Flavin-dependent halogenases increasingly attract attention as biocatalysts in organic synthesis, facilitating environmentally friendly halogenation strategies that require only FADH₂, oxygen, and halide salts. Different flavin-dependent tryptophan halogenases regioselectively chlorinate or brominate tryptophan’s indole moieties at C5, C6, or C7. Here, we present the first substrate-bound structure of a tryptophan 6-halogenase, namely Thal, also known as ThdH, from the bacterium Streptomyces albogriseolus at 2.55 Å resolution. The structure revealed that the C6 of tryptophan is positioned next to the ε-amino group of a conserved lysine, confirming the hypothesis that proximity to the catalytic residue determines the site of electrophilic aromatic substitution. Although Thal is more similar in sequence and structure to the tryptophan 7-halogenase RebH than to the tryptophan 5-halogenase PyrH, the indole binding pose in the Thal active site more closely resembled that of PyrH than that of RebH. The difference in indole orientation between Thal and RebH appeared to be largely governed by residues positioning the Trp backbone atoms. The sequences of Thal and RebH lining the substrate binding site differ in only few residues. Therefore, we exchanged five amino acids in the Thal active site with the corresponding counterparts in RebH, generating the quintuple variant Thal-RebH5. Overall conversion of L-Trp by the Thal-RebH5 variant resembled that of WT Thal, but its regioselectivity of chlorination and bromination was almost completely switched from C6 to C7 as in RebH. We conclude that structure-based protein engineering with targeted substitution of a few residues is an efficient approach to tailoring flavin-dependent halogenases.

Whereas chemical halogenation of aromatic compounds requires toxic elemental halogens and Lewis acids as catalysts, flavin-dependent halogenases (FDHs)³ can catalyze electrophilic aromatic substitutions with innoxious halide salts and O₂ as substrates and FADH₂ as co-factor (1, 2). FDHs are responsible for the regioselective chlorination or bromination of a vast variety of natural products (3) and have attracted a lot of attention due to their potential use as biocatalysts (2, 4, 5). Slow turnover and low stability still limit their large-scale application, but various strategies have been pursued to tackle these challenges (6–8). Bacterial tryptophan halogenases were the first FDHs to be described (9–11). In addition, FDHs were identified that halogenate substituted tryptophan (12, 13) or chlorinate tryptophan only when it is part of a substrate oligopeptide (14). FDHs acting on free substrates other than tryptophan were also described (15–18), including fungal enzymes that accept large substrates (19–21) or have rather broad substrate specificity (22, 23), raising the hope that diverse compounds can be enzymatically halogenated in vitro. Moreover, FDHs have been integrated into engineered biosynthetic pathways in vivo (24, 25). The first structure of an FDH showed that these enzymes consist of two subdomains, a conserved box-shaped subdomain harboring the FAD binding site and a variable subdomain that is pyramidal in tryptophan halogenases (26). Tryptophan is bound at the interface of the box and the pyramid.

In addition to using benign reactants under mild conditions, the regioselectivity of FDHs provides a further advantage of enzymatic halogenation in contrast to chemical approaches. Different FDHs are known that halogenate tryptophan selectively on C5, C6, or C7 of the indole moiety, whereas chemical halogenation preferably occurs at C2 (27). Thus, for the natural substrate L-Trp, the site of enzymatic halogenation by flavin-dependent tryptophan halogenases is not governed by electronic factors of the substrate but determined sterically by the orientation of the substrate in the active site. This model was suggested based on the structure of the tryptophan 7-halogenase PrnA (26) and confirmed by the structure of the tryptophan 5-halogenase PyrH (28), both in complex with their native substrate tryptophan. These structures showed that the carbon atom to be halogenated is positioned most closely to the ε-amino group of the catalytic lysine, which is conserved among FDHs. To achieve this positioning, tryptophan is bound with distinctly different poses by tryptophan 7-halogenases and by tryptophan 5-halogenases. The indole moiety of L-Trp bound to PyrH is flipped by about 180° relative to the orientation observed in the tryptophan 7-halogenases PrnA and RebH. Several FDHs have been described that halogenate tryptophan...
on C6, namely Thal (29) (also known as ThdH (30)), SttH (31), and thermophilic Th-Hal (32), which are bona fide tryptophan 6-halogenases, and KtzR, which produces 6-chlorotryptophan from tryptophan but has higher catalytic efficiency for converting 7-Cl-L-Trp to 6,7-di-Cl-L-Trp (12). BorH is another putative tryptophan 6-halogenase; however, its activity has not yet been confirmed in vitro (33). Based on sequence identity, SttH, Th-Hal, and KtzR are closely related to each other and to the tryptophan 5-halogenase PyrH, whereas on the other hand, Thal and BorH are closely related to each other and to the tryptophan 7-halogenases RebH and PrnA (Table 1). Crystal structures of SttH (34) and Th-Hal (32) have been published, but only in the absence of substrate. No structural data are available for Thal, BorH, or KtzR. Modeling the substrate into the active site of the empty enzyme is hampered by protein conformational changes that occur upon substrate binding (28, 35, 36). One would expect that in tryptophan 6-halogenases C6 of the indole moiety occupies a position roughly equivalent to the position of C7 or C5 in tryptophan 7-halogenases or 5-halogenases, respectively. This positioning of C6 could be achieved by slightly rotating the indole moiety, starting either from the binding pose observed in the tryptophan 7-halogenases PrnA and RebH or from the flipped binding pose observed in the tryptophan 5-halogenase PyrH (37). Thus, it remains elusive how L-Trp binds to the active site of tryptophan 6-halogenases.

