ARTICLE

Check for updates

Site-selective chlorination of pyrrolic heterocycles by flavin dependent enzyme PrnC

GuangRong Peh [1], Terence Tay[2], Lee Ling Tan[3], Elaine Tiong[3], Jiawu Bi [3], Yi Ling Goh[1], Suming Ye [1], Fu Lin [2], Cheryl Jia Xin Tan[4], Yong Zi Tan [3,4,5], Joel Wong [1], Huimin Zhao [2,6], Fong Tian Wong [1,3✉], Ee Lui Ang[2,7✉] & Yee Hwee Lim [1,7✉]

Halogenation of pyrrole requires strong electrophilic reagents and often leads to undesired polyhalogenated products. Biocatalytic halogenation is a highly attractive approach given its chemoselectivity and benign reaction conditions. While there are several reports of enzymatic phenol and indole halogenation in organic synthesis, corresponding reports on enzymatic pyrrole halogenation have been lacking. Here we describe the in vitro functional and structural characterization of PrnC, a flavin-dependent halogenase that can act on freestanding pyrroles. Computational modeling and site mutagenesis studies identified three key residues in the catalytic pocket. A moderate resolution map using single-particle cryogenic electron microscopy reveals PrnC to be a dimer. This native PrnC can halogenate a library of structurally diverse pyrrolic heterocycles in a site-selective manner and be applied in the chemoenzymatic synthesis of a chlorinated analog of the agrochemical fungicide Fludioxonil.

[1] Institute of Sustainability for Chemicals, Energy and Environment (ISCE2), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore. [2] Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore. [3] Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore. [4] Department of Biological Sciences, National University of Singapore, Singapore, Singapore. [5] Disease Intervention Technology Laboratory (DITL), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore. [6] Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [7] Synthetic Biology Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore. ✉email: wongft@imcb.a-star.edu.sg; ang_ee_lui@sifbi.a-star.edu.sg; lim_yee_hwee@isce2.a-star.edu.sg

Pyrroles are privileged structures in medicinal chemistry due to the bioactivities (antifungal, antibacterial, and anti-cancer) that are often associated with natural products possessing such scaffold (Fig. 1)[1]. In addition to their use as drugs, pyrrolic compounds have also found applications in agrochemicals[2-7], in advanced materials[8,9] (e.g., organic semiconductors, and solar cells), or as organic catalysts[10,11]. The high electron density of the pyrrole ring makes it susceptible to air oxidation when it is not substituted with electron-withdrawing substituents or part of a larger conjugated system[12,13]. As such, chemical halogenation of unsubstituted pyrroles often leads to uncontrolled poly-halogenation. Instead, halogenation of N-protected pyrroles with N-halosuccinimides[14] is a common approach though this method is notoriously unreactive, especially with less reactive N-chlorosuccinimide. Due to the higher resonance stability of the C2(C5) pyrrolic intermediate carbocation, Hammond's postulate implies pyrroles are more likely to undergo C2 halogenation (Fig. 2, also see Supplementary Materials and Methods, Section 1.8, Fig. S1). Therefore, direct C3-selective halogenation is even more challenging[15] in the absence of electrophilic-directing and/or steric hindering groups in the C2 position. In contrast, enzymatic pyrrole halogenation can predominantly yield site-selective mono-halogenation under mild buffer conditions without the need for any protective and/or directing groups[16-19], and avoids toxic halogenating reagents. The mono-halogenation feature further provides a useful handle for late-stage functionalization through a plethora of metal-catalyzed coupling reactions[20].

In nature, pyrrolic halogenases belong to the family of flavin-dependent halogenases (FDHs)[21-23]. Most of the well-studied flavin-dependent halogenases to-date focuses broadly on phenolic[24,25], aryl, or tryptophan-based[26-28] substrates corresponding to their native substrates. Of the known pyrrole halogenases, there are very limited reported examples of flavin-dependent halogenase acting on free-standing pyrrole substrates[29-32] (variant A FDHs) while most pyrrole halogenases act on carrier protein-tethered substrates (variant B FDHs)[33-37]. Consequently, PrnC is an attractive enzyme for investigation

because it can halogenate an unconfined pyrrole substrate with the potential to be applied in a flexible manner to other enzymatic or chemical transformations.

The *prnC* gene encoding PrnC is part of the *prnABCD* gene cluster responsible for the biosynthesis of a natural product, pyrrolnitrin (Fig. 2). Pyrrolnitrin was first reported by Arima et al.[38] in 1964. Its biosynthesis from L-tryptophan was proposed by Lively[39] and others[40], and finally confirmed by Van Pée and co-workers[41] through isotopic labeling using the *Pseudomonas fluorescens* strain BL915. In the proposed mechanism, the two chlorine atoms in **1** are introduced sequentially by the tryptophan halogenase PrnA, and the pyrrole halogenase PrnC that is hypothesized to regioselectively chlorinate the C3 position of the pyrrolic backbone. While PrnA is well studied[42], there are no reports of characterization and application of PrnC to date. Here,

Biosynthetic pathway of Pyrrolnitrin and site specific C-3 pyrrolic halogenation via the PrnC enzyme
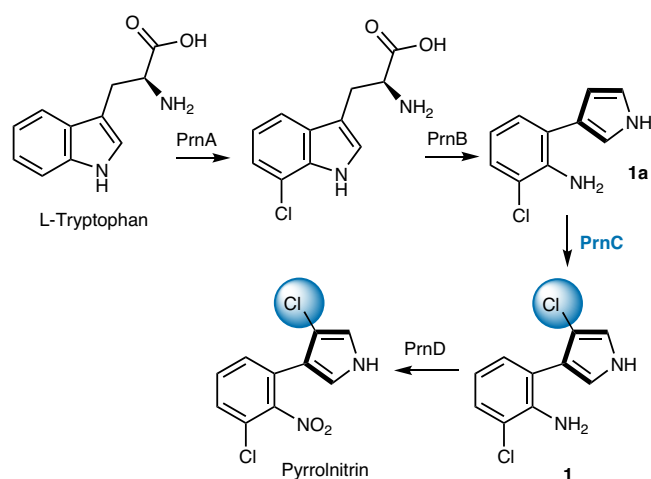
Representative chemical halogenation vs enzymatic site-selective halogenation of pyrroles (this work)

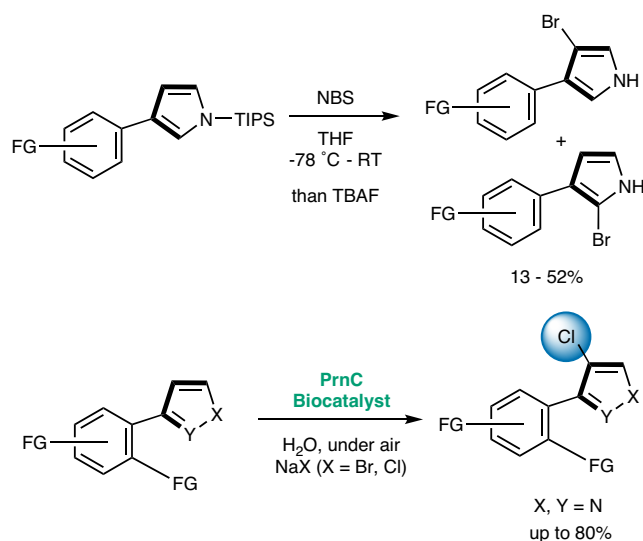**Fig. 2 The identification of the PrnC enzyme gene cluster and its application as a biocatalyst for the regioselective halogenation of pyrroles. Top:** Biosynthetic pathway of Pyrrolnitrin involving PrnABCD. **Bottom:** Comparison between chemical halogenation[63] and enzymatic halogenation of pyrroles (this work). NBS N-bromosuccinimide, TBAF tetra-n-butylammonium fluoride.

