Title

Biocatalytic strategy for the construction of sp3-rich polycyclic compounds from directed evolution and computational modelling.

Journal

Nature Chemistry, 16(5)

Authors

Vargas, David
Ren, Xinkun
Sengupta, Arkajyoti
et al.

Publication Date

2024-05-01

DOI

10.1038/s41557-023-01435-3

Peer reviewed

HHS Public Access

Author manuscript

Nat Chem. Author manuscript; available in PMC 2024 May 12.

Published in final edited form as:

Nat Chem. 2024 May ; 16(5): 817–826. doi:10.1038/s41557-023-01435-3.

Biocatalytic Strategy for Construction of *sp*³-Rich Polycyclic Compounds from Directed Evolution and Computational Modeling

David A. Vargas[1,‡], **Xinkun Ren**[2,‡], **Arkajyoti Sengupta**[3,‡], **Ledong Zhu**[4], **Satyajit Roy**[5], **Marc Garcia-Borràs**[6], **K. N. Houk**[3], **Rudi Fasan**[5]

[1]Process Research and Development, Merck & Co., Inc., Rahway, NJ, 07065, United States

[2]College of Engineering and Applied Sciences, Nanjing University, Nanjing, Jiangsu Province 210023, China

[3]Department of Chemistry and Biochemistry, University of California, Los Angeles, California, 90095, United States

[4]Environment Research Institute, Shandong University, Qingdao 266237, PR China

[5]Department of Chemistry and Biochemistry, University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, United States

[6]Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, 17003 Girona, Spain

Abstract

Catalysis with engineered enzymes has provided more efficient routes for the production of active pharmaceutical agents. However, the potential of biocatalysis toward assisting early-stage drug discovery campaigns remains largely untapped. Here we develop a biocatalytic strategy for the construction of sp^3-rich polycyclic compounds via an intramolecular cyclopropanation of benzothiophenes and related heterocycles. Two carbene transferases with complementary regioisomer selectivity were evolved to catalyze the stereoselective cyclization of benzothienyl substrates bearing diazo ester groups at either the C2 or C3 position of the heterocycle. The detailed mechanisms of these reactions were obtained 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 with new-to-nature transformations. The molecular scaffolds accessed here feature a combination of 3D and

Corresponding authors: R.F. (rudi.fasan@utdallas.edu) and K.N.H. (houk@chem.ucla.edu).

‡These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

D.V., X.R., S.R., and R.F. conceived the project, designed the experiments, and analyzed all the experiments with guidance from R.F. K.N.H. and M.G. mentored P.M., A.S. for molecular dynamics and quantum mechanics calculations and contributed to the writing of the mechanistic parts of the manuscript. D.V. and R.F wrote the manuscript with input from all of the authors. All authors discussed the results and contributed to the final manuscript.

COMPETING INTERESTS

The authors have no competing interests.

stereochemical complexity with ‘rule-of-3’ properties, which should make them highly valuable for fragment-based drug discovery campaigns.

Introduction

Biocatalysis with engineered enzymes has been covering an increasing important role toward enabling and streamlining the stereoselective synthesis of drug molecules and other high-value compounds.[1–6] In addition, the reaction scope of biocatalysis has been recently 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, 3D ‘fragments’ for fragment-based drug discovery (FBDD) campaigns (Figure 1a–b).[11] Over the past two decades, FBDD has represented a key strategy for drug discovery, yielding several candidates entering clinical trials, some of which have been approved for commercialization.[12] This drug discovery approach relies on the availability of libraries of small and diverse organic molecules (‘fragments’) which are screened to identify weak binders of a protein of interest and then linked to generate potent inhibitors of such targets.[12,13] In terms of physico-chemical properties, best suited fragments for FBDD applications are molecules that adhere to the ‘rule of three’ (Ro3), corresponding to a molecular weight below 300 Da, less than 3 rotatable bonds, and an n-octanol/water partition coefficient (cLogP) lower than 3.[14] Although several readily available molecules meet these requirements, medicinal chemists and chemical biologists have highlighted the need for more unique, stereochemically rich, and three-dimensional ‘fragments’ to expand opportunities for drug discovery via FBDD.[15–18] Motivated by this context, we have developed a biocatalytic strategy to access a series of previously inaccessible sp^3-rich sulfur-containing polycyclic scaffolds which exhibit these highly sought-after features and could thus serve as potentially valuable additions to fragment libraries for FBDD applications.

Heme-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 useful for this type of reactions have also been reported.[29–36] More recently, the scope of these strategies has been extended to include the first examples of enzyme-catalyzed intramolecular cyclopropanations.[37,38] Here we report a biocatalytic methodology based on engineered myoglobin variants for the asymmetric intramolecular cyclopropanation of diazoester-functionalized benzothiophenes. This reaction has no chemocatalytic precedent and attempts to realize similar transformations using organometallic catalysts were met with limited success.[39,40] Upon 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, three-dimensional scaffolds (Fsp^3: 0.36, 75% of which are stereogenic centers) with physical-chemical properties adhering to the rule of three (Avg. MW: 230 ± 31 Da, Avg. Rot. Bonds: 0.1 ± 0.3, Avg cLogP: 2.5 ± 0.5, Avg Polar Surf. Area: 28.4 ± 3.1; Figure 1c). Crystallographic analysis of the engineered enzymes in combination with computational studies based on DFT calculations and MD simulations, provide insights into the mechanism and origin of protein-mediated stereocontrol in these reactions, which

were further leveraged to enhance the activity and expand the substrate scope of these biocatalysts.

Results and Discussion

Biocatalyst Evolution for C2-functionalized Benzothiophene.

In initial studies, a diverse set of heme-containing enzymes and proteins was screened, including wild-type myoglobin (Mb), its distal histidine variant Mb(H64V), $P450_{BM3}$, CYP119, P411-CHF,[43] different cytochrome *c* proteins, and others, 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 (Figure 2a and Supplementary Table 1). Of note, 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(TPP)(Cl), and Co(TPP) (Figure 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 heme cofactors. While the large majority of these variant showed no activity, Mb(H64F) displayed basal activity for formation of the desired product **2a** (2% yield) with modest enantioselectivity (61% *ee*; Figure 2c) under reducing conditions. Mb(H64F) was thus selected as the parent scaffold for iterative rounds of directed evolution, in which active site residues most proximal to the iron center, i.e. 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 as purified protein in the reaction with **1a**.

Through this process, five beneficial mutations were accumulated resulting in the quintuple mutant Mb(F43I,F46L,H64F,V68G,I107A), called $Mb_{BTIC\text{-}C2}$, with significantly increased activity for generation of **2a** from **1a** (2%→75% yield). In addition, $Mb_{BTIC\text{-}C2}$ displays excellent enantioselectivity (>99% *ee*) producing a single stereoisomer whose configuration was determined to be 3a*S*,3b*S*,8a*R* via single-crystal X-ray diffraction (Figure 3c). Interestingly, structure-activity analyses of the $Mb_{BTIC\text{-}C2}$ lineage revealed a clear synergistic effect of mutations F43I and F46L on the catalytic activity (TON) of the enzyme (Figure 2c), without affecting enantioselectivity. These analyses also indicated a distinct benefit of a His→Phe mutation at the level of the distal histidine (His64) and large-to-small mutations at multiple sites within the heme pocket (Phe43→Ile/Leu; Phe46→Ile/Leu; Ile107→Ala) to favor the target reaction. The latter findings are generally consistent with the high steric demands inherent to mediating an intramolecular cyclopropanation reaction with the confines of myoglobin heme pocket.[37]

Cyclopropanation of C3-functionalized Benzothiophene.

Encouraged by the results above, we sought to apply the evolved $Mb_{BTIC\text{-}C2}$ biocatalyst to the cyclopropanation of C3-functionalized benzothiophene (**3a**), as we envisioned this reaction would provide access to an alternative tetracyclic scaffold (Figure 1c). Surprisingly, no detectable amount of the desired product **4a** was obtained, clearly indicating the need for

very different active site configurations for enabling the cyclization of **3a** vs. its regioisomer **1a** (Figure 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 S2). Unlike for **1a**, V68G (and other single mutations at position 68) had no beneficial effect toward transformation of **3a**, further highlighting the divergent catalyst requirement for the two reactions (Figure 2d). In contrast, Mb(H64V,V68A) yielded a viable biocatalyst for the intramolecular cyclopropanation of **3a**, producing **4a** in low yield (23% yield, Figure 2d) but with excellent enantioselectivity (>99% *ee*) toward formation of the 3a*S*,3b*S*,8b*S* enantiomer as determined by X-ray difraction (Figure 1c). Building upon these results, a protein engineering campaign was undertaken that resulted in the identification of Mb(L29F,H64V,V68A,I107L), named $Mb_{BTIC\text{-}C3}$, as an improved biocatalyst for this reaction, furnishing 3-fold improved yield of **4a** (23%→60% yield) while retaining excellent enantioselectivity (>99% *ee*) (Figure 2d). After optimization of the $Mb_{BTIC\text{-}C3}$-catalyzed reaction (i.e., 4°C, slow addition of **3a**, Supplementary Table 3), the cyclopropanation product **4a** could be obtained in quantitative yields and enantiopure form (>99% *ee*) (Figure 2d), with the improvement in cyclopropanation yield resulting primarily from a reduction of carbene dimerization side reaction. Under catalyst limited conditions, $Mb_{BTIC\text{-}C3}$ was found to support up to 440 TON, producing **4** in 87% yield using only 0.2 mol% catalyst (Supplementary Table 2). Moreover, both $Mb_{BTIC\text{-}C2}$ and $Mb_{BTIC\text{-}C3}$ could be applied in whole cells (OD_{600}:20 and OD_{600}:60, respectively) for the stereoselective synthesis of **4a** in 62% yield and of **2a** in 65% yield, respectively, demonstrating the compatibility of these enzymatic reactions with whole-cell biotransformations.