With their application as biocatalysts in mind, enzyme optimization has been performed on FDHs, aiming at increasing their lifetime and changing their substrate scope or regioselectivity (1, 2). A large amount of work has gone into both structure-based engineering and directed evolution (27). These efforts resulted in enzymes with higher stability (7, 8), higher activity (23, 38), different substrate scope (24, 37, 39, 40), or altered regioselectivity (34, 37, 40, 41). Despite the already achieved success, these tasks still pose formidable challenges. Directed evolution requires many rounds of high-throughput screening, and frequently it remains unclear how beneficial mutations exert their positive effect (37). Attempts to alter the regioselectivity with tryptophan as substrate have so far resulted in an incomplete change of selectivity so that only one-third of substrate was converted to a new regioisomer, whereas the majority was still halogenated on the initially preferred position (34, 41).

We embarked on a project to further elucidate the structural basis of regioselectivity of tryptophan halogenases. We aimed at crystallizing the tryptophan 6-halogenase Thal together with its native substrate L-Trp. Based on its structure, we designed a Thal variant with five amino acid substitutions in the active site with the goal to switch the regioselectivity from the 6- to the 7-position.

Results

Overall structure

The crystal structure of Thal without FAD, chloride, and tryptophan (apo-Thal) and a structure in complex with its substrate L-Trp (Trp-Thal) obtained by soaking contained two copies of Thal per asymmetric unit (Table S1). Amino acids (aa) 2–529 could be modeled in all chains except for chain B of the apo-Thal structure, which lacked interpretable electron density for residues 450–456. In the apo-Thal structure, the active site of chain A contains positive difference density, which we could not faithfully model. This difference density suggests that the substrate binding site of apo-Thal chain A is not truly empty. Clear electron density was observed for the substrate L-Trp in both chains of Trp-Thal.

In solution, Thal behaved as a monomer according to the elution volume from size-exclusion chromatography and its hydrodynamic radius as determined by dynamic light scattering (data not shown). Nevertheless, Thal appears to form a dimer in the crystal (Fig. S1) like other tryptophan halogenases (26, 28, 32, 34–36). The arrangement of protomers in the dimer is very similar to that of other tryptophan halogenases with a root mean square deviation (r.m.s.d.) between the apo-Thal dimer and the apo-RebH (PDB code 2OAM) dimer of only 1.4426 Å for 1027 structurally corresponding residues.

The structural similarity is even higher for the protomers with an r.m.s.d. of 0.9675 and 0.9013 Å for 521 and 519 structurally corresponding residues, respectively. This is not surprising, given the sequence identity between Thal and RebH of 64% with almost no insertions or deletions, except for Asp-101 and the C-terminal Ser-531 that are inserted in Thal and Asp-449 inserted in RebH (Fig. 1). Each monomer consists of a single domain, which folds into a larger rectangular box containing the FAD binding site, and a smaller triangular pyramid (Fig. 2), as described for PrnA (26) and other tryptophan halogenases (28, 32, 34–36). The substrate binding site is located between the box and the pyramid and occupies the same position as in RebH (Fig. S2a). Thal residues Lys-79 and Glu-358 align with the catalytic key residues Lys-79 and Glu-357 of RebH. The difference from the catalytic residues Lys-83 and Glu-362 previously reported for Thal (29) is due to a difference in numbering with an offset of 4. At the level of the backbone, just a few structural differences between Thal and RebH become apparent, mainly at the sites of insertions (Fig. S2a).

Sequence conservation between Thal and the tryptophan 5-halogenase PyrH or the tryptophan 6-halogenases Th-Hal and SttH is lower than between Thal and RebH (Table 1). Accordingly, structural differences between Thal and PyrH, Th-Hal or SttH are bigger than between Thal and RebH. As a L-Trp–bound structure is available only for PyrH, we will focus on a structural comparison of Thal with PyrH rather than SttH or Th-Hal. Thal (PDB code 6H44) and PyrH (PDB code 2WEU) protomers align with only 466 structurally corresponding residues and an r.m.s.d. value of 1.22 Å (Fig. S2b). For the dimer, the r.m.s.d. between Thal and PyrH is even higher with 1.73 Å for 928 structurally corresponding residues instead of the total of 1056 possible residues. Structural conservation is highest in the box-shaped FAD binding module. In the pyramidal subdomain, some secondary structure elements are rearranged or connected differently, which is mostly due to insertions or deletions (Fig. 1).
Regioselectivity switch of tryptophan 6-halogenase Thal

