

Biocatalytic strategy for the construction of sp^3 -rich polycyclic compounds from directed evolution and computational modelling

Received: 9 May 2022

Accepted: 20 December 2023

Published online: 13 February 2024

David A. Vargas^{1,7}, Xinkun Ren^{2,7}, Arkajyoti Sengupta^{3,7}, Ledong Zhu⁴, Satyajit Roy⁵, Marc Garcia-Borràs⁶, K. N. Houk³✉ & Rudi Fasan⁵✉

Catalysis with engineered enzymes has provided more efficient routes for the production of active pharmaceutical agents. However, the potential of biocatalysis to assist in early-stage drug discovery campaigns remains largely untapped. In this study, we have developed a biocatalytic strategy for the construction of sp^3 -rich polycyclic compounds via the intramolecular cyclopropanation of benzothiophenes and related heterocycles. Two carbene transferases with complementary regioisomer selectivity were evolved to catalyse the stereoselective cyclization of benzothiophene substrates bearing diazo ester groups at the C2 or C3 position of the heterocycle. The detailed mechanisms of these reactions were elucidated by a combination of crystallographic and computational analyses. Leveraging these insights, the substrate scope of one of the biocatalysts could be expanded to include previously unreactive substrates, highlighting the value of integrating evolutionary and rational strategies to develop enzymes for new-to-nature transformations. The molecular scaffolds accessed here feature a combination of three-dimensional and stereochemical complexity with ‘rule-of-three’ properties, which should make them highly valuable for fragment-based drug discovery campaigns.

Biocatalysis with engineered enzymes is playing an increasingly important role in enabling and streamlining the stereoselective synthesis of drug molecules and other high-value compounds^{1–6}. In addition, the reaction scope of biocatalysis has recently been expanded to include new-to-nature transformations^{7–10}. A highly attractive, but currently underexploited, role for biocatalysis lies in providing access to stereochemically rich, three-dimensional (3D) ‘fragments’ for fragment-based drug discovery (FBDD) campaigns (Fig. 1a,b)¹¹. Over the past two decades, FBDD has been a key strategy for drug discovery, yielding several candidates for clinical trials, some of which have been approved for

commercialization¹². This drug discovery approach relies on the availability of libraries of small, diverse organic molecules (‘fragments’) that are screened to identify weak binders for a protein of interest and then linked together to generate potent inhibitors of such targets^{12,13}. In terms of physicochemical properties, the most suitable fragments for FBDD applications are molecules that adhere to the ‘rule of three’ (Ro3), that is, a molecular weight below 300 Da, fewer than three rotatable bonds and an *n*-octanol–water partition coefficient (clogP) less than 3 (ref. 14). Although several readily available molecules meet these requirements, medicinal chemists and chemical biologists

¹Process Research and Development, Merck, Rahway, NJ, USA. ²College of Engineering and Applied Sciences, Nanjing University, Nanjing, China.³Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA. ⁴Environment Research Institute, Shandong University, Qingdao, People’s Republic of China. ⁵Department of Chemistry and Biochemistry, University of Texas at Dallas, Richardson, TX, USA. ⁶Institut de Química Computacional i Catàlisi (IQCC), Departament de Química, Universitat de Girona, Girona, Spain. ⁷These authors contributed equally: David A. Vargas, Xinkun Ren, Arkajyoti Sengupta. ✉e-mail: houk@chem.ucla.edu; rudi.fasan@utdallas.edu

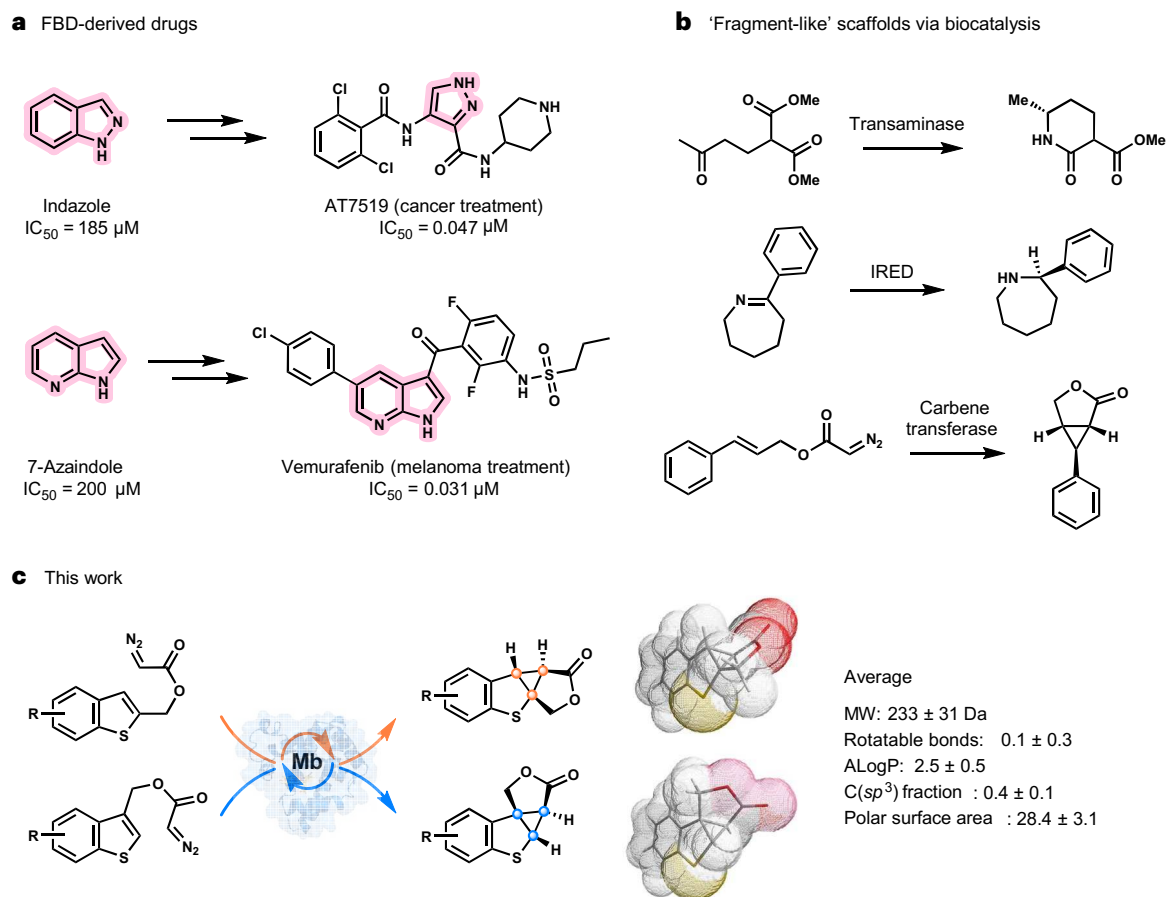


Fig. 1 | *sp*²- versus *sp*³-rich molecular 'fragments' for applications in FBDD.

a, Representative drug molecules discovered and developed via FBDD. IC_{50} , half-maximum inhibitory concentration. **b**, 'Fragment-like' scaffolds generated via biocatalysis include chiral piperidines⁵¹, azepanes⁵² and cyclopropane-

fused γ -lactones³⁷. IRED, imine reductase. **c**, Biocatalytic intramolecular cyclopropanation of C2- and C3-functionalized benzothiophene substrates to yield *sp*³-rich tetracyclic scaffolds with distinct 3D shapes and Ro3-compliant properties.

have highlighted the need for more unique, stereochemically rich, 3D 'fragments' to expand the opportunities for drug discovery via FBDD^{15–18}. In this context, we have developed a biocatalytic strategy to access a series of previously inaccessible *sp*³-rich sulfur-containing polycyclic scaffolds that exhibit these highly sought-after features and could thus serve as potentially valuable additions to fragment libraries for FBDD.

Haem-dependent enzymes and proteins have emerged as promising biocatalysts for mediating carbene transfer reactions such as olefin cyclopropanations that are not known in nature^{19–28}. Artificial metalloenzymes have also been reported to be useful for this type of reaction^{29–36}. More recently, the scope of these strategies has been extended to include the first examples of enzyme-catalysed intramolecular cyclopropanation reactions^{37,38}. Here we report a biocatalytic methodology based on engineered myoglobin variants for the asymmetric intramolecular cyclopropanation of diazo ester-functionalized benzothiophenes. This reaction has no chemocatalytic precedent and attempts to realize similar transformations using organometallic catalysts met with limited success^{39,40}. Following optimization of the biocatalyst via protein engineering, both C2- and C3-functionalized benzothiophenes, along with other heterocycles, could be cyclized with excellent enantioselectivity to furnish stereochemically rich, 3D scaffolds (fraction of *sp*³ carbon atoms = 0.36, 75% of which are stereogenic centres) with physicochemical properties adhering to the Ro3 (average MW = 233 ± 31 Da, average number of rotatable bonds = 0.1 ± 0.3 , average ALogP = 2.5 ± 0.5 , average polar surface area = 28.4 ± 3.1 ; Fig. 1c). Crystallographic analysis of the engineered

enzymes in combination with computational studies based on density functional theory (DFT) calculations and molecular dynamics (MD) simulations provided insights into the mechanism and origin of the protein-mediated stereocontrol in these reactions. This information was further leveraged to enhance the activity and expand the substrate scope of these biocatalysts.

Results and discussion

Biocatalyst evolution for C2-functionalized benzothiophene

Initially, a diverse set of haem-containing enzymes and proteins were screened, including wild-type myoglobin (Mb), its distal histidine variant Mb(H64V), cytochromes P450_{BM3}, CYP119 and P411_{CHF} (ref. 41), and various cytochrome c proteins, for their ability to promote the intramolecular cyclopropanation of benzo[*b*]thiophen-2-ylmethyl 2-diazoacetate (**1a**) to give the target tetracyclic scaffold **2a**. However, none of the enzymes produced any detectable amount of the desired product (Fig. 2a and Supplementary Table 1). Note that the same reaction also failed or gave minimal yields in the presence of several transition metal catalysts commonly used for carbene transfer reactions, such as Rh complexes, Cu(OTf)₂, Fe(tetraphenylporphyrinato)chloride (Fe(TPP)(Cl)), and Co(TPP) (Fig. 2a and Supplementary Table 1). We then extended our screening to an in-house collection of ~100 diverse Mb variants containing a range of single to quadruple mutations at residues surrounding the haem cofactor. While the large majority of these variants showed no activity, Mb(H64F) displayed basal activity for the formation of the desired product **2a** (2% yield) with modest enantioselectivity (61% e.e.; Fig. 2c) under reducing conditions. Mb(H64F)

