Enzymatic hydroxylation of an unactivated methylene C–H bond guided by molecular dynamics simulations

Alison R. H. Narayan[1], Gonzalo Jiménez-Osés[2], Peng Liu[2], Solymar Negretti[3], Wanxiang Zhao[4], Michael M. Gilbert[4], Raghunath O. Ramabhadran[2], Yun-Fang Yang[2], Lawrence R. Furan[2], Zhe Li[2], Larissa M. Podust[5], John Montgomery[3,4]★, K. N. Houk[2]★ and David H. Sherman[1,3,4,6]★

The hallmark of enzymes from secondary metabolic pathways is the pairing of powerful reactivity with exquisite site selectivity. The application of these biocatalytic tools in organic synthesis, however, remains under-utilized due to limitations in substrate scope and scalability. Here, we report how the reactivity of a monooxygenase (PikC) from the pikromycin pathway is modified through computationally guided protein and substrate engineering, and applied to the oxidation of unactivated methylene C–H bonds. Molecular dynamics and quantum mechanical calculations were used to develop a predictive model for substrate scope, site selectivity and stereoselectivity of PikC-mediated C–H oxidation. A suite of menthol derivatives was screened computationally and evaluated through in vitro reactions, where each substrate adhered to the predicted models for selectivity and conversion to product. This platform was also expanded beyond menthol-based substrates to the selective hydroxylation of a variety of substrate cores ranging from cyclic to fused bicyclic and bridged bicyclic compounds.

D irect C–H functionalization has the potential to streamline existing synthetic routes and also to provide access to novel, high-value compounds. Research in this area has grown exponentially over the past two decades, but the challenge of selectively targeting any given C–H bond for functionalization within a complex molecule has yet to be mastered[1]. Strategies for targeting the electronically weakest C–H bond or the most sterically accessible C–H bond have proven fruitful. A plethora of directing groups has also been developed to override the innate steric and electronic biases within a molecule to affect a reaction proximal to the directing group. Despite these efforts, certain types of C–H bonds remain difficult to target or entirely inaccessible, such as methylene C–H bonds remote from directing groups. Although some successes have been realized in site-selective methylene oxidations using small-molecule catalysts[2–4], the selective oxidation of a single C–H bond among several electronically and sterically similar methylene units remains one of the most challenging tasks in the field of C–H functionalization. Enzymes have the potential to offer orthogonal reactivity compared to small-molecule catalysts. In the C–H functionalization arena, cytochrome P450 biocatalysts have proven to be effective for the selective oxidation of unactivated methylene C–H bonds[5–9]. However, the development of these biocatalysts often requires extensive protein engineering to achieve selectivity on a given substrate or to change the substrate scope of the enzyme[10–13]. Here, we demonstrate a method for addressing this challenge that relies on a directing group distal to the site of functionalization and overrides steric or electronic effects inherent to the substrate with the ultimate goal of achieving a P450/directing group platform that can be broadly applied without requiring extensive protein engineering for each new substrate. While such a synthetic strategy using small-molecule catalysts originates from the pioneering early studies by Breslow[14,15] and has been applied
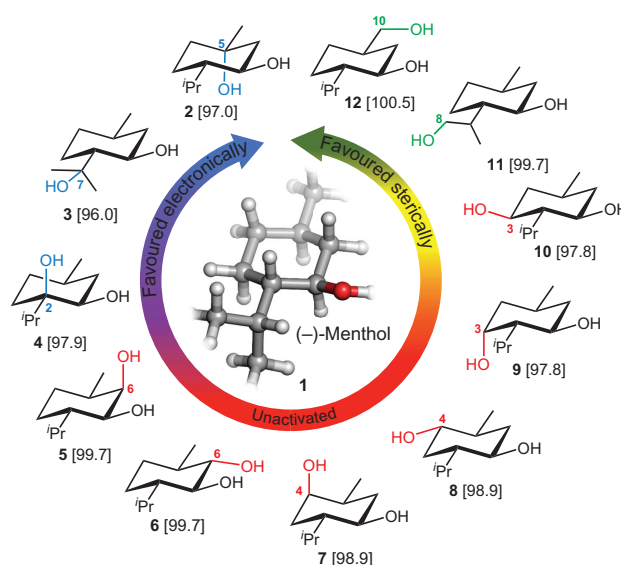

Figure 1 | Selected oxidation products of (–)-menthol and calculated bond dissociation energies for the corresponding C–H bonds (in kcal mol⁻¹). The calculated BDEs display a general trend in C–H bond strength, tertiary < secondary < primary.

[1]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109, USA. [2]Department of Chemistry, University of California, Los Angeles, California 90095, USA. [3]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. [4]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. [5]Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California, San Diego, California 92093, USA. [6]Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan 48109, USA. *e-mail: davidhs@umich.edu; houk@chem.ucla.edu; jmontg@umich.edu
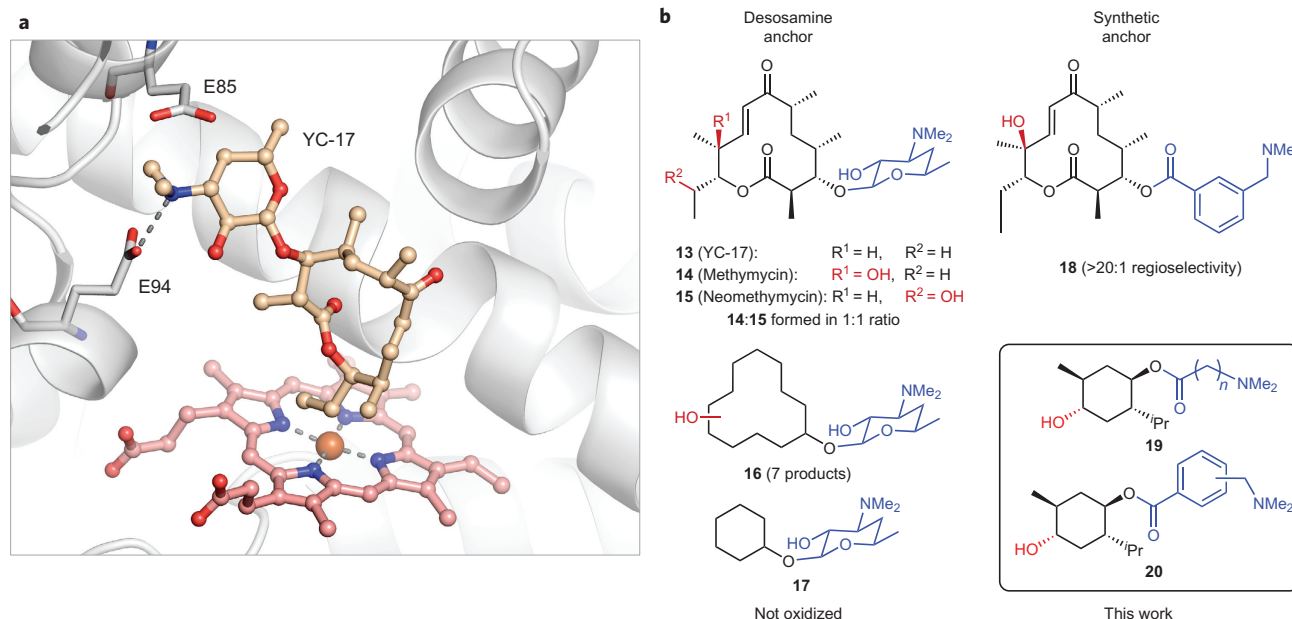
1