<table>
<thead>
<tr>
<th>Thal</th>
<th>β1</th>
<th>α1</th>
<th>β2</th>
<th>α2</th>
<th>β3</th>
<th>α3</th>
<th>β4</th>
<th>α4</th>
<th>β5</th>
<th>α5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>Rebl</td>
<td>β6</td>
<td>β7</td>
<td>α6</td>
<td>α7</td>
<td>β8</td>
<td>β9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrmA</td>
<td>β10</td>
<td>β11</td>
<td>β12</td>
<td>β13</td>
<td>β14</td>
<td>n1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrmA</td>
<td>200</td>
<td>210</td>
<td>220</td>
<td>230</td>
<td>240</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrmA</td>
<td>260</td>
<td>270</td>
<td>280</td>
<td>290</td>
<td>300</td>
<td>310</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PyCh</td>
<td>β15</td>
<td>β16</td>
<td>α8</td>
<td>α9</td>
<td>β17</td>
<td>β18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PyCh</td>
<td>320</td>
<td>330 *</td>
<td>340</td>
<td>350</td>
<td>360</td>
<td>370</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>StCh</td>
<td>β19</td>
<td>β20</td>
<td>β21</td>
<td>β22</td>
<td>β23</td>
<td>α10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>StCh</td>
<td>380</td>
<td>390</td>
<td>400</td>
<td>410</td>
<td>420</td>
<td>430</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>α11</td>
<td>α12</td>
<td>α13</td>
<td>α14</td>
<td>α15</td>
<td>α16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>440</td>
<td>450</td>
<td>460</td>
<td>470</td>
<td>480</td>
<td>490</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tryptophan binding site

The structure of apo-Thal, which we had determined first, revealed a substrate binding site very similar to that of RebH. All residues of the substrate binding site that differ between Thal and RebH are smaller in Thal, leaving more space for the substrate. As we were not able to determine the L-Trp binding pose or to deduce a reason for the different regioselectivity from the structure of apo-Thal, we determined the structure of Thal in complex with its native substrate L-Trp. As expected, L-Trp is oriented in such a way that its C6 position points toward the catalytic residues Lys-79 and Glu-358 (Fig. 3). The substrate binding site can be divided into three regions, namely the catalytic, the π-stacking, and the backbone-binding region (Fig. 4).

The catalytic region harbors the Lys-79 that is universally conserved among flavin-dependent halogenases and Glu-358 that is conserved among tryptophan halogenases (Fig. 4). In addition, we assign several small amino acids, including Val-52, Val-82, and Ser-360 to this region. The π-stacking region fixes the indole moiety (Fig. 4) and may shield the nonhalogenated aromatic carbon atoms of the indole from reactive HOCl. It consists of the aromatic residues His-110, Phe-112, Phe-465, and Trp-466. In addition, we assign Pro-53 and Pro-111 to this middle region (Fig. 4). The backbone-binding region, which is in part poorly ordered in the absence of substrate, contains residues that coordinate the α-amino and α-carboxyl group of L-Trp. It contains Gly-113 and Lys-114 from a loop that flips into position upon substrate binding; Tyr-454, Tyr-455, and Glu-461 from the substrate-binding loop; and other polar residues like Thr-467 and Ser-470 (Fig. 4).

Structural changes upon L-Trp binding

A comparison of apo-Thal and Trp-Thal indicates structural changes upon binding of L-Trp. The B factors of the loop covering the FAD binding site (FAD loop; aa 38–45) increase while the substrate-binding loop (aa 450–457) becomes more ordered. In chain B, Phe-112 undergoes a peptide flip, moving Gly-113 and Leu-114 closer to the substrate by 2.4 and 1.4 Å.

Table 1

Pairwise sequence identity of various tryptophan halogenases

The tryptophan 6-halogenases Thal and BorH are closely related to each other and to the tryptophan 7-halogenase RebH. The tryptophan 6-halogenases SttH and Th-Hal are closely related to each other and to the tryptophan 5-halogenase PyrH. The sequence identity is lower between Thal or BorH on the one hand and SttH or Th-Hal on the other hand. The sequence identity was derived from pairwise structure-based sequence alignments with the DALI server (53) except for BorH (asterisks), for which no structure is available.

<table>
<thead>
<tr>
<th>ID%</th>
<th>Thal</th>
<th>BorH*</th>
<th>RebH</th>
<th>PrnA</th>
<th>PyrH</th>
<th>SttH</th>
<th>Th-Hal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thal</td>
<td>100</td>
<td>70*</td>
<td>64</td>
<td>57</td>
<td>42</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>BorH</td>
<td>100</td>
<td>63*</td>
<td>56*</td>
<td>40*</td>
<td>39*</td>
<td>39*</td>
<td></td>
</tr>
<tr>
<td>RebH</td>
<td>100</td>
<td>55</td>
<td></td>
<td>41</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrnA</td>
<td>100</td>
<td>42</td>
<td>40</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PyrH</td>
<td>100</td>
<td>58</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SttH</td>
<td>100</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Th-Hal</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
respectively. A similar peptide flip involving the equivalent residue Phe-111 is observed for RebH and Trp-RebH (35, 36). Phe-112 in chain A of apo-Thal adopts the same conformation as chain A and B of Trp-Thal, presumably because the substrate binding site is not truly empty, as indicated by positive difference density. Upon substrate binding, residues of the backbone-binding region that form polar contacts with L-Trp become ordered and block the way out through the substrate tunnel to the dimer interface (Fig. S3). The α-amino group of L-Trp forms a salt bridge with the carboxyl group of Glu-461 and two H-bonds, one to the hydroxyl group of Tyr-454 and the other to the backbone carbonyl of Tyr-455 (Fig. 4d). The carboxyl group of L-Trp forms H-bonds to the hydroxyl group of Tyr-455 and to two conserved water molecules (see below). Upon substrate binding, several residues move out of the active site (Fig. 5). This makes room for L-Trp both in the catalytic region (Thr-51, 0.6 Å; Val-52, 0.7 Å; Glu-358, 1.0 Å) and in the π-stacking region (Pro-53, 0.7 Å; Phe-112, 1.0 Å; Trp-466, 0.6 Å).

