

ARTICLE

Received 15 Sep 2016 | Accepted 28 Apr 2017 | Published 15 Jun 2017

DOI: 10.1038/ncomms15804

OPEN

# Cryptic indole hydroxylation by a non-canonical terpenoid cyclase parallels bacterial xenobiotic detoxification

Susann Kugel<sup>1,\*</sup>, Martin Baunach<sup>1,\*</sup>, Philipp Baer<sup>2</sup>, Mie Ishida-Ito<sup>1</sup>, Srividhya Sundaram<sup>1</sup>, Zhongli Xu<sup>1</sup>, Michael Groll<sup>2</sup> & Christian Hertweck<sup>1,3</sup>

Terpenoid natural products comprise a wide range of molecular architectures that typically result from C-C bond formations catalysed by classical type I/II terpene cyclases. However, the molecular diversity of biologically active terpenoids is substantially increased by fully unrelated, non-canonical terpenoid cyclases. Their evolutionary origin has remained enigmatic. Here we report the *in vitro* reconstitution of an unusual flavin-dependent bacterial indoloterpenoid cyclase, XiaF, together with a designated flavoenzyme-reductase (XiaP) that mediates a key step in xiamycin biosynthesis. The crystal structure of XiaF with bound FADH<sub>2</sub> (at 2.4 Å resolution) and phylogenetic analyses reveal that XiaF is, surprisingly, most closely related to xenobiotic-degrading enzymes. Biotransformation assays show that XiaF is a designated indole hydroxylase that can be used for the production of indigo and indirubin. We unveil a cryptic hydroxylation step that sets the basis for terpenoid cyclization and suggest that the cyclase has evolved from xenobiotics detoxification enzymes.

<sup>1</sup>Department of Biomolecular Chemistry, Leibniz Institute for Natural Product Research and Infection Biology (HKI), Beutenbergstr. 11a, 07745 Jena, Germany.

<sup>2</sup>Center for Integrated Protein Science Munich (CIPSM), Department Chemie, Technische Universität München, Lichtenbergstr. 4, 85748 Garching, Germany.

<sup>3</sup>Natural Product Chemistry, Friedrich Schiller University, 07743 Jena, Germany. \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.G. (email: michael.groll@tum.de) or to C.H. (email: christian.hertweck@leibniz-hki.de).

It is a generally accepted model that the vast structural diversity of secondary metabolites has developed to address specific needs of the producing organism. Such natural products may, for example, serve as carrier of information or as chemical armour; they may be used to attract partners, to secure the niche, to protect against various types of stresses and to shape communities<sup>1,2</sup>. Natural products are also a valuable source of molecular tools and therapeutics that have been pre-optimized for particular biological targets over millions of years<sup>3</sup>. However, considering the impressive range of structures and functions, one of the remaining big questions is: how has this structural diversity evolved?

Research in the past decades has led to an impressive body of knowledge on biosynthetic pathways, their genetic basis and the biocatalysts involved<sup>4–6</sup>. Typically, there are various levels of diversification in each biosynthetic pathway, such as the choice and assembly of basic building blocks, the processing of various intermediates and various ways to tailor the individual scaffolds<sup>4,7</sup>. Typically, enzymes involved in secondary metabolite biosynthesis are rather specific<sup>8</sup>. They are often fine-tuned to direct reactive intermediates into specific reaction channels, and small changes may lead to dramatic changes in the metabolic profile<sup>9,10</sup>. This principle is well exemplified in terpene biosynthesis, where designated cyclases orchestrate a series of cation–olefin cyclizations and Wagner–Meerwein rearrangements of intermediates to produce diverse carbocyclic systems<sup>11,12</sup>. Beyond these classical type I/II terpene cyclases, there is a surprisingly broad range of non-canonical terpenoid cyclases<sup>13,14</sup>.

Various atypical terpenoid cyclizations and diversifications are involved in the pathway leading to the xiamycin family of indolosesquiterpenes (IST) (Fig. 1). These ‘plant-like’ terpenoids are actually produced by bacterial endophytes (*Streptomyces* spp.). Endowed with a range of antiviral<sup>15–17</sup> and antimicrobial activities<sup>18–24</sup> xiamycin and its congeners may aid in protecting the plant host from infections<sup>16,20</sup>.

To elucidate the molecular basis of IST diversification, we investigated the xiamycin (*xia*) biosynthetic gene cluster in the endophyte. We found that xiamycin (3), the major IST component of the complex<sup>20</sup>, is diversified by means of rare flavoenzyme-catalysed modifications to yield the cycloether oxiamycin (5)<sup>18,23</sup>, various *N,C*- and *N,N*-fused xiamycin dimers (such as 6a/6b)<sup>18,22,23</sup>, and sulfonyl-bridged xiamycin dimers (such as 7)<sup>19</sup> (Fig. 1). At the heart of the biosynthetic pathway, however, we discovered an unusual cyclization cascade involving the non-canonical terpenoid cyclases XiaE and XiaF<sup>22</sup>. The first cyclization yields the decalin system of indosespene (2), which is further cyclized into the spiro compound sespenine (4) or the carbazole ring system of xiamycin (Fig. 1)<sup>13,21,22,25,26</sup>.

This non-canonical terpenoid cyclization is mechanistically intriguing and has been the subject of several biosynthetic<sup>21,22,26</sup> and synthetic studies<sup>17,25</sup>. Through bioinformatics and mutational analysis, we found that XiaF is a flavin-dependent monooxygenase that catalyses the conversion of 2 to the carbazole scaffold of 3 (ref. 22). In an independent study, an orthologue from *Streptomyces* sp. SCSIO 02999 (XiaI, according to these authors’ nomenclature) was found to catalyse this reaction *in vitro*, albeit with unsatisfactory efficacy<sup>21</sup>. Yet, the detailed enzymatic mechanism of this rare cyclization reaction and the evolutionary origin of this extraordinary oxygenase/cyclase have remained elusive.

Here we report the *in vitro* reconstitution of XiaF and a designated flavoenzyme-reductase, which jointly mediate a cryptic indole hydroxylation step that has been proposed to be a key step in non-canonical terpenoid cyclization. By phylogenetic, functional and structural analyses, we also demonstrate the

relatedness of this rare terpenoid cyclase with broad-specific xenobiotic-metabolizing enzymes.

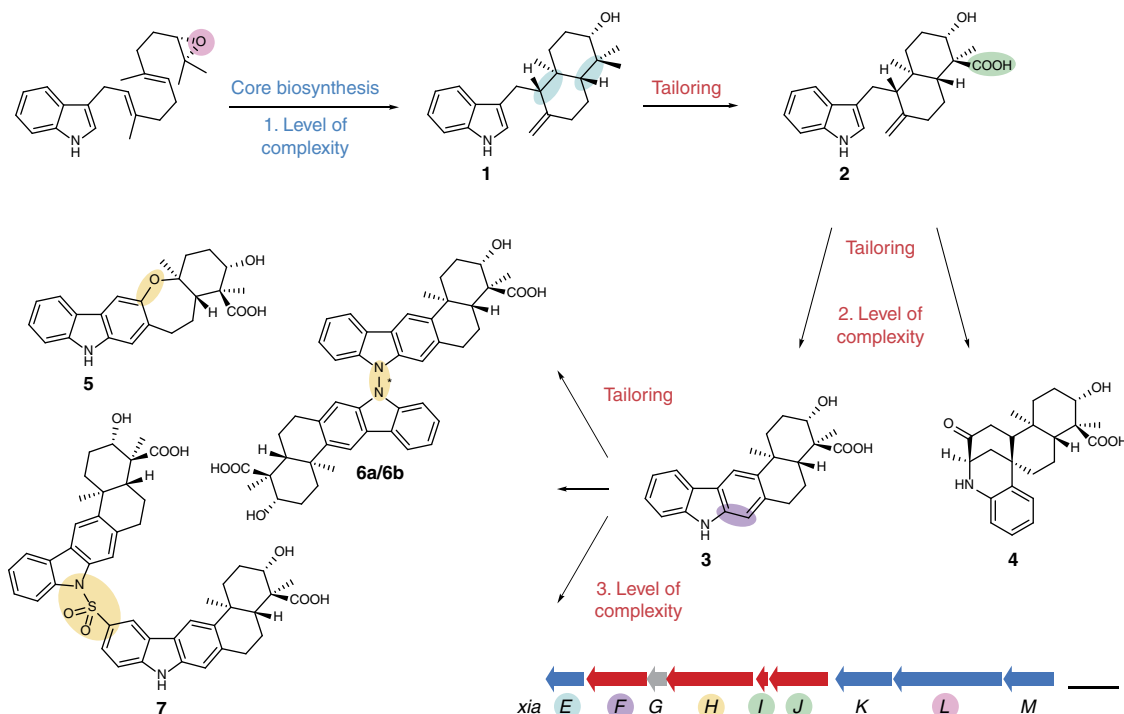
## Results

**Reconstitution of two-component terpenoid cyclization.** To gain insight into the biochemical function of the indosespene cyclase XiaF, we produced XiaF in *Escherichia coli*. Initially, the *in vitro* characterization of recombinant XiaF was hampered by its poor performance. Substantial biotransformation of 2 (100 μM) was only achieved after extended incubation times (12–24 h) with high concentrations of enzyme (25 μM) and cofactors FAD and NADH (1 and 2 mM). Similar observations were reported for enzyme assays using the XiaF orthologue XiaI from *Streptomyces* sp. SCSIO 02999 (refs 21,26). However, even under these optimized non-physiological reaction conditions, the catalytic performance of XiaF was unreliable.

To overcome these shortcomings, we investigated the effect of added NAD(P)H:flavin oxidoreductase (flavin reductase) for cofactor recycling. In fact, terpenoid cyclization was efficiently reconstituted *in vitro* using Fre, the flavin reductase from *E. coli* (Supplementary Fig. 2). This finding strongly suggested that XiaF naturally acts in a two-component system. It is conceivable that XiaF variants are orphan oxygenases<sup>27</sup> that are supported by unrelated flavin reductases derived from the central metabolism. Alternatively, XiaF variants may have a designated reductase partner. Thus, we revisited the *xia* gene locus<sup>22</sup> and related gene clusters of various other actinomycetes<sup>26</sup>. Genome mining revealed a cryptic gene cluster in *Amycolatopsis nigriscens* CSC17Ta-90 with high similarity to the *xia* gene locus. In contrast to the *xia* gene cluster in the endophytic *Streptomyces* sp. (HKI0576), a putative flavin reductase gene is located next to the *xiaF* orthologue (Fig. 2a). On the basis of this sequence, we screened the genome sequence of *Streptomyces* sp. (HKI0576) and discovered a homologous flavin reductase gene (named *xiaP*) located in the flanking region of the *xia* gene cluster (Fig. 2a; the *xiaP* orthologue from *A. nigriscens* CSC17Ta-90 was named *xiaP'*). The deduced gene products of both *xiaP* variants share homology with other flavin reductase partners of XiaF homologues<sup>28–34</sup> (Supplementary Fig. 5).

To test whether XiaP functions as flavin reductase partner of XiaF, we cloned and heterologously expressed *xiaP* from *Streptomyces* sp. (HKI0576) in *E. coli*. Flavin reduction activity assays of XiaP with NAD(P)H and FAD, FMN or riboflavin confirmed its proposed function. They revealed a dependency of XiaP on NADH, whereas NADPH seems to be no suitable cofactor. In contrast, XiaP seems to have no marked preference for any flavin substrate tested (Fig. 3c; for a detailed kinetic analysis of XiaP, see the Supplementary Note 2). When substituting Fre by XiaP in the XiaF enzyme assays, we found that XiaP substantially increases the performance of XiaF (Fig. 3b,I). Since the corresponding genes are found in the same gene locus, XiaP likely acts as the natural flavin reductase for XiaF, although its function may be substituted by other flavin reductases.

Using this robust enzyme assay, we set out to analyse the course of the reaction. In the current biosynthetic model, an oxidative cyclization sequence is proposed in which hydroxylation of 2 is presumed to trigger nucleophilic attack of the *exo*-methylene group followed by dehydration to yield 3 (refs 20–22). Furthermore, the immediate product of the XiaF-catalysed reaction has been proposed to be a dihydrocarbazole, which most likely undergoes spontaneous oxidation to give the carbazole 3 (ref. 21). By liquid chromatography high-resolution mass spectrometry (LC-HRMS) monitoring of the XiaF assay, we found evidence for the transient intermediate (denoted with an asterisk, Supplementary Fig. 2b,I;



**Figure 1 | Model of bacterial IST biosynthesis.** The biosynthetic scheme shows selected members of the xiamycin family of IST, their biogenetic relationship and their proposed biosynthetic pathway. IST biosynthesis comprises various levels of terpenoid diversification. An epoxidized linear IST-precursor has been proposed to be cyclized to yield preindosespene (**1**)<sup>21,22,26</sup>, which is subsequently oxygenated to the carboxylic acid indosespene (**2**)<sup>22,46</sup>. **2** has been proposed to be further cyclized into the spiro compound sespenine (**4**) or the carbazole ring system of xiamycin (**3**)<sup>21,22,26</sup>, of which the latter has been proposed to be further diversified to yield the cycloether oxiamycin (**5**)<sup>18,23</sup>, various *N,C*- and *N,N*-fused xiamycin dimers (such as **6a/6b**)<sup>18,22,23</sup>, and sulfonyl-bridged xiamycin dimers (such as **7**)<sup>19</sup>. Genes encoding enzymes that are involved in the core biosynthesis of the IST scaffold of preindosespene (**1**) are coloured in blue (*xiaEKL*M), genes for enzymes that tailor basic scaffolds to more complex ISTs are coloured in red (*xiaFGHIJ*) and genes with no proposed function are coloured in grey (*xiaG*). Atoms, bonds or functional groups that are added/are newly formed in biosynthetic reactions are highlighted in the same colour as the gene(s) encoding the gene product(s) that have been proposed to mediate these reactions<sup>18,19,22,26,46</sup>. The length of the scale bar represents a kilobase.