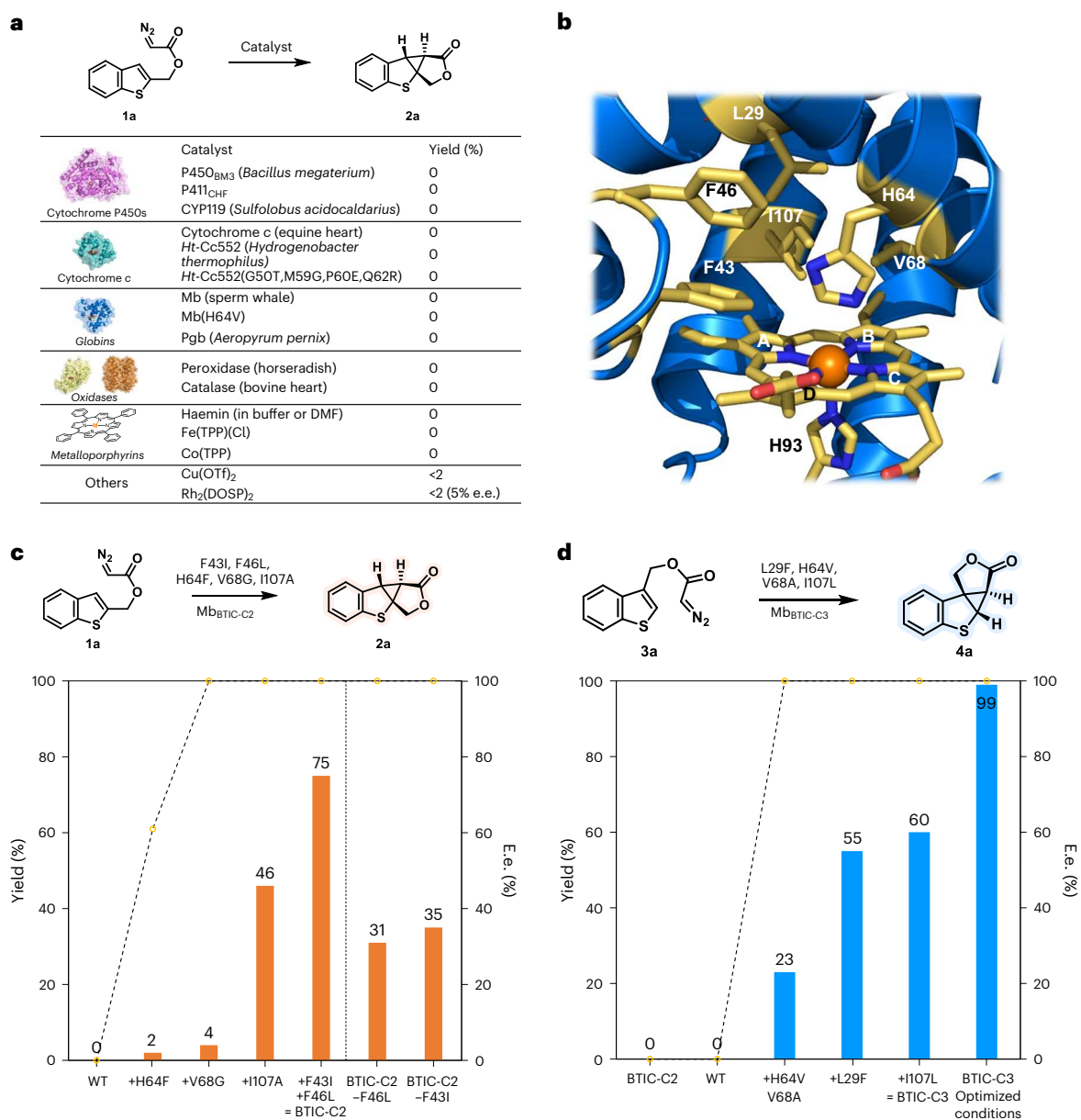


Fig. 2 | Biocatalytic intramolecular cyclopropanation of benzothiophene substrates. a, Activity of a diverse panel of haemoproteins (0.8 mol%) and chemical catalysts (10 mol%) in the intramolecular cyclopropanation of **1a** (see Supplementary Table 1 for the reaction conditions and additional catalysts). Representative crystal structures of a P450 (PDB: 1FAH), cytochrome c (PDB: 1HRC), globin (PDB: 2MBW), peroxidase (PDB: 1HCH) and catalase (PDB: 1TGU) are shown. **b**, Active site of sperm whale Mb (PDB: 1JW8) with the haem and surrounding amino acid residues highlighted as stick structures in gold. Haem

B pyrrole rings are labelled according to convention. The numbering of the pyrrole rings follows Fischer nomenclature. **c,d**, Activity and enantioselectivity of Mb variants in the intramolecular cyclopropanation of **1a** (**c**) and **3a** (**d**) along the evolutionary paths leading to Mb_{BTIC-C2} and Mb_{BTIC-C3}, respectively. The reported product yields were determined by gas chromatography analysis using calibration curves generated with isolated products. DMF, dimethylformamide; Pgb, protoglobin; Rh₂(DOSP)₂, dirhodium tetrakis((S)-N-(dodecylbenzenesulfonyl)proline); WT, wild type.

was thus selected as the parent scaffold for iterative rounds of directed evolution in which active site residues most proximal to the iron centre, that is, Leu29, Phe43, Phe46, Val68 and Ile107, were randomized via site-saturation mutagenesis, followed by library screening in multi-well plates and as whole cells under anaerobic conditions. The improved hits identified after each round were validated with the purified protein in the reaction with **1a**.

Through this process, five beneficial mutations were accumulated in the enzyme, resulting in the quintuple mutant Mb(F43I/F46L/H64F/V68G/I107A), denoted Mb_{BTIC-C2}, with largely increased activity towards the generation of **2a** from **1a** (2% → 75% yield). In addition, Mb_{BTIC-C2} displayed excellent enantioselectivity (>99% e.e.), producing a single

stereoisomer whose configuration was determined to be 3a*S*,3b*S*,8a*R* via single-crystal X-ray diffraction (Fig. 1c and Supplementary Fig. 13). Interestingly, structure–activity analysis of the Mb_{BTIC-C2} lineage revealed a clear synergistic effect of the mutations F43I and F46L on the catalytic activity (turnover number) of the enzyme (Fig. 2c), without affecting the enantioselectivity. This analysis also indicated the distinct benefits of the His→Phe mutation at the distal histidine (His64) and large-to-small mutations at multiple sites within the haem pocket (F43I, F46L and I107A) to favour the target reaction. These findings are generally consistent with the high steric demands inherent in mediating an intramolecular cyclopropanation reaction within the confines of a Mb haem pocket³⁷.

Cyclopropanation of C3-functionalized benzothiophene

Encouraged by the above results, we sought to apply the evolved Mb_{BTIC-C2} biocatalyst to the cyclopropanation of a C3-functionalized benzothiophene (**3a**) as we envisioned this reaction would provide access to an alternative tetracyclic scaffold (Fig. 1c). Surprisingly, no detectable amount of the desired product **4a** was obtained, clearly indicating the need for very different active site configurations to enable the cyclization of **3a** (Fig. 2d). To develop a biocatalyst for this reaction, we re-screened our in-house library of Mb variants against substrate **3a**, the large majority of which showed no activity (Supplementary Table 2). Unlike for **1a**, V68G (and other single mutations at position 68) had no beneficial effect on the transformation of **3a**, further highlighting the different catalyst requirements for the two reactions (Fig. 2d). By contrast, Mb(H64V/V68A) yielded a viable biocatalyst for the intramolecular cyclopropanation of **3a**, producing **4a** in low yield (23% yield; Fig. 2d), but with excellent enantioselectivity (>99% e.e.) for the formation of the 3aS,3bS,8bS enantiomer, as determined by X-ray diffraction (Fig. 1c). Building on these results, a protein engineering campaign was undertaken that resulted in the identification of Mb(L29F/H64V/V68A/I107L), denoted Mb_{BTIC-C3}, as an improved biocatalyst for this reaction, furnishing a threefold improved yield of **4a** (23% → 60% yield) while retaining excellent enantioselectivity (>99% e.e.; Fig. 2d). After optimization of the Mb_{BTIC-C3}-catalysed reaction (that is, 4 °C and slow addition of **3a**; Supplementary Table 3), the cyclopropanation product **4a** was obtained in quantitative yield and enantiopure form (>99% e.e.; Fig. 2d), with the improvement in cyclopropanation yield resulting primarily from a reduction of the carbene dimerization side reaction. Under catalyst-limited conditions, Mb_{BTIC-C3} was found to support a turnover number of up to 440, producing **4a** in 87% yield using only 0.2 mol% catalyst (Supplementary Table 2). Moreover, both Mb_{BTIC-C2} and Mb_{BTIC-C3} could be applied in whole cells (optical density at 600 nm (OD₆₀₀) = 20 and 60, respectively) for the stereoselective synthesis of **2a** in 65% yield and of **4a** in 62% yield, respectively, demonstrating the compatibility of these enzymatic reactions with whole-cell biotransformations.

Side-by-side comparison of the mutations in Mb_{BTIC-C3} and Mb_{BTIC-C2} revealed interesting similarities and differences between the two variants. Similarly to Mb_{BTIC-C2}, Mb_{BTIC-C3} incorporates space-creating mutations at positions 68 and 107 (Fig. 2b), although the nature of the optimal residue differs in each case (V68: Gly versus Ala; I107: Ala versus Leu). In contrast to Mb_{BTIC-C2}, however, mutations at positions Phe43 and Phe46 were detrimental to the cyclization of **3a**. In addition, for the reaction of **3a**, an aromatic residue (Phe) was beneficial at position 29 in place of leucine, whereas a mutation at this position provided no benefit for the cyclization of **1a**. These differences, along with the lack of reactivity of Mb_{BTIC-C3} towards **1a**, further evidenced the orthogonal active site requirements for enabling the Mb-catalysed cyclization of the two regioisomeric substrates.

Substrate scope of the evolved biocatalysts

Next, we investigated the substrate scope of the Mb_{BTIC-C2} biocatalyst using variously substituted benzothiophenes (Fig. 3a). These experiments showed that the enzyme can tolerate a range of electron-withdrawing and -donating groups at position C5, delivering the desired cyclopropane-fused lactone products (**2b–f**) in yields of 36–99% and with excellent enantioselectivity (≥99% e.e.; Fig. 3a). Similar results were obtained for a series of C6-substituted substrates, indicating a notable tolerance of the biocatalyst to variation at the C6 position to furnish **2g–j** in yields of 25–44% and with high enantiopurity (97–99% e.e.; Fig. 3a). Improved yields for the synthesis of **2g** and **2j** could be achieved using whole cells expressing the Mb_{BTIC-C2} variant (51–55% compared with 35–39%). The 7-fluoro-substituted substrate **1k** was also processed by Mb_{BTIC-C2} to give **2k** in good yield and with high enantiopurity (>99% e.e.). In contrast, substitution at the C4 site (for example, Br) was not tolerated, resulting in unreacted starting

material. Interestingly, only carbene dimerization was observed when benzothiophene was replaced with benzofuran in substrate **5** (Supplementary Fig. 1). A diazo ketone analogue of **1a** (compound **6**) was also prepared and tested (Supplementary Fig. 1), but it failed to cyclize using Mb_{BTIC-C2} or related variants.