**Comparison of substrate binding in Thal, RebH, and PyrH**

The indole moiety in Trp-Thal is almost parallel to that in Trp-RebH but flipped by 180°, so that the five-membered rings do not superpose at all (Fig. 3). In contrast, the halogenated carbon atoms (C6 in Trp-Thal and C7 in Trp-RebH) are only 0.5 Å apart. The same flip of ~180° relative to the orientation in RebH is present in PyrH. Hence, the binding pose of the substrate indole moiety in Thal is more like that in PyrH than that in RebH. Interestingly, l-Trp adopts different rotameric conformations in Thal and in RebH, whereas it is bound to PyrH in

---

**Figure 4. Depiction of the tryptophan binding site.** A, the binding site is divided into three regions: catalytic (light blue), π-stacking (light pink), and backbone-binding (light green). The substrate L-Trp is shown with yellow carbon atoms, and the halogenation position C6 is highlighted with a star. Direct hydrogen bonds or salt bridges between Thal and L-Trp are shown as dotted yellow lines. B, catalytic region. C, π-stacking region. D, backbone-binding region.
Regioselectivity switch of tryptophan 6-halogenase Thal

Figure 5. Structural flexibility of conserved residues in the π-stacking region. Overlay of the π-stacking region of chain A of apo-Thal (gray), Trp-Thal (dark blue), and Trp-RebH (PDB code 2E4G; violet). The substrate tryptophan is shown in yellow for Thal and in violet for RebH. The numbering of residues corresponds to that of Thal. RebH contains Ser-110 instead of Pro-111. In apo-Thal, Pro-53, Phe-112, and Trp-466 have moved into the substrate binding site relative to Trp-Thal. In chain B, the movement toward the empty substrate binding site of apo-Thal is even larger (see Fig. S7).

Along with the substrate L-Trp, some water molecules get trapped in the active site. The position of these waters is conserved between the two crystallographically independent Thal monomers and sometimes even among Thal and RebH. There are two prominent water molecules forming H-bonds with the L-Trp carboxyl group. Their positions are conserved between Thal and RebH, and here we call them W1 and W2 for simplicity (Fig. 6). W1 corresponds to A710 and B721 in Trp-Thal and to water A1031 and B1125 in Trp-RebH (PDB code 2E4G). W1 forms an H-bond with the carboxyl group of the substrate and one H-bond with the protein via the side chain carboxyl group of the Glu-461 and the backbone NH of Asn-54 (Fig. 6, a and c). The H-bond pattern between W1 and the protein is identical in Thal and RebH, and W1 is also present in chain A of apo-Thal (water A718) (Fig. 6b) and in apo-RebH (waters A777 and B646; PDB code 2OAM) (Fig. 6d). W2 corresponds to A752 and B712 in Trp-Thal and to waters A1080 and B1091 in Trp-RebH (PDB code 2E4G). In contrast to W1, W2 is missing in the apo-structures of Thal and RebH (Fig. 6, b and d). Like W1, W2 forms an H-bond with the carboxyl group of the substrate L-Trp. In addition, W2 forms an H-bond with the carbonyl oxygen of Gly-113 (Gly-112 in RebH). This H-bond in combination with the absence of W2 in the apo-structures may be one reason for the peptide flip of Phe-112 upon substrate binding. Further H-bonds between W2 and the protein differ between Thal and RebH. W2 forms two direct H-bonds with RebH via the side-chain amides of Asn-467 and Asn-470 (Fig. 6c). In contrast, W2 in Thal forms an H-bond with a secondary water molecule (A717 and B713, W3 in Fig. 6a), which in turn is bound to Thal via two H-bonds to the hydroxyl groups of Thr-467 and Ser-470. W3 is conserved between Trp-Thal monomers, but not among Trp-Thal and Trp-RebH. This difference between Thal and RebH is presumably due to the amino acid substitutions in the backbone-binding region of the active site.
Switching the regioselectivity by structure-guided mutagenesis

To test whether a few targeted mutations are sufficient to switch the regioselectivity, we substituted five residues in the substrate binding site of Thal by the corresponding amino acids of RebH. The resulting variant Thal-RebH5 harbored the substitutions V52I, V82I, S360T, G469S, and S470N already described above (Fig. S4). To assess the overall conversion of tryptophan, assay mixtures using purified protein and concomitant cofactor regeneration were analyzed via HPLC-MS (Fig. 7). After a reaction time of 86 h, 70% conversion of 5 mM L-Trp to L-chlorotryptophan was reached. This proved that these substitutions in the active site do not affect the overall conversion.

In addition to the main signal observed in the chromatogram, a less intense signal indicated the formation of a second product (Fig. 7c). As both signals were identified as monochlorinated tryptophan with the same m/z pattern from their corresponding mass spectra, we assumed that another regioisomer might have formed. HPLC-MS analysis showed that both isomers were formed in a ratio of 95:5, whereas comparison with authentic standards revealed that 7-chlorotryptophan was formed as the main product. The minor signal corresponds to 6-chlorotryptophan.