Fig. 3b,I), showing the expected exact mass ( $m/z$   $[M-H]^-$  364.1928 Da, calculated 364.1918 Da) (Supplementary Fig. 6). However, the proposed hydroxylated intermediate could not be detected *in vitro*.

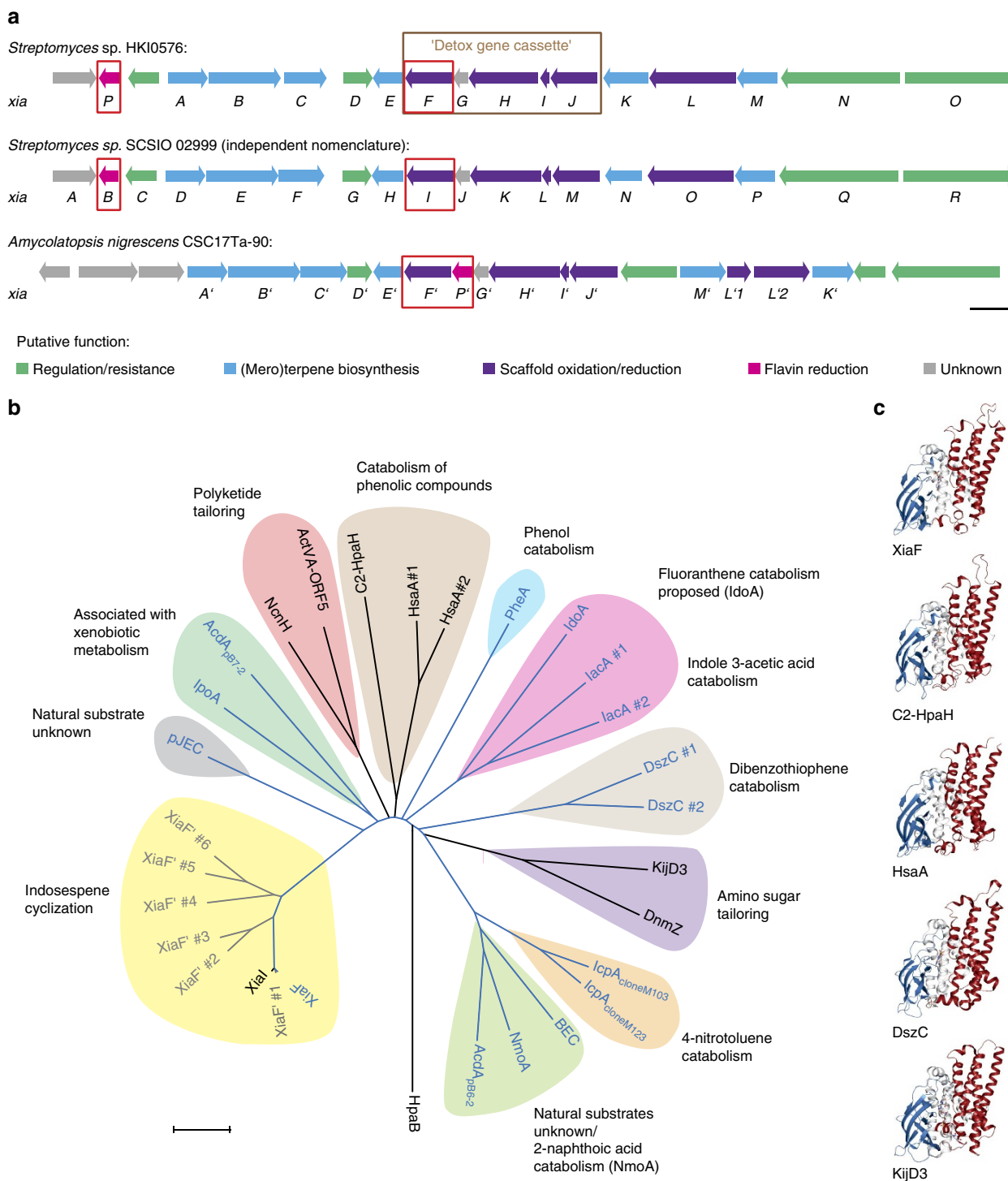
**Phylogenetic relationship and genetic context of XiaF.** To gain insight into the possible evolutionary origin of XiaF, we scrutinized related flavoenzymes, their catalytic function and biological role *in silico*. Through homology searches (BLAST), we compiled a diverse set of 28 XiaF homologues (Fig. 2b), most of which have been classified as group D flavin-dependent monooxygenases (FMOs). Group D FMOs are monooxygenases with an acetyl-CoA dehydrogenase (ACAD)-fold that receive reduced FAD/FMN from a NAD(P)H-dependent flavin reductase<sup>35</sup>. Various group D FMOs have been shown to catalyse aromatic hydroxylation and *N*-hydroxylation reactions<sup>35</sup>. Unexpectedly, apart from few secondary metabolite oxygenases, XiaF homologues identified in the homology search are involved in the catabolism of xenobiotics.

To investigate the phylogenetic position of XiaF, we used the neighbour-joining algorithm<sup>36</sup> to construct a phylogenetic tree of all sequences from evolutionary distance data (Fig. 2b). XiaF forms a clade together with its orthologues from *S. sp.* SCSIO 02999 (XiaI) and *A. nigriscens* CSC17Ta-90 (XiaF' #4; not yet characterized orthologues of XiaF were named XiaF' and consecutively numbered from 1 to 6), as well as with

five orthologues from other putative *xia* gene clusters from various actinomycetes. This subclade is part of a bigger clade including indigo-forming oxygenases pJEC, AcadA<sub>PB7-2</sub> and IpoA discovered through soil metagenomics<sup>27,37,38</sup>. Although substrates other than indole are unknown for these three enzymes, the induction of their expression by artificial pollution (AcadA<sub>PB7-2</sub>)<sup>27</sup> or the exposure to limonene (IpoA)<sup>38</sup> implies a role in xenobiotic metabolism for at least two of them.

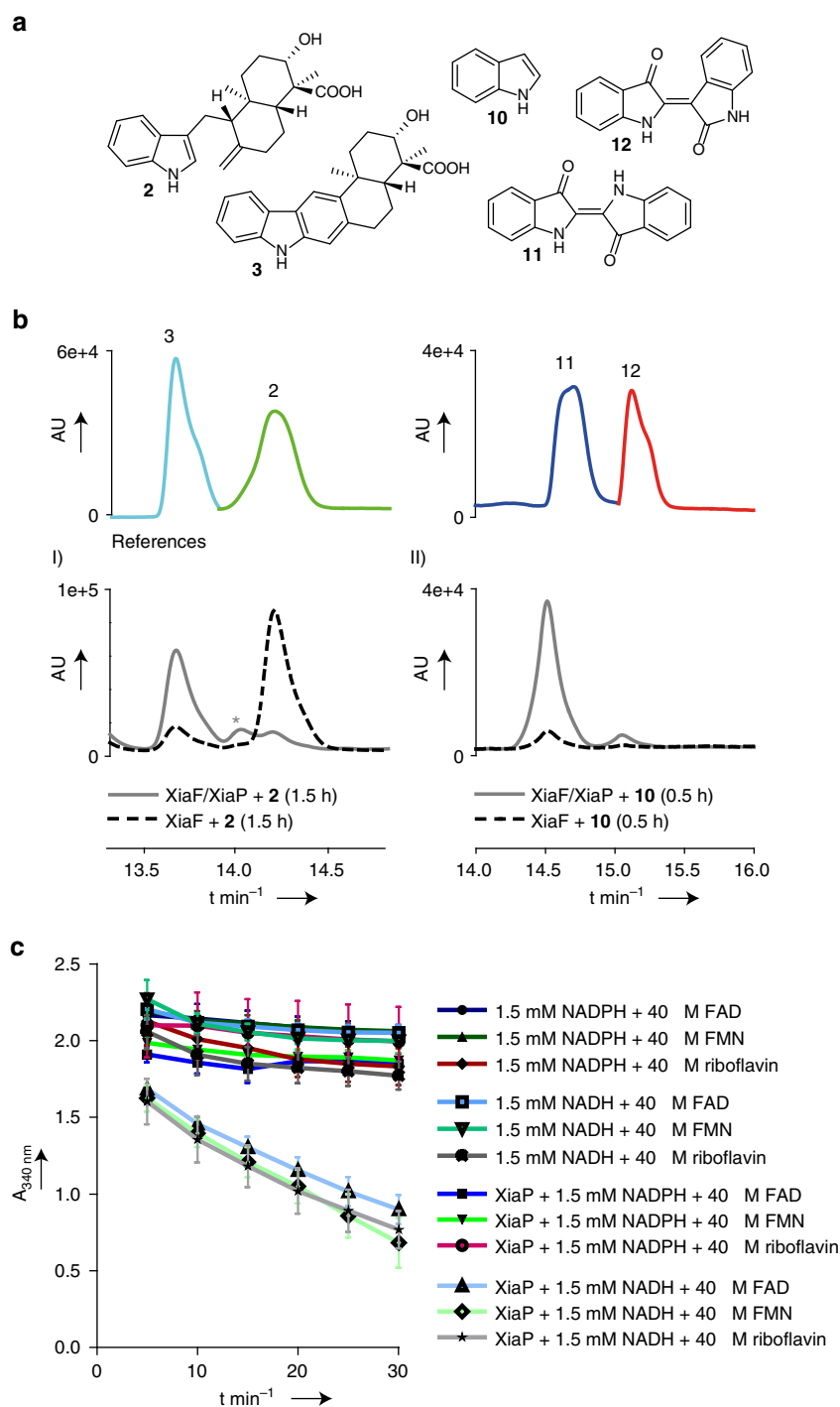
In the cladogram, we identified various additional groups of oxygenases that are specifically involved in xenobiotics degradation: dibenzothiophene monooxygenases from *Rhodococcus erythropolis* IGTS8 (DszC #1) and *Bacillus subtilis* (DszC #2)<sup>39,40</sup>; 4-nitrotoluene monooxygenases IcpA<sub>clone M103</sub> and IcpA<sub>clone M103</sub> discovered in a metagenome screening of activated sludge<sup>41</sup>; indole 3-acetic acid monooxygenases from *Acinetobacter baumannii* ATCC 19,606 (IacA #1) and *Pseudomonas putida* (IacA #2), and IdoA, for which the polycyclic aromatic hydrocarbon (PAH) fluoroanthrene was proposed to be the natural substrate<sup>42</sup>; 2-naphthoic acid monooxygenase NmoA<sup>30</sup>, 4-hydroxyphenylacetate 3-hydroxylase C2-HpaH<sup>43</sup> and phenol monooxygenase PheA<sup>44</sup>.

Surprisingly, also the genetic context of *xiaF* within the *xia* gene cluster of *S. sp.* (HKI0576) indicates its relationship to xenobiotics metabolism. On closer inspection, we found that *xiaF* forms a subcluster with genes (*xiaGHII*) that encode homologues of xenobiotic-metabolizing enzymes (Figs 1 and 2a). *XiaG*<sup>22</sup> encodes a protein that is homologous to limonene-1,2-epoxide hydrolase<sup>45</sup>, whereas XiaH<sup>18,19,22</sup> shares homology with flavoenzymes involved



**Figure 2 | Comparison of *xia* gene loci, and phylogeny of XiaF and its homologues.** (a) Comparison of the *xia* gene cluster of *Streptomyces* sp. (HKI0576) with the *xia* gene cluster of *Streptomyces* sp. SCSIO 02999 and a putative *xia* gene locus of *Amycolatopsis nigrescens* CSC17Ta-90. Genes for the two-component system XiaF/XiaP and orthologues are framed in red. The 'detox gene cassette' *xia*FGHIJ, which encodes homologues of xenobiotic-metabolizing enzymes that are mostly involved in IST tailoring reactions is framed in brown. The length of the scale bar represents a kilobase.