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

Substrate Scope of the Evolved Biocatalysts.

Next, we investigated the substrate scope of the $Mb_{BTIC\text{-}C2}$ biocatalyst using variously substituted benzothienyl molecules (Figure 3a). These experiments showed that the enzyme can tolerate a range of electron-withdrawing and electron-donating groups on position C5 delivering the desired cyclopropyl-lactone products (**2b-f**) in 36–99% yields and with excellent enantioselectivity (>99% *ee*) (Figure 3a). Similar results were obtained for a series of 6-substituted substrates indicating a significant tolerance of the biocatalyst to variation at the C6 position to furnish **2g-j** in 25–44% yields and high enantiopurity (97–99% *ee*, Figure 3a). Improved yields for the synthesis of **2g** and **2j,** respectively, could be achieved using whole cells expressing the $Mb_{BTIC\text{-}C2}$ variant (51–55% vs. 35–39%). The 7-fluoro-substituted substrate **1k** was also processed by $Mb_{BTIC\text{-}C2}$ to give **2k** in

good yield and high enantiopurity (>99% *ee*). In contrast, substitution at the C4 site (e.g., -Br) were not tolerated, resulting in unreacted starting material. Interestingly, only carbene dimerization was observed when the benzothiophene ring was changed to benzofuran in substrate **5** (Supplementary Figure 1). A diazoketone analog of **1a** was also prepared and tested (compound **6**; Supplementary Figure 1) but it failed to cyclize using $Mb_{BTIC\text{-}C2}$ or related variants.

Using a similar approach, we investigated the scope of the $Mb_{BTIC\text{-}C3}$-catalyzed reaction using a series of C3-functionalized benzothiophenes bearing substitutions in positions C5, C6, and C7 (Figure 3b). Compared to $Mb_{BTIC\text{-}C2}$, these experiments revealed a significantly different trend for $Mb_{BTIC\text{-}C3}$ in terms of the positional effect of substitutions on the enzyme activity. Specifically, substitutions on position C7 were well tolerated by $Mb_{BTIC\text{-}C3}$ producing **4g-i** with high efficiency (75%–85% yields; Figure 3b). For **3g** and **3i,** nearly quantitative yields (96%) could be achieved by using a higher catalyst loading (1.6 vs. 0.8 mol%). In contrast, only limited substitutions (e.g., -F, $-CH_3$) at position C5 and C6 of the substrate were accepted by the enzyme, with the higher yield of **4f** vs. **4b** (62% vs. 12%) suggesting a larger tolerance toward C6 vs. C5 substitution on the benzothiophene ring. Of note, all of the $Mb_{BTIC\text{-}C3}$-catalyzed reactions were found to proceed with excellent enantiocontrol, furnishing the desired cyclopropanation products in 97–99% *ee* (Figure 3b). Notably, the biocatalyst was found to withstand methyl substitution at the alpha position of the ester linkage, affording a polycyclic product containing as many as four new stereogenetic centers with high level of diastereoselectivity (**4l**; Figure 3b). The scope of the biocatalyst and reaction was further explored across different types of heterocycles (i.e., benzofurane, indole, furan) and types of carbene precursors such as diazoamides and diazoketones. Among them, the thiophenyl-based substrate **3j** and benzofurane-based substrate **3k** could be successfully cyclopropanated by $Mb_{BTIC\text{-}C3}$ to afford **4j** and **4k** in quantitative yields and 50–99% *ee* (Figure 3b). Diazoketone substrate **3m** could be also cyclized to obtain **4m** in 42% yield and high enantioselectivity (98% *ee*; Figure 3b). In contrast, benzofuran- and indole-based diazoamides as well as furan-based diazoesters (compounds **7**–**11**, Supplementary Figure 1) proved to be unviable substrates for cyclization by $Mb_{BTIC\text{-}C3}$ or its precursor variants, indicating that further enzyme engineering would be required to enable these transformations. These limitations notwistanding, these studies demonstrated that the biocatalytic transformation can be extended to thiophene- and benzofurane-based rings as well as benzothiophenyl diazoketones. Importantly, the molecular scaffolds made accessible through these $Mb_{BTIC\text{-}C2}$- and $Mb_{BTIC\text{-}C3}$-catalyzed reactions combine several highly sought-after features for FBDD,[15,16] namely 3D shape, stereochemical complexity (3 stereogenic centers), and Ro3-compliant physico-chemical properties (Figure 1c).

Mechanistic Studies.—Hemoprotein-catalyzed cyclopropanation reactions using diazo compounds involve the formation of a reactive iron porphyrin carbene (IPC) intermediate, which can engage the olefin to yield the cyclopropanation product via concerted or stepwise-diradical pathways.[35,44] To gain insight into the intrinsic mechanism of the reactions investigated here, these transformations were studied using density functional theory (DFT) with a truncated model of the ferrous hemoprotein (Fe(II) porphyrin with an

axial 4-methylimidazole ligand; see SI for further details) to facilitate computations. The reaction first involves the loss of N_2 and formation of a Fe-carbene IPC intermediate **Im1$_{C2}$** (open-shell singlet, ΔG = 3.6 kcal/mol) or **Im1$_{C3}$** (triplet, ΔG = 0.8 kcal/mol) from **1a** and **3a** respectively (Figures 4 and S2). The lowest Gibbs activation energy barriers for this step involve an open-shell singlet (OSS) transition state, **TS1$_{C2}$** (26.9 kcal/mol) and **TS1$_{C3}$** (27.7 kcal/mol) (Figures 4 and S8).

In the second step of the reaction, the reactive Fe-carbene (**Im1$_{C2}$** or **Im1$_{C3}$**) attacks C2, alpha to the sulfur atom, via **TS2$_{C2}$** (quintet, $\Delta G^{‡}$ = 14.7 kcal/mol), or C3 via **TS2$_{C3}$** (open-shell singlet, $\Delta G^{‡}$ = 16.2 kcal/mol), forming 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 Figure 1 and S2). These important geometric differences between optimized intermediates **Im2$_{C2}$** and **Im2$_{C3}$** also imply dramatically different geometric requirements for the benzothiophene ring when approaching to the carbene center during the cyclization **TS2$_{C2}$** and **TS2$_{C3}$** transition states (Extended Data Figure 1). This is in line with the experimentally observed divergent catalyst requirements for C2- vs. C3-linked substrates. Both intermediates (**Im2$_{C2}$** and **Im2$_{C3}$**, lowest in energy triplet state) undergo a very fast second C–C bond formation to generate the benzothienyl cyclopropyl products **2a** or **4a** with barriers of less than 2–4 kcal/mol (**TS3$_{C2}$**, triplet $\Delta G^{‡}$= 3.4 kcal/mol; **TS3$_{C3}$**, triplet $\Delta G^{‡}$= 1.4 kcal/mol). These studies imply the occurrence of spin-crossing events along the radical reaction pathway, leading to a multi-state mechanism.[45,46]

This proposed stepwise-diradical mechanism differs from the concerted mechanism–mediated by a closed-shell singlet IPC intermediate–previously established for the Mb-catalyzed *intermolecular* cyclopropanation of vinylarenes with diazo esters.[44] The concerted mechanism in the present intramolecular cyclopropanation is found to have slightly higher barrier (16.5 kcal/mol; Supplementary Figure 8). This difference can be attributed to the electronic properties of the thiophene and aromatic rings that stabilize radical formation at the benzylic position. However, it should be noted that our cluster model calculations only include the nearby active site residues. The consideration of the entire enzymatic scaffold in calculations could lead to changes in geometric or electrostatics features, and might favor the alternative concerted mechanism. Nonetheless, both the step-wise and concerted mechanisms highlight the multi-state reactivity of hemoproteins in carbene transfer reactions.[35]

With respect to substituted benzothiophene derivatives, the calculated activation barriers for the transition states, **TS1$_{C3}$**, of C5-, C6-, and C7-functionalized benzothiophenes are similar ($\Delta G^{‡}$ ~ 27 kcal/mol) with the Br substituent. Slightly higher activation barriers ($\Delta G^{‡}$ ~ 30 kcal/mol) for Me and Cl substituents are predicted (Supplementary Table 4). For all substituents at C5-, C6-, and C7-, the calculated activation barriers are within 1.6 kcal/mol, indicating the intrinsic substrate reactivity for this reaction step is independent of the position of substituents.

The activity of the Mb catalysts vis-à-vis the lack of reactivity for hemin in these reactions (Supplementary Tables 1–2) is also noteworthy, highlighting the beneficial effect

of the protein scaffold for catalyis. The calculated barriers for carbene formation with the benzothiophene reagents are in line with the equivalent calculated barrier of EDA reacting with Fe-porphyrin computational model (28.8 kcal/mol). Note that these barriers are with respect to the separated reactants, and hence incorporates an unfavorable entropy penalty (−TΔS) of about 12 kcal/mol regardless of the nature of the diazo substrate, which will be overcome by substrate binding to the enzyme ("entropy trap"). Since hemin can activate EDA in buffered conditions[47] and S-heterocycle are well known to bind metal centers, hemin's lack of reactivity in the current reactions may stem from catalyst inhibition and/or catalyst destruction after diazo activation, both of which are disfavored in the hemoprotein system.

