amongst them a less active variant with a β-methoxy-L-Asn residue[33]. Hence, this modification may represent a natural mechanism against AlbA-mediated transcriptional regulation.

Truncated AlbAS variants composed of a single domain were not sufficient to properly arrest albicidin: AlbAS$_{NTD}$ and AlbAS$_{CTD}$ were not able to neutralize albicidin in the agar diffusion assay (Fig. 2b) due to their much weaker binding affinities. Fluorescence quenching experiments yielded K_d values for albicidin binding of 3.4 ± 0.2 μM (NTD) and 0.6 ± 0.1 μM (CTD), respectively, indicating that recruitment of albicidin to AlbAS may be guided by the CTD. We propose that a binding pathway from the CTD toward the NTD of AlbAS would ensure that albicidin is first fully trapped in AlbAS before the binding signal is relayed to the N-terminal HTH domain of full-length AlbAL to induce DNA binding. The high affinity of AlbAS toward the C-terminal fragment of albicidin (compound 5) further supports the idea of the CTD as an initial anchoring point (Supplementary Figure 18).

Interestingly, when comparing AlbAS with AlbBS only key residues W162, Y169, and R181 of AlbAS are conserved in AlbBS (Supplementary Figure 2B). These residues interact with pMBA-6 (building block F) of albicidin indicating that AlbBS might recognize this C-terminal pMBA-motif, but that its cognate ligand might otherwise deviate from the albicidin structure. In light of the reduced affinity of AlbAS toward the cystobactamid-type compound 7, the missing binding cooperativity and the reduced stabilization effect we propose that AlbAS has been evolutionarily optimized to capture albicidin rather than the related cystobactamids or coralmycins[31,32]. Whether AlbBS represents a corresponding cystobactamid- or coralmycin-binding protein will be the subject of future investigations.

Sequence alignments and secondary structure predictions demonstrate the widespread occurrence of AlbAS homologs with tandem architecture among several ESKAPE pathogens (Supplementary Figure 3). These homologs are sequentially different (≤58% difference to AlbAL) but are predicted to adopt a similar α-helical repeat structure in which key residues important for albicidin binding are conserved. The AlbA-mediated sequestration of albicidin might thus operate in clinically relevant pathogens, which showed resistance against albicidin (MIC value ≥64 μg mL^{-1}). However, at this point it is unclear to which extent albicidin capture by AlbA homologs contributes to the overall antibacterial resistance in vivo, as several resistance strategies will act in concert. Likewise, it will be necessary to identify full-length AlbAL and the truncated AlbAS species in vivo, and to quantify their cellular abundances.

Given the ligand promiscuity of the TipAS system[27,36], and other MerR effector-binding domains (i.e., BmrR)[37], we studied whether AlbAS can accept structurally different albicidin derivatives by using an in vitro agar diffusion assay and an HSQC-based binding assay. While the former assay only provided information on the AlbAS-mediated neutralization of antibacterial derivatives, the NMR experiments gave a direct read-out

on the binding capabilities of AlbAS. The synthetic albicidin derivatives submitted to these assays probed steric as well as electronic requirements for effective binding by AlbAS.

As discussed above, the albicidin fragments could not efficiently fix the two domains of AlbAS. This observation supports the suggested clamping mechanism, for which albicidin derivatives are required that can span the entire binding tunnel, thus connecting the NTD and CTD (Supplementary Figure 18).

Structural variations at the N terminus of albicidin (building block A), like the withdrawal of hydrogen bond capabilities (compound 1) or the replacement of methyl coumaric acid with para-hydroxy benzoic acid (compound 2) had no significant effect on AlbAS binding. In addition, a polar L-Thr instead of L-Cya-3 (compound 2) was similarly tolerated by AlbAS (Supplementary Figure 16). Remarkably, as shown for the albicidin enantiomer (compound 3), inversion of the only stereocenter of the molecule was not detrimental to recognition by AlbAS. The agar diffusion assay clearly demonstrated the neutralizing effect of AlbAS toward compound 3 and the characteristic signal patterns in the HSQC spectra confirmed a binding mode equivalent to that of albicidin (Supplementary Figure 16). AlbAS binding of the enantiomeric compound 3 might be explained based on the orientation of albicidin in the crystal structure: the side chain of L-Cya-3 points out of the binding pocket toward the protein surface (Fig. 5a). The surrounding helices α1′, α4, and α6 would leave sufficient space for D-Cya-3 while preserving the overall binding mode of albicidin. These findings show that the rather flexible AlbAS tolerates structural modifications of the oligo-aromatic scaffold of albicidin.

A connection between AlbA and transcriptional regulator proteins has previously been suggested by Weng et al.[19] and could be corroborated by our bioinformatics analysis (Supplementary Figure 1). The structural homology to TipAS (Fig. 5c) suggests that AlbAS is a member of the MerR-type transcriptional regulator family. The N-terminal HTH DNA-binding domain requires future studies on the mechanism of transcriptional regulation by AlbAL. This includes questions on how signal transduction within AlbAL is triggered by albicidin binding to its drug-binding domain. Our studies on the full-length protein AlbAL (Supplementary Figure 5) already indicated that the binding mode of albicidin in AlbAS is equivalent to that in AlbAL (Supplementary Figure 5A and F). A major difference between the two systems is the dimerization of AlbAL crucial for DNA binding, in contrast to the monomeric state of AlbAS (Supplementary Figure 7). In both cases, quaternary structure is albicidin-independent illustrating that the binding stimulus is relayed from the NTD toward the HTH motif to trigger promoter binding and in consequence the expression of, e.g., MDR transporters[21].

For the TipAL system it has been suggested that the (partially) unfolded N-terminal region in unbound TipAS becomes folded upon thiopeptide binding. Formation of this α-helical lid (α6-α9) stabilizes the ligand and forwards the signal to the HTH DNA-binding domain[27]. Owing to the different ligand-binding sites and the different structural organizations of the TipAS and AlbAS N termini (Fig. 5c), the mechanisms of intramolecular signal transduction might differ for these two systems and need to be further investigated. Importantly, the proposed binding pathway proceeding from the CTD toward the NTD and finally the HTH motif may ensure a cooperative coupling between the drug-binding stimulus and DNA binding only when the cognate ligand is fully arrested. To avoid AlbA-mediated resistance, two cases therefore need to be differentiated in drug development of albicidin, i.e., analogs that (i) are not prone to AlbAS capture at all or (ii) are still bound by AlbAS but do not trigger transcriptional regulation (Supplementary Figure 18). Future studies are envisaged to answer these questions.

In summary, our findings contribute to a general understanding of resistance strategies of bacteria against antibiotics. The knowledge on AlbA and homologous members of this widespread class of binding proteins will be generally useful for future antibiotic-development efforts. Depending on their pharmacophores, potent antibiotics could be designed such that they either completely bypass capture or that they do not trigger regulative effects on a cellular level despite being partially arrested by AlbA-like drug-binding proteins.

Methods

Bioinformatics. The BLAST[38] was used for a protein sequence homology search and preparation of a phylogenetic tree[39]. Secondary structure predictions were performed with PSIPRED and PROMALS3D[40,41].

Cloning. A codon-optimized synthetic *albAS* (Supplementary Table 2) coding region (Thermo Scientific GmbH, Schwerte, Germany) was cloned into an Amp^R-pMAT vector system, using *Nde*I and *Not*I restriction enzyme sites. The plasmid was amplified in *E. coli* BL21 DH5α, extracted, double digested with *Nde*I and *Not*I, and the released fragment was cloned into expression vector pET-28a (+) for production of AlbAS bearing an N-terminal, TEV-cleavable His$_6$-tag. Codon-optimized synthetic *albAL* (Supplementary Table 2, Thermo Scientific GmbH) was cloned by Gibson cloning[42] into an expression vector pET-28a (+) for production of AlbAL bearing an N-terminal, TEV-cleavable His$_6$-tag.

The amino-acid substitutions (N75A, R181A, and Q205A) in the triple variant AlbAS$_M$ as well as the substitutions for the Trp-to-Phe variants of AlbAS (W5F, W27F, W56F, W110F, W133F, W162F, W195F, and W203F) were inserted by Gibson assembly cloning[42]. To this end, one primer pair (Supplementary Table 3) was designed to carry the point mutation in the respective base triplet and to amplify one-half of the gene. The other primer pair amplifies the remaining vector and possesses a 16 bp overhang matching to the 3′-end of the other linear construct. Both constructs were ligated with *Taq* DNA ligase (New England Biolabs GmbH, Frankfurt am Main, Germany) after purification with gel electrophoresis by incubation in the Gibson mixture. The mutated plasmid was then transformed into *E. coli* DH5α cells by heat shock. Transformants were selected on Luria Bertani (LB)-agar plates supplemented with 50 μg mL^{-1} kanamycin and confirmed later by sequencing of the gene.

The AlbAS$_{NTD}$ and AlbAS$_{CTD}$ constructs were cloned using the Gibson assembly method[42] with the primer pairs listed in Supplementary Table 3. The boundaries for the truncated AlbAS variants (AlbAS$_{NTD}$ M1-E132 and AlbAS$_{CTD}$ D128-E221) were based on bioinformatic analysis before solving the AlbAS crystal structure.

Protein synthesis and purification of AlbAS and AlbAL. All genes including *albAS*, *albAL*, *albAS$_M$*, *albAS$_{NTD}$*, *albAS$_{CTD}$*, and the Trp-to-Phe mutants were expressed in *E. coli* BL21-Gold (DE3) strain (Merck Millipore, Darmstadt, Germany) with the T7 promoter expression system pET-28a (+) bearing an N-terminal TEV-cleavable His$_6$-tag in LB medium. The expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Biomol, Hamburg, Germany) and the cells were further incubated at 18 °C, 180 rpm for 18 h. For isotope labeling, cells were grown in M9-minimal medium containing ^{15}N ammonium chloride (1 g L^{-1}), ^{13}C$_6$ D-glucose (3 g L^{-1}), and/or deuterium oxide. Cells were harvested by centrifugation and the resuspended cell pellet (50 mM Tris-HCl, 500 mM NaCl, and 5% glycerin, pH 7.5) was lysed by one passage through a high-pressure cell disrupter TS 0.75 KW (Constant Systems Limited, Königswinter, Germany) at 18 000 psi and incubated with DNAse (3 U per mL lysate). Subsequently, cell lysates were centrifuged at 55 000 × *g* for 30 min to separate the cell debris from soluble protein. The crude supernatant was filtered through a 45 μm filter membrane before loading on the purification column. AlbAS/AlbAL purification was performed by affinity chromatography on an ÄKTApurifier system (GE Healthcare, Munich, Germany) by making use of the N-terminal His$_6$-tag binding to a Nickel His-Trap Fast Flow Crude 1 mL column (GE Healthcare). The washing buffer consisted of 50 mM Tris-HCl and 300 mM NaCl at pH 7.5. Then AlbAS/AlbAL was eluted with a buffer containing 50 mM Tris-HCl, 300 mM NaCl, and 500 mM imidazole. To remove imidazole the protein fractions were diluted with the washing buffer and concentrated in centrifugal filter units (Merck, Darmstadt, Germany) with a cutoff of 10 000 nominal molecular weight limit. Cleavage of the His-tag was performed for 2 h at room temperature and 16 h at 4 °C with the addition of 100 μM of purified TEV protease. The separation of the protein from the cleaved His-tag was performed by a second purification step employing the Nickel His-Trap column as described before. Desalting and buffer exchange was accomplished by size-exclusion chromatography on a Superdex High Load 16/60 75 column (GE Healthcare) in the final buffer used for NMR studies (sodium hydrogen phosphate 50 mM and 100 mM NaCl, pH 6.8). The collected fractions were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and the protein concentration was determined on a nano-photometer P330

(Implen, Munich, Germany). For long-time storage, the aliquots (0.5 mL) of protein samples were shock-frozen in liquid nitrogen and stored at −80 °C.

Selenomethionine-labeled protein was produced in 500 mL M9-minimal medium with addition of 25 mg selenomethionine, 50 mg lysine, 50 mg threonine, 50 mg phenylalanine, 25 mg leucine, 25 mg isoleucine, and 25 mg valine shortly bevor induction with 0.2 mM IPTG (at OD$_{600}$ of 0.7–0.8). Cells were incubated at 18 °C, 180 rpm for 20 h[43]. Purification was carried out as described above for wild-type AlbAS. The final buffer contained 50 mM Tris (pH 7.5), 100 mM NaCl, and 4 mM dithiothreitol (DTT).

Preparation of GyrBA59. The *E. coli* fusion protein GyrBA59, combining the full-length GyrB and a C-terminally truncated GyrA segment (59 kDa), was cloned into pET-28a (+) by restriction/ligation (forward and reverse primers are listed in Supplementary Table 3) using the restriction enzyme sites *Nde*I and *Xho*I. GyrBA59 was expressed in *E. coli* BL21 DE3 over 18 h at 18 °C and 250 rpm using auto induction medium. The cells were harvested and resuspended in buffer 1A (20 mM Tris, 150 mM NaCl, 10 mM imidazole, 10% glycerol, 2 mM DTT, EDTA-free protease inhibitor cocktail (cOmplete, Roche), and DNase 2 U per mL lysate, pH 8.0). After cell lysis with 14 000 psi using a high-pressure cell disrupter TS 0.75 KW (Constant Systems Limited, Königswinter, Germany) the lysate was centrifuged at 50 000 × g at 4 °C for 30 min and subsequently incubated with DNAse (2 U per mL lysate) for another 30 min at 4 °C. The protein was purified in four steps as follows: (1) cell lysate was loaded onto a His-Trap HP column and washed back to baseline in buffer 1B (20 mM Tris, 500 mM NaCl, 10 mM imidazole, 10% glycerol, 2 mM DTT, and EDTA-free protease inhibitor cocktail, pH 8.0) before being eluted with a gradient from buffer 1B into buffer 1C (20 mM Tris, 300 mM NaCl, 300 mM imidazole, 10% glycerol, 2 mM DTT, and EDTA-free protease inhibitor cocktail, pH 8.0) over 15 column volumes. (2) Pooled fractions were equilibrated against buffer 1A and digested with TEV protease (1 mg per 10 mg of protein) overnight at 4 °C to cleave the N-terminal 6x His-tag. The mixture was loaded onto a His-Trap HP column and the tag cleavage protein was recovered from the flow through in buffer 2A (20 mM Tris, 10 mM NaCl, 5% glycerol, 2 mM DTT, and EDTA-free protease inhibitor cocktail, pH 8.0). (3) The protein fractions were equilibrated against buffer 2A and loaded onto a Heparin Sepharose column to be eluted with a gradient from buffer 2A into buffer 2B (20 mM Tris, 1 M NaCl, 10% glycerol, 2 mM DTT, and EDTA-free protease inhibitor cocktail, pH 8.0) over 15 column volumes. (4) Fractions from step 3 were concentrated using a 100 kDa cutoff filter unit, before being loaded onto a Superdex 200 HiLoad 16/60 column (GE Healthcare) equilibrated and eluted in buffer 3A (20 mM Tris, 100 mM KCl, 1 mM DTT, and 10% glycerol, pH 8.0). The collected fractions were analyzed by SDS-PAGE and the protein concentration was determined on a nano-photometer P330 (Implen, Munich, Germany). Post size-exclusion fractions were pooled, concentrated, and stored at −80 °C.

DNA cleavage assay. *E. coli* DNA gyrase cleavage assays were performed by incubating 250 ng of supercoiled pBR322 plasmid DNA (Inspiralis, Norwich, UK) with 25 nM of GyrBA59 protein in 20 μL reactions containing 35 mM Tris-HCl (pH 8.0), 24 mM KCl, 4 mM MgCl$_2$, 2 mM DTT, 1.8 mM spermidine, 1.4 mM ATP, 6.5% (w/v) glycerol, and 0.1 mg mL^{-1} bovine serum albumin in the presence or absence of various concentrations of albicidin (final concentration of 5% dimethyl sulfoxide (DMSO), incubation at 37 °C for 30 min). Linearized DNA was released by further incubation with 0.2% (w/v) SDS and 0.1 mg mL^{-1} proteinase K (Promega, Madison, USA) at 37 °C for 30 min. The reaction was stopped by addition of 20 μL of reaction stop buffer (STEB buffer) (40% (w/v) sucrose, 100 mM Tris-HCl (pH 8.0), 10 mM EDTA, and 0.5 mg mL^{-1} Bromophenol Blue) and precipitation using 30 μL of chloroform/isoamyl alcohol 24:1 (v:v). Samples were analyzed by electrophoresis in 1% w/v agarose gel containing 0.5 mg mL^{-1} ethidium bromide followed by photography under ultraviolet illumination.

MIC assay. The MIC assay was performed by Antiinfectives Intelligence GmbH (Rheinbach, Germany) using internationally standardized protocols and employing clinical isolates of *K. oxytoca*, *K. pneumoniae*, *R. ornithologica*, *P. aeruginosa*, *A. baumannii* and *E. cloacae*.

Chemical synthesis. Synthesis protocols and analytical data have been recently described for albicidin[9], compound **1**[8], compounds **3**, **5**, and **6**[14] as well as compounds **4** and **7**[33]. The coupling of the N-terminal building block and the global deprotection of compound **2** are described in the Supplementary Methods part 1. Thiostrepton A was purchased from Merck (Darmstadt, Germany).

Agar diffusion assay. To verify the neutralizing ability of AlbAS, AlbAL, AlbAS$_M$, AlbAS$_{CTD}$, or AlbAS$_{NTD}$ against different compounds, agar diffusion assays with the respective compounds were performed. An overnight culture of *E. coli* DH5α was diluted to an OD$_{600}$ of 0.05 with 0.75% LB-agar. The protein-albicidin reaction mixture consisted of a 1:1 mixture of 40 μM albicidin (or derivatives) with 40 μM protein (either AlbAS or its variants AlbAL, AlbAS$_M$, AlbAS$_{NTD}$, and AlbAS$_{CTD}$). The mixture was incubated for 20 min at room temperature in the dark. Reaction mixtures of 30 μL were transferred in triplicates of 2 mm holes on the inoculated

agar plate. The plates were evaluated concerning growth-inhibition zone after incubation for 18 h at 37 °C.

CD spectroscopy. For determination of the melting temperature and information on protein-ligand stability, CD measurements were performed on a J-815 CD spectrometer (Jasco, Groß-Umstadt, Germany). Protein samples (2.0 mM for AlbAS and 0.2 mM for AlbAL) were saturated with 1.5-fold albicidin and diluted to a protein concentration of 20 μM in order to ensure a final DMSO content of only 0.05% for AlbAS and 0.5% for AlbAL. A volume of 250 μL of the reaction mixture was transferred into a quartz cuvette with 1 mm path length. Six far-UV spectra were accumulated in a wavelength range of 195–300 nm (203–300 nm for AlbAL) at three different temperatures (10, 21, and 90 °C) with a scanning speed of 50 nm s^{-1}. For evaluation of the spectra the measured buffer blank (50 mM sodium hydrogen phosphate and 100 mM NaCl, pH 6.8) was subtracted from the averaged spectra. The temperature curves were recorded from 10–90 °C in 0.5 °C min^{-1} steps while monitoring the ellipticity at 215 nm (218 nm for ALbAL). The resulting values were recalculated into molar ellipticity [Θ]$_{MRW}$ with consideration of concentration, path length, number of amino acids, and molecular weight of the sample. For visualization purposes, the temperature curves of AlbAS–albicidin and AlbAS–compound **5** in Fig. 3b, c were scaled to that of AlbAS using scaling factors of 1.175 and 1.091, respectively. The melting temperatures of AlbAS, AlbAS–albicidin, AlbAS–compound **5**, AlbAS–compound **7**, AlbAL, and AlbAL–albicidin were determined by nonlinear regression of the measured data using Eq. (1). In case of AlbAS–albicidin a biphasic unfolding curve was observed, which we ascribed to the two-domain arrangement. The T_m was thus determined by fitting the second transition representing the stabilized domain/species. In all other cases, the protein unfolding curves appeared as monophasic.

$$\Theta(T) = \frac{(\Theta_N + m_N T) + (\Theta_U + m_U T)e^{-\frac{\Delta H_m^{v,H}}{RT}\left(\frac{T - T_m}{T_m}\right)}}{1 + e^{-\frac{\Delta H_m^{v,H}}{RT}\left(\frac{T - T_m}{T_m}\right)}} \qquad (1)$$

with Θ being the measured ellipticity, $\Delta H_m^{v,H}$ denoting the van't Hoff enthalpy at the transition point, Θ_N and m_N as well as Θ_U and m_U defining the pre-transition and post-transition phase.

NMR spectroscopy. For binding studies with albicidin and its derivatives, each ligand was dissolved in 100% d6-DMSO and added to ^{15}N-labeled AlbAS or ^{2}H, ^{15}N-labeled AlbAL in 50 mM sodium hydrogen phosphate (pH 6.8), 150 mM NaCl buffer (usually 0.24 mM AlbAS or AlbAL in 85% buffer, 5% d6-DMSO, and 10% D$_2$O) yielding a 1.5-fold excess of the ligand. After incubation for 20 min at room temperature in the dark, the sample was centrifuged at 20 000 × g for 3 min and the supernatant was transferred to a 5 mm NMR tube. The AlbAS blank sample (Supplementary Figure 14) was measured in 100% buffer without addition of ligands and DMSO. Spectra were recorded on a Bruker Avance III 700 MHz NMR spectrometer (Bruker BioSpin, Rheinstetten, Germany) with a TXI probe head. Standard Bruker pulse sequences were used. 2D ^{1}H-^{15}N HSQC spectra were recorded at 298 K with acquisition times of 90 and 22 ms for the ^{1}H and ^{15}N dimension, respectively. The experiments were accumulated with ≥16 scans depending on protein sample concentration. 2D ^{1}H-^{15}N TROSY spectra were accumulated with 256 scans and acquisition times of 90 and 22 ms for the ^{1}H and ^{15}N dimension, respectively. Several AlbAS–albicidin samples were measured with the same result and a representative spectrum is given for the figures in the main manuscript and the SI. NMR experiments with AlbAS and albicidin analogs were conducted once with accumulation of ≥16 scans depending on protein concentration. For alignment of HSQC and TROSY spectra, the TROSY spectrum was corrected by 1/2 $^1J_{NH}$ in both dimensions.

AlbAS–albicidin titration experiments were performed by means of ^{1}H-^{15}N HSQC experiments. Samples were prepared similarly as in the method described above: spectra were recorded at 303 K for AlbAS in buffer (50 mM sodium hydrogen phosphate and 150 mM NaCl, pH 6.8) at different protein-albicidin ratios. Albicidin was added to a 0.8 mM ^{15}N-labeled AlbAS sample to the following final protein-albicidin ratios: 0.1, 0.2, 0.5, 0.8, and 1. During each titration step, the DMSO concentration was increased and reached a final concentration of 7.3% (v/v). The experiment was conducted once with accumulation of 16 scans for each titration step. The spectra were recorded with acquisition times of 90 and 20 ms in the direct ^{1}H and indirect ^{15}N dimension, respectively.

For the competition experiment with the DNA gyrase construct GyrBA59, 1.7 mM ATP was incubated with 3.4 mM MgCl$_2$ for 10 min. Subsequently, 30 μM GyrBA59 and 3 μg supercoiled pBR322 plasmid DNA (Inspiralis) were added and incubated for another 10 min. Finally, 30 μM ^{15}N-labeled AlbAS were added to the mixture followed by 30 μM albicidin (10% D$_2$O and 5% DMSO final concentration). Owing to the low protein concentrations (to achieve a 1:1 ratio of gyrase and AlbAS), a ^{1}H-^{15}N band-selective optimized flip-angle short-transient heteronuclear multiple quantum coherence (SOFAST HMQC)[44] experiment was performed. The ^{1}H-^{15}N SOFAST HMQC was recorded with 2048 scans and with acquisition times of 45 and 38 ms for the ^{1}H and ^{15}N dimensions, respectively. Measurements were performed at 303.2 K. The control experiment was performed under the same conditions but without the addition of MgCl$_2$, ATP, pBR322 plasmid DNA, and GyrBA59.

For preliminary resonance assignments, three-dimensional HNCA and HNCO experiments were performed employing ^{13}C, ^{15}N-AlbAS (0.8 mM) saturated with albicidin. All NMR data were processed and analyzed with Bruker TopSpin 3.1 Software and SPARKY[45].

Size-exclusion chromatography with multi-angle light scattering. SEC-MALS experiments were performed at 18 °C in buffer containing 50 mM Tris (pH 7.5), 100 mM NaCl, and 0.02% NaN$_3$. A sample of 225 μg of AlbAS or AlbAL (with and without saturating amounts of albicidin) were loaded onto Superdex 75 or Superdex 200 increase 10/300 columns (GE Healthcare) that were coupled to a miniDAWN TREOS three-angle light-scattering detector (Wyatt Technology) in combination with a RefractoMax520 refractive index detector. For calculation of the molecular mass, protein concentrations were determined from the differential refractive index with a specific refractive index increment (dn/dc) of 0.185 mL g^{-1}. Data were analyzed with the ASTRA 6.1.4.25 software (Wyatt Technology).

Fluorescence assay. Fluorescence quenching measurements were performed by monitoring tryptophan fluorescence of AlbAS (20 nM), AlbAL (20 nM), AlbAS$_{NTD}$, and AlbAS$_{CTD}$ (100 nM) on a FluoroMax 2 spectrometer from Horiba (Potsdam, Germany). The excitation wavelength was set to $\lambda_{exc} = 280$ nm. Emission spectra were recorded from $\lambda_{em} = 290$–450 nm with a scanning speed of 1 nm s^{-1} and an integration time of 1 s. The excitation and emission slit widths were set to 8 and 4 nm, respectively. Proteins were solved in 50 mM sodium phosphate buffer with 150 mM NaCl and 0.01% DMSO at pH 6.8. All measurements were performed twice and the standard deviation is given for the K_d and the Hill factor n. Data evaluation was performed by integration of the emission band (318–410 nm for AlbAS, 316–412 nm for AlbAL, and 319–390 nm for AlbAS$_{NTD}$ and AlbAS$_{CTD}$) for every titration step. The area under the curve (AUC) of the buffer blank (buffer with 0.01% DMSO) was subtracted from the measured titration data. The quenching factor was determined by subtraction of AUCs for each titration step from the AUC of the protein in the absence of ligand. For determination of the AlbAS– and AlbAL–albicidin K_d values, the data were fitted to Eq. (2).

$$Y = B_{max} \frac{[L]^n}{K_d^n + [L]^n} \qquad (2)$$

where $[L]$, n, and B_{max} denote the total ligand concentration, the Hill coefficient, and the maximum binding capacity of the protein, respectively. For the determination of the K_d for AlbAS$_{NTD}$–albicidin and AlbAS$_{CTD}$–albicidin as well as for AlbAS–compound **5**, the Hill coefficient was constrained to $n = 1$ for the nonlinear regression as the binding curve showed no cooperativity.

For the determination of the K_d for AlbAS–compound **7** the data were fitted to Eq. (3).

$$Y = \frac{([P_t] + [L] + K_d) - \sqrt{([P_t] + [L] + K_d)^2 - 4[P_t][L]}}{2} \qquad (3)$$

where P_t is the total protein concentration.

Crystallization. Crystals of unlabeled and selenomethionine-labeled AlbAS (1.4 mM AlbAS saturated with 2.5-fold albicidin in 50 mM Tris (pH 7.5), 100 mM NaCl, 4 mM DTT, and 5% DMSO) were obtained by the sitting-drop vapor-diffusion method at 18 °C with a reservoir solution composed of 0.1 M Tris/HCl (pH 8.5) and 2.0 M ammonium sulfate. For crystallization 96-well MRC plates were used. The volume of the reservoir solution was 80 μL and the crystallization drop contained 0.5 μL reservoir solution and 0.5 μL protein solution. Crystals had dimensions of 200 μm × 150 μm × 50 μm and appeared within 3 days. Crystals were cryo-protected with a solution composed of equal volumes of mother liquor and 3.4 M sodium malonate at pH 7.0 and subsequently flash-cooled in liquid nitrogen.

Structure determination and refinement. Synchrotron diffraction data were collected at the beamline 14.1 and 14.2 of the MX Joint Berlin laboratory at BESSY (Berlin, Germany) at a wavelength of 0.97949 Å and temperature of 100 K. Diffraction data were processed with XDS[46]. Experimental phases were determined by single anomalous dispersion with the AUTOSOL routine in PHENIX[47] using PHASER[48] and SOLVE/RESOLVE[49] exploiting the selenomethionine-labeled AlbAS (Supplementary Table 1). There were 35 heavy-atom sites in the asymmetric unit (FOM: 0.413, BAYES-CC: 60.3). An initial model of AlbAS was built with the program AUTOSOL in PHENIX[47]. All dihedral angles were found in favored (99.8%) and allowed (0.2%) regions of the Ramachandran plot. Our structure contains no Ramachandran plot outliers. The structure was refined by maximum-likelihood restrained refinement using in PHENIX, including TLS refinement[50]. Model adjustment and water picking was performed with COOT[51]. Geometrical restraints used in the refinement of albicidin were generated by using the Grade Web Server (http://grade.globalphasing.org) that queries the Cambridge Structural Databank with optional quantum chemical regularization. Model quality was evaluated with MolProbity[52] and the JCSG validation server[53]. Secondary structure elements were assigned with DSSP[54], and ALSCRIPT[55] was used for secondary

structure-based sequence alignments. Figures were prepared using PyMOL[56] and channels were calculated with CAVER[57].