Using a similar approach, we investigated the scope of the Mb_{BTIC-C3}-catalysed reaction using a series of C3-functionalized benzothiophenes bearing substitutions at the positions C5–C7 (Fig. 3b). Compared with Mb_{BTIC-C2}, these experiments revealed a markedly different trend in terms of the positional effect of substitutions on the enzyme activity. Specifically, substitutions at position C7 were well tolerated by Mb_{BTIC-C3}, producing **4g–i** with high efficiency (75–86% yields; Fig. 3b). For **3g** and **3i**, nearly quantitative yields (96%) could be achieved by using a higher catalyst loading (1.6 mol% compared with 0.8 mol%). In contrast, the enzyme accepted only limited substitutions (for example, F and CH₃) at positions C5 and C6 of the substrate, with the higher yield of **4f** (62%) compared with **4b** (12%) suggesting a greater tolerance towards C6 substitution compared with C5 substitution. Note that all of these Mb_{BTIC-C3}-catalysed reactions were found to proceed with excellent enantiocontrol, furnishing the desired cyclopropanation products in 97–99% e.e. (Fig. 3b). Notably, the biocatalyst was found to withstand methyl substitution at the α-position of the ester linkage, affording a polycyclic product containing as many as four new stereogenic centres with a high level of diastereoselectivity (**4l**; Fig. 3b). The scope of the biocatalyst and reaction was further explored across different types of heterocycles, including benzofuran, indole and furan, and carbene precursors, such as diazo amides and diazo ketones. Among these, the thiophene **3j** and benzofuran **3k** could be successfully cyclopropanated by Mb_{BTIC-C3} to afford **4j** and **4k** in quantitative yields and 50–99% e.e. (Fig. 3b). Diazo ketone substrate **3m** could also be cyclized to obtain **4m** in 42% yield and with high enantioselectivity (98% e.e.; Fig. 3b). By contrast, benzothiophene- and indole-based diazoamides, furan-based diazoesters, and an indole diazoketone (compounds **7–11**, Supplementary Fig. 1) proved to be unviable substrates for cyclization by Mb_{BTIC-C3} or its precursor variants, indicating that further enzyme engineering would be required to enable these transformations. These limitations notwithstanding, these studies demonstrated that the biocatalytic transformation can be extended to thiophene and benzofuran substrates as well as benzothiophenyl diazo ketones. Importantly, the molecular scaffolds made accessible through these Mb_{BTIC-C2} and Mb_{BTIC-C3}-catalysed reactions combine several highly sought-after features for FBDD^{15,16}, namely 3D shape, stereochemical complexity (three stereogenic centres) and Ro3-compliant physicochemical properties (Fig. 1c).

Mechanistic studies

The haemoprotein-catalysed cyclopropanation reactions of diazo compounds involve the formation of a reactive iron porphyrin–carbene (IPC) intermediate in which the olefin reacts intramolecularly to yield the cyclopropanation product via concerted or stepwise diradical pathways^{35,42}. To gain an insight into the intrinsic mechanism of the reactions investigated here, these transformations were studied using DFT and a truncated model of the ferrous haemoprotein (Fe(II) porphyrin with an axial 4-methylimidazole ligand; see the Supporting Information for further details) to facilitate computations. Initially, the reaction involves the loss of N₂ and formation of the Fe–carbene IPC intermediate **Im1_{C2}** (open-shell singlet, ΔG = 3.6 kcal mol^{−1}) or **Im1_{C3}** (triplet, ΔG = 0.8 kcal mol^{−1}) from **1a** and **3a**, respectively (Fig. 4 and Supplementary Fig. 2). The lowest Gibbs activation energy barriers for this step involve open-shell singlet transition states **TS1_{C2}** (ΔG[‡] = 26.9 kcal mol^{−1}) and **TS1_{C3}** (ΔG[‡] = 27.7 kcal mol^{−1}), respectively (Extended Data Fig. 1 and Supplementary Fig. 8).

In the second step of the reaction, the reactive Fe–carbene (**Im1_{C2}** or **Im1_{C3}**) attacks C2, α to the sulfur atom, via **TS2_{C2}** (quintet, ΔG[‡] = 14.7 kcal mol^{−1}) or C3 via **TS2_{C3}** (open-shell singlet,

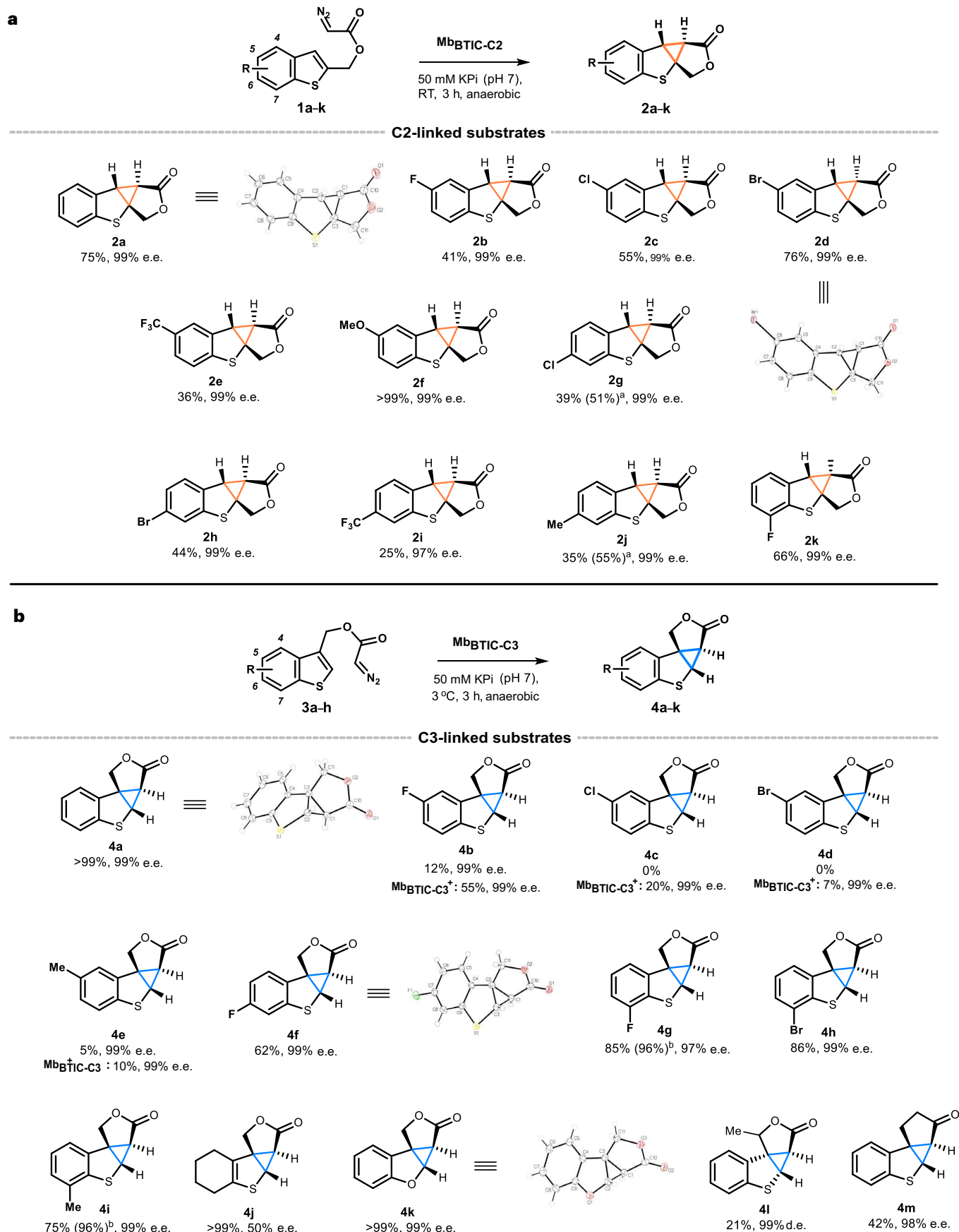


Fig. 3 | Intramolecular cyclopropanation reactions promoted by Mb biocatalysts. **a, b**, Substrate scope of the intramolecular cyclopropanation reactions promoted by Mb_{BTIC-C2} (**a**) and Mb_{BTIC-C3} (**b**) biocatalysts. Reagents and conditions: 20 μ M Mb catalyst, 2.5 mM diazo compound, 10 mM Na₂S₂O₄, 3 h,

anaerobic conditions. All of the reactions with Mb_{BTIC-C3} and Mb(L29F/H64V/V68A/I107V) (Mb_{BTIC-C3}⁺) were performed by slowly adding the diazo compound at 4 °C. ^aUsing whole cells (OD₆₀₀ = 40). ^bUsing 40 μ M Mb catalyst. KPi, potassium phosphate.

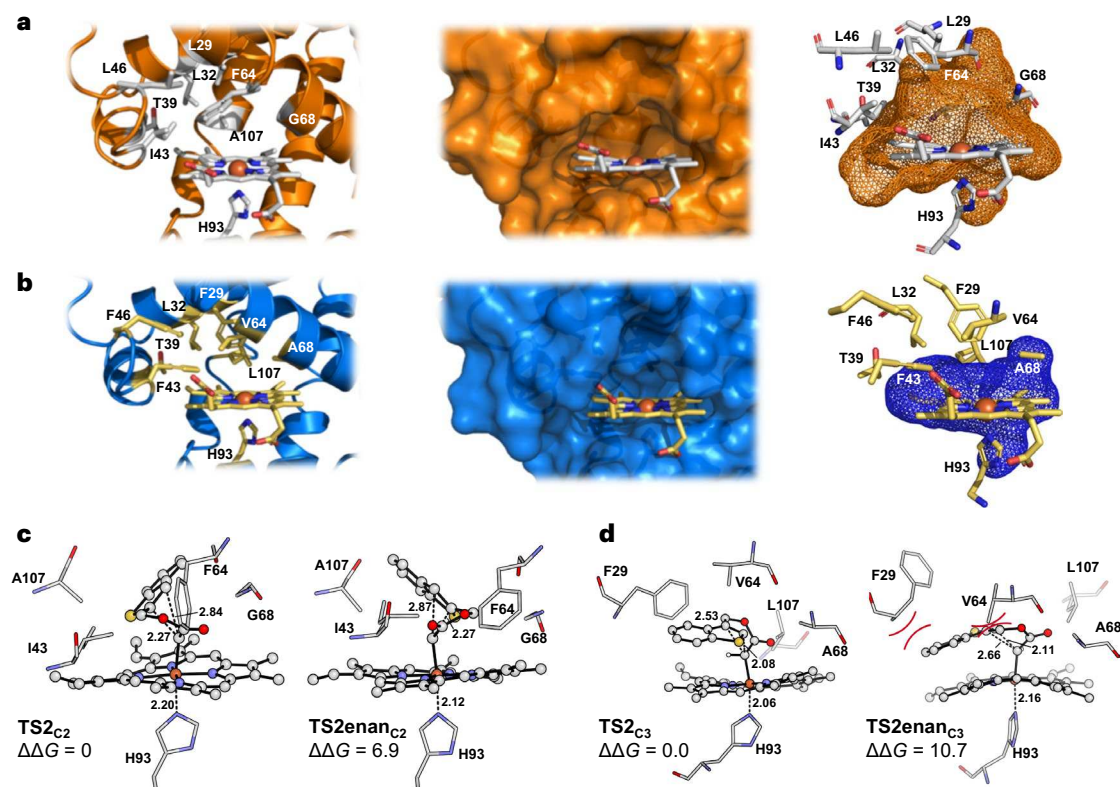


Fig. 4 | Crystal structures of Mb_{BTIC-C2} and Mb_{BTIC-C3} and the corresponding stereochemical models. a,b, High-resolution crystal structures of Mb_{BTIC-C2} (a) and Mb_{BTIC-C3} (b), showing the amino acid residues surrounding the haem cofactor (left), a molecular surface representation of the haem binding site (middle) and the active site volume (right). c,d, Optimized geometries of TS2_{C2}

and its enantiomer TS2_{enanC2} in the theozyme model of Mb_{BTIC-C2} (c) and TS2_{C3} and its enantiomer TS2_{enanC3} in the theozyme model of Mb_{BTIC-C3} (d). The relative Gibbs energies ($\Delta\Delta G$) are given in kcal mol⁻¹ and all distances are in angstroms (Å). See also Supplementary Fig. 4 and Supplementary Information for computational details.

