Halogenases for biosynthetic pathway engineering: Toward new routes to naturals and non-naturals

Binuraj R. K. Menon, Daniel Richmond & Navya Menon

To cite this article: Binuraj R. K. Menon, Daniel Richmond & Navya Menon (2020): Halogenases for biosynthetic pathway engineering: Toward new routes to naturals and non-naturals, Catalysis Reviews, DOI: 10.1080/01614940.2020.1823788

To link to this article: https://doi.org/10.1080/01614940.2020.1823788
Halogenases for biosynthetic pathway engineering: Toward new routes to naturals and non-naturals

Binuraj R. K. Menon, Daniel Richmond, and Navya Menon

Warwick Integrative Synthetic Biology Centre, School of Life Sciences, University of Warwick, Coventry, UK

ABSTRACT

Nature's repertoire of bio-halogenase enzymes is intriguing with halogenases from various natural product biosynthetic clusters that carry out site and region-specific halogenation of diverse bioactive precursors and molecules. Currently, we have a comprehensive catalogue of cryptic and non-cryptic halogenases that act on simple to complex aliphatic and aromatic molecular scaffolds. This will open up further synthetic and biosynthetic opportunities for C-H activation, ring formation and functionalization of different molecular structures. In fact, halogenases were exploited over the years for these potential applications, to replace traditional chemical halogenation chemistries toward creating economical and environmentally benign methodologies and also for biosynthetic pathways. This review will discuss our advances in utilizing bio-halogenases to generate both in vivo and in vitro biosynthetic pathways; summarizing all naturals and non-naturals that are synthesized with a direct bio-halogenase incorporation.

1. Introduction

The carbon-halogen bonds (C-F, C-Cl, C-Br and C-I) in biomolecules and pharmaceutical compounds have a profound role in improving their potency, efficacy, reaction selectivity, biophysical properties and molecular bonding interactions via halogen-bonds.[1–3] As a direct consequence, the last 50 years have witnessed a rapid increase in identifying new halogenated biomolecules from secondary metabolites and natural product discovery world.[4–6] A major proportion of all pharmaceuticals, therapeutic peptides, agrochemicals, detergents, pigments and polymers that are used in our modern life are halogenated.[7–10] In organic synthesis, C-H activation via halogenation followed by metalation of carbon-halogen bonds is of particular importance for selective chemical modification via metal-catalyzed cross-coupling reactions.[11–13]

Though halogenated compounds are widely used in many sectors, the chemical processes for site specific halogenation of small as well as complex...
organic molecular backbones often remains challenging.\textsuperscript{[14]} The traditional halogenation reactions are via noxious, environmentally hazardous processes, harsh reaction conditions and rely on toxic reagents and solvents.\textsuperscript{[15,16]} The chemical halogenation routes also encounter poor regioselectivity, lower space-time yields and formation of unwanted side products that requires tedious chemical separation methods.\textsuperscript{[17,18]} Enzymatic halogenation offers a more effective and ecological route to circumvent these problems, with a potential advantage of regio and stereo selectivity; eliminating the need for protecting or activating groups and biproduct separation.\textsuperscript{[5,19]} Thus, the current synthetic developments and endeavors of halogenation chemistries are to a large extent inspired by nature.\textsuperscript{[20]} In parallel, there is a growing interest in identifying, exploring and exploiting biohalogenases to create biosynthetic pathways of diverse molecular structures, including metabolites, complex natural products or natural product analogues.

Halogenation chemistry in nature is diverse with halogenases from biosynthetic gene clusters jointly responsible for halogenating more than 5000 halogenated biomolecules. It was half a century ago, the first report of an enzymatic halogenation – a heme-dependent haloperoxidase isolated from \textit{Caldariomyces fumago} – was published.\textsuperscript{[21,22]} Since then, many halogenases are discovered from different biosynthetic pathways that fall under one of the four main classes: 1) haloperoxidases (heme-dependent and vanadium-dependent), 2) flavin-dependent halogenases, 3) Fe (II)/2-(oxo)-glutarate-dependent halogenases and 4) S-adenosyl-L-methionine (SAM) dependent halogenases. The halogenation chemistry of these enzymes either follow electrophilic, nucleophilic or radical reaction mechanisms (Figure 1).\textsuperscript{[23]} Most reported biohalogenation reactions are with chlorine and bromine whereas halide preference for fluorine is limited to SAM fluorinases. The homologous enzymes of chlorinases present in marine organisms often display a higher bromide preference over chloride.\textsuperscript{[24]} Iodination reactions are reported from all class of halogenases except for Fe (II)/2-(oxo)-glutarate-dependent halogenases.\textsuperscript{[25–27]}

The halogenase enzymes from the above classes have a wider interest owing to its biocatalytic applications and potential utility for chemical and bio-manufacturing industries. Reflecting the critical importance, for the last many years, a veritable explosion has been observed in research articles that are dealing either with identification, engineering or developing potential applications with bio-halogenase enzymes (Table 1). There are several excellent reviews on mechanism and classification, engineering biohalogenases, halogenases from natural product pathways, application and development of halogenases etc.\textsuperscript{[20,23,60–64]} However, the use of halogenase enzymes for targeted installation of halogen moiety into both natural and synthetic scaffolds via pathway generation is rarely discussed. This is an attractive prospect for many specific synthetic biology applications and for the creation of bio-renewable routes to both naturals and non-naturals. Combined with the newest trend of heavy-metal-
based chemo-enzymatic cross-coupling chemistries, this opens up new ways of chemo-bio integrated synthetic exploitation of halogen handles for bio-orthogonal molecular derivatization and in designing larger molecular screening libraries. Here, we provide an overview of synthetic biology and biosynthetic applications of halogenases that are reported to date, highlighting the engineering efforts to incorporate halogenases in natural product and recreated biosynthetic pathways either in vivo or in vitro. The goal of this review is not to
## Table 1: List of key bio-halogenase enzymes identified and characterized from different biosynthetic pathways.

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cl</td>
<td>PmA</td>
<td>Tryptophan</td>
<td>Pyoluteorin</td>
<td>Pseudomonas fluorescens</td>
<td>2ARD</td>
</tr>
<tr>
<td>2</td>
<td>Cl</td>
<td>PyrH</td>
<td>Tryptophan</td>
<td>Pyrroindomycin</td>
<td>Streptomyces rugosporus</td>
<td>2WET</td>
</tr>
<tr>
<td>3</td>
<td>Cl</td>
<td>RebH</td>
<td>Tryptophan</td>
<td>Recbemycin</td>
<td>Lechevaleria aerocolonigenes</td>
<td>2O9Z</td>
</tr>
<tr>
<td>4</td>
<td>Cl</td>
<td>SthH</td>
<td>Tryptophan</td>
<td>Unknown NRPS cluster</td>
<td>Streptomyces toxytrici</td>
<td>5HYS</td>
</tr>
<tr>
<td>5</td>
<td>Cl</td>
<td>KtzQ &amp; KtzR</td>
<td>Tryptophan</td>
<td>Kutzneride</td>
<td>Kutzneria sp. 744</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>I</td>
<td>VirX1</td>
<td>Tryptophan</td>
<td>Unknown</td>
<td>Cyanophage Syn 10</td>
<td>6QGM</td>
</tr>
<tr>
<td>7</td>
<td>Cl</td>
<td>ClaH</td>
<td>Tryptophan</td>
<td>Cladoniamide</td>
<td>Streptomyces uncialis</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>Cl</td>
<td>Thal or ThdH</td>
<td>Tryptophan</td>
<td>Thienodolin</td>
<td>Streptomyces albogriseolus</td>
<td>6H44</td>
</tr>
<tr>
<td>9</td>
<td>Cl</td>
<td>XszenFHal</td>
<td>Tryptophan</td>
<td>Unknown</td>
<td>Xenorhabdus szentirmaii</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>Cl</td>
<td>Th_Hal</td>
<td>Tryptophan</td>
<td>Unknown</td>
<td>Streptomyces violaceus niger</td>
<td>5LV5</td>
</tr>
<tr>
<td>11</td>
<td>Cl</td>
<td>SpmH</td>
<td>Tryptophan</td>
<td>Spiroindimicins/indimicins</td>
<td>Streptomyces sp. SC500 01032</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>Cl</td>
<td>BorH</td>
<td>Tryptophan</td>
<td>Borregomycin A</td>
<td>DNA isolated from Anza-Borrego Desert soil</td>
<td>6UL3</td>
</tr>
<tr>
<td>13</td>
<td>Cl</td>
<td>PrnC</td>
<td>Monodechloroaminopyrrolnitrin</td>
<td>Pyrrolnitrin</td>
<td>Pseudomonas fluorescens</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>Br</td>
<td>BrnH</td>
<td>Indolea</td>
<td>Unknown</td>
<td>Brevundimonas sp. BAL3</td>
<td>6FRI</td>
</tr>
<tr>
<td>15</td>
<td>Cl</td>
<td>MalA</td>
<td>Premalbrancheamide</td>
<td>Malbrancheamide</td>
<td>Malbranchea aurantia</td>
<td>5WGR</td>
</tr>
</tbody>
</table>

### Carrier protein (CP) dependent Pyrrole Fl-Hal

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Cl</td>
<td>P1tA</td>
<td>Pyrrolyl-S-P1tL</td>
<td>Pyoluteorin</td>
<td>Pseudomonas fluorescens</td>
<td>5DB1</td>
</tr>
<tr>
<td>17</td>
<td>Cl</td>
<td>Clz5</td>
<td>Pyrrolyl-S-Clz18</td>
<td>Chiorizidine A</td>
<td>Streptomyces Sp CNH-287</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>Cl</td>
<td>Mpy16</td>
<td>Pyrrolyl-S-Mpy15</td>
<td>Marinopyrrole</td>
<td>streptomyces sp. CNQ-418</td>
<td>5BU8</td>
</tr>
<tr>
<td>19</td>
<td>Br</td>
<td>Bmp2</td>
<td>Pyrrolyl-S-Bmp1</td>
<td>Bromopyrroles</td>
<td>Pseudoaeteromonas sp.</td>
<td>SBVA</td>
</tr>
<tr>
<td>20</td>
<td>Cl</td>
<td>HrmQ</td>
<td>Pyrrole-S-Cpa</td>
<td>Hormaomycin</td>
<td>Streptomyces griseoflavus W-384</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>Cl</td>
<td>Pyr29</td>
<td>Pyrrole-S-Cpa</td>
<td>Pyrrolomycin</td>
<td>Actinosporangium vitaminophilum ATCC 31673</td>
<td>-</td>
</tr>
</tbody>
</table>

### Free standing Phenol Fl-Hal

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>Cl</td>
<td>RadH</td>
<td>Monocillin II</td>
<td>Radicicol</td>
<td>Chaetomium chiversi</td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>Cl</td>
<td>Rdc2</td>
<td>Monocillin II</td>
<td>Radicicol</td>
<td>Pachonia chlamydospora</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>Cl</td>
<td>AcOTAhal</td>
<td>Ochratoxin B</td>
<td>Ochratoxin</td>
<td>Aspergillus carbonarius</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>Cl</td>
<td>P1tM</td>
<td>Phloroglucinol</td>
<td>2,4-diacetylphloroglucinol</td>
<td>Pseudomonas protegens PF-5</td>
<td>6BZ2</td>
</tr>
<tr>
<td>26</td>
<td>Cl</td>
<td>ChiA</td>
<td>(2,4,6-trihydroxyphenyl)-1-hexan-1-one</td>
<td>Differentiation-inducing factor 1</td>
<td>Dictyostelium purpureum</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td>Cl</td>
<td>GedLb</td>
<td>Sulochrina</td>
<td>Dihydroglедин</td>
<td>Aspergillus terreus</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>Cl</td>
<td>Gsfβ</td>
<td>Dechlorogriseofulvina</td>
<td>Griseofulvin</td>
<td>Penicillium aethopicum</td>
<td>-</td>
</tr>
</tbody>
</table>

(Continued)
Table 1.: (Continued).

