doi.org/10.26434/chemrxiv.9674258.v1

Site-Selective C-H Halogenation using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling

Brian F. Fisher, Harrison M. Snodgrass, Krysten A. Jones, Mary C. Andorfer, Jared C. Lewis

Submitted date: 19/08/2019 • Posted date: 20/08/2019

Citation information: Fisher, Brian F.; Snodgrass, Harrison M.; Jones, Krysten A.; Andorfer, Mary C.; Lewis, Jared C. (2019): Site-Selective C-H Halogenation using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. ChemRxiv. Preprint.

Herein, we describe the use of a high-throughput mass spectrometry-based screen to evaluate a broad set of over one hundred putative FDH sequences drawn from throughout the FDH family. Halogenases with novel substrate scope and complementary regioselectivity on large, three-dimensionally complex compounds were identified. This effort involved far more extensive sequence-function analysis than has been accomplished using the relatively narrow range of FDHs characterized to date, providing a clearer picture of the regions in FDH sequence space that are most likely to contain enzymes suitable for halogenating small molecule substrates. The representative enzyme panel constructed in this study also provides a rapid means to identify FDHs for lead diversification via late-stage C-H functionalization. In many cases, these enzymes provide activities that required several rounds of directed evolution to accomplish in previous efforts, highlighting that this approach can achieve significant time savings for biocatalyst identification and provide advanced starting points for further evolution.

File list (2)

Site-Selective C-H Halogenation using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling

Brian F. Fisher,[1] Harrison M. Snodgrass,[1] Krysten A. Jones,[2] Mary C. Andorfer,[3] Jared C. Lewis*,[1]

Introduction

Enzymes can be powerful tools for the synthesis of fine chemicals, pharmaceuticals,[1] agrochemicals, and many other materials.[2] On the other hand, the very features that give rise to the selectivity and catalytic proficiency of enzymes acting on their native substrates often lead to high substrate specificity and thus poor activity on synthetic substrates. The dearth of enzymes available for reactions of interest can therefore be a major impediment to implementing enzymes in synthetic routes. Many of the enzymes commonly used today (e.g. ketoreductases, transaminases, cytochromes P450, etc.) were originally identified via arduous enzymology aimed at clarifying their native biological activities. Often, many rounds of directed evolution were also required to optimize these enzymes for synthetic applications. Expanding the number of known biocatalysts, with an emphasis on exploring broad sequence diversity within an enzyme family, could therefore greatly facilitate the use of enzymes in chemical synthesis.[3]

Numerous methods have been used to explore the functional diversity of naturally occurring enzymes in discrete genomes, metagenomic samples, and sequence databases.[4] Advances in DNA sequencing—in particular, metagenome sequencing— have resulted in an explosion of the size of protein sequence databases.[5] Coupled with the decreasing cost of gene synthesis,[6] mining these sequence databases for potential biocatalysts is becoming increasingly accessible to scientists.[7] Such approaches are most commonly used to identify enzymes that act on a substrate of interest, often a chromogenic probe compound chosen more for ease of screening than synthetic utility, but efforts to profile the activity of entire enzyme families on a range of substrates and identify biocatalysts with a collectively-broad substrate scope are far less common. Family-wide profiling efforts include investigations on phosphatases,[8] metallo-β-lactamases,[9] and glutathione-S-transferases.[10] Studies on dehalogenases,[11] esterases,[12] glycosyl transferases,[13] and imine reductases[14] highlight the potential synthetic utility of enzymes identified from such efforts. Comparable genome mining efforts on enzymes that functionalize C-H bonds have not been reported.

Our lab has extensively studied flavin-dependent halogenases (FDHs), which catalyze site-selective C-H halogenation of electron-rich aromatic compounds.[15] Our efforts have focused primarily on RebH, an FDH that was identified in studies aimed at elucidating the biosynthetic pathway of the antitumor compound rebeccamycin.[16] In this context, RebH catalyzes site-selective chlorination of tryptophan, and it has since been shown to halogenate a range of indoles and anilines.[17] Our group has also shown that directed evolution can be used to create RebH variants with improved thermal stability,[18] high activity on large, biologically active compounds,[19] and high selectivity for different sites on target compounds.[20] While effective, these efforts required 3-8 rounds of directed evolution due to the wild-type enzyme's modest stability, low activity on large substrates, and high regioselectivity.

While additional FDHs could therefore expand the utility of these enzymes for synthesis, only a relatively narrow set of FDHs have been investigated for biocatalysis.[15] FDHs that catalyze tryptophan chlorination, such as RebH, Thal,[21] SttH,[22] and PrnA,[23] in particular are over-represented. Fungal halogenases, such as Rdc2,[24] RadH,[25] and GsfI,[26] which natively chlorinate phenol-containing substrates, have also been shown to be active and selective biocatalysts. Literature reports on the collective substrate scopes of the FDHs reported to date suggest that they prefer chloride over other halides and that they act on electron-rich compounds similar to their native substrate.[15] On the other hand, the existence of a range of complex halogenated natural products implies that FDHs with unique substrate scopes might be found in less characterized halogenase subgroups.[27] We hypothesized that exploring uncharacterized FDHs found in protein sequence databases could, together with currently characterized enzymes, form a diverse starting toolkit for selective late-stage C-H halogenation.

Herein, we describe the use of a high-throughput mass spectrometry-based screen to evaluate a broad set of over one hundred putative FDH sequences drawn from throughout the FDH family. Halogenases with novel substrate scope and complementary regioselectivity on large, three-dimensionally complex compounds were identified. This effort involved far more extensive sequence-function analysis than has been accomplished using the relatively narrow range of FDHs characterized to date, providing a clearer picture of the regions in FDH sequence space that are most likely to contain enzymes suitable for halogenating small molecule substrates. The representative enzyme panel constructed in this study also provides a rapid means to identify FDHs for lead diversification via late-stage C-H functionalization. In many cases, these enzymes provide activities that required several rounds of directed evolution to accomplish in previous efforts, highlighting that this approach can achieve significant time savings for biocatalyst identification and provide advanced starting points for further evolution.

Results

Organization of Halogenase Sequence Similarity Network

A BLAST search of the UniProt sequence database using RebH as a query sequence and an *E*-value threshold of 10^{-5} generated 3,975 unique hits spanning a range of sequence and host diversity, including bacterial, archaeal, eukaryotic, and viral proteins. Nearly all (>90%) previously reported FDHs are present in this set. The dinucleotide-binding GxGxxG motif, characteristic of FAD-binding proteins, is present in 92% of the sequences, and the WxWxIP motif,[28] characteristic of FDHs but absent in monooxygenases, is found in 69% of the sequences. The latter value increases to 78% when motif variants WxWxI[R,G][29] are included. Collectively, these analyses suggest that the majority of the sequences examined are likely FDHs.

Sequence similarity networks (SSNs)[30] were then used to visualize functional relationships among putative FDH sequences. In this representation, protein sequences are illustrated as nodes in a network graph that are connected by edges (lines) to other sequences that exceed a specified pairwise sequence similarity. An SSN was generated for the entire FDH sequence set using a permissive edge detection threshold (≈30% sequence identity) in EFI-EST[31]. Previously reported data for 129 known enzymes found among the BLAST hits were mapped onto this Level 1 SSN to explore subnetwork co-localization of enzyme properties. The clearest defining features of the individual subnetworks are host domain and compound class—indole, phenol, or pyrrole—of native substrates for known FDHs within the subnetworks (Figure 1A), the latter suggesting that the SSN might provide a framework for identifying

enzymes that act on specific compound classes and for surveying regions of sequence space where substrate preference is unknown.

The largest subnetwork, comprising 2,270 sequences, contains FDHs that either natively halogenate tryptophan or have been shown to catalyze indole halogenation *in vitro*. All known tryptophan FDHs are found in this Indole Subnetwork, including tryptophan 5-, 6-, and 7-halogenases PyrH,[32] SttH,[33] and RebH.[34] BrvH, a halogenase identified from metagenomic analysis,[35] and three recently reported halogenases from *Xanthomonas campestris* are also in this subnetwork. Although the native substrates of these enzymes are not known, they have been shown to halogenate a variety of small indoles. A protein whose structure has been determined as part of structural genomics efforts (PDB: 2PYX)[36] is also present in this subnetwork, although its native activity is also unknown.

The second largest subnetwork comprises 438 sequences from bacteria and fungi and includes most known phenol FDHs that, collectively, halogenate a diverse range of phenol-containing substrates. For example, the bacterial halogenase TiaM chlorinates a large macrocyclic intermediate in the biosynthesis of tiacumicin B.[37] Bacterial halogenases VhaA and Tcp21 chlorinate PCP-tethered amino acids in the biosynthesis of the NRPS glycopeptide antibiotics vancomycin and teicoplanin, respectively.[38] The bacterial iodinase CalO3 is also present in the Phenol Subnetwork,[39] showcasing that substrate diversity also extends to halide specificity. All fungal phenol FDHs that have been studied as biocatalysts on diverse substrates, including Rdc2,[24] RadH,[25] and GsfI,[26] are also contained in this subnetwork.

The third largest subnetwork, with 212 sequences, contains FDHs that are involved in chlorinated pyrrole natural product biosynthesis. The six pyrrole halogenases in the Pyrrole Subnetwork have an average pairwise identity of 87%, and all are annotated in UniProt as PrnC, which halogenates a pyrrole small molecule intermediate in pyrrolnitrin biosynthesis.[40] The halogenase PltM natively halogenates a phenol, phloroglucinol, to produce chlorinated compounds that induce biosynthesis of a pyrrole-containing natural product, pyoluteorin.[41] Two proteins, Dox16 and Dox17, potentially halogenate phenolic moieties during the biosynthesis of pyrrolomycins,[42] pyrrole-containing compounds structurally similar to pyrrolnitrin. These observations suggest that the common ancestor to enzymes in this subnetwork diverged in substrate specificity to yield halogenases specialized for distinct roles in chlorinated pyrrole natural product biosynthesis.

Most of the smaller subnetworks contain uncharacterized proteins, but several include known FDHs with diverse native substrates. One small subnetwork contains several enzymes, including MibH,[43] MscL,[44] and KrmI,[45] that natively chlorinate peptidyl tryptophan sidechains in macrocyclic lanthipeptide and NRPS natural products. Other subnetworks include MalA and MalA', which are responsible for iterative chlorination in the biosynthesis of malbrancheamide,[29] ChlA, which chlorinates a phenol in DIF-1 biosynthesis,[46] and GetL, an enzyme suspected to be responsible for chlorinating histidine in the biosynthesis of tetrapeptide antibiotics[47]. Halogenases responsible for chlorinating ACP-tethered pyrroles (Variant B pyrrole halogenases) such as PltA[48] and Mpy16[49] occupied a subnetwork distinct from the larger subnetwork that included the Variant A pyrrole halogenase PrnC. A few subnetworks appear to contain enzymes that are not FDHs, including the flavin-dependent monooxygenases Qhpg and LodB, which hydroxylate tryptophan residues during cysteine tryptophylquinone cofactor biosynthesis[50] and in peptide substrates,[51] respectively. Another subnetwork, containing exclusively sequences from archaea,

appears to contain geranylgeranyl reductases (GGRs), based on GGR motif abundance and stronger sequence similarity to GGRs annotated in SwissProt.[52]

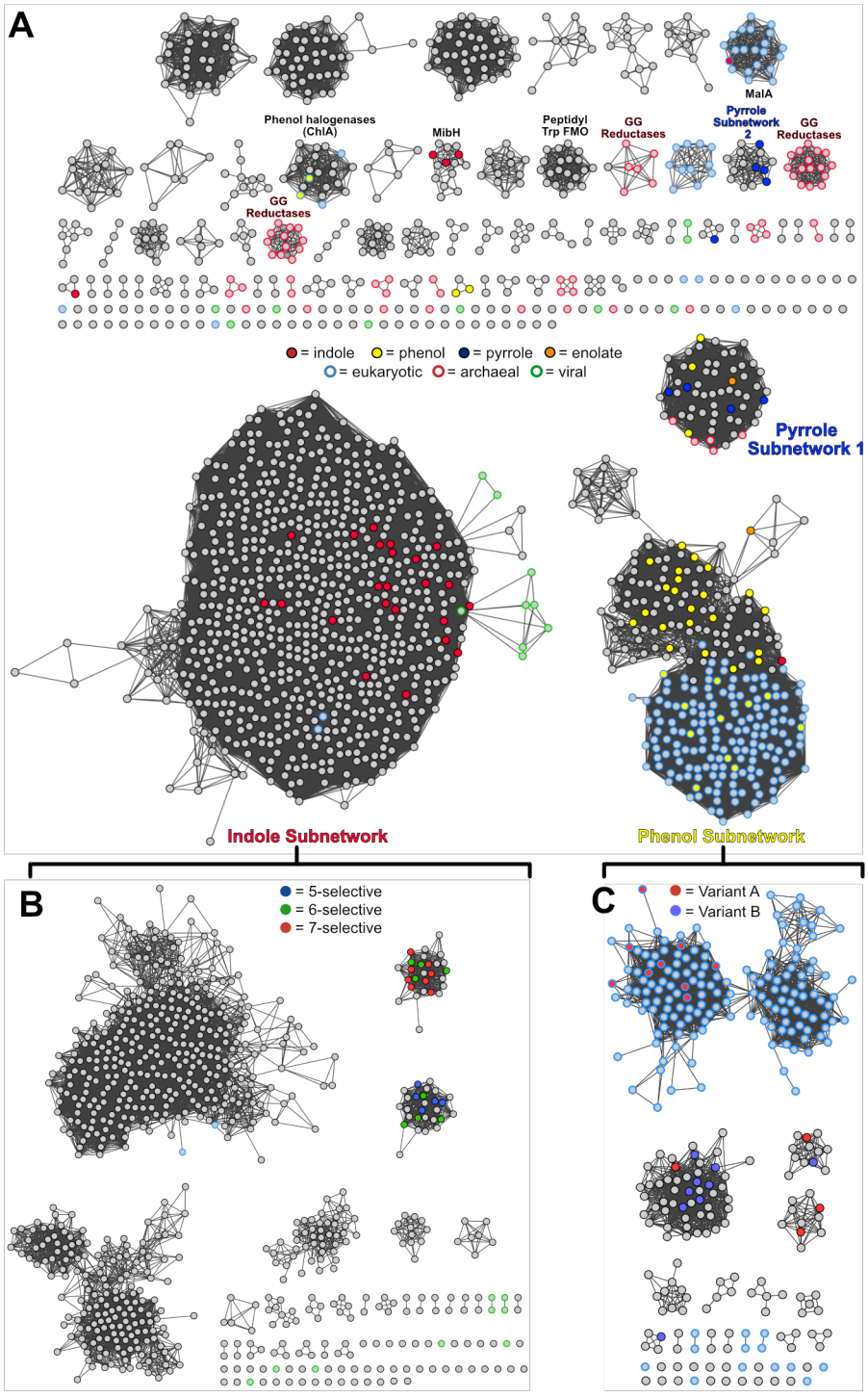


Figure 1. A) Sequence similarity network for flavin-dependent halogenases. Each circle is a representative node, grouping protein sequences with >50% sequence identity as determined by CD-HIT.[53] Edge detection threshold set at alignment score of 70 (≈30% sequence identity). Nodes are filled according to native substrate functional group of at least one sequence in the representative node; colored stroke indicates domain (thin black stroke = bacterial). B and C) Level 2 subnetworks formed from the Indole

(B) and Phenol (C) subnetwork using a stricter alignment score cutoff of 140 (≈40% sequence identity). For Indole Subnetwork sequences, regioselectivity of known tryptophan halogenases is colored. For Phenol Subnetwork sequences, variant of known sequences (A = free small molecule native substrate, B = ACP-tethered native substrate) is colored.

Subnetworks in SSNs can be explored in greater detail by increasing the stringency of the sequence similarity required for edge detection,[30] and notable co-localization of enzymes with similar function in the Level 1 FDH SSN became apparent using a cutoff corresponding to ≈40% pairwise sequence identity. For example, all known tryptophan halogenases localized to only two relatively small Level 2 subnetworks distinguished by their regioselectivity (Figure 1B). All tryptophan 5-halogenases, such as PyrH, localized into one of these, and all tryptophan 7-halogenases, including RebH and PrnA, were found in the other subnetwork. Interestingly, tryptophan 6-halogenases were found roughly evenly distributed between these two subnetworks. The two largest subnetworks within the Level 1 Indole Subnetwork have been sparsely characterized. Only two reports describe the substrate scopes of FDHs within the largest subnetwork in the Indole Subnetwork, which demonstrated that some enzymes in this subnetwork prefer bromination to chlorination.[35, 54] The second largest subnetwork contained the sequence of the structurally characterized but functionally uncharacterized protein 2PYX. Overall, the sparse coverage of the Indole Subnetwork demonstrates that, even among proteins that are most similar to the well-characterized tryptophan halogenases, there remains a vast amount of sequence space to be explored.

Closer inspection of the Phenol Subnetwork at the stricter identity cutoff shows separation of subnetworks based on domain and whether the FDH natively halogenates a free small molecule (variant A) or an acyl carrier protein-tethered small molecule (variant B) (Figure 1C). The largest subnetwork is composed entirely of eukaryotic sequences, and all experimentally characterized proteins within the subnetwork, such as Rdc2, are variant A halogenases. The second-largest subnetwork in this group contains only bacterial sequences, many of which, including VhaA,[55] catalyze chlorination in glycopeptide antibiotic biosynthesis. These halogenases were recently determined to halogenate PCP-tethered amino acids as their native substrates[38] and are therefore variant B. A single variant A halogenase, NcyM, which chlorinates a macrocyclic intermediate in nannocystin biosynthesis, is in this subnetwork.[56]

Expression of Genome-Mined Halogenases

The sequence similarity network outlined above was used as a framework to guide the selection of a diverse set of novel FDHs from each subnetwork. The Phenol Subnetwork was oversampled due to the high structural diversity of substrates natively halogenated by known enzymes in this subnetwork. Other sequences were sampled evenly from the rest of the SSN. Transcriptomic data for sequences from eukaryotes were analyzed using the JGI Mycocosm database to prioritize the synthesis of sequences in order of sequence model quality. Figure 2 depicts the SSN and treemap representations summarizing the distribution of different properties of enzymes in the different subnetworks.

A total of 128 putative halogenase sequences and RebH as a control were codon-optimized and co-expressed with chaperones from the plasmid pGro7 in *E. coli* BL21(DE3) to yield 87 new enzymes in sufficient soluble concentration for further characterization. Halogenases from throughout the entire SSN could be expressed with good titers, but solubility was not evenly distributed (Figure 2D). While 68% of enzymes were soluble, the Indole Subnetwork provided a significantly higher fraction of soluble enzymes compared to others (91%, 42 total). The halogenases in the Phenol Subnetwork had much lower solubility (49% overall, 17 total), which was not significantly influenced by the domain of the source organism (45%

soluble for eukaryotic, and 50% soluble for bacterial genes). An average number of Pyrrole Subnetwork halogenases were soluble (71%, 5 total), while several small subnetworks that were sampled provided no soluble halogenases under the expression conditions tested.

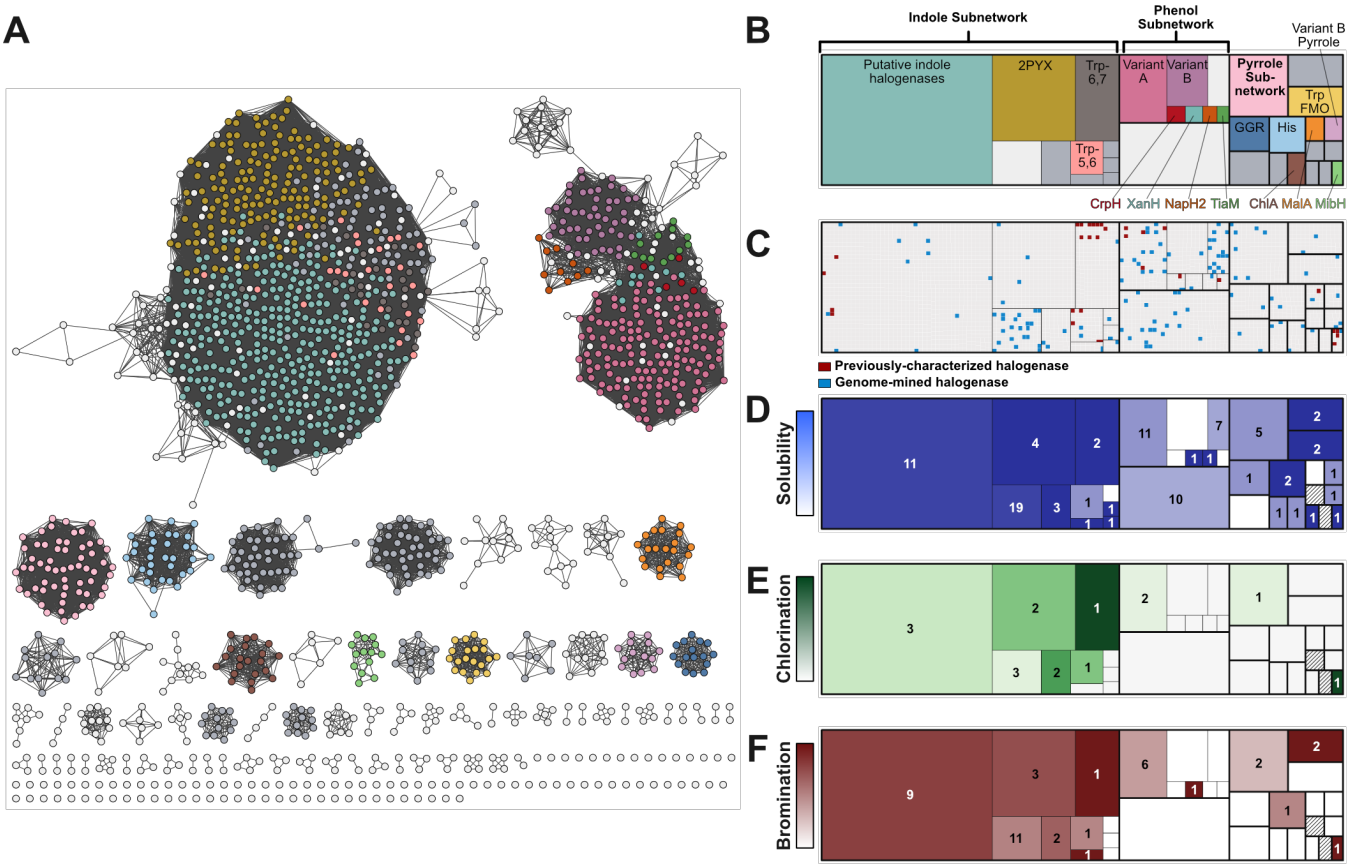


Figure 2. A) Sequence-similarity network for flavin-dependent halogenases, drawn at the less stringent edge detection threshold (≈30% identity), colored according to subnetwork. Subnetworks within the Indole and Phenol subnetworks at the more stringent threshold are colored differently. Subnetworks with fewer than 15 members are colored white; subnetworks without sequences of known or inferred function are colored light gray. B) Treemap illustrating the SSN with the same coloring as A. C) Treemap comparing FDHs previously studied as biocatalysts with FDHs investigated in this study. D) Treemap illustrating solubility of genome-mined enzymes in each subnetwork of the SSN. Color gradient represents fraction of enzymes within the subnetwork that was soluble; diagonal bars indicate subnetworks wherein no enzyme was tested. E and F) Treemaps illustrating the fraction of enzymes in each subnetwork that were capable of chlorinating (E) or brominating (F) at least one substrate in the high-throughput screen (8% conversion threshold).

Probe Substrate High-Throughput Screen

The set of 87 diverse, soluble FDHs was subjected to a high-throughput activity screen to evaluate which enzymes had detectable activity and, for active enzymes, to develop substrate activity profiles to better understand whether activity and subnetwork membership were related. For initial activity screens, a set of 12 probe substrates—4 indoles, 4 anilines, and 4 phenols—was selected from among the substrates previously found by our group to be reactive under FDH chlorination conditions (Figure 3A). The key hypothesis governing selection of these substrates was that their high inherent reactivity, reflected in their

high calculated halenium affinity values,[26, 57] would lead to detectable reactivity with active enzymes even if they exhibited poor binding within FDH active sites. Structural variation within the panel was used to facilitate the identification of viable substrates.[58] Initial screens evaluated both chlorination and bromination activities, the two most common halogenation reactions catalyzed by FDHs. The probe substrate screens required at least 2,040 independent experiments, not including replicates or controls; this heavy screening requirement prompted us to adopt a high-throughput LC-MS-based screen (Figure 3B).[59] Using this method, analysis throughput of up to ≈11 seconds per reaction was achieved, and ultimately ≈20,000 experiments were analyzed.

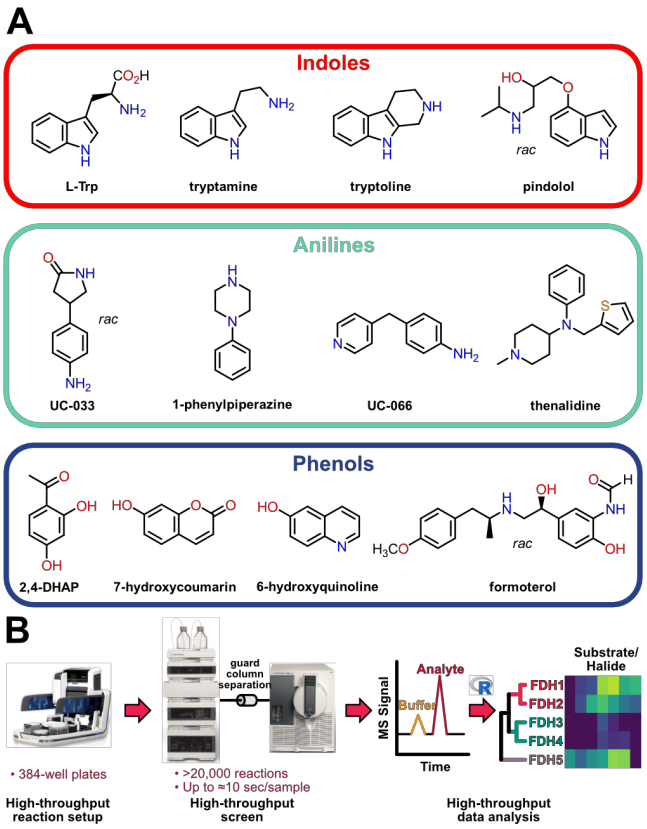


Figure 3. A) Probe substrates included in initial high-throughput screen. B) Scheme summarizing LC-MS-based high-throughput screening method employed.

A total of 39 new halogenases (45% of soluble enzymes) were able to halogenate at least one of the probe substrates. Halogenation of nearly the entire probe substrate panel was achieved by the genome-mined set of enzymes; only formoterol was not halogenated by at least one new halogenase. Overall, bromination activity was more prevalent than chlorination activity. All genome-mined enzymes that were active had brominase activity, but only 16% of the enzyme set had detectable chlorinase activity. Activity was unevenly distributed across the SSN; certain SSNs had a higher abundance of active enzymes than others (Figure 2E-F). The Indole Subnetwork had a particularly high percentage of active enzymes; of the 42 Indole Subnetwork enzymes screened, 27 (64%) were active. The fraction of active enzymes was similar for bacterial and eukaryotic proteins, with 48% of bacterial and 56% of eukaryotic enzymes screened having some activity on probe substrates. One of the three viral proteins tested was active, and none of the six archaeal proteins were active.

The high-throughput screening conversion data for each reaction were plotted as a heatmap, and hierarchical clustering analysis was used to characterize, separately, the similarity of activity profiles for substrates and for FDHs (Figure 4). Substrates tended to form clusters based on their compound class, consistent with the observed similarity of "enzyme-scope" of substrates within the same substrate class.[26] All phenols were present in two substrate clusters, one containing only chlorination reactions and the other containing only bromination reactions. Anilines and indoles were more mixed into the remaining two clusters, but still distinguishable. One of these clusters primarily included indole chlorination, dominated by the high indole chlorination activity of RebH and a highly similar enzyme, 1-B12. The other contained mostly aniline bromination reactions, high activity for which was more broadly distributed.

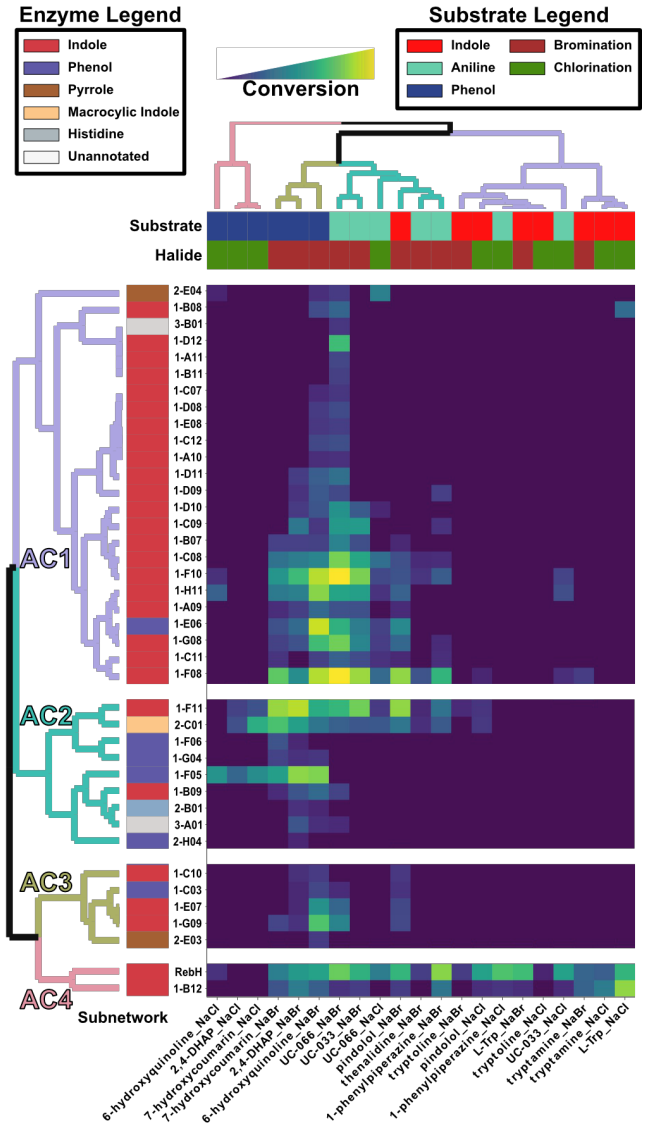


Figure 4. Heatmap of high-throughput screening results, with hierarchical clustering dendrograms for substrate/halide activity similarity and enzyme activity similarity at top and left, respectively. Substrate functional groups and halide used in the reaction are color-coded with bars at the tips of the dendrograms. Only reactions with >8% conversion were included.

Most importantly, enzymes in the same subnetwork tended to cluster together based on their activity profiles. Four activity clusters of enzymes (AC1-4) can be distinguished from the probe substrate high-throughput screening data. AC1, at the top of the heatmap, contained almost exclusively halogenases in the Indole Subnetwork. None of the Indole Subnetwork enzymes in this AC were in the Level 2 tryptophan subnetworks, however, and they were distinguished by their preference for bromination of phenols and anilines. Despite the fact that indoles are the most common substrates known to be halogenated by enzymes in the Indole Subnetwork, halogenase activity on indoles in AC1 was limited. Pindolol was the only indole halogenated by more than one enzyme, and only a single FDH, 1-F08 (34% identical to SttH), chlorinated more than one indole.

Activity cluster 2 (AC2) had similar bromination scope to AC1, but had higher breadth of phenol chlorination activity. Only two of the nine enzymes in this activity cluster were present in the Indole Subnetwork, whereas four were in the Phenol Subnetwork. Halogenase 1-F11, from an unannotated subnetwork within the Indole Subnetwork (38% identical to tryptophan-5 halogenase ClaH), and 2-C01, a halogenase in the same subnetwork as the lanthipeptide indole halogenase MibH (36% identical),[43] were capable of chlorinating multiple phenols, 2,4-dihydroxyacetophenone (2,4-DHAP) and 7-hydroxycoumarin. The FDH 2-C01 was particularly versatile in halide scope. UC-066, 7-hydroxycoumarin, and 2,4-DHAP were chlorinated and brominated by 2-C01 with similar yields. Enzyme 1-F05, from the Phenol Subnetwork (49% identical to ArmH4[60]), was similarly versatile in the halides it accepted, but its activity was specific for phenolic probe substrates. It had the broadest phenol substrate scope of any enzyme tested, but did not halogenate any aniline or indole.

Activity cluster 3 (AC3) was small and populated by low-activity enzymes only having bromination activity on the substrates that were most easily halogenated. AC4 contained only two enzymes, RebH and 1-B12, that had the broadest substrate scope, particularly on indole probe substrates. The high probe substrate scope of RebH was expected by design, since the indoles and anilines of the probe panel was assembled from substrates that were known to be chlorinated by RebH. The enzyme 1-B12 has high sequence similarity to RebH (64% identical) and a strongly similar substrate activity profile.

Activity and Selectivity of Mined Halogenases Toward Complex Substrates

Based on the remarkable activity with that our genome-mined halogenases exhibited toward probe substrates, we next wondered whether they might be capable of halogenating substrates that were not selected from a set of easily halogenated compounds. A total of 50 larger and more three-dimensionally complex additional substrates were selected for these activity studies (Figure 5A). Among the compounds in this expanded substrate set were yohimbine, a compound for which we previously evolved halogenase activity from RebH,[19] and premalbrancheamide, a compound natively halogenated by the FDH MalA.[29] Most of the substrates have not been reported as FDH substrates previously, including β-estradiol 17-(β-D-glucuronide), an estrogen metabolite, and cabergoline, an ergot alkaloid. A total of 48% of the more complex substrates tested were halogenated by at least one halogenase under the non-optimized conditions used in the high-throughput screen (Figure 5B). Hierarchical clustering was performed on the reaction data as above. However, the similarities between enzymes were substantially lower than in the clustering analysis of the probe substrate data, and activity clusters were consequently less defined.

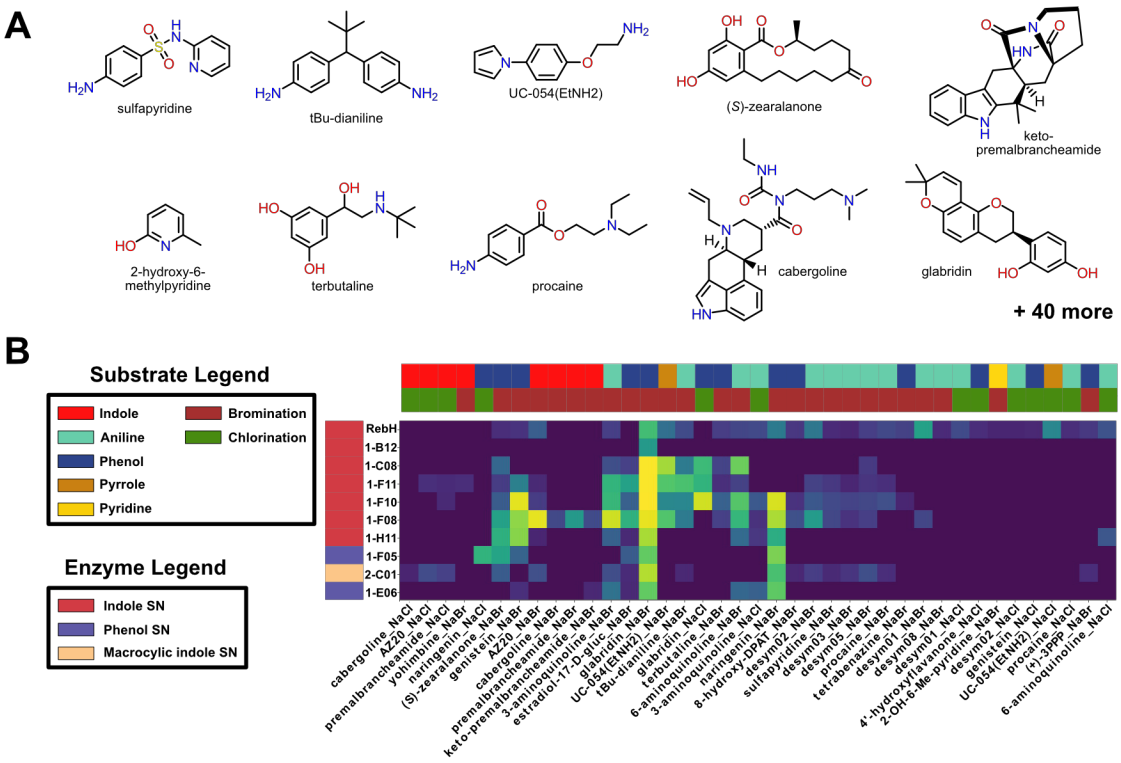


Figure 5. A) Representative compounds included in expanded high-throughput substrate screen, each of which was halogenated by at least one genome-mined FDH. B) Heatmap of expanded substrate screen data with ten of the most active enzymes from the probe high-throughput substrate screen.

Larger quantities of several of the most active genome-mined FDHs were expressed, purified, and used for preparative-scale bioconversions on a subset of the larger substrates evaluated (Figure 6). Premalbrancheamide is a compound natively dichlorinated by MalA at C5 and C6, and which has been shown to be halogenate at the C5 or C6 positions non-selectively using either chloride or bromide as halide sources.[29] The Indole Subnetwork FDH 1-F08 preferentially brominates premalbrancheamide at C5 in 51% isolated yield, and it also brominates AZ20, a selective ATR kinase inhibitor, in 28% isolated yield at the indole C3 position. A different Indole Subnetwork enzyme, 1-F11, was capable of brominating β-estradiol 17-(β-D-glucuronide), an estradiol metabolite, at the C4 position of the phenol in 57% yield and the 1-position of the carvedilol carbazole ring in 56% isolated yield.

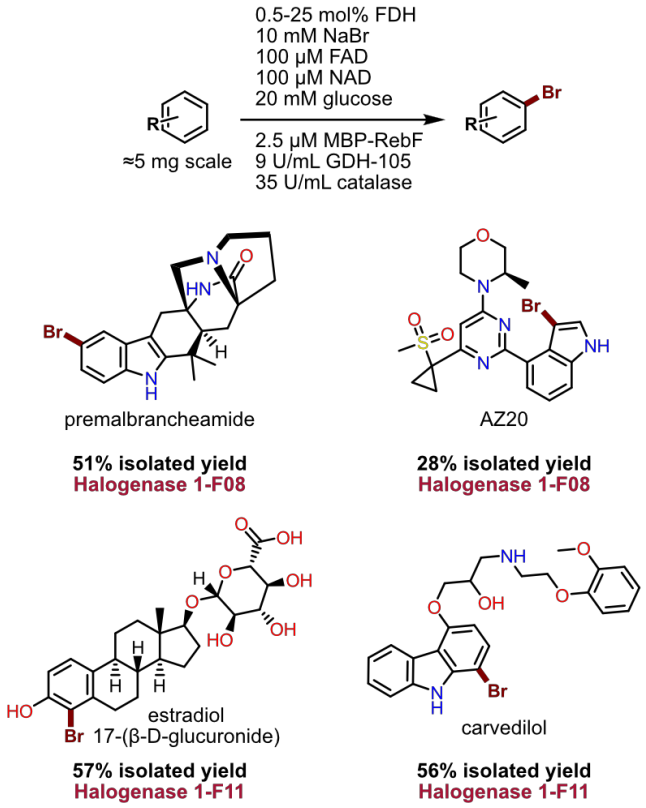


Figure 6. Preparative-scale bioconversions of larger substrates.