Marinoquinoline A
from *Ohtaekwangia kribbensis*
cytotoxic

Fludioxonil
fungicide

Celecoxib
anti-inflammatory

Slagenin A
from *Agelas nakamurai*
cytotoxic
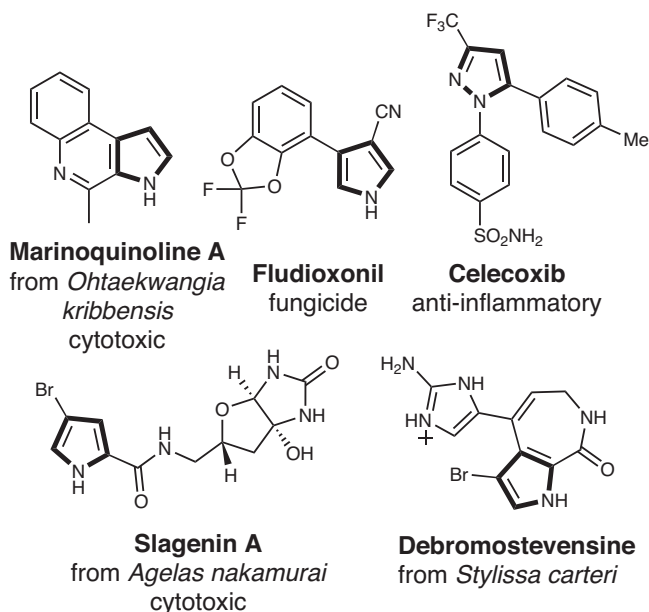
Debromostevensine
from *Stylissa carteri*

Fig. 1 Prevalence of the privilege pyrrolic framework. Pyrrole-based structures are evident in different domains, including Fludioxonil (agrochemical), celecoxib (pharmaceutical), and alkaloids such as slagenin A and debromostevensine (natural products).

we describe the first characterization of PrnC's activity and specificity towards structurally diverse pyrrolic heterocycles, expanding our understanding of pyrrolic halogenases. Effective protein engineering facilitated the establishment of a robust heterologous expression of PrnC, enabling us to conduct comprehensive in vitro characterization, including cryo-electron microscopy (cryo-EM) to visualize the apo-PrnC which revealed indications of a dimeric arrangement. A chemoenzymatic synthesis of a chlorinated analog of the agrochemical fungicide, Fludioxonil, was also demonstrated.

Results and discussion
Our early attempts to express PrnC heterologously in *Escherichia coli* (*E. coli*) yielded largely insoluble proteins with yields generally <1 mg/L. Initial optimization by co-expression with chaperone proteins such as GroES-GroEL (encoded in the pGro7 plasmid) only led to marginal improvements to the yields. Subsequent extensive optimization of the expression construct along with the addition of a recently described *N*-terminal (11 amino acids) solubility tag[43] finally enabled consistent production (~ 9 mg/L of culture, see Supplementary Materials and Methods, section 1.8.1, Fig. S2a, b) of soluble functional PrnC for in vitro functional assay. As the soluble protein yield increased, we also observed a slight enhancement in activity, which could be attributed to improved enzyme folding (see Supplementary Materials and Methods, section 1.8.2, Fig. S3a).

Reaction development. To determine the optimal conditions to characterize PrnC against a library of substrates, we investigated the impact of various co-factor combinations (Table 1). Our initial assay using the native substrate monodechloroaminopyrrolnitrin (**1a**) with purified PrnC, along with the co-factors FAD (50 mol% relative to PrnC), NADH (5.0 equiv) and *E. coli* flavin reductase (Fre, 0.5 mol%) at 30 °C for 18 h gave only 9% conversion to aminopyrrolnitrin (**1**) (entry 1). Introduction of the glucose dehydrogenase (GdHi)-NADH regeneration system with the intent to improve the turnover of the enzyme resulted in an appreciable improvement (33% conversion), suggesting that other parameters should be evaluated (entry 2). Interestingly, a ten-fold decrease in FAD concentration led to a substantial increase in conversion by 2.5-fold to 81% (entry 3). We rationalize that the lower FAD concentration could help in the following two ways: (1) to modulate the rate of generation of the reactive hypohalous acid intermediate for the site-selective halogenation, thus preventing large accumulation of highly reactive intermediates in the active site or HOCl leakage[44–46], (2) to modulate the rate of formation of PrnC:FADH⁻ complex reducing the likelihood of enzyme inactivation[47]. Further investigation was conducted to determine whether there is a fine balance between the co-factors' ratio for optimal monochlorination conversion. Indeed, doubling the Fre concentration (2×) to increase $FADH_2$ turnover led to a slight drop in monochlorination conversion (entry 4) along with small amount of dichlorinated product. Adding sub-stoichiometric amounts of NADH (0.05 equiv) (entry 5) does not adversely affect the conversion process, demonstrating efficient coupling of the (GdHi)-NADH regeneration system with other cofactors. Interestingly, doubling the loading of PrnC (entry 6) with the intention to limit non-specific HOCl within the substrate binding site resulted in higher production of di-chlorinated products. While the mechanism of the second halogenation event[48] is unclear at this moment, this observation along with other control experiments (see Supplementary Materials and Methods, section 1.8.2, Fig. S3c–d) suggest that the second halogenation is also facilitated by catalytic residues within the enzyme active site. The second

halogenation event appears to be less energetically favorable, likely due to the reduced electron density on the pyrrolic ring after the initial halogenation. To ascertain that the halogenation event was not caused by the formation of HOCl in futile cycles, another control experiment (entry 7) where co-factor catalase was introduced to the best conditions (entry 3) was conducted (also see Supplementary Materials and Methods, section 1.8.2, Fig. S3e, f). The similar conversion results (85%) to the conditions without catalase addition suggest that halogenation is indeed catalyzed by the enzyme PrnC. Removal of the glucose dehydrogenase (GdHi)-NADH regeneration system led to worse outcome, suggesting there is a tight kinetic balance between the PrnC enzyme and all its co-factors. Expectedly, the native PrnC enzyme has the highest preference for chloride followed by bromide. It does not accept iodide (see Supplementary Materials and Methods, section 1.8.2, Fig. S3g).

