

# UCLA

## UCLA Previously Published Works

### Title

Analysis of Laboratory-Evolved Flavin-Dependent Halogenases Affords a Computational Model for Predicting Halogenase Site Selectivity.

### Permalink

<https://escholarship.org/uc/item/28s1d67b>

### Journal

Chem Catalysis, 2(10)

### Authors

Andorfer, Mary  
Evans, Declan  
Yang, Song  
et al.

### Publication Date

2022-10-01

### DOI

10.1016/j.checat.2022.07.003

Peer reviewed



# HHS Public Access

Author manuscript

*Chem Catal.* Author manuscript; available in PMC 2023 October 01.

Published in final edited form as:

*Chem Catal.* 2022 October ; 2(10): 2658–2674. doi:10.1016/j.cheecat.2022.07.003.

## Analysis of Laboratory-Evolved Flavin-Dependent Halogenases Affords a Computational Model for Predicting Halogenase Site Selectivity

Mary C. Andorfer<sup>3,5,8</sup>, Declan Evans<sup>2,8</sup>, Song Yang<sup>2,6</sup>, Cyndi Qixin He<sup>2,7</sup>, Anna M. Girlich<sup>3</sup>, Jaylie Vergara-Coll<sup>3</sup>, Narayanasami Sukumar<sup>4</sup>, K. N. Houk<sup>2,\*</sup>, Jared C. Lewis<sup>1,9,\*</sup>

<sup>1</sup>Department of Chemistry, Indiana University, Bloomington, IN 47405, USA

<sup>2</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA

<sup>3</sup>Department of Chemistry, University of Chicago, Chicago, IL 60637, USA

<sup>4</sup>NE-CAT and Department of Chemistry and Chemical Biology, Cornell University, Argonne National Laboratory, Building 436E, Argonne, IL 60439, USA

<sup>5</sup>Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>6</sup>Present address: Merck Research Laboratories, South San Francisco, CA, USA

<sup>7</sup>Present address: Merck Research Laboratories, Kenilworth, NJ, USA

<sup>8</sup>These authors contributed equally

<sup>9</sup>Lead contact

### Summary

Flavin-dependent halogenases (FDHs) catalyze selective halogenation of electron-rich aromatic compounds without the need for harsh oxidants required by conventional oxidative halogenation reactions. Predictive models for halogenase site selectivity could greatly improve their utility for chemical synthesis. Toward this end, we analyzed the structures and selectivity of three halogenase

\*Correspondence: houk@chem.ucla.edu (K.N.H.), jcl3@iu.edu (J.C.L.).

#### Present Addresses

M.C.A. is now in the Department of Biology at Massachusetts Institute of Technology, Cambridge, MA

S.Y. is now at Merck Research Laboratories, South San Francisco, CA

C.Q.H. is now at Merck Research Laboratories, Kenilworth, NJ

Rationale for multiple corresponding authors: This interdisciplinary work involved calculations and experiments that were critical to the findings. K.N.H. oversaw all computations and J.C.L. oversaw all experiments.

#### Author Contributions

D.E., S.Y., and C.Q.H. carried out all computations, including DFT calculations, docking, and MD simulations with guidance from K.N.H. All experimental work was performed by M.C.A., A.M.G., J.V.C., and N.S. under the direction of J.C.L. Manuscript was written by J.C.L., M.C.A., and D.E. with contributions from all other authors.

#### Declaration of Interests

The authors declare no competing interests.

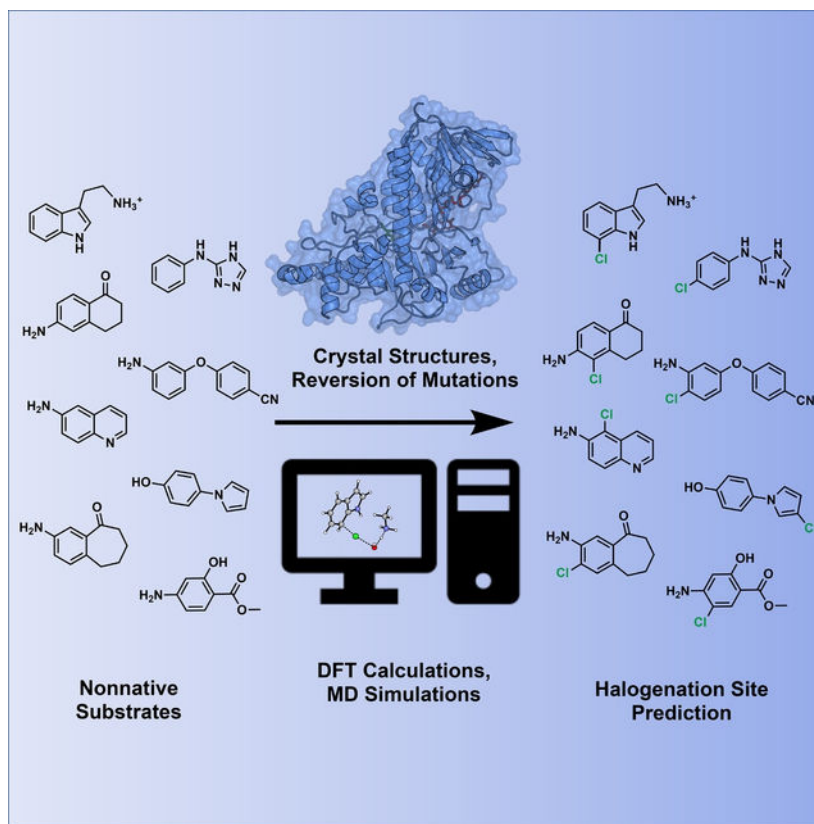
**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

variants evolved to halogenate tryptamine with orthogonal selectivity. Crystal structures and reversion mutations revealed key residues involved in altering halogenase selectivity. Density functional theory calculations and molecular dynamics simulations are both consistent with hypohalous acid as the active halogenating species in FDH catalysis. This model was used to accurately predict the site selectivity of halogenase variants toward different synthetic substrates, providing a valuable tool for implementing halogenases in biocatalysis efforts.

### eTOC:

Flavin-dependent halogenases (FDHs) can site-selectively halogenate a wide range of aromatic substrates. X-ray crystallography, site-directed mutagenesis, density functional theory, and molecular dynamics simulations were used to learn how site selectivity was previously altered through directed evolution. A predictive computational model for site selectivity of FDHs is presented.

### Graphical Abstract



### Keywords

halogenase; C-H functionalization; site-selective; directed evolution

## Introduction

Flavin-dependent halogenases (FDHs) natively catalyze selective halogenation of electron-rich aromatic compounds.<sup>1</sup> The utility of these enzymes stems from their unique mechanism, which involves the reaction of FDH-bound reduced flavin with O<sub>2</sub> to form a peroxyflavin species that reacts with proximally-bound chloride or bromide to generate HOX (X = Cl, Br).<sup>2</sup> This species migrates through the enzyme to the substrate binding site, where it forms a persistent intermediate thought to be a lysine-derived haloamine.<sup>3</sup> It is not known if this intermediate directly halogenates substrates<sup>3</sup> or serves as a reservoir of X<sup>+</sup> that can be used to reform HOX upon reaction with water<sup>4</sup>. In either case, halogenation proceeds via electrophilic aromatic substitution (EAS). Although conventional EAS requires harsh oxidants to generate reactive halonium (X<sup>+</sup>) formal equivalents,<sup>5</sup> FDHs use atmospheric oxygen and halide salts. Moreover, whereas EAS is typically subject to substrate-controlled site selectivity, substrate binding within an FDH can override substrate-based selectivity (Figure 1A/B),<sup>6,7</sup> enabling catalyst-controlled site selectivity on a range of substrates<sup>8,9</sup>.

The remarkable selectivity of FDHs has made them appealing starting points for enzyme engineering efforts. To assess the feasibility of tuning FDH selectivity, a number of groups have used targeted mutagenesis to alter the site of halogenation.<sup>10</sup> We used directed evolution to engineer variants of the FDH RebH (UniProtKB/Swiss-Prot: Q8KHZ8.1) that halogenate the C7, C6, and C5 positions of tryptamine with >90% selectivity (0S, 8F, and 10S, Figure 1C).<sup>11</sup> We also established that these enzymes demonstrate high conversion and selectivity on a range of other non-native substrates and therefore became interested in developing a general approach to predict FDH selectivity. Docking calculations revealed multiple substrate binding modes, only some of which were consistent with observed selectivity.<sup>8</sup> This inability to reliably predict relevant binding modes indicated that a more sophisticated approach was required to model FDH selectivity.

We reasoned that 0S, 8F, and 10S, which possess only 5–10 mutations relative to one another, would facilitate comparative analysis of FDH selectivity since wild-type (WT) enzymes with analogous selectivity preferences differ by hundreds of residues.<sup>1</sup> Crystal structures and reversion mutations were used to identify mutations responsible for the observed changes in 0S, 8F, and 10S yield and selectivity. To accurately model FDH selectivity, we computationally evaluated the energetic feasibility of both chloramine and HOCl halogenating agents. We then established whether FDH selectivity could be reliably predicted using molecular dynamics (MD) simulations involving energetically-relevant species in analogy to recent studies on native FDH/substrate pairs<sup>12,13</sup>. This approach provided insight into the mechanisms by which mutations in the RebH scaffold alter binding and selectivity on tryptamine and added support to the case that HOCl is the active halogenating species in FDH catalysis. Moreover, we established that density functional theory (DFT) calculations paired with docking calculations and MD simulations involving this species constitute a framework for accurately predicting FDH site selectivity on non-native substrates, improving the utility of these enzymes for biocatalysis.