Many examples of reactions in which different regioisomers were formed by different enzymes were also identified (Figure 7). Pindolol, which is brominated at C7 by RebH variants,[61] is also brominated at C7 by the Indole Subnetwork FDH 1-F11. The MibH subnetwork enzyme 2-C01, on the other hand, preferentially brominates at C2. This finding is notable since C2 is less electronically activated than C7 based on its 3 kcal/mol lower halenium affinity (HalA), a metric for computationally evaluating the reactivity of different positions of a molecule toward EAS.[57] Naringenin is brominated at two different positions using 1-F11 or Phenol Subnetwork enzyme 1-F05. Despite the negligible energetic differences in HalA for C6 and C8 (0.7 kcal/mol), 1-F05 was found to be >95% selective for C8, while 1-F11 and other FDHs were found to have only minor preferences in regioselectivity for C8 or C6. Trp-6,7 halogenase subnetwork enzyme 1-B12 halogenates the indole-containing compound methylergonovine at C7, which has a halenium affinity 4.2 kcal/mol lower than C2, the most nucleophilic aromatic C-H site on this compound. The FDH 2-C01, on the other hand, brominates methylergonovine at C2.

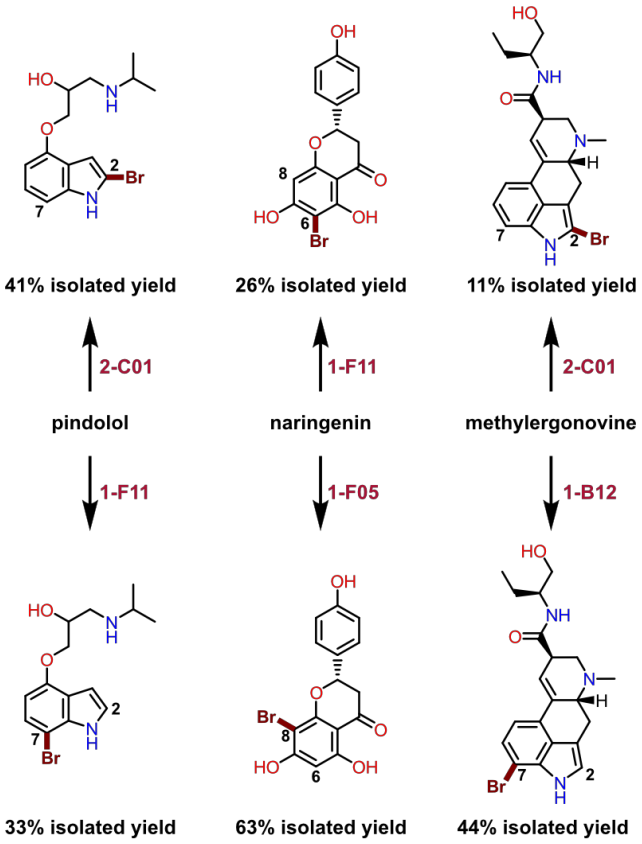


Figure 7. Regiocomplementary halogenation of large molecules.

Discussion

A Family-wide View of FDH Properties

Family-wide analysis of FDHs revealed several notable trends that are not apparent from prior studies. First, FDHs from diverse host organisms can be solubly expressed without significant optimization of expression conditions. Bacterial enzymes had the highest soluble expression success rate (76%), while a lower fraction (40%) of eukaryotic enzymes were soluble. Notably, however, the lower fraction of soluble eukaryotic FDHs reflects the poorer solubility of halogenases in the Phenol Subnetwork regardless of host organism domain. Nearly all (20/23) of the eukaryotic proteins evaluated were within the Phenol Subnetwork. Within this subnetwork, the soluble expression rate is generally low, but it is actually higher for proteins from eukaryotes (54%) relative to bacteria (44%). This finding indicates that eukaryotic FDHs can be readily expressed in *E. coli* and that genome mining efforts should be encouraged to include enzymes from eukaryotic species.[62]

Second, halogenase activity was also evenly distributed between enzymes from bacterial and eukaryotic organisms (48% and 56% active, respectively). FDH activity was not observed for any archaeal proteins evaluated, consistent with the strong possibility that most if not all archaeal sequences in the SSN are geranylgeranyl reductases. Interestingly, one viral FDH, a cyanophage auxiliary metabolic gene product,[63] was active, though its activity and substrate scope were low (conversion of <35% on only three probe substrates was observed). In general, the identification of such a high percentage of active halogenases, despite the use of non-native substrates for activity profiling and a lax homology requirement

for evaluation (*E*-value threshold of 10^{-5}), suggests that this family contains a large number of enzymes suitable for biocatalysis.

Third, bromination activity was much more widespread than chlorination activity within the FDHs surveyed. The majority of the FDH biocatalysis literature focuses on chlorination activity because most FDHs reported to date are involved in the biosynthesis of chlorinated natural products. Moore[64] has reported three flavin dependent brominases involved in the biosynthesis of brominated natural products, but these are more distantly related to enzymes comprising the SSN in the current study. These brominases have 17 ± 4% sequence identity to enzymes in the SSN; for comparison, RebH exhibits 29 ± 10% sequence identity to our genome-mined enzymes. Sewald[35, 54] reported flavin dependent halogenases (contained in the Indole Subnetwork of the FDH SSN) that prefer bromide over chloride when acting on the (presumably) non-native substrate indole. While this observation was taken to indicate specificity of these enzymes toward bromide, our findings indicate that a preference for bromination is common in FDHs. We suggest that the higher electrophilicity of bromine relative to chlorine in heteroatom-X species,[65] such as the proposed hypohalous acid or haloamine halogenating agents in FDH catalysis, leads to more facile bromination. For example, the native chlorinase RebH can brominate a greater range of non-native substrates than it can chlorinate. Preference for bromination over chlorination for non-native as well as native substrates is also observed when both Cl^- and Br^- are present in solution. In competition reactions including both NaCl and NaBr, RebH prefers bromide over chloride for L-tryptophan, 1-phenylpiperazine, pindolol, and 2,4-dihydroxyacetophenone halogenation.[52] It is therefore possible that enzymes with higher bromination than chlorination scope in our high-throughput screen could nevertheless natively catalyze chlorination reactions.

Analyzing FDH Activity using Sequence Similarity Networks and Activity Clustering

Sequence similarity networks provide an intuitive structure for exploring the protein sequence space of enzyme families. The FDH SSN contains Level 1 subnetworks comprising enzymes with similar native substrate preferences (indole vs. phenol, etc.). At a more stringent identity threshold cutoff, Level 2 subnetworks with finer functional distinction are revealed. Within the Level 1 Phenol Subnetwork, for example, different Level 2 subnetworks containing primarily either variant A or variant B halogenases, which natively halogenate free small molecules or PCP-tethered substrates respectively, are observed. The ability to distinguish enzyme such subclasses based on sequence alone is useful for focusing future genome mining efforts since our data indicate that neither of the variant B phenol halogenases examined were even soluble. Information on site selectivity could also be obtained directly from sequence information in some cases. For example, within the Level 1 Indole Subnetwork, separation of tryptophan 5- and 7-halogenases into distinct Level 2 subnetworks was apparent, though tryptophan 6-halogenases were roughly evenly distributed between these subgroups.

Of the Level 1 subnetworks examined, only six contained enzymes with measurable chlorination or bromination activity on our probe substrate set, but these subnetworks contained 78% of the FDHs within the SSN. No subnetwork with fewer than 15 members contained active enzymes; most enzymes in the Indole Subnetwork (66%) were active, several Phenol Subnetwork enzymes (42%) were active, and two of the five Pyrrole Subnetwork enzymes were active. Both enzymes in unannotated subnetwork 4 were active, and one of two enzymes in subnetwork 8 was active, but activity was low for these enzymes. FDH 2-C01 was the only member of the MibH subnetwork evaluated, but it was highly active. These

findings reflect the nature of the probe substrates chosen, but given the range of substrates examined and the similarity of these substrates to pharmaceuticals and other fine chemicals, they also highlight regions of FDH sequence space most likely to be of interest for biocatalysis.

Activity was present in most Level 2 subnetworks within the Indole Subnetwork. All enzymes in the Trp halogenase subnetworks were active, as was the only enzyme in the unannotated subnetwork 21. Most enzymes in the putative indole halogenase subnetwork, which includes BrvH, were active (84%). Three out of four enzymes in the subnetwork that includes the putative *S. frigidimarina* halogenase were active, and two of three enzymes, 1-C08 and 1-F11, in unannotated subnetwork 9 were active. In contrast with the finding that active enzymes in small Level 1 subnetworks is rare, slightly more than half (56%) of enzymes in subnetworks with fewer than 15 members within the Indole Subnetwork were active. The activity results within the Indole Subnetwork broadly show that a high fraction of these enzymes have potential as useful biocatalysts and highlight several underexplored regions in the FDH sequence space that merit further investigation.

Analysis of Level 2 Phenol Subnetworks also highlights regions with high potential for biocatalyst identification. The majority of the enzymes in the Variant A phenol halogenase subnetwork, including 1-F05, were active (66%). The low solubility of Phenol Subnetwork enzymes resulted in a small number of enzymes in smaller subgroups, making conclusions regarding these small Phenol Subnetwork subgroups difficult to make. The only enzyme in the large phenol halogenase subnetwork containing XanH was active. Both of the Variant B halogenases were insoluble under the conditions examined. The single enzyme in the NapH2 subnetwork was inactive, as were the six soluble enzymes that were either singletons or within small (<15 members) subnetworks. Overall, the Variant A subgroup within the Phenol Subnetwork shows clear promise as a source of novel biocatalysts, but further study of other subnetworks would be required to get a clearer picture of their potential.

Finally, functional characterization of enzymes across the FDH SSN demonstrated that enzymes within a Level 1 subnetwork have similar activity profiles on smaller probe substrates, but that this trend diminishes on more complex substrates. Not surprisingly, more closely-related enzymes possess more similar activity profiles. Highly similar substrate activity profiles are observed for the tryptophan 6,7-halogenase subnetwork enzymes RebH and 1-B12 (64% identical), as well as for BrvH subnetwork enzymes 1-H11 and 1-F10 (50% identical), suggesting an approximate %ID threshold for future genome-mining of new halogenases with similar substrate scopes. Because the gene selection process of this study intentionally favored diverse sequences to maximize the breadth of our search for new halogenases, however, there are few instances of such similar enzyme pairs in which both were soluble and highly active. The average %ID for the most closely-related enzyme in the genome-mined set was 41.2 ± 12.4%, likely too low for similarities in activity profiles among enzymes to result in consistent trends. More thorough genome mining of subnetworks with highly active FDHs could yield more concrete activity profiles and reveal more detailed information regarding enzyme substrate preferences.

Unique Activity and Selectivity of Mined Halogenases

The selectivity of FDHs on their native substrates has driven interest in these enzymes as biocatalysts. Because a given FDH may not provide the selectivity required for a particular application, a number of labs have explored the use of targeted mutations to alter FDH selectivity. While grafting key residues from one tryptophan halogenase into another has been used to switch selectivity on tryptophan,[21]

modest selectivity has generally been reported for efforts focused on non-native substrates (e.g. converting SttH from 90% 6-selective to 75% 5-selective for 3-indolepropionic acid chlorination).[22, 66] To address this issue, our lab established that directed evolution can be used to generate FDHs with high (>90%) regioselectivity for different sites on a single substrate (tryptamine), and that the resulting enzymes also had altered selectivity on a range of other substrates.[20] Several rounds of evolution were required to achieve this goal, so accelerating the identification of FDHs with complementary regioselectivity on non-native substrates remains an important goal.

Gratifyingly, a number of enzymes identified in our family-wide survey of FDH activity exhibited regiocomplementarity on a number of structurally complex substrates. For example, the enzyme 2-C01 often provided different regiochemical outcomes than other halogenases. This FDH is present in a subnetwork along with MibH, which natively chlorinates a tryptophan indole ring in a large lanthipeptide.[43] MibH has a large, hydrophobic binding pocket in order to accommodate its native substrate. The genome-mined halogenase 2-C01 may have a similar active site, which could accommodate large substrates in distinct binding poses. This finding suggests that 2-C01 could be a promising starting point for evolving FDHs to achieve late-stage functionalization of aryl C−H bonds with distinct regioselectivity relative to other FDHs characterized to date.

Comparison of the Genome-Mined Halogenase Library with Evolved Variants

Enzymes frequently require substantial modification before they are capable of being deployed as useful catalysts for organic synthesis. As the regiocomplementarity noted above illustrates, access to a diverse pool of enzymes that can serve as starting points for directed evolution can greatly expedite biocatalyst identification. While evolving a single enzyme can take a great deal of effort and may ultimately fail to provide the desired levels of improvement, a related enzyme may be better suited initially to the task and can drastically reduce the effort required to obtain a desired biocatalyst. This point is illustrated by several enzymes in our genome-mined set, which perform comparably to evolved RebH variants with increased thermal stability,[18] expanded substrate scope,[19] and altered regioselectivity[20].

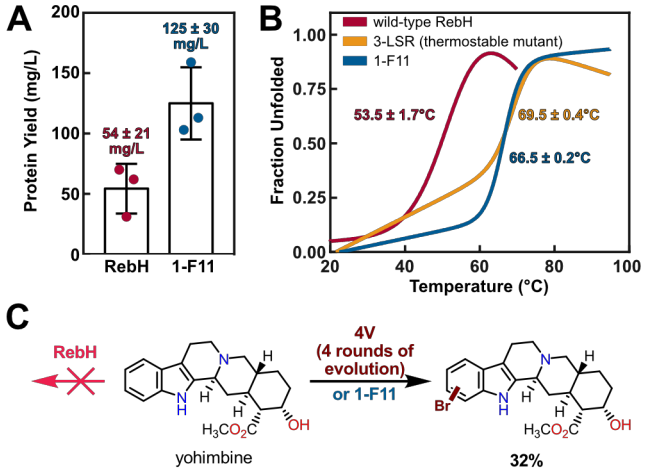


Figure 8. A) Comparison of isolated RebH and 1-F11 protein yields after Ni-NTA purification from 50 mL expression cultures. B) Comparison of CD thermal melts of RebH, thermostable RebH variant 3-LSR, and genome-mined halogenase 1-F11. Curves shown are best fit for thermal unfolding monitored at 222 nm using CDPal.[67] C) Wild-type RebH required several rounds of directed evolution before

yohimbine halogenation was detectable. Halogenase 1-F11 can halogenate yohimbine without directed evolution (HPLC conversion shown).

The enzyme 1-F11 provides an instructive example of a genome-mined FDH that holds significant promise as an advanced starting point for further evolution. FDH biocatalysis is often hampered by low protein expression yields,[45] therefore a more soluble starting enzyme for FDH directed evolution would be especially attractive. The FDH 1-F11 expression yield was 125 ± 30 mg/L from a 50 mL expression culture, higher than that of RebH, expression of which yielded 54 ± 21 mg/L enzyme under analogous expression conditions (Figure 9A). Higher halogenase activity in lysate on numerous substrates is also observed for 1-F11 compared with RebH.[52] Despite originating from a mesophilic sphingomonas species within an *A. thaliana* root microbiome, 1-F11 has comparable thermal stability (T_m = 66.5 °C) to RebH variant 3-LSR (T_m = 69.5 °C, Figure 9B), which was evolved over three rounds of directed evolution for improved thermal stability[18]. Since more stable enzymes can have a longer catalytic lifetime and can better tolerate random mutations,[68] 1-F11 provides a convenient starting point for directed evolution. 1-F11 also compares favorably in substrate scope with the RebH mutant 4V, which was evolved over four rounds of directed evolution for the late-stage C−H functionalization of yohimbine, a complex, biologically active molecule.[19] Because RebH has minimal activity on yohimbine, an substrate-walking directed evolution approach was required to evolve an enzyme that could halogenate this compound. Enzyme 1-F11, on the other hand, was capable of brominating yohimbine without any modification through directed evolution (Figure 9C). Given the broad substrate scope of 1-F11, we envision it could be an ideal starting point for directed evolution.

Conclusions

FDHs were first characterized in the mid-1990s. Since this time, most of the FDHs reported have come from either studies on individual biosynthetic pathways or genome mining efforts targeting specific organisms or metagenomic samples.[15] Figures 2C and 2E/F illustrate how these efforts have focused on a remarkably narrow range of FDH sequence space and missed out on large swaths of this space that contain functional enzymes, respectively. Family-wide activity analysis shows that similar fractions of FDHs from bacteria and fungi are soluble and active and that bromination is more commonly observed than chlorination. Broader sampling of this space has not only led to the identification of new enzymes with unique catalytic properties, but also highlighted regions of sequence space that are ripe for further exploration. As noted above, this is not to say that other regions might be suitable for very different types of substrates, but for the electron-rich aromatic compounds explored to date, these regions are clearly privileged. Moreover, the SSNs and substrate activity profiles developed in this study offer predictive ability for focusing biocatalyst selection or further genome mining efforts for particular applications. Extending this approach, involving SSN-guided selection of sequences from throughout an enzyme family, label-free high-throughput mass spectrometry screening using synthetic probe substrates, and activity profiling, to other enzymes has great potential for expediting biocatalyst identification.

Beyond these family-wide findings, a number of remarkably useful enzymes were identified in the representative set that was explored. Particularly notable in this regard are 1-F11, 1-F08, 2-C01, 1-F05, and 1-B12. Collectively, these enzymes enable C-H halogenation of previously inaccessible substrates, provide complementary site selectivity on complex biologically active substrates, and exhibit improved thermostability relative to a commonly reported FDH. Their sequences also differ significantly from other

FDHs that have been explored *in vitro*. With the exception of 1-B12, which is 64% identical to RebH, they are only 34-43% identical to FDHs that have been explored as biocatalysts. These novel and diverse halogenases therefore represent promising starting points for both directed evolution and additional genome mining aimed at identifying similarly-effective biocatalysts. The activity of these enzymes on complex natural products and pharmaceuticals also suggests that their native substrates could be similarly fascinating structures. It could therefore be interesting to examine the native function of these enzymes, reversing the enzymology-to-biocatalysis progression that has dominated biocatalyst development to date.[3b] This approach could provide a unique means of identifying new halogenated natural products and other unique compounds when extended to other enzyme classes.

Acknowledgements

This study was supported by the NIH (1R01GM115665) and by the NSF under the CCI Center for Selective C-H Functionalization (CCHF, CHE-1700982). B.F.F. was supported by an NIH F32 postdoctoral fellowship (F32GM123693). Halogenase genes were provided by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported under Contract No. DE-AC02-05CH11231, and we thank Drs. Sam Deutch, Yasuo Yoshikuni, Igor Grigoriev, Angela Tarver, and Sangeeta Nath for assistance with gene selection, transcriptomic analysis, and synthesis. We thank CCHF members for helpful discussions, Dr. Chris Welch for helpful comments regarding high-throughput screening, and Prof. Richmond Sarpong, Jose Roque, and Dr. David Stephens for providing samples of premalbrancheamide and keto-premalbrancheamide for halogenation studies.

Present Author Addresses

[1]Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States
[2]Department of Chemistry, University of Chicago, Chicago, Illinois 60637, United States
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

References

1. Devine, P. N.; Howard, R. M.; Kumar, R.; Thompson, M. P.; Truppo, M. D.; Turner, N. J., *Nature Reviews Chemistry* **2018**, 1.
2. Woodley, J. M., *Appl. Microbiol. Biotechnol.* **2019,** *103*, 4733.
3. (a) Truppo, M., *ACS Medicinal Chemistry Letters* **2017**; (b) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., *Nature* **2012,** *485*, 185.
4. Zaparucha, A.; Berardinis, V. d.; Vaxelaire-Vergne, C., Genome Mining for Enzyme Discovery. In *Modern Biocatalysis: Advances Towards Synthetic Biological Systems*, Williams, G.; Hall, M., Eds. 2018; pp 1.
5. Chen, C.; Huang, H.; Wu, C. H., Protein Bioinformatics Databases and Resources. In *Protein Bioinformatics: From Protein Modifications and Networks to Proteomics*, Wu, C. H.; Arighi, C. N.; Ross, K. E., Eds. Springer New York: New York, NY, 2017; pp 3.
6. Hughes, R. A.; Ellington, A. D., *Cold Spring Harbor Perspectives in Biology* **2017,** *9*, a023812.
7. Kamble, A.; Srinivasan, S.; Singh, H., *Molecular Biotechnology* **2019,** *61*, 53.
8. Huang, H.; Pandya, C.; Liu, C.; Al-Obaidi, N. F.; Wang, M.; Zheng, L.; Keating, S.; Aono, M.; Love, J. D.; Evans, B.; Seidel, R. D.; Hillerich, B. S.; Garforth, S. J.; Almo, S. C.; Mariano, P. S.; Dunaway-Mariano, D.; Allen, K. N.; Farelli, J. D., *Proc. Natl. Acad. Sci. USA* **2015,** *112*.

9. (a) Baier, F.; Tokuriki, N., *J. Mol. Biol.* **2014,** *426*, 2442; (b) Baier, F.; Chen, J.; Solomonson, M.; Strynadka, N.; Tokuriki, N., *ACS Chem. Biol.* **2015,** *10*, 1684.
10. Mashiyama, S. T.; Malabanan, M. M.; Akiva, E.; Bhosle, R.; Branch, M. C.; Hillerich, B.; Jagessar, K.; Kim, J.; Patskovsky, Y.; Seidel, R. D.; Stead, M.; Toro, R.; Vetting, M. W.; Almo, S. C.; Armstrong, R. N.; Babbitt, P. C., *PLoS Biology* **2014,** *12*, e1001843.
11. Chan, W. Y.; Wong, M.; Guthrie, J.; Savchenko, A. V.; Yakunin, A. F.; Pai, E. F.; Edwards, E. A., *Microbial Biotechnology* **2010,** *3*, 107.
12. Martínez-Martínez, M.; Coscolín, C.; Santiago, G.; Chow, J.; Stogios, P. J.; Bargiela, R.; Gertler, C.; Navarro-Fernández, J.; Bollinger, A.; Thies, S.; Méndez-García, C.; Popovic, A.; Brown, G.; Chernikova, T. N.; García-Moyano, A.; Bjergah, G. E. K.; Pérez-García, P.; Hai, T.; Pozo, M. V.; Stokke, R.; Steen, I. H.; Cui, H.; Xu, X.; Nocek, B. P.; Alcaide, M.; Distaso, M.; Mesa, V.; Pelaez, A. I.; Sánchez, J.; Buchholz, P. C. F.; Pleiss, J.; Fernández-Guerra, A.; Glöckner, F. O.; Golyshina, O. V.; Yakimov, M. M.; Savchenko, A.; Jaeger, K.-E.; Yakunin, A. F.; Streit, W. R.; Golyshin, P. N.; Guallar, V.; Ferrer, M., *ACS Chem. Biol.* **2017**.
13. Yang, M.; Fehl, C.; Lees, K. V.; Lim, E.-K.; Offen, W. A.; Davies, G. J.; Bowles, D. J.; Davidson, M. G.; Roberts, S. J.; Davis, B. G., *Nat. Chem. Biol.* **2018**, 1.
14. Wetzl, D.; Berrera, M.; Sandon, N.; Fishlock, D.; Ebeling, M.; Müller, M.; Hanlon, S.; Wirz, B.; Iding, H., *ChemBioChem* **2015,** *16*, 1749.
15. Latham, J.; Brandenburger, E.; Shepherd, S. A.; Menon, B. R. K.; Micklefield, J., *Chem. Rev.* **2018,** *118*, 232.
16. Yeh, E.; Garneau, S.; Walsh, C. T., *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 3960.
17. (a) Payne, J. T.; Andorfer, M. C.; Lewis, J. C., *Angew. Chem. Int. Ed.* **2013,** *52*, 5271; (b) Payne, J. T.; Andorfer, M. C.; Lewis, J. C., *Angew. Chem.* **2013,** *125*, 5379.
18. Poor, C. B.; Andorfer, M. C.; Lewis, J. C., *Chembiochem* **2014,** *15*, 1286.
19. (a) Payne, J. T.; Poor, C. B.; Lewis, J. C., *Angew. Chem.* **2015,** *127*, 4300; (b) Payne, J. T.; Poor, C. B.; Lewis, J. C., *Angew. Chem. Int. Ed.* **2015,** *54*, 4226.
20. Andorfer, M. C.; Park, H. J.; Vergara-Coll, J.; Lewis, J. C., *Chem. Sci.* **2016,** *7*, 3720.
21. Moritzer, A.-C.; Minges, H.; Prior, T.; Frese, M.; Sewald, N.; Niemann, H. H., *J. Biol. Chem.* **2019,** *294*, 2529.
22. Shepherd, S. A.; Menon, B. R. K.; Fisk, H.; Struck, A. W.; Levy, C.; Leys, D.; Micklefield, J., *ChemBioChem* **2016,** *17*, 821.
23. Lang, A.; Polnick, S.; Nicke, T.; William, P.; Patallo, E. P.; Naismith, J. H.; van Pée, K.-H. H., *Angew. Chem. Int. Ed.* **2011,** *50*, 2951.
24. Zeng, J.; Lytle, A. K.; Gage, D.; Johnson, S. J.; Zhan, J., *Bioorg. Med. Chem. Lett.* **2013,** *23*, 1001.
25. Menon, B. R. K.; Brandenburger, E.; Sharif, H. H.; Klemstein, U.; Shepherd, S. A.; Greaney, M. F.; Micklefield, J., *Angew. Chem. Int. Ed.* **2017,** *56*, 11841.
26. Andorfer, M. C.; Grob, J. E.; Hajdin, C. E.; Chael, J. R.; Siuti, P.; Lilly, J.; Tan, K. L.; Lewis, J. C., *ACS Catal.* **2017,** *7*, 1897.
27. Gribble, G. W., *Environmental Chemistry* **2015,** *12*, 396.
28. Podzelinska, K.; Latimer, R.; Bhattacharya, A.; Vining, L. C.; Zechel, D. L.; Jia, Z., *J. Mol. Biol.* **2010,** *397*, 316.
29. Fraley, A. E.; Garcia-Borràs, M.; Tripathi, A.; Khare, D.; Mercado-Marin, E. V.; Tran, H.; Dan, Q.; Webb, G. P.; Watts, K. R.; Crews, P.; Sarpong, R.; Williams, R. M.; Smith, J. L.; Houk, K. N.; Sherman, D. H., *J. Am. Chem. Soc.* **2017,** *139*, 12060.
30. Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C., *PLoS ONE* **2009,** *4*, e4345.

31. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L., *BBA – Proteins and Proteomics* **2015,** *1854*, 1019.
32. Zehner, S.; Kotzsch, A.; Bister, B.; Süssmuth, R. D.; Méndez, C.; Salas, J. A.; van Pée, K. H., *Chem. Biol.* **2005,** *12*, 445.
33. Zeng, J.; Zhan, J., *Biotechnol. Lett.* **2011,** *33*, 1607.
34. Sanchez, C.; Butovich, I. A.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A., *Chem. Biol.* **2002,** *9*, 519.
35. Neubauer, P. R.; Widmann, C.; Wibberg, D.; Schröder, L.; Frese, M.; Kottke, T.; Kalinowski, J.; Niemann, H. H.; Sewald, N., *PLOS ONE* **2018,** *13*, e0196797.
36. PDB ID: 2PYX. Joint Center for Structural Genomics (JCSG). Crystal structure of tryptophan halogenase (YP_750003.1) from Shewanella frigidimarina NCIMB 400 at 1.50 A resolution.
37. Xiao, Y.; Li, S.; Niu, S.; Ma, L.; Zhang, G.; Zhang, H.; Zhang, G.; Ju, J.; Zhang, C., *J. Am. Chem. Soc.* **2011,** *133*, 1092.
38. Kittilä, T.; Kittel, C.; Tailhades, J.; Butz, D.; Schoppet, M.; Büttner, A.; Goode, R. J. A.; Schittenhelm, R. B.; Pee, K.-H.; Süssmuth, R. D., *Chem. Sci.* **2017,** *8*, 5992.
39. Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Czisny, A.; Whitwam, R. E.; Farnet, C. M.; Thorson, J. S., **2002,** - *297*.
40. Kirner, S.; Hammer, P. E.; Hill, D. S.; Altmann, A.; Fischer, I.; Weislo, L. J.; Lanahan, M.; van Pée, K.-H.; Ligon, J. M., *J. Bacteriol.* **1998,** *180*, 1939.
41. Yan, Q.; Philmus, B.; Chang, J. H.; Loper, J. E., *eLife* **2017,** *6*, e22835.
42. Zhang, X.; Parry, R. J., *Antimicrob. Agents Chemother.* **2007,** *51*, 946.
43. Ortega, M. A.; Cogan, D. P.; Mukherjee, S.; Garg, N.; Li, B.; Thibodeaux, G. N.; Maffioli, S. I.; Donadio, S.; Sosio, M.; Escano, J.; Smith, L.; Nair, S. K.; Donk, W. A. v. d., *ACS Chem. Biol.* **2017,** *12*, 548.
44. Hoffmann, T.; Müller, S.; Nadmid, S.; Garcia, R.; Müller, R., *J. Am. Chem. Soc.* **2013,** *135*, 16904.
45. Smith, D. R. M.; Uria, A. R.; Helfrich, E. J. N.; Milbredt, D.; Pée, K.-H. v.; Piel, J. r.; Goss, R. J. M., *ACS Chem. Biol.* **2017,** *12*, 1281.
46. Neumann, C. S.; Walsh, C. T.; Kay, R. R., *Proc. Natl. Acad. Sci. USA* **2010,** *107*, 5798.
47. Binz, T. M.; Maffioli, S. I.; Sosio, M.; Donadio, S.; Müller, R., *J. Biol. Chem.* **2010,** *285*, 32710.
48. (a) Dorrestein, P. C.; Yeh, E.; Garneau-Tsodikova, S.; Kelleher, N. L.; Walsh, C. T., *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 13843; (b) Pang, A. H.; Garneau-Tsodikova, S.; Tsodikov, O. V., *Journal of Structural Biology* **2015,** *192*, 349.
49. El Gamal, A.; Agarwal, V.; Diethelm, S.; Rahman, I.; Schorn, M. A.; Sneed, J. M.; Louie, G. V.; Whalen, K. E.; Mincer, T. J.; Noel, J. P.; Paul, V. J.; Moore, B. S., *Proc. Natl. Acad. Sci. USA* **2016,** *113*, 3797.
50. Nakai, T.; Deguchi, T.; Frébort, I.; Tanizawa, K.; Okajima, T., *Biochemistry* **2014,** *53*, 895.
51. Chacón-Verdú, M. D.; Gómez, D.; Solano, F.; Lucas-Elío, P.; Sánchez-Amat, A., *Appl. Microbiol. Biotechnol.* **2014,** *98*, 2981.
52. See Supporting Information.
53. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W., *Bioinformatics* **2012,** *28*, 3150.
54. Ismail, M.; Frese, M.; Patschkowski, T.; Ortseifen, V.; Niehaus, K.; Sewald, N., *Adv. Synth. Catal.* **2019**.
55. Schmartz, P. C.; Zerbe, K.; Abou-Hadeed, K.; Robinson, J. A., *Org. Biomol. Chem.* **2014,** *12*, 5574.
56. Krastel, P.; Roggo, S.; Schirle, M.; Ross, N. T.; Perruccio, F.; Aspesi Jr., P.; Aust, T.; Buntin, K.; Estoppey, D.; Liechty, B.; Mapa, F.; Memmert, K.; Miller, H.; Pan, X.; Riedl, R.; Thibaut, C.; Thomas, J.; Wagner, T.; Weber, E.; Xie, X.; Schmitt, E. K.; Hoepfner, D., *Angew. Chem. Int. Ed.* **2015,** *54*, 10149.

57. Ashtekar, K.; Marzijarani, N.; Jaganathan, A.; Holmes, D.; Jackson, J. E.; Borhan, B., *J. Am. Chem. Soc.* **2014,** *136*, 13355.
58. Wood, W. J. L.; Patterson, A. W.; Tsuruoka, H.; Jain, R. K.; Ellman, J. A., *J. Am. Chem. Soc.* **2005,** *127*, 15521.
59. (a) Welch, C. J.; Gong, X.; Schafer, W.; Pratt, E. C.; Brkovic, T.; Pirzada, Z.; Cuff, J. F.; Kosjek, B., *Tetrahedron: Asymmetry* **2010,** *21*, 1674; (b) Zawatzky, K.; Barhate, C. L.; Regalado, E. L.; Mann, B. F.; Marshall, N.; Moore, J. C.; Welch, C. J., *J. Chromatogr. A* **2017,** *1499*, 211; (c) Santanilla, A. B.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D., *Science* **2015,** *347*, 49.
60. Wick, J.; Heine, D.; Lackner, G.; Misiek, M.; Tauber, J.; Jagusch, H.; Hertweck, C.; Hoffmeister, D., *Appl. Environ. Microb.* **2016,** *82*, 1196.
61. Durak, L. J.; Payne, J. T.; Lewis, J. C., *ACS Catal.* **2016,** *6*, 1451.
62. Mak, W. S.; Tran, S.; Marcheschi, R.; Bertolani, S.; Thompson, J.; Baker, D.; Liao, J. C.; Siegel, J. B., *Nat. Commun.* **2015,** *6*, 10005.
63. Breitbart, M.; Bonnain, C.; Malki, K.; Sawaya, N. A., *Nature Microbiology* **2018,** *3*, 754.
64. Agarwal, V.; El Gamal, A. A.; Yamanaka, K.; Poth, D.; Kersten, R. D.; Schorn, M.; Allen, E. E.; Moore, B. S., *Nat. Chem. Biol.* **2014,** *10*, 640.
65. Heeb, M. B.; Kristiana, I.; Trogolo, D.; Arey, J. S.; Gunten, U. v., *Water Res.* **2017,** *110*, 91.
66. Shepherd, S. A.; Karthikeyan, C.; Latham, J.; Struck, A.-W.; Thompson, M. L.; Menon, B. R. K.; Styles, M. Q.; Levy, C.; Leys, D.; Micklefield, J., *Chem. Sci.* **2015,** *6*, 3454.
67. Niklasson, M.; Andresen, C.; Helander, S.; Roth, M.; Kahlin, A.; Appell, M.; Mårtensson, L. G.; Lundström, P., *Protein Sci.* **2015,** *24*, 2055.
68. Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H., *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 5869.

Supporting Information for

Site-Selective C-H Halogenation using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling

Brian F. Fisher, Harrison M. Snodgrass, Krysten A. Jones, Mary C. Andorfer, Jared C. Lewis

Table of Contents

I. Materials and Instruments

A) Materials

Greiner Bio-One conical bottom 384-well plates (product number 781281) were purchased from Fisher Scientific International, Inc. (Hampton, NH). Skirted 96-well PCR plates (product number 82006-704) were purchased from VWR International (Radnor, PA). Eppendorf unskirted 96-well PCR plates (product number 951020362) were purchased from Fisher Scientific. Greiner Bio-One polypropylene 96-well V-bottom plates (product number 651201) were purchased from Fisher Scientific. Agilent 0.45 μm PVDF 96-well filter plates (product number 201276-100) were purchased from Agilent. Dialysis tubing (32 mm width; MWCO 6,000-8,000) was purchased from Fisher Scientific.

NAD, FAD, and antibiotics were purchased from Chem-Impex International Inc. (Wood Dale, IL). Antibiotics were prepared as 1000x stock solutions: 1000x chloramphenicol was prepared at 25 mg/mL in EtOH, and 1000x kanamycin was prepared at 50 mg/mL. Substrates were purchased from Sigma-Aldrich, Toronto Research Chemicals, Chem-Impex, or Santa Cruz Biotechnologies. Tryptamine was recrystallized from hot Et_2O before use.

GDH-105 (hereafter, GDH; 50 U/mg) was obtained from Codexis, Inc. (Redwood City, CA). Catalase from bovine liver was obtained from Millipore Sigma (2,000-5,000 U/mg; stock solutions were prepared assuming 2,000 U/mg; product number C9322). The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan). DH5α and BL21(DE3) *E. coli* were purchased from Invitrogen (Carlsbad, CA). Taq DNA polymerase and Phusion HF polymerase were purchased from New England Biolabs (Ipswitch, MA). Luria broth (LB) and Terrific broth (TB) media were purchased from Research Products International (Mt. Prospect, IL). Qiagen Miniprep Kits were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. Protein ladder (Blue Prestained Protein Standard, Broad Range (11-190 kDa); product number P7706) was purchased from New England Biolabs (Ipswitch, MA).

FDH genes were synthesized by the Joint Genome Institute (Department of Energy, USA), integrated into pET28b with a C-terminal 6x His-tag, and transformed into *E. coli* Top10 (Figure S1).