Crystallographic Studies.—To gain insights into the structures of the evolved enzymes $Mb_{BTIC\text{-}C2}$ and $Mb_{BTIC\text{-}C3}$, these proteins were crystallized in their ferric aquo-complex state and their X-ray structure solved to 1.3 Å resolution. Structural alignment of these structures with that of wild type Mb (pdb 1JW8)[48] yielded root-mean square deviation (RMSD) values for the protein backbone of 0.22 Å and 0.21 Å, respectively, indicating these proteins share a very similar fold. Further inspection of the crystal structures revealed however that the two variants exhibit a significantly different active site configuration as shaped by the respective mutations (Figure 4a,b). In particular, $Mb_{BTIC\text{-}C2}$ features an enlarged heme pocket compared to wild-type Mb (volume: 335 $Å^2$ vs. 125 $Å^2$ for) as a result of multiple space-creating mutations, in particular above the inner pyrrole rings A and B of the heme cofactor (Figure 4a). The Phe residue replacing the distal His64 is found in two possible orientations of similar occupancy (60:40), acting as a gating residue at the interface between the heme cavity and the solvent (Supplementary Figure 3). In stark contrast, the active site of MbC3 is characterized by a much smaller volume compared to $Mb_{BTIC\text{-}C2}$ (237 $Å^2$ vs. 335 $Å^2$; Figure 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 (i.e., above heme ring B and A, respectively), while it creates a cavity above the solvent exposed rim of the porphyrin cofactor (Figure 4b).

Origins of Enantiocontrol in the Enzymatic Reactions.

Using the x-ray structures, quantum mechanical DFT studies using the cluster model approach[49,50] were carried out to understand the basis for the high stereoselectivity exhibited by $Mb_{BTIC\text{-}C2}$ and $Mb_{BTIC\text{-}C3}$ in the respective reactions. Starting from $Mb_{BTIC\text{-}C2}$, DFT calculations were performed to reoptimize $\mathbf{TS2_{C2}}$ (Figure 4c), i.e. the TS leading to the experimentally generated enantiomer, as well as its enantiomeric transition state ($\mathbf{TS2\text{-}enan_{C2}}$), 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.[51,52] The final MD snapshot from 500 ns MD simulations were used as starting points for DFT cluster model optimizations (Figure 4c). In the optimized $\mathbf{TS2_{C2}}$/$Mb_{BTIC\text{-}C2}$ complex, the benzothiophene ring is placed within the heme pocket, with the heteroaryl ring oriented away from the Phe64 and Val107 residues and occupying the cavity above heme ring B created by the V68G mutation (Figure 4c). In contrast, in the optimized complex with $\mathbf{TS2\text{-}enan_{C2}}$, which would lead to the opposite enantiomer, the benzothiophene ring is situated

over ring A in close proximity to residues Phe64 and Ile43, resulting in steric clashes with these residues (Figure 4c). These unfavorable interactions result in a significantly higher energy for **TS2-enan$_{C2}$** vs. **TS2$_{C2}$** ($\Delta\Delta G^{\ddagger}$ = +6.9 kcal/mol), which can explain the high enantioselectivity of the enzyme observed experimentally.

We also investigated the origins for the high enantiocontrol exerted by Mb$_{BTIC\text{-}C3}$ in the cyclization of **3a**. In the DFT optimized Mb$_{BTIC\text{-}C3}$ cluster model structure of **TS2$_{C3}$**, the benzothiophene ring is oriented above ring D of the heme and sits below Val64 (as in the final MD snapshot), pointing toward the solvent-exposed side of the cofactor (Figure 4d). This configuration is largely dictated by the presence of Phe29 residue, which occupies the inner side of the heme pocket (Figure 4b). In the optimized structure of **TS2-enan$_{C3}$**, the benzothiophene ring sterically clashes with Val64 and displaces the Phe29 ring, resulting in a ΔG value of 10.7 kcal/mol higher than that of **TS2$_{C3}$**. This large energy difference can thus rationalize the excellent enantiomeric excess (>99% *ee*) observed for the 3a*S*,3b*S*,8b*S*-configured product generated by the Mb$_{BTIC\text{-}C3}$ variant.

Molecular Dynamics Simulations on Mb$_{BTIC\text{-}C2}$ system.

We performed molecular dynamics (MD) simulations to investigate how the protein scaffold and additional amino acid residues of Mb$_{BTIC\text{-}C2}$ contribute to stabilization of the **TS2$_{C2}$** transition state (Figure 5a). For these studies, the lowest energy spin-state **TS2$_{C2}$** transition state optimized from DFT calculations was docked into the full protein structure in the presence of explicit water as solvent. Two independent 500ns MD simulations starting from these optimized TS2$_{C2}$ docked near the enzyme "gate" resulted in the exploration of binding poses with the aromatic end of the substrate to deep inside the active site pocket along the MD trajectories. Steric clashes with residues Phe64, Ile43, Val107, and Leu29 induce a rotation of the benzothiophene ring away from the solvent exposed active site "gate" in the hemoprotein. Such conformational flexibility is consistent with the different orientations found for this residue in the crystal structure and further reveals its role as a "gatekeeper" residue between the heme pocket and the solvent.

To rationalize the structure-activity trends observed for the Mb$_{BTIC\text{-}C2}$ biocatalyst (Figure 3a), equivalent **TS2**-docked MD simulations were performed using 5-methoxy- (5OMe; **2f**) and 6-methyl-benzothiophene (6Me; **2j**) 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 (**2f**), rotation of the benzothienyl-ester moiety about the Fe-porphyrin ring occurs much faster along MD trajectories than observed with substrate **1a** (Figure 5b). In addition, during the MD trajectory the molecule reorients such that the thiophene ring lies above heme ring C (*vs.* ring B for **1a**) and the polar oxygen atom of the methoxy group is pointed toward the solvent (Figure 5c). In contrast, MD simulations with the 6Me-substituted substrate (**2j**) showed the benzothiophene ring as embedded deeply in the enzyme pocket and shielded away from the solvent (Figure 5c), adopting a conformation that differs from those observed with substrate **1a** and the 5-OMe substrate. Experimentally, the enzyme activity toward these substrates follows the order **2f** (5-OMe) > **1a** > **2j** (6-Me) (Figure 3a). Since this trend correlates with the degree of exposure of the benzothiophene ring to the solvent in the corresponding

TS as determined by 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 Molecular Dynamics Simulations of $Mb_{BTIC-C3}$.

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

MD simulations were then performed using the lowest in energy $\mathbf{TS2_{C3}}$ transition state docked into the $Mb_{BTIC-C3}$ active site as starting point. Interestingly, these simulations showed how, starting from the 'BT out' conformation, the heme-bound substrate briefly samples a slightly different conformation during the MD trajectory, in which the benzothiophene ring pushes away the nearby residue Phe43 (Figure 5f). However, the substrate quickly reorients itself to a final orientation that is nearly identical to the starting point, i.e., with the benzothiophenyl ring lying over ring D of the heme (Figure 5e,f). As made evident from these studies, both the accessible conformations and the preferred orientation of the heme-bound $\mathbf{TS2_{C3}}$ in $Mb_{BTIC-C3}$ (Figure 5e,f) differs completely from those found for the heme-bound $\mathbf{TS2_{C2}}$ in $Mb_{BTIC-C2}$ (Figure 5a,b), where the benzothiophene ring is pushed above ring B of the heme (vs. ring D for **3a**) and toward the inner side of the heme pocket by the concerted action of residues Leu29 and Phe64. These features can thus explain the orthogonal reactivity of these biocatalysts toward the cyclopropanation of the C2- vs. C3-functionalized benzothiophene substrates due to different active site requirements in order to stabilize the different key transition states $\mathbf{TS2_{C2}}$ and $\mathbf{TS2_{C3}}$.

Rational design of improved biocatalyst $Mb_{BTIC-C3}^{+}$.

Leveraging the mechanistic insights from the studies above, we sought to understand and potentially overcome the substrate scope limitations of $Mb_{BTIC-C3}$, which show very limited tolerance toward substitutions at the C5 position of the benzothiophene ring (Figure 3b). MD simulations with **TS2**-bound in **MbBTIC-C3** for the inactive 5-bromo-substituted substrate (C5-Br) showed that the **TS2** explores a major orientation in the active site, in which the ester group of the substrate lies at a close distance from Leu107. This suggests that the existence of these steric clashes disfavor the transition state **TS2**. Based on these analyses, we designed a variant, called $Mb_{BTIC-C3}^{+}$, in which a L107V mutation is introduced which

reduces such steric clashes facilitating the productive binding mode for **TS2** to take place, as confirmed by MD simulations with the C5-Br/$Mb_{BTIC\text{-}C3}^{+}$ system (Figure 5g). Gratifyingly, $Mb_{BTIC\text{-}C3}^{+}$ was determined to exhibit up to 5-fold improved activity over $Mb_{BTIC\text{-}C3}$ for the synthesis of C5-substituted substrates **4b** and **4e** (Figure 3b). Furthermore, albeit in low to moderate yields (7–20%), this biocatalyst is now able to produce the 5-chloro- and 5-bromo-substituted products **4c** and **4d**, which were unattainable using $Mb_{BTIC\text{-}C3}$ (Figure 3b). While expanding the substrate scope of the present method, these results highlight the value of combining evolutionary and rational strategies toward the development of these carbene transferases.

Conclusion

We report 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 fragment-based drug discovery, owing to their 3D and stereochemical complexity combined with Ro3-compliant properties (Figure 1c). Despite the inherent challenge posed by these transformations using chemical and biological carbene transfer catalysts alike (Figure 1a; Supplementary Table 1), two complementary biocatalysts could be evolved from myoglobin to 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 myoglobin-catalyzed intermolecular cyclopropanation with diazoesters,[44] highlighting the mechanistic plasticity of this hemoprotein in carbene transfer reactions. Crystallographic and computational (MD) studies have provided additional insights into 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 computational modelling information to design and expand the substrate scope of the $Mb_{BTIC\text{-}C3}$ enzyme to enable the transformation of a series of previously unreactive substrates. These results highlight the benefits of combining evolutionary and rational strategies for the development of enzymes to catalyze synthetically useful transformations not found so far in nature.