## Results

### Reversion of Mutations Reveals Residues that Improve Catalysis by Evolved RebH Variants

Relative to RebH, variants 0S, 8F, and 10S possess 1–11 total mutations that increase  $k_{\text{cat}}$  by 1.2–5.9-fold and significantly increase  $K_{\text{M}}$  for 8F and 10S (Table 1).<sup>11</sup> Each of the mutations in 10S and 8F was individually reverted to the WT residue, and the conversion and selectivity of each reversion variant toward tryptamine was evaluated (Figure 2).

The simplest RebH variant in the selectivity lineage, 0S, contains a single point mutation (N470S) that gives rise to a six-fold increase in  $k_{\text{cat}}$  with minimal change in  $K_{\text{M}}$  and selectivity for C7 (>99%). To establish if the beneficial effects of S470 only result from its smaller size, S470A was introduced into 0S. Although an increase in conversion compared to RebH was observed for this variant, 0S still outperformed S470A by approximately three-fold.

Variant 10S contains six mutations that enable it to halogenate tryptamine with 95% selectivity for C5. Individually reverting these mutations back to the WT residues showed that three of these six mutations are primarily responsible for the selectivity and/or conversion of 10S (Figure 2: H52I, C465F, and S470N). Of these three reversions, H52I and C465F are responsible for the observed C5 selectivity. Although the H52I variant has high conversion relative to 10S, the selectivity reverts to primarily C7. The C465F variant displays reduced selectivity for C5 as well as reduced overall conversion. As is observed in 0S, the S470N mutation affects conversion without significantly altering site selectivity. We wondered whether the precise identity of these three residues mattered, or whether similar residues would also confer the observed changes in selectivity and/or conversion. To investigate this possibility, we also made 10S variants H52F, C465S, and S470A (Figure 2). H52F led to a modest reduction in C5 selectivity relative to 10S and displayed higher overall conversion. Reduced conversions were observed for the C465S and S470A variants, indicating that C465 and S470 act as more than just a small replacement for the native Phe and Asn residues, respectively.

Variant 8F, which catalyzes 6-chlorination of tryptamine with 90% selectivity, contains 11 mutations. Individually reverting each of these mutations (Figure 2) again revealed that not all of the mutations acquired during the evolutionary campaign have a large effect on 8F activity or selectivity. Seven of the 11 reversions were found to affect the selectivity and/or conversion of 8F (Figure 2: M52I, P110S, L130S, P448S, L465F, S470N, and Q509R). Residues 52 and 465 were again found to influence selectivity to the greatest extent; however, variant P110S also displayed reduced site selectivity for the C6 position. Several reversions reduced 8F yield (P110S, L130S, P448S, L465F, S470N, and Q509R), but the largest effect was observed for the P448S variant.

### Crystal Structures of the 0S-Tryptamine-FAD, 10S-FAD, and 8F-FAD Complexes Provide Insight into the Effects of Key Mutations on Catalysis

The structures of 0S, 8F, and 10S were solved to resolutions of 2.6, 2.25, and 2.55 Å (PDB IDs: 7JU0, 6P00, and 6P2V), respectively. Consistent with previous structures, the

asymmetric unit contained two protein molecules. Aligning each structure to that of RebH (PDB ID: 2OA1)<sup>14</sup> revealed RMSD values of 0.283, 0.616, and 0.301 Å<sup>2</sup> for 0S, 8F, and 10S, respectively, indicating that subtle changes in structure are responsible for the observed changes in selectivity. Occupancy of different ligands typically observed in FDH structures (FAD and halogenation substrate) varied in the structures obtained. In all structures, electron density for the adenine base and phosphate moieties of FAD was observed, but the isoalloxazine ring showed very little density. The occupancy of the loop comprised of residues 449–455, often disordered in previous RebH structures,<sup>3,14</sup> also varied in the structures obtained.

In the structure of 0S, residues 2–528 were included in chain A and residues 2–529 in chain B with no missing regions in between (530 residues for full length enzyme). Density in the composite omit map was observed for the loop consisting of residues 449–455. Though crystals of 0S, 8F, and 10S were soaked in a solution containing tryptamine, occupancy for this substrate was only observed for 0S. This finding is consistent with the much lower  $K_M$  for tryptamine of 0S (10.6 μM) relative to 10S and 8F (160 and 1747 μM, respectively). Tryptamine was found to bind in the active site of 0S in a manner analogous to how L-tryptophan binds to RebH (Figure 3A). A hydrogen-bonding network between S470, H109, and E357, which is reminiscent of the catalytic triad found in serine hydrolases,<sup>15</sup> is also observed in the 0S structure, and this could alter substrate binding and access relative to the native network containing N470 in RebH.

The 10S structure consisted of residues 2–528 in chain A and residues 2–529 in chain B with the loop between residues 451–455 omitted in both chains. Although density is observed within this region, the main chain cannot be reliably modeled within this density. Numerous attempts were made to soak tryptamine into 10S crystals; however, even with high concentrations of tryptamine (>10 mM), no clear density for tryptamine was observed within the active site. Overlaying the structures of 10S and RebH containing L-Trp substrate shows a steric clash between H52 in 10S and L-Trp, suggesting that the combination of H52 and C465 must significantly alter substrate binding (Figure 3B).

The 8F structure consisted of residues 1–527 in chain A and residues 2–529 in chain B, but the loop between residues 447–456 was omitted in both chains. Seven mutations played a significant role in 8F (Figure 3C). Similar to 10S, when the structure of RebH containing L-Trp is overlaid with 8F, a steric clash is observed between N1 and C2 of the indole moiety of the substrate and the sulfur atom of M52 (3.1 and 2.6 Å, respectively). L465 opens the active site, potentially allowing the substrate to rotate or flip to accommodate the M52 mutation. P110 is flanked by two aromatic residues, H109 and F111, which sandwich tryptophan bound in the overlaid RebH structure (Figure 3D).

### Computational Analysis Suggests that Bound HOCl is the Active Halogenating Species in FDH Catalysis

As previously noted, a lysine-derived chloramine or bound HOCl, both of undetermined protonation state, have been proposed as the active halogenating species during FDH-catalyzed arene halogenation.<sup>3,4</sup> In either case, aromatic halogenation is believed to occur via EAS involving rate-limiting electrophilic attack to form a Wheland intermediate

followed by low-barrier proton loss to form the aromatic product. The lack of a deuterium kinetic isotope effect (KIE) for RebH-catalyzed tryptophan halogenation is consistent with this mechanism.<sup>11</sup> The  $G^\circ$  for formation of the Wheland intermediate ( $G^\circ_{\text{Whel}}$ ) has been found to correlate strongly with  $G^\ddagger$  for this reaction.<sup>16</sup> We therefore reasoned that  $G^\circ_{\text{Whel}}$  values calculated by DFT could provide insight into feasibility of different proposed halogenating agents within FDHs. Scheme 1 and Table 2 show the calculated  $G^\circ_{\text{Whel}}$  using several possible halogen donors for EAS at the the 7-position of indole.

Very high  $G^\circ_{\text{Whel}}$  values were calculated for reactions involving neutral methylchloramine and HOCl (72.4 and 21.2 kcal/mol; Table 2, Entries 1 and 2), making chlorination by these species unlikely. Protonation of either lowers these  $G^\circ_{\text{Whel}}$  values considerably (Table 2, Entries 3 and 4) due to the increased electrophilicity of the chlorine. The extremely negative  $G^\circ_{\text{rxn}}$  value calculated for  $\text{H}_2\text{OCl}^+$  calls into question the existence of such a highly reactive species in an enzyme. Instead, it has been suggested that K79 transiently protonates HOCl during the formation of the Wheland intermediate,<sup>12</sup> resulting in a slightly negative  $G^\circ_{\text{Whel}}$  (Table 2, Entry 5). Taken together, these energies support two possible chlorinating agents: a positively charged lysine-bound chloramine at K79, and HOCl with general acid catalysis by K79<sup>2,4,12</sup>.

Because DFT calculations suggested that two species were plausible halogenating intermediates, we next sought to compare these models using molecular dynamics (MD) simulations of tryptamine binding to RebH. Two systems were prepared to represent the two possible reactant complexes determined by DFT: one with the chlorine covalently bound to K79 (K79-Cl, Figure 4a), and one with HOCl held close to K79 (K79-HOCl, Figure 4b)<sup>13</sup> using restraints in AMBER (see methods). Each simulation was run in pentaplicate, and the distances between the chlorine and substrate carbons were measured (Figure S1). Consistently long C-Cl distances are observed with K79-Cl, and these distances vary greatly between replicas. In contrast, much shorter C-Cl distances are observed with K79-HOCl, and these distances were consistent across all replicas. Similar results are observed in the 0S mutant (Figure S2) and in previous simulations of tryptophan binding in PrnA and PyrH<sup>13</sup>. In the current simulations, the chlorine of HOCl remains closer to C7 than either C6 or C5, which is consistent with the experimentally observed selectivity of the enzyme. These simulations suggest that FDH catalysis likely involves halogenation by HOCl rather than a chloramine and that accounting for this distinction could improve predictive models of FDH selectivity.

### DFT Calculations and MD Simulations can be used to Model the Selectivity of RebH Variants

With a model for the reaction mechanism in hand, we next sought to investigate how mutations in 0S, 10S, and 8F give rise to the observed site selectivity of these variants. Transition state structures were calculated using DFT for the general acid-catalyzed electrophilic attack at the 5-, 6- and 7-positions of indole (Figure 5). These transition states are all structurally similar. A proton is shared between the methylammonium nitrogen and the HOCl oxygen with a nearly linear N-H-O angle. Simultaneously, the chlorine is shared between the HOCl oxygen and indole carbon with a nearly linear O-Cl-C angle.



Energetically, TS6 is most favorable with an activation energy of 18.2 kcal/mol with respect to the complex. TS5 was slightly higher in energy (19.0 kcal/mol) and TS7 was higher still (20.1 kcal/mol). The ability of RebH and OS to chlorinate the least favorable C7 highlights the importance of substrate binding (i.e. catalyst control) for halogenation of indole-containing substrates<sup>10</sup>.