References

1. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta* **1794**, 808–816 (2009).
2. Royer, M. et al. Albicidin pathotoxin produced by Xanthomonas albilineans is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes. *Mol. Plant Microbe Interact.* **17**, 414–427 (2004).
3. Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A. & Birch, R. G. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* **51**, 181–187 (2007).
4. Kretz, J. et al. Total synthesis of albicidin: a lead structure from Xanthomonas albilineans for potent antibacterial gyrase inhibitors. *Angew. Chem. Int. Ed. Engl.* **54**, 1969–1973 (2015).
5. Cociancich, S. et al. The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol.* **11**, 195–197 (2015).
6. Süssmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis-principles and prospects. *Angew. Chem. Int. Ed. Engl.* **56**, 3770–3821 (2017).
7. Vivien, E. et al. Heterologous production of albicidin: a promising approach to overproducing and characterizing this potent inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* **51**, 1549–1552 (2007).
8. Kerwat, D. et al. Synthesis of albicidin derivatives: assessing the role of N-terminal acylation on the antibacterial activity. *ChemMedChem* **11**, 1899–1903 (2016).
9. Grätz, S. et al. Synthesis and antimicrobial activity of albicidin derivatives with variations of the central cyanoalanine building block. *ChemMedChem* **11**, 1499–1502 (2016).
10. Süssmuth, R. et al. Albicidin derivatives, their use and synthesis. Patent WO2014125075 A1, PCT/EP2014/052922 (2014).
11. Birch, R. G., Pemberton, J. M. & Basnayake, W. V. Stable albicidin resistance in Escherichia-Coli involves an altered outer-membrane nucleoside uptake system. *J. Gen. Microbiol.* **136**, 51–58 (1990).
12. Fsihi, H., Kottwitz, B. & Bremer, E. Single amino-acid substitutions affecting the substrate-specificity of the Escherichia-Coli K-12 nucleoside-specific Tsx channel. *J. Biol. Chem.* **268**, 17495–17503 (1993).
13. Zhang, L. & Birch, R. G. The gene for albicidin detoxification from Pantoea dispersa encodes an esterase and attenuates pathogenicity of Xanthomonas albilineans to sugarcane. *Proc. Natl Acad. Sci. USA* **94**, 9984–9989 (1997).
14. Vieweg, L. et al. The albicidin resistance factor AlbD is a serine endopeptidase that hydrolyzes unusual oligoaromatic-type peptides. *J. Am. Chem. Soc.* **137**, 7608–7611 (2015).
15. Ramos, J. L. et al. The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **69**, 326–356 (2005).
16. Kahmann, J. D. et al. Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J.* **22**, 1824–1834 (2003).
17. Walker, M. J., Birch, R. G. & Pemberton, J. M. Cloning and characterization of an albicidin resistance gene from Klebsiella-oxytoca. *Mol. Microbiol.* **2**, 443–454 (1988).
18. Basnayake, W. V. S. & Birch, R. G. A gene from Alcaligenes-denitrificans that confers albicidin resistance by reversible antibiotic binding. *Microbiology* **141**, 551–560 (1995).
19. Weng, L. X. et al. Molecular and conformational basis of a specific and high-affinity interaction between AlbA and albicidin phytotoxin. *Appl. Environ. Microbiol.* **71**, 1445–1452 (2005).
20. Weng, L. X., Xu, J. L., Li, O., Birch, R. G. & Zhang, L. H. Identification of the essential histidine residue for high-affinity binding of AlbA protein to albicidin antibiotics. *Microbiology* **149**, 451–457 (2003).
21. Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* **27**, 145–163 (2003).
22. Newberry, K. J. & Brennan, R. G. The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus. *J. Biol. Chem.* **279**, 20356–20362 (2004).
23. Newberry, K. J. et al. Structures of BmrR-drug complexes reveal a rigid multidrug binding pocket and transcription activation through tyrosine expulsion. *J. Biol. Chem.* **283**, 26795–26804 (2008).

24. Watanabe, S., Kita, A., Kobayashi, K. & Miki, K. Crystal structure of the [2Fe-2S] oxidative-stress sensor SoxR bound to DNA. *Proc. Natl Acad. Sci. USA* **105**, 4121–4126 (2008).

25. Holmes, D. J., Caso, J. L. & Thompson, C. J. Autogenous transcriptional activation of a thiostrepton-induced gene in Streptomyces-lividans. *EMBO J.* **12**, 3183–3191 (1993).

26. Murakami, T., Holt, T. G. & Thompson, C. J. Thiostrepton-induced gene-expression in Streptomyces-lividans. *J. Bacteriol.* **171**, 1459–1466 (1989).

27. Habazettl, J. et al. Structural basis and dynamics of multidrug recognition in a minimal bacterial multidrug resistance system. *Proc. Natl Acad. Sci. USA* **111**, E5498–E5507 (2014).

28. Kumaraswami, M., Newberry, K. J. & Brennan, R. G. Conformational plasticity of the coiled-coil domain of BmrR is required for bmr operator binding: the structure of unliganded BmrR. *J. Mol. Biol.* **398**, 264–275 (2010).

29. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).

30. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).

31. Baumann, S. et al. Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed. Engl.* **53**, 14605–14609 (2014).

32. Kim, Y. J. et al. Isolation of coralmycins A and B, potent anti-gram negative compounds from the myxobacteria Corallococcus coralloides M23. *J. Nat. Prod.* **79**, 2223–2228 (2016).

33. von Eckardstein, L. Total synthesis and biological assessment of novel albicidins discovered by mass spectrometric networking. *Chemistry* **23**, 15316–15321 (2017).

34. Zhang, L. H., Xu, J. L. & Birch, R. G. High affinity binding of albicidin phytotoxins by the AlbA protein from Klebsiella oxytoca. *Microbiology* **144**, 555–559 (1998).

35. Bachas, S., Kohrs, B. & Wade, H. Unconventional coupling between ligand recognition and allosteric control in the multidrug resistance gene regulator, BmrR. *ChemMedChem* **12**, 426–430 (2017).

36. Chiu, M. L. et al. Broad spectrum thiopeptide recognition specificity of the Streptomyces lividans TipAL protein and its role in regulating gene expression. *J. Biol. Chem.* **274**, 20578–20586 (1999).

37. Bachas, S., Eginton, C., Gunio, D. & Wade, H. Structural contributions to multidrug recognition in the multidrug resistance (MDR) gene regulator, BmrR. *Proc. Natl Acad. Sci. USA* **108**, 11046–11051 (2011).

38. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

39. Dereeper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).

40. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

41. Pei, J. M., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).

42. Gibson, D. G. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–341 (2009).

43. Vanduyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin Fkbp-12 complexes with Fk506 and Rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).

44. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014–8015 (2005).

45. Goddard, T. D. & Kneller, D. G. SPARKY 3 (Univ. of California, San Francisco, 2001).

46. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

47. Terwilliger, T. C. et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 582–601 (2009).

48. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

49. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).

50. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 122–133 (2001).

51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

52. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

53. Yang, H. et al. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1833–1839 (2004).

54. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

55. Barton, G. J. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40 (1993).

56. DeLano, W. *The PyMOL Molecular Graphics System* http://www.pymol.org (2002).

57. Petrek, M., Kosinova, P., Koca, J. & Otyepka, M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* **15**, 1357–1363 (2007).

Acknowledgements

Author contributions

L.R., A.M., R.D.S., M.C.W., B.L., D.P., F.-J.S., and T.F. designed experiments. L.R., M.K., and D.P. performed molecular cloning, expression, and protein purification. L.R. performed NMR and CD spectroscopy experiments. F.J.S. and L.R. carried out fluorescence spectroscopic measurements. R.D., C.A., and B.L. performed crystallization, diffraction data collection, and structure elucidation. L.R. and M.K. performed gyrase competition experiments. L.v.E., D.K., and S.G. synthesized albicidin derivatives. L.R., A.M., and R.D.S. wrote the paper.

Additional information

Supplementary information

for

Molecular insights into antibiotic resistance - how a binding protein traps albicidin

by
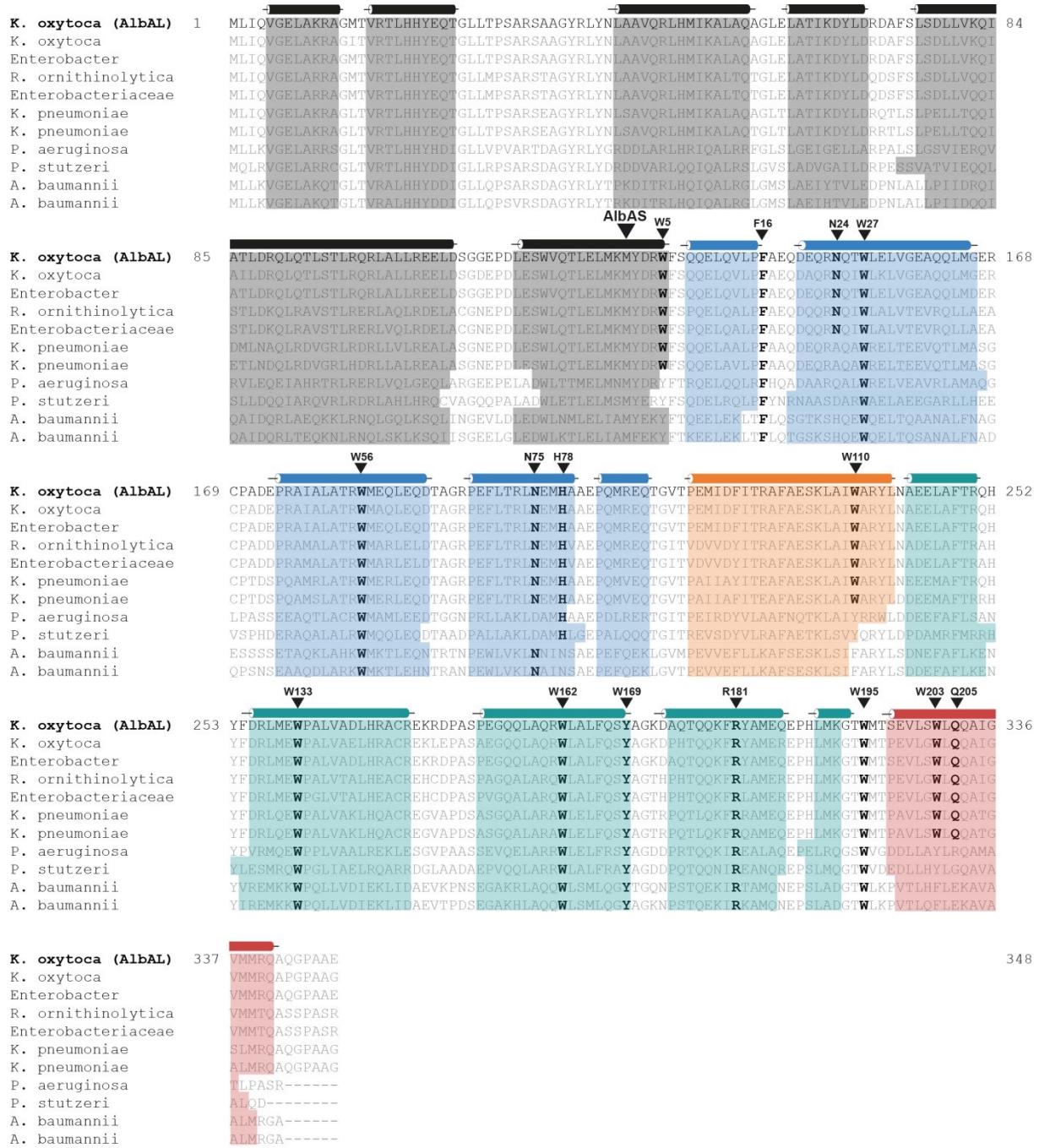
Rostock and Driller et al.

```
             1                                                                                    84
AlbAL    1   MLIQVGELAKRAGMTVRTLHHYEQTGLLTPSARSAAGYRLYNLAAVQRLHMIKALAQAGLELATIKDYLDRDAFSLSDLLVKQI   84
TipAL    1   MSYSVGQVAGFAGVTVRTLHHYDDIGLLVPSERSHAGHRRYSDADLDRLQQILFYRELGFPLDEVAALLDDPAADPRAHLRRQH   84

                                                                      AlbAS NTD
AlbAL   85   ATLDRQLQTLSTLRQRLALLREELDSGGEPDLESWVQTLELM------KMYDRWFSQQELQVLPFAE-------QDEQRNQTWL  155
TipAL   85   ELLSARIGKLQKMAAAVEQAMEARSMGINLTPEEKFEVFGDFDPDQYEEEVRERWGNTDAYRQSKEKTASYTKEDWQRIQDEAD  168
                                     TipAS #1

AlbAL  156   ELVGEAQQLMGERCPADEPRAIALATRWMEQLEQDT-AGRPEFLTRLNEMHAAEPQMREQTG-VTPEMIDFITRAFAESKL---  234
TipAL  169   ELTRRFVALMDAGEPADSEGAMDAAEDHRQGIARNHYDCGYEMHTCLGEMYVSDERFTRNIDAAKPGLAAYMRDAILANAVRHT  252

                                         AlbAS CTD
AlbAL  235   ------------AIWARYLNAE---ELAFTRQHYF--------------------DRLMEWPALVADLHRACREKRDPASPEGQQL  285
TipAL  253   PMGINLTPEEKFEVFGDFDPDQYEEEVRERWGNTDAYRQSKEKTASYTKEDWQRIQDEADELTRRFVALMDAGEPADSEGAMDA  336
                 TipAS #2

AlbAL  286   AQRWLALFQSYA-GKDAQTQQKFRYAMEQEPHLMKGT-WMTSEVLSWLQQAIGVM-MRQAQGPAAE             348
TipAL  337   AEDHRQGIARNHYDCGYEMHTCLGEMYVSDERFTRNIDAAKPGLAAYMRDAILANAVRHTP-----             397
```

Supplementary Figure 1 Sequence alignment of AlbAL with TipAL The TipAL sequence was artificially duplicated (red and orange bold letters) in order to match the repeat of drug-binding domains in AlbAS (black, bold letters). Regions corresponding to the N-terminal HTH DNA-binding domains are shown in plain letters. Start positions of AlbAS-NTD and AlbAS-CTD are indicated by arrows above and below the alignment, respectively. α-helices are indicated by cylinders above and below the respective amino acid sequence (black cylinders: HTH domain, blue: AlbAS NTD, orange: α6, cyan: AlbAS CTD, red: α6'and grey: TipAS).
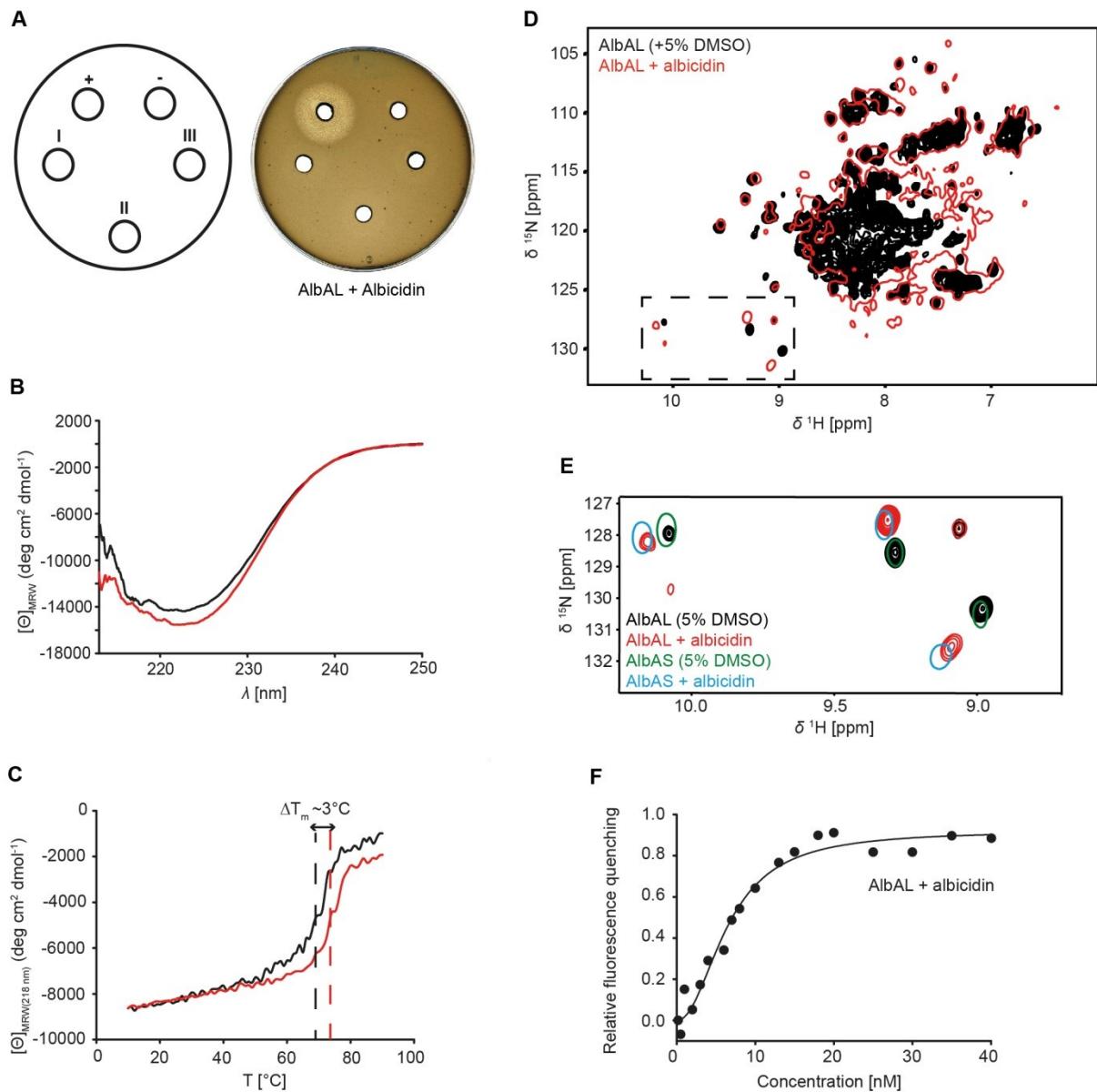
Supplementary Figure 2 Phylogenetic relationship of AlbAS and sequence alignment of AlbAS and AlbBS. A Phylogenetic relationship of AlbAS proteins based on a BLAST search with *Klebsiella oxytoca* AlbAS (red). **B** Sequence alignment of AlbAS and AlbBS. AlbAS secondary structure elements as seen in the AlbAS-albicin co-crystal structure are shown above the alignment (NTD - blue, α6 – orange, CTD – cyan, α6' - red). Predicted α-helical secondary structure elements of AlbBS are shown as cylinders below the alignment. Important AlbAS-albicidin-binding residues (triangles) are highlighted.
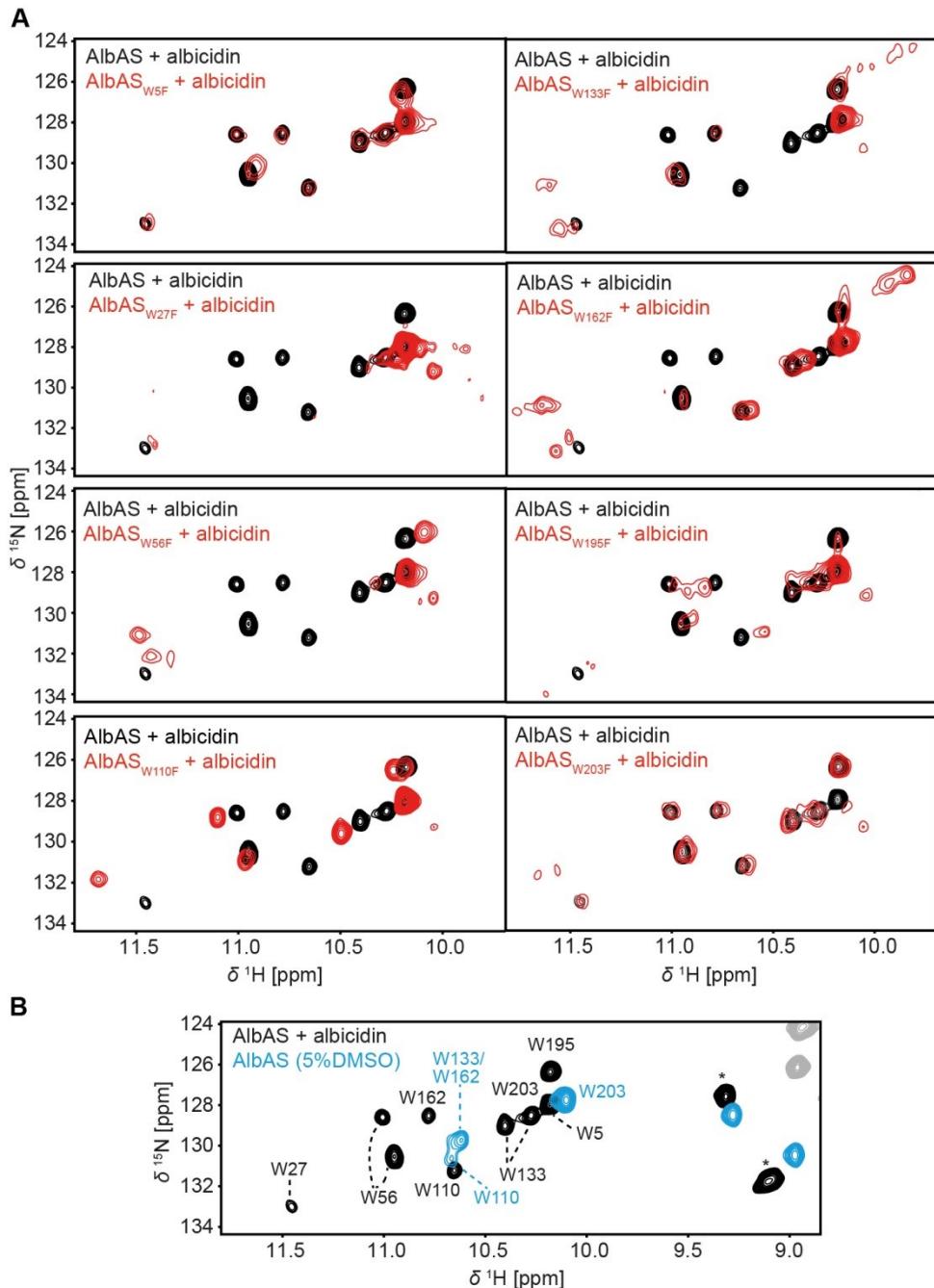
Supplementary Figure 3 Sequence alignments with secondary structure prediction of several AlbA homologs from selected ESKAPE pathogens. Predicted α-helical secondary structure elements (from PROMALS3D) are indicated by cylinders above the sequence (black cylinders: HTH domain, blue: AlbAS NTD, orange: α6, cyan: AlbAS CTD, red: α6'). Predicted α-helical secondary structure elements for the aligned MerR-family proteins are highlighted with the respective background color. Key residues are highlighted in bold letters and triangles. The percentage sequence identity for the aligned proteins to *Klebsiella oxytoca* AlbAL (100%, the protein sequence used in this work) is as followed: *Klebsiella oxytoca* 95%, *Enterobacter sp.* HMSC055A11 89%, *Raoultella ornithinolytica* 81%, *Enterobacteriaceae* 80%, *Klebsiella pneumoniae* 78% and 77%, *Pseudomonas aeruginosa* 48%, *Pseudomonas stutzeri* 47% and *Acinetobacter baumannii* both 42%.
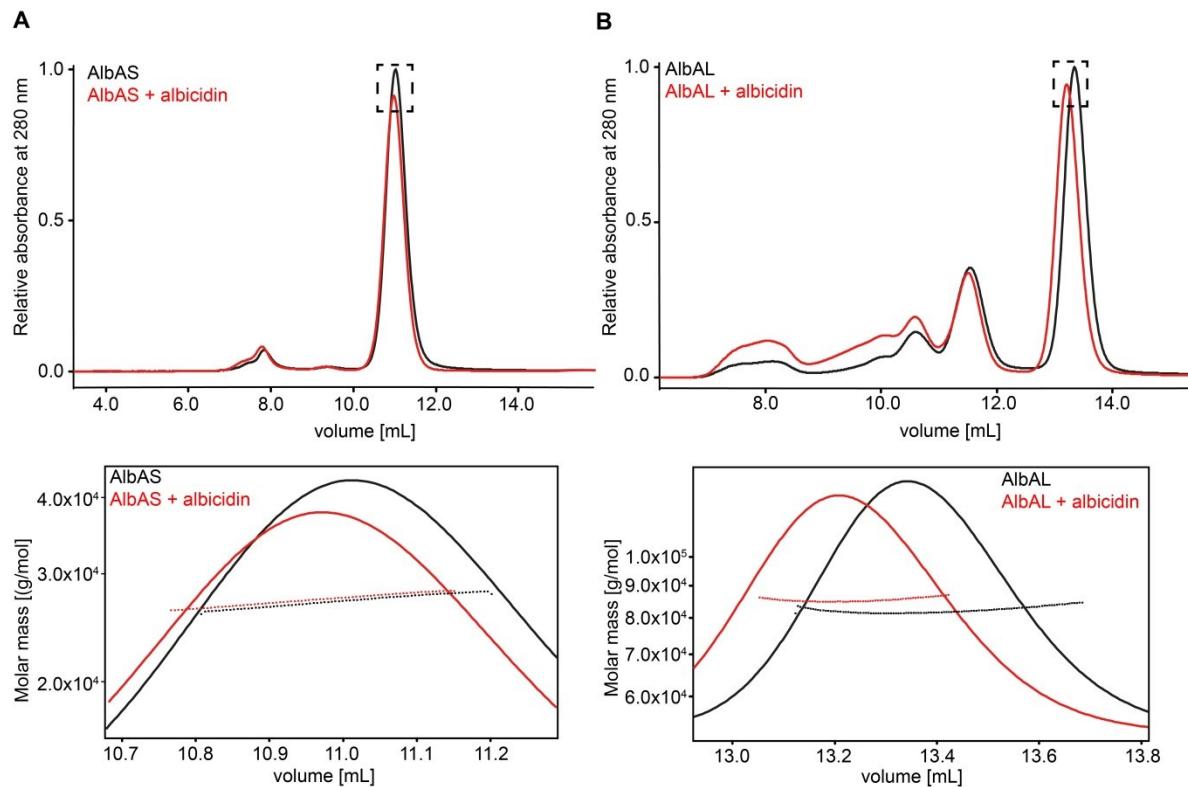
Supplementary Figure 4 Coomassie-stained SDS-PAGE of all proteins after TEV-cleavage. A AlbAS (25.9 kDa), AlbAS$_M$ after TEV-cleavage (25.7 kDa) and AlbAL (40 kDa) after TEV-cleavage. **B** AlbAS$_{CTD}$ (11.0 kDa) and AlbAS$_{NTD}$ (15.7 kDa) after TEV cleavage. M is the protein standard (10 kDa to 180 kDa).
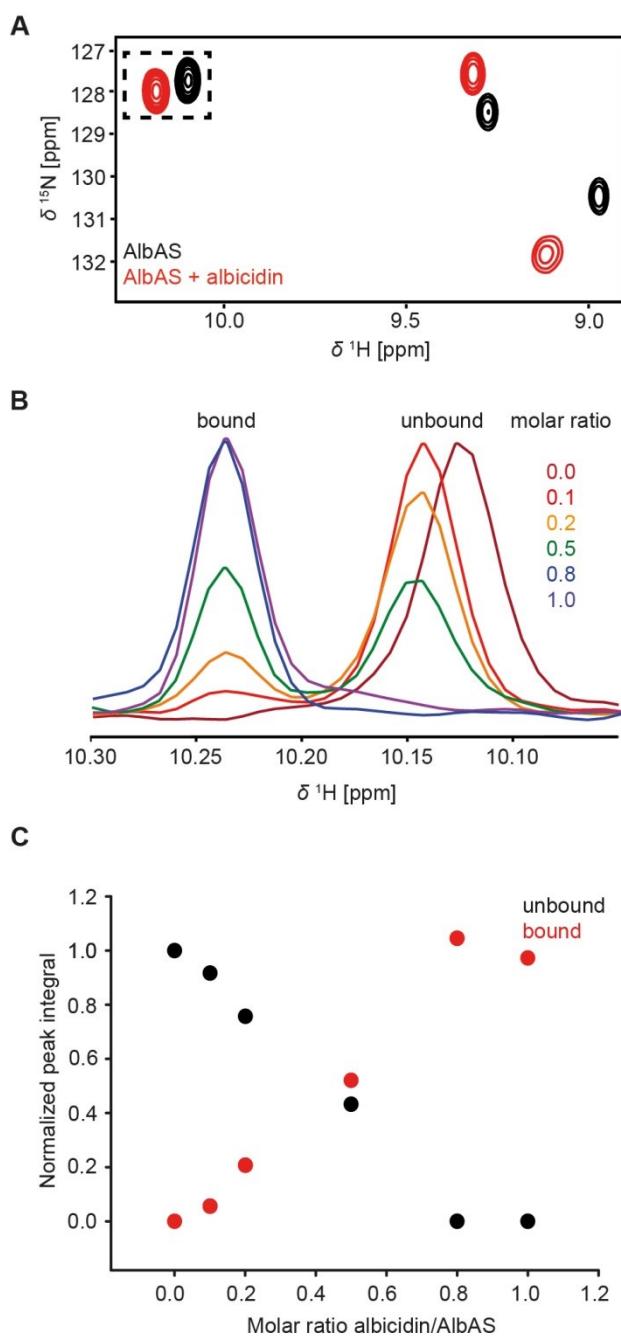
Supplementary Figure 5 Characterization of albicidin binding to AlbAL. A Agar diffusion assay with AlbAL (40 µM) and 40 µM albicidin (in triplicates I-III). The first plate illustrates the sample arrangement on the assay plates. Positive control with only 40 µM albicidin (+) and negative control with 40 µM protein in 5% DMSO and buffer (-) are shown on top. **B** Far-UV CD spectra of AlbAL in the absence (black) or presence (red) of albicidin (molar ratio 1.5: 1 albicidin:AlbAL) at 21 °C. **C** CD-based thermal unfolding of AlbAL in the absence (black) or presence (red) of albicidin. T_m values are indicated by dashed lines. **D** Overlay of ^{2}H, ^{15}N TROSY NMR spectra of AlbAL in the absence (black) or presence (red) of albicidin. Dashed lines highlight the reporter region enlarged in **E**. **F** Determination of binding affinities by monitoring quenching of fluorescence emission of AlbAL. Non-linear regression of fluorescence quenching data yielded a K_d of 7.4 ± 0.9 nM and a Hill coefficient, *n*, of 2.0 ± 0.23 for the interaction of albicidin with 20 nM AlbAL. The agar diffusion assay and CD experiments were conducted once. The fluorescence assay was conducted twice, and the standard deviation is given for the determined K_d. For the NMR experiments, one purified protein sample was used and spectra accumulated with ≥ 64 scans.
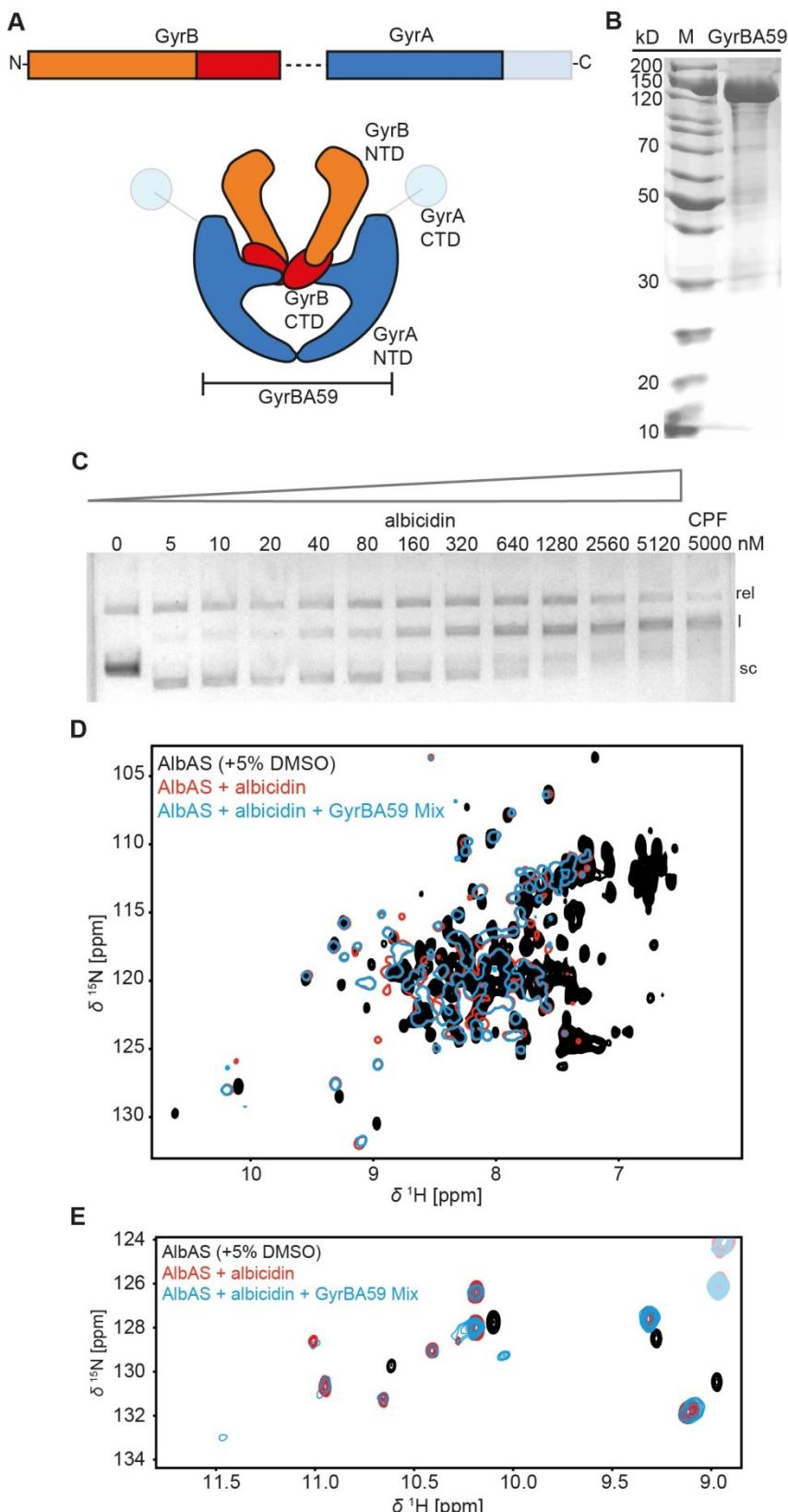
Supplementary Figure 6 Assignment of Trp side-chain signals. A Sections of ^{1}H-^{15}N HSQC spectra showing Trp Nε-Hε resonances of AlbAS (black) and the eight AlbAS Trp-to-Phe mutants (red) in the presence of albicidin. Since full backbone resonance assignments were hampered by the molecular weight and instability of AlbAS-albicidin over several days, we attempted to assign the Trp signals by generating eight single Trp-to-Phe mutants (W5F, W27F, W56F, W110F, W133F, W162F, W195F, W203F), thus anticipating the loss of a single HSQC cross peak for each mutant. The instability of the mutants, the inherent protein dynamics and the close vicinity of Trp residues in the complex usually caused the loss of more than one signal for each mutant and to non-trackable chemical shift changes, which impeded an unambiguous assignment. However, the combined information from all spectra and the knowledge of the relative positions of all Trp residues in the crystal structure helped us to tentatively assign the Trp signals as shown in **B**. The blue spectrum shows AlbAS in the absence of albicidin. Dashed lines indicate ambiguity. Signals marked with an asterisk represent backbone amides identified by 3D HNCO and HNCA experiments, whereas Trp Hε–Nε signals were further confirmed by tryptophan-selective 2D MUSIC experiments (Standard Bruker pulse sequences).

