***A Marine Viral Halogenase that Iodinates Diverse Substrates***

**Authors:** Danai S. Gkotsi,1Σ†§λ Hannes Ludewig,1†λϕ Sunil V. Sharma,1§ Jack Connolly,1λϕ Jagwinder Dhaliwal,1λ Yunpeng Wang,1λ William P. Unsworth,2∆ Richard J. K. Taylor,2∆ Matthew M.W. McLachlan,3,4 π Stephen Shanahan,3π James H. Naismith,5 and Rebecca J. M. Goss\*1Σ

**Affiliations:**

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

Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Fife, KY16 9ST, UK. RJMG@St-Andrews.ac.uk

2 Department of Chemistry, University of York, Heslington, York YO10 5DD, UK.

3 Syngenta, Jealott’s Hill International Research Centre, Bracknell, Berkshire, RG42 6EY, UK.

4 QEDDI, Staff House Road, The University of Queensland, Brisbane, QLD 4072, Australia.

5 Division of Structural Biology, Wellcome Trust Centre of Human Genomics, Roosevelt Drive, Oxford, OX3 7BN; Research Complex at Harwell, Rutherford Laboratory, Didcot, OX11 0FA; The Rosalind Franklin Institute, Didcot OX11 0FA

ΣThese authors determined the iodinase and carried out its biochemical investigation and substrate screening. †These authors carried out the structural analysis of the enzyme, under the guidance of JHN. §These authors explored the differential reactivity of the substrates with HOI, characterised the products of the iodinase and synthesised synthetic standards for comparison to products. ∆These authors synthesised a series of spiro-indolic compounds and derivatives, utilised as substrates by the enzyme. πThese authors contributed to the selection of compounds for the assaying of the iodinase. ϕThese authors carried out structural modelling of the iodinase. λThese authors assisted with cloning and protein production.

**Introductory Paragraph:**

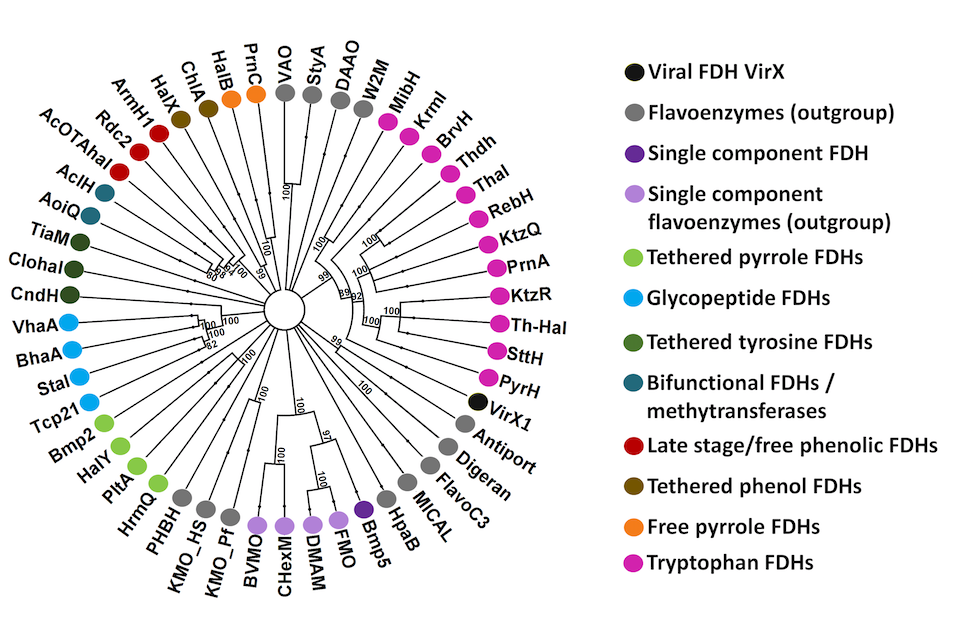
Enzymatic halogenation is associated with secondary metabolism, underpinning and modulating bioactivity and bioavailability of many natural products1-3. Here we report the first discovery and characterisation of a flavin-dependent halogenase (FDH) from cyanobacteriophage, and the first FDH with a preference for iodination. Exploration of this viral halogenase (VirX1) reveals it to be capable of regioselective halogenation of a diverse range of substrates, with a preference for forming highly reactive aryl iodide species; this has potential implications for the metabolism of the infected host. This wild type FDH is interesting from a biocatalytic perspective showing strikingly broad substrate flexibility. Until now, catalytically useful enzymatic iodination had been missing from the biocatalytic portfolio.