Figure 2 | Substrate anchoring mechanism employed by P450 PikC. a, Co-crystal structure of natural substrate, YC-17 and PikC (PDB ID 2CD8)[33] depicting the anchoring salt bridge between E94 and the dimethylamino group of YC-17. **b**, Evolution of the PikC substrate-engineering approach with the natural anchoring group, desosamine, and dimethylamine-containing synthetic anchoring groups.

in recent efforts to achieve *meta*-functionalizations of arenes[16,17], the architectural complexity and unique anchoring mechanism of biological catalysts provides what is perhaps the ideal platform for engineering directed reactions to functionalize remote and unactivated structural elements[18–20].

Menthol (**1**) was chosen as a model substrate based on its six-membered ring core, the variety of C–H bonds it contains (up to 18 potentially reactive positions, including primary, secondary and tertiary; Fig. 1) and the depth of information available on its use in C–H hydroxylation studies. Several groups have demonstrated the ability to hydroxylate the weaker methine C–H bonds at the C5 and C7 positions of menthol using Fe, Cr or Mn chemocatalysts[21–25] or biotransformations **2** and **3**, respectively (Fig. 1). The stronger but sterically less hindered methyl C–H bonds have also been successfully targeted for C–H oxygenation (**11** and **12**) through group transfer reactions[26,27] or biotransformation[28]. However, selective functionalization of any of the six methylene C–H bonds of menthol has remained an exceptional chemical and biochemical challenge. Biotransformation studies have established the feasibility of a biocatalytic methylene oxidation of menthol to afford mixtures of six or more regioisomeric products[29], but to our knowledge, the only example of site-selective methylene hydroxylation of menthol involved a directed evolution approach, in which Fasan and co-workers achieved a site- and stereoselective hydroxylation at C6 (see **5**) with a cytochrome P450 BM3 variant that was identified using a screening-based fingerprinting approach[8].

Our strategy was motivated by the selective oxygenation of unactivated C–H bonds catalysed by the cytochrome P450 monooxygenase PikC from the pikromycin natural product biosynthetic pathway[30] and exploits the native substrate anchoring mechanism used by this enzyme. PikC hydroxylates two types of endogenous substrates in *Streptomyces venezuelae* ATCC 15439. They include the 12- and 14-membered macrolides YC-17 (**13**) and narbomycin[31] to produce methymycin (**14**), neomethymycin (**15**), novamethymycin[32], pikromycin and neopikromycin. This level of substrate promiscuity is rare in secondary metabolic pathways, but is explained by the mechanism in which natural substrates bind within the PikC active site. Co-crystal structures of PikC with YC-17 (**13**) or narbomycin revealed salt bridge interactions between the protonated dimethylamino group of the substrate desosamine sugar and an exposed carboxylate moiety within the active site (Fig. 2a, E94 for YC-17 and E85 for narbomycin)[33,34]. To assess the ability of salt bridge-mediated anchoring of unnatural substrates in the PikC active site, several desosamine-containing substrates were synthesized and tested for oxidation with PikC (Fig. 2b)[35]. Macrocyclic substrates were hydroxylated to afford multiple products (**16**). However, no reaction was observed with smaller cores such as desosaminyl cyclohexanol **17**. We reasoned that the six-membered ring in **17** was too small to span the distance between the anchoring carboxylate residue and the PikC iron-oxo species where the reaction occurs. We previously demonstrated that the desosamine in the natural substrate YC-17 might be effectively replaced with synthetic alternatives (**18**, Fig. 2b)[36]. Thus, we envisioned that employing synthetic anchoring groups as well as introducing additional mutations in the biocatalyst could improve the activity of PikC with unnatural substrates.

