Selecting of a Cytochrome P450$_{cam}$ SeSaM Library with 3-Chloroindole and Endosulfan – Identification of Mutants that Dehalogenate 3-Chloroindole

Shaima Kammoonah,[a] Brinda Prasad,[a] Priyadarshini Balaraman, [a] Hemanshu Mundhada,[b,c] Ulrich Schwaneberg[b] and Erika Plettner[a]*

[a] *Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada – V5A 1S6.*

[b] *Institute of Biotechnology, RWTH Aachen University, Worringer Weg 3, 52074 Aachen, Germany*

[c] *Current address: Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs, Lyngby, Denmark*

* Author for correspondence:

e-mail: plettner@sfu.ca

Tel.: (+1)-778-782-3586

Abstract

Cytochrome P450$_{cam}$ (a camphor hydroxylase) from the soil bacterium *Pseudomonas putida* shows potential importance in environmental applications such as the degradation of chlorinated organic pollutants. Seven P450$_{cam}$ mutants generated from Sequence Saturation Mutagenesis (SeSaM) and isolated by selection on minimal media with either 3-chloroindole or the insecticide endosulfan were studied for their ability to oxidize of 3-chloroindole to isatin. The wild-type enzyme did not accept 3-chloroindole as a substrate. Mutant (E156G/V247F/V253G/F256S) had the highest maximal velocity in the conversion of 3-chloroindole to isatin, whereas mutants (T56A/N116H/D297N) and (G60S/Y75H) had highest k_{cat}/K_M values. Six of the mutants had more than one mutation, and within this set, mutation of residues 297 and 179 was observed twice. Docking simulations were performed on models of the mutant enzymes; the wild-type did not accommodate 3-chloroindole in the active site, whereas all the mutants did. We propose two potential reaction pathways for dechlorination of 3-chloroindole.

1. Introduction

Persistent organic pollutants (POPs), such as the insecticide endosulfan, pose a specific hazard because of four characteristics: 1) they are toxic; 2) they are persistent, resisting processes that break down contaminants; 3) they are lipid soluble and accumulate in the body fat of humans, marine mammals and other animals and are passed from mother to fetus; and 4) they can travel great distances through wind and water currents [1,2]. Even small quantities of POPs can wreak havoc in human and animal tissue, causing nervous system damage, diseases of the immune system, reproductive and developmental disorders, and cancers. One such POP in circulation today is the insecticide endosulfan **1** (Fig. 1). Endosulfan is banned under the Stockholm convention but is still used in India and China [3]. The reason endosulfan is so persistent is that it is difficult to biodegrade [3,4]. All the endosulfan metabolites reported so far still have the hexachlorinated bicyclic nucleus of the original pesticide intact [4]. The enzymatic degradation of the chlorinated end of endosulfan has not been reported according to our knowledge.

Cytochromes P450 are hemoproteins that catalyze the oxidation of hydrocarbons, among many other reactions [5]. The wild type P450$_{cam}$ (CYP101A1) hydroxylates 1(*R*) (+)-camphor **2** at the 5th position to form 5-*exo*-hydroxycamphor **3** and 5-ketocamphor **4** (Fig. 1)[6]. The crystal structure of P450$_{cam}$ with camphor bound has revealed that the 5-*exo* C-H bond or 1(R) (+)-camphor is poised very precisely over the proximal side of the iron porphyrin [7]. We

constructed an *in silico* model of wild-type $P450_{cam}$ with endosulfan **1** bound, because of the structural resemblance of endosulfan **1** with camphor **2**, but found that $P450_{cam}$ cannot position the polychlorinated end of endosulfan above the porphyrin, which is an essential requisite for its oxidation and subsequent dechlorination. We therefore predicted that mutating the active site of $P450_{cam}$ could lead to better positioning of endosulfan for oxidation at the chlorinated end.

We targeted to mutate the active site of $P450_{cam}$ to dehalogenate chlorinated substrates with at least one vinylic C-Cl bond. There are two approaches to altering the substrate selectivity of an enzyme: 1) site-directed mutagenesis based on structural models of the enzyme [8], or 2) directed evolution [9]. Directed evolution is a strategy to alter the properties of a protein by iterative rounds of random mutagenesis and subsequent screen for improved variants [10, 11].

For random mutagenesis of the gene, there are several options: a) error-prone PCR [12], b) multi-template PCR [13], c) Sequence Saturation Mutagenesis (SeSaM) [14, 15], d) DNA shuffling [16] and e) combinatorial active saturation test (CAST) [17]. We chose SeSaM because it can achieve up to 16.7% of consecutive nucleotide exchanges, which is remarkable. In comparison, the first two methods give recurring isolated mutations (hot-spots) [15]. The challenge with large libraries is the identification of mutants with the desired activity. Various selection/screening methods have been reported for large libraries of mutants. For instance, colorimetric/fluorometric assays, in which the product is converted to a colored or a fluorescent dye by reaction with an added reagent, is an easy method of detection. This method has been reported for a P450 BM3 variant 139-3, where fluorescent dye gamma-(4-nitrobenzyl) pyridine, is added to styrene oxide, the product produced upon oxidation of styrene [18]. One limitation of this assay is that the colorimetric agent (e.g., hydrazines used for ketone detection) may be inhibitory to P450 reactivity. In a different selection method (survival screen), bacteria that are expressing the enzyme are selected on media that contains the substrate of interest. The selection process can be improved in this case if the given substrate is toxic to bacteria and the metabolites of the substrate are less toxic than the substrate itself. However, this process is not useful if the product(s) formed is (are) more toxic than the substrate itself.

In this work, we used a survival assay to select members of the SeSaM library that dehalogenate two chlorinated substrates: 1) 3-chloroindole **5** and 2) endosulfan **1**. We chose 3-chloroindole **5**

because it has one vinylic C-Cl bond and, in analogy to indole **6,** could be oxidized to form the dye, isatin **7** (Fig. 1). Indole **6** has been reported to be oxidized to isatin **7** or indigo **8** by certain P450s or their mutants [19-21]. For example, mutagenesis of CYP2A6, a human P450, resulted in an increased catalytic efficiency for indole biotransformation to colored pigments and conferred a capacity to oxidize substituted indoles [19, 21]. To address the question of whether POPs such as endosulfan can be oxidized (or dehalogenated) by $P450_{cam}$ and its mutants, we used the substrate, 3-chloroindole **5**. The reasons we introduced a vinylic chloride on indole are: 1) endosulfan has two vinylic chlorides on its chlorinated end, 2) the C=C double bond in endosulfan is the most reactive site in the chlorinated portion of this molecule, and 3) oxidation/dechlorination of 3-chloroindole is expected to give the easily detected colored product isatin **7**. In contrast, oxidation of the C=C double bond in endosulfan does not produce easily detected colored products. Therefore, in this work, we investigate 3-chloroindole as a model substrate to find: 1) whether $P450_{cam}$ can be randomly mutated by SeSaM and active mutants can be selected by survival of bacteria that express them and 2) whether selected mutants can dechlorinate the vinylic carbon-chlorine bond of 3-chloroindole *via* oxidative dehalogenation/hydroxylation to form isatin. This work will set the platform for future work regarding endosulfan dehalogenation .

$P450_{cam}$ requires two electron transfer partners for catalysis: putidaredoxin (PdX) and putidaredoxin reductase (PdR) to transfer electrons donated by the NADH cofactor (Fig. 1)[22]. These electrons are used to couple the reduction of O_2 to the oxidation of the substrate. We used these electron transfer partners in our experiments with the $P450_{cam}$ mutants studied here. $P450_{cam}$ can also be oxidized, to the reactive intermediate that oxidizes the substrate, by chemical oxidants, such as *meta*-chloroperbenzoic acid, *m*-CPBA. This is known as the shunt pathway of P450 catalysis, and this does not require electron transfer partners. We used *m*-CPBA as a shunting agent during selection.
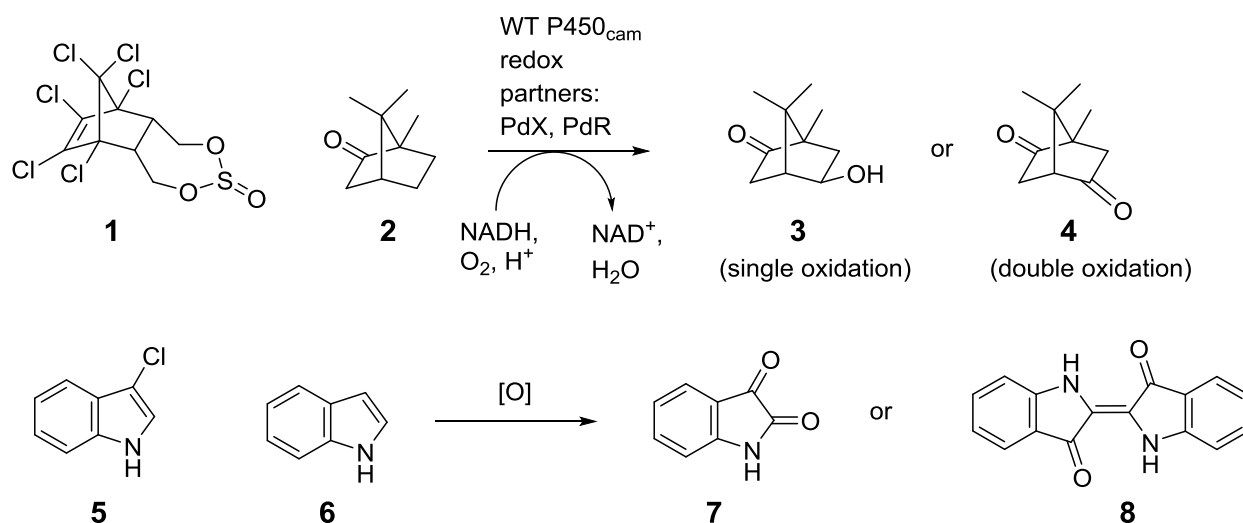
Fig. 1. Structures of the compounds presented in the text, along with the oxidative transformation of some of these compounds. Compound names: **1** = endosulfan, **2** = 1(*R*) (+)-camphor, **3** = 5-*exo* hydroxycamphor, **4** = 5-ketocamphor, **5** = 3-chloroindole, **6** = indole, **7** = isatin, **8** = indigo.