**Main Text:**

Here we report the identification, biochemical and structural characterisation of the first halogenase from a virus, and demonstrate unprecedented biotransformative potential of an enzyme from such a viral system. This viral FDH shows a high level of innate substrate flexibility. Furthermore, the enzyme displays preference for forming C-I bonds, and is, to the best of our knowledge, the first flavin-dependent iodinase demonstrated *in vitro*. These factors together render it an attractive and useful tool for synthesis.

The selective formation of C-X bonds is of great importance to the pharmaceutical and agrochemical industries1,2. The introduction of a halogen (X) into a molecule can be used to modulate bioactivity, bioavailability and metabolic stability1-3. Synthetic chemical approaches to halogenation employ highly reactive and toxic reagents, lack selectivity and generate harmful waste. Conversely, biosynthetic (enzymatic) halogenation is mild, highly selective and utilises simple salts such as NaCl or NH4Br as the source of halide1,3,4. Consequently, the discovery and structural characterization of FDHs capable of selectively forming C-Cl and C-Br bonds and the discovery of a SAM-dependent fluorinase, able to mediate C-F bond formation, attracted considerable attention4. An enzyme with a preference for the formation of C-I bonds remained to be discovered and characterized.

The FDHs that have previously been studied are predominantly drawn from a limited number of well-known phyla1,3,4. It may be seen, through branching analysis, that there is a correlation between protein sequence and substrate preference, including enzyme tethered or untethered substrate presentation (Fig. 1). As halogenases from a very limited series of organisms have been explored, so far, their natural substrate specificity has been limited. Almost all previously discovered FDHs have been determined due to their role in the generation of natural halogenated compounds. Many of these metabolites arise from well-studied bacteria, especially actinomycetes or from fungi. Significant effort has been invested into the rational redesign and directed evolution in order to expand or change the substrate specificity of these enzymes5.

**Fig. 1. Branching analysis reveals correlation between gene relatedness and substrate specificity, and enables unexplored areas of sequence space to be identified.**

Breaking from this trend in halometabolite-led identification of enzymes, we adopted a bioinformatics-based approach. Compiling data from all fully biochemically and structurally characterized FDH enzymes, sequence alignment enabled us to see both the known GxGxxG and WxWxIP motifs,4a as well as a previously unnoticed motif Fx.Px.Sx.G (see Supplementary Information), which we will refer to as the FloPpySeGment motif6.Whilst GxGxxG and WxWxIP motifs have previously been utilized to indicate the presence of FDHs, GxGxxG could as readily indicate Flavin Mono Oxygenase (FMOs) related enzymes and the WxWxIP motif is known to be absent in more unusual FDHs, such as Bmp57. We have demonstrated that the Fx.Px.Sx.G motif may be utilized, by itself, to mine for halogenases from uncurated genome sequences6.