Crystal structures of other FDH-tryptophan complexes show that each enzyme binds its substrate with the carbon to be halogenated closest to the active site lysine.<sup>2,3,14,17–19</sup> RebH is no exception, as C7 of tryptophan is closest to K79 in this structure.<sup>3</sup> The indole ring is held in place by pi stacking with H109 and F111, and the amine extends upward to hydrogen bond with Y454, E461, and F465 (Figure 6). The crystal structure of the OS-tryptamine complex shows that the N470S mutation does not affect this binding orientation, and the two substrates overlay well (Figure 6A). The crystal structures of 10S and 8F variants do not contain tryptamine, so, to probe binding within 10S and 8F, tryptamine was docked into the active site of these structures. Docked structures show that tryptamine binds to 10S and 8F in different orientations than that observed in RebH and OS (Figure 6B, C). In these variants, the indole ring flips to hydrogen bond with the backbone of S110 (P110 in 8F), and different carbons are projected towards the catalytic K79. In 10S, C5 is closest to K79 with a N-C5 distance of 4.3 Å; in 8F, C6 is closest to K79 with a N-C6 distance of 3.9 Å. These orientations are consistent with solved crystal structures of the tryptophan-5 halogenase PyrH and tryptophan-6 halogenases BorH and Thal (Figure S3)<sup>17</sup>. Thus, the crystal structure of the OS-tryptamine complex and structures of tryptamine docked into 10S and 8F show that the evolved variants all follow the established trend that substrates bind with the halogenation site closest to the active site lysine.

As previously noted, a correlation between halogenation site and proximity of that site to K79 in RebH was not observed in docking simulations of more structurally diverse substrates. We therefore sought to determine if MD simulations using HOCl-bound models of FDH variants could provide improved predictive value in this regard. Simulations of RebH in complex with HOCl and tryptamine show that the chlorine remains closer to C7 than either C5 or C6 for most of the simulation and simulations of OS show similar behavior (Figures S1–S2, Table S2). Likewise, previously reported simulations of tryptophan binding in PrnA and PyrH in complex with HOCl show that chlorine remains closer to C7 and C5, respectively, in these systems.<sup>13</sup> These results show that FDHs bind their substrates in an orientation that not only places halogenation sites proximal to the active site lysine, but also appears consistent with a near attack conformation (NAC),<sup>20</sup> those conformations that have the substrate and Cl of HOCl within a distance that is likely to lead to a transition state for reaction. Simulations of 8F and 10S were less clear (Figure S4); the carbon closest to chlorine changed frequently, and considerably longer C-Cl distances were observed. We reasoned that these C-Cl distances, which are significantly longer than those in the calculated transition state structures, are likely not relevant to the observed selectivity.

To precisely define a NAC for halogenation at each site of tryptamine, intrinsic reaction coordinate calculations were performed on the respective calculated transition state structures, and the reactant complexes were optimized by DFT. The resulting complexes show that a C-Cl distance of 3.0–3.1 Å is ideal for NAC formation. Distances this short were



rarely if ever observed in MD simulations using conventional force fields, so this threshold was extended to include any distance below 3.6 Å. The free energy of NAC formation ( $G_{\text{N}}^{\circ}$ ) in the enzyme can then be estimated by calculating the relative mole fraction of enzyme-NAC relative to enzyme-ground state in each simulation. Calculated  $G_{\text{N}}^{\circ}$  values provide a quantitative comparison to intrinsic  $G_{\text{TS}}^{\ddagger}$  calculated by DFT.

Using this approach, preferential NAC formation can be observed at C7 for RebH and OS (Table 3 and S3). This site is favored by roughly 2 kcal/mol in these enzymes, which contrasts with the 1.9 kcal/mol the site is disfavored from DFT-calculated  $G_{\text{TS}}^{\ddagger}$  values. While preferential NAC formation at C7 is slightly underestimated, the similarity between values obtained by MD simulations and DFT calculations shows that differences in NAC formation can be useful when gauging the enzymes' ability to override the intrinsic site selectivity of the reaction. In the variants 8F and 10S, there is a clear shift away from NAC formation at C7. Instead, NAC formation is roughly equivalent for both C5 and C6 in these enzymes. While C5 NAC formation is slightly underestimated in 10S, this metric provides a useful explanation for why these variants lack selectivity for the original halogenation site and correctly identifies two other sites where NAC formation is preferential.

### Predicting Site Selectivity in Non-native Substrates Beyond Tryptamine

Given successful recapitulation of RebH selectivity toward tryptamine using the methods described above, we next sought to evaluate their efficacy toward a subset of the compounds explored in our previous substrate-activity profiling effort.<sup>8</sup> Relative free energies of Wheland intermediate formation ( $G_{\text{whel}}$ ) were calculated at each  $sp^2$  hybridized carbon to gauge the intrinsic site selectivity for chlorination (Table 4). As previously noted, Wheland intermediate energies have been shown to correlate strongly with transition state barrier heights, while being easier to determine computationally. Comparison of  $G_{\text{whel}}$  on various sites of the indole ring confirm this to be the case (Figure S5). For many of the substrates, including **1**, **2**, **4**, **6**, and **7**, chlorination is observed at the most reactive carbon (lowest  $G_{\text{whel}}$ ), suggesting that the enzyme has little catalyst control over most non-native substrates. This is not always the case (e.g. substrates **3** and **5**), so further modeling of the enzyme is needed to reliably predict chlorination sites.

Docking the substrates into the RebH active site yielded several low-energy poses for each (Figures S6–S12). These different binding orientations often projected different halogenation sites towards the catalytic K79, making it difficult to predict the halogenation site based on these structures alone. In every case, docking poses that were consistent with experimentally-determined selectivity were observed, where the carbon to be halogenated is closest to K79. In the docking poses inconsistent with observed selectivity, a carbon predicted to be unreactive by DFT ( $G_{\text{whel}} > 15$  kcal/mol) was closest to K79. Although such poses may exist in the enzyme, selectivity on the native substrate, tryptophan, occurs on a site disfavored by 11.2 kcal/mol (C7 vs. C2), and it is unlikely RebH has a higher degree of catalyst control over non-native substrates. Limiting our search to only include reactive poses (where a site with  $G_{\text{whel}} < 11$  kcal/mol is closest to K79) the observed halogenation site was always the site closest to K79.

Finally, MD simulations were conducted to quantify catalyst control in the HOCl-bound enzyme (Table 4 and S4, Figure S13). For many of the substrates, including **1**, **3**, **6** and **7**, NAC formation was favored at the observed halogenation site. Moreover, in substrate **3**, preferential NAC formation correctly predicts chlorination at C3 over C5, the most reactive site calculated by DFT. Substrate **4** shows similar NAC formation for both C1 and C3; however, differences in intrinsic reactivity make C1 the only possible halogenation site for this substrate. Halogenation could not be correctly predicted with this method for two of the substrates tested. In substrate **2**, NAC formation predicts chlorination at C3, yet no chlorination is experimentally observed at this site. Instead, chlorination is observed at C1, which is calculated to be the most reactive site. For substrate **5**, poor substrate binding was observed, and no NACs were formed at any site. In this case, chlorination is observed at C9, which is less reactive than C7 but closer to K79 in docking calculations.

## Discussion

### Crystal Structures, Reversion Mutations, and MD Simulations Provide Molecular Insight into the Altered Selectivity of Evolved FDHs

The crystal structures of 0S, 8F, and 10S, reversion mutations in these enzymes, and a NAC model for selectivity was used to rationalize the effects of mutations identified via directed evolution. The crystal structure of 0S shows that the only mutation in this enzyme, N470S, is located in the active site but does not contact the substrate. Introducing N470A into RebH does not fully account for the increased conversion of 0S, suggesting that the hydroxyl group of serine, not just its small size, is important for the improved function of 0S. In WT RebH, N470 forms a hydrogen bond network with H109 and E357, and this network is maintained but less persistent in the N470S variant (Figure S14). Residue 470 is at the start of a large loop (residues 439–468) that regulates substrate access to the active site. We speculate that relaxed hydrogen bonding to S470 facilitates loop movements required for substrate entry, thus increasing  $k_{\text{cat}}$  6-fold,<sup>21</sup> while permitting native-like hydrogen bonding to maintain a  $K_M$  similar to that of WT RebH.

Residues 52 and 465 impart the largest changes in selectivity to both 10S and 8F. Overlaying the structures of 10S and RebH (PDB ID: 2OA1, Figure 3B) reveals that I52H sterically restricts one region of the active site whereas F465C expands a different section. Docking calculations show that 10S binds the indole moiety in a different orientation than RebH (Figure 6B) to avoid a steric clash between tryptamine and H52 if the native binding mode were adopted (Figure 3B). Just as in 10S, mutation of residue 52 to methionine within 8F would be expected to favor the flipped substrate binding conformation based on its size relative to the WT isoleucine residue, and this is supported by docking calculations (Figure 7).

While both H52 and M52 favor a flipped substrate binding mode, docking calculations suggest that residues 52 and 465 act in concert to rotate the flipped substrate such that C5 and C6 point toward K79 in 10S and 8F, respectively. H52 in 10S is positioned further into the active site than M52 in 8F due to hydrogen bonding between the indole nitrogen and the backbone carbonyl of E357. This orientation of H52 rotates tryptamine such that C5 is closer to K79 (Figure 7). Mutating residue 52 to Phe leads to a variant with reduced

selectivity for C5 (Figure 2, 85.5% vs 93.7%), perhaps due to its inability to H-bond with E357. The thioether sidechain of M52 in 8F, although bulky enough to cause substrate flipping, could allow the indole ring to rotate slightly closer to residue 52, which would cause C6 to project towards K79 (Figure 7).