To explore the sequence space around PrnC, a sequence similarity network was generated, which led to the identification of four other homologs with conserved active site residues and genomic neighborhoods of PrnAB (see Supplementary Materials and Methods, section 1.8.2, Fig. S4a–c). Similar to PrnC, these homologs exhibited solubility issues when heterologously expressed and optimized in *E. coli* hosts. Despite these challenges, we were able to obtain soluble expression of the homologs, and they were subsequently characterized against the native PrnC (see Supplementary Materials and Methods, section 1.8.2, Fig. S4d, e). Through these investigations, native PrnC was identified as the most effective halogenase, and therefore was utilized for the remainder of the study.

Native PrnC was found to have a relatively low k_{cat} of 0.46 min⁻¹ against its native substrate **1a** (Table S1). This is largely similar to other wild-type tryptophan-based FDHs such as RebH and PyrH (Table S1). PrnC also exhibits a moderate affinity for **1a**, with $K_m = 15.8 \pm 0.7$ µM. The catalytic efficiency of this native PrnC enzyme (K_{cat}/K_m) was found to be modest at 0.029 ± 0.002 µM⁻¹ min⁻¹ (see Supplementary Materials and Methods, section 1.8, Table S1 and section 1.8.2, Fig. S5).

Molecular docking and Cryo-EM analysis. To determine the residues that are essential for the catalytic reaction of the PrnC enzyme, a systematic computational modeling of the enzyme was performed and a three-dimensional model for the PrnC/substrate complex (see method: homology and docking models of PrnC) was achieved. A PrnC/substrate complex model (Fig. 3) indicates that K97 is close to **1a** and may initiate catalysis either directly by reacting with HOCl to form a chloramine intermediate[49] or indirectly by stabilizing the chloroamine intermediate through hydrogen bonds[50]; E129 is also postulated to form a salt bridge with K97 to stabilize its sidechain conformation and E60 may also form a hydrogen bond with **1a**. To validate the computationally proposed key residues, the site-directed mutagenesis of all three residues was conducted, and its results are shown in Fig. S6 (see Supplementary Materials and Methods, section 1.8.2). Mutation of either residue (K97, E60, or E129) to alanine abolished or significantly reduced the enzyme activity, indicating that these residues are important for catalysis, as suggested by the above hypothesis. Among, mutation of K97 resides also led to a significant drop in protein expression yield.

At the outset of our study (2020), we employed a closely related crystal structure of a FAD-dependent halogenase, PltM, for homology modeling, albeit with only ~30% sequence identity with PrnC (see Supplementary Materials and Methods, section 1.8.2, Fig. S7a, b). Subsequently, a more closely resembling apo structure of a halogenase (CtCP) has been reported[51]. Notably, this apo-CtCP structure exhibits significant structural

Table 1 Optimization study of PrnC biocatalyst on the native substrate 1a.
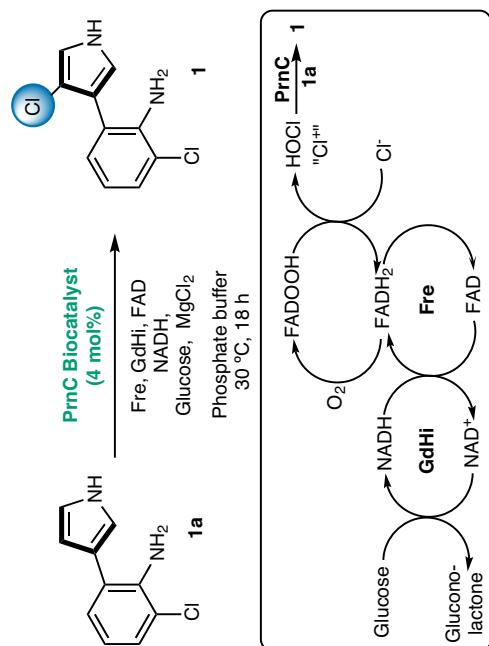
Entry	Variation from standard conditions						Conversion of 1^a(%)
	FAD (mol%)	GdHi (mol%)	Glucose (Equiv)	Fre (mol%)	NADH (Equiv)	PrnC (mol%)	
1.	2.0 (10×)	✘	✘	✓	✓	✓	9
2.	2.0 (10×)	✓	✓	✓	✓	✓	33
3.	✓	✓	✓	✓	✓	✓	81^{b,c}
4.	✓	✓	✓	1.0 (2×)	✓	✓	77^d
5.	✓	✓	✓	✓	0.05 (Less 100×)	✓	77
6.	✓	✓	✓	✓	0.05 (Less 100×)	8.0 (2×)	74^e
7.^{f}	✓	✓	✓	✓	✓	✓	85^c

^{a}Conversion is calculated based on calibration curves of mono-chlorinated product (**1**) and starting material (**1a**)(Calibration curve in Supplementary Materials and Methods, section 1.8.2, Fig. S3b).
^{b}Standard conditions: **1a** (0.5 mM), PrnC Biocatalyst (3 mol%), GdHi (0.5 mol%), FAD (0.2 mol%), Fre (0.5 mol%), NADH (5.0 equiv), Glucose (10.0 equiv), MgCl$_2$ (20.0 equiv), phosphate buffer (10 mM, pH 7.4).
^{c}≤5% di-chlorinated product.
^{d}~8% di-chlorinated product.
^{e}~12% di-chlorinated product.
^{f}Catalase (70 U/mL) added. Conversion of di-chlorinated product calculated based on relative HPLC trace between mono and di-chlorinated product peaks at 254 nm. Symbols: ✓ (tick) = as per standard conditions, ✘ (cross) = reagent/co-factor not used. FAD = flavin adenine dinucleotide, GdHi = glucose dehydrogenase, Fre = flavin reductase, NADH = nicotinamide adenine dinucleotide (reduced form).

resemblance to the initial PItm crystal structure employed in our modeling and optimization study. Consequently, the prediction utilizing PItM retains its validity. Both of these closely related halogenases, CtcP and PltM adopt a dimeric crystalline structure. The conserved dimeric interface observed in these structures also corroborates with the predicted structures generated by AlphaFold-Multimer (see Supplementary Materials and Methods,
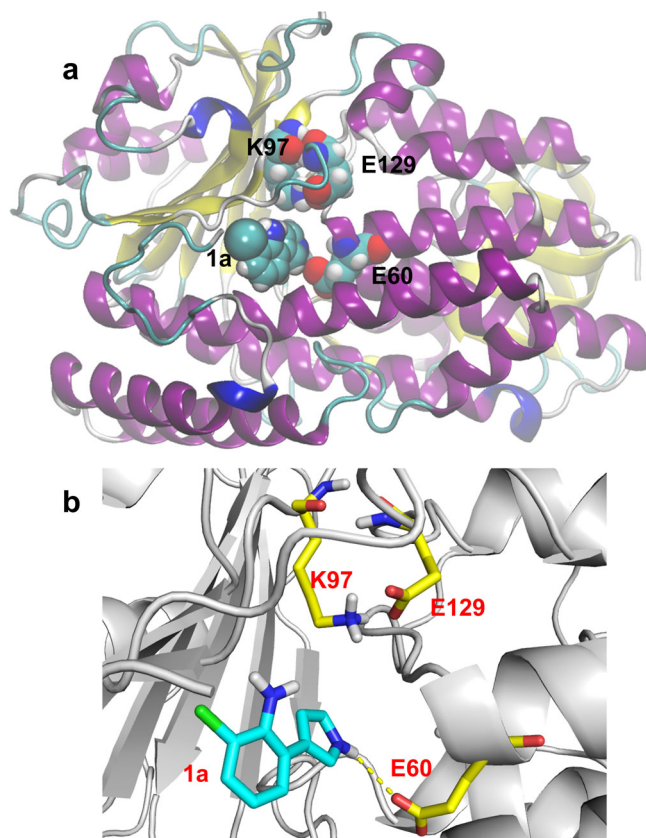