2. Materials and Methods

The chemicals were of analytical grade and purchased from Sigma-Aldrich Chemie (Taufkirchen Germany) or Sigma-Aldrich Canada (Oakville, Ontario), AppliChem (Darmstadt, Germany) or Carl Roth (Karlsruhe, Germany). The nucleotide analog dPTP was purchased from Jena Biosciences (Jena, Germany) and dPTP α S and dGTP α S were generously provided by Biology Life Science Institute (Bremen, Germany). All other nucleotides were purchased from Fermentas (St. Leon-Rot, Germany). Taq DNA-Polymerase was obtained from Qiagen (Hilden, Germany). All other DNA-polymerases, the terminal deoxynucleotidyltransferase (TdT) were obtained from New England Biolabs (Frankfurt, Germany). DNA was quantified using a NanoDrop photometer (NanoDrop Technologies, Wilmington, DE, USA). The PCRs were performed in 0.2 mL thin-walled PCR tubes from Sarstedt (Nuembrecht, Germany) using a Mastercycler gradient PCR-machine from Eppendorf (Hamburg, Germany). Volumes higher than 50 µl were distributed in multiple tubes. *m*-Chloroperbenzoic acid (*m*-CPBA), endosulfan (a 2:1 mixture of α and β isomers commonly used in the field) and tricarbonylchloro(glycinato)ruthenium (II) (CORM-3) were purchased from Sigma-Aldrich (Oakville, ON, Canada). *m*-CPBA was purified by reported

methods [23]. Centrifugations were carried out with a Beckmann Avanti J-26 XPI centrifuge (Mississauga, ON, Canada), equipped with JLA 8.1000 and JA 25.50 rotors. NMR spectra were obtained using a Bruker AVANCE II 400 MHz instrument. Gas Chromatography-mass spectrometry (GC-MS) was performed on a Varian Saturn CP3800 GC interfaced with a Saturn 2000 ion trap mass spectrometer. The GC was fitted with a temperature programmable split/splitless injector and a 30 m SPB-5 capillary column (0.2 mm i.d., 0.25 μm film thickness, Supelco, Missisauga, ON, Canada). GC analyses were performed in splitless mode. UV-VIS spectroscopy was performed on a Cary 300 UV-Visible double beam instrument at 246 nm for 15 min.

2.1. Construction and Cloning of SeSaM libraries

A transversion-biased SeSaM-Tv P/P mutant library was constructed by the known SeSAM-Tv P/P protocol (Fig. S1) [24]. The $P450_{cam}$ gene was cloned into the pALXtreme-1a vector [25] at restriction sites EcoRI and HindIII (Fig. S2). Taking this whole plasmid as a template, the $P450_{cam}$ SeSaM library acted as a megaprimer in a PCR, described as MEGAWHOP [26]. The PCR products were digested by *DpnI* to eliminate the methylated parental template plasmid. To check *DpnI* digestion efficiency, a 50 μL PCR reaction mixture (containing the megaprimer, template plasmid, dNTPs, Pfus enzyme and buffer) was taken as a control and digested with 10U of *DpnI* overnight. Colonies were not observed on this control plate, which suggested that the MEGAWHOP and the *DpnI* digestion processes worked efficiently. Twelve clones each from the library were checked (Fig. S3) for the presence of the insert (by culture and PCR), and the positive clones were grown in LB/kanamycin medium overnight. The isolated plasmids from these clones were sequenced at Eurofins MWG Operon, Germany. The SeSaM library was checked with the sequenced clones and glycerol stocks of the library were made for further experiments with 3-chloroindole.

2.2. Preparation of 3-chloroindole

The substrate, 3-chloroindole (Fig. S4, molecule 1), was prepared from indole (Sigma Aldrich) in two steps as reported previously [27].

6

N-chloroindole. To 0.585 g (1 mmol) of indole, 500 mL of *n*-pentane, 125 mL of aqueous sodium hypochlorite (reagent grade, 10-15% available chlorine) and 125 mL of water were added and stirred at 0°C for 3 h. The organic layer was separated and the aqueous layer was extracted with an additional 100 mL of *n*-pentane. The combined organic layers were dried over K_2CO_3. Pentane was distilled under vacuum and the oily residue was stored at 4 °C.

3-chloroindole. To the oily residue in a 2 L flask, 500 mL of *n*-butanol and 7.5 g of K_2CO_3 were added, and the reaction was refluxed for 2 h. The alcohol was removed under reduced pressure and to this, 200 mL of water and 200 mL of $CHCl_3$ were added. The layers were separated and the chloroform layer was dried over K_2CO_3 and distilled under vacuum. The oily residue was purified by column chromatography using hexane : ethyl acetate (8:1), as the solvent, and the isolated product (yield 65%), was characterized by NMR. ^{1}H NMR (400 MHz, $CDCl_3$): δ(ppm) 8.14 (s, NH), δ 7.55 (d, 1H, J=8.4 Hz), δ 7.39 (s, 1H), δ 7.20 (dd, 1H, J=2.8 Hz, 8.4 Hz), δ 7.09 (dd, 1H, J=2.8 Hz, 8.4 Hz), δ 6.54 (d, 1H, J=2.8 Hz), (Fig. S4).

2.3. IC_{50} experiments with 3-chloroindole and endosulfan

To determine the toxicity of 3-chloroindole or endosulfan on the expression hosts for the mutated $P450_{cam}$ variants, *E. coli* BL21(DE3) cells, a dose response was obtained as follows. An overnight culture of non-transformed BL21-(DE3) cells acted as an inoculum for 20 mL of LB medium (with no antibiotics). The culture was grown at 37°C for an hour until the O.D. was 0.1. For determination of 3-chloroindole IC_{50}, the culture was then diluted 1000× and to 3 mL of this diluted culture, 3-chloroindole or endosulfan was added in varying concentrations (from 5 μM to 300 μM for 3-chloroindole and from 70 nM to 300 μM for endosulfan), in 3 replicates per concentration. The cultures with 3-chloroindole or endosulfan were grown at 37°C for an hour and kept at 4°C overnight. An aliquot (20 μL) of each replicate was plated on LB-agar/non-antibiotic plates and incubated overnight at 37°C. The next day, colonies were counted, and IC_{50} values were obtained from the average colony-forming units (CFU)/mL of the three replicates, using the GraphPad Prism software (GraphPad Software, Inc., California, USA). Subsequent

selection protocols were conducted at $2 \times IC_{50}$ concentration for 3-chlroindole and at IC_{90} for endosulfan.

2.4. Selection of mutants from the SeSaM library with 3-chloroindole or endosulfan

The MEGAWHOP products (libraries and the wild type $P450_{cam}$) and the empty vector (pALXtreme-1a) (2μL) were transformed into *E. coli* BL21(DE3), plated on LB-agar (kanamycin) plates (50 μg/mL) supplemented with isopropyl thiogalactoside, IPTG, (80 μM), 3-chloroindole ($2 \times IC_{50}$ concentration, 128 μM) or endosulfan (IC_{90} 250 μM), δ-aminolevulinic acid (1 mM), vitamin B_1 (10 μM), $FeCl_2$ (0.1 μM) and *m*-chloroperbenzoic acid (*m*-CPBA, 0.3 μM). The *m*-CPBA allows the cytochrome to form the high-valent iron-oxo form (Compound I), without reducing O_2. This is known as the peroxide shunt [28, 29]. Plates were incubated overnight at 37°C. The active clones were identified by the orange color (isatin), and the plasmids were isolated from colored colonies. To check for the presence of inserts, sequencing of the isolated plasmids was performed at Macrogen, Korea.

2.5. Protein expression with the selected clones

Selected clones with the $P450_{cam}$ insert (WT or mutants) were grown individually in LB medium (6 L) with kanamycin (50 mg/L) at 37°C, with shaking at 250 rpm overnight with good aeration and the culture(s) were harvested at 7000 rpm (12,300 × g) for 30 minutes at 4°C the next day. Collected pellets were resuspended in fresh LB/kanamycin medium (6 L) and incubated for 2-3 h at 37°C, with shaking at 250 rpm. Supplements, IPTG (80 μM), δ-aminolevulinic acid (1 mM), vitamin B_1 (10 μM) and $FeCl_2$ (0.1 μM) were added when the OD had reached ~0.8 and the protein expression was carried forward at 27°C for 16 h. After induction was complete, the cells were harvested at 7000 rpm for 30 minutes and recovered in a minimum amount of lysis buffer (20 mM phosphate buffer, pH 7.4). Pellets can be stored overnight at this stage. Cell pellets were resuspended in lysis buffer (1 L) and EDTA-Na_2 (1.86 g/L, 0.1 mM) was added. The suspension was stirred for 15 minutes at 4 °C and the pH was adjusted to 7.4 using 0.1 M KOH. Camphor (1 mM, 152 mg in 1 mL of ethanol), 100 mg lysozyme and a protease inhibitor cocktail (40 mg of

PMSF (phenylmethane sulfonyl fluoride), 1 mg AEBSF (4-(2-aminoethyl)benzenesulfonyl fluoride), 0.2 mg leupeptin, 0.2 mg aprotinin) were added and stirring continued at 4°C for 40 minutes. Sonication was carried out in a Branson Ultrasonic sonicator at 50 % duty cycle for 10 minutes. $MgSO_4$ (1.2 g, 10 mM) was added, and the pH was readjusted to 7.4 using 0.1 M KOH. RNase (10 mg) and DNase (1 mg) were added, and the solution was stirred for at 4°C for 30 minutes, followed by sonication for 10 minutes. The lysate was homogenized with a Potter-Elvehjem tissue homogenizer, and then harvested at 7000 rpm (12,300 × g), 4°C for 30 minutes. Supernatant was collected (crude lysate), and absorbance was recorded at 280 nm, 392 nm, 410 nm and 454 nm. The lysate was dialyzed against potassium phosphate buffer (20 mM, pH 7.4, with 200 mM KCl, 1 mM DTT, and protease inhibitors 1 mM (+)-camphor) for 5 hours using 6-10 kDa dialysis tubing. The addition of camphor helps to stabilize the P450. The dialysate was subjected to a 20 % ammonium sulphate cut (to remove cell debris). Since the supernatant showed a band at ~ 410 nm (which corresponds to P450 Soret peak without substrate), it was further subjected to a 45 % ammonium sulphate cut (to isolate the remaining P450 protein). The 20-45 % ammonium sulphate saturated pellet was resuspended in T-100 buffer (50 mM Tris, 100 mM KCl, pH 7.4, supplied with 1 mM 1(R) (+)-camphor).

SDS-PAGE (Sodium dodecyl sulphate polyacrylamide gel electrophoresis) and UV-VIS analysis (at 280 nm, 392 nm, 410 nm and 454 nm) were performed to confirm the presence of cytochrome $P450_{cam}$ in lysate, dialysate, 20-45 % saturated pellet (crude protein). This crude protein was then checked by UV-VIS spectroscopy for the reduced CO difference spectrum (see section 2.10 below) and used for steady-state kinetic assays.

2.6. In vitro assays with the mutated and the wild type $P450_{cam}$ proteins

Enzymatic assays were performed with the $P450_{cam}$ mutant(s) (200 nM), 1 mM 3-chloroindole, 1 mM *m*-CPBA in 1 mL 50 mM potassium phosphate buffer (pH 7.4) with 200 mM KCl and three controls (in the absence of either the substrate, the shunt agent or the enzyme) were run in parallel and the reaction mixtures were incubated for 15 minutes at room temperature. Extraction was performed with $CHCl_3$/indanone (7.2 μM, internal standard) [30], the extract was dried over $MgSO_4$ and the presence of the metabolites was checked by GC-MS. Metabolites formed were compared to their respective standards.