Supplementary Figure 7 SEC-MALS measurements of AlbAS, AlbAS-albicidin, AlbAL and AlbAL- albicidin to determine the molecular mass. A AlbAS and AlbAS-albicidin form monomers in solution with molecular masses of 27.1 kDa and 27.2 kDa, respectively (theoretical masses of 25.8 kDa and 26.6 kDa) **B** Analysis of AlbAL and AlbAL-albicidin showed that the proteins form dimers in solution. The doted lines correspond to the average molar mass of 82.2 kDa (black) and 85.5 kDa (red) for AlbAL and AlbAL-albicidin, respectively (theoretical masses of 40.0 kDa and 40.8 kDa). The SEC-MALS experiments were performed once.
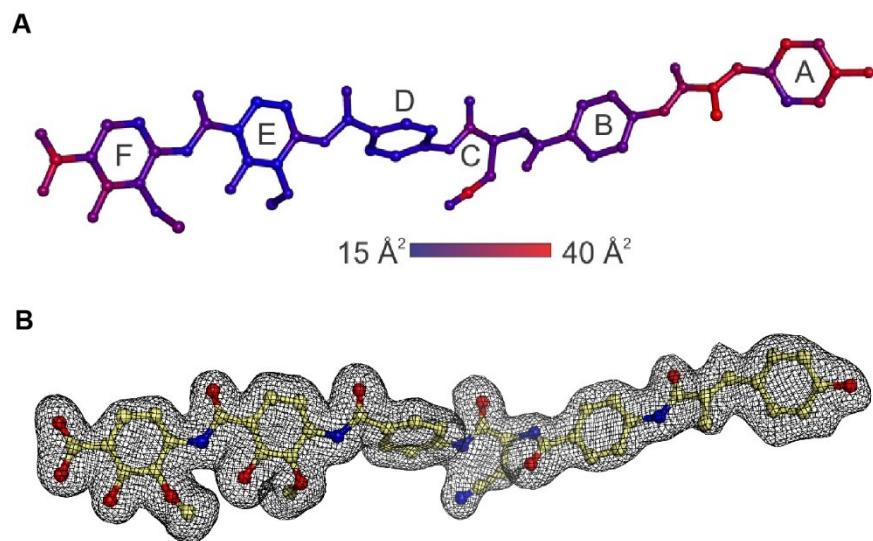
Supplementary Figure 8 NMR titration experiments. A Reporter region of ^{1}H-^{15}N HSQC spectra of AlbAS in the absence (black) and presence of albicidin (red) (final DMSO concentration of 5% in both cases). **B** Superposition of 1D slices extracted from ^{1}H-^{15}N HSQC spectra (dashed box in A) at different albicidin/AlbAS ratios. AlbAS without compound (0.0) contains no DMSO. **C** Plot of the normalized peak integrals of the bound (red) and unbound (black) populations as a function of the molar ratio of albicidin to AlbAS.
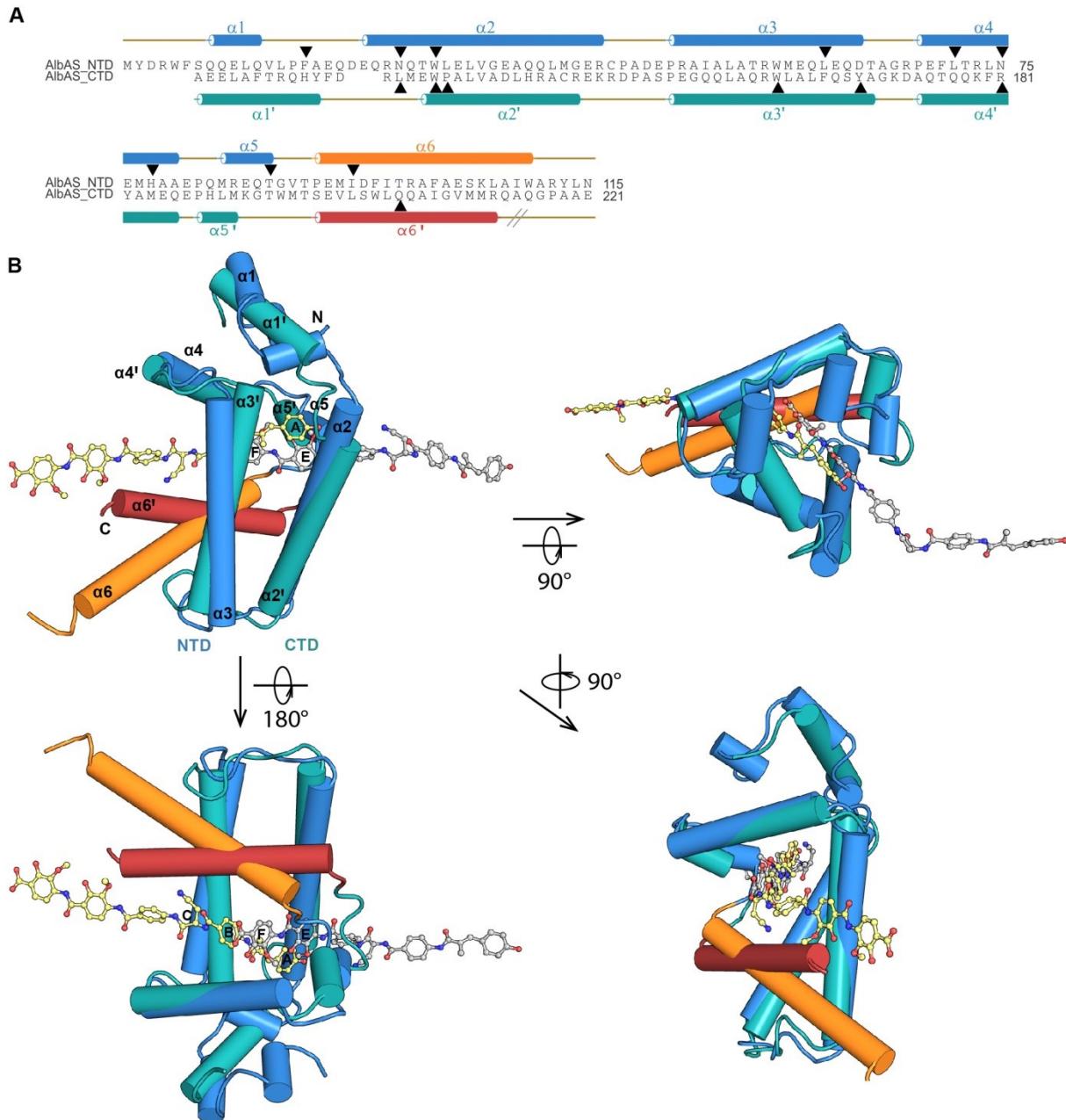
Supplementary Figure 9 Competition experiment with AlbAS and DNA gyrase construct GyrBA59. A Graphical scheme of the DNA gyrase with highlighted elements belonging to the artificial GyrBA59 fusion construct. **B** Purified GyrBA59 after TEV-cleavage (148.6 kDa) on a coomassie-stained SDS-PAGE. **C** Agarose gel-electrophoresis of the DNA-cleavage assay showing albicidin stabilizing the cleavage complex of *E.coli* GyrBA59 and pBR322 (linearized plasmid, l) over a wide range of albicidin concentrations (5 nM – 5120 nM) in presence of 1.4 mM ATP. Control reaction in presence of 5000 nM ciprofloxacin. rel: open nicked plasmid, l: linearized plasmid, sc: supercoiled plasmid. **D**
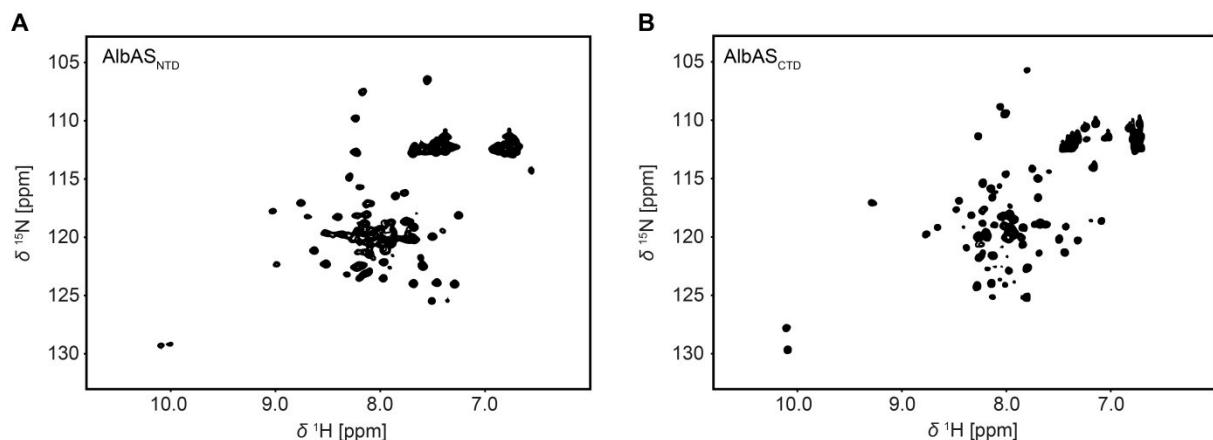
SOFAST-HMQC of ^{15}N-labelled AlbAS, purified GyrBA59 and albicidin with final concentrations of 30 µM each and addition of MgCl$_2$, ATP and pBR322 plasmid DNA. **E** Well-resolved reporter region of the SOFAST-HMQC in **D**. The NMR experiments were performed once.
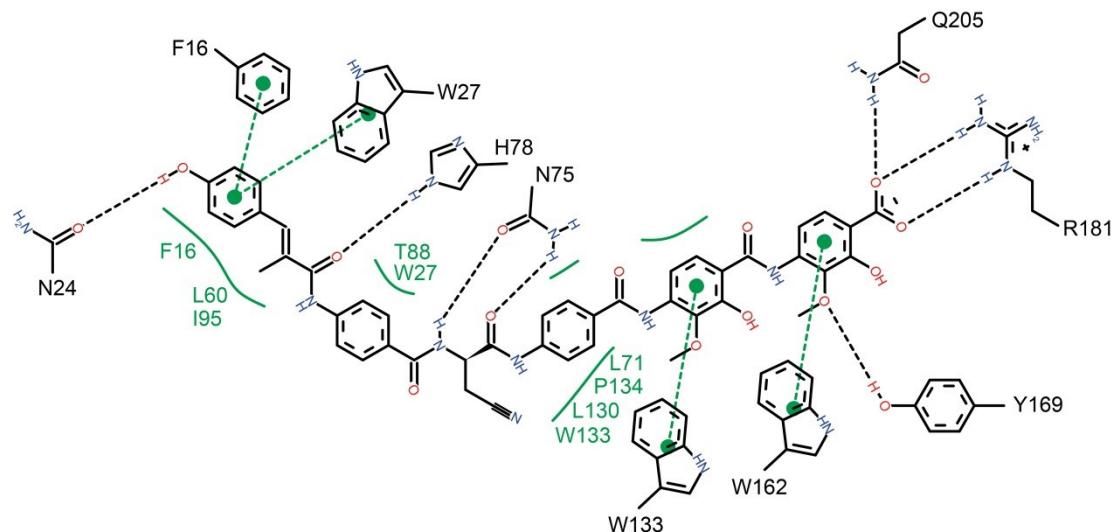
Supplementary Figure 10 Electron density and B-factor representation of albicidin. A AlbAS-bound albicidin with atoms colored by B-factors (low to high B-factors colored blue to red. **B** Polder electron density map of AlbAS-bound albicidin (yellow, with oxygen and nitrogen atoms colored red and blue, respectively) shown as a black mesh at a σ-level of 2.0. Orientation of albicidin is identical to Figure 5.

A

```
                          α1                    α2                      α3                    α4
AlbAS_NTD  MYDRWFSQQELQVLPFAEQDEQRNQTWLELVGEAQQLMGERCPADEPRAIALATRWMEQLEQDTAGRPEFLTRLN  75
AlbAS_CTD    AEELAFTRQHYFD    RLMEWPALVADLHRACREKRDPASPEGQQLAQRWLALFQSYAGKDAQTQQKFR  181
                          α1'                   α2'                     α3'                   α4'
```

```
                   α5          α6
AlbAS_NTD  EMHAAEPQMREQTGVTPEMIDFITRAFAESKLAIWARYLN  115
AlbAS_CTD  YAMEQEPHLMKGTWMTSEVLSWLQQAIGVMMRQAQGPAAE  221
                   α5'         α6'
```

B
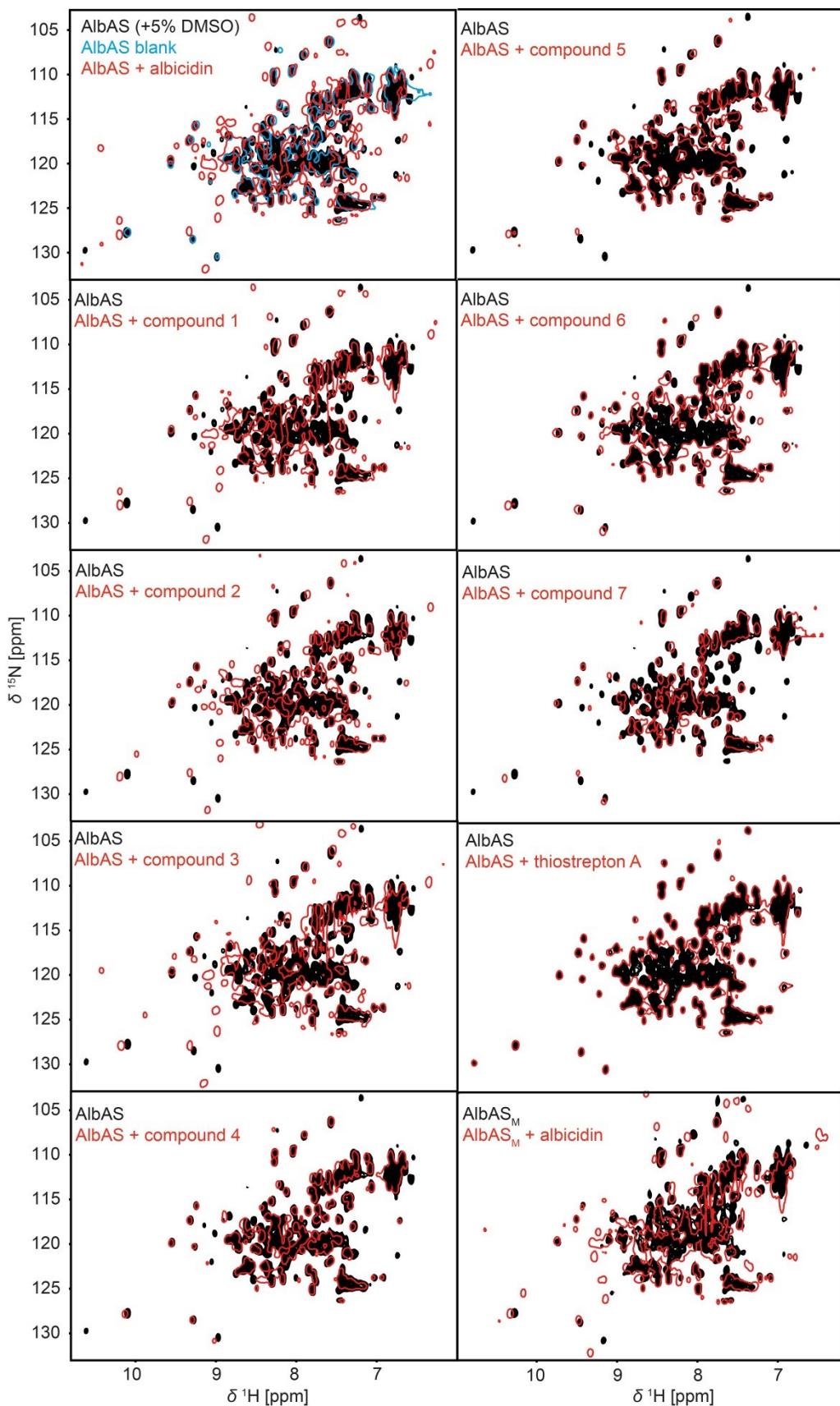
Supplementary Figure 116 Comparison of AlbAS NTD and AlbAS CTD. A Sequence alignment of the NTD (blue) and CTD (cyan). α-helical secondary structure elements as seen in the AlbAS-albicidin crystal structure are shown above and below the alignment as cylinders. Helices α6 (orange) and α6' (red) are highlighted. Important albicidin binding residues are marked with triangles. **B** Structural alignment of AlbAS NTD (blue) and CTD (cyan) showing the binding and entrance sites of albicidin used by the two repeat units. Albicidin is shown in ball-and-stick representation and colored grey (for the NTD) and yellow (for the CTD). Helices α6 (orange) and α6' (red) are highlighted. Different orientations are marked with arrows and rotation axis. The alignment in the left corner has the same orientation as in Figure 5B of the main text.
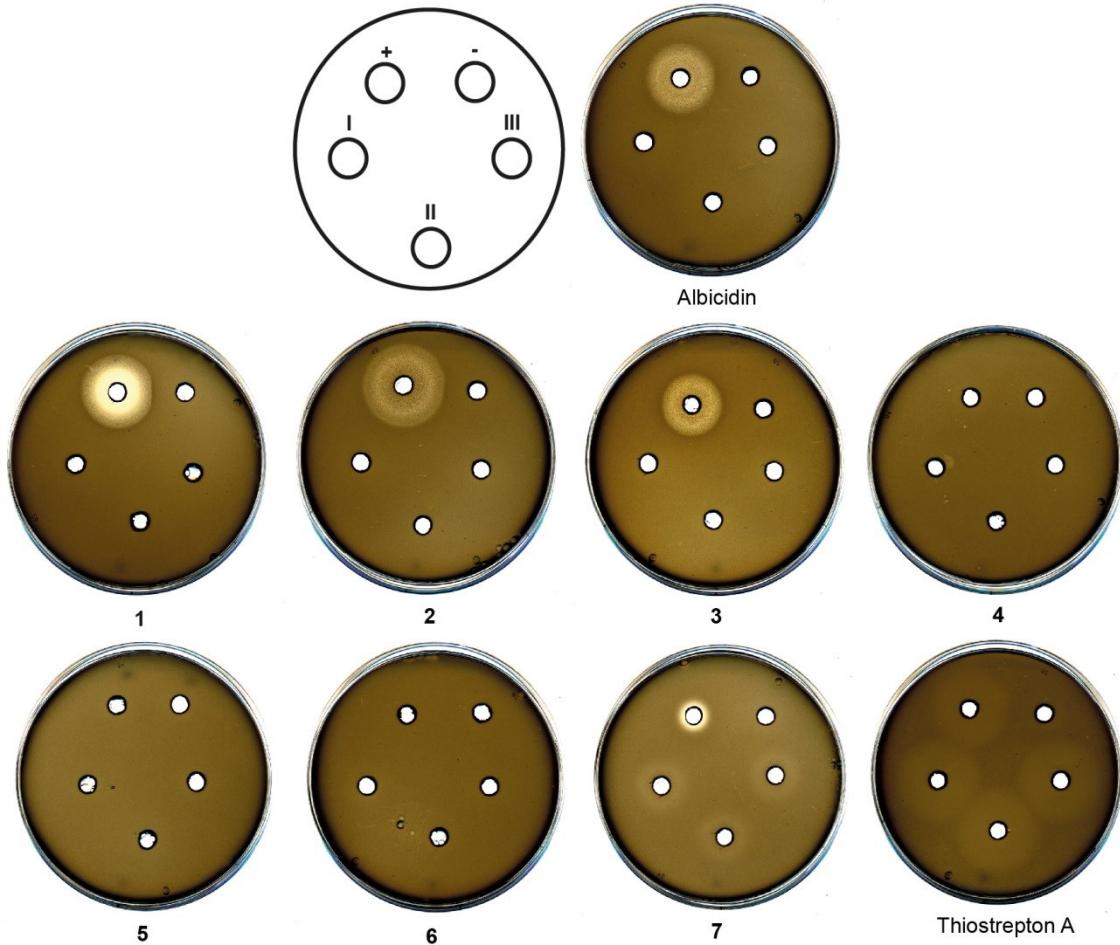
Supplementary Figure 12 ^{1}H-^{15}N **HSQC spectra of truncated AlbAS variants.** Spectra of **A** AlbAS$_{NTD}$ and **B** AlbAS$_{CTD}$ are shown in black. The NMR spectra were performed using one purified protein sample each and accumulated with 16 scans.
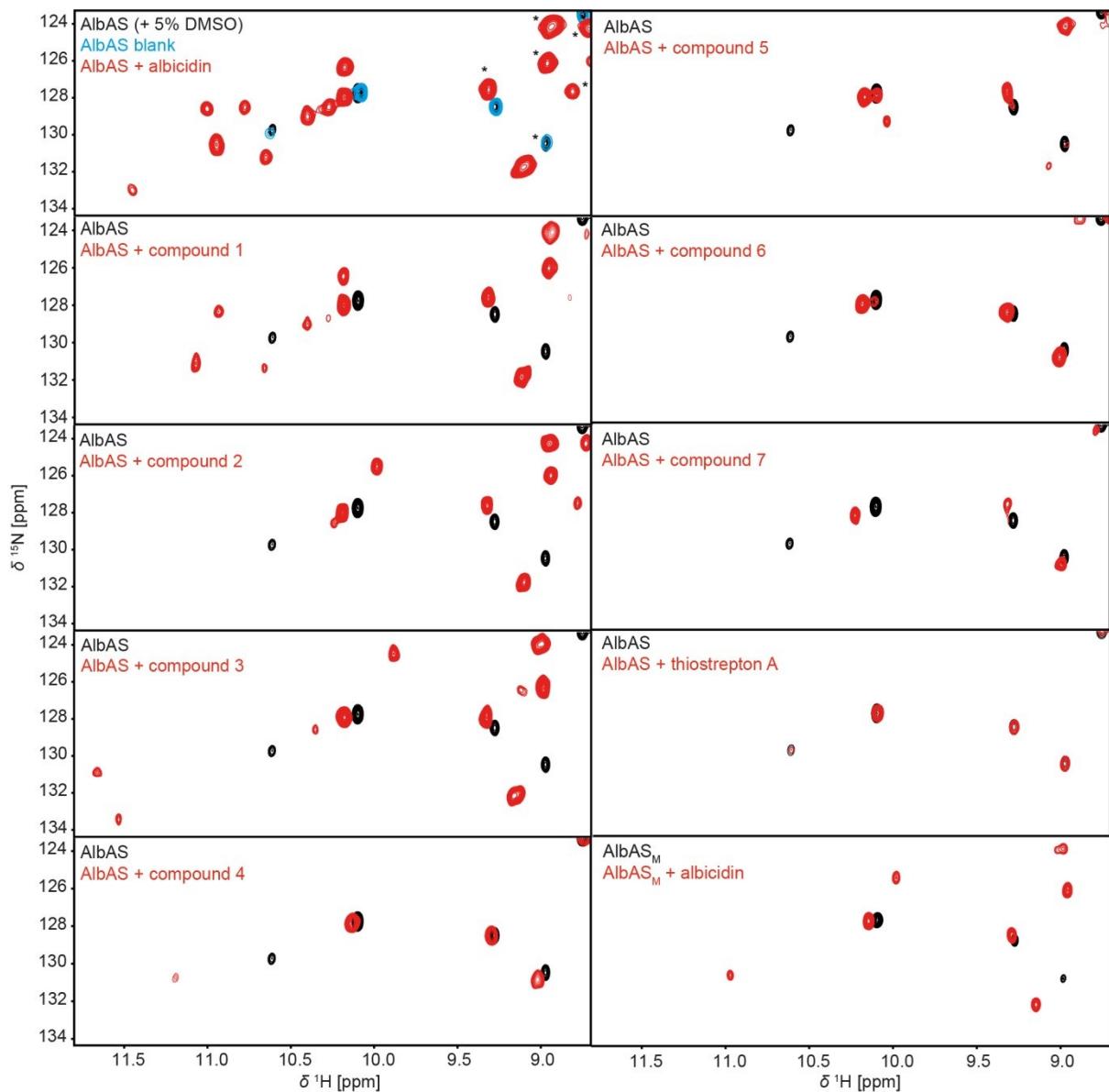
Supplementary Figure 13 Network of interactions in the albicidin binding pocket. Hydrogen bonds are illustrated as dashed lines. π-π stacking is highlighted with green dashed lines. Residues contributing to the hydrophobic binding tunnel are highlighted in green.
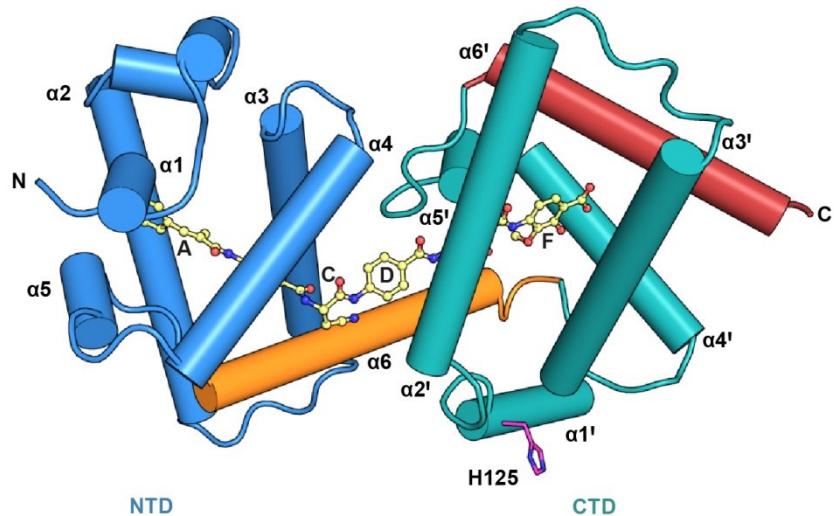
Supplementary Figure 14 Superposition of ^{1}H-^{15}N HSQC spectra of AlbAS with albicidin and different compounds. The first spectrum (same spectrum as presented in Figure 3E) shows AlbAS (black) in 95% buffer/5% DMSO, unbound- AlbAS blank in 100% buffer without compound (blue) and AlbAS with albicidin (red). The following spectra show AlbAS (black) with the corresponding titrated compound (red). AlbAS-albicidin measurements were performed on five different samples and a representative experiment is shown in the figure. The NMR measurements using albicidin derivatives were conducted once.
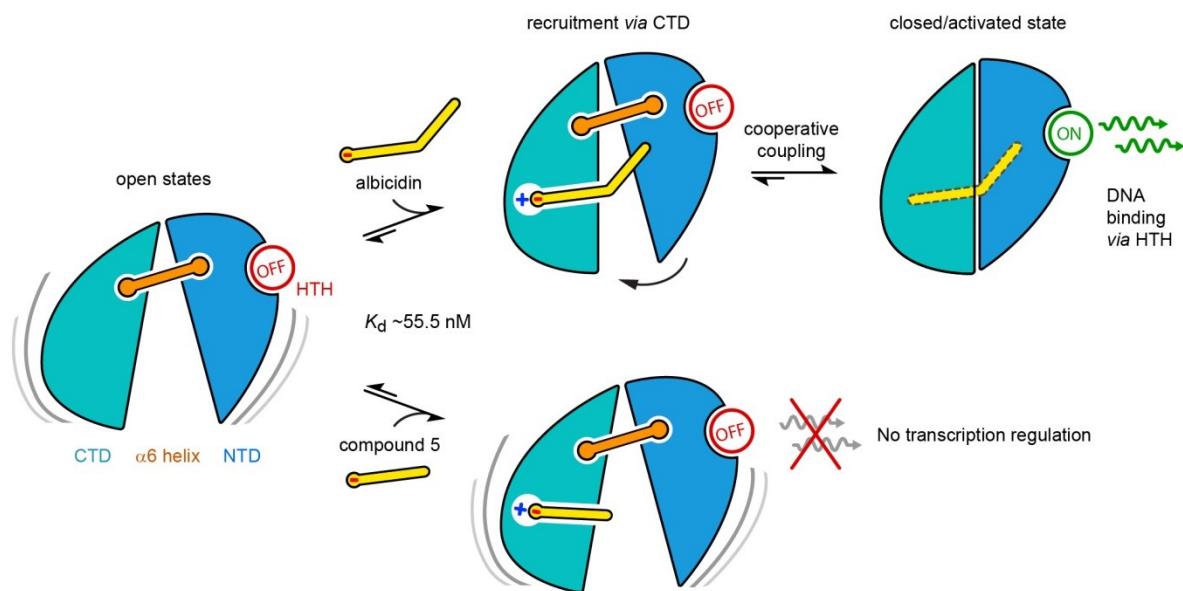
Supplementary Figure 15 Agar diffusion assay with different compounds. AlbAS (40 µM) with albicidin, thiostrepton A or compound **1**-**7** (40 µM each) in triplicates I-III. The first plate illustrates the sample arrangement on the assay plates. Positive control with only 40 µM albicidin (+) and negative control with 40 µM protein in 5% DMSO and buffer (-) are shown on top. Agar diffusion assays with AlbAS-albicidin were conducted three times and a representative result is shown in the figure. The assays with AlbAS and derivatives were performed once.

