Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Halogenases: structures and functions

Hannes Ludewig[1,2,4], Samuel Molyneux[4], Scarlet Ferrinho, Kai Guo, Rosemary Lynch, Danai S Gkotsi[3] and Rebecca JM Goss

Over 5000 halogenated natural products have been reported so far, many of these arising from the marine environment. The introduction of a halogen into a molecule can significantly impact its bioavailability and bioactivity. More recently enzymatic halogenation has been used to enable late stage functionalisation through site-selective halogenation and cross-coupling. Halogenases are becoming increasingly valued tools. This review outlines the various classes of halogenases that have been discovered, and examines these from both a structural and a mechanistic perspective, reflecting upon the many recent advances in halogenase discovery.

Address
School of Chemistry, University of St Andrews, North Haugh, St Andrews, Fife, KY16 9ST, UK

Corresponding author: Goss, Goss (rjmg@st-andrews.ac.uk)
[1] Present address: Department of Biochemistry, Brandeis University, Waltham, MA 02454, USA.
[2] Present address: Howard Huges Medical Institute, Brandeis University, Waltham, MA 02454, USA.
[3] Present Address: School of Molecular and Cellular Biology, University of Leeds, Leeds, Yorkshire, LS3 9JT, UK.
[4] These authors contributed equally.

Current Opinion in Structural Biology 2020, **65**:51–60

This review comes from a themed issue on **Catalysis and regulation**

Edited by **James H Naismith** and **Emily Parker**

For a complete overview see the Issue and the Editorial

Available online 30th June 2020

https://doi.org/10.1016/j.sbi.2020.05.012

Introduction

Halogenation of molecules can significantly impact their activity and bioavailability [1,2] and is a strategy employed by both nature, and in bioactive molecule design, with 27% of small molecule drugs and more than 80% of agrochemicals bearing a halogen [3,4]. Additionally, incorporation of Cl, Br, or I provides chemically reactive and orthogonal handles for selective modification through cross-coupling chemistry [5]. Today, over 5000 halometabolites, predominantly chlorinated and brominated metabolites, have been identified [6•]. Though more rare, approximately 100 iodinated natural metabolites and 6–13 natural, fluorinated metabolites have been found [6•]. This plethora of halometabolites show exciting structural diversity indicating the operation of a large variety of powerful enzymes [7].

Though initially biosynthetic halogenation was considered an artefact, the first enzyme capable of halogenation was discovered by Lowell Hager *et al.* in 1959 [8••]. This enzyme was named chloroperoxidase (CPO), a haloperoxidase, from the fungus *Calariomyces fumago* [9]. For several decades the haloperoxidases remained the only halogenases known [9]. However, the turn of the millennium corresponded to a significant acceleration in this research field, largely fuelled by investigations into the biogenesis of natural products from bacteria, fungi and plants. In the last 5 years, bioinformatics led approaches have enabled the identification of halogenases, even from viruses [10••,11••].

Halogenases may be broadly classified by their mechanisms: electrophilic, nucleophilic, or radical, and may be further subdivided according to their catalytic species. The vast majority of halogenases studied so far employ electrophilic halogenation, enabling incorporation of Cl, Br, and I [7]; in recent years increasing numbers of radical halogenases have been identified [12,13•,14•]. Nucleophilic halogenases remain the most rare [15,16•].

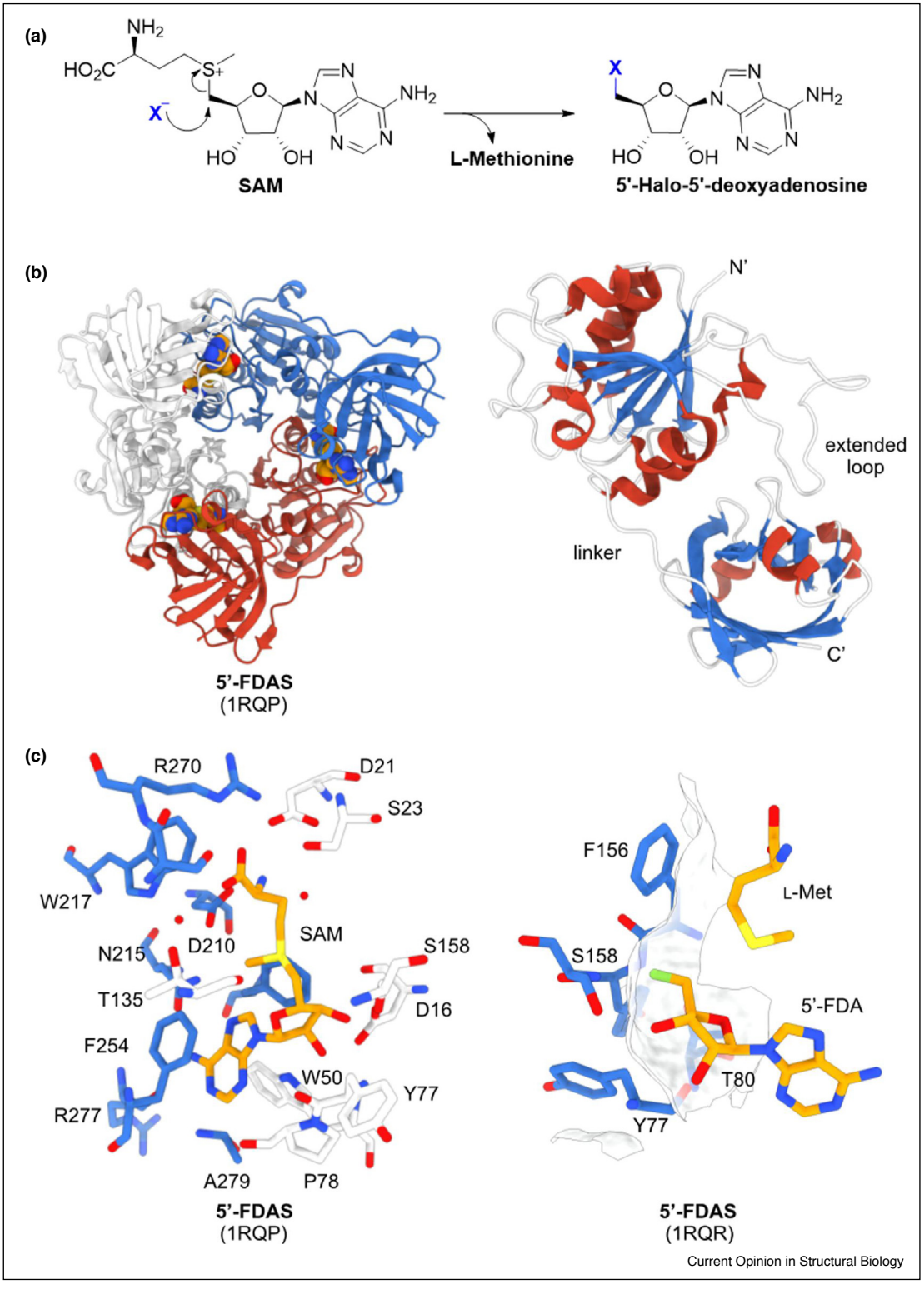
Here we overview the enzymatic mechanisms for each class of halogenase, reflecting upon the structures of the enzymes that mediate these chemistries.