METHODS

General Information.

All the chemicals and reagents were purchased from commercial suppliers (Sigma-Aldrich, Alfa Aesar, ACS Scientific, Acros, Ambeed, Combi-blocks) and used without any 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. ^{1}H and ^{13}C NMR spectra were measured on Bruker DPX-500 (operating at 500 MHz for ^{1}H and 125 MHz for ^{13}C) or Bruker DPX-400 (operating at 400 MHz for ^{1}H and 100 MHz for ^{13}C), ^{19}F was measured on Bruker DPX-400 (operating at 375 MHz). NMR instruments were operated using Topspin X (v. 1.3.10). Tetramethylsilane (TMS) (0 ppm) and/or $CDCl_3$ (7.26 ppm)

served as the internal standard for ^{1}HNMR, $CDCl_3$ was used as the internal standard (77.0 ppm) for ^{13}CNMR, and trifluorotoluene served as the internal standard (−63 ppm) 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.

Myoglobin (Mb) variants were cloned and expressed using pET22b(+) vectors (Novagen). The Mb gene was C-terminally fused to a 6xHis tag and it is under the control of an IPTG-inducible T7 promoter. Site-directed mutagenesis variants were prepared via 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 (Supplementary Table 13). Upon transformation of the PCR products into chemically competent *E. coli* DH5α cells, recombinant colonies were pooled, grown in 5 mL LB media containing 100 mg L^{-1} ampicillin and plasmid extracted using Qiagen mini-prep kit. 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-well or 96-well deep well plates and screened for activity as whole cells as described below. For crystallization, Mb were 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 (TB) medium (ampicillin, 100 mg L^{-1}) at 37 °C until OD_{600} reached 0.9–1.2. Cells were then induced with 0.25 mM β-d-1-thiogalactopyranoside (IPTG) and 0.3 mM δ-aminolevulinic acid. After induction, cultures were shaken at 27°C and harvested after 18–20 hours by centrifugation at 3,000 × *g* at 4 °C. After cell lysis by sonication, proteins were purified by Ni-affinity chromatography. Cell lysate was transferred to a Ni-NTA column equilibrated with Ni-NTA Lysis Buffer. The resin was washed with 50 mL of Ni-NTA Lysis Buffer and then 50 mL of Ni-NTA Wash Buffer (50 mM KPi, 250 mM, NaCl, 20 mM imidazole, pH 8.0). 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. Concentration of ferric myoglobin 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 myoglobin variants and 2.5 mM benzothienyl diazo compound. In a typical procedure, recombinant cells were grown in 24-well, 48-well plates, or 96-well plates and protein expression was performed as described above. After expression, cells were pelleted by centrifugation and transferred to an anaerobic chamber. The cell pellet in each well was resuspended in 390 μL of degassed potassium phosphate buffer (KPi, 50 mM, pH 7.0). The reactions were initiated by addition of 10 μL of a 100 mM stock solution of the diazo compound in EtOH or DMSO. The plates were covered with aluminum foil and left shaking for 3–16 hours at room temperature in the anaerobic chamber. The reactions were then analyzed following the Product Analysis procedure described below.

Enzymatic Reactions.

Analytical scale reactions (400 μL) were conducted using whole cells (final OD_{600} = 20 or 60) or purified myoglobin variant (final conc. 20 μM), 2.5 mM benzothienyl diazo compound, and 10 mM sodium dithionite ($Na_2S_2O_4$) 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 potassium phosphate buffer (KPi, 50 mM, pH 7.0) was added to the vessel to obtain the desired enzyme concentration, or cell density for whole cell reactions). To the reactions with purified proteins, protein was reduced by 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 addition of 10 μL of 100 mM stock solution of the diazo compound in EtOH or DMSO and shaken under magnetic agitation for 3–16 hours at room temperature. For reactions involving slow substrate addition, the diazo compound (100 mM stock solution in EtOH or DMSO) was added in four aliquots over one hour. For reactions at 3°C, the reaction vessel was kept on ice. The TON for the whole-cell reactions were calculated based on Mb concentration in the reaction mixture as measured via UV-vis spectroscopy (ε_{410} = 156 mM^{-1} cm^{-1}) after cell lysis. Preparative scale reactions were performed using similar procedures but using larger volumes (25–50 mL).

Product Analysis.

The reaction mixtures were added with 20 μL of internal standard (50 mM benzodioxole in EtOH) and extracted with 400 μL of dichloromethane. The organic layer was analyzed via GC-FID using a Shimadzu GC-2010 gas chromatograph equipped with an FID detector, and a chiral Cyclosil-B column (30 m × 0.25 mm × 0.25 μm film). Separation method: 1 μL injection, injector temperature: 300°C, detector temperature: 300°C. Gradient: column temperature set at 80°C for 2 min, then to 245°C at 40°C/min with a 10 min hold. Total run time: 16.1 min. Enantiomeric excess was determined via chiral GC-FID and calibration curves for assay yield determination were constructed with authentic standards prepared using enzymatic reactions in whole cells or with purified $Mb_{BTIC\text{-}C2}$ or $Mb_{BTIC\text{-}C3}$ as described in Synthetic Procedures in the Supplementary Information. All measurements were performed at least in duplicate. For each experiment, negative control samples containing no protein were included.

Crystallization, Data Collection, Processing, and Refinement.

Crystals of ferric $Mb_{BTIC\text{-}C2}$ and $Mb_{BTIC\text{-}C3}$ complexed with water were grown at room temperature using the hanging-drop vapor diffusion method by mixing 1 μL of reservoir buffer (2.4 M ammonium sulfate, 0.2 M Tris pH 9.0, 0.1 mM EDTA (pH 8.2 for $Mb_{BTIC\text{-}C2}$) with 1 μL of 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 1:1 mix of paratone and silicone oil and moving the xtals around until the solvent was replaced before being flash-cooled in liquid nitrogen. Data were collected remotely at Stanford Synchrotron Radiation Lightsource (SSRL) and using an in-house Rigaku XtaLAB Synergy-S microfocus X-ray diffraction system. Diffraction data for $Mb_{BTIC\text{-}C3}$ were reduced using XDS, CCP4 programs AIMLESS and POINTLESS via an autoXDS script; for $Mb_{BTIC\text{-}C2}$, CrysAlisPro (version 171.41), XDS and AIMLESS were used. The $Mb_{BTIC\text{-}C3}$ structure

was solved using the Fourier synthesis method beginning with the available structure of Mb(H64V,V68A) (PDB:6M8F) and this structure was also used to solve $Mb_{BTIC-C2}$ by molecular replacement in PHENIX, followed by further rounds of refinement and modeling building using PHENIX and COOT. Data processing and refinement statistics are presented in Supplementary Tables 6 and 7. Structure coordinates were deposited with the Protein Data Bank under the accession number 7SLH and 7SLI for $Mb_{BTIC-C3}$ and $Mb_{BTIC-C2}$, respectively. For comparative purposes (Figure 2), a structure of wild type sperm whale Mb (ferrous CO-complex) crystallized in the same space group (P6) was chosen in consideration of subtle conformational differences of Mb observed under different crystalline environments (PDB:1JW8). $Mb_{BTIC-C2}$ structure presents two alternate orientations of the distal phenylalanine residue (Phe64), both of which are shown in Supplementary Figure 3.

QM calculations.

All geometry optimizations, transition structure (TS) searches and single-point computations were carried out with the Gaussian 16 package. The periphery of the porphyrin moiety were replaced by H atoms. Additionally, the H93 was modeled by a 5-Me-imidazole. The geometry optimizations were performed with the unrestricted version of the hybrid B3LYP functional with D3 dispersion corrections. Stuttgart basis set and the associated relativistic pseudo potential were 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 def2TZVP basis set was used on all atoms to calculate single-point energies. To mimic the protein environment, single point energy calculations were performed Truhlar's SMD model was applied for solvent diethylether having dielectric constant, ε=4.33. 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. Different spin states for Fe(III) were considered (Supplementary Figure 8). The calculated Gibbs free energies in the main text are obtained after applying corrections using quasi-harmonic approximations using Goodvibes.[3] For the theozyme calculations, the input structure was used from the 500ns MD snapshot. The transition states $\mathbf{TS2_{C3}}$ and $\mathbf{TS2_{C2}}$ bound in the enzyme were reoptimized with selected amino acid residues with the backbone carbon atoms and atoms in amide bonds frozen in geometry optimizations.