Supplementary Figure 16 Sections of ^{1}H-^{15}N HSQC downfield region of AlbAS-albicidin and derivatives. The first spectrum shows AlbAS (black) in 95% buffer/5% DMSO, unbound-AlbAS blank in 100% buffer without compound (blue) and AlbAS with albicidin (red). Signals marked with an asterisk could be assigned as backbone amides. AlbAS-albicidin measurements were conducted five times and a representative experiment is shown in the figure. The NMR measurements for AlbAS with derivatives were conducted once.

Supplementary Figure 17 Position of residue H125 on the surface of the AlbAS-albicidin complex.

Supplementary Figure 18 Model of the binding mechanism of AlbAS. In the absence of albicidin, AlbAS undergoes inherent domain-domain motions on the micro-to-millisecond timescale. In this conformational ensemble, the N-terminal HTH domain is not competent for DNA binding. Upper lane: Initial anchoring of the C-terminal segment of albicidin is mediated via the CTD of AlbAS (the salt-bridge between R181 of AlbAS and building block F of albicidin is indicated by the charges). The subsequent binding of the NTD to the N-terminal segment of albicidin is highly cooperative giving an effective K_d for the entire binding process of ≤ 5.6 ± 0.2 nM. Only this mutual fixation of the NTD and CTD causes the HTH to adopt a DNA-binding competent state. Drug binding and transcriptional regulation of MerR-related genes are thus strongly coupled. Lower lane: C-terminal fragments of albicidin (here compound 5), which are still strongly captured by the CTD (K_d of approx. 55.5 ± 3.6 nM), do not or only inefficiently bridge the CTD and NTD. The HTH domain is thus decoupled from the drug-binding event and is not able to initiate transcription. Notably, in the case of AlbAL, these processes occur in a functional dimer capable of DNA binding after albicidin capture.

Supplementary Table 1 Data collection, phasing and refinement statistics (SAD-SeMet).

	SAD
Data collection	
Space group	$C222_1$
Cell dimensions	
a, b, c (Å)	54.07, 123.35, 159.2
a, b, g (°)	90, 90, 90
Resolution (Å)	50.00 - 1.70 (1.74 - 1.70)
R_{meas} (%)	11.9 (179.6)
I / sI	16.53 (1.49)
Completeness (%)	100 (100)
Redundancy	13.50 (13.62)
Refinement	
Resolution (Å)	50.00 - 1.70 (1.76 - 1.70)
No. reflections	58928 (5818)
R_{work} / R_{free}	0.177 / 0.204
No. atoms	
Protein	3789
Albicidin	124
Water	372
B-factors	
Protein	31.3
Albicidin	27.9
Water	35.4
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.02

*Number of xtals for each structure should be noted in footnote. *Values in parentheses are for highest-resolution shell.

[AU: Equations defining various *R*-values are standard and hence are no longer defined in the footnotes.]
[AU: Phasing data should be reported in Methods section.]
[AU: Ramachandran statistics should be in Methods section at end of Refinement subsection.]
[AU: Wavelength of data collection, temperature and beamline should all be in Methods section.]

Supplementary Table 2 Codon optimized DNA sequences.

Codon optimized AlbAS DNA sequence

CATATGTATGATCGTTGGTTTAGCCAGCAAGAACTGCAGGTTCTGCCGTTTGCAGAACAGGATG
AACAGCGTAATCAGACCTGGCTGGAACTGGTTGGTGAAGCACAGCAGCTGATGGGTGAACGTT
GTCCGGCAGATGAACCGCGTGCAATTGCCCTGGCAACCCGTTGGATGGAACAGCTGGAACAGG
ATACCGCAGGTCGTCCGGAATTTCTGACCCGTCTGAATGAAATGCATGCAGCAGAACCGCAGA
TGCGTGAACAGACCGGTGTTACTCCGGAAATGATTGATTTTATCACCCGTGCATTTGCCGAAAG
CAAACTGGCAATTTGGGCACGTTATCTGAATGCAGAAGAACTGGCATTTACCCGTCAGCATTAT
TTTGATCGTCTGATGGAATGGCCTGCACTGGTTGCCGATCTGCATCGTGCATGTCGTGAAAAAC
GTGATCCGGCAAGTCCGGAAGGTCAGCAGCTGGCACAGCGTTGGCTGGCACTGTTTCAGAGCT
ATGCAGGTAAAGATGCACAGACCCAGCAGAAATTTCGTTATGCAATGGAACAAGAACCGCATC
TGATGAAAGGCACCTGGATGACCAGCGAAGTGCTGAGCTGGCTGCAGCAGGCAATTGGTGTT
ATGATGCGTCAGGCACAGGGTCCGGCAGCAGAATAAGCGGCCGC
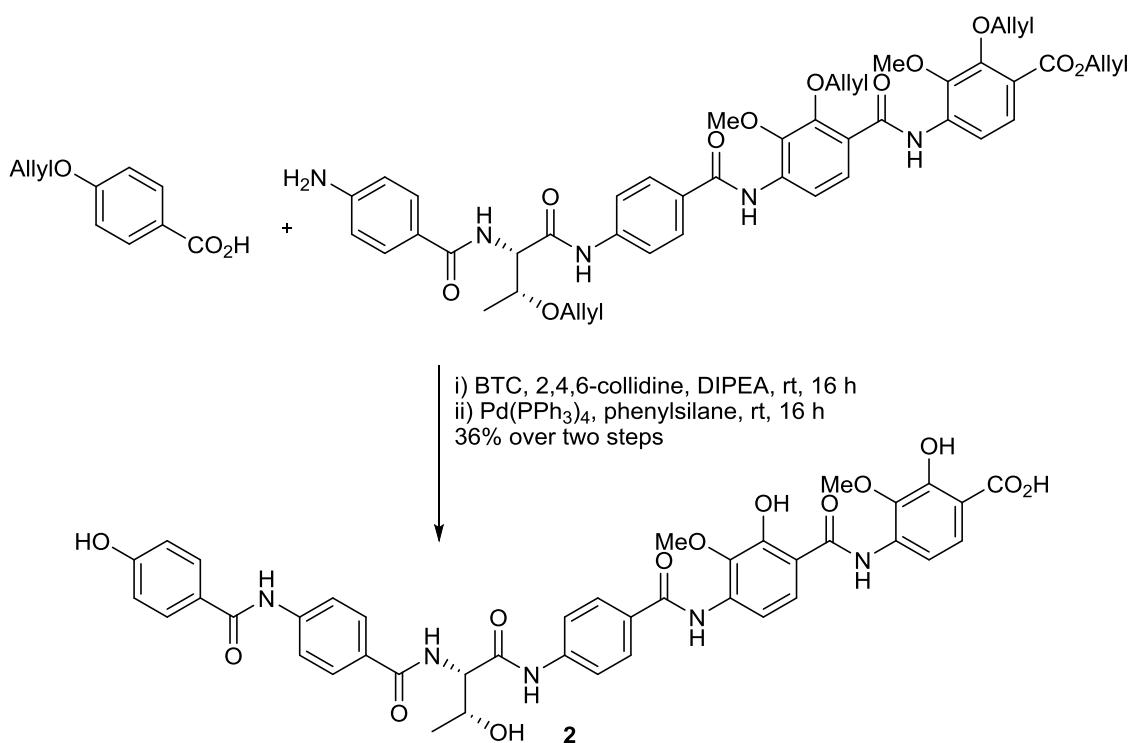
Codon optimized AlbAL DNA sequence

ATGCTGATCCAAGTGGGAGAGTTGGCGAAACGTGCGGGTATGACAGTGCGTACGCTGCATCAC
TACGAGCAAACCGGCCTGCTGACTCCGAGTGCGCGTTCTGCTGCTGGCTATCGCCTGTACAACC
TTGCAGCAGTTCAGCGTCTGCACATGATCAAAGCGTTAGCTCAGGCTGGCTTAGAGCTGGCAA
CCATCAAAGACTATCTGGATCGCGACGCATTTTCCTTGAGCGATCTGCTGGTGAAACAGATTGC
CACCTTAGATCGGCAGCTGCAGACTCTGTCGACCCTTCGCCAACGCCTGGCTCTGCTGCGCGAA
GAACTCGATAGTGGCGGTGAACCGGACCTGGAATCGTGGGTGCAGACCTTAGAGCTGATGAA
GATGTACGATCGCTGGTTTTCCCAACAGGAGCTTCAGGTCTTGCCGTTTGCCGAACAAGACGAA
CAGCGCAATCAGACCTGGCTCGAGCTTGTAGGCGAAGCGCAACAACTCATGGGCGAACGTTGC
CCTGCCGATGAACCCCGCGCGATTGCGCTGGCCACACGCTGGATGGAACAGCTGGAACAGGAT
ACAGCGGGACGTCCGGAATTCCTGACGCGCCTCAATGAGATGCATGCCGCTGAACCGCAAATG
CGTGAACAGACCGGTGTAACGCCCGAAATGATTGACTTCATCACCCGTGCGTTTGCCGAAAGCA
AATTAGCGATTTGGGCCCGTTATCTGAACGCGGAAGAACTGGCGTTTACGCGTCAGCACTATTT
CGATCGGCTGATGGAATGGCCAGCACTTGTTGCGGATTTGCATCGCGCCTGTCGGGAGAAACG
CGATCCAGCATCTCCGGAAGGGCAGCAATTAGCCCAACGCTGGTTGGCGCTCTTCCAGAGCTAT
GCCGGCAAAGACGCACAGACTCAGCAGAAGTTTCGCTACGCAATGGAACAGGAGCCGCATCTG
ATGAAAGGTACGTGGATGACCTCAGAAGTTCTGAGCTGGTTGCAGCAGGCCATTGGGGTCATG
ATGCGTCAAGCGCAAGGTCCTGCAGCTGAGTAA

Supplementary Table 3 Oligonucleotides used for cloning.

Oligonucleotide sequence (5' – 3')	Name
CTTGTATTTCCAGGGCCAATGGATCGTCTGATGGAATGGCC	AlbAS$_{CTD}$_fw
TCCACCAGTCATGCTAGCCATTTATTCTGCTGCCGGACCCTGTGCCT	AlbAS$_{CTD}$_rv
CTTGTATTTCCAGGGCCA ATGTATGATCGTTGGTTTAG	AlbAS$_{NTD}$_fw
TCCACCAGTCATGCTAGCCATTTAATTCAGATAACGTGCCCAAATTGC	AlbAS$_{NTD}$_rv
CGTATTACCGCCTTTGAGTGAGCTGATACC	Fw_all_AlbAS$_M$
GGTATCAGCTCACTCAAAGGCGGTAATACG	Rv_all_AlbAS$_M$
ATTTCTGACCCGTCTGGCAGAAATGCATGCAGCAG	AlbAS$_M$_N75A_fw
CAGACGGGTCAGAAATTCCGGACGACCTGCGG	AlbAS$_M$_N75A_rv
GACCCAGCAGAAATTTGCATATGCAATGGAACAAGAA	AlbAS$_M$_R181A_fw
AAATTTCTGCTGGGTCTGTGCATCTTTACCTGCATAG	AlbAS$_M$_R181A_rv
AGTGCTGAGCTGGCTGGCACAGGCAATTGGTGTTATG	AlbAS$_M$_Q205A_fw
CAGCCAGCTCAGCACTTCGCTGGTCATCCAGGTGCCT	AlbAS$_M$_Q205A_rv
CGTATTACCGCCTTTGAGTGAGCTGATACC	AlbAS_Trp_all_fw
GGTATCAGCTCACTCAAAGGCGGTAATACG	AlbAS_Trp_all_rv
CCATATGTATGATCGTTTTTTTAGCCAGCAAGAACTG	AlbAS_W5F_fw
ACGATCATACATATGGCCCTGGAAATACAAGTTTTCG	AlbAS_W5F_rv
GAACAGCGTAATCAGACCTTTCTGGAACTGGTTGGT	AlbAS_W27F_fw
GGTCTGATTACGCTGTTCATCCTGTTCTGCAAA	AlbAS_W27F_rv
GCCCTGGCAACCCGTTTTATGGAACAGCTGGAACAG	AlbAS_W56F_fw
ACGGGTTGCCAGGGCAATTGCACGCGGTTCATC	AlbAS_W56F_rv
GAAAGCAAACTGGCAATTTTTGCACGTTATCTGAATGCA	AlbAS_W110F_fw
AATTGCCAGTTTGCTTTCGGCAAATGCACGGGTGAT	AlbAS_W110F_rv
TTTGATCGTCTGATGGAATTTCCTGCACTGGTTGCCGAT	AlbAS_W133F_fw
TTCCATCAGACGATCAAAATAATGCTGACGGGTAAA	AlbAS_W133F_rv
AGCAGCTGGCACAGCGTTTTCTGGCACTGTTTCAGAGC	AlbAS_W162F_fw
ACGCTGTGCCAGCTGCTGACCTTCCGGACTTGC	AlbAS_W162F_rv
CCGCATCTGATGAAAGGCACCTTTATGACCAGCGAAGTGCTG	AlbAS_W195F_fw
GGTGCCTTTCATCAGATGCGGTTCTTGTTCCATTGCATA	AlbAS_W195F_rv
ACCAGCGAAGTGCTGAGCTTTCTGCAGCAGGCAATTGGT	AlbAS_W203F_fw
GCTCAGCACTTCGCTGGTCATCCAGGTGCCTTTCAT	AlbAS_W203F_rv
TAATACGACTCACTATAGGG	Sequencing_T7
TGCTAGTTATTGCTCAGCGG	Sequencing_T7-term
AACTTGTATTTCCAGGGCATGCTGATTCAGGTTGGT	pET28a_AlbAL_TEV_fw
GCTTCCTTTCGGGCTTTGTTATTCTGCGGCAGGACCCT	pET28a_AlbAL_TEV_rv
TGGATCCCATATGTCGAATTCTTATGACT	pET28a_GyrBA59_fw
GGTGCTCGAGGTTAGGCGGTGATTTCAGTACGA	pET28a_GyrBA59_rv
GCGAATATCGATATTAGCGACCTTGCG	GyrBA59_fusion_fw
CTCTCGCAAGGTCGCTAATATCGATATT	GyrBA59_fusion_rv

Supplementary methods 1 Chemical synthesis and analytical data of compound 2.

Synthesis of compound **2**:

To a solution of commercially available 4-(allyloxy)-benzoic acid (1.5 eq, 16 mg, 88 µmol), and triphosgene (0.6 eq, 11 mg, 35 µmol), dissolved in anhydrous THF (3 mL), was slowly added 2,4,6-collidine (8 eq, 62 µL, 0.47 mmol) via syringe. After 5 minutes of stirring, a solution of the corresponding allyl-protected aniline pentapeptide9 (1 eq, 50 mg, 59 µmol) and DIPEA (10 eq, 100 µL, 590 µmol), dissolved in anhydrous THF (2 mL), was added. After 16 h stirring at room temperature, the reaction was quenched by the addition of ice. The mixture was extracted with dichloromethane (3 x 20 mL). The combined organic extracts were washed with 1 N HCl (aq) (3 x 20 mL), sat. NaHCO3 (3 x 20 mL) and brine (1 x 20 mL), dried over Na2SO4, filtered and evaporated to dryness. The residue was dissolved in anhydrous THF (5 mL), and subsequently phenylsilane (8 eq, 58 µL, 0.47 mmol) and tetrakis(triphenylphosphine)palladium(0) (0.4 eq, 27 mg, 24 µmol) were added. The reaction was stirred for 16 h at room temperature under exclusion of light. The mixture was clarified via filtration through a 0.2 µm PTFE syringe filter and directly applied to a preparative RP-HPLC purification yielding compound 2 (17 mg, 21 µmol) as a white fluffy powder.

Analytical data for compound **2**:

^{1}H-NMR (500 MHz, DMSO-d6): δ (ppm) = 1.20 (d, J = 6.3 Hz, 3H), 3.78 (s, 3H), 3.91 (s, 3H), 4.15-4.24 (m, 1H), 4.59 (dd, J = 4.8 Hz, J = 8.1 Hz, 1H), 5.08 (s, 1H), 6.88 (d, J = 8.6 Hz, 2H), 7.56-7.62 (m, 2H),

7.80 (t, *J* = 8.5 Hz, 3H), 7.86-7.95 (m, 8H), 7.97 (d, *J* = 8.6 Hz, 2H), 8.04 (d, *J* = 8.9 Hz, 1H), 8.08 (d, *J* = 8.1 Hz, 1H), 9.66 (s, 1H), 10.16 (s, 1H), 10.22 (s, 1H), 10.37 (s, 1H), 11.16 (s, 1H), 11.53 (s, 1H).

HRMS (APCI): m/z calc. for $C_{41}H_{37}N_5O_{13}$ [M-H]$^-$: 806.2304; exp. 806.2336.

Publication V

Ronja Driller, Sophie Janke, Monika Fuchs, Evelyn Warner, Anil R. Mhashal, Dan Thomas Major, Mathias Christmann, Thomas Brück, Bernhard Loll (2018), Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis, *Nature Communications* 9:3971

Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis

Ronja Driller [1], Sophie Janke[2], Monika Fuchs [3], Evelyn Warner[2], Anil R. Mhashal [4], Dan Thomas Major [4], Mathias Christmann [2], Thomas Brück [3] & Bernhard Loll [1]

Terpenes constitute the largest and structurally most diverse natural product family. Most terpenoids exhibit a stereochemically complex macrocyclic core, which is generated by C–C bond forming of aliphatic oligo-prenyl precursors. This reaction is catalysed by terpene synthases (TPSs), which are capable of chaperoning highly reactive carbocation intermediates through an enzyme-specific reaction. Due to the instability of carbocation intermediates, the proteins' structural dynamics and enzyme:substrate interactions during TPS catalysis remain elusive. Here, we present the structure of the diterpene synthase CotB2, in complex with an in crystallo cyclised abrupt reaction product and a substrate-derived diphosphate. We captured additional snapshots of the reaction to gain an overview of CotB2's catalytic mechanism. To enhance insights into catalysis, structural information is augmented with multiscale molecular dynamic simulations. Our data represent fundamental TPS structure dynamics during catalysis, which ultimately enable rational engineering towards tailored terpene macrocycles that are inaccessible by conventional chemical synthesis.
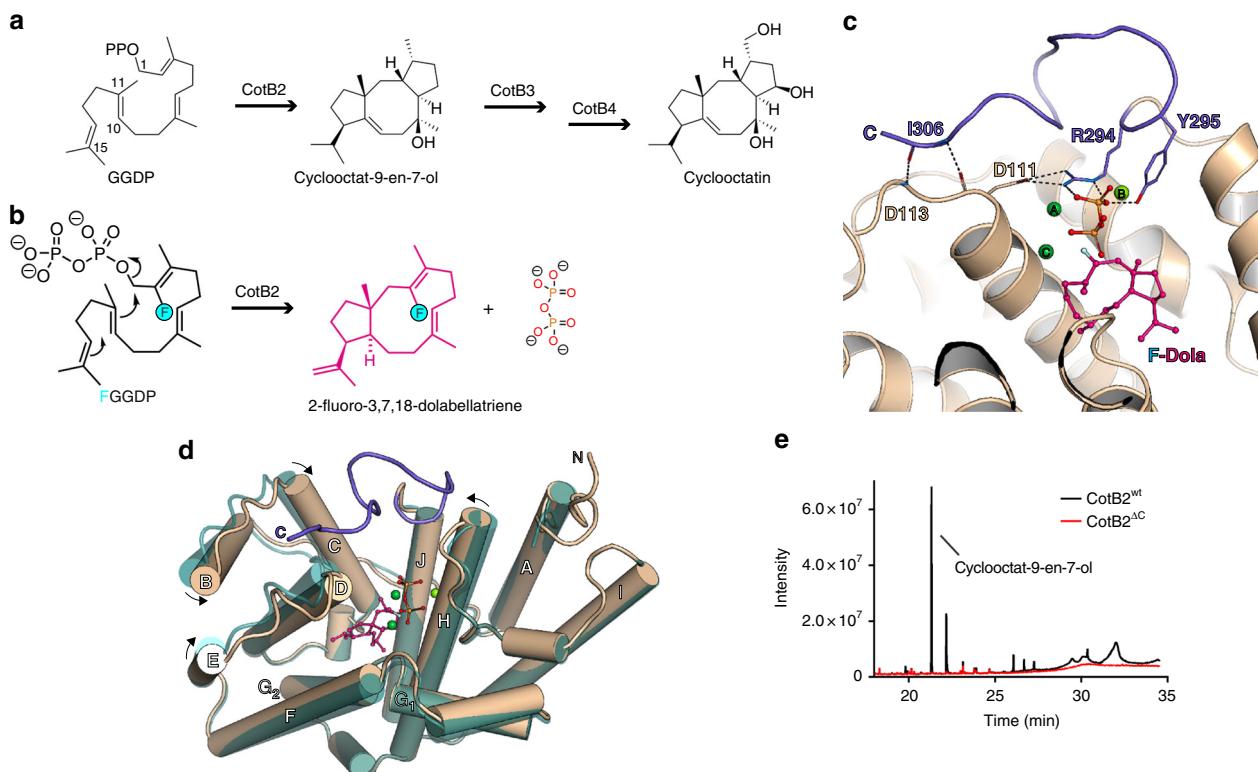
[1] Institut für Chemie und Biochemie, Strukturbiochemie, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany. [2] Institut für Chemie und Biochemie, Organische Chemie, Freie Universität Berlin, Takustr. 3, 14195 Berlin, Germany. [3] Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich (TUM), Lichtenbergstr. 4, 85748 Garching, Germany. [4] Department of Chemistry, Bar-Ilan University, Ramat-Gan 52900, Israel. Correspondence and requests for materials should be addressed to B.L. (email: loll@chemie.fu-berlin.de)
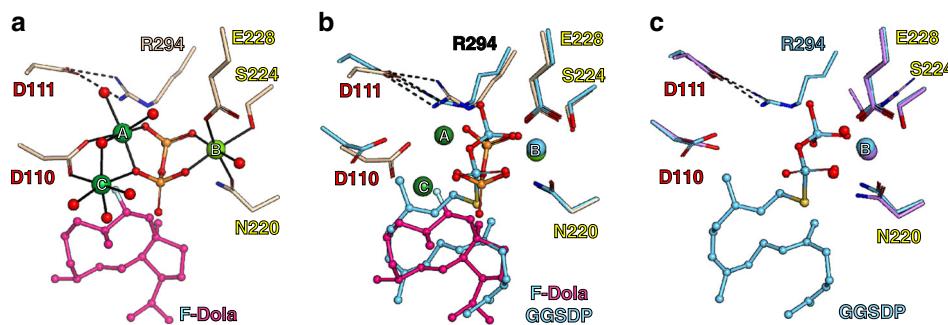
The intense structural diversity of terpenes is the molecular basis of diverse bioactivities, which render these compounds as development leads in the chemical, biotechnological, food and pharmaceutical industry[1,2]. Most terpenes constitute a macrocyclic core skeleton[3], which is generated by cyclisation of aliphatic oligo-prenyl diphosphates[3]. This reaction is catalysed by terpene synthases (TPSs). By contrast, chemically mimicking TPS reactions to generate tailored terpene-type macrocycles can be considered in its infancy, although promising directions have been proposed[4–9]. During the cyclisation event, the acyclic precursor has to be brought into a defined conformation that positions the leaving diphosphate group and the reactive alkene entities in proximity to initiate C–C bond forming reactions. Moreover, the substrate environment has to stabilise several propagating carbocations, to adjust to conformational changes of the substrate as well as assist in intramolecular atom transfer reactions, such as hydride or proton transfers and carbon shifts. Such enzyme guidance is necessary to control the reaction intermediates, although the inherent reactivity of carbocations is also an important component in terpene biocatalysis[10]. In addition to the complex processes in the enzyme's active site, global effects such as dynamic structural changes of the protein scaffold and inter-protein interactions during catalysis are poorly understood. A detailed understanding of dynamic, biochemical and structural effects during catalysis would ultimately enable rational engineering of this enzyme family. This capacity would be a game changing advantage in structure-activity-relation studies of

biologically relevant terpenes as even subtle structural modifications such as the additions of methyl groups to the native substrate (e.g., the "magic methyl effect")[11] or a simple H/D exchange may have a profound influence on the biological activity.

To elucidate the dynamic interactions of TPSs with their substrates during catalysis, we investigated the bacterial diterpene synthase CotB2 as a model system. This enzyme catalyses the cyclisation of the universal diterpene precursor *E,E,E*-geranylgeranyl diphosphate (GGDP) to cyclooctat-9-en-7-ol (Fig. 1a), which is subsequently elaborated to the bioactive compound cyclooctatin[12]. Cyclooctatin is a next-generation anti-inflammatory drug targeting a lysophospholipase that is upregulated in eosinophilic leucocytes rather than a cyclooxygenase inhibited by ibuprofen or aspirin[13]. The CotB2 protein sequence exhibits a modified aspartate-rich DDXD motif[12] deviating from the conventional DDXXD motif and the NSE triad NDXXSXX(R, K)(E,D) as found in bacteria and fungi[3,14], that is altered to a DTE triad in plants. Both motifs are involved in binding a conserved Mg^{2+} triad that is indispensable for precise GGDP orientation in the active site. Upon substrate and Mg^{2+} binding, TPSs undergo a discrete conformational change from an open, catalytically inactive one to a catalytically active closed conformation. Even though the active site cavity is already product-shaped in the open conformation, closure, accompanied by translation and rotation of secondary structure elements, is essential for catalysis. Catalysis is initiated by lysis of the GGDP

Fig. 1 Enzymatic reaction and the structure of the closed state of CotB2, revealing the importance of its C-terminus. **a** The linear substrate geranylgeranyl diphosphate (GGDP) is cyclised by CotB2 to a fusicoccane, with a 5-8-5 fused ring system, which is subsequently elaborated to the bioactive compound cyclooctatin by two cytochrome P450 enzymes CotB3 and CotB4, respectively. **b** 2-fluorogeranylgeranyl diphosphate (FGGDP) is converted to 2-fluoro-3,17,18-dolabellatriene (F-Dola). The fluorinated position of the substrate-analogue FGGPP is indicated by a light blue circle. **c** View into the active site of CotB2^{wt}•Mg^{2+}_3•F-Dola. CotB2 is shown in cartoon representation coloured in light brown. The bound intermediate is shown in magenta and Mg^{2+}-ions are shown in green. Folding of the C-terminus (purple) leads to the formation of several hydrogen bonds (dashed lines), allowing for sensing of the different catalytically important motives. **d** Structural superposition of CotB2^{wt} (open), shown in teal, and CotB2•Mg^{2+}_3•F-Dola (closed), shown in light brown. **e** GC/MS spectrum to monitor product formation by CotB2^{wt} (black) and CotB2^{ΔC} (red). Deletion of the C-terminus results in an inactive enzyme

Fig. 2 Structural comparison of different catalytic states of CotB2. Residues of the DDXD motif are labelled with red letters, residues of the NSE/DTE motif with yellow letters, respectively. Water molecules are depicted as red spheres. Hydrogen bonds are indicated by dashed lines. **a** Catalytic centre of CotB2^{wt}•Mg$^{2+}_3$•F-Dola. Mg^{2+}-ions are depicted as green spheres. Solid lines represent the coordination sphere of the Mg^{2+}-ions. **b** Comparison of CotB2^{wt}•Mg$^{2+}_3$•F-Dola shown in magenta and CotB2^{wt}•Mg$^{2+}_B$•GGSDP shown in cyan. **c** Comparison of CotB2^{F107A}•Mg$^{2+}_B$ shown in pale purple and CotB2^{wt}•Mg$^{2+}_B$•F-Dola shown in cyan

diphosphate moiety, a reaction mechanism characteristic for class I TPSs, generating a highly reactive, unstable carbocation[15]. This initial carbocation propagates intramolecular cyclisation, which is guided by amino acids lining the active site, where π-cation and other electrostatic interactions drive the carbocation through an enzyme-specific reaction cycle. The cyclisation cascade is terminated by lining amino acids mediated deprotonation or addition of a water molecule to the final carbocation [3,16].

Delineating interactions of lining amino acids with the substrate along the entire reaction trajectory is fundamental to understand the complexity of TPS catalysis. This information can be reliably obtained by an interdisciplinary approach combining biochemical, structural biology, computational and biocatalytic studies.

In 2014, we reported the first crystal structure of CotB2 in its open, inactive conformation without substrate. CotB2 adopts an α-helical bundle fold that is conserved among class I TPSs (PDB-ID 4OMG[17]). Single point mutations within the active site lead to pronounced structural product diversity (Supplementary Table 1 and Figs. 1, 2)[17–19].

Here, we report cumulative data that provide significant insights into the chemical mechanisms that constitute the catalytic conversion of GGDP to cyclooctat-9-en-7-ol. Remarkably, even though structural information of TPSs bound to a product or different substrate analogues as well as Mg^{2+} have been reported, a complete in crystallo cyclised product has not yet been observed, demonstrating that the previously reported substrate analogues were either not correctly positioned in the active site or catalytically inactive due to chemical modifications.

