

# Computational design of generalist cyclopropanases with stereodivergent selectivity

Received: 12 June 2025

Accepted: 5 January 2026

Published online: 26 January 2026

Check for updates

Zhuofan Shen<sup>1,7</sup>, Mary G. Siriboe<sup>2,7</sup>, Xinkun Ren<sup>3,6,7</sup>, Thakshila Dayananda<sup>3</sup>,  
Jermaine L. Jenkins<sup>4</sup>, Sagar D. Khare<sup>1,5</sup>✉ & Rudi Fasan<sup>2,3</sup>✉

Stereodivergent catalysis, whereby the full complement of stereoisomeric products is obtained through a set of stereocomplementary catalysts, represents a powerful tool for synthetic organic and medicinal chemistry. Despite recent progress in engineering biocatalysts for new-to-nature cyclopropanation reactions, cyclopropanases featuring a combination of stereodivergent selectivity with broad substrate scope have been elusive. Here, we report a mechanism-based, multi-state computational design workflow useful for the design of ‘generalist’ cyclopropanation biocatalysts with tailored selectivity. Using this strategy, cyclopropanases with high and predictable *trans*-(1*R*,2*R*), *cis*-(1*R*,2*S*), or *cis*-(1*S*,2*R*)-stereoselectivity in the transformation of a broad range of olefin substrates are designed based on three different hemoprotein scaffolds, including one (indoleamine 2,3-dioxygenase-1) not previously reported to support non-native carbene transfer reactions. Combined with a previously reported *trans*-(1*S*,2*S*)-stereoselective cyclopropanase, this biocatalytic toolbox provides access to a full set of cyclopropane stereoisomers from over 20 structurally diverse olefin substrates with high diastereo- and enantioselectivity (up to 99% *de*. and 99% *ee*). Crystal structures of a designed catalyst show good agreement with the computational model and highlight the role of subtle conformational heterogeneity in determining stereoselectivity. We envision that the present computational design methodology can guide the development of biocatalysts with tailored stereoselectivity for other carbene transfer reactions and enzymatic transformations.

Methodologies for stereodivergent synthesis, which allow the full complement of stereoisomers of a given product to be obtained from the same set of starting materials, are highly enabling tools in organic chemistry, particularly in the context of drug discovery and natural product synthesis<sup>1</sup>. It is well known that stereochemistry can influence

the interaction of bioactive molecules with their biomolecular targets, as well as their off-target effects, and absorption, distribution, metabolism, elimination, excretion, and toxicity (ADMET) properties<sup>2,3</sup>. As such, access to all possible stereoisomers of a chiral molecule is critical for enriching the screening library to identify small-molecule binders

<sup>1</sup>Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. <sup>2</sup>Department of Chemistry and Biochemistry, The University of Texas at Dallas, Richardson, TX, USA. <sup>3</sup>Department of Chemistry, University of Rochester, Rochester, NY, USA. <sup>4</sup>Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA. <sup>5</sup>Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. <sup>6</sup>Present address: College of Engineering and Applied Sciences, Nanjing University, Nanjing, Jiangsu, China. <sup>7</sup>These authors contributed equally: Zhuofan Shen, Mary G. Siriboe, Xinkun Ren. ✉e-mail: [sagar.khare@rutgers.edu](mailto:sagar.khare@rutgers.edu); [rudi.fasan@utdallas.edu](mailto:rudi.fasan@utdallas.edu)

of proteins that are hard to target<sup>4,5</sup> and distinguishing the bioactive stereoisomer (eutomer) with favorable ADMET properties from the less active or toxic stereoisomer (distomer)<sup>6</sup>. For approved drugs, efficient access to alternative stereoisomers represents a key step for drug repurposing via ‘chiral switching’<sup>7</sup>. Biocatalysis has emerged as a powerful strategy for the production of chiral molecules, including pharmaceutical leads and intermediates<sup>8</sup>. However, the development of stereodivergent biocatalysts capable of accepting a broad range of substrates and constructing molecules with multiple stereogenic centers remains a challenging endeavor.

Due to their distinct 3D structure, conformational rigidity, and high potential for stereochemical diversity (up to distinct 8 stereoisomers), cyclopropane rings are privileged scaffolds in medicinal chemistry<sup>9</sup>. Cyclopropanes are also recurring structural motifs in bioactive natural products and they can serve as valuable synthetic intermediates<sup>10,11</sup>. Over the past years, the development of engineered hemoproteins, including cytochrome P450s<sup>12–15</sup>, myoglobin (Mb)<sup>16,17</sup>, and others<sup>18–20</sup>, has made possible biocatalytic strategies for catalyzing ‘abiological’ olefin cyclopropanation reactions with diazo compounds with high degrees of activity and stereoselectivity<sup>21–25</sup>. While enantiocomplementary biocatalytic cyclopropanations have been achieved in some cases as a result of extensive screening<sup>17,18,21,24–27</sup>, cyclopropanation biocatalysts featuring both complementary stereoselectivity and broad substrate scope have been rare. For the prototypical cyclopropanation reaction of an olefin with ethyl diazoacetate (EDA), for example, four possible stereoisomers are produced, requiring a catalyst to exhibit both high diastereo- (*trans/cis*) and enantiocontrol for selective formation of only one of the four possible stereoisomeric products. An engineered myoglobin (Mb) variant, Mb (H64V, V68A)<sup>16</sup>, was previously reported that exhibits the rare combination of high stereoselectivity along with broad substrate scope, as demonstrated by its ability to cyclopropanate a broad range of vinyl arenes and other olefins (> 30 distinct substrates) with consistently high *trans*-(1*S*,2*S*)-stereoselectivity (90–99% *de* and *ee*)<sup>17,28–30</sup>. In contrast, generalist (i.e., with a broad substrate scope) biocatalysts for the stereoselective synthesis of the other three stereoisomers (*trans*-(1*R*,2*R*), *cis*-(1*R*,2*S*), and *cis*-(1*S*,2*R*)), in particular the less thermodynamically favorable *cis*-stereoisomers, have remained elusive<sup>12,18,26,27</sup>. Thus, methods that can result in bespoke biocatalysts with high stereoselectivity and broad substrate scope for a desired stereoisomer are highly desirable.

Computational enzyme design provides a complementary route to rationally tailor enzyme catalytic selectivity using *in silico* models. A promising structure-based strategy involves modeling enzyme–transition state (TS) complexes and designing sequences that preferentially stabilize the TS of a desired reaction, either by repurposing natural protein scaffolds or by generating *de novo* architectures. The *de novo* design of stereoselective Diels–Alderases represented a pioneering demonstration of stereochemical control by computational design<sup>31</sup>. However, these early designer enzymes exhibited low catalytic activities compared to natural enzymes. Other physics-based modeling approaches, including molecular dynamics (MD) simulations, have been applied to redesign existing enzymes and alter or improve their regio- and stereoselectivity<sup>32,33</sup>. While such methods can provide detailed mechanistic insights, their high computational cost prohibits their applicability for high-throughput exploration of multiple structurally diverse scaffolds and large mutant libraries, confining their use primarily to fine-tuning existing scaffolds with known catalytic activity. In principle, a computationally inexpensive (relative to MD) sequence and conformational sampling approach coupled with a sufficiently accurate determination of a multi-state design objective function<sup>34</sup> should allow sequence optimization on multiple properties including selectivity towards different substrates. Chica and co-workers used multi-state design to expand the substrate scope of an aminotransferase to include the non-native substrate L-phenylalanine<sup>35</sup>. More recently, deep learning-based

protein design frameworks have opened new avenues for the *de novo* design of biocatalysts, including stereoselective cyclopropanases<sup>36,37</sup>. Nonetheless, stereochemical outcomes of these enzymes were not explicitly optimized during computational design, and access to all possible product stereoisomers for a given reaction remained elusive. In parallel, machine learning–assisted directed evolution has emerged as a powerful hybrid strategy, aiding experimental screening with predictive modeling to accelerate the discovery of stereoselective enzymes<sup>38,39</sup>. Yet, these approaches still rely on pre-existing experimental datasets specific to a given enzyme class and reaction, limiting their generalizability.

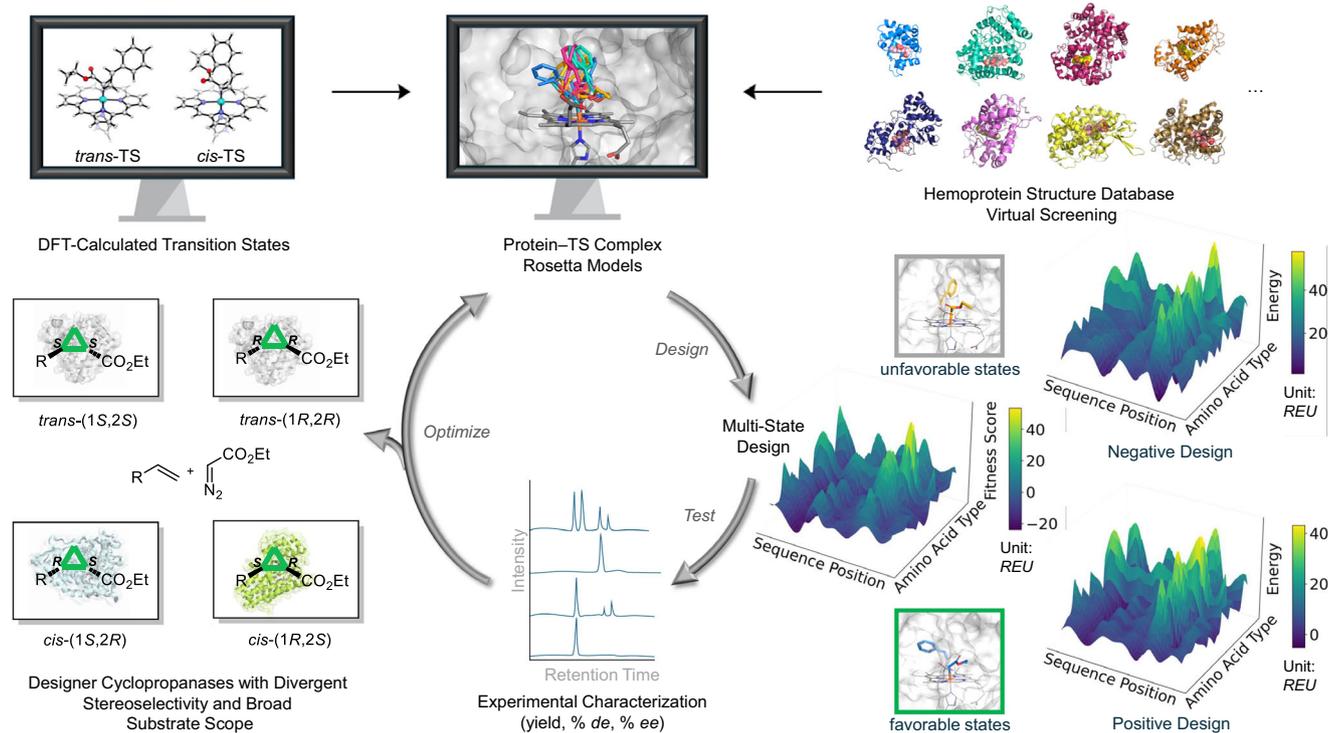
Targeting olefin cyclopropanation with EDA as a model reaction of synthetic relevance, we report here a mechanism-driven computational design strategy for developing biocatalysts capable of catalyzing this reaction with pre-defined and divergent diastereo- and enantioselectivity across a broad range of substrates (Fig. 1). This method entails a multi-state modeling approach that explicitly considers multiple transition state (TS) stereoisomers and conformers involved in the formation of the different stereoisomeric products and evaluates them in various configurational and conformational states within the active site of a hemoprotein. Using Rosetta FastDesign<sup>40</sup> and an evolutionary algorithm balancing both positive and negative design, a suitable hemoprotein scaffold(s) from the available hemoprotein structure database and appropriate active site mutations are identified to favor formation of a desired cyclopropane stereoisomer, while disfavoring formation of the other three possible products (Fig. 1). Through concise design-test-optimize cycles, this strategy enabled the computational design and experimental validation of generalist *trans*-(1*R*,2*R*), *cis*-(1*S*,2*R*), and *cis*-(1*R*,2*S*)-stereoselective cyclopropanases based on three different hemoprotein scaffolds, including one previously unexplored in carbene transfer reactions. Complementing the generalist *trans*-(1*S*,2*S*)-stereoselective cyclopropanase Mb\* reported previously<sup>16</sup>, this designer enzyme set constitutes a comprehensive biocatalytic platform for obtaining all four possible cyclopropane stereoisomers across a broad range of olefin substrates, including unactivated olefins, with high and tailored stereoselectivity (Fig. 1).

## Results

### Development and calibration of a multi-state modeling strategy for prediction of stereoselectivity

For computationally predicting stereopreference of hemoprotein-based cyclopropanation biocatalysts (Fig. 1), we utilized our previously developed method combining density functional theory (DFT) with Rosetta modeling<sup>41,42</sup>. Briefly, DFT calculations were used to compute the four TS stereoisomers, each corresponding to one of the four possible configurations of the 1,2-disubstituted cyclopropane product, for the attack of the styrene substrate on the iron-porphyrin-carbene intermediate (Fig. 2A)<sup>43</sup>. Several conformational isomers were modeled for each stereoisomeric TS (Supplementary Table S1, Supplementary Data 1). The DFT-calculated energy differences between the *trans/cis* stereoisomeric TSs and between the conformational isomers of these TSs are within 1–3 kcal·mol<sup>−1</sup> (Supplementary Table S2 and Figs. S1–S3), consistent with the expectation that non-covalent interactions with the protein matrix are the major determinants of stereoselectivity. Therefore, we anticipated that calculated protein-TS complex energy differences between the different stereoisomeric TSs calculated using Rosetta could predict experimentally observed stereoselectivities.