Our probe revealed the likely presence of a gene encoding a halogenase, which we named “VirX1,” within the cyanophage Syn10, a broad host range cyanophage with a 177 kbp genome known to infect *Prochlorococcus*and *Synechococcus*8.Aligned against known halogenases, VirX1 shows low similarity (30% amino acid identity) to PrnA, the well-studied tryptophan 7-halogenase from pyrrolnitrin biosynthesis. It may therefore be classified as being in the “Twilight Zone” of sequence similarity. Generation of a homology structural model of VirX1 using Phyre29 and comparative analysis of this to PrnA and all other structurally characterised halogenases revealed the characteristic, highly helical structure, of tryptophan halogenases with the pyramid and box shape4a,b. However, the model of VirX1 indicated participation of more loops in the packing of the C-terminus in VirX1, potentially enabling larger conformational changes upon substrate binding and enhanced substrate specificity compared to typical tryptophan halogenases. The active site could be identified from the position of the Fx.Px.Sx.G motif, as well as the co-factor binding site GxGxxG.

We set out to explore whether the encoded protein was indeed an active halogenase, and if so, whether it might show potential biocatalytic utility and substrate flexibility. To this end we screened the halogenase against a 400-member library containing substrates that would sterically and electronically challenge the potential catalyst, using LC-HRMS to validate the generation of new, halogenated products (see Supplementary Information). Under the conditions of our assay, the enzyme showed very poor activity in chlorinating substrates, but good bromination activity and a surprising preference for iodination. The wild type enzyme demonstrated a striking substrate flexibility. From this 400 member library, it was determined that 32 sterically and electronically diverse compounds could be accepted as substrates, by the wild type enzyme, with 1-95% conversion (Fig. 2). LC-HRMS analysis indicated that, biotransformative halogenation resulted in the formation of monoiodinated product only. For the majority of substrates only a single regioisomer was observed. Enzymatic halogenation could be seen to be mediated on several less reactive substrates, including bathophenanthroline (**30**), that were not readily iodinated using synthetic conditions at room temperature. Fourteen of these products were selected, based on both structural interest and conversion level, for further analysis including spectroscopic structural characterisation (see Supplementary Information). This analysis revealed that although halogenation often was mediated at the most chemically reactive position, this was not always the case. For example, enzymatic products **3**, **5**, **7**, **8**, **12** and **14** did not match with synthetic iodo-standards prepared via electrophilic iodination, indicating the possibility of different regioisomers being formed. Kinetic assessment of the enzyme catalysed reaction was carried out for twelve of these substrates.



**Fig. 2. Diverse substrate scope of VirX1, and kinetic assessment of 12 substrates.** Products **1** and **2** were isolated from VirX1 assays, whereas regiochemistry for **3** - **14** was determined by comparison with synthetic standards, \* denotes hypothesised halogenation site by VirX1 where the enzymatic product did not match with synthetic standards. The enzyme is shown to selectively mono-halogenate a diverse range of sterically and electronically different substrates. Notably, neither D or L tryptophan are accepted as substrates. Though VirX1 is capable of bromination a preference may be seen for iodination. Conversion levels, to the iodinated product, as estimated by LC-HRMS analysis, are reported. Fourteen substrates were selected (boxed, Fig. 2), based on both interest and stability of product, for further verification through either scale up and spectroscopic characterization of their product, or through comparison to synthetic standards that we generated and fully characterized (see Supplementary Information)10.

Kinetic analysis revealed the enzyme to have preference for iodide over bromide. This is unprecedented as there has been no previous characterisation of FDHs with a natural preference for iodination *in vitro*. Indeed, in halide competition assays with the halogenase RebH, it was shown that the introduction of NaI prevented the formation of any chlorinated product. Conversely, in the competition assays that we explored with VirX1 using equivalent concentrations of NaI, NaBr and NaCl, only the iodinated product was observed. (see Supplementary Information Fig. S19). Under the conditions of our assay, and with all substrates explored, the preference for the halide, shown by VirX1, is I > Br > Cl corresponding to the decreasing oxidative potential of the halide.