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>Cl</td>
<td>AcI&lt;sup&gt;H&lt;/sup&gt;</td>
<td>Dechloroastrilochlorinea</td>
<td>Aspilochlorine</td>
<td>Aspergillus oryze</td>
<td>– –</td>
</tr>
<tr>
<td>30</td>
<td>Cl</td>
<td>Caz&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Preciastromobromene</td>
<td>Chaetoverdin</td>
<td>Chaetomium globosum</td>
<td>– –</td>
</tr>
<tr>
<td>31</td>
<td>Cl</td>
<td>Asm&lt;sup&gt;2b&lt;/sup&gt;</td>
<td>Preanamitocina</td>
<td>Asansitocin</td>
<td>Actinosynnema pretiosum</td>
<td>– –</td>
</tr>
<tr>
<td>32</td>
<td>Cl</td>
<td>Nat&lt;sup&gt;1b&lt;/sup&gt;</td>
<td>Prenapthomycin</td>
<td>Napyradiomycin</td>
<td>Streptomyces sp. CNQ-525</td>
<td>– –</td>
</tr>
<tr>
<td>33</td>
<td>Cl</td>
<td>NapH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Naphthoquinone precursor</td>
<td>Napyradiomycin</td>
<td>Streptomyces sp. ATCC 53789</td>
<td>– –</td>
</tr>
<tr>
<td>34</td>
<td>Cl</td>
<td>Bmp5</td>
<td>4-hydroxybenzoic acid</td>
<td>Bromo phenols</td>
<td>Pseudoalteromonas leteoviolacea</td>
<td>– –</td>
</tr>
</tbody>
</table>

**Carrier protein (CP) dependent Phenol Fl-Hal**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>Cl</td>
<td>CndH</td>
<td>Tyrosyl-S-CPa</td>
<td>Chondrochloren</td>
<td>Chondromyces croatus</td>
<td>3E1T&lt;sup&gt;40&lt;/sup&gt;</td>
</tr>
<tr>
<td>36</td>
<td>Cl</td>
<td>SgcC3</td>
<td>Tyrosyl-SgcC2</td>
<td>Enediyne Antibiotic C-1027</td>
<td>Streptomyces globisporus</td>
<td>– –</td>
</tr>
<tr>
<td>37</td>
<td>Cl</td>
<td>BhaA</td>
<td>β-Hydroxy tyrosyl-S-CP</td>
<td>Balihimycin</td>
<td>Amycolatopsis bialyimycina</td>
<td>– –</td>
</tr>
<tr>
<td>38</td>
<td>Cl</td>
<td>ComH</td>
<td>Tyrosyl-S-CPa</td>
<td>Complestatin</td>
<td>Streptomyces lavendulae</td>
<td>– –</td>
</tr>
<tr>
<td>39</td>
<td>Cl</td>
<td>End3D</td>
<td>Hydroxyphenylglycine tethered CP</td>
<td>Enduracitin</td>
<td>Streptomyces fungicidicus</td>
<td>– –</td>
</tr>
<tr>
<td>40</td>
<td>Cl</td>
<td>Ram20</td>
<td>Hydroxyphenylglycine tethered CP</td>
<td>Ramoplanin</td>
<td>Actinoplanes subtrigicus</td>
<td>– –</td>
</tr>
<tr>
<td>41</td>
<td>Cl</td>
<td>Clo-Hal</td>
<td>Tyrosyl-S-CP or β-Hydroxy tyrosyl a</td>
<td>Chlorobiocin</td>
<td>Streptomyces roseochromogenus</td>
<td>– –</td>
</tr>
<tr>
<td>42</td>
<td>Cl</td>
<td>Tcp21</td>
<td>Tyrosyl-S-CP or β-Hydroxy tyrosyl a</td>
<td>Teicoplanin</td>
<td>Actinoplanes teichomyceticus</td>
<td>– –</td>
</tr>
<tr>
<td>43</td>
<td>Cl</td>
<td>CrpH</td>
<td>Tyrosyl-S-CPa</td>
<td>Cryptophycin</td>
<td>Nostoc sp. ATCC 53789</td>
<td>– –</td>
</tr>
<tr>
<td>44</td>
<td>Cl</td>
<td>TiaM&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Orsellinic acid moiety of tiacumicin-S-CPa</td>
<td>Tiacumicin</td>
<td>Dactylorosporangium aurantium</td>
<td>– –</td>
</tr>
<tr>
<td>45</td>
<td>Cl</td>
<td>VhaA</td>
<td>Hexapeptide-peptidy CP conjugate intermediate of vancomycin</td>
<td>Vancomycin</td>
<td>Amycolatopsis orientals DSM40040</td>
<td>– –</td>
</tr>
<tr>
<td>46</td>
<td>Cl</td>
<td>AerJ</td>
<td>β-Hydroxy phenylleic acid -S-AerAa</td>
<td>Aeruginosin</td>
<td>Microcystis aeruginosa NIES-98</td>
<td>– –</td>
</tr>
<tr>
<td>47</td>
<td>Ia</td>
<td>CalO3</td>
<td>2-Methoxyorssellinate-S-CP</td>
<td>Calicheamicin</td>
<td>Micromonospora echinospora</td>
<td>– –</td>
</tr>
</tbody>
</table>

**Aliphatic Fl-Hal**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>Cl</td>
<td>ClmS</td>
<td>CoA-thioester (ClmK)/Acetyl-CoAa</td>
<td>Chloramphenicol</td>
<td>Streptomyces venezuelae</td>
<td>3BL&lt;sup&gt;41&lt;/sup&gt;</td>
</tr>
<tr>
<td>49</td>
<td>Cl</td>
<td>AciQ</td>
<td>Diaporthin</td>
<td>Diaporthins</td>
<td>Aspergillus oryzae</td>
<td>– –</td>
</tr>
</tbody>
</table>

**Free standing α-KG-Hals**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>Cl</td>
<td>WelO5</td>
<td>12-epi-fischerindole U</td>
<td>Welwitindoline</td>
<td>Hapalosiphon welwitschii</td>
<td>5IQV&lt;sup&gt;42&lt;/sup&gt;</td>
</tr>
<tr>
<td>51</td>
<td>Cl</td>
<td>AmbOS</td>
<td>Ambiguine</td>
<td>Ambiguine</td>
<td>Fischerella ambiguia UTEX 1930</td>
<td>– –</td>
</tr>
<tr>
<td>52</td>
<td>Cl</td>
<td>BesD</td>
<td>L-lysine</td>
<td>β-ethylserine</td>
<td>Streptomyces cattleya</td>
<td>– –</td>
</tr>
<tr>
<td>53</td>
<td>Cl</td>
<td>AdeV</td>
<td>2-deoxyadenosine monophosphate</td>
<td>Adechlorin</td>
<td>Actinomadura sp. ATCC 39365</td>
<td>– –</td>
</tr>
<tr>
<td>54</td>
<td>Cl</td>
<td>DAH</td>
<td>Dechloroacutumine</td>
<td>(-) acutumine</td>
<td>Menispermaceae</td>
<td>– –</td>
</tr>
<tr>
<td>55</td>
<td>Cl</td>
<td>HalA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>L-lysine&lt;sup&gt;a&lt;/sup&gt;</td>
<td>L-lysine</td>
<td>Pseudomonas pseudoavonan</td>
<td>– –</td>
</tr>
<tr>
<td>56</td>
<td>Cl</td>
<td>HalB&lt;sup&gt;c&lt;/sup&gt;</td>
<td>L-lysine&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Unknown</td>
<td>Streptomyces iranensis</td>
<td>– –</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>Cl</td>
<td>HalC</td>
<td>L-lysine</td>
<td>Unknown</td>
<td>Legionella anisa</td>
<td>– –</td>
</tr>
<tr>
<td>58</td>
<td>Cl</td>
<td>HalD</td>
<td>L-Ornithine</td>
<td>Unknown</td>
<td>Pseudomonas kiliensis</td>
<td>– –</td>
</tr>
<tr>
<td>59</td>
<td>Cl</td>
<td>HalE</td>
<td>(L-leucine, L-isoleucine, L-norleucine)</td>
<td>Unknown</td>
<td>Pseudomonas sp. Root562</td>
<td>– –</td>
</tr>
</tbody>
</table>

**Carrier protein (CP) dependent a-KG-Hals**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>Cl</td>
<td>SyrB2</td>
<td>L-Thr-S-SyrB1</td>
<td>Syringomycin</td>
<td>Pseudomonas syringae</td>
<td>2FC[44]</td>
</tr>
<tr>
<td>61</td>
<td>Cl</td>
<td>KtzD</td>
<td>L-isoleucine-S-KtzC</td>
<td>Kutzneride</td>
<td>Kutzneria sp. 744</td>
<td>– –</td>
</tr>
<tr>
<td>62</td>
<td>Cl</td>
<td>KthP</td>
<td>(3S)-piperazyl-S-KtzC</td>
<td>kutzneride</td>
<td>Kutzneria sp. 744</td>
<td>– –</td>
</tr>
<tr>
<td>63</td>
<td>Cl</td>
<td>CmaB</td>
<td>L-allo-isoleucine-S-CmaD</td>
<td>Coronamic acid</td>
<td>Pseudomonas syringae</td>
<td>– –</td>
</tr>
<tr>
<td>64</td>
<td>Cl</td>
<td>BarB1</td>
<td>L-leucine-S-BarA</td>
<td>Barbamide</td>
<td>Lyngbya majuscule</td>
<td>– –</td>
</tr>
<tr>
<td>65</td>
<td>Cl</td>
<td>BarB2</td>
<td>L-leucine-S-BarA</td>
<td>Barbamide</td>
<td>Lyngbya majuscule</td>
<td>– –</td>
</tr>
<tr>
<td>66</td>
<td>Cl</td>
<td>HctB</td>
<td>Hexanoyl-S-CP</td>
<td>Hectochlorin</td>
<td>Lyngbya majuscule</td>
<td>– –</td>
</tr>
<tr>
<td>67</td>
<td>Cl</td>
<td>CytC3</td>
<td>Aminobutyryl-S-CytC</td>
<td>g, g-dichloroaminobutyrate</td>
<td>Streptomyces sp.</td>
<td>3GJB[45]</td>
</tr>
<tr>
<td>68</td>
<td>Cl</td>
<td>Thr3</td>
<td>L-Threonine-S-CPa</td>
<td>Unknown</td>
<td>Streptomyces sp.</td>
<td>– –</td>
</tr>
<tr>
<td>69</td>
<td>Cl</td>
<td>CurA</td>
<td>(S)-3-hydroxy-3-methylglutaryl-S-CP</td>
<td>Curacin A</td>
<td>Lyngbya majuscule</td>
<td>3NN[46]</td>
</tr>
<tr>
<td>70</td>
<td>Cl</td>
<td>JamE</td>
<td>(S)-3-hydroxy-3-methylglutaryl-S-CP</td>
<td>Jamaicamide</td>
<td>Lyngbya majuscule</td>
<td>– –</td>
</tr>
</tbody>
</table>

**SAM-dependent halogenases**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>71</td>
<td>F</td>
<td>FIA</td>
<td>S-adenosylmethionine</td>
<td>Fluoroacetate/fluorometabolites</td>
<td>Streptomyces cattleya</td>
<td>2CC[47]</td>
</tr>
<tr>
<td>72</td>
<td>F</td>
<td>FIA_homologue</td>
<td>S-adenosylmethionine</td>
<td>Fluoroacetate/fluorometabolites</td>
<td>Streptomyces sp. MA37</td>
<td>586[48]</td>
</tr>
<tr>
<td>73</td>
<td>F</td>
<td>NobA</td>
<td>S-adenosylmethionine</td>
<td>Fluoroacetate/fluorometabolites</td>
<td>Nocardia brasiliensis</td>
<td>– –</td>
</tr>
<tr>
<td>74</td>
<td>F</td>
<td>FIA_homologue</td>
<td>S-adenosylmethionine</td>
<td>Fluoroacetate/fluorometabolites</td>
<td>Actinoplanes sp</td>
<td>– –</td>
</tr>
<tr>
<td>75</td>
<td>F</td>
<td>FIA4</td>
<td>S-adenosylmethionine</td>
<td>Fluoroacetate/fluorometabolites</td>
<td>Streptomyces xinghaiensis NRRL B24674</td>
<td>– –</td>
</tr>
<tr>
<td>76</td>
<td>Cl</td>
<td>SalL</td>
<td>salinosporamide</td>
<td>Salinispora tropica CNB-44023</td>
<td>– –</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>Cl</td>
<td>CIA1</td>
<td>g-butyrolactone (GBL)a</td>
<td>Streptomyces algicola</td>
<td>– –</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>Cl</td>
<td>CIA2</td>
<td>S-adenosylmethionine</td>
<td>Unknown</td>
<td>Umezawaea tangerina NRRL B-24463</td>
<td>– –</td>
</tr>
</tbody>
</table>

**Vanadium-dependent haloperoxidases (non-regioselective)**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td>Br, I</td>
<td>Am_vHPO</td>
<td>Various substrates</td>
<td>Holoalkane production</td>
<td>Acaryochloris marina</td>
<td>5LPC[49]</td>
</tr>
<tr>
<td>80</td>
<td>Br, I</td>
<td>Sy_vHPO</td>
<td>Various substrates</td>
<td>Holoalkane production</td>
<td>Synechococcus sp. CC9311</td>
<td>– –</td>
</tr>
<tr>
<td>81</td>
<td>Br, I</td>
<td>Co_vHPO</td>
<td>Various substrates</td>
<td>Holoalkane production</td>
<td>Corallina officinalis</td>
<td>1QI9[50]</td>
</tr>
<tr>
<td>82</td>
<td>Br, I</td>
<td>Cp_vHPO</td>
<td>Various substrates</td>
<td>Holoalkane production</td>
<td>Corallina pilulifera</td>
<td>1UP[51]</td>
</tr>
<tr>
<td>83</td>
<td>Br, I</td>
<td>As_vHPO</td>
<td>Various substrates</td>
<td>Holoalkane production</td>
<td>Ascophyllum nodosum</td>
<td>1QR[52]</td>
</tr>
<tr>
<td>84</td>
<td>Cl, Br</td>
<td>CI_vHPO</td>
<td>Various substrates</td>
<td>Holoalkane production</td>
<td>Curvularia inaequalis</td>
<td>1VN[53]</td>
</tr>
<tr>
<td>85</td>
<td>Cl, Br</td>
<td>Zg_vHPO</td>
<td>Various substrates</td>
<td>Holoalkane production</td>
<td>Zobellia galactanivorans</td>
<td>4C[54]</td>
</tr>
</tbody>
</table>

(Continued)
Table 1.: (Continued).