Access to high-resolution co-crystal structures of PikC and its natural macrolide substrates has been instrumental in identifying the substrate–enzyme salt bridge at the centre of our substrate-engineering strategy. Moreover, ligand-free X-ray structures (PDB ID 2BVJ)[33] illustrate the differences in protein conformations attributed to a transition between the catalytically competent 'closed' conformation and the 'open' conformation, a prerequisite of the substrate binding and product release. However, these static conformation snapshots do not inform the dynamic nature of the protein–small molecule interactions. This has limited our view of the enzyme active site and the dynamic signature of the protein–ligand complex ultimately responsible for the unique site and stereoselectivity characteristics of the PikC biocatalyst. To address these experimental limitations we turned to molecular dynamics (MD) simulations, the powerful computational tool that has proven to be useful in determining the structural origins of enhanced activity in evolved enzymes[37]. We conceptualized a suite of (–)- and (+)-menthol derivatives bearing synthetic anchoring groups including linear alkyl chains of various lengths (**21–25**) as well as benzylic amines (**26–28**). The binding efficiency of selected substrates (**21**, **25** and **26**) to wild-type PikC and PikC$_{D50N}$ mutant was assessed by 0.5 µs MD simulations (Supplementary Methods).
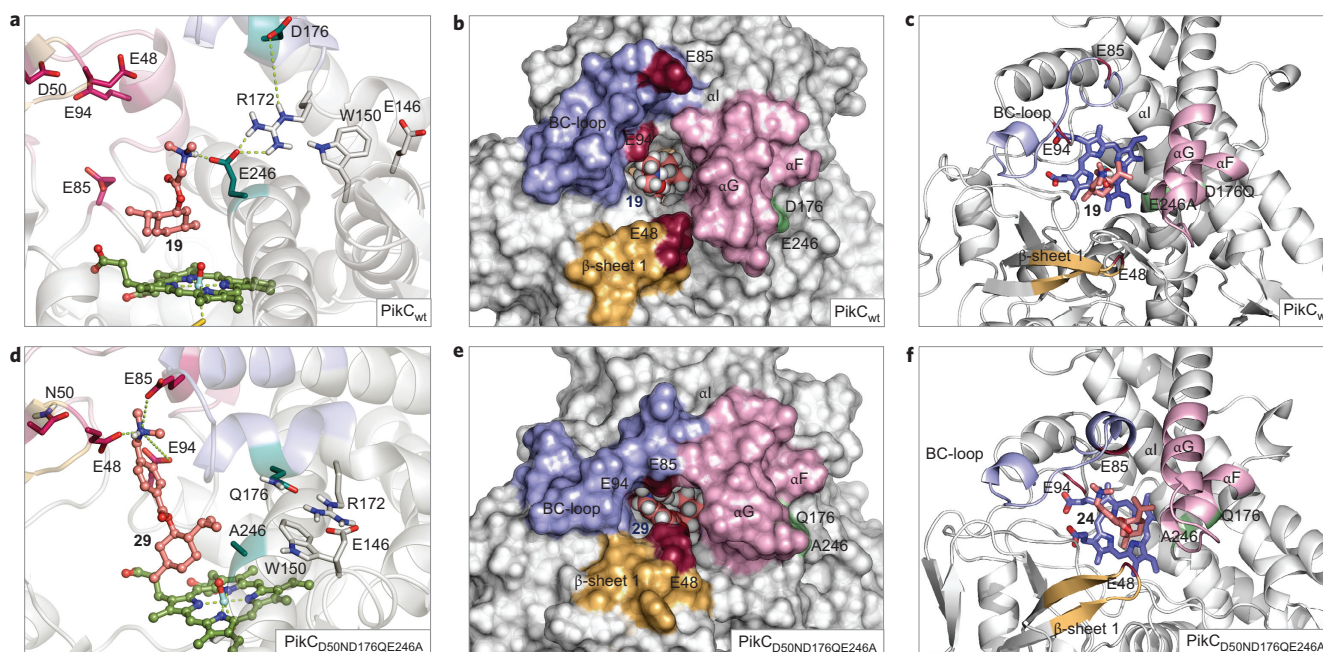
Figure 3 | Substrate binding and tertiary structure of PikC variants observed in MD simulations. a, Improper binding of substrate (−)-**21** to PikC$_{wt}$ showing detrimental interactions with E246, absence of contacts with E94, E85 or E48, and disruption of the hydrogen-bond network in the R172/E146/W150 motif. **b,c**, Space-filling representation (**b**) and ribbon representation (**c**) of PikC$_{wt}$ in open conformation showing an open active site with widely spread binding residues (E94, E85 and E48). **d**, Adequate binding of substrate (−)-**26** to PikC$_{D50ND176QE246A}$ showing interactions with the all three binding residues simultaneously and the undistorted R172/E146/W150 motif. **e,f**, Space-filling representation (**e**) and ribbon representation (**f**) of PikC$_{D50ND176QE246A}$ in closed conformation showing a narrow active site channel and a tight arrangement of the binding residues.
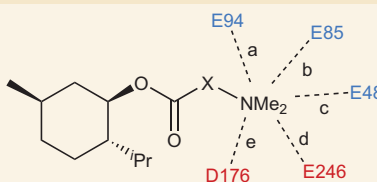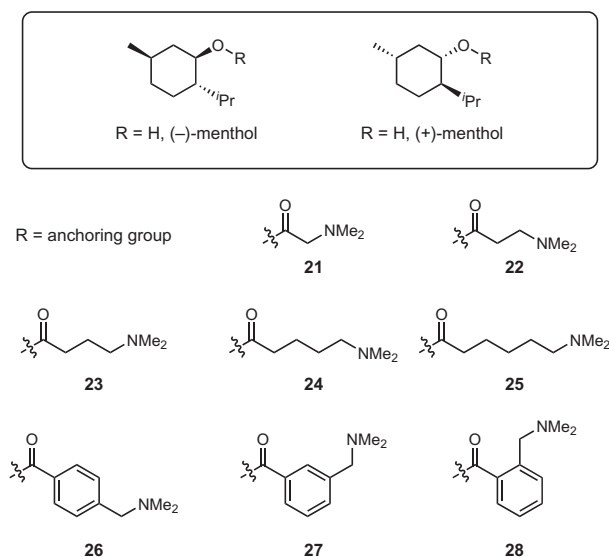

Results and discussion

MD simulations. MD simulations revealed a high dependence of the nature and persistence of the salt bridge contacts between the substrate and PikC$_{wt}$ on the structure of the anchoring group and the chirality of menthol (Supplementary Fig. 3). The (−)-menthol moiety was consistently retained in the hydrophobic pocket of the active site close to the haem cofactor. However, in the case of (−)-**21**, the two-carbon anchor was too short to establish strong interactions with the residues E94 and E85 typically mediating catalytically productive substrate binding[33,34]. Instead, detrimental salt bridge interactions were established with the neighbouring E246 located in the I-helix, with 58% frequency, throughout the

simulation (Table 1), unproductively positioning the menthol core with less reactive primary and tertiary C–H bonds oriented towards the iron-oxo site (Fig. 3a). The lack of contacts with E85 and E94 distorted the tertiary structure of the protein. As a result, the active site channel adopted an open conformation similar to that observed crystallographically in the *apo* state of PikC[33]. These mechanisms of inactivation revealed through MD simulations provided structural information to guide the optimization of catalysis through both substrate and protein engineering. The key objective was to maximize anchoring group-mediated substrate binding and maintain a closed, catalytically productive protein conformation.

The MD simulations revealed that less distortion of the PikC structure was observed with the substrate bearing the longest linear anchoring group ((−)-**25**). A new polar contact with E48, which had not been observed in the PikC complex bound to its natural macrolide substrates, appeared to be relevant for the engineered substrate binding. The anchoring group in (−)-**25** was sufficiently long to allow polar contacts between the dimethylamino group and binding residues located at the entrance of the active site (41%, 4% and 33% frequencies with E94, E85 and E48, respectively; Table 1) and to avoid interacting with E246. Analogously, with the more rigid substrate (−)-**26**, the structural integrity of the protein and substrate binding are further improved as a persistent, short salt bridge is established with E94 (79% frequency; Table 1). Similar behaviour was observed for the PikC$_{D50N}$ mutant, which has demonstrated an improved catalytic efficiency compared to PikC$_{wt}$ (Supplementary Fig. 4). We therefore predicted that these longer anchoring groups, (−)-**25** and (−)-**26**, should lead to improved enzymatic reactivities.

In contrast, the corresponding antipodal set of compounds showed inferior binding in the MD simulations (Supplementary Figs 5 and 6). Notably, the active site of PikC$_{wt}$ exhibited a higher shape complementarity towards (−)-menthol than (+)-menthol.

3

Table 1 | Persistence of the potential salt bridges (<4 Å) between a given substrate anchoring group and different PikC residues (percentages measured from MD simulations).

	Salt bridge persistence (%)												
	PikC$_{wt}$					PikC$_{D50N}$					PikC$_{D50ND176QE246A}$		
Substrate	a	b	c	d	e	a	b	c	d	e	a	b	c
(−)-21	0	3	0	58	0	4	0	0	0	0	4	0	0
(+)-21	38	0	4	0	0	0	0	0	0	0	6	0	0
(−)-25	41	4	33	0	0	60	1	0	5	0	75	15	0
(+)-25	84	7	0	0	0	95	0	0	0	0	57	17	0
(−)-26	79	0	0	0	0	95	0	0	0	0	13	9	31
(+)-26	7	0	6	0	0	3	5	32	0	0	1	5	32

(±) distinction refers to menthol absolute configuration.

