Published in final edited form as:

J Mol Biol. 2009 August 7; 391(1): 74–85. doi:10.1016/j.jmb.2009.06.008.

Structural Insights into Regioselectivity in the Enzymatic Chlorination of Tryptophan

Xiaofeng Zhu^{1,†}, Walter De Laurentis^{1,†}, Khim Leang^{1,†}, Julia Herrmann², Katja Ihlefeld², Karl-Heinz van Pée², and James H. Naismith^{1,*}

¹ Center for Biomolecular Sciences, University of St. Andrews, St. Andrews KY16 9ST, UK

²Allgemeine Biochemie, TU Dresden, D-01062 Dresden, Germany

Abstract

The regioselectively controlled introduction of chlorine into organic molecules is an important biological and chemical process. This importance derives from the observation that many pharmaceutically active natural products contain a chlorine atom. Flavin-dependent halogenases are one of the principal enzyme families responsible for regioselective halogenation of natural products. Structural studies of two flavin-dependent tryptophan 7-halogenases (PrnA and RebH) have generated important insights into the chemical mechanism of halogenation by this enzyme family. These proteins comprise two modules: a flavin adenine dinucleotide (FAD)-binding module and a tryptophan-binding module. Although the 7-halogenase studies advance a hypothesis for regioselectivity, this has never been experimentally demonstrated. PyrH is a tryptophan 5-halogenase that catalyzes halogenation on tryptophan C5 position. We report the crystal structure of a tryptophan 5-halogenase (PyrH) bound to tryptophan and FAD. The FAD-binding module is essentially unchanged relative to PrnA (and RebH), and PyrH would appear to generate the same reactive species from Cl⁻, O₂, and 1,5-dihydroflavin adenine dinucleotide. We report additional mutagenesis data that extend our mechanistic understanding of this process, in particular highlighting a strap region that regulates FAD binding, and may allow communication between the two modules. PyrH has a significantly different tryptophan-binding module. The data show that PyrH binds tryptophan and presents the C5 atom to the reactive chlorinating species, shielding other potential reactive sites. We have mutated residues identified by structural analysis as recognizing the tryptophan in order to confirm their role. This work establishes the method by which flavin-dependent tryptophan halogenases regioselectively control chlorine addition to tryptophan. This method would seem to be general across the superfamily.

Keywords

tryptophan 5-halogenase; regioselectivity; mechanism; FADH₂; enzyme

Introduction

More than 4000 natural organohalogenes (many of which have biological activity), such as the important antibiotics vancomycin and chloramphenicol and the antitumor agent rebeccamycin, have been discovered since the 1960s. Nature apparently accomplishes regioselective halogenation with ease; however, for the chemist, this remains a very difficult

problem. Harnessing enzymes as part of a synthetic strategy is well established, but it is clear that utility has been greatly enhanced by a thorough knowledge of the molecular basis of activity. As cells themselves become the synthetic chemistry laboratories for high-value molecules, the required genetic engineering must be based on a deep molecular understanding. The financial and health care importance of compounds such as vancomycin argues for a need to understand their synthesis. From the viewpoint of fundamental chemistry, the inherent chemical challenge of regioselective halogenation is very interesting, and the study of biological chemistry can often inspire or reveal new routes for conventional chemical organic synthesis.

Metal-dependent haloperoxidases were once thought to be solely responsible for the incorporation of halogen atoms into biological compounds. However, the halogenating agent in these enzymes is the freely diffusible hypohalous acid,^{2,3} which possesses only very limited substrate specificity and regioselectivity. Two novel halogenases, which apparently demonstrated both substrate specificity and regioselectivity, were detected and characterized in a pyrrolnitrin-producing *Pseudomonas fluorescens* strain. PrnA was shown to catalyze the regioselective chlorination of tryptophan to 7-chlorotryptophan (Fig. 1), the first step in pyrrolnitrin biosynthesis. The second halogenase, monodechloroaminopyrrolnitrin 3-halogenase (PrnC), catalyzes regioselective chlorination in the 3-position of the pyrrole ring and produces aminopyrrolnitrin.^{4,5} Both enzymes were shown to require 1,5-dihydroflavin adenine dinucleotide (FADH₂) as cofactor to utilize Cl⁻ (but Br⁻ can also be used) and O₂. Their dependence on FADH₂ led to their being named as FADH₂ (or flavin)-dependent halogenases.⁶ FADH₂-dependent halogenase genes have been identified in biosynthetic gene clusters for vancomycin group antibiotics, the chemotherapeutic agent rebeccamycin, and other halometabolites.⁷ FADH₂-dependent halogenases are thought to be responsible for the regioselective halogenation of aromatic or hetero-aromatic rings found in biology. Recently, a new family of nonheme Fe^{II} α -ketoglutarate-dependent and O₂-dependent halogenases has been identified to halogenate nonactivated aliphatic groups.⁸⁻¹⁰