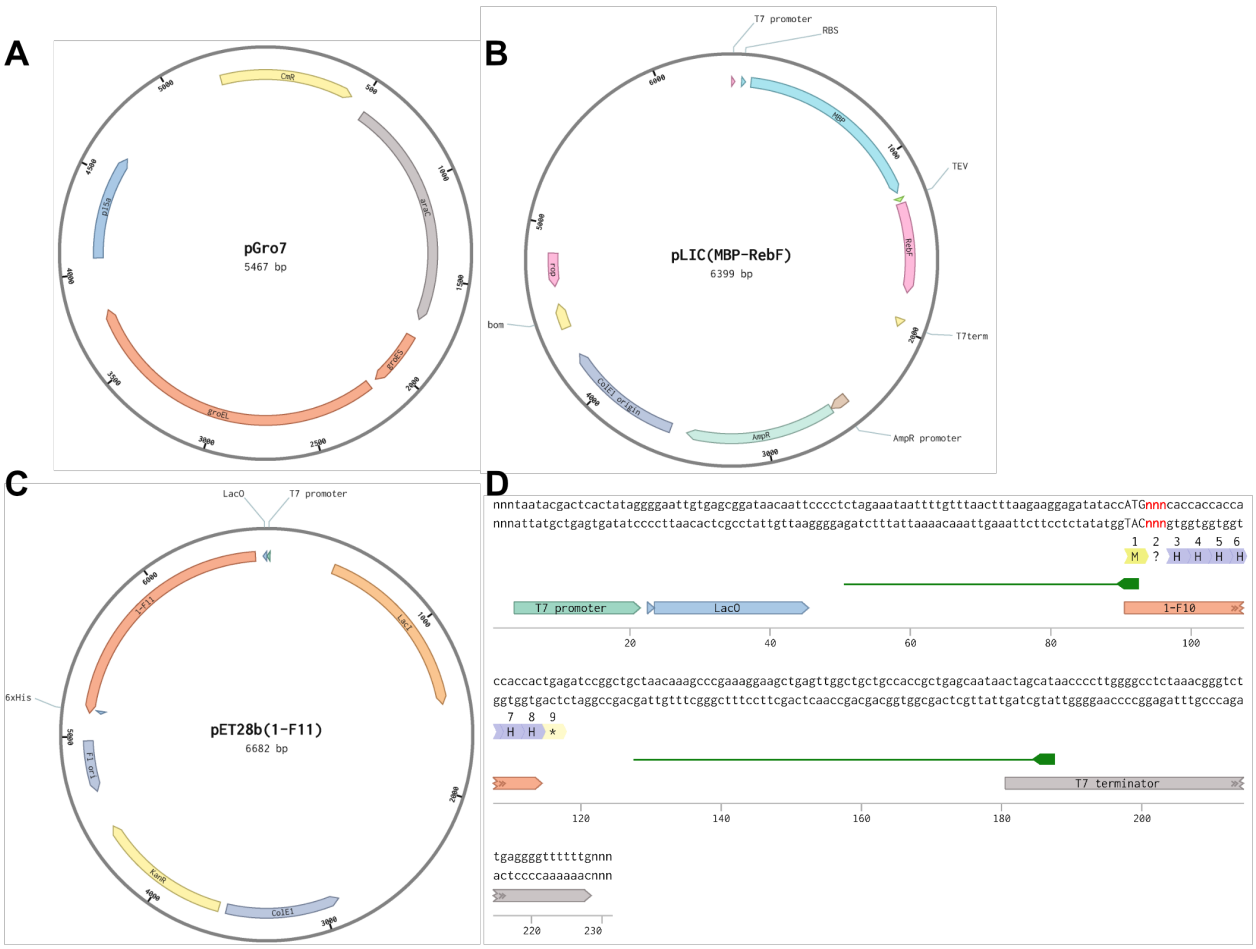


Figure S1. A) Plasmid map of pGro7 chaperone plasmid. B) Plasmid map of pLIC(MBP-RebF). C) Plasmid map of pET28b(1-F11), the expression vector for genome-mined halogenase 1-F11. D) Linear sequence of the insert region into pET28 of the genome-mined halogenases. The red "nnn" segment represents the position of the halogenase gene sequence.

Stock solutions of 10 mM NAD and 10 mM FAD were prepared in 25 mM HEPES pH 7.4 buffer (reaction buffer) and stored at -20 °C until use. Stock solutions of 1.5 M NaCl, 1.5 M NaBr, and 1 M glucose were prepared in reaction buffer and stored at 4 °C until use. Stock solutions of substrate were prepared at 30 mM in either water or DMSO; for the high-throughput screen, substrate stocks were then diluted to 1.67 mM in reaction buffer, manually arrayed into 96-well plates as described later and stored at -20 °C until use. RebF was expressed in *E. coli* BL21(DE3) as an MBP fusion from pLIC-MBP as described previously[1] and stored at 150 µM in reaction buffer with 10% glycerol at -20 °C. GDH was prepared as 180 U/mL stock solution in reaction buffer immediately before reaction setup. For high-throughput screening, stock solutions of 25 µM FDH were stored in reaction buffer with 10% glycerol and arrayed in 96-well unskirted PCR plates as described later. Catalase stock solutions were prepared at 7000 U/mL in reaction buffer immediately before reaction setup.

B) Instruments

Thermal plate sealing with aluminum foil was performed using either a Thermo Scientific ALPS-3000 automated plate sealer, a Packard MicroMate 496 manual plate sealer, or by pressing an oven-heated (≈200 °C) aluminum block onto the plate with foil atop it.

Automated reaction setup was performed using a custom automation setup controlled by Thermo Scientific Momentum software. 384-well reaction plates were loaded into a microplate carousel, and a Thermo Scientific Spinnaker robotic arm controlled plate movement across the deck. A Hamilton Nimbus liquid handler dispensed components of the reaction mixture.

Measurement of DNA/protein concentration was performed using a Tecan Infinite 200 PRO plate reader on a Tecan NanoQuant plate.

High-throughput LC-ESI-MS analysis was performed using an Agilent system equipped with a 1290 Infinity II Multisampler, a 1260 Infinity binary pump, and a 6130 single quadrupole mass spectrometer with an ESI/APPI multimode source.

Analytical-scale reactions were analyzed by LC-MS using an Agilent system equipped with a 1290 Infinity II Multisampler (dual-needle configuration), a 1290 Infinity II high-speed pump, a 1260 Infinity II diode array detector, and a 6135X single quadrupole mass spectrometer with an Agilent Jet Stream ESI source.

Preparative-scale bioconversions were purified using either: 1) a Biotage Isolera One with 12 g SNAP-KP-C18-HS columns and using 0.1% TFA in H_2O as the A solvent and 0.1% TFA in acetonitrile as the B solvent; 2) equipped with a Phenomenex Luna C18(2) semipreparative column (25 cm x 10 mm, 5 μm particle size, 100 Å pore size) or 3) An Agilent 1100 HPLC equipped with a Supelco Discovery C18 semipreparative column (25 cm x 10 mm, 5 μm particle size) and an Agilent 1260 Infinity II fraction collector using 0.1% formic acid in H_2O as the A solvent and 0.1% formic acid in acetonitrile as the B solvent.

Circular dichroism measurements were performed on a Jasco J-715 circular dichroism spectrometer.

C) Software

NMR spectra were processed using MestReNova 11.0.

Plots were generated using GraphPad Prism 7.0, Tableau 2018.3.5, or R.

Halenium affinity calculations were performed using Wavefunction Spartan '18 version 1.2.0 on a 2014 iMac (macOS 10.14.2) equipped with a 4 GHz Intel Core i7 processor, 32 GB 1600 MHz DDR3 RAM, and a 2 GB AMD Radeon R9 M290X GPU.

R code was run within RStudio (1.2.1335) using R version 3.6.0 with all packages updated current to May 2019 on a Windows 10 laptop (Intel Core i7-7500U 2.7 GHz CPU, 8 GB RAM). Example session information (for script used to generate heatmaps):

```
- Session info ---------------------------------------------------------------
-----------------------------------------
 setting  value
 version  R version 3.6.0 (2019-04-26)
 os       Windows 10 x64
 system   x86_64, mingw32
 ui       RStudio
 language (EN)
 collate  English_United States.1252
```

```
 ctype    English_United States.1252
 tz       America/Indianapolis
 date     2019-07-17

- Packages ------------------------------------------------------------------------
-------------------------------------------
 package      * version  date       lib source
 ash            1.0-15   2015-09-01 [1] CRAN (R 3.6.0)
 assertthat     0.2.1    2019-03-21 [1] CRAN (R 3.6.0)
 backports      1.1.4    2019-04-10 [1] CRAN (R 3.6.0)
 base64enc      0.1-3    2015-07-28 [1] CRAN (R 3.6.0)
 beeswarm       0.2.3    2016-04-25 [1] CRAN (R 3.6.0)
 bfish  * 0.1.0    2019-06-17 [1] Github (FishParade/bfish@5e117b3)
 bitops         1.0-6    2013-08-17 [1] CRAN (R 3.6.0)
 broom          0.5.2    2019-04-07 [1] CRAN (R 3.6.0)
 callr          3.2.0    2019-03-15 [1] CRAN (R 3.6.0)
 caTools        1.17.1.2 2019-03-06 [1] CRAN (R 3.6.0)
 cellranger     1.1.0    2016-07-27 [1] CRAN (R 3.6.0)
 cli            1.1.0    2019-03-19 [1] CRAN (R 3.6.0)
 clipr          0.6.0    2019-04-15 [1] CRAN (R 3.6.0)
 cluster        2.0.8    2019-04-05 [2] CRAN (R 3.6.0)
 codetools      0.2-16   2018-12-24 [2] CRAN (R 3.6.0)
 colorspace     1.4-1    2019-03-18 [1] CRAN (R 3.6.0)
 cowplot        0.9.4    2019-01-08 [1] CRAN (R 3.6.0)
 crayon         1.3.4    2017-09-16 [1] CRAN (R 3.6.0)
 data.table     1.12.2   2019-04-07 [1] CRAN (R 3.6.0)
 datapasta      3.0.0    2018-01-24 [1] CRAN (R 3.6.1)
 dendextend   * 1.12.0   2019-05-11 [1] CRAN (R 3.6.0)
 desc           1.2.0    2018-05-01 [1] CRAN (R 3.6.0)
 devtools       2.0.2    2019-04-08 [1] CRAN (R 3.6.0)
 digest         0.6.19   2019-05-20 [1] CRAN (R 3.6.0)
 dplyr        * 0.8.1    2019-05-14 [1] CRAN (R 3.6.0)
 eulerr       * 5.1.0    2019-02-04 [1] CRAN (R 3.6.0)
 evaluate       0.14     2019-05-28 [1] CRAN (R 3.6.0)
 extrafont      0.17     2014-12-08 [1] CRAN (R 3.6.0)
 extrafontdb    1.0      2012-06-11 [1] CRAN (R 3.6.0)
 fansi          0.4.0    2018-10-05 [1] CRAN (R 3.6.0)
 forcats      * 0.4.0    2019-02-17 [1] CRAN (R 3.6.0)
 foreach        1.4.4    2017-12-12 [1] CRAN (R 3.6.0)
 formattable  * 0.2.0.1  2016-08-05 [1] CRAN (R 3.6.0)
 fs             1.3.1    2019-05-06 [1] CRAN (R 3.6.0)
```

```
gclus          1.3.2    2019-01-07 [1] CRAN (R 3.6.0)
gdata          2.18.0   2017-06-06 [1] CRAN (R 3.6.0)
generics       0.0.2    2018-11-29 [1] CRAN (R 3.6.0)
ggalt        * 0.4.0    2017-02-15 [1] CRAN (R 3.6.0)
ggbeeswarm   * 0.6.0    2017-08-07 [1] CRAN (R 3.6.0)
ggplot2      * 3.1.1    2019-04-07 [1] CRAN (R 3.6.0)
ggpubr       * 0.2      2018-11-15 [1] CRAN (R 3.6.0)
ggthemes     * 4.2.0    2019-05-13 [1] CRAN (R 3.6.0)
glue         * 1.3.1    2019-03-12 [1] CRAN (R 3.6.0)
gplots         3.0.1.1  2019-01-27 [1] CRAN (R 3.6.0)
gridBase       0.4-7    2014-02-24 [1] CRAN (R 3.6.0)
gridExtra    * 2.3      2017-09-09 [1] CRAN (R 3.6.0)
gtable         0.3.0    2019-03-25 [1] CRAN (R 3.6.0)
gtools         3.8.1    2018-06-26 [1] CRAN (R 3.6.0)
haven          2.1.0    2019-02-19 [1] CRAN (R 3.6.0)
heatmaply    * 0.16.0   2019-05-11 [1] CRAN (R 3.6.0)
hms            0.4.2    2018-03-10 [1] CRAN (R 3.6.0)
htmltools      0.3.6    2017-04-28 [1] CRAN (R 3.6.0)
htmlwidgets    1.3      2018-09-30 [1] CRAN (R 3.6.0)
httpuv         1.5.1    2019-04-05 [1] CRAN (R 3.6.0)
httr           1.4.0    2018-12-11 [1] CRAN (R 3.6.0)
igraph         1.2.4.1  2019-04-22 [1] CRAN (R 3.6.0)
iterators      1.0.10   2018-07-13 [1] CRAN (R 3.6.0)
janitor      * 1.2.0    2019-04-21 [1] CRAN (R 3.6.0)
jsonlite       1.6      2018-12-07 [1] CRAN (R 3.6.0)
kableExtra   * 1.1.0    2019-03-16 [1] CRAN (R 3.6.0)
KernSmooth     2.23-15  2015-06-29 [2] CRAN (R 3.6.0)
knitr          1.23     2019-05-18 [1] CRAN (R 3.6.0)
labeling       0.3      2014-08-23 [1] CRAN (R 3.6.0)
later          0.8.0    2019-02-11 [1] CRAN (R 3.6.0)
lattice        0.20-38  2018-11-04 [2] CRAN (R 3.6.0)
lazyeval       0.2.2    2019-03-15 [1] CRAN (R 3.6.0)
lubridate      1.7.4    2018-04-11 [1] CRAN (R 3.6.0)
magrittr     * 1.5      2014-11-22 [1] CRAN (R 3.6.0)
maps           3.3.0    2018-04-03 [1] CRAN (R 3.6.0)
MASS           7.3-51.4 2019-03-31 [2] CRAN (R 3.6.0)
memoise        1.1.0    2017-04-21 [1] CRAN (R 3.6.0)
mime           0.7      2019-06-11 [1] CRAN (R 3.6.0)
modelr         0.1.4    2019-02-18 [1] CRAN (R 3.6.0)
munsell        0.5.0    2018-06-12 [1] CRAN (R 3.6.0)
```

```
nlme           3.1-139  2019-04-09 [2] CRAN (R 3.6.0)
pillar         1.4.1    2019-05-28 [1] CRAN (R 3.6.0)
pkgbuild       1.0.3    2019-03-20 [1] CRAN (R 3.6.0)
pkgconfig      2.0.2    2018-08-16 [1] CRAN (R 3.6.0)
pkgload        1.0.2    2018-10-29 [1] CRAN (R 3.6.0)
plotly       * 4.9.0    2019-04-10 [1] CRAN (R 3.6.0)
plyr           1.8.4    2016-06-08 [1] CRAN (R 3.6.0)
prettyunits    1.0.2    2015-07-13 [1] CRAN (R 3.6.0)
processx       3.3.1    2019-05-08 [1] CRAN (R 3.6.0)
proj4          1.0-8    2012-08-05 [1] CRAN (R 3.6.0)
promises       1.0.1    2018-04-13 [1] CRAN (R 3.6.0)
ps             1.3.0    2018-12-21 [1] CRAN (R 3.6.0)
purrr        * 0.3.2    2019-03-15 [1] CRAN (R 3.6.0)
R6             2.4.0    2019-02-14 [1] CRAN (R 3.6.0)
RColorBrewer   1.1-2    2014-12-07 [1] CRAN (R 3.6.0)
Rcpp           1.0.1    2019-03-17 [1] CRAN (R 3.6.0)
readr        * 1.3.1    2018-12-21 [1] CRAN (R 3.6.0)
readxl       * 1.3.1    2019-03-13 [1] CRAN (R 3.6.0)
registry       0.5-1    2019-03-05 [1] CRAN (R 3.6.0)
remotes        2.0.4    2019-04-10 [1] CRAN (R 3.6.0)
reshape2       1.4.3    2017-12-11 [1] CRAN (R 3.6.0)
rlang          0.3.4    2019-04-07 [1] CRAN (R 3.6.0)
rmarkdown      1.13     2019-05-22 [1] CRAN (R 3.6.0)
rprojroot      1.3-2    2018-01-03 [1] CRAN (R 3.6.0)
rstudioapi     0.10     2019-03-19 [1] CRAN (R 3.6.0)
Rttf2pt1       1.3.7    2018-06-29 [1] CRAN (R 3.6.0)
rvest          0.3.4    2019-05-15 [1] CRAN (R 3.6.0)
scales       * 1.0.0    2018-08-09 [1] CRAN (R 3.6.0)
seriation      1.2-7    2019-06-08 [1] CRAN (R 3.6.0)
sessioninfo    1.1.1    2018-11-05 [1] CRAN (R 3.6.0)
shiny          1.3.2    2019-04-22 [1] CRAN (R 3.6.0)
snakecase      0.11.0   2019-05-25 [1] CRAN (R 3.6.0)
stringi        1.4.3    2019-03-12 [1] CRAN (R 3.6.0)
stringr      * 1.4.0    2019-02-10 [1] CRAN (R 3.6.0)
tibble       * 2.1.1    2019-03-16 [1] CRAN (R 3.6.0)
tidyr        * 0.8.3    2019-03-01 [1] CRAN (R 3.6.0)
tidyselect     0.2.5    2018-10-11 [1] CRAN (R 3.6.0)
tidyverse    * 1.2.1    2017-11-14 [1] CRAN (R 3.6.0)
treemap      * 2.4-2    2017-01-04 [1] CRAN (R 3.6.0)
TSP            1.1-7    2019-05-22 [1] CRAN (R 3.6.0)
```

```
usethis        1.5.0   2019-04-07 [1] CRAN (R 3.6.0)
utf8           1.1.4   2018-05-24 [1] CRAN (R 3.6.0)
vctrs          0.1.0   2018-11-29 [1] CRAN (R 3.6.0)
vipor          0.4.5   2017-03-22 [1] CRAN (R 3.6.0)
viridis      * 0.5.1   2018-03-29 [1] CRAN (R 3.6.0)
viridisLite  * 0.3.0   2018-02-01 [1] CRAN (R 3.6.0)
webshot        0.5.1   2018-09-28 [1] CRAN (R 3.6.0)
withr          2.1.2   2018-03-15 [1] CRAN (R 3.6.0)
xfun           0.7     2019-05-14 [1] CRAN (R 3.6.0)
xml2           1.2.0   2018-01-24 [1] CRAN (R 3.6.0)
xtable         1.8-4   2019-04-21 [1] CRAN (R 3.6.0)
yaml           2.2.0   2018-07-25 [1] CRAN (R 3.6.0)
zeallot        0.1.0   2018-01-28 [1] CRAN (R 3.6.0)
```

II. Bioinformatics

A) Sequence Lists

Table S1. Flavin-dependent halogenases found in the SSN with structures deposited in the PDB.

FDH ID	PDB	Uniprot_ID
BrvH	6FRL	B4WBL8_9CAUL
CmlS	3I3L	Q9AL91_STRVZ
MalA	5WGR|5WGS|5WGT|5WGU|5WGV|5WGW|5WGX|5WGY|5WGZ	L0E155_9EURO
MibH	5UAO	W2EQU4_9ACTN
Mpy16	5BUK	J7H1A1_9ACTN
PltA	5DBJ	Q4KCZ0_PSEF5
PltM	6BZA|6BZI|6BZN|6BZQ|6BZT|6BZZ	Q4KCZ3_PSEF5
PrnA	2APG|2AQJ|2AR8|2ARD|2JKC|4Z43|4Z44	PRNA_PSEFL
PyrH	2WES|2WET|2WEU	A4D0H5_9ACTN
RebH	2E4G|2O9Z|2OA1|2OAL|2OAM|4LU6	REBH_NOCAE
S.frigidimarina-FDH	2PYX	Q085A0_SHEFN
SttH	5HY5	E9P162_9ACTN

Th-Hal	5LV9	A0A1L1QK36

A) Sequence Alignments

Pairwise sequence alignments were performed in R using the pairwiseAlignment function of the Biostrings package (available on Bioconductor[2]) with BLOSUM62 as the substitution matrix. Percent identity calculations were performed using the pid function in the Biostrings package with the %ID algorithm set to "PID1" (**Error! Reference source not found.**).

$$\%ID = 100 \times \frac{identical\ positions}{aligned\ positions + gap\ positions} \qquad \textbf{Equation S1.}$$

Multiple sequence alignments were performed using MUSCLE implemented in Geneious 4.8.5.

B) Sequence Similarity Network

Generation of SSN

Initial sequence analysis was conducted using the Enzyme Similarity Tool (EST) developed by the Enzyme Function Initiative.[3] The single sequence input option in the EST was used to conduct a BLAST search of the UniProt database with an expect value of 10^{-5} using the RebH sequence as a query. This search (conducted on May 10, 2016) returned 4380 sequences. These sequences, along with metadata from the UniProt database, were processed further in Excel. Duplicate sequences, sequences shorter than 285 residues, and sequences longer than 1038 residues were removed from this set, and Th-Hal, KrmI, and the C-terminal segment of KrmI were added manually to obtain the 3975 sequences used by the EST to generate a sequence similarity network (SSN) using of 70 (≈30% sequence identity). This Level 1 SSN was used to create a Level 2 SSN with a more stringent edge detection threshold at an alignment score of 140 (≈40% sequence identity). The Level 1 SSN contained 79 subnetworks, 22 of which had more than ten sequences, and the ten largest subnetworks contained between 30 and 2288 sequences. There were 90 singleton sequences. The Level 2 SSN contained 152 subnetworks, 35 of which had ten or more sequences, and the ten largest subnetworks contained between 63 and 1312 sequences. There were 297 singletons. Both SSNs were visualized in Cytoscape 3.6.1 using a 50% identity cutoff (computed by CD-HIT[4]) for grouping sequences into representative nodes. Node layouts were generated using the yFiles Organic algorithm.

A manually curated database of flavin-dependent halogenases (129 in total) was compared against the RebH BLAST hits. Of these, we found 14 that were not found in the RebH BLAST search. Of these 14, seven were not found because they were deposited in the UniProtKB after the date of the initial BLAST search: AcOTAHal, ArmH5, Bmp2 (the variant studied structurally by Moore and coworkers), Bmp5, ctcP, KrmI, and Th-Hal. The remaining seven were in the UniProtKB at the date of the RebH BLAST search but were not returned as hits: Ram20 (NRPS-tethered substrate), McnD, HrmQ, HalB (small molecule pyrrole substrate), Mpy10, Mpy11, ChlB4, ApdC, and AerJ.

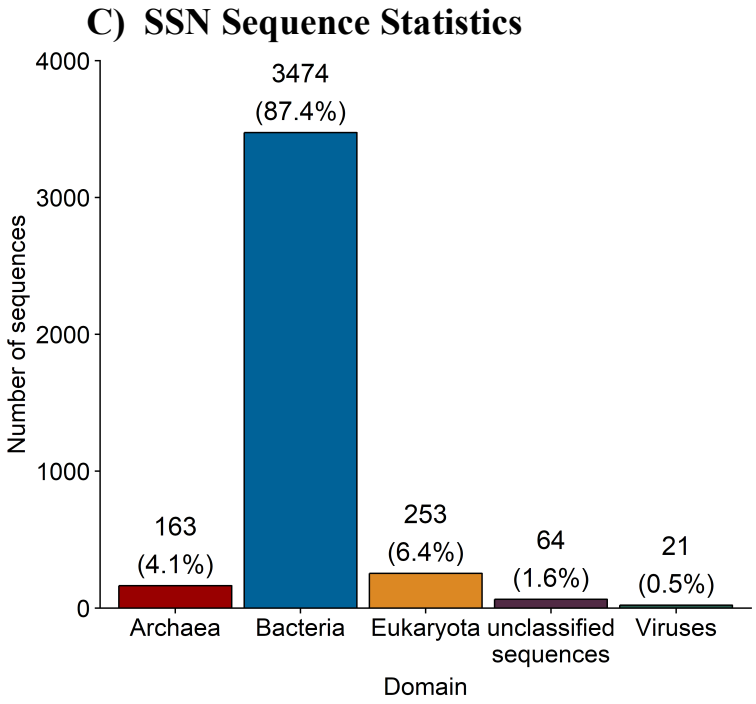


Figure S2. Distribution of host domains among sequences in the FDH sequence similarity network. Numbers above bars are the number of sequences and the percentage of the SSN in parentheses.

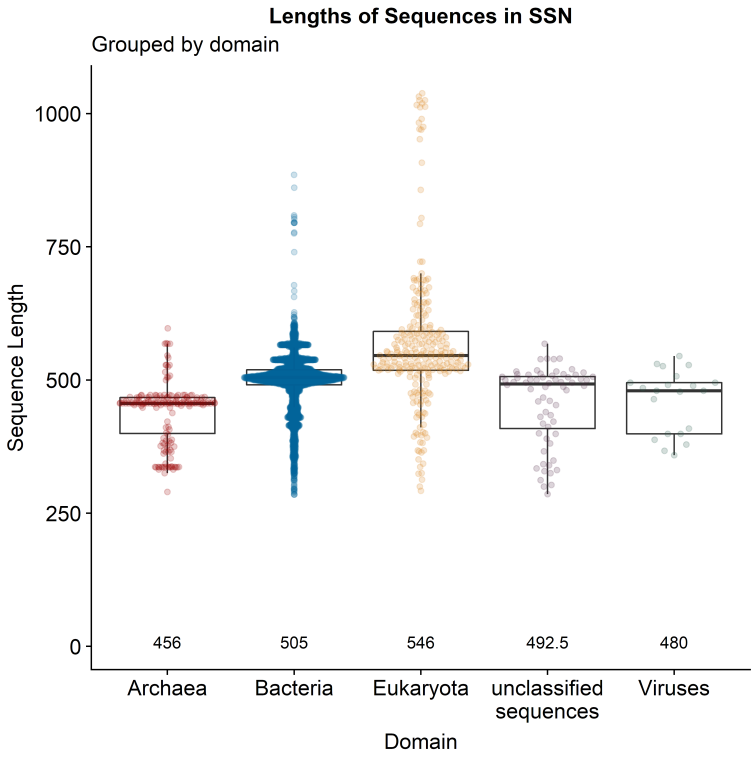


Figure S3. Distribution of sequence lengths among sequences in the FDH sequence similarity network, separated by host domain. Number at bottom of plot is the mean sequence length.

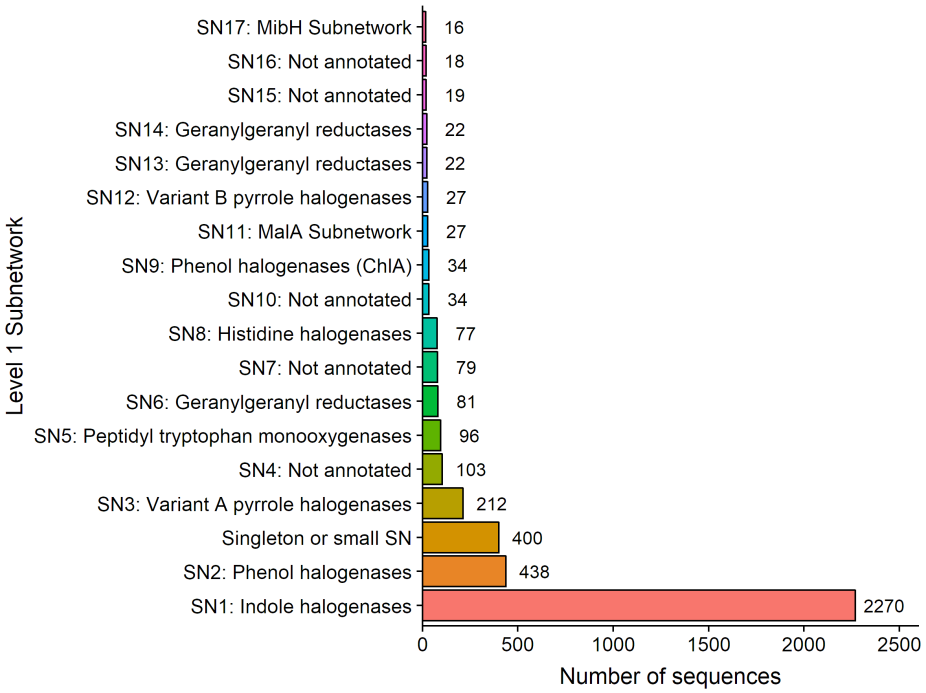


Figure S4. Distribution of sequences in different Level 1 subnetworks of the SSN. Only subnetworks with >15 members are shown.

Level 1 SSNs

Annotated by genome-mined vs. previously characterized

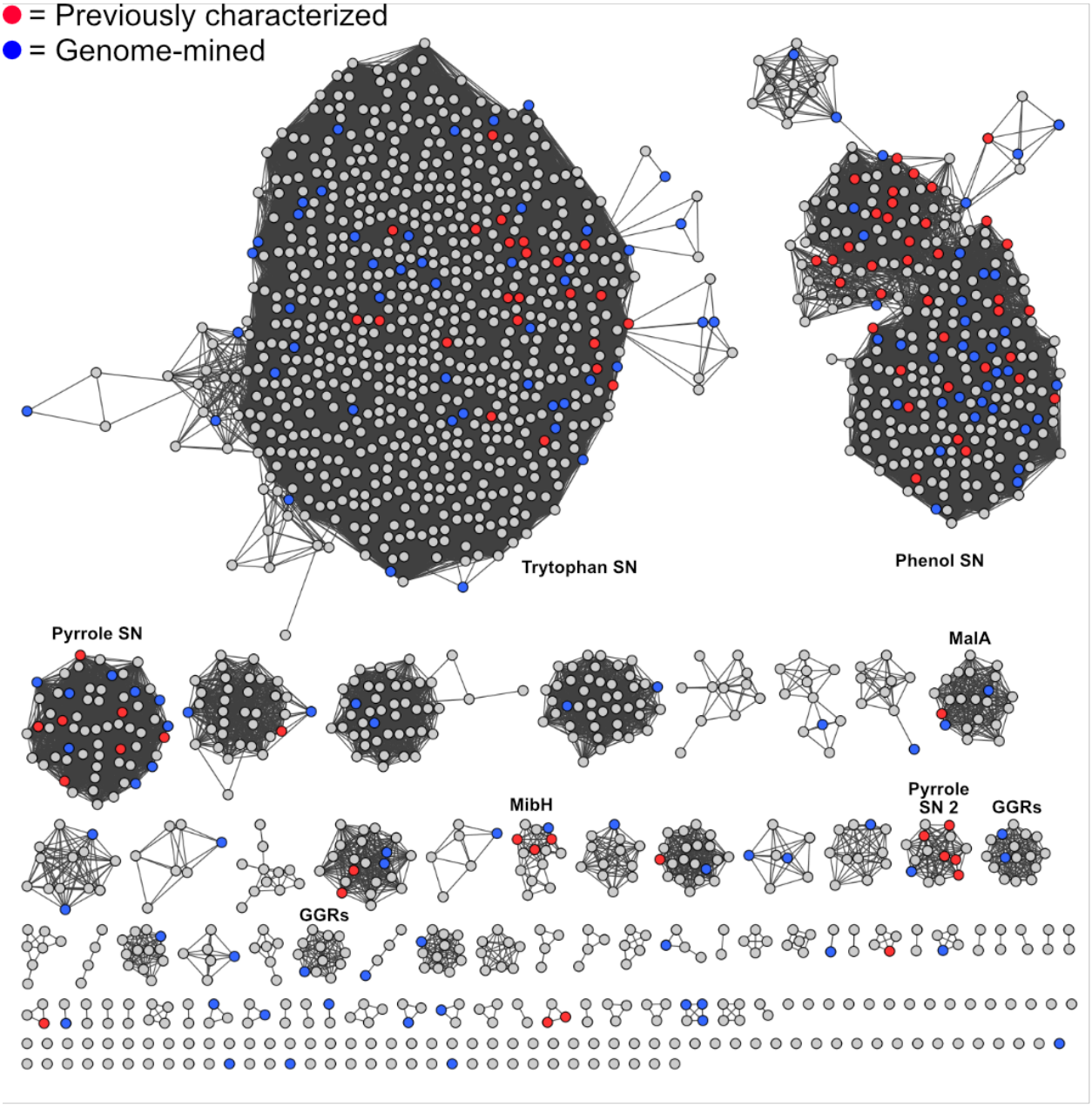


Figure S5. Level 1 SSN annotated with sequences that were characterized in this study in blue and FDHs that have been previously characterized in red. SSN drawn as 50% representative node network in which sequences sharing ≥ 50% sequence similarity as determined by CD-HIT implemented in EFI-EST. Only one node, in the phenol subnetwork, includes more than one genome-mined sequence (FDHs 1-B04 and 1-G04).

Annotated by solubility

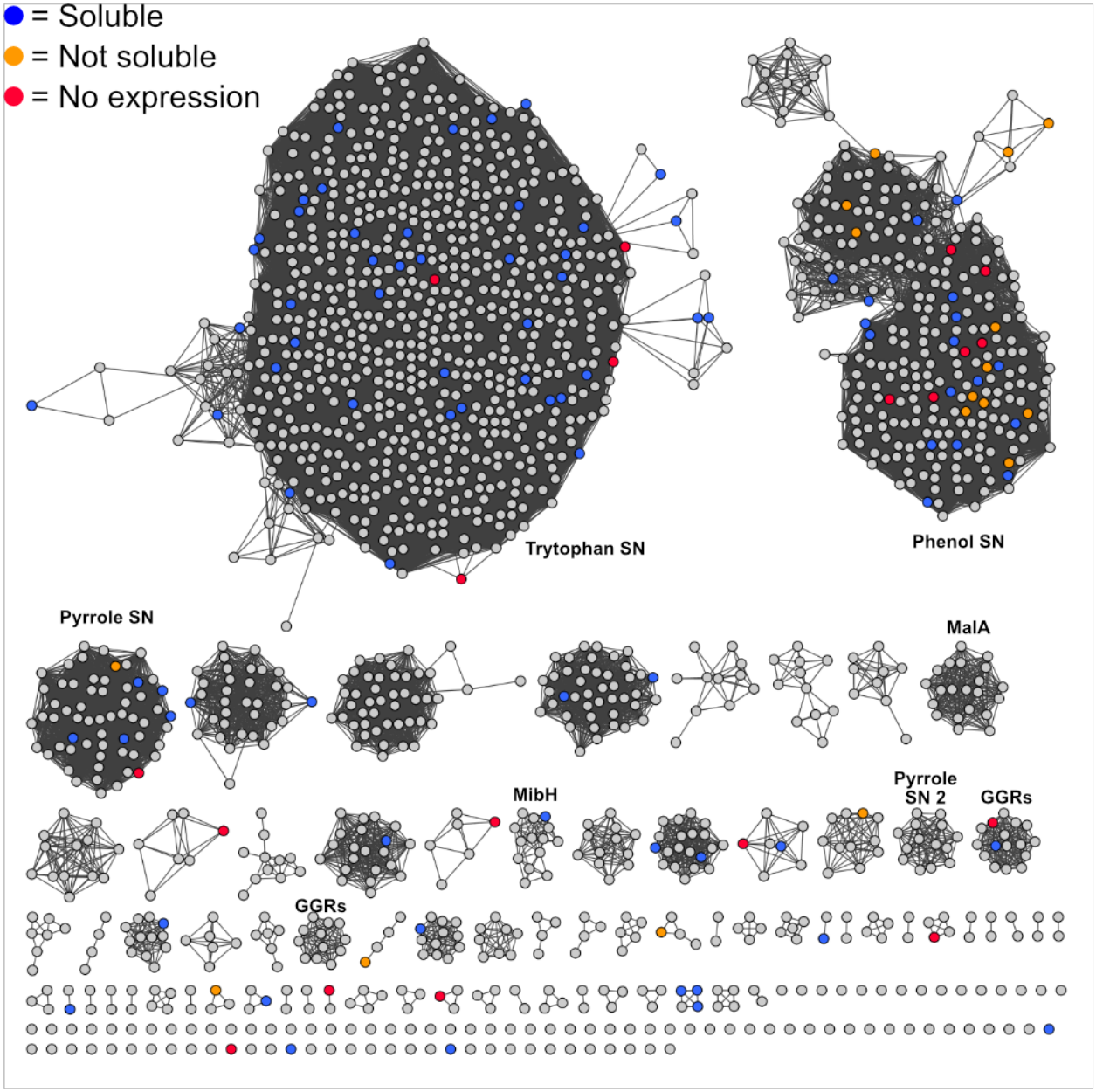


Figure S6. Level 1 SSN annotated with FDH solubility. SSN drawn as 50% representative node network in which sequences sharing ≥ 50% sequence similarity as determined by CD-HIT implemented in EFI-EST. Nodes colored blue contain a protein that was determined to be soluble from small-scale expression tests; nodes colored orange contain a protein that was determined to not be solubly expressed; nodes colored red contain a sequence for which expression could not be distinguished using SDS-PAGE. Only one node, in the phenol subnetwork, includes more than one genome-mined sequence (FDHs 1-B04 and 1-G04); both were solubly expressed.

Level 2 SSNs

Annotated by PDB

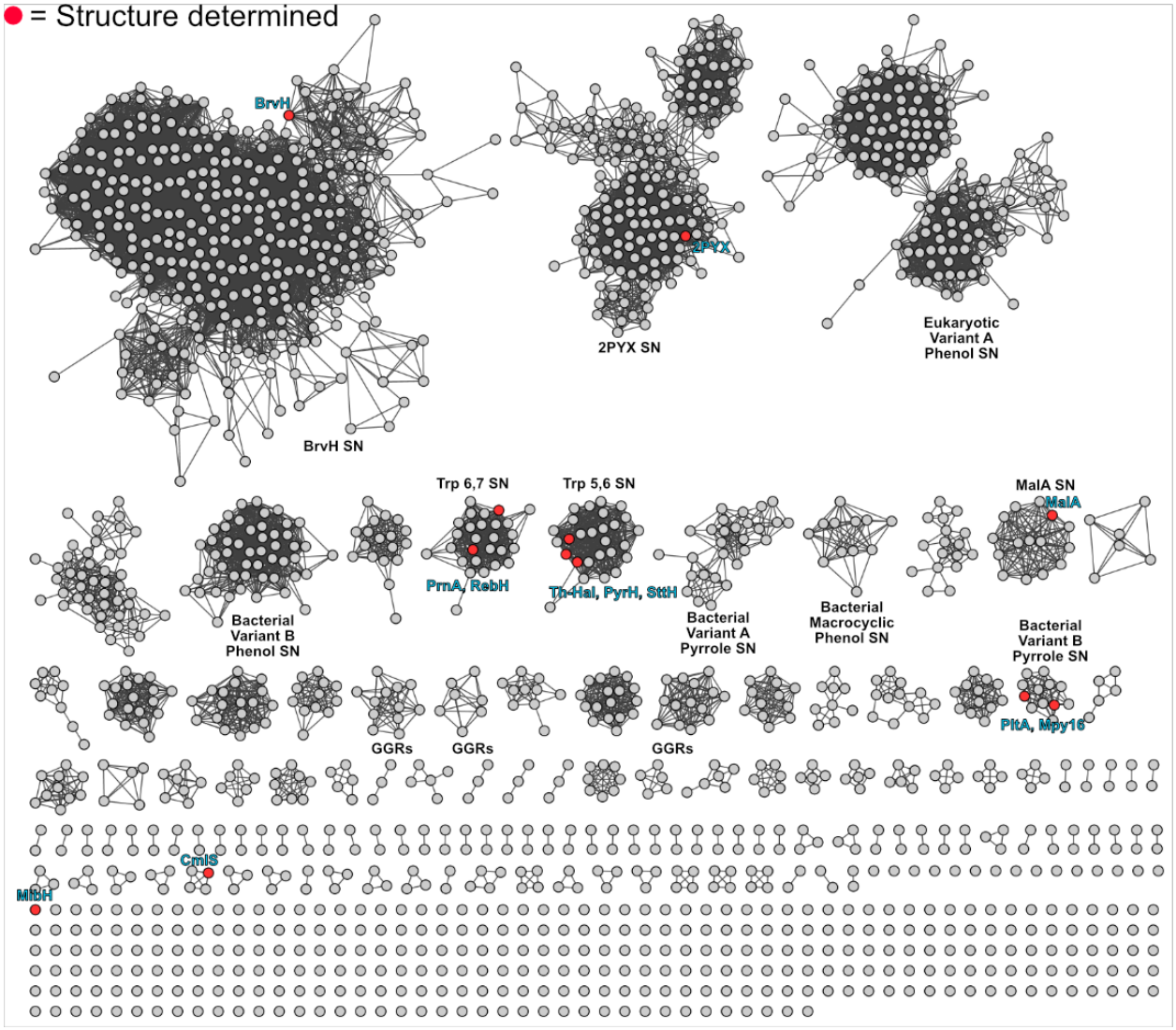


Figure S7. Level 2 SSN annotated with PDB-characterized FDHs. SSN drawn as 50% representative node network in which sequences sharing ≥ 50% sequence similarity as determined by CD-HIT implemented in EFI-EST. Nodes colored red have at least one sequence with a structure reported in the PDB.