For example, (+)-**25** completely abandoned the haem-containing hydrophobic pocket after ~100 ns, and (+)-**21** unproductively oscillates in and out of the hydrophobic pocket (average FeO⋯C4 distance, ~7 Å). MD simulations revealed that the (+)-menthol core is held proximal to the haem group only in (+)-**26**, albeit with relatively weak binding contacts to E94, E85 or E48 (>5 Å).
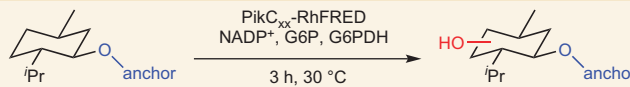
Enzymatic reactions. *In vitro* reactions of each substrate with PikC validated the MD predictions (Table 2). Each of the menthol compounds was tested as a substrate for PikC$_{wt}$-RhFRED and PikC$_{D50N}$-RhFRED[38] in the presence of the standard NADPH recycling system[39]. As observed with desosaminyl cyclohexanol (**17**), the desosamine derivative of (−)-menthol was not hydroxylated by PikC$_{wt}$-RhFRED or PikC$_{D50N}$-RhFRED. Similarly, short synthetic anchoring groups such as those in (−)-**21**, (−)-**22** and (−)-**23** showed low total turnover numbers (TTN = mol product/mol enzyme) with PikC$_{D50N}$-RhFRED, and no product was detected for the reactions of these compounds with the wild-type enzyme. However, increasing the length of the anchoring group by one or two additional methylene units improved catalysis with PikC$_{D50N}$-RhFRED, as shown by TTN of 24 and 43 for (−)-**24** and (−)-**25**, respectively (Table 2). Moreover, analysis of the benzylic amine substrates (−)-**26**, (−)-**27** and (−)-**28** demonstrated the superior nature of

the *para* derivative over the *meta* and *ortho* analogues (TTN 96, 41 and 10, respectively). Only (−)-**25** and (−)-**26** were converted to hydroxylated products by PikC-RhFRED[38], albeit in low conversion (TTN = 11 and 16, respectively).

As predicted by MD simulations, PikC reactions with the (+)-menthol substrate panel demonstrated lower catalytic efficiency and selectivity compared to the (−)-menthol series. For example, TTN of 2–10 were measured for (−)-**21**, (−)-**22**, (−)-**23** and (−)-**28**, while no product formation was observed for the (+)-menthol substrates with the same anchoring groups. Similarly, lower turnover numbers were observed for substrates (+)-**24**–**27**, ranging from 19 to 28 (Table 2).

MD-guided protein engineering. Although the improved binding of (−)-**25** and (−)-**26** positioned the menthol core close to the reactive Fe-oxo group, fluctuations between open and closed conformations of the active site were still observed by MD. We hypothesized that the protein can adopt the open conformation when that anchoring moiety of the substrate does not persistently bind E85 (Fig. 3b,c). Furthermore, we anticipated that dynamic binding of the anchoring group to the triad of binding residues (E48, E85 and E94), not isolated contacts with one or two residues, as observed crystallographically, would shift the open/ closed equilibrium in favour of the closed, active conformation. The E246 and D176 positions were readily identified as targets for mutagenesis to promote catalysis.

We therefore engineered the triple mutant monooxygenase PikC$_{D50ND176QE246A}$-RhFRED, lacking carboxylate groups at both D176 and E246, to eliminate the potential for substrate binding at these residues. As predicted, a significant increase in TTN was observed for all substrates tested with this PikC triple mutant variant (Table 2). For example, the TTN for (−)-**28** increased from 10 to 170, and substrates such as (−)-**23,** which showed no detectable product formation with PikC$_{D50N}$-RhFRED, had a TTN of 38 with the PikC$_{D50ND176QE246A}$-RhFRED biocatalyst. MD simulations of PikC$_{D50ND176QE246A}$-RhFRED alone and with substrate show a higher frequency of the closed conformation compared to PikC$_{wt}$-RhFRED and PikC$_{D50N}$-RhFRED (Fig. 3d–f and Supplementary Figs 7–9), which supports our hypothesis that these mutations shift the equilibrium in favour of the catalytically active closed conformation. A similar impact of enzyme evolution on the active site configuration has been proposed in other engineered proteins[37]. Through these mutations, the dependence of the enzyme reactivity on an induced-fit mechanism, proposed to operate for the larger and more polar macrolactones, is alleviated

Table 2 | Total turnover numbers (TTNs) for the hydroxylation of (−)- and (+)-menthol derived substrates using mutant forms of PikC.