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>86</td>
<td>Cl</td>
<td>NapH1</td>
<td>Naphthomevalin</td>
<td>Napyradiomycin</td>
<td>Streptomyces sp. CNQ-525</td>
<td>3W36[55]</td>
</tr>
<tr>
<td>87</td>
<td>Cl⁴</td>
<td>NapH3</td>
<td>Prenaphthomevalin</td>
<td>Napyradiomycin</td>
<td>Streptomyces sp. CNQ-525</td>
<td>--</td>
</tr>
<tr>
<td>88</td>
<td>Cl</td>
<td>NapH4</td>
<td>Napyradiomycin A1</td>
<td>Napyradiomycin</td>
<td>Streptomyces sp. CNQ-525</td>
<td>--</td>
</tr>
<tr>
<td>89</td>
<td>Cl</td>
<td>Md24</td>
<td>Premochlorin</td>
<td>Meroclorin</td>
<td>Streptomyces sp. CNH-189</td>
<td>--</td>
</tr>
<tr>
<td>90</td>
<td>Cl⁴</td>
<td>Md40</td>
<td>Meroclorin D</td>
<td>Meroclorin</td>
<td>Streptomyces sp. CNH-189</td>
<td>--</td>
</tr>
</tbody>
</table>

**Vanadium-dependent haloperoxidases (regioselective)**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>91</td>
<td>Cl, Br, I</td>
<td>Lf_HCPO</td>
<td>Various substrates</td>
<td>Caldariomycin</td>
<td>Caldariomycetes fumago or L. fumago</td>
<td>1CPO[56]</td>
</tr>
<tr>
<td>92</td>
<td>Cl, Br, I</td>
<td>C_MYP</td>
<td>Various substrates</td>
<td>Canine myeloperoxidase</td>
<td>Canis lupus familiaris</td>
<td>1MYP[57]</td>
</tr>
<tr>
<td>93</td>
<td>Cl, Br, I</td>
<td>H_MYP</td>
<td>Various substrates</td>
<td>Human myeloperoxidase</td>
<td>Homo sapiens</td>
<td>a1MHL[58]</td>
</tr>
<tr>
<td>94</td>
<td>Cl, Br, I</td>
<td>AaP</td>
<td>Various substrates</td>
<td>Unknown</td>
<td>Agrocybe aegerita</td>
<td>--</td>
</tr>
</tbody>
</table>

**Heme-dependent haloperoxidases**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>Br</td>
<td>CPO-A1</td>
<td>mono-chlorodimedone⁶</td>
<td>Unknown</td>
<td>Streptomyces aureofaciens</td>
<td>1A7U[59]</td>
</tr>
<tr>
<td>96</td>
<td>Br</td>
<td>CPO-L</td>
<td>mono-chlorodimedone⁶</td>
<td>Unknown</td>
<td>Streptomyces lividans,</td>
<td>1A86[64]</td>
</tr>
<tr>
<td>97</td>
<td>Br</td>
<td>CPO-F</td>
<td>mono-chlorodimedone⁶</td>
<td>Unknown</td>
<td>Pseudomonas fluorescens</td>
<td>1A85[59]</td>
</tr>
<tr>
<td>98</td>
<td>Br</td>
<td>CPO-P</td>
<td>mono-chlorodimedone⁶</td>
<td>Unknown</td>
<td>Pseudomonas pyrocina</td>
<td>--</td>
</tr>
<tr>
<td>99</td>
<td>Br</td>
<td>CPO-T</td>
<td>mono-chlorodimedone⁶</td>
<td>Unknown</td>
<td>S. aureofaciens Tü24</td>
<td>--</td>
</tr>
</tbody>
</table>

**Metal/cofactor free haloperoxidases**

<table>
<thead>
<tr>
<th>No</th>
<th>Halide</th>
<th>Abbreviation</th>
<th>Native substrate</th>
<th>Biosynthetic pathway</th>
<th>Native host</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Cl</td>
<td>CylC</td>
<td>Decanoyl-S-CylB</td>
<td>Cylindrycopophane</td>
<td>Cylindrospermum licheniforme</td>
<td>--</td>
</tr>
<tr>
<td>101</td>
<td>Cla</td>
<td>CoID</td>
<td>Unknown</td>
<td>Columbamide</td>
<td>Moorea bouillonii Pangs-198</td>
<td>--</td>
</tr>
<tr>
<td>102</td>
<td>Cla</td>
<td>CoIE</td>
<td>Unknown</td>
<td>Columbamide</td>
<td>Moorea bouillonii Pangs-198</td>
<td>--</td>
</tr>
<tr>
<td>103</td>
<td>Cla</td>
<td>BrtJ</td>
<td>Unknown</td>
<td>Bartolosides</td>
<td>Symehocystis salina LEGE 06155</td>
<td>--</td>
</tr>
</tbody>
</table>

---

a= Putative/consensus substrate or halide
b= In vitro substrate identification tests are not carried out. Included under each heading (free standing or carrier protein-dependent) with the limited information available.
c= Gene cluster abbreviation based on its sequence similarity to BesD enzyme.
d= No halogenation activity. Catalyze an isomerization reaction by acting on halogenated substrate.
e= Synthetic substrate for activity assays.
CP = Carrier Protein.
comprehensively summarize all halogenated natural products or mutasynthetically generated halogenated molecules; but to focus on naturals and non-naturals that are synthesized with a direct bio-halogenase incorporation in modified pathways.

2. Haloperoxidases

Haloperoxidases are the earliest class of enzymes identified with a halogenation activity (Figure 1). They utilize hydrogen peroxide produced within the active site to react with halides forming hypohalides. Hypohalides subsequently react with substrate and transfer halide moiety. In our environment, haloperoxidases often play a key role in biogenic organohalogen production. Haloperoxidases are widely used in organic synthesis for various bio-chemo synthetic applications. Reactions such as halohydroxylation, oxidative decarboxylation, halolactonisation using haloperoxidases were reported. The two main types of haloperoxidases currently known are heme-iron dependent and vanadium-dependent haloperoxidases (Figure 2). A class of haloperoxidases initially known as metal and cofactor-free haloperoxidases were reported previously by several groups.\cite{59,65} The structural studies reveal that this specific class of enzymes carry a catalytic triad (Ser-His-Asp) and an \( \alpha/\beta \) hydrolase fold found usually in either lipases, esterases, and serine proteases enzymes.\cite{66,67} Metal and cofactor-free haloperoxidases display brominase activity on mono-chlorodimedone 1 in the presence of acetate, hydrogen peroxidase and bromide. This haloperoxidase activity was later questioned by other groups and was proposed as a commonly unfavorable side reaction of hydrolases enzymes.\cite{68} Hence, further discussion on

![Figure 2. Structural comparison of vanadium and heme dependent haloperoxidases. a) Vanadium dependent haloperoxidases from Acaryochloris marina (PDB ID: 5LPC), b) heme dependent Canine myeloperoxidase (PDB ID:1MYP). Active site residues that interact with vanadium and heme cofactors are shown in the inset.](image-url)
artifact halogenase activity with the third class of haloperoxidase enzymes are not included in this review. Many focused reviews on haloperoxidases and its applications in organic synthesis, structure, sources, mechanisms, enzymology etc are available for further reading.^[69–75]

### 2.1. Heme-iron-dependent haloperoxidases (HPOs)

The heme-iron dependent haloperoxidases (HPOs) utilize free hypohalous acid released via halide oxidation of Fe (IV) oxo species generated during catalysis (Figure 1). Electron-rich substrates that are directly accessible to the heme centre or to the diffused hypohalous acid could undergo nonspecific halogenation. Thus, HPOs are halogenating catalysts without any substrate specificity or regioselectivity. The rapid enzyme inactivation is a common drawback for HPO reaction as the diffused hypohalous acid reacts with its own amino acids. The heme-dependent chloroperoxidases (HCPOs) could oxidize all halide ions except fluoride ions for halogenation.

HCPO has structural similarity to other heme proteins such as cytochrome P450s and heme peroxidases, enabling it to perform different oxidative hemoproteins reactions and ability to catalyze multiple oxidative functions. The proximal heme ligand of HCPO is either a cysteine thiolate or a histidine as in P450s and other hemoproteins, whereas the distal side contains polar amino acid residues including a glutamate (and/or aspartate) as in peroxidases (Figure 2).^[76] This distal ligand provides ‘peroxidase-like’ activity that can cleave heme peroxide O–O bonds during catalysis.^[56] Organic substrates interact via the distal face and it is the P450 and peroxidase-like structural features that provide peroxidase, catalase and P450-like reactivities to HCPO enzymes, along with the actual halogenation capabilities. HCPO reactions prefer low pH conditions, and halogenation is possible with a range of substrates including phenols, aromatic acids, flavonoids, hydrocarbons, lignins, biphenyls, steroids etc.^[77] The HPOs are widely distributed in nature, present in bacteria, fungi and even in higher organisms. The known mammalian haloperoxidases such as myeloperoxidase (MPO), lactoperoxidase (LPO) and eosinophil peroxidase (EPO) are also HPOs^[78–80].

#### 2.1.1. Heme-iron-dependent haloperoxidases from natural product pathways

The HCPO from the caldariomycin biosynthesis pathway in fungus *Caldariomyces fumago* was the first identified HPO involved in a halometabolite biosynthesis.^[22,81] *C. fumago* HCPO catalyzes the chlorination of β-ketoadipic acid to δ-chlorolevulinic acid during the formation of antibiotic caldariomycin 2.^[22] *C. fumago* HCPO could also dihalogenate 1,3-cyclopentandione 3, which is an intermediate in caldariomycin synthesis. The difference in halide preference of HPOs are normally related to the halide binding site than the halide oxidizing ability. It is also suggested that HCPOs
have specific chloride binding sites.\[^{82}\] The second fungal haloperoxidase discovered was from agaric mushroom Agrocybe aegerita. This enzyme has a strong brominating activity and was also shown to be a potential catalyst for different oxy-functionalization reactions. A. aegerita HPO could catalyze cyclohexane hydroxylation, which is a reaction that has never been observed with other HCPOs. Though there are several biosynthetic HCPOs such as bromoperoxidases from Pseudomonas aureofaciens and Penicillium capitatus that are known, no bioengineering or genomic engineering strategies were investigated with HCPOs. This could be due to the fact that these enzymes are prone to fast enzyme inactivation, along with other known issues such as no substrate specificity and non-regioselectivity.\[^{83–85}\]

### 2.2. Vanadium-dependent haloperoxidases (vHPO)

Vanadium-dependent haloperoxidases (vHPO) enzymes contain a vanadate prosthetic group and oxidizes halide ions (X\(^-\)) into a reactive hypohalite ('OX) intermediate by utilizing hydrogen peroxide (Figure 1). vHPOs are classified into vanadium-dependent chloroperoxidase (vCPO), vanadium-dependent bromoperoxidase (vBPO) and vanadium-dependent iodoperoxidase (vIPO), depending on the oxidation ability of halogen ions involved. vCPO oxidizes chloride, bromide and iodide. vBPO oxidizes bromide and iodide whereas vIPO only oxidizes iodide ions. vHPO and homologous metalloenzymes are widely distributed in nature and are present in algae, fungi, bacteria and in higher eukaryotes. They play a key role in the production of biogenic organohalogens in Earth’s atmosphere and in oceans.\[^{86,87}\] vBPOs present in marine micro algae are extracellular enzymes making it easily reactive to the hydrogen peroxide produced from marine plants during environmental stress. Hypobromic acid released from the active site of vBPO could halogenate the dissolved organic matter (mainly methane, halogenated methane and keto acids) in seawater. It is estimated that vBPOs from diatoms and algae are thus responsible for the production of over 2 million tons of bromoform and 56,000 tons of bromomethane released to the atmosphere annually.\[^{88}\]

The resolved crystal structure of vHPO exhibits striking similarities to each other with the arrangement of active site residues and co-ordination of vanadium in the active site (Figure 2). Vanadium is co-ordinated to a conserved histidine residue and the cofactor is in a trigonal bipyramidal geometry with bound oxygen atoms positioned in the equatorial direction.\[^{89,90}\] The active site is at the bottom of a 15–20-Å deep funnel-shaped channel that anchors an extensive and conserved hydrogen-bonding network to stabilize negative charges.\[^{91}\] Computational and spectroscopic studies have suggested the presence of a hydroxyl group or water molecule in the axial or apical position.\[^{92}\] The hydroxyl and/or the water molecule is thought to be stabilized by active site hydrogen-bonding interactions. During
catalysis, the axial water/hydroxyl molecule and a conserved histidine residue play an important role in the hydrogen peroxide co-ordination, followed by the formation of a peroxo intermediate.\[74\] This generates the active hypohalous acid for the nonselective halogenation catalysis on nucleophilic substrates that are in the close proximity to the active site.