Considerable progress has been made in understanding the mechanism of FADH₂-dependent halogenases based on the structural and biochemical characterization of the tryptophan 7-halogenases PrnA and RebH.¹¹⁻¹⁴ FADH₂ binds to the halogenase, where it reacts with O₂ to produce flavin hydroperoxide.¹² This highly reactive intermediate either decomposes to give peroxide (the so-called uncoupling or nonproductive turnover) or reacts with a bound chloride ion to generate hypochlorous acid.¹² Hypochlorous acid is not free to diffuse out of the flavin active site, but is guided to the substrate (site of halogenation or second active site), which is located 10 Å away. Here conserved lysine and glutamic acid are positioned adjacent to the tryptophan. A study of RebH shows that Lys forms a chloroamine intermediate by reacting with HOCl,¹³ suggesting that this, rather than HOCl, may be the chlorinating species. Another recent study has shown that glutamic acid (E346) is also absolutely essential for chlorination of tryptophan.¹⁴ The broad outlines of the mechanism seem very likely to be conserved across the superfamily. Recently, the structural determination of halogenase CndH from myxobacterial chondrochloren was reported;¹⁵ this halogenase has a complete flavin adenine dinucleotide (FAD) module but binds a second carrier protein containing the substrate. It was proposed that the carrier protein may be required to complete the second active site.¹⁵ The entry of halogenase CndH splits FADH₂-dependent halogenases into two variants to accept free small molecules or carrier-bound substrates for the reactions. Interestingly, those that react with free substrates directly can be further subdivided into two distinct subfamilies represented by PrnA and PrnC, respectively. The PrnA group catalyzes the halogenation of tryptophan or indole derivatives (which are less reactive), while the other group catalyzes the halogenation of phenols and pyrrole derivatives (which are more reactive).⁷ It seems likely that there will be some differences in

the precise method of halogenation that will reflect inherent differences in substrate reactivity.

The original structural and mechanistic studies of PrnA and RebH hypothesized that the controlled spatial presentation of HOCl (or chloroamine) to the substrate would be the basis for regioselective halogenation.¹¹⁻¹⁴ However, this remains an untested hypothesis. *In vitro* halogenase activity has only been demonstrated for tryptophan halogenases (Fig. 1): 6,12,16-19 a PrnC from pyrrolnitrin biosynthesis,¹⁶ a pyrrolyl-*S*-carrier protein halogenating enzyme,²⁰ a β -tyrosyl halogenase,²¹ and a halogenase (HalB) acting on another phenylpyrrole derivative.²² This limited data set has hindered a detailed analysis. Sequence analysis of 5-tryptophan, 6-tryptophan, and 7-tryptophan halogenases indicates that they operate by a common mechanism, but sequence divergence has made it impossible to produce reliable models for substrate binding. No experimental evidence for the basis of regioselectivity has been reported. We report the crystal structures of a mutant enzyme and of tryptophan 5-halogenase (apoPyrH) bound to FAD, tryptophan, and FAD with tryptophan. We have performed site-directed mutagenesis to test our structural mode. Our work experimentally establishes the basis of regioselectivity in flavin-dependent halogenases.

Results

Overall structure of PyrH

PyrH shares a 40% sequence identity with PrnA (Fig. 2), but the similarity is mainly confined to the flavin domain. Molecular replacement using this domain and noncrystallographic averaging of subsequent phases were used to solve the first apoPyrH crystal structure. Two native enzyme crystal structures and one mutant E46Q structure of apoPyrH have been solved at 2.4 Å, 1.7 Å, and 2.5 Å resolution, respectively (Table 1). All structures contain four protein subunits in the asymmetric unit. Four protein chains in the 2.4-Å native structure contain both FAD and chloride ion. One of the four monomers (chain B), in addition, also contains tryptophan (Fig. 3a and b). The four chains in the 1.7-Å native structure are essentially identical; each binds tryptophan to form a binary complex. All four subunits of the E46Q structure bind FAD only. Thus, we have binary structures of apoPyrH with FAD and apoPyrH with tryptophan, as well as the ternary complex of apoPyrH with FAD and tryptophan. The N-terminal His-tag is not visible in the electron density in any structure and is presumably disordered. Also not visualized in any structure are the surface loops D113-T117 and S146-L152. In the absence of FAD (where only Trp is bound), the loop G37-G43 (which is found at the FAD-binding site) is disordered. In the absence of tryptophan, the loop A153-A158 is disordered. Other than these regions, the structure is fully fitted to experimental electron density maps. At a gross level, the presence of substrate(s) does not impact upon the protein structure. Therefore, structural description will focus on the ternary complex (chain B) of the 2.4-Å structure. There are differences in structure between the complexes at the two active sites (Fig. 4a and c), and these are discussed in detail below.

The secondary structural arrangement of the FAD-binding modules of PyrH is conserved with PrnA (and RebH). Despite the divergence of sequence, tryptophan modules are very similar in secondary structural arrangement (Fig. 2). Thus, PyrH, similar to PrnA,¹¹ is a single-domain protein in which several helices and two large β -sheets form a “box”; stuck to the side of the “box” is a triangular pyramid dominated by helices (Fig. 3b). Gel filtration shows that PyrH is a dimer in solution and that a dimer is observed in all structures (Fig. 3c), as seen for PrnA¹¹ and RebH.¹³ The dimer interface buries about 4000 Å² of solvent-accessible area and contains 42 hydrogen bonds and 2 salt bridges. Inside the “box” of each monomer is the FAD binding motif typified by glutathione reductase.²³ The adenosine