Nucleophilic halogenases

A small handful of nucleophilic halogenases are known so far, the first of these, and the most well studied is the fluorinase or 5′-fluoro-5′-deoxyadenosine synthase (5′-FDAS) from *Streptomyces cattleya* [17]. Though fluorine is the most abundant halogen in the earth's crust, and the 13th most abundant element, few natural organofluorine compounds exist. This is perhaps dominated by three factors: the poor solubility of fluoride salts, the high enthalpy of hydration of fluoride (490 kJ mol^{-1}) and fluoride's electronegativity, rendering it a poor nucleophile.

5′-FDAS, is responsible for the first step in the biogenesis of fluoroacetate and 4-fluoro-L-threonine from *Streptomyces cattleya* [17]. It mediates the nucleophilic attack of fluoride onto S-adenosyl-L-methionine (SAM), generating 5′-deoxy-5′-fluoroadenosine (5′-FDA) and L-methionine (Figure 1a) [15]. Overcoming the substantial

Figure 1



Nucleophilic halogenation scheme and structures of 5′-FDAS. **(a)** Transformation of SAM and halide to 5′-XDA and L-Methionine, catalysed by a nucleophilic halogenase. **(b)** Cartoon representation of asymmetric unit content of 5′-FDAS crystal structure (PDB: 1RQP) in complex with SAM (spheres; carbon, oxygen, sulphur and nitrogen atoms are coloured orange, light red, yellow and dark blue, respectively). Monomers are coloured in blue, red and white (left) as well as cartoon representation of 5′-FDAS monomer (right), α-helices, β-strands and loops are coloured in blue, red and white, respectively. **(c)** (left) Stick representation of SAM binding site in 5′-FDA (1RQP). Carbon atoms belonging to each of two interfacing

enthalpy of hydration to unmask a naked fluoride nucleophile, is key.

A 1.8 Å crystal structure of 5′-FDAS (PDB: 1RQP), revealed a novel fold of Pfam superfamily PF01887 [15]. The N-terminal and C-terminal domains are connected by a 15-residue linker (Figure 1b). The N-terminal domain consists of a central ß-sheet sitting in between seven α-helices and an extensive loop connecting the last two ß-strands (see Figure 1b). The C-terminal domain is comprised of two anti-parallel ß-sheets sandwiched by three short helices (Figure 1b). Within the trimer found in the asymmetric unit (ASU) of the crystal structure, intermolecular domain interfaces are formed between N-terminal domains of neighbouring monomers (relying on the aforementioned extended loop (98–114)) as well as contacts between N-terminal domains and C-terminal domains of neighbouring monomers. A dimer of trimers may be seen within the crystal structure, consistent with the hexameric solution state suggested by analytical size exclusion chromatography [15]. Each of the three N-terminal to C-terminal domain interfaces harbour SAM, fully encapsulating the substrate (see Figure 1b). The ribose ring may be seen to adopt an unusual, high energy, eclipsed conformation (Figure 1c). Crystals of 5′-FDAS grown in the presence of potassium fluoride and SAM contained the product 5′-FDA (shown to reside with a relaxed ribose ring conformation) and L-methionine located within the SAM binding site (PDB: 1RQR; see Figure 1c). A small pocket, identified as the fluoride binding site, comprises of residues Y77, T80, F156 and S158 (Figure 1c), giving insight to the way in which the hydration sphere of the fluoride anion might be displaced [15].

Hydroxyl and amide moieties of the S158 are proposed to form two hydrogen bonds to fluoride, crucial for lowering the energetic cost for its necessary dehydration in order to produce a more reactive nucleophile. The *anti* positioning of fluoride with regards to the C5′ and -S bond is consistent with an S_N2 mechanism, and the observed inversion of configuration has been described by O'Hagan *et al.* [18]. Further detailed mechanistic studies demonstrate substrate inhibition by SAM, and requirement for fluoride binding to 5′-FDAS in advance of SAM, in order to generate a productive complex [19]. Dehydration is proposed to reach completion upon binding of SAM. This results in a more nucleophilic fluoride, potentially stabilised by T80 or the positively charged sulphur atom from SAM, optimally orientated for the S_N2 reaction, as suggested by the *in silico* studies of Vincent *et al.* [20].

The mechanistically related SalL, involved in the biosynthesis of the chlorinated metabolite salinosporamide A, cannot utilise fluoride. Whilst chloride is the native nucleophile, bromide and iodide, are shown to be poor but processable substrates *in vitro* (K_m of 45, 150 and 260 mM, respectively) [16•]. Structural investigations by Moore *et al.* [16•] revealed an identical monomeric fold compared to 5′-FDAS and a trimeric oligomeric state in both the crystal structure and in solution. Crystal structures of SalL show an enlarged halide binding pocket in which a glycine is in place of the serine (S158) of 5′-FDAS: absence of this serine results in the reduced ability to dehydrate the nucleophile and the inability to process fluoride (PDB: 2Q6I). The inability to exclude water, is exemplified by the presence of the proximal water molecule in the crystal structure [16•].