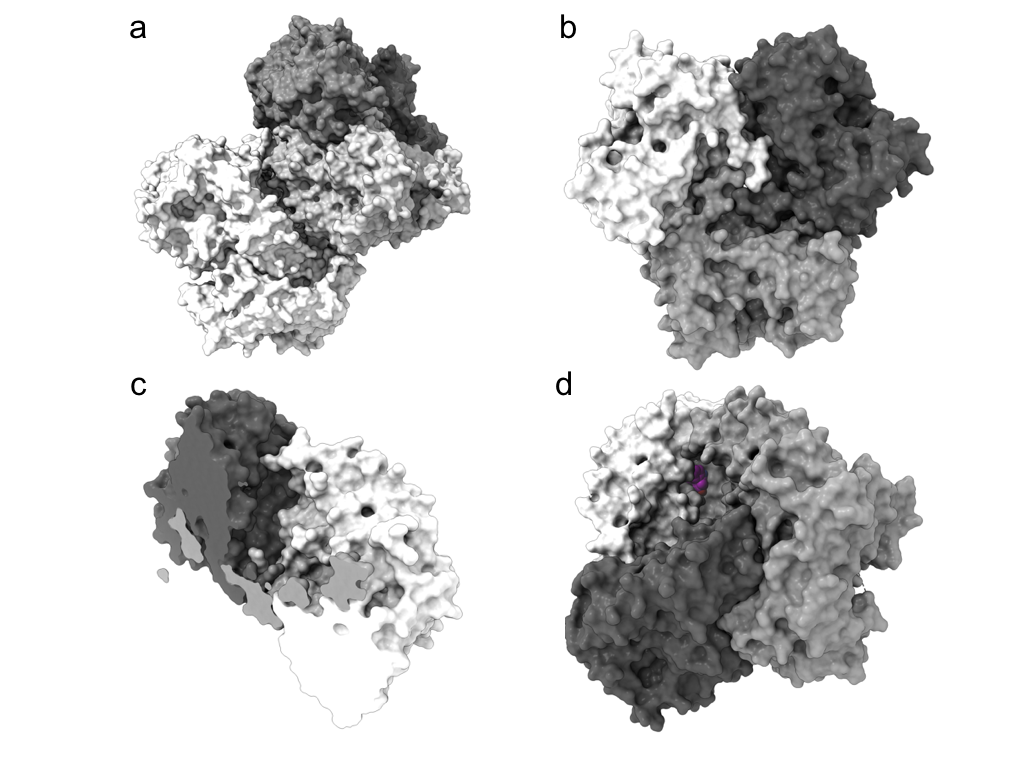
Recently the PltM halogenase11 has been reported revealing for the first-time permissible iodination, with a detectable level of iodinated product being released. However, levels of iodination of the compounds were too low to enable their spectroscopic characterisation or indeed the kinetics for the reaction11.

Though the benchmark FDH, PrnA, is known to have evolved to process tryptophan as its natural substrate, VirX1’s natural substrate is unknown, and its discovery is beyond the scope of present study. Nevertheless, VirX1 shows good catalytic activity. Iodination, by VirX1 of a modest/non-native substrate 6-azaindole (**1**) proceeds with kcat / KM 5 ± 0.5 whilst PrnA’s chlorination of its natural substrate is less catalytically efficient (kcat / KM 0.05 ± 0.005)4a*.* Synthetic iodination of 6-azaindole (**1**) with either free HOI or HOBr, at low concentration in phosphate buffer, unsurprisingly results in generation of di-iodinated species and series of other products, *e.g.* dimer of bromo-6-azaindole as observed by LCMS, conversely the enzyme yields only the monoiodinated product as a single regioisomer (see Supplementary Information, Scheme S1). We explored whether, well studied, PrnA might indeed mediate halogenation of substrates processed by VirX1. We saw that PrnA could indeed process many of the substrates, to a low level. Though PrnA could not iodinate its natural substrate tryptophan, excitingly with a number of its unnatural substrates, discernible levels of iodination were evident though typically at ~1% conversion (see Supplementary Information, Table S6).