The major advantage of vHPO over heme-dependent haloperoxidases is its high robustness in the presence of hydrogen peroxide and ability to create clean reaction products. The oxidation state of vanadium centre is maintained throughout the catalytic cycle, which avoids oxidative inactivation often displayed by heme-dependent haloperoxidases. vHPO shows excellent thermal stability as well as solvent and pH tolerance. In addition, few vHPOs are commercially available and it is also possible to recombinantly express the enzymes in industrial host cells such as E. coli and yeast. There are several important reviews already available on the role of vHPOs in natural product synthesis, organic synthesis as well as chemistry, mechanism and importance of vHPO enzymes.\[69,73,87,91\]

2.2.1. Reconstructed biosynthetic and natural product pathways with vHPO

For a long time, vHPOs are considered to have limited regio, chemo and enantioselectivity due to their mechanisms based on free hypohalous acid dissipation. This was challenged initially by the biotransformation carried out by vHPOs isolated from marine red algae such as Corallina officinalis, Laurencia pacifica and Plocamium cartilagineum.\[93\] C. officinalis vHPO was found to be very selective and asymmetrically brominate and cyclize sesquiterpenic molecules (E)-(+) -nerolidol 4 to bromo alcohols α-, β-, and γ-snyderol 5–7 (Figure 3). No snyderol diastereoisomer are formed in the absence of vHPO, indicating that oxidation of bromide by a peroxo–vHPO complex is necessary for the bromination and the subsequent cyclization. Similar results were also observed with monoterpenes nerol 8 and geraniol 9, where bromo cyclized products were only produced in the presence of vHPOs.\[94\] Later, the identification of regiospecific vHPOs and their homologous enzymes from natural product

![Figure 3. Substrates and products of non-regioselective vanadium dependent haloperoxidases.](image-url)
biosynthetic gene cluster for meroterpenoids paved the way for further details on stereo and regiospecific reactions that vHPO could perform.

Napyradiomycin and merochlorin are a group of intriguing molecules with a polyketide naphthoquinone–terpenoid hybrid scaffold, produced by different *Streptomyces* species through halo-functionalization and a followed ring closure involving alkene and arene units (Figure 4). The napyradiomycin biosynthetic
gene cluster in *Streptomyces aculeolatus* NRRL 18422 and *Streptomyces* sp. CNQ-525 contains three vHPO enzymes NapH1, H3, H4. For the first time, involvement of vHPO enzymes in this dedicated meroterpenoids biosynthetic pathway was investigated by studying the formation of a trichlorinated napyradiomycin molecules from napyradiomycin biosynthesis.\(^{95,96}\) This has revealed that the napyradiomycin structure is derived from a symmetrical 1,3,6,8-tetrahydroxynaphthalene (THN) 10 polyketide which is isoprenylated by the mevalonate pathway. The initial terpenoid fragments undergo a stereospecific chloronium ion induced cyclization by vHPO enzymes. The *nap* locus from the native host was engineered into non-producer *Streptomyces albus* for its heterologous production.\(^{55}\) The heterologous expression and assays in *S. albus* and *E. coli* cells indicated that NapH1 catalyzes the proposed halocyclization of a prenol moiety to form a 7-methylated napyradiomycin A1 derivative in the presence of hydrogen peroxide. NapH1 is a dual acting enzyme which catalyzes chlorination and etherification (to cyclize the prenol moiety) reactions at two distinct stages in the pathway.\(^{97}\) Using a heterologous soluble expression of NapH4 in *Streptomyces lividans* TK23, it was shown that NapH4 is involved in the halogenation-induced cyclization of a geranyl moiety to form an exomethylene-containing chlorinated cyclohexane ring of napyradiomycin B1.\(^{98}\) Though napH3 has above 50% sequence identity with NapH1 and napH4, the recombinant NapH3 enzyme exhibited no halogenation activity. Instead, it mediated the C4-to-C3 α-hydroxyketone rearrangement of the geranyl moiety with synthetic and pathway intermediate to form naphthomevalin 11. This indicates that though NapH3 shares homology with other two haloperoxidases, it has evolved into an exclusive and selective catalyst for the requisite isomerization reaction in the pathway; acting merely on a halogenated substrate.\(^{97}\)

The detailed understanding and identification of the activities of vHPO enzymes and two additional Mg\(^{2+}\)-dependent prenyltransferase (PTase) enzymes (NapT8 and T9) allowed to create a minimal engineered napyradiomycin biosynthetic pathway using recombinant and purified enzymes (Figure 4a). In this recreated pathway, the prenyltransferase NapT9 catalyzes the geranylation of 1,3,6,8-tetrahydroxynaphthalene at the nucleophilic 4-position, which is further oxidatively dearomatized and monochlorinated by NapH1. NapT8 catalyzed prenylation with dimethylallyl pyrophosphate (DMAPP) 14 and a followed NapH3-catalyzed α-hydroxyketone rearrangement forms naphthomevalin 11 molecule. The one pot chemo-enzymatic approaches using three vHPO and two PTase enzymes, in combination, produced napyradiomycin A1 and B1 12–13 up to a milligram scale. The total enantioselective chemical synthesis of these molecules is very challenging; often encountered with lower yields. The one pot chemo-enzymatic *in vitro* routes with minimal enzymes has greater advantage and it is highly likely that these methods could be engineered into a suitable host in near future for *in vivo* large-scale production.\(^{99}\)
Natural product merochlorin biosynthetic cluster, \textit{mcl}, identified from marine-derived bacterium \textit{Streptomyces} sp. CNH-189 contains two vCPO homologs (Mcl24 and Mcl40) and a PTase (Mcl23) (Figure 4c).\cite{100} The heterologous expression studies showed that merochlorins were assembled by chloroetherification reaction on THN scaffold similar to that observed in napyradiomycin biosynthesis. Mcl23 attach an isosesquilavandulyl pyrophosphate 15 to THN structure. Mcl22, a prenyl diphosphate synthase enzyme in the pathway initiate the coupling of DMAPP 14 and geranyl pyrophosphate (GPP) to form isosesquilavandulyl pyrophosphate 15. Mcl24 showed similarity to NapH1, which follows an intramolecular cyclization via a chloronium-induced oxidative dearomatization of the substrate pre-merochlorin 16 to create merochlorin A and B 17–18.\cite{38,101,102} Mcl40 is proposed to be involved in the construction of merochlorin C 19, a 15-membered cyclic ether ring from merochlorin D, via one of the largest natural olefin halofunctionalization reaction (Figure 4b). Mcl24 displayed no activity toward simple terpene alcohols and phenolic and naphtholic substrate analogues, indicating that this halogenation is substrate as well as regiospecific. The merochlorin pathway was in vitro constituted using Mcl17 (THN polyketide synthases), Mcl22, Mcl23 and Mcl24 with external addition of DMAPP, GPP and malonyl CoA metabolites. This has produced both merochlorin A 17 and merochlorin B 18 in analytical quantities along with premerochlorin through a one pot total enzymatic synthesis for the first time.\cite{38}

Though not experimentally proven, the presence of a conserved lysine residue in the vicinity of vanadate cofactor that forms a chloramine type intermediate as in the case of Fl-Hals were speculated for site-specific halogenation of merochlorin and napyradiomycin vHPOs. Identification and reconstruction of new vHPOs from other natural product biosynthetic gene clusters such as recently identified vHPOs from naphterin and marinone biosynthetic cluster and from unexplored cyanobacterial species will shed more light in this direction.\cite{103,104,105,106}

3. Flavin-dependent halogenases (Fl-Hals)

The discovery of flavin-dependent halogenase from pyrrolnitrin 19 biosynthesis demonstrated the existence of an alternative regiospecific halogenation mechanism in nature for the first time.\cite{107} Fl-Hals are evolved from a superfamily of flavin-dependent monooxygenases (FMOs) that activates molecular oxygen using reduced flavin (FADH$_{2}$) to generate hydroperoxy flavin (FAD-OOH) (Figure 1). Fl-Hals based halogenation chemistry utilizes this initial activation mechanism, followed by the reaction of halide anion with hydroperoxy flavin to produce hypohalous acid.

The overall structural similarity of Fl-Hals to flavin-dependent monooxygenases (FMOs) and the existence of FMO like regions in Fl-Hals were
known from the earliest resolved crystal structures of Fl-Hal enzymes.\textsuperscript{[28]} However, the resolved crystal structures of Fl-Hals showed that substrate and flavin-binding sites in Fl-Hals are spatially distinct and are separated by over 10 Å thus avoiding any direct substrate-flavin interaction as observed in FMOs (Figure 5).\textsuperscript{[28]} The hypohalous acid produced from cofactor is thought to be guided inside a channel toward an active site lysine to form a long-lived enzyme-chloride adduct.\textsuperscript{[108]} The chemical nature of this adduct is still under debate, which could either be a covalent chloramine species or a lysine-hypochlorous acid adduct formed via hydrogen bonding.\textsuperscript{[109]} The substrate halogenation site is often oriented toward the active site lysine residue and it has a vital role in reaction chemistry and in deciding regioselectivity of the substrate halogenation position. A conserved glutamate residue in the active site is responsible for the abstraction of a proton from the Wheland intermediate formed by substrate interaction with electrophilic chlorine species. This evolved and finely controlled enzymatic halogenation mechanism for regioselectivity makes the Fl-Hals to differ from other halogenase enzymes, making it extremely useful for biocatalytic and biosynthetic applications.\textsuperscript{[20]}

The specific roles played by structural dynamics and active site residues in the regio selective halogenation mechanism were being investigated by several
recent computational and spectroscopic studies. The quantum mechanics (QM)/molecular mechanics (MM) approaches supported the key atomistic interactions in the reaction path confirming the respective proton acceptor and donor roles of active site lysine and glutamate residues. A structurally conserved and highly flexible ‘strap’ region, present in flavin-dependent tryptophan halogenases, links the FAD and substrate-binding domain and is also known to be involved in FAD interactions and catalysis. The computational studies also revealed the potential energy and free energy profiles along with geometric features by which lysine and glutamate residue activates the hypohalous acid.

The highly conserved nucleoside binding GxGxxG (of flavin binding) and a structural WxWxIP (which blocks monooxygenase-specific functions) signature motifs were used to identify putative Fl-Hals in genome mining from NRPS and PKS natural product biosynthetic clusters. To date, many Fl-Hals from both bacterial and fungal species involved in complex biosynthetic pathways have been identified and characterized. All Fl-Hals share a structurally similar flavin-binding domain region whereas the main difference arises from the substrate-binding end terminal domains. The recruitment of different substrate-binding regions contributes to the substrate diversity within Fl-Hals enzymes. Some Fl-Hals also require a substrate activation or tethering via specific carrier proteins leading to additional structural differences. For example, the crystal structure of mycobacterial chondrochloren halogenase – CndH – has an unstructured C-terminal region with surface-exposed hydrophobic patches that are predicted to be interacting with an acyl carrier protein (ACP) that bound to the substrate (Figure 5). CndH also lack catalytically relevant Wheland glutamate base, which is thought to be supplemented by a surrogative residue from the carrier protein. A recently resolved structure of PltA halogenase – a carrier protein and flavin-dependent halogenase – also suggests large conformational changes within C-terminus regions for substrate positioning in Fl-Hal proteins.

Different classes of Fl-Hals based on the substrate chemical scaffold, such as indolic, phenolic and pyrrolic halogenases are currently isolated from natural pathways and heterologously characterized. Many Fl-Hals are known for halogenating complex biomolecular natural product structures either via an early stage halogenation of smaller starter units or by late stage tailoring activities of matured complex molecules. Though most of the Fl-Hal substrates belong to the category of electron-rich aromatic groups: regio and stereoselective aliphatic halogenation reactions are also reported. In vitro activity studies and direct evolution approaches showed that the active site of Fl-Hals in general are easily modifiable to adapt and display a wider substrate activity toward small and complex non-natural substrates.
3.1. Integration of Fl-Hals into non-native biosynthetic pathways

There are several examples from the natural product world, in which Fl-Hals are incorporated and utilized in combinatorial biosynthesis of natural product analogues. Combinatorial biosynthesis of two indolocarbazole alkaloids, rebeccamycin and staurosporine in *Streptomyces albus*, are the seminal works in this area. Rebeccamycin and staurosporine scaffolds 20–21 are formed from two tryptophan-derived units via decarboxylative fusion, followed by the attachment of a sugar moiety (Figure 6). Earlier studies had shown that the genomic cluster of rebeccamycin and staurosporine could be heterologously cloned into non-producers like *S. albus*. For the combinatorial biosynthesis of rebeccamycin in *S. albus*, a collection of rebeccamycin (*reb*) gene combinations were constructed using genes from different

---

**Figure 6.** *In vivo* combinatorial biosynthesis of indolocarbazole alkaloids, rebeccamycin and staurosporine molecular analogues in *S. albus*. The tryptophan Fl-Hals, RebH, Thal and PyrH used here to install regio selective halogen handles in these molecules.
rebeccamycin producers like *Lechevalieria aerocolonigenes*, *Streptomyces long-isporoflavus*, *Streptomyces albogriseolus* and *Streptomyces rugosporus*.\[^{122}\]

The *rebO* (amino acid oxidase) and *rebD* (chromopyrrolic acid synthase) genes are responsible for producing 3,4-bis(indol-3-yl)pyrrole-2,5-dicarboxylic acid or chromopyrrolic acid. Co-expression of two additional genes *rebC* (FAD-containing monooxygenase) and *rebP* (a P450 oxygenase) was required for producing indolopyrrolocarbazole core structure that is identical to arcyriaflavin A \[^{22}\]. Other combination involves *rebG* (N-glycosyltransferase) and *rebM* (sugar O-methyltransferase) which are responsible for glycosylation and methylation of arcyriaflavin A sugar moiety. The main structural difference of the indolopyrrolocarbazole core structure in rebeccamycin and staurosporine is the absence of a C-7 carbonyl function in staurosporine. This originates from the differential oxidative activity of *rebC* oxygenase compared to its counter-part *staC* in staurosporine cluster. In order to increase the size of combinatorial libraries via various gene combinations arising from all the aforementioned genes, the tryptophan 7-halogenase enzyme RebH present in this cluster was replaced with either a tryptophan 6-halogenase (Thal from thienodolin biosynthesis in *S. albogriseolus*) or tryptophan 5-halogenase (PyrH from pyrroindomycin biosynthesis in *S. rugosporus*) yielding different site specifically mono or dihalogenated analogues.