Reverting C465 and L465 to Phe in 10S and 8F, respectively, also results in lower selectivity and reduced conversion. In 10S, F465C creates more space in the active site, perhaps making room for the bulkier H52 residue in 10S. Variant 10S C465S exhibits similar site selectivity but 8-fold lower conversion relative to 10S, indicating that any improvements resulting from the reduced size and hydrogen bonding capability of C465 cannot be recapitulated even by the relatively conservative S465 variant. In 8F, docking calculations suggest that L465 leads to a steric clash with the primary amine of tryptamine, leading to rotation of tryptamine to achieve C6 selectivity. This repulsion also breaks a hydrogen bond between the backbone carbonyl of residue 465 and the primary amine of tryptamine that is found in both 0S and 10S, which is consistent with the significantly higher  $K_M$  of 8F compared with 10S (1747 vs. 160  $\mu\text{M}$ , respectively). The subtle changes in tryptamine positioning by residues 52 and 465 highlight how these residues act together to tune the selectivity of 10S and 8F.

Although several additional residues modulate the yield and selectivity of 8F, reversions P110S and P448S have the largest effects on 8F yield. Residues H109 and F111 contribute to substrate binding via  $\pi$ -stacking, and MD simulations show that P110 rigidifies F111 to enforce this interaction<sup>13</sup> (Figure S15) while shifting the backbone carbonyl to better engage the indole N-H of tryptamine in a hydrogen bond to favor the flipped substrate orientation. Residue 448 is in the active site access loop noted above, so P448 could modulate the dynamics of this loop to facilitate substrate entry.

Comparing the effects of critical mutations in 10S and 8F to those present in natural halogenases reveals interesting parallels. First, the three reported 5-tryptophan halogenases possess F52,<sup>22–24</sup> which, as noted above, was nearly as effective as H52 in the engineered 5-tryptamine halogenase 10S. O'Connor and coworkers showed that introducing I52F into RebH results in an inactive variant.<sup>25</sup> Introducing I52H both alone and in combination with F465C into RebH also resulted in inactive variants in our hands, indicating that while an aromatic residue at this position can lead to a 5-halogenase, other mutations are required to enable such a change in RebH. Previously reported 6-tryptophan halogenases possess either V52 and greater sequence similarity to 7-tryptophan halogenases<sup>18,19</sup> or F52 and more sequence similarity to 5-tryptophan halogenases<sup>26–28</sup>. Our finding that I52M in 8F confers selectivity for the tryptamine 6-position thus represents a unique solution to achieving this selectivity relative to those observed in natural FDHs. Similarly, all known 5- and 6-Trp-FDHs contain proline at the residue corresponding to residue 110 in RebH, but while P110 emerged in the 8F lineage, it is not present in 10S. Finally, all known 5- and 6-Trp-FDHs contain serine at the site corresponding to residue 470 in RebH, and all 7-Trp-FDHs contain asparagine, yet S470 increases the catalytic activity of 0S without changing its selectivity. In short, laboratory evolution of FDHs to alter site selectivity led to mutations that, in some cases, are similar to those that emerged via natural selection,<sup>29</sup> but novel solutions to altered selectivity were also obtained.

## Analysis of Laboratory-Evolved Enzymes Affords a Computational Model for FDH Selectivity

Both HOCl and an active site lysine-derived haloamine have been proposed as the active halogenating species in FDH catalysis.<sup>3,4</sup> DFT calculations show halogenation by either of these species, when protonated, is exergonic. Calculated transition state energies show that K79 can activate HOCl by general acid catalysis, and this finding is consistent with mutagenesis experiments that show K79 is essential for activity<sup>4</sup>. Furthermore, the catalytic lysine of many halogenases has been calculated to be much more acidic than free lysine in solution<sup>30,31</sup>. We calculate pK<sub>a</sub> values of 7.1, 7.0, 6.9, and 6.9 for K79 in RebH, 0S, 10S, and 8F, respectively. These low pK<sub>a</sub> values result largely from hydrophobic residues, including A50, I52, I82, and F111, that are proximal to K79. Furthermore, these values suggest that at physiological pH, K79 can readily donate a proton to HOCl to assist in halogenation. MD simulations of WT RebH using K79-HOCl also give much shorter C-Cl distances than K79-Cl, and these distances are consistent with the observed selectivity. Taken together, these data suggest that chlorination by HOCl provides a better model for site selectivity in RebH than a lysine-derived chloramine.

Calculated transition state energies show that chlorination of tryptamine at C6 is most favored and chlorination at C7 is least favored. Nevertheless, RebH and 0S chlorinate C7 with high selectivity, which highlights the importance of substrate binding. Indeed, all tryptophan FDHs crystalized with substrates to date bind their substrates in orientations that project the C<sub>sp2</sub>-H bond to be halogenated proximal to the active site lysine residue.<sup>2,3,14,17-19</sup> Although 10S and 8F were not crystallized with substrate, docking calculations show these enzymes also bind tryptamine following this trend. MD simulations using HOCl show that each enzyme preferentially forms near attack conformations for the observed halogenation site. Comparison of  $\Delta G^\ddagger$  values determined by MD simulations to  $\Delta G^\ddagger$  values calculated by DFT shows that preferential NAC formation accounts for almost all the catalyst control over the reaction.

### The FDH Selectivity Model Accurately Predicts Selectivity on Non-Native Substrates

Beyond providing insights into the selectivity of 0S, 8F, and 10S toward tryptamine, computational tools also provided improved ability to predict the selectivity of other non-native FDH-substrate pairs. We previously attempted to use docking calculations alone to model tryptamine binding in 0S, 10S, and 8F<sup>11</sup> and to model the binding poses of a panel of substrates in RebH<sup>8</sup>. In both cases, poses consistent with the observed halogenation selectivity were observed, but so too were poses inconsistent with the observed selectivity and scope.

Calculating the relative energies of Wheland intermediates provided a wealth of predictive ability. Similar metrics, like RegioSQM,<sup>32</sup> and HalA,<sup>33</sup> have been used previously, and we find that these methods perform comparably (Figure S16). For almost every substrate, the site identified as most reactive was the experimentally observed halogenation site. This is perhaps unsurprising, as these enzymes would not necessarily be expected to exhibit high catalyst control over reactions involving non-native substrates. Docking provided additional insight into substrate-enzyme interactions. As with our previous studies,<sup>8,11</sup> many docking

poses could be identified, and each projected a different halogenation site proximal to K79. Including the  $G_{\text{whel}}$  values calculated from DFT allowed us to exclude docking poses that do not represent catalytically active binding conformations (i.e. NACs) and provided a much better predictor for site selectivity. Furthermore, MD simulations show that NAC formation can provide a quantitative comparison between the intrinsic site selectivity of the substrate and the catalyst control of the enzyme.

Density functional theory calculations, docking, and molecular dynamics simulations all provided insight into site selectivity, but no one method was able to reliably predict the halogenation site every time. Instead, the three methods used together provide a more complete picture for substrate and catalyst control over the reaction to identify sites where halogenation is likely. Perfect predictive ability, while ideal, is not completely necessary for a model to provide value. The ability to predict one or a few sites out of many can identify regions of the substrate where chlorination is possible and can be paired with subsequent directed evolution to engineer highly selective catalysts for the desired site.

## Conclusions

Improved models for rationalizing and predicting FDH selectivity toward aromatic substrates would substantially improve the utility of these enzymes for biocatalysis. Our recent finding that FDHs can also catalyze non-native olefin halocyclization suggests that such models would have utility beyond even the native function of these enzymes.<sup>34</sup> We therefore evaluated the energetics of electrophilic aromatic substitution (EAS) by hypohalous acid and chloramine species that have been proposed as active halogenating intermediates in FDH catalysis. The first step of EAS is highly endergonic unless these species are protonated. A low pKa value was calculated for K79, which is more consistent with its role as a proton donor than chloramine source. Moreover, MD simulations showed that HOCl provided a better model for site selectivity in RebH.

Crystal structures and reversion mutations for three previously evolved RebH variants with orthogonal selectivity were next used to identify mutations that altered variant activity and selectivity. Notably, RebH variants containing only those mutations that led to large changes in selectivity in the evolved variants were inactive, highlighting the importance of permissive mutations, identified via random mutagenesis, for achieving the high selectivities observed. Because these enzymes only possess a handful of mutations, they constitute ideal test cases for predictive models of FDH selectivity. Indeed, docking calculations and MD simulations show that each enzyme binds tryptamine in an orientation that places the carbon to be chlorinated closest the catalytic lysine. This binding orientation allows the enzyme to overcome the intrinsic site selectivity that favors chlorination at the 6-position of tryptamine. Similar calculations also recapitulated the selectivity for several non-native FDH/substrate pairs, which provides a general framework for modeling and predicting FDH selectivity for biocatalysis efforts.