Fig. 3 The homology model of PrnC reveals the catalytic triad accountable for its site-specific halogenation function. a The optimized homology model for PrnC is based on the template crystal structure of halogenase PltM (PDB code: 6BZA). **b** The substrate binding site of the PrnC-**1a** complex from molecular modeling and docking. K97, E129, and E60 are proposed to be the key residues.

section 1.8.2, Fig. S7c). We also employed single-particle cryogenic electron microscopy (cryo-EM[52]) to determine the experimental structure of PrnC. When the apo-structure of PrnC was imaged, 2D class averages resembling dimers were observed (Fig. 4a). Based on the overall medium resolution 3D reconstruction obtained at around 9 Å, two copies of PrnC can be fitted into the density map when it is docked with the model (Fig. 4b). Several prior examples[49,50,53–55] have suggested that tryptophan based flavin-dependent halogenases exist as homodimers in crystalline state and as monomers in solution. Examination of chemically cross-linked NT11-PrnC in solution also suggests that PrnC exists as multi-mers (see Supplementary Materials and Methods, section 1.8.2, Fig. S8). However, we were not able to unambiguously assign the dimer interfaces at this resolution and thus, it cannot be ascertained at this stage whether this dimeric form of PrnC is critical for function.

Substrate scope. To determine the scope and specificity of PrnC, we tested against a panel of substituted pyrroles (Fig. 5) featuring (a) different substituted aryl or heteroaryl groups, or cyclic and acyclic alkyl groups at C-3 to the pyrrolic fragment (see Supplementary Materials and Methods, section 1.7.1, scheme S1-2), (b) C-2 positional linkage of the pyrrolic fragment instead of C-3 (see Supplementary Materials and Methods, section 1.7.1, scheme S3), and c) other five-membered heterocycles instead of pyrroles (see Supplementary Materials and Methods, section 1.7.1, scheme S4). By employing the optimized conditions (Table 1, entry 3), PrnC was found to be able to accommodate a range of aryl or heteroaryl substituents at the C-3 of the pyrrole (products **2–9**) giving conversion ranging between 15% and 87% (see Supplementary Materials and Methods, section 1.8.2, Fig. S10 and section 1.8.3 Figs S11–24). Protecting the free amino group with acetanilide in product **2** significantly reduces the conversion yield (15%), which may be due to the limited space for an extra acetyl group in the active site (Fig. 3). For **4, 5**, and **6**, the simultaneous replacement of the ortho-phenyl -NH2 group and removal of the phenyl 3-Cl atom had a moderate impact on the conversion (22%, 41%, and 26%), respectively. These results also confirm that prior installation of the 3-Cl atom by PrnA has no bearing on the subsequent step of halogenation. The total removal of substituents from the aryl fragment **7** (65%) or substitution with a heterocycle pyridine **8** (53%) or quinoline **9** (51%) yielded good conversion, indicating that the hydrophobic interaction of these phenyl groups embedded in the hydrophobic patch is highly relevant (see Supplementary Materials and Methods, section 1.8.2, Fig. S7a). These findings further suggest that the amino
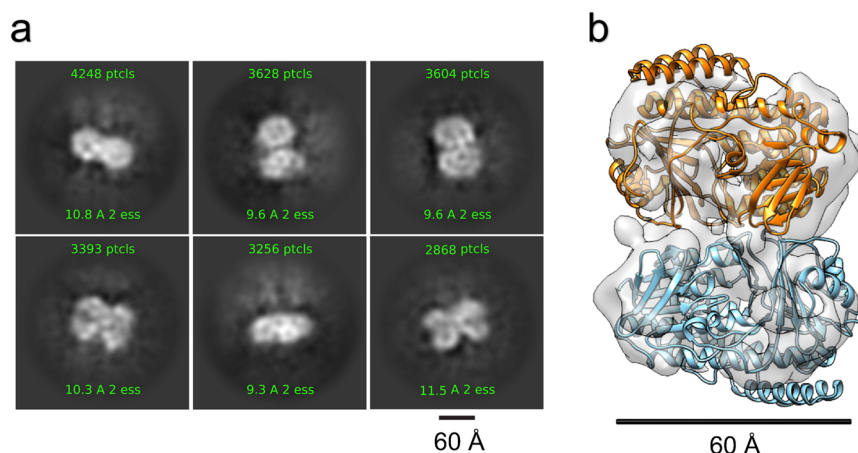

Fig. 4 Cryo-electron microscopy for the empirical structural elucidation of PrnC. Single-particle cryo-EM analysis of PrnC reveals that the protein exists as dimers in solution, as seen from both **a** the 2D class averages and **b** docking of the homology model into the 3D reconstruction.
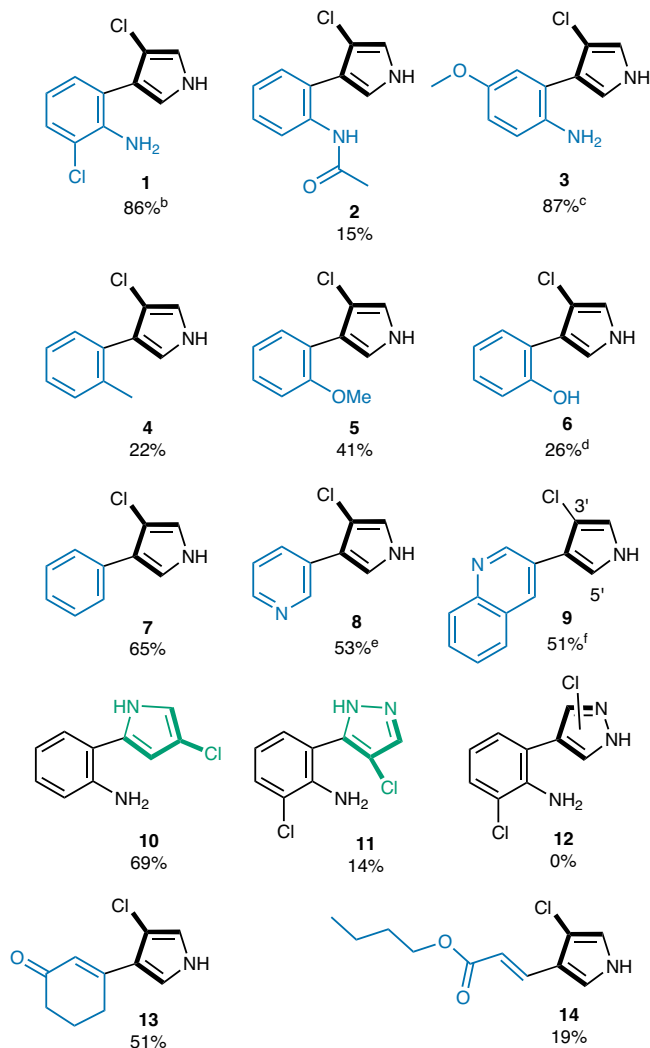
Fig. 5 Scope of PrnC mediated in vitro chlorination[a]. [a]
Conditions:Substrate (0.5 mM), PrnC Biocatalyst (4 mol%), GdHi (0.5 mol
%), FAD (0.2 mol%), Fre (0.5 mol%), NADH (5.0 equiv), Glucose
(10.0 equiv), MgCl$_2$ (20.0 equiv), phosphate buffer (10 mM, pH 7.4). The
conversion (%) represents the area under the peak for desired product and
remaining starting material (SM) relative to the area of a control SM
standard. Average of duplicate measurements determined by LC/MS from
UV absorbance at 254 nm. [b] Based on product standard calibration curve.
[c] Mono-Chlorinated isomer: Di-Chlorinated isomer (7.3:1). [d] Mono-
Chlorinated isomer: Di-Chlorinated isomer (4.2:1). [e] Mono-Cl: Di-Cl
isomers (17:1). [f] Major isomer shown. Mono-Chlorinated 3'-isomer and
5'-isomer (1.9:1).