In addition, chlorination of L-Trp was conducted on a preparative scale using *Escherichia coli* lysate containing overexpressed halogenase. After 26 h, a maximum conversion of 30%
into the expected monochlorinated products (95:5 ratio) was observed via HPLC-MS. Separation and purification of the main product by reversed-phase HPLC finally allowed the identification of L-7-chlorotryptophan by one- and two-dimensional NMR spectroscopy (Figs. S8–S13). This proves the successful and nearly quantitative change of the regioselectivity from position 6 to 7 by substituting five residues of Thal by the corresponding RebH residues.

Enzymatic bromination reactions were performed on a preparative scale analogously to the chlorination described above. After a reaction time of 48 h, HPLC-MS data revealed 15% conversion to monobrominated L-Trp. Like the chlorination, two product signals in a ratio of 95:5 were observed. As the overall conversion was lower, the minor isomer was hardly visible in the UV trace, but its quantity was sufficient for MS detection. After purification, C7-brominated L-Trp was isolated as the major bromination product of Thal-RebH5, which was also confirmed by NMR analysis (Figs. S14–S17). Corresponding to the previous observations made for Thal-RebH5–catalyzed chlorination, 6-bromotryptophan had been formed in negligible quantity, as identified via an authentic HPLC standard.

**Structure of Thal-RebH5**

Thal-RebH5 crystallized like the WT protein, and the 2.12 Å structure (Table S1) showed clear electron density for the five mutated residues. The rotamers of substituted residues in Thal-RebH5 correspond to those of RebH except for Ser-469. An overlay of apo-Thal, apo-Thal-RebH5, and apo-RebH showed that in Thal-RebH5, most active-site residues, including many conserved residues, adopt an intermediate conformation between Thal and RebH (Fig. S5). In some cases, like Thr-360 or Phe-465, Thal-RebH5 more closely resembles RebH than WT Thal. We also soaked L-Trp into Thal-RebH5 crystals and obtained a structure at 2.55 Å resolution. There was clear positive difference density in the active site of chain A (Fig. S6a) and some positive difference density in the active site of chain B. The difference density in chain A is akin to the shape of l-Trp but cannot be modeled faithfully. The best interpretation that we achieved is a superposition of L-Trp as it is bound to Thal (conformer A) and to RebH (conformer B) (Fig. S6b). Occupancy refinement converged to values of about 0.5 for conformer A and about 0.36 for conformer B. If only one conformer was modeled, the occupancy refined to values of about 0.8–0.9. Placing a single conformer, however, resulted in positive differ-
ence density in the region occupied by the five-membered ring of the other conformer (Fig. S6, c and d), supporting the presence of the second conformer. While the plane of the indole moiety was well-defined, each conformer could be rotated or translated to a certain degree in this plane. Positional refinement of L-Trp did not converge to a uniquely defined solution, presumably due to the partial occupancy and the limited resolution. Therefore, we decided not to deposit this structure in the PDB but to provide coordinates and structure factors as supporting information. The Trp-Thal-RebH5 structure strongly suggests that Thal-RebH5 is permissive to binding L-Trp with its C7 positioned most closely to the /H9280-amino group of the catalytic Lys-79.

Discussion

Here, we present the first experimental structure of a tryptophan 6-halogenase in complex with its native substrate L-Trp. The structure confirms the model of regioselectivity put forward by the Naismith and van Peé groups (26, 28). As in the tryptophan 7-halogenases PrnA and RebH and the tryptophan 5-halogenase PyrH, the halogenated aromatic carbon atom is positioned most closely to the e-amino group of the catalytic lysine in Thal. It clearly appears to be the binding pose of the substrate that determines the regioselectivity.

Our apo- and L-Trp-bound structures of Thal suggest that the empty active site of the enzyme may not be well-suited to reliably model the substrate binding pose for three reasons. First, in the absence of L-Trp, several conserved residues in the catalytic region and the π-stacking region move toward the center of the substrate binding site. If one places L-Trp from the Trp-Thal structure into structurally aligned apo-Thal, there would be severe clashes (Fig. S7). Second, major conformational changes between the apo-Thal and Trp-Thal structure affect key residues for L-Trp binding in the backbone-binding region (Fig. 4d and Fig. S3). Finally, two water molecules that are conserved between Thal and RebH mediate interactions between the L-Trp carboxylate and the protein, but only one of these waters is present in the substrate-free structures of Thal and RebH (Fig. 6).

Given the high degree of sequence and structural similarity between Thal and RebH, also in the active site, it may have come as some surprise to find in Thal an indole binding pose that more closely resembles that of PyrH than that of RebH. Altogether, this work underscores the importance of experimental structure determination, even in cases where structures of highly homologous proteins are available.

To gain more insight into which residues may be key to determine the regioselectivity in RebH and Thal, we performed a structural overlay of Trp-Thal and Trp-RebH (PDB code 2E4G) and copied the L-Trp from Trp-RebH to Trp-Thal and vice versa. We then checked for clashes indicating which residues might prevent binding of L-Trp to Thal in the same binding pose as in RebH and vice versa.