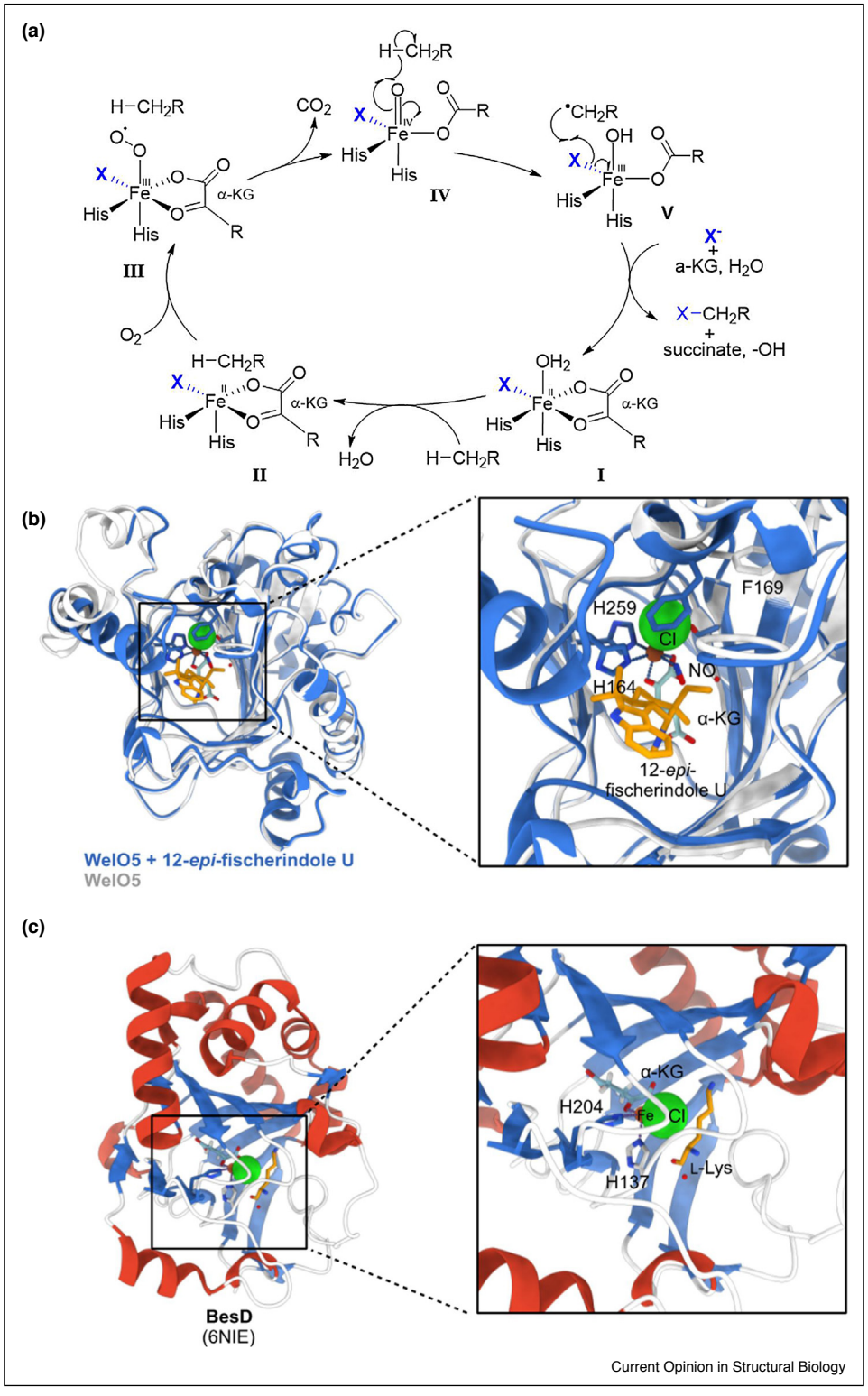
Radical halogenases

All known enzymatic radical halogenation reactions are catalysed by non-heme-iron α-ketoglutarate (KG)-dependent enzymes; proteins which are able to selectively halogenate an unactivated aliphatic carbon centre. All Fe(II)/α-KG halogenation proceeds by the same general mechanism. The enzyme resting state (Figure 2a. **I**) generally consists of two active site histidine residues, a bidentate α-KG, a halide anion, and a weakly bound water (which is displaced by substrate (Figure 2a. **II**)) coordinating the Fe(II) centre. Dioxygen addition to the iron centre (Figure 2a. **III**) is followed by a decarboxylative event between the bound oxygen and α-KG, which leads to the formation a high energy Fe(IV) = O intermediate (Figure 2a. **IV**) [22]. A radical deprotonation of the substrate C–H bond ensues (Figure 2a. **V**). The resultant alkyl radical reacts with the bound halide, generating the alkyl halide. Before the discovery of WelO5, [21] all radical halogenases required substrates to be presented bound to a separate protein.

WelO5 is responsible for halogenation of 12-epi-fischerindole U in *Hapalosiphon welwischii*, inspection of its structure provides key insights into how these halogenases can bind free molecules. It was the first Fe(II)/α-KG enzyme structure to be solved (PDB: 5IQT) with its substrate bound in the active site [12]. WelO5's 2.4 Å crystal structure was solved to show three monomers in the ASU, with the protein acting as a monomer in solution. Overall the structure consists of the 8 stranded β-sandwich topology, or the β jelly roll motif of Fe(II)/α-KG proteins in the cupin superfamily, which is surrounded by external α helicies. The Fe(II)/α-KG cofactor was shown to sit within this sandwich, as with other cupin enzymes. A notable difference between WelO5 and halogenases in which the substrate is protein bound, is the presence of an external α helix in the C-terminus (215–232), which forms additional interactions

monomers are coloured in blue or white respectively. Carbon atoms belonging to SAM are coloured orange. Oxygen, sulphur and nitrogen atoms are coloured light red, yellow and dark blue, respectively; (right) fluoride binding pocket (surface; translucent white) in 5′-FDAS product complex structure (PDB: 1RQR) represented as sticks. 5′-FDAS carbon atoms are coloured in blue. Carbon atoms belonging to L-Met and 5′-FDA are depicted in orange. Oxygen, sulphur, fluorine and nitrogen atoms are coloured light red, yellow, green and dark blue, respectively.

Figure 2



Mechanism and structures of Fe(II)/α-KG halogenases WelO5 and BesD. **(a)** General enzymatic mechanism for radical halogenation by Fe(II)/α-KG halogenases. **(b)** Structural overlay of WelO5 complex structures with and without 12-*epi*-fischerindole U (5IQS (coloured light grey) and 5IQT (coloured blue), respectively). Active site components are depicted as spheres (chloride) and sticks. α-KG carbons are coloured in turquoise and 12-*epi*-fischerindole U in orange. Oxygen, nitrogen, chloride and iron atoms are coloured light red, dark blue, green and brown, respectively. **(c)**

upon binding of the substrate and moves 9 Å closer to the active site, acting as a 'lid'. This helix is postulated to influence substrate acceptance (this helix differs most in sequence between the more promiscuous AmbO5, and more substrate-specific WelO5*) [23•]. Substrate binding also causes conformational change causing F169 to block the large site expected for substrate. Instead F169 sits over H164 in a hydrophobic pocket closed by this external helix. The loop (residues 75–82) moves in to form a H-bond from A82 nitrogen to NC carbon of the substrate, whilst an α helix methionine (M225) binds the substrate's NH (Figure 2b). These interactions position the free substrate over H164, such that the iron centre is 4.5 Å from the carbon to be chlorinated, but a considerable distance away from the bound chloride (5.7 Å). Subsequent rearrangement of ligands around iron in the solved WelO5 structure, mediated by O_2 binding, is postulated to position the oxo ligand such that chlorination is preferred to hydroxylation [12]. The NO bound structure (PDB: 5IQV) shows this ligand orientation. Outer sphere effects may also lead to specific halogenation activity: S189 is theorised to form interactions with the Fe(IV) = O oxygen, and a mutation to alanine has been shown to reduce halogenation selectivity, possibly as a consequence of altered ligand positioning. Interestingly a different orientation of α-KG is observed in free substrate radical halogenases, in which the C3–C4 bond points out of the plane of the α-KG, chloride ion, and histidine, to substrate bound variants [22]. This could explain the need for a ligand rearrangement in WelO5.