Results

Comparison of the open and closed structure of CotB2. In this study, we intended to probe the structural dynamics of the CotB2 reaction using X-ray crystallography of the catalytically relevant protein states in the presence of a substrate-analogue. Remarkably, we could capture the closed conformation of CotB2 by co-crystallisation with the substrate-analogue 2-fluoro GGDP (FGGDP) and determined the crystal structure at 1.8 Å (CotB2^{wt}•Mg$^{2+}_3$•F-Dola (2-fluoro-3,7,18-dolabellatriene); Figs. 1c, d, 2a, Supplementary Table 2). FGGDP is functionalised with a fluorine atom at the C2 position of its hydrophobic tail (Fig. 1b) that interferes with the propagation of the generated carbocation (s). A direct comparison of the open and closed conformation is possible, since our crystals are isomorphous. The overall structure of the open and closed conformation reveals significant differences. Substrate binding induces a translation and rotation of α-helices B, C, D, F and H towards the active site (Fig. 1d) to accurately

position the DDXD Mg^{2+} binding motif and to bring the active site into product-shaped conformation. Upon binding of FGGDP the 12 C-terminal residues are structured to fold over the active site, resulting in the fully active and closed conformation (Fig. 1c, d). Surprisingly, we did not observe the entire FGGDP molecule. Instead, we detected clearly resolved electron density for a single diphosphate moiety and for a ring-like structure positioned in the active site that we interpret as the reaction intermediate 2-fluoro-3,7,18-dolabellatriene (F-Dola; Figs. 1b–d, 2a and Supplementary Fig. 3).

With the native substrate GGDP such in crystallo snapshots would not be feasible due to the highly unstable nature of the carbocation intermediates. Most likely the fluorinated substrate alters propagation of the carbocation and hence we do not observe the native product. The possibility of cyclisation reactions of fluorinated substrates has been demonstrated for 2-fluorofarnesyl diphosphate[20,21]. To our knowledge such a crystal structure with a trapped intermediate has never been described before for other TPSs crystallised in presence of isoprene substrates. In these reported structures, the diphosphate is bound, but not cleaved off and consequently no cyclisation reaction has occurred (Supplementary Discussion).

The diphosphate is coordinated by three Mg^{2+}-ions with Mg$^{2+}_B$ bound to the NSE motif and Mg$^{2+}_A$ and Mg$^{2+}_C$ being coordinated by D110 of the non-canonical DDXD motif (Fig. 2a and Supplementary Discussion and Supplementary Fig. 11). Additionally, active site residues R227 and R294 make contact with the oxygen atoms of the diphosphate moiety to compensate its negative charge. Upon binding of Mg$^{2+}_A$ and Mg$^{2+}_C$, D110 stretches out towards the active site, causing helix D movement thereby reducing the volume of the active site and trapping of the substrate. Moreover, this helix D movement positions the C-terminus for folding over the active site to shield it from bulk solvent. The closure by the C-terminus is initiated by formation of a salt-bridge between D111 and R294. In our previous studies we showed that D111E is an inactive mutant[17]. With our current data set, we can now explain this observation with the exchange at position 111 that disturbs the salt-bridge formation emphasising its catalytic relevance. Additionally, Y295 establishes hydrogen bonds to the diphosphate moiety derived from FGGDP as well as to N220 of the NSE motif. The latter RY-pair (R294/Y295) has been previously described for other TPSs[22]. Folding of the C-terminus is completed by formation of a short anti-parallel β-strand at the end of the C-terminus (Fig. 1c). Folding over of the C-terminus does not seem to be driven by strong ionic interactions, but is dominated by hydrogen bonds and van der Waals contacts.

To verify that the full closure of the active site and the folded C-terminus is not due to crystal contacts, we crystallised CotB2 with (4-amino-1-hydroxy-1-phosphonobutyl)phosphonate (alendronate (AHD); Supplementary Table 2), a compound that mimics the diphosphate group of GGDP. The CotB2^{wt}•Mg$^{2+}_3$•AHD structure has the same overall structure and active site architecture as described for CotB2^{wt}•Mg$^{2+}_3$•F-Dola, but crystallised in a different space group and clearly demonstrates that the folded C-terminus is not a crystallographic artefact caused by crystal contacts (Supplementary Fig. 4).

Importance of C-terminal region. In order to biochemically confirm the importance of the entire C-terminal region for catalysis, we generated a C-terminal truncation of CotB2 terminating at R294 (CotB2$^{\Delta C}$). CotB2$^{\Delta C}$ lacks the last 12 C-terminal residues, corresponding to the lid as observed in the structure of CotB2^{wt}•Mg$^{2+}_3$•F-Dola and CotB2^{wt}•Mg$^{2+}_3$•AHD. Indeed, CotB2$^{\Delta C}$ was an inactive mutant, not capable of catalysis as shown by GC/MS (Fig. 1e). Co-crystallisation of CotB2$^{\Delta C}$ with FGGDP resulted in the open-state conformation with no substrate bound (Supplementary Table 2). Additionally, this data provides further evidence that the salt-bridge between D111 and R294 is not sufficient to fully close the active site. Interestingly, there are two Mg^{2+}-ions bound in the CotB2$^{\Delta C}$ structure. One is fully occupied Mg$^{2+}_B$ and the other one is partially occupied Mg$^{2+}_C$, which can only be fully occupied if properly coordinated by the DDXD motif upon substrate binding and the accompanying conformational changes of the active site.

Closure of the active site seems to be an essential process in TPS catalysis. The implementation, however, appears to be very specific for each individual enzyme class. In several sesquiterpene synthases unstructured loop regions become ordered upon ligand binding, initiating active site closure[23,24]. In case of the plant-derived bornyl diphosphate synthase, a monoterpene synthase, the N-terminus together with a loop segment close the active site, and the enzyme was co-crystallised with the final product bornyl diphosphate[25]. By contrast, in the structure of the sesquiterpene aristolochene synthase from a fungus it is merely a loop segment that closes the active site[26]. Interestingly, molecular dynamic (MD) simulations of the closed state of the plant derived taxadiene synthase indicated that the N-terminus might be involved in closure of the active site[27,28]. Therefore, the mechanism of active site closure may be very specific to the reaction type and taxonomic origin of the respective TPS.

Given the importance of the C-terminus of CotB2 for catalysis, we screened the PDB for other diterpene synthases for which structural information for the open and closed conformation is available and which are structurally most related according to a DALI search[29] (Supplementary Table 3). In the closed structure of the labdane-related diterpene synthase (PDB-ID 5A0K) its C-terminus is folded as well. The RY-pair (Supplementary Fig. 5) is at the same position and engaging identical interactions with the diphosphate. Analysing protein sequences of other bacterial diterpene synthases for the presence of the RY-pair[22] and their flanking regions, we found a conserved tryptophan six amino acids upstream. Therefore, we suggest, that the WXXXXXRY ("X" any amino acid) motif is relevant for other diterpene synthases (Supplementary Fig. 5). If the tryptophan of latter motif is exchanged in CotB2 for glycine (W288G), the product becomes 3,7,18-dolabellatriene (Supplementary Table 1 and Supplementary Fig. 2)[17]. In this study, we introduced the less stereochemically demanding W288F mutation, which reduced CotB2's activity, but did not change the product profile. Thus, the aromatic character of the amino acid at position 288 in CotB2 is important for product formation and any drastic exchange to a non-aromatic side chain does interfere with the propagation of the carbocation. In the bacterial sesquiterpene synthase pentalenene synthase the corresponding mutation W308F leds to a product mixture[30]. However, in the plant-derived epi-aristolochene synthase mutations of W273 resulted in total loss of enzymatic function[23]. Therefore, tryptophan residues at this strategic position within the active site are of general importance in TPS to guide product formation.
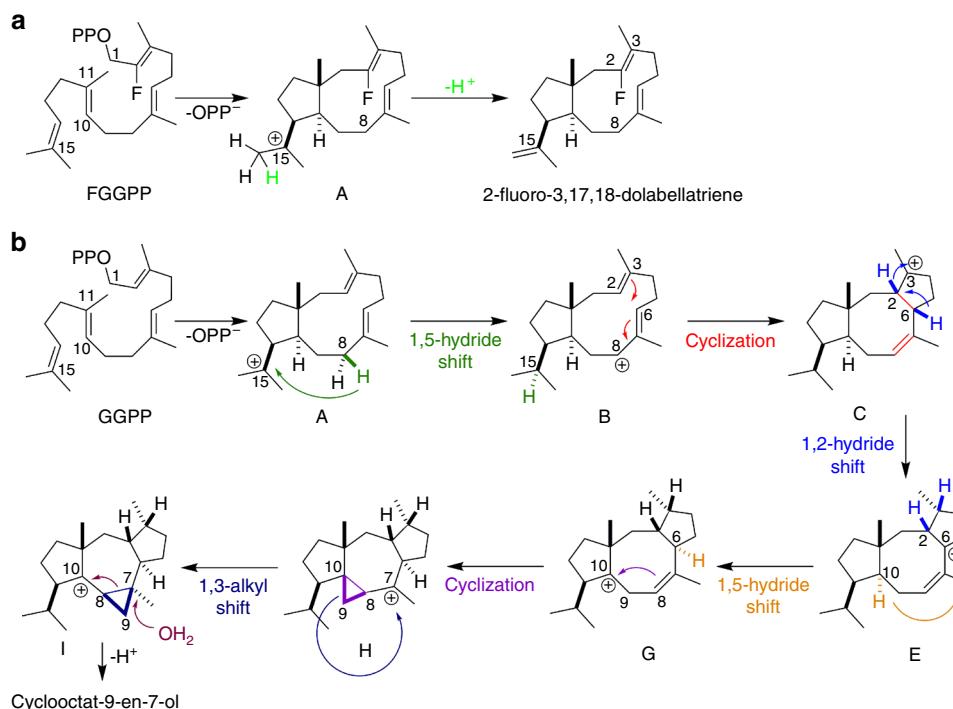
Pre-catalytic states of CotB2. Very recently, Tomita et al. published a co-crystal structure of CotB2 bound to the inert substrate-analogue geranylgeranyl thiodiphosphate (GGSDP) (CotB2^{wt}•Mg$^{2+}_B$•GGSDP, PDB-ID 5GUE[19]), which was described as representing the closed state of CotB2. Surprisingly, the diphosphate moiety of GGSDP is coordinated merely by a single magnesium-ion (Mg$^{2+}_B$; Fig. 2c), in contradiction to the accepted principle of a trinuclear magnesium cluster required for catalysis. Notably, beside the two missing Mg^{2+}, even though the salt-bridge between D111 of the DDXD motif and R294 is present (Supplementary Fig. 6b), the remaining C-terminus could not be resolved[19]. From our point of view this indicates that the structure of CotB2^{wt}•Mg$^{2+}_B$•GGSDP represents rather a pre-catalytic state of the enzyme reflecting a snapshot from the open to the fully closed conformation (Supplementay Discussion).
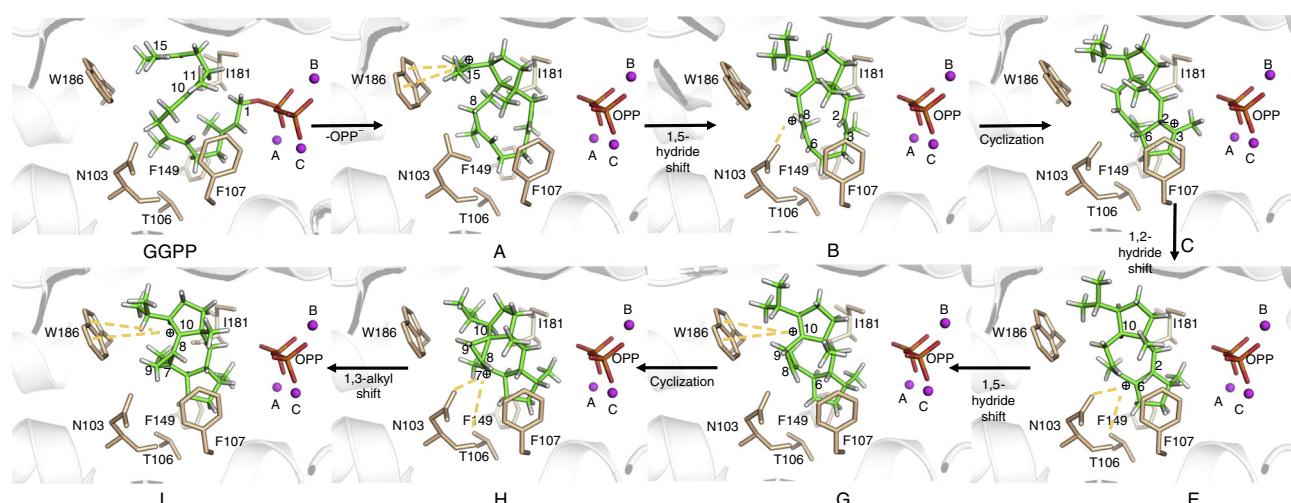
To comprehend the state of the CotB2^{wt}•Mg$^{2+}_B$•GGSDP structure, we determined the structure of CotB2^{F107A} under a Mg^{2+}-rich condition (CotB2^{F107A}•Mg$^{2+}_B$; Supplementary Table 2). Interestingly, a single magnesium-ion (Mg$^{2+}_B$) was bound to the active site, coordinated by the residues N220, S224 and E228 of the NSE motif and two water molecules (Fig. 2c and Supplementary Fig. 6a). It has been proposed for the sesquiterpene synthase aristolochene synthase, that Mg$^{2+}_B$ only binds the active site after the ligand has entered[31]. However, this does not seem to be the case for CotB2. Therefore, we suggest the binding of Mg$^{2+}_B$ to be the first step in preparing CotB2 for substrate binding.

The position of the Mg$^{2+}_B$ matches perfectly with that in the CotB2^{wt}•Mg$^{2+}_B$•GGSDP structure (Fig. 2c). Moreover, the conformation of the Mg$^{2+}_B$ coordinating amino acids are oriented in the same manner as there are two water molecules in our structure that mimic the oxygens of the diphosphate group in CotB2^{wt}•Mg$^{2+}_B$•GGSDP (Fig. 2b and Supplementary Fig. 6b). In the structure of CotB2^{F107A}•Mg$^{2+}_B$ the salt-bridge of D111 and R294 has not yet been established (Supplementary Fig. 6a) and the remaining C-terminus could not be resolved. This suggests, that the accurate positioning of the diphosphate moiety is required for the formation of the latter salt-bridge. The overall structure of CotB2^{wt}•Mg$^{2+}_B$•GGSDP is most similar to CotB2^{wt} and CotB2^{F107A}•Mg$^{2+}_B$ as indicated by a low root-mean-square deviation (Supplementary Table 4). Hence, the structures are likely different pre-catalytic states of CotB2.

CotB2 mechanism and structure-function dynamics. The in crystallo capture of the reaction intermediate F-Dola provides a static snapshot along the reaction coordinate in CotB2 (Fig. 3a). The mechanism in CotB2 has previously been delineated experimentally by Meguro et al. and Sato et al., and theoretically by Hong and Tantillo[32–34]. To explain mechanistic details of the entire catalytic cascade, we turned to in silico multiscale modelling. Modelling of the CotB2 reaction commenced with the bound 3,7,18-dolabellatriene (Figs. 3b, 4). Due to the high resolution of the crystal structure that captured CotB2 in a fully closed, catalytically competent state, QM/MM modelling yielded insight into the possible role of the enzyme in stabilising the reaction intermediates. We generated all-trans GGDP in a pre-folded

Fig. 3 Proposed mechanism of CotB2[wt]. **a** Conversion of FGGDP to F-Dola. **b** Proposed catalytic mechanism in CotB2, as obtained in the literature from isotope labelling experiments

Fig. 4 Mechanistic model of initial and intermediate states of CotB2. The figure shows mechanistic steps corresponding to those in Fig. 3b. The models were obtained from QM/MM simulations. The dashed yellow lines indicate interactions between the intermediate cation and CotB2

conformation, ready for catalysis, as well as all intermediate steps until the final carbocation. In the initial pre-folded GGDP state, the C1–C11 distance is 3.9 Å and the C10–C14 distance is 3.3 Å (Fig. 4). Following initial C–O cleavage and presumably concerted C1–C11 and C10–C14 bond formations, dolabellatrienyl cation (**A**) is formed, with the cation at C15 being stabilised by π-cation interactions with W186 (distance 4.6 Å; Fig. 4). The cation dislocates to C8 via a 1,5-hydride transfer reaction to yield cation **B**, as shown by deuterium labelling[32], with the cation being stabilised by an oxygen hole formed by N103 (OD1), N103 (O) and T106 (OG1) (Fig. 4).

The carbocation intermediate **B** then undergoes ring-closure via C2–C6 bond formation to yield cation **C**, with the cation located at C3 where it is stabilised by π-cation interactions, as it is

sandwiched between F107 and F149 (Fig. 4). Further stabilisation is likely provided by additional electrostatic interactions, as the cation migrates towards the diphosphate moiety during the **B→C** transformation step. Subsequently, a proposed 1,3-hydride transfer takes place to generate cation **E**. However, Hong and Tantillo[34] suggested that a series of two 1,2-hydride transfers would be energetically favoured in the gas-phase. Based on the modelled intermediate states in this study, there seems to be significant steric hindrance to a direct 1,3-hydride transfer. Therefore, two sequential 1,2-hydride transfers to provide intermediate **E** appear plausible (Fig. 3b). Identical two 1,2-hydride shifts were also experimentally reported for tsukubadiene synthase[35]. As suggested[34], these different mechanistic possibilities can be differentiated via isotope labelling as shown by Sato

et al[33]. In intermediate **E**, the cation is stabilised by F107 and the backbone carbonyl of I181 (distance 4.2 Å; Fig. 4). The following step involves another 1,5-hydride transfer to give carbocation **G**, as shown by deuterium labelling[32], with the cation located at C10, which is partly stabilised by N285 (distance 5.4 Å) and W186 (C10–W186 distance 5.3 Å; Fig. 4). A subsequent ring-closure generates a propyl-ring in intermediate **H**, while moving the cation to C7, which is stabilised by N103 (OD1), N103 (O) and T106 (OG1) (distances 3.9, 5.6, and 5.0 Å respectively; Fig. 4). The subsequent step entails a 1,3-alkyl shift to yield **I**, as established by ^{13}C isotope labelling[32], which moves the cation back to C10 (C10–N285 distance 4.7 Å, C10–W186 distance 5.8 Å), and shifts the cyclopropyl moiety along the central octa-cyclic ring (Fig. 4). Mutation of this residue to W186H results in formation of 3,7,18-dolabellatriene[19]. This is conceptually very similar to the taxadiene synthase W753H mutation, which yields $(-)-(R)$-cembrene A[27]. The approximate energetic contribution of the above-discussed interactions between active site moieties and the carbocation intermediates were quantified using density-functional theory cluster calculations (Supplementary Tables 6 and 7). Inspection of these interaction energies clearly show that the enzyme stabilises the evolving carbocations, with values between ca. −10 and −20 kcal/mol. The reaction concludes with attack by a water molecule at position C7, followed by deprotonation, to give the final product cyclocotat-9-en-7-ol. A potential hydroxylating water molecule was not observed in the crystal structure. But based on the current mechanistic model, it could be located between N103 (OD1), N103 (O) and T106 (OG1). The role of Asn in binding hydroxylating water molecules has been stressed in the context of 1,8-cineole synthase[36]. The above mechanistic model suggests that the CotB2 active site architecture has evolved to chaperone the changing cation along the entire reaction coordinate. Future QM/MM calculations will address the complete free energy surface for the CotB2 catalysed reaction. We note that the role of the enzyme environment in guiding the carbocations along the reaction coordinate is in line with our earlier work[37–39] and that of Peters and co-workers[40,41].

Discussion

Co-crystallising CotB2 with FGGDP enabled us to capture an intermediate state of the reaction mechanism. In combination with structural and biochemical data we could point out the importance of the properly folded C-terminus for the enzymatic activity. Our findings allowed us to propose a complete catalytic mechanism of CotB2 at a structural level by combining crystallography, biochemical tools and theoretical modelling (Supplementary Fig. 7). The enzyme mechanistic cascade commences with Mg^{2+}_B binding to the NSE motif, although this binding event is not accompanied by any conformational changes of the main structural core of CotB2. Subsequently, the substrate binds with its diphosphate moiety to Mg^{2+}_B. This binding event induces the formation of the salt-bridge between D111 and R294. Binding of the ions Mg^{2+}_A and Mg^{2+}_C to the diphosphate moiety of the substrate causes the largest conformational change in the enzyme, including rigid helix motion, which reduces the active site volume, and assures proper coordination of the substrate. At this point, the complete C-terminus can fold over the active site, thereby establishing a hydrogen network between the two catalytic motifs, and shields the active site from bulk solvent. Once all the pre-catalytic binding events and conformational changes are established, GGDP is correctly positioned to subsequently undergo cyclisation, which in the case of FGGDP results in premature quenching of the cyclisation reaction, as shown by the co-crystallisation of CotB2 with FGGDP. The closure of the C-terminus is the final trigger that initiates the chemical reaction cascade in CotB2, and possibly in additional class I TPSs.

Using theoretical modelling, we modelled the reaction using the natural substrate, GGDP. Importantly, based on in silico multiscale modelling we conclude that the three-dimensional active site architecture is such that the evolving carbocation is stabilised by precisely positioned amino acids and the bound diphosphate at each stage of the reaction cascade. Product release is possibly facilitated by the opening/unfolding of the C-terminus, which is entropically favourable, and might be the first stage in active site opening and release of the final product.

In summary, the cumulative experimental data, in conjunction with computational simulations, clearly demonstrate an active role of the protein scaffold in guiding the carbocation driven reaction towards a definitive end product. In line with our direct crystal structure evidence, there is mounting indirect evidence that the enzyme plays a crucial role in: binding the substrate, folding the substrate, selectively stabilising highly reactive carbocations to guide the reaction cascade towards the desired product and finally deprotonate or hydroxylate the final product by careful positioning of active site moieties. Therefore, the data delineated from the abrupt product trapped in the closed CotB2 structure provide fundamental advance in the understanding of TPS structural dynamics during catalysis.

Moreover, we identified universal structural features, which can be exploited for rational engineering of other bacterial TPS towards tailor-made terpene macrocycles. This information will pave the way for future semi-synthetic drug development strategies. The proof of concept for this strategy has already been well documented[42]. With respect to CotB2, site-directed mutagenesis of plasticity residues in the active site altered the enzymes product portfolio, which strongly argues for an active involvement of the protein scaffold in guiding the catalytic transformation of carbocations into distinct products[17]. The CotB2 mutation F107A afforded cembrene A as a product. The cembrene-type diterpenes display diverse bioactivities, including cytotoxic, insect deterrent and antimicrobial. Relevant work on insecticidal cembrenes has recently been reported[43]. Similarly, the single mutation of W288G alters to product to $(1R,3E,7E,11S,12S)-3,7,18$-dolabellatriene. Its structure has been reported to have potential antibiotic activity against multidrug-resistant *Staphylococcus aureus*. However, most of the targeted mutagenesis strategies were designed based on the open, non-catalytically active CotB2 structure. Therefore, the catalytic relevance of targeted amino acids in the active site could not be predicted, which often leads to generation of non-productive mutants. Moreover, one could not predict the structure of the cyclisation product. On the basis of our current model with a catalytically relevant intermediate in the active site it is feasible to trace and predict productive mutations. Furthermore, future expansion of our data to other TPSs may also enable prediction of the resulting terpene cyclisation product when targeting a specific amino acid residue(s).

Methods

proteolytic step after Ni^{2+}-NTA affinity chromatography. The amino-terminal His_6-tag of CotB2$^{\Delta Cter}$ was cleaved overnight by tobacco etch virus protease while dialysing in buffer B.

Crystallisation. CotB2^{F107A} was concentrated to 28 mg/ml as measured by the absorbance at 280 nm. Crystals were obtained by the sitting-drop vapour-diffusion method at 18 °C with a reservoir solution composed of 17.5% (v/v) polyethylene glycol 4000, 100 mM Tris/HCl at pH 8.5 and 100 mM $MgCl_2$. For co-crystallisation experiments, CotB2^{wt} was concentrated to 20 mg/ml and incubated in a fivefold molar excess of FGGDP for 30 min on ice. Crystals were obtained by the sitting-drop vapour-diffusion method at 18 °C with a reservoir solution composed of 30% (v/v) polyethylene glycol 400, 100 mM HEPES/NaOH at pH 7.5 and 200 mM $MgCl_2$. CotB2$^{\Delta Cter}$ was concentrated to 27.9 mg/ml and incubated in a 15-fold molar excess of FGGDP for 30 min on ice. Crystals were obtained by the sitting-drop vapour-diffusion method at 18 °C with a reservoir solution composed of 16% (v/v) polyethylene glycol 4000, 100 mM Tris/HCl at pH 8.5 and 100 mM $MgCl_2$. For co-crystallisation experiments with alendronate (Alfa Aesar, Germany), CotB2^{wt} was incubated in a twofold molar excess for 30 min on ice. Crystals were obtained by the sitting-drop vapour-diffusion method at 18 °C with a reservoir solution composed of 26% (v/v) polyethylene glycol 4000, 100 mM Tris/HCl at pH 8.5 and 150 mM $MgCl_2$. All crystals were cryo-protected with 25% (v/v) 2-methyl-2,4-pentanediol supplemented to the reservoir resolution and subsequently flash-cooled in liquid nitrogen.

Structure determination and refinement. Synchrotron diffraction data were collected at the beamline 14.2 of the MX Joint Berlin laboratory at BESSY II (Berlin, Germany) and beamline P14 or beamline P11 of PETRA III (Deutsches Elektronen Synchrotron, Hamburg, Germany) at 100 K and wavelengths as followed: 0.9184 Å (CotB2^{wt}•Mg^{2+}_3•F-Dola and CotB2^{wt}•Mg^{2+}_3•AHD), 1.0332 Å (CotB2$^{\Delta C}$•Mg^{2+}_B) and 0.9763 Å (CotB2^{F107A}•Mg^{2+}_B). Diffraction data were processed with XDS[44] (Supplementary Table 2). The structures were determined by molecular replacement with the coordinates of CotB2^{wt} (PDB-ID: 4OMG[17]) as search model using PHASER[45]. The structure was refined by maximum-likelihood restrained refinement in PHENIX[46,47]. Model building and water picking was performed with COOT. Geometrical restraints, used in the refinement of the fluorinated intermediate, were generated by using the PRODRG Web Server or ELBOW[48]. Model quality was evaluated with MolProbity[49] and the JCSG validation server (JCSG Quality Control Check v3.1). Secondary structure elements were assigned with DSSP[50]. Figures were prepared using PyMOL[51]. The Ramachandran plot shows that 98.64% (CotB2^{wt}•Mg^{2+}_3•F-Dola), 98.43% (CotB2^{wt}•Mg^{2+}_3•AHD), 98.73% (CotB2$^{\Delta C}$•Mg^{2+}_B) and 99.64% (CotB2^{F107A}•Mg^{2+}_B) of all residues are in favoured regions. The MolProbity Clashscores are 5.48 (CotB2^{wt}•Mg^{2+}_3•F-Dola), 4.57 (CotB2^{wt}•Mg^{2+}_3•AHD), 3.03 (CotB2$^{\Delta C}$•Mg^{2+}_B) and 2.87 (CotB2^{F107A}•Mg^{2+}_B).

Computational modelling. The modelling commenced with construction of cation **A** based on 2-F-3,7,18-dolabellatriene trapped during crystallisation of FGGDP in CotB2 (monomer A of PDB-ID 6GGI). The substrate GGDP was then constructed by performing the reverse reaction in silico[16]. Hydrogen atoms were added using the HBUILD facility of CHARMM. Subsequently, the enzyme was soaked in a pre-equilibrated TIP3P water[52] box of size about $80 \times 80 \times 80$ Å^3, and the effect of ionic buffer environment and system charge neutralisation was accounted for by addition of 29 Na^+ and 20 Cl^- ions[53]. Subsequently, we relaxed the system by performing step-wise energy minimisation, followed by MD simulations. The MD simulations entailed a 25 ps heating period up to 298 K followed by 5 ns of equilibration. During the first 1 ns of the equilibration simulations, weak harmonic restraints (1 kcal/mol Å^2) were imposed on the backbone C_α atoms (residues 24–290), as well as nuclear Overhauser effect (NOE) restraints on selected hydrogen bonds in the protein and between the protein and cofactor diphosphate and Mg^{2+} ions. The MD heating was performed in the NVT ensemble, while equilibration was performed in the NPT ensemble[54]. Long-range electrostatics were treated via particle-mesh Ewald summation.

Following MD equilibration, we performed hybrid QM/MM energy minimisations on the complete system, with the QM region defined as GGDP and the three Mg^{2+} ions. Employing a combination of NOE restraints and general distance restraints, we generated the following cations along the reaction pathway[34]: **A, B, C, E, G, H** and **I** (Figs. 3b, 4). All MM calculations used the CHARMM protein[34] and nucleic acid force fields[55], in conjunction with parameters for GGDP[56], while QM/MM calculations[57] employed the SCCDFTB method[58] for the QM region. All simulations used the CHARMM simulations programme [59].

Synthesis of isprenoid ligand. Detailed experimental procedures are provided in the Supplementary Methods. Intermediate reactions steps and the final product were characterised by ^{1}H-NMR (Supplementary Table 8A), ^{13}C-NMR (Supplementary Table 8B), ^{19}F-NMR (Supplementary Table 9A) and ^{31}P-NMR (Supplementary Table 9B). The precise molecular weight was confirmed by mass spectrometry (Supplementary Table 10).