A panel of 40 engineered Mb variants previously characterized for their activity and stereoselectivity in the cyclopropanation of styrene with EDA was used for method calibration<sup>17</sup>. This enzyme set included the *trans*-(1*S*,2*S*)-stereoselective Mb variant Mb (H64V, V68A) (denoted as Mb\*)<sup>16</sup>, Mb (L29T, H64V, V68L) (denoted as RR2)<sup>17</sup>, a representative member of the first-generation “RR” variants (RR1–RR5) exhibiting *trans*-(1*R*,2*R*)-stereoselectivity in this reaction, albeit with narrow



**Fig. 1 | Overall workflow for the computational design and experimental validation of generalist cyclopropanases with stereodivergent selectivity.** DFT-calculated TS stereoisomers and Rosetta modeling/design are used to identify (i) an optimal hemoprotein scaffold among known carbene transferases or within the entire Protein Data Bank (PDB) and (ii) design optimal active site mutations to

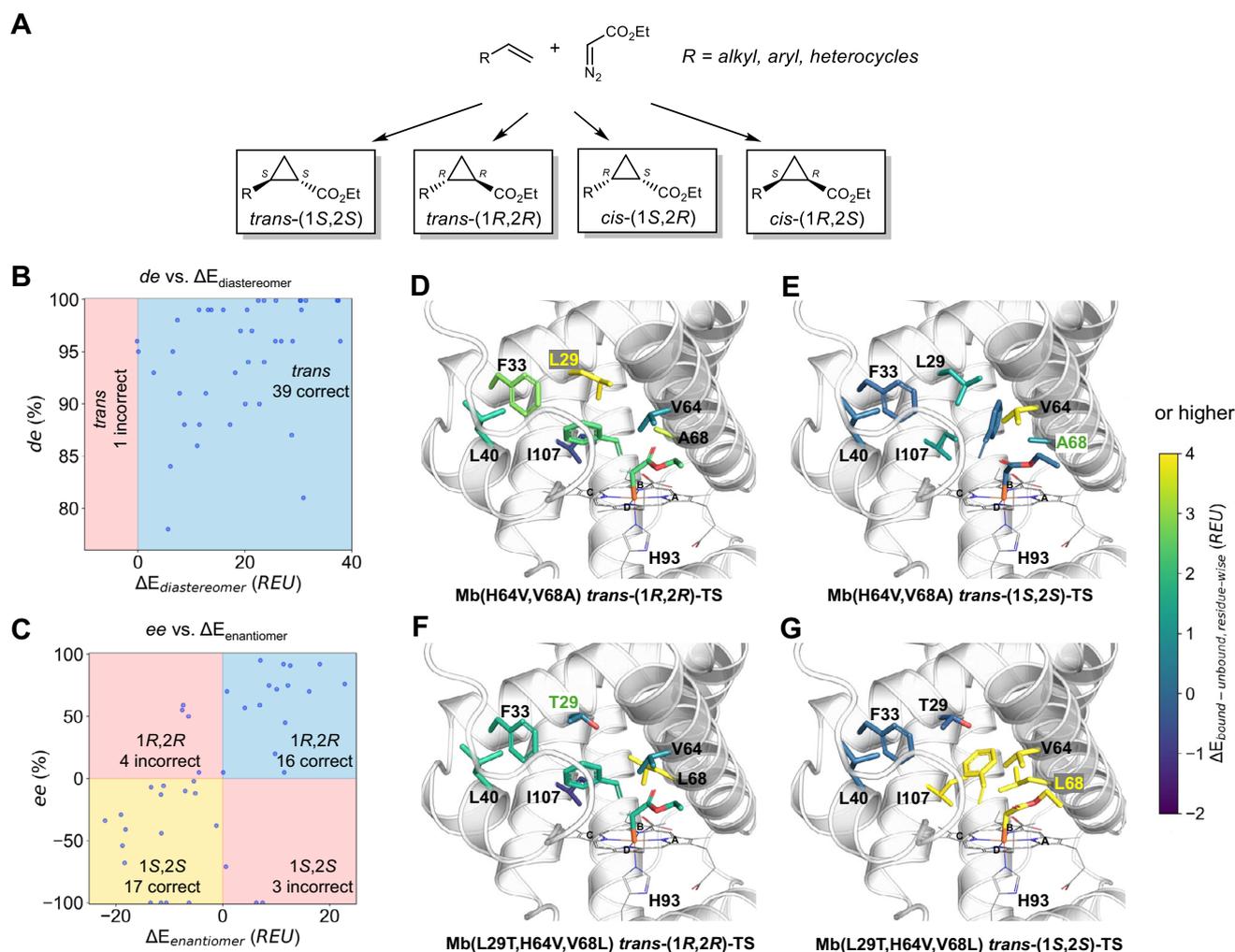
stabilize the desired TS stereoisomer in the hemoprotein active site, while simultaneously disfavoring interaction with competing TS stereoisomers. Upon experimental testing in the target reaction (styrene cyclopropanation with EDA), another cycle of modeling/design can be carried out to enhance the desired diastereo- and enantioselectivity and/or expand substrate tolerance of the biocatalyst.

substrate scope, and 38 additional Mb variants, each exhibiting good activity (>30% yield) and *trans*-diastereoselectivity (76–99% *de*) for the cyclopropanation of styrene with EDA but varying degree of enantioselectivity (from –99% to 95% *ee*)<sup>16,17</sup>. To computationally evaluate the stereopreference of these cyclopropanases, the TS isomers were superimposed (based on the porphyrin ring of the heme group) in various configurational and conformational states into a crystal structure of Mb. Active site mutations were modeled into the substrate-docked Mb structures, and the protein structures were optimized via Rosetta-based modeling (Fig. 1; see Methods and SI for DFT and Rosetta modeling details)<sup>41,42</sup>. For each Mb variant, the stereopreference was predicted based on the calculated relative Rosetta energies of the four stereoisomeric TS configurations. Comparison between the calculated energies and experimental data revealed that *trans*-diastereoselectivity was qualitatively well-predicted for 39 out of the 40 Mb variants (Fig. 2B). The calculated energies also accurately predicted (1*R*,2*R*)-enantioselectivity for 16 out of 20 variants (Fig. 2C; Supplementary Table S3) and (1*S*,2*S*)-enantioselectivity for 17 out of 20 variants (Fig. 2C; Supplementary Table S4), indicating good overall predictive power of our structure-based modeling protocol with respect to stereopreference.

Mb\* and RR2 models were analyzed to identify the likely origin of the stereopreference via structure and per-residue energy comparisons between the *trans*-(1*R*,2*R*)-TS and *trans*-(1*S*,2*S*)-TS-bound Mb models. According to the Rosetta models, for both *trans*-stereoisomers, the polar carbene-borne ester moiety projects towards the solvent-exposed side of the distal cavity, while the non-polar styrene is buried in the hydrophobic pocket (Fig. 2D–G). However, distinct orientations of the styrene moiety in these structures are responsible for either (1*R*,2*R*) or (1*S*,2*S*)-enantioselectivity. In Mb\*, two tightly packed sidechains from Leu29 and Phe33 are destabilized upon binding to the *trans*-(1*R*,2*R*)-TS (Fig. 2D; Supplementary Table S5).

Meanwhile, the *trans*-(1*S*,2*S*)-TS is stabilized by the V68A substitution, which creates a cavity for accommodating the styrene moiety of the pro-*trans*-(1*S*,2*S*)-TS and forms favorable hydrophobic interactions with it (Fig. 2E). However, these beneficial protein-styrene contacts are missing in the *trans*-(1*R*,2*R*)-TS-bound state, resulting in the higher energy associated with residue Ala68 (from V68A) in this complex (Fig. 2E; Supplementary Table S5). In contrast, the L29T substitution in RR2 creates sufficient space for accommodating the pro-*trans*-(1*R*,2*R*)-TS within the active site (Fig. 2F), and residues Leu68 (from V68L), Val64 (from H64V), and Ile107 render the RR2-bound *trans*-(1*S*,2*S*)-TS energetically unfavorable (Fig. 2G; Supplementary Table S6). Altogether, these studies revealed how individual mutations in these 40 variants together shape the active site to not only bind to the favored TS but also destabilize the competing TS enantiomer, resulting in the experimentally observed enantiodivergence in this reaction.

We also examined possible origins of the relatively few (7 out of 40) incorrect enantioselectivity predictions (Fig. 2C). (1*S*,2*S*)-selective Mb variants that were incorrectly predicted as (1*R*,2*R*)-selective variants include Mb(F43V,V68F), Mb(L29T, H64V,V68F, I107A), and Mb(V68F), all of which contain V68F, a substitution that more frequently appears in (1*R*,2*R*)-selective variants. We postulate that our models may overestimate the destabilization effect of V68F on the *trans*-(1*S*,2*S*)-TS in these variants, possible due to limited sampling of backbone degrees of freedom in our Rosetta calculations. (1*R*,2*R*)-selective Mb variants that were incorrectly predicted as (1*S*,2*S*)-selective variants include Mb(F43S,H64V), Mb(H64V,I107Y), Mb(L29G,H64V), and Mb(L29T,H64V). These four variants all exhibit low to moderate (1*R*,2*R*)-enantioselectivity (5–59% *ee*), which may contribute to the difficulty of correctly predicting their enantioselectivity. Indeed, the calculated  $\Delta E_{\text{enantiomer}}$  gaps for these variants are also small (4.3–7.5 REU). Finally, Mb(L29G,H64V) was incorrectly predicted as *cis*-diastereoselective (Fig. 2B). In this case, the under-



**Fig. 2 | Modeling of Mb biocatalysts-catalyzed cyclopropanation to predict experimentally measured stereoselectivity data.** A Reaction scheme for the hemoprotein-catalyzed olefin cyclopropanations with EDA. Scatter plot of experimentally measured *de* (B) and *ee* (C) values vs. Rosetta-calculated diastereomer and enantiomer energy differences for 40 laboratory-evolved Mb variants. Reaction conditions: 20  $\mu\text{M}$  catalyst, 10 mM substrate, 20 mM EDA, 10 mM  $\text{Na}_2\text{S}_2\text{O}_4$ , in potassium phosphate buffer (50 mM, pH = 7.0), anaerobic. Rosetta

model of Mb\* (= Mb (H64V, V68A)) in complex with *trans*-(1*R*,2*R*)-TS (D) and *trans*-(1*S*,2*S*)-TS (E) vs. RR2 (= Mb (L29T, H64V, V68L)) in complex with *trans*-(1*R*,2*R*)-TS (F) and *trans*-(1*S*,2*S*)-TS (G) of the cyclopropanation of styrene with EDA. Individual residues are colored according to their calculated per-residue binding energy  $\Delta E_{\text{bound}} - \text{unbound}$ , with darker colors representing lower energy (more stable) and lighter colors representing higher energy (less stable).

estimation of (1*R*,2*R*)-TS stability may be linked to substantial changes of backbone conformation as a result of the L29G mutation, which is not precisely captured by the Rosetta-based modeling approach.

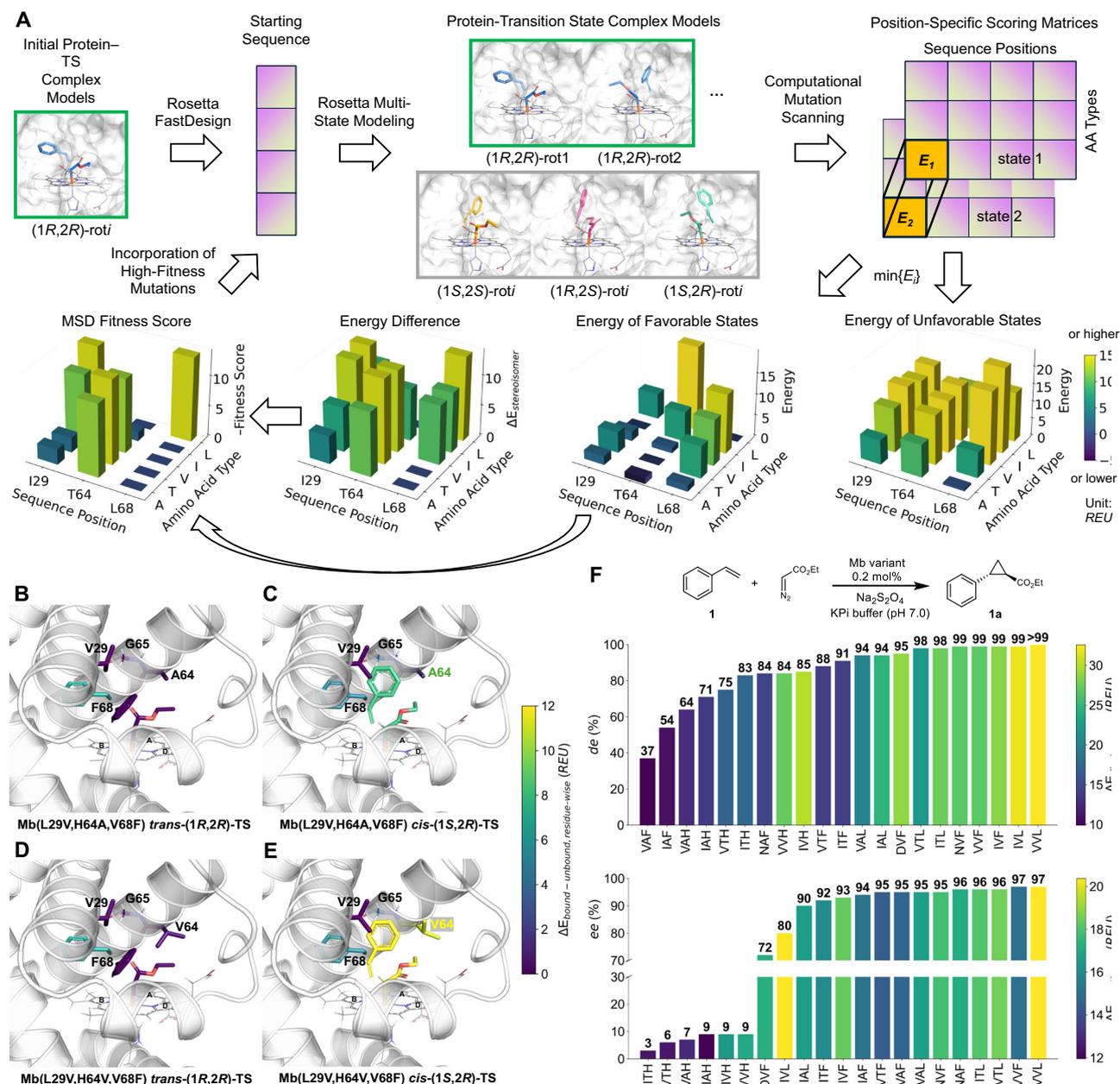
### Multi-state design of *trans*-(1*R*,2*R*)-stereoselective Mb cyclopropanases

Based on the benchmarking studies above, we envisioned that designing variants with pre-defined stereoselectivity would require not only positive design, i.e., maximizing favorable interactions to accommodate the desired stereoisomeric TS while maintaining enzyme stability, but also negative design, i.e., disfavoring interactions with the other competing TS (Fig. 3A). Accordingly, positive and negative design were integrated into one comprehensive fitness function, which measures the favorability of the 'positive state(s)' as well as the difference between positive and the most competitive negative state(s) based on calculated energies:

$$\text{Fitness} = (E_{\text{favorable, mut}} - E_{\text{favorable, WT}}) + (E_{\text{favorable, mut}} - \min(E_{\text{unfavorable, mut}}))$$

(See SI for a detailed discussion of the fitness function).

As a first test of our design approach, we applied it toward the design of a generalist *trans*-(1*R*,2*R*)-stereoselective cyclopropanase based on the Mb scaffold. The most *trans*-(1*R*,2*R*)-stereoselective Mb-based cyclopropanase obtained from previous efforts, Mb (L29T, F43W, H64V, V68F) (denoted as RR5)<sup>17</sup>, exhibits high (1*R*,2*R*)-enantioselectivity (53% yield, 99% *de*, 95% *ee*) for the cyclopropanation of styrene with EDA, but with narrow substrate scope. Starting from the available crystal structure of Mb\*<sup>44</sup>, we applied Rosetta FastDesign to redesign the distal cavity residues to accommodate the *trans*-(1*R*,2*R*)-TS (positive design; Fig. 3A). Three active mutations, namely L29I, H64A and V68G, were identified for accommodating the *trans*-(1*R*,2*R*)-TS (Supplementary Table S7). Using the TS-bound Mb (L29I, H64A, V68G) models as a starting point, a computational site-saturation mutation scanning for selectivity was performed at the three modified active site residues (i.e., 29, 64 and 68), followed by the calculation of the fitness scores and an adaptive search through the calculated position-specific scoring matrices (Fig. 3A; Supplementary Fig. S4). Site-saturation at position 68 with the multi-state design fitness function favored bulkier sidechains (e.g., Leu) over Gly due to its



**Fig. 3 | Multi-state design (MSD) protocol for designing *trans*-(1R,2R)-stereoselective Mb biocatalysts.** **A** Schematic representation of the computational multi-state design approach employed for achieving cyclopropanases with tailored stereoselectivity. A lower energy for the favorable stereoisomer and a higher energy for the unfavorable stereoisomer results in a higher  $\Delta E_{\text{stereoisomer}}$  value, which corresponds to predicted higher stereoselectivity. A mutation with a higher fitness score is more favorable. Rosetta models of Mb (L29V, H64A, V68F) in complex with *trans*-(1R,2R)-TS (**B**) and *trans*-(1S,2S)-TS (**C**) vs. RR10 (= Mb (L29V, H64V, V68F)) in complex with *trans*-(1R,2R)-TS (**D**) and *cis*-(1S,2R)-TS (**E**) of the cyclopropanation of styrene with EDA. Individual residue sidechains are colored

according to their calculated per-residue binding energy  $\Delta E_{\text{bound-unbound}}$ , with darker colors representing lower energy (more stable) and lighter colors representing higher energy (less stable). **F** Experimentally measured *de* and *ee* values ( $n \geq 2$ ) of the *trans*-(1R,2R)-stereoselective cyclopropanation of styrene (**1**) with EDA vs. Rosetta-calculated diastereomer and enantiomer energy differences for computationally designed triple mutant Mb variants. Designed variants are named using a three-letter code representing the amino acid identities at positions 29, 64, and 68. Reaction conditions: 20  $\mu\text{M}$  catalyst, 10 mM styrene (**1**), 20 mM EDA, 10 mM Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub>, in potassium phosphate (KPi) buffer (50 mM, pH = 7.0), anaerobic, 16 h.

significantly improved predicted (1R,2R)-enantiopreference. Subsequent iterations identified H64T and H64V for enhanced *trans*-diastereopreference compared to H64A (Fig. 3B–E; Supplementary Tables S8, S9). Overall, several high-fitness amino acid substitutions were identified at each of the three target positions, namely, L29  $\rightarrow$  I/N/V/D, H64  $\rightarrow$  T/V/A, and V68  $\rightarrow$  L/F/H (Supplementary Fig. S5). These predicted beneficial substitutions were combined to give thirty-six novel triple mutant Mb variants with computed energy gaps larger than 10 Rosetta Energy Units between the *trans*-(1R,2R)-state

and competing *trans*-(1S,2S)-states, as well as *cis*-states (Supplementary Table S7).

### Experimental validation of designed *trans*-(1R,2R)-stereoselective Mb cyclopropanases

All 36 designed variants were tested for activity and stereoselectivity in the cyclopropanation of styrene with EDA in whole cells. Gratifyingly, all of them displayed cyclopropanation activity, nearly all of them (32/36 = 88%) showed good to excellent *trans*-diastereoselectivity

(70–99% *de* (AVG: 85% *de*)), and nearly half of them (14/36 = 40%) exhibited high to very high (1*R*,2*R*)-enantioselectivity in the reaction (85–96% *ee* (AVG: 92% *ee*)) (Supplementary Table S10). The 21 well-expressing designed variants (expression yield > 5 mg·L<sup>-1</sup> culture) were further validated in reactions with purified protein, showing comparable or slightly higher degrees of *trans*-(1*R*,2*R*)-stereoselectivity compared to the whole-cell reactions (Supplementary Table S11). Importantly, a qualitative relationship between the experimentally obtained *de* values and the Rosetta-calculated diastereoisomer energy gap was observed (Fig. 3F). Notably, 12 designed biocatalysts compare favorably with RR5 in terms of *trans*-(1*R*,2*R*)-stereoselectivity (Designs: AVG 95% *de* and 94% *ee* vs. RR5: 99% *de* and 95% *ee*), while offering significantly improved catalytic activities (92% AVG yield vs. 53% yield) under identical reaction conditions (Supplementary Table S11).

### Design of generalist *trans*-(1*R*,2*R*)-stereoselective Mb cyclopropanases

The six best performing *trans*-(1*R*,2*R*)-stereoselective cyclopropanases derived from the workflow above (called RR6-RR11) were selected for further substrate scope studies using a panel of six *ortho*-, *meta*- and *para*-substituted styrene derivatives (Supplementary Tables S12–S18). These variants produced the desired *trans*-(1*R*,2*R*)-cyclopropane product 1a in high yields (69–99%) and *trans*-diastereoselectivity (94–99% *de*) but varying levels of (1*R*,2*R*)-enantioselectivity (24–97% *ee*), in particular with *para*-substituted styrene substrates. While several members of the RR6-RR11 enzyme set already outperformed the previous best catalyst among the RR1-RR5 panel<sup>17</sup> for each of the tested substrates, we sought to further expand their generality toward a broader range of olefin substrates. To this end, another round of computational design was carried out using the bulkier substrate *p*-CF<sub>3</sub>-styrene (11) as the model substrate. We first modeled several selected Mb cyclopropanases (RR6, RR7, RR10, RR11) in complex with the TS for the cyclopropanation of 11 with EDA and calculated their energies (Supplementary Table S19). Energy comparison on a per-residue basis of proteins in complex with the *trans*-(1*R*,2*R*)-TS vs. unbound proteins revealed three hotspots of energy difference corresponding to F33, L40 and L32 (Supplementary Table S20), all of which are located away from the heme moiety (Fig. 4A). In particular, in the *trans*-(1*R*,2*R*)-TS-bound Mb models, the F33 sidechain adopts a different solvent-exposed conformation (Fig. 4B) compared to the buried rotamer in the unbound protein (Fig. 4A) in order to alleviate unfavorable close contacts with the *para* substituent group on 11. Computational design results (Supplementary Figs. S5–S7) indicated that F33 → A/V/L substitutions enlarge the active-site pocket to better accommodate the bulkier, *p*-CF<sub>3</sub>-substituted substrate 11 (Fig. 4C), thereby stabilizing the binding of the *trans*-(1*R*,2*R*)-TS binding. This trend was consistently observed for *para*-substituted substrate 11 (Supplementary Table S19, S21), unsubstituted styrene 1 (Supplementary Table S22, S23), and other structurally diverse non-styrenyl olefin substrates 16, 17, 18, 23, and 25 (Supplementary Table S24). A summary comparing the differences in key designed residues within the active-site pocket is provided in Supplementary Table S25.

Based on these analyses, we designed, expressed, and tested twelve second-generation RR designs incorporating the F33A/V/L mutations into RR6, RR7, RR10, or RR11 as backgrounds (called RR12-RR23; Supplementary Tables S19, S22). Ten of the twelve designs exhibited excellent (1*R*,2*R*)-enantioselectivity (>99% *ee*) in the cyclopropanation of styrene with EDA (Fig. 4D; Supplementary Table S26), and five variants (i.e., RR12, RR14, RR17, RR22, RR23) show excellent (1*R*,2*R*)-enantioselectivity (>99% *ee*) in the presence of *p*-CF<sub>3</sub>-styrene (11) as the substrate (Fig. 4E; Supplementary Table S27). Based on these results, four of these five variants were challenged against a broader set (8) of substituted styrene derivatives

(Supplementary Fig. S8). While all these variants yielded desired *trans*-(1*R*,2*R*)-cyclopropane products in good yields (AVG: 77% yield) and high (1*R*,2*R*)-enantioselectivity (AVG: 94% *ee*), RR22 (= Mb (L29V, F33V, H64V, V68F)) emerged as the most general biocatalyst, giving 90–99% *ee* across all the tested substrates.

To further probe the generality of RR22 as a *trans*-(1*R*,2*R*)-stereoselective cyclopropanase, this variant was tested across a broad panel of 25 different olefins, which included *ortho/meta/para* and *alpha*-substituted styrene substrates (2–15 in Fig. 4F), 2-vinylnaphthalene (16), along with a structurally diverse set of electron-deficient, heteroatom-substituted and aliphatic olefins (17–25). RR22-catalyzed reactions were found to proceed with high to excellent *trans*-(1*R*,2*R*)-stereoselectivity (66–99% *de* (AVG: 95% *de*); 90–99% *ee* (AVG: 97% *ee*)) and 57% average yield across the entire panel of styrene derivatives to give products 1a–16a (Fig. 4F; Supplementary Fig. S9). X-ray crystallography of 16a confirmed the absolute configuration of this compound as the expected stereoisomer *trans*-(1*R*,2*R*) (Supplementary Tables S28, 29; Supplementary Fig. S10A). In addition, RR22 efficiently catalyzes cyclopropanations of electron-deficient, aliphatic and cyclic olefins to give the corresponding *trans*-(1*R*,2*R*)-cyclopropane products 17a–25a with consistently high *trans*-(1*R*,2*R*)-stereoselectivity (99% *de* except 21a (AVG: 90% *de*); 66–99% *ee* (AVG: 94% *ee*)) and 61% average yield. It is worth noting that electron-deficient olefins such as 17–20 or unactivated olefins such as 23–24 represent very challenging substrates for carbene transfer catalysts due to the typical electrophilic nature of the reactive metal-carbene intermediates involved in these reactions<sup>28,45</sup>. Cyclopropanation of benzofuran (25) is also difficult with iron-based catalysts and this transformation previously required a dedicated protein engineering campaign<sup>29</sup>, further highlighting these challenges along with the difficulty of achieving general, *trans*-(1*R*,2*R*)-stereoselectivity in this reaction. In contrast, the RR5 reactions featured high *trans*-(1*R*,2*R*)-stereoselectivity only with 3 out of the 25 substrates tested, showing little to no detectable activity on non-styrenyl substrates (17–25). Altogether, these results support the effectiveness of our computational framework in guiding the design of cyclopropanation biocatalysts with tailored *trans*-(1*R*,2*R*)-stereoselectivity.

### Comparison between crystal structures and Rosetta models of RR22 reveals substrate binding-induced conformational changes

We solved two crystal structures of RR22, with either water (aquo-complex) or imidazole bound to the distal axial coordination site of the heme cofactor (Fig. 4G; Supplementary Table S30). Overall, both structures show high agreement with the Rosetta model of RR22 (backbone RMSD = 0.2 Å). In particular, the sidechain rotamers of Val33 (F33V mutation), Phe46, and Val64 (H64V mutation) in the *trans*-(1*R*,2*R*)-TS-bound RR22 Rosetta model are identical to those observed in the RR22 aquo-complex crystal structure (Fig. 4H). No significant backbone rearrangements were detected between the *trans*-(1*R*,2*R*)-TS-bound RR22 model and the aquo-complex structure. Only minor sidechain adjustments in Phe68 (from V68F mutation), Val29 (from L29V mutation), Phe43, and Leu61 are necessary to accommodate the TS. In contrast, when RR22 binds the less energetically favorable *trans*-(1*S*,2*S*)-TS, pronounced conformational adjustments are observed in Phe43 and Phe46 sidechains, as well as in the backbone region surrounding the E helix residue H64V (Fig. 4I). These observations indicate that the RR22 active site is better pre-organized for binding the bulky substrate 11 in the *pro-trans*-(1*R*,2*R*)-state than in the energetically unfavorable *pro-trans*-(1*S*,2*S*)-state.

Compared to the aquo-complex structure (white, Fig. 4G), Phe46 and Leu61 in the imidazole-bound RR22 crystal structure are able to switch between an aquo-complex-like rotamer state (violet, Fig. 4G) and an alternative sidechain packing mode (green, Fig. 4G). The CD loop region (Phe43 to Phe46) and the E helix segment (Leu61 to Val64)



also exhibit local backbone flexibility<sup>46</sup>, collaboratively accommodating the bound imidazole (Fig. 4G). Similarly, in the Rosetta model of RR22 complexed with the *trans*-(1S,2S)-TS, comparable outward displacement of the E helix residue Val64 from the heme center, along with major Phe43 and Phe46 sidechains repacking to alleviate unfavorable contacts between the *para*-trifluoromethyl group of II and the CD loop, is also evident compared to the aquo-complex crystal structure (Fig. 4I). However, the accompanying CD loop backbone displacement observed in the imidazole-bound structure (Fig. 4G) was not captured in the Rosetta model (Fig. 4I). This omission may account for the discrepancies between Rosetta-calculated  $\Delta E_{\text{enantiomer}}$  values and experimentally measured *ee* values for the bulkier substrate II (Fig. 4E), compared to the unsubstituted I (Fig. 4D), for which such CD loop interactions and the associated modeling inaccuracies are absent.

Taken together, comparative analysis between the crystal structures and Rosetta models highlights the overall accuracy of RR22 computational models and demonstrates the successful design of an active site preorganization in RR22 that favors binding of the pro-*trans*-(1*R*,2*R*)-substrate. Moreover, the computational model successfully captured the dynamic structural adaptations of RR22 upon substrate binding, as confirmed by the crystal structures. These observations also suggest potential directions for improving computational modeling methods to more accurately describe enzymes undergoing transitions that involve both backbone and sidechain rearrangements during the catalytic cycle.

### Computational design of generalist *cis*-(1*S*,2*R*)-stereoselective cyclopropanase from cytochrome P450<sub>cam</sub>

Encouraged by the success with the designer *trans*-(1*R*,2*R*)-stereoselective Mb cyclopropanases, we targeted the design of hemoprotein-based biocatalysts with tailored stereoselectivity for the formation of the two *cis*-cyclopropane stereoisomers. Similar to the *trans*-(1*R*,2*R*)-stereoselective cyclopropanation, no general biocatalysts are available for these transformations<sup>12,26</sup>. To this end, we evaluated a panel of structurally diverse hemoprotein scaffolds previously reported to feature carbene transfer reactivity, namely myoglobin, dehaloperoxidase, cytochrome *c*, ascorbate peroxidase and P450s (e.g., P450<sub>cam</sub>, CYP119, P450<sub>BM3</sub>) using our multi-state design approach. These design calculations identified cytochrome P450 camphor 5-monooxygenase (known as P450<sub>cam</sub> or CYP101, Fig. 5A) as a promising lead for achieving *cis*-(1*S*,2*R*)-stereoselectivity. Rosetta FastDesign suggested that a V247L substitution (i.e., removal of C $\beta$ -branching) can enlarge the substrate binding pocket to better accommodate the *cis*-(1*S*,2*R*)-TS (Fig. 5B). In addition, a D297T mutation is predicted to create a less polar environment to accommodate the carbene-borne ester group while maintaining an electrostatic interaction between residue 297 and one heme propionate group, which is hydrogen-bonded to D297 in the wild-type enzyme (Fig. 5B).

Starting from this positive design-generated P450<sub>cam</sub> (V247L, D297T) sequence, we performed iterative computational mutation scanning at eleven first-shell positions to explore beneficial mutations for stabilizing *cis*-(1*S*,2*R*)-TS while disfavoring other TS stereoisomers. Mutations I395L and V396L were predicted to be effective toward promoting tighter hydrophobic sidechain packing with *cis*-(1*S*,2*R*)-TS and V295 (Fig. 5B), while destabilizing competing TS stereoisomers (Fig. 5C–E; Supplementary Table S31). In addition, several selectivity-enhancing mutations were identified at positions T101, F87, and Y96 (Supplementary Figs. S11–S12), which were incorporated, along with positive design-selected substitutions V247L and D297T, into a combinatorial variant library of designed *cis*-(1*S*,2*R*)-stereoselective P450<sub>cam</sub> cyclopropanases for selectivity prediction (Supplementary Figs. S13–S16). Balancing predicted catalytic activity and selectivity with sequence diversity to ensure broad coverage of different active-site mutation types and to avoid overrepresentation of potentially detrimental mutations that could compromise all tested variants, we

selected twenty-four designed sequences bearing three to seven active site substitutions for experimental validation (called SRI–SR24; Supplementary Table S32). A summary comparing the differences in key designed catalytic residues within the active-site pocket is provided in Supplementary Table S33.

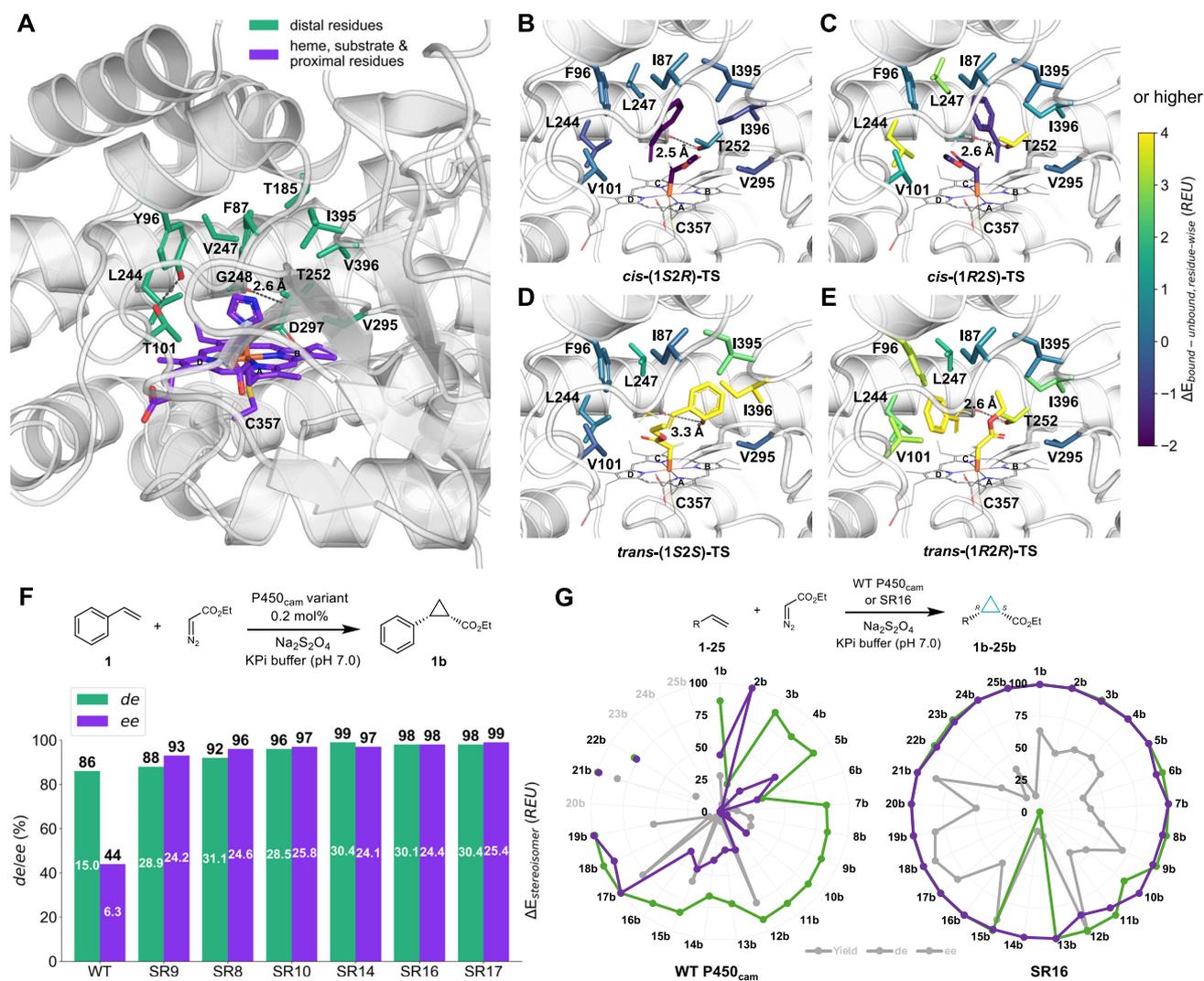
### Experimental characterization of designed *cis*-(1*S*,2*R*)-stereoselective cyclopropanases

Cyclopropanation reactions with styrene (1) and EDA showed that, while wildtype P450<sub>cam</sub> exhibits modest *cis*-(1*S*,2*R*)-stereoselectivity (86% *de*, 44% *ee*), as many as 33% (8/24) of the P450<sub>cam</sub> designs produced the desired *cis*-(1*S*,2*R*)-cyclopropane product 1b with excellent (1*S*,2*R*)-enantioselectivity (exceeding 96% *ee*) as well as good to excellent *cis*-diastereoselectivity (67–92% *de*; Supplementary Table S34) in whole-cell transformations. Among them, the six best-performing catalysts (called SR8, SR9, SR10, SR14, SR16, SR17) were selected for further characterization as purified proteins (Fig. 5F; Supplementary Table S35). Under these conditions, all of the six P450<sub>cam</sub> variants yielded the *cis*-(1*S*,2*R*)-cyclopropane product 1b with high to excellent *cis*-diastereoselectivity and (1*S*,2*R*)-enantioselectivity (88–99% *de* (AVG: 95% *de*); 93–99% *ee* (AVG: 97% *ee*); Supplementary Table S35), along with significantly higher activity compared to the wild-type enzyme (60–74% yield (64% AVG yield) vs. 28% yield).

Among the experimentally tested P450<sub>cam</sub> variants, SR16 (= P450<sub>cam</sub> (F87I, Y96F, T101V, V247L, D297T, V396I)) exhibited the highest yield and *cis*-(1*S*,2*R*)-stereoselectivity both in whole cells and as purified protein (Supplementary Tables S34 and S35). SR16 was also predicted to retain consistent *cis*-(1*S*,2*R*)-stereoselectivity across different structurally diverse substrates including *p*-CF<sub>3</sub>-substituted styrene (11) and several other non-styrenyl olefin substrates 16, 17, 18, 23, and 25 according to our computational models (Supplementary Table S36; Fig. S17). Therefore, the generality of SR16 against the panel of 25 structurally diverse olefin substrates was further assessed experimentally. Overall, and in stark contrast to wildtype P450<sub>cam</sub>, SR16 was found to show good to excellent *cis*-(1*S*,2*R*)-stereoselectivity (84–99% *de* (AVG: 98% *de*); 86–99% *ee* (AVG: 97% *ee*); 54% average yield) in cyclopropanations of a diverse panel of styrenyl substrates (1–16 except 14) including electron-rich and electrodeficient *alpha*, *para*, *meta* and *ortho*-substituted derivatives (Fig. 5G; Supplementary Fig. S18). As the only exception, moderate *trans*-diastereoselectivity (–22% *de*) was observed for 14b. Notably, SR16 exhibited high *cis*-(1*S*,2*R*)-stereoselectivity also across a panel of non-styrenyl olefin substrates (16–25), the majority of which were cyclopropanated with 99% *de* and *ee* (AVG: 99% *de*; AVG: 98% *ee*; 54% average yield). The latter include substrates such as aliphatic olefins (23, 24), vinylsuccinamide (20) and benzofuran (25), on which P450<sub>cam</sub> shows none to negligible activity (Fig. 5G; Supplementary Fig. S18). X-ray crystallography of 16b confirmed the absolute configuration of this compound as the expected isomer *cis*-(1*S*,2*R*) (Supplementary Tables S37, 38; Supplementary Fig. S10B). Overall, these results demonstrate that SR16 constitutes a generalist *cis*-(1*S*,2*R*)-stereoselective cyclopropanase.

### Virtual screening against the hemoprotein database leads to the identification of novel *cis*-(1*R*,2*S*)-stereoselective IDO1 cyclopropanases

Encouraged by the success with the *trans*-(1*R*,2*R*) and *cis*-(1*S*,2*R*)-stereoselective cyclopropanases, we next sought to assess the effectiveness of the strategy outlined in Fig. 1 in its more general format, namely toward designing a chosen (e.g., *cis*-(1*R*,2*S*)) stereoselectivity into a hemoprotein scaffold with previously unknown carbene transfer activity. Accordingly, we created a nonredundant set of 655 annotated hemoprotein structures from the Protein Databank (PDB), categorized by their heme types, structural properties, and biological functions. Starting from these structures, we performed Rosetta FastDesign as a rapid screen to evaluate the compatibility of each hemoprotein for



**Fig. 5 | Design of generalist *cis*-(1*S*,2*R*)-stereoselective P450<sub>cam</sub> biocatalysts.**

**A** Crystal structure of P450<sub>cam</sub> (PDB ID: 2H7Q). Rosetta models of SR16 (=P450<sub>cam</sub> (F87I, Y96F, T101V, V247L, D297T, V396I)) in complex with *cis*-(1*S*,2*R*)-TS (**B**), *cis*-(1*S*,2*R*)-TS (**C**), *trans*-(1*S*,2*S*)-TS (**D**), and *trans*-(1*R*,2*R*)-TS (**E**) of the cyclopropanation of **1** with EDA. Individual residue sidechains are colored according to their calculated per-residue binding energy  $\Delta E_{\text{bound}} - \text{unbound}$ , with darker colors representing lower energy (more stable) and lighter colors representing higher energy (less stable). **F** Experimentally measured *de* and *ee* values of the *cis*-(1*S*,2*R*)-cyclopropanation of **1** with EDA vs. Rosetta-calculated diastereomer and enantiomer energy differences for computationally designed P450<sub>cam</sub> variants. A higher

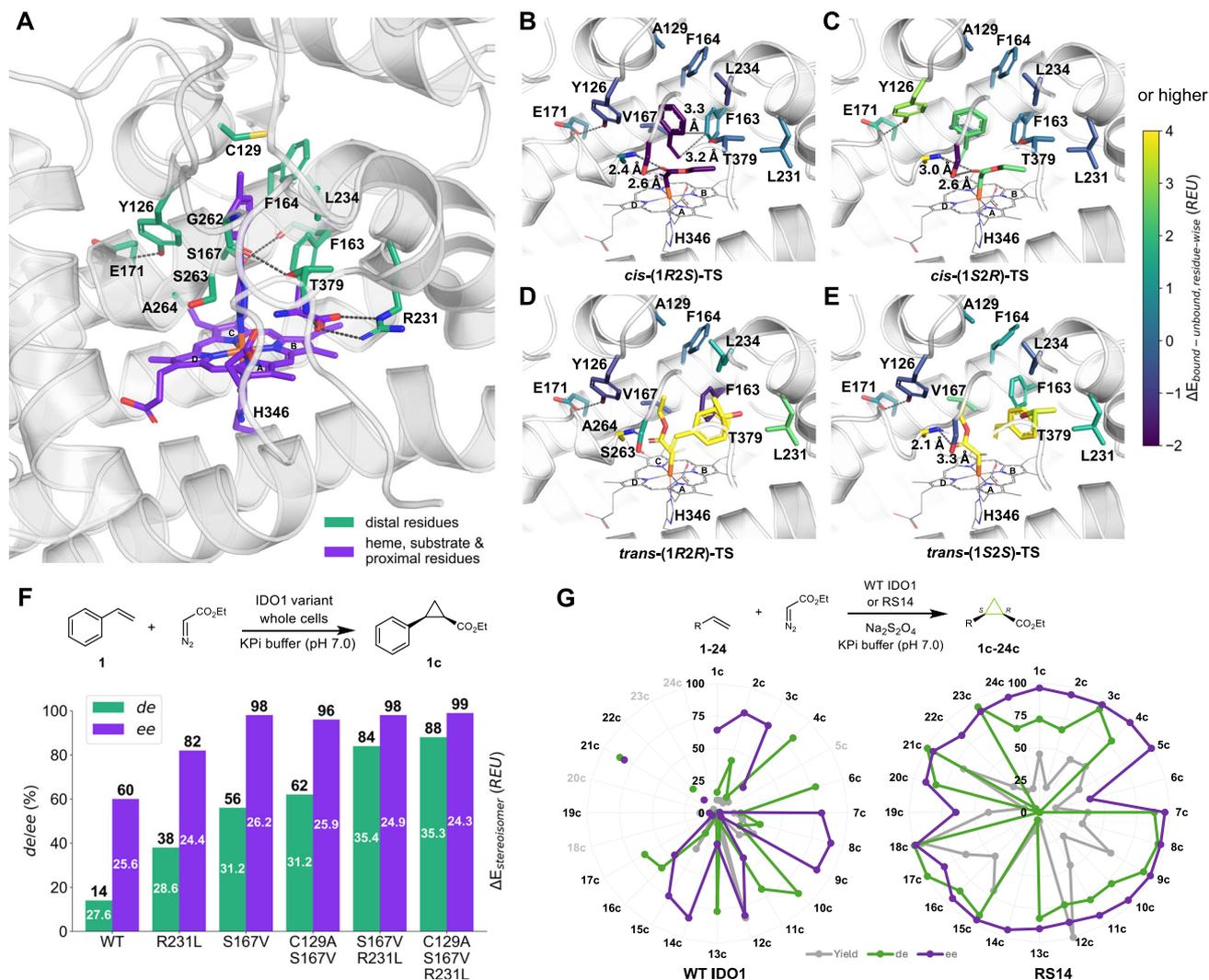
$\Delta E_{\text{stereoisomer}}$  value corresponds to predicted higher stereoselectivity. Reaction conditions: 20  $\mu\text{M}$  catalyst, 10 mM styrene (**1**), 20 mM EDA, 10 mM Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub>, in KPI buffer (50 mM, pH = 7.0), anaerobic, 16 h. **G** Substrate scope for olefin cyclopropanations with EDA using wild type P450<sub>cam</sub> and SR16. Yield, *de* and *ee* refer to *cis*-(1*S*,2*R*)-cyclopropane products **1b**-**25b**. Cyclopropanation products inaccessible to wild type P450<sub>cam</sub> are greyed out. Reaction conditions: 60  $\mu\text{M}$  catalyst, 2.5 mM olefin substrate, 20 mM EDA, 10 mM Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub>, in KPI buffer (50 mM, pH = 7.0), anaerobic, 4 h. Yields were determined by GC analysis using calibration curves generated with isolated products ( $n \geq 2$ ; SE < 10%). Diastereoselectivity and enantioselectivity were determined via chiral GC and SFC ( $n \geq 2$ ; SE < 3%).

binding to the target *cis*-(1*R*,2*S*)-TS within their distal cavity. Multiple globins, P450s, and other hemoproteins emerged as potential candidates for *cis*-(1*R*,2*S*)-stereoselective cyclopropanases (see Supplementary Data 2-3). Among these candidates, indoleamine 2,3-dioxygenase-1 (IDO1, Fig. 6A), a heme-containing enzyme involved in the oxidative catabolism of tryptophan to kynurenine<sup>47,48</sup>, was selected as the parent scaffold based on favorable active site interactions with the modeled *cis*-(1*R*,2*S*)-TS, and its novelty as a biocatalyst for abiotic reactions, including carbene transfer reactions.

In modeling studies with the wild-type IDO1, the carbene-borne carbonyl group of *cis*-(1*R*,2*S*)-TS was found to form favorable short hydrogen bonding interactions (below 2.7 Å) with two active site residues, namely the sidechain hydroxyl group of Ser263 and the backbone amide N-H atom of Ala264 (Fig. 6B). These hydrogen bonding interactions are either deformed, yielding suboptimal

(elongated or clashing) bond geometries, or entirely abolished in models of alternative TS stereoisomers bound to IDO1 (Fig. 6C-E). The competing enantiomeric *cis*-(1*S*,2*R*)-TS was predicted to be destabilized due to steric clashes between the styrene substrate and Ala264 (Fig. 6C; Supplementary Table S39). Iterations of the multi-state design protocol suggested that substituting Ser167 with I, T, V, C, or A and substituting Cys129 with alanine may further enhance *cis*-(1*R*,2*S*)-stereoselectivity (Supplementary Table S40; Supplementary Fig. S19), while the C129A substitution may facilitate accommodation of *para*-substituted styrene derivatives (Supplementary Table S41).

Based on these analyses, wild-type IDO1 and an initial set of double mutant IDO1 variants (i.e., C129A + S167I/T/V/C/A; called “RS1-RS5”) were prepared and experimentally tested in whole cells. As predicted by our modeling protocol, wild-type IDO1 was found to favor



**Fig. 6 | Design of generalist *cis*-(1*R*,2*S*)-stereoselective IDO1 biocatalysts.**

**A** Crystal structure of wild type IDO1 (PDB ID: 6FOA). Rosetta model of RS14 (= IDO1 (C129A, S167V, R231L)) in complex with *cis*-(1*R*,2*S*)-TS (**B**), *cis*-(1*S*,2*R*)-TS (**C**), *trans*-(1*R*,2*R*)-TS (**D**), and *trans*-(1*S*,2*S*)-TS (**E**) of the cyclopropanation of **1** with EDA. Individual residue sidechains are colored according to their calculated per-residue binding energy  $\Delta E_{\text{bound}} - \text{unbound}$ , with darker colors representing lower energy (more stable) and lighter colors representing higher energy (less stable).

**F** Experimentally measured *de* and *ee* values of the *cis*-(1*R*,2*S*)-cyclopropanation of **1** with EDA vs. Rosetta-calculated diastereomer and enantiomer energy differences for RS14 and variants bearing reversion mutation(s). A higher  $\Delta E_{\text{stereoisomer}}$  value

corresponds to predicted higher stereoselectivity. Reaction conditions: *E. coli* whole cells ( $OD_{600} = 20$ ), 10 mM styrene (**1**), 20 mM EDA, in KPi buffer (50 mM, pH = 7.0), anaerobic, 4 h. **G** Substrate scope for olefin cyclopropanations with wild type IDO1 vs. RS14. Yield, *de* and *ee* refer to *cis*-(1*R*,2*S*)-cyclopropane products **1c**-**24c**. Cyclopropanation products inaccessible to wild type IDO1 are greyed out. Reaction Conditions: 20  $\mu\text{M}$  catalyst, 2.5 mM olefin substrate, 20 mM EDA, 10 mM  $\text{Na}_2\text{S}_2\text{O}_4$ , in KPi buffer (50 mM, pH = 7.0), anaerobic, 4 h. Product yields were determined by GC analysis using calibration curves generated with isolated products ( $n \geq 2$ ;  $\text{SE} < 10\%$ ). Diastereoselectivity and enantioselectivity were determined via chiral GC and SFC ( $n \geq 2$ ;  $\text{SE} < 3\%$ ).

formation of the desired *cis*-(1*R*,2*S*)-cyclopropane product **1c** (60% *ee*) in the reaction with styrene and EDA, albeit with low *cis*-diastereoselectivity (14% *de*) and activity (1% yield). Of note, all of the designed double mutant IDO1 variants showed improved (1*R*,2*S*)-enantioselectivity (68–96% *ee* (AVG: 88% *ee*)), and four of them exhibited enhanced *cis*-diastereoselectivity (16–62% *de* (AVG: 40% *de*)) compared to the wild-type enzyme (Supplementary Table S42). Among them, IDO1 (C129A, S167V) (named “RS4”) displayed the highest values for both *de* and *ee* (62% *de*, 96% *ee*) as well as slightly higher activity (4% yield) compared to wild-type IDO1.

To further improve the catalytic activity and *cis*-diastereoselectivity of these IDO1-derived cyclopropanases, a second round of computational active site mutagenesis was carried out starting from the RSI-RSS series, which suggested multiple potentially beneficial mutations at the active site positions Ser263, Ala264, Gly265 and Thr379 (Supplementary Fig. S20). Among them, the R231L mutation

was found to lead to the highest fitness in the multi-state design procedure, due to its ability to stabilize the *cis*-(1*R*,2*S*)-TS (Fig. 6B) by removing suboptimal interactions of the Arg231 sidechain with the ethyl ester group of the carbene, as well as destabilizing the two *trans*-TS diastereomers (Fig. 6D–E). These predicted beneficial mutations were combined to generate ten second-generation designs (RS6 through RS15; Supplementary Table S40). In addition, an active site mutation F163G was introduced to alleviate close contacts between F163 and pro-*cis*-(1*R*,2*S*)-styrene (with two non-bonded C–C distances of 3.2 Å and 3.3 Å; Fig. 6B). Starting from IDO1 (F163G, S167C) and computationally predicted beneficial mutations at positions F164 and L234 (Supplementary Fig. S20), a second set of triple mutant IDO1 variants (variants RS16–RS21) were designed (Supplementary Table S40). A summary comparing the differences in key designed catalytic residues within the active-site pocket is provided in Supplementary Table S43.

In experimental tests, the majority of these second-generation IDO1 designs show greatly improved *cis*-diastereoselectivity (AVG: 45% *de* vs. 14% *de*) as well as (1*R*,2*S*)-enantioselectivity compared to wild-type IDO1 (AVG: 79% *ee* vs. 60% *ee*), with about half of them exhibiting 99% *ee* in formation of the target *cis*-(1*R*,2*S*) product 1c from the reaction with styrene and EDA (Supplementary Table S42). In addition, all of these highly (1*R*,2*S*)-enantioselective variants showed also greatly improved activity compared to IDO1 (32–46% vs. 10% yield). Among these variants, the triple mutant IDO1 (C129A, S167V, R231L), named RS14, exhibited the highest stereoselectivity in the model reaction with styrene and EDA (88% *de*, 99% *ee*) (Supplementary Table S42). In terms of structure-activity relationships, reversal of any of the three active site mutations in RS14 led to lower *cis*-(1*R*,2*S*)-stereoselectivity, highlighting their essential contributions towards high stereoselectivity (Fig. 6F; Supplementary Table S44). Computational modeling further suggested the consistent *cis*-(1*R*,2*S*)-stereopreference of RS14 for the cyclopropanation of the bulkier substrate *p*-CF<sub>3</sub>-styrene (11) and several other non-styrenyl, structurally diverse olefin substrates 16, 17, 18, 23, and 25 (Supplementary Table S45). Therefore, RS14 was selected as the most promising variant to serve as generalist, *cis*-(1*R*,2*S*)-stereoselective cyclopropanase.

Indeed, when challenged against a diverse panel of olefin substrates, RS14 exhibited cyclopropanation activity on 24 out of 25 substrates compared to only 19 for wild-type IDO1 (Fig. 6G; Supplementary Figs. S21, S22). In addition, for nearly all of the viable substrates (19/24) the designed enzyme produces the desired *cis*-(1*R*,2*S*)-cyclopropane products in good to high *cis*-diastereoselectivity (66–99% *de* (AVG: 85% *de*)) and (1*R*,2*S*)-enantioselectivity (64–98% *ee* (AVG: 94% *ee*); Fig. 6G). Of note, such reactivity and high *cis*-(1*R*,2*S*)-stereopreference extends well beyond styrenyl substrates 1–16 to include the diverse group of non-styrenyl and unactivated olefins 17–24. As an exception, benzofuran (25) was not accepted by the enzyme. Despite these limitations, and generally more modest catalytic activity compared to the other stereocomplementary catalysts described above, these results demonstrate the generalist nature and predictable *cis*-(1*R*,2*S*)-stereopreference of the IDO1-based biocatalyst RS14.

## Discussion

Focusing on hemoprotein-catalyzed olefin cyclopropanation as the target transformation, we introduced and validated here a mechanism-based computational design workflow useful for developing a suite of biocatalysts with high and complementary diastereo- and enantioselectivity across a broad range of substrates, a feature often exhibited by small molecule catalysts but rarely found in natural or engineered enzymes<sup>49</sup>. The central component of the present strategy (Fig. 1) is a multi-state design protocol through which the active site of a parent hemoprotein scaffold is (iteratively) optimized to accommodate and differentially stabilize the DFT-calculated TS models leading to the desired product stereoisomer, while destabilizing the competing TS stereoisomers. In the reaction targeted here (styrene cyclopropanation with EDA), this translated into performing iterative rounds of computational mutation scanning using a composite fitness function to identify mutations that stabilize one of the four possible TS stereoisomers, while destabilizing other three TS stereoisomers. The computational design process was extended to exemplify bulkier target substrates (i.e., *para*-trifluoromethyl-styrene vs. styrene) to ensure that the designed enzyme(s) can accept a broad range of substrates along with tailored stereoselectivity. The computational strategy was experimentally tested and validated across three design problems of increasing complexity: (i) an engineered myoglobin-based carbene transferase with good *trans*-(1*R*,2*R*)-stereoselectivity but very narrow substrate scope (i.e., RR5)<sup>17</sup> was re-designed to obtain a generalist cyclopropanase capable of processing a panel of structurally diverse olefin substrates with high and consistent *trans*-(1*R*,2*R*)-stereoselectivity (RR22; Fig. 4F); (ii) a heme enzyme (cytochrome P450<sub>cam</sub>)

featuring moderate *cis*-(1*S*,2*R*)-stereopreference along with narrow substrate scope in cyclopropanation, was redesigned to obtain a generalist *cis*-(1*S*,2*R*)-stereoselective cyclopropanase with high and consistent *cis*-(1*S*,2*R*)-stereoselectivity across a broad substrate scope (SR16; Fig. 5G); (iii) a generalist cyclopropanase with high *cis*-(1*R*,2*S*)-stereoselectivity (RS14; Fig. 6G) was obtained through the computational identification and optimization of a hitherto biocatalytically unutilized hemoprotein scaffold from the protein structure database. It is worth noting that the three protein scaffolds from which the stereodivergent cyclopropanases were derived (i.e., Mb, P450<sub>cam</sub>, IDO1), share minimal sequence and structural similarity (Supplementary Fig. S23, S24), highlighting the broad applicability and generalizability of our computational approach across distinct protein folds. The development of these effective catalysts required experimentally testing only a handful and progressively smaller number of variants as our methodology matured: 48 variants (in two rounds) were tested for the development of RR22, 24 variants (in a single round) for SR16, and 22 variants (in two rounds) for RS14, with design success rates ranging 33–50% (defined as > 60% *de* and > 80% *ee* in whole-cell cyclopropanation of styrene with EDA; reversions and other negative controls used for model validation are excluded from success rate estimates). Taken together, these results demonstrate the generality, effectiveness and resource-efficiency of the present strategy toward discovering and developing new biocatalysts with tailored stereoselectivity and broad substrate scope for a non-native transformation.

Computational protein design<sup>50</sup>, biomolecular deep learning<sup>51,52</sup>, and machine learning-guided directed evolution<sup>38,39,53</sup> have become increasingly important tools for the de novo design of enzymes<sup>54</sup> and guiding the engineering of extant enzymes for altered and improved activity and selectivity<sup>38</sup>. Despite this progress, accurately predicting enzyme interactions with transient species along catalytic reaction pathways, particularly in scenarios involving stereochemistry and conformational dynamics, remains challenging for purely data-driven methods, likely due to the limited availability of relevant training data. In the context of carbene transferases, computational modeling studies have been largely limited to retrospectively explaining experimental reactivities<sup>41–44</sup> or improving existing functions via deep learning<sup>38,39</sup>. While this work was in progress/review, computational design has been successfully applied for the de novo design of heme-containing enzymes displaying carbene transferase activity<sup>36,37,55</sup>, but none of these enzymes outperformed carbene transferases previously obtained by protein engineering/directed evolution. In addition, neither broad substrate scope nor a specific stereopreference was programmatically designed or optimized in these systems. Our approach, which utilizes physically realistic and exhaustive modeling of TS isomers and explicit multi-state design, can be readily extended for fine-tuning and further improving de novo designed hemoproteins to achieve desired biocatalytic functionalities.

Crystal structures of RR22 showed good agreement with the computational model, provided insights into the structural basis of active site pre-organization underlying improved stereoselectivity and broadened substrate scope. These structures revealed subtle, dynamic displacements of the CD loop and E helix upon substrate binding, suggesting a potential avenue for more quantitative stereoselectivity predictions. It is also noteworthy that, in our initial stereoselectivity prediction study, the only variant Mb(L29G,H64V) for which diastereoselectivity was incorrectly predicted contains a glycine residue. This is likely to introduce substantial backbone conformational changes that are not well captured in Rosetta modeling. Future improvements in computational design accuracy may come from sampling protein conformational ensembles using MD simulations and/or deep learning models to better capture protein flexibility and dynamics, thereby enabling more accurate predictions of mutational effects on enzyme–substrate complexes. A second avenue for improvement is more accurate scoring of the reaction energetics and the energy gaps

between competing stereoisomeric TSs in the context of the protein scaffold using advanced modeling methods such as quantum mechanics/molecular mechanics (QM/MM) simulations. Improved scoring and sampling may lead to more quantitative predictions of stereoselectivity and yields, compared to our current qualitative predictions. Regardless of the specific modeling practices employed, the overarching design philosophy of achieving an optimal balance between positive and negative design factors remains applicable for computational stereoselectivity design, and with further developments, should lead to general and improved protocols for computational design of stereodivergent biocatalysts in a 'zero-shot' manner.

From a synthetic standpoint, when combined with the previously reported *trans*-(1*S*,2*S*)-stereoselective Mb\*<sup>16</sup>, the computationally designed *trans*-(1*R*,2*R*), *cis*-(1*S*,2*R*), and *cis*-(1*R*,2*S*)-stereoselective biocatalysts developed here make available a comprehensive biocatalytic platform for obtaining all four possible stereoisomers of cyclopropanation products across a broad range of olefin substrates including electron-deficient and unactivated olefins. Although enantiodivergent variants with *trans*-(1*R*,2*R*)-stereoselectivity for this reaction have been also obtained via re-engineering of the Mb scaffold, a panel of five different variants was required to cover the same substrate panel of the generalist, *trans*-(1*S*,2*S*)-stereoselective Mb\* catalyst<sup>17</sup>. Similarly, the development of both general and highly stereoselective cyclopropanases for the formation of less thermodynamically favorable *cis*-cyclopropane products has proven difficult<sup>12,18,26,27</sup>. Stereodivergence in biocatalytic cyclopropanations was previously achieved in only a few (1-5) substrates after experimentally screening >40 different types of hemoproteins followed by multiple rounds of directed evolution<sup>18,26,27</sup>. Our methods also have the potential to aid directed evolution approaches by pinpointing stereoselectivity-determining residue positions for focusing library design. Access to cyclopropanation biocatalysts with consistent and predictable stereodivergent selectivity across a broad substrate scope (>20 substrates) is expected to render them powerful tools for medicinal chemistry and other synthetic applications, including the execution of detailed structure-activity relationships on bioactive molecules and/or the preparation of stereoisomeric small-molecule libraries for drug discovery campaigns<sup>1-3,6</sup>. Finally, we expect the present strategy to be readily extendable to other types of stereoselective reactions catalyzed by both de novo designed and naturally occurring metalloenzymes. As such, it holds promise for expediting the development of novel biocatalysts with high and tailored stereochemical control as well as broad substrate tolerance for various types of synthetic transformations.

## Methods

### Procedures for DFT Analysis of the Cyclopropanation Reaction Mechanism

Previous studies proposed that the hemoprotein-catalyzed styrene and ethyl diazoacetate (EDA) cyclopropanation involves a multi-step reaction mechanism<sup>43</sup>. First, EDA occupies the open coordination site of the heme iron and releases a nitrogen molecule (carbenoid-forming TS) to form a heme-bound carbenoid reaction intermediate (Supplementary Fig. S1). Geometry optimization and relaxed potential energy surface scanning were performed on the carbenoid intermediate. By scanning the Fe-C<sub>carbene</sub> bond, our calculation identified multiple differently orientated but stable carbenoid rotation states (noted by rot1, rot2, etc.) as flat local minima on the potential energy surface (Supplementary Fig. S2). Besides, the carbene-borne ester can adopt two energetically stable C<sub>carbene</sub>-C<sub>ester</sub> bond rotation states by flipping its carbonyl oxygen and the ethoxy group relative to the heme plane, with the  $\angle\text{FeC}_{\text{carbene}}\text{C}_{\text{ester}}\text{O}_{\text{carbonyl}}$  dihedral being either roughly +90° or -90° (noted by + or -, respectively) (Supplementary Fig. S3). The calculated rotational energy barriers of these two dihedrals are small, indicating that the carbenoid can interconvert rapidly between

different conformational isomers during the reaction process. According to the Curtin-Hammett principle, the rapid conformational interconversion of the carbenoid intermediate implies that the reaction stereoselectivity is solely influenced by the disparities in total energy among various cyclopropanation TS stereoisomers. In the following carbenoid-styrene addition step (cyclopropanation TS; Supplementary Fig. S1), the styrene can attack the sp<sup>2</sup> carbenoid carbon from either one of the two prochiral faces, leading to divergent *R/S*-configuration of this carbon atom in the final cyclopropane adduct. Besides, the styrene-borne benzyl group and the carbenoid-borne ester group can also be on the same side (*cis*) or separated by the cyclopropane plane (*trans*), resulting in the formation of the second chiral center (i.e., the substituted olefin carbon). In combination, these positional variations give rise to four possible configurations of the cyclopropane adduct, namely *trans*-(1*R*,2*R*), *trans*-(1*S*,2*S*), *cis*-(1*S*,2*R*), and *cis*-(1*R*,2*S*). Detailed DFT calculation settings are provided in the Supplementary Methods. Cartesian coordinates of DFT-optimized molecular structures are available in Supplementary Data 1.

### Rosetta-based multi-state modeling protocol

Crystal structures of the previously engineered sperm whale myoglobin-based biocatalyst Mb (H64V, V68A) (PDB ID: 6M8F), the wild-type cytochrome P450<sub>cam</sub> complexed with imidazole (PDB ID: 2H7Q), and the human wild-type indoleamine 2,3-dioxygenase 1 bound to a triazole inhibitor and alanine molecule (PDB ID: 6FOA) were used as initial structures for building Rosetta models. After stripping solvent and ligand molecules from crystal structures, the Rosetta FastRelax mover<sup>40</sup> was applied to pre-relax the entire protein. Then, DFT-generated TS models were superimposed onto the heme cofactor of pre-relaxed structures using the following stepwise procedure to generate initial hemoprotein-substrate complex structures. (i) The axial coordinating imidazole and the majority of the porphyrin ring were stripped from the DFT-generated TS to generate a truncated TS core consisting of only the olefin, the carbenoid, the iron center, and the four iron-coordinating sp<sup>2</sup> pyridine nitrogen atoms of the porphyrin. The iron center and iron-coordinating nitrogen atoms were set as virtual atoms in both.pdb files and.params files. (ii) The iron-coordinating nitrogen atom that has the closest-to-zero  $\angle\text{NFeC}_{\text{carbene}}\text{C}_{\text{ester}}$  dihedral was indexed as N1. The remaining three nitrogen atoms were indexed as N2, N3, and N4 in a counter-clockwise manner from the TS-binding side of the porphyrin plane. (iii) The virtual N1, N2, N3, and N4 atoms were superimposed onto nitrogen atoms of heme pyridine rings A, B, C, and D, respectively, to generate the rot1 conformer. By performing rotation operations on the heme-overlaid TS around the porphyrin C4 axis, virtual N1, N2, N3, and N4 atoms can be differently aligned to heme rings A, B, C, and D, resulting in total of four TS rotation states (denoted by rot1, rot2, rot3, and rot4, respectively). These were combined with the two carbenoid-borne ester rotational states (noted by + or -, respectively). In combination, for each of the four TS stereoisomers, eight different conformational isomers in complex with the hemoprotein were explicitly modeled in the subsequent multi-state modeling process. Using the Rosetta FastRelax protocol, amino acid substitutions were introduced into each substrate-docked hemoprotein structures along with the unbound hemoprotein and structural optimizations were performed on the mutation sites and their surrounding neighboring residues to resolve the structural perturbations associated with the introduced substitutions. Conformational sampling on the hemoprotein-bound TS was also performed to find the optimal substrate binding pose for each hemoprotein-bound TS state. All the rotatable bond dihedrals of the carbenoid and styrene were allowed to be sampled by the FastRelax mover except the two forming TS bonds between the alkene and the carbenoid. Dihedral constraints were applied to the iron-carbenoid Fe-C<sub>carbene</sub> bond ( $\angle\text{NFeC}_{\text{carbene}}\text{C}_{\text{ester}}$ ), the ester C<sub>ester</sub>-O<sub>ester</sub> bond ( $\angle\text{C}_{\text{carbene}}\text{C}_{\text{ester}}\text{O}_{\text{ester}}\text{C}_{\text{ethyl}}$ ), and the styrene C<sub>olefin</sub>-C<sub>benzyl</sub> bond within

the TS to appropriately reflect the rotational energy barriers obtained from DFT calculations (Supplementary Figs. S2, S3). A modified version of the Rosetta energy function “ref2015\_cst”<sup>56</sup> was used to better reflect the intramolecular interactions within ligands (weight of the score term “fa\_intra\_atr\_nonprotein” was set to 1.0 and “fa\_intra\_rep\_nonprotein” was set to 0.545, respectively). Coordinate constraints were applied to protein backbone atoms to prevent large deviations of atomic positions from the crystal structure. For each modeled structure, 50 independent Rosetta simulations were carried out to get the minimum energy value over all 50 trajectories. Since the model was partially repacked and minimized based on the pre-relaxed structure, several top trajectories would usually converge to a single minimum energy. The final energy score was calculated by subtracting the coordinate constraint score from the original total score. For a given hemoprotein biocatalyst, its diastereo- and enantioselectivity were predicted by calculating the energy difference between TS diastereomers and enantiomers in complex with the protein.

### Computational hemoprotein database screening protocol

For the first round of computational screening, a nonredundant set of 655 hemoprotein crystal structures were docked with eight differently aligned conformers of the *cis*-(1*R*,2*S*)-TS using the aforementioned procedure, generating a total of 5240 hemoprotein–TS complex structures. The Rosetta FastDesign protocol<sup>40</sup> was then applied to design the active site residues of these complexes. A residue type constraint of 3 REU was imposed to favor retention of native amino acid identities. Final energy scores were calculated by subtracting both the residue type and coordinate constraint penalties from the total Rosetta energy. Two quantitative metrics were used to assess the designability of each hemoprotein scaffold. The total energy change ( $\Delta\Delta G$ ), calculated as the difference between the total energy of the designed hemoprotein–TS complex and that of the unbound wild-type hemoprotein, was used as an indicator of scaffold designability. In addition, the Rosetta residue energy of the TS within the designed complex,  $E(\text{TS})$ , was used to evaluate the degree of substrate stabilization.  $\Delta\Delta G$  values and  $E(\text{TS})$  scores for all designs are summarized in Supplementary Data 2. Designed complexes satisfying the filtering criteria  $\Delta\Delta G < 0$  and  $E(\text{TS}) < 0$  were retained for further analysis. Sequences of the filtered hemoprotein–TS complexes were collected, and the corresponding mutations were reintroduced into the original wild-type hemoprotein structures complexed with *trans*-(1*R*,2*R*)-, *trans*-(1*S*,2*S*)-, *cis*-(1*R*,2*S*)- or *cis*-(1*S*,2*R*)-TS models using the Rosetta FastRelax protocol<sup>40</sup>. The resulting designed variants were ranked according to their calculated *cis*-(1*R*,2*S*)-stereoselectivity, which is summarized in Supplementary Data 3.

### Rosetta-based multi-state design (MSD) protocol

Starting from the Rosetta-generated multi-state models, computational mutation scanning was performed on selected active site residue positions. Using the Rosetta FastRelax protocol<sup>40</sup>, single amino acid substitutions were introduced into substrate-docked hemoprotein structures along with the unbound hemoprotein. Structural optimizations were performed on the mutation site and its surrounding neighboring residues as well as the substrate binding pose to resolve structural perturbations associated with the introduced point mutation. Each round of computational mutation scanning generated a position-specific scoring matrix calculated by using the multi-state design fitness function described in Supplementary Methods. High-ranking beneficial mutations emerged from computational mutation scanning were incorporated into the starting sequence to generate a combinatorial variant library. The newly generated sequences were computationally validated and filtered based on their calculated energies. These validated sequences were then used as starting points for subsequent rounds of computational sequence refinement, iterating until the MSD fitness of the designed sequences was maximized.

### General information for experimental procedures

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. <sup>1</sup>H and <sup>13</sup>C NMR spectra were measured on Bruker DPX-500 (operating at 500 MHz for <sup>1</sup>H and 125 MHz for <sup>13</sup>C) or Bruker DPX-400 (operating at 400 MHz for <sup>1</sup>H and 100 MHz for <sup>13</sup>C), <sup>19</sup>F was measured on Bruker DPX-400 (operating at 375 MHz). Tetramethylsilane (TMS) (0 ppm) and/or CDCl<sub>3</sub> (7.26 ppm) served as the internal standard for <sup>1</sup>H NMR, CDCl<sub>3</sub> was used as the internal standard (77.0 ppm) for <sup>13</sup>C NMR, and trifluorotoluene served as the internal standard (-63 ppm) for <sup>19</sup>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.

### Molecular cloning

pET22b(+) vector (Novagen) was used as the recipient plasmid vector for expression of all of the Mb, P450cam, and IDO1 variants. The gene of interest was prepared synthetically (Genscript) and cloned into pET22 vector with a C-terminal hexa-histidine tag and under the control of an IPTG-inducible T7 promoter. Site-directed mutagenesis was performed using a QuickChange mutagenesis protocol and oligonucleotides listed in Table S46 as primers. KOD Hot Start DNA polymerase from Merck was employed and chemically competent *E. coli* DH5 $\alpha$  cells were used for plasmid amplification.

### Protein expression and purification

Engineered purified Mb, P450cam, and IDO1 variants were expressed in *E. coli* C41(DE3) cells as follows. Briefly, cells were grown in TB medium (ampicillin, 100 mg/L) at 37 °C (170 rpm) until OD<sub>600</sub> reached 0.9–1.2. Cells were then induced with 0.3 mM  $\gamma$ -aminolevulinic acid (ALA) and 0.25 mM IPTG. After induction, cultures were shaken at 150 rpm at the corresponding temperatures (27 °C for Mb and IDO1 variants, 24 °C P450cam variants) and harvested after 18–20 h by centrifugation at 4000 rpm at 4 °C. After cell lysis by sonication, the proteins were purified by Ni-affinity chromatography. The 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 and 30 KDa Centricon filters. Protein concentration was determined using the following extinction coefficients: Myoglobin (Fe(III))  $\epsilon_{408} = 157 \text{ mM}^{-1} \text{ cm}^{-1}$ , P450cam (Fe(II)CO)  $\epsilon_{450} = 100 \text{ mM}^{-1} \text{ cm}^{-1}$ , IDO1 (Fe(III))  $\epsilon_{405} = 159 \text{ mM}^{-1} \text{ cm}^{-1}$ .

### Protein expression and purification for crystallization

Mb RR22 was cloned in His-tag-free form in the *Nde*I/*Xho*I cassette of a pET22 vector. Freshly transformed BL21(DE3) cells expressing Mb RR22 were used to inoculate 5 mL LB medium containing ampicillin (100 mg·L<sup>-1</sup>), followed by growth overnight at 37 °C with shaking (200 rpm). The overnight culture was used to inoculate 1 L of TB medium containing ampicillin (100 mg·L<sup>-1</sup>) followed by incubation at 37 °C (150 rpm) until OD<sub>600</sub> reached 1.0–1.2. Cells were then induced with 0.25 mM isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) and 0.3 mM  $\delta$ -aminolevulinic acid (ALA). After induction, cultures were shaken at 150 rpm and 27 °C and harvested after 18–20 h by centrifugation at 4000 rpm at 4 °C. Cell pellets were washed with ddH<sub>2</sub>O and resuspended in 5 mM KPi (pH 7.0).

After cell lysis by sonication and clarification of the lysate via centrifugation (14000 rpm, 4 °C, 40 min), the protein was purified using a two-step purification process. The first step was strong cation

exchange chromatography (SP Sepharose Fast Flow Resin) under a gradient of KPi at pH 7.0. 10 mM and 50 mM Kpi were used as wash buffer and elution buffer, respectively. After concentration using 10 kDa Centricon filters, protein was further purified via gel filtration chromatography using a Superdex 75 10/300 GL column (GE Healthcare). The protein was eluted under isocratic elution with 50 mM KPi (pH 7.0) at flow rate of 0.5 mL/min, following absorbance at 409 nm and 280 nm. The purity of collected fractions were confirmed by SDS-PAGE and the purified protein was buffer exchanged against 20 mM Tris-HCl buffer (pH 8.0) containing 0.1 mM EDTA, to a final concentration of 4 mM. In the case of RR22 bound to imidazole, the Tris-HCl buffer contained 0.1 mM EDTA and 10 mM.

### Protein crystallization

A crystal of the ferric Mb RR22 complexed with water was grown at room temperature using the hanging-drop vapor-diffusion method over a total reservoir volume of 1 mL, by mixing 1  $\mu$ L of reservoir buffer (2.3 M ammonium sulfate, 20 mM Tris-HCl, 0.1 mM EDTA, pH 8.8) with 1  $\mu$ L of protein in crystallization buffer (20 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). RR22 bound to imidazole was also crystallized under similar conditions. Crystals were formed within a week. The crystals were cryoprotected by soaking in a drop containing drop containing a 1:1 mix of Paratone and Silicone oil prior to being flash cooled in liquid nitrogen.

### Anaerobic reactions

Analytical reactions were conducted with purified Mb, P450cam, and IDO1 variants (20  $\mu$ M), 10 mM olefin substrate, 10 mM EDA and 10 mM sodium dithionite ( $\text{Na}_2\text{S}_2\text{O}_4$ ) using crimp vials producing a final volume of 400  $\mu$ L. In a typical procedure, the selected vessel containing the corresponding amount of purified biocatalyst were introduced to an anaerobic chamber. Then, a corresponding amount of degassed potassium phosphate buffer (KPi, 50 mM, pH 7.0) was added to the vessel producing a 20  $\mu$ M myoglobin solution followed by the addition of 40  $\mu$ L of a freshly prepared sodium dithionite solution (100 mM stock solution) in KPi (50 mM, pH 7.0). The reactions were initiated by the addition of 10  $\mu$ L olefin substrate (from a 400 mM stock solution in EtOH), 10  $\mu$ L EDA (from an 800 mM stock solution in EtOH). The vessels were capped and left under magnetic agitation for 3–16 h at room temperature. The reactions were then analyzed outside of the chamber following the Product Analysis protocol reported below.

### Product analysis

After completion of the desired reaction times, the vessel (crimp vials) of the reaction was open to air. The reactions were then analyzed by the addition of 20  $\mu$ L of internal standard (50 mM benzodioxole in EtOH) to the reaction mixture, followed by extraction with 400  $\mu$ L of  $\text{CH}_2\text{Cl}_2$ . After thorough mixing, the solutions were spun down at 14,000 rpm for 5 minutes. The organic layer was extracted via pipette, placed in a GC vial containing a glass insert, and capped tight. The GC vials were analyzed by chiral GC-FID using a Shimadzu GC-2010 gas chromatograph equipped with a FID detector, a chiral Cyclosil-B column (30 m  $\times$  0.25 mm  $\times$  0.25  $\mu$ m film) and HPLC. GC Separation Method 1: 1  $\mu$ L injection, injector temperature: 250  $^\circ\text{C}$ , detector temperature: 300  $^\circ\text{C}$ . Gradient: column temperature set at 140  $^\circ\text{C}$  for 3 min, then to 160  $^\circ\text{C}$  for 1.8  $^\circ\text{C}/\text{min}$ , then 165  $^\circ\text{C}$  for 1.0  $^\circ\text{C}/\text{min}$ , then 245  $^\circ\text{C}$  at 25  $^\circ\text{C}/\text{min}$  with a 6 min hold. Total run time: 28.3 min. GC Separation Method 2: 1  $\mu$ L injection, injector temperature: 320  $^\circ\text{C}$ , detector temperature: 320  $^\circ\text{C}$ . Gradient: column temperature set at 80  $^\circ\text{C}$  for 1 min, then to 320  $^\circ\text{C}$  at 20  $^\circ\text{C}/\text{min}$  with a 7 min hold. Total run time: 20.0 min. GC-FID calibration curves for quantification of the different cyclopropanation products were constructed with authentic standards prepared using our catalyst as described in Synthetic Procedures in the Supporting Information. All measurements were

performed at least in duplicate. For each experiment, negative control samples containing no protein were included.

### Reporting summary

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

### Data availability

Data supporting the findings of this study are included in this published article and provided in the Supplementary Information/Source Data file, and are available from the corresponding author(s) upon request. Protein crystal structures reported in this manuscript have been deposited in the Protein Data Bank (PDB) under accession codes 9PIE (RR22) and 9PIF (RR22-imidazole complex). Crystallographic data for the structures reported in this Article have been deposited at the Cambridge Crystallographic Data Center, under deposition numbers CCDC 2463873 (16a) and CCDC 2463872 (16b). Copies of the data can be obtained free of charge via <https://www.ccdc.cam.ac.uk/structures/>. Source data are provided with this paper.

### Code availability

An open-source implementation of the hemoprotein biocatalysts multi-state design protocol is available at <https://github.com/ZhuofanShen/Rosetta-Enzyme-Design-Pipeline>.

### References

1. Nguyen, L. A., He, H. & Pham-Huy, C. Chiral drugs: an overview. *Int. J. Biomed. Sci.* **2**, 85–100 (2006).
2. Mehvar, R., Brocks, D. R. & Vakily, M. Impact of stereoselectivity on the pharmacokinetics and pharmacodynamics of antiarrhythmic drugs. *Clin. Pharmacokinet.* **41**, 533–558 (2002).
3. Smith, S. W. Chiral toxicology: it's the same thing only different. *Toxicol. Sci.* **110**, 4–30 (2009).
4. Gerry, C. J., Wawer, M. J., Clemons, P. A. & Schreiber, S. L. DNA barcoding a complete matrix of stereoisomeric small molecules. *J. Am. Chem. Soc.* **141**, 10225–10235 (2019).
5. Bassi, G. et al. A single-stranded DNA-encoded chemical library based on a stereoisomeric scaffold enables ligand discovery by modular assembly of building blocks. *Adv. Sci.* **7**, 2001970 (2020).
6. McVicker, R. U. & O'Boyle, N. M. Chirality of new drug approvals (2013–2022): trends and perspectives. *J. Med. Chem.* **67**, 2305–2320 (2024).
7. Agranat, I., Caner, H. & Caldwell, J. Putting Chirality To Work: The Strategy Of Chiral Switches. *Nat. Rev. Drug Discov.* **1**, 753–768 (2002).
8. Yang, Y. & Arnold, F. H. Navigating the unnatural reaction space: directed evolution of heme proteins for selective carbene and nitrene transfer. *Acc. Chem. Res.* **54**, 1209–1225 (2021).
9. Talele, T. T. The "Cyclopropyl Fragment" is a versatile player that frequently appears in preclinical/clinical drug molecules. *J. Med. Chem.* **59**, 8712–8756 (2016).
10. Ma, S., Mandalapu, D., Wang, S. & Zhang, Q. Biosynthesis of cyclopropane in natural products. *Nat. Prod. Rep.* **39**, 926–945 (2022).
11. Lebel, H., Marcoux, J.-F., Molinaro, C. & Charette, A. B. Stereoselective cyclopropanation reactions. *Chem. Rev.* **103**, 977–1050 (2003).
12. 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).
13. Coelho, P. S. et al. A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins. *Nat. Chem. Biol.* **9**, 485–487 (2013).
14. Heel, T., McIntosh, J. A., Dodani, S. C., Meyerowitz, J. T. & Arnold, F. H. Non-natural olefin cyclopropanation catalyzed by diverse

- cytochrome P450s and other hemoproteins. *ChemBioChem* **15**, 2556–2562 (2014).
- Wang, Z. J. et al. Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran. *Angew. Chem. Int. Ed.* **53**, 6810–6813 (2014).
  - 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).
  - Bajaj, P., Sreenilayam, G., Tyagi, V. & Fasan, R. Gram-scale synthesis of chiral cyclopropane-containing drugs and drug precursors with engineered myoglobin catalysts featuring complementary stereo-selectivity. *Angew. Chem. Int. Ed.* **55**, 16110–16114 (2016).
  - Knight, A. M. et al. Diverse engineered heme proteins enable stereodivergent cyclopropanation of unactivated alkenes. *ACS Cent. Sci.* **4**, 372–377 (2018).
  - Siriboe, M. G., Vargas, D. A. & Fasan, R. Dehaloperoxidase catalyzed stereoselective synthesis of cyclopropanol esters. *J. Org. Chem.* **88**, 7630–7640 (2023).
  - Fu, W., Liu, A. & Yang, Y. Diastereo- and enantioselective chemoenzymatic synthesis of chiral tricyclic intermediate of anti-HIV drug lenacapavir. *ACS Catal* **15**, 2045–2052 (2025).
  - 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).
  - 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).
  - Kim, T. et al. Hemoprotein-catalyzed cyclopropanation en route to the chiral cyclopropanol fragment of grazoprevir. *ChemBioChem* **20**, 1129–1132 (2019).
  - Ren, X. et al. Highly stereoselective and enantiodivergent synthesis of cyclopropylphosphonates with engineered carbene transferases. *Chem. Sci.* **13**, 8550–8556 (2022).
  - Roy, S. et al. Stereodivergent synthesis of pyridyl cyclopropanes via enzymatic activation of pyridotriazoles. *J. Am. Chem. Soc.* **146**, 19673–19679 (2024).
  - Gober, J. G. et al. Mutating a highly conserved residue in diverse cytochrome P450s facilitates diastereoselective olefin cyclopropanation. *ChemBioChem* **17**, 394–397 (2016).
  - Schau, L. et al. Protoglobin-catalyzed formation of cis-trifluoromethyl-substituted cyclopropanes by carbene transfer. *Angew. Chem. Int. Ed.* **62**, e202208936 (2023).
  - 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).
  - Vargas, D. A., Khade, R. L., Zhang, Y. & Fasan, R. Biocatalytic strategy for highly diastereo- and enantioselective synthesis of 2,3-dihydrobenzofuran-based tricyclic scaffolds. *Angew. Chem. Int. Ed.* **58**, 10148–10152 (2019).
  - Carminati, D. M., Decaens, J., Couve-Bonnaire, S., Jubault, P. & Fasan, R. Biocatalytic strategy for the highly stereoselective synthesis of CHF<sub>2</sub>-containing trisubstituted cyclopropanes. *Angew. Chem. Int. Ed.* **60**, 7072–7076 (2021).
  - Siegel, J. B. et al. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. *Science* **329**, 309–313 (2010).
  - Dodani, S. C. et al. Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. *Nat. Chem.* **8**, 419–425 (2016).
  - Zhang, B. et al. In silico prediction of a multimutational stereoselective alcohol dehydrogenase. *ACS Catal* **15**, 16633–16642 (2025).
  - Leaver-Fay, A., Jacak, R., Stranges, P. B. & Kuhlman, B. A generic program for multistate protein design. *Plos One* **6**, e20937 (2011).
  - St-Jacques, A. D. et al. Computational remodeling of an enzyme conformational landscape for altered substrate selectivity. *Nat. Commun.* **14**, 6058 (2023).
  - Kalvet, I. et al. Design of heme enzymes with a tunable substrate binding pocket adjacent to an open metal coordination site. *J. Am. Chem. Soc.* **145**, 14307–14315 (2023).
  - Hou, K. P. et al. De novo design of porphyrin-containing proteins as efficient and stereoselective catalysts. *Science* **388**, 665–670 (2025).
  - Ding, K. R. et al. Machine learning-guided co-optimization of fitness and diversity facilitates combinatorial library design in enzyme engineering. *Nat. Commun.* **15**, 6392 (2024).
  - Yang, J. S. et al. Active learning-assisted directed evolution. *Nat. Commun.* **16**, 714 (2025).
  - Leaver-Fay, A. et al. in *Methods Enzymol.* Vol. 487 (eds Johnson, M. L. & Ludwig Brand) 545–574 (Academic Press, 2011).
  - Nam, D. et al. Enantioselective synthesis of  $\alpha$ -trifluoromethyl amines via biocatalytic N–H bond insertion with acceptor-acceptor carbene donors. *J. Am. Chem. Soc.* **144**, 2590–2602 (2022).
  - Couture, B. M. et al. Radical-mediated regiodivergent C(sp<sup>3</sup>)-H functionalization of N-substituted indolines via enzymatic carbene transfer. *Chem Catal* **4**, 101133 (2024).
  - 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).
  - Tinoco, A. et al. Origin of high stereocontrol in olefin cyclopropanation catalyzed by an engineered carbene transferase. *ACS Catal* **9**, 1514–1524 (2019).
  - Brandenberg, O. F. et al. Stereoselective enzymatic synthesis of heteroatom-substituted cyclopropanes. *ACS Catal* **8**, 2629–2634 (2018).
  - Guallar, V., Jarzecki, A. A., Friesner, R. A. & Spiro, T. G. Modeling of ligation-induced helix/loop displacements in myoglobin: toward an understanding of hemoglobin allostery. *J. Am. Chem. Soc.* **128**, 5427–5435 (2006).
  - Pallotta, M. T. et al. Indoleamine 2,3-Dioxygenase 1 (IDO1): an up-to-date overview of an eclectic immunoregulatory enzyme. *The FEBS Journal* **289**, 6099–6118 (2022).
  - Alexandre, J. A. C. et al. New 4-Amino-1,2,3-triazole inhibitors of indoleamine 2,3-dioxygenase form a long-lived complex with the enzyme and display exquisite cellular potency. *ChemBioChem* **19**, 552–561 (2018).
  - Zhu, C. D., Das, S., Guin, A., De, C. K. & List, B. Organocatalytic regio- and stereoselective cyclopropanation of olefins. *Nat. Catal.* **8**, 487–494 (2025).
  - Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed.* **52**, 5700–5725 (2013).
  - Abramson, J. et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* **630**, 493–500 (2024).
  - Krishna, R. et al. Generalized biomolecular modeling and design with RoseTTAFold all-atom. *Science* **384**, ead12528 (2024).
  - Yang, J., Li, F.-Z. & Arnold, F. H. Opportunities and challenges for machine learning-assisted enzyme engineering. *ACS Cent. Sci.* **10**, 226–241 (2024).
  - Lauko, A. et al. Computational design of serine hydrolases. *Science* **388**, eadu2454 (2025).
  - 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).

56. Alford, R. F. et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).

## Acknowledgements

This work was supported by the U.S. National Institute of Health Grants R01GM098628 and R35GM158365 (R.F.) and National Science Foundation grants CBET-1929256 (R.F.) and CBET 1929237 (S.D.K.). R.F. acknowledges chair endowment support from the Robert A. Welch Foundation (Chair, AT-0051-20221212). The authors are grateful to the UTD Center for High-Throughput Reaction Discovery & Synthesis supported by grant RR230018 from the Cancer Prevention and Research Institute of Texas, and Rutgers High Performance Computing resources provided by the Office of Information Technology. M.G.S. acknowledges support from the NIH Predoctoral Training Grants T32GM118283 and T32GM145461. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institute of General Medical Sciences (P30GM133894).

## Author contributions

Z.S., X.R., S.D.K. and R.F. conceived the project; Z.S. performed computational studies under the supervision of S.D.K.; X.R. carried out the bulk of the protein engineering studies, M.G.S. performed the bulk of the synthetic work and enzyme activity and selectivity characterization studies, and T.D. and J.L.J. crystallized and solved the crystal structure of RR22, under the supervision of R.F.; Z.S., M.S., S.D.K. and R.F. wrote the manuscript with input from all other authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-68327-1>.

**Correspondence** and requests for materials should be addressed to Sagar D. Khare or Rudi Fasan.

**Peer review information** *Nature Communications* thanks Yu-Fei Ao, Ying-wu Lin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2026