BesD from *Streptomyces cattleya*, is the most recent Fe(II)/α-KG halogenase to be structurally characterised. This 'cryptic' halogenase catalyses conversion of free lysine to 4-Cl-lysine, an intermediate in the biogenesis of propargyl amino acids [22]. Notably these radical halogenases come from the same bacterium as 5′-FDAS [24]. The solved 1.95 Å structure (PDB: 6NIE) revealed four BesD monomers in the ASU, the classic β sandwich, and an active site around iron are consistent with other Fe(II)/α-KGs [25••]. The most striking difference is the binding site of the lysine substrate: a small polar pocket containing multiple hydrogen bonds to the carboxylate of lysine (H134, W238, R74) and to the amine (E120, N219, T221, D140) (Figure 2c). This corresponds to substrate utilisation: free and polar lysine rather than a protein bound amino acid or the large hydrophobic hapalindole.

The positioning of substrate within BesD was again shown critical for activity; a mutation of H134A removes a H-bond between H134 and the lysine carbonyl, resulting in markedly more hydroxylation that halogenation [25••]. In a structure similar to WelO5 α helix lid, a tryptophan residue (W239) sits on the top of the bound lysine forming hydrophobic interactions and closing off the active site. However, the distinct α helix lid is itself absent. Second sphere interactions are also at play-N219 is postulated to control positioning of Fe(IV) = O (analogous to WelO5's S189) with N219A leading to decreased halogenation and increased hydroxylation activity [25••]. Ligand rearrangement allows for formation of Fe(IV) = O at this desired position, and is thought to proceed by binding of O_2 followed by α-KG decarboxylation.

Electrophilic halogenation

Haloperoxidases

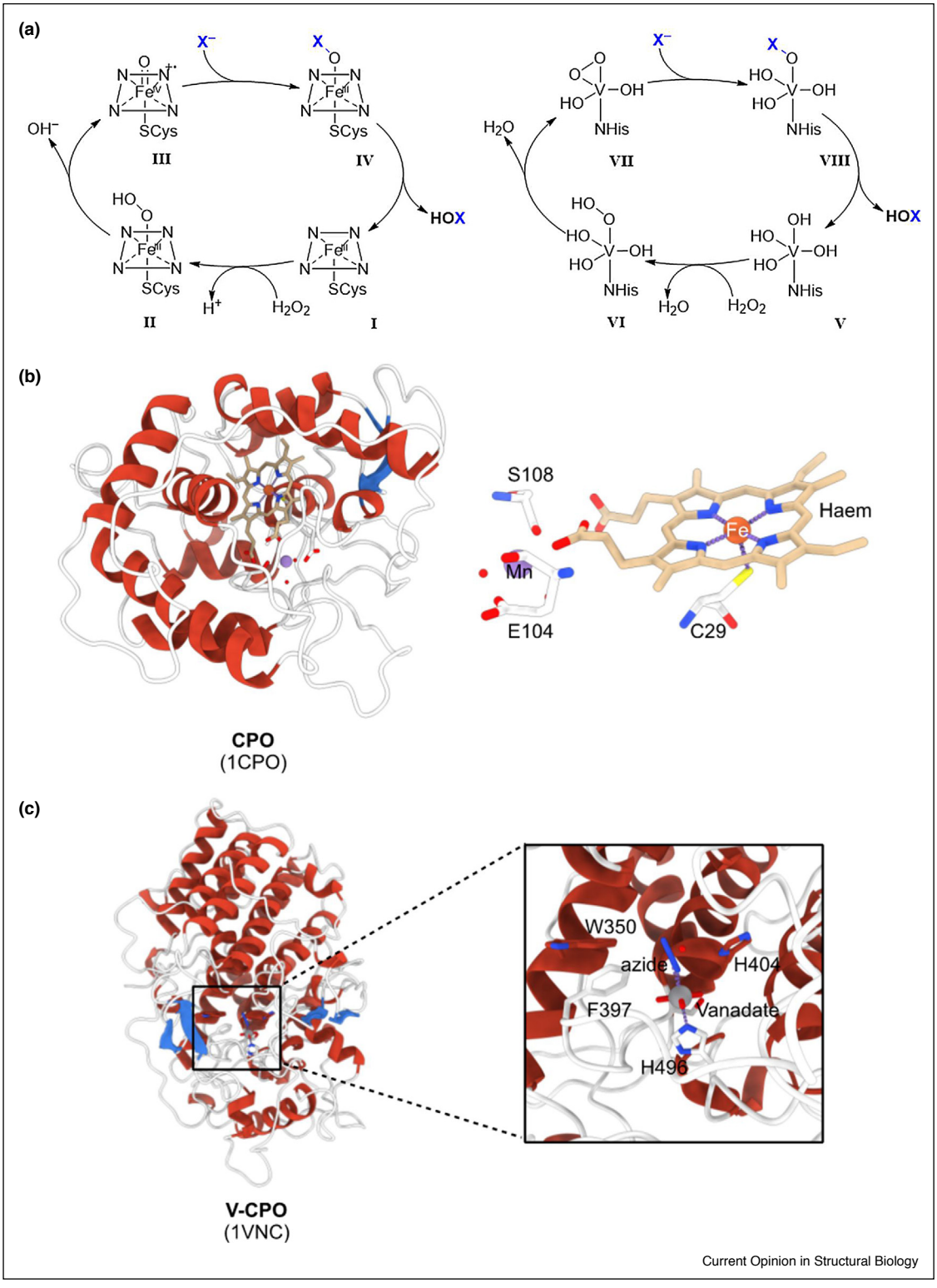
Haloperoxidases (HPOs) are often considered as two main classes, in accordance with the prosthetic groups utilised in the active site: haem iron haloperoxidases (such as CPO), and vanadium-dependent haloperoxidases (V-HPOs). The mechanism of halogenation employed by these enzymes is similar; both (generally) are considered to produce free hypohalous acid, HOX, which reacts with the substrate. In the case of the haem-dependent haloperoxidases, resting state haem (Figure 3a. **I**) reacts with H_2O_2, (Figure 3a. **II**). And, following loss of hydroxide, a Fe^{IV} -porphyrin compound (Figure 3a. **III**) is generated to oxidise the halide. (Figure 3a. **IV**) For the V-HPOs, a peroxovanadate species (Figure 3a. **VII**) enables halide oxidation (Figure 3a. **V–VIII**). Until the early 2000s these enzymes were considered to lack regioselective control, as they generate free HOX. Intriguingly, a small series of bacterial haloperoxidases are reported to display regiochemical control [26••,27,28], though the rational for this control within these outlying examples remains unclear [28].