References

1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **79**, 629–661 (2016).
2. Brück, T., Kourist, R. & Loll, B. Production of macrocyclic sesqui- and diterpenes in heterologous microbial hosts: a systems approach to harness nature's molecular diversity. *ChemCatChem* **6**, 1142–1165 (2014).
3. Christianson, D. W. Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* **117**, 11570–11648 (2017).
4. Pronin, S. V. & Shenvi, R. A. Synthesis of highly strained terpenes by non-stop tail-to-head polycyclization. *Nat. Chem.* **4**, 915–920 (2012).
5. Zhang, Q. & Tiefenbacher, K. Terpene cyclization catalysed inside a self-assembled cavity. *Nat. Chem.* **7**, 197–202 (2015).
6. Bartels, F. et al. Bioinspired synthesis of pentacyclic onocerane triterpenoids. *Chem. Sci.* **8**, 8285–8290 (2017).
7. McCulley, C. H., Geier, M. J., Hudson, B. M., Gagne, M. R. & Tantillo, D. J. Biomimetic platinum-promoted polyene polycyclizations: influence of alkene substitution and pre-cyclization conformations. *J. Am. Chem. Soc.* **139**, 11158–11164 (2017).
8. Brill, Z. G., Condakes, M. L., Ting, C. P. & Maimone, T. J. Navigating the chiral pool in the total synthesis of complex terpene natural products. *Chem. Rev.* **117**, 11753–11795 (2017).
9. Zhang, Q., Rinkel, J., Goldfuss, B., Dickschat, J. S. & Tiefenbacher, K. Sesquiterpene cyclizations catalysed inside the resorcinarene capsule and application in the short synthesis of isolongifolene and isolongifolenone. *Nat. Catal.* **1**, 609–615 (2018).
10. Tantillo, D. J. Importance of inherent substrate reactivity in enzyme-promoted carbocation cyclization/rearrangements. *Angew. Chem. Int. Ed. Engl.* **56**, 10040–10045 (2017).
11. Schönherr, H. & Cernak, T. Profound methyl effects in drug discovery and a call for new C–H methylation reactions. *Angew. Chem. Int. Ed. Engl.* **52**, 12256–12267 (2013).
12. Kim, S. Y. et al. Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two p450 hydroxylases. *Chem. Biol.* **16**, 736–743 (2009).
13. Aoyagi, T. et al. Cyclooctatin, a new inhibitor of lysophospholipase, produced by Streptomyces melanosporofaciens MI614-43F2. Taxonomy, production, isolation, physico-chemical properties and biological activities. *J. Antibiot.* **45**, 1587–1591 (1992).
14. Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106**, 3412–3442 (2006).
15. Aaron, J. A. & Christianson, D. W. Trinuclear metal clusters in catalysis by terpenoid synthases. *Pure Appl. Chem.* **82**, 1585–1597 (2010).
16. Major, D. T., Freud, Y. & Weitman, M. Catalytic control in terpenoid cyclases: multiscale modeling of thermodynamic, kinetic, and dynamic effects. *Curr. Opin. Struct. Biol.* **21**, 25–33 (2014).
17. Janke, R., Görner, C., Hirte, M., Brück, T. & Loll, B. The first structure of a bacterial diterpene cyclase: CotB2. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1528–1537 (2014).
18. Görner, C., Hauslein, I., Schrepfer, P., Eisenreich, W. & Brück, T. Targeted engineering of cyclooct-9-en-7-ol synthase: a stereospecific access to two new non-natural fusicoccane-type diterpenes. *ChemCatChem* **5**, 3289–3298 (2013).
19. Tomita, T. et al. Structural Insights into the CotB2-catalyzed cyclization of geranylgeranyl diphosphate to the diterpene cyclooct-9-en-7-ol. *Acs. Chem. Biol.* **12**, 1621–1628 (2017).
20. Miller, D. J., Yu, F. & Allemann, R. K. Aristolochene synthase-catalyzed cyclization of 2-fluorofarnesyl-diphosphate to 2-fluorogermacrene A. *ChemBioChem* **8**, 1819–1825 (2007).
21. 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).
22. Rabe, P., Schmitz, T. & Dickschat, J. S. Mechanistic investigations on six bacterial terpene cyclases. *Beilstein J. Org. Chem.* **12**, 1839–1850 (2016).
23. 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).

24. Harris, G. G. et al. Structural Studies of geosmin synthase, a bifunctional sesquiterpene synthase with aa domain architecture that catalyzes a unique cyclization-fragmentation reaction sequence. *Biochemistry* **54**, 7142–7155 (2015).

25. 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).

26. Chen, M. et al. Mechanistic insights from the binding of substrate and carbocation intermediate analogues to aristolochene synthase. *Biochemistry* **52**, 5441–5453 (2013).

27. Schrepfer, P. et al. Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. *Proc. Natl Acad. Sci. USA* **113**, E958–E967 (2016).

28. Freud, Y., Ansbacher, T. & Major, D. T. Catalytic control in the facile proton transfer in taxadiene synthase. *ACS Catal.* **7**, 7653–7657 (2017).

29. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).

30. Seemann, M. et al. Pentalenene synthase. Analysis of active site residues by site-directed mutagenesis. *J. Am. Chem. Soc.* **124**, 7681–7689 (2002).

31. Shishova, E. Y. et al. X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal ion binding sequence for catalysis. *J. Biol. Chem.* **283**, 15431–15439 (2008).

32. Meguro, A. et al. An unusual terpene cyclization mechanism involving a carbon–carbon bond rearrangement. *Angew. Chem. Int. Ed.* **54**, 4353–4356 (2015).

33. Sato, H. et al. "Cation-Stitching Cascade": exquisite control of terpene cyclization in cyclooctatin biosynthesis. *Sci. Rep.* **5**, 18471 (2015).

34. Hong, Y. J. & Tantillo, D. J. The energetic viability of an unexpected skeletal rearrangement in cyclooctatin biosynthesis. *Org. Biomol. Chem.* **13**, 10273–10278 (2015).

35. Rabe, P. et al. Mechanistic investigations of two bacterial diterpene cyclases: spiroviolene synthase and tsukubadiene synthase. *Angew. Chem. Int. Ed. Engl.* **56**, 2776–2779 (2017).

36. Karuppiah, V. et al. Structural basis of catalysis in the bacterial monoterpene synthases linalool synthase and 1,8-cineole synthase. *ACS Catal.* **7**, 6268–6282 (2017).

37. Weitman, M. & Major, D. T. Challenges posed to bornyl diphosphate synthase: diverging reaction mechanisms in monoterpenes. *J. Am. Chem. Soc.* **132**, 6349–6360 (2010).

38. Major, D. T. & Weitman, M. Electrostatically guided dynamics—the root of fidelity in a promiscuous terpene synthase? *J. Am. Chem. Soc.* **134**, 19454–19462 (2012).

39. Major, D. T. Electrostatic control of chemistry in terpene cyclases. *ACS Catal.* **7**, 5461–5465 (2017).

40. Xu, M., Wilderman, P. R. & Peters, R. J. Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proc. Natl Acad. Sci. USA* **104**, 7397–7401 (2007).

41. Zhou, K. & Peters, R. J. Electrostatic effects on (di)terpene synthase product outcome. *Chem. Commun.* **47**, 4074–4080 (2011).

42. Chang, M. C. & Keasling, J. D. Production of isoprenoid pharmaceuticals by engineered microbes. *Nat. Chem. Biol.* **2**, 674–681 (2006).

43. Mischko, W. et al. Modular biomanufacturing for a sustainable production of terpenoid-based insect deterrents. *Green. Chem.* **20**, 2637–2650 (2018).

44. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010). https://doi.org/10.1107/S0907444909047337

45. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

46. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

47. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).

48. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009).

49. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

50. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

51. DeLano, W. The PyMOL Molecular Graphics System http://www.pymol.org. (2002).

52. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

53. Beglov, D. & Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063 (1994).

54. Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. Constant-pressure molecular-dynamics simulation—the Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).

55. Foloppe, N. & MacKerell, A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21**, 86–104 (2000).

56. Kamp, M. Wvd, Sirirak, J., Żurek, J., Allemann, R. K. & Mulholland, A. J. Conformational change and ligand binding in the aristolochene synthase catalytic cycle. *Biochemistry* **52**, 8094–-88105 (2013).

57. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions—dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).

58. Gaus, M., Cui, Q. & Elstner, M. DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCCDFTB). *J. Chem. Theory Comput.* **7**, 931–948 (2011).

59. Brooks, B. R. et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

Acknowledgements

Author contributions

B.L. conceived the study. R.D. performed cloning, protein expression, purification, crystallisation, diffraction data collection and refinement. R.D. and B.L. analysed the structural data. M.F., with guidance from T.B., performed and analysed the activity assays. A.R.M. and D.T.M. performed MD calculations. S.J. and E.W. with guidance from M.C. conceived and performed organic synthesis. B.L., T.B., D.T.M. and M.C. supervised the project. R.D, B.L., T.B., D.T.M. and M.C. wrote the manuscript.

Additional information

Supplementary Information

Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis

Driller *et al.*

Supplementary Discussion

FGGDP co-crystallized with other TPSs

FGGDP has previously been co-crystallised with taxadiene synthase (TXS; PDB-ID 3P5R ref[1]), where FGGDP is bound in the active site with its diphosphate group still attached to the aliphatic tail. The reported TXS construct however was N-terminally truncated, stalling the enzyme in the catalytically inactive open conformation. Therefore, the mere binding of the substrate to the open TXS structure does not promote the complete closure of the active site. Consequently, FGGPP is not correctly positioned and cannot undergo cyclisation.

The non-canonical DDXD motif in comparison to other TPSs with the canonical DDXXD motif

The amino acid sequence of CotB2 shows an unconventional aspartate-rich ^{110}DDMD^{113} motif residing on helix D (Supplementary Fig. 11a). In the structure of CotB2^{wt}•Mg$^{2+}_3$•F-Dola D110 is directly involved in the coordination of Mg$^{2+}_A$ and Mg$^{2+}_C$, whereas D111 forms a salt-bridge with R294 of the RY-pair. D113 is surface exposed and points away from the active site (Supplementary Fig 11a and c). The difference in the protein sequence of the aspartate-rich motif has structural consequences (Supplementary Figure 11c). In the structure of epi-isozizaene synthase in complex with diphosphate, 3 Mg^{2+} ions and N-benzyl-N,N-diethylethanaminium,[2] with a canonical aspartate-rich motif ^{99}DDRHD^{103}, the α-helix carrying the motif is extended compared to CotB2 (Supplementary Fig. 11b). The first aspartate of the aspartate–rich motif resides at an identical position to coordinate Mg$^{2+}_A$ and Mg$^{2+}_C$. The second aspartate establishes as well a salt-bridge with the RY-pair. In contrast the third aspartate points towards the active site approaching water molecules in the water network around the Mg^{2+} ions.

Comparison of CotB2^{wt}•Mg$^{2+}_3$•F-Dola and CotB2^{wt}•Mg$^{2+}_B$•GGSDP

Tomita and colleagues stated they crystallised the fully closed structure of CotB2, the C-terminus ending with K296, hypothesising that the salt-bridge is the driving force of active site closure.[3] However, the remaining C-terminal residues have not been modelled due to their flexibility. In contrast, in

CotB2^{wt}•Mg$^{2+}_3$•F-Dola we observe the complete C-terminus, ending with the C-terminal Q307. This structural feature results in a higher rmsd between the open and closed structure (Supplementary Table 4), manifesting the importance of the C-terminus for proper closure of the enzyme. Since in CotB2^{wt}•Mg$^{2+}_B$•GGSDP two Mg^{2+}-ions are missing,[3] the diphosphate moiety is not properly coordinated (Fig. 2b), resulting in significant substrate flexibility within the catalytically active site. Moreover, in the reported structure the missing Mg^{2+} ions prevent helix D movement towards the active site thereby prohibiting folding over of the C-terminus. Interestingly, the hydrophobic tail of GGSDP adopts a position similar to the partly cyclised intermediate in our CotB2^{wt}•Mg$^{2+}_3$•F-Dola structure (Fig. 2b), as determined by the internal shape of the active site. In summary, our observations indicate that the crucial step of the initiation of the cyclisation reaction is the precise orientation and binding of the diphosphate moiety. The structure of CotB2^{wt}•Mg$^{2+}_B$•GGSDP on the other hand represents a pre-catalytic state of the enzyme reflecting a snapshot from the open to the fully closed conformation.

Supplementary Methods

Cloning. DNA manipulations and cloning procedures were performed according to standard protocols. The C-terminal deletion variant of CotB2 (CotB2$^{\Delta Cter}$) was designed by adding restriction sites *Nco*I at the 5'-end and *Hind*III at the 3'-end to the sequence (Supplementary Table 5; primer 1 and 2). The digested *Nco*I/*Hind*III fragment was ligated into the target-vector pETM-11 with a TEV-cleavable N-terminal hexa-histidine-tag. Single point mutations V80L, M189C, V150A, L281V and W288F of CotB2 were introduced by quick change site-directed mutagenesis (Supplementary Table 4; primer 3 to 12). Correctness of the amplified gene sequences was assessed by DNA sequencing.

Synthesis of (*2Z*,*6E*,*10E*)-2-Fluor-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraen-1-yl diphosphate (2-fluoro-geranylgeranyl diphosphate) tetrabutylammonium salt.

Materials were purchased from commercial suppliers and used without further purification. Synthesis was performed as described by Roe *et al.* [4] and some reactions were optimised.[5,6] The products of intermediate reaction steps were analysed by NMR and HRMS. The NMR spectra (Supplementary Figs. 8 and 9) and HRMS data (Supplementary Fig. 10) of the final product, 2-fluoro-geranylgeranyl diphosphate tetrabutylammonium salt, are presented below:

R$_f$ = 0.73 (*n*-Pentan)

^{1}H NMR (700 MHz, CD$_3$OD): δ = 5.15–5.03 (m, 3H, H-6, H-10, H-14), 4.65 (dd, J = 23.5, 5.3 Hz, 2H, H-1), 3.25–3.19 (m, 24H, 12xCH$_2$, NBu$_4$), 2.10–2.00 (m, 8H, H-4, H-5, H-9, H-13), 1.99–1.90 (m, 4H, H-8, H-12), 1.72 (d, J = 2.9 Hz, 3H, H-16), 1.69–1.60 (m, 24 H, 12xCH$_2$, NBu$_4$), 1.60–1.55 (m, 12H, H-17, H-18, H-19, H-20), 1.40 (tq, J = 14.8, 7.4 Hz, 24H, 12xCH$_2$, NBu$_4$), 1.00 (t, J = 7.4 Hz, 36H, 12xCH$_3$, NBu$_4$) ppm.

^{13}C NMR (175 MHz, CD$_3$OD): δ = 151.3 (d, J = 233.0 Hz, C-2), 135.1 (C-7 or C-11), 134.5 (C-7 or C-11), 130.7 (C-15), 124.1 (C-6 or C-10), 123.7 (C-6 or C-10), 117.2 (d, J = 15.8 Hz, C-3), 110.0 (C-14), 58.1 (d, J = 6.8 Hz, C-1), 39.5 (C-8 or C-12), 39.4 (C-8 or C-12), 29.7 (d, J = 7.1 Hz, C-4), 26.4 (C-5), 26.3 (C-9 orr C-13), 24.5 (C-9 or C-13), 23.4 (C-16), 19.3 (d, J = 3.5 Hz, C-17), 14.7 (C-17 or C-18 or C-19), 14.6 (C-17 or C-18 or C-19), 12.6 (C-17 or C-18 or C-19) ppm.

19F NMR (376 MHz, CD$_3$OD): δ = −118.7 (t, *J* = 23.9 Hz) ppm.

31P NMR (162 MHz, CD$_3$OD): δ = −6.8 (s, organic monophosphate), −8.9 (d, *J* = 19.6 Hz), −9.4 (d, *J* = 19.4 Hz) ppm.

HRMS (ESI-TOF, *m/z*): (C$_{20}$H$_{34}$FO$_7$P$_2^-$) [M-H]$^-$ = 467.1728 (calculated: 467.1768).

Bacterial strains, genes and vectors for diterpene production. Cloning, protein expression and production of cyclooct-9-en-7-ol was conducted as previously described.[7] In brief a pET Vector harbouring *cotB2* (GenBank: BAI44338.1) or its mutants and a pACYC vector harbouring polycistronic genes of *dxs*-pathway[8] were transformed to *E. coli* HMS174(DE3) (Merk Millipore). Cultivation was carried out in BasalR-Media with 3 % glycerol as carbon source. A pre-culture was diluted to an OD of 0.1 and protein expression induced with 0.5 mM IPTG. After cultivation at 28° C for three days, the products were extracted with ethyl acetate and analysed by GC/MS.

Analytics of diterpenes. Diterpenes as products of the CotB2 enzyme and its variants were analysed by a Trace GC Ultra with DSQII (Thermo Scientific). 1 µl sample was loaded by a TriPlus auto-sampler onto a SGE BPX5 column (30 m, I.D 0.25 mm, film 0.25 µm). Initial column temperature was set to 50 °C and maintained for 2.5 min before a temperature gradient at 10 °C*min^{-1} up to 320 °C was applied. The final temperature was kept for additional 3 min. MS data were recorded at 70 eV (EI) and m/z (rel. intensity in %) as total ion current (TIC). The recorded m/z range was in between 50 to 650.

Supplementary Table 1. CotB2[wt] and CotB2 variants as well as the altered products. The numbering of compounds is according to Supplementary Figure 2.

Variant	Product	Compound	Reference
wild-type	cyclooctat-9-en-7-ol	1	9
V80L	no product	-	this study
N103A	3,7,12-dolabellatriene	2	3
F107A	*R*-Cembrene A	3	7
F107Y	cyclooctat-1,7-diene	5	7
F107L	cyclooctat-9-en-7-ol	1	7
	3,7-Dolabelladiene-9-ol	5	3
	cyclooctat-6-en-8-ol	6	
F149L	cyclooctat-7-en-3-ol	7	7
V150A	cyclooctat-9-en-7-ol	1	this study
F185A	cyclooctat-9-en-7-ol	1	3
	cyclooctat-6-en-8-ol	6	
W186L	cyclooctat-9-en-7-ol	1	3
	cembrane A,	3	
	3,7,18-dolabellatriene	8	
W186F	cyclooctat-9-en-7-ol	1	3
	cyclooctat-6-en-8-ol	6	
	3,7-dolabelladiene-9-ol	5	
	cyclooctat-7-en-3-ol	7	
W186H	3,7,18-dolabellatriene	8	3
	cyclooctat-7-en-3-ol	7	
M189C	cyclooctat-9-en-7-ol	1	this study
L281V	cyclooctat-9-en-7 ol	1	this study
W288F	cyclooctat-9-en-7-ol	1	this study
W288G	3,7,18-dolabellatriene	8	10

Supplementary Table 2. Crystallographic data collection and refinement statistics.

	CotB2^{wt}•Mg$^{2+}_3$•F-Dola	CotB2^{wt}•Mg$^{2+}_3$•AHD	CotB2$^{\Delta C}$•Mg$^{2+}_B$	CotB2^{F107A}•Mg$^{2+}_B$
Data collection				
Space group	$P2_12_12_1$	$P1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, b, c (Å)	62.9 98.7 105.9	53.4 56.5 57.0	61.0 100.0 107.7	61.4 100.1 107.9
α, β, γ (°)	90.0 90.0 90.0	91.8 101.7 115.9	90.0 90.0 90.0	90.0 90.0 90.0
Resolution (Å)	44.74 - 1.80 (1.87 - 1.80)	20.00 - 2.10 (2.18 - 2.10)	50.00 - 2.15 (2.23 - 2.15)	47.50 – 1.90 (1.95 – 1.90)
R_{meas} (%)	11.4 (114.6)	14.0 (92.2)	33.3 (208.7)	12.2 (164.6)
$I / \sigma I$	14.4 (1.9)	7.6 (1.4)	6.2 (1.0)	12.2 (1.4)
Completeness (%)	99.6 (98.8)	99.1 (99.1)	99.8 (99.0)	99.9 (99.9)
Redundancy	6.6 (6.7)	3.1 (2.8)	6.6 (6.7)	7.3 (7.4)
Refinement				
Resolution (Å)	44.74 – 1.80	19.83 – 2.10	49.86 – 2.15	47.50 – 1.90
No. reflections	61142	33590	36480	53118
R_{work} / R_{free}	0.17 / 0.20	0.18 / 0.24	0.20 / 0.24	0.28 / 0.21
No. atoms				
Protein	2439 (A), 2484 (B)	2320 (A), 2438 (B)	2273 (A), 2271 (B)	2309 (A), 2297 (B)
Ligand/ion	21 (EXW), 9 (PPV)	14 (AHD)	-	-
Water	283	223	197	246
B-factors				
Protein	24.4 (A), 25.8 (B)	36.3 (A), 33.7 (B)	32.9 (A), 33.4 (B)	32.2 (A), 32.7 (B)
Ligand/ion	28.0 (EXW), 17.8 (PPV)	30.90 (AHD)	-	-

Water	31.1	33.0	34.4	37.3
R.m.s. deviations				
Bond lengths (Å)	0.012	0.007	0.007	0.007
Bond angles (°)	1.291	0.845	0.806	0.776

*Values in parentheses are for highest-resolution shell.

Supplementary Table 3. Results of a DALI search[11] with the structure of CotB2^{wt}•Mg$^{2+}_3$•F-Dola.

PDB ID	rmsd [Å]	sequence identity [%]	Z-score	protein	TPS family	Ligand in active site	resolution [Å]	Lit.
5gue	0.8	100	46.5	cyclooctat-9-en-7-ol synthase (CotB2)	diterpene	GGSDP, 1 Mg^{2+}	1.80	[12]
2oa6	3.4	15	20.3	aristolochene synthase	sesquiterpene	PP$_i$, 3 Mg^{2+}	2.15	[13]
3kb9	3.2	11	18.8	epi-isozizaene synthase	sesquiterpene	PP$_i$, 3 Mg^{2+}, N-benzyl-N,N-diethylethanaminium	1.60	[2]
4okm	3.4	15	18.4	selinadiene synthase	sesquiterpene	PP$_i$, 3 Mg^{2+}	2.10	[14]
5a0i	3.4	11	18.0	labdane-related diterpene synthase (LrdC)	diterpene	PP$_i$, 2 Mg^{2+}	2.57	[15]
5dz2	3.3	12	17.8	germacradienol/ geosmin synthase	sesquiterpene	alendronate, 3 Mg^{2+}	2.11	[16]
5erm	3.0	17	17.2	fusicoccadiene synthase	diterpene	pamidronate, 3 Mg^{2+}	2.30	[17]
1jfa	3.8	13	17.1	trichodiene synthase	sesquiterpene	no ligand	2.50	[18]
4la6	3.6	10	16.0	2-methylisoborneol synthase	monoterpene	2-fluorolinalyl diphosphate, 2 Mg^{2+}	2.00	[19]
4xlx	3.6	10	15.7	*ent*-kaurene synthase (BjKS)	diterpene	no ligand	2.00	[20]
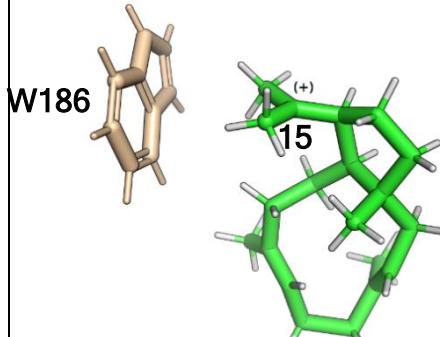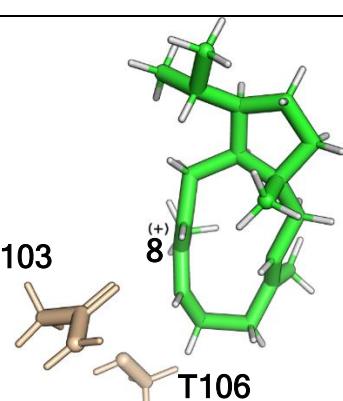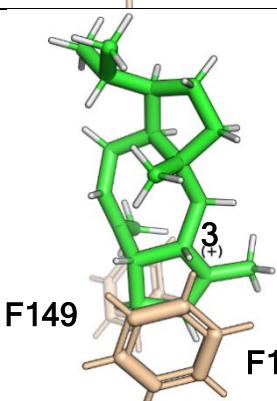
Supplementary Table 4. Root-mean-square deviation (rmsd) given in Å for superposition of different CotB2 structures.
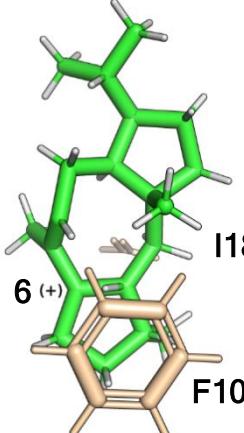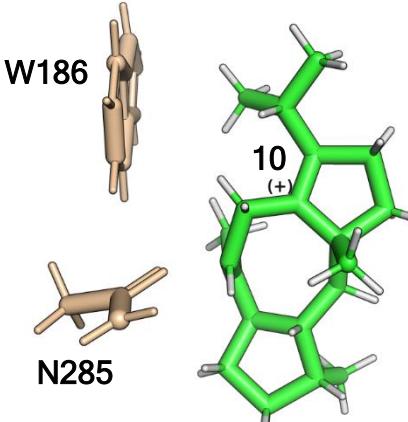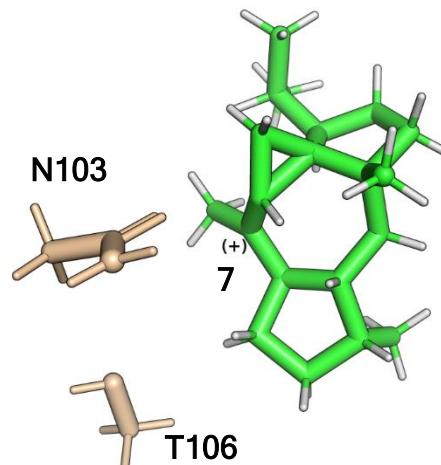
	CotB2^{wt}	CotB2^{wt}•Mg$^{2+}_3$•F-Dola	CotB2^{F107A}•Mg$^{2+}_B$	CotB2^{wt}•Mg$^{2+}_B$•GGSDP	CotB2$^{\Delta C}$
CotB2^{wt} (open)	-	0.7	0.5	0.3	0.4
CotB2^{wt}•Mg$^{2+}_3$•F-Dola (closed)	0.7	-	0.4	0.7	0.5
CotB2^{F107A}•Mg$^{2+}_B$	0.5	0.4	-	0.5	0.2
CotB2^{wt}•Mg$^{2+}_B$•GGSDP	0.3	0.7	0.5	-	0.4
CotB2$^{\Delta C}$	0.4	0.5	0.2	0.4	-
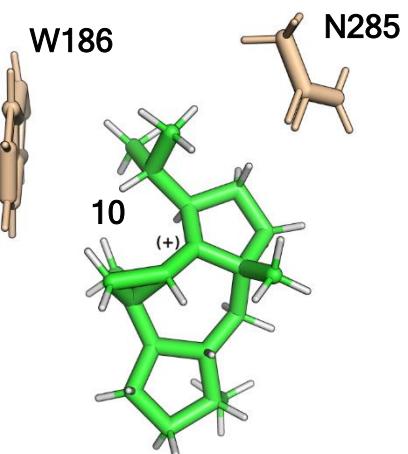
Supplementary Table 5. Oligonucleotides used for cloning.

number	description	oligonucleotide sequence (5' → 3')
1	fw Cter deletion	TATACCATGGCAGGCGCACAGGATATTG
2	rv Cter deletion	TATAAAGCTTTTAACGTTTGTTGGAGGTGGTC
3	fw V80L	GATTAGCTATGTTGGTTTAGTTCTGTGGTC
4	rv V80L	CAACATAGCTAATCCAACGTTCATCACTAAC
5	fw V150A	CAGCACGTGCATTTGCTACCAGCGATCAC
6	rv V150A	CAAATGCACGTGCTGCTTCATAGGCAACTTC
7	fw M189C	GATTTTTGGATGAAATGTAGCTATCCGATTTATC
8	rv M189C	CAAAAATCCACGCCAATATCGGTAAC
9	fw L281V	GATGTTTTTCTGGATGTGATTTATGGCAATTTTG
10	rv L281V	CAGAAAAACATCCTGTGTCAGCTGATC
11	fw W288F	CACCTCCAACAAACGTTATAAAAC
12	rv W288F	GTTGGAGGTGGTAAACACAAAATTG

Supplementary Table 6. Total and interaction energies[a] (kcal/mol) between CotB2 active site amino acid moieties and carbocation intermediates.

	systems	total energies (Hartree)	counterpoise corrected total energies (Hartree)	complexation energies (kcal/mol)
A		-1145.025566110	-1145.043565256	-11.3
B		-1106.162596787	-1106.191610934	-18.2
C		-1245.649659045	-1245.665518468	-10.0

E		-1127.952930905	-1127.978105188	-15.8
G		-1354.174380348	-1354.205374990	-19.4
H		-1106.197299196	-1106.217665068	-12.8

I		-1354.204624524	-1354.229503270	-15.6
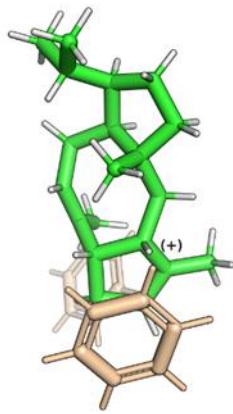
a The interaction energies were computed as the energy of the complex relative to the energy of the

individual molecular fragments. The structures were taken from the QM(SCCDFTB)/MM simulations

after energy minimization, as described in the Methods section. The interaction energies were obtained

from M06-2X/6-31+G(d,p) single point calculations, and include counter-poise correction to remove

basis-set superposition error.