## Experimental Procedures

### Resource Availability

**Lead Contact:** Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jared C. Lewis (jcl3@iu.edu).

**Materials Availability:** This study did not generate new unique reagents.

**Data and Code Availability:** Crystallography data have been deposited at the PDB under the database identifiers 7JU0 (PDB DOI: [10.2210/pdb7JU0/pdb](https://doi.org/10.2210/pdb7JU0/pdb)), 6P00 (PDB DOI: [10.2210/pdb6P00/pdb](https://doi.org/10.2210/pdb6P00/pdb)), and 6P2V (PDB DOI: [10.2210/pdb6P2V/pdb](https://doi.org/10.2210/pdb6P2V/pdb)). Any additional information required to reanalyze the data reported in this paper is available in the supplemental information, or from the lead contact upon request.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments and Funding Sources

This study was supported by the NIH (R01 GM115665 to JCL) and by the NSF (CHE-1764328 to KNH). Additional support by the NSF under the CCI Center for Selective C–H Functionalization (CCHF, CHE-1205646, JCL and KNH). M.C.A. was supported by an NIH Chemistry and Biology Interface training grant (T32 GM008720), an NSF predoctoral fellowship (DGE-1144082), a University of Chicago Department of Chemistry Helen Sella-Beretvas Fellowship, and an ARCS Scholar Award. D.E. was supported by the UCLA Chemistry-Biology Interface training grant (USPHS National Research Service Award 5T32GM008496). This study is based on research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on 24-ID-E is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Calculations were performed on the Hoffman2 cluster at UCLA, and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

### References

1. Pée K.-H. van, and Patallo E (2006). Flavin-Dependent Halogenases Involved in Secondary Metabolism in Bacteria. *Applied Microbiology And Biotechnology* 70, 631–641. [PubMed: 16544142]
2. Dong C, Flecks S, Unversucht S, Haupt C, Pée K.-H. van, and Naismith JH (2005). Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science* 309, 2216–2219. [PubMed: 16195462]
3. Yeh E, Blasiak LC, Koglin A, Drennan CL, and Walsh CT (2007). Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases. *Biochemistry* 46, 1284–1292. [PubMed: 17260957]
4. Flecks S, Patallo EP, Zhu X, Ernyei AJ, Seifert G, Schneider A, Dong C, Naismith JH, and Pée K-H van (2008). New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angewandte Chemie International Edition in English* 47, 9533–9536.
5. Podgoršek A, Zupan M, and Iskra J (2009). Oxidative Halogenation with “Green” Oxidants: Oxygen and Hydrogen Peroxide. *Angewandte Chemie International Edition in English* 48, 8424–8450.
6. Latham J, Brandenburger E, Shepherd SA, Menon BRK, and Micklefield J (2017). Development of Halogenase Enzymes for Use in Synthesis. *Chemical Reviews* 118, 232–269. [PubMed: 28466644]



7. Weichold V, Milbredt D, and Pée K-H van (2016). Specific Enzymatic Halogenation-From the Discovery of Halogenated Enzymes to Their Applications In Vitro and In Vivo. *Angewandte Chemie International Edition in English* 55, 6374–6389.
8. Andorfer MC, Grob JE, Hajdin CE, Chael JR, Siuti P, Lilly J, Tan KL, and Lewis JC (2017). Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catalysis* 7, 1897–1904. [PubMed: 28989809]
9. Fisher BF, Snodgrass HM, Jones KA, Andorfer MC, and Lewis JC (2019). Site-Selective C–H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. *ACS Cent Sci* 5, 1844–1856. [PubMed: 31807686]
10. Andorfer MC, and Lewis JC (2018). Understanding and Improving the Activity of Flavin-Dependent Halogenases via Random and Targeted Mutagenesis. *Annual Review of Biochemistry* 87, 159–185.
11. Andorfer MC, Park H-J, Vergara-Coll J, and Lewis JC (2016). Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. *Chemical Science* 7, 3720–3729. [PubMed: 27347367]
12. Karabencheva-Christova TG, Torras J, Mulholland AJ, Lodola A, and Christov CZ (2017). Mechanistic Insights into the Reaction of Chlorination of Tryptophan Catalyzed by Tryptophan 7-Halogenase. *Scientific Reports* 7, 1–15. [PubMed: 28127051]
13. Ainsley J, Mulholland AJ, Black GW, Sparagano O, Christov CZ, and Karabencheva-Christova TG (2018). Structural Insights from Molecular Dynamics Simulations of Tryptophan 7-Halogenase and Tryptophan 5-Halogenase. *ACS Omega* 3, 4847–4859. [PubMed: 31458701]
14. Bitto E, Huang Y, Bingman CA, Singh S, Thorson JS, and Phillips GN (2008). The Structure of Flavin-Dependent Tryptophan 7-Halogenase RebH. *Proteins-Structure Function And Bioinformatics* 70, 289–293.
15. Hedstrom L (2002). Serine Protease Mechanism and Specificity. *Chemical Reviews* 102, 4501–4524. [PubMed: 12475199]
16. Liljenberg M, Brinck T, Herschend B, Rein T, Rockwell G, and Svensson M (2010). Validation of a Computational Model for Predicting the Site for Electrophilic Substitution in Aromatic Systems. *Journal of Organic Chemistry* 75, 4696–4705. [PubMed: 20552984]
17. Zhu X, Laurentis WD, Leang K, Herrmann J, Ihlefeld K, Pée K.-H. van, and Naismith JH (2009). Structural Insights into Regioselectivity in the Enzymatic Chlorination of Tryptophan. *Journal Of Molecular Biology* 391, 74–85. [PubMed: 19501593]
18. Moritzer A-C, Minges H, Prior T, Frese M, Sewald N, and Niemann HH (2019). Structure-based switch of regioselectivity in the flavin-dependent tryptophan 6-halogenase Thal. *Journal Of Biological Chemistry* 294, 2529–2542. [PubMed: 30559288]
19. Lingkon K, and III JJB (2019). Structure and Activity of the Thermophilic Tryptophan-6 Halogenase BorH. *ChemBioChem* 21, 1121–1128. [PubMed: 31692209]
20. Bruice TC, and Lightstone FC (1999). Ground State and Transition State Contributions to the Rates of Intramolecular and Enzymatic Reactions. *Accounts Of Chemical Research* 32, 127–136.
21. Lee H-L, Chang C-K, Jeng W-Y, Wang AH-J, and Liang P-H (2012). Mutations in the substrate entrance region of  $\beta$ -glucosidase from *Trichoderma reesei* improve enzyme activity and thermostability. *Protein Eng Des Sel* 25, 733–740. [PubMed: 23077275]
22. Zehner S, Kotzsch A, Bister B, Sussmuth R, Mendez C, Salas JA, and Pée K-H van (2005). A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chemistry & Biology* 12, 445–452. [PubMed: 15850981]
23. Chang F-Y, and Brady SF (2011). Cloning and Characterization of an Environmental DNA-Derived Gene Cluster That Encodes the Biosynthesis of the Antitumor Substance BE-54017. *J Am Chem Soc* 133, 9996–9999. [PubMed: 21542592]
24. Ryan KS (2011). Biosynthetic Gene Cluster for the Cladoniamides, Bis-Indoles with a Rearranged Scaffold. *Plos One* 6, e23694. [PubMed: 21876764]
25. Glenn WS, Nims E, and O'Connor SE (2011). Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *Journal of the American Chemical Society* 133, 19346–19349. [PubMed: 22050348]



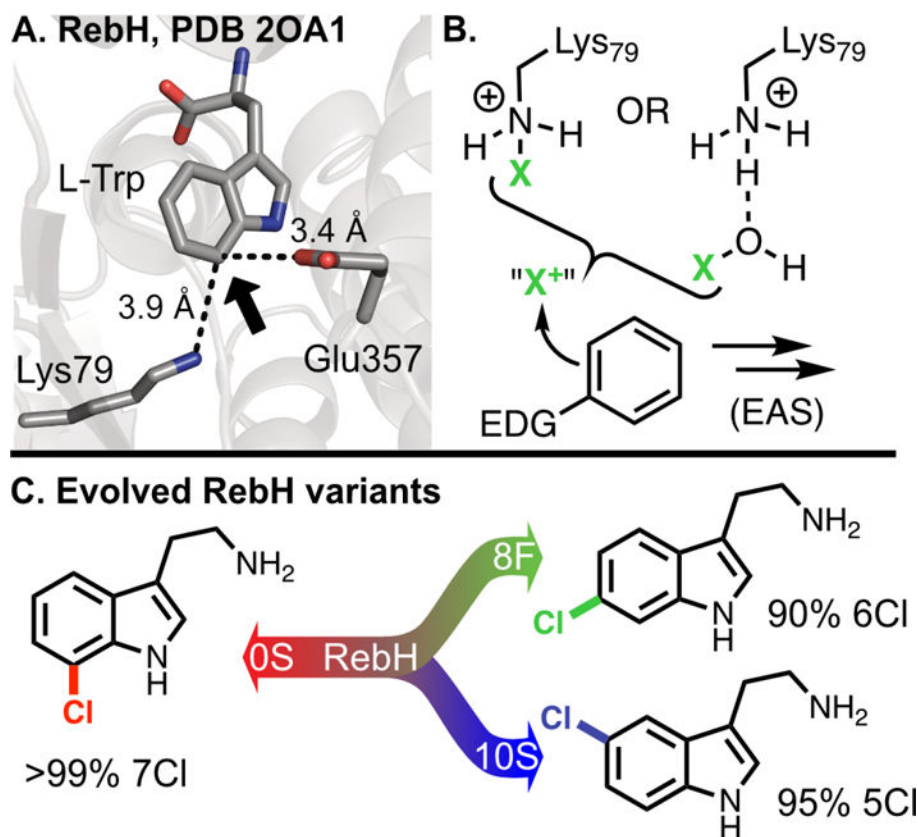
26. Shepherd SA, Menon BRK, Fisk H, Struck A-W, Levy C, Leys D, and Micklefield J (2016). A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase. *ChemBioChem* 17, 821–824. [PubMed: 26840773]
27. Zeng J, and Zhan J (2011). Characterization of a tryptophan 6-halogenase from *Streptomyces toxytricini*. *Biotechnol Lett* 33, 1607–1613. [PubMed: 21424165]
28. Heemstra JR, and Walsh CT (2008). Tandem Action of the O-2- and FADH(2)-Dependent Halogenases KtzQ and KtzR Produce 6,7-Dichlorotryptophan for Kutzneride Assembly. *Journal of the American Chemical Society* 130, 14024–+. [PubMed: 18828589]
29. Romero PA, and Arnold FH (2009). Exploring Protein Fitness Landscapes by Directed Evolution. *Nature reviews. Molecular cell biology* 10, 866–876. [PubMed: 19935669]
30. Fraley AE, Garcia-Borràs M, Tripathi A, Khare D, Mercado-Marin EV, Tran H, Dan Q, Webb GP, Watts KR, Crews P, et al. (2017). Function and Structure of MalA/MalA', Iterative Halogenases for Late-Stage C–H Functionalization of Indole Alkaloids. *Journal of the American Chemical Society* 139, 12060–12068. [PubMed: 28777910]
31. Phintha A, Prakinee K, Jaruwat A, Lawan N, Visitsathawong S, Kantiwiriyanitch C, Songsunthong W, Trisrivirat D, Chenprakhon P, Mulholland AJ, et al. (2020). Dissecting the low catalytic capability of flavin-dependent halogenases. *Journal Of Biological Chemistry*, jbc.RA120.016004–36.
32. Kromann JC, Jensen JH, Kruszyk M, Jessing M, and Jørgensen M (2017). Fast and accurate prediction of the regioselectivity of electrophilic aromatic substitution reactions† †Electronic supplementary information (ESI) available. See DOI: 10.1039/c7sc04156j. *Chem Sci* 9, 660–665. [PubMed: 29629133]
33. Ashtekar KD, Marzijarani NS, Jaganathan A, Holmes D, Jackson JE, and Borhan B (2014). A New Tool To Guide Halofunctionalization Reactions: The Halenium Affinity ( HalA) Scale. *Journal of the American Chemical Society* 136, 13355–13362. [PubMed: 25152188]
34. Mondal D, Fisher BF, Jiang Y, and Lewis JC (2021). Flavin-dependent halogenases catalyze enantioselective olefin halocyclization. *Nat Commun* 12, 3268. [PubMed: 34075034]