The overall 1.90 Å structure of CPO (PDB: 1CPO), is highly helical, with the haem cofactor nestled between a proximal α helix and distal α helix, and a single monomer in the ASU. This 'distal' side of the haem cofactor is accessible from the surface of the protein by a 'narrow' channel and a 'wide' channel [29••]. The proximal helix contains a cysteine (C29) that acts as an axial ligand to the iron centre, and the distal containing resides for catalytic peroxide activation (E183, H105, and D106). Whilst a cysteine axial to the haem is often seen in cytochrome P450s, CPO is one of the few haem iron haloperoxidases, characterised so far, that has this feature. The glutamic acid residue (E183) acts as an acid-base catalyst, deprotonating a free peroxide molecule. H105 and D106 are proposed to assist with this deprotonation step [30].

Haem-dependent haloperoxidases have also been discovered in multiple mammalian groups; thyroid peroxidase

Cartoon representation of BesD (6NIE) and magnification of the active site. α-helices are coloured in red, β-strands are coloured in blue and loops are coloured in white. Active site components are depicted as spheres (chloride) and sticks. α-KG carbons are coloured in turquoise and lysine in orange. Oxygen, nitrogen, chloride and iron atoms are coloured light red, dark blue, green and brown, respectively.

Figure 3



Mechanism and structure of haem- and vanadium dependent- haloperoxidases. **(a)** Overview of haloperoxidase reaction mechanisms. **(b)** Cartoon representation CPO (PDB: 1CPO) and close up of the haem binding site; α-helices are coloured in red, β-strands are coloured in blue and loops are coloured in white. Active site components are depicted as spheres (iron) and sticks. Carbon atoms corresponding to protein are coloured according to their secondary structure and haem carbon atoms are coloured in light brown. Oxygen, nitrogen, iron, and manganese atoms are coloured light red, dark blue, orange, and violet respectively. **(c)** Cartoon representation V-CPO (PDB: 1VNC) and close up of the haem binding

(TPO), from *Homo sapiens*, is notable [31]. This enzyme catalyses iodination of tyrosine residues on thyroglobulin enabling production of thyroxine and triiodothyronine. The crystal structure of TPO is yet to be solved, however structures of human peroxidases able to utilise hypohalous acid, showing moderate sequence similarity to TPO, such as myeloperoxidase (MPO) and lactoperoxidase (LPO) have been studied extensively. Both MPO [32] (PDB: 3F9P) and LPO (PDB: 2R5L) [33] have structures dominated by 20 α-helices (19 in MPO) [34] and have haems accessible through a wide channel from the surface, as with CPO. The proteins, unlike CPO, are covalently linked, through esters, to their haem cofactor (E242 and D94 in MPO and E258 and D108 in LPO). Rather than a catalytic glutamic acid residue involved in peroxide deprotonation, these mammalian enzymes contain a catalytic histidine in the distal site, more reminiscent of haem peroxidases [35]. Both MPO and LPO have a conserved histidine as the axial ligand, again distinct from CPO's axial cysteine, and similar to the active site structure of classified peroxidases.

The core structure of *Curvularia inaequalis* V-CPO, 2.1 Å with one monomer per ASU, (PDB: 1VNC) [36••] consists mainly of tight α-helices, with two four-helix bundles as the main structural feature. This general structure is shared with vanadium-dependent bromoperoxidases (V-BPO) [37,38] and iodoperoxidases (V-IPO) [39]. This highly compact α helical structure is thought to be, in part, responsible for these enzymes' stability. It has also been postulated that the stabilising nature of vanadate and calcium ions play a key role [36••,40]. The reactive site of these enzymes is shown to lie at the bottom of a long, narrow channel in the core of the four-helix bundle with the vanadium bound as $V(V)O_3$ via a histidine residue (H496 in V-CPO). This histidine is essential for vanadate cofactor positioning, (H496A mutation leads to a complete loss of halogenation activity) [36••] and as such is conserved across V-HPOs. A catalytic histidine H404 residue is also conserved above the vanadate and, much like haem haloperoxidases, acts as an acid-base catalyst for deprotonation of incoming peroxide. The three oxygen atoms are located in the equatorial plane and the anionic $V(V)O_3$ group is stabilised by the surrounding hydrogen-bonding network of amino acids, this active site binding remains almost identical in all V-HPOs (Figure 3c).

Flavin-dependent halogenases (FDHs)

FDHs are generally subdivided into two groups: those acting on free substrates (variant a) and those acting on substrates tethered to a carrier protein, via a flexible phosphopantatheine linkage (variant b). Both contain a flavin-binding domain, akin to flavin-dependent oxygenases. The most well characterised FDHs act on tryptophan and often display a 'box' architecture of the flavin binding domain, and 'pyramid' architecture of the substrate binding domain [7]. $FADH_2$ is generated from FAD (Figure 4a. **IV**) by a concomitant flavin reductase [7]. Once generated, $FADH_2$ (Figure 4a. **I**) undergoes reaction with molecular oxygen forming flavin-C(4a)-hydroperoxide (Figure 4a. **II**) (in analogy to the monooxygenases). A halide anion then reacts with $FADH_2$ forming the corresponding hypohalous acid (HOX). The flavin binding site and the substrate binding site are separated by a 10 Å tunnel; hypohalous acid migrates down this tunnel to the substrate binding site, where a catalytic lysine residue is believed to form a haloamine species (Figure 4a. **V**) governing regioselectivity [7].