2.7. In vivo screening assay for the oxidation of 3-chloroindole to form isatin

To avoid having to use a shunting agent, we generated an *E. coli* BL21(DE3) strain that contained the entire wild-type P450 system comprised of P450$_{cam}$, putidaredoxin (PdX) and putidaredoxin reductase (PdR) in the pET-22b(+) vector. The three genes are adjacent to each other on the CAM plasmid in *Pseudomonas putida* strain ATCC 17453, the strain from which P450$_{cam}$ was originally isolated [31].

P450$_{cam}$ mutants pre-transformed into *E. coli* strain with pET-22b(+) vector and a tricistronic construct of P450$_{cam}$-PdX-PdR were used in this screening experiment. This strain expresses the WT P450$_{cam}$ enzyme and the two natural redox partners of P450$_{cam}$. As a result, the enzyme does not need to be shunted with an oxidizing agent (such as a peracid) for the oxidation of the substrate. The mutants (in pET-28a(+) vector) have been co-transformed into the *E. coli* strain containing the tricistronic expression vector and double-selected with ampicillin (the resistance marker on pET-22b(+)) and kanamycin (the resistance marker on pET-28a(+) vector)).

In a sterile culture tube with a vented cap, 5 mL of LB media was added with 10 μL of 50 mg/mL ampicillin/kanamycin solution [only ampicillin was added for the WT, whereas for the mutants, both ampicillin and kanamycin were added]. These pre-cultures were allowed to grow at 37°C, 250 rpm overnight. The next day, (Day 0), the starter cultures were centrifuged at 10,000 x g, 4°C for 10 minutes and the pellets were suspended in fresh LB-ampicillin (WT) and LB-ampicillin-kanamycin (mutants) liquid media containing IPTG (1.0 mM) and δ-aminolevulinic acid (1.0 mM). The induced cultures were incubated at 29°C, 250 rpm for 4 hours and then centrifuged at 10, 000 x g, 4°C for 10 minutes. The supernatants were discarded and the pellets were suspended into 1 mL M9 minimal media (Na$_2$HPO$_4$.7H$_2$O, KH$_2$PO$_4$, NaCl, MgSO$_4$, and NH$_4$Cl) and transferred into sterile Eppendorf tubes.

One of the two WT samples was boiled thoroughly for 10 minutes (to be used as a control, see below). The first set was started by adding 300 μL of the appropriate starter culture to tubes that pre-contained 4183 μL M9 minimal media with 17 μL 3-chloroindole stock (62 mM in DMSO, 0.25 mM). The concentration of 3-chloroindole in the culture was 0.25 mM. At time zero, the tube was vortexed and a 1 mL sample was transferred into a sterile Eppendorf tube. The 1 mL

samples were centrifuged thoroughly at 10,000 × g, 4°C for 10 minutes. The supernatants were carefully transferred into new Eppendorf tubes and monitored for 3-chloroindole and isatin via UV-VIS spectroscopy. The remaining 3.5 mL culture in each tube was placed in an incubator at 27 °C, with shaking at 250 rpm until the next set on days 3 and 6.

During the screening experiment, the appearance of isatin was measured at 315 nm in the cell supernatant, Fig. S7. Three replicates were done per set whereby each set consisted of the 19 mutants, the live WT, the boiled control, and the M9 media-only control. Measurements were taken at time zero (just after setup), then 3, and 6 days after setup.

2.8. Transformation and expression of P450$_{cam}$ mutants into an E.coli strain that co-expresses putidaredoxin and putidaredoxin reductase

Construction of bicistronic expression vector

A bicistronic recombinant expression vector containing putidaredoxin reductase (PdR, CamA) and putidaredoxin (PdX, CamB) was constructed as described below. The forward primer consisted of the vector complementary (pET-30 Xa/LIC) overhang and the N-terminal sequence of PdR (starting with the codon GTG). The reverse primer consisted of the C-terminal PdX sequence with its stop codon (TAA) replaced by alanine, a linker consisting of Factor Xa protease recognition sequence (IEGRS), His$_6$ sequence with a stop codon (TAA) followed by GG (for T4 DNA polymerase activity) and additional sequence complementary to the vector overhang. Primer details are tabulated in the supplementary information (Table S1).

The open reading frames (NCBI accession # D00528.1) of PdR (CamA, 1269 bp) and PdX (Cam B, 324 bp) were amplified by colony PCR using *P. putida* (ATCC 17453) colonies and the primers (Table S1). PCR amplification was done using 2 mM MgCl$_2$, 0.2 mM dNTP mix, 0.5 µM primers and DNA polymerase enzyme using Roche (Expand High Fidelity, # 11732641001) PCR system. The PCR program was: 3 cycles of denaturation at 94 °C for 3 min, annealing at 67 °C for 1 min, and extension at 76 °C for 2 min., followed by 18 cycles of 94 °C for 45 seconds, 70.3 °C for 1.25 min, and 76 °C for 1.55 min. The resulting 1716 bp long PCR amplified product was agarose gel purified using QIAquick® gel extraction kit (# 28704, Qiagen). Purified PCR product (0.01 pmol/µL) was then treated with T4 DNA polymerase (0.02 µL/µL reaction, #70099, Novagen) in the presence of dGTP (2.5 mM/20 µL reaction, # 70026, Novagen) and DTT (5 mM/µL reaction, # 69842, Novagen) at 22 °C for 30

min to create vector complementary overhangs (sticky ends). T4 DNA polymerase treated PCR product (2 µL or 0.02 pmol) was then annealed with the linearized pET-30Xa/LIC vector (1 µL, # 70022-1UG, Novagen) with the complementary sticky ends at 22 °C for 5 min, followed by addition of 25 mM EDTA (1 µL, # 70098, Novagen) and continued incubating at 22 °C for 30 min. The annealed vector (1 µL) was then transformed into 50 µL of NovaBlue GigaSinglesTM competent cells (# 71227, Novagen) for initial cloning, as follows: the mixture was incubated on ice for 5 min, followed by heat shock for 30 s at 42 °C and immediate incubation on ice for 2 min. To the mixture SOC medium (# 69319, Novagen, 250 µL) was added at room temperature and incubated at 37 °C with shaking (250 rpm) for 2 hours. This culture (50 µL) was then plated on a SOC-agar plate with kanamycin (30 mg/L) and incubated overnight at 37 °C in an incubator. A single colony was transferred to 5 mL SOC liquid medium with kanamycin (30 mg/L) and incubated overnight at 37 °C and 250 rpm. The overnight grown culture was used as an inoculum for a 50 mL LB liquid medium/kanamycin (30 mg/L) and incubated at 37 °C, 250 rpm, for 15 hours. The cells from this culture were used to isolate pure recombinant vector (bicistronic Pdr+Pdx) using the Qiagen (QIAprep® Spin miniprep, # 27104) kit. The isolated plasmid DNA was then sequenced (Eurofins Genomics, Louisville, KY) to verify the sequence of the bicistronic system. Isolated plasmid was then transformed into the expression host, *E. coli* BL21(DE3) competent cells (# 69450, Novagen) as described earlier.

Plasmids of P450$_{cam}$ mutants were transformed into a strain of *E. coli* BL21(DE3) cells that contain a bicistronic construct consisting of putidaredoxin (PdX) and putidaredoxin reductase (PdR) from *Pseudomonas putida* using the heat shock method.

BL21(DE3) competent cells with PdR and PdX (100 µL) were prepared and10 µL of 1 ng/µL of P450$_{cam}$ plasmid was supplemented. The sample was incubated on ice (4°C) for 45 minutes to allow the plasmid to absorb on the surface of the cells. The sample was heat-shocked using a water bath set at exactly 42°C for 30 seconds. Then, the sample was placed on ice for 2 minutes. SOC medium (1 mL) was added to the tube and the sample was incubated at 37°C, 250 rpm for 3 hours. Aliquots with different volumes, ranging from 100 µL to 200 µL, were plated on LB agar plates containing 50 µg/mL of the antibiotic (ampicillin/kanamycin) and incubated overnight at 27°C, 250 rpm. Individual colonies were picked with a toothpick, grown in 2 mL LB medium containing the antibiotic, and stored as glycerol stocks.

The *E. coli* strains with P450 mutants, PdX and PdR were checked for expression of all three proteins by Western blotting with anti PdX, anti PdR and anti P450$_{cam}$ antibodies (polyclonal, produced in rabbit), with pre-immune serum controls [33]. All strains were found to express all three proteins.

Each *E.coli* strain containing the mutated cytochrome P450$_{cam}$ plasmid, putidaredoxin (PdX) and putidaredoxin reductase (PdR) was grown in LB medium (6 L) with kanamycin (50 mg/L) at 37°C, 250 rpm overnight until an optical density of ~0.8 - 0.9 was reached. Cells were harvested by centrifugation at 7000 rpm, 4°C, for 30 minutes. Collected pellets were resuspended in fresh LB-kanamycin medium (6 L) and incubated for 2 - 3 hours at 37°C, 250 rpm. IPTG (1.0 mM) and trace additives (FeCl$_2$ (0.1 μM), 5-aminolevulinic acid (1.0 mM), vitamin B$_1$ (10 μM)), riboflavin (10 μM) were added [33]. Incubation was continued for 4 - 5 hours at 27° C, with shaking at 250 rpm. After induction was complete, the cells were harvested at 7000 rpm (12300 $\times$ g) for 30 minutes and stored in a minimum amount of lysis buffer (20 mM phosphate buffer, pH 7.4) overnight.

Stored cell pellets were resuspended in lysis buffer (1 L), and disodium EDTA (0.1 mM) was added. The suspension was stirred for 15 minutes at 4°C and the pH was adjusted to 7.4 using 0.1 M KOH. (+)-Camphor (152 mg in 1 mL of ethanol, 1 mM) and protease inhibitor cocktail: PMSF (100 μL of 40 mg/mL in EtOH), 100 μL of a mixture of 1 mg AEBSF, 0.2 mg leupeptin, 0.2 mg aprotinin, and 100 mg lysozyme were added and stirring continued at 4°C for 40 minutes.

The lysate was sonicated, homogenized and dialyzed as explained in section 2.5 SDS-PAGE (Sodium dodecyl sulphate polyacrylamide gel electrophoresis) and UV analysis (at 280 nm, 392 nm, 410 nm and 454 nm, see section 2.10 below) were performed to confirm the presence of cytochrome P450$_{cam}$ in the dialyzed lysate. The crude protein was then used for steady-state kinetic assays.