(b) Phylogenetic tree of XiaF and related flavoenzymes. Enzymes that have been reported to catalyse indigo formation are coloured in blue; putative XiaF orthologues of other putative *xia* gene clusters that yet have not been characterized are coloured in grey. The evolutionary history was inferred using the neighbour-joining method<sup>36</sup>. The oxygenase component HpaB of the 4-hydroxyphenylacetate 3-monooxygenase from *Thermus thermophilus* HB8 (ref. 90) was used as an outgroup. The length of the scale bar represents an evolutionary distance of 0.1 amino acid differences per site. (For further details, see the Supplementary Note 3 and Supplementary Figs 8–11.) (c) Structural comparison of XiaF (PDB ID: 5LVW), the 4-hydroxyphenylacetate 3-hydroxylase C2-HpaH (PDB ID: 2JBT)<sup>48</sup>, the secosteroid hydroxylase HsaA from *Mycobacterium tuberculosis* (PDB ID: 3AFE)<sup>32</sup>, the dibenzothiophene monooxygenase DszC from *Rhodococcus erythropolis* D-1 (PDB ID: 3X0Y)<sup>52</sup> and the *N*-oxygenases KijD3 involved in *D*-kijanose biosynthesis (PDB ID: 4KCF)<sup>53</sup>. The structural alignment was calculated by the Dali server<sup>51</sup> and shows the superposition of the different monomers. (For a more detailed structural comparison, see Supplementary Fig. 7.)



**Figure 3 | *In vitro* reconstitution and characterization of XiaF and XiaP.** (a) Structures of substrates and products of XiaF enzyme assays. (b) HPLC-HRMS monitoring (PDA 300 nm) showing the results of (I) XiaF enzyme assays with indosespene (**2**) as the substrate in the presence and absence of XiaP (the asterisk denotes the proposed transient intermediate prexiamycin (**8**); for details, see Supplementary Fig. 6); (II) XiaF enzyme assays with indole (**10**) as the substrate in the presence and absence of XiaP. The profiles showing the reference compounds indosespene (**2**) and xiamycin (**3**) as well as indigo (**11**) and indirubin (**12**) are composed of different measurements of pure compounds (colour coded). (c) *In vitro* activity test of flavin reductase XiaP (0.1  $\mu$ M). Conversion rates were followed by measuring the decrease in absorbance at 340 nm over time. Data represent mean values of three independent experiments, each conducted as duplicates. Error bars indicate s.d.

in the degradation of phenolic compounds<sup>18</sup>. The deduced gene product of *xiaJ* is a cytochrome P450 monooxygenase (CYP450)<sup>22,46</sup>, which is closely related to hydroxylases implicated in bacterial detoxification processes. Except for XiaG, all enzymes encoded in this subcluster are proposed to be involved in IST tailoring reactions<sup>18,19,22,26,46</sup> (Fig. 1).

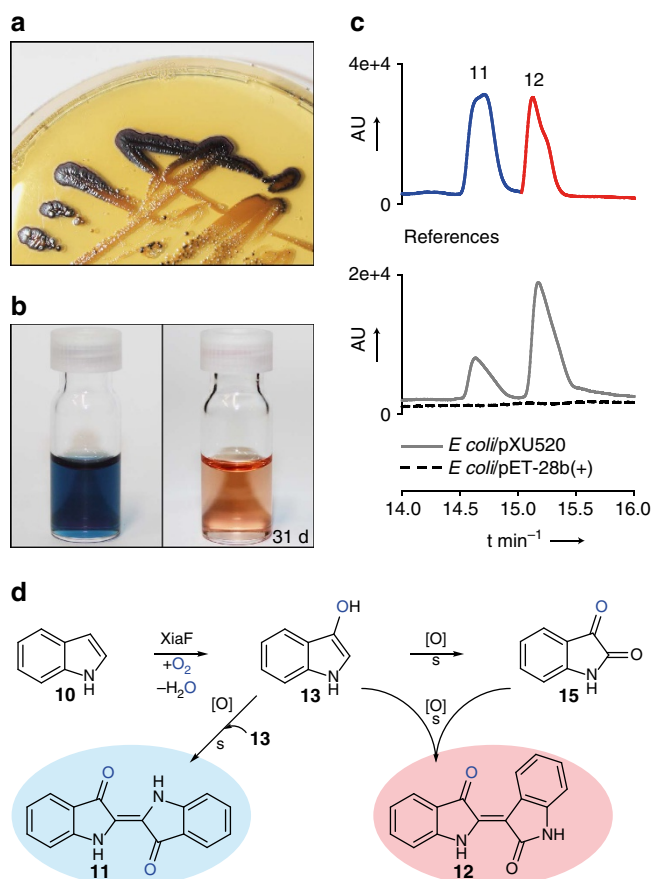
**XiaF catalyses a cryptic indole hydroxylation.** To test XiaF's ability to detoxify typical xenobiotics, we added various aromatic compounds to the *in vitro* enzyme assay. We selected reported substrates of the XiaF homologues, 4-nitrotoluene, 4-hydroxyphenylacetate, 2-naphthoic acid, 3-hydroxybiphenyl, indole 3-acetic acid and indole. However, from this range of aromatic

compounds, only indole (**10**) was transformed into an oxygenated product. LC-MS monitoring of the XiaF *in vitro* assay supplemented with indole gave diagnostic peaks for indigoid pigments indigo (**11**) and indirubin (**12**). Both compounds are known to result from oxidative dimerization of indoxyl (**13**), the immediate product of indole hydroxylation. Notably, these pigments were not formed in the negative control experiment using Fre and heat-denatured XiaF (Supplementary Fig. 2b,III). Likewise, no pigments could be detected in the chromatographic profile after 1.5 h incubation of XiaF alone; only when the assay was supplemented with the reductase partners Fre or XiaP, **11** as well as **12** were formed in considerable amounts (Fig. 3b,II and Supplementary Fig. 2b,IV). These results underpin that XiaF works in a two-component system together with a flavin reductase.

In addition to these *in vitro* experiments, we found that the same indigoid pigments are also formed *in vivo*. On extended incubation of *E. coli* plate cultures that harbour the *xiaF*-expression plasmid (*E. coli*/pXU520), we noted a slight colouration of the colonies. Blue pigment production was substantially enhanced when *E. coli*/pXU520 was plated on agar supplemented with tryptophan (Fig. 4a). Since the colouration appeared without coexpression of *fre* or *xiaP* in *E. coli*, XiaF obviously recruits native *E. coli* enzymes for flavin reduction. The blue pigmentation was completely absent in control strains harbouring the empty vector only. For a full characterization of the blue pigments, cells were grown in liquid tryptophan medium<sup>47</sup>. HPLC-HRMS analysis with authentic references unequivocally showed that the pigments are identical with indigo (**11**) and indirubin (**12**) (Fig. 4b,c).

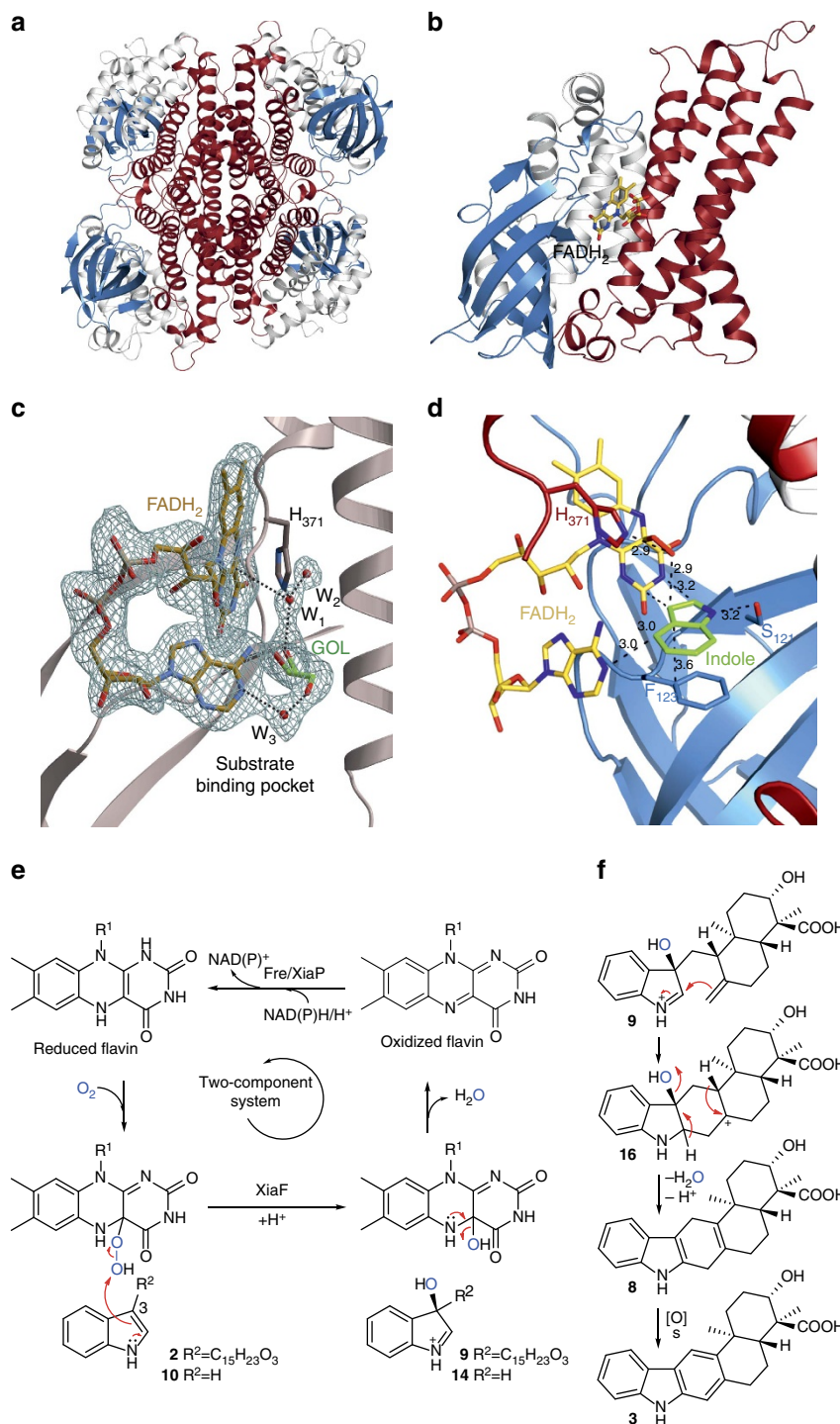
**XiaF crystal structure and model of cryptic hydroxylation.** To further characterize the non-canonical terpenoid cyclase and the cryptic hydroxylation reaction, we carried out protein crystallization and modelling experiments. The structure of XiaF was solved at 2.4 Å resolution applying molecular replacement ( $R_{\text{free}} = 22.9\%$ , search model: 4-hydroxyphenylacetate 3-hydroxylase, PDB ID: 2JBS<sup>48</sup>). XiaF (PDB entry codes: 5LVU, 5LVW and 5MR6) assembles as a tetrahedral oligomer, which is in agreement to the molecular mass of 176 kDa as determined by size-exclusion chromatography. The intermolecular interactions between the individual subunits are solely formed between the C-terminal domains, whereas the N-terminal as well as the central  $\beta$ -sheet domains are directed towards the corners of the tetrahedron (Fig. 5a). The N-domain spans from Met1-Lys118, the  $\beta$ -sheet from Leu119-Ser213 and the C domain from Gly214-Leu397 (Fig. 5b), which describes the typical architecture of a member of the ACAD superfamily (Supplementary Fig. 7).

Cocrystallizations of XiaF with flavin proved to be challenging; initially no FAD bound to crystals could be observed when the oxidized cofactor was used. A similar result has been reported for FMN and the XiaF homologue C2-HpaH (4-hydroxyphenylacetate 3-hydroxylase), which has been shown to bind reduced FMN considerably better than oxidized FMN<sup>48</sup>. Eventually, on incubation of XiaF with its cognate reductase XiaP under anaerobic conditions, we succeeded in binding reduced FADH<sub>2</sub> to the active site (PDB entry code: 5LVW and 5MR6, Fig. 5c and Supplementary Fig. 12). This result corroborates the finding that XiaF needs a reductase that supplies the enzyme with reduced cofactor. The strong three-fold coordination of FADH<sub>2</sub>'s  $\beta$ -phosphate by Tyr144 (2.1 Å), Arg266 (2.0 Å) and Ala350 (3.0 Å) substantially contributes to the binding of the cofactor to the active site (Supplementary Fig. 14a). The hydrophobic part of the isoalloxazin ring is embedded within a hydrophobic pocket mainly comprised by Trp87 (3.1 Å), Tyr144 (4.2 Å) and Trp186



**Figure 4 | Indigo production catalysed by XiaF.** (a) Blue pigment production of *xiaF*-expressing *E. coli* (*E. coli*/pXU520) plated on agar supplemented with tryptophan. (b) Dimethylformamide extract of *E. coli*/pXU520, which turns red after several days of incubation; reduction of indigo yields indigo white and reveals the red indirubin that is no longer overlaid by the dark blue indigo. (c) HPLC-HRMS monitoring (extracted ion chromatograms) of extracts from cultures of *E. coli*/pXU520. *E. coli* carrying the empty vector (*E. coli*/pET-28b(+)) was used as a negative control. The profiles showing the reference compounds indigo (**11**) and indirubin (**12**) are composed of different measurements of pure compounds (colour coded). (d) Hydroxylation of indole (**10**) catalysed by XiaF yields indoxyl (**13**), which is spontaneously oxidized to yield indigo (**11**). Alternatively, isatin (**15**) is formed from indoxyl as a side reaction, which can condense with indoxyl to yield indirubin (**12**), the structural isomer of **11**.