$\Delta G^\ddagger = 16.2$ kcal mol⁻¹) to form a five-membered spirocycle intermediate (Im2_{C2}) or a six-membered ring (Im2_{C3}), respectively, both with a radical localized and stabilized at the benzylic position (Extended Data Fig. 1 and Supplementary Fig. 2). These important geometric differences between the optimized intermediates Im2_{C2} and Im2_{C3} also imply dramatically different geometries of the benzothiophene moiety in the transition states TS2_{C2} and TS2_{C3} during the cyclization step (Extended Data Fig. 1). This is in line with the diverging catalyst requirements observed experimentally for the C2- and C3-substituted substrates. Both intermediates (Im2_{C2} and Im2_{C3}, both triplets) undergo a very fast second C–C bond formation to generate the cyclopropane-fused benzothiophene products 2a and 4a with barriers of less than 4 kcal mol⁻¹ (triplet TS3_{C2}, $\Delta G^\ddagger = 3.4$ kcal mol⁻¹; triplet TS3_{C3}, $\Delta G^\ddagger = 1.4$ kcal mol⁻¹). These results suggest that spin-crossing events occur along the radical reaction pathway, leading to a multi-state mechanism^{43,44}.

This proposed stepwise diradical mechanism differs from the concerted mechanism mediated by a closed-shell singlet IPC intermediate previously established for the Mb-catalysed intermolecular cyclopropanation of vinylarenes with diazo esters⁴². A concerted mechanism for the intramolecular cyclopropanation reaction studied here was calculated to have a slightly higher barrier ($\Delta G^\ddagger = 16.5$ kcal mol⁻¹; Supplementary Fig. 9). This difference can be attributed to the electronic properties of the thiophene and aromatic rings that stabilize the radical formation at the benzylic position. However, it should be noted that our truncated cluster model calculations only included the nearby active site residues. Consideration of the entire enzymatic scaffold in calculations could lead to different geometric or electrostatic features that might favour the alternative concerted mechanism. Nonetheless, both the stepwise and concerted mechanisms highlight the multi-state reactivity of haemoproteins in carbene transfer reactions³⁵.

With respect to the substituted benzothiophene derivatives, similar activation barriers were calculated for the transition state TS1_{C3} of the C5-, C6- or C7-brominated benzothiophenes ($\Delta G^\ddagger \approx 27$ kcal mol⁻¹). Slightly higher activation barriers ($\Delta G^\ddagger \approx 30$ kcal mol⁻¹) were predicted for Me and Cl substituents (Supplementary Table 4). The calculated activation barriers are within 1.6 kcal mol⁻¹ for each substituent at C5, C6 or C7, indicating that the intrinsic substrate reactivity of this reaction step is independent of the position of substituents.

The activity of the Mb catalysts compared with the lack of reactivity of haemin in these reactions (Supplementary Tables 1 and 2) is also noteworthy, highlighting the beneficial effect of the protein scaffold in the catalysis. The calculated barriers for carbene formation with the benzothiophene reagents are in line with the equivalent calculated barrier of ethyl diazoacetate reacting with the Fe porphyrin computational model ($\Delta G^\ddagger = 28.8$ kcal mol⁻¹). Note that these barriers were calculated with respect to the separated reactants, and hence they incorporate an unfavourable entropy ($-\Delta S$) penalty of about 12 kcal mol⁻¹, regardless of the nature of the diazo substrate, which will be overcome by the binding of the substrate to the enzyme ('entropy trap'). As haemin can activate EDA in buffered conditions⁴⁵ and S heterocycles are well known to bind metal centres, the lack of reactivity of haemin in the current reactions may stem from catalyst inhibition and/or catalyst destruction after diazo activation, both of which are disfavoured in the haemoprotein system.

Crystallographic studies

To gain insights into the structures of the evolved enzymes Mb_{BTIC-C2} and Mb_{BTIC-C3}, these proteins were crystallized in their ferric-aquo complex state and their X-ray structures solved at a resolution of 1.3 Å. Structural alignment of these structures with that of wild-type Mb (Protein Data

Bank (PDB): 1JW8)⁴⁶ yielded root-mean-square deviations for the protein backbone of 0.22 and 0.21 Å, respectively, indicating that these proteins share very similar folding patterns. Further inspection of the crystal structures revealed, however, that the two variants exhibit a notably different active site configuration that is shaped by the respective mutations (Fig. 4a,b). In particular, Mb_{B_{TIC}-C₂} features a larger haem pocket than wild-type Mb (volumes of 335 and 125 Å³, respectively) as a result of multiple space-creating mutations, particularly above the inner pyrrole rings A and B of the haem cofactor (Fig. 4a). The Phe residue replacing the distal His64 in Mb_{B_{TIC}-C₂} has two possible orientations of similar occupancy (60:40), acting as a gating residue at the interface between the haem cavity and the solvent (Supplementary Fig. 3). In stark contrast, the active site of Mb_{B_{TIC}-C₃} is characterized by a much smaller volume than Mb_{B_{TIC}-C₂} (volumes of 237 and 335 Å³, respectively; Fig. 4b), as dictated by the presence of Phe29 (from the L29F mutation) and its packing against the side chain of Leu107 (from the I197L mutation). Together with the H64V mutation, this active site configuration completely obstructs and reduces the space available in the ‘inner side’ of the pocket (that is, above haem rings B and A, respectively), while it creates a cavity above the solvent-exposed rim of the porphyrin cofactor (Fig. 4b).

Origin of the enantiocontrol in the enzymatic reactions

Using the X-ray structures, quantum mechanical DFT studies were carried out using the truncated cluster model approach^{47,48} to understand the basis for the high stereoselectivity exhibited by Mb_{B_{TIC}-C₂} and Mb_{B_{TIC}-C₃} in their respective reactions. Starting from Mb_{B_{TIC}-C₂}, DFT calculations were performed to reoptimize TS₂_{C₂} (Fig. 4c), which is the transition state leading to the experimentally generated enantiomer, as well as its enantiomeric transition state (TS₂en_{C₂}) in the presence of nearby amino acid residues in the protein active site, building theoretical enzyme (‘theozyme’) models previously used by our group and others^{49,50}. The final MD snapshot from 500 ns MD simulations were used as starting points for DFT optimized truncated cluster model optimizations (Fig. 4c). In the optimized structure of TS₂_{C₂} in Mb_{B_{TIC}-C₂}, the benzothiophene moiety is placed within the haem pocket, with the heteroaryl ring oriented away from the Phe64 and Ala107 residues and occupying the cavity above haem ring B created by the V68G mutation (Fig. 4c). In contrast, in the optimized complex with TS₂en_{C₂}, which would lead to the opposite enantiomer, the benzothiophene moiety is situated over ring A in close proximity to residues Phe64 and Ile43, resulting in steric clashes with these residues (Fig. 4c). These unfavourable interactions result in a higher energy for TS₂en_{C₂} compared with TS₂_{C₂} ($\Delta\Delta G^\ddagger = +6.9$ kcal mol⁻¹), which explains the high enantioselectivity shown by the enzyme observed experimentally.

We also investigated the origins of the high enantiocontrol exerted by Mb_{B_{TIC}-C₃} in the cyclization of 3a. In the DFT-optimized Mb_{B_{TIC}-C₃} truncated cluster model structure of TS₂_{C₃}, the benzothiophene moiety is oriented above ring D of the haem and sits below Val64 (as in the final MD snapshot), pointing towards the solvent-exposed side of the cofactor (Fig. 4d). This configuration is largely dictated by the presence of the Phe29 residue, which occupies the inner side of the haem pocket (Fig. 4b). In the optimized structure of TS₂en_{C₃}, the benzothiophene moiety clashes sterically with Val64 and displaces the Phe29 ring, resulting in a ΔG value that is 10.7 kcal mol⁻¹ higher than that of TS₂_{C₃}. This large energy difference thus rationalizes the excellent enantiomeric excess (>99% e.e.) observed for the 3aS,3bS,8bS-configured product generated by the Mb_{B_{TIC}-C₃} variant.

MD simulations of the Mb_{B_{TIC}-C₂} system

We performed MD simulations to investigate how the protein scaffold and additional amino acid residues of Mb_{B_{TIC}-C₂} contribute to the stabilization of the TS₂_{C₂} transition state (Fig. 5a). In these studies, the lowest-energy TS₂_{C₂} transition state optimized by DFT calculations was docked into the full protein structure in the presence of explicit

water as solvent. Two independent 500 ns MD simulations starting from the optimized TS₂_{C₂} docked near the enzyme ‘gate’ allowed the exploration of binding poses with the aromatic end of the substrate deep inside the active site pocket along the MD trajectories. Steric clashes with residues Phe64, Ile43, Ala107 and Leu29 induce a rotation of the benzothiophene moiety away from the solvent-exposed active site ‘gate’ in the haemoprotein. Such conformational flexibility is consistent with the different orientations found for residue Phe64 in the crystal structure and further reveals its role as a ‘gatekeeper’ residue between the haem pocket and the solvent.

To rationalize the structure–activity trends observed for the Mb_{B_{TIC}-C₂} biocatalyst (Fig. 3a), equivalent TS₂-docked MD simulations were performed using 5-methoxybenzothiophene (5OMe, 1f) and 6-methylbenzothiophene (6Me, 1j) derivatives as the substrates. While the overall structural features of the resulting TS₂_{C₂} complexes were similar to those obtained with 1a, important differences were also evident. With the 5OMe substrate (1f), rotation of the benzothiophenyl ester moiety around the Fe–porphyrin ring was much faster along MD trajectories than was observed with substrate 1a (Fig. 5b). In addition, during the MD trajectory, the molecule reorients such that the thiophene ring lies above haem ring C (compared with ring B for 1a) and the polar oxygen atom of the methoxy group points towards the solvent (Fig. 5c). In contrast, MD simulations of the 6Me substrate (1j) showed the benzothiophene moiety to be embedded deep inside the enzyme pocket and shielded from the solvent (Fig. 5c), adopting a conformation that differs from those observed with substrates 1a and 1f. Experimentally, the activity of the enzyme towards these substrates decreases in the order 1f (5OMe) > 1a > 1j (6Me; Fig. 3a). As this trend correlates with the degree of exposure of the benzothiophene moiety to the solvent in the corresponding transition state as determined by TS₂-docked MD simulations, we hypothesize that substrates capable of adopting more solvent-exposed orientations during the cyclopropanation step may be more efficiently processed by the enzyme.

