# Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination

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

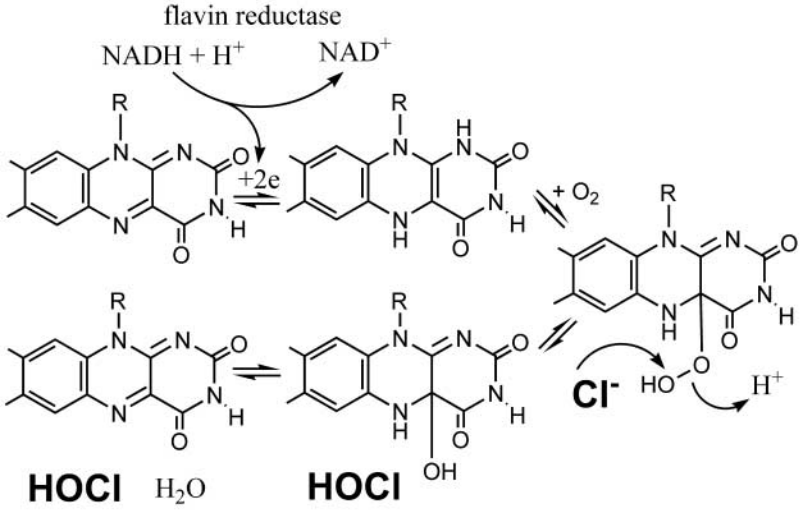
Chlorinated natural products include vancomycin and cryptophycin A. Their biosynthesis involves regioselective chlorination by flavin-dependent halogenases. We report the structural characterization of tryptophan 7-halogenase (PrnA), which regioselectively chlorinates tryptophan. Tryptophan and flavin adenine dinucleotide (FAD) are separated by a 10 angstrom–long tunnel and bound by distinct enzyme modules. The FAD module is conserved in halogenases and is related to flavin-dependent monooxygenases. On the basis of biochemical studies, crystal structures, and by analogy with monooxygenases, we predict that $FADH_2$ reacts with $O_2$ to make peroxyflavin, 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 4000 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 to produce 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*). Members of this halogenase superfamily have been identified in the biosynthetic pathways of the antibiotics balhimycin (*16*) and 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 has been explicitly discounted because of its reactivity and regioselectivity (*11*, *23*).

PrnA was purified from a *P. fluorescens* expression system, and its structure was determined to 1.95 Å. 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 $Thr^{348}$ (T348) and $Gly^{349}$ (G349) (*25*). 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 (*26*). Complexes with tryptophan and 7-chlorotryptophan were obtained by incubating the protein, before




**Scheme 1.**

[1]Centre for Biomolecular Sciences, EaStchem, University of St. Andrews, St. Andrews KY16 9ST, UK. [2]Institut für Biochemie, Technische Universität Dresden, D-01062 Dresden, Germany.

*To whom correspondence should be addressed.

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 substantial change in the protein structure upon tryptophan binding; the most notable difference 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 (*25*), 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 (*27*, *28*). These enzymes are reduced to give $FADH_2$, which then binds molecular oxygen to form a spectroscopically characterized highly reactive peroxide-linked flavin (*27*, *28*). This intermediate is decomposed by the nucleophilic attack of an adjacent phenolate substrate, resulting in oxygen atom transfer. Consistent with this established mechanism (*27*, *28*), 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 10 Å 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 10 Å 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 because large conformational changes could take place. However, we postulate that the chlorinating agent created from flavin peroxide moves 10 Å 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

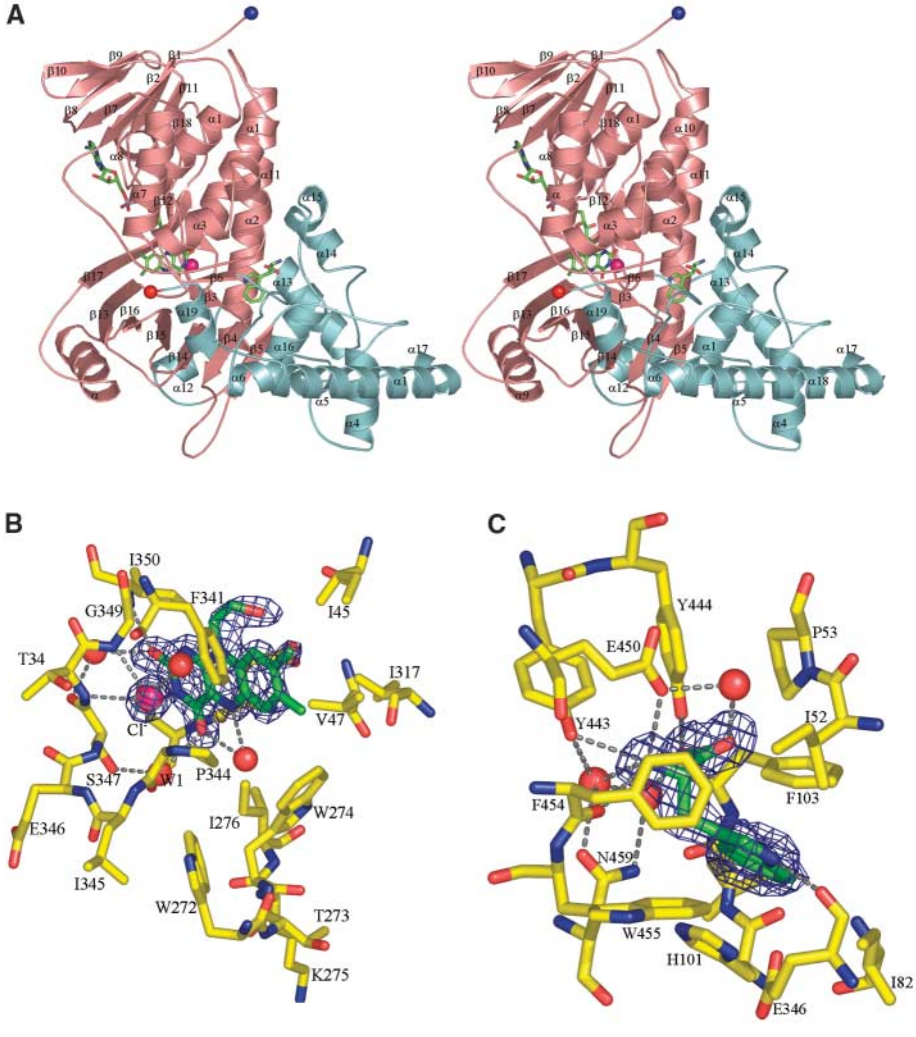



**Fig. 1.** (**A**) Stereoimage of the monomer of PrnA, a flavin-dependent halogenase; secondary structure elements are numbered (*25*). Residues 1 and 519 to 538 are disordered. The N terminus is denoted by a blue sphere and C terminus by a red one. FAD and 7-chlorotryptophan are shown as sticks; carbon, green; oxygen, red; nitrogen, blue; chlorine, purple; and phosphorus, magenta. The $Cl^-$ ion is shown as a pink sphere. The box-shaped flavin binding module (residues 1 to 102 and 159 to 401) is colored salmon and has dimensions of 30 Å by 30 Å by 60 Å. The pyramid-shaped substrate module (residues 103 to 158 and 402 to 518) is colored cyan; each side of the pyramid is about 35 Å. The flavin binding module has two large β sheets, one parallel consisting of β1, β2, β7, β11, and β18 at the top of the figure and one mainly antiparallel β4, β5, and β13 to β17 at the bottom of the figure. In addition, it has two smaller β sheets (β8 to β10 and β3 and β6) and an α-helical bundle (α1 to α3 and α7 to α11). The substrate binding module is entirely helical (α4 to α6 and α12 to α19). (**B**) The FAD and $Cl^-$ ion binding sites. The figure is oriented 180° around the vertical axis relative to (A). An unbiased Fo-Fc map is 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. S2, B 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 (A), 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 (B). The interactions with the protein are identical for 7-chlorotryptophan (the product). All residues within 4 Å of tryptophan are shown and labeled. A stereo version of the figure is shown as fig. S5A.

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).

PrnA incubated with 5-methylindole results in the formation of 3-chloro-5-methylindole (*29*) (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, which is much more reactive than 5-methylindole, is chlorinated by free HOCl and is used to assay haloperoxidases (*5*, *30*) (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 toward 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 (*25*) (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 toward 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, it has been shown that only *p*-substitution occurs (*31*).