2.9. Steady-state kinetic assays for 3-chloroindole hydroxylation with the P450$_{cam}$ mutant(s) with a NADH regeneration system

Steady-state kinetic assays were performed in 1 mL potassium phosphate buffer (50 mM, 150 mM K$^+$, pH 7.4) with varying concentrations of 3-chloroindole (ranging from 10 μM to 1 mM),

200 nM of crude protein P450$_{cam}$, 2 mM glucose, 0.05 units glucose dehydrogenase, 0.1 mM NAD$^+$ and 1 μM NADH. The reaction mixtures were monitored by UV-VIS spectroscopy for 15 min at 246 nm for the formation of isatin. The concentrations of the product, isatin **7**, were estimated using extinction coefficient (2.3 mM^{-1}cm^{-1}), determined by us using a pure standard. Kinetic parameters such as V_{max} and K_M were obtained using nonlinear regression with GraphPad Prism software. Three controls were run: 1) in the absence of the substrate, 2) without NADH regeneration system, or 3) without P450$_{cam}$.

The kinetic assay was performed by first adding the phosphate buffer, glucose, glucose dehydrogenase, NADH, P450$_{cam}$ protein, then bubbling with O$_2$, taking a blank reading, then adding the 3-chloroindole and recording the absorbance at 246 nm for 15 min.

2.10. Determination of Fe-CO absorbance at 450 nm with Carbon Monoxide Releasing Molecule-3, CORM-3

To 1 mL phosphate buffer (50 mM, 150 mM K$^+$, pH 7.4), mutant P450$_{cam}$ (200 nM or 9 μM) and sodium dithionite (1 mM) or NADH (5 mM) was added. The reduced baseline spectrum was recorded between 400-500 nm in 1 nm intervals. After adding a few grains of CORM-3, (tricarbonylchloro (glycinato) ruthenium (II), Sigma-Aldrich) or bubbling CO gas, the mixture was scanned 5 to 6 times until the Fe^{II}-CO complex at 450 nm was identified (Fig. S5). All measurements were done in triplicate.

2.11. Docking simulations

Docking simulations were performed using Molecular Operating Environment (MOE, Chemical Computing Group, Montreal, Canada). The amino acid sequence and crystal structure information of P450$_{cam}$ (CYP101A1) was obtained from Protein Data Bank. P450$_{cam}$ with PDB codes 1DZ9 (iron-oxo complex of P450$_{cam}$) and 3L63 (closed conformation of P450$_{cam}$ in the presence of a substrate) were used for the docking studies [34,35]. The protein was imported into MOE as a PDB structure, each residue protonated at pH 7, 298 K and 0.1 M salt and the charges were assigned according to the default setting under "Compute| Pronate 3D"[36]. Ligands 3-chloroindole, 3-chloro-3-hydroxy-indole, (2S,3S)-2,3-epoxy-3-chloroindole, (2R,3R)-

2,3-epoxy-3-chloroindole and 3-ketoindole (Fig. S7) were constructed in MOE using "builder" and imported into the MOE database as ".mdb" files. Prior to docking, camphor was deleted and potential docking sites were identified on the protein (by setting "all atoms" (protein and heme) as the receptor) using the option "site finder." Dummy atoms were placed at each site targeted for docking and these dummy atoms were markers used by the docking algorithm. Ligands were docked using an induced fit protocol and all atoms of the P450 polypeptide and the porphyrin chosen as the receptor. Triangle matcher was used for placement, rescored by London dG, force field for refinement, rescored using GBVI/WSA dG; a maximum of 30 poses were retained. The populations of poses—structures that result and are retained from the docking trajectory, were weighed in importance according to their preponderance if all poses retained were to equilibrate. The distribution of poses was estimated based on each structure's energy. We assumed that the energy from MOE is approximately equal to the free energy.

$$\Delta E_{conf} = (E_{i+1} - E_i)cal \times 4.184\ ^{J}/_{cal}$$

$$lnK_{\frac{i}{i+1}} = -\left(\Delta E_{conf}\right) \div (8.314\ \frac{J}{mol \times K} \times T\ (K)\ )$$

$$K_{\frac{i}{i+1}} = e^{lnK_{\frac{i}{i+1}}}$$

% of preponderance all poses obtained from normalization of the above function with all poses.

frequency of pose, F = % of preponderance of each pose $\times$ E$_{conf}$

weighted average distance $= \Sigma\ F$

Conformational energy (E$_{conf}$) is obtained from docking results database in kcal.

To distinguish two possible reaction pathways of 3-chloroindole dechlorination (see Fig. 2), which differ in the site of initial substrate oxidation by Compound I of P450, we measured the distance between Fe and N1 of 3-chloroindole, as well as between Fe and carbons 2 and 3 of 3-chloroindole for all retained poses of each mutant. Average Fe-N1, Fe-C2 and Fe-C3 distances were compared by one-way ANOVA using GraphPad Prism V 5.04 (San Diego, CA).

3. Results

3.1. Selection of the SeSaM library with 3-chloroindole or endosulfan

Error prone PCR methods are highly biased towards transition mutations which generally lead to chemically conserved amino acid substitutions, whereas transversion mutations lead to chemically different amino acid substitutions [37]. Transversion biased SeSaM-Tv-II [15] and

SeSaM-Tv P/P [24] methods, developed in Schwaneberg group, explore such protein sequence space. Directed evolution campaigns employing the above SeSaM methods have been successful in engineering proteins for non- natural cofactors [38] and thermal resistance [39, 40]. Since enzyme activity was desired on a non-natural substrate, the tranversion-enriched SeSaM-Tv P/P method [24] was employed for construction of the mutant library. The mutant $P450_{cam}$ library ($\sim 10^6$ variants) [9] with a mutation frequency (substitution/base: 3×10^{-3}) was selected with 3-chloroindole as described in section 2.4 of the experimental section (Fig. S8). We have found that 3-chloroindole was toxic to bacteria (IC_{50} 64 µM), as was endosulfan (IC_{50} 27 µM) (Fig. S9).

Subjecting of the mutant library (with approximately 3 mutations/clone, based on the sequencing results of 12 clones) to selection with 120 µM 3-chloroindole ($\sim 2\times$ the IC_{50} concentration) under shunt conditions, using *m*-CPBA, resulted in ~ 30 colonies per plate out of which at least 15 were orange or orange with a blue rim, due to the formation of isatin **7** or indigo **8**, respectively. In the absence of shunt or the expression additives, and with the wild-type control, the colonies formed (~ 5-7) were only white. Plasmids isolated from the orange (10) and the white colonies (2) were sequenced and the mutants obtained from colored colonies are listed in Table 1. These mutants of $P450_{cam}$ were coded "IND".

The $P450_{cam}$ mutant library was also subjected to selection with endosulfan (at IC_{90} 250 μ M) using shunt conditions (*m*-CPBA). Mutants that survived were able to metabolize endosulfan. The surviving seven mutant colonies in treatment with endosulfan were sequenced. These mutants of $P450_{cam}$ were coded "ES". In the control without endosulfan, $(9.2 \pm 0.2) \times 10^5$ CFU/mL were obtained. One of the ES mutants (ES1) was included in steady-state kinetic analysis with 3-chloroindole (see next section) to compare its activity with that of IND mutants, because IND4 and ES1 both had a mutation of residue 297 (Table 1). The sequence results and study of endosulfan biodegradation by $P450_{cam}$ mutants will be described elsewhere.

In vitro assays were performed with all the sequenced mutants, and isatin was detected as a major product with the mutants using UV-VIS spectroscopy (Fig. S10). Isatin from these reactions was isolated by reverse-phase HPLC and analyzed by GC-MS (Fig. 2), ^{1}H and ^{13}C NMR (Figs. S11-S13). The compound matched an authentic isatin standard.

16

3.2. Steady-state kinetic analysis of the isolated P450$_{cam}$ wild type and mutants

The rates of 3-chloroindole oxidation were measured with the selected mutants and normalized for the amount of P450$_{cam}$ (nmol) taken in the reaction (Table 1). The rates of isatin formation were highest for the mutants IND1 (E156G/V247F/V253G/F256S), IND6 (G93C/K314R/L319M) and IND2 (G60S/Y75H). The WT enzyme was not active (Table 1), yet in the *in vivo* screen we did see some isatin forming in the WT culture, particularly on days 3 and 6 (Fig. S10). A possible reason for this is the slow utilization of endogenous indole in other oxidation processes. From the kinetics (Table 1) and the docking simulations (see section 3.4) it is clear that the WT did not accept 3-chloroindole as a substrate.

The ranking of the IND mutants, according to their maximal velocity, V_{max}, seen here in assays with the electron transfer partners, NADH and O_2, generally parallels the rates found in assays with *m*-CPBA shunt [6]. Thus, IND1 had the highest rate with the shunting agent and also the highest V_{max} in this study (Table 1), and IND5 was lowest in both cases. The only exception to this parallel ranking was IND6: it had the second highest V_{max} here (Table 1), but had the lowest rate (along with IND5) with a shunting agent [6]. Consistent with previous data with the shunting agent [6], with the kinetic data (Table 1) and docking simulations obtained here (see below), the WT did not show activity with 3-chloroindole as substrate.

| Table 1. Kinetic data from oxidation of 3-chloroindole to isatin by P450$_{cam}$ mutants and wild type. [a]					
Clone Name	Mutations	V_{max} (M/s)	K_M (M)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1}s^{-1})
WT	none	N/D[b]	N/D	N/D	N/D
IND1	E156G/V247F/V253G/F256S	3.73 E-8	4.36 E-4	1.86 E-1	4.27 E+2
IND2	G60S/Y75H	1.21 E-8	4.06 E-5	6.03 E-2	1.49 E+3
IND3	D97F/P122L/Q183L/L244Q	2.17 E-9	9.03 E-5	1.08 E-2	1.20 E+2
IND4	G120A/Y179H/G248S/D297H	1.07 E-9	4.64 E-6	5.34 E-3	1.15 E+3
IND5	Y179H	2.70 E-10	2.64 E-5	1.35 E-3	5.11 E+1
IND6	G93C/K314R/L319M	1.24 E-8	1.47 E-4	6.20 E-2	4.23 E+2
ES1	T56A/N116H/D297N	5.40 E-9	1.61 E-5	2.70 E-2	1.68 E+3
[a] The assay used was the one with the NADH regeneration system (see section 2.9 above) [b] N/D = not detected					

*3.3. Proposed pathways for the formation of isatin*

Routes A and B, shown in Fig. 2, are proposed as plausible reaction pathways for the oxidation of 3-chloroindole to isatin by cytochrome $P450_{cam}$ mutants. In both cases two consecutive oxidations by $P450_{cam}$ generate isatin. In pathway A, we propose that the substrate 3-chloroindole **5** forms an epoxide upon oxidation by $P450_{cam}$ mutants (Fig. 2). The epoxide intermediate **9** then ring-opens to form 3-chlorohydrin **10**. In pathway B, the N of 3-chloroindole is oxidized by Compound I to give a radical cation and Compound II (Fig. 2). This radical cation protonates compound II, which then transfers the hydroxyl radical to C-3 to furnish 3-chlorohydrin **10**. In both pathways, chlorohydrin **10** then loses hydrogen chloride to form 3-ketoindole **11**. A hydroxylation of **11** at C-2 by $P450_{cam}$ mutants forms the minor tautomer of isatin **7**. The major and minor tautomers of isatin can be clearly seen in the NMR of this compound (Figs. S12 & S13). Pathway A would require C2-C3 of the substrate to be positioned closest to the heme iron. Pathway B would require NH to be closest to the heme iron. Furthermore, since both pathways involve 3-ketoindole hydroxylation, the various mutants could also differ in this part of the reaction, and this would require that more active mutants have C-2 of 3-ketoindole closer to Fe than less active ones.