Docking and MD simulations of Mb_{B_{TIC}-C₃}

To better understand the reactivity of Mb_{B_{TIC}-C₃} and its lineage, substrate 3a as well as the carbene derived from 3a were docked into the crystal structure of Mb_{B_{TIC}-C₃}. In both cases, the lowest-energy binding pose from docking predictions shows that the benzothiophene (BT) ring is exposed to the solvent (‘BT out’ conformation) and binds to a crevice on the protein surface created by the H64V mutation (Fig. 5d). Interestingly, similar docking studies on Mb(H64V,V68A), the earliest intermediate in the Mb_{B_{TIC}-C₃} evolutionary lineage (Fig. 2c), predict an alternative conformation (‘BT in’ conformation) in which the benzothiophene moiety is buried inside the protein (Supplementary Fig. 5). In Mb_{B_{TIC}-C₃}, the latter conformation is prevented by steric clashes with the bulky Phe residue at position 29, introduced through the highly beneficial L29F mutation (Fig. 4b). Thus, in addition to dictating high enantioselectivity in the cyclopropanation step (Fig. 4d), the L29F mutation contributes to favour the ‘BT out’ conformation, which is associated with more efficient catalysis, as described earlier.

MD simulations were then performed using the lowest-energy TS₂_{C₃} transition state docked into the Mb_{B_{TIC}-C₃} active site as the starting point. Interestingly, these simulations showed how, starting from the ‘BT out’ conformation, the haem-bound substrate briefly samples a slightly different conformation during the MD trajectory in which the benzothiophene moiety pushes away the nearby residue Phe43 (Fig. 5f). However, the substrate quickly reorients itself to a final orientation that is nearly identical to the starting point, that is, with the benzothiophenyl moiety lying over ring D of the haem (Fig. 5e,f). As is evident from these studies, both the accessible conformations and the preferred orientation of the haem-bound TS₂_{C₃} in Mb_{B_{TIC}-C₃} (Fig. 5e,f) differ completely from those found for the haem-bound TS₂_{C₂} in Mb_{B_{TIC}-C₂} (Fig. 5a,b), where the benzothiophene moiety is pushed above ring B of the haem (compared with ring D for 3a) and towards

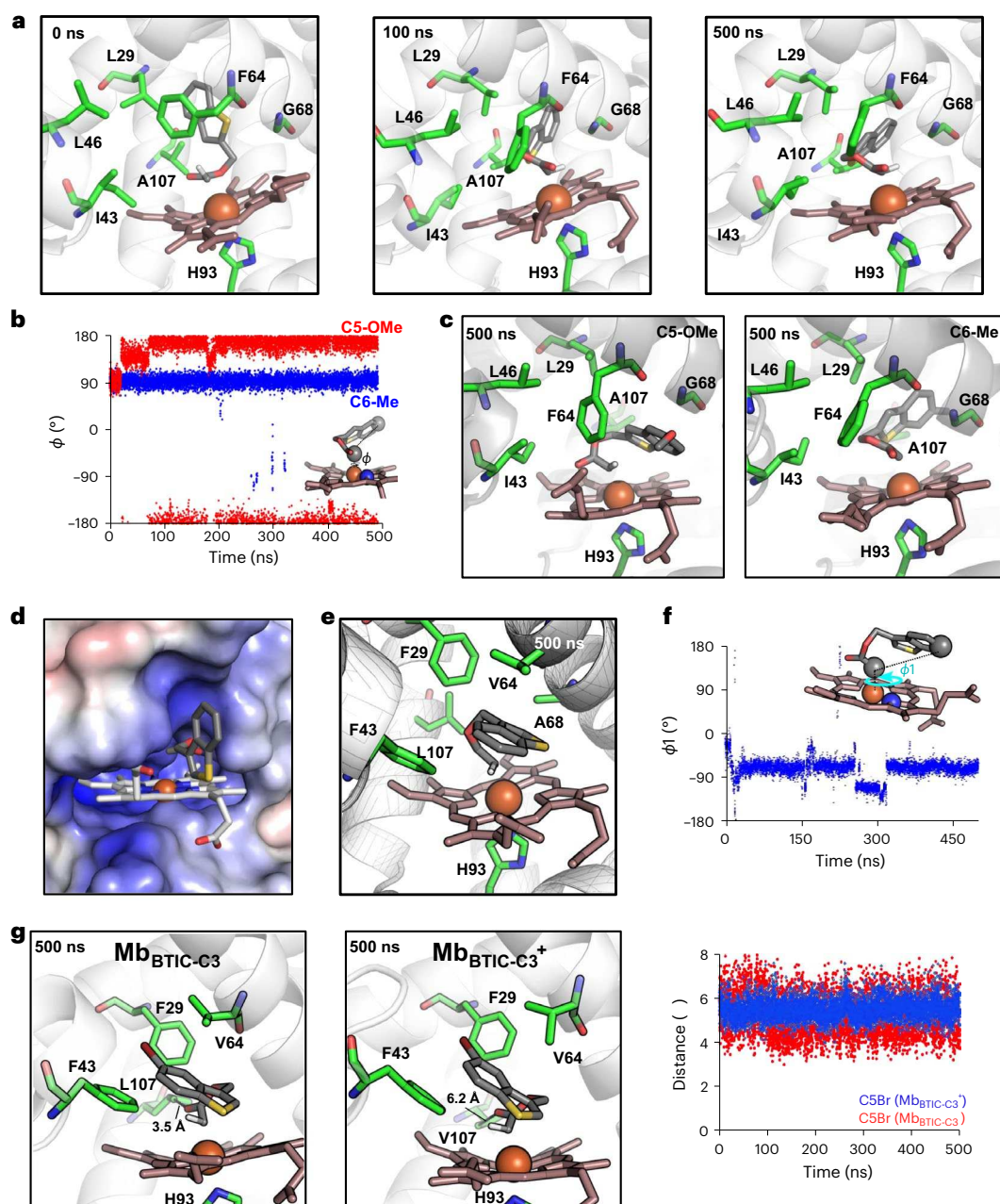


Fig. 5 | Analysis of the Mb_{BTIC-C2}, Mb_{BTIC-C3} and Mb_{BTIC-C3}⁺-catalysed cyclization reactions by MD simulations. **a**, MD snapshots of TS_{2C2} docked into Mb_{BTIC-C2} protein at 0, 100 and 500 ns. **b**, Variation of the pseudo-dihedral angle (ϕ) in the MD simulations of the haem-carbene intermediates corresponding to the 5OMe and 6Me substrates (that is, compounds **1f** and **1j**) in Mb_{BTIC-C2}. The atoms used for the measurement of the ϕ angle are displayed in the structure as spheres. **c**, End-of-simulation (500 ns) poses for the TS_{2C2} transition states corresponding to the 5OMe and 6Me substrates in Mb_{BTIC-C2}. See Supplementary Fig. 5 for additional time points. **d**, Lowest-energy pose for the diazo substrate **3a** after molecular docking into the structure of Mb_{BTIC-C3}, showing the benzothiophene moiety in the 'BT out' conformation. **e**, End-of-simulation (500 ns) pose for the TS_{2C3} complex in Mb_{BTIC-C3}. See Supplementary Fig. 6 for additional time points.

f, Variation of the pseudo-dihedral angle (ϕ_1) across the different geometries of the haem-bound TS_{2C3} complex sampled during the 500 ns MD simulation. The reference atoms for the ϕ_1 angle are displayed in the structure as spheres. **g**, MD snapshots at 500 ns of the TS_{2C3} analogues of the C5-Br-substituted substrate (compound **3d**) in Mb_{BTIC-C3} (left) and Mb_{BTIC-C3}⁺ (middle), which is the Mb_{BTIC-C3} variant containing the rationally designed L107V mutation for enhanced reactivity towards 5-substituted benzothiophenes. The closest distance between the carbonyl (C=O) oxygen atom and residue 107 in these complexes is indicated. The graph (right) compares the variation of these distances for the C5-Br-substituted substrates in Mb_{BTIC-C3} and Mb_{BTIC-C3}⁺ throughout the 500 ns MD simulation.

the inner side of the haem pocket by the concerted action of residues Leu29 and Phe64. These features thus explain the orthogonal reactivity of these biocatalysts towards the cyclopropanation of the C2- and C3-functionalized benzothiophene substrates due to different active site requirements to stabilize the different key transition states TS_{2C2} and TS_{2C3}.

Rational design of the improved biocatalyst Mb_{BTIC-C3}⁺
Leveraging the mechanistic insights from the studies above, we sought to understand and potentially overcome the limited substrate scope of Mb_{BTIC-C3}, which shows very low tolerance towards substitutions at the C5 position of benzothiophene (Fig. 3b). MD simulations of the transition state TS₂ of the inactive 5-bromo-substituted substrate (C5Br)

bound in Mb_{B_{BTIC-C3}} showed that **TS2** mainly explores an orientation in the active site in which the ester group of the substrate lies close to Leu107. This suggests that these steric clashes disfavour the transition state **TS2**. Based on these analyses, we designed a variant, denoted Mb_{B_{BTIC-C3}}⁺, in which an L107V mutation was introduced that reduces these steric clashes, facilitating a productive binding mode for **TS2**, as confirmed by MD simulations of the C5Br–Mb_{B_{BTIC-C3}}⁺ system (Fig. 5g). Gratifyingly, Mb_{B_{BTIC-C3}}⁺ exhibited up to fivefold improved activity compared with Mb_{B_{BTIC-C3}} for the synthesis of C5-substituted substrates **4b** and **4e** (Fig. 3b). Furthermore, albeit in low-to-moderate yields (7–20%), this biocatalyst also produced the 5-chloro- and 5-bromo-substituted products **4c** and **4d**, respectively, which could not be attained using Mb_{B_{BTIC-C3}} (Fig. 3b). While expanding the substrate scope of the present method, these results highlight the value of combining evolutionary and rational strategies for the development of these carbene transferases.