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 (*32*, *33*). 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 K79→A79 mutant, consistent with K79 playing a key role in guiding and activating HOCl. The $k_{cat}$ for PrnA E346Q (an E346→Q346 mutant) 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.

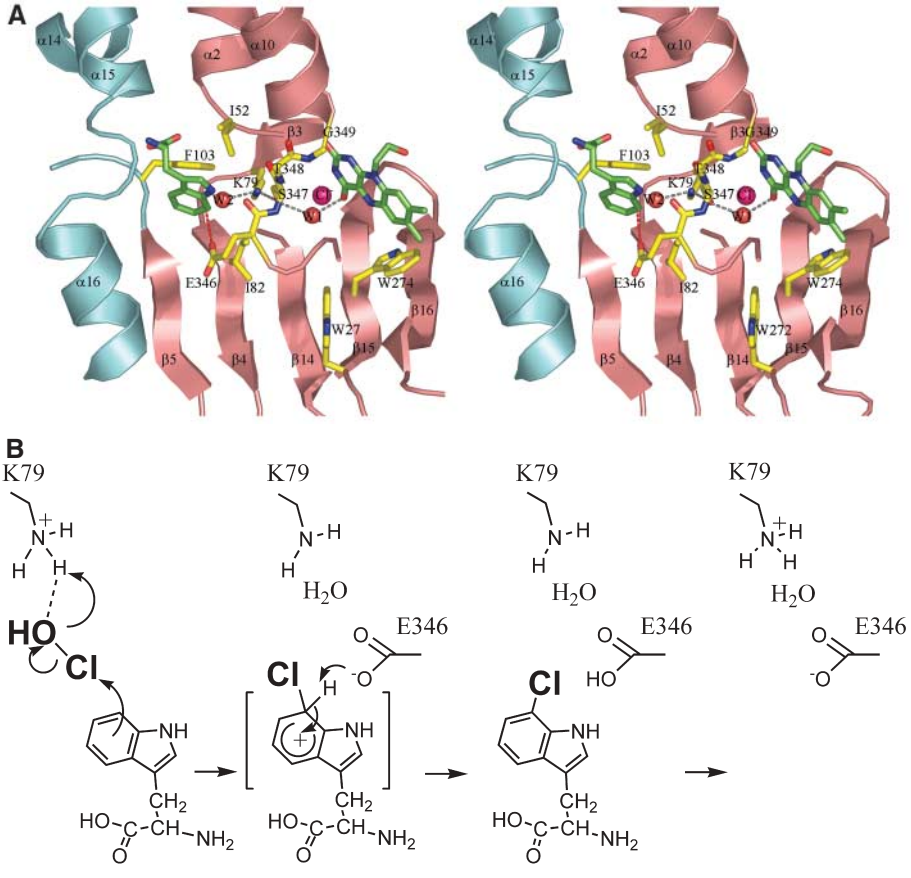



**Fig. 2.** (**A**) A tunnel connects the FAD and tryptophan binding sites. The side chain of T348 (*25*) is omitted and only a truncated FAD molecule is shown for clarity. Part of the protein structure is removed for clarity; secondary structure elements are labeled. W1, shown as a red sphere, is found in all structures. W2 is a second water molecule found only in the native structure. It has been placed in this image to illustrate the path through the tunnel in the protein that 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 the 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 square brackets and would be stabilized by interaction with E346. K79 in the apo structure makes three 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.



### References and Notes

1. G. W. Gribble, *Chemosphere* **52**, 289 (2003).
2. D. H. Williams, B. Bardsley, *Angew. Chem. Int. Ed.* **38**, 1172 (1999).
3. C. Sanchez *et al.*, *Chem. Biol.* **9**, 519 (2002).
4. B. Sailler, K.-W. Glombitza, *Nat. Toxins* **7**, 57 (1999).