When placing L-Trp from RebH in the Trp-Thal structure (Fig. 8a), clashes mainly occur with conserved residues. This
Regioselectivity switch of tryptophan 6-halogenase Thal

may not be surprising, given the fact that all sequence differences in the active sites of Thal and RebH have smaller side chains in Thal. Several conserved residues of the catalytic and the π-stacking region move toward the substrate in Trp-Thal compared with Trp-RebH. Accordingly, there would be close contacts between l-Trp from RebH and Thal residues Pro-53, Phe-112, and Trp-466. Reasons why these residues move into the active site in Thal are not easily apparent, as even the shell of residues surrounding Pro-53, Phe-112, and Trp-466 is highly conserved between Thal and RebH. We suspect that the indole binding pose in Thal may not be determined primarily by steric hindrance from residues surrounding the indole moiety but by the different positioning of l-Trp Cα due to the different H-bond pattern between the l-Trp carboxylate and the backbone-binding region of the protein.

In contrast, moving l-Trp from Trp-Thal to Trp-RebH results in distinct clashes between RebH and l-Trp (Fig. 8b). Two RebH residues would form particularly short contacts with l-Trp. First, the side chain of Asn-470 would clash with C2 and C3 of the indole moiety. Second, the aromatic ring of Tyr-455 would clash with the l-Trp carboxyl group. Our conclusions are consistent with the previous notion of the van Peé and Naismith groups that in PrnA the equivalent residues Tyr-444 and Asn-459 prevent l-Trp from adopting the same orientation as that seen in PyrH (41). Intriguingly, even a random mutagenesis approach to select RebH variants that can halogenate larger substrates picked up the N470S substitution (39).

These observations suggest that the N470S substitution (RebH → Thal) may be of central importance to facilitate the indole binding pose observed in Trp-Thal. Interestingly, a serine is also present at this position in BorH (Ser-469), SttH (Ser-464), Th-Hal (Ser-460), and PyrH (Ser-455). In PyrH, however, mutating this Ser to Asn would not result in clashes with the l-Trp indole moiety. Still, the N470S substitution may be an enabling mutation in RebH that allows flipping the indole moiety relative to the orientation found in Trp-RebH. Due to the higher activity of the N470S variant on tryptamine, Andorfer et al. (37) used RebH N470S as starting point for their directed evolution of RebH toward C5 or C6 selectivity. Interestingly, RebH N470S retained regioselectivity for the 7-position. It produced >99% 7-chlorotryptamine and only minute amounts of 5- and/or 6-chlorotryptamine. However, all of the later variants that halogenated up to 95% of tryptamine on C5 or C6 had this founding N470S substitution.

The second RebH residue that would clash with l-Trp from Thal, namely Tyr-455, is conserved between RebH and Thal. Despite sequence conservation, its spatial arrangement is markedly different between RebH and Thal. In RebH, Tyr-455 is pushed toward the substrate relative to Thal. This displacement may be caused by the only insertion in RebH relative to Thal. A structure-based sequence alignment identifies Asp-449 as the inserted residue (Fig. 1). This different location of Tyr-455 decreases the space available to the l-Trp backbone atoms in RebH. Again, the extensive mutagenesis work on RebH by Andorfer et al. (37) provides circumstantial support for this hypothesis, as their RebH substitution variants with good halogenation activity on C5 or C6 of tryptamine had no activity toward the larger l-Trp.

Previous studies demonstrated that halogenase regioselectivity can be altered using different mutagenesis strategies. In a rational approach, a partial change of regioselectivity yielding a 2:1 ratio of 7- and 5-bromotryptophan was achieved for PrnA (41). In an approach more similar to ours, Micklefield and coworkers (34) tried to switch the regioselectivity of the tryptophan 6-halogenase SttH to position 5. Their work is based on structural differences between SttH and the closely related tryptophan 5-halogenase PyrH. A triple variant of SttH showed activity like WT SttH with relaxed regioselectivity, leading to 32% 5-Cl-l-Trp. A larger switch in regioselectivity was observed for the smaller substrate 3-indolepropionic acid, yielding 75% 5-chlorinated product compared with just 10% 5-chlorination for WT SttH. In a directed evolution approach, Lewis and co-workers (37) used random mutagenesis in combination with structure-guided mutation of active site residues to modify the regioselectivity of RebH. Their efforts to evolve the tryptophan 7-halogenase RebH into a C5- or C6-regioselective biocatalyst resulted in the generation of two new variants capable of chlorinating tryptamine with up to 95 and 90% regioselectivity for positions 5 and 6, respectively. For 2-methyltryptamine and N-methyltryptamine, the selectivity was even higher. Both the C5- and the C6-selective variant were inactive on the native substrate l-Trp.

In summary, previous studies achieved nearly complete selectivity switches for nonnative substrates like 3-indolepropionic acid and tryptamine both by rational, structure-based engineering (34) and by (random) mutagenesis coupled to high-throughput screening (37). For l-Trp, however, structure-based mutagenesis of both PrnA (41) and SttH (34) shifted regioselectivity to only about one-third of 5-Cl-l-Trp from 100% each of 7-Cl-l-Trp and 6-Cl-l-Trp, respectively. Our Thal-RebH5 variant exhibited a 95% change in regioselectivity toward C7 halogenation for both chlorination and bromination of l-Trp, constituting a nearly quantitative switch that has not been reported for any tryptophan halogenase until now. Both 3-indolepropionic acid and tryptamine are smaller than l-Trp and should have more flexibility to accommodate in the active site, offering a potential explanation why switching regioselectivity for smaller substrates was more successful. This may be particularly relevant for RebH, because, as pointed out before, the insertion of Asp-449 in RebH pushes the conserved Tyr-455 into the active site compared with Thal. Thus, Thal, which provides more space for the l-Trp backbone atoms, may represent a better starting point than RebH to evolve or engineer regioselectivity on the larger substrate l-Trp.