moiety of FAD contacts the C-termini of the parallel β -sheets, while the isoalloxazine ring sits above one face of the anti-parallel β -sheets and the N-termini of $\alpha 1$ and $\alpha 9$ surround the riboflavin moiety, in which T356-I358 of $\alpha 9$ provide a positive dipole to stabilize the negatively charged N1 and (C2)O2 during catalysis, as seen in *p*-hydroxybenzoate hydroxylase.²⁴ The chloride ion is located in a pocket centered on T356 on the *re*-face of the isoalloxazine ring. Tryptophan binds in the same general location as seen in PrnA,¹¹ on the edge of the anti-parallel β -sheet 9 Å away from the isoalloxazine ring. The signature motifs of the flavin-dependent halogenases GxGxxG (residues 9–14) and WxWxIP (residues 279–284) are located in the FAD module, although these motifs themselves do not interact with FAD directly. Apart from some small changes in peripheral loops, the FAD active site is essentially identical with that described first for PrnA.¹¹ Two key catalytic residues revealed in the studies of PrnA—K79 and E346—are conserved in PyrH as K75 and E354. Although the secondary structural elements of the tryptophan module are conserved, there are major differences in the connecting loops when comparing PyrH and PrnA¹¹ (RebH¹³). In PyrH, these changes map to the 5-residue deletion between E85 and P86, the 18-residue insertion from S142 to Q160, and two short deletions totaling 11 residues from S436 to M459 (PyrH sequence numbering) (Figs. 2 and 3d).

Flexibility at the FAD-binding module

FAD cofactor binds in an extended conformation in a solvent-exposed groove, as observed in PrnA¹¹ and RebH.¹³ A network of hydrogen bonds and water-mediated interactions stabilizes the complex between FAD and apoPyrH. These interactions are almost exclusively observed with main-chain amide or carbonyl groups and thus depend on structure rather than sequence. This rationalizes the observation that the FAD binding fold found in halogenases, although very divergent in sequence, remains the same as the exemplar glutathione reductase.²³ The position of the isoalloxazine ring is secured by two hydrogen bonds (isoalloxazine N3 and O2) with carbonyl of A47 and amide of I358 and is constrained by bulky residues, namely, W281, I283 V44, A47, I325, M327, F349, P352, and I358 (Fig. 4b). Interestingly, neither tryptophan residue of the characteristic motif is essential, as both can even be substituted simultaneously in PrnA with other bulky residues with retention of activity, and the radical mutant W272A of PrnA (W279 in PyrH) is as active as wild-type PrnA while W274A mutant in PrnA (W281 in PyrH) is inactive.¹⁴ In apoPyrH monomers with FAD bound, the region between R41 and E46 forms a strap that sits across the isoalloxazine ring; notably, G45 sits above the *si*-face of the central ring. The side chain of the conserved residue E46 points away from the isoalloxazine ring, instead making salt contacts with R228 and R257. The conformation of this strap varies significantly between PyrH, PrnA (Fig. 4b), and RebH.^{13,25} In RebH, the strap adopts a “relaxed” conformation, and FAD is solvent exposed.^{13,25} Structural variation suggests the inherent structural flexibility of this region.

Although the structure of the surrounding protein and the orientation of FAD are the same in each monomer where it is bound, in the tryptophan complex, the bound FAD and the “strap” residues appear less well ordered in electron density maps. This can be seen in two ways: there is some residual negative $F_o - F_c$ density around the flavin (see Supporting Information, Fig. S4), and the average *B*-factor of the strap is nearly 1.4 times the average for the monomer compared to only 1.1 times the monomer average where only FAD is bound. Similar disturbances to the flavin site were observed in RebH^{13,25} and PrnA¹¹ structures when tryptophan was bound.

In monomers that do not have bound FAD, the side chain of E46 flips in position by almost 180° and points towards the vacant isoalloxazine binding position. As a consequence, G45 has moved by more than 5 Å, and residues G37-G43 have become disordered. We made the mutants E46Q and E46D to probe the role of this conserved residue. Both mutants exhibit a

decreased k_{cat} value against wild-type PyrH (3.6–0.06 min⁻¹). We were able to obtain a structure for E46Q mutant bound to FAD that shows that the mutated side chain adopts an orientation similar to those in monomers without bound FAD (Fig. 4b).

The tryptophan-binding site

The secondary structure elements $\beta 3$, $\beta 5$, and $\beta 6$; the sixth 3_{10} -helix; and $\alpha 13$ in apoPyrH form the tryptophan-binding pocket. As a result of insertions and deletions, the binding pocket for tryptophan has been substantially rearranged when compared to PrnA (Fig. 3d). In PyrH, insertion means that an extra loop of structure (F144-P165) sits adjacent to the tryptophan. The loop from D440 to F451 in PyrH is shorter due to deletion rather than the corresponding loop (L433-W455) in PrnA. Crudely, insertion and deletion can be thought to cancel each other out, occupying approximately the same volume of structural space but with a different three-dimensional structure. Consequently, the tryptophan is bound differently in PyrH to PrnA11 (RebH13) (Fig. 5a and b). The tryptophan indole rings in both PyrH and PrnA are essentially coplanar, but flipped relative to each other such that the six-membered rings—but not the five-membered rings nor the main-chain atoms—superimpose (Fig. 5a and b). In PyrH, the benzene portion of the indole ring is sandwiched between the π systems of H92 and F94. These residues and π -stacking interaction are conserved in PrnA and RebH. The π -stacking interaction seen for the five-membered ring of the indole in PrnA11 (RebH13) is missing in PyrH. The indole nitrogen is hydrogen bonded to the hydroxyl of Y454 and the carbonyl of P93. The amine of the tryptophan makes hydrogen bonds to two water molecules and the carbonyl of F451. The two carboxylate oxygen atoms make a bifurcated hydrogen bond to the side chain of S50, and each makes a second separate hydrogen bond to the side chain of Q163 and the amide nitrogen of S50, respectively (Supporting Information, Fig. S1). The orientations of tryptophan seen in PyrH and PrnA are mutually exclusive. In PyrH, F49, F94, and F451 would clash with the tryptophan orientation seen in PrnA. While in PrnA, Y444 and E450 would clash with a PyrH tryptophan orientation.