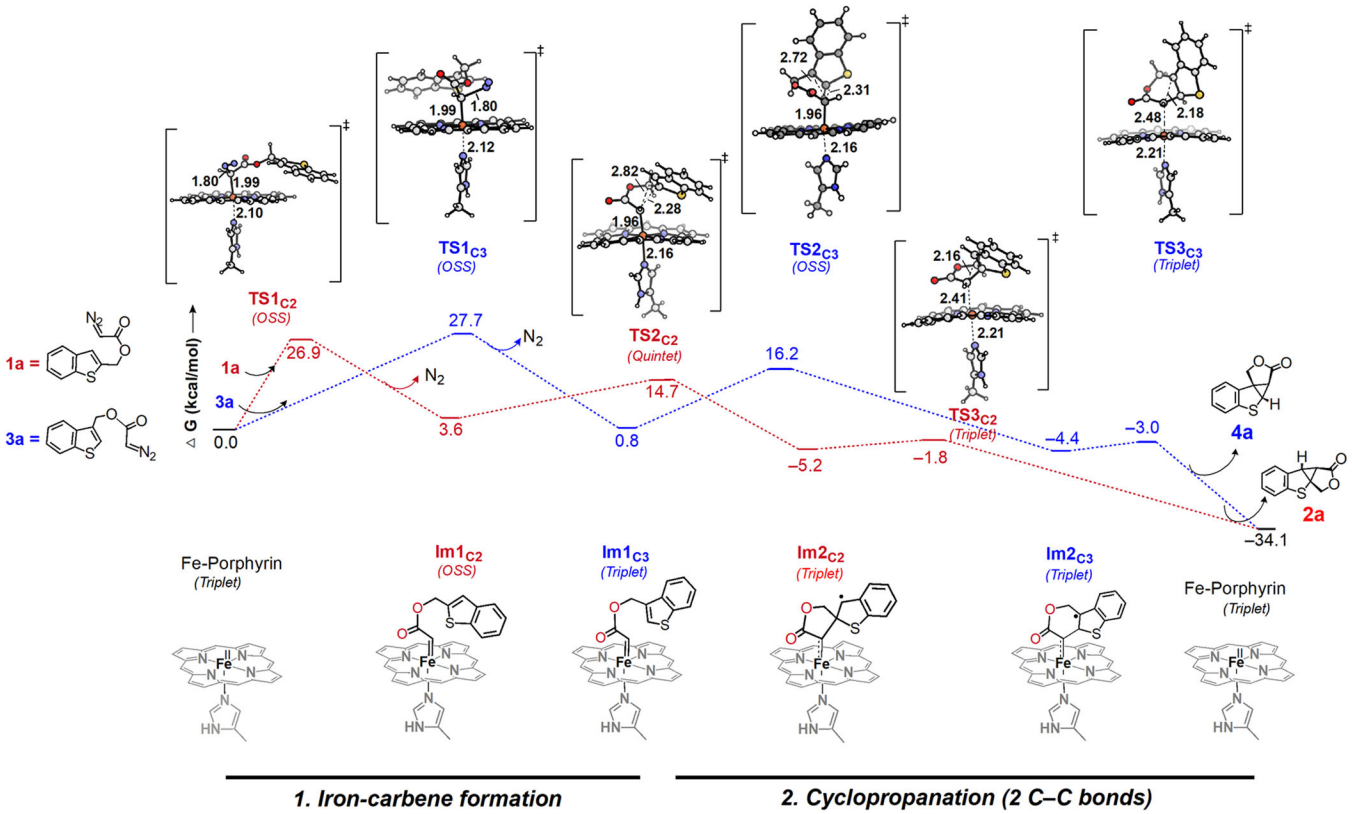
MD simulations.

$\mathbf{TS2_{C3}}$ and $\mathbf{TS2_{C2}}$ are important in the Mb catalyzed cyclopropanation of C3- and C2-benzothiophene-yl-diazoacetate to study the rate-determining step and enantioselectivity in the reaction. MD simulations on these species and on substituent analogues of $\mathbf{TS2_{C3}}$ and $\mathbf{TS2_{C2}}$ were performed to understand the catalytic activity of the enzyme. Simulations were performed using the GPU code (pmemd) of the AMBER 18 package. Parameters for the substrates were generated within the antechamber module using the general AMBER force field (gaff), with partial charges set to fit the electrostatic potential generated at the HF/6–31G(d) level by the RESP model. The charges were calculated according to the Merz–Singh–Kollman scheme using the Gaussian calculations. Since we are interested in the effects of amino acids on the catalytic activity, we have restrained the C–C distance and the

torsion angle $C_3C_2C_8O$ throughout the MD simulations as that 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 TIP3P water molecules using the leap module, resulting in the addition of around 6727 solvent molecules. The systems were neutralized by addition of explicit counter ions (Na^+ and Cl^-). All subsequent calculations were done using the widely used Stony Brook modification of the Amber 14 force field (ff14sb).

A four-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 10 kcal mol^{-1} Å^{-2}. The second, third, and fourth stage allows to relax the transition state bound to enzyme slowly with restrain values 5.0, 2.0, and 0.5 kcal mol^{-1} Å^{-2}. 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 modeled using the particle-mesh-Ewald method. A 10 Å cutoff 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 restraints 10, 5.0, 2.5, and 1.0 kcal mol^{-1} Å^{-2} on the enzyme-substrate complex with 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 1.5.7. Substrates were docked into the active site of Mb(H64V,V68A) and Mb$_{BTIC\text{-}C3}$ using AutoDock with a grid cube of length 50 Å and with Fe atom as the center of cube.

Extended Data



Extended Data Figure 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, ε=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 Figure 8 and Supplementary Table 5 in Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the U.S. National Institute of Health grant 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.V. acknowledges support from the National Science Foundation Graduate Fellowship Program. K. N.H. and A.S. acknowledges 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 Ramón y Cajal program via the RYC 2020–028628-I fellowship. The authors are grateful to Drs. William Brennessel and Jermaine Jenkins (University of Rochester) for assistance with crystallographic

analyses. MS and X-ray instrumentation at the University of Rochester are supported by U.S. National Science Foundation grants CHE-0946653 and CHE-1725028 and the U.S. National Institute of Health grant S10OD030302.

DATA AVAILABILITY

Protein crystal structure coordinates were deposited with the Protein Data Bank under the accession number 7SLH ($Mb_{BTIC\text{-}C3}$) and 7SLI ($Mb_{BTIC\text{-}C2}$). 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/.

REFERENCES

1. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC & Robins K Engineering the third wave of biocatalysis. Nature 485, 185–194, (2012). [PubMed: 22575958]
2. Devine PN, Howard RM, Kumar R, Thompson MP, Truppo MD & Turner NJ Extending the application of biocatalysis to meet the challenges of drug development. Nat. Rev. Chem. 2, 409–421, (2018).
3. Slagman S & Fessner WD Biocatalytic routes to anti-viral agents and their synthetic intermediates. Chem. Soc. Rev. 50, 1968–2009, (2021). [PubMed: 33325938]
4. Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, Devine PN, Huisman GW & Hughes GJ Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. Science 329, 305–309, (2010). [PubMed: 20558668]
5. Huffman MA, Fryszkowska A, Alvizo O, Borra-Garske M, Campos KR, Canada KA, Devine PN, Duan D, Forstater JH, Grosser ST, Halsey HM, Hughes GJ, Jo J, Joyce LA, Kolev JN, Liang J, Maloney KM, Mann BF, Marshall NM, McLaughlin M, Moore JC, Murphy GS, Nawrat CC, Nazor J, Novick S, Patel NR, Rodriguez-Granillo A, Robaire SA, Sherer EC, Truppo MD, Whittaker AM, Verma D, Xiao L, Xu Y & Yang H Design of an in vitro biocatalytic cascade for the manufacture of islatravir. Science 366, 1255–1259, (2019). [PubMed: 31806816]
6. McIntosh JA, Benkovics T, Silverman SM, Huffman MA, Kong J, Maligres PE, Itoh T, Yang H, Verma D, Pan WL, Ho HI, Vroom J, Knight AM, Hurtak JA, Klapars A, Fryszkowska A, Morris WJ, Strotman NA, Murphy GS, Maloney KM & Fier PS Engineered Ribosyl-1-Kinase Enables Concise Synthesis of Molnupiravir, an Antiviral for COVID-19. ACS Centr. Sci. 7, 1980–1985, (2021).
7. Brandenberg OF, Fasan R & Arnold FH Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. Curr. Opin. Biotech. 47, 102–111, (2017). [PubMed: 28711855]
8. Schwizer F, Okamoto Y, Heinisch T, Gu YF, Pellizzoni MM, Lebrun V, Reuter R, Kohler V, Lewis JC & Ward TR Artificial Metalloenzymes: Reaction Scope and Optimization Strategies. Chem. Rev. 118, 142–231, (2018). [PubMed: 28714313]
9. Ren X & Fasan R Engineered and Artificial Metalloenzymes for Selective C-H Functionalization. Curr. Opin. Green Sustain. Chem. 31, 100494, (2021). [PubMed: 34395950]
10. Liu Z & Arnold FH New-to-nature chemistry from old protein machinery: carbene and nitrene transferases. Curr. Opin. Biotech. 69, 43–51, (2021). [PubMed: 33370622]
11. Ramsden JI, Cosgrove SC & Turner NJ Is it time for biocatalysis in fragment-based drug discovery? Chem. Sci. 11, 11104–11112, (2020). [PubMed: 34094353]
12. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W & Jhoti H Twenty years on: the impact of fragments on drug discovery. Nat. Rev. Drug Discov. 15, 605–619, (2016). [PubMed: 27417849]
13. Erlanson DA, Davis BJ & Jahnke W Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. Cell Chem. Biol. 26, 9–15, (2019). [PubMed: 30482678]