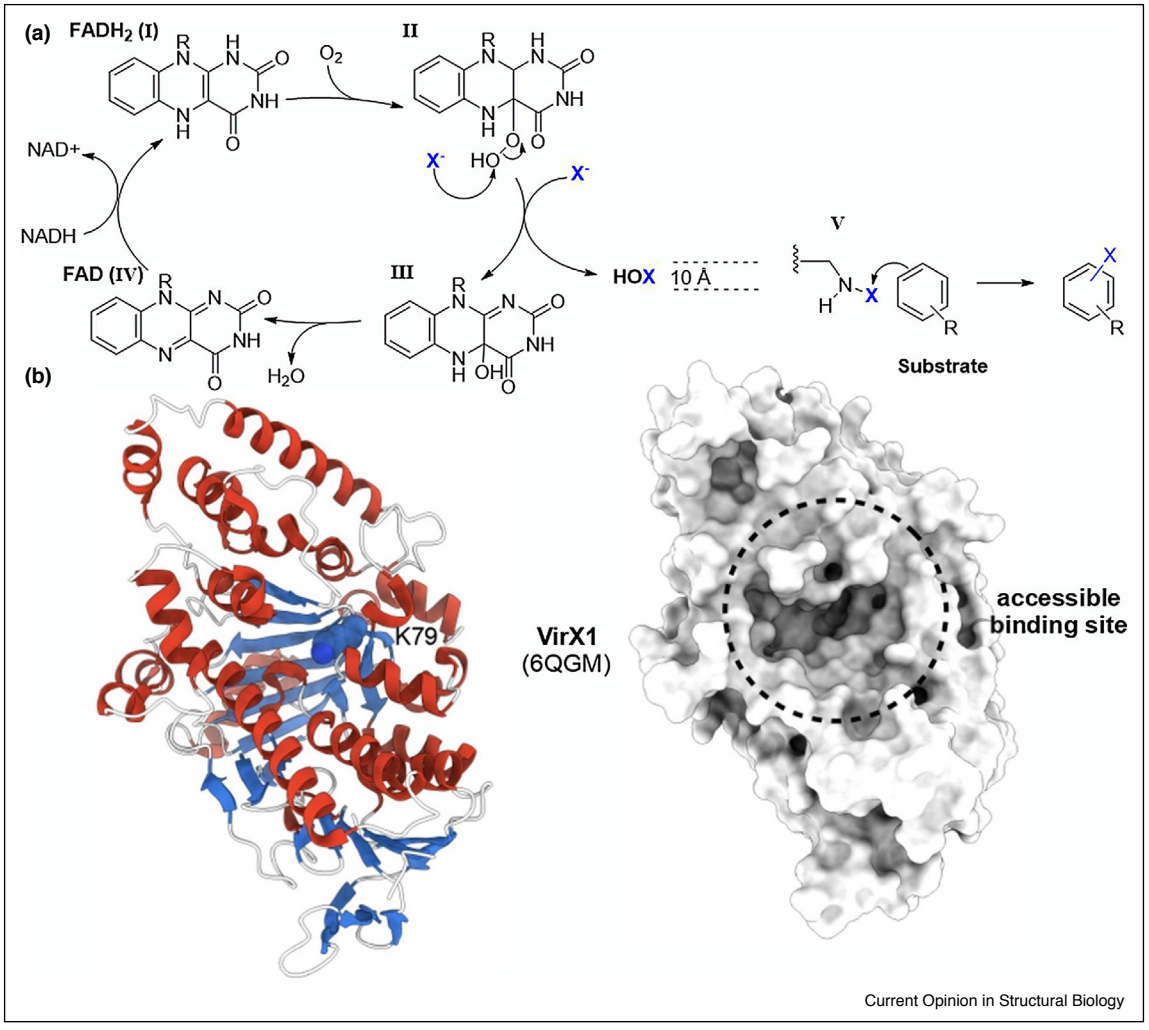
Regioselectivity in FDHs is governed by both electrophile and substrate positioning. PyrH is a tryptophan-5-halogenase, whilst ThaI and BorH are tryptophan-6-halogenases [41–43]. Analysis of the X-ray crystal structures of these three enzymes reveals that PyrH has a distinctly different tryptophan binding site, to that within the other two enzymes. As a result of altered indole positioning, in PyrH C-5 of indole is placed proximal to the catalytic lysine residue, postulated to direct the halogenating species. This results in the observed regioselectivity of halogenation [41–43]. Recent molecular dynamics simulations on PrnA and PyrH, carried out by Karabencheva-Christova and team [44], further corroborated the imposed binding orientation of tryptophan dictating regio-specificity.

Recently, the tryptophan FDH related halogenase BvrH was identified by meta-genome mining, it not only has a more accessible and open substrate binding site, but also accepts indole instead of tryptophan. The halide binding site for BvrH is identical to that observed in RebH [45]. However, the halogen preference by BvrH was for bromide over chloride (fluoride and iodide were not tested).

FDHs which accept tethered substrates, in contrast to free substrates, display a different substrate binding domain. This has been shown for PltA (PDB: 5DBJ) [46,47]. The dimeric PltA has a smaller C-terminal substrate domain compared to RebH, PrnA and PyrH, but the putative substrate binding site is sealed, indicating a requirement for domain re-arrangement for tethered substrate (pyrrolyl-*S*-PltL) to bind for dichlorination. It was speculated that this closed off structure would limit off-target chlorination of small molecules [47]. PltA is the first FDH revealing a second halide

site. α-helices are coloured in red, β-strands are coloured in blue and loops are coloured in white. Active site components are depicted as spheres (vanadium) and sticks. Carbon atoms corresponding to protein are coloured according to their secondary structure. Oxygen, nitrogen and vanadium atoms are coloured light red, dark blue and grey, respectively.

Figure 4



Mechanism of FDHs and structure of VirX1. **(a)** Catalytic mechanism for FDHs, with proposed halolysine mediated halogenation. In FDHs without halolysine formation, electrophilic species remains as free HOX. **(b)** Cartoon representation of VirX1 (PDB: 6QGM). α-helices are coloured in red, β-strands are coloured in blue and loops are coloured in white. Surface representation of VirX1 all coloured in shades of grey.

binding site, distant (25 Å) to the FAD, the functional relevance of this secondary chloride binding site is not clear. It could perhaps be a crystallisation artefact, or perhaps function as a reservoir locally increasing chloride concentration [47].

Another recently discovered phenolic FDH with peculiar substrate scope and substrate binding domain is PltM (PDB: 6BZN) [48•]. This FDH mono-halogenates and di-halogenates phenolic compounds (such as phloroglucinol) using chloride, bromide and iodide as substrates, with a preference for chloride and bromide over iodide. PltM crystal structures not only highlighted multiple FAD binding modes, shedding light on the conformational control of $FADH_2$ binding and FAD release, crucial for FAD recycling, but also revealed a small novel helical substrate binding domain [48•].

A flavin-dependent halogenase from a Marine virus

Following a bioinformatics led approach that revealed a previously unnoticed motif Fx.P×.Sx.G with which to identify halogenases, *in silico*, discriminating them from monooxygenases, the first halogenase from a virus and the first FDH with a preference for iodination VirX1 was found [10••,49]. This FDH was found in the genome of the marine cyanophage Syn10. Not only is the viral origin of VirX1 puzzling, but its wide substrate scope and preference for iodination over bromination raises questions on its native function. Its broad substrate flexibility and ability to generate aryl iodides makes it an attractive biotechnological tool. Structural elucidation of this enzyme revealed a trimeric quaternary structure including the commonly found flavin binding domain and a substrate binding domain reminiscent of PrnA. The VirX1 crystal structure (PDB: 6QGM) (Figure 4b) and

in silico studies reveal the substrate binding site in VirX1 to be much more accessible than in other FDHs that act on free substrates (such as PrnA, RebH and BvrH), providing explanation of the diverse and unusual substrate scope of this FDH. The halide binding site of VirX1 shows a similar architecture to the halide binding site of PrnA. VirX1's considerably larger and more accessible active site is likely to be the reason for its observed broad substrate specificity and ability to use the largest, but electronically most readily oxidised of the halides.