III. FDH Expression

A) Cloning

Competent cells

Chemically-competent BL21(DE3) *E. coli* (Invitrogen) transformed with pGro7 were used for generation of the expression strains. A 5 mL culture of BL21(DE3) + pGro7 was grown in LB medium supplemented with chloramphenicol overnight at 37 °C shaking at 250 rpm. The entire overnight culture was used to inoculate 500 mL LB containing chloramphenicol, which was incubated at 37 °C at 250 rpm. Once the

culture reached OD_{600} of ≈0.4, the culture was centrifuged at 3600 rpm at 4 °C for 10 minutes, and the supernatant was discarded. The cell pellet was resuspended in 20 mL cold Ca^{2+}/glycerol buffer (60 mM $CaCl_2$, 10 mM HEPES, 15% glycerol, pH 7.0, sterile filtered) by gentle swirling. The resuspended pellet was again centrifuged at 3600 rpm at 4 °C for 10 minutes, and the supernatant was discarded. The cell pellet was again resuspended in 20 mL cold Ca^{2+}/glycerol buffer by gentle swirling and centrifuged at 3600 rpm at 4 °C for 10 minutes. The supernatant was discarded, and the cell pellet was resuspended in 6 mL cold Ca^{2+}/glycerol buffer. The cell suspension was dispensed in 100 µL aliquots into chilled PCR tubes, snap frozen in liquid N_2, and stored at -80 °C until transformation.

Transformation

Chemically-competent BL21(DE3) *E. coli* cells containing the pGro7 chaperone plasmid were transformed with pET28b containing FDH insert. Aliquots of competent cells were transferred to 5 mL polypropylene culture tubes on ice, to which approximately 10 ng of plasmid was added. Competent cells were incubated with plasmid on ice for 30 min, and the cells were heat shocked at 42 °C in water bath for 45 seconds. The culture tubes were afterward incubated on ice for 2 minutes, then 350 µL SOC medium was added. The culture tubes were transferred to an incubator at 37 °C shaking at 250 rpm for 1 hr to recover. After recovery, 100 µL cells were added to agar plates (with 50 µg/mL kanamycin and 25 µg/mL chloramphenicol) and spread using 3 mm glass beads. After drying, the agar plates were transferred to a 37 °C incubator and grown overnight. Single colonies were picked and grown in 300 µL LB in a 1 mL deep well 96-well plate with kanamycin and chloramphenicol overnight. To each well of this 96-well plate was then added 200 µL autoclaved 50% glycerol, and this glycerol stock plate was kept at -80 °C.

B) Protein Expression

Small-scale FDH expression

96-well plates containing glycerol stocks of BL21(DE3) + pGro7 FDH expression cultures were stamped into autoclaved 2 mL 96-well deep well plates containing 1 mL LB with kanamycin and chloramphenicol. These inoculated plates were sealed with an AeraSeal adhesive film and incubated overnight at 37 °C, 235 rpm overnight. The overnight cultures (20 µL) were used to inoculate 2 mL antibiotic-containing LB media in 6 mL 24-well deep well plates. Inoculated expression cultures were sealed with an AeraSeal adhesive film and incubated at 37 °C, 235 rpm until OD_{600} ≈0.6-0.8, at which point the incubator was cooled to 15 °C. After the cultures were cooled sufficiently (≈15 min) protein expression was induced with 2 mg/mL L-arabinose and 10 µM IPTG. Protein expression proceeded for 20 hr, at which point the cells were pelleted by centrifugation at 3600 rpm, 4 °C for 15 min. The supernatant was discarded, and the cell pellets were resuspended in 200 µl lysis buffer (0.75 mg/mL lysozyme, 25 mM HEPES, pH 7.4). Cells were incubated in lysis buffer at 37 °C, 250 rpm for 30 min. After lysis, the suspensions were frozen by immersing the 24-well plates in liquid N_2, thawed at room temperature for 15 min, then transferred to a warm water bath. Once thawing was complete, 20 µL DNAse buffer (1 mg/mL DNAse, 25 mM HEPES, pH 7.4) was added, and the 24-well plates were incubated at 37 °C, 250 rpm for 15 min. The insoluble fraction of the cell lysate was pelleted by centrifugation at 3600 rpm, 4 °C for 15 min. The supernatant was isolated, and the insoluble fraction of the cell lysate was washed by resuspending in 300 µL 25 mM HEPES, pH 7.4, then centrifugation at 3600 rpm, 4 °C for 15 min. This supernatant wash was discarded.

Medium-scale FDH expression

14 mL culture tubes containing 5 mL LB with kanamycin and chloramphenicol were inoculated with a glycerol stock of BL21(DE3) + pGro7 + pET28b(FDH) and incubated overnight at 37 °C, 250 rpm. The

next day, 50 mL TB with antibiotics in a 250 mL Erlenmeyer was inoculated with 500 µL of the overnight cultures. The inoculated expression cultures were incubated at 37 °C, 250 rpm until OD_{600} ≈0.6-0.8, at which point the incubator was cooled to 15 °C. Once the liquid cultures were cool (about 15 min), protein expression was induced with 2 mg/mL L-arabinose and 100 µM IPTG, and the expression cultures were incubated for 20 hr. Once protein expression was complete, the expression cultures were transferred to 50 mL centrifuge tubes and centrifuged at 3600 rpm, 4 °C for 15 min. The supernatant media was discarded, and the cell pellets were resuspended in 10 mL 25 mM HEPES, pH 7.4. Cell suspensions were sonicated on ice using a QSonica S-4000 with a 0.5" horn at 40W using 1 min on/1 min off cycles for 5 min total cycle time. Cell lysates were clarified by centrifuging at 15,000 rpm in a high-speed fixed-angle rotor for 40 min at 4 °C. The soluble fraction of the lysate was decanted into a new 50 mL centrifuge tube, then transferred to 10 mL polypropylene frit-bottomed spin columns capped at the bottom and containing 500 µL Ni-NTA resin pre-equilibrated with equilibration buffer (20 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4). The columns were capped on top, inverted a few times to mix evenly, and transferred back to the centrifuge tubes which were then capped. Protein was bound to resin by gentle mechanical inversion of these centrifuge tubes for 1 hr at 4 °C. After the binding step, the Ni-NTA suspensions were transferred back to uncapped spin columns and allowed to drain by gravity into a waste basin. 5 mL of wash buffer (20 mM phosphate, 300 mM NaCl, 25 mM imidazole, pH 7.4) was added to the columns, which were allowed to drain by gravity into a waste basin. The spin columns were nested within new 50 mL centrifuge tubes, and 5 mL elution buffer (20 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4) was added and allowed to drain into the centrifuge tubes by gravity. The eluted protein solutions were transferred to 4 mL Amicon Ultra 30K MWCO spin filters and concentrated to ≈500 µL by centrifugation at 4000G for ≈15 min at 4 °C. Protein solution was diluted with 25 mM HEPES, pH 7.4, and centrifuged again. Buffer exchange in this manner was performed 3-5 times, after which glycerol was added for a final concentration of 10% v/v. Protein solutions were centrifuged at 13.2 krpm at 4 °C for 3 min prior to measuring concentration[5] using absorbance at 280 nm using a Tecan NanoQuant plate with protein extinction coefficients using ProtCalc v3.4 (http://protcalc.sourceforge.net/), and these data were used to compute protein titers in mg/L (Figure S8).

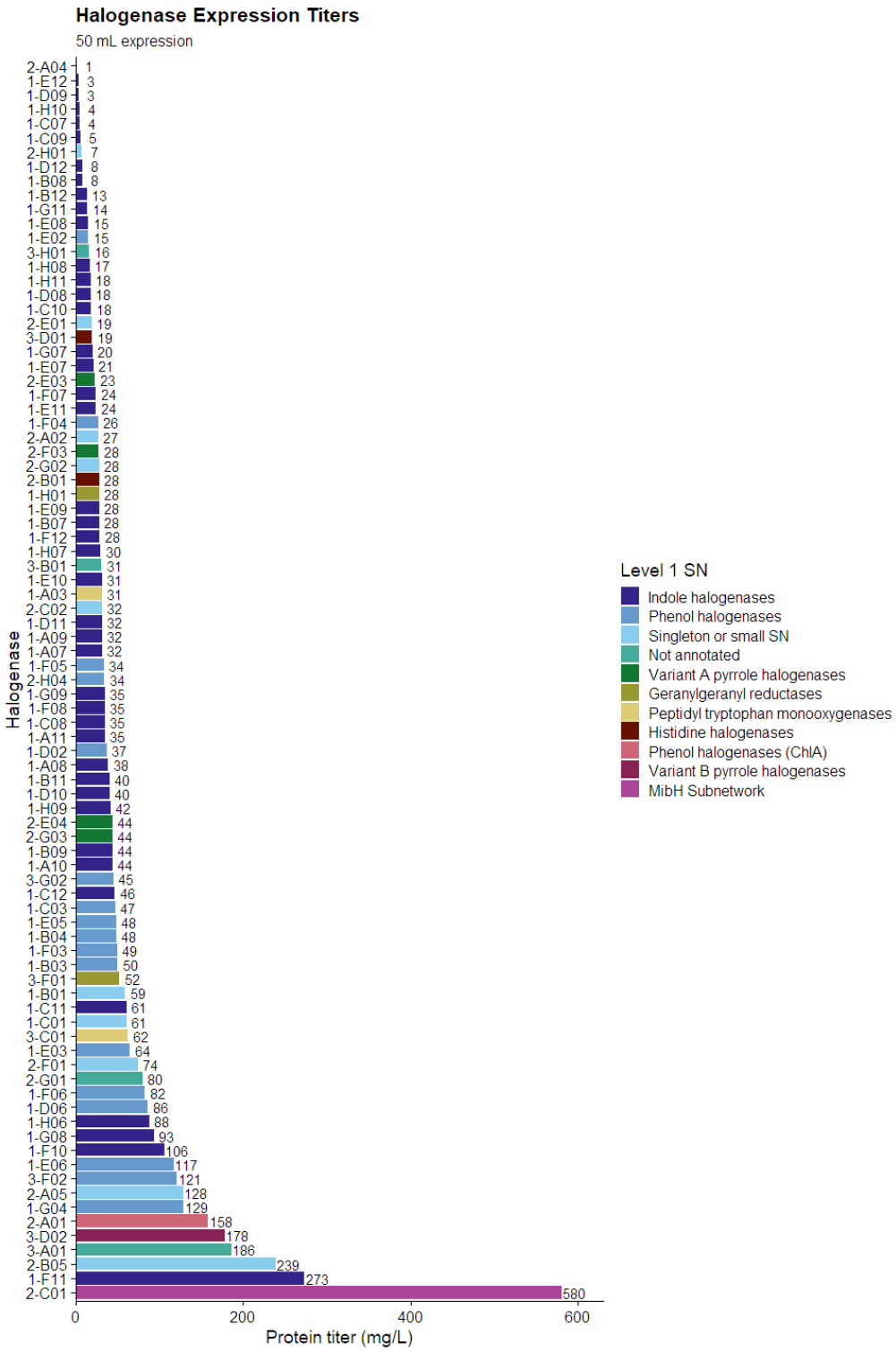


Figure S8. Protein expression titers determined by A_{280} from the 50 mL expressions of genome-mined FDHs.

Large-scale FDH expression

14 mL culture tubes containing 5 mL LB with kanamycin and chloramphenicol were inoculated with a glycerol stock of BL21(DE3) + pGro7 + pET28b(FDH) and incubated overnight at 37 °C, 250 rpm. The next day, 750 mL TB with antibiotics in a 2.8 L Fernbach flask was inoculated with the entire overnight culture. The inoculated expression cultures were incubated at 37 °C, 250 rpm until OD_{600} ≈0.6-0.8, at which point the incubator was cooled to 15 °C. Once the liquid cultures were cool (about 15 min), protein expression was induced with 2 mg/mL L-arabinose and 100 μM IPTG, and the expression cultures were incubated for 20 hr. Once protein expression was complete, the expression cultures were transferred to 750 mL centrifuge bottles and centrifuged at 3600 rpm, 4 °C for 15 min. The supernatant media was discarded, and the cell pellets were resuspended in 30 mL 25 mM HEPES, pH 7.4. Cell suspensions were sonicated on ice using a QSonica S-4000 with a 0.5" horn at 40W using 1 min on/1 min off cycles for 5 min total cycle time. Cell lysates were clarified by centrifuging at 15,000 rpm in a high-speed fixed-angle rotor for 40 min at 4 °C. The soluble fraction of the lysate was decanted into a new 50 mL centrifuge tube, then transferred to 10 mL polypropylene frit-bottomed spin columns capped at the bottom and containing 3 mL Ni-NTA resin pre-equilibrated with equilibration buffer (20 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4). The columns were capped on top, inverted a few times to mix evenly, and transferred back to the centrifuge tubes which were then capped. Protein was bound to resin by gentle mechanical inversion of these centrifuge tubes for 1 hr at 4 °C. After the binding step, the Ni-NTA suspensions were transferred back to uncapped spin columns and allowed to drain by gravity into a waste basin. 5 mL of wash buffer (20 mM phosphate, 300 mM NaCl, 25 mM imidazole, pH 7.4) was added to the columns, which were allowed to drain by gravity into a waste basin. The spin columns were nested within new 50 mL centrifuge tubes, and 5 mL elution buffer (20 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4) was added and allowed to drain into the centrifuge tubes by gravity. The eluted protein solutions were transferred to 15 mL Amicon spin filters Ultra 30K MWCO spin filters and either buffer exchanged as described previously or concentrated prior to overnight dialysis at 4 °C against 25 mM HEPES, pH 7.4.

SDS-PAGE

SDS-PAGE gels, 1 mm thick, were cast with 10% or 12% acrylamide using standard procedures,[6] but including 0.5% trichloroethanol in the separating portion of the gel to enable fluorescent detection of Trp-containing proteins.[7] 8 μL of the soluble or insoluble fraction of the cell lysate, 16 μL water, and 12 μL 4x SDS-PAGE loading buffer (200 mM Tris-Cl (pH 6.8), 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol) were combined in Eppendorf tubes, which were heated at 95 °C in a dry incubator for 10 min. Samples were cooled to room temperature, then centrifuged for a few seconds in a mini centrifuge to collect liquid at the bottom. SDS-PAGE gels were loaded into the gel chamber, and 8 μL of each sample was loaded onto the gels. SDS-PAGE gels were run at 130 V for 90 minutes, after which they were removed from the chambers and glass slides, washed three times with deionized water, and placed onto the UV stage of a gel imager. Gels were irradiated at 302 nm for 2 min to develop the photochemically modified, fluorescent Trp residues, and images were taken. Serendipitously, the chaperone groEL, one of the chaperones expressed by pGro7 that unfortunately has a molecular weight close to that of many halogenases, has zero Trp residues, and is not stained by this procedure, allowing for easy identification of the expressed FDHs. Gels were then stained using Coomassie Brilliant Blue R-250 staining solution[8] for 30 min at 50 °C with gentle shaking at 60 rpm. Staining solution was returned to its stock contained, and gels were washed three times with deionized water. Destaining solution[9] was added to the gels, which were then incubated at 50 °C for 30 min with gentle shaking at 60 rpm. Destaining solution was discarded, gels were washed three times with deionized water, and images of the gels were acquired using a gel imager.

IV. High-Throughput LC-MS Bioconversion Screen

A) Reaction Setup

FDH preparation

Stock solutions of each purified FDH (25 μM in reaction buffer with 10% glycerol) were manually arrayed into three adjacent wells (i.e. A1, A2, A3, etc.) of a 96-well non-skirted PCR plate such that arraying this 96-well plate into the four quadrants of a 384-well plate would dispense each FDH into 12 wells of the 384-well plate, each containing a different substrate.

Substrate preparation

Stock solutions of substrate (1.67 mM in reaction buffer) were manually arrayed into every third column of a 96-well PCR plate, such that each plate contained three substrates. For each round of screening, four substrate plates were prepared in this manner, for a total of 12 substrates. Aliquots of 3 μL from each substrate plate were arrayed using a Hamilton Nimbus fitted with a 96-channel pipetting head into a different quadrant of a 384-well plate, forming blocks of 2 x 6 blocks (rows x columns) containing all 12 different substrates. Substrate-containing 384-well plates were prepared in advance, sealed with adhesive film, and stored at -20 °C until reaction setup.

Reaction setup

	Chlorinations	**Brominations**
NaX	100 mM	10 mM
glucose	20 mM	20 mM
NAD	100 μM	100 μM
FAD	100 μM	100 μM
GDH	9 U/mL	9 U/mL
MBP-RebF	2.5 μM	2.5 μM
FDH	5 μM	5 μM
substrate	500 μM	500 μM
Rxn Volume	10 μL	10 μL

A mixture of the small molecule components of the FDH bioconversions (NaX where X=whichever halide is being tested, glucose, NAD, and FAD) was prepared in reaction buffer such that 3 μL of the small molecule mix in 10 μL reactions would yield the working concentration of each component (see above). This small molecule mix was manually arrayed into 96-well skirted PCR plates and placed at the proper location on the Nimbus automation deck.

A mixture of the cofactor regen components of the FDH bioconversions (GDH, MBP-RebF) was prepared in reaction buffer such that 2 μL of the cofactor regen mix in 10 μL reactions would yield the working concentration of each component (see above). The cofactor regen mix was manually arrayed into 96-well skirted PCR plates and placed at the proper location on the Nimbus automation deck.

Six substrate-containing 384-well plates were removed from a -20 °C freezer and centrifuged at 2000 rpm at room temperature for 2 min. The adhesive plate seals were removed, and the 384-well plates were placed into the carousel of our lab's automation system. House-fashioned cooling blocks for the three FDH stock plates had been made by cutting out the gel from a freezer pack, squeezing it into the space at the bottom of a skirted 96-well PCR plate until full, wrapping the PCR plate in aluminum foil, and leaving the plate at -20 °C overnight. The homemade FDH cooling blocks were placed at the proper position on the Nimbus deck, and FDH stock plates were nested into the cooling blocks. The cooling blocks were able to keep the FDH stock solutions cool over the duration of the long (40 min to 2.5 hr) automation sequences.

A Thermo Scientific Spinnaker robot was used to move two substrate-containing 384-well plates onto the Nimbus deck. Hamilton 10 µL tips were used to transfer 3 µL NaCl small molecule mix to each well of the first 384-well plate, then 3 µL NaBr small molecule mix to each well of the second plate. Then 2 µL FDH stock from the first FDH stock plate was transferred to each well of both plates. To initiate the halogenation reactions, 2 µL cofactor regen mix was added to each well of both plates. The plates were returned to the carousel, and two more substrate-containing 384-well plates were transferred from the carousel to the Nimbus deck. The above process was repeated, using the second and third FDH plates for the successive cycles. Between all liquid handling steps, the tips were washed twice in a pool of water constantly refreshed with a peristaltic pump and drained by vacuum, then washed once with EtOH. To dry the tips between liquid transfers, the tips were ejected back into the tip rack, which was connected to house vacuum. The tips were dried in the rack for 12 seconds, then liquid handling resumed. After all reactions had been prepared, the 384-well plates were sealed with adhesive film and shaken on a plate shaker at 750 rpm for 18 hr.

Reaction workup – probe substrate screens

Reactions were quenched and proteins precipitated by adding 40 µL MeOH to each well of the 384-well reaction plate using a Hamilton Nimbus liquid handler. Reaction plates were then centrifuged at 3600 rpm, 4 °C for 3 min to pellet precipitated protein. Using a Hamilton Nimbus liquid handler, 30 µL of supernatant from the reaction plates were transferred to new 384-well plates retaining the well layout of the previous plate (this process removed solids sufficiently for LC analysis; system pressure of the LC-MS did not increase significantly after >10,000 injections of these reactions). The new plates were heat sealed with aluminum foil and stored at -20 °C until analysis later in the day.

Reaction workup – four-reaction pooled screens

After 18 hr, reactions were quenched and proteins precipitated by adding 30 µL MeOH to each well of the 384-well reaction plate using a Hamilton Nimbus liquid handler. Reaction plates were then centrifuged at 3600 rpm, 4 °C for 3 min to pellet precipitated protein. Using a Hamilton Nimbus liquid handler, 20 µL of supernatant from each reaction plate were re-arrayed to a single quadrant of a new 384-well plate, pooling four reactions together for analysis. Up to four reaction plates were re-arrayed into a single new 384-well plate in this manner. The new plates were heat sealed with aluminum foil and stored at -20 °C until analysis later in the day.

Reaction workup – two-reaction pooled screens

After 18 hr, reactions were quenched and proteins precipitated by adding 40 µL MeOH to each well of the 384-well reaction plate using a Hamilton Nimbus liquid handler. Reaction plates were then centrifuged at 3600 rpm, 4 °C for 3 min to pellet precipitated protein. Using a Hamilton Nimbus liquid handler, 30 µL of supernatant from the upper two quadrants of each reaction plate were re-arrayed to the upper left

quadrant of a new 384-well plate, pooling two reactions together for analysis. The lower two quadrants of the reaction plate were re-arrayed into the upper right quadrant of the new 384-well plate. Up to two reaction plates were re-arrayed into a single new 384-well plate in this manner. The new plates were heat sealed with aluminum foil and stored at -20 °C until analysis later in the day.

B) High-Throughput Screen Reaction Data Analysis

EIC chromatograms for each reaction were Agilent data files from the high-throughput screen were converted to CSV chromatograms using OpenChrom.[10] These CSV files were imported into R using the data.table package, and this full LCMS dataset was saved as a Feather file for faster access later. EIC chromatograms of each starting material, single halogenation product isotopomer, and double halogenation product were extracted for each experiment ($[M+H]^+$ *m/z* for compounds detected in positive mode; $[M-H]^-$ for compounds detected in negative mode). EIC chromatograms were baseline-corrected using the minimum rolling average of the intensity (center alignment; 13 point window) computed with the RcppRoll package (Figure S9). The maximum intensity in each EIC—the peak height—was used to quantify amount of each compound. EIC chromatograms of singly-halogenated product isotopomers that included contaminant peaks (e.g. $[HEPES+H]^+$ *m/z* = 238, which overlaps with $[^{79}Br\text{-tryptamine}+H]^+$) were excluded for analysis, instead, the non-excluded isotopomer peak height was multiplied to account for the excluded abundance (e.g. ^{81}Br peak multiplied by two if ^{79}Br peak was excluded; ^{35}Cl peak multiplied by 1.33 if ^{37}Cl was excluded; etc.). EIC chromatograms were manually reviewed to remove false positives. Conversion for each experiment was computed by adding the product peak heights and dividing by the sum of the starting material peak height and product peak heights. The mean conversion in replicate experiments was used as the conversion for each reaction in generating the heatmap and performing hierarchical clustering analysis of the activity data.

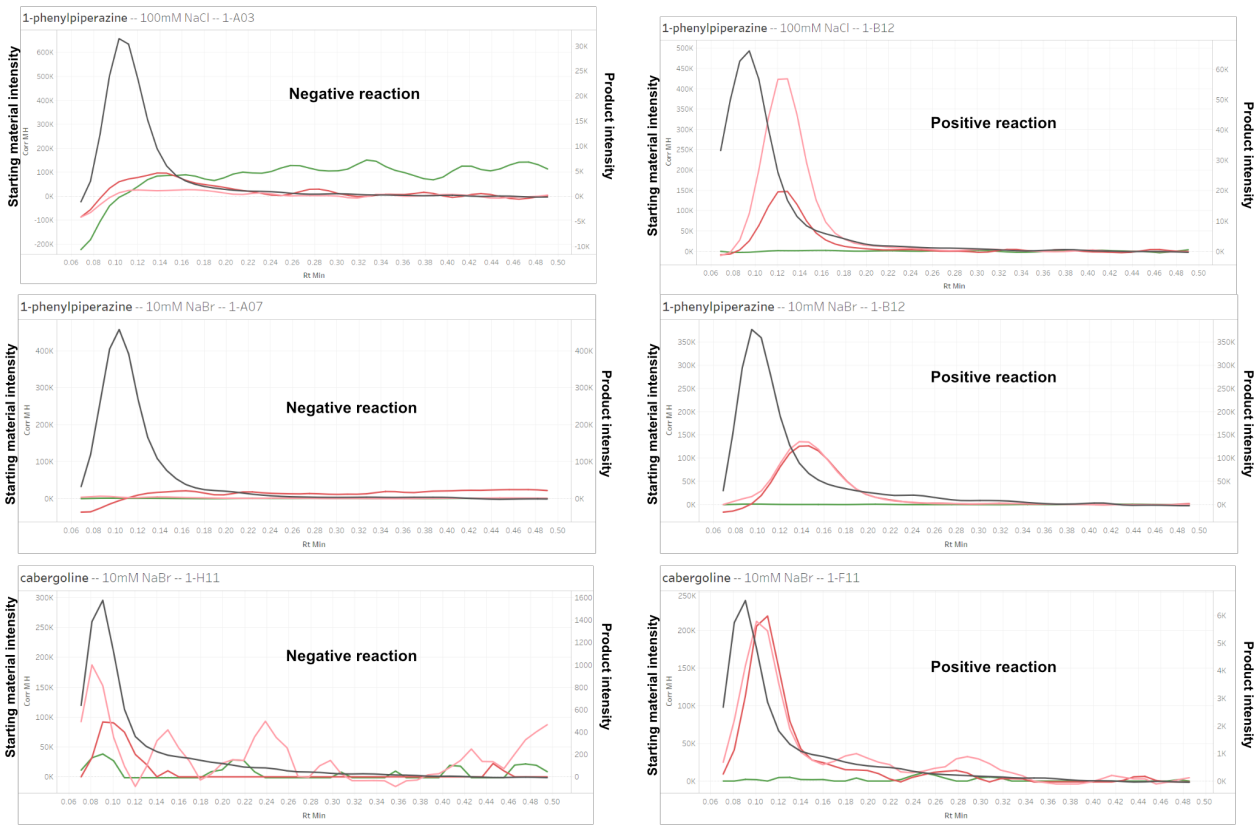
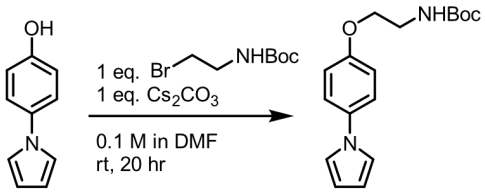


Figure S9. Example EIC chromatograms for select reactions. Dark gray curve is the EIC chromatogram for starting material; pink is the EIC chromatogram for the monohalogenated product from the lighter halogen isotope (^{35}Cl or ^{79}Br); red is the EIC chromatogram for the monohalogenated product from the heavier halogen isotope (^{37}Cl or ^{81}Br); green is the EIC chromatogram for doubly-halogenated product (the most abundant isotopomer).

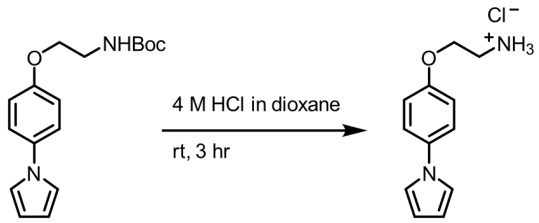
C) Heatmap/Clustering Analysis

The full high-throughput screen dataset, a dataset containing information about each substrate (molecular weight, etc.) and a dataset containing a manually-curated list of experimentally characterized FDHs as well as information about each enzyme in the genome-mined set obtained from UniProt were imported into R for further analysis. Any conversion value from the HTS that was less than 8% was set to 0. Substrates that were not chlorinated or brominated beyond 5% conversion by any tested FDH were omitted from the analysis, and any FDH that did not halogenate any substrate beyond 5% conversion was also omitted. Heatmap generation and hierarchical clustering analysis was performed using the heatmaply package[11] with "pearson" as the distfun argument; other (non-aesthetic) arguments were left at their default values. Other distance functions were evaluated, but we found that Pearson distance best conveyed the correlation in activity trends (rather than similarity in overall conversions) among enzymes and among substrate/halide pairs. The number of clusters to color was selected manually.

V. Substrate Synthesis



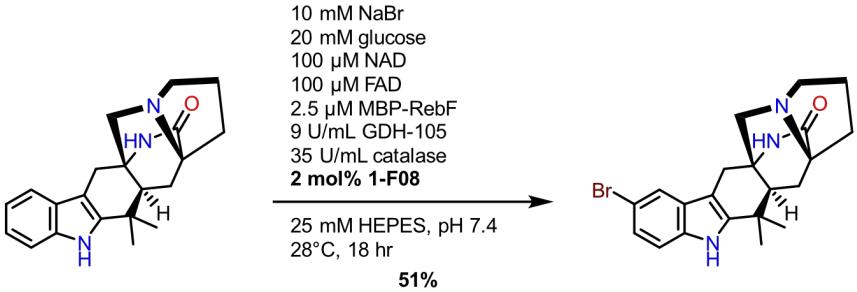
Starting material **S1** (4-(1H-pyrrol-1-yl)phenol; 1 eq., 100 mg, 0.63 mmol) was dissolved in DMF in a 100-mL round-bottom flask charged with a stirbar, after which Cs_2CO_3 (1 eq., 205 mg, 0.63 mmol) was added. 2-(Boc-amino)ethyl bromide (1 eq., 141 mg, 0.63 mmol) was added at room temperature, and the reaction was allowed to stir overnight. After completion (20 hr), the reaction was diluted with EtOAc, and the organic phase was washed with 0.5% LiCl, 0.5 M HCl, saturated $NaHCO_3$, and brine. The washed organic phase was dried with Na_2SO_4 then concentrated by rotary evaporation to yield an off-white solid as the crude material. The crude product was purified by flash silica gel chromatography eluting with 5:1 hexanes:EtOAc + 2% Et_3N to yield 108 mg **S2** as a white power (57% yield). **^{1}H NMR:** (500 MHz, $CDCl_3$) δ 7.23 (d, J = 8.9 Hz, 2H), 6.92 (t, J = 2.2 Hz, 2H), 6.86 (d, J = 8.6 Hz, 2H), 6.25 (t, J = 2.2 Hz, 2H), 4.92 (bs, 1H), 3.96 (t, J = 5.2 Hz, 2H), 3.47 (q, J = 5.5 Hz, 2H), 1.39 (s, 9H). **^{13}C NMR:** (126 MHz, $CDCl_3$) δ 155.58, 154.85, 133.77, 121.13, 118.60, 114.20, 108.89, 78.58, 66.52, 39.09, 27.37. **HRMS:** *m/z* (ESI) calc. for $[C_{17}H_{22}O_3N_2Na]^+$ ($[M + Na]^+$): 325.1523; found 325.1524.



Starting material **S2** (45.5 mg, 0.15 mmol) was added to a 20 mL scintillation charged with a stir bar, then 3 mL 4.0 M HCl in dioxane was added dropwise. The reaction was stirred, during which time a solid precipitate formed, until TLC analysis indicated complete consumption of starting material (3 hr). Solvent was removed under a stream of N_2. Material was used without further purification. **HRMS:** *m/z* (ESI) calc. for $[C_{12}H_{15}ON_2]^+$ ($[M + H]^+$): 203.1179; found 203.1178.

VI. Preparative Scale Bioconversions

Premalbrancheamide bromination with 1-F08



Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: premalbrancheamide (30 mM DMSO stock, 500 µM final

concentration 12 µmol, 4.0 mg), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-F08 (2 mol%, 10 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and finally GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated in a VWR 1585 incubator and agitated at 250 rpm at 28°C. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (Supelco Discovery C18, 3.0mL/min, 20%B 0-5 min, 20-35%B 5.01-23 min, 35%B 23.01-24 min, 95%B 24.01-29 min, 20%B 28.01-30 min). The fractions found to contain product via LC-MS were collected and concentrated by rotary evaporation to produce product in 51% yield. **^{1}H NMR:** (600 MHz, Methylene Chloride-d_2) δ 8.03 (s, 1H), 7.49 (d, J = 1.6 Hz, 1H), 7.30 (d, J = 8.3 Hz, 1H), 7.19 (dd, J = 8.4, 1.7 Hz, 1H), 6.40 (s, 1H), 3.42 – 3.35 (m, 1H), 3.35 – 3.28 (m, 1H), 2.95 (d, J = 17.1 Hz, 1H), 2.88 (d, J = 17.1 Hz, 1H), 2.78 (d, J = 10.8 Hz, 1H), 2.65 – 2.59 (m, 1H), 2.51 (q, J = 9.0 Hz, 1H), 2.39 – 2.31 (m, 1H), 2.28 (dd, J = 10.0, 3.6 Hz, 1H), 2.01 (d, J = 4.5 Hz, 2H), 1.96 (dd, J = 13.7, 3.6 Hz, 1H), 1.65 – 1.55 (m, 1H), 1.29 (s, 3H), 1.19 (s, 3H). **^{13}C NMR:** (126 MHz, DMSO-d_6) δ 173.07, 143.63, 137.74, 126.28, 121.32, 119.74, 113.69, 113.45, 103.74, 64.72, 62.45, 55.00, 53.22, 46.09, 34.48, 31.54, 27.94, 27.51, 27.37, 24.17, 23.01. **HRMS (ESI-MS):** calc. for $[C_{21}H_{26}BrN_3O]^+$ ($[M+H]^+$): 414.1176 and 416.1157, found 414.1179 and 416.1159.

Note: A second brominated product peak was found using the above HPLC method. When attempting to isolate this product the apparent brominated material degraded, resulting in a compound with m/z = 352, corresponding to [starting material+H+16]. Literature precedent for chemical bromination of this molecule suggests that the 3-brominated material is formed, which rearranges to form a spirooxindole product (**Figure S10**),[12] but we were unable to produce enough material to confirm this hypothesis.

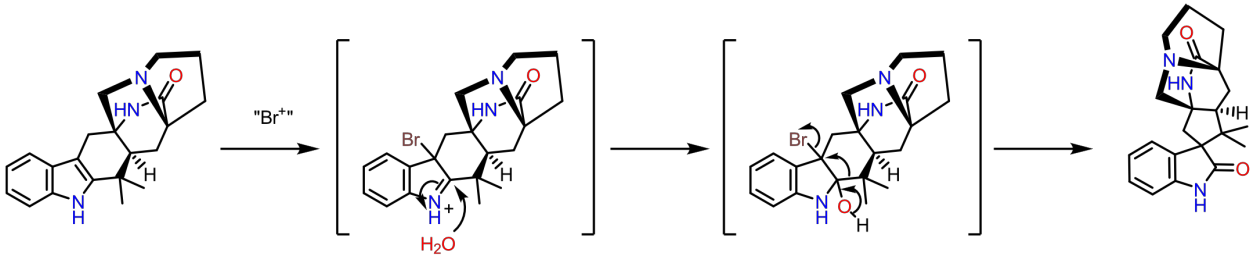
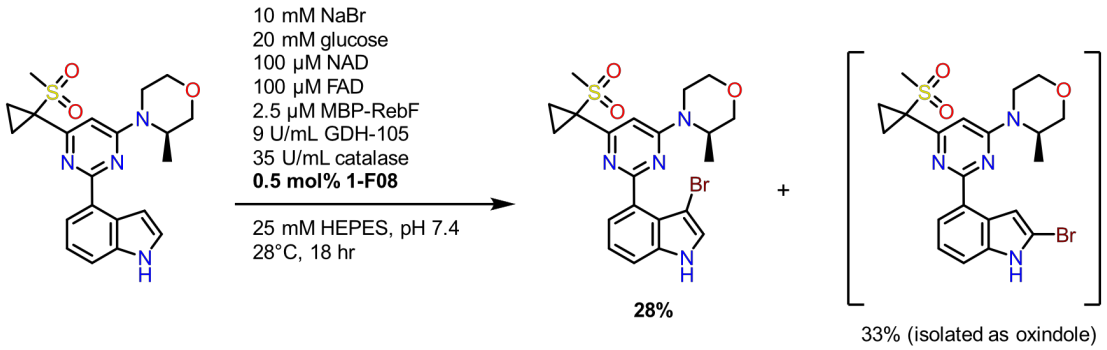


Figure S10: Proposed mechanism formation of side product spiro-oxindole from 1-F08 bromination of premalbrancheamide.

AZ20 bromination with 1-F08



Ten 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: AZ20 (30 mM DMSO stock, 500 µM final concentration 15 µmol, 6.2 mg), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration), and 200 µL DMSO (8% cosolvent). 1-F08 (0.5 mol%, 2.5 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated in a VWR 1585 incubator and agitated at 250 rpm at 28°C. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (Supelco Discovery C18 0-5 min 32%B, 5.01-20 min 32-38%B, 20-25 min 38%B, 25.01-28 min 95%B, 28.01-30 min 32%B). Two non-starting material peaks were found in the semiprep HPLC chromatogram, and each was collected separately and solvent removed by rotary evaporation. One fraction contained the 3-brominated AZ20 product, isolated in 28% yield, and the other contained an oxindole likely formed from hydrolysis of the 2-brominated AZ20 product, isolated in 33% yield.

3-Br-AZ20:

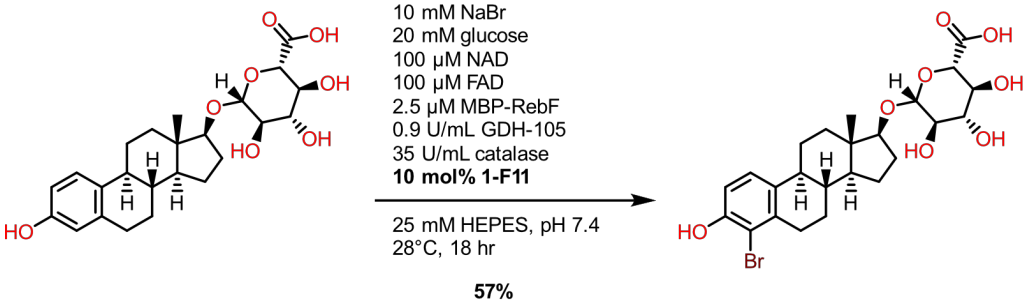
^{1}H NMR (500 MHz, Methylene Chloride-d_2) δ 8.62 (s, 1H), 7.50 (dd, J = 5.8, 3.4 Hz, 1H), 7.32 – 7.31 (m, 1H), 7.29 (s, 1H), 7.29 – 7.28 (m, 1H), 6.88 (s, 1H), 4.51 – 4.41 (m, 1H), 4.11 (d, J = 13.5 Hz, 1H), 4.00 – 3.92 (m, 1H), 3.76 (d, J = 11.5 Hz, 1H), 3.71 (dd, J = 11.6, 3.2 Hz, 1H), 3.56 (td, J = 11.9, 3.1 Hz, 1H), 3.30 (td, J = 12.9, 4.0 Hz, 1H), 3.02 (s, 3H), 1.76 (dd, 7.9 Hz, 4.7 Hz, 2H), 1.57 (dd, 7.9 Hz, 4.7 Hz, 2H), 1.32 (d, J = 6.8 Hz, 3H). **^{13}C NMR:** (126 MHz, DMSO-d_6) δ 164.98, 161.57, 161.22, 136.71, 132.88, 127.33, 122.89, 121.82, 121.60, 113.22, 101.34, 88.20, 70.68, 66.56, 46.62, 46.44, 40.81, 13.92, 12.65. One peak at ~39ppm is covered by the DMSO-d_6 peak. **HRMS (ESI-MS):** calc. for $[C_{21}H_{24}BrN_4O_3S]^+$ ($[M+H]^+$): 491.0747 and 493.0728, found 491.0745 and 493.0723.

Oxindole-AZ20:

^{1}H NMR: (400 MHz, Methylene Chloride-d_2) δ 8.04 (d, J = 8.0 Hz, 1H), 7.64 (s, 1H), 7.32 (t, J = 7.9 Hz, 1H), 6.96 (d, J = 7.7 Hz, 1H), 6.84 (s, 1H), 4.52 – 4.42 (m, 1H), 4.10 – 4.00 (m, 2H), 3.97 (s, 2H), 3.82 (d, J = 11.6 Hz, 1H), 3.73 (dd, J = 11.6, 3.2 Hz, 1H), 3.59 (td, J = 11.9, 3.1 Hz, 1H), 3.34 (td, J = 12.8,

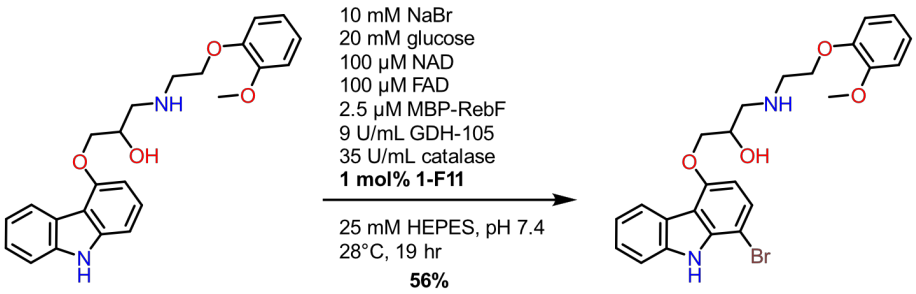
4.0 Hz, 1H), 3.03 (s, 3H), 1.85 – 1.78 (m, 2H), 1.33 (d, J = 6.8 Hz, 3H). Methylene hydrogens concealed under residual H_2O peak. **^{13}C NMR:** (126 MHz, DMSO-d_6) δ 177.09, 163.18, 162.33, 161.85, 145.07, 134.44, 127.99, 125.93, 121.99, 111.26, 102.16, 70.67, 66.47, 46.95, 46.56, 40.58, 38.83, 13.89, 12.85. **HRMS (ESI-MS):** calc. for $[C_{21}H_{25}N_4O_4S]^+$ ($[M+H]^+$): 429.1591, found: 429.1588.

β-Estradiol 17-(β-D-glucuronide) bromination with 1-F11



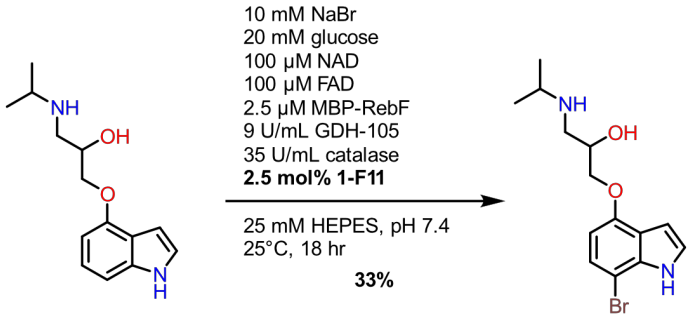
Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: β-Estradiol 17-(β-D-glucuronide) sodium salt (30 mM DMSO stock, 500 µM final concentration 12 µmol, 5.6 mg sodium salt), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-F11 (10 mol%, 50 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (0.9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated in a VWR 1585 incubator and agitated at 250 rpm and 28°C. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH 3 using 10% citric acid. The compound was then extracted into 10 mL of a 3:1 $CHCl_3$/^{i}PrOH mixture five times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC. (Supelco Discovery C18; 3.0mL/min, 35%B 0-5 min, 35-45%B 5.01-23 min, 45%B 23.01- 25 min, 95%B 25.01-32 min, 35%B 32.01-35 min). Purified fractions were collected and solvent was removed by rotary evaporation to produce product in 57% yield. **^{1}H NMR:** (400 MHz, Methanol-d_4) δ 7.14 (d, J = 8.3 Hz, 1H), 6.72 (d, J = 8.4 Hz, 1H), 4.59 (s, 2H), 4.42 (d, J = 7.8 Hz, 1H), 3.91 (d, J = 8.8 Hz, 1H), 3.58 (s, 1H), 3.24 (d, J = 8.8 Hz, 1H), 2.93 (s, 1H), 2.32 (s, 1H), 2.26 – 2.09 (m, 2H), 1.99 (s, 1H), 1.71 (s, 2H), 1.46 (s, 2H), 1.33 (d, J = 13.8 Hz, 3H), 0.90 (s, 3H). **^{13}C NMR:** (126 MHz, DMSO-d6) δ 172.68, 152.54, 136.65, 132.74, 125.35, 113.70, 113.00, 103.17, 86.65, 77.48, 74.26, 74.15, 72.72, 49.82, 43.96, 43.27, 37.83, 37.50, 31.23, 28.80, 27.34, 26.60, 22.87, 11.79. **HRMS (ESI-MS):** calc. for $[C_{24}H_{30}BrO_8]^-$ $[M-H]^-$: 525.1130 and 527.1113, found 525.1122 and 527.1101.

Carvedilol bromination with 1-F11



A volume of reaction buffer (25 mM HEPES, pH 7.4) was added to a 50 mL centrifuge tube such that addition of the following components as stock solutions would result in the given final concentration of each reaction component. Carvedilol (1 eq., 1 mM final con, 15 µmol) was added from a 30 mM stock solution in DMSO. Solutions in reaction buffer of NaBr (10 eq., 10 mM final concentration), glucose (20 eq., 20 mM final concentration), NAD (0.1 eq., 100 µM final concentration), and FAD (0.1 eq., 100 µM final concentration) were added. Freshly prepared GDH (9 U/mL final concentration) and catalase (35 U/mL final concentration) were added next, followed by MBP-RebF (2.5 µM final concentration) and 1-F11 (1 mol%, 10 µM final concentration). The centrifuge tube was capped and inverted gently twice, then 5 mL aliquots distributed into 20 mL scintillation vials. The vials were covered in foil poked with holes and transferred to an incubator at 25 °C shaking at 250 rpm. Periodically, 10 µL aliquots were removed and analyzed by UPLC to assess conversion. Near-complete conversion was observed after 19 hr, and the reaction aliquots were pooled into a 50 mL centrifuge tube and centrifuged at 3600 rpm, 4 °C for 10 minutes. The supernatant was filtered through a cotton plug into a separatory funnel, to which was added EtOAc and saturated $NaHCO_3$ until basic. The aqueous phase was removed, and the organic phase was washed with brine. The solvent of the organic phase was removed by rotary evaporation, and the crude material was stored at −20 °C until purification. The crude material was purified by reverse-phase chromatography (Biotage 35% to 60%B over 12 column volumes) to yield 5.0 mg 1-bromocarvedilol (TFA salt) as a white film (56% yield). **^{1}H NMR:** (500 MHz, Methanol-d_4) δ 8.24 (d, J = 7.8 Hz, 1H), 7.53 (d, J = 8.1 Hz, 1H), 7.46 (d, J = 8.4 Hz, 1H), 7.37 (t, J = 7.7 Hz, 1H), 7.09 (t, J = 7.5 Hz, 1H), 7.04 – 6.96 (m, 3H), 6.92 (td, J = 7.6, 6.9, 1.7 Hz, 1H), 6.70 (d, J = 8.5 Hz, 1H), 4.58 – 4.52 (m, 1H), 4.39 (dd, J = 9.9, 4.6 Hz, 1H), 4.35 – 4.27 (m, 3H), 3.79 (s, 3H), 3.65 (dd, J = 12.7, 3.1 Hz, 1H), 3.60 – 3.55 (m, 2H), 3.49 (dd, J = 12.7, 9.8 Hz, 1H). **^{13}C NMR:** (126 MHz, Methanol-d_4) δ 154.00, 149.51, 146.90, 139.49, 139.25, 137.93, 127.90, 125.14, 122.79, 122.58, 122.21, 120.96, 119.23, 114.93, 113.51, 111.83, 110.56, 101.85, 95.56, 69.74, 65.30, 64.63, 55.02, 50.22. **HRMS:** *m/z* (ESI) calc. for $[C_{24}H_{26}BrN_2O_4]^+$ ($[M+H]^+$ with ^{79}Br) 485.1070, found 485.1073.

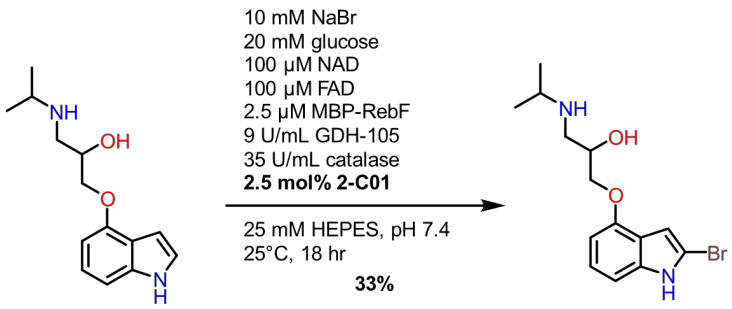
Pindolol bromination with 1-F11: 7-bromination



Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 5 mL of buffered solution was as follows: Pindolol (30 mM from DMSO stock, 1 mM final concentration, 40 μmol, 9.9 mg), and solutions in reaction buffer of NaBr (10 eq., final concentration 10 mM), glucose (20 eq., final concentration 20 mM), NAD (0.1 eq., 100 μM), FAD (0.1 eq., 100 μM final concentration). 1-F11 (2.5 mol%, 25 μM final concentration) and MBP-RebF (2.5 μM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated at room temperature (25°C) and agitated at 250 rpm in a VWR 1585 incubator. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated onto Celite. This was then purified via reverse-phase chromatography (Biotage) as previously reported.[13] The fractions were collected and concentrated, then basified with 1M NaOH to precipitate trifluoroacetate, which was removed by filtration. Product distribution: 1:15.8 2-brominated:7-brominated product. Solvent from the filtrate was removed by rotary evaporation to produce product in 33% isolated yield. **^{1}H NMR:** (400 MHz, Methanol-d_4) δ 7.21 (d, J = 3.1 Hz, 1H), 7.17 (d, J = 8.3 Hz, 1H), 6.65 (d, J = 3.2 Hz, 1H), 6.49 (d, J = 8.3 Hz, 1H), 4.31 (d, J = 7.3 Hz, 1H), 4.20 (dd, J = 10.0, 5.0 Hz, 1H), 4.12 (dd, J = 10.0, 5.8 Hz, 1H), 3.51 – 3.41 (m, 1H), 3.35 (dd, J = 12.7, 3.0 Hz, 0H), 3.20 (dd, J = 12.7, 9.5 Hz, 1H), 1.38 (d, J = 6.4 Hz, 6H). **^{13}C NMR:** (126 MHz, Methanol-d_4) δ 151.78, 135.72, 123.67, 120.09, 101.33, 99.63, 96.10, 70.62, 67.98, 49.30, 48.84, 23.84, 20.54.

HRMS (ESI-MS): calc. for $[C_{14}H_{20}BrN_2O_2]^+$ ($[M+H]^+$): 327.0703 and 329.0683, found 327.0708 and 329.0687.

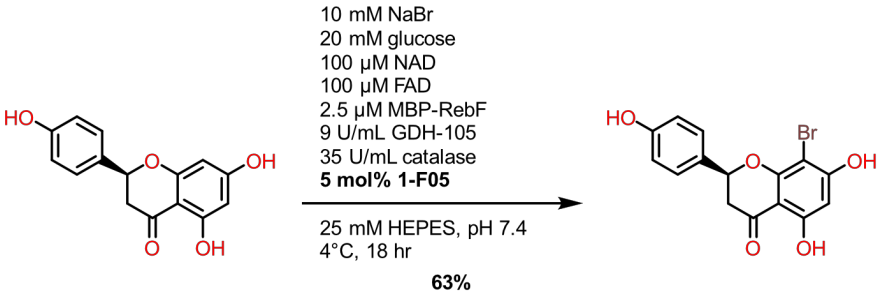
Pindolol bromination with 2-C01: 2-bromination



Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 5 mL of buffered solution was as follows: Pindolol (30 mM DMSO stock, 1 mM final concentration, 40 μmol, 9.9 mg), NaBr (10 eq. in reaction buffer, final concentration 10 mM), glucose (20 eq. in reaction buffer, final concentration 20 mM), NAD (0.1 eq. in reaction buffer, 100 μM), FAD (0.1 eq. in reaction buffer, 100 μM final concentration). 2-C01 (2.5 mol%, 25 μM final concentration) and MBP-RebF (2.5 μM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated at room temperature (25°C) agitated at 250 rpm in a VWR 1585 incubator. After 18

hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (Supelco Discovery C18; 3 mL/min, 19%B 0-5 min, 19-45%B 5.01-25 min, 45%B 25-27 min, 95%B 27.01 min-32 min, 19%B 32.01 min-35 min). Product distribution: 15.7:1.4:1 2-brominated:7-brominated:dibrominated product. The fractions found to contain product via LC-MS were collected and concentrated by rotary evaporation to produce product in 41% isolated yield. **^{1}H NMR:** (500 MHz, Methanol-d_4) δ 7.13 (s, 1H), 7.04 (t, J = 7.9 Hz, 1H), 7.00 (d, J = 8.2 Hz, 1H), 6.53 (d, J = 7.5 Hz, 1H), 4.19 (m, 1H), 4.14 (dd, J = 9.4, 4.6 Hz, 1H), 3.99 (dd, J = 9.3, 6.5 Hz, 1H), 3.20 (dd, J = 12.2, 3.4 Hz, 1H), 2.96 (hept, J = 6.5 Hz, 1H), 2.86 (dd, J = 12.2, 8.7 Hz, 1H), 1.15 (d, J = 6.2 Hz, 3H), 1.14 (d, J = 6.2 Hz, 3H). **^{13}C NMR** (126 MHz, Methanol-d_4) δ 154.47, 139.92, 125.48, 124.97, 118.07, 107.33, 102.84, 88.74, 72.40, 70.13, 51.88, 50.95, 22.91.
. **HRMS (ESI-MS):** calc. for $[C_{14}H_{20}BrN_2O_2]^+$ ($[M+H]^+$): 327.0703 and 329.0683, found 327.0708 and 329.0688.

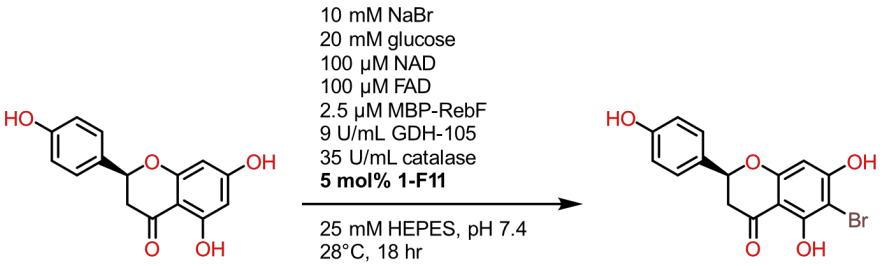
Naringenin bromination with 1-F05: 8-bromination



Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3mL of buffered solution was as follows: Naringenin (30 mM DMSO stock, 500 µM final concentration 12 µmol, 3.3 mg), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-F05 (5 mol%, 25 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were placed into a crystallization dish, this crystallization dish was taped to the top of an Eppendorf Thermomixer R, and the reactions were incubated in a 4°C cold room held and agitated at 300 rpm using the Thermomixer. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL 3:1 $CHCl_3$/^{i}PrOH three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated by rotary evaporation. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semi-prep HPLC (Supelco Discovery C18; 3.0 mL/min, 45%B 0-5 min, 45-53%B 5.01-23 min, 95%B 23.01-28 min, 45%B 28.01 min-30 min). Product distribution: 27.8:1:4.8 8-brominated:6-brominated:dibrominated product. Fractions with product were collected, and solvent was removed by rotary evaporation to produce product, which matched the ^{1}H-NMR for 8-brominated naringenin,[14] in 63% isolated yield. **^{1}H NMR** (500 MHz, DMSO-d_6) δ 12.15 (s, 1H), 9.62 (bs, 1H), 7.34 (d, J = 8.3 Hz, 2H), 6.82 (d, J = 8.6 Hz, 2H), 6.13 (s, 1H), 5.59 (dd, J = 12.4, 3.1

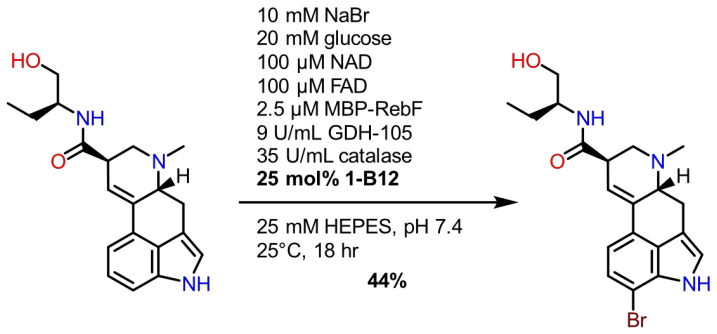
Hz, 1H), 3.32 (dd, J = 17.2, 12.4 Hz, 1H), 2.82 (dd, J = 17.2, 3.2 Hz, 1H). Note: one phenol hydroxyl peak appears to be broadened into baseline. **^{13}C NMR:** (126 MHz, DMSO-d_6) δ 196.83, 163.53, 162.41, 159.46, 158.22, 128.96, 128.62, 115.70, 102.86, 96.48, 88.74, 79.43, 41.90. **HRMS (ESI-MS):** calc. for $[C_{15}H_{10}BrO_5]^-$ ($[M-H]^-$): 348.9717; found 348.9711.

Naringenin bromination with 1-F11: 6-bromination



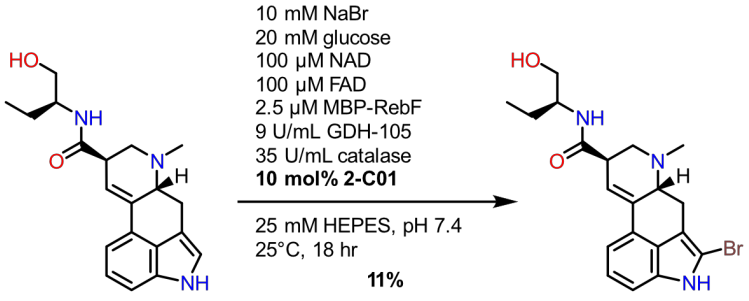
Ten 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: Naringenin (30 mM DMSO stock, 500 µM final concentration 15 µmol, 4.1 mg), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-F11 (5 mol%, 25 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated in VWR 1585 incubator and agitated at 250 rpm at 28°C. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL 3:1 $CHCl_3$/^{i}PrOH eight times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 300 µL DMSO, then purified via via reverse-phase semiprep HPLC (Supelco Discovery C18, 3.0mL/min, 45%B 0-5 min, 45-53%B 5.01-23 min, 95%B 23.01-28 min, 45%B 28.01 min-30 min). Product distribution: 1:2.4:1.3 8-brominated:6-brominated:dibrominated product. Purified fractions were collected, and solvent was removed by rotary evaporation to produce product in 26% isolated yield. **^{1}H NMR:** (500 MHz, DMSO-d_6) δ 8.16 (s, 1H), 7.31 (d, J = 8.6 Hz, 2H), 6.79 (d, J = 8.5 Hz, 2H), 5.98 (s, 1H), 5.41 (dd, J = 12.7, 3.0 Hz, 1H), 3.24 (dd, J = 17.1, 12.7 Hz, 1H), 2.68 (dd, J = 17.1, 3.1 Hz, 1H). Note: two phenolic resonances are broadened into baseline. **^{13}C NMR:** (126 MHz, DMSO-d_6) δ 195.62, 163.81, 161.79, 160.34, 158.19, 129.31, 128.77, 115.62, 101.47, 96.21, 90.73, 78.83, 42.02. **HRMS (ESI-MS):** calc. for $[C_{15}H_{10}BrO_5]^-$ ($[M-H]^-$): 348.9717 and 350.9698; found 348.9712 and 350.9690.

Methylergonovine bromination with 1-B12: 7-bromination



Four 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: Methylergonovine maleate salt (30 mM DMSO stock, 500 µM final concentration 6 µmol, 2.7 mg of maleate salt), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-B12 (25 mol%, 100 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated at room temperature (25°C) agitated at 250 rpm in a VWR 1585 incubator. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (SUPELCO Discovery C18; 3.0 mL/min, 10%B 0-5 min, 10-38%B 5.01-23 min, 38%B 23.01-24 min, 95%B 24.0-28 min, 19%B 28.01-30 min). Product distribution: 1 : 9.4 7-brominated to 2-brominated product. Purified fractions were collected, and solvent was removed by rotary evaporation to produce product in 44% isolated yield. **^{1}H NMR:** (600 MHz, Methanol-d_4) δ 8.34 (bs, 1H), 7.24 (dd, J = 7.8, 2.2 Hz, 1H), 7.11 (dd, J = 7.8, 2.2 Hz, 1H), 7.07 (s, 1H), 6.53 (s, 1H), 3.83 (dd, J = 9.5, 5.1 Hz, 1H), 3.71 – 3.52 (m, 5H), 3.38 – 3.32 (m, 1H), 3.04 (t, J = 11.1 Hz, 1H), 2.79 (s, 3H), 2.75 (t, J = 12.8 Hz, 1H), 1.67 (dt, J = 13.8, 7.3 Hz, 1H), 1.47 (dt, J = 14.6, 7.7 Hz, 1H), 0.96 (t, J = 7.4 Hz, 3H). Note: slow hydrogen-deuterium exchange appears to have resulted in the amide hydrogen resonance not disappearing. **^{13}C NMR** (126 MHz, DMSO-d_6) δ 171.91, 134.61, 132.89, 127.54, 125.19, 121.97, 121.07, 113.42, 110.83, 102.05, 63.48, 62.72, 56.03, 52.63, 49.06, 43.81, 43.34, 27.11, 24.16, 10.94.27.11, 24.16, 10.94. **HRMS (ESI-MS):** calc. for $[C_{20}H_{25}BrN_3O_2]^+$ ($[M+H]^+$): 418.1125 and 420.1107, found 418.1126 and 420.1105.

Methylergonovine bromination with 2-C01: 2-bromination



Seven 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: Methylergonovine maleate salt (30 mM DMSO stock, 500 μM final concentration 10.5 μmol, 4.8 mg maleate salt), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 μM), FAD (0.2 eq. in reaction buffer, 100 μM final concentration). 2-C01 (10 mol%, 50 μM final concentration) and MBP-RebF (2.5 μM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated at room temperature (25°C) and agitated at 250 rpm in a VWR 1585 incubator. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 μL DMSO, then purified via reverse-phase semiprep HPLC (Supelco Discovery C18; 3.0mL/minute, 10%B 0-5 min, 10-38%B 5.01-23 min, 38%B 23.01-24 min, 95%B 24.01-28 min, 19%B 28.01-30 min). Product distribution: 2.1:1 2-brominated to 7-brominated product. The fractions found to contain product via LC-MS were collected and concentrated by rotary evaporation to produce product in 11% isolated yield. **^{1}H NMR:** (500 MHz, DMSO-d_6) δ 11.48 (s, 1H), 8.14 (s, 2H), 7.73 (d, J = 8.5 Hz, 1H), 7.13 (dd, J = 8.5 Hz, 4.1 Hz, 1H), 7.09 (d, J = 5.1 Hz, 1H), 6.40 (s, 1H), 3.66 (tt, J = 10.4, 5.2 Hz, 1H), 3.52 – 3.45 (m, 1H), 3.40 (dd, J = 10.7, 5.3 Hz, 1H), 3.34 (dd, J = 10.7, 6.0 Hz, 1H), 3.29 (dd, J = 14.8, 5.8 Hz, 1H), 3.19 – 3.13 (m, 1H), 3.11 (dd, J = 11.2, 5.3 Hz, 1H), 2.64 (t, J = 11.0 Hz, 1H), 2.53 (s, 3H), 2.42 (dd, J = 14.7, 11.2 Hz, 1H), 1.65 – 1.53 (m, 1H), 1.33 (ddt, J = 16.1, 14.5, 7.4 Hz, 1H), 0.84 (t, J = 7.4 Hz, 3H).**^{13}C NMR:** (126 MHz, DMSO-d_6) δ 171.65, 163.62, 134.68, 134.12, 126.80, 126.17, 123.33, 121.75, 112.48, 109.71, 109.50, 104.28, 63.46, 62.17, 55.70, 52.69, 43.50, 42.99, 26.28, 24.16, 10.94. **HRMS (ESI-MS):** calc. for $[C_{20}H_{23}BrN_3O_2]^-$ ($[M-H]^-$): 416.0979 and 418.0959, found 416.0973 and 418.0952.

VII. Halenium Affinity Calculations

A) Halenium Affinity Calculation Setup

Procedure for HalA calculations was adapted from Borhan *et al.*[15] Structures of all substrates and all plausible Wheland intermediates formed by EAS chlorination or bromination of aryl carbons of the substrates were constructed in Chemdraw and saved as .sdf files separately for substrates and for products. More complex substrates were simplified by removing aliphatic or remote groups unlikely to exert substantial influence over the inherent gas-phase energetics of the bond formation. Initial geometries were generated by opening the .sdf file in Chem3D. Initial geometries were refined by first performing MMFF molecular mechanics minimization, then performing an MMFF equilibrium conformer search. Lowest-energy conformers were then subjected to DFT geometry optimization and frequency calculations (B3LYP/6-31G*; uncharged for starting material, and +1 charge for product). The gas-phase standard enthalpy of formation for each compound (H°; 0.952 scale factor for vibrational frequency; 298.15 K; 1 atm) was exported to Microsoft Excel.

Calculations for halenium (Cl^+ and Br^+) were performed in Spartan comparably using the triplet ground state.

HalA values were computed in Microsoft Excel using Error! Reference source not found..

$$HalA = -H°_{Sub-X} + H°_{Sub} + H°_{X+} + \frac{5}{2}RT$$ **Equation S2**

B) Halenium Affinity Data

VIII. Halide Selectivity Experiments

A) Reaction Setup

Stock solutions of substrate, FDH, NaX/glucose/NAD/FAD, and MBP-RebF/catalase/GDH were prepared in 25 mM HEPES, pH 7.4 such that, when combined in wells of a V-bottom 96-well microtiter plate to initiate the reaction, the final concentrations and reaction volume were as listed in the table below.

NaX	10 mM total
glucose	20 mM
NAD	100 μM
FAD	100 μM
GDH	9 U/mL
MBP-RebF	2.5 μM
Catalase	35 U/mL
FDH	25 μM
substrate	1 mM
Rxn Volume	45 μL

Concentrations of halide were varied for NaCl and NaBr.

Concentration NaCl	Concentration NaBr
10 mM	0 mM
9 mM	1 mM
5 mM	5 mM
1 mM	9 mM
0 mM	10 mM

Reactions were quenched after 16 hr by addition of 125 μL MeOH. Precipitated protein was pelleted by centrifugation at 3600 rpm at 4°C for 10 min. Supernatant was transferred to a 0.2 μm PVDF filter plate,

and filtered into a new flat-bottom polystyrene 96-well microtiter plate by centrifugation at 3600 rpm at 4°C for 10 min. The microtiter plate was heat-sealed with aluminum foil and kept at -20°C until analysis.

Reactions were analyzed by LC-MS eluting through an Agilent StableBond C18 (1.8 μm particle size, 2.1 x 5.0 mm) with 0.1% formic acid in H_2O as the A solvent and 0.1% formic acid in acetonitrile as the B solvent using methods shown below. MS data were collected in dual polarity mode.

Substrate	**LC Gradient Method**
L-Trp	10-50%B over 4 min
1-phenylpiperazine	20-50%B over 4 min
Pindolol	20-45%B over 4 min
2,4-dihydroxyacetophenone	40-65%B over 4 min

B) Halide Selectivity Analysis

Mass spectrometry data for the halide selectivity experiments were converted to Agilent MassHunter format using an Agilent utility and processed using MassHunter. Integration of the extracted ion chromatograms for the different species (starting material, monohalogenated product, and dihalogenated product) were used to obtain conversions. Comparison to conversions obtained from UV chromatogram integration (220 nm) demonstrated that the two methods were qualitatively indistinguishable (data not shown). For monochlorinated L-Trp, 1-phenylpiperazine, and pindolol, the EIC chromatograms consist of the $[M^{35}Cl+H]^+$ and $[M^{37}Cl+H]^+$ signal added; for monochlorinated 2,4-dihydroxyacetophenone, the EIC chromatograms consist of the $[M^{35}Cl-H]^-$ and $[M^{37}Cl-H]^-$ signal added, where MCl represents the monoisotopic molecular mass of a product isotopomer. For monobrominated L-Trp, 1-phenylpiperazine, and pindolol, the EIC chromatograms consist of the $[M^{79}Br+H]^+$ and $[M^{81}Br+H]^+$ signal added; for monobrominated 2,4-dihydroxyacetophenone, the EIC chromatograms consist of the $[M^{79}Br-H]^-$ and $[M^{81}Br-H]^-$ signal added, where MBr represents the monoisotopic molecular mass of a product isotopomer. For dichlorinated L-Trp, 1-phenylpiperazine, and pindolol, the EIC chromatograms consist of the $[M^{35}Cl^{35}Cl+H]^+$ and $[M^{35}Cl^{37}Cl+H]^+$ signals added; for 2,4-dihydroxyacetophenone, the EIC chromatograms consist of the $[M^{35}Cl^{35}Cl-H]^-$ and $[M^{35}Cl^{37}Cl-H]^-$ signals added. For dibrominated L-Trp, 1-phenylpiperazine, and pindolol, the EIC chromatograms consist of the $[M^{79}Br^{81}Br+H]^+$ and $[M^{79}Br^{79}Br+H]^+$ signals added; for 2,4-dihydroxyacetophenone, the EIC chromatograms consist of the $[M^{79}Br^{81}Br-H]^-$ and $[M^{79}Br^{79}Br-H]^-$ signals added. An analysis method was developed in MassHunter for the integration of all listed signals, and these were used to compute conversion values illustrated in plots below.

RebH

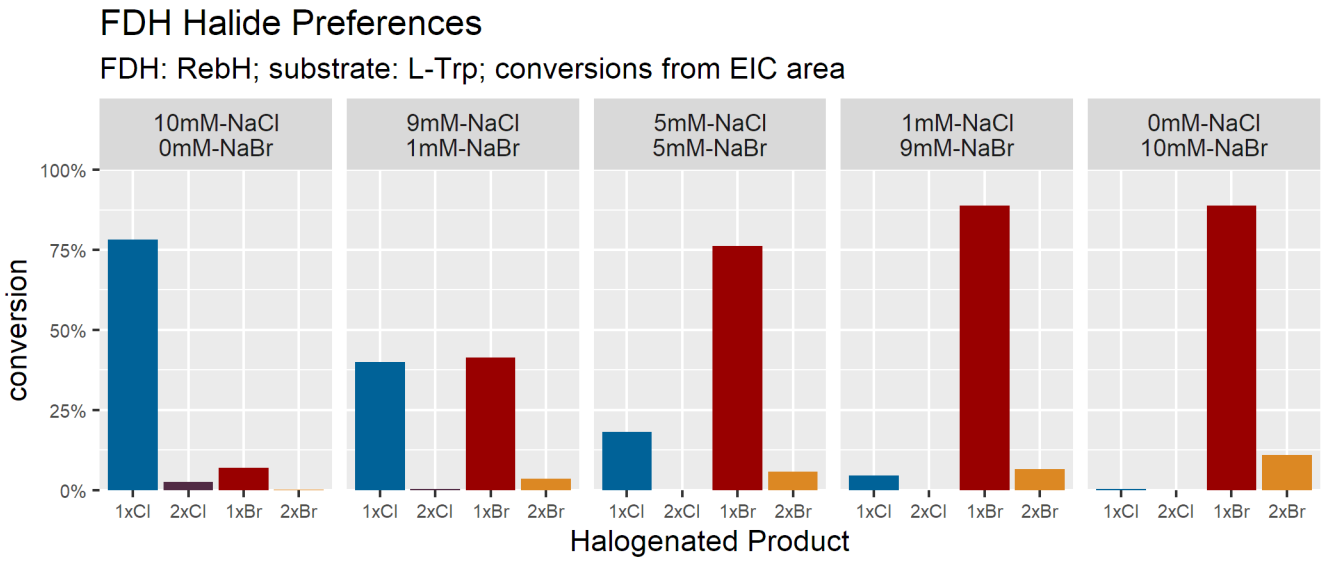


Figure S11. Halide selectivity of RebH in L-Trp halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

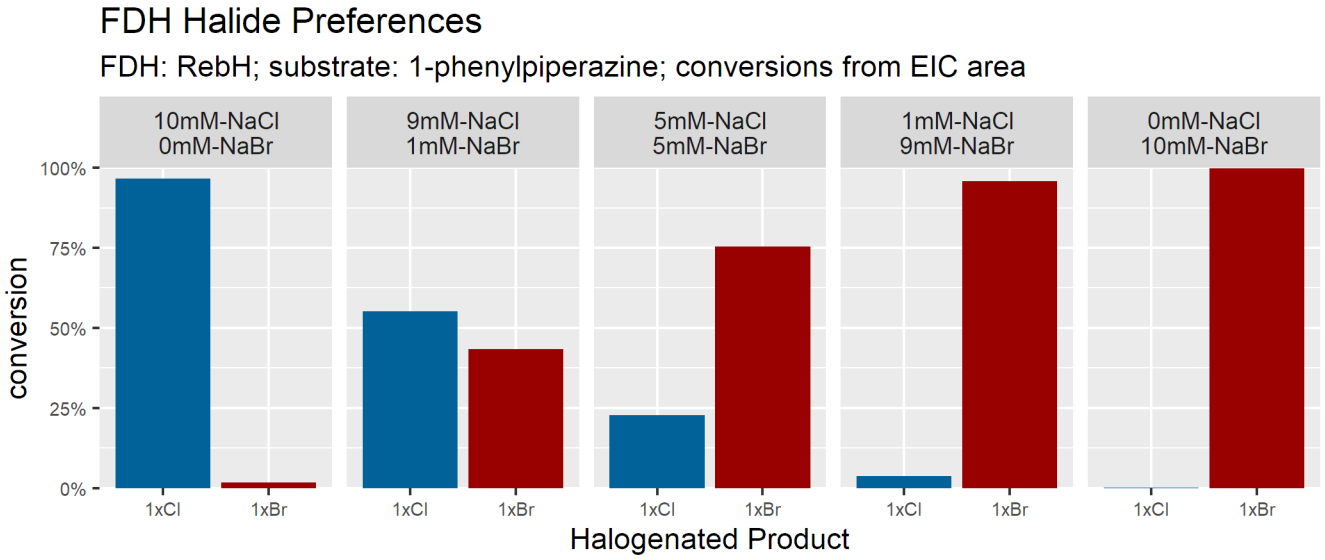


Figure S12. Halide selectivity of RebH in 1-phenylpiperazine halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

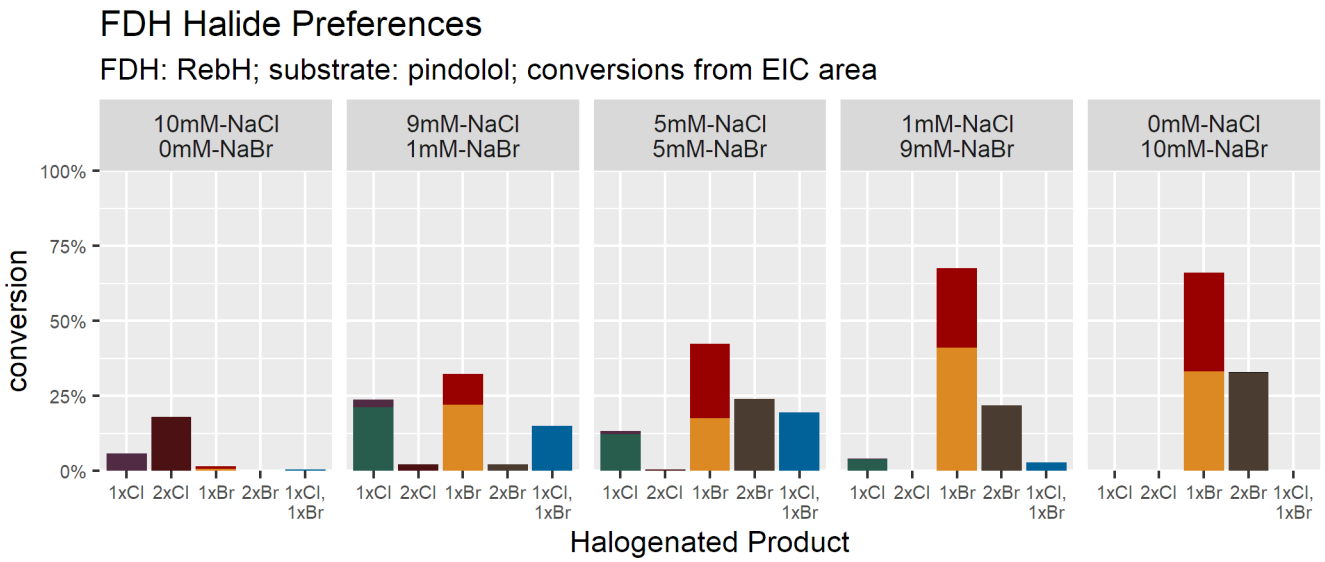


Figure S13. Halide selectivity of RebH in pindolol halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product *m/z*).

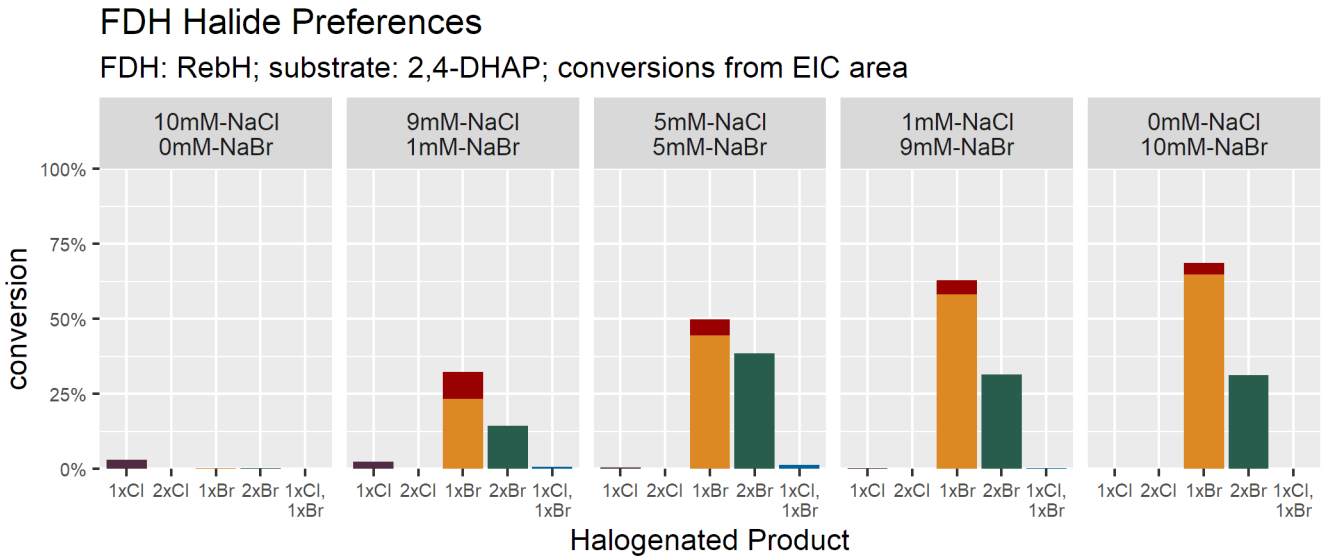


Figure S14. Halide selectivity of RebH in 2,4-dihydroxyacetophenone halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product *m/z*).

1-B12

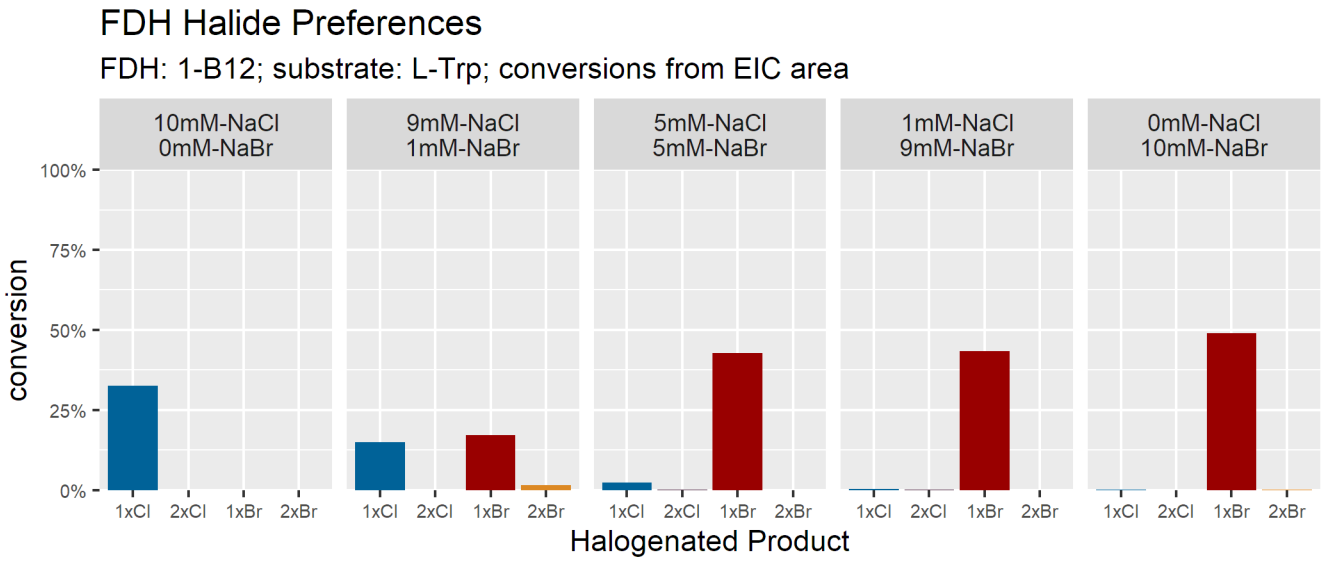


Figure S15. Halide selectivity of 1-B12 in pindolol halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

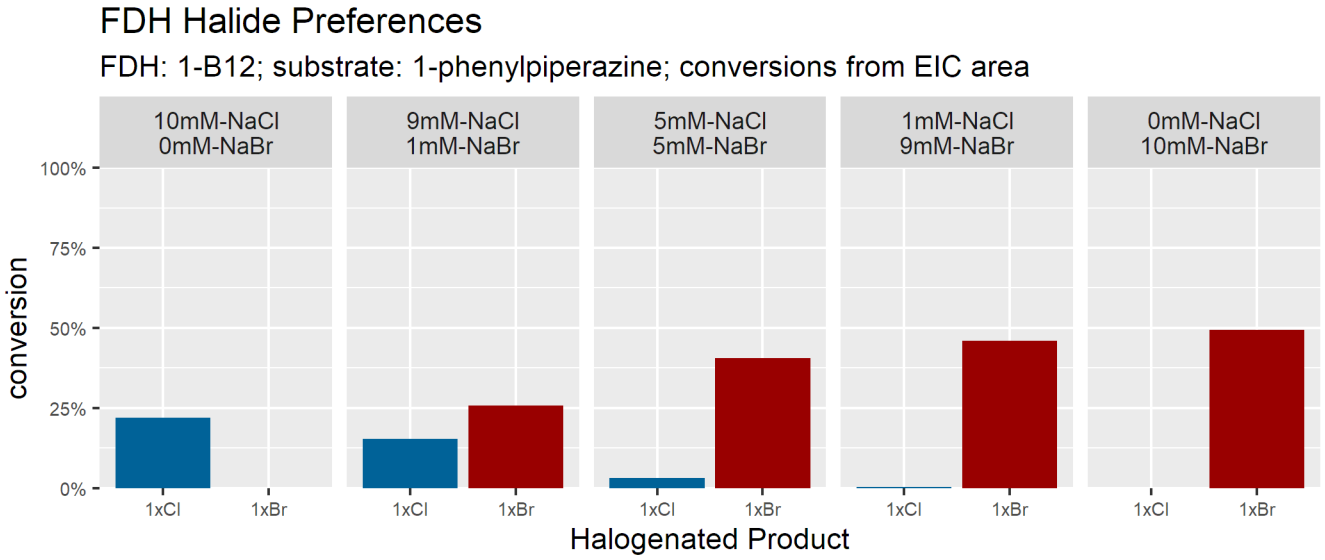


Figure S16. Halide selectivity of 1-B12 in 1-phenylpiperazine halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

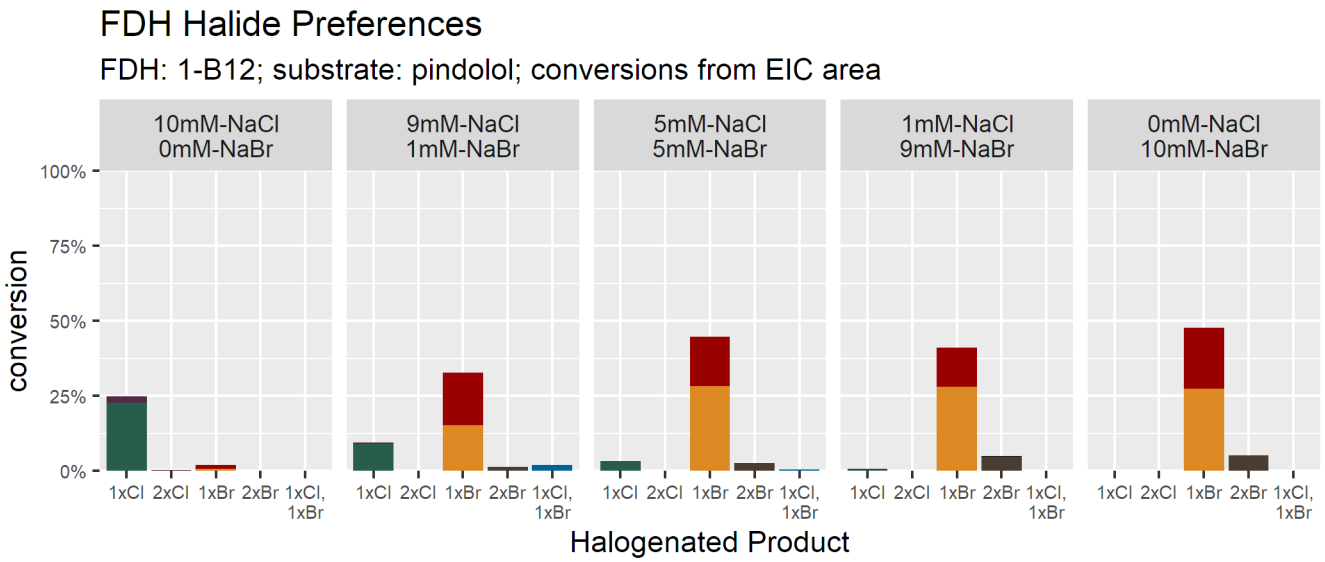


Figure S17. Halide selectivity of 1-B12 in pindolol halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product *m/z*).

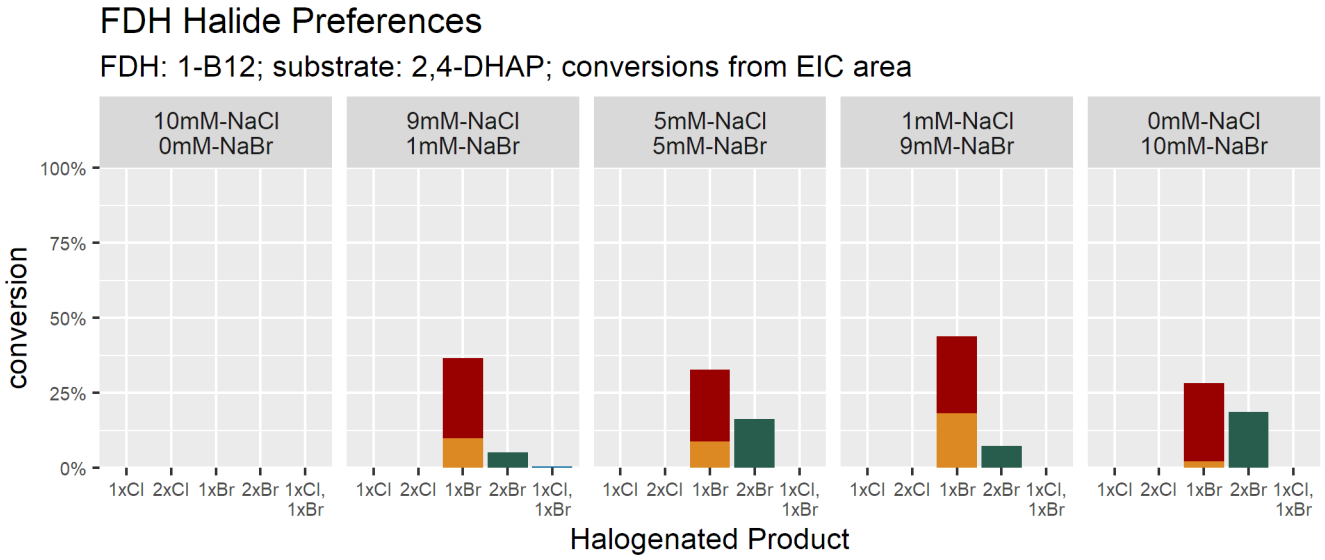


Figure S18. Halide selectivity of 1-B12 in 2,4-dihydroxyacetophenone halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product *m/z*).

2-C01

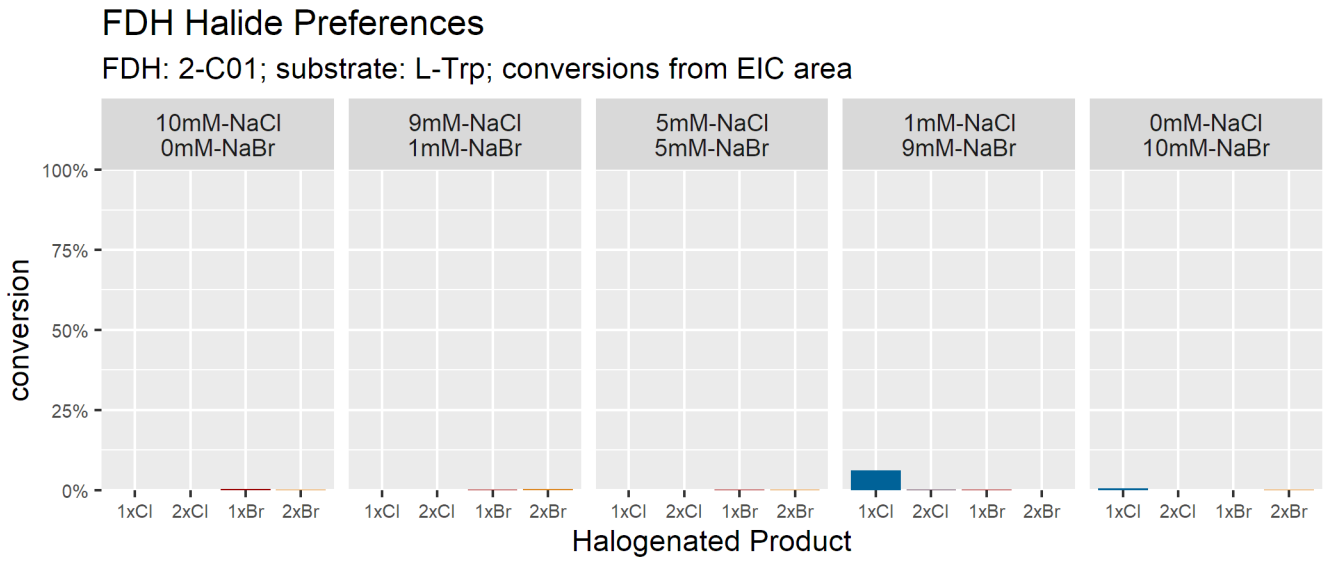


Figure S19. Halide selectivity of 2-C01 in L-Trp halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

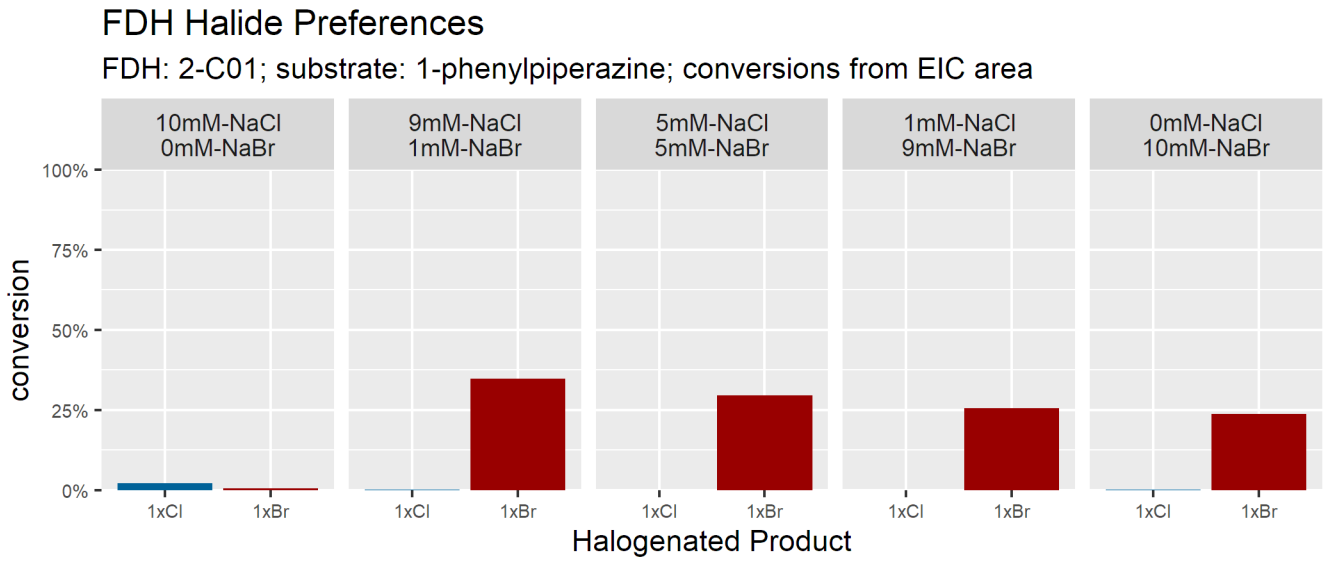


Figure S20. Halide selectivity of 2-C01 in 1-phenylpiperazine halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

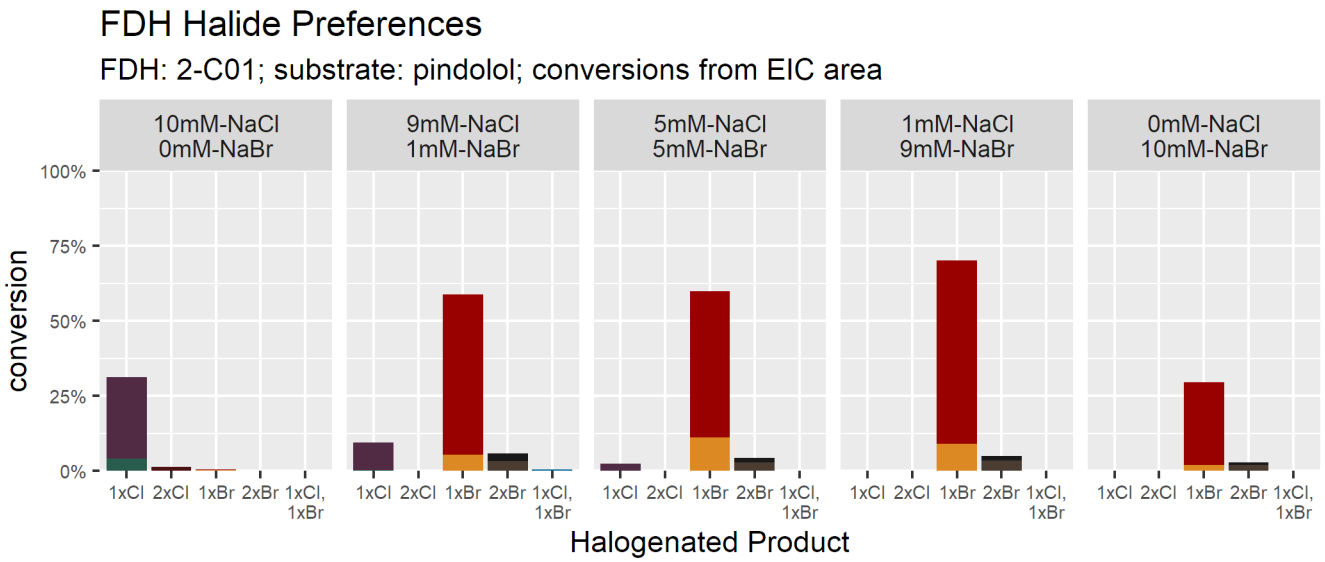


Figure S21. Halide selectivity of 2-C01 in pindolol halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product *m/z*).

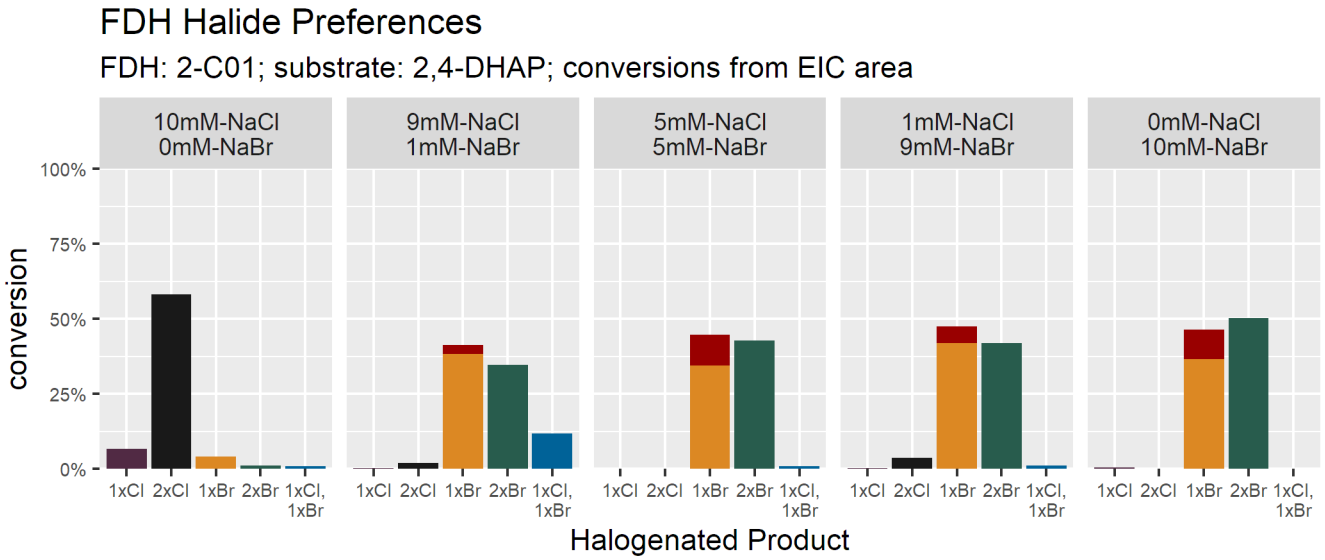


Figure S22. Halide selectivity of 2-C01 in 2,4-dihydroxyacetophenone halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product *m/z*).

IX. Analytical Scale Bioconversions

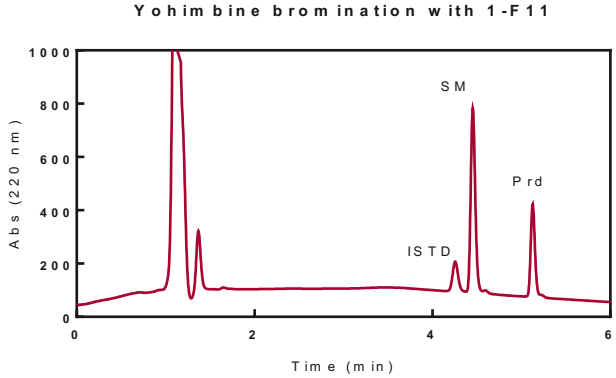
A) Yohimbine Bioconversion

Analytical-scale bioconversion of 1-F11 bromination of yohimbine for LC analysis were prepared using substrate/small molecule/FDH/cofactor regen mixtures similarly to the high-throughput screen reaction setup.

NaBr	10 mM
glucose	20 mM
NAD	100 µM
FAD	100 µM
GDH	9 U/mL
MBP-RebF	2.5 µM
Catalase	35 U/mL
FDH	25 µM
substrate	1 mM
Rxn Volume	45 µL

Bioconversion was quenched after 20 hr with 45 µL of 1 mM internal standard phenol in MeOH. Precipitated protein was pelleted by centrifugation at 3600 rpm at 4 °C for 3 min. The supernatant was transferred to a 0.45 µm PVDF 96-well filter placed atop a V-bottom 96-well plate, and the filter plate was centrifuged at 3600 rpm at 4 °C for 3 min. The filtered quenched reaction plate was heat-sealed with aluminum foil then kept at -20 °C until analysis.

Analysis was performed using an Agilent 1100 HPLC equipped with an Agilent Eclipse Plus C18 column (3.5 µm particle size; 4.6 x 150 mm), eluting with a gradient of 20-80%B over 6 min.



B) HTS Validation Study

Although the high-throughput LC-MS-based assay we chose for this work has been extensively applied and validation for quantitation of reaction conversions,[16] we were interested in determining whether our method gave good quantitative correlation with conversions obtained through more conventional HPLC methods that rely on integration of UV chromatogram peaks. To this end, we ran a subset of reactions (Table S2) under conditions that matched the conditions used for the high-throughput screen. After quenching, reactions were diluted with 100 µL MeOH and filtered through 0.2 µm PVDF syringe filters into high recovery LC vials. Reactions were analyzed by Agilent 1100 HPLC, Agilent Eclipse Plus C18 column (3.5 µm particle size; 4.6 x 150 mm), monitoring at 220 nm. Peaks were integrated manually for product and starting material to determine conversions. Afterward, these samples were analyzed using the high-throughput LC-MS assay described above. Conversions for each reaction were added to a scatterplot where X values indicated conversion by HPLC and Y values indicated conversion by LC-MS (Figure S23). Excellent agreement was found between the two conversion values, validating our screening approach.

Table S2. Substrates and FDHs tested to validate LC-MS-based high-throughput screen.

Substrates	**FDHs**
L-Trp	BSA (negative control protein)
UC-066	RebH
Pindolol	1-F07 (inactive negative control)
	1-F08
	1-F10
	1-F11
	1-B12
	1-H11

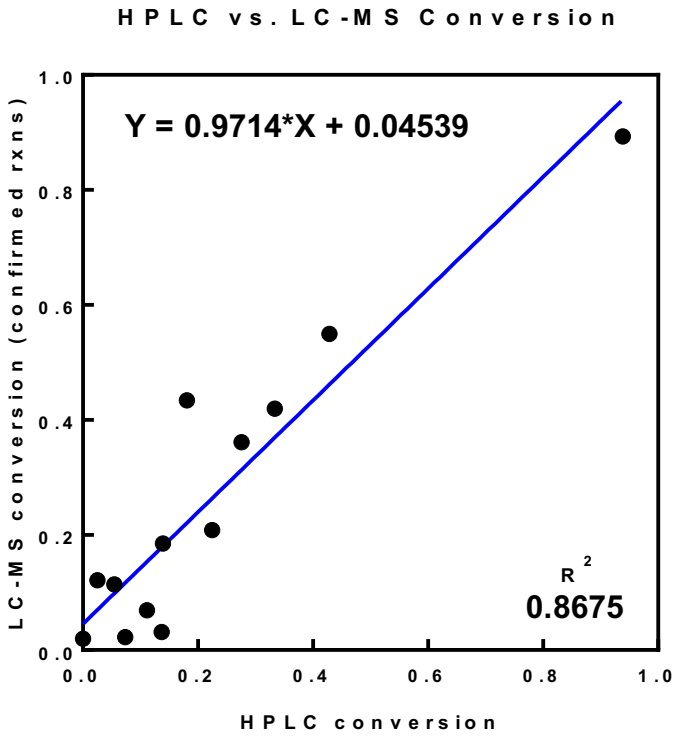


Figure S23. Comparison of conversions obtained from integrating UV chromatogram peaks by HPLC (220 nm) and conversions obtained from the high-throughput LC-MS screen.

X. SDS-PAGE Gels

A selection of SDS-PAGE gels used to assess the solubility of genome-mined enzymes is shown below in Figure S24.

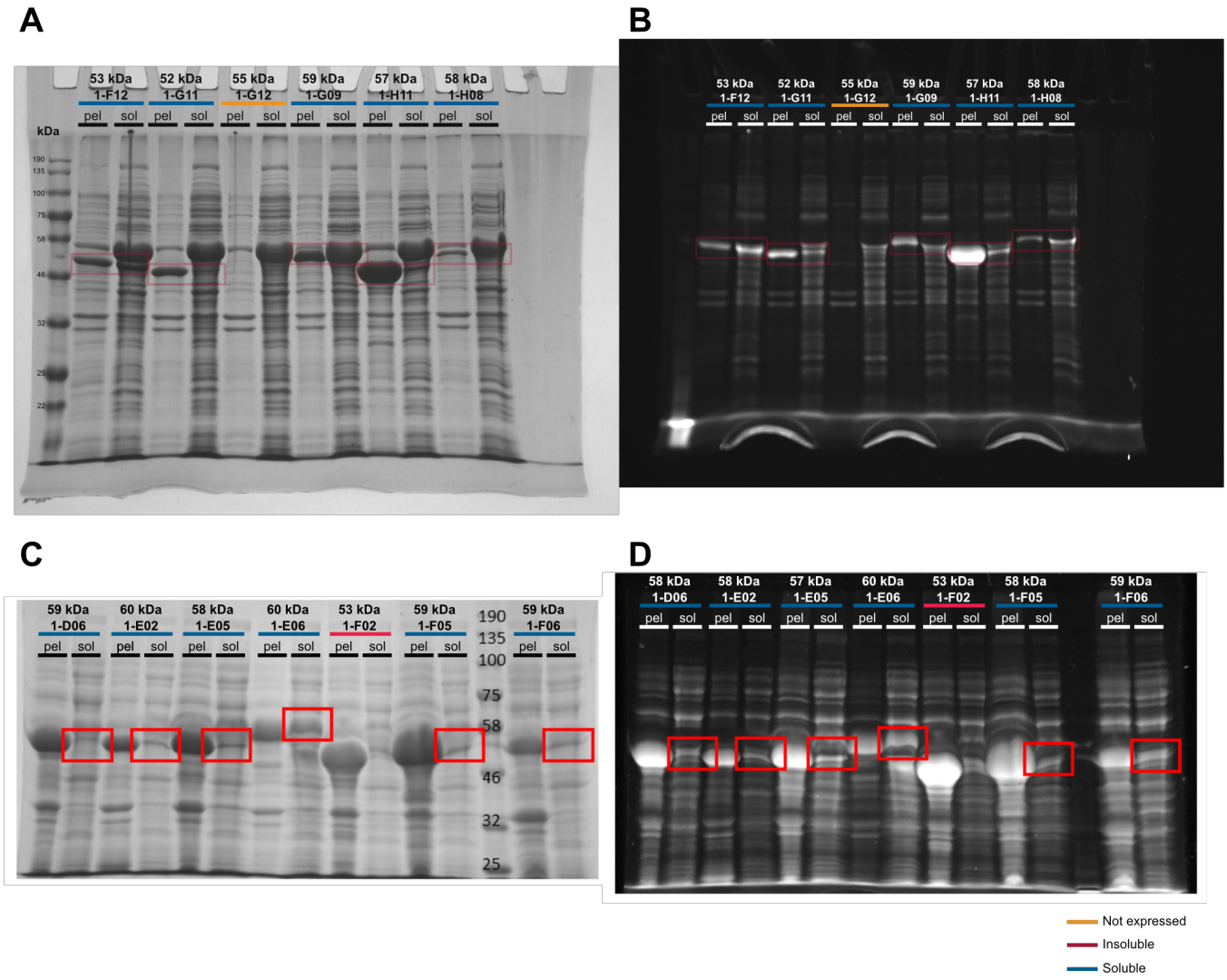


Figure S24. Selection of SDS-PAGE gels used to assess solubility of genome-mined FDHs. A, B) SDS-PAGE gels using gels cast with 12% acrylamide. A) Coomassie-stained gel, with proteins of interest highlighted in red boxes. B) Trichloroethanol fluorescence-stained gel, with proteins of interest highlighted in red boxes. Note the shadow-like band that appears above some of the FDH bands—this is groEL, which has zero Trp residues and is not stained by the fluorescence staining procedure. C, D) SDS-PAGE gels using gels cast with 10% acrylamide. C) Coomassie-stained gel, with proteins of interest highlighted in red boxes D) Trichloroethanol fluorescence-stained gel, with proteins of interest highlighted in red boxes.

XI. Circular Dichroism

A) CD Data Collection and Analysis

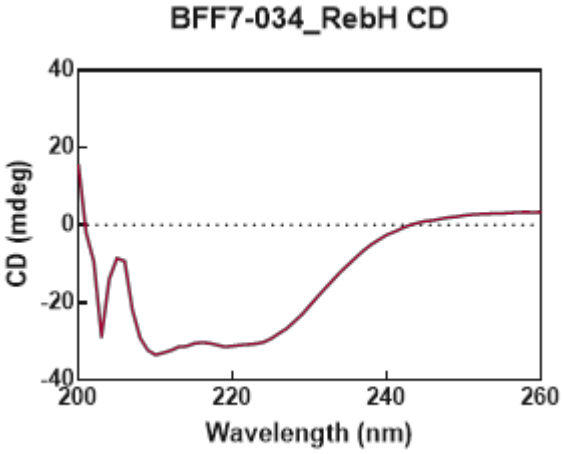
Frozen stocks of enzyme in 10% glycerol, 25 mM HEPES, pH 7.4 stored at −80°C were thawed on ice, then buffer exchanged twice in 4 mL Amicon Ultra 30K MWCO spin filters into 25 mM HEPES, pH 7.4. Enzyme stocks were transferred to ice cold Eppendorf tubes, which were centrifuged at 13.2 krpm, 4°C for 3 minutes. Concentration of enzyme solutions were measured by A_{280}, then 10 µM stock solutions of each enzyme were prepared using cold 25 mM HEPES, pH 7.4. For each enzyme, 10 µM enzyme stock was transferred to a 3 mm pathlength cuvette, which was placed into the CD spectrometer at 20°C. A

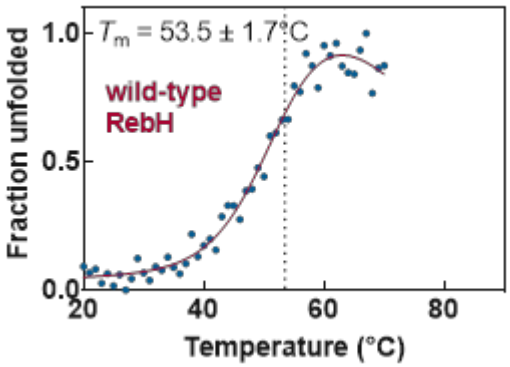
wavelength scan was acquired between 190 and 260 nm at 100 nm/min and a 1 nm wavelength width. Variable temperature CD was performed monitoring at 222 nm with temperature increasing at 1°C/min, data points collected at each 0.5°C interval. After the thermal melt, another wavelength scan was acquired to confirm thermal unfolding hysteresis.