The combinatorial biosynthesis by using halogenases that act on carrier protein tethered substrates in native host systems are explored with the antibiotic biosynthetic clusters of hormaomycin \[^{23}\] and aminocoumarin antibiotic clorobiocin \[^{24}\].\[^{123}\] Hormaomycin produced in *Streptomyces griseoflavus* contains a 5-chloro 1-hydroxypyrrole moiety generated by the halogenation activity of *hrmQ*, a pyrrol Fl-Hal in the biosynthetic pathway.

![Figure 7](image_url)

*Figure 7.* In vivo combinatorial biosynthesis of novoclorobiocin analogues in *S. roseochromogenes* using Fl-Hal that act on a carrier protein tethered substrate.
On the other hand, clorobiocin produced by a *Streptomyces roseochromogenes* variant contains a 5-methyl pyrrol moiety. The heterologous expression of hrmQ gene in *Streptomyces roseochromogenes* after genetically inactivating pyrrole-5-methyl-transferase (CloN6) produced two halogenated clorobiocin derivatives 25–26 in quantifiable yields (Figure 7). Though the strategy worked

*Figure 8.* Natural products that are biosynthesized by a flavin-dependent halogenase enzyme.
for above-mentioned case, there are examples in which different substrate specificity of Fl-Hals hindered the halogenase replacement between different biosynthetic clusters. The effort to functionally replace Clo-hal which halogenates position 8 in aminocoumarin of clorobiocin with a BhaA Fl-Hal from balhimycin 27 cluster did not produce the expected product (Figure 8).\[^{124}\] Though both Fl-Hals could act on a carrier protein-bound β-hydroxytyrosine moiety, this indicates that the substrate specificity of Fl-Hals is an important factor that could contribute to additional levels of complexity for replacement strategies.

Incorporating free-standing Fl-Hals directly into existing biosynthetic pathways where no natural halogenation biosynthetic machinery is present is an attractive strategy for non-naturals. This was attempted by Goss group in pacidamycin biosynthetic cluster. In this approach, chlorinated pacidamycin were produced from *Streptomyces coeruleorubidus* by genetically introducing a tryptophan-7-halogenase (PrnA) from the pyrrolnitrin biosynthetic pathway.\[^{125}\] Incorporation of Fl-Hals into medicinal plant secondary metabolism for the *de novo* biosynthesis of halogenated natural products with improved bioactivities has many potential applications. Compared to biosynthetic genomes found as clusters in micro-organisms, plant biosynthetic gene clusters are more scattered, thereby making gene manipulations using halogenase an easy and viable approach for combinatorial biosynthesis. This was demonstrated in medicinal plant *Catharanthus roseus* by including PyrH (tryptophan 5-halogenase) and RebH (tryptophan 7-halogenase) to produce site specifically chlorinated tryptophans, which was then shuttled into monoterpene indole alkaloid

![Figure 9. Introducing Fl-Hals into plant secondary metabolism. a) Modifying indole alkaloid biosynthesis in *C. roseus* by Fl-Hal PyrH (tryptophan 5-halogenase). b) Metabolic engineering of akuammicine scaffold with a variant of Fl-Hal enzyme RebH (tryptophan 7-halogenase) that act on tryptamine. c) Pd-based Suzuki-Miyaura cross coupling chemistries for further derivatization of akuammicine scaffold.](image-url)
metabolism to yield chlorinated alkaloids (10- chloroaajmalicine 28, 15-chlorotabersonine 29, 12-chloro-19,20- dihydroakuammicine 30) from its root cultures (Figure 9a-b). The inefficiency of tryptophan decarboxylase to act on chlorinated tryptophan to give required chlorotryptamine was found to be the metabolic bottle neck of this reaction which was addressed later by incorporating a RebH-Y455W variant capable of selectively converting 7-chlorotryptophan to 7-chlorotryptamine. The work was further extended to tobacco chloroplasts by using SttH (tryptophan 6-halogenase) and RebH in combinations to produce dihalogenated tryptamines.

Like for like gene swap experiments were also reported with Fl-Hal genes in naphthomycins 32 and ramoplanin 33 biosynthetic clusters (Figure 8). Naphthalenic antibiotic naphthymin and ansamitocin 34 has structural similarity, with both biosynthetic clusters possessing a free-standing phenolic Fl-Hal. The Asm12; an Fl-Hal from ansamitocin cluster of Actinosynnema pretiosum has over 78% sequence identity with Nat1 of naphthymin biosynthetic cluster present in Streptomyces sp. CS. Introduction and complementation of Asm12 in a Nat1 mutant strain restored the production of naphthymin indicating their functional similarity within the biosynthetic clusters. The substrate-tethered halogenases that act on hydroxyphenylglycine (Hpg) residues of the lipopeptide antibiotics ramoplanin and enduracidin 35 biosynthetic clusters present in soil bacterium Streptomyces fungicidicus were also studied for halogenase gene swap experiments (Figure 8). End30 halogenase from enuracidin dichlorinates enduracidin Hpg13 whereas Ram20 halogenase only monohalogenates the Hpg17 residue of ramoplanin. Ram20 knockout strain when complemented with end30 resulted in monohalogenation of Hpg17. When ramoplanin mannosyltransferase was knocked out in the complemented strain, a dihalogenation was restored on the Hpg13 residue in ramoplanin, indicating that both enzymes have different active site structure. A similar replacement of enduracidin halogenase End30 with Ram20 in knockout strain, produced dideschloroenduracidins A. The site-specific halogenation of End30 and Ram20 also helped to synthesize new halogenated non-natural molecules in enduracidin wild-type producer when Ram20 is recombinantly expressed. These proto-type examples of like for like gene swap and joint recombinant expression approaches indicate its potential applications not only for increasing the library of non-natural molecules, but also to improve the production titers.

3.2. **Fl-Hals for non-native biosynthetic pathways in industrial hosts**

Free-standing phenolic Fl-Hals from fungal species are potential targets for many biocatalytic and biosynthetic pathway applications. The halogenase Rdc2 and RadH involved in the biosynthesis of resorcylic acid lactone (RCL) polyketides in fungus *Pochonia chlamydosporia* and *Chaetomium chiversii* respectively.
were well exploited toward this direction.\textsuperscript{113,135} Radicicol 36 production in \textit{P. chlamydospora} involves two iterative polyketide synthases (IPKs) Rdc5 and Rdc1. The complete reconstitution of the two IPKs in \textit{Saccharomyces cerevisiae} (both \textit{in vitro} and \textit{via in vivo} routes) by Yi Tang group had displayed the formation of a resorcylic acid lactone intermediate, (R)-monocillin II 37 (Figure 10a). Introducing Rdc2 Fl-Hal into the pathway produced site-specifically halogenated molecule pochonin D 38 (approx. 15 mg/L).\textsuperscript{135} Utilizing industrial host strains such as \textit{E. coli} and yeast for heterologous fungal biosynthetic genomes presents many major challenges, including the known inactivity of endogeneous 4′-phosphopantetheine (pPant) transferase to act up on fungal ACP domains that leave fungal iterative polyketide synthases as unmodified. This was addressed with a genetically modified \textit{S. cerevisiae} BJ5464-NpgA strain that carry pPant transferase gene from \textit{Aspergillus nidulans} which was previously used for fungal pathway reconstruction studies in yeast.\textsuperscript{136,137} Yi Tang group had also introduced Rdc2 in \textit{S. cerevisiae} along with other two IPKs (Hmp3 and Hmp8) that are present in the RCL hypothemycin 39 biosynthesis cluster of fungus \textit{Hypomyces subiculosus}.\textsuperscript{138} In the presence of Rdc2 halogenase, the reconstituted yeast strain produced 6-chloro 7,8-dehydrozearalenol 40, an RAL related to monocillin II, with a different stereochemistry for the lactone alcohol group (Figure 10b). The structural difference between monocillin and 7,8-dehydrozearalenol arises from the strict stereospecificity of Hmp8 β-ketoreductase domain to form D-hydroxyl product which is reduced to form L-alcohol, whereas in Rdc5 an opposite stereochemistry is followed. \textit{In vitro} studies with RadH Fl-Hal showed that although it halogenates monocillin II, it is

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure10.png}
\caption{Recreating fungal resorcylic acid lactone pathways in yeast \textit{S. cerevisiae}. a) Radicicol pathway recreated with Rdc2 halogenase and minimal iterative PKS pathway. enzymes b) Hypothemycin pathway for the production chlorinated dehydrozearalenol in yeast.}
\end{figure}
inactive toward zearalenone 41 or curvularin 42, suggesting a different active site structure despite having a high sequence similarity (Figure 10). However, RadH is more promiscuous than Rdc2 with a potential scope to replace Rdc2 in reconstructed monocillin pathway toward many radicicol analogues via halospecific chemical coupling reactions.[139]

The late-stage macrolactone scaffold halogenation and ability to halogenate different smaller phenolic compounds suggested that RadH and Rdc2 are ideal enzymes for recreating secondary metabolite biosynthetic pathways in E. coli. Rdc2 are used for recreating phenylpropanoid acid-based secondary metabolite biosynthetic pathways in E. coli.[140] Phenylpropanoid acids based pathways lead to variety of natural products in plants with some of the pathways previously reconstructed in E. coli. By the recombinant expression of TAL (Tryptophan Ammonia Lyase from Saccharothrix espanaensis), 4Cl (4-coumarate:CoA ligase from Arabidopsis thaliana) and STS (stilbene synthase from Arachis hypogaea) from different sources, bioactive natural stilbenoid molecule resveratrol 43 was combinatorially bio-synthesized in E. coli. Introducing Rdc2 into reconstituted resveratrol pathway affords the production of a new non-natural molecule, 2-chlorinated resveratrol 44, with improved antimicrobial and antioxidant activity (Figure 11a).[140,141] In vitro studies with RadH showed that umbelliferone 45, a central intermediate of coumarin-based natural molecules produced in plants, is one of the highly active non-native substrate for RadH enzyme.[113] This had led to the reconstruction of an umbelliferone pathway in E. coli with TAL and 4 Cl along with an additional enzyme feruloyl CoA 6′-hydroxylase (F6′

**Figure 11.** In vivo biosynthesis of halogenated molecules in E. coli. a) Biosynthesis of chlororesveratrol and B) chloro umbelliferone by utilizing phenylpropanoid acid based biosynthetic pathways in E. coli. c) Chlorinated anthraquinone derivatives in E. coli with type II PKSs proteins from P. luminescens.
The umbrellaferone formed was then halogenated to non-natural molecule chloro umbrellaferone 46 (approx. 1 mg/L) by incorporating RadH into this reconstructed pathway (Figure 11b). \[113\]

The wider substrate scope that Fl-Hals exhibit makes it possible to target complex polyketide biosynthetic pathways reconstructed in E. coli. The typical example is the anthraquinone pathway from Photorhabdus luminescens recombinantly expressed recently in E. coli.\[142\] Though type I modular polyketide synthases (PKSs) are expressed previously in E. coli, the incompatibility of type II PKSs with E. coli cell line makes it more elusive.\[143,144\] The soluble expression of a ketosynthase (KS) and chain length factor (CLF) pairs of P. luminescens from anthraquinone pathway had provided a minimal type II PKS genomic components to synthesize anthraquinone in E. coli (Figure 11c). The anthraquinone derivative AQ256 47 thus formed could be in vivo halogenated with RadH to its monochlorinated AQ256 analogue neocatamidomycin (or 1,3,8-trihydroxy-monochloroanthraquinone) 48 (approx. 1 mg/L). The chlorinated anthraquinone derivatives are well known for its ATP citrate lyase inhibiting activity and in modulating mammalian fatty acid synthesis. This implies that the biosynthetic approaches toward chloro anthraquinone molecules will have a wider application in the near future.\[145,146\]