Concluding thoughts

As indicated by compound structural diversity, so far only a very small proportion of existing halogenases have been explored. *In silico* approaches for the discovery of novel halogenases enable the identification of wildly different enzymes. Using such approaches, major advances have been seen in recent years including discovery of radical and FDH halogenases with biotechnologically interesting substrate scope.

Though the field is progressing rapidly, numerous exciting central questions remain to be answered. The postulated lysine haloamine, within the FDHs and speculation over its exact role as either the definitive halogenating species or as a reservoir of Cl^+ remains to be fully resolved. The regioselective HPOs hold considerable promise; what are the underpinning mechanisms by which they mediate their control? What are the key structural features of substrate free radical halogenases, what factors influence stability? Currently there is too little structural data to robustly address these questions, this is likely to change over the next few years. It is an exciting time for enzymatic halogenation.

Conflict of interest statement

Nothing declared.

Acknowledgements

We thank ERC GenoChemetics (FP7/2007-2013/ERC consolidator grant GCGXC grant agreement no 614779 RJMG) for funding, CSC and EPSRC CRITICAT EP/L016419/1 for studentship support (KG & SM respectively).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Kirk KL, Filler R: **Recent advances in the biomedicinal chemistry of fluorine containing compounds**. In *Biomedical Frontiers of Fluorine Chemistry*. Edited by Ojima I, McCarthy JR, Welch JT. American Chemical Society; 1996:1-24.

2. Herrera-Rodriguez LN, Khan F, Robins KT, Meyer H-P: **Perspectives on biotechnological halogenation part I: halogenated products and enzymatic halogenation**. *Chim Oggi-Chem Today* 2011, **29**:31-33.

3. Xu Z, Yang Z, Liu Y, Lu Y, Chen K, Zhu W: **Halogen bond: its role beyond drug–target binding affinity for drug discovery and development**. *J Chem Inf Model* 2014, **54**:69-78.

4. Jeschke P: **The unique role of halogen substituents in the design of modern agrochemicals**. *Pest Manag Sci* 2010, **66**:10-27.

5. Mahoney KPP, Smith DRM, Bogosyan EJA, Goss RJM: **Access to high value natural and unnatural products through hyphenating chemical synthesis and biosynthesis**. *Synthesis* 2014, **46**:2122-2132.

6. • Gribble G: **A recent survey of naturally occurring organohalogen compounds**. *Environ Chem* 2015, **12**:396-405
A helpful overview of the naturally occurring organohalides discovered so far.

7. Gkotsi D, Dhaliwal J, McLachlan M, Mulholand K, Goss R: **Halogenases: powerful tools for biocatalysis (mechanisms applications and scope)**. *Curr Opin Chem Biol* 2018, **43**:119-126.

8. •• Shaw PD, Hager LP: **Biological chlorination- peroxidative nature of enzymatic chlorination**. *J Am Chem Soc* 1959, **81**:6527-6528
First example of biosynthetic chlorination.

9. Hager LP, Mirris DR, Brown FS, Eberwein H: **Chloroperoxidase. II. Utilization of halogen anions**. *J Biol Chem* 1966, **8**:1769-1777.

10. •• Gkotsi DS, Ludewig H, Sharma SV, Connolly JA, Dhaliwal J, Wang Y, Unsworth WP, Taylor RJK, McLachlan MMW, Shanahan S *et al.*: **A marine viral halogenase that iodinates diverse substrates**. *Nat Chem* 2019, **12**:1091-1097
First FDH with a preference for iodination, first halogenase from a virus. Extensive breadth of substrates accepted by an FDH. An unusual trimeric structure.

11. •• Fisher BF, Snodgrass HM, Jones KA, Andorfer MC, Lewis JC: **Site-selective C-H halogenation using flavin-dependent halogenases identified via family-wide activity profiling**. *ACS Cent Sci* 2019, **5**:1844-1856
Extensive breadth of substrates accepted by an FDH.

12. Mitchell AJ, Zhu Q, Maggiolo AO, Ananth NR, Hillwig ML, Liu X, Boal AK: **Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5**. *Nat Chem Biol* 2016, **8**:636-640.

13. • Chang Z, Flatt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH: **The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit**. *Gene* 2002, **1–2**:235-247
Indication of operation of a radical halogenase.

14. • 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 U S A.* 2005, **29**:10111-10116
Discovery of first radical halogenase.

15. Dong C, Huang F, Deng H, Schaffrath C, Spencer JB, O'Hagan D, Naismith JH: **Crystal structure and mechanism of a bacterial fluorinating enzyme**. *Nature* 2004, **6974**:561-565.

16. • Eustáquio AS, Pojer F, Noel JP, Moore BS: **Discovery and characterization of a marine bacterial SAM-dependent chlorinase**. *Nat Chem Biol* 2008, **1**:69-74
Discovery of a rare nucleophilic halogenases, with striking similarity to 5'-FDAS.

17. Sanada M, Miyano T, Iwadare S, Williamson JM, Arison BH, Smith JL, Douglas AW, Liesch JM, Inamine E: **Biosynthesis of fluorothreonine and fluoroacetic acid by the thienamycin producer, *Streptomyces cattleya***. *J Antibiot* 1986, **2**:259-265.

18. O'Hagan D, Goss RJ, Meddour A, Courtieu J: **Assay for the enantiomeric analysis of [2H1]-fluoroacetic acid: insight into the stereochemical course of fluorination during fluorometabolite biosynthesis in streptomyces cattleya**. *J Am Chem Soc* 2003, **2**:379-387.

19. Zhu X, Robinson DA, McEwan AR, O'Hagan D, Naismith JH: **Mechanism of enzymatic fluorination in *Streptomyces cattleya***. *J Am Chem Soc* 2007, **47**:14597-14604.

20. Vincent MA, Hillier IH: **The solvated fluoride anion can be a good nucleophile**. *Chem Commun* 2005, **47**:5902-5903.

21. Hillwig ML, Liu X: **A new family of iron-dependent halogenases acts on freestanding substrates**. *Nat Chem Biol* 2014, **11**:921-923.

22. Marchand JA, Neugebauer ME, Ing MC, Lin CI, Pelton JG, Chang MCY: **Discovery of a pathway for terminal-alkyne amino acid biosynthesis**. *Nature* 2019, **7748**:420-424.

23. • Hillwig ML, Zhu Q, Ittiamornkul K, Liu X: **Discovery of a promiscuous non-heme iron halogenase in ambiguine alkaloid biogenesis: implication for an evolvable enzyme family for late-stage halogenation of aliphatic carbons in small molecules**. *Angew Chem Int Ed Engl* 2016, **19**:5780-5784
Paper reporting the discovery of a second radical halogenase able to act on a free-standing substrate, AmbO5, with an increased substrate scope compared to WelO5.