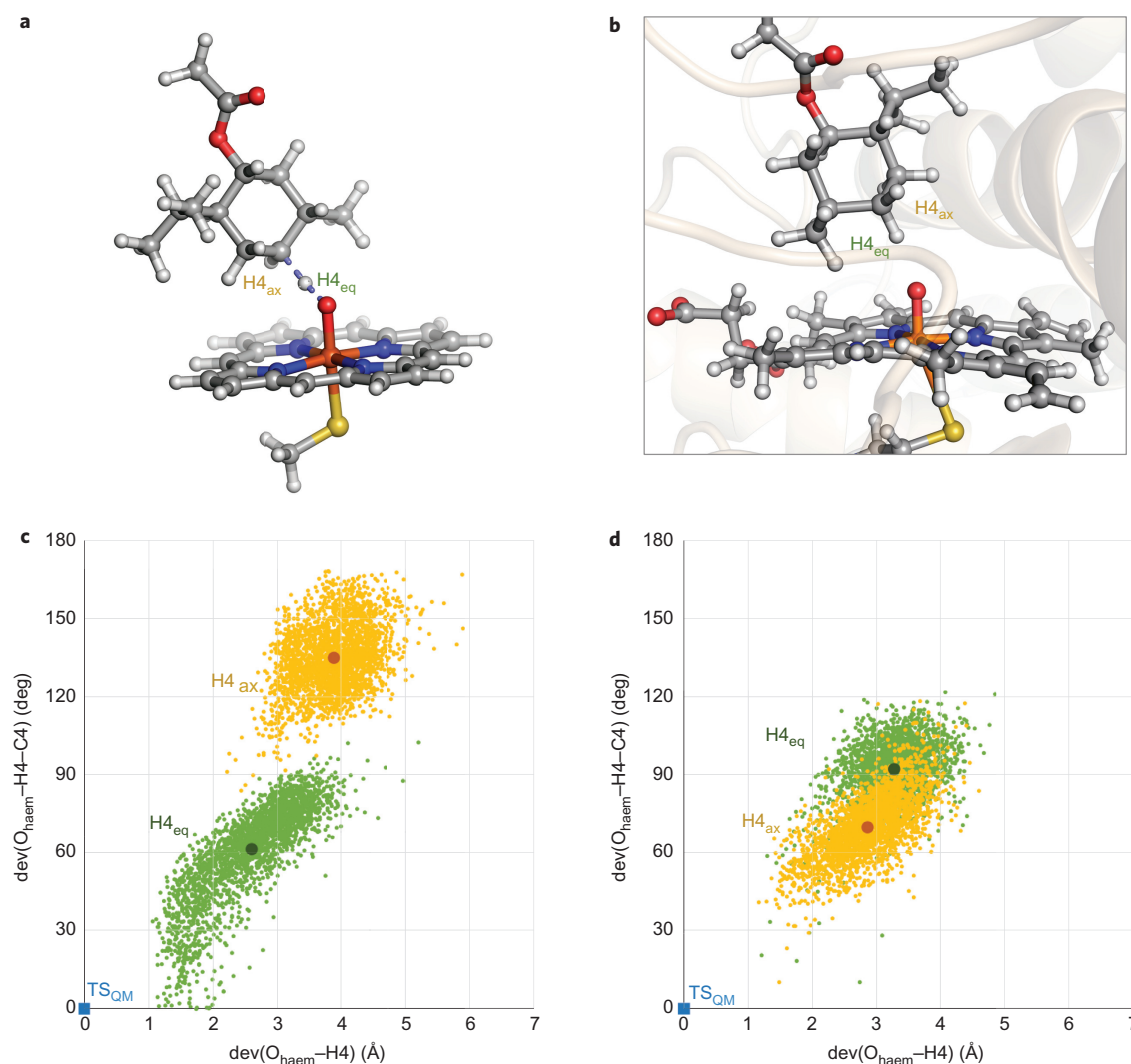| Substrate	PikC$_{wt}$-RhFRED		PikC$_{D50N}$-RhFRED		PikC$_{D50ND176QE246A}$-RhFRED	
	(−)	(+)	(−)	(+)	(−)	(+)
21	0	0	2.0	0	95	26
22	0	0	2.3	0	89	20
23	0	0	7.8	0	182	38
24	0	9.6	22	19	179	90
25	11	14	43	27	186	129
26	16	0	96	28	137	72
27	0	0	41	13	153	47
28	0	0	10	0	170	26

TTN given in mol product/mol enzyme.

Figure 4 | Stereoselectivity in the C4–H abstraction catalysed by PikC. a, QM optimized transition for hydrogen abstraction of $H4_{eq}$ in (–)-menthol (Supplementary Fig. 2). **b**, Representative snapshot of a MD trajectory of $PikC_{D50N}$ bound to (–)-**25** showing the substrate in close proximity to the catalytic haem iron-oxo species (compound I). **c,d**, Orientation of $H4_{eq}$ (in green) and $H4_{ax}$ (in yellow) of (–)-**25** (**c**) and (+)-**25** (**d**) relative to the iron-haem oxo group in $PikC_{D50N}$, derived from 0.5 μs MD trajectories. Each point represents a simulation snapshot. Deviations of the distances (*x* axis) and angles (*y* axis) from the QM optimized transition structure (in blue) with a hypothetical model haem catalyst are shown. The greater separation between the two sets of points ($H4_{eq}$ and $H4_{ax}$) in (–)-menthol (**c**) is interpreted as a potential higher diastereoselectivity in the associated transition states and thus the hydroxylation reaction.

by stabilizing the tertiary structure of the protein in a conformation that facilitates the binding in a productive orientation of the smaller and less polar unnatural ligands.

Site selectivity. Once the catalytic efficiency of PikC had been enhanced, we turned our attention to the selectivity of the hydroxylation reaction. Thus, in addition to driving the substrate and protein engineering on the PikC platform, we investigated the site- and stereoselectivity of this enzymatic hydroxylation reaction using a combination of quantum mechanical (QM) and MD data. Density functional theory (DFT) transition-state calculations of a model of compound I reacting with (–)-menthol predicted that the methylene C–H bonds at the C4 position of menthol would be functionalized (Supplementary Methods and Supplementary Table 1). Interestingly, the previously reported positions for site-selective C–H oxidation, including the sterically unhindered methyl C–H bonds and electronically activated methine C–H bonds, all involve higher activation barriers. Using this model system, the predicted selectivity between the axial and equatorial C4–H bonds is subtle, and abstraction of the axial hydrogen is

favoured by only 0.7 kcal mol^{-1} (3:1 kinetic ratio at 25 °C). We anticipated this selectivity to be modulated by the chiral environment of the enzyme, such that the C–H bond in closer proximity and more properly oriented towards the haem iron-oxo group in the preferred binding pose of the substrate in the active site will be preferentially oxidized. Considering the transition structure for C4–H_{eq} abstraction (Fig. 4a) as a reference, we screened all the binding poses sampled for each substrate bound to $PikC_{D50N}$ in the MD trajectories (compound (–)-**25** is shown in Fig. 4b) in order to detect the most favourably arranged C–H bond for oxidation. The hydrophobic active site of PikC consistently exposes the equatorial C4–H_{eq} bond of (–)-menthol in a more appropriate orientation towards the iron-oxo group, irrespective of the anchoring group used and overriding the preference for C4–H_{ax} predicted with the model non-enzymatic catalyst (Fig. 4c and Supplementary Fig. 10). In contrast, PikC was calculated to be unable to recognize any particular binding mode of the (+)-menthol substrates, leading to continuous reallocation of the cyclohexane ring inside the active site and equally exposing $H4_{eq}$ and $H4_{ax}$ to the iron-oxo centre (Fig. 4d
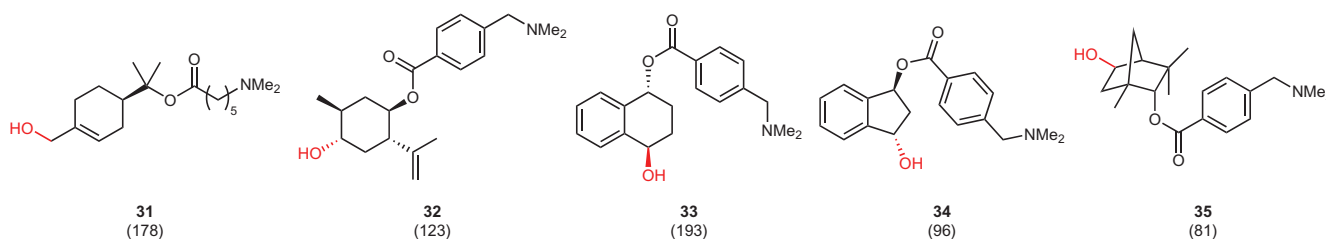
Figure 5 | Site-selective oxidation with PikC$_{D50ND176QE246A}$-RhFRED. C–O bonds highlighted in red indicate a hydroxyl group introduced in the P450 reaction. Numbers given in parentheses are TTN values determined in triplicate. Reaction conditions: 1 mM substrate, 5 μM enzyme, 1 mM NADP$^+$, 5 mM glucose-6-phosphate, 1 U ml^{-1} glucose-6-phosphate dehydrogenase.

and Supplementary Fig. 11). Simulations of the most active triple mutant PikC$_{D50ND176QE246A}$ gave analogous results (Supplementary Figs 12 and 13). The combined QM and MD results predicted a high regioselectivity for both enantiomers of menthol at the C4 position and a high stereoselectivity towards the equatorial C–H bond[4] for the (−)-enantiomer.