**Highlights:**

- Flavin-dependent halogenases site-selectively halogenate under mild conditions
- Selectivity of evolved variants explored using crystallography and mutagenesis
- Data are consistent with hypohalous acid as the active halogenating species
- Predictive computational model for halogenase site selectivity

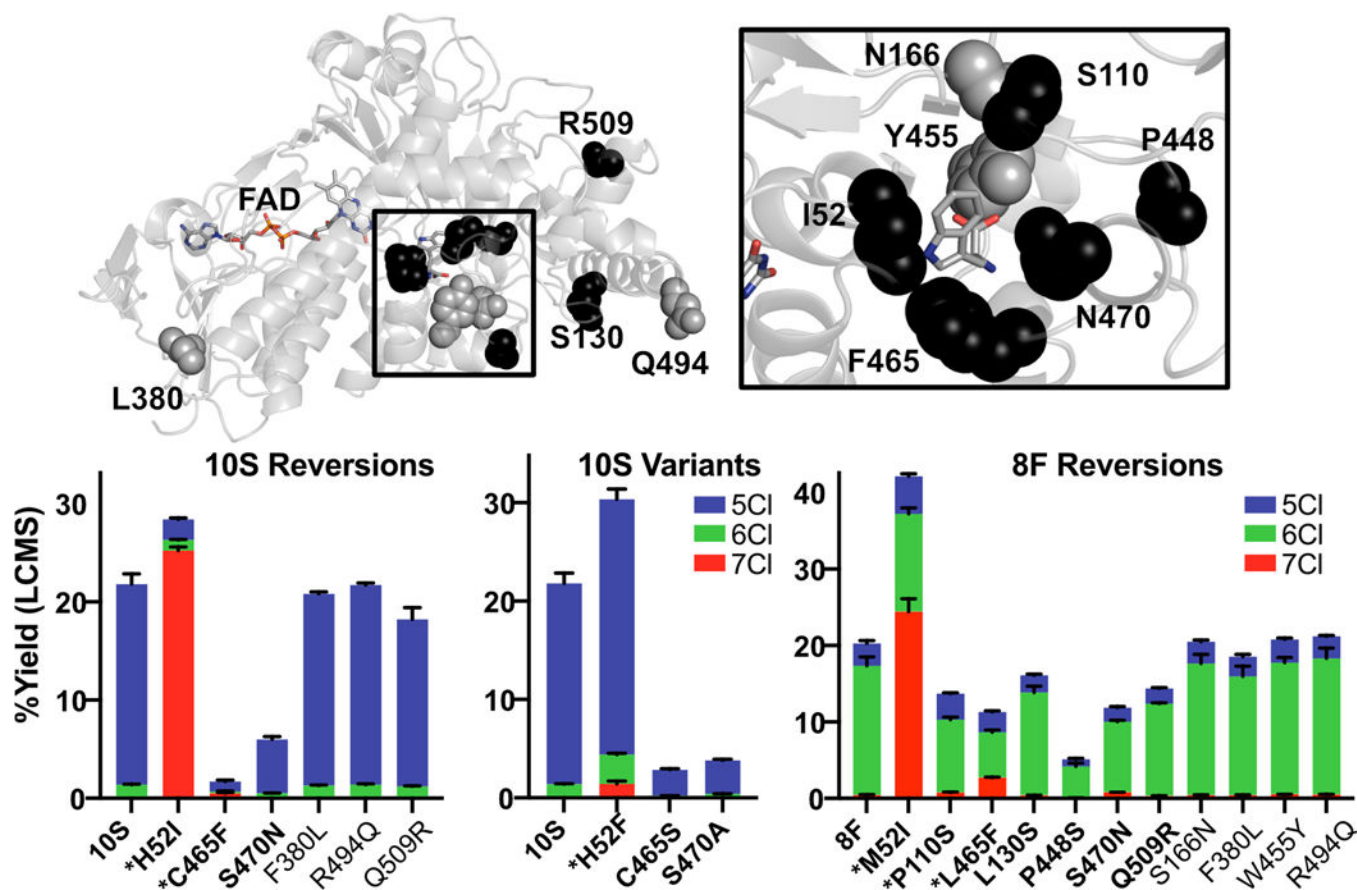
**Bigger Picture Statement:**

Flavin-dependent halogenases (FDH) have been widely studied for their ability to site-selectively halogenate under mild conditions (i.e. water as solvent, NaX as a halogen source, and O<sub>2</sub> as an oxidant). Previously, we and others have demonstrated the tunability of substrate scope and site selectivity through directed evolution. Moving forward, the ability to predict site selectivity for different FDHs on a range of substrates would be invaluable to biocatalysis efforts. Here, we characterize FDH variants evolved for orthogonal selectivities using X-ray crystallography and reversion mutations to probe the mechanism by which selectivity was altered. We use density functional theory calculations and molecular dynamics simulations to explore the feasibility of potential halogenating species within the activated FDHs. Lastly, we develop a computational model for predicting site selectivity within FDHs.



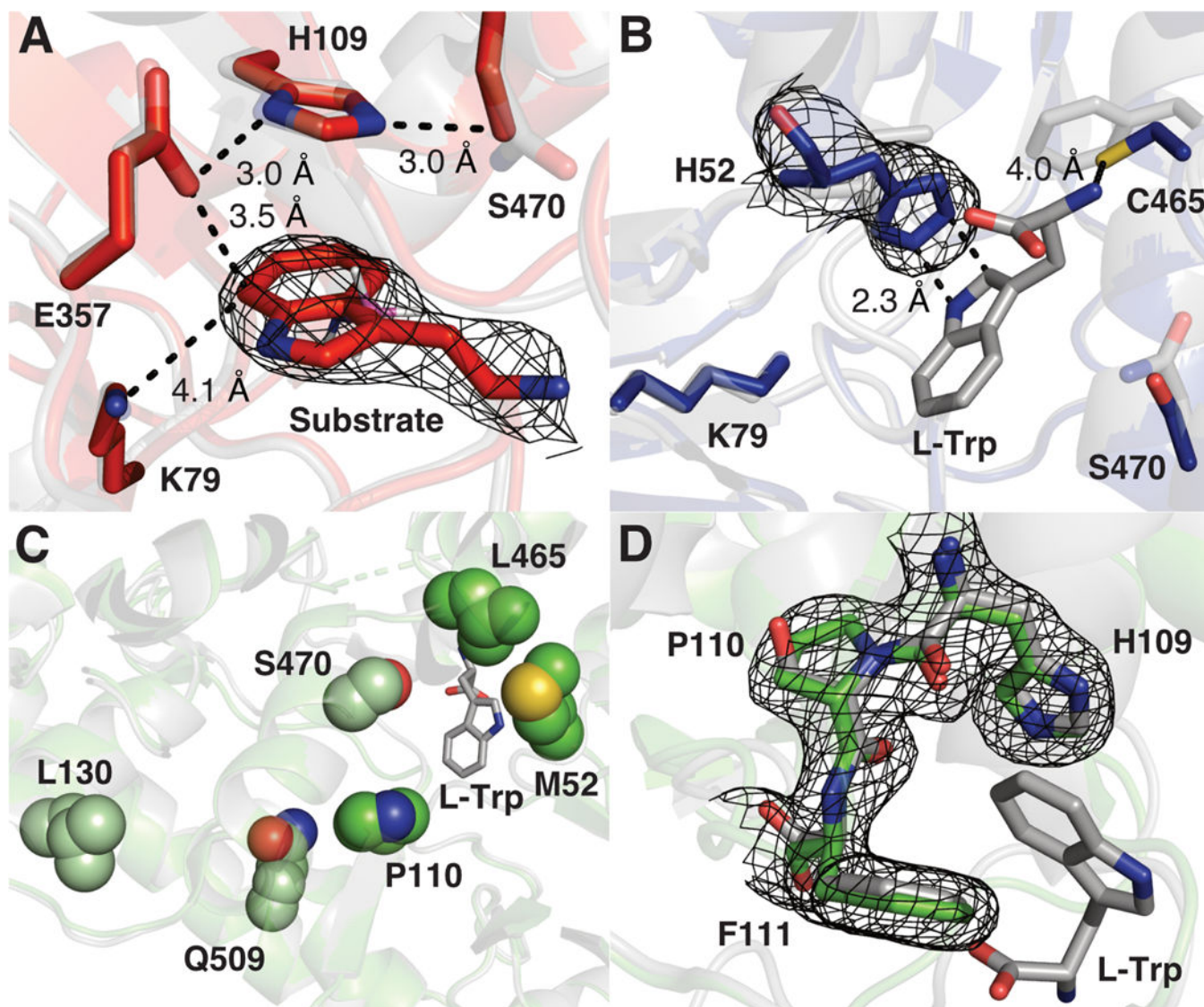
**Figure 1. Substrate positioning within flavin-dependent halogenases allows for site-selective halogenation.**

A) Substrate positioning within FDHs controls selectivity. B) An electron-rich aromatic ring, shown with an electron-donating group (EDG), undergoes electrophilic aromatic substitution (EAS). Selective EAS catalysis requires an active site lysine residue thought to form a reactive chloramine intermediate or activate HOCl. C) Directed evolution was used to engineer RebH variants that halogenate tryptamine at C5, C6, or C7 with high site selectivity.