24. O'Hagan D, Schaffrath C, Cobb SL, Hamilton JT, Murphy CD: **Biochemistry: biosynthesis of an organofluorine molecule**. *Nature* 2002, **6878**:279.

25. •• Neugebauer ME, Sumida KH, Pelton JG, McMurry JL, Marchand JA, Chang MCY: **A family of radical halogenases for the engineering of amino-acid-based products**. *Nat Chem Biol* 2019, **10**:1009-1016
Disovery of a new family of radical halogenases acting on non-protein bound amino acids.

26. •• Bernhardt P, Okino T, Winter JM, Miyanaga A, Moore BS: **A stereoselective vanadium-dependent chloroperoxidase in bacterial antibiotic biosynthesis**. *J Am Chem Soc* 2011, **12**:4268-4270
First demonstration of stereoselectivity by a HPO.

27. Kaysser L, Bernhardt P, Nam SJ, Loesgen S, Ruby JG, Skewes-Cox P, Jensen PR, Fenical W, Moore BS: **Merochlorins A-D, cyclic meroterpenoid antibiotics biosynthesized in divergent pathways with vanadium-dependent chloroperoxidases**. *J Am Chem Soc* 2012, **29**:11988-11991.

28. Miles ZD, Diethelm S, Pepper HP, Huang DM, George JH, Moore BS: **A unifying paradigm for naphthoquinone-based meroterpenoid (bio)synthesis**. *Nat Chem* 2017, **12**:1235-1242.

29. •• Sundaramoorthy M, Terner J, Poulos TL: **The crystal structure of chloroperoxidase : a heme peroxidase-cytochrome P450 functional hybrid**. *Structure* 1995, **3**:1367-1377
Crystal structure of first halogenase CPO, shows its unique structure and haem active site.

30. Sundaramoorthy M, Terner J, Poulos TL: **Stereochemistry of the chloroperoxidase active site: crystallographic and molecular-modeling studies**. *Chem Biol* 1998, **9**:461-473.

31. Ruf J, Carayon P: **Structural and functional aspects of thyroid peroxidase**. *Arch Biochem Biophys* 2006, **2**:269-277.

32. Carpena X, Vidossich P, Schroettner K, Calisto BM, Banerjee S, Stampler J, Soudi M, Furtmüller PG, Rovira C, Fita I, Obinger C: **Essential role of proximal histidine-asparagine interaction in mammalian peroxidases**. *J Biol Chem* 2009, **38**:25929-25937.

33. Singh AK, Singh N, Sharma S, Singh SB, Kaur P, Bhushan A, Srinivasan A, Singh TP: **Crystal structure of lactoperoxidase at 2.4 A resolution**. *J Mol Biol* 2008, **4**:1060-1075.

34. Sharma S, Singh AK, Kaushik S, Sinha M, Singh RP, Sharma P, Sirohi H, Kaur P, Singh TP: **Lactoperoxidase: structural insights into the function, ligand binding and inhibition**. *Int J Biochem Mol Biol* 2013, **3**:108-128.

35. Poulos TL: **Heme enzyme structure and function**. *Chem Rev* 2014, **7**:3919-3962.

36. •• Messerschmidt A, Wever R: **X-ray structure of a vanadium-containing enzyme: chloroperoxidase from the fungus *Curvularia inaequalis***. *Proc Natl Acad Sci U S A* 1996, **1**:392-396
First structure determination of a vanadium dependent halogenase giving insights into stability that comes from cofactor binding and overall structure.

37. Weyand M, Hecht H, Kiess M, Liaud M, Vilter H, Schomburg D: **X-ray structure determination of a vanadium-dependent haloperoxidase from *Ascophyllum nodosum* at 2.0 A resolution**. *J Mol Biol* 1999, **3**:595-611.

38. Isupov MN, Dalby AR, Brindley AA, Izumi Y, Tanabe T, Murshudov GN, Littlechild JA: **Crystal structure of dodecameric vanadium-dependent bromoperoxidase from the red algae *Corallina officinalis***. *J Mol Biol* 2000, **4**:1035-1049.

39. Fournier JB, Rebuffet E, Delage L, Grijol R, Meslet-Cladière L, Rzonca J, Potin P, Michel G, Czjzek M, Leblanc C: **The vanadium iodoperoxidase from the marine Flavobacteriaceae species *Zobellia galactanivorans* reveals novel molecular and evolutionary features of halide specificity in the vanadium haloperoxidase enzyme family**. *Appl Environ Microbiol* 2014, **80**:7561-7573.

40. Frank A, Seel CJ, Groll M, Gulder T: **Characterization of a cyanobacterial haloperoxidase and evaluation of its biocatalytic halogenation potential**. *Chembiochem* 2016, **21**:2028-2032.

41. Zhu X, De Laurentis W, Leang K, Herrmann J, Ihlefeld K, van Pée KH, Naismith JH: **Structural insights into regioselectivity in the enzymatic chlorination of tryptophan**. *J Mol Biol* 2009, **1**:74-85.

42. Moritzer AC, Minges H, Prior T, Frese M, Sewald N, Niemann HH: **Structure-based switch of regioselectivity in the flavin-dependent tryptophan 6-halogenase Thal**. *J Biol Chem* 2019, **7**:2529-2542.

43. Lingkon K, Bellizzi JJ 3rd: **Structure and activity of the thermophilic tryptophan-6 halogenase BorH**. *Chembiochem* 2019, **21**:1121-1128.

44. Ainsley J, Mulholland AJ, Black GW, Sparagano O, Christov CZ, Karabencheva-Christova TG: **Structural insights from molecular dynamics simulations of tryptophan 7-halogenase and tryptophan 5-halogenase**. *ACS Omega* 2018, **5**:4847-4859.

45. Neubauer PR, Widmann C, Wibberg D, Schröder L, Frese M, Kottke T, Kalinowski J, Niemann HH, Sewald N: **A flavin-dependent halogenase from metagenomic analysis prefers bromination over chlorination**. *PLoS One* 2018, **5**:e0196797.

46. 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, **2**:520-530.

47. Pang AH, Garneau-Tsodikova S, Tsodikov OV: **Crystal structure of halogenase PltA from the pyoluteorin biosynthetic pathway**. *J Struct Biol* 2015, **3**:349-357.

48. • Mori S, Pang AH, Chandrika NT, Garneau-Tsodikova S, Tsodikov OV: **Unusual substrate and halide versatility of phenolic halogenase PltM**. *Nat Commun* 2019, **1**:2053
Important studies into the substrate acceptance of PltM.

49. Goss RJM, Gkotsi, DS: **Discovery and utilisation of wildly different halogenases, powerful new tools for medicinal chemistry**: patent application number GB1803491.8.