Pathway A would require a general acid near the epoxy oxygen, to open the epoxide to the chlorohydrin. Both pathways would require a general base catalyst near the chlorohydrin OH to produce 3-ketocamphor.

Reaction Pathway A
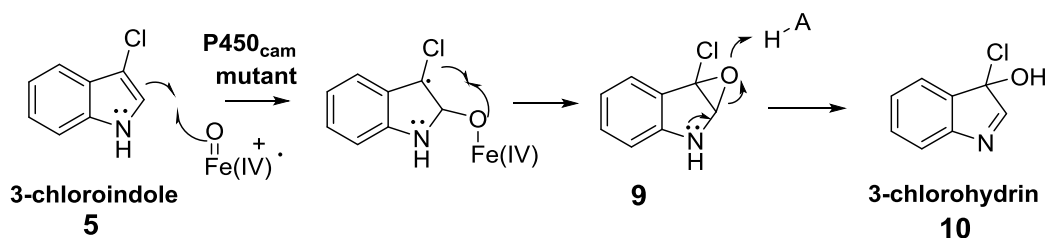
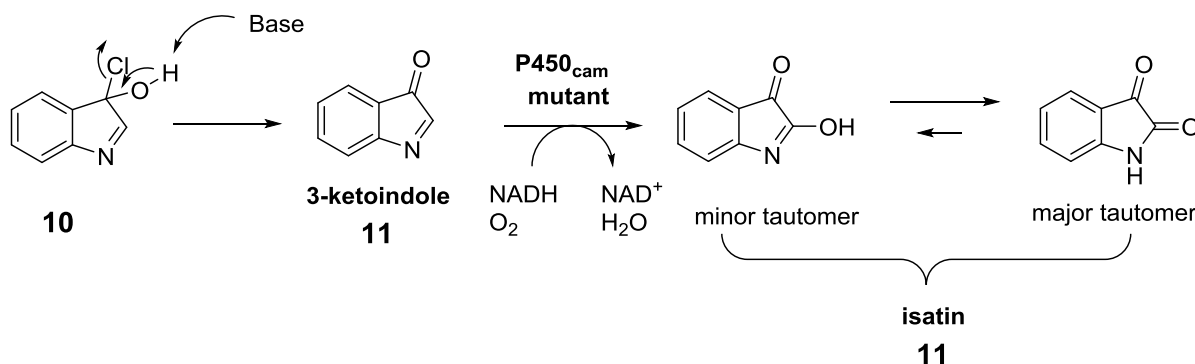**Reaction Pathway B**
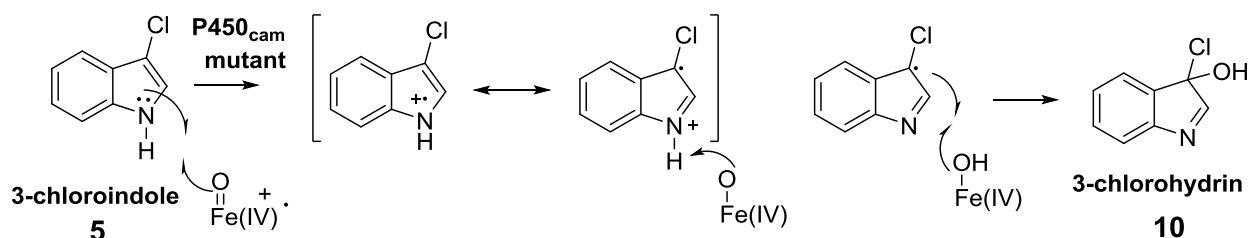
Both Pathways

Figure 2. Two proposed reaction pathways (A, B) in the oxidation of 3-chloroindole by P450$_{cam}$ mutants to form chlorohydrin **10**. The chlorohydrin has to eliminate HCl in both cases to give 3-ketoindole **11**, and this could be hydroxylated at C-2 to give isatin.

3.4. Docking of 3-chloroindole in the active site of P450$_{cam}$

To gain insight into how 3-chloroindole fits into the active sites of mutants and wild type P450$_{cam}$, *in silico* docking calculations were performed. First, 3-chloroindole was docked into the active site of P450$_{cam}$ (WT as well as mutant IND1 E156G/V247F/V253G/F256S, which had the highest V_{max}). Amino acid positions V247, V253, and F256 can be seen close to the active site

heme of WT P450$_{cam}$. The WT enzyme showed no activity with 3-chloroindole conversion to isatin in the *in vitro* assays, and the docking suggests that 3-chloroindole does not fit in the active site of WT P450$_{cam}$. V247F, F256S and V253 mutations create a more open cavity in the active site (see below). For example, it is interesting to note that in the model of the mutant, F247 and the 3-chloroindole are exactly orthogonal, (Fig. S14), suggesting that a sp^2 C-H/π interaction may play a role in positioning 3-chloroindole above the porphyrin.

Superposition of the model of each mutant on that of WT revealed that many of the mutations seen in this collection were outside of the active site (Fig. 3). A few mutations occurred within the active site: V247F, V253G, F256S of IND1, L244Q of IND3, G248S, D297H of IND4 and D297N of ES1. Among mutations on the exterior of the protein, Y179H is of particular interest, as it occurred twice (once in IND4 with three other changes and once alone in IND5). The side chain of residue 179 lines the external site where 3-chloroindole docked in the WT P450$_{cam}$ (Fig.3).
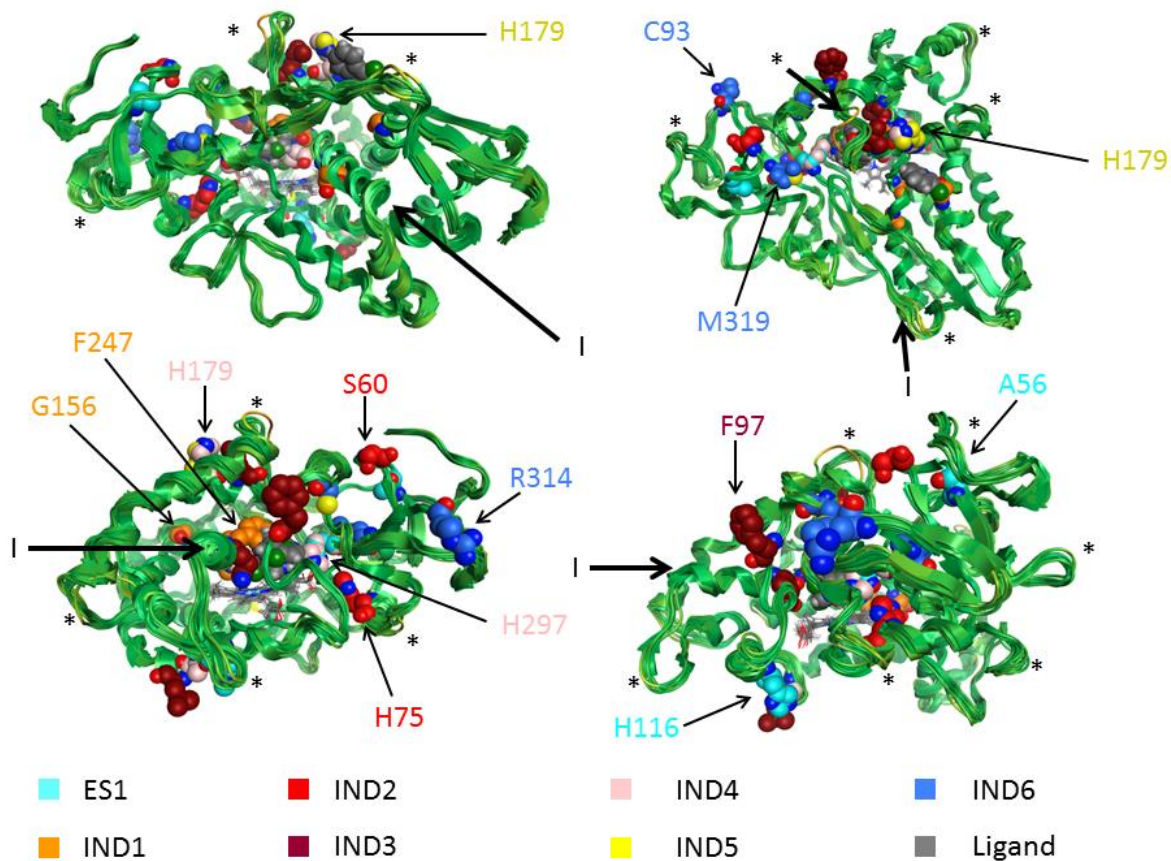
Figure 3. Location of the mutations in mutants IND1, IND2, IND3, IND4, IND5, IND6 and ES1. Different views of superimposed, energy minimized models of the mutants and the wild-type. The carbon atoms of selected mutated residues (shown as space-filling structures) are shown in the color indicated for each mutant. The porphyrin and cys357 are shown in stick models and the ligand (3-chloroindole) has its carbon atoms colored gray. The I-helix is indicated with "I". The backbone ribbon is colored according to the RMSD (for comparison of each mutant to wild-type). Portions of backbone that do not differ significantly from WT are dark green, whereas portions that differ are light green. Loops that changed most in one or more of the mutants are marked with "*".

Table 2. Results from superposition of WT and individual mutant structures with 3-chloroindole docked				
Superposition	Overall RMSD (Å) [a]	Residues with RMSD > 2.0 [a]	Residues with altered side chain positions [b]	Residues with small side chain position alteration [b]
WT and IND1 E156G/V247F/V253G/F256S	0.961	A36, Y78, G93, P215, G248	F87[c], **D251**[e], F163, <u>K178</u>[e], E183, <u>D182</u>[e], **T252**[c]	Y96, Y179
WT and IND2 G60S/Y75H	0.924	S35, A36	Y96, F98[d], F163, <u>D182</u>[e], L244, H352[d]	F87, E152, <u>K178</u>[e], Y179, T185, D188, **D251**[e], I367
WT and IND3 D97F/P122L/Q183L/L244Q	0.814	D188, G189, S190, M191	F87[c], Y96, F98, E152, P175, <u>K178</u>[e], Y179, <u>D182</u>[e], T185, **D251**[e], I367,	Y75, F163, L294, H352
WT and IND4 G120A/Y179H/G248S/D297H	0.922	P106, G120, M121, P122, P215, G386	F87[c], V119, E152, F163, <u>K178</u>[e], <u>D182</u>[e], F193, L246, **D251**[d,e], **T252**	Y75, Y96, F98, P100, *R112*, D156, T185 F256, F350
WT and IND5 Y179H	0.773	P215, H308, G386	F87[d], Y96, F98, P100, Q108, *R112*, V119, E152, F163, L166, <u>K178</u>[e], T185, H355[d], L358, E366	F256, F350, H352
WT and IND6 G93C/K314R/L319M	0.812	P13, S35, D188, G189, G386	F87[c], Y96, F98[d], P100, <u>K178</u>[e], <u>D182</u>[e], T185, **T252**, L362, I367	P86, F163, **D251**[e] F256, F350, V396
WT and ES1 T56A/N116H/D297N	0.934	S32, A36, E76, Y78, D188, G386	Y75, F87[c], P100, E152, E156[d], <u>K178</u>[e], F193, V247, H355[d]	Q108, F163, <u>D182</u>[e], **D251**[e], F256, L289, F350, H352, L362

[a] Root Mean Square Deviation of backbone atoms, determined during superposition
[b] Determined through manual inspection of the site around the docked ligand and porphyrin (through the site view function in MOE)
[c] Large movement of the side chain

[d] 90[o] rotation of the functional group in the side chain (phenyl ring in F, imidazole in H, carboxylate in D and E).
[e] Residue is part of a salt bridge.
Underlined residues have coupled movements due to a strong interaction (e.g. salt bridge or H-bond).
Bold residues are involved in the catalytic cycle.
Italicized residues are known to be involved in electron transfer.