Structural analysis reveals a number of quite significant changes upon tryptophan binding in apoPyrH (Fig. 4a and c). The loop L152-A158 becomes ordered upon tryptophan binding, and there are large changes in the main-chain position between E159 and A167. The main chains of F164 and P165 have rotated by almost 180°. The side chain of Q163 moves over by 7 Å to make contact with the tryptophan carboxylate. A large shift occurs in the main chain of T48-V52, with S50 moving over 2.5 Å and with F49 adopting a different rotamer in order to create space for tryptophan to bind. To probe the recognition, we constructed the single mutants F49A and Y454F, as well as the double mutant. The two single mutants show a reduction in activity compared to wild-type PyrH. The double mutant F49A/Y454F, based on its near-UV (260–320 nm) circular dichroism spectrum (Supporting Information, Fig. S2), is unfolded.

Discussion

Mechanistic implications

During the catalytic cycle, halogenase consumes two electrons from FADH₂, creating FAD. The next cycle requires FADH₂ to displace FAD from the binding site, which presumably needs to be flexible enough to allow the exchange of such a large molecule. Spectroscopic studies¹² suggest additional conformation states of the FADH₂ protein complex. Our structural data suggest that the point of this flexibility is the *si*-face “strap” of residues that holds flavin to the protein. Structural data show that it can exist in multiple conformations that interact differently with the isoalloxazine ring. Although the conserved E46 residue within the strap has no apparent role in catalysis, it undergoes a large conformational change

upon FAD binding. In both conformations, it makes important salt links or hydrogen bonds. Even a conservative mutation of the residue has substantial impacts on both structure and catalysis. We suggest that E46 is important in regulating the exchange of FAD and FADH₂.

The presence or the absence of tryptophan does not significantly perturb the rate of FADH₂ oxidation.^{12,13} Therefore, halogenases have a tendency to uncouple (i.e., the majority of FADH₂ is “wasted” on making hydrogen peroxide rather than the product). Despite the uncoupling, mutation at K79 of RebH does slow the oxidation of FADH₂ even though the two active sites are almost 10 Å apart.¹³ Thus, there must indeed be some communication between the active sites. Our analysis establishes a structural basis for this. The marked conformation changes we see on the *si*-face strap, particularly the movement of the absolutely conserved G45 and E46 that interact with the isoalloxazine ring, are mirrored by profound changes in T48-V52 at the tryptophan-binding site. We propose that this short stretch of amino acids (G45-V52) is the conduit for information transfer between the two active sites.

Origin of regioselectivity

The key feature of flavin-dependent halogenases is their regioselectivity. The tryptophan-binding module of PyrH binds its substrate in such a way as to expose the C5 position to the key lysine and glutamate residues at the site of chlorination. The C5 of tryptophan in the PyrH complex occupies the same position (relative to critical active-site residues) as the C7 of tryptophan in the PrnA complex within 0.4 Å (Fig. 5b). The other sites on tryptophan are all protected from the chlorinating agent by the structure. The use of bulky protecting groups in organic chemistry essentially mimics the strategy used by halogenases. The PyrH structure shows that despite a quite significant sequence variation, the core structure of the tryptophan-binding domain is conserved. The changes in structure are confined to loops that have insertions and deletions, which have presumably evolved to control regio-selectivity within a conserved structural framework. We have identified and mutated two of the key residues that control recognition, and the reduction activity of these mutants (Table 2) supports our structural model. The structure of PyrH provides a reliable template for modeling and rationalizing the regioselectivity of the tryptophan 6-halogenases.

Conclusions

A picture of halogenase with two chemically independent modules emerges. The halogenase from myxobacterial chondrochloren takes this to the extreme in that the second module appears to be a separate protein.¹⁵ The production of the active chlorinating species occurs within a relatively well-conserved FAD-binding module. The requirement for the chlorinating agent to be sequestered from solvent means that the substrate has to be bound close to the site of production of the chlorinating agent. However, in order to control the regiochemistry of reaction, the substrate itself not only has to be recognized but also has to be bound in such a way that protects the other sites of potential reactivity while exposing the desired one. Separating the two modules, as is seen in flavin-dependent halogenases, allows the substrate-binding module to evolve to recognize a variety of different substrates without compromising the production of the active chlorinating agent. Sequence analysis suggests that the chemical reactivity of the substrate may manifest itself in sequence differences at the site of halogenation, and this is hinted at by the myxobacterial chondrochloren enzyme.¹⁵ Future studies will be required to address this.