(–NH$_2$) moiety on the adjacent aryl ring does not appear to play
an active directing function in the halogenation step. Interest-
ingly, positional isomeric or isosteric alternatives in the pyrrolic
fragment were also accepted with 2'-pyrrole **10** and 5'-pyrazole
11 giving moderate conversions of 69% and 14%, respectively.
Other five-membered heterocycles such as 4'-pyrazole **12**, furan,
and thiophene or N-protected pyrroles (see Supplementary
Materials and Methods, section 1.8.2, Fig. S9), however, were not
tolerated by the PrnC enzyme. It is hypothesized that substrates
without the free pyrrolic (N-H) functionality group are not
favored by PrnC due to abrogating the effective hydrogen bond
with the crucial residue E60, as illustrated in Fig. 3, despite having
electron-rich rings that are susceptible to electrophilic substitu-
tion. Surprisingly, replacing the aromatic group on C3 of the

pyrroles with a non-aryl cyclohexanone (see Supplementary
Materials and Methods, section 1.7.1, scheme S5) is well-tolerated
by the PrnC enzyme, suggesting the feasibility of accommodating
alternative polar groups (entry 13, 51%). Notwithstanding,
complete replacement of the aromatic substituent to an acyclic
acrylate side chain (see Supplementary Materials and Methods,
section 1.7.1, scheme S5) is also accepted, albeit with diminished
conversion (entry 14, 19%). The regioselectivity of the bio-
halogenation has been analogously assigned to be on the back-
bone of the pyrrolic fragment based on our computational model
with further collaboration by 2D-NMR structural confirmation of
the products (see Supplementary Data 1).

Typically, substituted pyrroles are synthesized through the Van
Leusen [3 + 2] cycloaddition reaction between tosylmethyl
isocyanides (TosMICs)[56,57] and pre-functionalized electron-
deficient substrates, followed by deprotection. To showcase the
applicability and scalability of the biocatalytic halogenation by
PrnC, we applied the developed protocol towards the late-stage
protecting-group free chlorination of a decyano-derivative **15** of
an agrochemical fungicide, Fludioxonil. Substrate **15** was
synthesized from **13** in two-steps using a Pd-catalyzed Suzuki
coupling followed by tetra-n-butylammonium fluoride deprotec-
tion. Under both optimized as well as in vivo biosynthesis
conditions, substrate **15** was selectively chlorinated at the C-3
position to give the chlorinated analog **16** of Fludioxonil in an
isolated yield of 58% (Fig. 6).

Conclusions

In conclusion, we report the first characterization of the activity and
specificity of the pyrrolic halogenase PrnC towards its native sub-
strate and a mini library of substituted pyrroles. This in vitro work
was made possible by robust heterologous expression of PrnC via
an 11-amino acid NT-11 solubility tag. Using cryo-EM, we also
obtained the apo-structure of this enzyme as a dimeric state in
solution though its significance is yet to be fully understood. The
key residues that are responsible for PrnC's catalytic activity have
been identified by molecular modeling and validated by mutagen-
esis experiments. Leveraging the PrnC site-selective chlorination on
the pyrrole fragment of the substrate, we applied this halogenase
towards the chemoenzymatic synthesis of a C-3 chlorinated analog
of the agrochemical fungicide, Fludioxonil. This work highlights the
complementarity between chemical and biocatalytic methodologies
to access certain chemical spaces or reactive positions. Further
directed evolution studies are being conducted in our laboratories to
improve the enzyme's stability and promiscuity.

Methods

Halogenase cloning, expression, and purification. NT11-PrnC-
6His construct in pET-28a(+) was ordered from Twist Bios-
ciences as a clonal construct and transformed into T7 Express E.
coli (NEB). The resulting strain was cultured in 1 L LB media at
37 °C. When OD600 reached 0.4, 0.1 mM IPTG was used to
induce for overnight expression at 16 °C. After expression, the
cultures were centrifuged at 10,000 g for 10 min at 4 °C. The
resulting pellets were resuspended in 20 mL of 100 mM sodium
phosphate pH 7, 10 mM imidazole, 150 mM sodium chloride
before sonication. After sonication, the resulting lysate was then
centrifuged at 19,000 g for 1 h at 4 °C. The supernatant was
incubated with Ni-NTA agarose for 1 h at 4 °C. The resin was
washed with 20 mL 100 mM sodium phosphate pH 7, 80 mM
imidazole, 150 mM sodium chloride and the bound protein was
eluted with 5 mL of 100 mM sodium phosphate pH 7, 500 mM
imidazole, 50 mM sodium chloride. The elution was buffer
exchanged and concentrated with 50 mM sodium phosphate pH
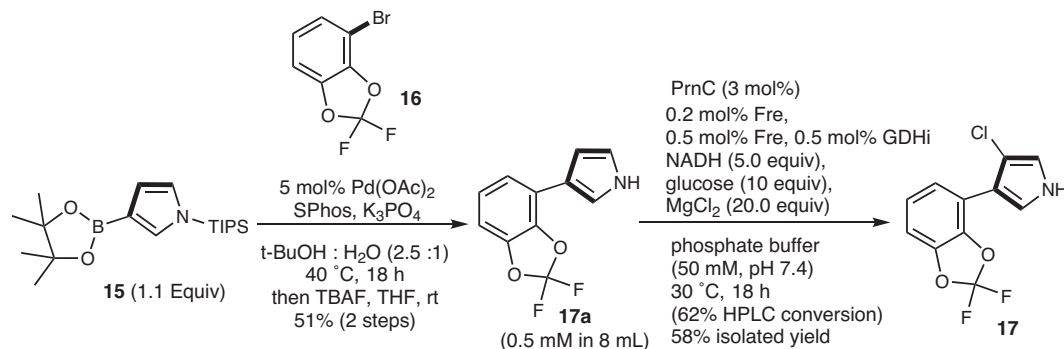7, 10% glycerol.