Experimentally, one major regioisomer was consistently formed from the (−)-menthol substrates, whereas at least five isomeric products were detected upon reaction of (+)-**24–27** with PikC$_{D50N}$-RhFRED (Supplementary Section III). To determine the structure of the major hydroxylation product, preparative-scale reactions were performed with (−)-**25** and (−)-**26**, the two substrates with the highest TTN. Following enzymatic hydroxylation, the anchoring group was removed quantitatively (LiAlH$_4$ reduction) for subsequent analysis. Diol **8** (Fig. 1), the result of C4-H$_{eq}$ oxidation, was obtained as the major product for both (−)-**25** and (−)-**26**, confirming the prediction made through the QM and MD studies.

To evaluate the transferability of this approach to a structurally diverse panel of substrates, we tested the ability of the engineered P450, PikC$_{D50ND176QE246A}$-RhFRED, to oxidize non-menthol small-molecule cores bearing synthetic anchoring groups. As shown in Fig. 5, six-membered ring substrates derived from (+)-α-terpineol and (−)-isopulegol are hydroxylated at C–H bonds remote from the anchoring group to afford **31** and **32**, respectively. More sterically demanding fused bicyclic and bridged bicyclic substrates were also hydroxylated with high site and stereoselectivity to deliver alcohols **33–35**. Notably, the relationship between the C–H bond oxidized in each case and the attachment point of the anchoring group is conserved, with C–H bonds remote to the anchoring group favoured for hydroxylation. This directing effect can override electronic effects as more activated C–H bonds are untouched; for example, a methylene C–H bond is hydroxylated over allylic C–H bonds to give **32**. Furthermore, the products generated consistently possess an *anti* relationship between the newly introduced hydroxyl group and the anchoring group, as was predicted computationally and observed in the (−)-menthol model system.

Conclusions

This work provides the foundation for a new method to site- and stereoselectively oxygenate unactivated methylene C–H bonds, something that has eluded previous efforts using chemical catalysts, and with a substrate scope that exceeds what is typical of biocatalytic methods. Computation-driven substrate and protein engineering enabled the expansion of the substrate scope of PikC P450 monooxygenase from 12- and 14-membered macrolides to a six-membered ring scaffold. The efficiency of PikC oxidation of a given substrate, as well as the selectivity, was iteratively evaluated through QM and MD calculations with excellent correlation to experimental data. The optimized protocol devised through this initial study is described in Supplementary Fig. 14. This tool is currently being applied for the development of new anchoring groups to target specific classes of substrates, the prediction and improvement of enzyme selectivity,

and the rational design of additional monooxygenases to generate superior C–H functionalization biocatalysts.

Methods

QM calculations. Full geometry optimizations, transition structure (TS) searches and single-point computations were carried out in the Gaussian 09 package[40]. All geometry optimizations were carried out with the unrestricted version of the hybrid B3LYP functional[41], which has been widely used in the iron haem literature[42,43]. The relativistic LANL2DZ pseudopotential[44], and associated basis set, was used for Fe. For S, O, N, C and H, the double-zeta basis set 6-31G(d) was employed to obtain the geometries, and the larger 6-311+G(d,p) basis set was used to calculate single-point energies. Additional single-point energy calculations using functionals able to account for dispersion forces such as B3LYP-D3 with Becke–Johnson damping[45] and M06[46] in conjunction with the larger 6-311+G(2d,p) basis set, and the SDD pseudopotential[47] were performed (Supplementary Table 1). Thermal and entropic corrections to energy were calculated from vibrational frequencies. The nature of the stationary points was determined in each case according to the appropriate number of negative eigenvalues of the Hessian matrix from the frequency calculations. Scaled frequencies were not considered. In some problematic systems, the SCF convergence criterion was lowered (scfcon = 6) to locate the transition states. Different spin states for Fe(III) and multiple conformers of the isopropyl and ester groups in the menthol derivatives were considered (Supplementary Figs 1 and 2). Mass-weighted intrinsic reaction coordinate calculations were carried out for a representative transition structure using the Gonzalez and Schlegel scheme[48,49] to ensure that they indeed connected the appropriate reactants and products.

For the calculation of bond dissociation energies (BDEs)[50], single-point energy calculations using the correlated *ab initio* SCS-MP2 methods, in combination with the cc-pVTZ basis set[51], were performed on the B3LYP/6-31G(d) optimized geometries. BDEs were defined as the difference in zero-point energies between menthol and the sum of the isolated radicals generated upon homolytic C–H cleavage (Supplementary Table 2).

Cartesian coordinates, electronic energies, entropies, enthalpies, Gibbs free energies and lowest frequencies of the optimized structures are provided in Supplementary Table 1.

MD simulations. The haem iron(IV)–oxo complex involved in the cytochrome-catalysed oxidative hydroxylation cycle (compound I) was used to model the active form of the PikC cofactor. Simulations were performed using the GPU code (*pmemd*)[52] of the AMBER 12 package[53]. The Amber-compatible parameters developed by Cheatham *et al.*[54] were used for compound I and its axial Cys ligand. Parameters for menthol substrates were generated within the *antechamber* module using the general AMBER force field (*gaff*)[55], with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model[56]. The charges were calculated according to the Merz–Singh–Kollman scheme[57,58] using the Gaussian 09 package[40]. Each protein was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P[59] water molecules using the *leap* module, resulting in the addition of around 15,000 solvent molecules. The systems were neutralized by addition of explicit counter ions (Na$^+$ and Cl$^-$). All subsequent calculations were done using the widely tested Stony Brook modification of the Amber 99 force field (*ff99sb*)[60]. A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on the solute by a harmonic potential with a force constant of 500 kcal mol^{-1} Å^{-2} and the second stage minimizes all the atoms in the simulation cell except those involved in the harmonic distance restraint. The systems were gently heated using six 50 ps steps, incrementing the temperature by 50 K for each step (0–300 K) under constant-volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between the hydrogen atoms was kept fixed. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method[61]. An 8 Å cutoff was applied to Lennard–Jones and electrostatic interactions. Harmonic restraints of 30 kcal mol^{-1} were applied to the solute and the Andersen equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for

2 ns with a 2 fs time step at a constant volume. Production trajectories were then run for an additional 500 ns under the same simulation conditions.