Superposition of each mutant onto the structure of WT $P450_{cam}$ with 3-chloroindole docked revealed regions where the backbone changed (Table 2, residues with RMSD > 2.0) and residues with altered side-chain positions (Table 2). Most notably, the mutations all caused movement in the side chain of F87 (Table 2), which is coupled to the neighboring side chains of Y96 and F98. In the WT, the phenyl ring of F87 projects deeply into the pocket above the heme. In the mutants, the phenyl group of F87 rotated or moved out of the site, such that the flat 3-chloroindole fits into the pocket above heme (Fig. 4). The mutations also changed side chain positions in P100, K178, D251, and T252 (Table 2, Fig. 4). Together, these motions likely changed the overall shape and flexibility of the active site, as well as the flexibility of the substrate access channel.

The various mutants differed in the manner in which they positioned 3-chloroindole above heme (IND1, IND3 and IND5 stacked on heme; IND4 and ES1 perpendicular to heme; IND2 and IND6 at an intermediate angle relative to heme). Importantly, all the mutants were able to position 3-chloroindole in their active site, whereas WT was not.
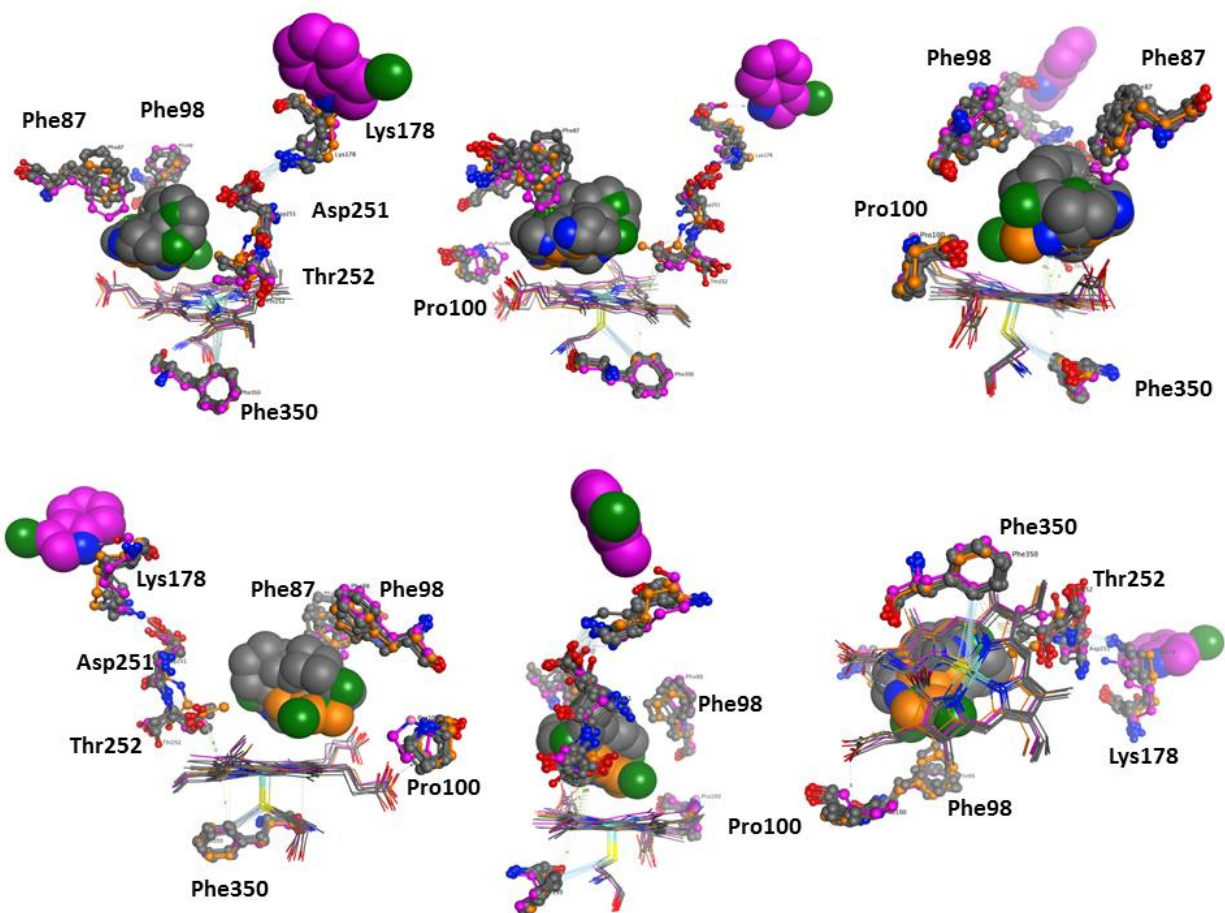
Figure 4. Six rotated views of the active sites of superimposed mutants and WT, with residues whose side chain positions differed significantly between the mutants and the WT: Phe87, Phe98, Pro100, Lys178, Asp251 and Thr252. Phe350 is shown for comparison; it is a residue whose side chain did not change very much between WT and mutants. The heme and cys357 are shown as stick models, the residues of interest are shown as ball-and-stick models and the ligand (3-chloroindole) is shown as space-filling models. The structures belonging to the WT are shown in fuchsia; structures belonging to IND1 are shown in orange.

We docked 3-chloroindole, the two enantiomers of 2-3-epoxy-3-chloroindole, the chlorohydrin and 3-ketoindole into the WT and all mutant enzymes. We then measured distances from Fe to N, C-2 and C-3 for the simulations with 3-chloroindole. Each simulation yielded multiple retained poses, so the distances measured were weighted according to the energy of the poses, and an average distance was determined for that simulated population of poses. The most important result of these *in silico* docking simulations is that the WT showed ≥ 2 times the distances between Fe and N, C-2 and C-3 of 3-chloroindole and all intermediates docked than were found for the mutants, and this was significant (one-way analysis of variance, ANOVA,

Tukey's test, $p < 0.05$). Thus, the steady-state kinetic assays and the docking studies show that WT P450$_{cam}$ does not accept 3-chloroindole as substrate. The second result is that the mutants did not differ significantly in the Fe-N, Fe-C-2 or Fe-C-3 distances for 3-chloroindole (one-way analysis of variance, Tukey's test, $p > 0.1$). Therefore, it is not possible to state which pathway in Fig. 2 is applicable to the P450$_{cam}$ mutants studied here.

For simulations with epoxy intermediates, 3-chloro-3-hydroxyindole and 3-ketoindole (Fig. S6) we focused on residues that interact with the epoxide or chlorohydrin moieties recurrently, in the majority of poses obtained. Since both pathways involve 3-ketoindole, we also checked the distance of C-2 of ketoindole to the Fe porphyrin, since that position would need to be hydroxylated to obtain the isatin hydroxyl tautomer. The most important result was that IND1 had the shortest Fe-C-2 distance (Fig. 5).
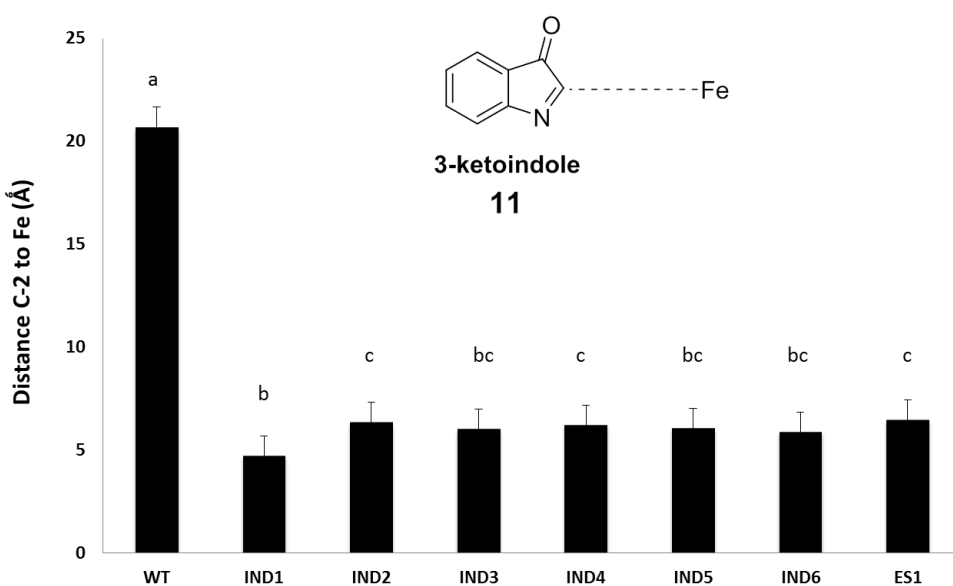

Figure 5. Average distances for all poses of 3-ketoindole docked into the active sites of WT and mutant P450$_{cam}$. Bars show the mean ± S. E. of the distance of C-2 to Fe in Ångstrom. Columns marked with different letters differed significantly (ANOVA, Tukey $p < 0.05$).

4. Discussion

One of the main difficulties in the alteration of P450$_{cam}$ towards new substrates remains its poor affinity for non-natural substrates [41]. The catalytic cycle involves the expulsion of the axially coordinated water molecule from the porphyrin upon substrate binding, and this is believed to shift the spin equilibrium, as well as raise the redox potential of the P450$_{cam}$ [34]. Poor substrate binding causes an incomplete spin transition of the heme in the active site, which decreases the catalytic efficiency of the enzyme [42]. Therefore, to apply P450$_{cam}$ to new substrates, there is need for engineering the active site of the enzyme to enhance the substrate binding affinity, which then enhances the catalytic turnover of the enzyme. Researchers have attempted to design P450$_{cam}$ variants by either increasing the active site volume for larger substrate molecules [43], or enhancing the substrate binding affinity by complementary non-covalent interactions between the substrate and the suitable amino acid chains in the active site [42].