Materials and Methods

Expression of prnH and purification of recombinant apoPyrH and its mutants

His-tagged apoPyrH was expressed in *P. fluorescens* BL915 using the 22.7-kb plasmid pSZexM2,18 which contains the *Streptomyces rugosporus* LL-42D005 tryptophan 5-halogenase (PyrH) coding sequence inside the *Pseudomonas* expression vector pCIB-His. 6,11,26 In order to cleave the His-tag, we inserted a tobacco etch virus (TEV) cleavage site before the PyrH encoding sequence. The encoding sequence was cloned using the following PCR primers containing the TEV cleavage site (marked by a box):

5' ATTCGTAGCGGATCCAGAAAACCTGTATTTTCAGGGCATGATCCGATCTGTGGTGATCGTGGGTGGTGGC-3'

5'-GTATGCTACCAAGCTTCTATCATTGGATGCTGGCGAGGTACT-3'.

The resultant PCR fragment was cloned back to pCIB-His vector using BamHI and HindIII (cleavage sites are underlined). The pCIBHisTEV-PyrH construct was expressed in *P. fluorescens* BL915Δ1–4, as previously described.6,11,26

For site-directed mutagenesis, the above PCR fragment was cloned into the pFastBac-HTa vector (Invitrogen Ltd., UK) as template using BamHI and HindIII. After the mutation had been introduced using the Quick-Change site-directed mutagenesis kit (Stratagene), the insert was then transferred back into the pCIB-His vector for expression. Use of the following primers produced the mutations:

E46D: 5'-dGATCGGGGTCGGCGATGCGACCTTCAG-3' and 5'-dCTGAAGGTCGCATCGCCGACCCCGATC-3'

E46Q: 5'-dGATCGGGGTCGGCCAGGCGACCTTCAGCA-3' and 5'-dTGCTGAAGGTCGCCGGCCGACCCCGATC-3'.

The F49A and Y454F single mutants and the F49A/Y454F double mutant were constructed by overlap extension PCR using the following primers:

F49A: 5'-GTCGGCGAAGCGACCGCCAGCACGGTCCGCCAC-3' and 5'-GTGGCGGACCGTGCTGGCGGTGCTTCGCCGAC-3'

Y454F: 5'-CACGGCTTCGAGACCTTTTCGTGGATCACCATG-3' and 5'-CATGGTGATCCACGAAAAGGTCTCGAAGCCGTG-3'.

The template for the single mutants was pSZexM2pyrH, and that for the double mutant was pBSKpyrHY454F. The DNA fragments were sequenced by MWG sequencing services (Eurofins MWG Operon Co.) and cloned into pCIB-His. After the transformation of *P. fluorescens* BL915, ΔORF1–4 all genes were overexpressed. Proteins were purified using Ni-chelating chromatography. The assay mixture used routinely for the determination of halogenating activity consisted of flavin reductase²⁷ (10 mU), NaCl (25 mM), catalase (520 U), FAD (5 μM), L-tryptophan (0.25 mM), halogenase, and potassium phosphate buffer (10 mM, pH 7.2) in a total volume of 200 μL. The reaction was started by adding NADH (5 mM), and the reaction mixture was incubated for 45 min at 30 °C. The reaction was stopped by incubation at 96 °C for 5 min, and product formation was analyzed by HPLC.⁶ For the determination of K_m values, the tryptophan concentration was varied between 10 μM and 60 μM; concentrations above 60 μM led to substrate inhibition.

For crystallization, the enzyme was purified using a 5-mL HisTrap column, a 5-mL HiTrap Q-FF column, and a HiLoad 16/60 Superdex200 column (GE Healthcare Life Sciences, UK). Despite several attempts, we were unable to cleave the His-tag and used the intact

construct in crystallization experiments. The final protein sample appeared slightly yellow due to partial FAD binding and was suspended in 20 mM Tris–HCl (pH 7.2) prior to use. E46D and E46Q were verified to fold properly by near-UV and far-UV circular dichroism spectroscopies.²⁸ The double mutant F49A/Y454F was shown to be incorrectly folded.

Structural biology

Purified apoPyrH was concentrated to 10 mg/mL using Centricon-10 centrifugal concentrators (Amicon). Crystallization trials were set up using the sitting-drop vapor-diffusion method at 20 °C. One crystallization condition was identified from commercial sparse matrix screens. Optimization of crystallization conditions and subsequent preliminary X-ray analysis identified diffraction-quality crystals. The best crystals were obtained by equilibrating a mixture of 2 μ L of protein solution (supplemented with 1 mM FAD and 5 mM tryptophan) and 2 μ L of precipitant solutions [0.01 M MgCl₂, 0.05 M Na-cacodylate (pH 5.8), and 1.5 M Li₂SO₄]. Diamond-shaped crystals formed within a few hours and grew to a maximum size of 0.2 mm $\times$ 0.2 mm $\times$ 0.4 mm within 3 weeks. A cryoprotectant solution was made by supplementing the crystallization precipitant solution with 20% glycerol, 1 mM FAD, and 5 mM L-tryptophan. Crystals were mounted in Hampton Research cryoloops and flash-frozen at 100 K. Data were collected from a single crystal in a single pass using 0.2° (Daresbury SRS station 10.1) or 0.3° (ESRF ID14-2) oscillations at 2.4 Å and 1.7 Å resolution, respectively. Given the differing results from the two data collections, we reexamined the crystallization conditions and identified no specific difference that resulted in different space groups, resolutions, and structures. The crystals came from different batches of proteins and were collected over 18 months apart. A 2.5-Å-resolution data set of the mutant E46Q was collected at Diamond Light Source beamline I02 at 0.4° oscillation, which was crystallized under the same conditions as described above.

