The structure of tryptophan 7-halogenase (PrnA) suggests a mechanism for regioselective chlorination

Changjiang Dong[1], **Silvana Flecks**[2], **Susanne Unversucht**[2], **Caroline Haupt**[2], **Karl-Heinz van Pée**[2], and **James H Naismith**[1],*

[1]Centre for Biomolecular Sciences, EaStchem, The University, St Andrews, UK KY16 9ST

[2]Institut für Biochemie, TU Dresden, D-01062 Dresden, Germany

Abstract

Chlorinated natural products include vancomycin and cryptophycin A. Their biosyntheses involves regioselective chlorination by flavin-dependent halogenases. We report the structural characterization of tryptophan 7-halogenase (PrnA), which regioselectively chlorinates tryptophan. Tryptophan and FAD are separated by a 10Å-long tunnel and bound by distinct enzyme modules. The FAD module is conserved in halogenases and is related to flavin-dependent monooxygenases. Based on biochemical studies, crystal structures and by analogy with monooxygenases, we predict $FADH_2$ reacts with O_2 making peroxy-flavin which is decomposed by Cl^-. The resulting HOCl is guided through the tunnel, to tryptophan, where it is activated to participate in electrophilic aromatic substitution.

In addition to man-made chemicals, there are nearly four thousand chlorinated and brominated natural products (1), including drugs such as vancomycin (2), rebeccamycin (3) and cryptophycin A (4). The *de-novo* chemical synthesis of complex natural products is often too expensive or too difficult to be practical. Their production relies on fermentation and introducing diversity in such molecules requires protein engineering. This has been hampered by a lack of understanding of the molecular basis of the biological regioselective halogenation mechanism.

Metal-dependent haloperoxidases were once thought to catalyze all halogenation reactions in biology and fall into two classes. Both, heme-iron-dependent enzymes (5) and vanadium-dependent enzymes (6), have been structurally characterized. Although different in structure, both form a metal-bound hydrogen peroxide, which reacts with halide ions producing a metal-bound hypohalite ion. This ion dissociates from the metal as hypohalous acid (5,6) where, in solution, it reacts with substrate. Such halogenation lacks regioselectivity and substrate specificity (7). Perhydrolases are now recognized not to be involved in halometabolite biosynthesis (8). A new halogenase was reported by Dairi *et al.* identifying the gene for the chlorinating enzyme in chlorotetracycline biosynthesis (9). The gene product showed no similarity to haloperoxidases. Studies of the antifungal compound pyrrolnitrin from *Pseudomonas fluorescens* identified two related genes (*prnA* and *prnC*), coding for two halogenating enzymes (10) (fig. S1). Both contain a flavin binding site (9-13) and exhibit weak sequence homology to flavin-dependent monooxygenase enzymes (14). PrnA catalyzes the regioselective chlorination of the 7-position of tryptophan (15). Turnover requires that FAD is reduced to $FADH_2$ (by flavin reductase) and that O_2 is present (11).

---

*To whom correspondence should be addressed. .

<u>One sentence</u>: The crystal structure of tryptophan 7-halogenase indicates that the halogenation of important metabolites in bacteria proceeds by the novel use of hypohalous acid to achieve regioselective substitution.

Members of this halogenase superfamily have been identified in the biosynthetic pathways of the antibiotics balhimycin (16), vancomycin (2), the antitumor agent rebeccamycin (3) and other halometabolites (7, 17-21). It is likely that regiospecific halogenation reactions carried out by bacteria are predominately catalyzed by flavin-dependent halogenases. To account for the regioselectivity, two mechanisms have been proposed. In one, the substrate is oxidized to an epoxide, which is decomposed by a nucleophilic attack of Cl^- (11). In the other, direct chlorination by a high energy flavin hypochlorite intermediate occurs (22,23). The use of HOCl as the halogenating agent was explicitly discounted on the grounds of reactivity and regioselectivity (11,23).

PrnA was purified from a *P. fluorescens* expression system and its structure determined to 1.95Å. The structure consists of residues 2 to 518 (Fig. 1A) and is a dimer (fig. S2A). Each monomer is a single domain, shaped like a "box" with a triangular "pyramid" stuck to one face (Fig. 1A). The "box" which we identify as the FAD binding module, is dominated by two large β sheets (Fig 1A). FAD is bound in a solvent-exposed groove adjacent to the large parallel β sheet (Fig. 1A). The C4-N5 edge of the isoalloxazine ring sits above one face of the mainly antiparallel β sheet (Fig. 1, A and B). Sequence alignment shows that only the flavin binding module is conserved in flavin-dependent halogenases (fig. S3). Conserved residues map to this module and some of these residues are conserved in the monooxygenase enzymes (fig. S3). The flavin binding module of PrnA is structurally similar to *p*-hydroxybenzoate hydroxylase (PHBH) (24), a monooxygenase (fig. S4).

Cl^- is bound in a pocket on one face of the isoalloxazine ring (Fig. 1B) and makes contacts with the amide nitrogen atoms of T348 and G349. No other Cl^- is experimentally located despite the presence of 50 mM NaCl, and we identify this as the Cl^- binding site. The amide backbone is known to bind negative ions and is proposed to bind F^- ion in the fluorinase structure (25). Complexes with tryptophan and 7-chlorotryptophan were obtained by incubating the protein, prior to crystallization, with each. Tryptophan (Fig. 1C) and 7-chlorotryptophan (fig. S5B) are bound by the "pyramid", which we denote as the substrate binding module (Fig. 1A). The protein structures are similar, with the compounds bound in an essentially identical manner. The discussion focuses on the higher resolution tryptophan complex. There is no significant change in the protein structure upon tryptophan binding; the most notable differences is increased disorder in the loop between β2 and α2 at G48, adjacent to the isoalloxazine ring (fig. S5C). FAD and Cl^- are bound at the same positions as in the native enzyme. One face of the tryptophan ring stacks against H101 and W455, the other face against F103 (Fig. 1C). The indole nitrogen is hydrogen bonded to the backbone carbonyl of E346. The amino group of tryptophan makes a salt link with E450 and hydrogen bonds to other residues. The carboxylate group is also hydrogen bonded to the protein (Fig. 1C). We have obtained an $FADH_2$ structure by reducing the enzyme with dithionite. The crystals, obtained in the absence of Cl^-, have an empty Cl^- binding site. The isoalloxazine ring is bent and slightly changes its interaction with surrounding residues (fig. S5. C and D). The loop at G48 undergoes a conformational change and the helices capping the tryptophan binding site become disordered (fig. S5C). Together with the observation that in substrate and product complexes the G48 loop is less well ordered, it suggests there is communication between the modules regulating the enzyme.

The mechanism of flavin-dependent monooxygenases is well established (26,27). These enzyme are reduced to give $FADH_2$, which then binds molecular oxygen to form a spectroscopically characterized highly reactive peroxide-linked flavin (26,27). This intermediate is decomposed by the nucleophilic attack of an adjacent phenolate substrate, resulting in oxygen atom transfer. Consistent with this established mechanism (26,27), we and others (11,23) propose that in PrnA, $FADH_2$ binds O_2 forming the same peroxide-linked isoalloxazine ring (Scheme 1). There is no moiety in PrnA that appears capable of binding

or interacting with O_2 other than $FADH_2$. In PrnA, unlike PHBH, there is no room for an organic molecule to bind adjacent to the isoalloxazine ring and the substrate is 10Å distant. We see no indication of a large conformation change that could bring tryptophan and flavin together. Furthermore, the Cl⁻ ion binding site is over 10Å from 7-chlorotrytophan and no other Cl⁻ binding site has been located. These observations seem to disfavor mechanisms involving direct transformation of substrate by oxygen linked to flavin (11) or flavin hypochlorite (22,23). We cannot entirely eliminate these mechanisms as large conformational changes could take place. However, we postulate that the chlorinating agent created from flavin peroxide moves 10Å to the tryptophan. In PrnA, Cl⁻ is bound on the opposite face of FAD from the solvent and near to the entrance of the tunnel leading to the tryptophan (Fig. 1B). Cl⁻ is positioned to make a nucleophilic attack on the flavin peroxide, resulting in the formation of hydroxylated FAD and HOCl (Scheme 1). The formation of HOCl from peroxides is well known in chemistry. Close to the expected site of this reaction a water molecule is hydrogen bounded to S347 (Fig. 1B).

5-Methylindole incubated with PrnA results in the formation of 3-chloro-5-methylindole (28) (fig. S6). We have established that in solution, the substrate mimic 5-methylindole reacts with HOCl to give 3-chloro-5-methylindole (fig. S7). In order to confirm the enzyme reaction was not due to adventitious HOCl production, we tested PrnA for its ability to chlorinate monochlorodimedone (MCD). MCD is chlorinated by free HOCl (much more reactive than 5-methylindole) and is used to assay haloperoxidases (5,29) (fig. S6). MCD does not inhibit and is not chlorinated by PrnA (fig. S8), indicating that MCD does not access the tryptophan binding site. It establishes that production of 3-chloro-5-methylindole by enzyme does not arise from the presence of HOCl in solution. 5-Methylindole inhibits PrnA, consistent with it accessing the tryptophan binding site. The chlorinating agent at the tryptophan site has identical reactivity to that of HOCl (with respect to 5-methylindole). These data are consistent with the proposal that HOCl is created and channeled by PrnA. Modeling suggests PrnA binds 5-methylindole orienting its 3-position towards the tunnel (fig. S9). We propose HOCl is generated by the conserved flavin binding module and is therefore general to all flavin-dependent halogenases, irrespective of substrate or halogen (Cl or Br).

The 10 Å tunnel is lined by the polypeptide main chain and the side chains of residues I82, K79, I52 and S347 (Fig. 2A). These would not be oxidized or chlorinated by HOCl. W272 and W274 are remote from the site of HOCl formation and in any event, on its own tryptophan is not chlorinated by HOCl (23). We suggest that HOCl, after formation, is prevented from diffusing into solvent by the protein structure and instead enters this tunnel, moving towards tryptophan. In the native PrnA structure, K79 makes a hydrogen bond with a water molecule located at the end of the tunnel adjacent to the Cl atom of 7-chlorotryptophan (Fig. 2A). K79 may hydrogen bond to HOCl and thus position it to react with tryptophan. In the absence of this water molecule, there is a cavity centered on the 7-position of tryptophan (Fig. 2A). We propose the basis for regioselective halogenation is the controlled spatial presentation of HOCl to substrate. There is a chemical precedent for this; in solution HOCl reacts with anisole resulting in *p*- and *o*-substitution. By first adding cyclodextrin, which is thought to wrap around anisole masking the *o*-positions, only *p*-substitution occurs (30).

Electrophilic addition of chlorine to tryptophan proceeds through a Wheland intermediate (Fig. 2B). This intermediate would be stabilized by interaction with E346, reminiscent of, but reversed in polarity from, π cation interactions (31,32). E346 is positioned to deprotonate the intermediate leading to product (Fig. 2, A and B). Tryptophan is less reactive than other aromatics and in solution is not chlorinated by HOCl (23). A more electrophilic source of Cl than HOCl is known to be required (23). If K79 does hydrogen

bond to HOCl it would activate Cl by increasing its electrophilicity (Fig. 2B), allowing it to chlorinate tryptophan. Such a spatially constrained activating step may not only accomplish halogenation of less reactive substrates, but could protect nearby tryptophan residues. In our assay the native enzyme has a low k_{cat} of 0.1 min^{-1} (fig. S10), however rebeccamycin halogenase is also slow with a k_{cat} of 1.4 min^{-1}(23). We find no detectable activity for a K79A mutant, consistent with K79 playing a key role in guiding and activating HOCl. The k_{cat} for PrnA E346Q is decreased by about two orders of magnitude; K_M is unchanged (fig. S11). This supports our proposed role for E346 (Fig. 2B) in stabilization and deprotonation of the intermediate. Halogenation using HOCl is well known in nature and organic chemistry, yet it lacks the reactivity and regioselectivity required for biosynthetic pathways. We suggest that with halogenases, biology has evolved a remarkable mechanism of generating, activating and controlling HOCl.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and notes

1. Gribble GW. Chemosphere. 2003; 52:289. [PubMed: 12738253]

2. Williams DH, Bardsley B. Angew. Chem. Int. Ed. 1999; 38:1173.

3. Sanchez C, et al. Chem. Biol. 2002; 9:519. [PubMed: 11983340]

4. Sailler B, Glombitza K-W. Natural Toxins. 1999; 7:57. [PubMed: 10495467]

5. Sundaramoorthy M, Terner J, Poulos TL. Chem. Biol. 1998; 5:461. [PubMed: 9751642]

6. Messerschmidt A, Prade L, Wever R. J. Biol. Chem. 1997; 378:309.

7. van Pee KH, Unversucht S. Chemosphere. 2003; 52:299. [PubMed: 12738254]

8. Hofmann B, et al. J. Mol. Biol. 1998; 279:889. [PubMed: 9642069]

9. Dairi T, Nakano T, Aisaka K, Katsumata R, Hasegawa M. Biosci. Biotechnol. Biochem. 1995; 59:1099. [PubMed: 7612997]

10. Hammer PE, Hill DS, Lam ST, van Pee KH, Ligon JM. Appl. Environ. Microbiol. 1997; 63:2147. [PubMed: 9172332]

11. Keller S, et al. Angew. Chem. Int. Ed. 2000; 39:2300.

12. Hammer PE, Burd W, Hill DS, Ligon JM, van Pee KH. FEMS Microbiol. Lett. 1999; 180:39. [PubMed: 10547442]

13. Ligon JM, et al. Pest Manag. Sci. 2000; 56:688.

14. Ziegler DM. Drug Metabol. Rev. 2002; 34:503.

15. Hohaus K, et al. Angew. Chem. Int. Ed. 1997; 36:2012.

16. Puk O, et al. Chem. Biol. 2002; 9:225. [PubMed: 11880037]

17. Piraee M, White RL, Vining LC. Microbiology. 2004; 150:85. [PubMed: 14702400]

18. Eustaquio AS, et al. Chem. Biol. 2003; 10:279. [PubMed: 12670542]

19. Chiu HT, et al. Proc. Natl. Acad. Sci. U S A. 2001; 98:8548. [PubMed: 11447274]

20. Weitnauer G, et al. Chem. Biol. 2001; 8:569. [PubMed: 11410376]

21. Liu W, Christenson SD, Standage S, Shen B. Science. 2002; 297:1170. [PubMed: 12183628]

22. Hubbard BK, Walsh CT. Angew. Chem. Int. Ed. 2003; 42:730.

23. Yeh E, Garneau S, Walsh CT. Proc. Natl. Acad. Sci. U S A. 2005; 102:3960. [PubMed: 15743914]

24. Wierenga RK, Drenth J, Schulz GE. J. Mol. Biol. 1983; 167:725. [PubMed: 6876163]

25. Dong CJ, et al. Nature. 2004; 427:561–565. [PubMed: 14765200]

26. Sheng DW, Ballou DP, Massey V. Biochemistry. 2001; 40:11156. [PubMed: 11551214]