(3.8 Å). The ring N5 nitrogen is coordinated by the hydroxyl group of Thr146 (3.1 Å), whereas N3 and O4 are both stabilized by the carbonyl group of Ser121 (both 3.0 Å). Interestingly, the XiaF structure in complex with FADH<sub>2</sub> and glycerol (PDB entry code: 5MR6) displays a unique bent conformation of the cofactor's adenine moiety, thus blocking the substrate-binding channel. The phosphoribosyl linker is hereby exposed to the solvent (Fig. 5c and Supplementary Fig. 12). A similar conformation of FAD has been previously reported for an FAD synthase from yeast<sup>49</sup>. This bent conformation is missing the distinct interactions between the FADH<sub>2</sub>'s  $\beta$ -phosphate and the amino acids Tyr144, Arg266 and Ala350. Instead, three new intramolecular H-bonds between the ribose and adenine moieties of the cofactor are formed (3.0, 2.6 and 3.0 Å; Supplementary Fig. 14b). However, control assays with FMN, which lacks the adenine moiety, and the characterization of XiaF's orthologue XiaI (ref. 21) have shown that FAD and FMN can both be used by



**Figure 5 | Structural analysis of XiaF and scheme of FADH<sub>2</sub>-mediated indole hydroxylation.** (a) Ribbon structure of the homotetramer (PDB entry code: 5LVU). (b) Close-up cartoon representation of a monomer with bound FADH<sub>2</sub> (PDB entry code: 5LVW). The N-terminal domain is coloured in white, the  $\beta$ -sheet domain in blue and the C-terminal domain in red. (c) Ribbon diagram of XiaF with the enlarged active site. The electron density represents a  $2F_o - F_c$  map with the displayed FADH<sub>2</sub>, glycerol (GOL) and three water molecules (W<sub>1-3</sub>) omitted for phasing (PDB ID: 5MR6). The amino acid is labelled by the one-letter code and numbered according to the XiaF sequence. Black dashed lines indicate hydrogen bonds. (d) Docking of indole (carbon atoms coloured in green), using AutoDock Vina<sup>50</sup>, with XiaF in complex with the modelled reaction intermediate FAD-OOH. Distances in Å. (e) Catalytic cycle of XiaF consuming one molecule of oxygen and of reduced flavin provided by cognate reductase, yielding a C-3-hydroxylated indole derivative. The oxidized cofactor is transferred back to the reductase where it is recycled in a NAD(P)H/H<sup>+</sup>-dependent reaction. (f) Nucleophilic attack of the exo-methylene group on 3-hydroxyiminium cation **9** results in the pentacyclic intermediate **16**. Deprotonation and loss of an equivalent of water yields prexiamycin (**8**), which aromatizes to yield **3**. Spontaneous reactions are indicated by 's'.

the enzymes, a feature that is also shared by XiaF's homologues C2-HpaH and HsaA from *Mycobacterium tuberculosis*<sup>32,48</sup>. Therefore, it is unclear whether this unique bent conformation has a biological function. As described previously for the XiaF homologue C2-HpaH<sup>48</sup>, the highly conserved His371 (3.7 Å from C4) sits above the FADH<sub>2</sub> ligand, perfectly positioned for stabilizing, orienting and protonating the FAD-hydroperoxide intermediate.

In course of the reaction, the distal electrophilic oxygen would be attacked by the double bond (C2–C3) of the indole moiety, thus leading to a hydroxylation of the C3 position. In case of the XiaF:FADH<sub>2</sub>:glycerol complex structure, the glycerol molecule exactly occupies the substrate position, as shown by a structural alignment with the XiaF homologue 4-hydroxyphenylacetate 3-hydroxylase (Supplementary Fig. 13, PDB ID: 2JBT). Furthermore, a water molecule is directly placed between the C4 of FADH<sub>2</sub> (3.7 Å) and His371 (2.7 Å), indicating the position of the hydroperoxide (Fig. 5c and Supplementary Fig. 14b). Thus, we performed docking experiments using AutoDock Vina<sup>50</sup> to place indole within the active site (Supplementary Fig. 15). This modelling highlights the role of Ser121 (3.2 Å) and Phe123 (3.6 Å) in positioning indole. Especially the  $\pi$ -stacking between the indole ring and Phe123 could be a key interaction catalyzing the specific hydroxylation, which ultimately leads to both indigo and xiamycin formation. Moreover, we used this structure to model the FAD–OOH reaction intermediate to highlight the hydroperoxide's potential binding geometry relative to the indole ligand (Fig. 5d). This modelling approach suggests that the hydroperoxide is located between His371 (2.9 Å) and the C3 position of the indole ring (2.9 Å), thus explaining XiaF's high specificity for hydroxylating indole.

A conservation mapping using the Dali server<sup>51</sup> revealed that XiaF is structurally closely related to the 4-hydroxyphenylacetate 3-hydroxylase C2-HpaH (root mean square deviation of C $\alpha$ -carbon atoms (r.m.s.d.) = 2.3 Å, PDB ID: 2JBT)<sup>48</sup>, the secosteroid hydroxylases HsaA from *Mycobacterium tuberculosis* (r.m.s.d. = 1.8 Å, PDB ID: 3AFE)<sup>32</sup>, the dibenzothiophene monooxygenase DszC from *Rhodococcus erythropolis* D-1 (r.m.s.d. = 2.6 Å, PDB ID: 3X0Y)<sup>52</sup> and the *N*-oxygenase KijD3 involved in *D*-kijanose biosynthesis (r.m.s.d. = 2.4 Å, PDB ID: 4KCF)<sup>53</sup> (Fig. 2c; for a primary sequence alignment of XiaF and crystallized homologues, see Supplementary Fig. 10). The superposition of these homologous two-component monooxygenases reveals a strong structural conservation of the overall topology, whereas the individual substrate-binding pockets comprise a diverse set of amino acids (Supplementary Fig. 7). In case of the *N*-oxygenase KijD3, the overall substrate-binding channel is uniquely orientated compared to the above-mentioned homologous enzymes, highlighting the enzymes' flexible active site architecture. It is noteworthy that the FADH<sub>2</sub> coordinating residues are mainly located on the  $\beta$ -sheet domain, while the substrate-binding pocket forming amino acids predominantly derive from the C domain. This structural separation of the cofactor and the ligand-binding elements might theoretically allow a structural plasticity of the substrate-binding pocket without disturbing the fine-tuned cofactor binding. In case of XiaF, the substrate-binding pocket is comprised by residues N91, L98, S121, F123, I237, M240, H371 and M373 (Supplementary Fig. 7a).

## Discussion

In this study, we have gained insights into the structure, function and evolutionary relationship of a non-canonical terpenoid cyclase. Our findings illuminate the mechanism of the key step of this rare bacterial indoloterpenoid pathway, provide a

biocatalyst for the production of indigoid pigments and shed more light on the evolution of metabolic diversity.

*In vitro* characterization, bioinformatics and structural investigations revealed that XiaF is a new member of the group D FMOs, monooxygenases that receive reduced FAD/FMN from a NAD(P)H-dependent flavin reductase and assemble in a tetrameric ACAD-fold composed of N- and C-terminal helical domains and a central  $\beta$ -sheet domain<sup>35</sup>. Genome mining and biochemical analyses revealed the designated flavin reductase partner of XiaF in a two-component system. The successful *in vitro* reconstitution of the two-component system has laid the foundation for studying bacterial indoloterpenoid biosynthesis in more detail. In the current model for xiamycin biosynthesis, it was proposed that a hydroxylated intermediate would set the basis for terpenoid cyclization<sup>20–22</sup>. While bioinspired synthetic studies support this model<sup>25,54</sup>, experimental evidence for an enzymatic hydroxylation as trigger for the cyclization reaction has been missing. Using indole as substrate surrogate, we unveiled the 'cryptic hydroxylation' at C-3 of the indole ring, which is clearly indicated by the formation of indigoid pigments. Hydroxylation of indole followed by deprotonation of intermediate **14** (Fig. 5e) results in the formation of indoxyl, which is spontaneously oxidized to yield indigo (Fig. 4d). Alternatively, in an oxygen-rich environment isatin (**15**) is formed from indoxyl as a side reaction. Condensation of indoxyl with isatin produces indirubin, which is a common by-product of indigo biosynthesis<sup>55</sup>. These findings have important implications for the XiaF-catalysed reaction mechanism. Accordingly, in the terpenoid pathway C-3 hydroxylation of the indole ring of indosespene by C4a-hydroperoxyflavin as the electrophilic oxygenating species yields a reactive intermediate (most likely the 3-hydroxyiminium cation **9**, although deprotonation to yield an hydroxyindolenine intermediate is also conceivable) that is prone to nucleophilic attack of the *exo*-methylene group, resulting in the pentacyclic intermediate **16** (Fig. 5f). Deprotonation and loss of an equivalent of water yields prexiamycin, which spontaneously aromatizes to yield **3**.

Beyond these important mechanistic implications, our findings are also relevant from a biotechnological point of view. Bacterial indigo production is of considerable interest as a sustainable and environmentally friendly substitute for the synthetic production of indigo that depends on coal and oil and generates toxic byproducts<sup>56–59</sup>. Also, the second indigoid pigment, indirubin, is a valuable natural product<sup>58</sup>. Indirubin is not only an antimicrobial agent<sup>37,60</sup>, but also has therapeutic value against diseases and pathological conditions like cancer, Alzheimer's disease and inflammation<sup>61</sup>. In addition, indirubin is the main active ingredient of the plant mixture 'Danggui Longhui Wan', which has been used for millennia in traditional Chinese medicine to treat chronic diseases, likely because of its potent inhibition of cyclin-dependent kinases<sup>62,63</sup>. Because of these important bioactivities, recently there has been a renaissance of oxygenase screening activities towards indirubin production<sup>37,47,60,64,65</sup>. However, these enzymes generally seem to be rather promiscuous detoxification enzymes, not specialized indole oxygenases. In stark contrast, XiaF is to all appearance uniquely selective in hydroxylating the indole ring, since aromatic model xenobiotics have not been accepted as substrates. There seems to be the general trend that broad substrate acceptance comes at the price of low reaction turnover numbers<sup>66,67</sup>. Thus, a specialized indole oxygenase could be of particular interest for biotechnological applications, especially regarding large-scale indigoid pigment production. There is an unabated demand for indigo dye<sup>57</sup>, which is supplied by an annual world production of 80,000 tons of pigment<sup>68</sup>.



From an evolutionary point of view, the most intriguing discovery is that a terpenoid cyclase is functionally and structurally related to various enzymes from bacterial xenobiotic-degradation pathways. Xenobiotic-metabolizing enzymes have existed for more than 2.5 billion years<sup>69</sup>. Primordial cells have likely been exposed to many hostile chemical challenges and naturally occurring xenobiotics (for example, from volcano eruptions)<sup>70,71</sup> and the requirement of protective mechanisms against cytotoxic and genotoxic chemicals are certainly older than secondary metabolism. Moreover, besides their role in detoxification, xenobiotic-metabolizing systems have evolved in organisms like bacteria and fungi to tap new chemical resources by degradation of foreign aromatic hydrocarbons<sup>72</sup>. Therefore, one may conceive a metabolic crosstalk of the myriad of ancient broad-specific detoxification/degradation enzymes that are constitutively employed to guard against xenobiotics and/or utilize alternative sources of energy and elements on endogenous biosynthetic pathways.

Indole hydroxylation and subsequent indigo formation is a side reaction of many xenobiotic-metabolizing enzymes<sup>58,73</sup> that is often regarded as an indicator for general substrate tolerance towards aromatic hydrocarbons<sup>27,41,74</sup>. On the other hand, this reaction could also be the result of natural selection. In recent years, evidence is mounting that indole hydroxylation catalysed by these enzymes is of physiological relevance, since indole is toxic to microorganisms, and indigo formation by indole oxygenases represents a detoxification mechanism<sup>75–77</sup>. Moreover, the catalysis of indigo formation provides the missing link that connects XiaF with most of these catabolic oxygenases at a functional level.