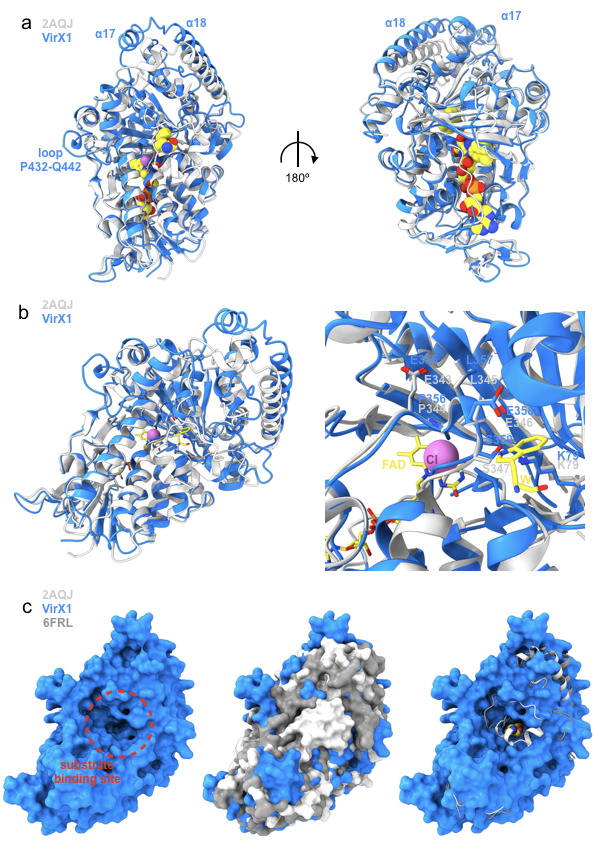
We next sought to structurally investigate VirX1. Purification of the protein enabled X-ray crystallographic analysis of the apo structure revealing it to comprise of six monomers of VirX1 in the asymmetric unit (Fig. 3a); all showing highest structural similarity to the PrnA E450K structure (DALI search; highest z-score: 47.1, pdb code: 4Z43) sharing box and pyramid architecture; agreeing with our bioinformatics predictions4a,b,12. The six monomers in the asymmetric unit are predicted by jsPISA 2.0.513 to form two stable trimeric assemblies (Fig. 3b; *G*diss. ~ 42.6 kcal mol-1) with protein-protein interfaces covering an area of 10,486.4 Å2 per trimer (see Supplementary Information for details)14-23. The trimeric state of VirX1 was confirmed by analytical size exclusion chromatography revealing an apparent molecular weight of VirX1 of ~ 180 kDa in solution, as well as by SEC MALS (Supplementary Information Table S4).

The VirX1 structure shows many similarities to other FDHs with the box shaped flavin binding module, required for flavin binding and generation of the hypohalous acid electrophile, being almost identically folded. Both K79 and E358, implicated in halo-amine formation and in deprotonation of the Wheland intermediate, respectively, occupy similar positions to K79 and E346 in the active site of PrnA. The positioning of F99 and Y97 is similar to that seen for F103 and H101, in PrnA complexing Trp, Cl, FAD (PDB ID 2AQJ) that are believed to stabilise the Wheland intermediate4a,b. Assay with monochlorodimedone (MCD) indicated that the halogenating species must be enzyme bound. Site directed mutagenesis was carried out to identify key catalytic residues, and whilst the mutants K79A, K79R, F353A, S359A and P356A all yielded soluble protein, which eluted in an identical manner to WT VirX1 indicating correct folding, each of these mutants were determined to be completely inactive in catalysing halogenation of azaindole (**1**) with either NaI or NaBr, reinforcing the importance of these residues for the enzyme’s catalytic activity (Supplementary Information Fig. S18).