For carrier protein substrate dependent Fl-Hals, heterologous expression of halogenase containing genomic clusters in industrial hosts could be challenging. The seminal example was reported with a tetrabromo pyrrole biosynthetic cluster (bmp) from marine bacteria Pseudoalteromonas and Marinomonas mediterranea MMB-1 genome. The bromopyrrol biosynthesis in Pseudoalteromonas starts from loading of L-proline 49 onto an ACP phosphopantetheine arm by an adenylation enzyme (bmp4), followed by oxidation of the prolyl ring to a pyrrole by a dehydrogenase enzyme (bmp3) (Figure 12a). The thio templated halopyrrole can further proceed down biosynthetic assembly lines toward complex marine natural products like marinopyrrole 50, chlorizidine 51 etc (Figure 12b). Bmp2 is the brominase enzyme that acts on a pyrroly-2-carboxylate substrate attached to the Bmp1 carrier protein via thioesterase link. Having found that Bmp1 to 4 enzymes are key in the biosynthetic pathway to produce 2,3,4-tribromo pyrrole 52 and 3,4,5-tribromo-pyrrole-2-carboxylic acid 53; bmp1 to bmp4 protein cluster was heterologously expressed in E. coli. The expected brominated molecules (2,3,4-tribromo pyrrole and 3,4,5-tribromo-pyrrole-2-carboxylic acid) were produced in the culture when bromide was present. This was also confirmed by in vitro reconstitution of minimal pathway using bmp1 to bmp4 purified enzymes. The pyrrole 2-carboxylate moiety is suggested to be offloaded from the ACP by a non-enzymatic hydrolysis and a further decarboxylation of terminal bromide yields bromopyrrole. As there are several natural product molecules derived from L-proline starting units that follow a similar initial synthetic machinery, this example could provide further insights into future design of divergent and interesting molecular scaffolds in industrial hosts.
3.3. **Fl-Hals for biosynthetic pathways with integrated chemo-catalysts**

The possibility to replace carbon-halogen bond with carbon-carbon, carbon-nitrogen, carbon-fluorine or carbon-oxygen bonds via transition metal-based complex chemo-catalysts are known for decades. Carbon-halogen bonds present in both natural and non-naturals created by biosynthetic and biotransformation protocols provide an opportunity to derivatize the molecular scaffolds using heavy metal-based cross coupling chemistries. The effective catalyst control for C-H activation with traditional chemical halogenation methods is often impossible and imposes many challenges. The existence of specific regioselective Fl-Hals that act on electron-rich aromatic substrates makes it logical to have enzymatic routes to create halogenated moieties for these approaches. Such methods are slowly gaining wider interest from synthetic chemists and chemical biologists, and as a result the past few years have seen several reports on the development of compatible chemo-enzymatic reaction platforms using Fl-Hals.

The first known example of an *in vivo* chemo-enzymatic approach from the natural product world was by Goss group using pacidamycin 54 biosynthetic cluster in *S. coerleorubidis*. The chloro pacidamycin 55 produced by the
introduction of Fl-Hal PrnA was selectively functionalized using Pd-mediated Suzuki-Miyaura cross-coupling reactions (Figure 13a). The synthetic derivatization of a complex molecule like pacidamycin is quite challenging with the presence of reactive functional groups, its insolubility in organic solvents and poor thermal stability. This was addressed by seeking and imposing milder cross coupling reaction conditions and via screening for efficient chemo-catalysts for biocompatibility and bio-orthogonality. Few pacidamycin analogues are thus created via cross-coupling using aromatic boronic acid substrates, Na₂Cl₃Pd-SPhos catalyst and K₂CO₃ as base in water-acetonitrile solvent system. The chemo coupling reactions were performed with aqueous extracts of the fermentation broth, or with semi-purified cultures by heating at 80°C in a microwave. Another notable early example was for synthesizing monoterpene indole alkaloid analogues from plant C. roseus. 12-chloro 19,20-dihydroakuammicine and 12-bromo 19,20-dihydroakuammicine 30 are produced in C. roseus root cultures by introducing Fl-Hal RebH and its flavin reductase partner RebF via genetic engineering. The aryl and heteroaryl analogues of 19,20-
dihydroakuammicine 31 was synthesized by a subsequent Pd-catalyzed Suzuki-Miyaura cross-coupling reaction (Figure 9c). In these approaches, lower activity of halogenases seems to be the limiting factor for large scale synthesis and scale-ups, which was often addressed by traditional mutasynthesis or precursor-directed biosynthesis (not discussed under the scope of this review).\textsuperscript{148,149}

The development of a concurrent method in which bio-engineered halogen moieties are reactively coupled with Pd based chemo catalysis in a living system has many potential applications in \textit{in situ} labeling, affinity tagging of proteins and biomolecules. This possibility was explored by Goss group with a living ‘GenoChemetics’ approach (Figure 13b).\textsuperscript{150} The living cell culture of an engineered \textit{E. coli} PHL644 strain that carry PrnA Fl-Hal to produce 7-brominated tryptophan was used for the synchronized production and cross-coupling, after initial optimization of growth media and coupling conditions. Next, these approaches were extended to pacidamycin biosynthesis, using a bromide salt tolerant \textit{S. coelicolor} M1154 strain and after genetically introducing \textit{S. coeruleorubidus} pacidamycin cluster along with \textit{prnA}. The cell culture grown in a defined media was capable of producing sufficient-brominated pacidamycin D 56 which was derivatized to desired cross coupling product \textit{p}-tolyl-pacidamycin 57 using a corresponding aromatic boronic acid substrate (Figure 13b). The \textit{in vivo} generated halo-metabolites could be potentially applied to other types of Pd or heavy metal-based cross coupling chemistries, especially for Heck coupling reactions. The earliest attempt in this direction was with the bromo pacidamycin D 56 engineered \textit{S. coelicolor} M1154 strain. The crude culture extract containing brominated pacidamycin D was cross coupled with fluoro styrene to form fluoro-styrene pacidamycin D 58 in the presence of a tris(2,4-dimethyl-5-sulfophenyl)phosphine trisodium (TXPTS) based palladium catalyst.\textsuperscript{151} Though the brominated molecule was fully converted, scale up and separation of the product from cell culture wasn’t attained. These synchronized, continuous one-pot halo-metabolite production and a followed cross coupling approaches have many advantages, including improving the metabolite flux toward final products, absence of intermediate purification steps etc. Fluorogenic cross-coupling of halogenated natural products and bio-synthesized small molecules in living or culture supernatants could also be used as a potential tool for molecular identification and screening. This approach was already applied by other groups for high-throughput-directed evolution of halogenase enzymes to increase its reactivity and thermal stability.\textsuperscript{152,153}

4. Nonheme Fe (II)-\textit{α}-ketoglutarate-dependent halogenases (\textit{α}-KG-Hals)

\textit{α}-KG-Hals belong to the superfamily of \textit{α}-ketoglutarate-dependent oxygenases (\textit{α}-KGOs), a diverse family of enzymes that catalyzes hydroxylation, epoxidation, epimerization, demethylation, ring formation, C-C bond
cleavage and desaturation reactions.\textsuperscript{154,155} α-KG-Hals are structurally and mechanistically related to α-KG-oxygenases, which follow a radical mechanism via formation of a high valent and short-lived Fe\textsuperscript{IV} = O ferryl intermediate (Figure 1).\textsuperscript{156,157} This intermediate is a powerful oxidant that abstracts an hydrogen from un-activated carbon-hydrogen bond creating a substrate radical and another Fe\textsuperscript{III}-hydroxyl (or Chloride) intermediate. The chloride then rebounds to substrate radical giving halogenated product in α-KG-Hals, whereas in hydroxylases, it is the hydroxyl species that rebounds.\textsuperscript{157}

The main structural difference in the active site of hydroxylases and α-KG-Hal is within the HxG(A) motif, in which a glycine or an alanine residue is replaced by a carboxylate residue allowing direct coordination of a halide ion in α-KG-Hals.\textsuperscript{44} The placement of substrate relative to active site Fe\textsuperscript{IV} = O ferryl intermediate has a major role in the halogenation mechanism including various other factors such as oxidative and dissociative properties of Fe center.\textsuperscript{158,159,160,161} The electronic structural properties and oxyl character of the reactive ferryl intermediate during the transition state formation in the α-KG-dependent enzymes were analyzed and reported by several computational and spectroscopic studies.\textsuperscript{162,163,164} These studies also revealed the fine details in which the highly oxidative ferryl intermediate could also lead to other oxidative reaction environments within the active site. For example, in α-KG-Hals side reactions such as aliphatic hydroxylation, nitration and azidation reactions were also reported.\textsuperscript{43,165}

An earlier observation on the aliphatic carbon centers of nonribosomal peptides barbamide 59 and syringomycin 60, which are halogenated without the presence of any known halogenases, led to the identification of α-KGOs dependent enzymatic halogenase machinery from these clusters. α-KG-Hals known as SyrB2 (from the syringomycin genome cluster) BarB1 and BarB2 (from barbamide biosynthesis cluster) were subsequently identified (Figure 14).\textsuperscript{166,167} The first α-KG-Hal enzyme characterized from a biosynthetic cluster was SyrB2 in 2005.\textsuperscript{168} Similar to most of the initially known α-KG-Hals, these enzymes utilize substrates that are covalently tethered to acyl or peptidyl carrier proteins with a phosphopantetheine arm. The carrier protein tethered L-threonine (L-Thr-S-SyrB1) 61 acts as a substrate for the SyrB2 reaction, whereas for BarB1 and BarB2-mediated chlorination reactions the substrate is L-leucine (L-Leu-S-BarA) 62. Several carrier protein substrate-dependent α-KG-Hals are discovered over the years, including CytC3 (from γ,γ-dichloroaminobutyrate antibiotic producer), KtzD and KthP (from Kutzneride 63 biosynthetic cluster), HctB (from hectochlorin 64), CmaB (from coronamic acid 65 biosynthetic cluster) etc (Figures 14 & 15).\textsuperscript{45,169,170,171,172}

The first free-standing α-KG-Hal known as WelO5 was discovered from the cyanobacterium Hapalosiphon welwitschii.\textsuperscript{42} WelO5, present within the welwitindoline biosynthetic cluster, could regioselectively monochlorinate
aliphatic carbon of 12-epi-fischerindole U 66 and 12-epi-hapalindole C 68 to afford 12-epi-Fischerindole G 67 and 12-epi-Hapalindol E 69 respectively, creating a new stereocenter in the molecules (Figures 14–15). Genome searching of similar late stage free-standing α-KG-Hals led to the discovery of AmbO5 from an ambiguine indole alkaloid producing cyanobacterium
Fischerella ambigua UTEX 1930.\(^{[173]}\) Compared to limited substrate scope of WelO5, the identified AmbO5 enzyme exhibited a wider substrate scope by selectively modifying structurally distinct ambiguine 70, fischerindole 66 and hapalindole 68 alkaloids.\(^{[174]}\)

The resolved crystal structures of carrier protein tethered and free standing \(\alpha\)-KG-Hals revealed the existence of conformational dynamics that are coupled to the halogenation mechanism present in both class of \(\alpha\)-KG-Hal enzymes. For example, the crystal structure of CurA halogenase was resolved in different ligand states that corresponds to open and closed conformations of
the enzyme. The substrate recognition and binding of 3-S-hydroxy-3-methylglutaryl-ACP in the presence of αKG and chloride leads to chlorination in the closed form of CurA halogenase. It is also proposed from the crystal structures that for the carrier protein-dependent α-KG-Hals, substrates are positioned close to the ligand, whereas for free standing α-KG-Hals it is either on the same side or opposite to the putative oxygen-binding location (Figure 16). The resolved crystal structure of WelO5 also provided more insights into a dynamic C-terminal sequence motif – an α-helix region which plays an important role in substrate tolerance and specificity of these class of freestanding α-KG-Hals. Henceforth, the C-terminal α-helical region in WelO5 was substituted and modified to be structurally similar to AmbO5, creating a WelO5- AmbO5 chimera protein, which regained broader substrate scope as of wild-type AmbO5.

Recently Chang group published another class of freestanding α-KG-Hal called BesD that acts on amino acids. BesD is involved in the β-ethylserine biosynthesis of Streptomyces cattleya and chlorinates β-carbon of L-lysine. There are several additional α-KG-Hals with sequence similarity to BesD identified and clustered into HalA to HalH groups. The reaction profile of individual HalA-H enzymes catalogue them to different groups of regioselective halogenating catalysts for of L-lysine, L-ornithine, L-leucine, L-isoleucine, L-norleucine etc (Figures 14–16). The BesD class of α-KG-Hal has a low sequence similarity with WelO5 and AmbO5 and are more related to α-KGOs indicating that both these classes are evolved independently from oxygenases in two lineages. Recently, two additional-free standing α-KG-Hals were characterized from the dechlorin 75 biosynthetic cluster from Actinomadura sp. ATCC 39365 and from acutumine 77 biosynthetic cluster in Menispermaceae plants. AdeV halogenase from dechlorin cluster halogenates C2’ position of 2’-deoxyadenosine monophosphate and it is the first example of a α-KG-Hal that acts on nucleotides. Dechboroacutumine halogenase (DAH) catalyzes the terminal chlorination step in the biosynthesis of acutumine by chlorinating

![Figure 16. Comparison of active site structures of carrier protein dependent and free standing α-KG-Hals. Active site of SyrB2, WelO5 and BesD are shown in figure A, B and C respectively. The figures are based on the deposited PDB structures of SyrB2 (PDB ID: 2FCT), WelO5(PDB ID: 5IQT) and BesD (PDB ID: 6NIE).](image-url)
dechloroacutumine 79 to acutumine 77.[176] Both the enzymes have a low sequence similarity with other α-KG-Hals which has potential utility in expanding chemo diversity of nucleotides and acutumine type scaffolds. There are some excellent reviews on the enzymology and halogenation mechanism of α-KG-Hal and α-KGO superfamily of enzymes for further reading.[154,155,177]

### 4.1. α-KG-Hals for generating non-naturals via in vitro and in vivo routes

The earlier attempts for in vitro characterization and biosynthetic engineering of α-KG-Hal was often hampered by its low activity and the pre-requisite of substrate tethering with specific carrier proteins. This was partly addressed by utilizing soluble compatible tethering units from adjacent biosynthetic clusters present within the same organism. For KthP halogenase, which is involved in kutzneride 63 biosynthesis in Kutzneria sp. 74, a stand-alone thiolation domain protein, KtzC, from a nearby cluster was used to tether the piperazate substrate 80 and to halogenate it (Figure 14b).[169] Previously, KtzC was known to act as a substrate for α-KG-Hal KtzD when tethered with an amino acid (Ile-S-pantetheinyl) prosthetic group in the same Kutzneria species.[178] Though KthP has a low sequence similarity to KtzD, the biosynthetic cluster units are interchangeable in relation to α-KG-Hals.