On the basis of our structural data and docking experiments with indole, a potential binding mode of indole within the active site of XiaF could be revealed (Fig. 5d and Supplementary Fig. 13). A two-fold coordination of the substrate via an interaction of indole's N1 and Ser121 as well as  $\pi$ -stacking between the aromatic substrate and Phe123 seems to be most important for molecular recognition of this chemical group. A modelling of the FAD–OOH reaction intermediate suggests a coordination sphere of indole, which places the hydroperoxide exactly above and in close distance to indole's C3 position (Fig. 5d). This unique architecture may rationalize XiaF's specificity for hydroxylating indole. From a structural point of view, the potential of XiaF to utilize highly diverse substrates like indole and indosespene is explained by the fact that only a small, conserved proportion of the ligand is bound to the active site, close to the highly reactive FAD–OOH group. The remaining part of indosespene is bound to an open substrate channel (Supplementary Fig. 7). In case of the XiaF homologue 4-hydroxyphenylacetate 3-hydroxylase, the active site is limited by Arg263 (ref. 48). This amino acid is replaced in XiaF by Ile237, which opens up the binding channel for accepting the elongated indosespene substrate. Thus, the flexible active site architecture of these enzymes presumably builds the basis for the enzymes' potential utilization of various substrates.

We suggest that the promiscuous indole-metabolizing reaction, shared by many catabolic XiaF homologues might be the evolutionary origin of the cyclization reaction of indosespene (2). The most parsimonious explanation would be that XiaF has evolved from a broad-specific xenobiotic-metabolizing oxygenase that was able to accept the more complex indole compound 2, besides its ability to metabolize indole. In contrast to the formation of indoxyl, which results in the formation of indigoid pigments, hydroxylation of 2 at position C-3 triggers cyclization. In the course of evolution, XiaF likely specialized on this particular reaction. This is indicated by experiments to test the substrate flexibility of XiaF, for which

typical xenobiotic substrates of other XiaF homologues were not transformed.

Surprisingly, among the XiaF homologues are also the tentative polyketide-tailoring hydroxylases ActVA-ORF5 (ref. 78), and NcnH<sup>79</sup> and the glycoside-modifying *N*-oxygenases KijD3 (ref. 53) and DnmZ<sup>80</sup>. Thus, it appears that various enzymes that catalyse secondary metabolite tailoring reactions could have evolved from detoxification pathways, yet this relationship has been overlooked.

Finally, our hypothesis that *xiaF* originates from xenobiotic metabolism is substantiated by the fact that it clusters with genes (*xiaGHJ*) that encode homologues of xenobiotic-metabolizing enzymes. Therefore, it is tempting to speculate that this 'detox gene cassette' (Fig. 2a), which interrupts the core biosynthesis genes *xiaE* and *xiaKLM* originates from genes for broad-specific xenobiotic-metabolizing enzymes that were recruited for the diversification of the IST core scaffold preindosespene (1) (Fig. 1). The fact that xenobiotic-metabolizing enzymes are usually encoded on mobile genetic elements, such as transposons and plasmids, which are frequently shared by microorganisms via horizontal gene transfer within microbial communities<sup>81,82</sup>, together with the ability of these catabolic genomic islands to integrate into the chromosome after transfer<sup>82</sup> makes such a scenario even more plausible. Overall, these considerations along with the structural and phylogenetic analyses, the biochemical characterization and biotransformation experiments indicate that the specialized terpenoid cyclase presumably has evolved from the functionally related detoxification enzymes.

## Methods

**Organisms, cultivation conditions and plasmids.** The *E. coli* strains Top 10 (Life Technologies Corporation), BL21 (DE3) and BL21-CodonPlus(DE3)-RP (Agilent Technologies GmbH) were used for cloning and heterologous expression experiments. Electrocompetence of cells was achieved by repeated washing with diluted glycerol. Transformation was carried out in 2 mm cuvettes at 2.5 kV in a Gene Pulser Xcell electroporator (Bio-Rad) or according to the manual for commercially available chemical competent Top 10 cells (Life Technologies Corporation). Cells were grown in Luria broth (LB) medium, with appropriate antibiotics, whenever needed.

Plasmids used in this study were the pET-28a(+)-based expression vector (Merck KGaA) pET28-fre carrying a NAD(P)H-dependent flavin reductase gene (*fre*) combined with a 5'-sequence coding for a hexahistidine-tag, the pET-28b(+)-based expression vector pXU520 (for the production of recombinant XiaF with N-terminal histidine-tag), the pET-28c(+)-based expression vector pET28c-xiaP (for the production of recombinant XiaP with N-terminal histidine-tag) and the pET-26b(+)-based expression vector pET26b(+)-xiaF (for the production of recombinant XiaF with N-terminal histidine-tag used for crystallization).

**DNA preparation and analysis.** PCR or digested products were separated on agarose gels and purified from the gel using innuPREP Gel Extraction Kit (Analytic Jena). Digestion and ligation reactions were carried out by standard methods. Isolation of plasmid DNA was accomplished with standard alkaline lysis principle.

**Construction of expression vectors for *xiaF*, *xiaP* and *fre*.** For cloning the coding sequences of *xiaF* and *xiaP* downstream of the T7 promoter in pET28 vectors, the genes were PCR-amplified from cosmid 04b02 (from a genomic cosmid library of the endophyte) using primer sets XiaF\_NdeI\_fw (5'-GGAGTGGTG-CATATGACAGA-3', *NdeI* restriction site underlined) and XiaF\_rv (5'-CGA-CATCGGGATGAGAAAG-3') for *xiaF* and SK01 (5'-GCCATATGATCG-ACCGCGAGACCTTCG-3') and SK02 (5'-CCGGATCCCTCATCAACC-GCCGCGGC-3', *NdeI* and *BamHI* restriction sites underlined) for *xiaP*. For crystallization of XiaF, the gene sequence was cloned into vector pET-26b(+). For that, *xiaF* was PCR-amplified using pXU520 as template and SK14 (5'-GATCA-TATGCATCATCACCACCACATACAGACATCCGATCGGAG-3') and SK15 (5'-CTAGGATCCCTCATCAGAGGCTGGGCACGATGG-3', *NdeI* and *BamHI* restriction sites underlined) as primers. PCR was carried out with KAPAHiFi (PEQLAB) for *xiaF* and Phusion High-Fidelity DNA Polymerase (NEB) for *xiaP*. The amplicons were subcloned into the vector pCR-Blunt (Life Technologies) and the ligation mixture was introduced into electrocompetent *E. coli* Top 10 cells. Afterwards, plasmids were isolated from positive transformants and verified by sequencing. Excised DNA from correct plasmids was cloned by using the *NdeI* and *BamHI* restriction sites of pET-28b(+) or pET-26b(+) (for recombinant *xiaF*) and pET-28c(+) (for recombinant *xiaP*) to generate pXU520 and

pET26b(+)-xiaF (for the production of recombinant XiaF with N-terminal histidine-tag) and pET28c-xiaP (for the production of recombinant XiaP with N-terminal histidine tag). After introducing the ligation mixture into competent Top 10 cells, plasmids were isolated from positive transformants and verified by sequencing. Correct plasmids were introduced into *E. coli* BL21-CodonPlus(DE3)-RP for heterologous expression.

Amplification and cloning of *fre* from BL21 (DE3) were carried out as described by Zeng *et al.*<sup>83</sup> leading to a pET-28a(+)-based expression vector (pet28a-*fre*), which was introduced into electrocompetent *E. coli* BL21-CodonPlus(DE3)-RP cells for overexpression.

**Production and purification of recombinant XiaF and XiaP.** For heterologous protein production, *E. coli* BL21-CodonPlus(DE3)-RP carrying the plasmid pXU520 or pET26b(+)-xiaF for recombinant XiaF or pET28c-xiaP for recombinant XiaP were grown in LB medium supplemented with 34  $\mu\text{g ml}^{-1}$  chloramphenicol and 50  $\mu\text{g ml}^{-1}$  kanamycin at 30 °C and 180 r.p.m. to an OD<sub>600</sub> of 0.4–0.5. After induction with 0.1 mM isopropyl-1-thio- $\beta$ -D-galactoside (IPTG) for XiaF or 0.15 mM IPTG (for XiaP), cells were cultivated for 16 h at 15 °C and orbital shaking at 150 r.p.m. On collecting, the pellet from a 1.6-L-culture was resuspended in 50 ml cold XiaF lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 0.2 mg ml<sup>-1</sup> lysozyme, 110  $\mu\text{g ml}^{-1}$  PMSE, 0.1% Triton X-100, 1 mM DTT) or XiaP lysis buffer (10 mM Tris-HCl pH 9.0, 100 mM NaCl, 10% glycerol, 0.3 mg ml<sup>-1</sup> lysozyme, 0.5 mM PMSE, 0.1% Triton X-100, 1 mM DTT, 10 mM imidazole) and incubated for 30 min at 37 °C. Subsequently, the suspension was lysed by ultra-sonication with a Sonopuls HD 2,200 (BANDELIN). To separate cellular debris from soluble protein, the lysate was centrifuged.

For XiaF, soluble fraction was filtered (pore size 0.45  $\mu\text{m}$ ), diluted 1:1 with equilibration buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10% glycerol, 1 mM DTT, 10 mM imidazole) and subjected to a FPLC machine (Äkta pure25, GE Healthcare Life Science) equipped with a Protino Ni-NTA column (Macherey-Nagel GmbH & Co. KG). His-tagged XiaF was eluted with 250 mM imidazole in the same buffer. Fractions containing XiaF were concentrated and buffer was exchanged against assay buffer (20 mM Tris-HCl pH 8.0, 10% glycerol) using ultrafiltration in Amicon Ultra-15 Centrifugal Filter (MWCO 3,000 Da) (Merck Millipore) at 22 °C before either treatment with thrombin to cleave the hexahistidine-tag or a size-exclusion chromatography was performed. For cleavage, 10 U thrombin were added per mg of XiaF and the reaction was incubated for 16 h at 4 °C in cleavage buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 25 mM CaCl<sub>2</sub>). The cleaved His-tag was separated by gravity-flow column purification with Protino Ni-NTA agarose resin (Macherey-Nagel GmbH & Co. KG). Thrombin was removed with benzamide sepharose (GE Healthcare) in manual handling. Subsequently, tag-cleaved XiaF was concentrated in an Amicon Ultra-15 Centrifugal Filter (MWCO 3,000 Da) and dialysed with assay buffer at 22 °C. For size-exclusion chromatography (SEC) of His-tagged XiaF (44.5 kDa) and untagged XiaF (42.4 kDa), the protein solution was loaded onto a HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare Life Science) by using FPLC. Elution was carried out with 1.5 column volumes (180 ml) of XiaF SEC buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl). Immediately on elution, XiaF fractions were supplemented with 10% glycerol to avoid precipitation, concentrated by ultrafiltration in Amicon Ultra-15 Centrifugal Filter (MWCO 3,000 Da) and dialysed with assay buffer. Calibration of SEC was performed according to the manual of Gel Filtration Calibration Kits (GE Healthcare), which implicated that XiaF natively occurs as a homotetrameric protein.

For XiaP, the supernatant was filtered (pore size 0.45  $\mu\text{m}$ ) and loaded onto a gravity-flow column filled with 3 ml of Protino Ni-NTA agarose resin (Macherey-Nagel GmbH & Co. KG) in a manual handling. Elution was obtained with elution buffer (10 mM Tris-HCl pH 9.0, 100 mM NaCl, 10% glycerol, 500 mM imidazole). Fractions containing XiaP were desalted with PD-10 columns (GE Healthcare) and resolved in XiaP storage buffer (100 mM Na<sub>2</sub>HPO<sub>4</sub>, 100 mM Na<sub>2</sub>HPO<sub>4</sub>, 10% glycerol, pH 7.0). For SEC of His-tagged XiaP, the protein solution was subjected onto a HiLoad 16/600 Superdex 200 prep grade column using a FPLC machine. Elution was carried out with 1.5 column volumes (180 ml) of XiaP SEC buffer (100 mM Na<sub>2</sub>HPO<sub>4</sub>, 100 mM Na<sub>2</sub>HPO<sub>4</sub>, 2% glycerol, pH 7.0). The protein was then concentrated in Amicon Ultra-15 Centrifugal Filter (MWCO 3,000 Da) and dialysed with XiaP storage buffer.

The purity of both proteins was checked on SDS-PAGE (Supplementary Fig. 1) and sequences were confirmed with MALDI-TOF-MS(MS) (UltrafleXtreme, Bruker) after in-gel tryptic digestion. For that, bands were excised and digested with Trypsin/Lys-C Mix (Promega) using the corresponding protocol. Short-term storage at 4 °C was possible for 3 days (XiaF) or 2 weeks (XiaP). For long-time storage, proteins were shock frozen in liquid nitrogen and stored at -80 °C.