PikC$_{xx}$-RhFRED analytical-scale enzymatic reactions. The standard assay contained 5 μM PikC$_{xx}$-RhFRED, 1 mM substrate, 1 mM NADP$^+$, 0.05 units of glucose-6-phosphate dehydrogenase and 5 mM glucose-6-phosphate for NADPH regeneration in reaction buffer (50 mM NaH$_2$PO$_4$, pH 7.3, 1 mM EDTA, 0.2 mM dithiothreitol and 10% glycerol) with a total volume of 50 μl. The reaction was carried out at 30 °C for 3 h and quenched by the addition of 150 μl MeOH. The resulting mixture was briefly vortexed and centrifuged at 10,000g for 10 min. The subsequent liquid chromatography mass spectrometry (LC-MS) analysis was performed on an Agilent Q-TOF HPLC-MS (Department of Chemistry, University of Michigan) equipped with a high-resolution electrospray mass spectrometry (ESI-MS) source and a reverse-phase HPLC system using a Waters XBridge C18 column 3.5 μm, 2.1 × 150 mm, under the following conditions: mobile phase (A = deionized water + 0.1% formic acid, B = acetonitrile + 0.1% formic acid), 10–100% B over 15 min, 100% B for 4 min; flow rate, 0.2 ml min^{-1}. Reactions were scanned for [M + 16] (monohydoxylation) and [M + 32] (dihydroxylation). The percent conversion was determined as outlined by Li and co-authors[35]. Briefly, the percent conversion was calculated using AUC$_{total\ products}$/(AUC$_{total\ products}$ + AUC$_{unreacted\ substrate}$) by assuming that the ionization efficiencies of the substrate and hydroxylated products were the same, because the ionization site of this series of compounds is presumed to be the dimethylamino group.