27. Entsch, et al. J. Biol. Chem. 1976; 251(1976):2550. [PubMed: 816794]

28. Hölzer M, Burd W, Reissig HU, van Pee KH. Adv. Synth. Catal. 2001; 343:591.

29. Hager LP, Morris DR, Brown FS, Eberwein H. J Biol. Chem. 1966; 241:1769. [PubMed: 5945851]

30. Breslow R, Campbell P. J. Am. Chem. Soc. 1969; 91:3085.

31. Dougherty DA. Science. 1996; 271:163. [PubMed: 8539615]

32. Zacharias N, Dougherty DA. Trends Pharmacol. Sci. 2002; 23:281. [PubMed: 12084634]
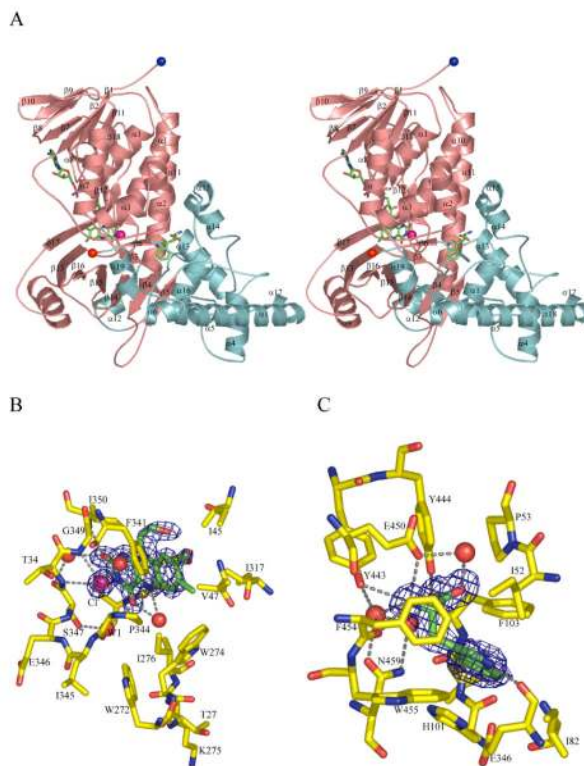
Fig. 1.
(A) The monomer of PrnA, a flavin-dependent halogenase shown in cartoon format, secondary structure elements are numbered. Residues 1 and 519-538 are disordered. The N-terminus is denoted by a blue sphere and C-terminus by a red one. FAD and 7-chlorotryptophan are shown as sticks, carbon green, oxygen red, nitrogen blue, chlorine purple and phosphorous magenta. The Cl⁻ ion is shown as a pink sphere. The "box" shaped flavin binding module (residues 1-102, 159-401) is colored salmon and has dimensions 30 Å × 30 Å × 60 Å. The "pyramid" shaped substrate module (residues 103-158, 402-518) is colored cyan, each side of the pyramid is approx 35 Å. The flavin binding module has two large β sheets one parallel consisting of β1,2,7,11,18 at the top of the figure and one mainly antiparallel β4,5,13-17 at the bottom of the figure. In addition it has two smaller β-sheets (β8-10 & β3,6) and an α-helical bundle (α1-3,7-11). The substrate binding module is entirely helical (α4-6, 12-19). **(B)** The FAD and Cl⁻ ion binding sites. The figure is oriented 180° around the vertical axis relative to Fig. 1A. An unbiased Fo-Fc map contoured at 3σ (blue) and at 8σ (cyan). Only the isoalloxazine of FAD is shown for clarity. A stereo version of the figure and the full molecule are shown in fig. S2B and C. The Cl⁻ ion is at over 10σ in the Fo-Fc map. In the protein molecule, carbon atoms are colored yellow, all other atoms are colored as in Fig. 1a and water molecules are shown as red spheres. One water