Fig. 6 Synthesis of a chlorinated analog 17 of Fludioxonil. Our chemoenzymatic approach involves a palladium-mediated Suzuki coupling to construct the biaryl pyrrolic precursor **17a**, which then undergoes a late-stage site-selective chlorination to yield an analogue of the Fludioxonil fungicide, **17**.

Homology and docking models of PrnC. Homology modeling was performed in Modeler v10 program and generated 100,000 homology models based on the template crystal structure of halogenase PltM (PDB code: 6BZA) whose sequence was aligned with that of PrnC (Fig. S5) with the sequence identity of 36%. Subsequently, all the models were subjected to the backbone, sidechain, and loop optimization, and then the top 50 optimized models with the lowest DOPE scores were used for the subsequent molecular docking with the native substrate (**1**) by using GOLD v2018 program with the optimal docking parameters (the binding pocket is defined by the ligand copied from the crystal structure of 6BZA with the spherical radius of 8.0 Angstrom; scoring function is GoldScore; population size is 500; the number of operations is 500000; number of island is 10; crossover frequency is 95%; mutation frequency is 95%; migration frequency is 20%; the number of output docking solutions is 3), which afforded 150 docking solutions in total.

In order to achieve an optimal model for the PrnC/**1a** complex, 150 docking solutions were subjected to further inspection based on the following two criteria: (1) there must be a nearby lysine residue stretching towards the pyrrole ring in **1a**, because a lysine is required for the catalysis; (2) the hydrogen on the pyrrolic nitrogen of **1a** should form a hydrogen bond with the hydrogen acceptor of a residue to stabilize the intermediate during the catalytic reaction. After the manual and visual examination, a reasonable model of PrnC/**1a** complex, which satisfied both criteria above, was harvested and shown in Fig. 3 and Fig. S7a (in Supplementary Materials and Methods, section 1.8.2).

Cryo-EM. Single-particle Cryo-EM sample vitrification. Purified PrnC (in the buffer of 50 mM sodium phosphate pH 7, 10% glycerol) was concentrated using a 10-kDa molecular weight cutoff filter concentrator to 2.8 mg mL^{-1}. A total of 2 µL of sample was added to a glow discharged (JEOL DATAM HDT-400) was applied onto the copper side of nanofabricated gold grids[58] and blotted using filter paper on one side for 2 s using the Leica GP plunger system before plunging immediately into liquid ethane for vitrification. The plunger was operating at 5 °C with >80% humidity to minimize evaporation and sample degradation.

Single-particle Cryo-EM Data Collection and Processing. Images were recorded on a Titan Krios electron microscope (FEI) equipped with a BioQuantum K3 direct detector with energy filter operating at 0.8341 Å per pixel in electron counting mode using the SerialEM software package[59]. Pixel size was calibrated using apoferritin. Slit width was 20 eV. Data collection was performed using a dose of ~64 e$^-$ Å^{-2} across 48 frames (125 ms per frame) at a dose rate of ~7.4 e$^-$ pix^{-1} s^{-1}, using a set defocus range of -0.8 to -1.8 µm. In all, 100-µm objective aperture was used. A total of 2763 micrographs were recorded

over one day using an image beam shift data collection strategy[60]. Data processing was done using cryoSPARC 2.0[61]. Patch motion correction was applied to the movies, and patch CTF estimation was done. Blob picker was used to pick out the particles, and 2D classification was done to remove junk particles. Thereafter, particles looking like protein were put through ab initio with multiple models to further clean up the good particles. A final stack of 65,768 good particles was put through local motion correction, and then homogenous refinement to obtain a resolution of around 9 Å. The homology models of PrnC were docked into this final map using UCSF Chimera[62].

Analytical scale biotransformations. In a solution containing the pyrrolic derivative starting material (0.5 mM), MgCl$_2$/MgBr$_2$ (10 mM), glucose (5.0 mM), FAD (1.0 µM), NT-11 PrnC (12.5 µM), Fre (2.5 µM) and Gdhi (2.5 µM) in 10 mM potassium phosphate buffer, NADH (2.5 mM) was added to a total volume of 200 µL. After an overnight incubation of 30 °C and orbital shaking at 350 rpm, reactions were quenched with an equivalent volume of MeOH, pelleted by centrifugation (15000 rpm for 10 min), and the supernatant analyzed by HPLC-MS using the analytical HPLC method.

Determination of kinetic parameters for NT-11 PrnC and PrnC mutant assay. Kinetic analysis of PrnC (2.5 µM) activity against MDA was performed over a 5-250 µM substrate concentration range. The assay reaction was supplemented with Fre (2.5 µM), FAD (1 µM), and MgCl$_2$ (10 mM) in 20 mM Tris buffer, pH 7.4. NADH (2.5 mM) was added last for reaction initiation. The products formed were measured at 120, 300, and 600 s via a Kinetex XB-C18 reversed-phased column (2.6 µm, 150 × 4.6 mm) on a Shimadzu LC-20AD HPLC. Absorbance at $\lambda = 254$ nm was used to monitor product formation during an isocratic flow rate of 0.6 mL/min (50% MeCN/H$_2$O + 0.1% TFA) over 10 min. Kinetic parameters were determined by nonlinear fitting of a Michaelis-Menten curve using the GraphPad Prism software. Activity assays for PrnC mutants were performed with the respective over-expressed protein variant. Wild-type PrnC enzyme was used as a positive control and reaction conditions were similar to the 18 h assay method described above.

General HPLC and LC-MS methods. General considerations. Spectroscopic grade solvents were purchased from Sigma Aldrich. Low-resolution LC-MS spectra were recorded on an Agilent LCMS machine with dual MM-APCI-ES. High-resolution mass spectra (HRMS) were recorded on an Agilent ESI-TOF mass spectrometer at 3500 V emitter voltage. Exact m/z values are reported in Daltons.

Semi-Preparative HPLC method. 900 µL of the crude mixture dissolved in H_2O/MeCN was injected into a Phenomenex Jupiter® semi-preparative C18 HPLC column (90 Å, 5 µm packing, 250 × 10 mm) and purified using reverse phase chromatography. Gradient starting conditions of 5% MeCN/H_2O (+ 0.1% Formic acid) to 25% MeCN/H_2O over 10 min, followed by 25% MeCN/H_2O into 50% MeCN/H_2O over 20 min, followed by 50% MeCN/H_2O into 75% MeCN/H_2O over 10 min, followed by 75% MeCN/H_2O into 95% MeCN/H_2O over 5 min, followed by a hold at 95% MeCN/H_2O for 5 min. Column condition was equilibrated back to starting conditions over 2 min post-run. Flow rates were kept constant at at 3 mL/min. UV absorbance was monitored at 220 nm, 254 nm, and 280 nm.