Supplementary Table 7. Cartesian coordinates of complexes presented in Supplementary Table 6.
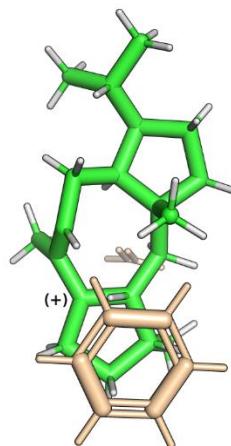
intermediate A				intermediate B			

atom	x	y	z	atom	x	y	z
C	5.229	-1.115	-4.849	C	-2.108	-0.451	-1.760
C	5.743	0.050	-5.367	H	-2.065	0.227	-2.606
H	6.606	0.572	-4.968	H	-3.082	-0.340	-1.291
N	5.051	0.385	-6.507	H	-2.033	-1.468	-2.139
H	5.313	1.017	-7.223	C	2.018	-3.588	-0.640
C	4.082	-0.539	-6.739	H	2.784	-3.008	-1.150
C	4.147	-1.503	-5.707	H	2.418	-4.578	-0.438
C	3.237	-2.558	-5.694	H	1.819	-3.103	0.314
H	3.241	-3.298	-4.908	C	-0.011	-3.243	2.384
C	2.309	-2.651	-6.733	H	-0.562	-3.558	3.270
H	1.612	-3.478	-6.745	H	0.288	-2.209	2.519
C	3.158	-0.628	-7.763	H	0.889	-3.861	2.330
H	3.148	0.096	-8.562	C	-1.092	-0.964	0.546
C	2.274	-1.706	-7.753	H	-0.209	-0.773	1.161
H	1.555	-1.813	-8.552	H	-1.945	-0.582	1.124
H	5.576	-1.641	-3.960	C	-1.301	-2.435	0.359
C	-1.895	-1.089	-1.862	H	-1.958	-2.724	-0.459
H	-1.713	-0.654	-2.840	C	-0.867	-3.431	1.169
H	-2.963	-1.034	-1.666	C	-1.241	-4.849	0.837
H	-1.623	-2.137	-1.906	H	-2.220	-4.892	0.358
C	2.537	-4.315	-1.353	H	-1.267	-5.470	1.735
H	3.072	-4.584	-2.265	C	-0.206	-5.563	-0.147
H	2.489	-5.203	-0.733	H	-0.617	-6.540	-0.379
H	3.130	-3.558	-0.833	H	0.752	-5.670	0.367
C	0.377	-3.144	2.063	C	-0.124	-4.745	-1.343
H	0.285	-3.677	3.010	H	-0.957	-4.804	-2.046
H	0.497	-2.085	2.278	C	0.763	-3.671	-1.463
H	1.291	-3.506	1.586	C	0.288	-2.590	-2.203
C	-1.434	-0.977	0.661	H	-0.639	-2.748	-2.754
H	-0.660	-0.664	1.367	H	0.209	0.490	-4.053
H	-2.377	-0.537	1.009	C	0.923	-1.284	-2.316
C	-1.590	-2.469	0.628	H	0.773	-0.988	-3.359
H	-2.441	-2.822	0.046	H	2.000	-1.380	-2.140
C	-0.816	-3.417	1.201	C	0.436	-0.088	-1.392
C	-1.103	-4.877	0.899	H	1.131	-0.110	-0.542
H	-2.145	-4.976	0.582	C	-0.998	-0.107	-0.756
H	-0.957	-5.496	1.792	C	-1.152	1.388	-0.363
C	-0.191	-5.414	-0.244	H	-0.566	1.590	0.537
H	-0.656	-6.299	-0.685	H	-2.189	1.645	-0.141
H	0.766	-5.728	0.172	C	-0.596	2.163	-1.548
C	0.008	-4.340	-1.272	H	-0.279	3.171	-1.270
H	-0.916	-3.856	-1.591	H	-1.372	2.271	-2.305
C	1.166	-3.796	-1.685	C	0.594	1.331	-2.073
C	1.134	-2.417	-2.325	H	1.515	1.768	-1.655
H	0.197	-2.259	-2.854	C	0.774	1.335	-3.626
H	1.939	-2.288	-3.058	C	2.258	1.158	-3.964
C	1.306	-1.421	-1.145	H	2.831	2.009	-3.601
H	2.341	-1.067	-1.140	H	2.400	1.091	-5.038
H	1.187	-1.983	-0.210	H	2.679	0.265	-3.511

	x	y	z
C	0.408	-0.210	-0.951
H	0.770	0.251	-0.018
C	-1.135	-0.335	-0.759
C	-1.564	1.153	-0.812
H	-1.276	1.656	0.116
H	-2.641	1.275	-0.928
C	-0.797	1.745	-1.998
H	-0.682	2.827	-1.907
H	-1.327	1.545	-2.932
C	0.600	1.086	-1.999
H	1.325	1.713	-1.458
C	1.201	0.684	-3.251
C	2.672	0.620	-3.337
H	3.167	1.146	-2.525
H	3.029	0.990	-4.299
H	2.979	-0.432	-3.293
C	0.447	0.292	-4.448
H	-0.510	-0.168	-4.208
H	1.023	-0.363	-5.099
H	0.219	1.193	-5.034

	x	y	z
C	0.229	2.568	-4.342
H	-0.849	2.643	-4.244
H	0.457	2.498	-5.400
H	0.677	3.483	-3.972
C	-0.028	-6.523	-5.529
H	-0.371	-7.334	-6.171
H	0.366	-6.937	-4.601
C	-1.198	-5.590	-5.215
O	-1.056	-4.597	-4.504
N	-2.377	-5.899	-5.802
H	-3.158	-5.278	-5.688
H	-2.451	-6.711	-6.371
H	0.757	-5.964	-6.039
C	1.137	-8.889	0.094
H	0.968	-8.161	0.888
O	1.171	-8.224	-1.159
H	1.323	-8.908	-1.836
H	2.087	-9.396	0.263
H	0.330	-9.621	0.096

intermediate C

intermediate E

atom	x	y	z
C	-1.357	-0.620	-2.309
H	-0.929	-0.142	-3.185
H	-2.377	-0.255	-2.195
H	-1.418	-1.687	-2.506
C	1.578	-3.587	0.203
H	2.546	-3.108	0.118
H	1.730	-4.661	0.336
H	1.085	-3.225	1.113
C	-1.874	-2.640	2.499
H	-1.841	-3.308	3.352
H	-2.865	-2.160	2.476
H	-1.182	-1.810	2.675
C	-0.855	-1.221	0.156
H	0.036	-1.293	0.788
H	-1.613	-0.719	0.767
C	-1.427	-2.670	-0.051
H	-2.424	-2.551	-0.511
C	-1.606	-3.328	1.239
C	-1.435	-4.790	1.161
H	-2.376	-5.279	1.447
H	-0.715	-5.123	1.923
C	-0.989	-5.116	-0.267
H	-1.771	-5.645	-0.805
H	-0.115	-5.766	-0.269
C	-0.688	-3.777	-0.961
H	-1.119	-3.753	-1.967
C	0.739	-3.321	-1.002

atom	x	y	z
C	-4.331	-5.963	-1.091
C	-4.893	-5.116	-0.118
H	-5.300	-5.538	0.792
C	-4.963	-3.735	-0.329
H	-5.395	-3.090	0.426
C	-4.500	-3.193	-1.532
H	-4.588	-2.132	-1.708
C	-3.854	-5.404	-2.286
H	-3.446	-6.045	-3.054
C	-3.942	-4.026	-2.510
H	-3.596	-3.614	-3.447
H	-4.267	-7.037	-0.919
C	3.224	-1.267	2.636
O	2.554	-1.372	1.612
H	2.798	-0.791	3.519
H	4.244	-1.650	2.661
C	-2.230	-0.604	-1.758
H	-2.426	0.112	-2.548
H	-3.007	-0.483	-1.006
H	-2.326	-1.599	-2.179
C	2.125	-3.077	-1.525
H	2.580	-2.552	-2.359
H	2.598	-4.057	-1.434
H	2.339	-2.528	-0.609
C	-0.610	-3.030	2.669
H	0.458	-3.177	2.525
H	-0.934	-3.639	3.508

C	1.045	-2.435	-1.986	H	-0.792	-1.986	2.915
H	0.377	-2.406	-2.845	C	-0.656	-1.253	0.156
H	1.173	0.164	-3.888	H	0.350	-1.055	0.542
C	1.924	-1.250	-1.827	H	-1.343	-0.864	0.915
H	2.386	-0.974	-2.774	C	-0.814	-2.765	0.190
H	2.729	-1.431	-1.111	H	-2.445	-3.149	1.562
C	1.024	-0.074	-1.274	C	-1.395	-3.441	1.420
H	1.400	0.114	-0.258	C	-1.283	-4.955	1.111
C	-0.533	-0.279	-1.051	H	-2.251	-5.330	0.786
C	-0.896	1.174	-0.630	H	-0.990	-5.528	1.989
H	-0.540	1.356	0.389	C	-0.263	-5.107	-0.024
H	-1.975	1.349	-0.634	H	-0.553	-5.852	-0.769
C	-0.156	2.062	-1.623	H	0.727	-5.403	0.341
H	0.080	3.040	-1.192	C	-0.156	-3.739	-0.612
H	-0.796	2.247	-2.483	H	-1.327	-3.306	-0.765
C	1.133	1.297	-2.020	C	0.643	-3.231	-1.748
H	1.987	1.831	-1.577	C	-0.085	-2.627	-2.713
C	1.431	1.181	-3.556	H	-1.136	-2.905	-2.766
C	2.935	1.385	-3.779	H	0.205	0.763	-4.626
H	3.233	2.389	-3.486	C	0.278	-1.329	-3.342
H	3.190	1.255	-4.824	H	-0.490	-1.039	-4.060
H	3.526	0.683	-3.195	H	1.236	-1.351	-3.863
C	0.641	2.136	-4.454	C	0.370	-0.332	-2.142
H	-0.426	1.940	-4.414	H	1.245	-0.656	-1.561
H	0.954	2.004	-5.486	C	-0.855	-0.339	-1.125
H	0.803	3.178	-4.189	C	-0.864	1.137	-0.629
C	-4.445	-5.976	-1.221	H	-0.082	1.277	0.123
C	-4.962	-5.066	-0.279	H	-1.818	1.409	-0.173
H	-5.347	-5.432	0.663	C	-0.538	1.952	-1.873
C	-5.015	-3.695	-0.564	H	-0.246	2.977	-1.632
H	-5.410	-3.004	0.168	H	-1.409	2.009	-2.529
C	-4.569	-3.225	-1.805	C	0.611	1.170	-2.527
H	-4.603	-2.169	-2.030	H	1.531	1.472	-2.002
C	-4.009	-5.489	-2.463	C	0.854	1.425	-4.039
H	-3.635	-6.179	-3.205	C	2.309	1.102	-4.385
C	-4.067	-4.122	-2.754	H	2.987	1.797	-3.892
H	-3.729	-3.766	-3.716	H	2.472	1.175	-5.455
H	-4.384	-7.039	-0.990	H	2.585	0.100	-4.067
C	5.253	-7.383	3.291	C	0.511	2.849	-4.458
C	5.069	-6.003	3.464	H	-0.541	3.053	-4.288
H	5.933	-5.356	3.509	H	0.706	2.984	-5.516
C	3.781	-5.467	3.606	H	1.106	3.574	-3.908
H	3.650	-4.406	3.762				
C	2.663	-6.308	3.564				
H	1.673	-5.893	3.689				
C	4.120	-8.214	3.220				
H	4.251	-9.276	3.078				
C	2.832	-7.684	3.369				
H	1.971	-8.336	3.339				
H	6.256	-7.804	3.213				
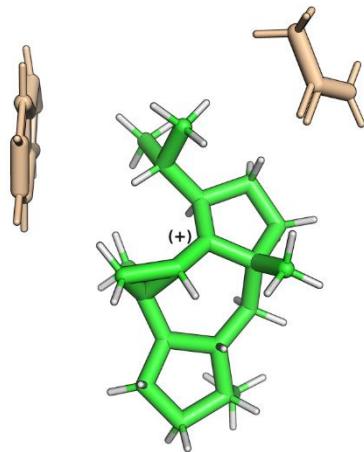
intermediate G	intermediate H

atom	x	y	z
C	-0.048	-6.537	-5.631
H	-0.391	-7.476	-6.066
H	0.439	-6.736	-4.676
C	-1.245	-5.609	-5.410
O	-1.113	-4.486	-4.928
N	-2.434	-6.060	-5.870
H	-3.228	-5.446	-5.829
H	-2.503	-6.967	-6.272
H	0.661	-6.061	-6.309
C	5.154	-1.230	-4.856
C	5.638	-0.141	-5.538
H	6.536	0.405	-5.269
N	4.845	0.098	-6.633
H	5.023	0.717	-7.388
C	3.831	-0.808	-6.669
C	3.980	-1.666	-5.555
C	3.047	-2.678	-5.332
H	3.127	-3.337	-4.480
C	1.984	-2.814	-6.227
H	1.241	-3.584	-6.061
C	2.784	-0.949	-7.561
H	2.694	-0.293	-8.411
C	1.860	-1.965	-7.322
H	1.030	-2.100	-7.995
H	5.580	-1.677	-3.957
C	-2.525	-1.196	-1.626
H	-2.949	-0.542	-2.384
H	-3.209	-1.235	-0.782
H	-2.425	-2.194	-2.046
C	2.357	-2.102	-1.528
H	2.599	-1.739	-2.525
H	3.208	-2.670	-1.154
H	2.245	-1.238	-0.872
C	-0.342	-3.169	2.696
H	0.737	-3.094	2.574
H	-0.538	-3.814	3.549
H	-0.730	-2.178	2.924
C	-0.699	-1.447	0.174
H	0.267	-1.042	0.486
H	-1.430	-1.124	0.921
C	-0.656	-2.980	0.152
H	-2.098	-3.733	1.608
C	-1.010	-3.744	1.453
C	-0.525	-5.184	1.128
H	-1.374	-5.810	0.851
H	-0.067	-5.650	2.004
C	0.475	-5.088	-0.047

atom	x	y	z
C	-0.044	-6.373	-5.594
H	-0.428	-7.259	-6.099
H	0.434	-6.666	-4.659
C	-1.197	-5.414	-5.299
O	-1.016	-4.338	-4.732
N	-2.406	-5.789	-5.776
H	-3.181	-5.157	-5.688
H	-2.512	-6.663	-6.238
H	0.686	-5.878	-6.235
C	1.161	-8.394	0.040
H	0.999	-7.639	0.809
O	1.131	-7.781	-1.239
H	1.279	-8.491	-1.886
H	2.131	-8.867	0.194
H	0.376	-9.148	0.101
C	-2.752	-0.326	-1.994
H	-2.884	0.465	-2.724
H	-3.571	-0.254	-1.284
H	-2.833	-1.281	-2.507
C	2.062	-2.839	-1.564
H	2.437	-2.598	-2.553
H	2.389	-3.845	-1.295
H	2.525	-2.165	-0.838
C	-1.324	-2.501	2.635
H	-0.235	-2.459	2.624
H	-1.632	-3.047	3.521
H	-1.702	-1.486	2.724
C	-1.420	-0.955	0.112
H	-0.561	-0.621	0.700
H	-2.312	-0.617	0.648
C	-1.436	-2.493	0.071
H	-2.966	-3.192	1.455
C	-1.870	-3.188	1.391
C	-1.333	-4.640	1.247
H	-2.150	-5.344	1.101
H	-0.810	-4.947	2.156
C	-0.381	-4.666	0.041
H	-0.871	-5.086	-0.840
H	0.526	-5.248	0.224
C	-0.062	-3.170	-0.198
H	-2.151	-2.826	-0.694
C	0.584	-2.776	-1.483
C	-0.269	-2.231	-2.446
H	-1.315	-2.508	-2.336
H	2.530	-0.575	-2.782
C	-0.063	-1.146	-3.441
H	-0.899	-1.029	-4.123

H	0.039	-5.506	-0.958
H	1.401	-5.631	0.150
C	0.732	-3.586	-0.209
H	-1.376	-3.337	-0.598
C	1.127	-2.959	-1.521
C	0.207	-2.938	-2.518
H	-0.651	-3.604	-2.460
H	0.273	0.436	-4.718
C	0.033	-1.691	-3.334
H	-0.817	-1.791	-4.024
H	0.894	-1.435	-3.958
C	-0.244	-0.604	-2.359
H	1.463	-3.276	0.548
C	-1.151	-0.659	-1.149
C	-1.289	0.842	-0.725
H	-0.643	1.042	0.134
H	-2.310	1.082	-0.425
C	-0.828	1.656	-1.924
H	-0.441	2.637	-1.649
H	-1.647	1.807	-2.629
C	0.263	0.766	-2.566
H	1.141	0.835	-1.887
C	0.757	1.123	-4.009
C	2.270	0.926	-4.105
H	2.796	1.642	-3.479
H	2.603	1.065	-5.128
H	2.562	-0.073	-3.786
C	0.368	2.539	-4.419
H	-0.710	2.658	-4.438
H	0.740	2.741	-5.417
H	0.792	3.273	-3.740

H	0.904	-1.056	-3.926
C	-0.166	-0.557	-2.077
H	0.632	-2.863	0.601
C	-1.415	-0.167	-1.266
C	-1.142	1.327	-0.938
H	-1.591	1.603	0.019
H	-1.601	1.958	-1.699
C	0.379	1.511	-0.936
H	0.753	1.754	0.057
H	0.661	2.345	-1.581
C	1.014	0.187	-1.441
H	1.346	-0.369	-0.551
C	2.268	0.396	-2.336
C	3.446	0.811	-1.454
H	3.211	1.707	-0.885
H	4.325	1.014	-2.062
H	3.687	0.022	-0.745
C	2.058	1.381	-3.489
H	1.129	1.189	-4.019
H	2.878	1.292	-4.196
H	2.034	2.409	-3.138

intermediate I
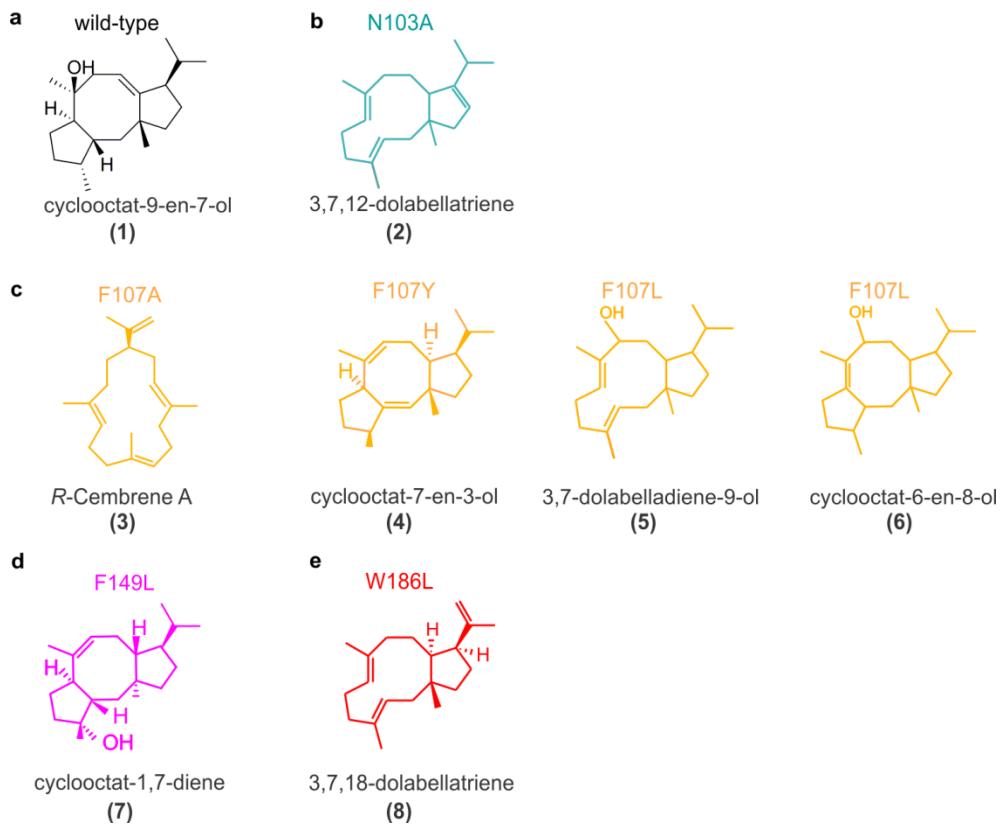
atom	x	y	z
C	5.122	-1.122	-4.851
C	5.613	-0.027	-5.518
H	6.530	0.492	-5.260
N	4.791	0.267	-6.578
H	4.967	0.887	-7.331
C	3.751	-0.610	-6.606
C	3.918	-1.510	-5.529
C	2.979	-2.520	-5.320
H	3.084	-3.220	-4.505
C	1.889	-2.605	-6.189
H	1.143	-3.377	-6.041
C	2.672	-0.695	-7.466
H	2.566	-0.002	-8.285
C	1.735	-1.703	-7.236
H	0.871	-1.786	-7.873
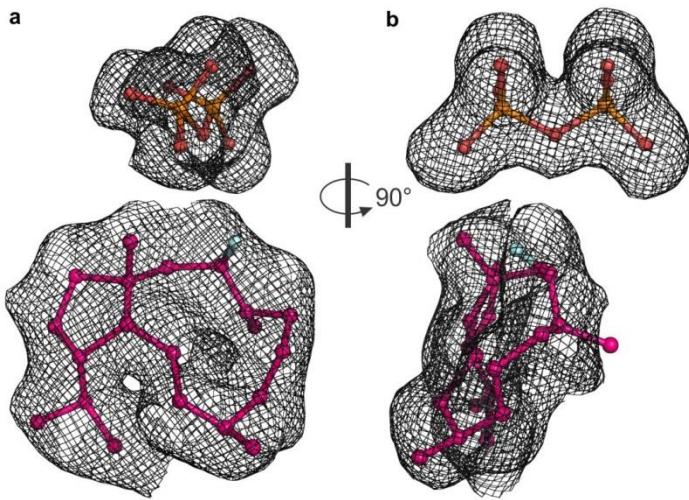
|			
H	5.561	-1.603	-3.977
C	-1.292	5.421	-4.699
H	-0.827	6.377	-4.457
H	-2.122	5.584	-5.386
C	-1.810	4.768	-3.419
O	-2.375	3.674	-3.455
N	-1.652	5.447	-2.262
H	-2.044	5.046	-1.439
H	-1.238	6.365	-2.273
H	-0.556	4.768	-5.168
C	-2.695	0.396	-0.789
H	-2.861	1.163	-1.542
H	-3.212	0.685	0.120
H	-3.141	-0.533	-1.137
C	1.867	-2.085	-1.562
H	2.221	-1.581	-2.459
H	2.452	-3.001	-1.454
H	2.086	-1.472	-0.691
C	-0.486	-2.916	2.753
H	0.547	-2.969	2.413
H	-0.616	-3.655	3.540
H	-0.655	-1.929	3.179
C	-0.943	-0.792	0.677
H	0.091	-0.688	1.017
H	-1.580	-0.464	1.504
C	-1.240	-2.272	0.386
H	-2.488	-3.025	2.007
C	-1.472	-3.184	1.621
C	-1.333	-4.621	1.044
H	-2.307	-5.105	0.983
H	-0.716	-5.243	1.697
C	-0.698	-4.497	-0.354
H	-1.455	-4.597	-1.136
H	0.064	-5.257	-0.538
C	-0.112	-3.078	-0.352
H	-2.165	-2.342	-0.203
C	0.417	-2.487	-1.646
C	-0.788	-1.444	-2.272
H	-1.750	-1.786	-1.899
H	1.169	-0.251	-3.886
C	-0.133	-2.563	-3.019
H	-0.845	-3.349	-3.238
H	0.561	-2.336	-3.821
C	-0.412	-0.215	-1.726
H	0.745	-3.081	0.336
C	-1.188	0.243	-0.506
C	-0.554	1.622	-0.149
H	0.104	1.513	0.714
H	-1.313	2.360	0.112
C	0.253	2.038	-1.376
H	1.100	2.677	-1.124
H	-0.378	2.582	-2.079
C	0.724	0.709	-2.004
H	1.561	0.358	-1.369
C	1.250	0.765	-3.471
C	2.728	1.160	-3.479
H	2.859	2.164	-3.086
H	3.116	1.144	-4.492
H	3.322	0.476	-2.875
C	0.432	1.704	-4.358
H	-0.633	1.551	-4.217
H	0.673	1.537	-5.403
H	0.651	2.745	-4.130
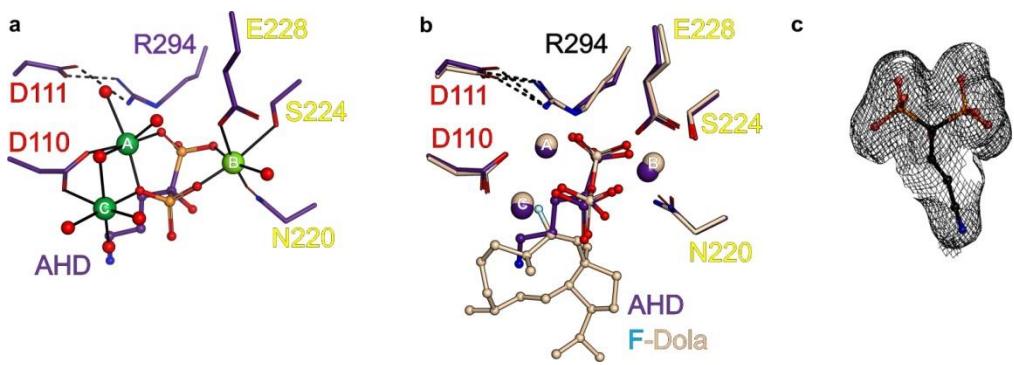
Supplementary Figure 1. Secondary structure elements and variants of CotB2. a, On top of the primary sequence of CotB2 the secondary structure elements are drawn. Slashed lines at the C-terminus indicate the terminal residue of the open state of CotB2. Missing residues have not been built due to their flexibility. The aspartate-rich motif (^{110}DDXD^{113}) is highlighted in red and the ^{218}NSE^{228} motif in yellow. CotB2 variants, as listed in Supplementary Table 1 are indicated by circles. For colour-coding see inlet. The structured C-terminus of CotB2 in the closed conformation is shown in purple. **b,** The open structure of CotB2^{wt} (PDB-ID 4OMG) is shown in gray cartoon representation. The ^{110}DDXD^{113} motif is shown in red and the ^{218}NSE^{228} motif in yellow. CotB2 variants are indicated by colored spheres. Color-coding of variants as in (a). **c,** View into the hydrophobic pocket of CotB2^{wt}. View rotated by 90 ° in respect to (b).

a wild-type

OH

H,,

H

cyclooctat-9-en-7-ol
(1)

b N103A

3,7,12-dolabellatriene
(2)

c F107A

R-Cembrene A
(3)

F107Y

H

H,,

cyclooctat-7-en-3-ol
(4)

F107L

OH

3,7-dolabelladiene-9-ol
(5)

F107L

OH

cyclooctat-6-en-8-ol
(6)

d F149L

H

H,

H

OH

cyclooctat-1,7-diene
(7)

e W186L

H

''H

3,7,18-dolabellatriene
(8)

Supplementary Figure 2. Overview of products of CotB2[wt] and its variants. Some colour coding of variants and their products as in Supplementary Table 1 and Supplementary Figure 1. **a,** cyclooctat-9-en-7-ol (1) produced by wild-type CotB2. **b,** Variant N103A producing 3,7,12-dolabellatriene (2) **c,** Variant F107A producing *R*-Cembrene A (3), F107Y producing cyclooctat-1,7,-diene (4), and F107L producing a product mixture composed of cyclooctat-9-en-7-ol (1), 3,7-dolabelladiene-9-ol (5), and cyclooctat-6-en-8-ol (6). **d,** Variant F149L producing cyclooctat-6-en-8-ol (7). **e,** Variant W186L producing cyclooctat-9-en-7-ol (2) and *R*-Cembrene A (3) in addition to 3,7,18-dolabellatriene (8).

Supplementary Figure 3. Polder electron density maps[21] shown as mesh at a σ-level of 2.0. Molecules are presented as ball-stick-models with carbon atoms colored in magenta, oxygen in red, phosphorous in orange and fluorine in light blue. **a,** Structure of CotB2^{wt}•Mg$^{2+}_3$•F-Dola with a polder electron density map around the F-Dola and diphosphate molecule. **b**, view in a, rotated by 90°. The polder electron density of the diphosphate and intermediate is clearly not connected, arguing for a cleaved diphosphate moiety.

Supplementary Figure 4. a, In the structure of CotB2^{wt}•Mg$^{2+}_3$•AHD, the bisphosphonate moiety of alendronate (AHD) is coordinated by three Mg^{2+}-ions. The strong binding of AHD to the active site is reflected in low B-factors of 24 Å^2. The entire C-terminus is observed and folds over the active site. Residues of the DDXD motif are labelled with red letters, residues of the NSE motif with yellow letters respectively. **b,** Superposition of CotB2^{wt}•Mg$^{2+}_3$•AHD and CotB2^{wt}•Mg$^{2+}_3$•F-Dola. The coordination of the Mg^{2+} ions is identical in both structures. **c,** Structure of CotB2^{wt}•Mg$^{2+}_3$•AHD with a polder electron density map around AHD molecule shown as mesh at a σ-level of 2.0. The AHD molecule is presented as ball-stick-model with carbon atoms colored in black, oxygen in red, phosphorous in orange and nitrogen in blue.