Conclusion

We have reported here the development of a biocatalytic strategy for the construction of *sp*³-rich tetracyclic compounds via the intramolecular cyclopropanation of benzothiophenes. These molecular scaffolds, which are unexplored in medicinal chemistry, embody several highly sought-after features for use as ‘fragments’ in FBDD owing to their 3D and stereochemical complexity combined with Ro3-compliant properties (Fig. 1c). Despite the inherent challenge posed by these transformations using chemical and biological carbene transfer catalysts alike (Fig. 1a and Supplementary Table 1), we have developed two complementary biocatalysts by directed evolution from Mb that enable the highly stereoselective cyclization of both C2- and C3-functionalized benzothiophenes and related heterocycles. Our mechanistic and DFT studies demonstrate a diradical mechanism for these intramolecular cyclopropanation reactions, which contrasts with the concerted mechanism previously observed for Mb-catalysed intermolecular cyclopropanation with diazo esters⁴², highlighting the mechanistic plasticity of this haemoprotein in carbene transfer reactions. Crystallographic and computational (MD) studies have provided additional insights into the structural features of the two evolved biocatalysts that are responsible for the high level of stereocontrol in the corresponding reactions as well as for their orthogonal (and complementary) reactivity. We also used the results of computational modelling studies to engineer a further biocatalyst to expand the substrate scope of the Mb_{B_{BTIC-C3}} enzyme to enable the transformation of a series of previously unreactive substrates. Our results highlight the benefits of combining evolutionary and rational strategies for the development of enzymes to catalyse synthetically useful transformations not found in nature.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41557-023-01435-3>.

References

- Bornscheuer, U. T. et al. Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
- Devine, P. N. et al. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2**, 409–421 (2018).
- Slagman, S. & Fessner, W. D. Biocatalytic routes to anti-viral agents and their synthetic intermediates. *Chem. Soc. Rev.* **50**, 1968–2009 (2021).
- Savile, C. K. et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).
- Huffman, M. A. et al. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* **366**, 1255–1259 (2019).
- McIntosh, J. A. et al. Engineered ribosyl-1-kinase enables concise synthesis of molnupiravir, an antiviral for COVID-19. *ACS Cent. Sci.* **7**, 1980–1985 (2021).
- Brandenberg, O. F., Fasan, R. & Arnold, F. H. Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. *Curr. Opin. Biotechnol.* **47**, 102–111 (2017).
- Schwizer, F. et al. Artificial metalloenzymes: reaction scope and optimization strategies. *Chem. Rev.* **118**, 142–231 (2018).
- Ren, X. & Fasan, R. Engineered and artificial metalloenzymes for selective C–H functionalization. *Curr. Opin. Green Sustain. Chem.* **31**, 100494 (2021).
- Liu, Z. & Arnold, F. H. New-to-nature chemistry from old protein machinery: carbene and nitrene transferases. *Curr. Opin. Biotechnol.* **69**, 43–51 (2021).
- Ramsden, J. I., Cosgrove, S. C. & Turner, N. J. Is it time for biocatalysis in fragment-based drug discovery? *Chem. Sci.* **11**, 11104–11112 (2020).
- Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619 (2016).
- Erlanson, D. A., Davis, B. J. & Jahnke, W. Fragment-based drug discovery: advancing fragments in the absence of crystal structures. *Cell Chem. Biol.* **26**, 9–15 (2019).
- Jhoti, H., Williams, G., Rees, D. C. & Murray, C. W. The ‘rule of three’ for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discov.* **12**, 644 (2013).
- Murray, C. W. & Rees, D. C. Opportunity knocks: organic chemistry for fragment-based drug discovery (FBDD). *Angew. Chem. Int. Ed.* **55**, 488–492 (2016).
- Morley, A. D. et al. Fragment-based hit identification: thinking in 3D. *Drug Discov. Today* **18**, 1221–1227 (2013).
- Over, B. et al. Natural-product-derived fragments for fragment-based ligand discovery. *Nat. Chem.* **5**, 21–28 (2013).
- Yu, B., Zheng, Y. C., Shi, X. J., Qi, P. P. & Liu, H. M. Natural product-derived spirooxindole fragments serve as privileged substructures for discovery of new anticancer agents. *Anticancer Agents Med. Chem.* **16**, 1315–1324 (2016).
- Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307–310 (2013).
- Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744–1748 (2015).
- Tinoco, A., Steck, V., Tyagi, V. & Fasan, R. Highly diastereo- and enantioselective synthesis of trifluoromethyl-substituted cyclopropanes via myoglobin-catalyzed transfer of trifluoromethylcarbene. *J. Am. Chem. Soc.* **139**, 5293–5296 (2017).
- Key, H. M. et al. Beyond iron: iridium-containing P450 enzymes for selective cyclopropanations of structurally diverse alkenes. *ACS Cent. Sci.* **3**, 302–308 (2017).
- Chandgude, A. L. & Fasan, R. Highly diastereo- and enantioselective synthesis of nitrile-substituted cyclopropanes by myoglobin-mediated carbene transfer catalysis. *Angew. Chem. Int. Ed.* **57**, 15852–15856 (2018).
- Knight, A. M. et al. Diverse engineered heme proteins enable stereodivergent cyclopropanation of unactivated alkenes. *ACS Cent. Sci.* **4**, 372–377 (2018).
- Wittmann, B. J. et al. Diversity-oriented enzymatic synthesis of cyclopropane building blocks. *ACS Catal.* **10**, 7112–7116 (2020).
- Nam, D., Steck, V., Potenzino, R. J. & Fasan, R. A diverse library of chiral cyclopropane scaffolds via chemoenzymatic assembly and diversification of cyclopropyl ketones. *J. Am. Chem. Soc.* **143**, 2221–2231 (2021).

27. Vargas, D. A., Tinoco, A., Tyagi, V. & Fasan, R. Myoglobin-catalyzed C–H functionalization of unprotected indoles. *Angew. Chem. Int. Ed.* **57**, 9911–9915 (2018).
28. Brandenburg, O. F., Chen, K. & Arnold, F. H. Directed evolution of a cytochrome P450 carbene transferase for selective functionalization of cyclic compounds. *J. Am. Chem. Soc.* **141**, 8989–8995 (2019).
29. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat. Commun.* **6**, 7789 (2015).
30. Sreenilayam, G., Moore, E. J., Steck, V. & Fasan, R. Metal substitution modulates the reactivity and extends the reaction scope of myoglobin carbene transfer catalysts. *Adv. Synth. Catal.* **359**, 2076–2089 (2017).
31. Dydio, P. et al. An artificial metalloenzyme with the kinetics of native enzymes. *Science* **354**, 102–106 (2016).
32. Ohora, K. et al. Catalytic cyclopropanation by myoglobin reconstituted with iron porphycene: acceleration of catalysis due to rapid formation of the carbene species. *J. Am. Chem. Soc.* **139**, 17265–17268 (2017).
33. Villarino, L. et al. An artificial heme enzyme for cyclopropanation reactions. *Angew. Chem. Int. Ed.* **57**, 7785–7789 (2018).
34. Zhao, J. M., Bachmann, D. G., Lenz, M., Gillingham, D. G. & Ward, T. R. An artificial metalloenzyme for carbene transfer based on a biotinylated dirhodium anchored within streptavidin. *Catal. Sci. Technol.* **8**, 2294–2298 (2018).
35. Carminati, D. M. & Fasan, R. Stereoselective cyclopropanation of electron-deficient olefins with a cofactor redesigned carbene transferase featuring radical reactivity. *ACS Catal.* **9**, 9683–9697 (2019).
36. Stenner, R., Steventon, J. W., Seddon, A. & Anderson, J. L. R. A de novo peroxidase is also a promiscuous yet stereoselective carbene transferase. *Proc. Natl Acad. Sci. USA* **117**, 1419–1428 (2020).
37. Chandgude, A. L., Ren, X. & Fasan, R. Stereodivergent intramolecular cyclopropanation enabled by engineered carbene transferases. *J. Am. Chem. Soc.* **141**, 9145–9150 (2019).
38. Ren, X. K., Liu, N. Y., Chandgude, A. L. & Fasan, R. An enzymatic platform for the highly enantioselective and stereodivergent construction of cyclopropyl- δ -lactones. *Angew. Chem. Int. Ed.* **59**, 21634–21639 (2020).
39. Yong, K., Salim, M. & Capretta, A. Intramolecular carbenoid insertions: reactions of α -diazo ketones derived from furanyl-, thienyl-, (benzofuranyl)-, and (benzothienyl)acetic acids with rhodium(II) acetate. *J. Org. Chem.* **63**, 9828–9833 (1998).
40. Padwa, A., Wisnieff, T. J. & Walsh, E. J. Synthesis of cycloalkenones via the intramolecular cyclopropanation of furanyl diazo ketones. *J. Org. Chem.* **51**, 5036–5038 (1986).
41. Zhang, R. K. et al. Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp^3 C–H functionalization. *Nature* **565**, 67–72 (2019).
42. Wei, Y., Tinoco, A., Steck, V., Fasan, R. & Zhang, Y. Cyclopropanations via heme carbenes: basic mechanism and effects of carbene substituent, protein axial ligand, and porphyrin substitution. *J. Am. Chem. Soc.* **140**, 1649–1662 (2018).
43. de Visser, S. P., Ogliaro, F., Harris, N. & Shaik, S. Multi-state epoxidation of ethene by cytochrome P450: a quantum chemical study. *J. Am. Chem. Soc.* **123**, 3037–3047 (2001).
44. Schroder, D., Shaik, S. & Schwarz, H. Two-state reactivity as a new concept in organometallic chemistry. *Acc. Chem. Res.* **33**, 139–145 (2000).
45. Sreenilayam, G. & Fasan, R. Myoglobin-catalyzed intermolecular carbene N–H insertion with arylamine substrates. *Chem. Commun.* **51**, 1532–1534 (2015).
46. Kondrashov, D. A., Zhang, W., Aranda, R. T., Stec, B. & Phillips, G. N. Jr. Sampling of the native conformational ensemble of myoglobin via structures in different crystalline environments. *Proteins* **70**, 353–362 (2008).
47. Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R. Z. & Siegbahn, P. E. M. Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* **114**, 3601–3658 (2014).
48. Himo, F. Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* **139**, 6780–6786 (2017).
49. Lind, M. E. S. & Himo, F. Quantum chemistry as a tool in asymmetric biocatalysis: limonene epoxide hydrolase test case. *Angew. Chem. Int. Ed.* **52**, 4563–4567 (2013).
50. Liao, R. Z. & Thiel, W. On the effect of varying constraints in the quantum mechanics only modeling of enzymatic reactions: the case of acetylene hydratase. *J. Phys. Chem. B* **117**, 3954–3961 (2013).
51. Ryan, J. et al. Transaminase triggered aza-Michael approach for the enantioselective synthesis of piperidine scaffolds. *J. Am. Chem. Soc.* **138**, 15798–15800 (2016).
52. Zawodny, W. et al. Chemoenzymatic synthesis of substituted azepanes by sequential biocatalytic reduction and organolithium-mediated rearrangement. *J. Am. Chem. Soc.* **140**, 17872–17877 (2018).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

Methods

General information

All the chemicals and reagents used in this study were purchased from commercial suppliers (Sigma-Aldrich, Alfa Aesar, ACS Scientific, Acros, Ambeed and Combi-blocks) and used without further purification. All dry reactions were carried out under argon in flame-dried glassware with magnetic stirring using standard gas-tight syringes, cannula and septa. ^1H and ^{13}C NMR spectra were measured on a Bruker DPX-500 (operating at 500 MHz for ^1H and 125 MHz for ^{13}C) or Bruker DPX-400 spectrometer (operating at 400 MHz for ^1H and 100 MHz for ^{13}C). ^{19}F NMR spectra were measured on a Bruker DPX-400 spectrometer (operating at 375 MHz). The NMR spectrometers were operated using Topspin X (version 1.3.10) software. Tetramethylsilane ($\delta = 0$ ppm) and/or CDCl_3 ($\delta = 7.26$ ppm) served as internal standards for ^1H NMR, CDCl_3 ($\delta = 77.0$ ppm) was used as the internal standard for ^{13}C NMR and trifluorotoluene ($\delta = -63$ ppm) served as the internal standard for ^{19}F NMR. Silica gel chromatography purifications were carried out using AMD silica gel 60 (230–400 mesh). Thin layer chromatography (TLC) was carried out using Merck Millipore TLC silica gel 60 F254 glass plates.