References

1. Newhouse, T. & Baran, P. S. If C–H bonds could talk: selective C–H bond oxidation. *Angew. Chem. Int. Ed.* **50,** 3362–3374 (2011).
2. Gormisky, P. E. & White, M. C. Catalyst-controlled aliphatic C–H oxidations with a predictive model for site-selectivity. *J. Am. Chem. Soc.* **135,** 14052–14055 (2013).
3. Chen, M. S. & White, M. C. Combined effects on selectivity in Fe-catalyzed methylene oxidation. *Science* **327,** 566–571 (2010).
4. Chen, K., Eschenmoser, A. & Baran, P. S. Strain release in C–H bond activation. *Angew. Chem. Int. Ed.* **48,** 9705–9708 (2009).
5. Roiban, G. D., Agudo, R. & Reetz, M. T. Cytochrome P450 catalyzed oxidative hydroxylation of achiral organic compounds with simultaneous creation of two chirality centers in a single C–H activation step. *Angew. Chem. Int. Ed.* **53,** 8659–8663 (2014).
6. Kille, S., Zilly, F. E., Acevedo, J. P. & Reetz, M. T. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nature Chem.* **3,** 738–743 (2011).
7. Zhang, K. D., Shafer, B. M., Demars, M. D., Stern, H. A. & Fasan, R. Controlled oxidation of remote sp^3 C–H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity. *J. Am. Chem. Soc.* **134,** 18695–18704 (2012).
8. Zhang, K. D., El Damaty, S. & Fasan, R. P450 fingerprinting method for rapid discovery of terpene hydroxylating P450 catalysts with diversified regioselectivity. *J. Am. Chem. Soc.* **133,** 3242–3245 (2011).
9. Lewis, J. C., Coelho, P. S. & Arnold, F. H. Enzymatic functionalization of carbon–hydrogen bonds. *Chem. Soc. Rev.* **40,** 2003–2021 (2011).
10. Fasan, R. Tuning P450 enzymes as oxidation catalysts. *ACS Catal.* **2,** 647–666 (2012).
11. Agudo, R., Roiban, G. D. & Reetz, M. T. Achieving regio- and enantioselectivity of P450-catalyzed oxidative C–H activation of small functionalized molecules by structure-guided directed evolution. *ChemBioChem* **13,** 1465–1473 (2012).
12. Acevedo-Rocha, C. G., Hoebenreich, S. & Reetz, M. T. in *Methods in Molecular Biology* 2nd edn, Vol. 1179 (eds Gillam, E. M. J., Copp, J. N. & Ackerley, D. F.) 103–128 (Humana Press, 2014).
13. Jung, S. T., Lauchli, R. & Arnold, F. H. Cytochrome P450: taming a wild type enzyme. *Curr. Opin. Biotechnol.* **22,** 809–817 (2011).
14. Yang, J., Gabriele, B., Belvedere, S., Huang, Y. & Breslow, R. Catalytic oxidations of steroid substrates by artificial cytochrome P-450 enzymes. *J. Org. Chem.* **67,** 5057–5067 (2002).
15. Breslow, R. *et al.* Remote oxidation of steroids by photolysis of attached benzophenone groups. *J. Am. Chem. Soc.* **95,** 3251–3262 (1973).
16. Leow, D., Li, G., Mei, T. S. & Yu, J. Q. Activation of remote meta-C–H bonds assisted by an end-on template. *Nature* **486,** 518–522 (2012).
17. Tang, R. Y., Li, G. & Yu, J. Q. Conformation-induced remote meta-C–H activation of amines. *Nature* **507,** 215–220 (2014).
18. Larsen, A. T., May, E. M. & Auclair, K. Predictable stereoselective and chemoselective hydroxylations and epoxidations with P450 3A4. *J. Am. Chem. Soc.* **133,** 7853–7858 (2011).
19. Menard, A., Fabra, C., Huang, Y. & Auclair, K. Type II ligands as chemical auxiliaries to favor enzymatic transformations by P450 2E1. *ChemBioChem* **13,** 2527–2536 (2012).
20. Polic, V. & Auclair, K. Controlling substrate specificity and product regio- and stereo-selectivities of P450 enzymes without mutagenesis. *Bioorg. Med. Chem.* **22,** 5547–5554 (2014).
21. Vermeulen, N. A., Chen, M. S. & White, M. C. The Fe(PDP)-catalyzed aliphatic C–H oxidation: a slow addition protocol. *Tetrahedron* **65,** 3078–3084 (2009).
22. Olivo, G., Lanzalunga, O., Mandolini, L. & Di Stefano, S. Substituent effects on the catalytic activity of bipyrrolidine-based iron complexes. *J. Org. Chem.* **78,** 11508–11512 (2013).
23. Yazerski, V. A. *et al.* Making Fe(BPBP)-catalyzed C–H and C=C oxidations more affordable. *Org. Biomol. Chem.* **12,** 2062–2070 (2014).
24. Lee, S. & Fuchs, P. L. An efficient C–H oxidation protocol for alpha-hydroxylation of cyclic steroidal ethers. *Org. Lett.* **6,** 1437–1440 (2004).
25. Ottenbacher, R. V., Samsonenko, D. G., Talsi, E. P. & Bryliakov, K. P. Highly efficient, regioselective, and stereospecific oxidation of aliphatic C–H groups with H$_2$O$_2$, catalyzed by aminopyridine manganese complexes. *Org. Lett.* **14,** 4310–4313 (2012).
26. Chow, Y. L., Hayasaka, T. & Tam, J. N. S. Photoreaction of nitroso compounds in solution. 13. Photo-oxidation of alkyl nitrites. *Can. J. Chem.* **48,** 508–511 (1970).
27. Petrovic, G., Saicic, R. N. & Cekovic, Z. Regioselective free radical phenylsulfonation of a non-activated delta-carbon atom by the photolysis of alkyl benzenesulfenate. *Tetrahedron* **59,** 187–196 (2003).
28. Miyazawa, M., Kumagae, S. & Kameoka, H. Biotransformation of (+)- and (−)-menthol by the larvae of common cutworm (*Spodoptera litura*). *J. Agric. Food Chem.* **47,** 3938–3940 (1999).
29. Atta ur, R. *et al.* Fungal transformation of (1R,2S,5R)-(−)-menthol by *Cephalosporium aphidicola*. *J. Nat. Prod.* **61,** 1340–1342 (1998).
30. Xue, Y. Q., Zhao, L. S., Liu, H. W. & Sherman, D. H. A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *Proc. Natl Acad. Sci. USA* **95,** 12111–12116 (1998).
31. Xue, Y. Q., Wilson, D., Zhao, L. S., Liu, H. W. & Sherman, D. H. Hydroxylation of macrolactones YC-17 and narbomycin is mediated by the PikC-encoded cytochrome P450 in *Streptomyces venezuelae*. *Chem. Biol.* **5,** 661–667 (1998).
32. Zhang, Q. B. & Sherman, D. H. Isolation and structure determination of novamethymycin, a new bioactive metabolite of the methymycin biosynthetic pathway in *Streptomyces venezuelae*. *J. Nat. Prod.* **64,** 1447–1450 (2001).
33. Sherman, D. H. *et al.* The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *J. Biol. Chem.* **281,** 26289–26297 (2006).
34. Li, S. Y., Ouellet, H., Sherman, D. H. & Podust, L. M. Analysis of transient and catalytic desosamine-binding pockets in cytochrome P-450 PikC from *Streptomyces venezuelae*. *J. Biol. Chem.* **284,** 5723–5730 (2009).
35. Li, S. Y. *et al.* Selective oxidation of carbolide C–H bonds by an engineered macrolide P450 mono-oxygenase. *Proc. Natl Acad. Sci. USA* **106,** 18463–18468 (2009).
36. Negretti, S. *et al.* Directing group-controlled regioselectivity in an enzymatic C–H bond oxygenation. *J. Am. Chem. Soc.* **136,** 4901–4904 (2014).
37. Jiménez-Osés, G. *et al.* The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nature Chem. Biol.* **10,** 431–436 (2014).
38. Li, S. Y., Podust, L. M. & Sherman, D. H. Engineering and analysis of a self-sufficient biosynthetic cytochrome P450 PikC fused to the RhFRED reductase domain. *J. Am. Chem. Soc.* **129,** 12940–12941 (2007).
39. Wong, C-H. & Whitesides, G. M. Enzyme-catalyzed organic synthesis: NAD(P)H cofactor regeneration by using glucose-6-phosphate and glucose-6-phosphate dehydrogenase from *Leuconostoc mesenteroides*. *J. Am. Chem. Soc.* **103,** 4890–4899 (1981).
40. Frisch, M. J. *et al.* Gaussian 09 (Gaussian, 2009).
41. Lee, C., Yang, W. & Parr, R. G. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37,** 785–789 (1988).
42. Shaik, S. *et al.* P450 enzymes. Their structure, reactivity, and selectivity—modeled by QM/MM calculations. *Chem. Rev.* **110,** 949–1017 (2009).
43. Rydberg, P., Sigfridsson, E. & Ryde, U. On the role of the axial ligand in heme proteins: a theoretical study. *J. Biol. Inorg. Chem.* **9,** 203–223 (2004).
44. Hay, P. J. & Wadt, W. R. *Ab initio* effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.* **82,** 270–283 (1985).
45. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32,** 1456–1465 (2011).
46. Zhao, Y. & Truhlar, D. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120,** 215–241 (2008).
47. Dolg, M., Wedig, U., Stoll, H. & Preuss, H. Energy-adjusted *ab initio* pseudopotentials for the first row transition elements. *J. Chem. Phys.* **86,** 866–872 (1987).
48. Gonzalez, C. & Schlegel, H. B. Reaction path following in mass-weighted internal coordinates. *J. Phys. Chem.* **94,** 5523–5527 (1990).

49. Gonzalez, C. & Schlegel, H. B. An improved algorithm for reaction path following. *J. Chem. Phys.* **90,** 2154–2161 (1989).
50. Gerenkamp, M. & Grimme, S. Spin-component scaled second-order Moller–Plesset perturbation theory for the calculation of molecular geometries and harmonic vibrational frequencies. *Chem. Phys. Lett.* **392,** 229–235 (2004).
51. Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90,** 1007–1023 (1989).
52. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on Gpus. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* **9,** 3878–3888 (2013).
53. Case, D. A. *et al.* AMBER 12 (UCSF, 2012).
54. Shahrokh, K., Orendt, A., Yost, G. S. & Cheatham, T. E. Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **33,** 119–133 (2012).
55. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **25,** 1157–1174 (2004).
56. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A Well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97,** 10269–10280 (1993).
57. Besler, B. H., Merz, K. M. & Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **11,** 431–439 (1990).
58. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5,** 129–145 (1984).
59. 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).
60. Wang, J. M., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21,** 1049–1074 (2000).
61. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an *N*-log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

Author contributions

A.R.H.N. and G.J-O. contributed equally to this work. A.R.H.N., G.J-O., P.L., S.N., J.M., K.N.H. and D.H.S. conceived, designed and supervised the project. G.J-O., P.L., R.O.R., Y-F.Y., L.R.F. and Z.L. performed the computational experiments. A.R.H.N., S.N., W.Z. and M.M.G. synthesized substrates and characterized products. A.R.H.N. conducted the biochemical experiments. A.R.H.N., G.J-O., P.L., L.M.P., J.M., K.N.H. and D.H.S. wrote and edited the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M., K.N.H. and D.H.S.

Competing financial interests

The authors declare no competing financial interests.