> clooctat-9-en-7-ol synthase (WP_093468823), *Streptomyces melanosporofaciens*, PDB-ID 6GGI

MTTGLSTAGAQDIGRSSVRPYLEECTRRFQEMFDRHVVTRPTKVELTDAELREVIDDCNAAVAPLGKTVSDERWISY
VGVVLWSQSPRHIKDMEAFKAVCVLNCVTFVW**DDMD**PALHDFGLFLPQLRKICEKYYGPEDAEVAYEAARAFVTSDH
MFRDSPIKAALCTTSPEQYFRFRVTDIGVDFWMKMSYPIYRHPFTEHAKTSLAARMTTRGLTIV**NDFYSYDRE**VSL
GQITNCFRLCDVSDETAFKEFFQARLDDMIEDIECIKAFDQLTQDVFLDLIYGNFV**W**TTSNK**RY**KTAVNDVNSRIQ

>labdane-related diterpene synthase (WP_019525557), *Streptomyces*, PDB-ID: 5A0K

MTDTDDGGTMLPLPDFTATFPEPFPAGPHSERTEHRLLDWLEEHPLLPSAKAKAVLVNITSHGASRTFPTADADDLL
LFAELLLWLTAF**DDVHAE**GNGVGGPAALVDRASELMLVLAGGNPPRAMSPFPAVLHDLLARFRARASAAAYHRLAAS
LRDTLMALVWEAHHVAKPEGVALATYLAMRPHTVFIKTITAAGEILLGYELTDTQRALAAVRNLETAVANLAGWI**ND**
LASYEREMQRGRGQPLSLPTLLHARHGGTIEEAFTRASSMCENEAAVARRGITHLAHASPNALTAHARALEDITRSF
I**W**HTSHA**RY**QGIRPNRGSSTSSPAR

> iso-elsabellatriene synthase(WP_003963279), *Streptomyces clavuligerus*

MTISVPQLDCPLSRPVHPEGERADAYAVEWLRGVGLMADEADAAPVLAVGLGRLAACYVDENASWDTLAFMTILLAW
YAEY**DDRAID**STGAIDGLTDAEVAELHRALGEILRDRPAPDPSDPVQRGLADVWRTLNGLASDWDRAAFVDTTLRYF
EANRYERVNIRRGIPPTPSAHIGMRRHGGHVYGMYILGAAVNGYRPERRVLDHAAVRELETLAANYTSWA**NDLHSFA**
REHRMGQVNNLVWSVHHHEGLTFQQAADRVADLCDKELAAYLELRQTLPELGIPLTGATGRHVRFLEDMMWSMVD**W**S
ARSA**RY**DVVPEAA

> terpene cyclase (WP_046708564.1), *Streptomyces europaeiscabiei*

MGDAGLAHIPEIHCPFPYRVNPHADRARAHLADWVQRTGLVHRESARKRFDQADFGWFAALVYPTASLRHLELMADW
FAWLFLV**DDQLD**DGSAGRSPDRMQQMVAGMHDVLGSPDFGVSLLRDPDVPPAVASLAELWQRTAVDAPAHWRRRFIR
HLRDCLVTATVWEGGNRIRGIVPDEVTYIENRRHTGAIYVCMDLIDIVERLDVPEELYDSAEFTRALDAACNVVCWT
NDVYSLDKERSLGEVHNIAYLAQYHRGLDREQALAEVCAATSAETERFLAAERHLLSARPGQSAVLAPYLAGMRTWI
RGNLD**W**SRRTK**RY**QAGTAVSWARPADYVERALIGVDQ

> diterpene synthase(BAP82252), *Streptomyces avermitilis*

MNVIDFPQVDLGTEATISPDVDKAGEHLKTWSRATGIVLDGPNGDAFDALAQHRLAAWTYPAATGAELDLMADWIGW
LFAF**DDVFD**ESIDGCDQDFAIATAAATNTVYTGLPSAPPSPVVRPYVVALEDLWERTTQGMPAYWCRRLANDMVDYV
NSYRSHALINASRIALDEQSYRAHRLISSAVFITLDLGEAAARRALSESLLVHPYIRAAREAANNIVSWS**NDLYSAP**
KELSLGDLCNYIAVLQSQDDLTTAQAADRVSQYLHEEIARFDQTKQLIHTHLIPRLPTEDQQGVISMLDTCSNWITG
NVA**W**SLETA**RY**ASRAIDMEREKEVFGDLRT

> diterpene synthase (BAP82229), Streptomyces sp. ND90

MPFVPDFTTPFRYRLNPHLAEVTPRARQWMLESDLVDDTHMLEYEMARIPELMAAAYPGASADDLLLSCDLMGIMFA
I**EDED**CGSHPRHSVAGIATRCKAMIQVMGGVDPGADDPVVLAFSDTWHRLCDGMSDTWVVRHRSSWKDFLDNHHTWE
PVVVEKRGMPTLEDYLWERAYSSGMYVLYDWSERFSADRAEIPQPALEDPRLATLHRNCIYTIIAI**NDTHSLERE**IR
RNDPVPNLLKVLMHHERLTVEQSVERAKKMLADAIEEYLEVEFEYLAHWWRSGLPPRQMGAVEQRLRDMRNWTSSNC
R**W**HCIVP**RY**DHVARDPERDRPGLPTQPAGRQGKDT

> Hydropyrene synthase (EFG04252), *Streptomyces clavuligerus* ATCC 27064

MTISVPQLDCPLSRPVHPEGERADAYAVEWLRGVGLMADEADAAPVLAVGLGRLAACYVDENASWDTLAFMTILMAW
YAEY**DDRAID**STGAIDGLTDAEVAELHRALGEILRDRPAPDPSDPVQRGLADVWRTLNGLASDWDRAAFVDTTLRYF
EANRYERVNIRRGIPPTPSAHIGMRRHGGHVYGMYILGAAVNGYRPERRVLDHAAVRELETLAANYTSWA**NDLHSFA**
REHRMGQVNNLVWSVHHHEGLTFQQAADRVADLCDKELAAYLELRQTLPELGIPLTGATGRHVRFLEDMM**W**SMVDWS
ARSA**RY**DVVPEAA

> terpene cyclase (WP_084744884.1), *Kitasatospora albolonga*

MDSELPDIYCPFPQRTNPHVGHTRGHLATWIRQTGLVHRESAMHRFEQADFGAFVGMVYPTAGPEHLDLVADWFVWL
FLV**DDQLD**DGHLGRSPELVRAVVERMRAVVDGSAPAPLPGEELPAAVIALMDLWERTTPNAAPHWRTRFAWHLVTYL
TTATTWEAGNRADDVVPSEETYIAKRRHTGAIHVCMDLIEIVTGIEAPESLHNDPRFITALEAACNHVCWA**NDVYSF**


24
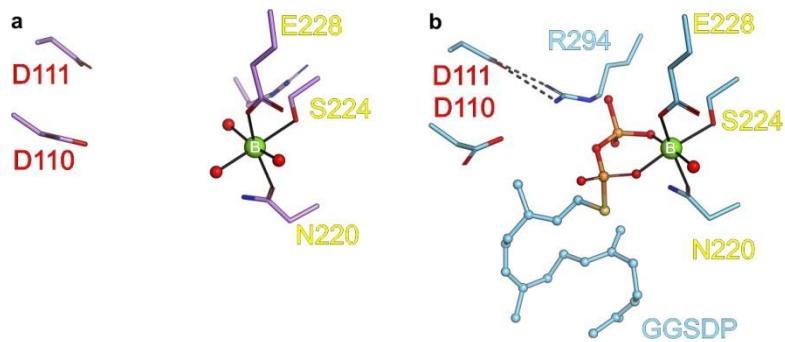
EKEQVLGEIHNLVHLVRHHRGLGEQQALDHVAERLALETERFLTAEDELLELYPELTGLLVPYLEGMRSWMRGNLD**W**
SRQTP**RY**NPADVGQYEEPQQYLEETVLGIAPDHAEAAPCAAEAPRQG

> tsukubadiene synthase (EIF90392), *Streptomyces tsukubaensis* NRRL 18488
MIEVPPFWCPLPIAIHPAADQAEKDARAWAERYGVRLRIADQVQPGRLGAYWAPHGTYEGMLAVGCWNFWAFAF**DDH**
LDEPLPLDVPVTTSLVQQAVDIPSPPITDDPWAAGAQAVFNMFRDLATPTQVRYCADNHRRWLHGACWRHSNHVNRR
LPPLAEYIPLRMQDAAAQATCLIAVLIGSDISVPEQEMDSPRVRALLETASWTATID**SDLHSFQLE**DTQRPVSQHIV
SVLMHERGIGVDEALRQSVALRDRFMTRFLHLQQECARTGSSELARFAHTLGYVISGYLQ**W**AVDTS**RY**GQTEATFSF
TDTPRDDTPEPPPGIPSVEWLWTL
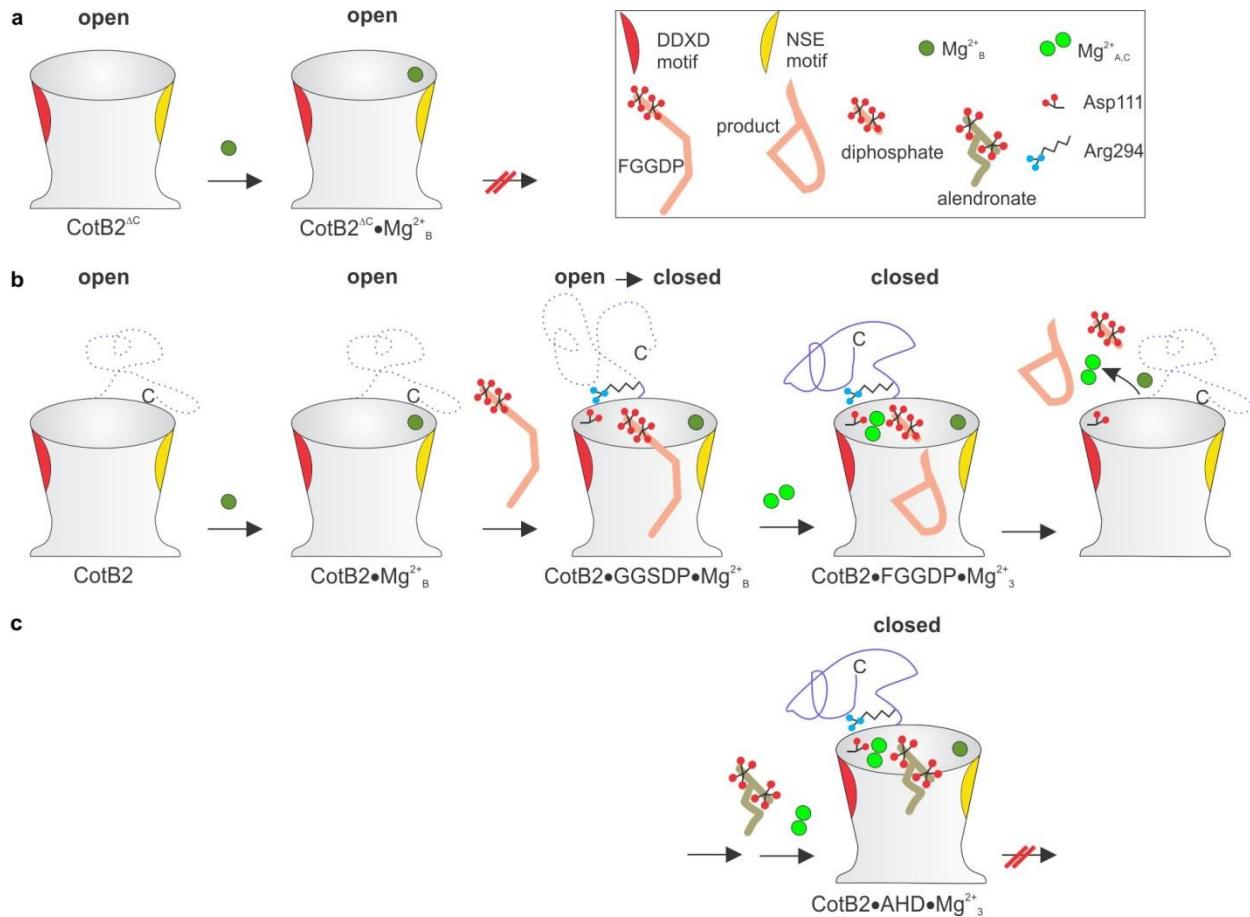
> terpene synthase(ZP_00085244), *Pseudomonas fluorescens* PfO-1
MRKFRQRFYMDLSSLHRNRVGEHLNRPREKDHPMNQSSSARTPRSATAPFIVRAVRCPPPTRIDEALGQEVNERLME
WISNIGIFAGKEEKIRASDFGRYAMLCHADTNNPDRLLLVAQCFAALFAV**DDHYCD**DQSLGGRPETVAESLSFALTA
IDPVYLPSPFDKELLKQQMCDPVIRGLLAYMKRVAQFCTPSQVARVRQITIAMFVTMAAEGPWRLYGTQPTVAEYLA
SRQVNSFWPCLVLIDLIGGYEVPANTYSRPDIHHVTALASLATTLV**NDLYSAYKE**HLNETGDFKLPYLLAARHNCSL
QEAIDLAADIHDAVMEEYERLHATLMKGTRSPVLRRYLTGLSTWIGGNLE**W**HKHSA**RY**HI

> spiroalbatene synthase (WP_030426588.1), *Allokutzneria albata*
MPKLGLSALVPGFTEPSTPPVNPLAERAEDEVVAWLWRIGFLTSQAQEQHLRSFRFGLYHGIATPELDLPALVLGMK
WFCWGSLA**DDQYD**NYDWGDRDARMRSVIRSARTILGGGAVPRTDPVIRGLAEFWPSLVAGMSPAARRRVTRNFLDYL
DAVRFQNRFHAKGDIPDAATFLGLRRHTIAMIFQADVLEALSSLDIPAVLRGHRMFRELVSCFADITAWH**NDVYGLE**
KDIADGQLCNTVLVVSAGEECSTEVAVSRVVERAKERQRLFLGIEAELPWLAEELGLGPEAVASALVLTRQLRAYAY
ANLV**W**IGQTR**RY**DLDLPRIRGTFDDVLCDG

Supplementary Figure 5. Protein sequences of diterpene TPS from different bacteria with accession numbers in parenthesis. Highlighted are the Mg^{2+} binding motifs in red and yellow as well as the WXXXXXRY motif in light blue. Underlined sequences refer to crystal structures of the respective diterpene TPS in the closed conformation, with a structured C-terminus.
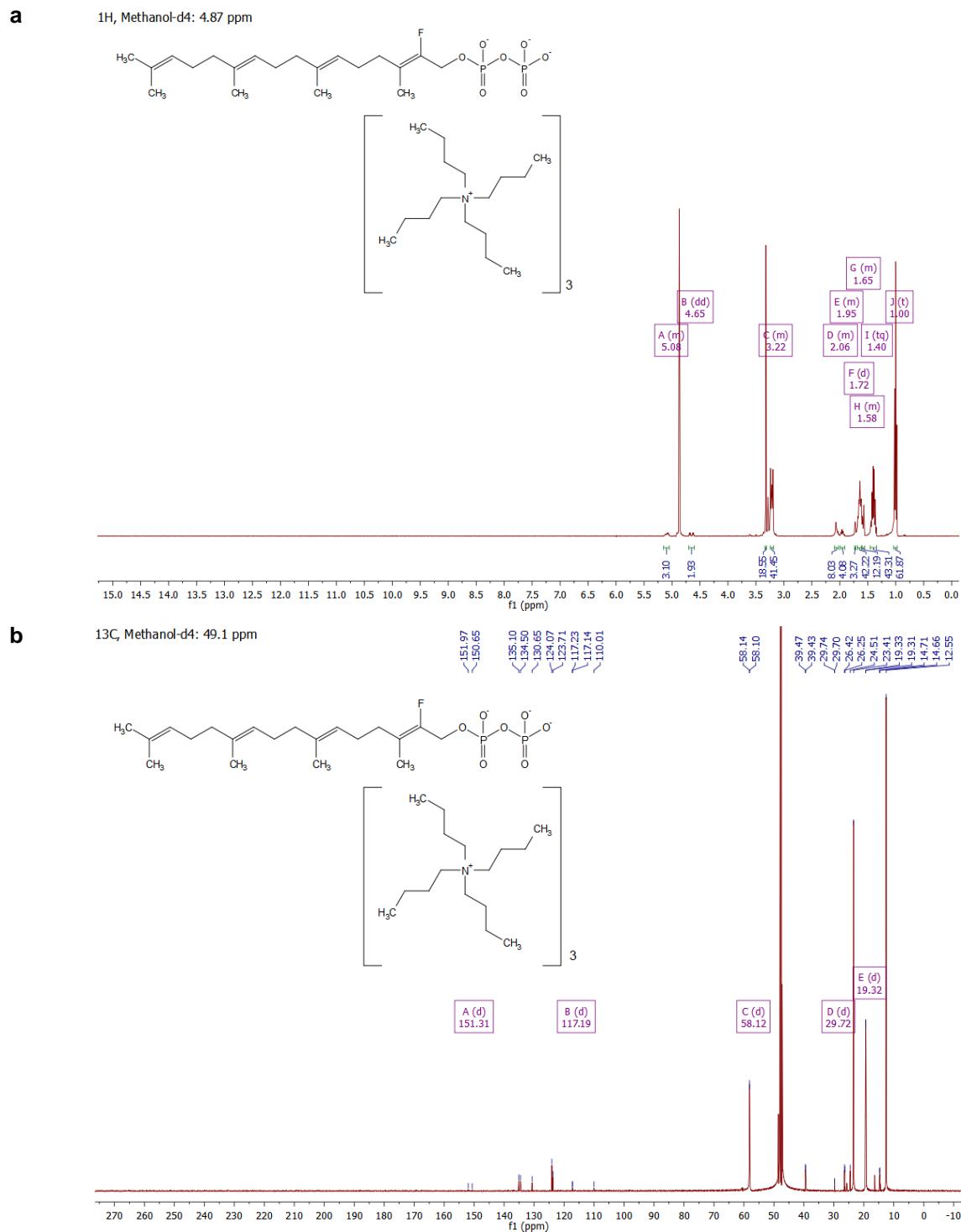
Supplementary Figure 6. CotB2 in different catalytic states. Residues of the DDXD motif are labelled with red letters, residues of the NSE motif with yellow letters, respectively. Mg^{2+}-ions are depicted as green spheres and water molecules as red spheres. Hydrogen bonds are indicated by dashed lines. Solid lines represent the coordination sphere of the Mg^{2+}-ions. **a,** Catalytic centre of CotB2^{F107A}•Mg^{2+}_B. **b,** Catalytic centre of CotB2^{wt}•Mg^{2+}_B•GGSDP.[3]
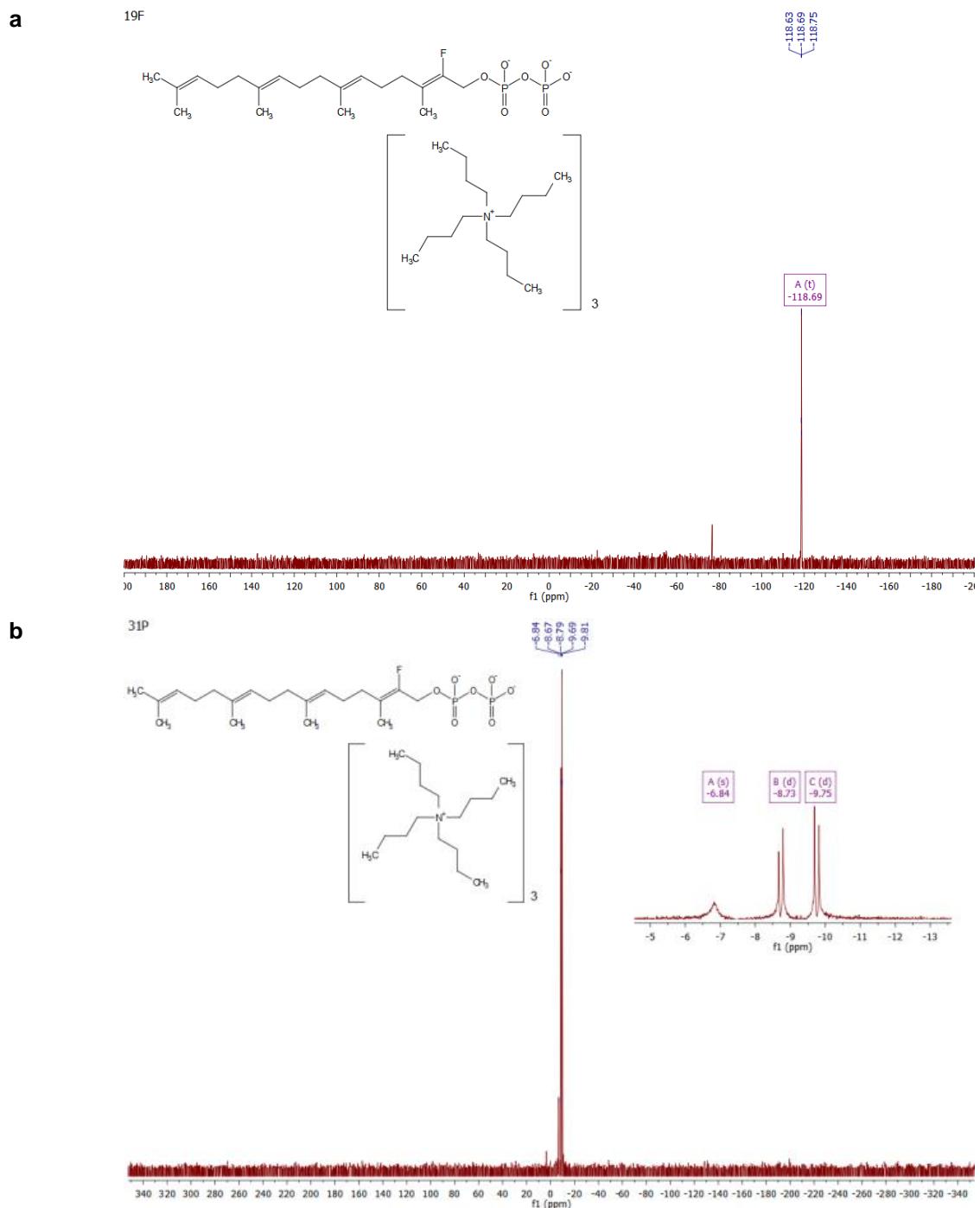
**Supplementary Figure 7. Schematic drawing of the different structural snapshots of CotB2.** CotB2 is schematically drawn as a gray cup. The Mg^{2+} binding motifs are indicated in red ($^{110}DDXD^{113}$ motif) and yellow ($^{218}NSE^{228}$ motif). The unfolded C-terminus of the open state of CotB2 is indicated by a dashed line and in the folded, closed state as a solid purple line. **a,** CotB2$^{\Delta C}$ is still capable of binding to Mg^{2+}_B (dark green circle) by the NSE motif (relevant structure CotB2$^{\Delta C} \cdot Mg^{2+}_B$; PDB-ID 6GGK). But the missing C-terminus does not allow neither to bind nor to cyclize the educt. **b,** In the open structure of CotB2 the C-terminus is not folded (relevant structure CotB2^{wt}; PDB-ID 4OMG ref[10]). As the first binding event, Mg^{2+}_B is bound by the NSE motif (relevant structure CotB2$^{F107A} \cdot Mg^{2+}_B$; PDB-ID 6GGL). Subsequently the substrate GGDP and the C-terminus starts to fold over the active site. A salt-bridge between R294 and D111 of the DDXD motif is at this stage already established (relevant structure CotB2$^{wt} \cdot Mg^{2+}_B \cdot GGSDP$; PDB-ID 5GUE ref[3]), but most of the C-terminus remains unfolded. Binding of $Mg^{2+}_{A,C}$ (light green circles) to the DDXD motif finally leads to a fully structured C-terminus that now completely closes the active site. Now, CotB2 is ready to initiate the cyclisation reaction and to cleave the diphosphate moiety (relevant structure CotB2$^{wt} \cdot Mg^{2+}_3 \cdot F$-Dola; PDB-ID 6GGI). Finally the cyclised product is released. **c,** Alendronate (AHD) is recognised by CotB2 like the substrate-analogue FGGDP, but the diphosphate
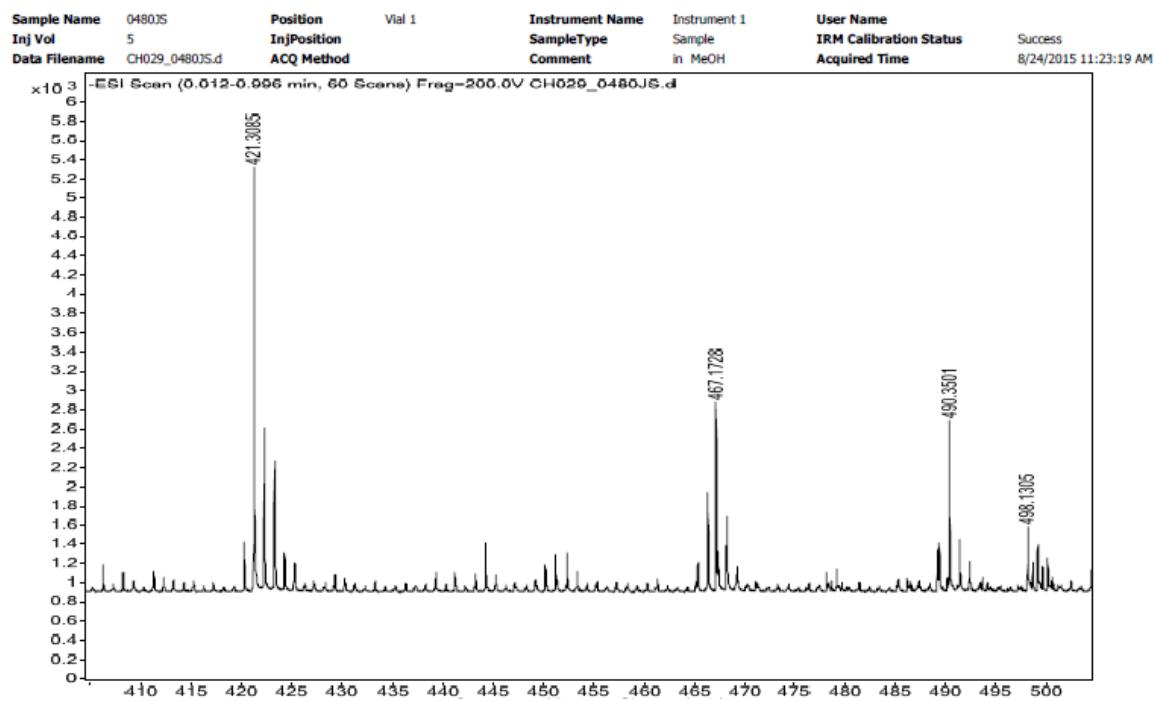
moiety cannot be cleaved (relevant structure CotB2^{wt}•Mg$^{2+}_3$•AHD; PDB-ID 6GGJ). The diphosphate moiety of FGGDP and AHD occupies an identical position in both structures. The C-terminus is folded as in the closed structure of CotB2^{wt}•Mg$^{2+}_3$•F-Dola. Since no cyclisation occurs, the AHD remains captured in the active site.

a 1H, Methanol-d4: 4.87 ppm

b 13C, Methanol-d4: 49.1 ppm
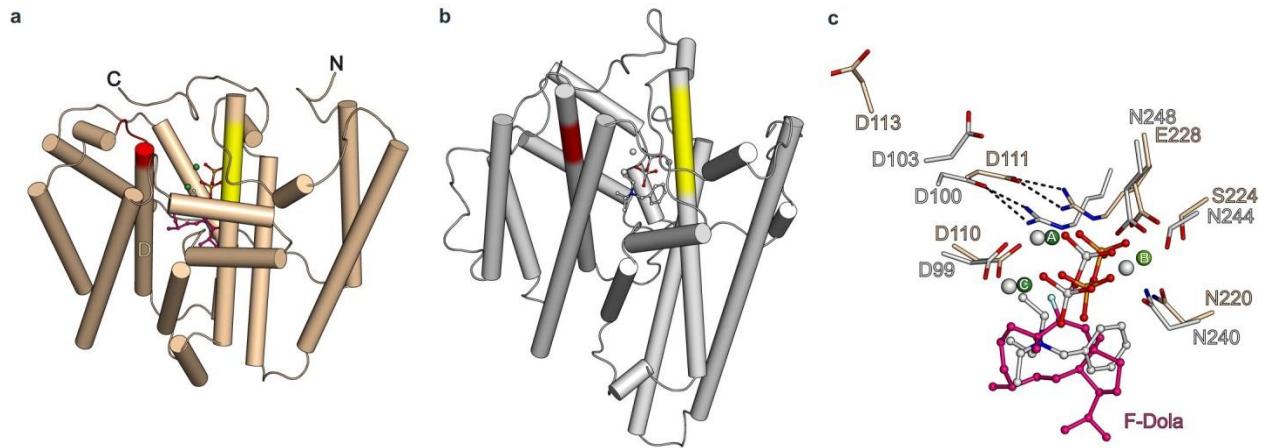
Supplementary Figure 8. NMR analysis of the tetrabutylammonium salt of 2-fluoro-geranylgeranyl diphosphate. a, ^{1}H-NMR **b,** ^{13}C-NMR.

Supplementary Figure 9. NMR analysis of the tetrabutylammonium salt of 2-fluoro-geranylgeranyl diphosphate. a, ^{19}F-NMR **b,** ^{31}P-NMR.

Supplementary Figure 10. HRMS analysis of tetrabutylammonium salt of 2-fluoro-geranylgeranyl diphosphate. (ESI-TOF, m/z): ($C_{20}H_{34}FO_7P_2^-$) [M-H]$^-$ = 467.1728 (calculated 467.1768).

31

Supplementary Figure 11. The non-canonical DDXD motif in comparison to other TPSs with the canonical DDXXD motif. a. The overall structure of CotB2^{wt}•Mg$^{2+}_3$•F-Dola with α-helices drawn as cylinders. The ^{110}DDMD^{113} motif residing on helix D is highlighted in red and the NSE motif in yellow **b.** The overall structure of epi-isozizaene synthase (PDB ID 3KB9 ref[2]) with α-helices drawn as cylinders. The ^{99}DDRHD^{103} motif is highlighted in red and the DTE motif in yellow. **c.** Superposition of the catalytic motifs in CotB2^{wt}•Mg$^{2+}_3$•F-Dola and epi-isozizaene synthase. Zoom into the active site. Same view as in a and b. The Mg^{2+} in the CotB2^{wt}•Mg$^{2+}_3$•F-Dola structure are drawn as green spheres and of epi-isozizaene synthase as gray spheres. The bound F-Dola is shown in magenta and the bound N-benzyl-N,N-diethylethanaminium in gray stick representation, respectively. For clarity water molecules have been omitted.

Supplementary References

1   Koksal, M., Jin, Y., Coates, R. M., Croteau, R. & Christianson, D. W. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* **469**, 116-120, (2011).

2   Aaron, J. A., Lin, X., Cane, D. E. & Christianson, D. W. Structure of epi-isozizaene synthase from Streptomyces coelicolor A3(2), a platform for new terpenoid cyclization templates. *Biochemistry* **49**, 1787-1797, (2010).

3   Tomita, T. *et al.* Structural Insights into the CotB2-Catalyzed Cyclization of Geranylgeranyl Diphosphate to the Diterpene Cyclooctat-9-en-7-ol. *ACS Chem. Biol.* **12**, 1621-1628, (2017).

4   Roe, S. J., Oldfield, M. F., Geach, N. & Baxter, A. A convergent stereocontrolled synthesis of [3-(14) C]solanesol. *J. Labelled Comp. Radiopharm.* **56**, 485-491, (2013).

5   Long, J. *et al.* Total syntheses of parthenolide and its analogues with macrocyclic stereocontrol. *J. Med. Chem.* **57**, 7098-7112, (2014).

6   Cane, D. E., Yang, G., Xue, Q. & Shim, J. H. Trichodiene synthase. Substrate specificity and inhibition. *Biochemistry* **34**, 2471-2479, (1995).

7   Görner, C., Hauslein, I., Schrepfer, P., Eisenreich, W. & Brück, T. Targeted Engineering of Cyclooctat-9-en-7-ol Synthase: A Stereospecific Access to Two New Non-natural Fusicoccane-Type Diterpenes. *Chemcatchem* **5**, 3289-3298, (2013).

8   Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330**, 70-74, (2010).

9   Kim, S. Y. *et al.* Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two p450 hydroxylases. *Chemistry & biology* **16**, 736-743, (2009).

10 Janke, R., Görner, C., Hirte, M., Brück, T. & Loll, B. The first structure of a bacterial diterpene cyclase: CotB2. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1528-1537, (2014).

11 Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549, (2010).

12 Tomita, T. *et al.* Structural Insights into the CotB2-Catalyzed Cyclization of Geranylgeranyl Diphosphate to the Diterpene Cyclooctat-9-en-7-ol. *ACS chemical biology* **12**, 1621-1628, (2017).

13 Shishova, E. Y., Di Costanzo, L., Cane, D. E. & Christianson, D. W. X-ray crystal structure of aristolochene synthase from Aspergillus terreus and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry* **46**, 1941-1951, (2007).

14 Baer, P. *et al.* Induced-fit mechanism in class I terpene cyclases. *Angew Chem Int Ed Engl* **53**, 7652-7656, (2014).

15 Serrano-Posada, H. *et al.* Crystallization and X-ray diffraction analysis of a putative bacterial class I labdane-related diterpene synthase. *Acta Crystallogr F Struct Biol Commun* **71**, 1194-1199, (2015).

16 Harris, G. G. *et al.* Structural Studies of Geosmin Synthase, a Bifunctional Sesquiterpene Synthase with alpha alpha Domain Architecture That Catalyzes a Unique Cyclization-Fragmentation Reaction Sequence. *Biochemistry* **54**, 7142-7155, (2015).

17 Chen, M., Chou, W. K., Toyomasu, T., Cane, D. E. & Christianson, D. W. Structure and Function of Fusicoccadiene Synthase, a Hexameric Bifunctional Diterpene Synthase. *ACS chemical biology* **11**, 889-899, (2016).

18 Rynkiewicz, M. J., Cane, D. E. & Christianson, D. W. Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade. *P Natl Acad Sci USA* **98**, 13543-13548, (2001).

19 Koksal, M., Chou, W. K., Cane, D. E. & Christianson, D. W. Unexpected reactivity of 2-fluorolinalyl diphosphate in the active site of crystalline 2-methylisoborneol synthase. *Biochemistry* **52**, 5247-5255, (2013).

20 Liu, W. *et al.* Structure, function and inhibition of ent-kaurene synthase from Bradyrhizobium japonicum. *Sci Rep* **4**, 6214, (2014).

21 Liebschner, D. *et al.* Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D Biol. Crystallogr.* **73**, 148-157, (2017).