The success of our protein engineering is based on a careful structural analysis of substrate binding in closely related tryptophan halogenases with different regioselectivity. As such, our approach most closely resembles that of Shepherd et al. who also based their mutations on structures of two closely related tryptophan halogenases (34). Our results show that of the roughly 200 sequence differences between RebH and Thal, only a handful are required to switch regioselectivity. All of these substitutions affect active site residues. Substitution of residues distant from the substrate binding site that are frequently picked up in random mutagenesis but almost impossible to predict in rational structure-based approaches was not required.
Experimental procedures

**Plasmids and molecular cloning**

The pET-28a plasmid containing a codon-optimized gene for *E. coli* encoding Thal (Uniprot ID: A1E280) was reported earlier (43). The genes encoding Thal and Thal- RebH5 were separately subcloned into the vector pETM-11 (Gunter Stier, Heidelberg) using the Ncol and NotI restriction sites with primers listed in Table S2. The plasmid vector pET-21-adh encoding alcohol dehydrogenase from *Rhodococcus ruber* (RR-ADH) was kindly provided by Prof. Dr. Werner Hummel, Bielefeld University, and pCIBhis-prnF encoding the flavin reductase PrnF from *Pseudomonas fluorescens* was donated by Prof. Dr. Karl-Heinz van Pée, TU Dresden. Plasmid pGro7 was purchased from TaKaRa. Plasmids were purchased from Novagen. Generation of the Thal- RebH5 quintuple variant was performed according to the QuikChange II site-directed mutagenesis kit with primers listed in Table S2. Mutations were confirmed by DNA sequencing (GATC Biotech).

**Expression and purification of recombinant Thal for crystallization**

N-terminal His$_6$-tagged Thal was overexpressed in *E. coli* BL21(DE3) using a pETM-11 vector. The culture was grown in lysogeny broth medium with kanamycin (30 mg/liter$^{-1}$) at 37 °C until an $A_{600}$ of 0.6 was achieved. After that, the culture was induced with 0.1 mM isopropyl-$\beta$-D-thiogalactopyranoside and shaken for an additional 20 h at 25 °C. The cells were harvested by centrifugation (30 min, 4 °C, 5180 × g), resuspended in lysis buffer (50 mM sodium phosphate, pH 7.4, 300 mM NaCl) with additional DNase I and a Complete™ EDTA-free protease inhibitor tablet (Roche Applied Science), and disrupted using a SPCH high-pressure homogenizer (Stansted Fluid Power Ltd.). The lysate was clarified by centrifugation (30,966 × g, 30 min, 4 °C), and the supernatant was filtered through a 0.2-μm Whatman filter. The cleared lysate was loaded on HisTALON agarose affinity resin (bed volume of 1.5 ml of resin, flow rate, 0.5 ml/min$^{-1}$), followed by washing steps (10 ml; flow rate, 1 ml/min$^{-1}$) with 50 mM Na$_2$HPO$_4$, pH 7.4, 300 mM NaCl and with 50 mM Na$_2$HPO$_4$, pH 7.4, 300 mM NaCl, 10 mM imidazole. The fusion protein was eluted by the addition of 300 mM NaCl, 300 mM imidazole, 50 mM Na$_2$HPO$_4$, pH 7.4 (flow rate, 1 ml/min$^{-1}$). Fractions of 0.5 ml were collected, and protein concentration was determined by Nano-Drop UV spectroscopy.

**Crystallization and data collection**

Purified Thal was crystallized using the sitting-drop vapor-diffusion method at 20 °C with a drop ratio of 2:1 of protein solution (~15 mg/ml$^{-1}$) and reservoir solution (0.1 M Tris, pH 8.5, 1.5 M K$_2$HPO$_4$ /KH$_2$PO$_4$). To obtain an L-Trp-bound structure, Thal crystals were soaked in reservoir solution additionally containing 5 mM L-tryptophan. The Thal crystals for soaking experiments were obtained from a different crystallization condition (reservoir solution: 0.1 M Bicine, pH 9.0, 1.6 M K$_2$HPO$_4$ /Na$_2$HPO$_4$; drop ratio, 1:1). Thal-RebH5 was crystallized as the WT but using another reservoir solution (0.1 M Bicine, pH 8.4, 1.3 M K$_2$HPO$_4$ /KH$_2$PO$_4$). PrnF-Thal-RebH5 crystals were obtained by soaking Thal-RebH5 crystals in reservoir solution saturated with L-tryptophan for 75 min. Crystals appeared as hexagonal prisms within 7 days. For cryoprotection, the crystals were transferred to reservoir solution supplemented with 25% glycerol, before being flash-cooled and stored in liquid nitrogen. All data were collected at a temperature of 100 K at beamline P13 operated by EMBL Hamburg at the PETRA III storage ring at DESY, Hamburg, Germany (44).