Circular dichroism thermal melts were processed using CDPal using the Autofit All function.[17]

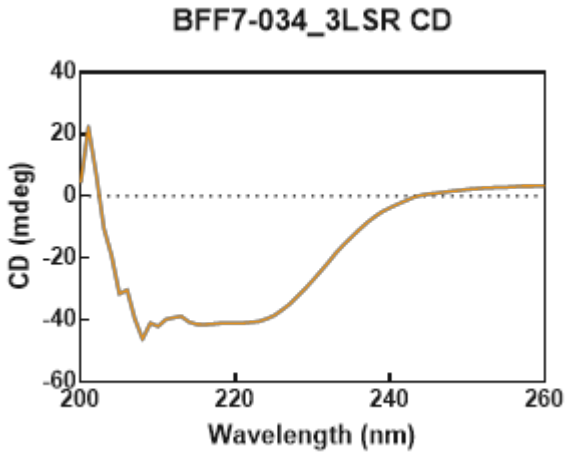
B) CD Data

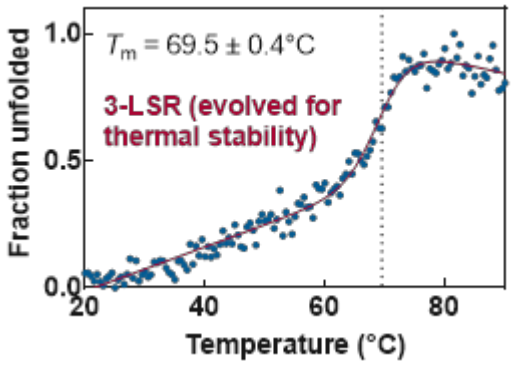
RebH



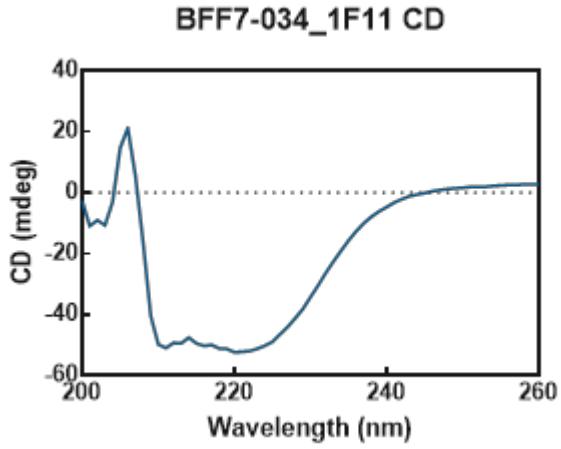


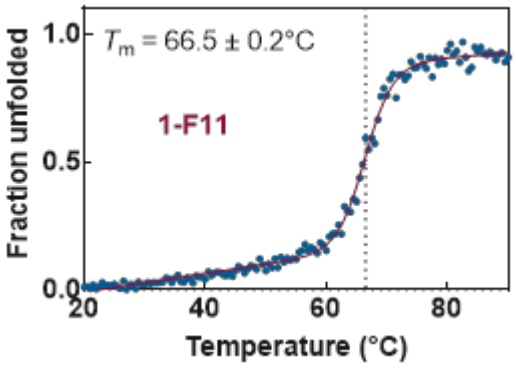
3-LSR





1-F11

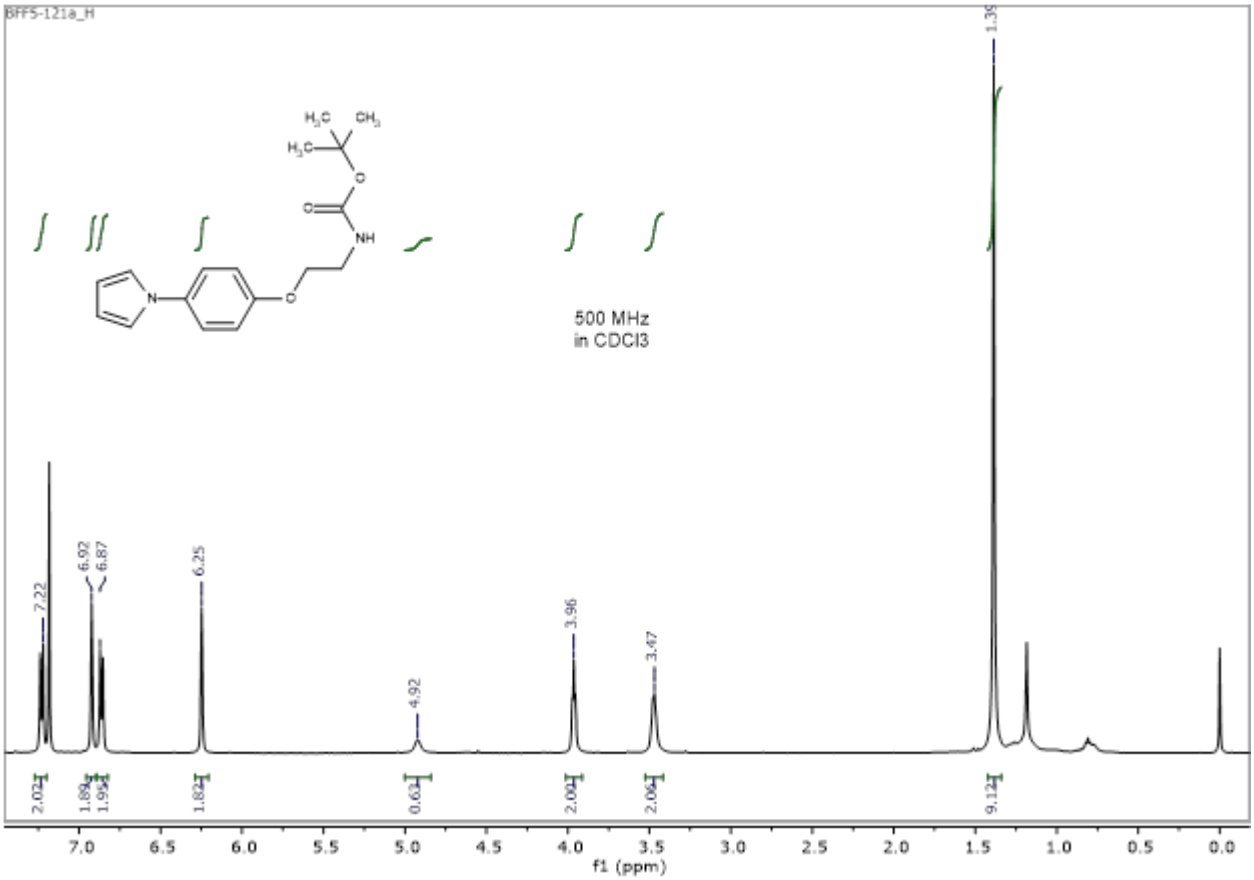




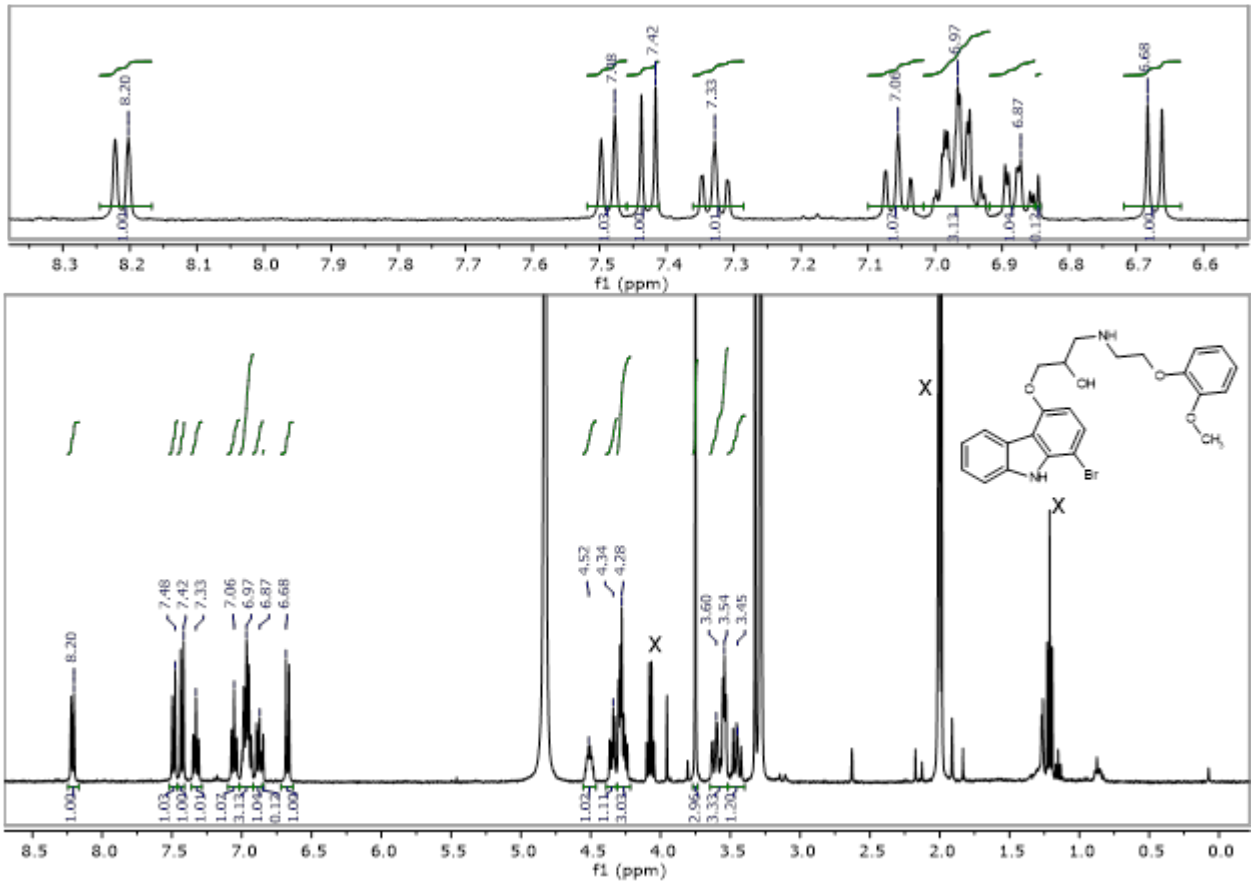
XII. NMR

A) ^{1}H-NMR

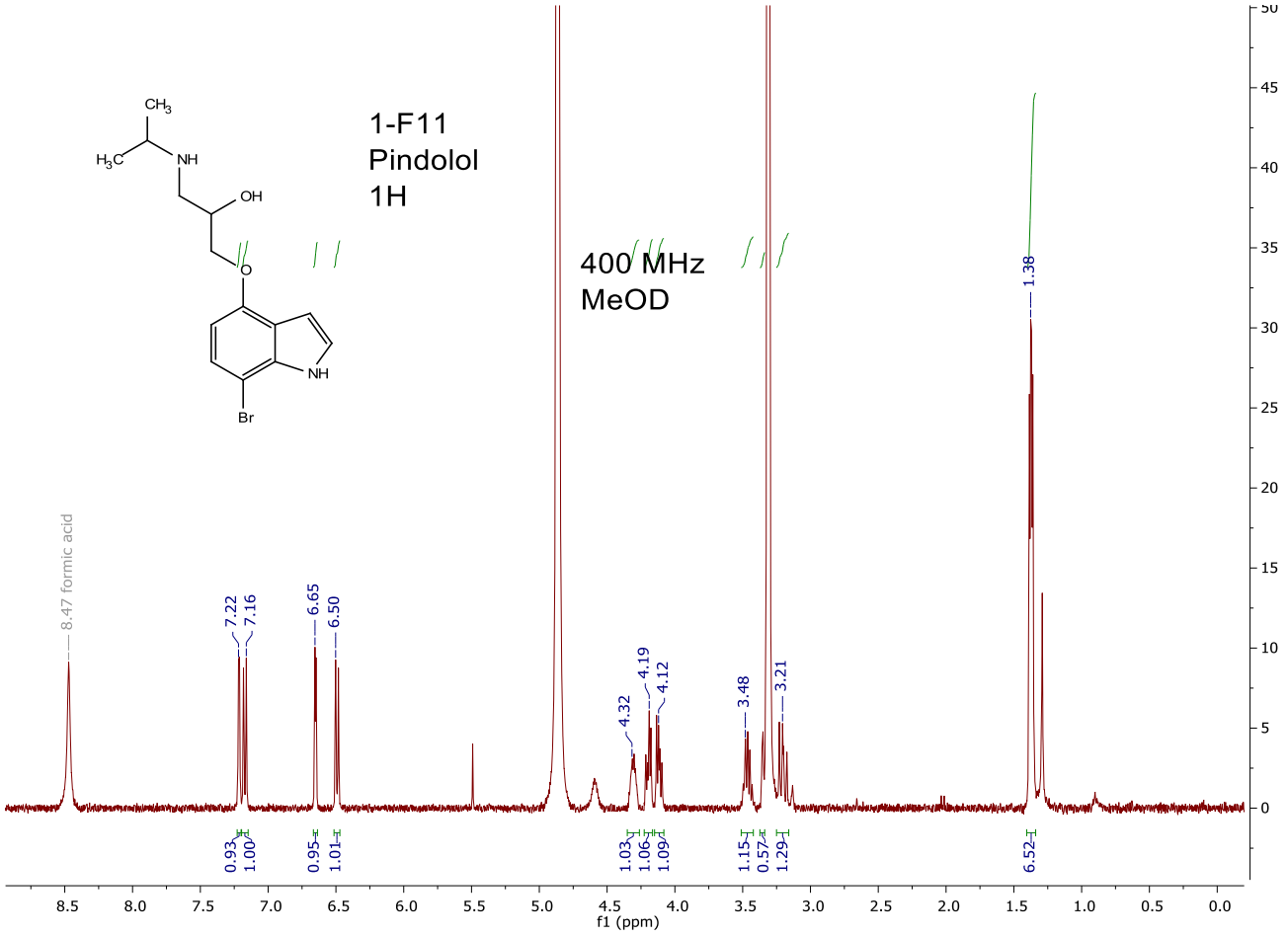
S2



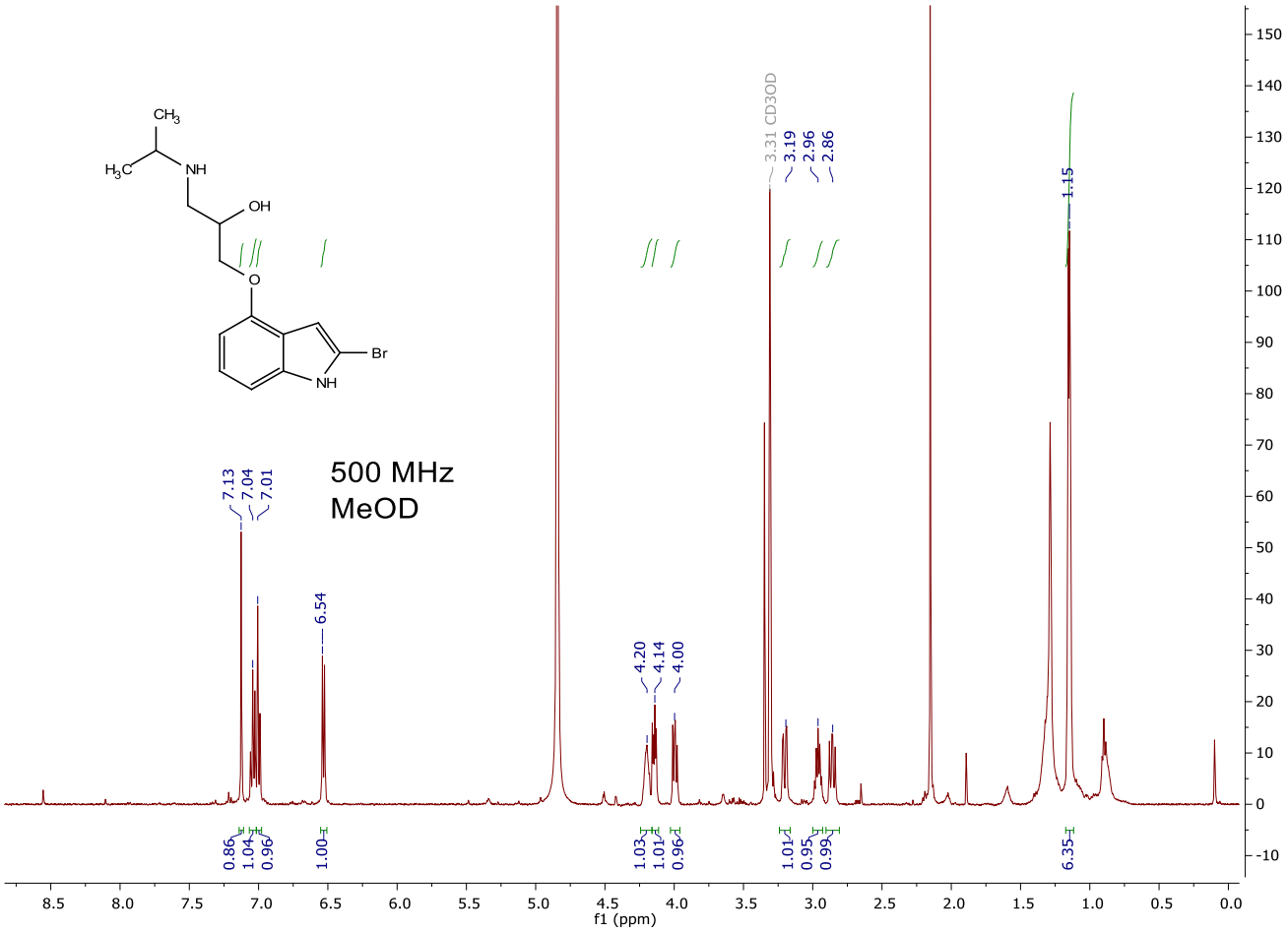
Carvedilol – 1-F11



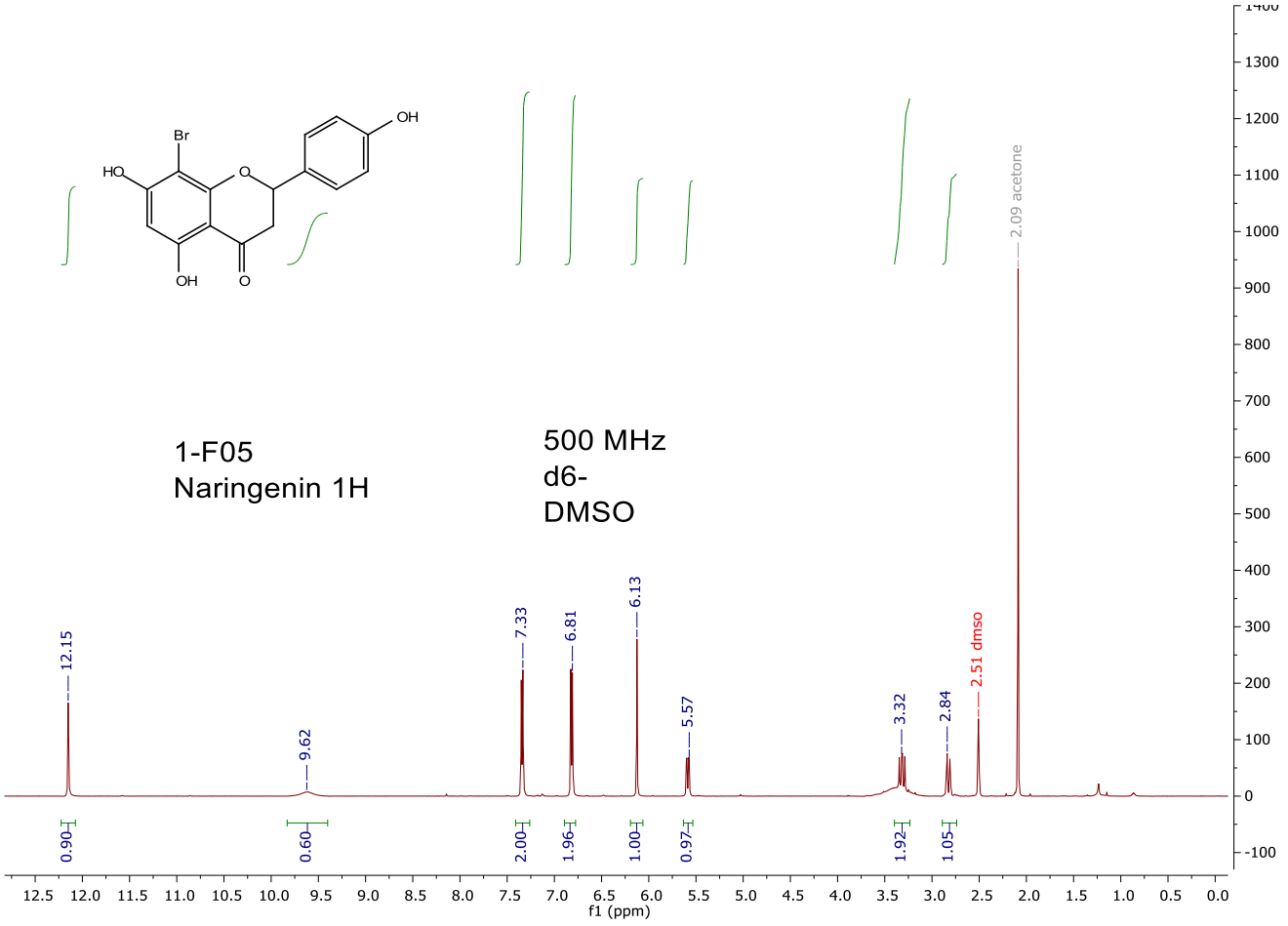
Pindolol – 1-F11



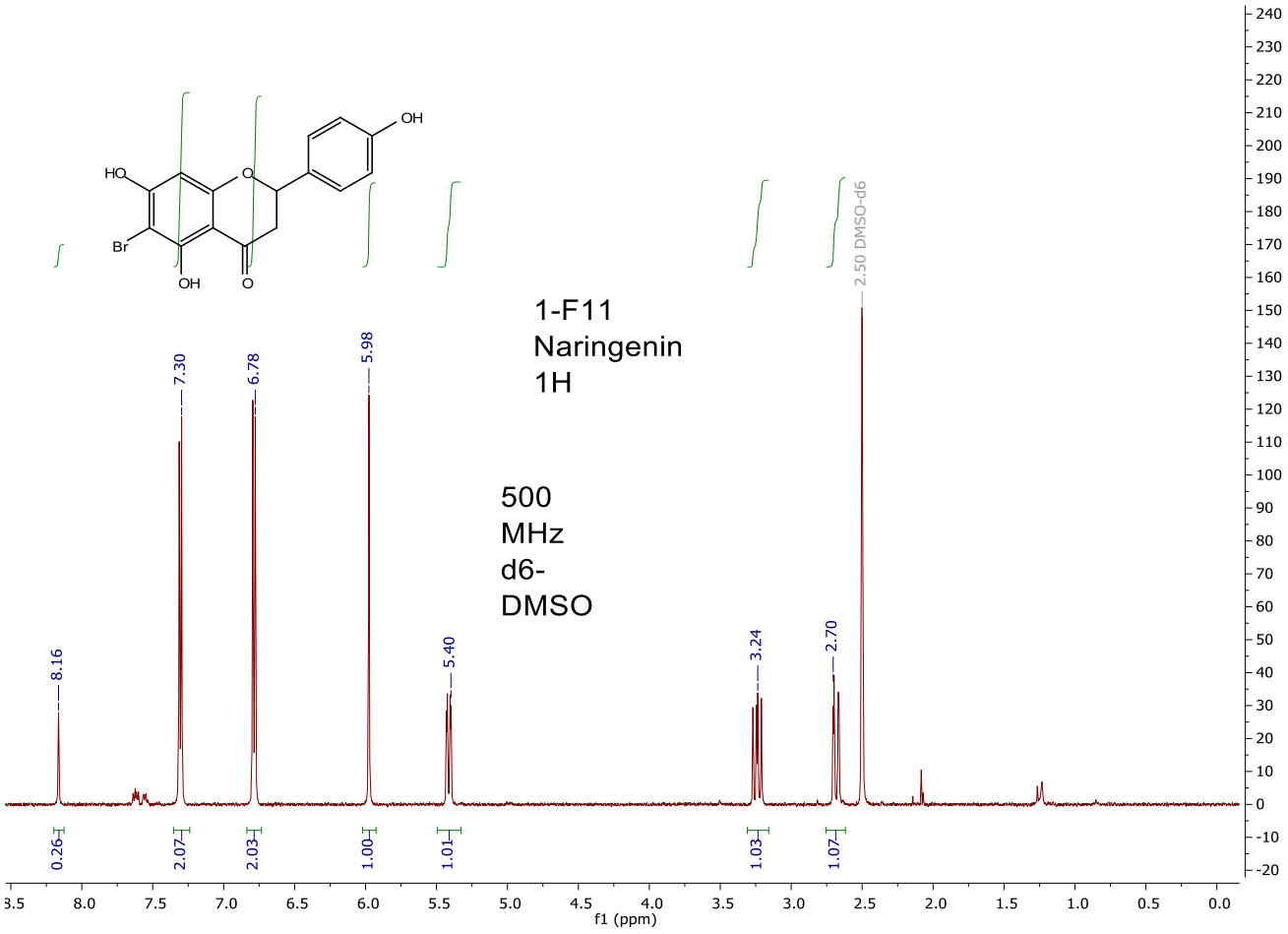
Pindolol – 2-C01



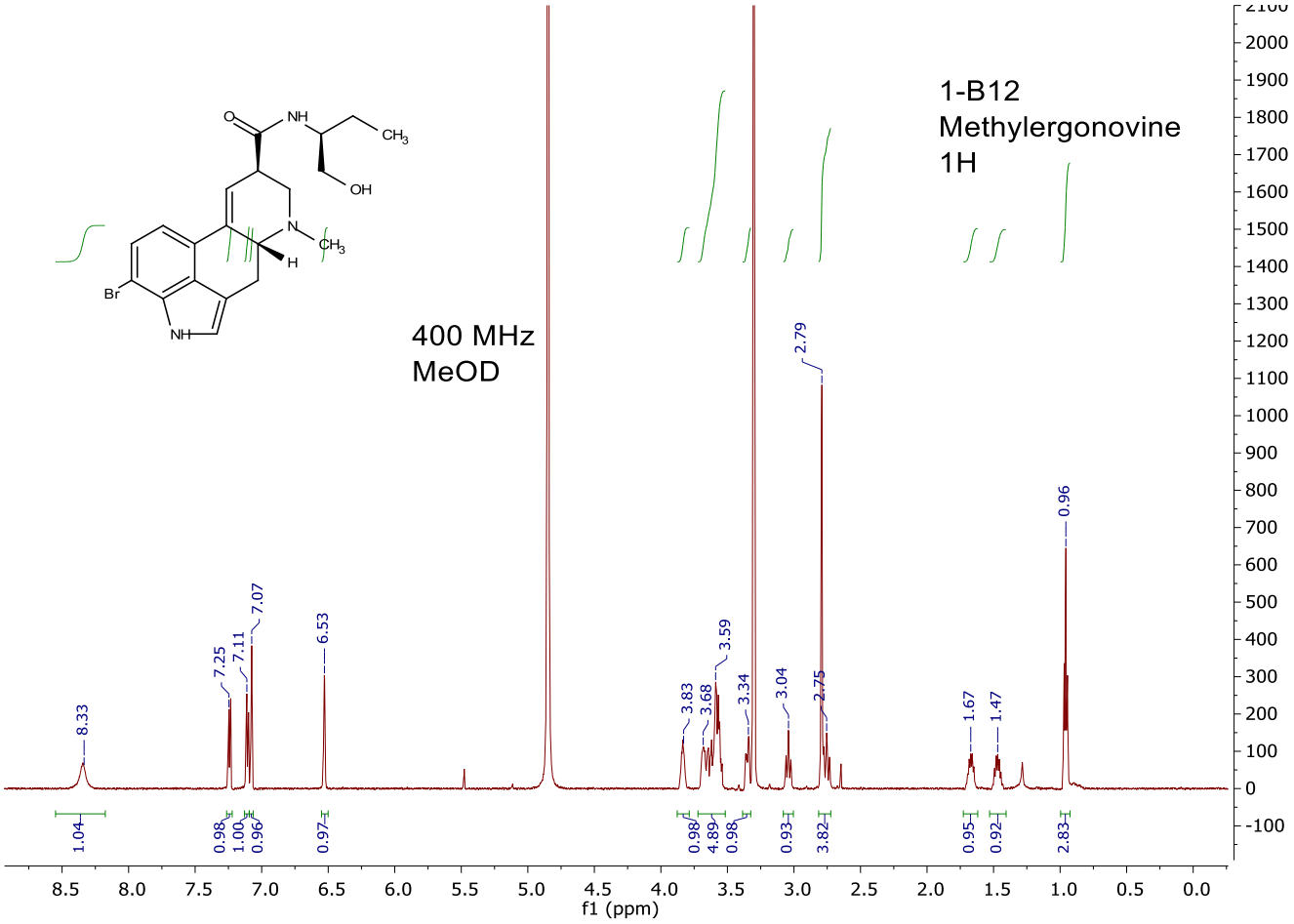
Naringenin – 1-F05



Naringenin – 1-F11

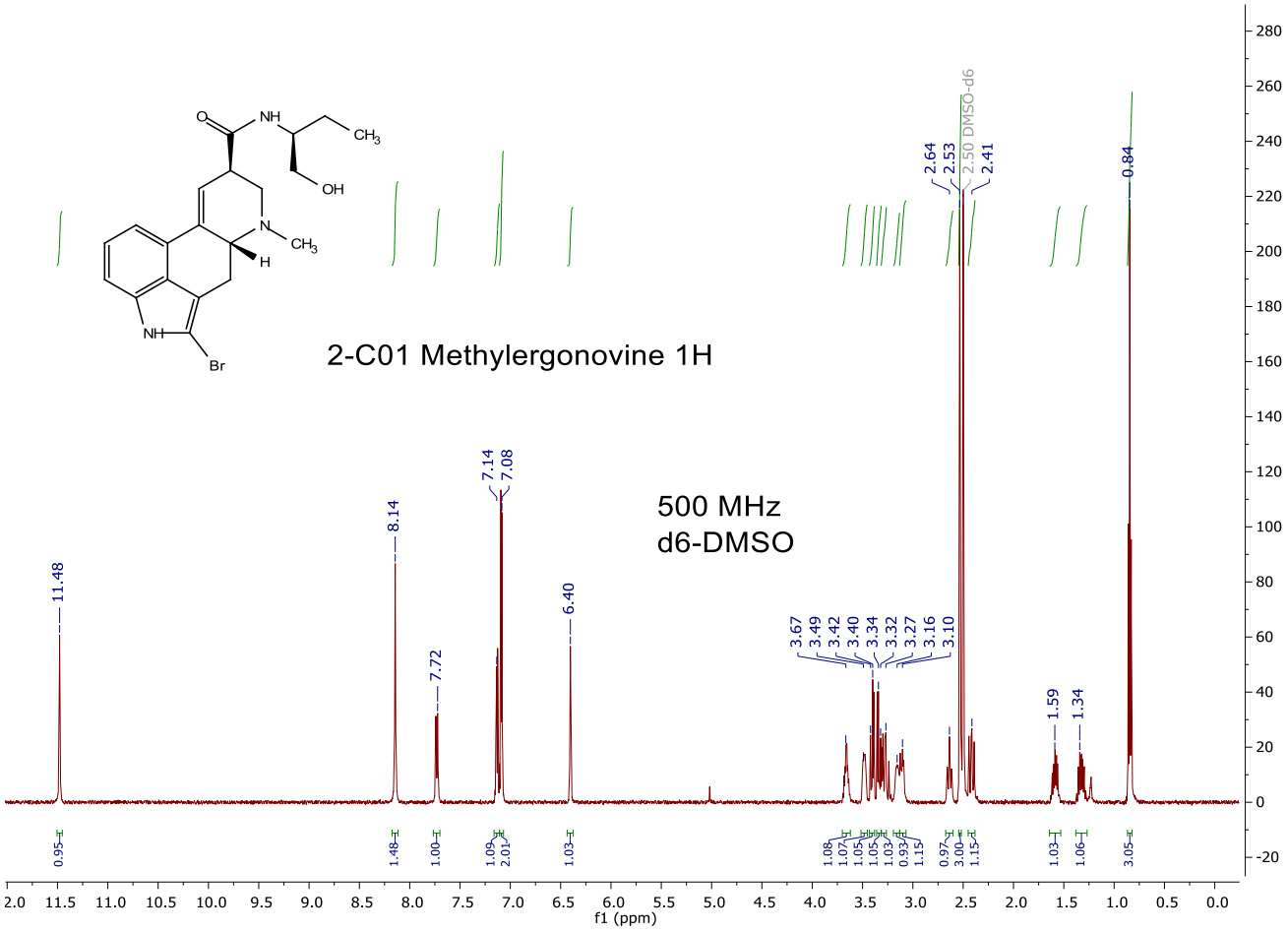


Methylergonovine – 1-B12

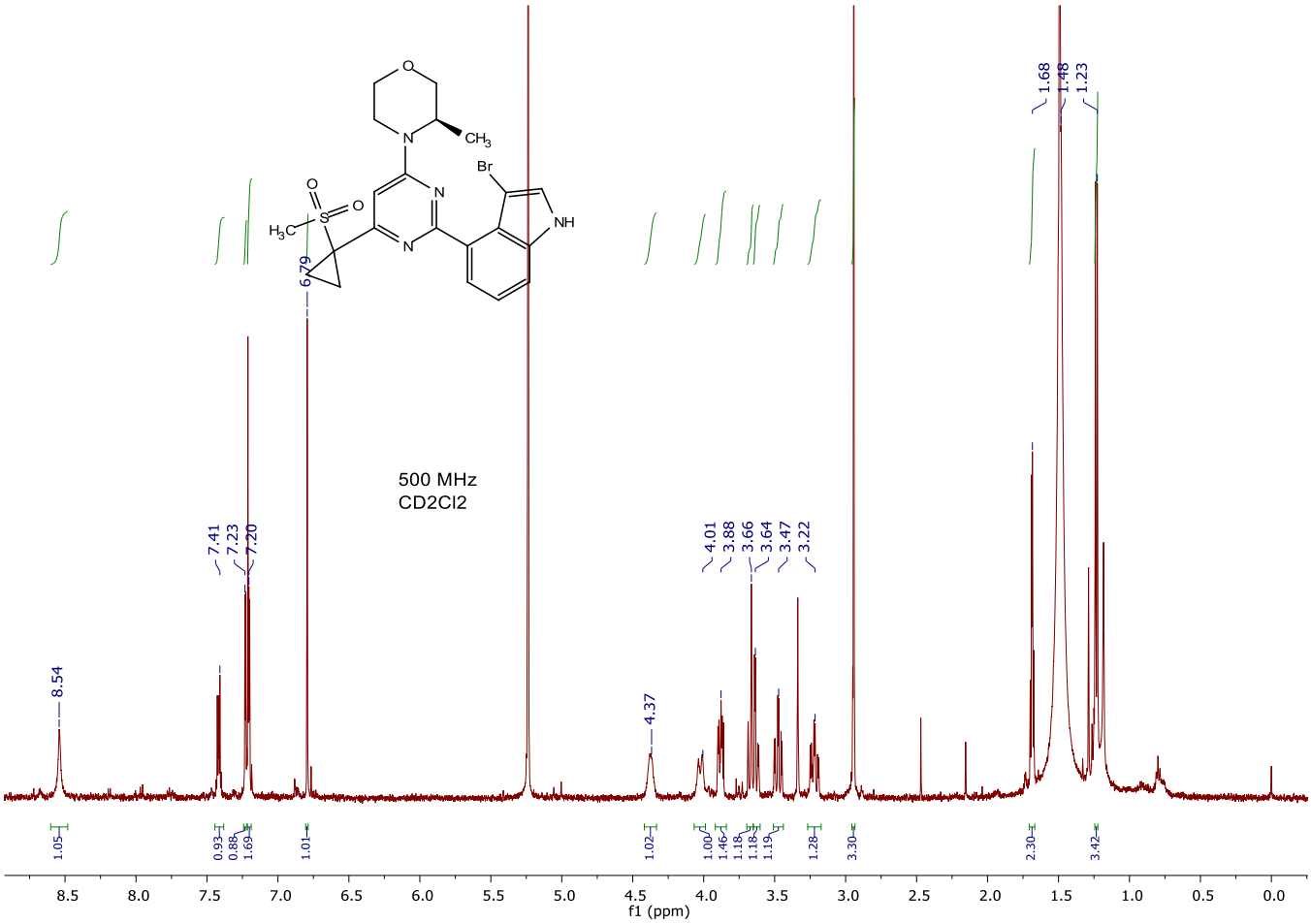


Note: the steric environment of the secondary amide appears to slow down hydrogen-deuterium exchange sufficiently that the N−H resonance appears in the ^{1}H-NMR spectrum in CD_3OD.

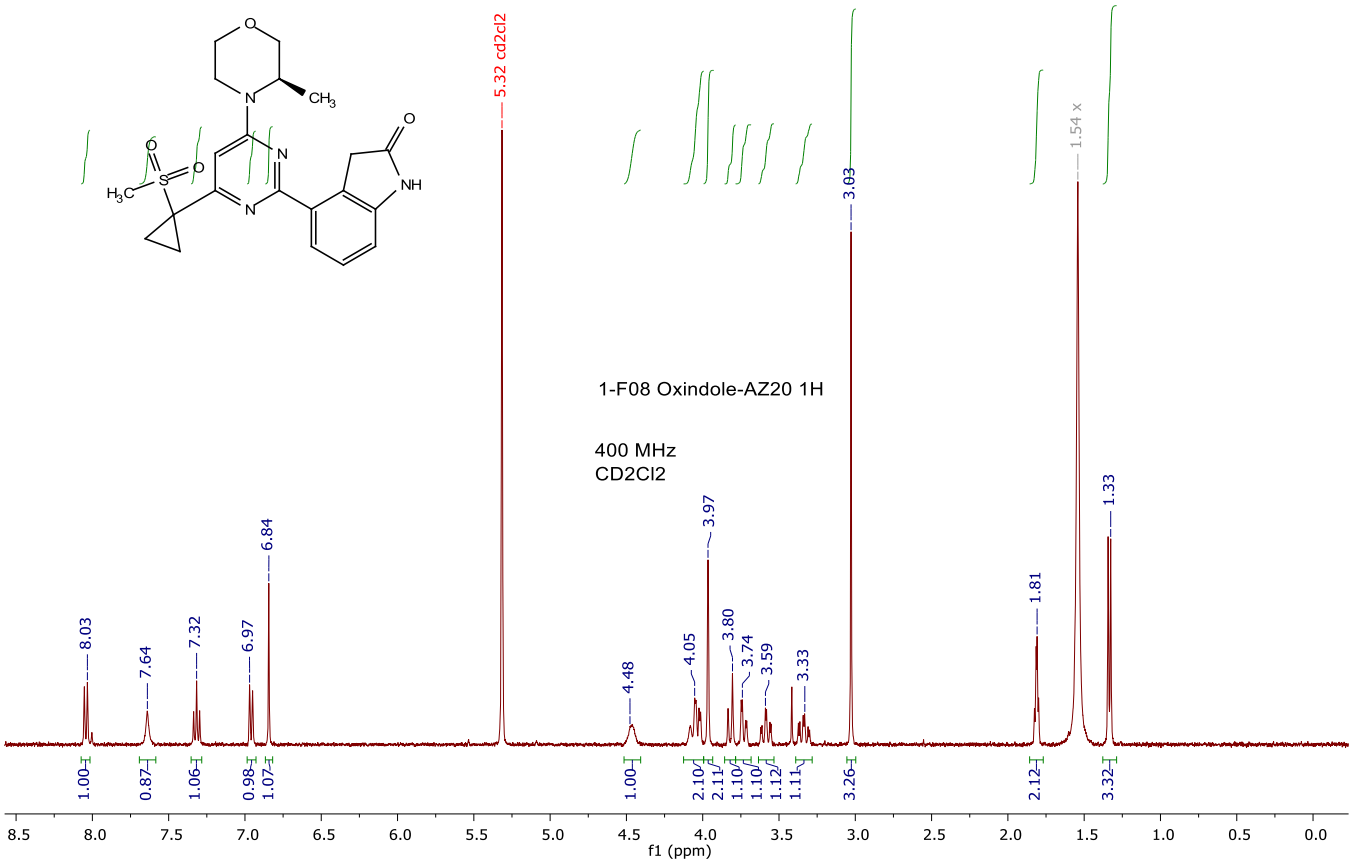
Methylergonovine – 2-C01



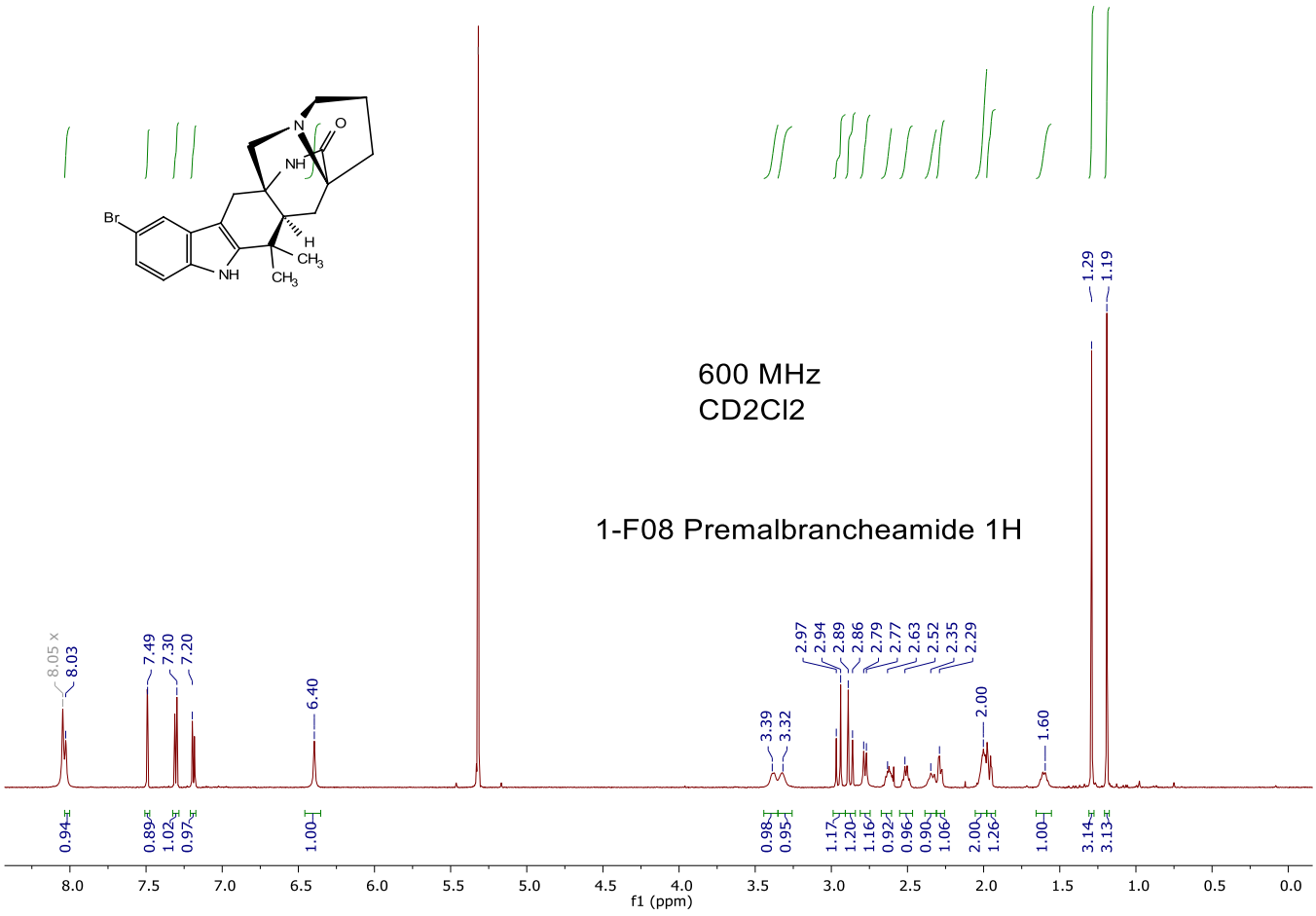
AZ20 – 1-F08


500 MHz
CD2Cl2
8.54
7.41
7.23
7.20
6.79
4.37
4.01
3.88
3.66
3.64
3.47
3.22
1.68
1.48
1.23
f1 (ppm)

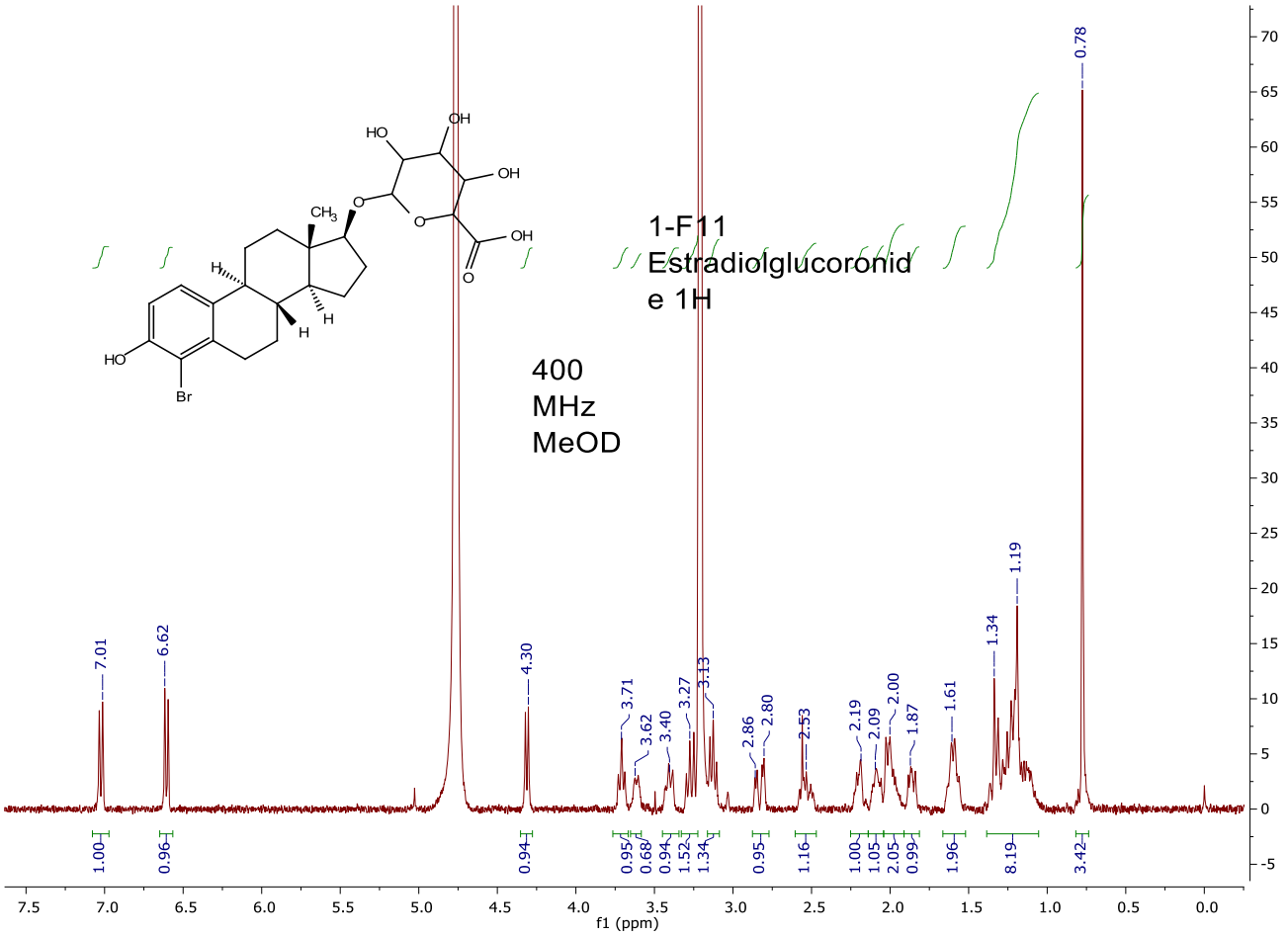
AZ20 – 1-F08 (oxindole product from C2 bromination)