Previous literature suggests that the wild type P450$_{cam}$ has no activity towards indole [44]. Site-directed mutagenesis of P450$_{cam}$ active site was attempted before by researchers to obtain activity towards indole [19, 20, 45]. The amino acids, F87, Y96, T101 and L244 in the active site of the enzyme were mutated, and the mutants were assayed with indole and various substituted indoles [42], but not with 3-chloroindole. The hydrophobic residue V247, located above heme, was mutated to leucine, to change the shape of the cavity above heme and detect oxidation of a variety of unnatural substrates that included α-pinene, aromatic halides and short-chain alkanes [46-50]. Here, we report that IND1 (E156G/V247F/V253G/F256S) has the highest rate of conversion (V_{max}) of 3-chloroindole to isatin per second. IND1 carries a novel mutation (V247F) that can play a role in altering the activity of P450$_{cam}$ towards the dehalogenation of 3-chloroindole. The additional mutations, V253G and F256S create space in the active site. Residue F247 and 3-chloroindole are orthogonal in the *in-silico* models, suggesting a sp^2 C-H/π interaction (Fig. S14). It has been reported before that unnatural substrates may not fit well into the P450$_{cam}$ active site, either because they are too small and do not fill the cavity or because they are too large and cannot position above the Fe, and are therefore not turned over [46]. In our case, 3-chloroindole was well positioned in the mutant active site for turnover.

Kinetic results show that IND2 has one of the highest catalytic efficiencies (k_{cat}/K_M), (Table 1, Fig. S15). Both residues that were mutated, Gly60 and Tyr75, lie far away from the active site,

along the substrate access channel. Hence, the new residues Ser60 and His75 may play a role in accommodating 3-chloroindole into the active site, by causing changes in backbone and side-chain positions within the active site (Table 2, Fig. 4).

IND3 exhibited a 17-fold decrease in activity (V_{max}) compared to the IND1 mutant. Here, the mutated residues were along the K^+ binding site. D97F is close to B' helix end. Gln244 interacts with the chlorine atom of 3-chloroindole, (Fig. S16). L244Q is on the I helix and interacts with the chlorine atom of 3-chloroindole. P122L is on the C' helix and is a surface residue of $P450_{cam}$. In the literature, Pro122 is among the few residues that interacts with putidaredoxin (PdX) residue Trp106 [51]. Hence, P122L may affect the PdX-induced structural changes in $P450_{cam}$. D97F is close to the B' helix end. It has been reported that residues Asp97 and Lys197 form a salt link via hydrogen bonding interactions, and are reported to play a role in the ligand egress pathway [52]. Q183L is on the F helix of $P450_{cam}$, on the exterior of the protein; it lines the site where 3-chloroindole docked in the wild-type.

IND4 had the highest binding affinity (lowest K_M). Residues Ser248 and His297 form a cage around 3-chloroindole, as shown in Fig. S17. In the wild type, the carboxylate side chain of Asp297 forms a hydrogen bond to a propionate of the heme prosthetic group. Mutation of Tyr 179 to His also occurred alone in mutant IND5 (Y179H). Kinetic results showed that IND5 still converted 3-chloroindole to isatin, but had the lowest rate of formation at saturating conditions. Residue 179 lies on the F-helix of $P450_{cam}$ (Fig. S18) and lines the site where 3-chloroindole docked in the WT protein.

It has been reported that residues D251, K178, D182 and R186 belong to a salt link tetrad in the wild-type enzyme [53]. Salt bridges occur frequently in proteins and can contribute to molecular recognition and catalysis [54]. The replacement of a nearby tyrosine residue at position 179 by histidine (mutants IND4 and IND5) caused a significant change in the activity of $P450_{cam}$ and in its ability to accommodate 3-chloroindole as a substrate. From inspection of the docked models and the superposition of WT and mutants it is apparent that mutation of Y179 to H caused changes in the salt link tetrad, particularly in K178, D182 and D251.

IND6 (G93C/K314R/L319M) had the lowest binding affinity (highest K_M) towards 3-chloroindole and the second highest V_{max} (Fig.S19). Mutation at residue K314 was one of the recurring mutations that were observed in the current P450$_{cam}$ SeSaM procedure and selection of clones. Here, lysine was substituted for another positively charged and amphipathic amino acid, arginine. The guanidinium group of arginine can bind negatively charged groups on carboxylates and can form multiple hydrogen bonds [55]. Hence, this K314R mutation could enhance protein stability. Interestingly, IND6 had the lowest rate of isatin formation in experiments with *m*-CPBA shunt (see above) [6].

Residue 93 (mutated G93C in IND6) is part of the B' helix, which is strongly perturbed by the binding of reduced putidaredoxin (PdX) to the reduced substrate-bound P450$_{cam}$. Residues G93, E84, E94 and Y96 are part of the K$^+$ binding site of P450$_{cam}$ [56]. Residue 319 (mutated L319M in IND6) lies on the surface of P450$_{cam}$, and thus may not be directly involved in catalysis but may play a role in substrate binding and recognition [55].

Substitutions at position D297 occurred in two cases: IND4 (G120A/Y179H/G248S/D297H) and ES1 (T56A/N116H/D297N). The carboxylate group of the D297 side chain is reported to have specific interactions with one of the heme propionate groups and the mutation, D297M is reported to have reduced oxidation activity with butane and propane [46]. In our case, D297N, along with T56A and N116H in ES1 mutant resulted in the highest catalytic efficiency and rate of formation at low substrate concentration as reflected in the value of k_{cat}/K_M (Table 1, Fig. S19). At residue 297, aspartic acid was replaced by asparagine. In the literature, it was shown that the structure of a D297N P450$_{cam}$ mutant superimposes well with that of wild type P450$_{cam}$. On the other hand, substitution of D297 with Ala or Leu were different, especially in terms of the behavior of T101 upon substrate binding [57]. In the wild type and D297N mutant, T101 hydrogen bonds to Y96 in the camphor-free form, and upon substrate binding, T101 forms a hydrogen bond with the heme-7-propionate. With the mutant D297N, the amide group of N298 side chain hydrogen bonds to the heme-7-propionate, and forms a tetrad hydrogen bonding network with Q322 and R299, which allow P450$_{cam}$ to bind with the substrate and block the access of bulk water to the active site [57]. This hydrogen bonding interaction between N297

side chain and the heme was observed in the molecular docking simulations with the ES1 (T56A/N116H/D297N) mutant (Fig. S20).

In the literature, it is shown that putidaredoxin (PdX) residue W106 forms hydrogen bonds with P450$_{cam}$ residues R108 and N116 [58]. This interaction is the key driving force in the closed to open switch of P450$_{cam}$. These hydrogen bonds force the C helix up toward PdX by about 2-3 Å, and this motion is coupled with movements in the B', I, F and G helices, all of which are involved with substrate access. This movement also leads to large motions of the F and G helices and the F/G loop which opens the active site to bulk solvent. Residue 116 (mutated N116H in ES1) lies on the C helix of P450$_{cam}$. Histidine 116 in the mutant could still form a hydrogen bonding interaction with W106 of PdX.

Threonine 56 is located close to the five P450$_{cam}$ tryptophan residues that lie close to the heme center (W42, W55, W63, W374 and W406). It is not clear if the mutation T56A has any effect on the activity of P450$_{cam}$ in the conversion of 3-chloroindole to isatin but the replacement of threonine to alanine at position 56 is not a conservative mutation—Thr56 is a polar residue while Ala56 is nonpolar.

From the docking studies it was not possible to distinguish between the early steps of the two proposed reaction pathways of 3-chloroindole dechlorination (Fig. 2): A) via a C2-C3 epoxide that opens to give a chlorohydrin that loses HCl to give 3-ketoindole, or B) via a oxidation of N1 of indole that leads to a chlorohydrin at position 3 via OH radical rebound at C3 (Fig. 2). In the docking simulations of all mutants, the Fe-N1 distance did not differ significantly from the Fe-C2 or F3-C3 distance. In future, this could be addressed with isotope labeled variants of 3-chloroindole.

5. Conclusions

We have succeeded in randomly mutating P450$_{cam}$ by using SeSaM and in selecting P450$_{cam}$ mutants that dechlorinate 3-chloroindole. P450$_{cam}$ mutants (IND1-IND6, ES1) are promising new catalysts for the oxidation and biotransformation of unnatural substrates with a vinylic chloride. The steady-state kinetic assays showed that WT P450$_{cam}$ is completely devoid of

catalytic activity in the oxidation of 3-chloroindole to isatin. No isatin was observed with the WT. Mutant IND1 (E156G/V247F/V253G/F256S) is the most active in the oxidation of 3-chloroindole to isatin with the highest V_{max}. IND1 is 138-fold faster than mutant IND5 (Y179H), the least active one found. Similarly, mutant IND6 (G93C/K314R/L319M) is 50-fold faster than the IND5 mutant. The mutant with the highest catalytic efficiency, k_{cat}/K_M, is ES1 (T56A/N116H/D297N). Mutant ES1 is about 33-fold more catalytically efficient than the IND5 mutant, the least efficient mutant we found. This example shows that even a single mutation can alter the activity of the WT P450$_{cam}$. However, the steady-state kinetic results suggest that multiple substitutions, whether in the active site or close to the P450$_{cam}$ surface, may play a cooperative role in improving P450$_{cam}$ activity.

Molecular modeling results showed that distance measurements could help determine if a substrate can be accepted into a certain enzyme. To determine the effect of specific residues on the activity of the enzyme, the docked structures must be analyzed for specific ligand interaction-to-amino acid or amino acid-to-amino acid interactions. Two plausible reaction pathways were proposed for the oxidation or 3-chloroindole to isatin. One pathway involved forming an epoxide intermediate, and another involved hydroxylation at the N-1 position. It was not possible to distinguish the two possibilities by the combination of kinetics and modeling used here.

P450$_{cam}$ variants that are efficient in the oxidation of vinylic chlorides could be used in the oxidative dehalogenation of more complex organochlorine substrates such as endosulfan which is classified as a persistent organic pollutant.

References
[1]   K. C. Jones, , P. de Voogt Persistent organic pollutants (POPs): state of the science, Environ. Pollut. 100 (1999) 209–221.

[2]   B. Gevao, H. Alegria M. F. Jaward, M. Beg, *Persistent Organic Pollutants*. (Blackwell Publishing, 2010).

[3]   J.Weber, *et al.* Endosulfan, a global pesticide: A review of its fate in the environment and occurrence in the Arctic. *Sci. Total Environ.* **408,** 2966–2984 (2010).

[4]   Kataoka, R. & Takagi, K. Biodegradability and biodegradation pathways of endosulfan and endosulfan sulfate. *Appl. Microbiol. Biotechnol.* **97,** 3285–3292 (2013).

[5]   Chefson, A. & Auclair, K. Progress towards the easier use of P450 enzymes. *Mol. Biosyst.* **2,** 462–469 (2006).