**Data processing, structure determination, and refinement**

The data sets were processed with XDS and scaled with XSCELE (45). All crystals belonged to the space group P6$_3$ and showed no indication of twinning using the program Aimless (46) from the CCP4 package (47). A summary of the data processing is shown in Table S1. The structure of apo-Thal was determined by molecular replacement using the program
Regioselectivity switch of tryptophan 6-halogenase Thal

Phaser (48). The search model consisted of an ensemble of structures of the tryptophan halogenases RebH (PDB code 2OAM), PrnA (PDB code 2AQJ), PyrH (PDB code 2WEU), and SttH (PDB code 5HY5). Phaser placed two molecules per asymmetric unit with no clashes and rotation function z-scores of 21.0 and 9.3 and translation function z-scores of 42.2 and 84.8. The Trp-Thal structure was determined using the apo-Thal structure as a search model and was used as search model for the apo-Thal-RebH5 and Trp-Thal-RebH5 structures.

The first layout of the apo-Thal structure was generated with AutoBuild (49) from the PHENIX suite (50) and restrained refinement in Refmac5 (52) using noncrystallographic symmetry restraints or, in the case of Thal-RebH5, improved by phenix.refine using TLS refinement and noncrystallographic symmetry restraints. All structures were optimized until R values converged (Table S1). The occupancy of L-Trp in Trp-Thal refined to a value of 1. Figures of the final structures were generated using PyMOL. All r.m.s.d. values were calculated with the CCP4 program TOPP with the setting “superpose using secondary structure matching.”

Analytical reversed-phase HPLC

A Shimadzu CBM-20A chromatography system was employed for analytical measurements equipped with a diode array detector SPD-M20A IVDD (Shimadzu), pump Nexera XR LC LC-20ADXR (Shimadzu), column oven CTO-20A (Shimadzu), and an autosampler, Nexera XR SIL-20A (Shimadzu). For separation, a Hypersil Gold C18 column (3 μm, 150 × 2.1 mm, from Thermo Scientific, with a flow rate of 0.5 ml/min) was used. Absorbance was measured simultaneously at 220, 254, and 280 nm. Eluents were as follows: A, water/TFA (99.9:0.1); B, acetonitrile/TFA (99.9:0.1). Elution profile at 40°C column temperature was as follows: A%, 0–5.5 min, linear gradient from 95 to 5%; 5.5–6 min, 5%; 6–6.1 min, gradient from 5 to 95%; 6.1–9.0 min, 95%.

Preparative reversed-phase HPLC

A Merck-Hitachi LaChrom HPLC was used for purification consisting of interface D-7000, pump L-7150, and detector L-7420 equipped with a Hypersil Gold C18 8-μm column (250 × 2.1 mm, Thermo Scientific): flow rate, 10 ml/min; injection volume, 2.00 ml; detection at λ = 280 nm; solvents: A, water/acetonitrile/TFA (94.9:5:0.1); B, water/acetonitrile/TFA (5:94.9:0.1); gradient elution, A, 0%–5 min, 100%; 5–105 min, linear gradient from 100 to 0%; 105–110 min, gradient from 0 to 100%; 110–115 min, 100%.

HPLC MS (LC-MS)

Electrospray ionization (ESI) mass spectra were acquired employing an Agilent 6220 TOF mass spectrometer from Agilent Technologies (Santa Clara, CA) in extended dynamic range mode containing a Dual-ESI source, operating with a nitrogen generator, NGM 11. The mass axis was externally calibrated with ESI-L tuning mix (Agilent Technologies, Santa Clara, CA) as calibration standard. Samples were introduced from an HPLC outlet, and separation was performed with a 1200 HPLC system consisting of an autosampler, degasser, binary pump, column oven, and diode array detector (Agilent Technologies, Santa Clara, CA) using a C18 Hypersil Gold column (3 μm, 150 × 2.1 mm; Thermo Scientific). HPLC solvent A contained water/acetonitrile/formic acid (94.9:5:0.1); HPLC solvent B contained water/acetonitrile/formic acid (5:94.9:0.1). Gradient elution was as follows: A%, 0–10 min, gradient from 100 to 60%; 10–11 min, gradient from 60 to 2%; 11–11.5 min, gradient from 2 to 100%; 11.5–15 min, 100% with a flow rate of 0.3 ml/min and column temperature of 40°C.

NMR spectroscopy

NMR spectra were recorded on a Bruker Avance 500 (1H: 500 MHz; 13C: 126 MHz), Avance 500 HD (1H: 500 MHz; 13C: 126 MHz), or Avance 600 (1H: 600.13 MHz; 13C: 150.92 MHz) spectrometer. Chemical shifts (δ) were reported in ppm relative to tetramethylsilan (TMS) (Δδ, TMS = 0 ppm) and referenced to the residual nondeuterated solvent signal as internal standard (DMSO-d6: 1H: 2.50 ppm; 13C: 39.5 ppm). Coupling constants (J) were reported in Hertz (Hz) with 0.05 Hz resolution. Multiplicities were described as singlet (s), doublet (d), triplet (t), quartet (q), or multiplet (m). The assignment of 13C and 1H
NMR signals was supported by two-dimensional NMR techniques (COSY, ROESY, and HSQC).

**Acknowledgments**—The synchrotron MX data were collected at beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We thank Dr. Guillaume Pompidor, Dr. Isabel Hakanpää for assistance in using the beamline. We thank Christiane Widmann and Dominic Gilzer for carefully reading the manuscript.

**References**


Regioselectivity switch of tryptophan 6-halogenase Thal

42. Deleted in proof