One of the available examples in which an α-KG-Hal including pathway was partly reconstructed in vitro is from curacin 81 and jamaicamide 82 biosynthetic clusters (Figure 17).[154,179] α-KG-Hals CurA from curacin and JamE from jamaicamide biosynthetic pathways in the cyanobacterium Lyngbya

![Diagram](image-url)
has over 92% sequence identity. CurA is responsible for the production of β-branched cyclopropane, whereas JamE forms a vinyl chloride group, a distinct chemical motif. It was shown that in these two biosynthetic assemblies, structural diversification occurs after the halogenation step as both halogenease utilize (S)-3-hydroxy-3-methylglutaryl-S-ACP (HMG-S-ACP) 83 as native substrate. In curacin pathway, the halogenation is followed by a dehydratase (by ECH1 domain of CurE) and decarboxylase (by ECH2 domain of CurF) activities on the 4-chloro-3-methylcrotonyl-ACP 84 halogenated molecule. The final cyclopropane ring is formed by the enoyl reductase (ER) activity of CurF ER domain on 4-chloro-3-methylcrotonyl-ACP. Thus, the cryptic chlorination of CurA led to the cyclopropane building block of curacin A. In the jamaicamide pathway the ECH2 domain of JamJ produces a β,γ-enoyl thioester intermediate in contrary to α,β-enoyl thioester produced in curacin. The structural modification of HMG-S-ACP in both cases was proven by setting up one-pot in vitro assays using CurA- ECH1-ECH2, ECH1-ECH2-ER and CurA- ECH1-ECH2-ER combinations, where ECH1, ECH2, and ER were either from curacin or jamaicamide biosynthetic clusters (Figure 17). The proteins for the assay were purified via heterologous expression in E. coli cells. Though in vivo production of modified curacin or jamaicamide molecules were not carried out in E. coli expression system, the one-pot assays indicate that multienzyme pathways with α-KG-Hal could be reconstituted in heterologous hosts and non-producers for molecular diversification.

Many natural product biosynthetic gene clusters carry cryptic α-KG-Hals like CurA, that generates nonproteinogenic amino acids. The cyclopropyl amino acid ring structure of coronamic acid is formed by CmaABCDE enzymes in Pseudomonas syringae (Figure 14b). L-allo-isoleucine, which is covalently attached to CmaD acts as a substrate for halogenation with CmaB. This is followed by the reaction of CmaC which catalyzes the formation of the cyclopropyl ring from γ-Cl-L-allo-isoleucine. Similar types of nonribosomal peptides clusters that contain allo coronamic acid cassettes were also found in biosynthetic pathways of other organisms. In kutzneride biosynthesis, the gene cluster contains α-KG-Hal KtzD and a flavin-dependent acyl-CoA dehydrogenase-like protein KtzA. The other proteins in the ktzABCD gene cassette are an adenylylating protein KtzB and a carrier protein KtzC. The in vitro reconstitution of ktzABCD with the substrate L-allo-Isoleucine tethered to KtzC showed that the γ-chloroisoleucyl intermediate formed by KtzD is cyclized by KtzA enzyme, forming the cyclopropane ring (Figure 14b). The final product was identified as (1S, 2R)-allocoronamic acid bound in thioester linkage to KtzC, which is incorporated to kutzneride by downstream enzymes. The stericly rigid cyclopropane amino acids are important synthetic intermediates that have been previously incorporated into therapeutic peptides, synthetic hormones, and enzyme
inhibitors to avoid enzymatic degradation and to improve the biophysical properties.[182] Thus, the reconstruction of allo coronamic acid-like gene cassettes in heterologous host systems have a commercially valuable proposition in providing novel routes to highly demanding non-natural cyclo propane amino acid precursors and its derivatives.

The structural similarity between α-KG-Hals and α-KG-hydroxylases at the Fe-binding active site had provided more deducible information for functional switching within these two classes of enzymes using site-directed mutagenesis approaches. Initially, standalone hydroxylases were targeted to perform hydroxylase activity via mutagenesis, with an obvious advantage to avoid the need of a tethered substrates. It was assumed that replacing Asp or Glu of HxD(E) with Gly or Ala will mimic halogenase HxG motif. When this was applied to taurine dioxygenase and prolyl-4-hydroxylase, the wildtype hydroxylase activity was lost with no observable halogenation.[183,184] However, hydroxylase activity was switched to halogenase for N-succinyl-L-leucine hydroxylase enzyme SadA from Burkholderia ambifaria when HxG mutation was installed.[185,186] Similar approaches were also attempted to switch halogenase activity to hydroxylase activity in substrate tethered α-KG-Hal-SyrB2 and for free standing α-KG-Hal-WelO5.[42,44] HxA to HxD (and to HxE) mutants in SyrB2 inactivated the enzyme, whereas HxG to HxD mutant in WelO5 displayed hydroxylase activity with annulated halogenation. It was later identified that for hydroxylation, substrate positioning close to Fe$^{IV} = O$ ferryl intermediate is necessary. L-norvaline which has an extended structure due to an additional methylene group than the native substrate L-threonine 85 was hydroxylated by SyrB2 due to its close proximity to oxo group.[167,185]

Remodelling α-KG-hydroxylases to α-KG-Hals and vice versa with single-site substitutions in HxD motif at the native genome is one of the easiest approaches to introduce non-natural halogen or hydroxylase moiety to many natural product scaffolds. Though there are no proven examples currently available from natural product world in this direction, such a possibility still exists in several biosynthetic systems.

5. SAM-dependent halogenases

SAM-dependent halogenases were discovered from the initial efforts to identify origin and biosynthetic route to the formation of fluoroacetate toxins in many plant and Streptomyces bacterial species.[187,188] 5′-fluoro-5′-deoxyfluoroadenosine (5′-FDA) 86, a molecule derived from S-adenosylmethionine (SAM) 87 cofactor, was found to be the fluorinated precursor in Streptomyces cattleya, which is further converted to fluoroacetate 88 and 4-fluorothreonine 89 by downstream activity of an aldehyde dehydrogenase and transaldolase, respectively (Figure 18a-b).[189,190,191,192] Compared to other halogenases, the SAM-dependent fluorinase (FIA or FIA1) from
Figure 18. Fluorinated natural product pathways and other routes toward fluoro metabolites. 

a) Engineered halosalinosporamide biosynthetic pathway in S. tropica. b) The 4-fluorotheonine forming in vitro pathway reconstructed using fluorinase (FIA), purine nucleoside phosphorylase (FIB), fuculose aldolase and 4-fluoro theonine transaldolase enzymes. c) In vitro reconstituted pathway for 5-fluoro-2,3,4-trihydroxypentanoic acid (5-FHPA). d) Reported mutasynthetic approaches using fluoromalonyl-CoA and fluoroacetyl-CoA molecules.
S. cattleya has a strong substrate preference for fluoride ion over chloride. However, the equilibrium of this reversible halogenation is positioned toward the reactant side for chlorination reactions. Additional enzymes, such as amino acid oxidase which acts on L-methionine product is required to shift the equilibrium to the product side (5’-chloro-5’-deoxyadenosine (5’-CIDA) and L-methionine) for chlorination reactions.

Later on, a homologous chlorinase enzyme (SalL) was identified from salinosporamide 90 biosynthetic pathway in Salinispora tropica which had shown chlorinase activity and low preference for fluoride as a nucleophile (Figure 18a). A recent computational study indicated that the high preference for chloride ion exhibited by SalL is also linked to the increased potential energy requirements to cross the transition states that can lead to catalytic inactivity in the presence of fluoride. The natural product, Salinosporamide is produced by a hybrid polyketide synthase–nonribosomal peptide synthetase (PKS-NRPS) gene cluster in S. tropica. The chlorination units in salinosporamide are formed from a chloroethylmalonyl-CoA 91 PKS building block. 5’-CIDA formed via SalL halogenation acts as a precursor for downstream enzymes to produce chloroethylmalonyl-CoA. A few homologous of FIA and SalL halogenases are known to date, which are either characterized or annotated in the genome depository.

The SAM-dependent halogenases have a higher sequence homology to a superfamily of proteins from the duf-62 gene, called SAM hydroxide adenosyltransferase, that mediates hydrolytic cleavage of SAM to generate adenosine and L-methionine. The crystal structure of SAM-dependent halogenases also showed that these enzymes are structurally homologous to duf-62 proteins, essentially formed of an amino acid N-terminal and a carboxy terminal domain region joined by a long-extended loop region. The active site of SAM-dependent halogenase and duf-62 enzymes are usually formed and confined at the interface of adjacent monomers. The SAM is also positioned at an interface of monomers, between C and N terminal domains originated from two adjacent and interacting monomers (Figure 19). The crystal structures also predicted that these protein-SAM contacts would drive the domain dynamics in SAM-dependent halogenases which are coupled to halogenation by forming a closed state, inferring the unique quaternary structure that play a crucial role in substrate binding and halogenation reaction chemistry. The main structural differences between SAM-dependent chorinase (SalL) and fluorinase (FlA) is the absence of a 23-residue loop at the N-terminal domain of chorinases (Figure 19). This loop region is thought to decrease the surface area and interactions around the active site thus reducing the fluorine affinity in SalL.

Duf-62 proteins utilize a conserved amino acid triad (Asp-Arg-His) which has a role in activating water to hydroxide ions. SAM-dependent halogenases operate with a very similar mechanism to the hydroxylases. However, in SAM-dependent halogenases, hydroxide replaces the halide ion via SN2 substitution reactions and
also the conserved catalytic triad is absent. Due to strong solvation energy of halide ions (especially for fluoride) in water, the SAM-dependent halogenases should overcome a higher kinetic energy barrier to convert halides into a potent nucleophile within the active site (Figure 1). As a consequence, SAM halogenases are sluggish enzymes in nature. Along with lower turnover rates, the substrate scope of SAM-dependent halogenases are also limited to substrates SAM analogues, mainly 2-deoxy analogues. This is often considered as one of the main reasons for not having many biosynthetic pathway engineering works reported with SAM-dependent halogenases. The fluorination chemistry in natural products, engineering fluorinases and applications of fluorinase enzymes are discussed elaborately in various recommended reviews.

5.1. Engineering natural product pathways with SAM-dependent halogenases

In 2010, Moore group had reported the first example where a SAM-dependent halogenase gene could be genetically engineered into another biosynthetic genomic cluster in a different organism for a ‘like for like’ gene swap. This was by incorporating he fla fluorinase gene into salinosporamide biosynthetic cluster of Salinispora tropica CNB-44023 by replacing the native sall chlorinase. Low levels of fluorinated salinosporamide were formed when the fluoride ion donor was added at later stages of the small-scale fermentation protocol. The extreme sensitivity of S. tropica toward fluoride donors, and the
lack of fluoride toxic mitigation mechanisms were found to hinder the fluoro metabolite formation in \textit{S. tropica}, making large-scale production technically impossible. The fluoro tolerant \textit{S. cattleya} has an evolved mechanism to dissipate toxic fluoroacetate and 4-fluorothreonine which is lacking in \textit{S. tropica}. The fluoroacetyl-CoA hydrolase (FlK) and a trans acting fluoro-threonyl-tRNA deacylase (FthB) is part of this enzymatic machinery in \textit{S. cattleya}.\textsuperscript{[209,210,211]} FlK enzyme hydrolyze fluoroacetyl-CoA, whereas FthB removes the misacylation and incorporation of fluorothreonyl-tRNA instead of L-threonine into proteins during the synthesis.