14. Jhoti H, Williams G, Rees DC & Murray CW The ‘rule of three’ for fragment-based drug discovery: where are we now? Nat. Rev. Drug Discov. 12, 644–+, (2013). [PubMed: 23845999]
15. Murray CW & Rees DC Opportunity Knocks: Organic Chemistry for Fragment-Based Drug Discovery (FBDD). Angew. Chem. Int. Ed. 55, 488–492, (2016).
16. Morley AD, Pugliese A, Birchall K, Bower J, Brennan P, Brown N, Chapman T, Drysdale M, Gilbert IH, Hoelder S, Jordan A, Ley SV, Merritt A, Miller D, Swarbrick ME & Wyatt PG Fragment-based hit identification: thinking in 3D. Drug Discovery Today 18, 1221–1227, (2013). [PubMed: 23906694]
17. Over B, Wetzel S, Grutter C, Nakai Y, Renner S, Rauh D & Waldmann H Natural-product-derived fragments for fragment-based ligand discovery. Nat. Chem. 5, 21–28, (2013). [PubMed: 23247173]
18. Yu B, Zheng YC, Shi XJ, Qi PP & Liu HM Natural Product-Derived Spirooxindole Fragments Serve as Privileged Substructures for Discovery of New Anticancer Agents. Anti-Cancer Agents in Medicinal Chemistry 16, 1315–1324, (2016). [PubMed: 26522954]
19. Coelho PS, Brustad EM, Kannan A & Arnold FH Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. Science 339, 307–310, (2013). [PubMed: 23258409]
20. 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).
21. 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). [PubMed: 28366001]
22. Key HM, Dydio P, Liu ZN, Rha JYE, Nazarenko A, Seyedkazemi V, Cark DS & Hartwig JF Beyond Iron: Iridium-Containing P450 Enzymes for Selective Cyclopropanations of Structurally Diverse Alkenes. ACS Centr. Sci. 3, 302–308, (2017).
23. Chandgude AL & 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).
24. Knight AM, Kan SBJ, Lewis RD, Brandenberg OF, Chen K & Arnold FH Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. ACS Cent. Sci. 4, 372–377, (2018). [PubMed: 29632883]
25. Wittmann BJ, Knight AM, Hofstra JL, Reisman SE, Kan SBJ & Arnold FH Diversity-Oriented Enzymatic Synthesis of Cyclopropane Building Blocks. ACS Catal. 10, 7112–7116, (2020). [PubMed: 33282460]
26. Nam D, Steck V, Potenzino RJ & 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). [PubMed: 33497207]
27. Vargas DA, Tinoco A, Tyagi V & Fasan R Myoglobin-Catalyzed C-H Functionalization of Unprotected Indoles. Angew. Chem. Int. Ed. 57, 9911–9915, (2018).
28. Brandenberg OF, Chen K & Arnold FH Directed Evolution of a Cytochrome P450 Carbene Transferase for Selective Functionalization of Cyclic Compounds. J. Am. Chem. Soc. 141, 8989–8995, (2019). [PubMed: 31070908]
29. Srivastava P, Yang H, Ellis-Guardiola K & Lewis JC Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. Nat. Commun. 6, 7789, (2015). [PubMed: 26206238]
30. Sreenilayam G, Moore EJ, Steck V & Fasan R Metal substitution modulates the reactivity and extends the reaction scope of myoglobin carbene transfer catalysts. Adv. Synth. Cat. 359, 2076–2089, (2017).
31. Dydio P, Key HM, Nazarenko A, Rha JY, Seyedkazemi V, Clark DS & Hartwig JF An artificial metalloenzyme with the kinetics of native enzymes. Science 354, 102–106, (2016). [PubMed: 27846500]
32. Ohora K, Meichin H, Zhao LM, Wolf MW, Nakayama A, Hasegawa J, Lehnert N & Hayashi T 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). [PubMed: 29148750]

33. Villarino L, Splan KE, Reddem E, Alonso-Cotchico L, de Souza CG, Lledos A, Marechal JD, Thunnissen AMWH & Roelfes G An Artificial Heme Enzyme for Cyclopropanation Reactions. Angew. Chem. Int. Ed. 57, 7785–7789, (2018).
34. Zhao JM, Bachmann DG, Lenz M, Gillingham DG & Ward TR An artificial metalloenzyme for carbene transfer based on a biotinylated dirhodium anchored within streptavidin. Catal. Sci. Technol. 8, 2294–2298, (2018).
35. Carminati DM & Fasan R Stereoselective Cyclopropanation of Electron-Deficient Olefins with a Cofactor Redesigned Carbene Transferase Featuring Radical Reactivity. ACS Catal. 9, 9683–9697, (2019). [PubMed: 32257582]
36. Stenner R, Steventon JW, Seddon A & Anderson JLR A de novo peroxidase is also a promiscuous yet stereoselective carbene transferase. Proc. Natl. Acad. Sci. USA 117, 1419–1428, (2020). [PubMed: 31896585]
37. Chandgude AL, Ren X & Fasan R Stereodivergent Intramolecular Cyclopropanation Enabled by Engineered Carbene Transferases. J. Am. Chem. Soc. 141, 9145–9150, (2019). [PubMed: 31099569]
38. Ren XK, Liu NY, Chandgude AL & Fasan R An Enzymatic Platform for the Highly Enantioselective and Stereodivergent Construction of Cyclopropyl-delta-lactones. Angew. Chem. Int. Ed. 59, 21634–21639, (2020).
39. Yong K, Salim M & Capretta A Intramolecular carbenoid insertions: Reactions of alpha-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 TJ & Walsh EJ Synthesis of Cycloalkenones Via the Intramolecular Cyclopropanation of Furanyl Diazo Ketones. J. Org. Chem. 51, 5036–5038, (1986).
41. Ryan J, Siauciulis M, Gomm A, Macia B, O'Reilly E & Caprio V Transaminase Triggered Aza-Michael Approach for the Enantioselective Synthesis of Piperidine Scaffolds. J. Am. Chem. Soc. 138, 15798–15800, (2016). [PubMed: 27960354]
42. Zawodny W, Montgomery SL, Marshall JR, Finnigan JD, Turner NJ & Clayden J Chemoenzymatic Synthesis of Substituted Azepanes by Sequential Biocatalytic Reduction and Organolithium-Mediated Rearrangement. J. Am. Chem. Soc. 140, 17872–17877, (2018). [PubMed: 30521324]
43. Zhang RK, Chen K, Huang X, Wohlschlager L, Renata H & Arnold FH Enzymatic assembly of carbon-carbon bonds via iron-catalysed sp(3) C-H functionalization. Nature 565, 67–72, (2019). [PubMed: 30568304]
44. 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). [PubMed: 29268614]
45. de Visser SP, 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). [PubMed: 11457014]
46. Schroder D, Shaik S & Schwarz H Two-state reactivity as a new concept in organometallic chemistry. Acc. Chem. Res. 33, 139–145, (2000). [PubMed: 10727203]
47. Sreenilayam G & Fasan R Myoglobin-catalyzed intermolecular carbene N-H insertion with arylamine substrates. Chem. Commun. 51, 1532–1534, (2015).
48. Kondrashov DA, Zhang W, Aranda R. t., Stec B & Phillips GN Jr. Sampling of the native conformational ensemble of myoglobin via structures in different crystalline environments. Proteins 70, 353–362, (2008). [PubMed: 17680690]
49. Blomberg MRA, Borowski T, Himo F, Liao RZ & Siegbahn PEM Quantum Chemical Studies of Mechanisms for Metalloenzymes. Chem. Rev. 114, 3601–3658, (2014). [PubMed: 24410477]
50. Himo F Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. J. Am. Chem. Soc. 139, 6780–6786, (2017). [PubMed: 28493715]
51. Lind MES & Himo F Quantum Chemistry as a Tool in Asymmetric Biocatalysis: Limonene Epoxide Hydrolase Test Case. Angew. Chem. Int. Ed. 52, 4563–4567, (2013).

52. Liao RZ & 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). [PubMed: 23517056]

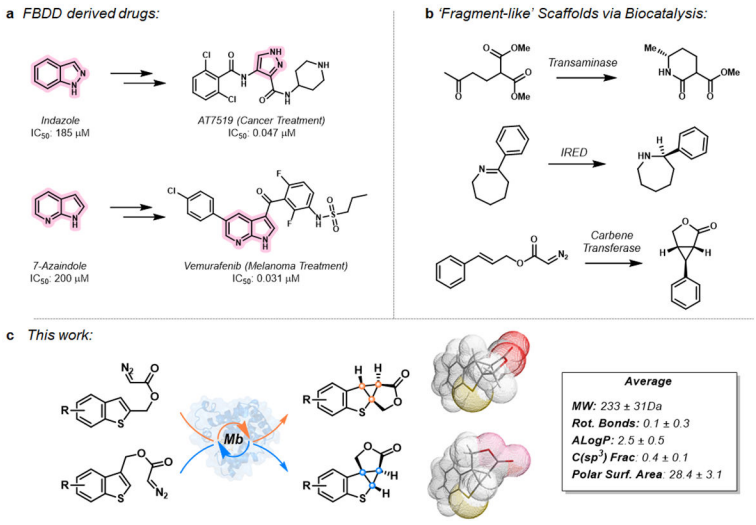


Figure 1. Sp^2 vs. sp^3-rich molecular 'fragments' for applications in fragment-based drug discovery.

(a) Representative drug molecules discovered and developed via fragment-based drug discovery (FBDD). **(b)** 'Fragment-like' scaffolds generated via biocatalysis include chiral piperidines,[41] azepans,[42] and cyclopropyl-γ-lactones.[37] **(c)** Biocatalytic intramolecular cyclopropanation of C2- and C3-functionalized benzothienyl substrates to give sp^3-rich tetracyclic scaffolds with distinct 3D shapes and Ro3-compliant properties.

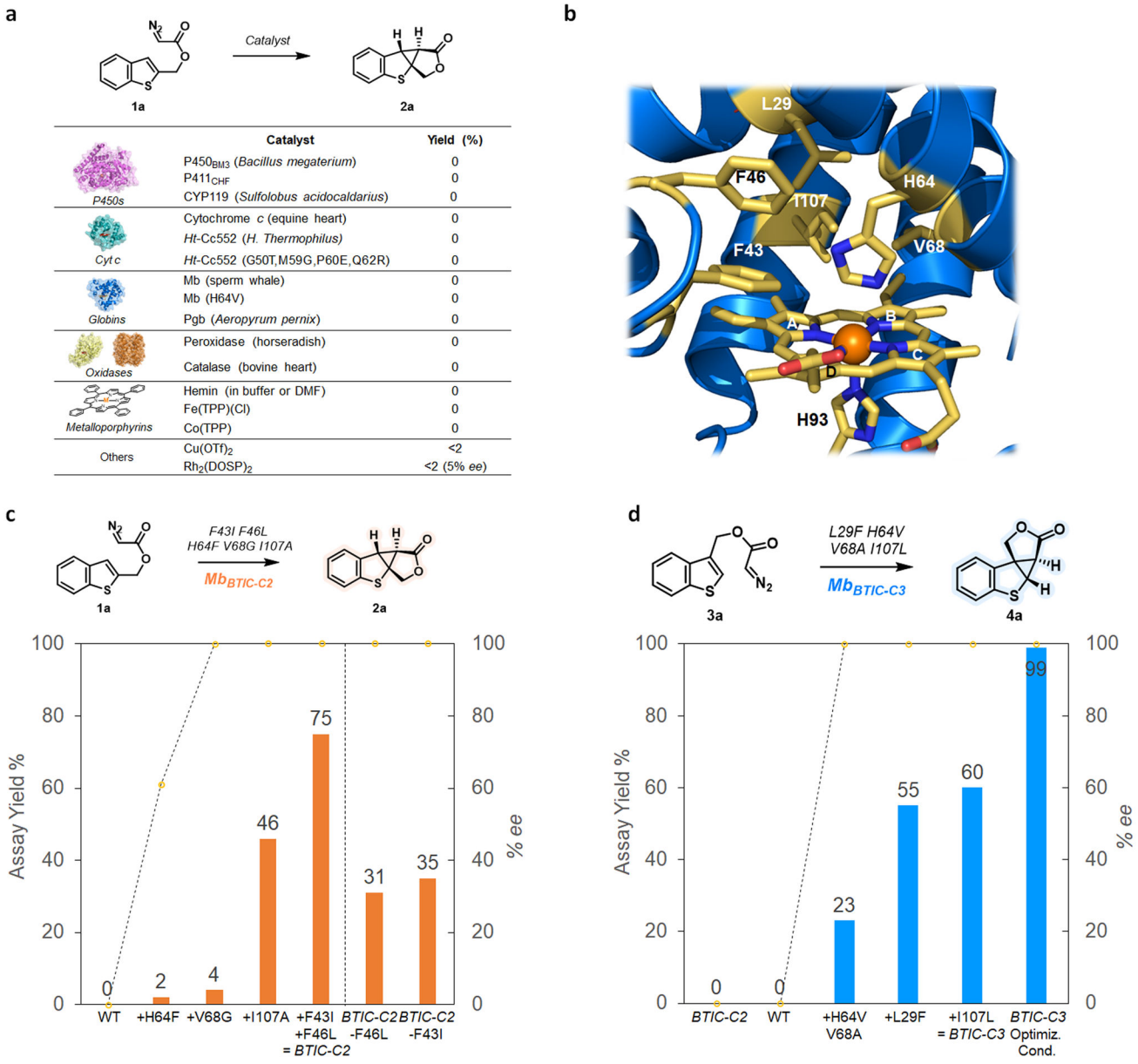


Figure 2. Biocatalytic intramolecular cyclopropanation of benzothiophenyl substrates.

a) Activity of a diverse panel of hemoproteins (0.8 mol%) and chemical catalysts (10 mol%) toward the intramolecular cyclopropanation of **1a** (see Supplementary Table 1 for reaction conditions and additional catalysts). Representative crystal structures of P450s (pdb 1FAH), cytochrome *c* (pdb 1HRC), globin (pdb 2MBW), peroxidase (pdb 1HCH), catalase (pdb 1TGU) are shown. **b)** Active site of sperm whale myoglobin (Mb; pdb 1JW8) with the heme and surrounding amino acid residues highlighted as sticks in gold. Heme b pyrrole rings labeled according to convention. Numbering of pyrrole rings follows Fischer nomenclature. **c-d)** Activity and enantioselectivity of Mb variants in the intramolecular cyclopropanation of **1a** or **3a** along the evolutionary paths leading to $Mb_{BTIC-C2}$ (**c**) and $Mb_{BTIC-C3}$ (**d**). Reported

yields represent the product yields as determined by GC analysis using calibration curves generated with isolated products.

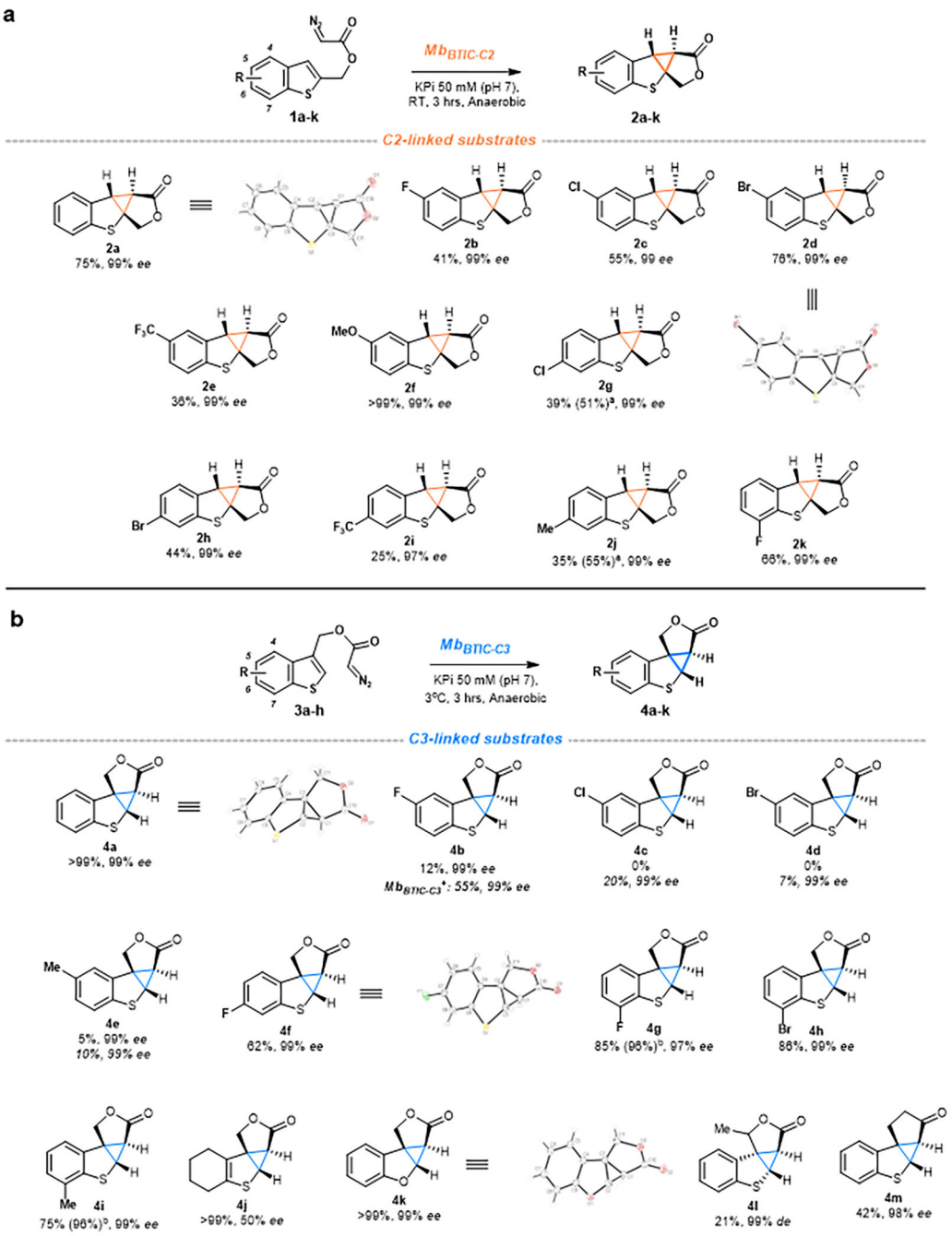


Figure 3. Substrate scope of $Mb_{BTIC-C2}$ and $Mb_{BTIC-C2}$ biocatalysts.

Reaction conditions: 20 μM Mb catalyst, 2.5 mM diazo compound, 10 mM $Na_2S_2O_4$, 3 hrs, anaerobic conditions. All reactions with $Mb_{BTIC-C3}$ and Mb(L29F,H64V,V68A,I107V) (= $Mb_{BTIC-C3}^{+}$) were carried out using slow diazo addition at 3°C. [a] Using whole cells (OD_{600} = 40). [b] Using 40 μM Mb catalyst. [c] Enantiomers not resolvable.

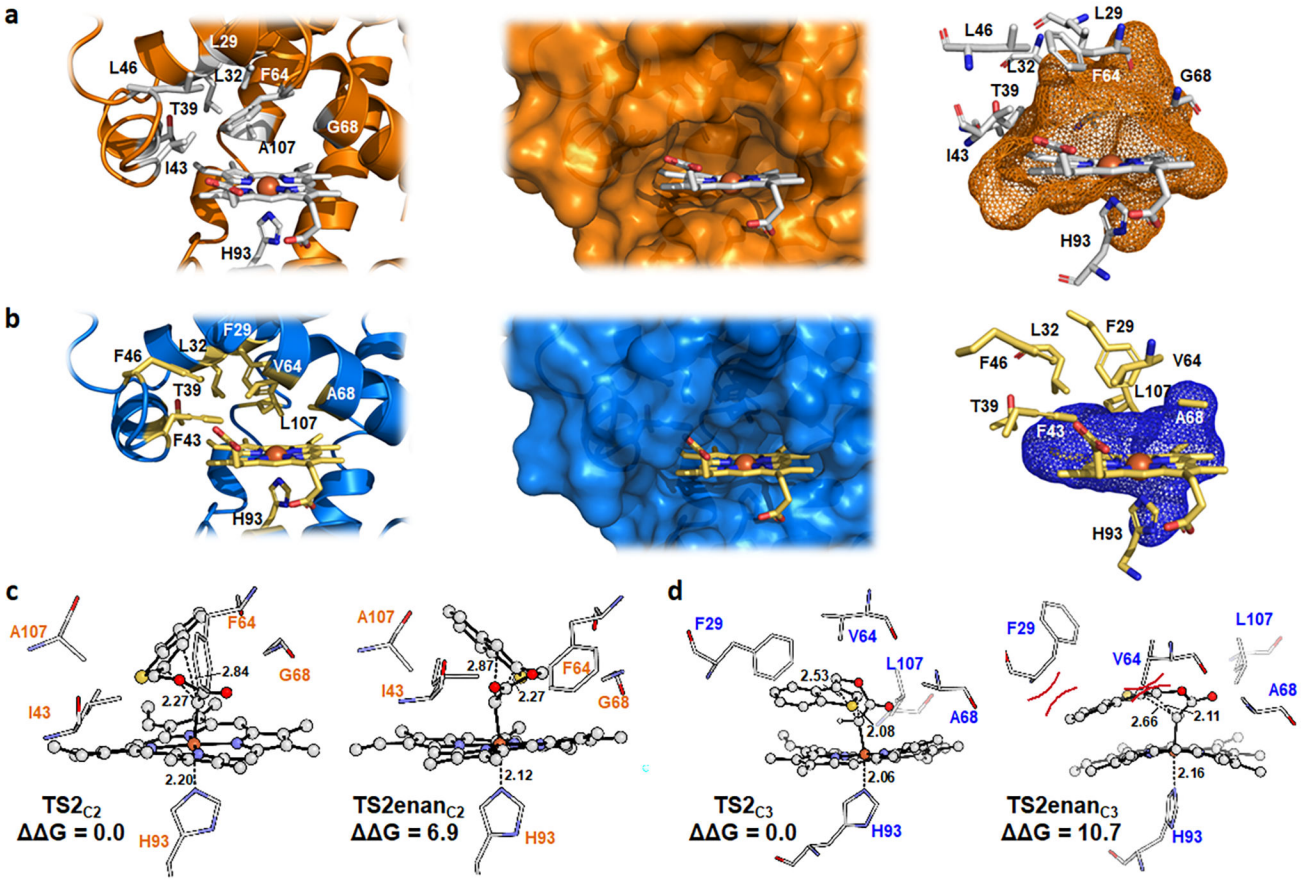


Figure 4. Crystal structures of $Mb_{BTIC\text{-}C2}$ and $Mb_{BTIC\text{-}C2}$ and corresponding stereochemical models.

a-b, High-resolution crystal structure of (**a**) $Mb_{BTIC\text{-}C2}$ and (**b**) $Mb_{BTIC\text{-}C3}$, highlighting the amino acid residues surrounding the heme cofactor (*left panel*), molecular surface representation of the heme binding site (*middle panel*), and the active site volume (*right panel*). **c-d,** Optimized geometries of (**a**) **$TS2_{C2}$** and its enantiomeric transition state (**$TS2enan_{C2}$**) in the theozyme model of $Mb_{BTIC\text{-}C2}$, and **(d) $TS2_{C3}$** and its enantiomeric transition state (**$TS2enan_{C3}$**) in the theozyme model of $Mb_{BTIC\text{-}C3}$. Relative Gibbs energies (ΔΔG) are given in kcal·mol^{-1}, and all distances are in angstroms (Å). See also Supplementary Figure 4 and computational details in the Supplementary Information.

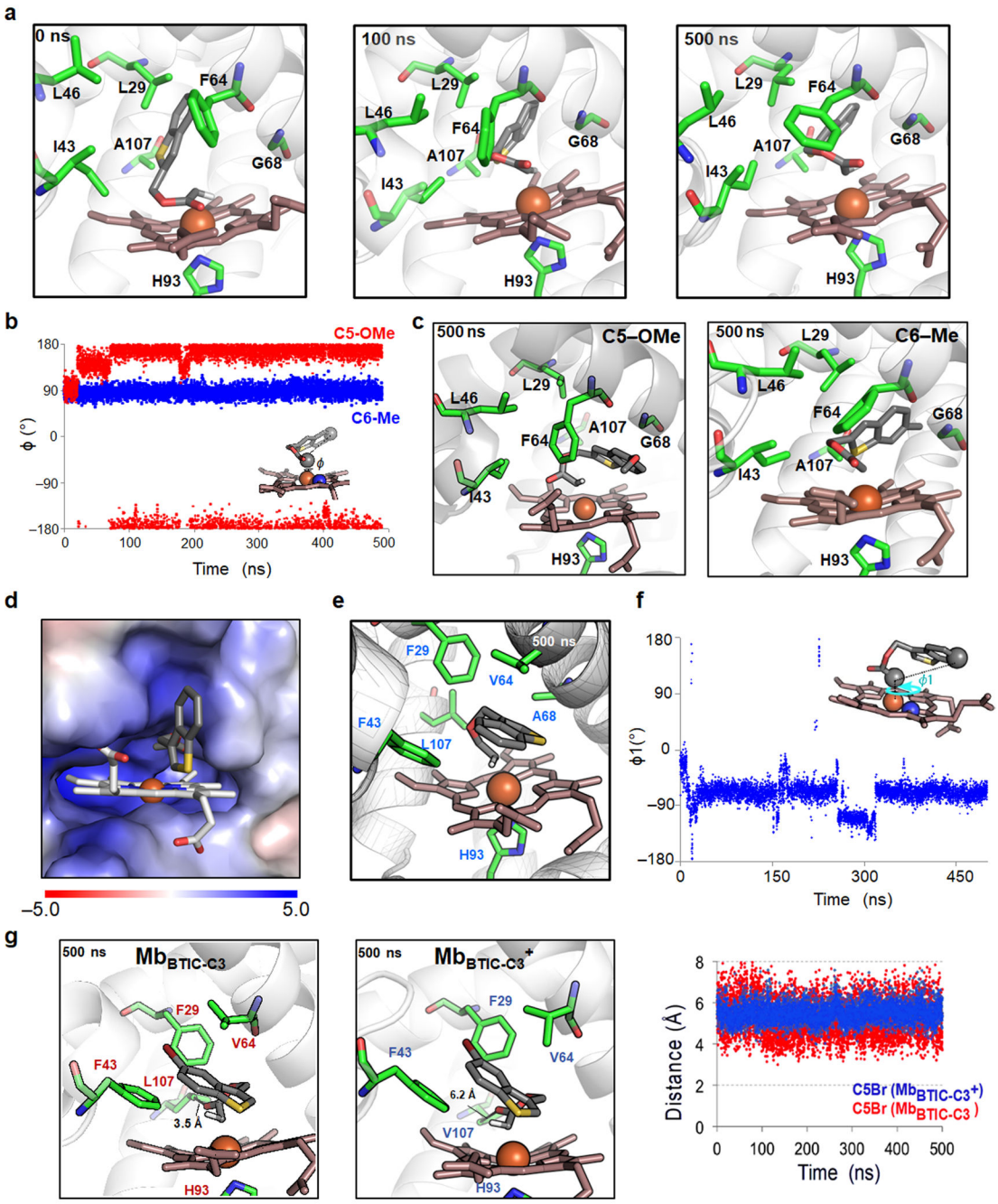


Figure 5. Analysis of the $Mb_{BTIC\text{-}C2}$, $Mb_{BTIC\text{-}C3}$, and MbBTIC-C3+ catalyzed cyclization reactions via molecular dynamics (MD) simulations.

a, MD snapshots for the **$TS2_{C2}$** docked in $Mb_{BTIC\text{-}C2}$ protein at 0, 100, and 500ns. **b,** Variation of the pseudo dihedral angle (ϕ) in the MD simulation for the heme-carbene intermediates corresponding to the 5-OMe and 6-Me substrates (i.e., compounds **2f** and **2j**) in $Mb_{BTIC\text{-}C2}$. The atoms used for measurement of the ϕ angle are displayed as spheres. **c,** End-of-simulation (500 ns) poses for the **$TS2_{C2}$** transition states corresponding to the 5-OMe and 6-Me substrates in $Mb_{BTIC\text{-}C2}$. See Supplementary Figure 5 for additional time

points. **d,** Lowest-energy pose for the diazo substrate **3a** after molecular docking into the structure of $Mb_{BTIC\text{-}C3}$, showing the benzothiophene ring in the 'BT out' conformation. **e,** End-of-simulation (500 ns) pose for the **TS2$_{C3}$** complex in $Mb_{BTIC\text{-}C3}$. See Supplementary Figure 6 for additional time points. **f,** Variation of pseudo dihedral angle (ϕ!) across the different geometries of the heme-bound **TS2$_{C3}$** complex sampled during the 500 ns MD simulation. The reference atoms for the ϕ! angle are displayed as spheres. **g,** MD snapshots at 500 ns of the **TS2$_{C3}$** analogues of the 5-Br substituted substrates (compound **3d**) in $Mb_{BTIC\text{-}C3}$ vs. $Mb_{BTIC\text{-}C3}^{+}$, which is the $Mb_{BTIC\text{-}C3}$ derived variant containing the rationally designed L107V mutation for enhanced reactivity toward 5-substituted benzothiophenes. The closest distance between the carbonyl (C=O) oxygen atom and the Leu107 residue in these complexes is indicated. The graph compares the variation of these distances for the C5-Br substrates in $Mb_{BTIC\text{-}C3}$ vs. $Mb_{BTIC\text{-}C3}^{+}$ throughout the 500 ns MD simulation.