Cloning and library construction

Mb variants were cloned and expressed using pET22b(+) vectors (Novagen). The Mb gene was fused to the carboxy terminal of a 6xHis tag and controlled by an β -D-1-thiogalactopyranoside (IPTG)-inducible T7 promoter. Site-directed mutagenesis variants were prepared via the QuickChange mutagenesis protocol using KOD Hot Start DNA polymerase. Site-saturation mutagenesis libraries were constructed using a similar PCR procedure and degenerate primers containing NNK codons (where N = A, G, C or T, and K = G or T; Supplementary Table 13). After transformation of the PCR products into chemically competent *Escherichia coli* DH5 α cells, recombinant colonies were pooled, grown in 5 mL LB media containing 100 mg l $^{-1}$ ampicillin and the plasmid extracted using a Qiagen mini-prep kit. Next, 1 μl of the DNA library solution was transformed into chemically competent *E. coli* C41(DE3) cells for protein expression. Single colonies were grown in 48- or 96-well deep well plates and screened for activity as whole cells as described below. For crystallization, Mb was expressed and purified without the His tag.

Protein expression and purification

Engineered Mb variants were expressed in *E. coli* C41(DE3) cells in Terrific Broth supplemented with 100 mg l $^{-1}$ ampicillin at 37 °C until the OD $_{600}$ reached 0.9–1.2. The cells were then induced with 0.25 mM IPTG and 0.3 mM δ -aminolevulinic acid. After induction, the cultures were shaken at 27 °C and collected after 18–20 h by centrifugation at 3,000 g at 4 °C. After cell lysis by sonication, the proteins were purified by Ni-affinity chromatography. The cell lysate was transferred to a Ni-nitrilotriacetate (Ni-NTA) column equilibrated with Ni-NTA lysis buffer. The resin was first washed with 50 mL Ni-NTA lysis buffer and then 50 mL Ni-NTA wash buffer (50 mM KPi, 250 mM NaCl, 20 mM imidazole, pH 8.0). The proteins were eluted with Ni-NTA elution buffer (50 mM KPi, 250 mM NaCl, 250 mM histidine, pH 7.0). After elution, the proteins were buffer-exchanged against 50 mM KPi buffer (pH 7.0 or 8.0) using 10 KDa Centricon filters. The concentration of ferric Mb was determined using an extinction coefficient at 410 nm (ϵ_{410}) of 157 mM $^{-1}$ cm $^{-1}$.

Library screening

Reactions for library screening were carried out at a 400 μl scale using C41(DE3) *E. coli* cells expressing the Mb variants and 2.5 mM diazo-benzothiophene. In a typical procedure, recombinant cells were grown in 24-, 48- or 96-well plates, and protein expression was performed as described above. After expression, the cells were pelleted by centrifugation and transferred to an anaerobic chamber. The cell pellet in each well was resuspended in 390 μl degassed KPi buffer (50 mM KPi, pH 7.0). The reactions were initiated by the addition of 10 μl of a 100 mM stock solution of the diazo compound in ethanol or dimethylsulfoxide

(DMSO). The plates were covered with aluminium foil and shaken for 3–16 h at room temperature in the anaerobic chamber. The reactions were then analysed following the procedure described in the Product analysis section below.

Enzymatic reactions

Analytical scale reactions (400 μl) were conducted using whole cells (final OD $_{600}$ = 20 or 60) or purified Mb variant (final concentration 20 μM), 2.5 mM diazo-benzothiophene and 10 mM sodium dithionite ($\text{Na}_2\text{S}_2\text{O}_4$) in crimp vials. In a typical procedure, a vial containing the cell suspension or purified Mb was transferred into a Coy anaerobic chamber. Then, degassed 50 mM KPi buffer (pH 7.0) was added to the vessel to obtain the desired cell density for whole-cell reactions or enzyme concentration. For the reactions with purified proteins, the protein was diluted by the addition of 40 μl of freshly prepared sodium dithionite stock solution (100 mM in 50 mM KPi, pH 7.0). The reactions were initiated by the addition of 10 μl of 100 mM stock solution of the diazo compound in ethanol or DMSO and shaken under magnetic agitation for 3–16 h at room temperature. For reactions involving slow substrate addition, the diazo compound (100 mM stock solution in ethanol or DMSO) was added in four aliquots over the course of 1 h. For reactions at 3 °C, the reaction vessel was kept on ice. The turnover number for the whole-cell reactions was calculated on the basis of the Mb concentration in the reaction mixture as measured by UV-visible spectroscopy ($\epsilon_{410} = 156 \text{ mM}^{-1} \text{ cm}^{-1}$) after cell lysis. Preparative-scale reactions were performed following similar procedures but using larger volumes (25–50 mL).

Product analysis

The reaction mixtures were mixed with 20 μl internal standard (50 mM benzodioxole in ethanol) and extracted with 400 μl dichloromethane. The organic layer was then analysed by gas chromatography with flame ionization detection (GC-FID) using a Shimadzu GC-2010 gas chromatograph equipped with an FID detector and a chiral Cyclosil-B column (30 m $\times$ 0.25 mm $\times$ 0.25 μm film). Separation was achieved with 1 μl injection, an injector temperature of 300 °C and a detector temperature of 300 °C; the column temperature was set at 80 °C for 2 min, then raised to 245 °C at 40 °C min $^{-1}$ with a 10 min hold for a total run time of 16.1 min. The enantiomeric excesses were determined by chiral GC-FID and calibration curves to determine product yields were constructed using authentic standards prepared via enzymatic reactions in whole cells or with purified Mb_{BTIC-C2} or Mb_{BTIC-C3}, as described in the Supplementary Information. All measurements were performed at least in duplicate. For each experiment, negative control samples without protein were included.

Crystallization, data collection, processing and refinement

Crystals of ferric Mb_{BTIC-C2} and Mb_{BTIC-C3} complexed with water were grown at room temperature using the hanging-drop vapour diffusion method by mixing 1 μl reservoir buffer (2.4 M ammonium sulfate, 0.2 M Tris, pH 9.0, 0.1 mM EDTA, pH 9 for Mb_{BTIC-C3} and pH 8.2 for Mb_{BTIC-C2}) with 1 μl protein in buffer (4 mM, 20 mM Tris, pH 8.0, 1 mM EDTA). The crystals were cryoprotected by transferring them to a drop containing a 1:1 mixture of paratone and silicone oil, moved around until the solvent was replaced and then flash-cooled in liquid nitrogen. Data were collected remotely at Stanford Synchrotron Radiation Light-source and in house using a Rigaku XtaLAB Synergy-S microfocus X-ray diffraction system. Diffraction data for Mb_{BTIC-C3} were reduced using XDS software and the CCP4 programs AIMLESS and POINTLESS via an autoXDS script; the diffraction data for Mb_{BTIC-C2} were reduced using CrysAlisPro (version 171.41), XDS and AIMLESS. The Mb_{BTIC-C3} structure was solved using the Fourier synthesis method beginning with the available structure of Mb(H64V,V68A) (PDB: 6M8F); this structure was also used to solve Mb_{BTIC-C2} by molecular replacement in PHENIX, followed by further rounds of refinement and model building using PHENIX and COOT. The data processing and refinement statistics are presented in

Supplementary Tables 6 and 7. Structural coordinates have been deposited at the Protein Data Bank under the accession numbers **7SLH** and **7SLI** for Mb_{BTIC-C3} and Mb_{BTIC-C2}, respectively. For comparison purposes (Fig. 2), a structure of wild-type sperm whale Mb (ferrous CO complex), which crystallizes in the same space group (*P6*), was chosen owing to the subtle conformational differences observed in Mb under different crystalline environments (PDB: **IJW8**). The structure of Mb_{BTIC-C2} presents two alternative orientations of the distal phenylalanine residue (Phe64), both of which are shown in Supplementary Fig. 3.

Quantum mechanical calculations

All geometry optimizations, transition-state searches and single-point computations were carried out using Gaussian 16, revision C.01. The periphery of the porphyrin moiety was replaced by H atoms. In addition, His93 was modelled by 5-methylimidazole. The geometries were optimized using the unrestricted version of the B3LYP hybrid functional with D3 dispersion corrections. The Stuttgart basis set and associated relativistic pseudopotential were used for Fe, the double-zeta basis set 6-31G+(d) was used for S, O, N, C and H, and the larger def2TZVP basis set was used for all atoms to calculate single-point energies. Single-point energy calculations were performed to mimic the protein environment. Truhlar's continuum solvation model based on density (SMD) was applied for solvent diethyl ether with a dielectric constant of 4.33. Thermal and entropic corrections to energies 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. Different spin states for Fe(III) were considered (Supplementary Fig. 8). The calculated Gibbs free energies reported in the main text were obtained after applying corrections using quasi-harmonic approximations in Goodvibes³. For the theozyme calculations, the MD snapshot at 500 ns was used as the input structure. For the theozyme calculations, the final snapshots of 500 ns restrained MD simulations of **TS2_{C3}** (and **TS2_{C2}**) were used to generate input geometry for the catalytic site, transition state and nearby amino acid residues. The backbone carbon atoms and amide bonds of the residues were kept fixed during DFT reoptimizations.

MD simulations

TS2_{C3} and **TS2_{C2}** are important in the Mb-catalysed cyclopropanation of 3- and 2-benzothienylmethyl diazoacetates to study both the rate-determining step and enantioselectivity of the reaction. MD simulations were performed on these species and substituted analogues of **TS2_{C3}** and **TS2_{C2}** to better understand the catalytic activity of the enzymes. Simulations were performed using the graphical processing unit (GPU) code (pmemd) of the AMBER18 package. Substrate parameters were generated within the antechamber module using the general AMBER force field, with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level using the RESP model. The charges were calculated according to the Merz–Singh–Kollman scheme using Gaussian 16. As we were interested in the effects of amino acids on the catalytic activity, we restrained the C–C distances and C₃–C₂–C₆–O torsion angle throughout the MD simulations to those in the optimized geometry of **TS2_{C2}** leading to the cyclopropanation using a harmonic restraint of 120 kcal mol^{−1} Å^{−2}. Each protein was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of three-site water model (TIP3P) using the leap module, resulting in the addition of around 6,727 solvent molecules. The systems were neutralized by the addition of explicit counterions (Na⁺ and Cl[−]). All subsequent calculations were carried out using the widely used Stony Brook modification of the Amber 14 force field (ff14sb).