**Production and purification of recombinant Fre.** For overexpression of *fre*, a protocol based on Zeng *et al.*<sup>83</sup> was used. The transformant (BL21-CodonPlus (DE3)-RP-pET28a-*fre*) was grown in LB medium supplemented with 25  $\mu\text{g ml}^{-1}$  chloramphenicol and 35  $\mu\text{g ml}^{-1}$  kanamycin at 30 °C and 200 r.p.m. orbital shaking to an OD<sub>600</sub> of 0.4–0.6. Subsequently, cultures were induced with IPTG at a final concentration of 0.2 mM and incubated at 20 °C for 16 h. After collecting of cells, the pellet was resuspended in cold Fre lysis buffer (up to 50 ml for 1.6 l culture; 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 2 mM DTT) and

lysed by means of Sonopuls HD 2,200 (BANDELIN). Separated soluble fraction was subjected to 3 ml of Protino Ni-NTA agarose resin (Macherey-Nagel GmbH & Co. KG), and the mixture was shaken at 4 °C for 30 min before it was loaded into a gravity-flow column. Fre was eluted with 25–250 mM imidazole in Fre storage buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM EDTA, 2 mM DTT). Fractions containing Fre protein were concentrated using Amicon Ultra-15 Centrifugal Filter Units (MWCO 3,000 Da) (Merck KGaA) and dialysed with Fre storage buffer at 22 °C. The purity of His-tagged Fre (28.4 kDa) was checked on SDS-PAGE (Supplementary Fig. 1) and confirmed using in-gel tryptic digestion with MALDI-TOF-MS(MS). For long-term storage, protein was frozen at -80 °C in 50% glycerol.

**Activity assays of Fre and XiaP.** The recombinant enzymes Fre (derived from *E. coli*) and XiaP (derived from *Streptomyces* sp. HKI0576) were tested for NAD(P)H:flavin oxidoreductase activity. Reactions were set up in XiaF assay buffer (20  $\mu\text{M}$  Tris-HCl pH 8.0, 10% glycerol) containing 500  $\mu\text{M}$  NAD(P)H (Sigma-Aldrich), 50  $\mu\text{M}$  FAD/FMN/riboflavin (Sigma-Aldrich) and 1  $\mu\text{M}$  flavin reductase at room temperature, unless stated otherwise. Reaction rates were monitored by following the decrease of NAD(P)H absorbance at 340 nm caused by the oxidation to NAD(P) on the Tecan Safire<sup>2</sup> multi-detection microplate reader. Assays were carried out in a volume of 100  $\mu\text{l}$ . Reactions without enzyme or flavin served as negative controls. Blanc measurements were performed with XiaF assay buffer and subtracted from each value.

**In vitro biotransformation assays with XiaF.** To test the catalytic activity of XiaF, initial assays were performed with 25  $\mu\text{M}$  XiaF, 2 mM NADH, 1 mM FAD and 100  $\mu\text{M}$  indosespene in XiaF assay buffer at 28 °C for 24 h, which were supplemented with 10  $\mu\text{M}$  Fre during the course of the study. Control assays contained heat-denatured XiaF (10 min at 99 °C). To study the influence of a reductase partner on the substrate conversion by XiaF assays were performed, which contained 500  $\mu\text{M}$  NADH, 50  $\mu\text{M}$  FAD, 100  $\mu\text{M}$  indosespene and 10  $\mu\text{M}$  XiaF and 10  $\mu\text{M}$  Fre or 5  $\mu\text{M}$  XiaF and 0.5  $\mu\text{M}$  XiaP, in 100  $\mu\text{l}$  or 50  $\mu\text{l}$  total volumes. Reactions were carried out for at least 5 min at 28 °C. Control assays contained no flavin reductase supplement. Extraction of the assay was performed three times with two volumes of ethyl acetate each time. The combined extracts were dried by evaporation of the solvent with nitrogen. The residues were dissolved in methanol and analysed with HPLC-HRMS (Exactive Plus Orbitrap Mass Spectrometer, Thermo Scientific) operating with a 150  $\times$  2.1 mm 3  $\mu$  Betasil C18 column (Thermo Scientific) and a gradient from 5 to 98% solvent B in 15 min (solvent A, 0.3% (v/v) HCOOH in water; solvent B, 0.1% (v/v) HCOOH in acetonitrile) at a flow rate of 200  $\mu\text{l min}^{-1}$  and UV/HRMS detection. As alternative substrates, indole (conditions identical to indosespene assays) and substrates from homologous enzymes (4-nitrotoluene, 4-hydroxyphenylacetate, 2-naphthoic acid, indole 3-acetic acid, 3-hydroxybiphenyl) were tested in XiaF biotransformation assays. In the latter case, reaction volumes were increased up to 300  $\mu\text{l}$ . Maximum concentrations used were 50  $\mu\text{M}$  XiaF, 10  $\mu\text{M}$  XiaP and 2 mM substrates. For analysis, the reactions were stopped with one volume of acetonitrile (indole, 4-nitrotoluene) or methanol (indole 3-acetic acid), dried under reduced pressure, dissolved in acetonitrile or methanol and subjected to HPLC-HRMS. Reaction mixtures with 4-hydroxyphenylacetate, 2-naphthoic acid and 3-hydroxybiphenyl as substrates were extracted with ethyl acetate and analysed identical to indosespene assays. Control experiments were carried out without XiaF or with heat-denatured XiaF.

To test XiaF for flavin reductase activity reactions containing 15  $\mu\text{M}$  XiaF, 500  $\mu\text{M}$  NADH, 50  $\mu\text{M}$  FAD and 100  $\mu\text{M}$  indosespene and were performed at room temperature in XiaF assay buffer. Reaction rate was followed by the decrease in NADH absorbance at 340 nm on the Tecan Safire<sup>2</sup> multi-detection microplate reader. Controls were carried out without XiaF or with Fre. Blanc measurements were performed with XiaF assay buffer and subtracted from each value.

**Indigoid pigment production in vivo.** For pigment production on plate, an overnight culture of *E. coli* BL21-CodonPlus(DE3)-RP carrying the plasmid pXU520 (*E. coli*/pXU520) was streaked out on LB-agar plates supplemented with 0.1 mM IPTG, 10 mM tryptophan and 50  $\mu\text{g ml}^{-1}$  kanamycin. The plates were incubated at room temperature for several days until blue pigmentation occurred. *E. coli* BL21-CodonPlus(DE3)-RP carrying pET28c(+) (*E. coli*/pET-28b(+)) was used as a control. For pigment production in liquid culture, *E. coli*/pXU520 was grown in 100 ml tryptophan medium (2 g l<sup>-1</sup> tryptophan, 5 g l<sup>-1</sup> yeast extract, 10 g l<sup>-1</sup> NaCl)<sup>47</sup> supplemented with 50  $\mu\text{g ml}^{-1}$  kanamycin at 30 °C and 200 r.p.m. After reaching an OD<sub>600</sub> of 0.4, cells were induced with a final concentration of 0.5 mM IPTG and cultivated for 48 h. Centrifugation of the production culture resulted in a dark blue pellet. Extraction of this pellet with dimethylformamide yielded a dark blue extract, which turned reddish after several days of incubation. In addition, for pigment characterization, blue pellet was dried and extracted with acetonitrile. The extract was filtered and analysed by means of HPLC-HRMS (Exactive Plus Orbitrap Mass Spectrometer, Thermo Scientific) operating with a 150  $\times$  2.1 mm 3  $\mu$  Betasil C18 column (Thermo Scientific) and a gradient from 5 to 98% solvent B in 15 min (solvent A, 0.3% (v/v) HCOOH in water; solvent B, 0.1% (v/v) HCOOH in acetonitrile) at a flow rate of 200  $\mu\text{l min}^{-1}$

and UV/HRMS detection. Authentic references were used for comparison. *E. coli* PET28b(+) was used as a control.

**Determination of kinetic parameters of XiaP.** Flavin reductase (19,220 Da, sequence see below) activity was determined aerobically in XiaF assay buffer (20 mM Tris-HCl pH 8.0, 10% glycerol) at 28 °C in the dark. The reaction mixtures (100  $\mu$ l or 120  $\mu$ l) contained constant NAD(P)H and varying flavin concentrations (FAD, FMN, riboflavin), or *vice versa*, and 0.1  $\mu$ M XiaP. His-tagged XiaP was purified in three independent experiments by affinity chromatography (IMAC) followed by a gel filtration technique (active appearance as homodimer). Concentrations of all XiaP solutions were determined and equalized using Pierce BCA Protein Assay Kit (Thermo Scientific). Assays were initiated by addition of enzyme. The activity was measured by monitoring the NAD(P)H absorbance at 340 nm in 96-well plates on a Varioskan LUX plate reader (Thermo Scientific). Initial decrease in NAD(P)H concentration over time was used to calculate the conversion rate ( $\mu$ M min<sup>-1</sup> as  $v_0$ ). Apparent kinetic parameters were determined from double reciprocal plotting of initial velocities ( $v_0^{-1}$ ) and substrate concentration ( $[S]_0^{-1}$ ) and fitted by linear regression calculated with GraphPad Prism5. Three independent experiments with two technical replicates were used for calculation of kinetic parameters. For the determination of apparent kinetic values of NADH, 42  $\mu$ M FMN or riboflavin were added to 4.1, 12.5, 25, 41.6, 58.3, 83.3, 208, 416, 833, 1,250, 1,500, 1,667, 2,083 and 2,500  $\mu$ M of NADH ( $V_{\text{total}} = 120 \mu$ l) and absorbance was monitored for maximal 24 min. The apparent kinetic values of FAD, FMN or riboflavin were determined with varying concentrations of flavins (5, 10, 20, 40, 80, 100 and 150  $\mu$ M) and 1.5 mM NADH. Absorbance measurement was started 5 min after enzyme addition for 25 min. For each absorbance value, buffer absorbance was subtracted and final absorbance was converted into NAD(P)H concentration by a calibration curve. Since autocatalytic NAD(P)H oxidation in control assays (without XiaP) was minute in comparison to the XiaP enzyme assays (Supplementary Fig. 4a,d) its influence was neglected for the determination of the kinetic parameters of XiaP. Moreover, the impact of the different kinds of flavins on the absorbance was considered to be equal and was neglected for the determination of the kinetic parameters of XiaP as well. Flavins that were putatively bound to XiaP after purifications (noticeable by a very slight yellow colour of the solution) were neglected in the determination since control measurements of XiaP with added NAD(P)H did not lead to a decrease in absorbance as well as none of the heat-denatured enzyme solutions revealed a detectable absorbance at 450 nm.

Amino-acid sequence of His-tagged XiaP used for determination of kinetic parameters: MGSSEHHHHHSSGLVPRGSHMIDRETFVELMSGVCAPVT-VTTTATADGRPHGSTVSSFTSLSLDPLASFDLDRGSGLLPHLHQGARVG-VNIIAAHQHALAASFARRCPGTGSKFDGITWTRAGLPHLPDSAGWTAGRV-ERHVP GGDHVLVISVEEAHSTPAAPLV YLRRSFGTYSGLPAGG.

**Phylogenetic analysis.** The evolutionary history was inferred using the neighbour-joining method<sup>36</sup> and tested with the bootstrap test (1,000 replicates)<sup>84</sup>. The tree (Fig. 2b) is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method<sup>85</sup> and are in the units of the number of amino-acid differences per site. Evolutionary analyses were conducted in MEGA6 (ref. 86). The analysis involved the following enzymes (NCBI/PDB accession number in brackets): XiaF from *Streptomyces* sp. HKI0576 (CCH63732)<sup>22</sup>, XiaI from *Streptomyces* sp. SC5IO 02999 (AFK78075)<sup>21</sup>, XiaF' #1 from *Streptomyces olivaceus* (WP\_031042714)<sup>26</sup>, XiaF' #2 from *Streptomyces* sp. NRRL S-15 (WP\_031097971)<sup>26</sup>, XiaF' #3 from *Streptomyces fradiae* (WP\_043467134)<sup>26</sup>, XiaF' #4 from *Amycolatopsis nigrescens* CSC17Ta-90 (WP\_020673280)<sup>26</sup>, XiaF' #5 from *Streptomyces* sp. AA0539 (Streptomyces sp. AA0539)<sup>26</sup>, XiaF' #6 from *Streptomyces* sp. NRRL F-2890 (WP\_030727840)<sup>26</sup>, pJEC from uncultured bacterium JEC54 (AA54299)<sup>37</sup>, IpoA from *Rhodococcus* sp. T104 (AAS88875)<sup>38</sup>, AcdA<sub>pB7-2</sub> (BAP28583)<sup>27</sup>, AcdA<sub>pB6-2</sub> (BAP28542)<sup>27</sup>, NcnH from *Streptomyces arenae* (AAG44124)<sup>79</sup>, ActVA-ORF5 from *Streptomyces coelicolor* A3(2) (WP\_011030043)<sup>33</sup>, C2-HpaH from *Acinetobacter baumannii* (Q6Q272)<sup>43</sup>, HsaA #1 from *Mycobacterium tuberculosis* (WP\_003900101)<sup>32</sup>, HsaA #2 from *Rhodococcus jostii* RHA1 (Q05811)<sup>34</sup>, PheA from *Geobacillus stearothermophilus* (AAA85688)<sup>44</sup>, IdoA from *Pseudomonas alcaligenes* PA-10 (AF474963)<sup>42</sup>, IacA #1 from *Acinetobacter baumannii* ATCC 19,606 (EEX04702)<sup>87</sup>, IacA #2 from *Pseudomonas putida* (ABY62757)<sup>88</sup>, DszC #1 from *Rhodococcus erythropolis* IGTS8 (AAA99484)<sup>39</sup>, DszC #2 from *Bacillus subtilis* (BAC20182)<sup>40</sup>, KijD3 from *Actinomodura kijaniata* (4KCF A)<sup>53</sup>, DnmZ from *Streptomyces peucetius* (4ZXV A)<sup>80</sup>, IcpA<sub>clone M103</sub> (BAH29956)<sup>41</sup>, IcpA<sub>clone M123</sub> (BAH29955)<sup>41</sup>, BEC from *Cupriavidus necator* (AAK38074)<sup>89</sup>, NmoA from *Burkholderia* sp. JT1500 (CAD97420)<sup>30</sup> and HpaB from *Thermus thermophilus* HB8 (WP\_011228328) as an outgroup<sup>90</sup>. The result of the phylogenetic tree was confirmed with the minimal evolution method<sup>91</sup>, which was used as an alternative method. For more detailed cladograms, see Supplementary Figs 8 and 9.

**Crystallization and structure determination.** Crystals of XiaF were grown at 20 °C within 2 days to their final size applying the hanging drop (1  $\mu$ l  $\times$  1  $\mu$ l) vapour diffusion method. For XiaF:FADH<sub>2</sub> crystals, XiaF (127  $\mu$ M), XiaP (79  $\mu$ M),

NADH (500  $\mu$ M) and FAD (150  $\mu$ M) were anaerobically mixed inside a glove box and incubated for 25 min before setting up the crystallization trials. The XiaF:FADH<sub>2</sub> crystals appeared in 1.6 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100 mM HEPES/NaOH, pH 7.5 and 100 mM NaCl. Subsequently, crystals were transferred in 4  $\mu$ l drops containing the crystallization buffer plus 30% of glycerol. After 2 min, crystals were vitrified and stored in liquid nitrogen. For XiaF:FADH<sub>2</sub>:glycerol crystals, XiaF (193  $\mu$ M), XiaP (80  $\mu$ M), NADH (2.75 mM), FAD (480  $\mu$ M) and Na-dithionite (20 mM) were mixed and incubated for 60 min at 10 °C. Next, 1.1 mM of indospene was added and incubated for 15 min. This mixture was used for setting up the crystallization trials. The XiaF:FADH<sub>2</sub>:glycerol crystals appeared in 1.6 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100 mM HEPES/NaOH, pH 7.5, 100 mM NaCl and 5 mM Na-dithionite. Subsequently, crystals were transferred in 4  $\mu$ l drops containing the crystallization buffer plus 30% of glycerol and 5 mM of Na-dithionite. Diffraction data sets were recorded using synchrotron radiation of  $\lambda = 1.0 \text{ \AA}$  at the beamline X06SA, Swiss Light Source (SLS), Paul Scherrer Institut, Villigen, Switzerland. Collected reflections were processed using the programme package XDS. The statistics are summarized in Supplementary Table 3. Determination of the crystal structure was performed by molecular replacement using the programme PHASER. Coordinates of 4-hydroxyphenylacetate 3-hydroxylase (C2-HpaH) (PDB ID: 2JBR)<sup>48</sup> were applied as a suitable Patterson search model for the XiaF (apo) structure elucidation. The refined coordinates of XiaF in turn were employed for the XiaF structure in complex with FADH<sub>2</sub> as well as the complex with FADH<sub>2</sub> and glycerol. Notably, successful processing of the latter data set was only possible in the space group P1. Model building has been carried out with the graphic programme MAIN<sup>92</sup> and finalized applying REFMAC5 (ref. 93) by conventional crystallographic rigid body, positional and anisotropic temperature factor refinements. Water molecules were placed automatically in the 2F<sub>o</sub>-F<sub>c</sub>-electron density maps (contoured at 1 $\sigma$ ) using the software package ARP/wARP<sup>94</sup>. The achieved crystal structures possess crystallographic values of at least  $R_{\text{free}} = 22.9\%$ , r.m.s.d. bond lengths of 0.012  $\text{\AA}$ , and r.m.s.d. bond angle of 1.6° (Supplementary Table 3). The crystal structures were deposited at the RCSB Protein Data Bank under the accession codes 5LVU for XiaF (apo), 5LVW for XiaF (FADH<sub>2</sub>) and 5MR6 for XiaF (FADH<sub>2</sub> + glycerol).

**Docking of the indole ligand into the active site of XiaF.** For docking of the indole ligand, the XiaF structure in complex with FADH<sub>2</sub> and glycerol was chosen (PDB ID: 5MR6). The position of the glycerol ligand served as search centre for the docking experiment. The docking grid was restricted to the active site, which is limited by FADH<sub>2</sub>'s adenine moiety. The indole ligand was built and energetically minimized using the Avogadro software<sup>95</sup>. The docking simulation was performed using AutoDock Vina<sup>50</sup> as a plugin from the structure analysis software suite Chimera<sup>96</sup>. To prepare XiaF for the docking experiment, the plugin Dock Prep from the Chimera software suite was used. Hereby, solvent molecules were deleted and Gasteiger charges and hydrogens were added. For the latter ones, hydrogen bonds were considered as well, non-polar hydrogens were merged. Next, AutoDock Vina was run. Therefore, hydrogens were added to the indole ligand as well. The search was centred around the glycerol ligand coordinates ( $x = 49.1286$ ,  $y = -47.6138$ ,  $z = -0.4335$ ), the docking grid had a size of 10.0586  $\text{\AA} \times 9.5815 \text{ \AA} \times 7.8537 \text{ \AA}$ . For the receptor (XiaF), all of the following options were selected to be true: Merge charges and remove non-polar hydrogens, merge charges and remove lone pairs, ignore waters, ignore chains of non-standard residues, ignore all non-standard residues. For the ligand (indole), the following settings were selected to be true: Merge charges and remove non-polar hydrogens, merge charges and remove lone pairs. The number of binding modes was set to 10, the exhaustiveness of search was set to 8 and the maximum energy difference (kcal mol<sup>-1</sup>) was set to 3. In summary, 10 different conformations were computed, ranking from -5.3 (best hit) to -4.8 kcal mol<sup>-1</sup> (worst hit). The best hit was used for subsequent XiaF analysis (Fig. 5d and Supplementary Fig. 15). The choice to use this specific indole conformation was not solely based on its docking score, but also considering its orientation relative to the H<sub>2</sub>O molecule W<sub>2</sub> (Fig. 5c), which presumably complies with the orientation of the hydroperoxide group during catalysis. The second best score, for example, flips the ligand vertically, increasing the distance between the C3 position of indole and W<sub>2</sub>. The XiaF:FADH<sub>2</sub>:glycerol structure docked with indole was further analysed by modelling the reaction intermediate FAD-OOH using the graphic programme MAIN<sup>92</sup> (Fig. 5d).

**Data availability.** The coordinates and structure factors of XiaF have been deposited in the Protein Data Bank (<http://www.rcsb.org/pdb>) with accession numbers 5LVU, 5LVW and 5MR6. Any additional data not shown in this published article or in its Supplementary Information are available from the corresponding authors on reasonable request.

## References

- Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408–414 (2007).
- Davies, J. E. Specialized microbial metabolites: functions and origins. *J. Antibiot.* **66**, 361–364 (2013).
- Welsch, M. E., Snyder, S. A. & Stockwell, B. R. Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.* **14**, 347–361 (2010).

4. Clardy, J. & Walsh, C. Lessons from natural molecules. *Nature* **432**, 829–837 (2004).
5. Dewick, P. M. *Medicinal Natural Products: a Biosynthetic Approach* 550 (John Wiley & Sons, 2009).
6. Medema, M. H. et al. Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015).
7. Hertweck, C. Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology. *Trends Biochem. Sci.* **40**, 189–199 (2015).
8. Floss, H. G. Combinatorial biosynthesis—potential and problems. *J. Biotechnol.* **124**, 242–257 (2006).
9. Herrera, J. B. R., Wilson, W. K. & Matsuda, S. P. T. A tyrosine-to-threonine mutation converts cycloartenol synthase to an oxidosqualene cyclase that forms lanosterol as its major product. *J. Am. Chem. Soc.* **122**, 6765–6766 (2000).
10. Morrone, D., Xu, M. M., Fulton, D. B., Determan, M. K. & Peters, R. J. Increasing complexity of a diterpene synthase reaction with a single residue switch. *J. Am. Chem. Soc.* **130**, 5400–5401 (2008).
11. Roach, J. J. & Shenvi, R. A. Supramolecular catalysis: terpenes in tight spaces. *Nat. Chem.* **7**, 187–189 (2015).
12. Dickschat, J. S. Bacterial terpene cyclases. *Nat. Prod. Rep.* **33**, 87–110 (2016).
13. Baunach, M., Franke, J. & Hertweck, C. Terpenoid biosynthesis off the beaten track: unconventional cyclases and their impact on biomimetic synthesis. *Angew. Chem. Int. Ed.* **54**, 2604–2626 (2015).
14. Li, S. et al. Hapalindole/ambiguine biogenesis is mediated by a cope rearrangement, C–C bond-forming cascade. *J. Am. Chem. Soc.* **137**, 15366–15369 (2015).
15. Kim, S. H., Ha, T. K., Oh, W. K., Shin, J. & Oh, D. C. Antiviral indolosesquiterpenoid xiamycins C–E from a halophilic actinomycete. *J. Nat. Prod.* **79**, 51–58 (2016).
16. Ding, L. et al. Xiamycin, a pentacyclic indolosesquiterpene with selective anti-HIV activity from a bacterial mangrove endophyte. *Bioorg. Med. Chem. Lett.* **20**, 6685–6687 (2010).
17. Meng, Z. et al. Total synthesis and antiviral activity of indolosesquiterpenoids from the xiamycin and oridamycin families. *Nat. Commun.* **6**, 6096 (2015).
18. Baunach, M., Ding, L., Bruhn, T., Bringmann, G. & Hertweck, C. Regiodivergent N–C and N–N aryl coupling reactions of indoloterpenes and cycloether formation mediated by a single bacterial flavoenzyme. *Angew. Chem. Int. Ed.* **52**, 9040–9043 (2013).
19. Baunach, M., Ding, L., Willing, K. & Hertweck, C. Bacterial synthesis of unusual sulfonamide and sulfone antibiotics by flavoenzyme-mediated sulfur dioxide capture. *Angew. Chem. Int. Ed.* **54**, 13279–13283 (2015).
20. Ding, L., Maier, A., Fiebig, H. H., Lin, W. H. & Hertweck, C. A family of multicyclic indolosesquiterpenes from a bacterial endophyte. *Org. Biomol. Chem.* **9**, 4029–4031 (2011).
21. Li, H. et al. Identification and characterization of xiamycin A and oxiamycin gene cluster reveals an oxidative cyclization strategy tailoring indolosesquiterpene biosynthesis. *J. Am. Chem. Soc.* **134**, 8996–9005 (2012).
22. Xu, Z., Baunach, M., Ding, L. & Hertweck, C. Bacterial synthesis of diverse indole terpene alkaloids by an unparallelled cyclization sequence. *Angew. Chem. Int. Ed.* **51**, 10293–10297 (2012).
23. Zhang, Q. et al. N–N-coupled indolo-sesquiterpene atropo-diastereomers from a marine-derived actinomycete. *Eur. J. Org. Chem.* **2012**, 5256–5262 (2012).
24. Takada, K., Kajiwara, H. & Imamura, N. Oridamycins A and B, anti-*Saprolegnia parasitica* indolosesquiterpenes isolated from *Streptomyces* sp. KS84. *J. Nat. Prod.* **73**, 698–701 (2010).
25. Sun, Y. et al. Bioinspired total synthesis of sespenine. *Angew. Chem. Int. Ed.* **53**, 9012–9016 (2014).
26. Li, H. et al. Elucidating the cyclization cascades in xiamycin biosynthesis by substrate synthesis and enzyme characterizations. *Org. Lett.* **17**, 306–309 (2015).
27. Nagayama, H. et al. Isolation of oxygenase genes for indigo-forming activity from an artificially polluted soil metagenome by functional screening using *Pseudomonas putida* strains as hosts. *Appl. Microbiol. Biotechnol.* **99**, 4453–4470 (2015).
28. Omokoko, B., Jantges, U. K., Zimmermann, M., Reiss, M. & Hartmeier, W. Isolation of the *phe*-operon from *G. stearothermophilus* comprising the phenol degradative *meta*-pathway genes and a novel transcriptional regulator. *BMC Microbiol.* **8**, 197 (2008).
29. Li, Q. et al. Purification and characterization of a flavin reductase from the biodesulfurizing bacterium *Mycobacterium goodii* X7B. *Process Biochem.* **47**, 1144–1149 (2012).
30. Deng, D., Li, X., Fang, X. & Sun, G. Characterization of two components of the 2-naphthoate monooxygenase system from *Burkholderia* sp. strain JT1500. *FEMS Microbiol. Lett.* **273**, 22–27 (2007).
31. Gray, K. A. et al. Molecular mechanisms of biocatalytic desulfurization of fossil fuels. *Nat. Biotechnol.* **14**, 1705–1709 (1996).
32. Dresen, C. et al. A flavin-dependent monooxygenase from *Mycobacterium tuberculosis* involved in cholesterol catabolism. *J. Biol. Chem.* **285**, 22264–22275 (2010).
33. Valtou, J., Filisetti, L., Fontecave, M. & Niviere, V. A two-component flavin-dependent monooxygenase involved in actinorhodin biosynthesis in *Streptomyces coelicolor*. *J. Biol. Chem.* **279**, 44362–44369 (2004).
34. Van der Geize, R. et al. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proc. Natl Acad. Sci. USA* **104**, 1947–1952 (2007).
35. Huijbers, M. M., Montersino, S., Westphal, A. H., Tischler, D. & van Berkel, W. J. Flavin dependent monooxygenases. *Arch. Biochem. Biophys.* **544**, 2–17 (2014).
36. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
37. Lim, H. K. et al. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl. Environ. Microbiol.* **71**, 7768–7777 (2005).
38. Choi, K. Y. et al. Molecular cloning and identification of a novel oxygenase gene specifically induced during the growth of *Rhodococcus* sp. Strain T104 on limonene. *J. Microbiol.* **42**, 160–162 (2004).
39. Cocco, W. M. et al. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* **19**, 354–359 (2001).
40. Furuya, T., Takahashi, S., Ishii, Y., Kino, K. & Kirimura, K. Cloning of a gene encoding flavin reductase coupling with dibenzothiophene monooxygenase through coexpression screening using indigo production as selective indication. *Biochem. Biophys. Res. Commun.* **313**, 570–575 (2004).
41. Kimura, N., Sakai, K. & Nakamura, K. Isolation and characterization of a 4-nitrotoluene-oxidizing enzyme from activated sludge by a metagenomic approach. *Microbes Environ.* **25**, 133–139 (2010).
42. Alemayehu, D., Gordon, L. M., O’Mahony, M. M., O’Leary, N. D. & Dobson, A. D. Cloning and functional analysis by gene disruption of a novel gene involved in indigo production and fluoranthene metabolism in *Pseudomonas alcaligenes* PA-10. *FEMS Microbiol. Lett.* **239**, 285–293 (2004).
43. Thotsaporn, K., Sucharitakul, J., Wongratana, J., Suadee, C. & Chaiyen, P. Cloning and expression of *p*-hydroxyphenylacetate 3-hydroxylase from *Acinetobacter baumannii*: evidence of the divergence of enzymes in the class of two-protein component aromatic hydroxylases. *Biochim. Biophys. Acta* **1680**, 60–66 (2004).
44. Kim, I. C. & Oriel, P. J. Characterization of the *Bacillus stearothermophilus* BR219 phenol hydroxylase gene. *Appl. Environ. Microbiol.* **61**, 1252–1256 (1995).
45. Arand, M. et al. Structure of *Rhodococcus erythropolis* limonene-1,2-epoxide hydrolase reveals a novel active site. *EMBO J.* **22**, 2583–2592 (2003).
46. Zhang, Q. et al. Carboxyl formation from methyl via triple hydroxylations by XiaM in xiamycin A biosynthesis. *Org. Lett.* **14**, 6142–6145 (2012).
47. Han, G. H., Gim, G. H., Kim, W., Seo, S. I. & Kim, S. W. Enhanced indirubin production in recombinant *Escherichia coli* harboring a flavin-containing monooxygenase gene by cysteine supplementation. *J. Biotechnol.* **164**, 179–187 (2012).
48. Alfieri, A. et al. Structure of the monooxygenase component of a two-component flavoprotein monooxygenase. *Proc. Natl Acad. Sci. USA* **104**, 1177–1182 (2007).
49. Leulliot, N. et al. Crystal structure of yeast FAD synthetase (Fad1) in complex with FAD. *J. Mol. Biol.* **398**, 641–646 (2010).
50. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
51. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
52. Guan, L. J. et al. Crystal structures of apo-DszC and FMN-bound DszC from *Rhodococcus erythropolis* D-1. *FEBS J.* **282**, 3126–3135 (2015).
53. Thoden, J. B., Branch, M. C., Zimmer, A. L., Bruender, N. A. & Holden, H. M. Active site architecture of a sugar *N*-oxygenase. *Biochemistry* **52**, 3191–3193 (2013).
54. Sun, Y. et al. A concise total synthesis of sespenine, a structurally unusual indole terpenoid from *Streptomyces*. *Org. Chem. Front.* **3**, 368–374 (2016).
55. Maugard, T., Enaud, E., Choisy, P. & Legoy, M. D. Identification of an indigo precursor from leaves of *Isatis tinctoria* (woad). *Phytochemistry* **58**, 897–904 (2001).
56. Berry, A., Dodge, T. C., Pepsin, M. & Weyler, W. Application of metabolic engineering to improve both the production and use of biotech indigo. *J. Ind. Microbiol. Biotechnol.* **28**, 127–133 (2002).
57. Chavan, R. B. in *Denim: Manufacture, Finishing and Applications* (ed. Paul, R.) Ch. 3, 37–67 (Woodhead Publishing, 2015).
58. Han, X., Wang, W. & Xiao, X. Microbial biosynthesis and biotransformation of indigo and indigo-like pigments. *Chin. J. Biotechnol.* **24**, 921–926 (2008).

59. O'Connor, K. E., Dobson, A. D. & Hartmans, S. Indigo formation by microorganisms expressing styrene monooxygenase activity. *Appl. Environ. Microbiol.* **63**, 4287–4291 (1997).
60. MacNeil, I. A. *et al.* Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* **3**, 301–308 (2001).
61. McClay, K., Boss, C., Keresztes, I. & Steffan, R. J. Mutations of toluene-4-monooxygenase that alter regiospecificity of indole oxidation and lead to production of novel indigoid pigments. *Appl. Environ. Microbiol.* **71**, 5476–5483 (2005).
62. Hoessel, R. *et al.* Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependent kinases. *Nat. Cell Biol.* **1**, 60–67 (1999).
63. Wang, Y., Hoi, P. M., Chan, J. Y. & Lee, S. M. New perspective on the dual functions of indirubins in cancer therapy and neuroprotection. *Anticancer Agents Med. Chem.* **14**, 1213–1219 (2014).
64. Zhang, X. *et al.* Production of indirubin from tryptophan by recombinant *Escherichia coli* containing naphthalene dioxygenase genes from *Comamonas* sp. MQ. *Appl. Biochem. Biotechnol.* **172**, 3194–3206 (2014).
65. Rui, L., Reardon, K. F. & Wood, T. K. Protein engineering of toluene *ortho*-monooxygenase of *Burkholderia cepacia* G4 for regiospecific hydroxylation of indole to form various indigoid compounds. *Appl. Microbiol. Biotechnol.* **66**, 422–429 (2005).
66. Carbonell, P., Lecoindre, G. & Faulon, J.-L. Origins of specificity and promiscuity in metabolic networks. *J. Biol. Chem.* **286**, 43994–44004 (2011).
67. Jakoby, W. B. & Ziegler, D. M. The enzymes of detoxication. *J. Biol. Chem.* **265**, 20715–20718 (1990).
68. Ghaly, A. E., Ananthashankar, R., Alhattab, M. & Ramakrishnan, V. V. Production, characterization and treatment of textile effluents: a critical review. *J. Chem. Eng. Proc. Technol.* **5**, 1–18 (2014).
69. Nebert, D. W. & Dieter, M. Z. The evolution of drug metabolism. *Pharmacology* **61**, 124–135 (2000).
70. Coleman, J. O. D., BlakeKalfif, M. M. A. & Davies, T. G. E. Detoxification of xenobiotics by plants: chemical modification and vacuolar compartmentation. *Trends Plant Sci.* **2**, 144–151 (1997).
71. Menzie, C. A., Potocki, B. B. & Santodonato, J. Exposure to carcinogenic PAHs in the environment. *Environ. Sci. Technol.* **26**, 1278–1284 (1992).
72. Mishra, V., Lal, R. & Srinivasan. Enzymes and operons mediating xenobiotic degradation in bacteria. *Crit. Rev. Microbiol.* **27**, 133–166 (2001).
73. O'Connor, K. E. & Hartmans, S. Indigo formation by aromatic hydrocarbon-degrading bacteria. *Biotechnol. Lett.* **20**, 219–223 (1998).
74. Fernández-Arrojo, L., Guazzaroni, M.-E., López-Cortés, N., Beloqui, A. & Ferrer, M. Metagenomic era for biocatalyst identification. *Curr. Opin. Biotechnol.* **21**, 725–733 (2010).
75. Lee, J., Jayaraman, A. & Wood, T. K. Indole is an inter-species biofilm signal mediated by SdiA. *BMC Microbiol.* **7**, 42 (2007).
76. Lee, J. H. & Lee, J. Indole as an intercellular signal in microbial communities. *FEMS Microbiol. Rev.* **34**, 426–444 (2010).
77. Lin, G. H., Chen, H. P. & Shu, H. Y. Detoxification of indole by an indole-induced flavoprotein oxygenase from *Acinetobacter baumannii*. *PLoS ONE* **10**, e0138798 (2015).
78. Taguchi, T. *et al.* Biosynthetic conclusions from the functional dissection of oxygenases for biosynthesis of actinorhodin and related *Streptomyces* antibiotics. *Chem. Biol.* **20**, 510–520 (2013).
79. Brunker, P., Sterner, O., Bailey, J. E. & Minas, W. Heterologous expression of the naphthocyclinone hydroxylase gene from *Streptomyces arenae* for production of novel hybrid polyketides. *Anton. Leeuw.* **79**, 235–245 (2001).
80. Al-Mestarihi, A., Romo, A., Liu, H. W. & Bachmann, B. O. Nitrosynthase-triggered oxidative carbon-carbon bond cleavage in baumycin biosynthesis. *J. Am. Chem. Soc.* **135**, 11457–11460 (2013).
81. Diaz, E. Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *Int. Microbiol.* **7**, 173–180 (2004).
82. Springael, D. & Top, E. M. Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol.* **12**, 53–58 (2004).
83. Zeng, J. & Zhan, J. A novel fungal flavin-dependent halogenase for natural product biosynthesis. *ChemBioChem.* **11**, 2119–2123 (2010).
84. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
85. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics* (Oxford University Press, 2000).
86. Tamura, K., Stecher, G., Peterson, D., Filipksi, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
87. Lin, G. H. *et al.* Identification and characterization of an indigo-producing oxygenase involved in indole 3-acetic acid utilization by *Acinetobacter baumannii*. *Anton. Leeuw.* **101**, 881–890 (2012).
88. Leveau, J. H. & Gerards, S. Discovery of a bacterial gene cluster for catabolism of the plant hormone indole 3-acetic acid. *FEMS Microbiol. Ecol.* **65**, 238–250 (2008).
89. Drevlo, S., Bramer, C. O., Madkour, M., Mayer, F. & Steinbuechel, A. Cloning and expression of a *Ralstonia eutropha* HF39 gene mediating indigo formation in *Escherichia coli*. *Appl. Environ. Microbiol.* **67**, 1964–1969 (2001).
90. Kim, S. H. *et al.* Crystal structure of the oxygenase component (HpaB) of the 4-hydroxyphenylacetate 3-monooxygenase from *Thermus thermophilus* HB8. *J. Biol. Chem.* **282**, 33107–33117 (2007).
91. Rzhetsky, A. & Nei, M. A simple method for estimating and testing minimum-evolution trees. *Mol. Biol. Evol.* **9**, 945–967 (1992).
92. Turk, D. MAIN software for density averaging, model building, structure refinement and validation. *Acta Crystallogr. D* **69**, 1342–1357 (2013).
93. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60**, 2184–2195 (2004).
94. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
95. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 1–17 (2012).
96. Pettersen, E. F. *et al.* UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

## Acknowledgements

We thank A. Perner for HRMS measurements, M. Poetsch for MALDI experiments and the staff of the beamline X06SA at the Paul Scherrer Institute, SLS, Villigen (Switzerland) for assistance during data collection. We are grateful for financial support by the SFB749 (M.G.) and the International Leibniz Research School for Microbial and Biomolecular Interactions (ILRS, to S.S.).

## Author contributions

All authors were involved in the design of the research; S.K., M.B., S.S., M.I.-I. and Z.X. performed biochemical experiments, M.B. and P.B. performed bioinformatic analyses; P.B. and M.G. performed protein crystallography and analysed data; P.B. performed docking simulations; all authors contributed to the early drafts of the manuscript; C.H. wrote the final version.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Kugel, S. *et al.* Cryptic indole hydroxylation by a non-canonical terpenoid cyclase parallels bacterial xenobiotic detoxification. *Nat. Commun.* **8**, 15804 doi: 10.1038/ncomms15804 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017