As anticipated, with all key catalytic residues in place, the α-helical pyramid-shaped substrate module displays a different arrangement of secondary structure elements to its closest structural homologue PrnA. Key differences include the absence of the α-helical lid required for tryptophan halogenases (T435-W455 in PrnA) closing off the tryptophan binding site (as seen in BrvH)24*.* The substrate binding site of VirX1 displays an even wider opening than BrvH, providing a possible explanation for the wide substrate scope, and the ability to accommodate a sterically bulky halide. Additional expansion of the putative substrate binding site is enabled by the increased distance of α17 and α18 from the catalytic residues within it. A further, key difference to PrnA, RebH and BrvH is an additional loop (P432-Q442) taking part in the trimer interface through a H-bonding network, opening up access to substrate binding site (Fig. 3).



**Fig. 3: Crystal structure of VirX1.** a) ASU content of VirX1 (surface representation) crystal structure. Chains a, b, and c are coloured in increasing grey scale and the second trimer is coloured white. b) Trimeric VirX1, bottom view. c) Trimeric VirX1 angled top view, cut through to highlight cavity formed by the three monomers. d) Trimeric VirX1 with substrate **14** (spheres; carbon magenta, nitrogen blue and oxygen red) docked to all three substrate binding sites.



**Fig. 4: Structural comparison of VirX1 to PrnA (2AQJ) and BrvH (6FRL).** **a)** Structural

alignment (r.m.s.d. 1.166 Å) of VirX1 monomer in blue and the PrnA complex structure (2AQJ) in light grey, both represented as cartoons. PrnA ligands (FAD, Trp and Cl) are represented as spheres. **b)** View of ligand binding sites in VirX1. The right panel provides a zoom into the substrate and halogen binding site. Residue side chains, FAD and Trp are represented in stick form. c) Surface representation of VirX1 (all panels) as well as 2AQJ and 6FRL in surface (middle panel) and cartoon style, including Trp (spheres; right panel), focusing on increased accessibility of VirX1 substrate binding site compared to PrnA and BrvH. Nitrogen, oxygen phosphorus and chlorine atoms are coloured yellow, blue, red, orange and magenta respectively. Carbons within proteins are coloured according to backbone colour of respective macromolecule.

To further explore the VirX1 substrate binding mode, many attempts were made to crystallise VirX1 in complex with various substrates (Supplementary Information Table S1). However, these did not yield sufficiently diffracting crystals for structural analysis, therefore *in silico* methods were explored instead. Given the subjectivity of *in silico* small molecule protein docking experiments, several complementary approaches were explored, with both rigid and flexile models of VirX1. Compounds **1** - **14** were docked within a 47.25 × 47.25 × 47.25 Å (105,488 Å3) search space centred on rigid VirX1 (chain a), which identified a substrate binding cleft in similar position to other FDHs harbouring K79 and E358. These results informed further docking experiments restricted to this cleft, conducted with rigid and flexible amino acid residues (Supplementary Information Section Y). The flexible docking experiments yielded conformations with a decreased binding energy (~ 5 kcal mol-1) for all substrates compared with docking into the rigid VirX pocket (Supplementary Information Table S2). The choice of hypothetical binding modes (Supplementary Information Fig. S1c) was mainly informed by proximity of the putatively halogenated carbon to the K79 that would form and position the iodamine for reaction with the substrate. Through this analysis, three residues were identified to be involved in substrate binding across all docked substrates: Y97, P98 and F99, with the two aromatic resides often sandwiching the substrate (Supplementary Information Fig. S1 – S13 and “heat plot” Fig. S14). Other residues such as L53, I82 and G100 were also frequently involved in binding docked compounds, with the predicted binding mainly characterized by hydrophobic contacts. The docking experiments support binding in a cleft with access to K79 and E358, similarly to characterized complex structures of enzyme-substrate complexes and have identified several further residues that could hypothetically be relevant to binding and reinforced the extensive active site cavity enabling the strikingly broad substrate flexibility. It is likely that it is this enlarged cavity that enables halogenation via a bulky iodamine species, and the oxidation potential dictates the preference halide utilization.