Data were indexed and integrated using MOSFLM²⁹ and scaled using Scala.^{30,31} POINTLESS helped determine the space group as $P4_32_12$ for the two lower-resolution data sets.³² The 1.7-Å-resolution data were indexed and scaled to $P2_12_12_1$ by XDS.³³ PHENIX Xtriage³⁴ showed that it was a twin; gave the twinning operator as $k, h, -l$; and estimated the twin fraction of 0.25. Detwinned data were obtained using DETWIN.³¹ The data processing statistics are summarized in Table 1. The 2.4-Å-resolution data were solved using Phaser,³⁵ with the monomer of PrnA structure (after removal of water and ligands) as search model [Protein Data Bank (PDB) entry 2AQJ]. We initially searched for up to six monomers based on reasonable Matthews numbers.³⁶ An unambiguous solution was found for four monomers (Matthews number of 3.23; 62% solvent content), which, after rigid-body refinement, gave an R -factor of 38.9%. Further rebuilding and refinement were carried out using Coot³⁷ and Refmac5³⁸ until R_{free} failed to decrease. Chain D in this structure shows obvious disorder, and tight NCS restraints were applied during the refinement. Chain B binds FAD and tryptophan, while the other three chains only bind FAD. The mutant E46Q structure was solved using rigid-body refinement starting with the 2.4-Å apoPyrH structure. The four monomers in the E46Q structure are essentially identical as the binary complex with the bound FAD. The 1.7-Å structure was solved using the 2.4-Å monomer from chain B. Two dimers were identified with only tryptophan bound. We investigated the use of anisotropic thermal factor refinement for the 1.7-Å-resolution data set. The improvement in R_{free} was less than 0.5%, and so isotropic thermal factors were used. The data processing and refinement statistics are summarized in Table 1. Unbiased $F_o - F_c$ and $2F_o - F_c$ electron density maps for all ligands are shown in Supporting Information (Figs. S3 and S4). All structures were validated by MolProbity.³⁹ All pictures were produced using PyMOL (DeLano Scientific, San Carlos, CA), unless otherwise indicated.

Accession numbers

Coordinates and structure factors have been deposited in the PDB with accession numbers 2WES, 2WET, and 2WEU.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Biotechnology and Biological Science Research Council (J.H.N.) and Deutsche Forschungsgemeinschaft (K.-H.v.P.).

Abbreviations used

FAD	flavin adenine dinucleotide
PrnC	monodechloroaminopyrrolnitrin 3-halogenase
FADH₂	1,5-dihydroflavin adenine dinucleotide
TEV	tobacco etch virus
PDB	Protein Data Bank.

References

1. van Pée K-H, Patallo EP. Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl. Microbiol. Biotechnol.* 2006; 70:631–641. [PubMed: 16544142]
2. Messerschmidt A, Prade L, Wever R. Implications for the catalytic mechanism of the vanadium-containing enzyme chloroperoxidase from the fungus *Curvularia inaequalis* by X-ray structures of the native and peroxide form. *Biol. Chem.* 1997; 378:309–315. [PubMed: 9165086]
3. Sundaramoorthy M, Turner J, Poulos TL. Stereochemistry of the chloroperoxidase active site: crystallographic and molecular-modeling studies. *Chem. Biol.* 1998; 5:461–473. [PubMed: 9751642]
4. Hammer PE, Hill DS, Lam ST, van Pée K-H, Ligon JM. Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Appl. Environ. Microbiol.* 1997; 63:2147–2154. [PubMed: 9172332]
5. Kirner S, Hammer PE, Hill DS, Altmann A, Fischer I, Weislo LJ, et al. Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. *J. Bacteriol.* 1998; 180:1939–1943. [PubMed: 9537395]
6. Keller S, Wage T, Hohaus K, Hölzer M, Eichhorn E, van Pée K-H. Purification and partial characterization of tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew. Chem. Int. Ed.* 2000; 39:2300–2302.
7. van Pée K-H. Microbial biosynthesis of halometabolites. *Arch. Microbiol.* 2001; 175:250–258. [PubMed: 11382220]
8. Vaillancourt FH, Yeh E, Vosburg DA, O'Connor SE, Walsh CT. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature.* 2005; 436:1191–1194. [PubMed: 16121186]
9. Vaillancourt FH, Yin J, Walsh CT. SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O₂-dependent halogenase. *Proc. Natl Acad. Sci. USA.* 2005; 102:10111–10116. [PubMed: 16002467]
10. Blasiak LC, Vaillancourt FH, Walsh CT, Drennan CL. Crystal structure of the nonhaem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature.* 2006; 440:368–371. [PubMed: 16541079]