Analytical HPLC Method. 10 µL of the supernatant injected onto SecurityGuard™ column (KJ0-4282) with a (4.0 mm × 3.0 mm) guard cartridge before separation using a Phenomenex Gemini® C18 analytical column (5 µm packing, 150 mm × 4.6 mm). Gradient starting conditions of 5% MeCN/H_2O (+ 0.1% Formic acid) were held for 1 min before development into 50% MeCN/H_2O over 3 min, followed by development into 95% MeCN/H_2O over 3 min. 95% MeCN/H_2O was held for 1 min before equilibration back to starting conditions over 1 min. Starting conditions were held for 1 min followed by another 2 min post-run. Flow rates were kept constant at 1 mL/min. Column temperature was kept constant at 30 °C. UV absorbance was detected at 220 nm, 254 nm, and 210 nm throughout the run.

General LC-MS Method. 10 µL of the supernatant was separated using the appropriate analytical HPLC method described above. Detection was performed using an Agilent® single quadrupole LC/MSD system.

Data availability

Electronic Supplementary Supplementary Information contains comprehensive experimental procedures for the synthesis and characterization of new compounds as well as kinetic studies and LCMS analysis. Supplementary Data 1 contains the ^{1}H, ^{13}C, and 2D NMR data spectra of the isolated new compounds.

References

1. Singh, N. et al. Recent progress in the total synthesis of pyrrole-containing natural products (2011–2020). Organic Chemistry Frontiers 8, 5550–5573 (2021).
2. Bhardwaj, V., Gumber, D., Abbot, V., Dhiman, S. & Sharma, P. Pyrrole: a resourceful small molecule in key medicinal hetero-aromatics. RSC Adv. 5, 15233–15266 (2015).
3. Gholap, S. S. Pyrrole: An emerging scaffold for construction of valuable therapeutic agents. Eur. J. Med. Chem. 110, 13–31 (2016).
4. Huffman, J. W. & Padgett, L. W. Recent developments in the medicinal chemistry of cannabimimetic indoles, pyrroles and indenes. Curr. Med. Chem. 12, 1395–1411 (2005).
5. Wang, M.-Z. et al. Design, synthesis and antifungal activities of novel pyrrole alkaloid analogs. Eur. J. Med. Chem. 46, 1463–1472 (2011).
6. Santo, R. D. et al. Pyrrolnitrin and related pyrroles endowed with antibacterial activities against Mycobacterium tuberculosis. Bioorg. Med. Chem. Lett. 8, 2931–2936 (1998).
7. Kuznietsova, H. et al. Pyrrole derivatives as potential anti-cancer therapeutics: synthesis, mechanisms of action, safety. J. Drug Target. 28, 547–563 (2020).
8. Hagfeldt, A., Boschloo, G., Sun, L., Kloo, L. & Pettersson, H. Dye-sensitized solar cells. Chem. Rev. 110, 6595–6663 (2010).
9. Nishide, H. & Oyaizu, K. Toward flexible batteries. Science 319, 737–738 (2008).
10. Rather, I. A., Wagay, S. A., Hasnain, M. S. & Ali, R. New dimensions in calix[4]pyrrole: the land of opportunity in supramolecular chemistry. RSC Adv. 9, 38309–38344 (2019).
11. Eliseenko, S. S. & Liu, F. Switchable pyrrole-based hydrogen bonding motif in enantioselective trifunctional organocatalysis. Tetrahedron 75, 518–526 (2019).
12. Ji Ram, V., Sethi, A., Nath, M. & Pratap, R. in The Chemistry of Heterocycles, eds. V. Ji Ram, A. Sethi, M. Nath and R. Pratap, Elsevier, 2019, https://doi.org/10.1016/B978-0-08-101033-4.00005-X, pp. 149–478.
13. Young, I. S., Thornton, P. D. & Thompson, A. Synthesis of natural products containing the pyrrolic ring. Nat. Prod. Rep. 27, 1801–1839 (2010).
14. Veisi, H., Ghorbani-Vaghei, R. & Zolfigol, M. A. Recent progress in the use ofN-halo compounds in organic synthesis. Org. Preparations Proced. Int. 43, 489–540 (2011).
15. in Organic Chemistry: A Series of Monographs, eds. R. A. Jones and G. P. Bean, Academic Press, 1977, vol. 34, pp. 115–207.
16. Gkotsi, D. S., Dhaliwal, J., McLachlan, M. M. W., Mulholand, K. R. & Goss, R. J. M. Halogenases: powerful tools for biocatalysis (mechanisms applications and scope). Curr. Opin. Chem. Biol. 43, 119–126 (2018).
17. Latham, J., Brandenburger, E., Shepherd, S. A., Menon, B. R. K. & Micklefield, J. Development of halogenase enzymes for use in synthesis. Chem. Rev. 118, 232–269 (2018).
18. Agarwal, V. et al. Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. Chem. Rev. 117, 5619–5674 (2017).
19. Weichold, V., Milbredt, D. & van Pée, K.-H. Specific enzymatic halogenation—from the discovery of halogenated enzymes to their applications in vitro and in vivo. Angew. Chem. Int. Ed. 55, 6374–6389 (2016).
20. Crowe, C. et al. Halogenases: a palette of emerging opportunities for synthetic biology–synthetic chemistry and C–H functionalisation. Chem. Soc. Rev. 50, 9443–9481 (2021).
21. Andorfer, M. C. & Lewis, J. C. Understanding and improving the activity of flavin-dependent halogenases via random and targeted mutagenesis. Annu. Rev. Biochem. 87, 159–185 (2018).
22. Payne, J. T., Andorfer, M. C. & Lewis, J. C. Methods in enzymology. Methods Enzymol. 575, 93–126 (2016).
23. Büchler, J., Papadopoulou, A. & Buller, R. Recent advances in flavin-dependent halogenase biocatalysis: sourcing, engineering, and application. Catalysts 9, 1030 (2019).
24. Gkotsi, D. S. et al. A marine viral halogenase that iodinates diverse substrates. Nat. Chem. 11, 1091–1097 (2019).
25. Mori, S., Pang, A. H., Thamban Chandrika, N., Garneau-Tsodikova, S. & Tsodikov, O. V. Nat. Commun. 10, 1255 (2019).
26. Latham, J. et al. Integrated catalysis opens new arylation pathways via regiodivergent enzymatic C–H activation. Nat. Commun. 7, 11873 (2016).
27. Sana, B. et al. Engineered RebH halogenase variants demonstrating a specificity switch from tryptophan towards novel indole compounds. ChemBioChem. 22, 2791–2798 (2021).
28. Menon, B. R. K. et al. RadH: a versatile halogenase for integration into synthetic pathways. Angew. Chem. Int. Ed. 56, 11841–11845 (2017).
29. Wynands, I. & van Pée, K. H. FEMS Microbiol. Lett. 237, 363–367 (2004).
30. Hammer, P. E., Hill, D. S., Lam, S. T., Pée, K. H. V. & Ligon, J. M. Four genes from Pseudomonas fluorescens that encode the biosynthesis of pyrrolnitrin. Appl. Environ. Microbiol. 63, 2147–2154 (1997).
31. Andorfer, M. C. et al. Understanding flavin-dependent halogenase reactivity via substrate activity profiling. ACS Catal. 7, 1897–1904 (2017).
32. Fisher, B. F., Snodgrass, H. M., Jones, K. A., Andorfer, M. C. & Lewis, J. C. Site-selective C–H halogenation using flavin-dependent halogenases identified via family-wide activity profiling. ACS Cent. Sci. 5, 1844–1856 (2019).
33. Agarwal, V. et al. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. Nat. Chem. Biol. 10, 640–647 (2014).
34. Heide, L. et al. Use of a halogenase of hormaomycin biosynthesis for formation of new clorobiocin analogues with 5-chloropyrrole moieties. ChemBioChem. 9, 1992–1999 (2008).
35. Yamanaka, K., Ryan, K. S., Gulder, T. A. M., Hughes, C. C. & Moore, B. S. Flavoenzyme-catalyzed atropo-selective N,C-bipyrrole homocoupling in marinopyrrole biosynthesis. J. Am. Chem. Soc. 134, 12434–12437 (2012).
36. Jaremko, M. J., Lee, D. J., Opella, S. J. & Burkart, M. D. Structure and substrate sequestration in the Pyoluteorin type II peptidyl carrier protein PltL. J. Am. Chem. Soc. 137, 11546–11549 (2015).
37. Dorrestein, P. C., Yeh, E., Garneau-Tsodikova, S., Kelleher, N. L. & Walsh, C. T. Dichlorination of a pyrrolyl-S-carrier protein by FADH2-dependent halogenase PltA during pyoluteorin biosynthesis. Proc. Natl Acad. Sci. USA 102, 13843–13848 (2005).
38. Arima, K., Imanaka, H., Kousaka, M., Fukuta, A. & Tamura, G. Pyrrolnitrin, a new antibiotic substance, produced by Pseudomonas. Agric. Biol. Chem. 28, 575–576 (1964).
39. Gorman, M. & Lively, D. H., in Biosynthesis, eds. Gottlieb, D. & Shaw, P. D. Springer Berlin Heidelberg, Berlin, Heidelberg, 1967, https://doi.org/10.1007/978-3-662-38441-1_32, pp. 433–438.