**Figure 2. Summary of reversion mutations.**

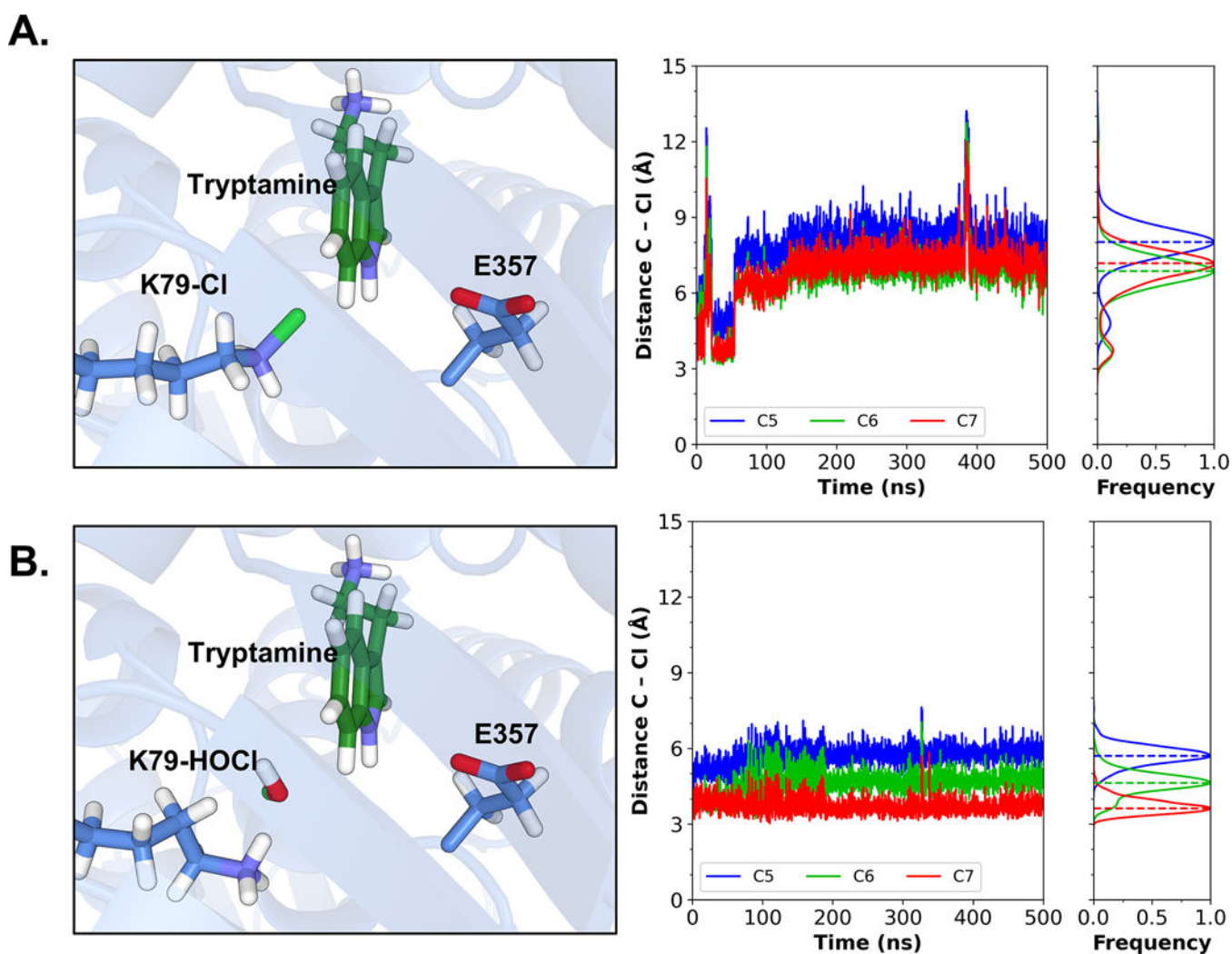
The locations of the 11 sites mutated during the evolutionary campaign to alter site selectivity are shown as spheres within the RebH crystal structure (PDB ID: 2OAI)<sup>14</sup>. Dark grey residues were found to have no effect on yield or selectivity. Black residues were found to impact yield and/or selectivity. FAD and L-Trp are shown as sticks. Zoom-in on active site is shown in the boxed panel to the right. Graphs indicate yield and selectivity of reversion mutations. Starred residues were found to alter site selectivity. N=3.



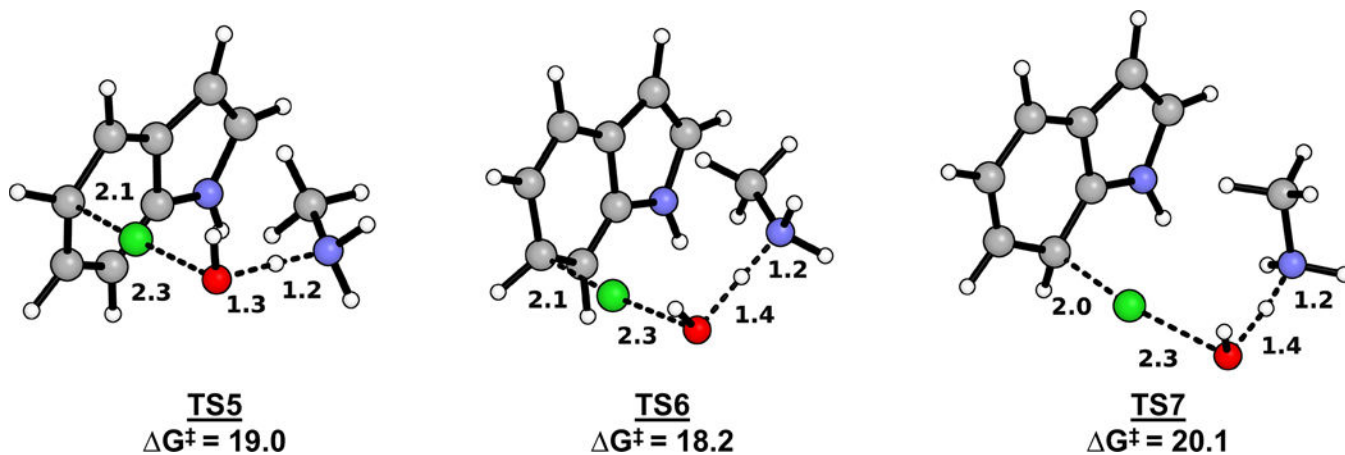
**Figure 3. Comparison of crystal structures between variants.**

Crystal structures of variants 0S (red), 10S (blue), and 8F (green) overlaid with RebH (grey, PDB ID: 2OA1). A) Tryptamine bound in the active site of 0S. B) H52 in 10S clashes with L-Trp in RebH. C) Mutations affecting 8F selectivity are shown as darker green spheres. Mutations affecting only 8F yields are shown as light green spheres. D) 8F mutation P110 alters 8F selectivity and yield, perhaps by adding rigidity to neighboring aromatic residues.