[6]   Prasad, B. The Borneol Cycle of Cytochrome P450$_{cam}$ and Evolution of the Enzyme for New Applications. *PhD Thesis.*Simon Frasesr University. (2013).

[7]   Poulos, T. L., Finzel, B. C. & Howard, A. J. High-resolution crystal structure of Cytochrome P450$_{cam}$. *J. Mol. Biol.* **195,** 687–700 (1987).

[8]   Bell, S. G. *et al.* Engineering cytochrome P450$_{cam}$ into an alkane hydroxylase. *Chem. Commun.* **11**, 2133-2140 (2003).

[9]   Wong, T. S., Roccatano, D., Zacharias, M. & Schwaneberg, U. A statistical analysis of random mutagenesis methods used for directed protein evolution. *J. Mol. Biol.* **355,** 858–871 (2006).

[10]   Arnold, F. H., Wintrode, P. L., Miyazaki, K. & Gershenson, A. How enzymes adapt: Lessons from directed evolution. *Trends Biochem. Sci.* **26,** 100–106 (2001).

[11]   Wong, T. S., Zhurina, D. & Schwaneberg, U. The diversity challenge in directed protein evolution. *Comb. Chem. High Throughput Screen.* **9,** 271–88 (2006).

[12]   Cirino, P. C., Mayer, K. M. & Umeno, D. Generating mutant libraries using error-prone PCR. *Methods Mol. Biol.* **231,** 3–9 (2003).

[13]   Kanagawa, T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* **96,** 317–323 (2003).

[14]   Wong, T. S., Tee, K. L., Hauer, B. & Schwaneberg, U. Sequence saturation mutagenesis with tunable mutation frequencies. *Anal. Biochem.* **341,** 187–189 (2005).

[15]   Mundhada, H., Marienhagen, J., Scacioc, A., Schenk, A., Roccatano, D., Schwaneberg, U. SeSaM-Tv-II Generates a Protein Sequence Space that is Unobtainable by epPCR. ChemBioChem 12, 1595-1601 (2011).

[16]   Stemmer, W. P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370,** 389–

391 (1994).

[17] Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. & Vogel, A. Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew. Chemie - Int. Ed.* **44,** 4192–4196 (2005).

[18] Alcalde, M., Farinas, E. T. & Arnold, F. H. Colorimetric High-Throughput Assay for Alkene Epoxidation Catalyzed by Cytochrome P450 BM-3 Variant 139-3. *J. Biomol. Screen.* **9**(2), 141–146 (2004).

[19] Gillam, E. M. J., Notley, L. M., Cai, H., De Voss, J. J. & Guengerich, F. P. Oxidation of indole by cytochrome P450 enzymes. *Biochemistry* **39,** 13817–13824 (2000).

[20] Kelly, P. P. *et al.* Active site diversification of P450cam with indole generates catalysts for benzylic oxidation reactions. *Beilstein J. Org. Chem.* **11,** 1713–1720 (2015).

[21] Sansen, S., Hsu, M.-H., Stout, C. D. & Johnson, E. F. Structural insight into the altered substrate specificity of human cytochrome P450 2A6 mutants. *Arch. Biochem. Biophys.* **464,** 197–206 (2007).

[22] Sevrioukova, I. F. & Poulos, T. L. Structural biology of redox partner interactions in P450cam monooxygenase: A fresh look at an old system. *Archives of Biochemistry and Biophysics* **507,** 66–74 (2011).

[23] Armarego, W. L. . & Perrin, D. . Purification of Laboratory Chemicals. *Molecules* **2,** 152 (1997).

[24] Zhao, J., Kardashliev, T., Ruff, A. J., Bocola, M., Schwaneberg U. Lessons From Diversity of Directed Evolution Experiments by an Analysis of 3,000 Mutations. Biogechnol. & Bioeng. **111**, 2380-2388 (2014)

[25] Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J., Schwaneberg, U. Phosphorothioate-based ligase-independent gene cloning (:LICing): An enzyme-free and sequence-independent cloning method. Anal. Biochem. **406**, 141-146 (2010).

[26] Miyazaki, K. *MEGAWHOP cloning: A method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids*. *Methods in Enzymology* **498,** (Elsevier Inc., 2011).

[27] De Rosa, M. & Triana Alonso, J. L. Studies of the mechanism of chlorination of indoles. Detection of N-chloroindole and 3-chloro-3H-indole as intermediates. *J. Org. Chem.* **43,** 2639–2643 (1978).

[28]   Jung, C. The mystery of cytochrome P450 Compound I: A mini-review dedicated to Klaus Ruckpaul. *Biochim. Biophys. Acta - Proteins Proteomics* **1814,** 46–57 (2011).

[29]   Westphal Christina, Anne Konkel, and W.-H. S. *Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450*. **851,** (2015).

[30]   Prasad, B., Rojubally, A. & Plettner, E. Identification of Camphor Oxidation and Reduction Products in Pseudomonas putida: New Activity of the Cytochrome $P450_{cam}$ System. *J. Chem. Ecol.* **37,** 657–667 (2011).

[31]   Gunsalus, I.C.; Katagiri, Masayuki; Ganguli, B. N. A Soluble Cytochrome P-450 Functional in Methylene Hydroxylation. *J. Biol. Chem.* 243, 3543-3546 (1968).

[32]   Rojubally, A. Linking cytochrome $P450_{cam}$ (CYP101) to its redox partner putidaredoxin and probing new reactions of the $P450_{cam}$ system. MSc.Thesis. Simon Fraser University. (2008).

[33]   Unger, B.P., Gunsalus, I.C., Sligar, S. G. Nucleotide Sequence of the Pseudomonas putida Cytochrome $P-450_{cam}$, Gene and Its Expression in *Escherichia coli*. J. Biol. Chem. **261,** 1158-1163 (1986).

[34]   Schlichting, I. *et al.* The catalytic pathway of cytochrome $P450_{cam}$ at atomic resolution. *Science* **287,** 1615–1622 (2000).

[35]   Lee, Y. T., Wilson, R. F., Rupniewski, I. & Goodin, D. B. P450cam visits an open conformation in the absence of substrate. *Biochemistry* **49,** 3412–3419 (2010).

[36]   Labute, P. Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins Struct. Funct. Bioinforma.* **75,** 187–205 (2009).

[37]   Wong, T. S., Roccatano, D., Schwaneberg, U. Are Transversion mutations better? A Mutagenesis Assistant Program analysis on P450 BM-3 heme domain. Biotechnol. J. **2**, 133-142 (2007).

[38]   Gutierrez, E. A., Mundhada, H., Meier, T., Deufel, H., Bocola, M., Schwaneberg, U. Reengineered glucose oxidase for amperometric glucose determination in diabetes analytics. Biosens. & Bioelectr. **50**, 84-90 (2013).

[39]   Martinez, R., Jakob, F., Tu, R., Stiegert, P., Maurer, K.-H., Schwaneberg, U. Increasing Activity and Thermal Resistance of Bacillus gibsonii Alkaline Protease (BgAP) by Directed Evolution. Biotechnol. & Bioeng. **110**, 711-720 (2013).

40]   Shivange, A. V., Serwe, A., Dennig, A., Roccatano, D., Haefner, S., Schwaneberg, U.

Directed evolution of a highly active Yersinia mollaretii phytase. Appl. Microbiol. Biotechnol. 95, 405-418 (2012).

[41]   Guengerich, F. P. Cytochrome P450 oxidations in the generation of reactive electrophiles: Epoxidation and related reactions. *Arch. Biochem. Biophys.* **409,** 59–71 (2003).

[42]   Manna, S. K. & Mazumdar, S. Tuning the substrate specificity by engineering the active site of cytochrome P450$_{cam}$: A rational approach. *Dalt. Trans.* **39,** 3115–3123 (2010).

[43]   Wu, Z. L., Podust, L. M. & Guengerich, F. P. Expansion of substrate specificity of cytochrome P450 2A6 by random and site-directed mutagenesis. *J. Biol. Chem.* **280,** 41090–41100 (2005).

[44]   Chowdhury, G. *et al.* Structural identification of diindole agonists of the aryl hydrocarbon receptor derived from degradation of indole-3-pyruvic acid. *Chem. Res. Toxicol.* **22,** 1905–1912 (2009).

[45]   Çelik, A. *et al.* Identification of broad specificity P450CAM variants by primary screening against indole as substrate. *Chem. Commun. (Camb).* 3652–4 (2005).

[46]   Bell, S. G. *et al.* Butane and propane oxidation by engineered cytochrome P450cam. *Chem. Commun. (Camb).* 490–491 (2002).

[47]   Xu, F., Bell, S. G., Rao, Z. & Wong, L. L. Structure-activity correlations in pentachlorobenzene oxidation by engineered cytochrome P450cam. *Protein Eng. Des. Sel.* **20,** 473–480 (2007).

[48]   Bell, S. G., Chen, X., Xu, F., Rao, Z. & Wong, L.-L. Engineering substrate recognition in catalysis by cytochrome P450cam. *Biochem. Soc. Trans.* **31,** 558–62 (2003).

[49]   Bell, S. G., Sowden, R. J. & Wong, L. L. Engineering the haem monooxygenase cytochrome P450(cam) for monoterpene oxidation. *Chem. Commun.* 635–636 (2001).

[50]   Jones, J. P., O'Hare, E. J. & Wong, L. L. Oxidation of polychlorinated benzenes by genetically engineered CYP101 (cytochrome P450$_{cam}$). *Eur. J. Biochem.* **268,** 1460–1467 (2001).

[51]   Yin, M. NMR Study of Interaction between Cytochrome P450$_{cam}$ and Putidaredoxin and Structural Study of Cytochrome P450 3A4. (2009).

[52]   Winn, P. J., Lüdemann, S. K., Gauges, R., Lounnas, V. & Wade, R. C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc. Natl. Acad.*

Sci. U. S. A. **99,** 5361–6 (2002).

[53] Lüdemann, S. K., Lounnas, V. & Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J. Mol. Biol.* **303,** 813–30 (2000).

[54] Donald, J. E., Kulp, D. W. & Degrado, W. F. Salt Bridges: Geometrically Specific, Designable Interactions. *Biochemistry* **79,** 898–915 (2012).

[55] Betts, M. J. & Russell, R. B. Amino-Acid Properties and Consequences of Substitutions. *Bioinforma. Genet. A Bioinforma. Prim. Anal. Genet. Data Second Ed.* **4,** 311–342 (2007).

[56] OuYang, B., Pochapsky, S. S., Pagani, G. M. & Pochapsky, T. C. Specific Effects of Potassium Ion Binding on Wild-Type and L358P Cytochrome P450cm. *Biochemistry* **45,** 14379–14388 (2006).

[57] Sakurai, K. Studies on roles of the amino acid residues in the vicinity of the active site in cytochrome P450cam. *PhD Thesis, Osaka Univ.* (2009).

[58] Poulos, T. L. Heme Enzyme Structure and Function. *Chem. Rev.* **114**, 3919-3962 (2014).