5. M. Sundaramoorthy, J. Terner, T. L. Poulos, *Chem. Biol.* **5**, 461 (1998).
6. A. Messerschmidt, L. Prade, R. Wever, *Biol. Chem.* **378**, 309 (1997).
7. K. H. van Pee, S. Unversucht, *Chemosphere* **52**, 299 (2003).
8. B. Hofmann *et al.*, *J. Mol. Biol.* **279**, 889 (1998).
9. T. Dairi, T. Nakano, K. Aisaka, R. Katsumata, M. Hasegawa, *Biosci. Biotechnol. Biochem.* **59**, 1099 (1995).
10. P. E. Hammer, D. S. Hill, S. T. Lam, K. H. van Pee, J. M. Ligon, *Appl. Environ. Microbiol.* **63**, 2147 (1997).
11. S. Keller *et al.*, *Angew. Chem. Int. Ed.* **39**, 2300 (2000).
12. P. E. Hammer, W. Burd, D. S. Hill, J. M. Ligon, K. H. van Pee, *FEMS Microbiol. Lett.* **180**, 39 (1999).
13. J. M. Ligon *et al.*, *Pest Manag. Sci.* **56**, 688 (2000).
14. D. M. Ziegler, *Drug Metab. Rev.* **34**, 503 (2002).
15. K. Hohaus *et al.*, *Angew. Chem. Int. Ed.* **36**, 2012 (1997).
16. O. Puk *et al.*, *Chem. Biol.* **9**, 225 (2002).
17. M. Piraee, R. L. White, L. C. Vining, *Microbiology* **150**, 85 (2004).
18. A. S. Eustaquio *et al.*, *Chem. Biol.* **10**, 279 (2003).
19. H. T. Chiu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8548 (2001).
20. G. Weitnauer *et al.*, *Chem. Biol.* **8**, 569 (2001).
21. W. Liu, S. D. Christenson, S. Standage, B. Shen, *Science* **297**, 1170 (2002).
22. B. K. Hubbard, C. T. Walsh, *Angew. Chem. Int. Ed.* **42**, 730 (2003).
23. E. Yeh, S. Garneau, C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3960 (2005).
24. R. K. Wierenga, J. Drenth, G. E. Schulz, *J. Mol. Biol.* **167**, 725 (1983).
25. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
26. C. Dong *et al.*, *Nature* **427**, 561 (2004).
27. D. Sheng, D. P. Ballou, V. Massey, *Biochemistry* **40**, 11156 (2001).
28. B. Entsch, D. P. Ballou, V. Massey, *J. Biol. Chem.* **251**, 2550 (1976).
29. M. Hölzer, W. Burd, H. U. Reissig, K. H. van Pee, *Adv. Synth. Catal.* **343**, 591 (2001).
30. L. P. Hager, D. R. Morris, F. S. Brown, H. Eberwein, *J. Biol. Chem.* **241**, 1769 (1966).
31. R. Breslow, P. Campbell, *J. Am. Chem. Soc.* **91**, 3085 (1969).
32. D. A. Dougherty, *Science* **271**, 163 (1996).
33. N. Zacharias, D. A. Dougherty, *Trends Pharmacol. Sci.* **23**, 281 (2002).


34. J.H.N. is a Biotechnology and Biological Sciences Research Council (BBSRC) career development fellow. We thank J. Farack for chlorination of 5-methylindole with HOCl and M. Dorward for technical assistance. This work was supported by grants from the BBSRC (J.H.N.), the Wellcome Trust (J.H.N.), the Deutsche Forschungsgemeinschaft [Pe 348/17-1 and 17-2 (K.H.V.P.)], the Sächsische Staatsministerium für Umwelt und Landwirtschaft (K.H.V.P.) and the Fonds der Chemischen Industrie (K.H.V.P.). We thank G. Taylor, D. O'Hagan, and S. Chapman for a critical review of the manuscript. The native structure was determined by the Scottish Structural Proteomics Facility funded by Scottish Higher Education Funding Council and BBSRC. We acknowledge the use of BM14 (UK) and thank M. Walsh for data collection assistance. The use of European Synchrotron Radiation Facility beamlines is acknowledged, as is the BBSRC circular dichroism spectroscopy center in Glasgow. Coordinates and data have been deposited in the Protein Database with codes 2APG, 2AQJ, 2ARD, and 2AR8.