Our results reveal, for the first time catalytically notable iodination by an FDH, with an enzyme demonstrating a preference for iodination. Notably, we also show for the first time, that by extending the substrate scope of PrnA iodination within this system may be permissible. We demonstrate that VirX1 possesses a very broad substrate flexibility and is able to process sterically and electronically demanding substrates in addition to series of N and O heterocycles. This may be attributed the relatively large and accessible active site revealed by the crystal structure, and the high level of conformational flexibility predicted by our modelling studies. Excitingly we reveal a system that may be utilised as a biocatalytic tool enabling enzyme catalysed iodination for the first time.

With the recently postulated role of marine viruses in manipulating the metabolism of the cyanobacteria that they infect, the discovery of a viral halogenase and demonstration of its catalytic efficiency and broad substrate flexibility is interesting. Halogenation of a molecule can significantly perturb its bioactivity and bioavailability, and iodination can render a molecule highly chemically reactive. Though the levels of iodide in the oceans are very low compared to those for bromide and chloride (0.05 ppm vs. 65 ppm, and 18980 ppm) marine algae and bacteria have been shown to accumulate and utilise iodide25. The naturally broad substrate specificity of this enzyme, encoded in a virus which naturally infects the two most abundant photosynthetic organisms on the planet, is fascinating, and leads one to wonder as to what its natural role might be.

**Acknowledgements** We thank the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013/ERC grant agreement no 614779 GenoChemetics to RJMG), Syngenta and Wellcome ISSF 204821/Z/16/Z (DSG) for generous financial support. We thank Gemma Harris and Miriam Weckener (Harwell) for SEC-MALS analysis. We thank all our colleagues, in particular, Professor Terry Smith and our wonderful colleagues in Biology and Chemistry at St Andrews for all of the help that they have afforded us in the aftermath of the BMS fire.

**References and Notes:**

1. **a)** V. Agarwal *et al.*, Enzymatic Halogenation and Dehalogenation Reactions: Pervasive and Mechanistically Diverse, *Chem. Rev.*, ***117***, 5619-5674 (2017). **b**) D. S. Gkotsi *et al.*, [Halogenases: powerful tools for biocatalysis (mechanisms applications and scope)](javascript:void(0)), *Curr. Opin. Chem. Biol*., ***43***, 119-126 (2018).

### 2. The Complete List and Analysis of the Best Selling Drugs in 2017. All the Top Products in the Pharmaceutical and Biotechnology Industry with detailed performance and future trends https://www.igeahub.com/2018/04/07/20-best-selling-drugs-2018/(accessed August 2018).

3. C.S. Neumann, D. G. Fujimori and C. T. Walsh*,* Halogenation strategies in natural product biosynthesis, *ChemBiol.,* ***15***, 99-109, (2008).

4. **a**) C. Dong *et al.*, Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science*, ***309***, 2216-2219 (2005). **b**) L. C. Blasiak *et al.*, Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis, *Nature,* ***440***, 368-371 (2006).  **c**) C. J. Dong *et al.*, Crystal structure and mechanism of a bacterial fluorinating enzyme, *Nature,* ***427***, 561-565, (2004).

5. For examples see **a**) A. Lang *et al.*, Changing the Regioselectivity of the Tryptophan 7-Halogenase PrnA by site directed mutagenesis, *Angew. Chem. Int. Ed.,* ***50***, 2951-2951(2011). **b**) W. S. Glenn, E. Nims and S. E. O’Connor, Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *J. Am. Chem. Soc.*, ***133***, 19348-19349 (2011).