A four-stage geometry optimization was performed. In the first stage, the positions of solvent molecules and ions were minimized, imposing positional restraints on the solute by a harmonic potential with a force constant of 10 kcal mol^{−1} Å^{−2}. In the second, third and fourth stages, the transition state bound to the enzyme was allowed to relax slowly with restraint values of 5.0, 2.0 and 0.5 kcal mol^{−1} Å^{−2} in

the simulation cell, except for those involved in the harmonic distance restraint. The systems were gently heated in six 50 ps steps, incrementing the temperature by 50 K in 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. A 10 Å cut-off was applied to Lennard-Jones and electrostatic interactions. Harmonic restraints of 20 kcal mol^{−1} Å^{−2} 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. The systems were then equilibrated in a stepwise manner with each step equilibrated with restraints of 10, 5.0, 2.5 and 1.0 kcal mol^{−1} Å^{−2} on the enzyme–substrate complex for each step of 2 ns simulation with a 1 fs time step at a constant pressure. Production trajectories were then run for an additional 500 ns with a 1 fs time step at a constant pressure. Docking files were prepared using AutoDock Tools (version 1.5.7). Substrates were docked into the active site of Mb(H64V,V68A) and Mb_{BTIC-C3} using AutoDock with a grid cube of length 50 Å and with Fe atom at the centre of the cube.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Protein crystal structure coordinates have been deposited with the Protein Data Bank (PDB) under accession numbers **7SLH** (Mb_{BTIC-C3}) and **7SLI** (Mb_{BTIC-C2}). Crystallographic data for the structures reported in this Article have been deposited at the Cambridge Crystallographic Data Centre under CCDC deposition numbers **2157009 (2a)**, **2157011 (2d)**, **2157007 (4a)**, **2157010 (4f)** and **2157008 (4k)**. Copies of the data can be obtained free of charge via <https://www.ccdc.cam.ac.uk/structures/>.

Acknowledgements

This work was supported by the US National Institutes of Health (grant no. GM098628, R.F.). R.F. acknowledges support from the Cancer Prevention and Research Institute of Texas (CPRIT RR230018) and Robert A. Welch Foundation (Chair, AT-0051). D.A.V. acknowledges support from the National Science Foundation Graduate Fellowship Program. K.N.H. and A.S. acknowledge support from the National Science Foundation (CHE-1764328). M.G.-B. acknowledges support from the Spanish Ministerio de Ciencia e Innovación (MICINN; project PID2019-111300GA-I00) and the Ramón y Cajal programme via the RYC 2020-028628-I fellowship. The authors are grateful to W. Brennessel and J. Jenkins (University of Rochester) for assistance with crystallographic analyses. MS and X-ray instrumentation at the University of Rochester are supported by the US National Science Foundation (grant nos. CHE-0946653 and CHE-1725028) and the US National Institutes of Health (grant no. S10OD030302).

Author contributions

D.A.V., X.R., S.R. and R.F. conceived the project, designed the experiments and analysed all the experiments with guidance from R.F. K.N.H. and M.G.-B. mentored L.Z. and A.S. for MD and quantum mechanics calculations, and contributed to the writing of the mechanistic parts of the paper. D.A.V. and R.F. wrote the paper with input from all of the authors. All authors discussed the results and contributed to the final paper.

Competing interests

The authors declare no competing interests.

Additional information

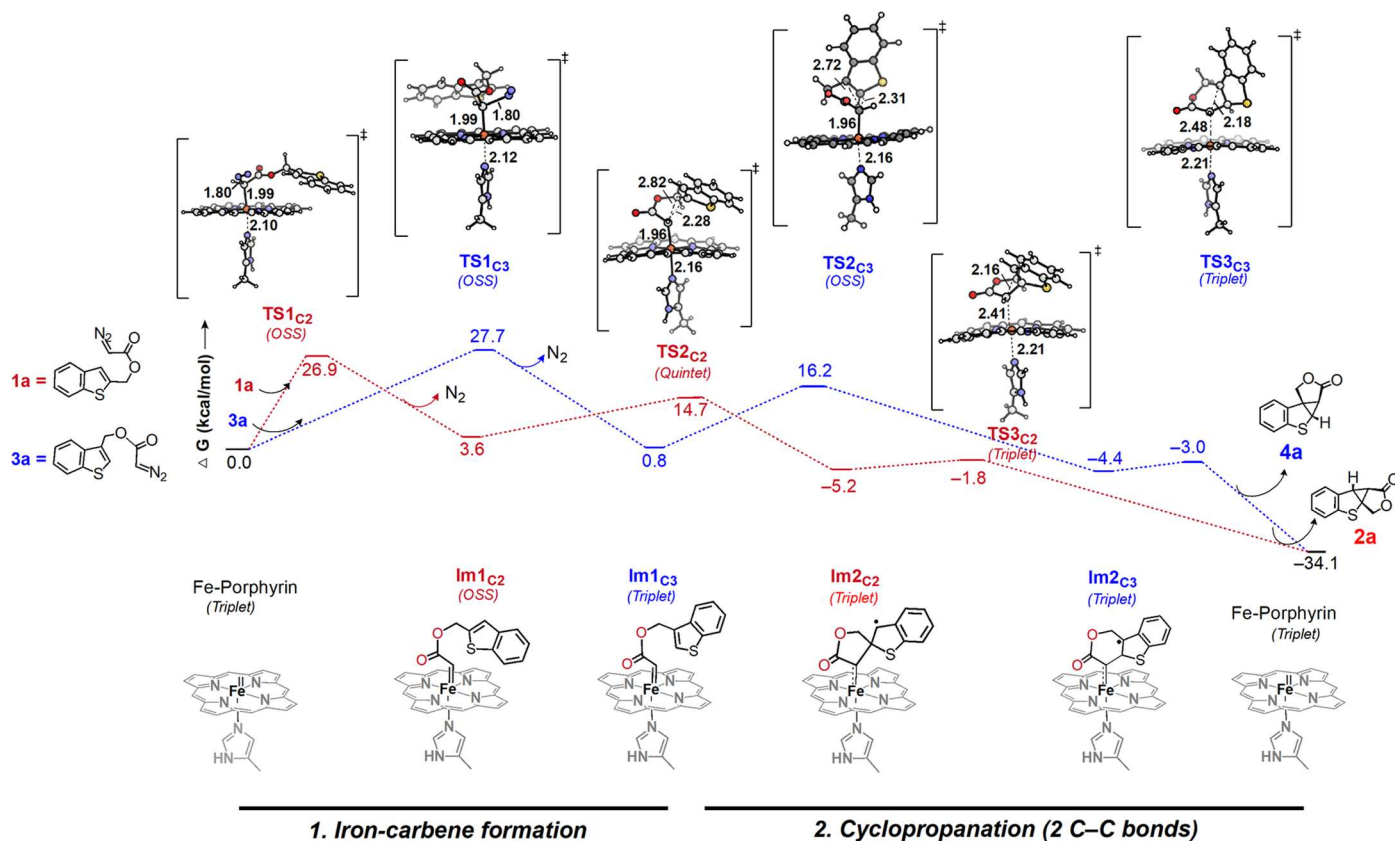
Extended data is available for this paper at <https://doi.org/10.1038/s41557-023-01435-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41557-023-01435-3>.

Correspondence and requests for materials should be addressed to K. N. Houk or Rudi Fasan.

Peer review information *Nature Chemistry* thanks Adrian Mulholland and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Calculated reaction pathways for C2- and C3-benzothienyl diazo ester substrates. DFT analysis of the reaction mechanism for intramolecular cyclopropanation of C2- and C3-benzothiophenyl-methyl-diazoacetate catalyzed by a truncated iron porphyrin with an axial 4-methylimidazole ligand as a simplified model called **Fe-Porphyrin**. The reaction proceeds via a (1) iron-carbene formation

followed by (2) cyclopropanation (2 C–C bonds). ΔG values are calculated at the B3LYP-D3BJ/def2TZVP (SMD, $\epsilon=4$) // B3LYP-D3BJ/6-31G(d)+SDD (Fe) level. For each stationary point, the Gibbs free energy is provided for its lowest energy spin state. Detailed free energy profiles and additional data are provided in Supplementary Fig. 8 and Supplementary Table 5 in Supporting Information.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

- NMR data were acquired on a Bruker DPX-500 (500 MHz) and Bruker DPX-400 (400 MHz) spectrometers running Topspin X version 1.3.10
- GCMS data were acquired using GCSolution version 2.50 SU3 (Shimadzu)
- GC-FID data were acquired using GCSolution version 231.00 (Shimadzu)
- UV-VIS absorption data were acquired using UVProbe version 2.10 (Shimadzu)
- Molecular Dynamics simulations were performed using the GPU code (pmemd) of the AMBER 18 package. All Density Functional Theory (DFT) calculations were carried out using Gaussian16, Revision C.01 software.
- Density Functional Theory (DFT) calculations were carried out using Gaussian16.
- Protein crystallography was acquired remotely at Stanford Synchrotron Radiation Lightsource (SSRL) and an in-house source a Rigaku XtaLAB Synergy-S microfocus X-ray diffraction system.
- Detailed information regarding all data collection are provided in the Supplementary Information.

Data analysis

- NMR spectra were analyzed using MNova version 14.2.1-27684
- GC-MS data were analyzed using GCsolution version 2.50 SU3
- GC-FID data were analyzed using GCsolution version 231.00 (Shimadzu)
- UV-VIS absorption data were acquired using UVProbe version 2.10 (Shimadzu)
- Molecular Dynamics trajectories were analyzed using the set of programs and scripts included in the AmberTools module from Amber18 package.
- Protein crystallographic data were analyzed using XDS, CCP4 programs AIMLESS and POINTLESS via an autoXDS script and CrysAlisPro (version 171.41).
- Molecular replacement was done using Phaser (2.8.3) and model building and refinement was done using Phenix (1.19.2), and Coot (0.8.9).
- Detailed information regarding all data collection is provided in the Supporting Information.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Protein crystal structure coordinates were deposited with the Protein Data Bank under the accession number 7SLH (MbBTIC-C3) and 7SLI (MbBTIC-C2).

Crystallographic data for the structures reported in this Article have been deposited at the Cambridge Crystallographic Data Centre, under deposition numbers CCDC 2157009 (2a), 2157011 (2d), 2157007 (4a), 2157010 (4f), 2157008 (4k). Copies of the data can be obtained free of charge via <https://www.ccdc.cam.ac.uk/structures/>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Enzymatic reactions found positive in the primary screen were re-tested at least in duplicate from independent experiments with different batches of enzyme. Assay yields were typically within 10% deviation between independent runs, and enantioselectivity was almost identical in duplicate runs (within 1% deviation). All presented data are representative results for at least two experiments that were performed independently on different days.
Data exclusions	No data collected in this data were excluded.
Replication	All samples were measured as technical replicates ($n \geq 2$) and are representative results for at least two experiments that were performed independently on different days and that successfully replicated the presented results.
Randomization	Randomization was not applicable to this study. All reagents and catalysts were selected and the reaction conditions were carefully designed.
Blinding	Blinding was not applicable to this study. All experimental data were acquired using automated equipment and analyzed using computational software, eliminating human error.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging