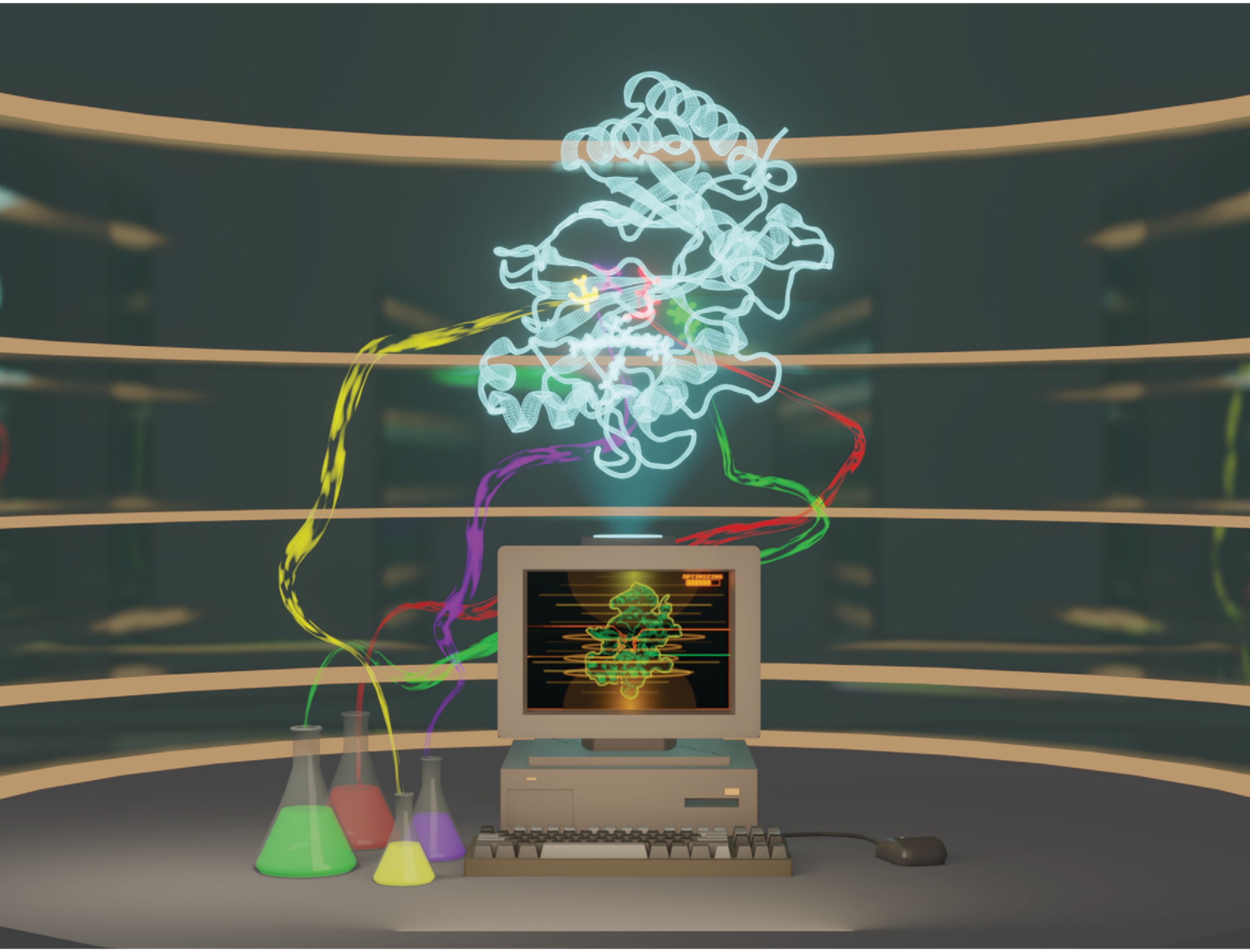
Volume 14
Number 18
21 September 2024
Pages 5117–5476

Catalysis Science & Technology

rsc.li/catalysis



ISSN 2044-4761

ROYAL SOCIETY OF **CHEMISTRY**

PAPER
Nicolai Lehnert *et al.*
Combined experimental and molecular dynamics approach towards a rational design of the YfeX biocatalyst for enhanced carbene transferase reactivity

Catalysis Science & Technology

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Catal. Sci. Technol.*, 2024, **14**, 5218

Combined experimental and molecular dynamics approach towards a rational design of the YfeX biocatalyst for enhanced carbene transferase reactivity†

Victor Sosa Alfaro, ‡ Hannah Palomino,‡ Sophia Y. Liu, Cybele Lemuh Njimoh and Nicolai Lehnert *

Received 26th October 2023,
Accepted 19th June 2024

DOI: 10.1039/d3cy01489d

rsc.li/catalysis

Pharmaceutical synthesis has been a driving force behind advancements in the field of biocatalysis. In this paper, we report the further optimization of the heme protein YfeX for regio- and stereoselective carbene transfer reactions using mutagenesis to explore the role of important amino acids in the active site for catalysis. In this way, we identified YfeX variants that are efficient and selective carbene transferases towards primary and secondary amines, olefins, and indoles. Molecular dynamic simulations reveal that mutations within the second coordination sphere induce distinct alterations in the conformation and electrostatic properties within the active site. These changes, in turn, affect substrate positioning both within the active site and at its entrance, which explains the distinct and sometimes surprising variations in selectivity observed experimentally between select YfeX variants. Our results show that the I230A single variant identified here is one of the most active N–H insertion catalysts known, producing >90% yields *in only 1 hour* (typical reaction times in the literature are 8–24 hours). On the other hand, the R232A variant catalyzes the C–H insertion of unprotected indole in up to 21% yield. The capacity to selectively act on unprotected indole offers a cost-effective, environmentally-friendly approach for late-stage functionalization of indoles and similar precursors in pharmaceuticals. In addition, YfeX is an efficient and fast biocatalyst that shows no structural degradation or heme loss during turnover, underscoring YfeX's robustness as a viable biocatalyst for both industrial and academic applications.

Introduction

Carbene transfer biocatalysis has become an important area of study with immense potential for the development of new industrial processes, especially in the pharmaceutical industry. The development of biocatalysts with carbene transfer reactivity has been inspired by heme-dependent cytochrome P450 monooxygenase enzymes (Cyt P450), which are metalloporphyrin catalysts that play a key role in fundamental oxidative transformations in biology.[1–4] Various model metalloporphyrin complexes have since been developed for their O-atom transfer capability in C–H bond activation and epoxidation reactions.[1,2,5] In addition, an influential paper by Arnold and coworkers demonstrated that Cyt P450$_{BM3}$ and other heme-containing proteins can actually catalyze the cyclopropanation of styrene using ethyl diazoacetate (EDA) as a carbene precursor.[6] Their work on directed evolution to engineer enzymes paved the way to discover novel reactions in the field of biocatalysts. Utilizing further engineering strategies, various groups have now established a variety of carbene insertion reactions, with optimized selectivity, and broadened the substrate scope of the reactions to generate high value products.[7–15] While carbene transfer biocatalysis has found applications in the synthesis of various compounds, including pharmaceuticals, natural product derivatives and fine chemicals, traditional rhodium and palladium small-molecule catalysts remain the industrially employed standard for these transformations.[16–18] However, for the synthesis of complex drug molecules such as beclabuvir and paritaprevir that contain many stereocenters, researchers are turning to biocatalysts, taking advantage of their protein scaffolds to fulfill the task of producing these drugs in a stereo- and enantioselective way.[17] Here, the chiral

---

Department of Chemistry and Department of Biophysics, University of Michigan, Ann Arbor, Michigan 48109-1055, USA. E-mail: lehnertn@umich.edu

† Electronic supplementary information (ESI) available: CD data for YfeX variants in buffer and with methanol added, further details about the MD simulations, GC/MS chromatograms and calibration data for product analysis. See DOI: **https://doi.org/10.1039/d3cy01489d**

‡ V. S. A. and H. P. are co-first authors.

environment of metalloenzymes provides the means to tune the selectivity of the reactions by orienting the substrate in the active site in a specific way that then determines selectivity, as demonstrated in many biosynthetic pathways in nature.[19–23]

Heme proteins are pertinent and attractive catalysts for group-transfer transformations such as carbene or nitrene transfer to generate hard to synthesize C–C or C–N bonds, respectively.[23] Recently, we have demonstrated that peroxidases, like YfeX, have great potential for the development of new biocatalysts.[24,25] YfeX is a novel heme protein, found in the extracellular matrix of *E. coli*, that has been shown to have native dye-decolorizing-type (DyP) peroxidase-like activity (Fig. 1).[26,27] DyPs are a superfamily of peroxidases that contain sequence similarities around 15%.[28] Structurally, they are usually comprised of two-domains, an α and β ferredoxin-like fold motif, that are distinct from other peroxidase superfamilies and that are organized in an α-helical fold (Fig. 1). In YfeX, the heme cofactor is buried in a large active site that is connected to the protein surface *via* a tunnel, which, in the native peroxidase reaction of YfeX, is proposed to guide hydrogen peroxide to the active site. The first DyP, identified relatively recently in the fungus *Geotrichum candidum*, was characterized with respect to its ability to catalyze the degradation of anthraquinone-based industrial dyes, and hence, this enzyme family was subsequently termed dye-decolorizing peroxidases.[29] The versatility of DyP-type peroxidases has attracted attention for their potential applications in bioremediation, lignin degradation, and other industrial processes.[30–32] In recent years, DyPs have also emerged as promising biocatalysts for carbene transfer reactions, due to their ability to generate reactive carbene species at their heme porphyrin active sites under mild reaction conditions.[24,25] However, the efficiency and selectivity of DyP-mediated carbene transfer reactions can be limited by their fundamental active-site design, optimized for their natural substrate specificity and catalytic activity.

We have previously shown that wild-type (WT) YfeX is a highly reactive and versatile catalyst for carbene transfer reactions.[24,25] A substrate scope study of WT YfeX for N–H insertion and cyclopropanation reactions using ethyl diazoacetate (EDA) revealed that WT YfeX is reactive towards a variety of *para*-substituted anilines and styrenes with electron-donating and electron-withdrawing substituents. However, WT YfeX favors electron-rich and polar substrates, and hence, produces lower yields with nonpolar substrates like styrene. In addition, WT YfeX shows high stability against organic solvents and the use of co-solvents such as methanol (MeOH) and dimethyl sulfoxide (DMSO) increases reactivity with nonpolar substrates such as styrene in cyclopropanation reactions.[25] To achieve higher product yields with nonpolar substrates, further insight into the effects of specific amino acid residues in the active site on catalysis is necessary. Rational enzyme design is a powerful approach that can be used to modify enzymes for specific applications.[9,13,33] By using computational and experimental methods to identify and optimize key amino acid residues in an enzyme and its substrate binding site, further considering its catalytic mechanism, it is possible to rationally engineer enzymes with enhanced or novel catalytic properties.

To further understand YfeX carbene transfer reactivity and selectivity,[25] and to elucidate the function of critical amino acids in the second coordination sphere (SCS) within the YfeX active site, we investigated several YfeX single variants in this study. Important amino acids

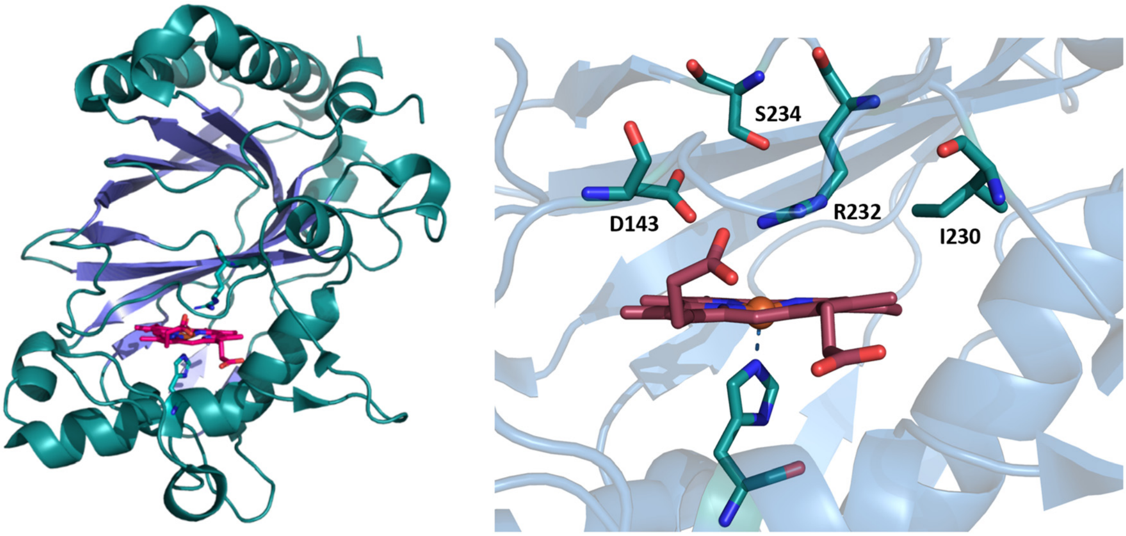


Fig. 1 PyMol representation of the tertiary structure of YfeX (left). The cartoon shows α-helices (dark green), β-sheets (blue), and the heme active site (red). Pymol image of the WT YfeX active site (right). The important second coordination sphere amino acids are highlighted. Images were generated with PDB code: 5GT2.

around the active site and the entrance channel were mutated, to probe for their roles during substrate binding and carbene transfer reactions. The overall mechanism of these types of reactions has been studied previously,[11,25,34–36] and it has been shown that after the initial formation of the iron-porphyrin carbene (IPC) intermediate, the next step during catalysis is substrate entering the active site. In the case of YfeX, the substrate enters by passing through a unique channel that is formed *via* a mobile loop. Once bound in the active site of YfeX, the substrate borders a hydrophobic pocket inside the active site, near the entrance channel, which could help guide nonpolar substrates towards the active site. Prominent polar amino acids that border the active site are R232, S234 and D143, which may contribute to positioning substrates in the active site, but also unfavorable interactions with nonpolar substrates (Fig. 1).[25] To probe these interactions, we modified these polar amino acids to alanine, a nonpolar, smaller amino acid. The resulting R232A, S234A, and D143A single variant YfeX proteins were then examined experimentally and computationally. In this regard, note that Pfanzagl *et al.* proposed that both D143 and R232 have an important role in stabilizing hydrogen peroxide, the natural substrate, and "Compound I" during the natural peroxidase reaction of the dye decolorizing peroxidase *Kp*DyP found in the pathogenic bacterium *Klebsiella pneumoniae*.[37] These polar amino acids and their interactions with substrate may affect carbene transfer reactivity. Finally, another key amino acid investigated here is I230, which has a bulky, non-polar side chain that is part of the entrance channel. Not only does adding an alanine increase the pocket size, but it also changes the electrostatic interactions by enhancing the hydrophobicity of the active site. Therefore, the R232A, S234A, and D143A and I230A variants could potentially increase substrate binding, and hence, increase reactivity. To test the catalytic performance of these four variants, we used a diverse set of substrates with varying steric and electronic attributes. Additionally, we tested the pharmaceutically relevant building block indole as a substrate.

These experimental studies are complemented by computational analysis, to better understand the global and local protein structure and dynamics in the variants, and in this way, help rationalize the experimentally observed, often surprising changes in product yields and selectivity. Using computational methods, protein stability in the variants was also analyzed. These results provide important guidelines for the ongoing development of YfeX variants with enhanced carbene transferase activity. In particular, the I230A variant identified here is one of the most active N–H insertion catalysts known, producing >90% yields *in only 1 hour*. Despite the propensity of YfeX to activate N–H bonds, we observe a strong preference for C–H bond activation with indole as the substrate, with the R232A variant catalyzing this reaction in up to 21% yield.

Methods

General procedures

All chemicals and reagents were purchased from commercial suppliers (Sigma-Aldrich, Fisher Scientific, Acros, Frontier Scientific) and used without further purification unless otherwise noted.

UV-visible spectra were recorded on an Analytic Jena Specord S600 spectrophotometer using sealed quartz cuvettes.

Gas chromatography/mass spectrometry (GC/MS) analyses were performed using a Shimadzu QP-2010 GC/MS equipped with a 30 m long DB-5 column with 0.25 mm ID. Separation method: 1 μL injection, injector temperature: 200 °C, detector temperature: 250 °C. Gradient: column temperature set to 60 °C for 3 min, then to 250 °C at 20 °C min^{-1} and held at 250 °C for 2.5 min (7.5 min for derivatives). Total run time was 15.00 min for the aniline and styrene reactions, and 20.00 min for the reactions with the aniline and styrene derivatives.

Supercritical fluid chromatography (SFC) was used to resolve the enantiomer products for the cyclopropanation reactions. SFC was performed using a Waters SFC instrument equipped with a column oven (35 °C), photodiode array detector, a backpressure regulator (12.0 MPa), and a carbon dioxide pump. The sample injection volume was 3 μL. A Daicel Chiralpak IC column (5 μm, 4.6 × 250 mm analytical chiral column) was used for separation of the enantiomers and % ee determination. All samples were eluted using an isocratic solvent system with the indicated modifier in liquid CO_2, containing 10% isopropanol as co-solvent, at an elution rate of 4 mL min^{-1} and detection at $\lambda = 250$ nm, with a total run time of 6.0 min.

YfeX mutants protein expression and purification

Protein expression and purification followed a protocol by the Weissenborn group (Martin-Luther-University Halle-Wittenberg in Halle (Saale), Germany).[24] The YfeX genes (pCA24N) were transformed into chemocompetent *E. coli* BL21 (DE3) cells by heat shock. Freshly plated transformants were grown overnight in 5 ml Terrific Broth (TB) medium containing 50 μg ml^{-1} chloramphenicol. 2 ml of the pre-cultures were inoculated in 400 ml TB autoinduction medium containing 50 μg ml^{-1} chloramphenicol. Cells were incubated at 37 °C and 120 rpm. After 4 h of cultivation, aqueous solutions of $FeCl_3$/5-aminolevulinic acid (final concentration: 100 μM) were added, the temperature was reduced to 30 °C and the cells were incubated for another 16.5 h. Cells were harvested by centrifugation (3000 × *g*, 20 min, 4 °C). The supernatant was discarded, and the pellets were resuspended in binding buffer (50 mM KPi, pH = 7.4, 200 mM NaCl). Cells were lysed by sonication (Bandelin Sonoplus HD3100: 6 × 30 s, 70% amplitude, pulse mode) and the lysate was incubated with hemin (final concentration: 600 μM) at room temperature for 30 min with a final DMSO concentration of 2%. The cell debris was removed by centrifugation for 45 min at 4 °C and 6000 × *g*. Excess of hemin and traces of

DMSO were removed during the protein purification steps. The proteins contain hexahistidine-tags and were purified by IMAC (immobilized metal ion affinity chromatography) using 1 mL His GraviTrap TALON columns (GE Healthcare Europe GmbH, Freiburg, DE). Washing buffer consisting of 50 mM Kpi (pH = 7.4), 200 mM NaCl and 5 mM imidazole was used to remove unspecific proteins. Protein was finally eluted by the addition of elution buffer (50 mM Kpi (pH = 7.4), 200 mM NaCl and 250 mM imidazole). PD-10 desalting columns (GE Healthcare Europe GmbH, Freiburg, DE) were used for buffer exchange of the pooled elution fractions, yielding the purified protein in 50 mM KPi (pH = 7.4) and 10% glycerol (v/v). These stocks were flash-frozen in liquid N_2 and stored at −20 °C. Protein and heme b amounts were determined in duplicates using the BSA protein assay and the pyridine hemochromagen assay (reduced form, $\varepsilon = 34.7\ \text{mM}^{-1}\ \text{cm}^{-1}$ at 557 nm).

Protein mass spectrometry analyses were performed using an Agilent Q-TOF HPLC-MS equipped with a Poroshell 300SB-C8 column. Separation method: 20 μL injection, flow rate: 0.5 mL min^{-1}, gradient: 95 : 5 water/acetonitrile for 3 min, followed by an increase over 10 min to 100% acetonitrile to elute the protein. Total run time was 13 min. Porphyrin mass spectrometry analyses were performed using an Agilent 6230 TOF HPLC-MS.

N–H insertion reactions

Reactions were performed in triplicates as previously described,[7,25] at a ~250 μL scale using 20 μM YfeX, 20 mM aniline (or aniline derivative), 40 mM ethyl diazoacetate (EDA), and 10 mM $Na_2S_2O_4$. The sodium dithionite (100 mM stock solution) in potassium phosphate buffer (100 mM, pH 7.0) was purged by bubbling nitrogen through the solution for 10 min in a sealed vial. All reagents and proteins were brought into a Coy vinyl anaerobic chamber (10–30 ppm of O_2, 1.5–3.0% H_2) before running the reactions. The as-isolated, ferric YfeX was first reduced to ferrous YfeX through addition of 40 μL of $Na_2S_2O_4$ solution, followed by the addition of 4 μL of aniline or aniline derivatives from a 1 M stock solution in methanol, and 8 μL of EDA (2 M stock solution in methanol), which translates to a total MeOH content of 2.65%. The reactions were left under magnetic agitation for 1 h at room temperature (or as indicated).

Cyclopropanation reactions

Reactions were performed in triplicates the same manner as the N–H insertion reactions, using 20 mM styrene (or styrene derivatives) instead of aniline. All 1 M styrene solutions were prepared in methanol.

Product analysis

The reactions were analyzed, as previously described,[25] by addition of 20 μL of internal standard (2-phenyl-ethanol, 1 M in methanol) to the reaction mixture, followed by extraction with 1.5 or 3 mL of ethyl acetate, depending on the reaction. The organic layers were dried with magnesium sulfate and later filtered and analyzed by GC/MS and SFC (see General procedures section for details on GC/MS and SFC analyses). Calibration curves for quantification of the N–H insertion products of aniline and aniline derivatives and the cyclopropanation products of styrene and styrene derivatives were constructed using authentic standards produced synthetically (using 1–2 mol% $Rh_2(OAc)_4$ as the catalyst[38]). These products were used to generate calibration curves for quantification of the aniline, styrene, and indole products in the catalytic reactions with YfeX. The reactions to produce authentic standards were conducted according to the same methods previously reported by our laboratory.[25] All yields and % selectivities reported in the tables in the Results section were obtained in triplicate. Negative control experiments were performed in the absence of YfeX.

Computational methods: system preparation and MD simulations

Molecular dynamics simulations were performed using the GPU code (pmemd) of the AMBER 22 package. The YfeX crystal structure available from the PDB (PDB code: 5GT2) was used as the starting point, in its monomeric form.[27] Parameters for the variants were generated using the same procedure as that for WT. Initial structures for the YfeX variants R232A, S234A, D143A and I230A were generated using the PyMOL mutagenesis tool. The protonation states of the ionizable side chains were accessed with Propka software,[39] while the axial His215 that is coordinated to the Fe center of the heme was assigned a protonation state based on visual inspection of its local environment. The active center parameters were generated using Metal Center Parameter Builder (MCPB.py),[40] as implemented in Amber18. Parameters for the iron-carbenoid and substrates were generated within the *antechamber* and MCPB.py modules in the AMBER22 package using the general AMBER force field (*gaff*). The bond and angular force constants were derived using the Seminario method,[41] while point charge parameters for the electrostatic potentials were obtained using the ChgModB method. Remaining protein residues were described using the Amber FF14SB force field. All the missing hydrogen atoms in the crystal structure were added with the Leap module in Amber and charges were neutralized using Na^+ counter ions. The system was solvated with TIP3P water molecules in a rectangular box within a distance of 10 Å from the protein surface.[42] Several MD and QM/MM studies on both heme and non-heme Fe-containing enzymes have successfully used the parameters generated *via* this procedure to study both the dynamics and the catalytic mechanisms of Fe-containing proteins.[11,34,43,44] A two-stage minimization of the geometries was first performed using MM to eliminate clashes of atoms. The first stage minimizes the positions of solvent molecules and ions, while positional restraints were imposed on the solute molecules by a harmonic potential of 500 kcal mol^{-1} Å^{-2}. The second stage minimizes all the atoms without any restraints. Minimization was performed using the CPU version of the Sander in Amber by subjecting the system to 5000 steps of steepest descent, followed by 5000 steps of conjugate gradient energy minimization. The minimized system was first slowly heated by restraining the solute molecules (harmonic potential of 50 kcal mol^{-1} Å^{-2}) from 0 to 300 K for 50 ps in an NVT ensemble using a Langevin thermostat.[45] To achieve uniform density, the heated system was further simulated with a weak restraint on the solute molecules at constant temperature of 300 K for 1 ns in an NPT ensemble. After that, the system was equilibrated for 3 ns in an NPT ensemble at a fixed temperature and pressure of 300 K and 1 bar, respectively, without any restraints on solute molecules. After equilibration, unrestrained production trajectories were run under the NVT ensemble and periodic boundary conditions for 1000 ns (1 μs). Unrestrained MD simulations with the iron-carbenoid intermediate were set up for the WT, I230A, and R232A variants by first selecting the most populated clusters (based on backbone clustering analysis using cpptraj) obtained from holo MD simulations as starting points to build the iron-carbenoid bound structures. Each iron-carbenoid bound system was run for 1000 ns. Unrestrained MD simulations of WT enzyme with styrene involved styrene placement in the active site through visual inspection, followed by the standard pre-production processing steps mentioned above with the exception of a distance restraint that was added between the center of mass of the styrene C–C double bond and the central C1 atom of the iron-carbenoid (3.0–3.2 Å), which was defined by adding a harmonic potential with $k = 100$ kcal mol^{-1} Å^{-2} to this coordinate during the equilibration. Trajectories were processed and analyzed using the cpptraj module from Ambertools utilities. All structures used for analysis are central structures of the most presented clusters.

Results and analysis

Earlier observations with the YfeX R232A variant demonstrated increased reactivity in N–H insertion as well as shifts in substrate selectivity patterns. However, explaining these outcomes, including this variant's vulnerability to heme loss, proved challenging.[25] To further this investigation, we extended our research to YfeX variants R232A, S234A, and D143A, to discern the effects of modifying key polar amino acids in the active site to alanine, a smaller, nonpolar amino acid, on the carbene transfer reactivity. Additionally, we studied the YfeX variant I230A, seeking to understand how diminishing the steric hindrance at the entrance of the active site influences reactivity. Substrates were analyzed across a range of steric, polarity, and functional group profiles to facilitate an understanding of the contrasting reactivity pattern between wild-type (WT) YfeX and its variants. To contextualize our findings, we utilized molecular dynamics (MD) techniques to examine the impact of each mutation on the YfeX protein structure and dynamics. Analyzing the global structural dynamics allowed us to gauge how the mutations alter the overall stability of the protein. Concurrently, local structural dynamics analyses allowed us to observe the effects of the amino acid substitutions on the structure and electrostatic properties of the active site. This comprehensive approach allowed us to rationalize our experimental findings and strategically guide experimental design.

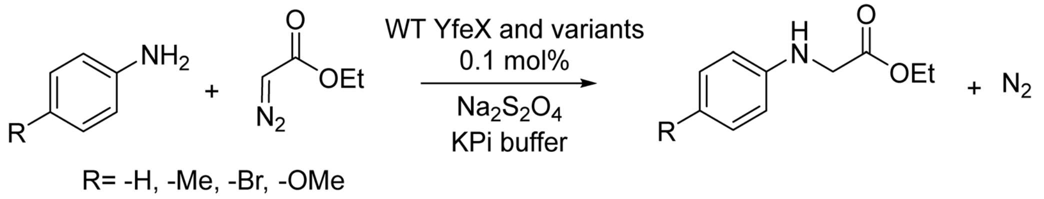
1. N–H insertion reactivity of YfeX variants

Active site engineering: the S234 residue is positioned above the heme in the active site, 7.6 Å away from the iron center. The removal of this serine near the active site may alter the hydrogen bonding network within the active site and could potentially play an important role in substrate binding. However, the N–H insertion with aniline and aniline derivatives for YfeX S234A is counterproductive: the reactivity for aniline is decreased, giving only up to 54% yield, and other *para*-substituted aniline substrates have moderate yields as well. There is also no drastic difference in selectivity between *ortho*-, *meta*-, and *para*-toluidine compared to WT YfeX (see Tables S2† and 1). These results show that Ser234 plays a subtle role in reactivity, and its removal is more deleterious than productive.

Active site engineering: the D143 residue sits right above the heme, 8.2 Å away from the Fe center, and might restrict substrate mobility in the active site. We therefore replaced D143 with alanine, to investigate its effect on carbene transfer catalysis. The N–H insertion with aniline and aniline derivatives for YfeX D143A reveals interesting results: the reactivity for aniline decreases compared to WT YfeX, giving up to 50% yield. However, *para*-substituted aniline substrates have relative high yields, especially seen with 4-bromo-aniline, generating the highest yield among WT and all variants studied here (see Tables S3† and 1). There is also an interesting difference in selectivity between *ortho*-, *meta*- and *para*-toluidine compared to WT YfeX. It was shown that *para*-toluidine is the preferred substrate in WT YfeX, whereas for the D143A variant, the *meta*-substituted toluidine derivative generates the highest yield (88%). For D143A, *para*-toluidine gives the lowest yield (54%) amongst the three derivatives. These results show that D143 plays an important role in substrate orientation in the active site, and that removal of this group generally improves yields for substituted anilines. The D143A variant therefore shows potential for the use of larger substrates, as generally required for the synthesis of drug precursors.

Active site entrance engineering: the I230 residue. Initial investigations into altering the active site entrance of YfeX focused on mutating I230, a bulky, nonpolar amino acid that sits in the entrance to the active site, to the smaller, nonpolar amino acid alanine. We hypothesized that the I230A variant would be able to accommodate larger substrates as a result of decreased steric hindrance in the channel while maintaining nonpolar character. Excitingly, the I230A variant shows a significant increase in yield for the N–H insertion reaction with

Table 1 Results for the N–H insertion reaction of anilines with WT YfeX and variants[a]



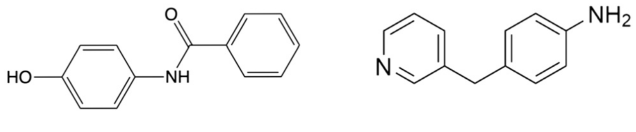
Substrate	WT[25]	R232A[25]	I230A	S234A	D143A
Aniline	72.4 ± 3.2	75.6 ± 6.4	92.2 ± 8.7	55.7 ± 6.2	49.8 ± 1.5
4-Bromoaniline	40.4 ± 6.2	43.4 ± 8.7	36.7 ± 4.7	53.4 ± 6.9	84.1 ± 7.0
p-Anisidine	57.0 ± 8.7	54.4 ± 2.3	45.8 ± 1.7	47.9 ± 3.1	40
o-Toluidine	27.9 ± 3.0	50.2 ± 0.8	73.7 ± 4.9	61.5 ± 1.8	78.7 ± 2.3
m-Toluidine	37.4 ± 1.9	51.2 ± 9.6	93.5 ± 1.7	60.4 ± 3.1	88.1 ± 4.2
p-Toluidine	57.1 ± 1.9	40.8 ± 0.4	44.7 ± 1.9	42.9 ± 1.0	54.0 ± 3.1
4-Trifluoromethyl aniline	51.0 ± 2.5	50.5 ± 6.7	—	—	76.4 ± 5.2
N-(4-Hydroxyphenyl)benzamide	70.6 ± 2.9	51.3 ± 2.0	69.4 ± 6.7	—	—
4-(4-Pyridylmethyl) aniline	63.9 ± 8.3	60.5 ± 0.4	59.3 ± 11.7	—	—

[a] Reaction conditions: 20 μM YfeX variant (0.1 mol%), 20 mM aniline and derivatives, 40 mM EDA, 10 mM dithionite, 1 hour reaction time, and reactions were performed in triplicates. Yields are based on GC/MS analysis (see ESI† for original data). Based on studies by Fasan *et al.*, Fe(PpIX) (free heme) is reported to have a yield for N–H insertion with aniline of 18%.[38]

aniline, 92%, compared to WT enzyme. Additionally, to demonstrate the viability of the I230A variant for scale-up, we increased the scale of the reaction with aniline to 3 ml and isolated the desired product. We were able to isolate a total of 22.6 mg of product, which corresponds to 99.9% yield (after 16 h) and illustrates how exceptional this variant is for future applications in N–H insertion reactions. This variant also shows increased yields for aniline derivatives (see Tables S4† and 1). Interestingly, there is a difference in selectivity between *ortho*-, *meta*-, and *para*-toluidine compared to WT YfeX and R232A. Previously, it was shown that *para*-toluidine is the preferred substrate in WT YfeX, whereas, surprisingly, toluidine selectivity is low in R232A. In contrast, with the I230A mutation, the *meta*-toluidine derivative now generates the highest yield (93%) for N–H insertion, whereas *para*-toluidine gives the lowest yield (44%) among the three derivatives. As evident from Table 1, the I230A variant discriminates strongly against *para*-substituted anilines. Finally, we also used two larger substrates, see Scheme 1, which also give good yields with YfeX and variants.

2. Functionalizing indoles

To expand on WT, R232A, and I230A YfeX's high N–H insertion reactivity, we conducted reactions using indole, a prevalent synthon in medicinal chemistry.[46] Previous studies by Fasan and coworkers have demonstrated the selective C–H functionalization of indoles by myoglobin variants through the formation of a carbene intermediate with EDA.[47] Additionally, the Weissenborn group has shown that WT, I230A, and R232A YfeX exhibit reactivity for C–H insertion using protected *N*-methylindole in combination with the more reactive carbene precursor diazoacetonitrile.[48] They also showed reactivity with WT and I230A YfeX using *N*-methylindole and EDA, as well as I230A YfeX with unprotected indole and diazoacetonitrile. All of these experiments showed relatively low reactivity (yielding ≤3%), leading us to test the unprotected indole with EDA. Interestingly, despite the availability of the N–H site in indole, we observed that WT YfeX exhibits low reactivity with indole and EDA, resulting in low yields of carbene insertion products (∼1%; see Table 2). Furthermore, the I230A and R232 variants show a preference for C–H insertion at the C3 position under standard reaction conditions. Excitingly, the R232A variant reaches 12% yield for this reaction. To further investigate the reactivity of the R232A variant, we examined its reactivity with other indole derivatives, an electron withdrawing indole (5-bromoindole) and an electron donating indole (5-methylindole). Interestingly we observed 21% yield for the 5-methylindole, but no quantifiable yield for the 5-bromoindole. Although we did not observe the single N–H insertion product, we did detect small amounts of the N–H and C–H double insertion product for the R232A variant. Collectively, we see that YfeX exhibits a preference to catalyze C–H insertion over N–H insertion of indole with less reactive diazo compounds like EDA, adding to its unique reactivity profile.



N-(4-Hydroxyphenyl)benzamide

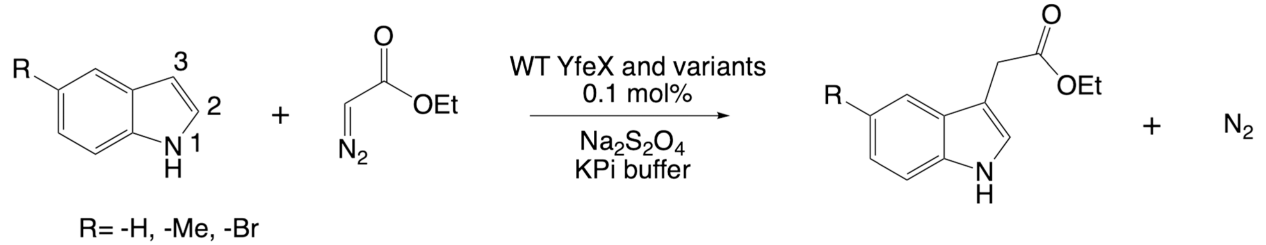
4-(4-Pyridylmethyl) aniline

Scheme 1

3. Cyclopropanation reactivity of YfeX variants

Active site engineering: the S234 residue. Cyclopropanation reactions with YfeX S234A show no

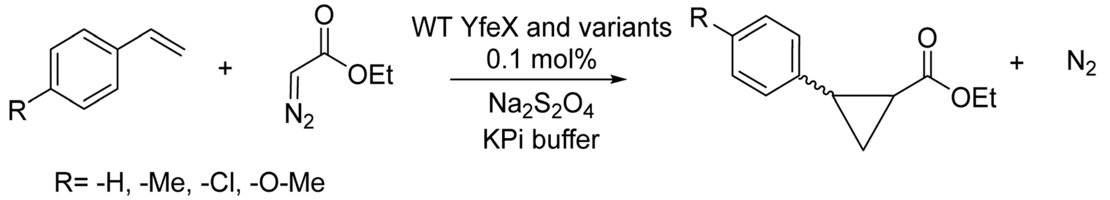
Table 2 Results for the C–H insertion reaction of indole with WT YfeX, and R232A and I230A variants[a]



Substrate	WT	R232A	I230A
Indole	1.4 ± 0.04	12.0 ± 0.01	1.6 ± 1.3
5-Bromoindole	—	<1	—
5-Methylindole	—	21.2 ± 2.1	—

[a] Reaction conditions: 20 μM YfeX variants (0.1 mol%), 20 mM indole, 40 mM EDA, 10 mM dithionite, 1 hour reaction time, and reactions were performed in triplicates. Yields are based on GC/MS analysis and calculated as (total product concentration)/(limiting reagent concentration). Based on studies by Fasan *et al.*, Fe(PpIX) (free heme) is reported to catalyze the reaction with indole with <1% yield.[47]

Table 3 Results for the cyclopropanation reactivity of styrenes with WT YfeX and variants



Substrate	WT[25]	R232A[25]	I230A	S234A	D143A
Styrene[a]	26.8 ± 7.2	11.9 ± 1.2	12.8 ± 1.4	8.5 ± 0.7	7.1 ± 1.6
Styrene[b]	47.4 ± 1.5	8.9 ± 1.2	7.0 ± 0.8	9.6 ± 0.9	—
4-Methylstyrene	48.8 ± 3.0	4.1 ± 0.5	5.3 ± 0.2	4.1 ± 0.4	—
Vinyl anisole	29.7 ± 8.4	4.6 ± 1.0	6.7 ± 0.8	5.4 ± 0.3	—
4-Chlorostyrene	29.3 ± 3.0	5.0 ± 0.2	5.9 ± 1.0	3.3 ± 0.1	—

[a] Reaction conditions: 20 μM WT YfeX and variants (0.1 mol%), 20 mM styrene and derivatives, 40 mM EDA, 10 mM dithionite, 1 hour reaction time, and reactions were performed in triplicates. Yields are based on GC/MS analysis. [b] Reactions were run in 30% methanol. Based on studies by Fasan *et al.*, Fe(PpIX) (free heme) is reported to catalyze the cyclopropanation of styrene with a yield of 13%.[35]

enhanced reactivity compared to WT enzyme, but in fact, a decrease of the yields is actually observed (see Tables S5† and 3). To improve styrene solubility in buffer, 30% methanol co-solvent was used, similar to WT YfeX, but this approach did not improve catalysis with the S234A variant. Circular dichroism (CD) data show that there is a difference in protein folding for YfeX S234A, with an increased β-sheet contribution to the folding (Fig. S1†). Hence, despite S234's nonpolar nature and moderate steric profile, substitution of this amino acid causes structural changes around the active site that are disadvantageous for catalysis, potentially by freeing up space on top of the active site, thus allowing catalytically unproductive substrate orientation in the active site. The loss of preference for the toluidine substrates in the S234A variant compared to WT enzyme supports these ideas.

Active site engineering: the D143 residue. Cyclopropanation reactions with YfeX D143A show no enhanced reactivity compared to WT YfeX, but a decrease of the yields is again observed (see Tables S6† and 3). The CD data show that there is a small difference in protein folding for YfeX D143A, but not as drastic as observed for the variants R232A, I230A, and S234A. Overall, there does seem to be an increased β-sheet contribution to protein folding (see Fig. S2†). These results show that YfeX D143A has only minor structural changes compared to WT, and D143 is therefore a good target amino acid for further engineering moving forward. In addition, the positively-charged active site of this variant (see MD results below) might help with the binding of anilines, especially those with electron withdrawing substituents.

Active site engineering: the I230 residue. Opposite to the high N–H insertion reactivity, cyclopropanation reactions with YfeX I230A show no enhanced reactivity compared to WT YfeX, but in fact, a slight decrease of the yields is observed (see Tables S7† and 3). To help increase styrene solubility in buffer, 30% methanol co-solvent was then used. This approach significantly enhances WT YfeX cyclopropanation reactivity,[25] but unfortunately, this method did not improve catalysis with YfeX I230A, similar to what is seen for the other variants. The CD data show that there is a difference in protein folding between WT and YfeX I230A, with a slightly increased β-sheet contribution to the fold in the I230A variant (Fig. S3†). To investigate if the putative decrease in stability of I230A YfeX in the presence of 30% methanol affects carbene transfer catalysis, we performed the N–H insertion reaction with aniline under these conditions as well. Surprisingly, we observed no significant decrease in YfeX activity for the N–H insertion of aniline. Furthermore, we observed an increase in the formation of the double N–H insertion product (Table S4†). Hence, in the presence of higher concentrations of methanol co-solvent, I230A YfeX becomes less selective for the single N–H insertion product, likely because methanol assists in allowing the less polar, single insertion product to enter the active site and react again. *This result therefore illustrates that the lack of styrene reactivity of YfeX I230A and other variants, in buffer with and without methanol co-solvent, is an intrinsic property of the YfeX protein that must somehow relate to the active-site properties of the enzyme*, and not so much to the lack of styrene access to the active site. This aspect was further analyzed computationally (see below).

To further investigate the lower cyclopropanation yields observed in YfeX and the YfeX variants (I230A and R232A), we conducted a study altering the EDA to styrene ratio in the reactions. Traditionally, an excess of EDA is often used in carbene transfer catalysis; however, by modifying this ratio, we observed a noteworthy increase in cyclopropanation reactivity in the I230A variant, with styrene to EDA ratios of 1 : 2. For WT protein, a 1 : 3 styrene to EDA ratio improves yields. In contrast, we did not observe a significant difference in reaction yields with altered ratios of EDA and styrene in

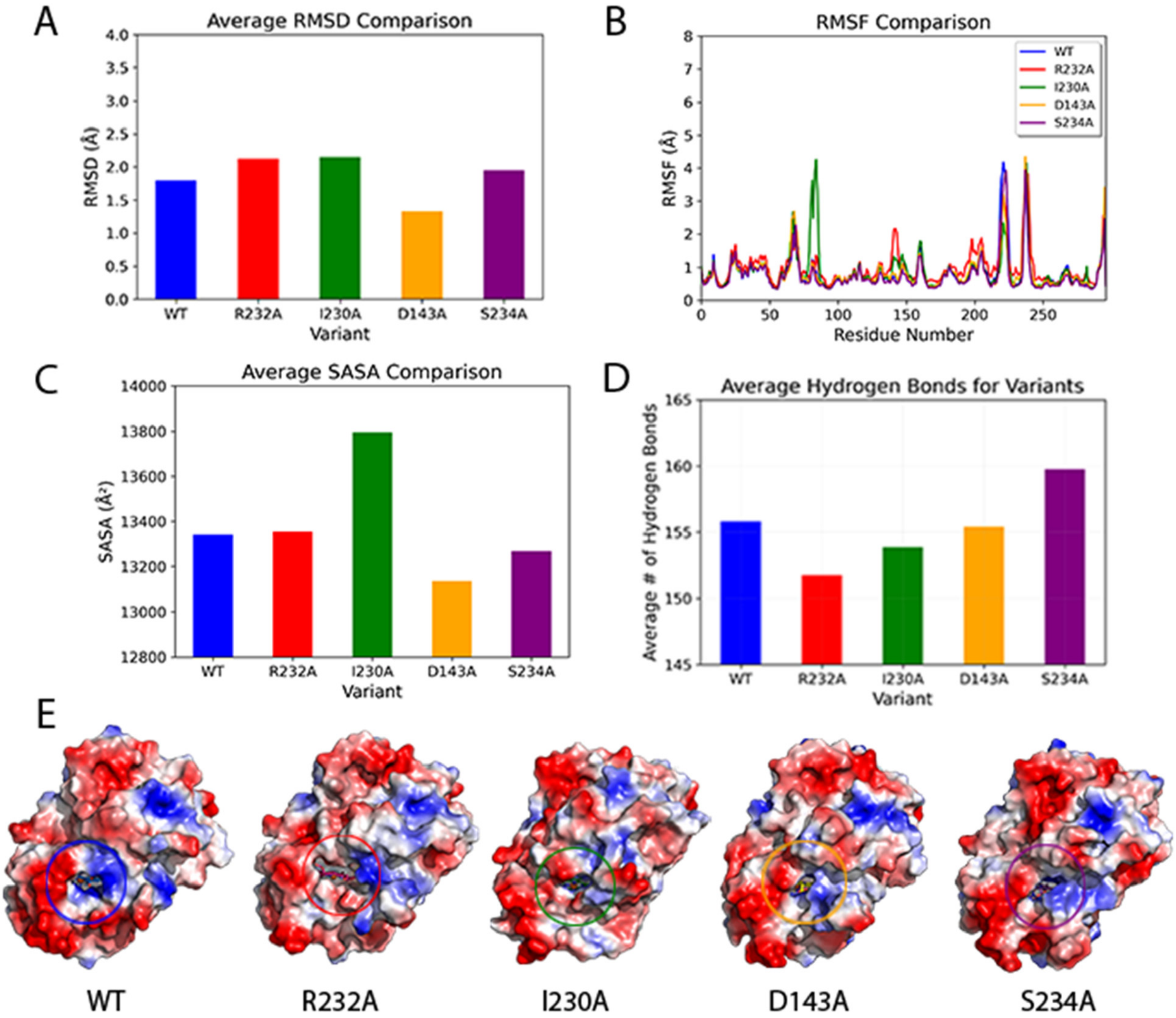


Fig. 2 Global molecular dynamics (MD) analysis from 1 μs MD simulations of WT YfeX and the R232A, I230A, D143A and S234A variants. (A) The average RMSD values for the MD simulations. (B) The RMSF comparison for different dynamic regions of the proteins. (C) The average solvent accessible surface area (SASA) comparison of the overall proteins. (D) The comparison of the average number of total hydrogen bonds present in the protein. (E) The electrostatic representation of the proteins denoting positive charge (blue), negative charge (red) and neutral charge (white) with the active site circled.

the R232A variant. These results are summarized in Table S1.†

4. Molecular dynamics simulations

In this section, we present a comprehensive analysis of molecular dynamics (MD) simulations focusing on WT YfeX and its four variants R232A, I230A, D143A and S234A. Our objective was to investigate and compare the impact of these point mutations on both global and local structural dynamics, aiming to gain insights into the overall stability of the variants and the changes occurring in the active site environment. To achieve this, we performed several types of analyses including root mean square deviation (RMSD), root mean square fluctuation (RMSF), solvent accessible surface area (SASA), hydrogen bond analysis, heme dihedral angle measurements, and evaluation of the electrostatic potential.

Global structural dynamics of YfeX. As an initial step, RMSD was utilized to gauge the global structural dynamics and the effect of point mutations on the overall stability of the proteins (Fig. 2A). WT, D143A and S234A YfeX all stabilized with deviations from the initial structure of less than 2 Å. We also observed no significant structural changes between the relaxed structure of WT YfeX and the crystal structure (Fig. S10†). In contrast, R232A and I230A required an additional 1000 ns for the structures to stabilize with deviations around 2.2 Å. These results suggest that there are no major global structural changes in the proteins; however, there are greater structural changes for our two highest performing N–H insertion variants, R232A and I230A, compared to WT and the other variants.

To complement the insights revealed from the RMSD analysis, we assessed the SASA for the entire proteins (Fig. 2C). This assessment is valuable in helping to identify global folding or unfolding events within the protein structure, which can be impacted by single amino acid modifications and exert a substantial impact on protein functionality. Here, higher values could suggest instability or unfolding while the lower values could suggest the variant is becoming more compact. For reference, the WT enzyme exhibits an overall SASA of 13 280 Å. Among the variants, D143A displays a marginally reduced SASA in comparison to WT at 13 132 Å. However, S234A, R232A, and I230A present with elevated SASA values of 13 441 Å, 13 544 Å and 13 932 Å, respectively. This evaluation highlights I230A as the variant with the most significant increase in protein SASA, suggesting structural rearrangements that could cause structural instability. In contrast to the slight reduction in SASA for D143A compared to WT, all other variants demonstrate increased values, prompting further inquiry into the specific protein regions contributing to these alterations.

To compare the dynamics of specific regions of the proteins, we calculated the RMSF for all residues (Fig. 2B). This enabled us to precisely identify regions whose flexibility was altered by mutations. The RMSFs of D143A and S234A do not reveal any significant deviations when compared to WT YfeX. However, the R232A variant displays elevated fluctuations proximal to residues 140–145, which are residues that are associated with catalysis in the second coordination sphere (SCS) of the active site. Conversely, the I230A variant is shown to have increased rigidity in the vicinity of residues 220–225, which are also near the active site. It is worth noting that the most pronounced disparity in mobility is observed in the I230A variant, encompassing residues 75–83, although this region is remote from the active site. By visual inspection, it represents a loop region that swings away from the bulk of the protein (see Fig. S11†) and explains the significantly increased SASA we observe for the I230A variant.

To complete our analysis of the global structural dynamics, we compared the average number of hydrogen bonds to understand the overall folding and interactions within each protein (Fig. 2D). The average number of hydrogen bonds for WT is 155.8, which is about the same as observed for the D143A variant with an average number of 155.4. Both the R232A and I230A variants have a slightly lower hydrogen bond number of 151.8 and 153.9, respectively. The most notable deviation from WT is seen in the S234A variant, which exhibits an average number of approximately 159.8 hydrogen bonds. Interestingly, we note that a decrease in hydrogen bonding coupled with an increase in SASA could indicate a decrease in protein stability, especially in the R232A and I230A variants.

Local structural dynamics. To investigate the differences in the size of their active sites, we examined the SASA of the heme cofactor (Fig. 3A) in WT enzyme and the variants. This analysis provides insight into the alterations in the active site's size and accessibility across different variants, which can be critical for substrate binding and catalysis. In WT enzyme, the heme SASA was measured at 115.33 Å^2, a value that is in a comparable range with the 120.14 Å^2 and 95.50 Å^2 observed for the D143A and S234A variants, respectively. In contrast, the R232A variant exhibits a substantially larger active site, with a heme SASA of 170.31 Å^2, and also a much larger channel into the active site. We suspected the larger active site would help increase catalytic activity, especially for larger substrates; however, we also associate this feature with difficult heme reconstitution and stability, which we observed specifically for the R232A variant. Another factor that likely contributes to heme instability in the R232A variant is the drastically changed H-bonding network in the active site. Although the N–H insertion reactivity increases in the I230A variant, we surprisingly observed a significantly smaller active site, measuring a heme SASA of just 47.02 Å^2. In this case, the smaller active site may be advantageous by "forcing" the right type of substrate into a position close to the iron-carbene intermediate, and in this way, promote catalysis. However, this also means that I230A is likely not able to work with larger substrates, and the types of substrates that work with this variant might be more limited. This is in agreement with the finding that the I230A variant does not accept indole as a substrate, despite its very high

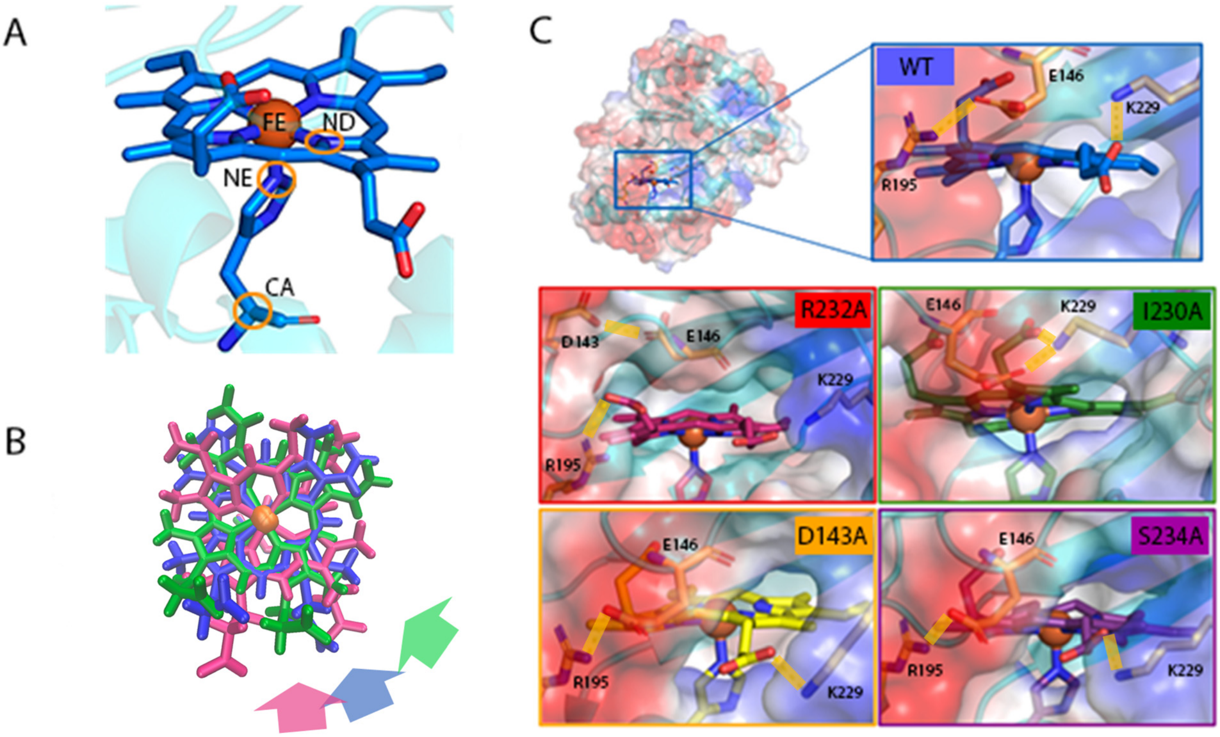


Fig. 3 Local molecular dynamics (MD) analysis of WT YfeX and the R232A, I230A, D143A and S234A variants. (A) The active site of WT YfeX is depicted with circles around the atoms used to calculate the heme dihedral angle. (B) An overlay of the heme cofactor in WT (blue), R232A (pink) and I230A (green) aligned to the iron, proximal histidine and the protein backbone. The colored arrows indicate the entrance to the active site of each protein, respectively. (C) The electrostatic representations of the active sites of each of the proteins, featuring key amino acids comprising the entrance to the active site, along with notable hydrogen bonds shown as yellow rectangles.

N–H insertion reactivity. Correspondingly, further engineering of this variant is necessary going forward.

To complement the insights from the SASA analysis with respect to steric hindrance in the active site, a visual inspection of the electrostatic potential maps and an overall hydrogen bonding analysis of the proteins was conducted. By comparing the electrostatic potential maps of WT YfeX and variants, we discerned that there are two amino acids, E146 and K229, that play a key role in the size and structure of the active site. In WT enzyme, E146 forms a hydrogen bond with R195 creating a negative charge in the entrance channel to the active site, whereas K229 is responsible for the positive charge on the opposite side of the channel, as shown in Fig. 3C. In the R232A variant, the active site entrance looks much different because E146 moves across the distal face of the heme and forms a hydrogen bond with D143 in the space created by the absence of the hydrogen bond between residue R232 and the heme propionate. In addition, the heme propionate previously hydrogen bonded to R232 undergoes a substantial rotation of approximately 150 degrees away from the center of the active site, and forms a new hydrogen bond with residue R195, as shown in Fig. 3C. In the variant I230A the active site seems to be significantly constricted due to a hydrogen bond formed between E146 and K229, leaving the entrance less accessible and facing towards the proximal side of the heme. The D143A variant has a similar entrance compared to WT with E146 forming a hydrogen bond with R195; however, K229 sits on the proximal side of the heme and hydrogen bonds with a heme propionate, compared to K229 in WT that sits on the distal side of the heme. This change in orientation makes the D143A variant entrance orientation different than that of WT, although the sizes remain similar. Lastly, the S234A variant has a nearly identical active site entrance compared to the D143A variant with the hydrogen bond between E146 and R195, along with the K229 hydrogen bond with the heme propionate arm located on the proximal side of the heme. With the K229 residue on the proximal side of the heme, the entrance opens more directly towards the distal side of the heme where reactivity takes place. We suspect that polar residues K229 and E146 in the active site entrance are critical for substrate selectivity and could explain the lower cyclopropanation yields with YfeX, due to low styrene affinity.

The differences in electrostatic interactions in the SCS of the active site between WT and the variants affect the orientation of the heme within the active site, which, in turn, could critically affect its reactivity. To understand how mutations affect heme orientation, we measured the dihedral angle of the heme using the atoms ND–FE–NE–CA, denoted in Fig. 3A, and compared this property between the WT enzyme and variants. For reference, WT enzyme has a dihedral angle of 100.2°, while R232A, I230A and S234A all have lower dihedral angles of 95.4°, 88.3° and 88.1°, respectively; however, D143A has a larger dihedral angle of

111.6°. Changes in dihedral angles are linked to alterations in hydrogen bonding networks in the active site, especially the positioning of heme propionate arms relative to surrounding residues. The spatial arrangement of the heme can impact substrate binding and affinity, and those changes in heme environment are directly reflected in UV-vis experiments, which show minor shifts of the heme's Soret band among the variants. In the R232A variant, which has the rotated heme (depicted in Fig. 3B) and changes in the heme hydrogen bonding environment, we observe a blue-shifted Soret band at 406 nm (compared to 410 nm in WT enzyme), reflecting the changes in the SCS of the heme.

The iron-porphyrin carbene (IPC) intermediate. To investigate the impact of substrates on protein dynamics, we introduced the reactive carbene intermediate into WT YfeX and its R232A and I230A variants. By performing MD simulations on the iron-porphyrin carbene (IPC) complexes, we aimed to gain insight into the dynamics of the system during carbene-transfer reactions. We observed that the IPC does not lead to significant global structural changes compared to the system without the carbene. Our analysis focused on two aspects: the local dynamics of the active site and the orientation of the IPC.

First, we examined the SASA of the heme to test if the IPC changed the structure and subsequent accessibility of the active site. Interestingly, we see that WT SASA increases from 115.3 $Å^2$ without the IPC to 139.9 $Å^2$ with the IPC. This change is attributable to the repositioning of L221, which shifts further from the active site entrance due to an enhancement in alpha-helical character within the residue region 219–221. On the other hand, we do see smaller changes for the R232A and I230A variants: the heme SASA of the R232A IPC complex increases by only 9.8 $Å^2$ to 180.1 $Å^2$, and that of the I230A IPC complex increases by 15.1 $Å^2$ to 62.1 $Å^2$. The data reveal that WT enzyme demonstrates a unique increase in active site size and structural changes with the addition of the IPC, potentially pointing towards an increased active site flexibility compared to the variants, enhancing the ability of WT enzyme to accommodate a wider range of substrates such as styrene for cyclopropanation reactivity. In the R232A variant, with the largest active site, it can be expected that incoming substrate will not be restricted by as much steric hindrance as in the other proteins, which helps explain the decrease in selectivity between the toluidine derivatives with the R232A variant (see Discussion). On the other hand, we were surprised to see that the I232A variant, with its substantially smaller active site, is able to accommodate the IPC without undergoing a large SASA increase. This finding supports our conclusion that the I230A variant is not deleterious, and it suggests that the size of the active site may not play a significant role in promoting efficient N–H insertion reactivity with the smaller substrates (aniline derivatives) we employed, but that larger substrates like indole are affected by the I232A substitution, as observed experimentally.

Next, we evaluated the dihedral angle of the IPC using the C1–C2–FE–ND angle to assess the orientation of the carbene in the active site, as depicted in Fig. 4A. We observe similar dihedral angles of −100 and −120 degrees for the WT and I230A IPC complexes, respectively, as shown in Fig. 4B and C

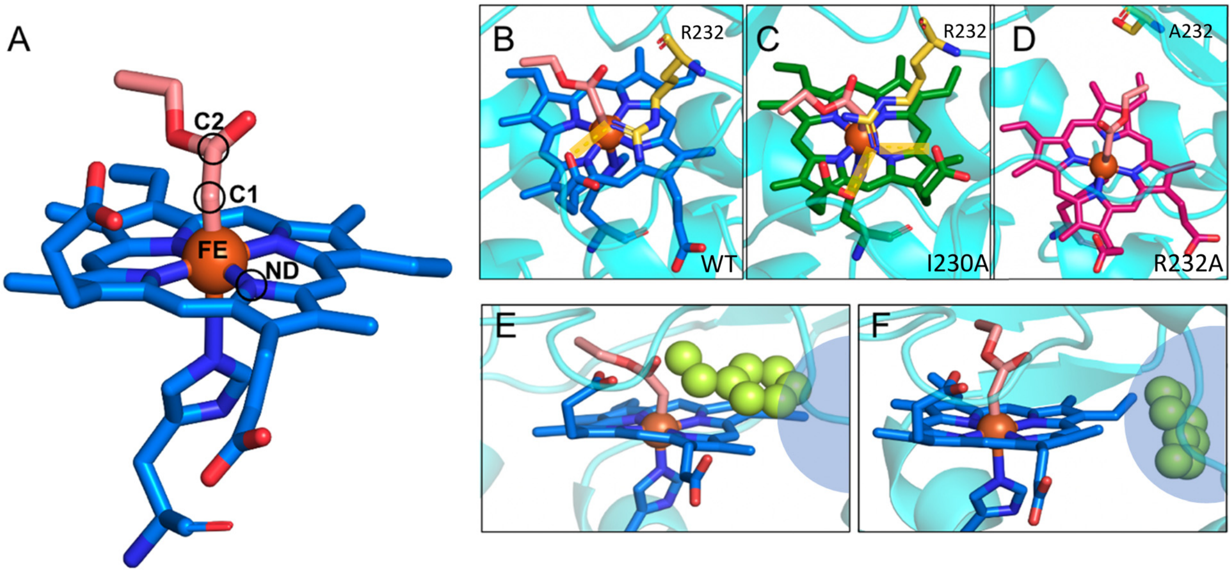


Fig. 4 Active site analysis with IPC and styrene. (A) The WT IPC complex denoting the atoms C2–C1–FE–ND used to calculate the dihedral angle of the heme in the active site. (B) The active site of WT enzyme with the IPC complex showing the position of the IPC and the hydrogen bond (yellow rectangle) between R232 and the heme propionate. (C) The active site of the I230A IPC complex showing the position of the IPC and the hydrogen bonds (yellow rectangles) between R232 and both heme propionate arms. (D) The active site of the R232A IPC complex showing the position of the IPC and the mutated A232 residue with no hydrogen bond. (E) An initial snapshot of the active site of WT enzyme with IPC and styrene in its initial position next to the IPC, on the distal side of the heme. The blue circles indicate a hydrophobic region of the YfeX active site. (F) An end point snapshot of the active site of WT enzyme with IPC and styrene migrated to the edge of the heme, into the hydrophobic region.

and S26.† R232A shows a significant change with a dihedral angle of 50 degrees, rotating the IPC to nearly the opposite side of the active site, situated closer to the active site entrance, as shown in Fig. 4D. This result emphasizes the importance of the bulky R232 residue in determining the orientation of the IPC in the active site: without the R232 residue forming a hydrogen bond across the distal side of the heme with the heme propionate, this opens up space for the IPC to rotate into that region of the active site that is inaccessible in both WT and the I230A variant. Despite the significant difference in IPC orientation in the R232A variant, it does not appear to influence N–H insertion reactivity, although it may further contribute to the observed lack of selectivity among the toluidine derivatives. Conversely, the I230A variant, which has an IPC orientation similar to WT, still exhibits distinct selectivity and reactivity patterns. This suggests that IPC orientation may not be a key determinant in influencing overall reactivity, but that it has a strong effect on regio- and potentially stereoselectivity when more bulky substrates are used.

Styrene positioning in the active site. To take another step towards rationalizing the low reactivity of the IPC towards styrene, we next performed MD simulations incorporating styrene into the active site of the relaxed WT IPC complex to better understand what behaviors are responsible for the low experimental yields. The simulation was run for 100 ns with a low weight constraint on styrene and the overall structure stabilized with an RMSD of 2 Å. We did not observe any major global or local dynamic changes in the WT ICP styrene system. Interestingly, we found that during the simulation styrene moved from its initial position above the heme, shown in Fig. 4E, to a hydrophobic region next to the heme, shown in Fig. 4F. The hydrophobic region consists of residues F248, F172, L262, I258, which are located in the central region of YfeX, forming an opening adjacent to the heme. By observing the behavior of styrene within the active site of WT YfeX, we conclude that styrene has a high affinity for the hydrophobic pocket away from the IPC, impeding efficient cyclopropanation reactivity of YfeX. Future work will now focus on engineering this hydrophobic pocket to improve styrene cyclopropanation yields with YfeX.

Discussion

Carbene transfer reactions are widely used in synthetic organic chemistry, as carbene intermediates are reactive and highly versatile in chemical transformations such as cyclopropanation and N–H insertion. The use of enzymes as biocatalysts in carbene transfer reactions has emerged as an attractive alternative to traditional chemical synthesis methods due to their high efficiency, selectivity, and sustainability. In addition, biocatalysts can be engineered and optimized for specific carbene transfer reactions, which can be difficult to achieve using chemical catalysts, especially when it comes to the late-stage stereo- and enantioselective functionalization of complex organic molecules.

Our detailed experimental studies show that YfeX and its variants investigated here form highly reactive carbene intermediates, which is evident from the very high yields observed for N–H insertion reactions with aniline derivatives (see Table 4).[25] Additionally, through the investigation of the various derivatives, it was proven that YfeX does N–H insertion also on secondary N–H bonds. Variants R232A and I230A generally exhibit elevated N–H insertion reactivity compared to WT enzyme, *with the I230A variant setting a new record, providing >90% yield for the N–H insertion of aniline in only 1 hour reaction time*, compared to typical reaction times of 8–24 hours used in the literature for these types of reactions. We further illustrate this capability here, by performing a scale up of the reaction to 3 ml of aniline, obtaining an isolated yield of 99.9% of the N–H insertion product. On the other hand, the D143A and S234A variants do not exhibit increases in N–H insertion, suggesting that these amino acids may play a more subtle role in reactivity. What is surprising are the variations in yields that we observe with toluidine derivatives (see overview in Table 1). In the case of WT enzyme, the yields with *para*-, *meta*-, and *ortho*-toluidine decrease sequentially compared to aniline, indicating that the enhanced steric hindrance in the toluidine derivatives lowers catalytic activity. Yields with toluidine derivatives are also lower for R232A compared to aniline, but the lack of selectivity among the substituents, each yielding between 40–50%, was unexpected. For the I230A variant, the yield sequence deviates,

Table 4 Summary of N–H insertion and cyclopropanation results with YfeX variants and other relevant proteins. For YfeX and variants, all reactions were run for 1 hour

	WT YfeX[25]	YfeX R232A[25]	YfeX I230A	YfeX S234A	YfeX D143A	Cyt. P450 (H2-5-F10)[a] [6,49]	WT Mb[b] [38,50]	Mb (H64V, V68A)[b] [38,50]
Aniline	72	76	92	56	50	68	21	>99
4-Bromoaniline	40	43	37	53	84	—	—	—
p-Anisidine	57	54	46	48	40	50	—	87
o-Toluidine	28	50	74	62	79	—	—	>99
m-Toluidine	37	51	94	60	88	—	—	>99
p-Toluidine	57	41	45	43	54	70	—	>99
Styrene	27	12	13	9	7	59	36	>99

[a] Reaction conditions: protein (20 μM), EDA (10.0 mM), aniline (20 mM) and $Na_2S_2O_4$ (10 mM). [b] Reaction conditions: 10 mM aniline, 10 mM EDA, 10 mM $Na_2S_2O_4$ with (a) 20 μM (0.2 mol%) and (b) 1 μM (0.01 mol%) hemoprotein. Reaction times: Cyt. P450 (H2-5-F10) with all substrates: 12 h; WT Mb with aniline: 12 h and styrene: 16 h; Mb (H64V, V68A) with styrene: 16 h, all other substrates: 12 h.

with *para*-toluidine providing the lowest yield, followed by *ortho*- and *meta*-toluidine, the latter surprisingly parallels the yield of aniline of around 90%. The S234A variant is unique with no significant yield differences between *ortho*- and *meta*-toluidine, yet a substantial decrease is observed with the *para* compound. Intriguingly, D143A emerges as the only variant that witnesses a yield surge with toluidine derivatives, from *para*- to *ortho*- and then with the *meta* compound yielding a remarkable 88% yield. Similarly, reactivity differences between electron donating and electron withdrawing anilines with the different variants studied here are observed. These changes in selectivity are hard to rationalize and pointed us towards the use of computational methods to better understand our experimental findings.

Whereas YfeX is one of the best biocatalysts for N–H insertion reactions, with great potential for further development for the synthesis of pharmaceuticals, the yields for the cyclopropanation reaction are surprisingly low.[25] Compared to WT YfeX, we observed a further decrease in cyclopropanation yields with styrene for the S234A, D143A, and I230A variants, whereas the R232A variant maintains the same cyclopropanation activity as WT enzyme. Since the carbene intermediate of YfeX is highly reactive, this indicates that non-polar substrates like styrene are likely the problem, but not the reactivity of the carbene intermediate itself. To better solubilize non-polar substrates, use of organic cosolvents for WT YfeX does indeed show drastically enhanced yields for cyclopropanation.[25] However, addition of 30% MeOH cosolvent decreases the cyclopropanation yields for all variants investigated here. CD data reveal differences in folding between the variants and WT enzyme, so we investigated the possibility of lower stability of the variants in the presence of cosolvent. In this regard, note that R232A shows an unexpected propensity for heme loss during expression and purification, further indicating some stability issue with this variant. Unexpectedly, we found that in 30% MeOH cosolvent, I230A YfeX still exhibits equal reactivity for the N–H insertion with aniline compared to the standard conditions (3% MeOH). *This suggests that the low cyclopropanation reactivity may not be due to the instability of the YfeX variants, but an intrinsic affinity or positioning issue with styrene in the active site of YfeX.*

Reactivity studies with indole further demonstrate that YfeX has a chemo-selective preference for the formation of the C–H insertion product at the C3 position of this substrate, compared to the same reaction catalyzed by $Rh_2(OAc)_4$, which yields nearly equal amounts of N–H and C–H insertion products. Notably, WT, R232A and I230A YfeX are mostly selective (>95%) towards the C–H insertion product, with R232A giving a notable 21% overall yield. Interestingly, Weissenborn and coworkers observed selective C–H functionalization at the C3 position of an unprotected indole when reacted with diazoacetonitrile, using I230A YfeX and resulting in a 3% yield,[48] consistent with the selectivity seen herein. Due to the high reactivity of YfeX towards aniline, we were surprised that for indole, the C–H insertion is preferred. However, this is likely to be attributed to the nucleophilicity at the C3 position of indole.[51] Note that a mechanism for this C–H insertion reaction was previously proposed by Fasan and coworkers for myoglobin.[47] While we believe the nucleophilicity at the C3 position of indole lends to this selectivity, it is possible that this observed trend could also be affected by the increased difficulty of accessing the secondary amine of indole compared to the primary amine of aniline. Our results are also comparable to the WT YfeX and I230A YfeX reactions that Weissenborn and coworkers performed with *N*-methylindole and EDA, yielding <3% and 3% product, respectively.[24] Overall, our results show improved indole functionalization activity with the YfeX R232A variant, which therefore forms the basis for the future development of a highly active, C3-selective catalyst for the functionalization of indole derivatives.

When the amino acid substitutions performed here are evaluated in a "static" way, by simply substituting the amino acids in the crystal structure using PyMol, none of these findings can be understood. This is because even simple amino acid substitutions can have secondary effects that influence structural dynamics, hydrogen bonding networks and other second coordination sphere (SCS) interactions that cause larger, unexpected structural changes around a protein's active site. To begin, we performed MD simulations on WT YfeX and our 4 variants in the absence of substrate to analyze the global and local structural dynamics of these proteins.

In the R232A variant, removal of R232 has a significant effect on the H-bonding network and structure of the active site. Most notably, the hydrogen bond between one of the propionate arms and R232 is removed, leading to a drastic reorientation of the heme propionates in the active site, and a shift of residues D141, D135 and E144 into the space created by the absence of R232. This leads to a much larger entrance to the active site in R232A compared to WT YfeX, as shown in Fig. 3C, and a generally larger heme SASA value in the variant. Moreover, the absence of the hydrogen bond between the heme propionate and R232 leads to increased space above the heme in the distal pocket. Notably, this expanded active site in R232A is large enough to accommodate the IPC intermediate with just minor adjustments. The large active site in R232A explains the higher N–H insertion yields and loss of selectivity for the toluidine derivatives observed experimentally, compared to WT enzyme. In addition, the larger entrance channel and/or active site in R232A compared to WT and I230A explains the enhanced reactivity of R232A with a bulkier substrate like indole. On the other hand, the unexpected formation of a polar salt-bridging network in the distal site may further disfavor nonpolar substrates such as styrene. These results also explain the gradual loss of heme over time and the resultant complications in heme reconstitution in the R232A variant. We deduce this to be a consequence of the expanded active site entrance revealed in the MD simulations, which exposes the heme to more solvent. Moreover, the hydrogen bond between the heme propionate and R232 presumably provides additional stabilization to the heme that is lacking in the R232A variant.

Contrary to our expectations, removal of the entrance channel I230 residue in the I230A variant does not lead to an increase in the size of the channel into the active site, but it actually reduces the diameter of the channel and greatly decreases the heme SASA value, along with altering the orientation of the entrance in relation to the heme. Notably, the I230A variant has the highest N–H insertion reactivity with aniline, again contradicting our expectations considering the narrowed size of its active site entrance channel and reduced volume of the active site. We speculate that the smaller active site forces aniline derivatives in a catalytically productive conformation in the active site, thus increasing N–H insertion reactivity, while at the same time preventing the reaction with the larger substrate indole. The narrow active site further means that charged residues E146 and K246 are closer together, likely presenting issues for nonpolar, hydrophobic substrates such as styrene to enter the active site, thus contributing to the lower cyclopropanation yields we observe in the I230A variant. The size of the active site does not increase as much upon the addition of the IPC as compared to WT, which is again surprising, considering its smaller initial size. Increased steric hinderance in the smaller active site of I230A further explains the unusual trends in selectivity among the substituted toluidine substrates (especially the discrimination against *para*-substituted anilines).

As summarized above, the cyclopropanation reactivity across all variants is markedly low. Our MD simulations suggest that a hydrophobic region in the active site, where styrene preferentially migrates, might be the root cause. This region could impede the active site's interaction with nonpolar substrates. Despite our initial assumption that the larger active site in R232A would enhance styrene accessibility, this did not materialize experimentally in a higher yield for cyclopropanation. Our computational analysis shows that in R232's absence, residue E146 shifts directly above the active site, which could increase the site's polarity, thereby affecting nonpolar substrate reactivity. Furthermore, the IPC intermediate in R232A is seen to stabilize towards the active site's hydrophobic region, suggesting that the substrate would need to bind from the distal site's polar region, which might not be favorable for nonpolar substrates. In addition, the active site entrances of WT and all variants studied here are primarily polar and charged due to the lysine and glutamic acid residues framing them. These residues present intriguing opportunities for active site engineering, potentially aiding in the introduction of nonpolar substrates into the active site. Based on these results, we propose that a more hydrophobic entrance channel and an active site with less steric hindrance will help substrate affinity and positioning for improved cyclopropanation reactivity. At the same time, the hydrophobic region in the active site needs to be disrupted, to avoid trapping hydrophobic substrates in this pocket, which prevents them from reacting with the iron-carbene intermediate.

In summary, in this paper we show that single amino acid substitutions in the YfeX active site have an unexpectedly complex effect on overall protein stability, carbene transfer reactivity, and selectivity. Above all, YfeX and its variants are excellent catalysts for the N–H insertion reaction, operating on a large variety of aniline derivatives with electron-donating or withdrawing and bulky substituents. Detailed computational studies help rationalize the experimental observations, also with respect to YfeX's surprisingly low cyclopropanation reactivity. In order to improve reactivity with non-polar substrates like styrene, we hypothesize that mutating the charged active site entrance residues E146 and K229 to alanine would increase cyclopropanation yields by encouraging nonpolar substrate to enter the active site. In addition, mutating residue L262 to phenylalanine or tryptophan would increase steric hindrance in the hydrophobic region of the active site and help to direct styrene closer to the carbene intermediate, thus potentially increasing cyclopropanation yields. These amino acid substitutions will be explored in future studies.

Data availability

The data supporting this article, in particular GC/MS chromatogram and MS data, supercritical fluid chromatogram and MS data, and calibration curves have been included as part of the ESI.† UV-vis and CD data are included in the ESI;† the corresponding raw data are available from the corresponding author upon request. Original research data pertaining to the MD simulations are available from the University of Michigan – Deep Blue Data. **https://doi.org/10.7302/ear2-rd86**.

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

VSA acknowledges support from a Rackham Merit Fellowship and a One-Term Dissertation Fellowship (both University of Michigan). The authors thank Prof. Martin Weissenborn (Martin-Luther-University Halle-Wittenberg, Germany) and his graduate student Dr. Anja Knorrscheidt for providing YfeX variant plasmids. The authors further acknowledge Prof. Christo Christov (Michigan Technological University) and his graduate students Midhun George Thomas and Sodiq Waheed for the guidance and expertise provided to perform the MD simulations and analyses on WT YfeX and YfeX variants.

References

1 C.-M. Che, V. K.-Y. Lo, C.-Y. Zhou and J.-S. Huang, Selective functionalisation of saturated C–H bonds with

metalloporphyrin catalysts, *Chem. Soc. Rev.*, 2011, **40**, 1950–1975.

2 H. Lu and X. P. Zhang, Catalytic C–H functionalization by metalloporphyrins: recent developments and future directions, *Chem. Soc. Rev.*, 2011, **40**, 1899–1909.

3 A. B. McQuarters, M. W. Wolf, A. P. Hunt and N. Lehnert, 1958-2014: After 56 Years of Research, Cytochrome P450 Reactivity Finally Explained, *Angew. Chem., Int. Ed.*, 2014, **53**, 4750–4752.

4 J. Rittle and M. T. Green, Cytochrome P450 compound I: capture, characterization, and CH bond activation kinetics, *Science*, 2010, **330**, 933–937.

5 M. M. Simões, D. T. Gonzaga, M. F. Cardoso, L. d. S. Forezi, A. T. Gomes, F. D. C. Da Silva, V. F. Ferreira, M. G. Neves and J. A. Cavaleiro, Carbene transfer reactions catalysed by dyes of the metalloporphyrin group, *Molecules*, 2018, **23**, 792.

6 P. S. Coelho, E. M. Brustad, A. Kannan and F. H. Arnold, Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes, *Science*, 2013, **339**, 307–310.

7 A. Tinoco, V. Steck, V. Tyagi and R. Fasan, Highly diastereo- and enantioselective synthesis of trifluoromethyl-substituted cyclopropanes via myoglobin-catalyzed transfer of trifluoromethylcarbene, *J. Am. Chem. Soc.*, 2017, **139**, 5293–5296.

8 O. F. Brandenberg, R. Fasan and F. H. Arnold, Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions, *Curr. Opin. Biotechnol.*, 2017, **47**, 102–111.

9 M. W. Wolf, D. A. Vargas and N. Lehnert, Engineering of RuMb: Toward a Green Catalyst for Carbene Insertion Reactions, *Inorg. Chem.*, 2017, **56**, 5623–5635.

10 K. Oohora, H. Meichin, L. Zhao, M. W. Wolf, A. Nakayama, J.-Y. Hasegawa, N. Lehnert and T. Hayashi, Catalytic cyclopropanation by myoglobin reconstituted with iron porphycene: Acceleration of catalysis due to rapid formation of the carbene species, *J. Am. Chem. Soc.*, 2017, **139**, 17265–17268.

11 M. Garcia-Borràs, S. J. Kan, R. D. Lewis, A. Tang, G. Jimenez-Osés, F. H. Arnold and K. N. Houk, Origin and Control of Chemoselectivity in Cytochrome c Catalyzed Carbene Transfer into Si–H and N–H bonds, *J. Am. Chem. Soc.*, 2021, **143**, 7114–7123.

12 R. Stenner, J. W. Steventon, A. Seddon and J. R. Anderson, A de novo peroxidase is also a promiscuous yet stereoselective carbene transferase, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 1419–1428.

13 M. Pott, M. Tinzl, T. Hayashi, Y. Ota, D. Dunkelmann, P. R. Mittl and D. Hilvert, Noncanonical heme ligands steer carbene transfer reactivity in an artificial metalloenzyme, *Angew. Chem.*, 2021, **133**, 15190–15195.

14 J. Huang, A. Quest, P. Cruz-Morales, K. Deng, J. H. Pereira, D. Van Cura, R. Kakumanu, E. E. Baidoo, Q. Dan and Y. Chen, Complete integration of carbene-transfer chemistry into biosynthesis, *Nature*, 2023, 1–6.

15 Y. Gu, S. N. Natoli, Z. Liu, D. S. Clark and J. F. Hartwig, Site-Selective Functionalization of (sp3) C– H Bonds Catalyzed by Artificial Metalloenzymes Containing an Iridium-Porphyrin Cofactor, *Angew. Chem.*, 2019, **131**, 14092–14098.

16 E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna and E. Romero, Biocatalysis, *Nat. Rev. Methods Primers*, 2021, **1**, 46.

17 J. C. Sharland, B. Wei, D. J. Hardee, T. R. Hodges, W. Gong, E. A. Voight and H. M. Davies, Asymmetric synthesis of pharmaceutically relevant 1-aryl-2-heteroaryl-and 1, 2-diheteroarylcyclopropane-1-carboxylates, *Chem. Sci.*, 2021, **12**, 11181–11190.

18 J. Wang, C.-M. Che and M. P. Doyle, *Transition Metal-Catalyzed Carbene Transformations*, John Wiley & Sons, 2021.

19 C. Zeymer and D. Hilvert, Directed evolution of protein catalysts, *Annu. Rev. Biochem.*, 2018, **87**, 131–157.

20 X. Zhang and S. Li, Expansion of chemical space for natural products by uncommon P450 reactions, *Nat. Prod. Rep.*, 2017, **34**, 1061–1089.

21 Y. Peng, C. Gao, Z. Zhang, S. Wu, J. Zhao and A. Li, A chemoenzymatic strategy for the synthesis of steroid drugs enabled by P450 monooxygenase-mediated steroidal core modification, *ACS Catal.*, 2022, **12**, 2907–2914.

22 R. Ushimaru and I. Abe, Unusual Dioxygen-Dependent Reactions Catalyzed by Nonheme Iron Enzymes in Natural Product Biosynthesis, *ACS Catal.*, 2022, **13**, 1045–1076.

23 M. Diéguez, J. E. Bäckvall and O. Pàmies, *Artificial Metalloenzymes and MetalloDNAzymes in Catalysis: From Design to Applications*, Wiley, 2018.

24 M. J. Weissenborn, S. A. Löw, N. Borlinghaus, M. Kuhn, S. Kummer, F. Rami, B. Plietker and B. Hauer, Enzyme-Catalyzed Carbonyl Olefination by the E. coli Protein YfeX in the Absence of Phosphines, *ChemCatChem*, 2016, **8**, 1636–1640.

25 V. S. Alfaro, S. O. Waheed, H. Palomino, A. Knorrscheidt, M. Weissenborn, C. Z. Christov and N. Lehnert, YfeX–A New Platform for Carbene Transferase Development with High Intrinsic Reactivity, *Chem. – Eur. J.*, 2022, e202201474.

26 H. A. Dailey, A. N. Septer, L. Daugherty, D. Thames, S. Gerdes, E. V. Stabb, A. K. Dunn, T. A. Dailey and J. D. Phillips, The Escherichia coli protein YfeX functions as a porphyrinogen oxidase, not a heme dechelatase, *MBio*, 2011, **2**, DOI: 10.1128/mbio.00248-11.

27 X. Liu, Z. Yuan, J. Wang, Y. Cui, S. Liu, Y. Ma, L. Gu and S. Xu, Crystal structure and biochemical features of dye-decolorizing peroxidase YfeX from Escherichia coli O157 Asp143 and Arg232 play divergent roles toward different substrates, *Biochem. Biophys. Res. Commun.*, 2017, **484**, 40–44.

28 C. Chen, R. Shrestha, K. Jia, P. F. Gao, B. V. Geisbrecht, S. H. Bossmann, J. Shi and P. Li, Characterization of dye-decolorizing peroxidase (DyP) from Thermomonospora curvata reveals unique catalytic properties of A-type DyPs, *J. Biol. Chem.*, 2015, **290**, 23447–23463.

29 S. J. Kim and M. Shoda, Decolorization of molasses and a dye by a newly isolated strain of the fungus Geotrichum candidum Dec 1, *Biotechnol. Bioeng.*, 1999, **62**, 114–119.

30 A. Falade and T. Ekundayo, Emerging biotechnological potentials of DyP-type peroxidases in remediation of lignin

wastes and phenolic pollutants: a global assessment (2007–2019), *Lett. Appl. Microbiol.*, 2021, **72**, 13–23.

31 D. I. Colpa, M. W. Fraaije and E. van Bloois, DyP-type peroxidases: a promising and versatile class of enzymes, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 1–7.

32 N. Lončar, N. Drašković, N. Božić, E. Romero, S. Simić, I. Opsenica, Z. Vujčić and M. W. Fraaije, Expression and Characterization of a Dye-decolorizing Peroxidase from Pseudomonas fluorescens Pf0-1, *Catalysts*, 2019, **9**, 463.

33 P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold and E. M. Brustad, A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo, *Nat. Chem. Biol.*, 2013, **9**, 485–487.

34 R. D. Lewis, M. Garcia-Borràs, M. J. Chalkley, A. R. Buller, K. Houk, S. J. Kan and F. H. Arnold, Catalytic iron-carbene intermediate revealed in a cytochrome c carbene transferase, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 7308–7313.

35 Y. Wei, A. Tinoco, V. Steck, R. Fasan and Y. Zhang, Cyclopropanations via heme carbenes: Basic mechanism and effects of carbene substituent, protein axial ligand, and porphyrin substitution, *J. Am. Chem. Soc.*, 2017, **140**, 1649–1662.

36 R. K. Zhang, X. Huang and F. H. Arnold, Selective CH bond functionalization with engineered heme proteins: new tools to generate complexity, *Curr. Opin. Chem. Biol.*, 2019, **49**, 67–75.

37 V. Pfanzagl, K. Nys, M. Bellei, H. Michlits, G. Mlynek, G. Battistuzzi, K. Djinovic-Carugo, S. Van Doorslaer, P. G. Furtmüller and S. Hofbauer, Roles of distal aspartate and arginine of B-class dye-decolorizing peroxidase in heterolytic hydrogen peroxide cleavage, *J. Biol. Chem.*, 2018, **293**, 14823–14838.

38 G. Sreenilayam and R. Fasan, Myoglobin-catalyzed intermolecular carbene N–H insertion with arylamine substrates, *Chem. Commun.*, 2015, **51**, 1532–1534.

39 M. H. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.

40 P. Li and K. M. Merz Jr, MCPB.py: A Python Based Metal Center Parameter Builder, *J. Chem. Inf. Model.*, 2016, **56**, 599–604.

41 J. M. Seminario, Calculation of intramolecular force fields from second-derivative tensors, *Int. J. Quantum Chem.*, 1996, **60**, 1271–1277.

42 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.

43 S. O. Waheed, S. S. Chaturvedi, T. G. Karabencheva-Christova and C. Z. Christov, Catalytic mechanism of human ten-eleven translocation-2 (tet2) enzyme: Effects of conformational changes, electric field, and mutations, *ACS Catal.*, 2021, **11**, 3877–3890.

44 R. Ramanan, S. O. Waheed, C. J. Schofield and C. Z. Christov, What Is the Catalytic Mechanism of Enzymatic Histone N-Methyl Arginine Demethylation and Can It Be Influenced by an External Electric Field?, *Chem. – Eur. J.*, 2021, **27**, 11827–11836.

45 R. L. Davidchack, T. Ouldridge and M. Tretyakov, New Langevin and gradient thermostats for rigid body dynamics, *J. Chem. Phys.*, 2015, **142**, 144114.

46 K. Urbina, D. Tresp, K. Sipps and M. Szostak, Recent advances in metal-catalyzed functionalization of indoles, *Adv. Synth. Catal.*, 2021, **363**, 2723–2739.

47 D. A. Vargas, A. Tinoco, V. Tyagi and R. Fasan, Myoglobin-Catalyzed C–H Functionalization of Unprotected Indoles, *Angew. Chem., Int. Ed.*, 2018, **57**, 9911–9915.

48 K. J. Hock, A. Knorrscheidt, R. Hommelsheim, J. Ho, M. J. Weissenborn and R. M. Koenigs, Tryptamine synthesis by iron porphyrin catalyzed C− H functionalization of indoles with diazoacetonitrile, *Angew. Chem., Int. Ed.*, 2019, **58**, 3630–3634.

49 Z. J. Wang, Cytochrome P450-catalyzed insertion of carbenoids into N–H bonds, *Chem. Sci.*, 2014, **5**, 598–601.

50 M. Bordeaux, V. Tyagi and R. Fasan, Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts, *Angew. Chem.*, 2015, **127**, 1764–1768.

51 R. J. Sundberg, Electrophilic substitution reactions of indoles, in *Heterocyclic Scaffolds II: Reactions and Applications of Indoles*, 2010, pp. 47–115.