This also means that in order to install new fluorination strategies and to maximize the targeted molecular production via SAM-dependent fluorinases; additional chemistries should be provided via complex engineering strategies to biosynthetic gene clusters in other organisms. As fluorinated salinosporamide was formed via fluoroethylmalonyl-CoA building blocks, adapting these complex engineering strategies could also be beneficial for producing many fluorinated non-natural molecules via other PKS systems where the extender unit is an ethylmalonyl-CoA.\textsuperscript{[212]} The potential applications of this was already demonstrated by the mutasynthesis of tetraketides in \textit{E. coli} cells via feeding synthetic fluoromalonyl-CoA 92 molecules (Figure 18d).\textsuperscript{[213]} An earlier attempt to pathway intervention via fluoroacetyl-CoA 93 was also reported in the natural product world.\textsuperscript{[214]} \textit{In vitro} incubation of fluoroacetyl-CoA with reconstituted minimum components of the artinochadin PKS machinery has shown that fluorinated octaketide intermediate could be formed. Incubation of fluoroacetyl-CoA with phosphopantetheinyl transferase could convert holo-acyl carrier protein (ACP) to fluoroacetyl-ACP to act as a substrate for downstream enzyme in biosynthetic pathways. Here, this has led to the selective formation of final aromatic polyketide (Fluoro-SEK4b 94) using \textit{in vitro} assays. In order to form fluoro-SEK4b molecule, an \textit{in vitro} assay was performed using a purified keto synthase complexed to a chain length factor (KS–CLF) protein from artinochadin PKS cluster (Figure 18d).

\textbf{5.2. SAM-dependent halogenases for in vivo and in vitro biosynthetic pathways}

Engineering fluoro-metabolites production in industrial microorganisms via incorporating SAM-dependent fluorinase is often considered as a challenging task. Successful outcomes in this direction was reported only recently; via a rational multistage genetic engineering approach in \textit{E. coli} cells.\textsuperscript{[215]} For this, the \textit{crcB} gene that encodes for a fluoride specific ion channel was deleted initially to increase the intracellular fluoride concentrations in \textit{E. coli}. A second gene deletion was carried out to remove the \textit{deoD} gene, which encodes a purine nucleoside phosphorylase (PNP) to avoid 5′-FDA degradation. The 5′-FDA degradation was observed in \textit{S. cattleya} via formation of
a 5-fluoro-5-deoxy-d-ribose 1-phosphate (5-FDRP) intermediate. An additional modification was to include a transmembrane SAM transporter protein from *Rickettsia prowazekii*, to increase the pool of intracellular SAM concentration. The engineered *E. coli* cells were capable of catalyzing the *in vivo* reaction of FlA halogenase to produce 5ʹ-FDA from SAM.

Combining fluorinase with nucleoside-replacing enzymes like PNP, pyrimidine nucleoside phosphorylase (PyNP) and thymidine phosphorylase (TP) was used previously as an attractive strategy to synthesize fluorinated nucleosides and radio (18F) labelled nucleoside analogues.[216,217] The highly reversible reaction of 5ʹ-FDA with PNP that affords 5ʹ-FDRP formation could be used to base swap adenine with another purine or pyrimidine base. PyNP and TP display different nucleoside specificity that adds wider synthetic scope for base-swap methods. This was used by O’Hagan group to develop one-pot fluorinase and phosphorylase combined reactions to generate different pyrimidine, uridine and thymidine nucleosides (Figure 20).[218] As fluorinated nucleosides are important compounds in antiviral and anticancer treatment, these one-pot *in vitro* approaches have wider synthetic applications. Different 18F had labelled nucleoside analogues and 18F labelled nucleoside-peptide conjugates are affordable via these reactions for positron emission tomography (PET) in medical imaging and scanning.[219,220] Another interesting example of one-pot base swap methods was the synthesis of analogues of ribavirin, a broad-spectrum antiviral compound (Figure 20). A dehydroxyfluoro

![Figure 20](image-url)
analogue of ribavirin 95 was synthesized in a one-pot enzymatic cascade using 1,2,4-triazole 3-caboxamide 96 base by O’Hagan group.\textsuperscript{218}

The success of these approaches could offer possibilities to generate many downstream metabolites, such as fluororibose, fluoroacetate, 4-fluorothreonine etc from 5′-FDA in \textit{E. coli}, by reconstructing 5′-FDA degradation pathway enzymes.\textsuperscript{221,222} It was shown previously that 4-fluorothreonine could be formed from SAM through an \textit{in vitro} pathway reconstitution of four overexpressed enzymes, a fluorinase (FIA), purine nucleoside phosphorylase (FLB), an isomerase fuculose aldolase and a PLP-dependent 4-fluoro threonine transaldolase.\textsuperscript{215,222} Fluoroacetaldehyde is converted into 4-fluorothreonine by PLP-dependent transaldolase in the final step, which could be replaced by any acetaldehyde dehydrogenase to produce fluoroacetate. A chemo-enzymatic route to fluoroacetate is also reported via biohalogenation followed by an oxidative degradation method.\textsuperscript{223} Identification of another fluoro metabolite, (2R3S4S)-5-fluoro-2,3,4-trihydroxypentanoic acid (5- FHPA) 97, from \textit{Streptomyces} sp. MA37 indicated that several other routes of 5′-FDA degradation also exist in nature (Figure 18c).\textsuperscript{224} In \textit{vitro} assay of over-expressed short-chain dehydrogenase protein FdrC demonstrated that the NAD\textsuperscript{+} dependent enzyme can oxidize 5-fluoro-D-ribose (5-FDR) 98 to its corresponding lactone followed by hydrolysis to generate 5-FHPA (Figure 18c).\textsuperscript{225} Though this pathway was not engineered inside microbial cells, it could lead to many synthetic biology approaches toward branched halometabolite pathway generation in the coming years.

The structural similarity of fluoroacetate to the common acetate metabolite would even provide further opportunities to create other metabolic pathways that utilize fluoroacetate as the synthetic starting unit. In fact, the presence of ω-fluoro-fatty acids from seeds of the shrub \textit{Dichapetalum toxicarium} formed from the fluoroacetate starting units has been known for many decades.\textsuperscript{207,226} For these approaches, an initial activation of fluoroacetate to fluoroacetyl-CoA is required for the fluoro acetate to be taken as a substrate by fatty acid synthase. The key information by which plants effectively avoid fluoroacetyl-CoA conversion back to fluoroacetate via hydrolysis is currently lacking. It is highly logical to assume that these plants could be executing an evolved metabolism to selectively incorporate fluoroacetyl-CoA at higher concentrations to fatty acid biosynthesis.

\textbf{6. Future perspective}

Nature has its own intriguing methods to synthesize highly bioactive-halogenated molecules and natural products.\textsuperscript{5} This is well reflected within the literatures that covers the earliest halogenase identification and characterization to the recent ones.\textsuperscript{227} Though once enzymatic halogenation was considered to be a byproduct of uncontrolled hypohalous acid generation within the active site; halogenases that carry fully evolved and well-orchestrated halogenation machinery to afford more complex regio-
stereoselective transformations are currently known. The cryptic halogenases in many biosynthetic clusters in fact display diverse strategies that exist in nature for C-H activation, ring formation and functionalization of natural molecules. In many cases, this inspires synthetic chemists to invent more viable, alternative and even parallel chemical reactions.

The α-KG Hals, Fl-Hals and vHPOs exhibit exquisite reaction control with a defined active site and could be tuned for halogenating different structural scaffolds. This is one of the main reasons to consider the halogenase enzymes as most promising candidates for biosynthetic pathway generation. SAM-dependent halogenases, though with a limited substrate scope, presents the only class of enzymes capable for introducing fluorine to organic molecules. The biosynthetic generation of halogenated molecules via halogenases and their further exploitation via chemo or bio routes have a great potential impact to chemical synthesis and to all sectors of chemical industry. Here, we have only covered most of the recent and known examples in which halogenases are utilized explicitly for in vivo and in vitro pathway generations. Several practical and experimental challenges still remain unaddressed in this field, even though we could claim that an enormous stride have already been made over the years.

Halogenase enzymes are non-essential for the host survival or growth, hence they have undergone less evolutionary pressure to catalyze reactions more efficiently and with fast reaction kinetics. The known catalytic turn over number for most active flavin-dependent and SAM-dependent halogenases are only within the range of 0.5–4.0/min and 0.05–0.15/min, respectively. For a general comparison of the catalytic turnover with highly used industrial enzymes such as tryptophan synthases (70/sec) and phenylalanine ammonia lyase (115/sec), this is around 1500 times slower. As a consequence, the low turnovers results in biosynthetic bottlenecks when generating in vivo pathways using halogenase enzymes, which are otherwise difficult to resolve without improving the halogenase conversion rates. Along with this, in most cases halogenation will only introduce a minor structural change with the halogenated product competing with the substrate, inhibiting the enzymatic reaction.

Several attempts have been made to improve stability and activity of halogenases through optimization, protein engineering, directed evolution, chimeric protein design etc. The information about structure–function and structure–reactivity relationship of halogenases and pathway enzymes are crucial and this is currently lacking for many novel halogenation reactions. Most of the reported approaches were by targeting the substrate-binding domains with an aim to change conformational and binding flexibility to increase the catalytic turn over and to explore the substrate specificity. The oxidative nature of the halogenation reaction mechanism often indicated that the active cofactor (FAD, Fe, heme or vanadium) itself is labile to oxidative damages that can lead to decay of the active intermediate. For example, the
formation of a decayed FAD(C4a)−OH intermediate from the active FAD (C4a)−OOH species were reported for Fl-Hals.\[109\] This indicates that there are unknown and currently untargeted conformational events within cofactor binding regions that are important to catalysis. New targeted engineering approaches are necessary to alleviate the low turn overs by comparing cofactor chemistry and reaction mechanism of evolutionary-related enzymes. Therefore, the other reported kinetic and mechanistic studies on flavin-dependent monoxygenases, α-KG-dependent hydroxylases, heme hydroxylases and oxidases, SAM-dependent hydrolases etc could shed more light in this direction for each class of halogenases.

The key developments in this area should also need to be translated to creating complex molecular scaffolds via pathway generation. Surprisingly, few halogenated natural product pathways, like chlorosulfolipids isolated from algae, are still not explored.\[238,239\] Recently, a putative diiron-carboxylate type halogenase from cylindrocyclophane biosynthesis in Cylindrospermum lichenesiforme was identified, which undergo a unique C-C bond formation.\[228\] Though not fully characterized, these currently unknown chemistries have potential applications in biocatalysis and metabolic engineering.

As mentioned previously, identifying and engineering metabolic bottleneck is one of the common ways to improve productivity in biosynthetic pathways and this also applies to the halogenase pathways.\[127\] The low bromide and fluoride tolerance of host organism is often found to be problematic for large-scale fermentation processes. The alternative is to either engineer host organisms with additional toxicity mitigation mechanisms (‘built-in’ high resistance) or selecting halide tolerant host systems. This was already attempted in SAM-dependent fluorinases by creating an engineered E. coli cell.\[215\] Whereas a higher titer of brominated tryptophan was fermented using a bromide tolerant recombinant Corynebacterium glutamicum strain that carried tryptophan Fl-Hals.\[240\] In phenylpropanoid acid-based pathways, the reconstruction of non-native enzyme could introduce pathway intermediates such as cinnamic and coumaric acids that are toxic to the cell growth.\[241\] The low catalytic efficiency of down-stream pathway enzymes often seems to be an issue with many early stage halogenation metabolic pathways. The down-stream enzymes in natural product pathways have strict substrate specificity for halogenated molecule, which limits the possibilities to alter halogenation sites via combinatorial synthesis using different regioselective halogenases. This is exemplified in an earlier attempt to replace tryptophan 7-halogenase with a 6-halogenase in pyrrolnitrin combinatorial biosynthesis, which ended up with the accumulation of 6-chlorinated intermediate with no downstream enzyme activity.\[116\] The new developments in synthetic biology on microbial chassis engineering, transcription control via synthetic promotors, designing metabolic division of labor in cells etc could also be adapted here to address many of the metabolic bottleneck issues.\[242,243,244,245\]
The emergence of heavy-metal-based cross-coupling chemistries under mild and aqueous conditions pushed advance developments in the chemo-bio catalytic integrative reaction platforms. Several halogenated natural product and biosynthetic pathways could be employed for direct chemo-bio catalytic integration, as carbon–halogen bond is the key substrate for these integrated reactions. This merges synthetic biology with synthetic chemistry, providing a powerful approach toward molecular diversification using specifically installed halogenation tags. Integration was found to be effective when bio and chemo catalytic components were compartmentalized for in vitro biotransformation using purified enzymes.\cite{230} The whole cell reactions with engineered biotransformation have a greater advantage here, with the cellular components acting toward the required natural compartmentalization. The living geno-chemetics approaches were shown to be a way forward for synchronous biosynthesis, where biocatalysis and chemocatalysis could perform in a sideways manner.\cite{150} However, many parallel lessons from the chemo-catalyst research world including catalyst immobilization, separation, recycling, biphasic reactions etc are rarely applied to halogenase-based integrated catalytic platforms. This could be a way forward to enhance the currently limited product titers in such applications. In fact, the recent developments so far in this area have already set a stage for other transition metal-catalyzed reactions, additional regioselective transformations, abiotic reactions in microbial living culture and many un-attempted chemocatalytic reactions in biological context using halogenase enzymes.

**Disclosure statement**

This is to acknowledge that there is no financial interest or benefit that has arisen from the direct applications of this research. The authors declare no conflict of interest.

**Funding**

This work was supported by the UKRI Synthetic Biology for Growth Programme (SBfG) under Grant BB/M017982/1 and by the Royal Society under Grant RGS\(\text{R2}\)180317.

**ORCID**

Binuraj R. K. Menon [http://orcid.org/0000-0001-6852-8053](http://orcid.org/0000-0001-6852-8053)

**References**