**c**) J. T. Payne, C. B. Poor and J. C. Lewis, Directed evolution of RebH for site selective halogenation of large biologically active molecules.*Angew. Chem. Int. Ed.,* ***54***, 4226-4230 (2015). **d**) B. R. K. Menon *et al.*, RadH: A Versatile Halogenase for Integration into Synthetic Pathways. *Angew. Chem. Int. Ed.,* ***56***, 11841-11845, (2017). e) C. Schnepel *et al.*, A high-throughput fluorescence assayto determine the activity of tryptophan halogenases*, Angew. Chem. Int. Ed.,* ***55***, 14159-14163, (2017).

6. R. J. M. Goss and D. S. Gkotsi,Discovery and utilisation of wildly different halogenases, powerful new tools for medicinal chemistry: patent application number GB1803491.8.

7. V. Agarwal *et al.*, Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Bio.,* ***10***, 640-647 (2014).

### 8. M. B. Sullivan, J. B. Waterbury and S. W. Chrisholm*.,* Cyanophage infecting the oceanic cyanobacterium Prochlorococcus, *Nature*, *424*, 1047–1051(2003).

9. L. A. Kelley *et al.*, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* ***10***, 845-58 (2015).

10. a) M. J. James *et al.*, Silver(I) or Copper(II)-Mediated Dearomatisation of Aromatic Ynones: Direct Access to Spirocyclic Scaffolds, *Angew. Chem. Int. Ed.,****54***, 7640-7643 (2015). b) S. J. Chambers *et al.*, From Heteroaromatic acids and Imines to Azaspirocycles: Stereoselective Synthesis and 3D Shape Analysis; *Chem. Eur. J.,****22***,6496-6500 (2016)*.*

11. S. Mori *et al.*, Unusual substrate and halide versatility of phenolic halogenase PltM, *Nat. Commun.*, ***10***, 1255 (2019).

12. L. Holm and L. M. Laakso, Dali server update, *Nucl. Acids Res.,* ***44***, W351-W355 (2016).

13. E. Krissinel, Stock-based detection of protein oligomeric states in jsPISA, *Nucl. Acids Res.*, ***43***, W314-W319 (2015).

14. C. Vonrhein *et al.*, Data processing and analysis with the autoPROC toolbox, *Acta Cryst.,* ***D67***, 293-302 (2011).

15. W. Kabsch, XDS, *Acta Cryst.*, ***D66***, 125-132 (2010).

16. P. R. Evans, Scaling and assessment of data quality, *Acta Cryst.*, ***D62***, 72-82 (2006).

17. P. R. Evans and G. N. Murshudov, How good are my data and what is the resolution?, *Acta Cryst.*, ***D69***, 1204-1214 (2013).

18. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments, *Acta. Cryst.*, ***D67***, 235-242 (2011).

19. I. J. Tickle, STARANISO, Cambridge, United Kingdom: Global Phasing Ltd. (2016).

20. R. M. Keegan and M. D. Winn, Automated search-model discovery and preparation for structure solution by molecular replacement, *Acta Cryst.* ***D63***, 447 - 457 (2007).

21. P. Emsley *et al.*, Features and development of Coot, *Acta Cryst.*, ***D66***, 486-501 (2010).

22. G. N. Murshudov, Refinement of Macromolecular Structures by the Maximum-Likelihood method, *Acta Cryst.*, ***D53***, 240-255 (1997).

23. V. B. Chen *et al.*, MolProbity: All-Atom Structure Validation for Macromolecular Crystallography, *Acta Cryst.,* ***D66***, 12–21 (2010).

24. P. R. Neubauer *et al.*, A flavin-dependent halogenase from metagenomic analysis prefers bromination over chlorination, *PLoS ONE,* ***13***, e0196797 (2018).

25. See for example: **a**) S. Amachi, Microbial contribution to global iodine cycling: volatilization, accumulation, reduction, oxidation and sorption of iodine*, Microbes. Environ.* *23*, 269-276 (2008), and references therein b) S. J. Crockford, Evolutionary roots of iodine and thyroid hormones in cell-cell signalling, *Integr. Com. Biol.* *49*, 155-166 (2009) and references therein.