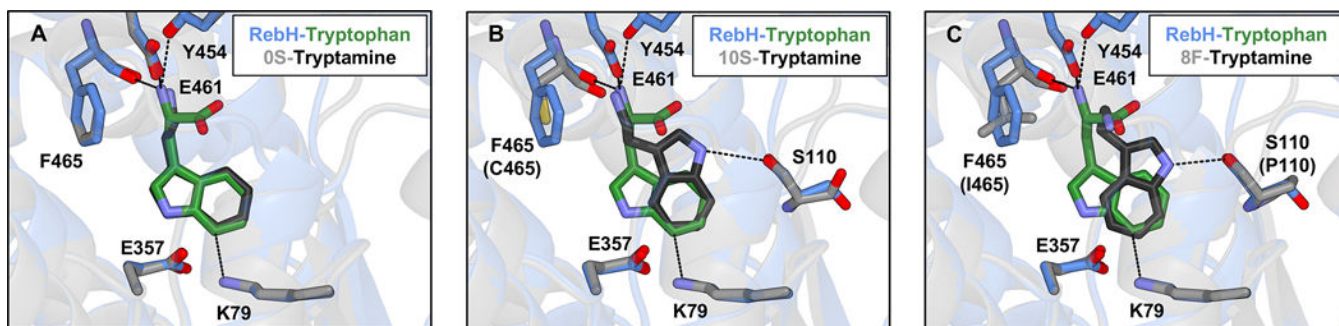






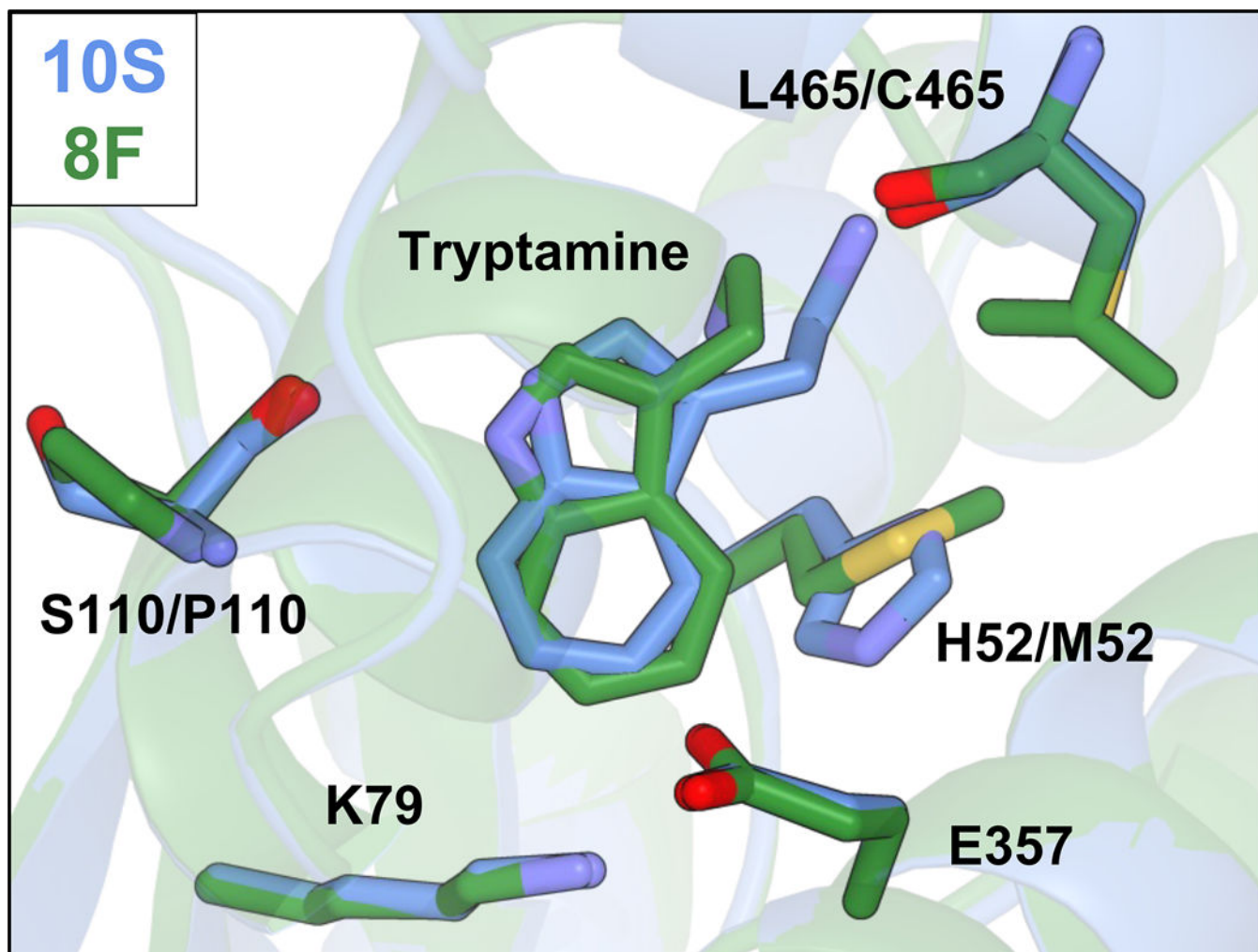


**Figure 5. Transition state structures for the acid-catalyzed chlorination of indole.** DFT-calculated transition state structures are shown for acid-catalyzed chlorination at the 5-, 6- and 7-positions of indole. Geometric similarities and energetic differences between structures are highlighted. Chlorination at C6 is most favorable, while C7 is least favorable according to the calculated activation energy.



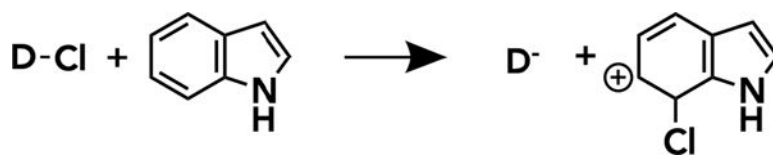
**Figure 6. Substrate binding poses.**

Substrate binding poses obtained from crystal structures and docking calculations are shown overlaid with the crystal structure of RebH bound to tryptophan (2OA1). Overlays of A) 0S B) 10S and C) 8F binding poses.



**Figure 7. Comparison of substrate binding in variants.**

Docked binding orientations for in 10S (blue) and 8F (green). Tryptamine positioning is affected by residue 52. H52 in 10S projects C5 toward K79, and M52 in 8F projects C6 toward K79.



Scheme 1. Electrophilic attack of different halonium donors on the 7-position of indole.

**Table 1.**Kinetic parameters of RebH and variants.<sup>11</sup>

Enzyme	$K_m$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{min}^{-1}$ )	$k_{\text{cat}}/K_m$ ( $\text{min } \mu\text{M}^{-1}$ )	Number of Mutations
RebH	9	0.023	$2.6 \times 10^{-3}$	0
0S	10.6	0.135	$1.3 \times 10^{-2}$	1
8F	1747	0.037	$2.1 \times 10^{-5}$	11
10S	160	0.028	$1.8 \times 10^{-4}$	6

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2.**Thermodynamic parameters for electrophilic attack of different  $\text{Cl}^+$  donors on indole.

Entry	Donor-Cl	Donor <sup>-</sup>	$G^\circ_{\text{Whel}}$
1	$\text{CH}_3\text{NHCl}$	$\text{CH}_3\text{NH}^-$	72.4
2	$\text{HOCl}$	$\text{HO}^-$	21.2
3	$\text{CH}_3\text{NH}_2\text{Cl}^+$	$\text{CH}_3\text{NH}_2$	1.7
4	$\text{H}_2\text{OCl}^+$	$\text{H}_2\text{O}$	-57.8
5	$\text{HOCl} +$	$\text{H}_2\text{O} +$	-3.3
	$\text{CH}_3\text{NH}_3^+$	$\text{CH}_3\text{NH}_2$	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**Substrate ( $G_{TS}^\ddagger$ ) vs catalyst ( $G_N^\circ$ ) control over site selectivity of tryptamine in RebH and variants.

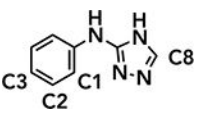
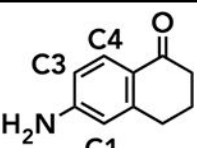
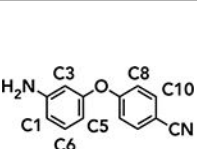
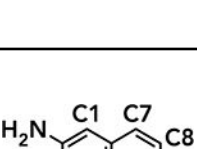
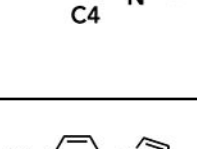
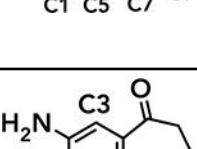
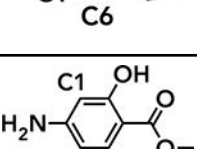
FDH	C <sub>n</sub>	$G_{TS}^\ddagger$ <sup>a</sup> (kcal/mol)	$G_N^\circ$ <sup>b</sup> (kcal/mol)	$G_{TS}^\ddagger$ (kcal/mol)	$G_N^\circ$ (kcal/mol)	Obs. <sup>c</sup> (%)
RebH	5	19.0	-	0.8	-	0
	6	18.2	2.2	0.0	1.9	0
	7	20.1	0.3	1.9	0.0	99
0S	5	19.0	4.2	0.8	3.5	0
	6	18.2	2.8	0.0	2.1	0
	7	20.1	0.7	1.9	0.0	99
8F	5	19.0	2.2	0.8	0.0	11
	6	18.2	2.3	0.0	0.1	89
	7	20.1	4.0	1.9	1.8	0
10S	5	19.0	4.2	0.8	0.0	94
	6	18.2	4.2	0.0	0.0	6
	7	20.1	-	1.9	-	0

<sup>a</sup>Transition state free energy barrier for chlorination in solution calculated by DFT.<sup>b</sup>Free energy of NAC formation in enzyme estimated by MD simulations. Empty values indicate no NAC is formed at this site.<sup>c</sup>Percent of different isomers as determined by <sup>1</sup>H NMR spectroscopy



Table 4.

Substrate ( $G_{\text{whel}}$ ) vs catalyst ( $G_{\text{N}^\circ}$ ) control over site selectivity of non-native substrates in RebH.

Entry	Compound	C <sub>n</sub>	$G_{\text{whel}}^a$ (kcal/mol)	$G_{\text{N}^\circ}^b$ (kcal/mol)	Obs. <sup>c</sup> (%)
1		1	5.8	-	0
		2	20.9	2.4	0
		3	0.0	0.0	100
		8	2.7	-	0
2		1	0.0	-	100
		3	0.4	0.0	0
		4	29.1	4.1	0
3		1	1.8	0.0	100
		3	5.1	-	0
		5	0.0	-	0
		6	34.4	1.6	0
		8	28.7	-	0
		10	48.0	-	0
4		1	0.0	0.3	100
		3	18.7	0.0	0
		4	26.6	2.3	0
		7	20.6	-	0
		8	29.7	-	0
		9	15.2	-	0
5		1	25.2	-	0
		5	26.0	-	0
		7	0.0	-	0
		9	6.2	-	100
6		1	0.0	0.0	93
		3	0.7	-	7
		6	20.8	2.5	0
7		1	1.7	-	6
		3	0.0	0.0	94
		4	35.4	-	0

<sup>a</sup>Relative free energy of Wheland intermediate formation at the given site calculated by DFT.<sup>b</sup>Relative free energy of NAC formation in RebH estimated by MD simulations. Empty values indicate no NAC is formed at this site.<sup>c</sup>Percent of different isomers as determined by <sup>1</sup>H NMR spectroscopy