11. Dong C, Flecks S, Unversucht S, Haupt C, van Pée K-H, Naismith JH. Tryptophan 7-halogenase (PnA) structure suggests a mechanism for regioselective chlorination. *Science*. 2005; 309:2216–2219. [PubMed: 16195462]
12. Yeh E, Cole LJ, Barr EW, Bollinger JM Jr, Ballou DP, Walsh CT. Flavin redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. *Biochemistry*. 2006; 45:7904–7912. [PubMed: 16784243]
13. Yeh E, Blasiak LC, Koglin A, Drennan CL, Walsh CT. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry*. 2007; 46:1284–1292. [PubMed: 17260957]
14. Flecks S, Patallo EP, Zhu X, Ernyei AJ, Seifert G, Schneider A, et al. New insights into the mechanism of enzymatic chlorination of tryptophan. *Angew. Chem. Int. Ed.* 2008; 47:9533–9536.
15. Buedenbender S, Rachid S, Müller R, Schulz GE. Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases. *J. Mol. Biol.* 2009; 385:520–530. [PubMed: 19000696]
16. Hohaus KAA, Burd W, Fischer I, Hammer PE, Hill DS, Ligon JM, van Pée K-H. NADH-dependent halogenases are more likely to be involved in halometabolite biosynthesis than haloperoxidases. *Angew. Chem. Int. Ed. Engl.* 1997; 36:2012–2014.
17. Yeh E, Garneau S, Walsh CT. Robust *in vitro* activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl Acad. Sci. USA*. 2005; 102:3960–3965. [PubMed: 15743914]
18. Zehner S, Kotsch A, Bister B, Sussmuth RD, Mendez C, Salas JA, van Pée K-H. A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* 2005; 12:445–452. [PubMed: 15850981]
19. Seibold C, Schnerr H, Rumpf J, Kunzendorf A, Hatscher C, Wage T, et al. A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocatal. Biotransform.* 2006; 24:401–408.
20. Dorrestein PC, Yeh E, Garneau-Tsodikova S, Kelleher NL, Walsh CT. Dichlorination of a pyrrolyl-*S*-carrier protein by FADH₂-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc. Natl Acad. Sci. USA*. 2005; 102:13843–13848. [PubMed: 16162666]
21. Lin S, Van Lanen SG, Shen B. Regiospecific chlorination of (*S*)-beta-tyrosyl-*S*-carrier protein catalyzed by SgcC3 in the biosynthesis of the enediyne antitumor antibiotic C-1027. *J. Am. Chem. Soc.* 2007; 129:12432–12438. [PubMed: 17887753]
22. Wynands I, van Pée K-H. A novel halogenase gene from the pentachloropseudilin producer *Actinoplanes* sp. ATCC 33002 and detection of *in vitro* halogenase activity. *FEMS Microbiol. Lett.* 2004; 237:363–367. [PubMed: 15321684]
23. Schulz GE, Schirmer RH, Pai EF. FAD-binding site of glutathione reductase. *J. Mol. Biol.* 1982; 160:287–308. [PubMed: 7175934]
24. Wierenga RK, Drenth J, Schulz GE. Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of *p*-hydroxybenzoate hydroxylase with the FAD- as well as NADPH-binding domains of glutathione reductase. *J. Mol. Biol.* 1983; 167:725–739. [PubMed: 6876163]
25. Bitto E, Huang Y, Bingman CA, Singh S, Thorson JS, Phillips GN Jr. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins*. 2008; 70:289–293. [PubMed: 17876823]
26. Dong C, Kotsch A, Dorward M, van Pée K-H, Naismith JH. Crystallization and X-ray diffraction of a halogenating enzyme, tryptophan 7-halogenase, from *Pseudomonas fluorescens*. *Acta Crystallogr. Sect. D*. 2004; 60:1438–1440. [PubMed: 15272170]
27. Fontecave M, Eliasson R, Reichard P. NAD (P)H:flavin oxidoreductase of *Escherichia coli*. A ferric iron reductase participating in the generation of the free radical of ribonucleotide reductase. *J. Biol. Chem.* 1987; 262:12325–12331. [PubMed: 3305505]
28. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochim. Biophys. Acta*. 2005; 1751:119–139. [PubMed: 16027053]
29. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. *Jt. CCP4 ESF-EAMCB Newsl. Protein Crystallogr.* 1992; 26:1–10.
30. Evans PR. Scala. *Jt. CCP4 ESF-EACBM Newsl.* 1997; 33:22–24.

31. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D.* 1994; 50:760–763. [PubMed: 15299374]
32. Evans P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D.* 2006; 62:72–82. [PubMed: 16369096]
33. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 1993; 26:795–800.
34. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. Sect. D.* 2002; 58:1948–1954. [PubMed: 12393927]
35. McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. Sect. D.* 2007; 63:32–41. [PubMed: 17164524]
36. Matthews BW. Solvent content of protein crystals. *J. Mol. Biol.* 1968; 33:491–497. [PubMed: 5700707]
37. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D.* 2004; 60:2126–2132. [PubMed: 15572765]
38. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D.* 1997; 53:240–255. [PubMed: 15299926]
39. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 2007; 35:W375–W383. [PubMed: 17452350]
40. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* 2003; 31:3320–3323. [PubMed: 12824317]

Fig. 1.
The reactions catalyzed by tryptophan 7-halogenases, 5-halogenases, and 6-halogenases.

Fig. 2.

Sequence alignment of PyrH (AAU95674.1, tryptophan 5-halogenase, *S. rugosporus*), PrnA (AAB97504, tryptophan 7-halogenase, *P. fluorescens*), and Thal (ABK79936.1, tryptophan 6-halogenase, *Streptomyces albobrisesolus*). Identical residues are boxed and shaded in red. Secondary structure elements were assigned according to the available crystal structures of PyrH and PrnA. The two signature motifs of tryptophan halogenases are marked in bold boxes. An asterisk marks the two key residues that have been identified as being absolutely required for PrnA activity. The large insertions and deletions of PyrH are shown boxed in yellow. The diagram was produced using ESPrnt.40

Fig. 3.
 (a) Unbiased $F_o - F_c$ electron density map at 3σ for the ligands refined in the structures: Trp, FAD + Cl⁻, and FAD + Cl⁻ + Trp (from left to right). The ligands are shown in stick

presentation and colored according to atom type, with carbon in yellow, oxygen in red, nitrogen in blue, and chlorine in green. (b) Stereo cartoon diagram of chain B of PyrH, with FAD, Cl⁻, and tryptophan bound. The main secondary structure elements are labeled. The bound ligands and ion are shown in spheres and colored according to atom type, with carbon in yellow, oxygen in red, nitrogen in blue, and chlorine in green. (c) Dimer of PyrH comprising chains B and C. Chain B is colored as in Fig. 2, while chain C is shown in cyan. (d) Stereo diagram of C^α traces of PyrH (chain B) and PrnA ternary complex with FAD and tryptophan. Chain B of PyrH is colored as in (b), while PrnA is colored according to atom type, with carbon in gray, oxygen in red, and nitrogen in blue. The rearranged loops of the tryptophan-binding module are highlighted in darker colors.

Fig. 4.

(a) Comparison of C α traces of PyrH binary complexes with tryptophan (yellow) or FAD (cyan). The loops showing significant changes are highlighted in darker colors. The bound ligands are shown in spheres and colored as in (a). Q163, F164, and P165 are labeled by red asterisks to simplify the figure. The short stretch of amino acids G45-V52 connects both active sites and undergoes large conformational changes, which depend on the presence or the absence of FAD or tryptophan. (b) Stereo diagram of the "strap" region of PyrH covering the bound FAD along with the superimposed E46Q mutant and PrnA. PyrH is colored as in Fig. 3, carbons of PrnA are shown in gray, and carbons of E46Q are shown in green. The residues of PyrH are labeled. (c) The tryptophan-binding site of PyrH undergoes large conformational changes in order to accommodate tryptophan. Carbon atoms belonging to the tryptophan complex are shown in yellow, while those belonging to the apo structure are shown in cyan.

Fig. 5.

(a) Stereo diagram comparing the tryptophan-binding sites of PyrH and PrnA. The residues in PyrH are labeled in boldface, with carbons in yellow, while the residues in PrnA are labeled in italic, with carbons in gray. Tryptophan is highlighted in a darker shade of yellow for PyrH and in a darker shade of gray for PrnA. The halogenated atom is marked by a red asterisk. The hydrogen bonds of Trp are shown in broken lines. (b) Stereo diagram showing the spatial conservation of the C5 and C7 atoms of tryptophan relative to the active-site residues. The color scheme is the same as in (a).

Table 1

Crystallographic statistics

	ApoPyrH+FAD (+Trp)	ApoPyrH+Trp	ApoPyrHE46Q+FAD
<i>Data collection</i>			
Space group	$P4_32_12$	$P2_12_12_1$	$P4_32_12$
Cell dimensions			
a, b, c (Å)	137.5, 137.5, 307.8	142.7, 141.1, 122.9	137.6, 137.6, 309.1
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
R_{merge}	0.094 (0.340)	0.078 (0.221)	0.069 (0.352)
$I/\sigma I$	16.9 (3.7)	12.6 (4.3)	43.6 (6)
Completeness (%)	98.3 (92.4)	98.6	100 (99.9)
Redundancy	7.0 (4.6)	4.7 (4.8)	9.6 (9.4)
<i>Refinement</i>			
Resolution (Å)	48.1–2.4 (2.5–2.4)	100.5–1.7 (1.8–1.7)	48.2–2.50 (2.6–2.5)
Number of reflections	108,080	228,784	98,115
$R_{\text{work}}/R_{\text{free}}$	0.197/0.235	0.216/0.250	0.188/0.233
Number of atoms			
Protein	16,017	16,007	16,160
Ligand/ion	255	60	216
Water	747	1456	968
<i>B-factors</i>			
Protein	44.6	19.97	39.0
Ligand/ion	47.6	15.35	35.4
Water	41.6	31.3	40.5
r.m.s.d.			
Bond lengths (Å)	0.013	0.013	0.012
Bond angles (°)	1.421	1.303	1.317
PDB code	2WET	2WEU	2WES

Table 2

Steady-state kinetic parameters for PyrH and mutants

Sample	k_{cat} (min^{-1})	K_{m} (μM)	$k_{\text{cat}}/K_{\text{m}}$ ($\text{mM}^{-1} \text{min}^{-1}$)
PyrH	3.56 ± 1.1	109 ± 44	32.54
F49A	0.24 ± 0.05	146 ± 5	1.64
Y454F	0.71 ± 0.39	235 ± 77	3.01
E46D*	0.06 ± 0.005	104 ± 13	0.58
E46Q*	0.06 ± 0.002	26 ± 2	2.31

The steady-state kinetic parameters for PyrH and mutants were obtained from Hanes plots, except those of mutants E46D* and E46Q*, which were obtained by least-squares fitting of the initial velocity against substrate concentrations according to the Michaelis–Menten equation.