molecule (labeled W1) is conserved in all structures and its relationship to the mechanism is discussed. Hydrogen bonds are shown as dotted lines. The isoalloxazine ring has essentially the same interaction with the protein in all structures. **(C)** The substrate binding site. An unbiased Fo-Fc map contoured at 3σ is shown for the tryptophan substrate. Atoms are colored as in Fig. 1B. The interactions with the protein are identical for 7-chlorotrptophan (the product). All residues within 4Å of tryptophan are shown and labeled. A stereo version of the figure is shown fig. S5A.
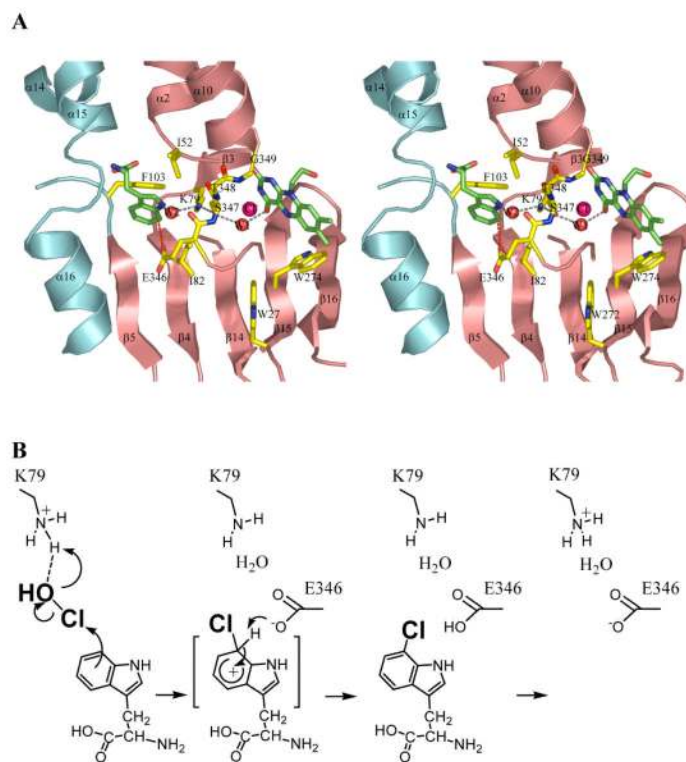
Fig. 2.
(A) A tunnel connects the FAD and tryptophan binding sites. The side chain of T348 is omitted and only a truncated FAD molecule is shown for clarity. Part of the protein structure is removed for clarity; secondary structure elements are labeled. W1, shown as a red sphere, is found in all structures. W2 is a second water molecule found only in the native structure. It has been placed in this figure to illustrate the path through the tunnel in the protein which we suggest HOCl follows, from Cl⁻ to substrate. W2 is absent in both co-complexes because it would sterically clash with ligand. The interaction of E346 with 7-position of tryptophan is shown as a red dotted line. **(B)** Halogenation of tryptophan proceeds by electrophilic aromatic substitution at the 7-position. The Wheland intermediate is shown in [] and would be stabilized by interaction with E346. K79 in the apo structure makes 3 hydrogen bonds suggesting a polar environment for the NZ atom, consistent with a protonated NZ atom. The indole ring stacks with W455 and H101 on one face and F103 on the other, which may further stabilize the intermediate.
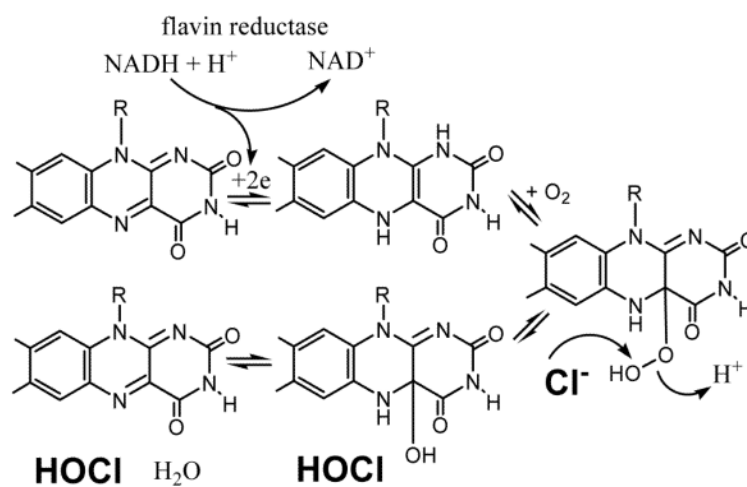
Scheme 1.
The formation of the peroxy-flavin molecule is known to occur in the related monooxygenase enzymes (26,27). Formation of HOCl by PrnA requires nucleophilic attack of Cl^- on the peroxy-flavin.