**Supporting Online Material**
www.sciencemag.org/cgi/content/full/309/5744/2216/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 to S2
References

23 June 2005; accepted 19 August 2005
10.1126/science.1116510

# Rev1 Employs a Novel Mechanism of DNA Synthesis Using a Protein Template

**Deepak T. Nair,[1] Robert E. Johnson,[2] Louise Prakash,[2] Satya Prakash,[2] Aneel K. Aggarwal[1]***

The Rev1 DNA polymerase is highly specialized for the incorporation of C opposite template G. We present here the crystal structure of yeast Rev1 bound to template G and incoming 2′-deoxycytidine 5′-triphosphate (dCTP), which reveals that the polymerase itself dictates the identity of the incoming nucleotide, as well as the identity of the templating base. Template G and incoming dCTP do not pair with each other. Instead, the template G is evicted from the DNA helix, and it makes optimal hydrogen bonds with a segment of Rev1. Also, unlike other DNA polymerases, incoming dCTP pairs with an arginine rather than the templating base, which ensures the incorporation of dCTP over other incoming nucleotides. This mechanism provides an elegant means for promoting proficient and error-free synthesis through $N^2$-adducted guanines that obstruct replication.

Rev1, a member of the eukaryotic Y family DNA polymerases, is highly specific for incorporating a C opposite template G (*1*, *2*). In this respect, Rev1 differs not only from the replicative and repair polymerases (Pols), which incorporate the correct nucleotide opposite all four template bases with nearly equivalent catalytic efficiencies, but it differs

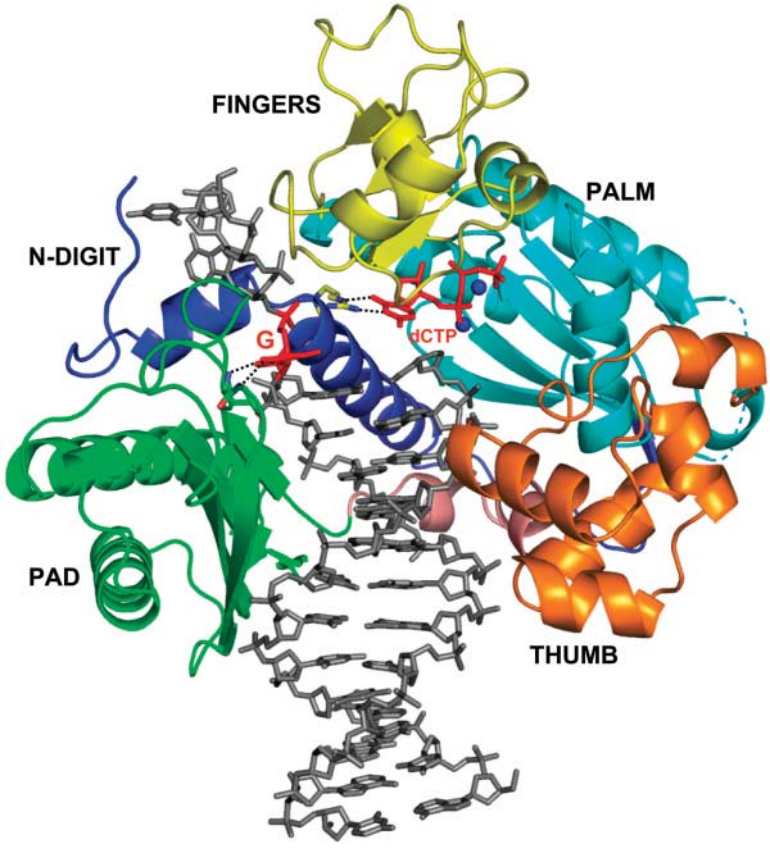



**Fig. 1.** Structure of the Rev1-DNA-dCTP ternary complex. The palm, fingers, and thumb domains and the PAD are shown in cyan, yellow, orange, and green, respectively. The linker joining the thumb to the PAD is shown in pink. The N-digit in Rev1 is shown in dark blue. DNA is in gray, template G and incoming dCTP are in red, and the putative $Mg^{2+}$ ions are in dark blue. Black dashed lines depict hydrogen bonds between dCTP and $Arg^{324}$, and between template G and a loop in the PAD. Cyan dashed line indicates an unstructured loop in the palm domain.

# Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination

Changjiang Dong, Silvana Flecks, Susanne Unversucht, Caroline Haupt, Karl-Heinz van Pée, and James H. Naismith

*Science*, 309 (5744), • DOI: 10.1126/science.1116510