40. Zhou, P., Mocek, U., Siesel, B. & Floss, H. G. Biosynthesis of pyrrolnitrin. Incorporation of 13C,15N double-labelled D- and L-tryptophan. *J. Basic Microbiol.* **32**, 209–214 (1992).

41. Hölzer, M., Burd, W., Reißig, H.-U. & Pée, K.-H. V. *Adv. Synth. Catal.* **343**, 591–595 (2001).

42. Keller, S. et al. Purification and partial characterization of tryptophan 7-halogenase (PrnA) from Pseudomonas fluorescens. *Angew. Chem. Int. Ed.* **39**, 2300–2302 (2000).

43. Nguyen, T. K. M., Ki, M. R., Son, R. G. & Pack, S. P. The NT11, a novel fusion tag for enhancing protein expression in *Escherichia coli. Appl. Microbiol. Biotechnol.* **103**, 2205–2216 (2019).

44. Phintha, A. et al. Dissecting the low catalytic capability of flavin-dependent halogenases. *J. Biol. Chem.* **296**, 100068 (2021).

45. Zhang, Y. et al. Aryl C-H iodination: are there actual flavin-dependent iodinases in nature. *Sci. China Chem.* **64**, 1730–1735 (2021).

46. Mondal, D., Fisher, B. F., Jiang, Y. & Lewis, J. C. Flavin-dependent halogenases catalyze enantioselective olefin halocyclization. *Nat. Commun.* **12**, 3268 (2021).

47. Prakinee, K. et al. Mechanism-guided tunnel engineering to increase the efficiency of a flavin-dependent halogenase. *Nat. Catal.* **5**, 534–544 (2022).

48. Frese, M., Guzowska, P. H., Voß, H. & Sewald, N. Regioselective enzymatic halogenation of substituted tryptophan derivatives using the FAD-dependent halogenase RebH. *ChemCatChem.* **6**, 1270–1276 (2014).

49. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* **46**, 1284–1292 (2007).

50. Dong, C. et al. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309**, 2216–2219 (2005).

51. Yin, L. Crystal structure determination of the halogenase CtcP from Streptomyces aureofaciens. *Acta Crystallogr. Sect. F* **78**, 270–275 (2022).

52. Chua, E. Y. D. et al. Better, faster, cheaper: recent advances in cryo–electron microscopy. *Annu. Rev. Biochem.* **91**, 1–32 (2022).

53. Zhu, X. et al. Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J. Mol. Biol.* **391**, 74–85 (2009).

54. Moritzer, A.-C. et al. Structure-based switch of regioselectivity in the flavin-dependent tryptophan 6-halogenase Thal. *J. Biol. Chem.* **294**, 2529–2542 (2019).

55. Minges, H. et al. Targeted enzyme engineering unveiled unexpected patterns of halogenase stabilization. *ChemCatChem.* **12**, 818–831 (2020).

56. Mathiyazhagan, A. D. & Anilkumar, G. Recent advances and applications of p-toluenesulfonylmethyl isocyanide (TosMIC). *Org. Biomol. Chem.* **17**, 6735–6747 (2019).

57. Kumar, K. TosMIC: a powerful synthon for cyclization and sulfonylation. *ChemistrySelect* **5**, 10298–10328 (2020).

58. Marr, C. R., Benlekbir, S. & Rubinstein, J. L. Fabrication of carbon films with ~500 nm holes for cryo-EM with a direct detector device. *J. Struct. Biol.* **185**, 42–47 (2014).

59. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

60. Schorb, M., Haberbosch, I., Hagen, W. J., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. methods* **16**, 471–477 (2019).

61. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. methods* **14**, 290–296 (2017).

62. Pettersen, E. F. et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J. computational Chem.* **25**, 1605–1612 (2004).

63. Ayats, C., Soley, R., Albericio, F. & Álvarez, M. Synthesis of the pyrrolo[2,3-c] carbazole core of the dictyodendrins. *Org. Biomol. Chem.* **7**, 860–862 (2009).

Acknowledgements

Author contributions

Y.H.L., E.L.A., and F.T.W. conceptualized the project and analyzed the experimental data. G.P. and Y.H.L. wrote the manuscript with inputs from all other authors. G.P., Y.L.G., S.Y., and J.W. ran the bioassays and chemistry analytical experiments. T.T., L.L.T., E.T., and J.B. performed protein purification and fermentation experiments. F.L. and Y.Z.T. performed the computational docking studies. C.J.X.T. and Y.Z.T. performed the cryo-EM experiments and analysis. E.L.A., F.T.W., H.Z., and Y.H.L. carried out funding acquisition and supervision.

Competing interests

G.P., T.T., F.T.W., E.L.A and Y.H.L are inventors for PCT/SG2023/050347, which includes some of the data disclosed in this paper.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42004-023-01083-1.

Correspondence and requests for materials should be addressed to Fong Tian Wong, Ee Lui Ang or Yee Hwee Lim.

Peer review information *Communications Chemistry* thanks Rebecca Goss and the other, anonymous, reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.