Premalbrancheamide – 1-F08

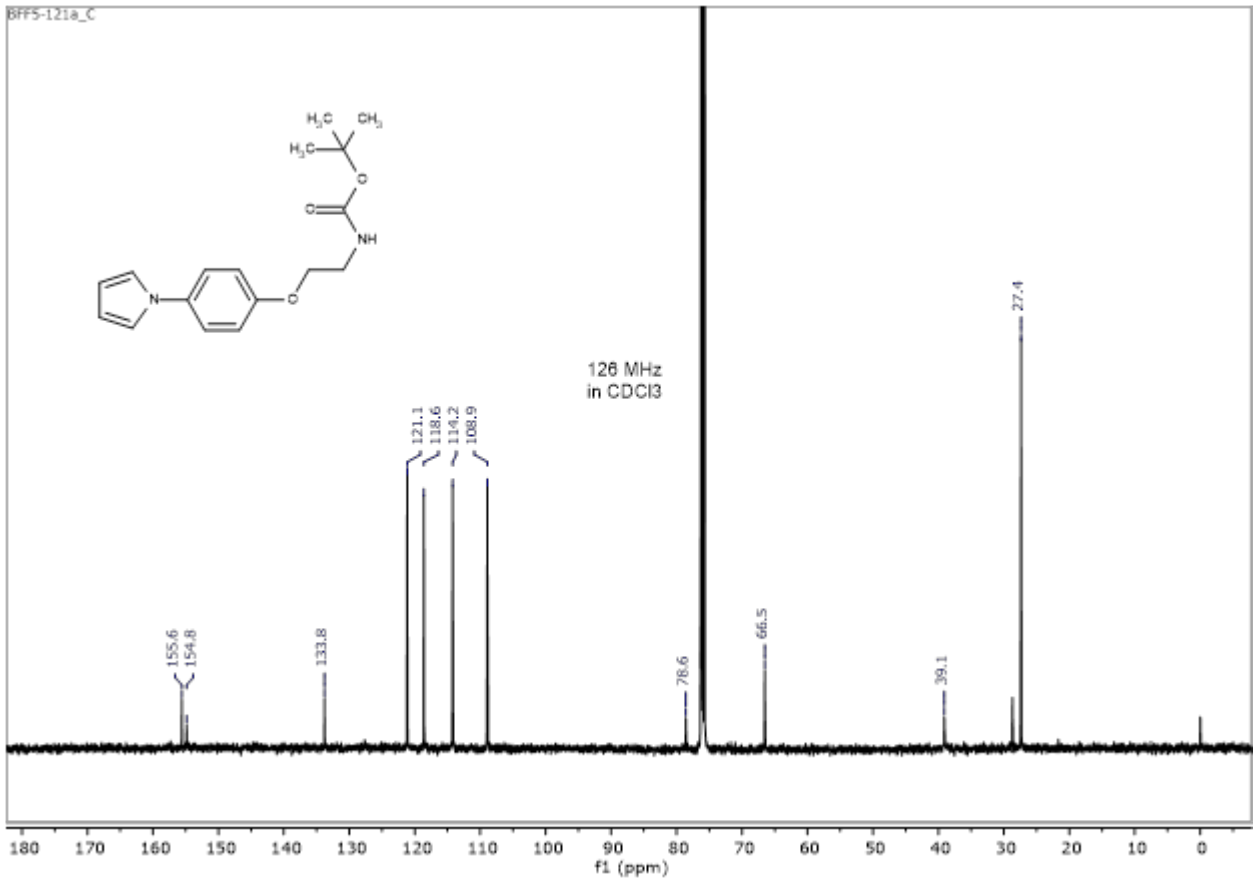


Estradiol 17β-D-glucuronide – 1-F11

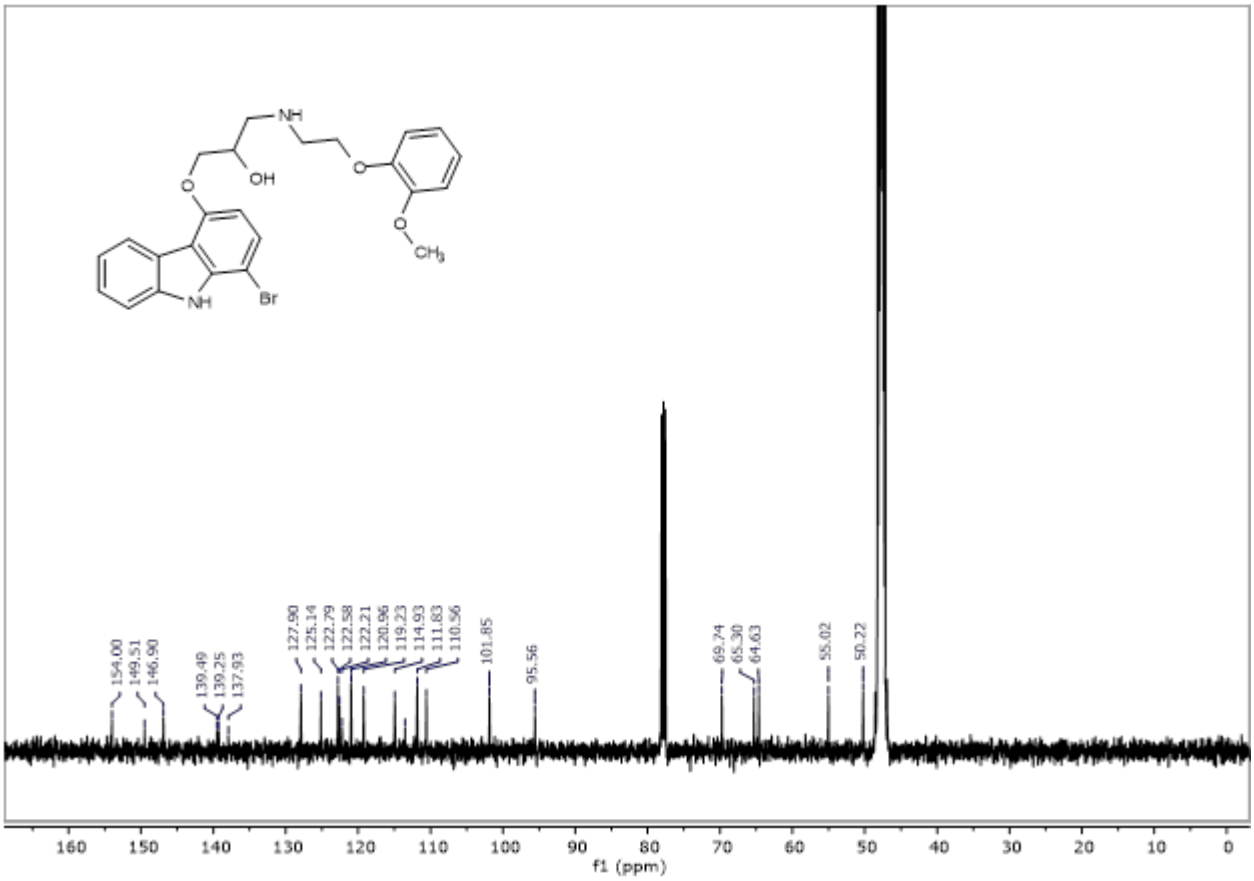


B) ^{13}C-NMR

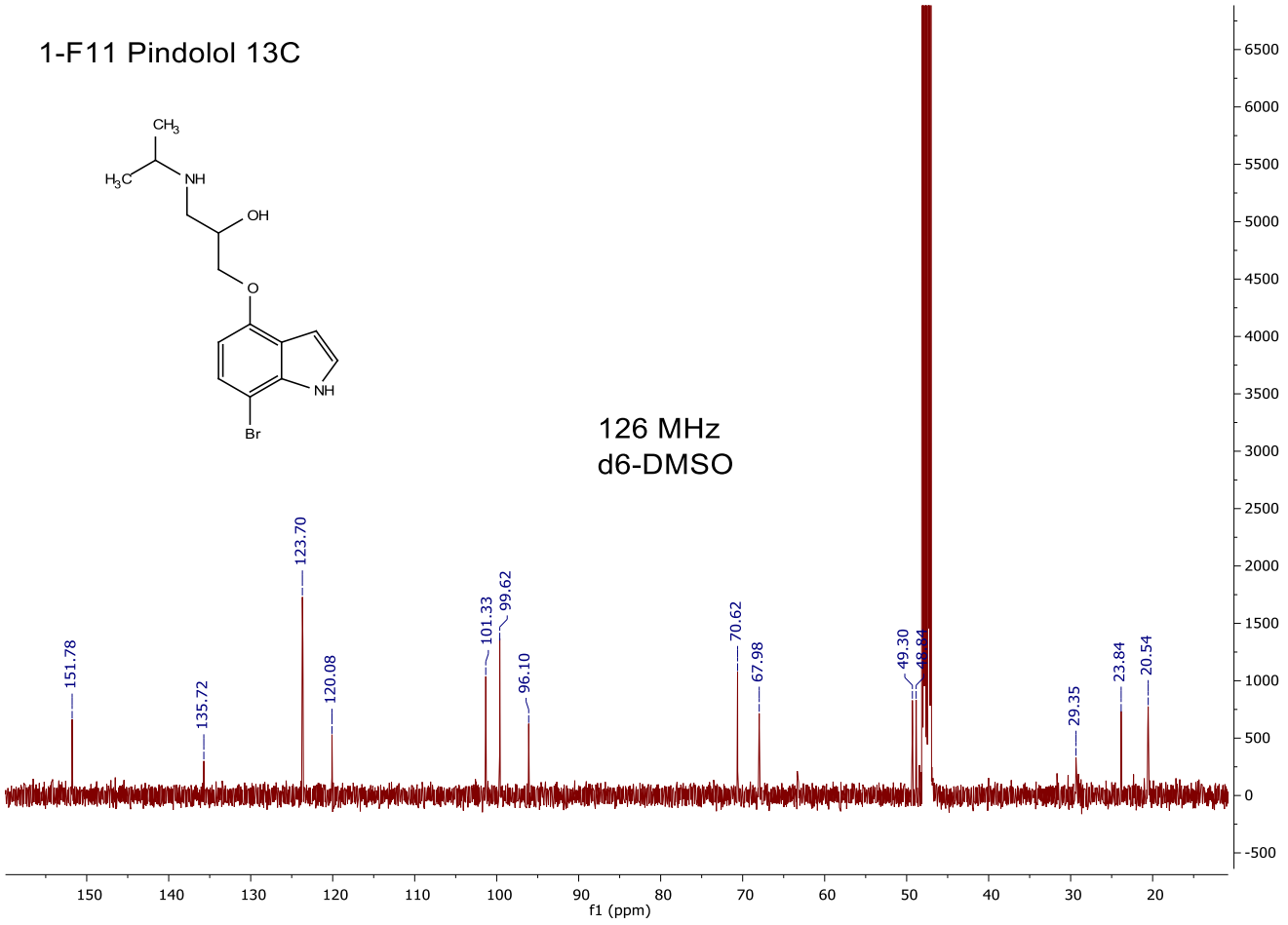
S2


BFF5-121a_C
126 MHz
in CDCl3
155.6
154.8
133.8
121.1
118.6
114.2
108.9
78.6
66.5
39.1
27.4
f1 (ppm)

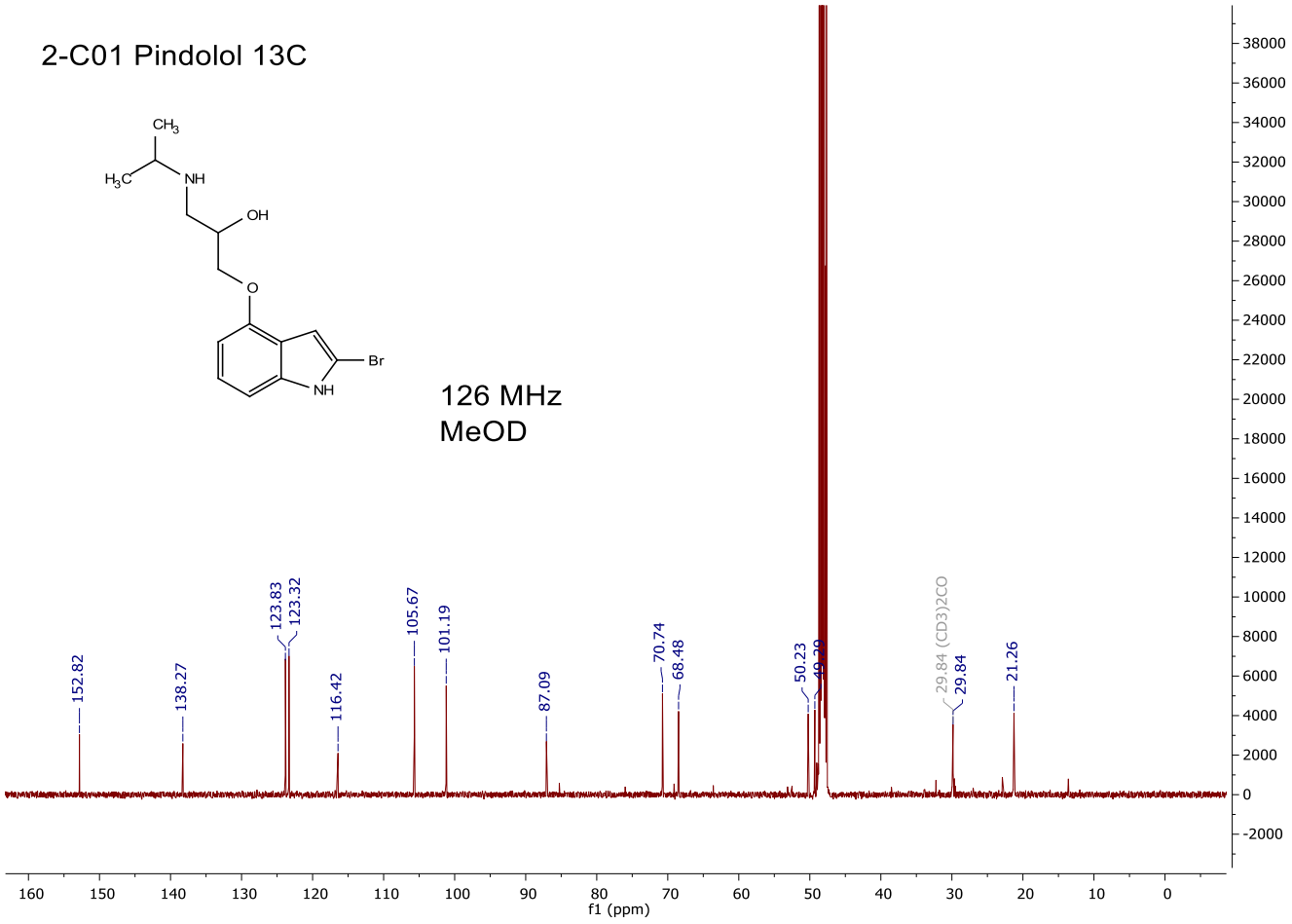
Carvedilol – 1-F11



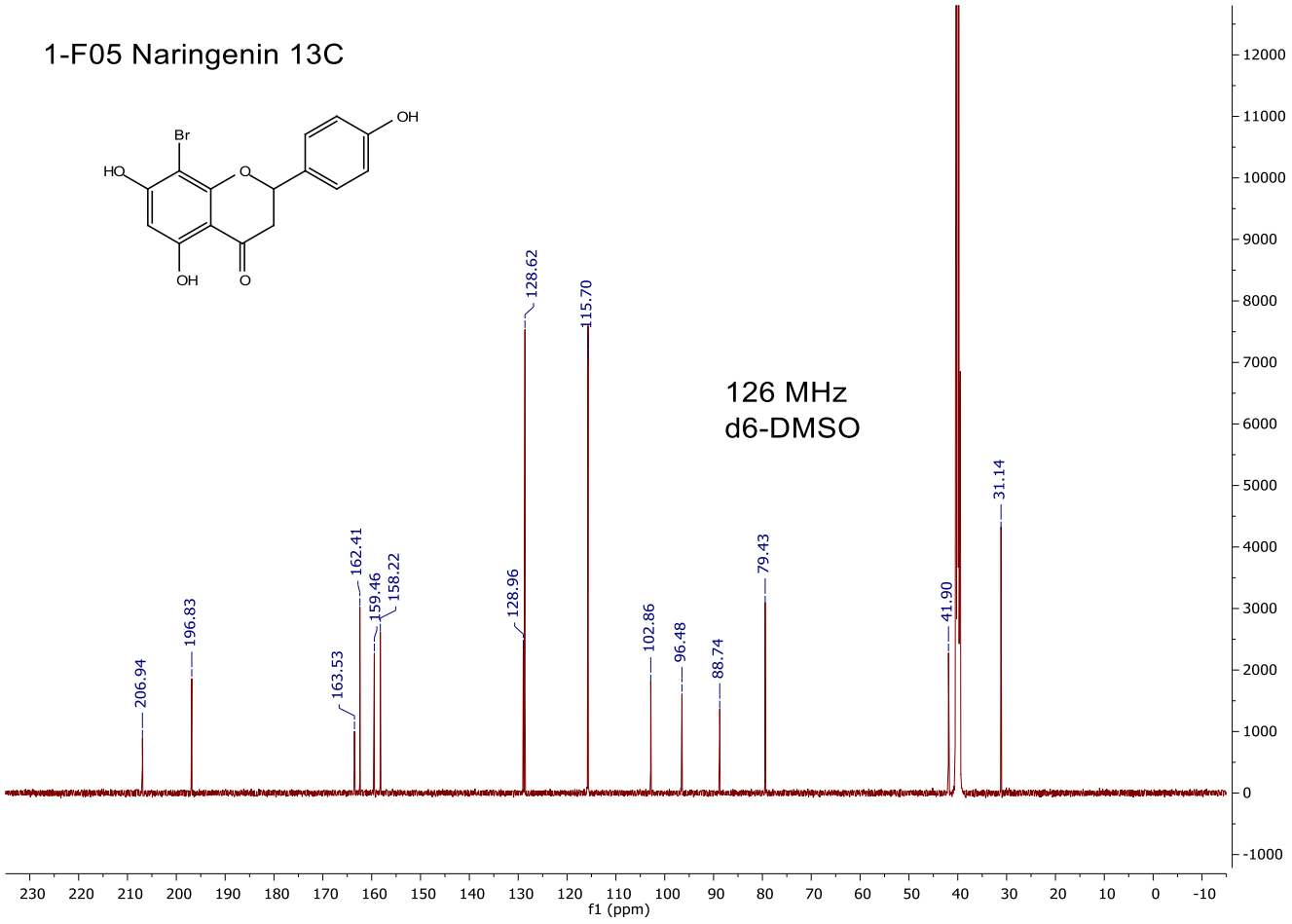
Pindolol – 1-F11



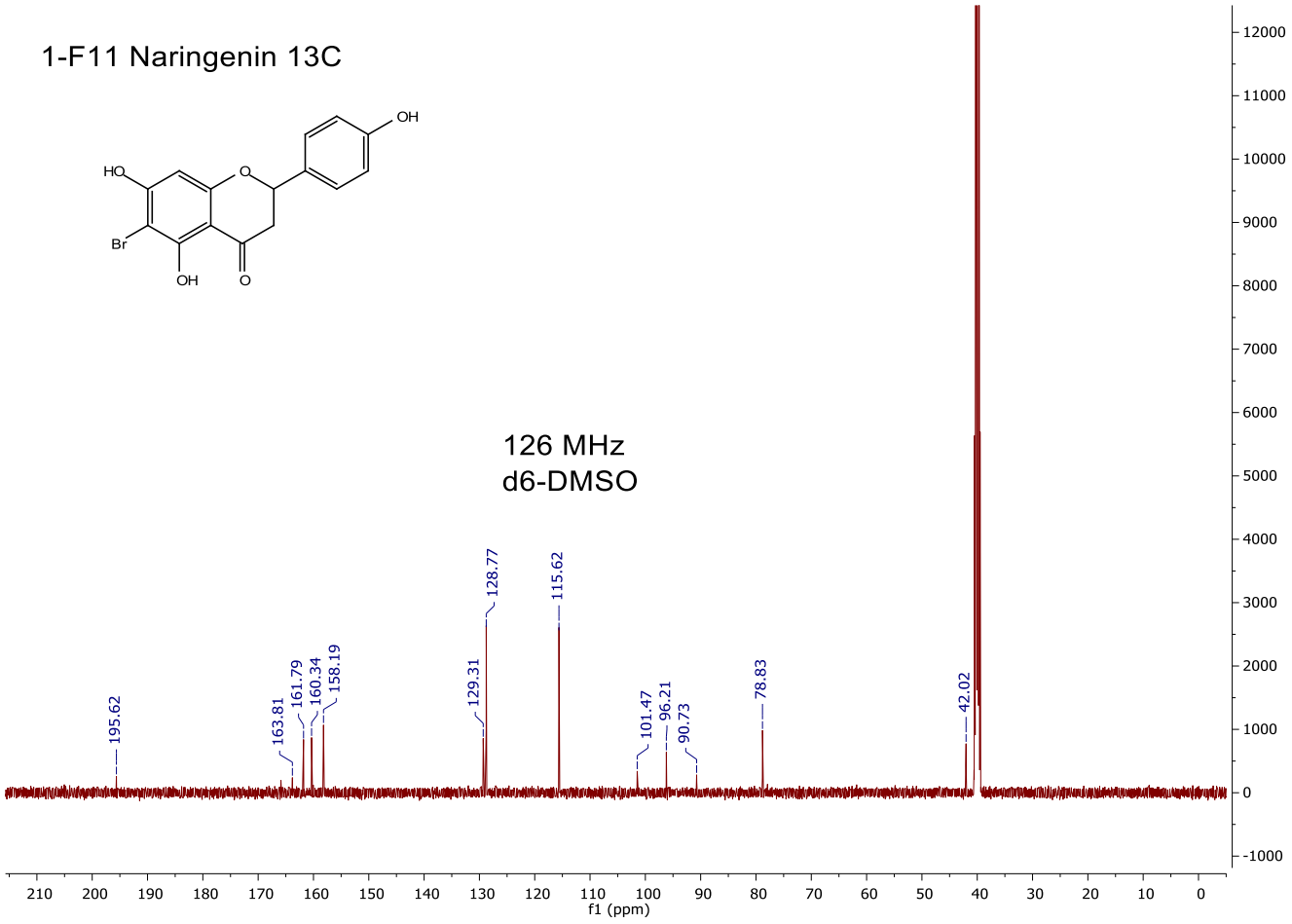
Pindolol – 2-C01



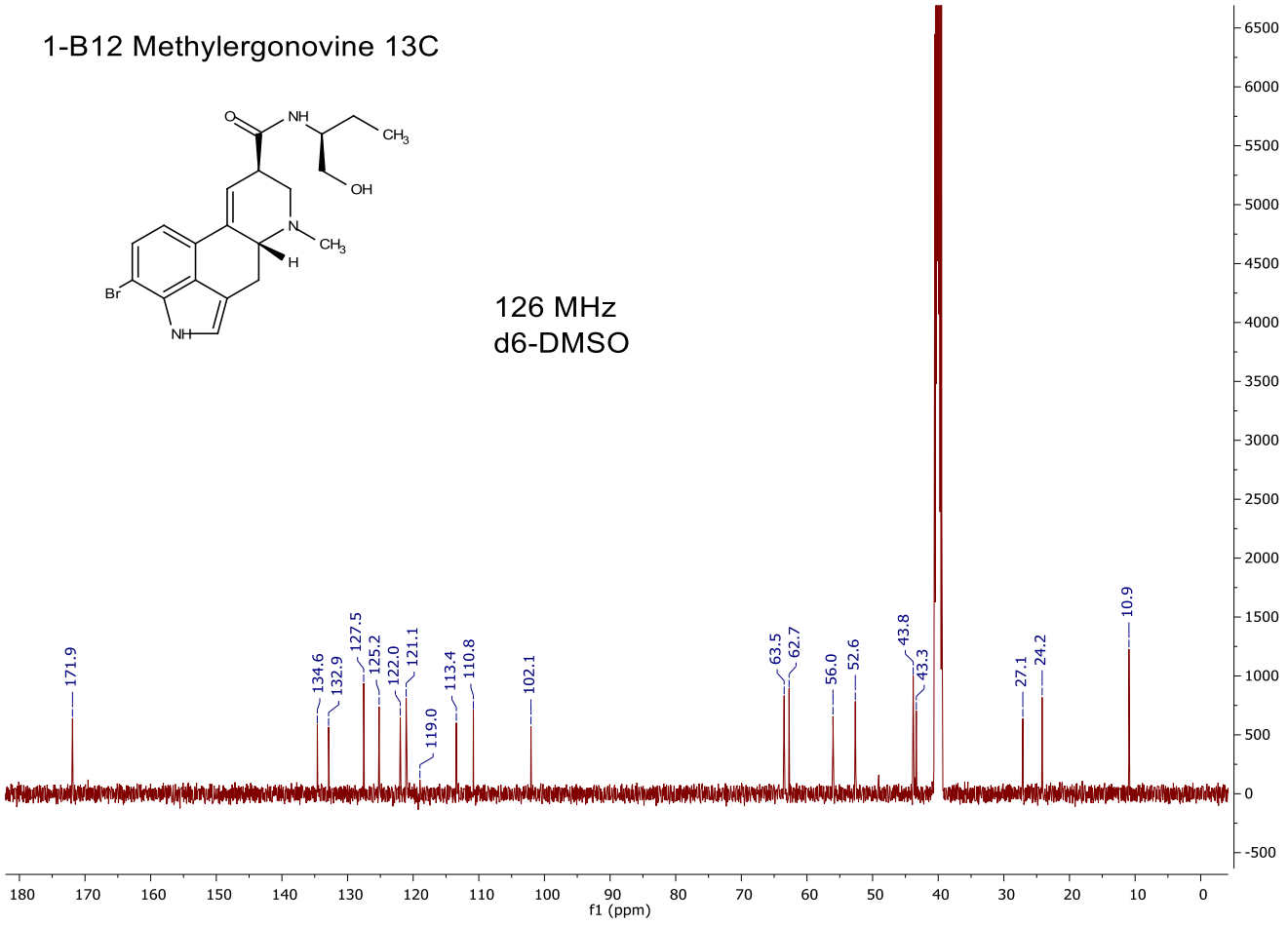
Naringenin – 1-F05



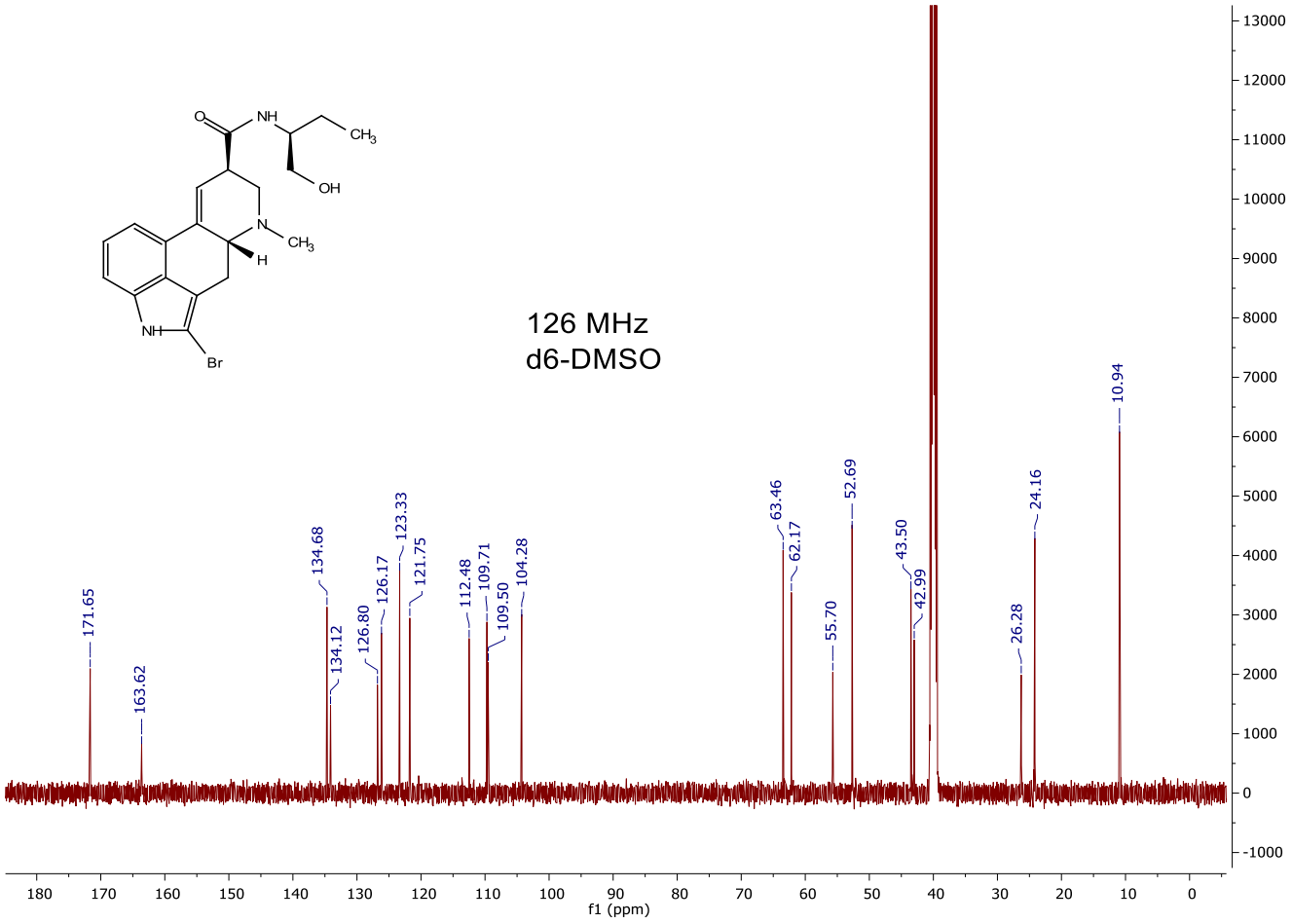
Naringenin – 1-F11



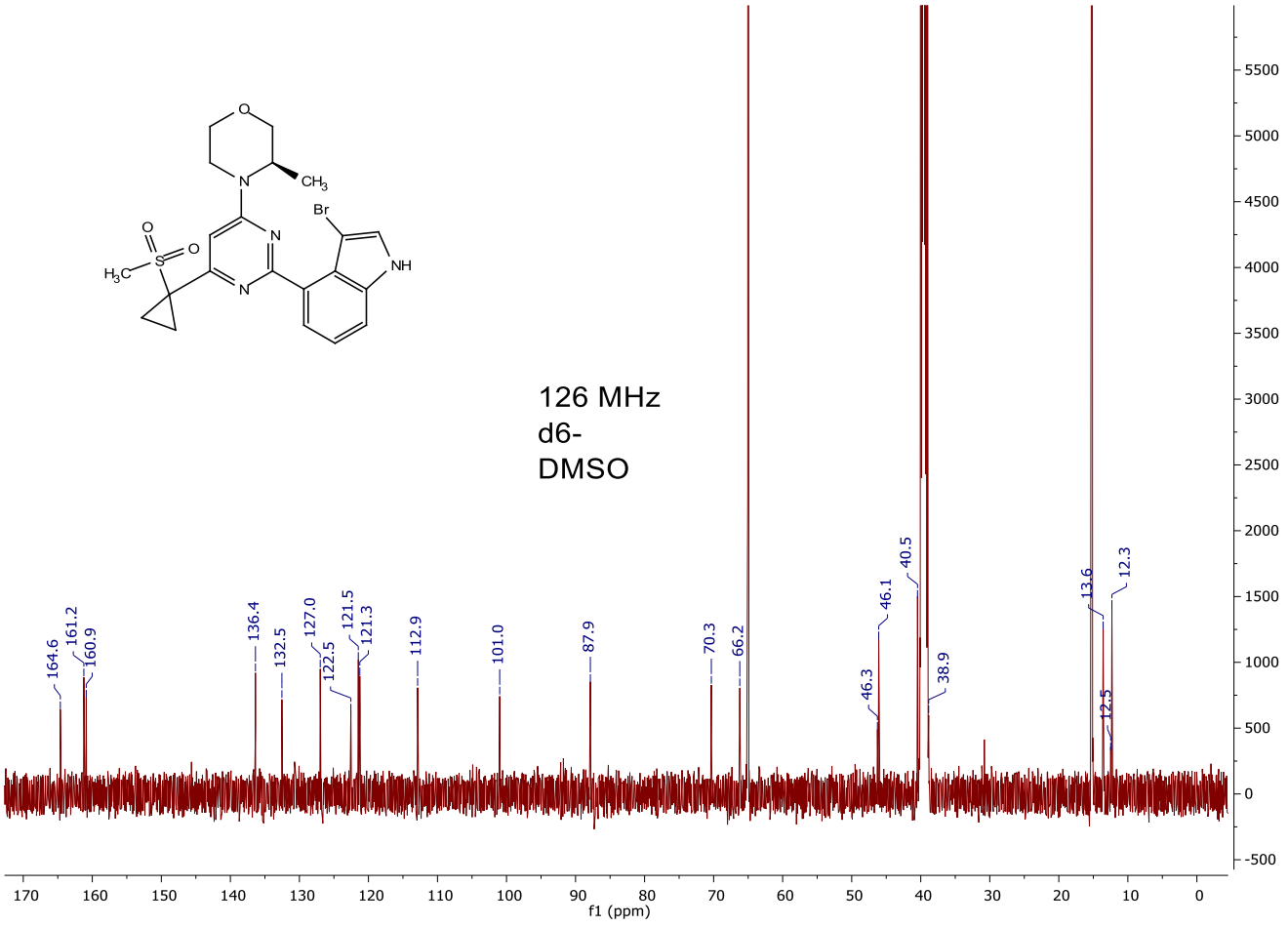
Methylergonovine – 1-B12



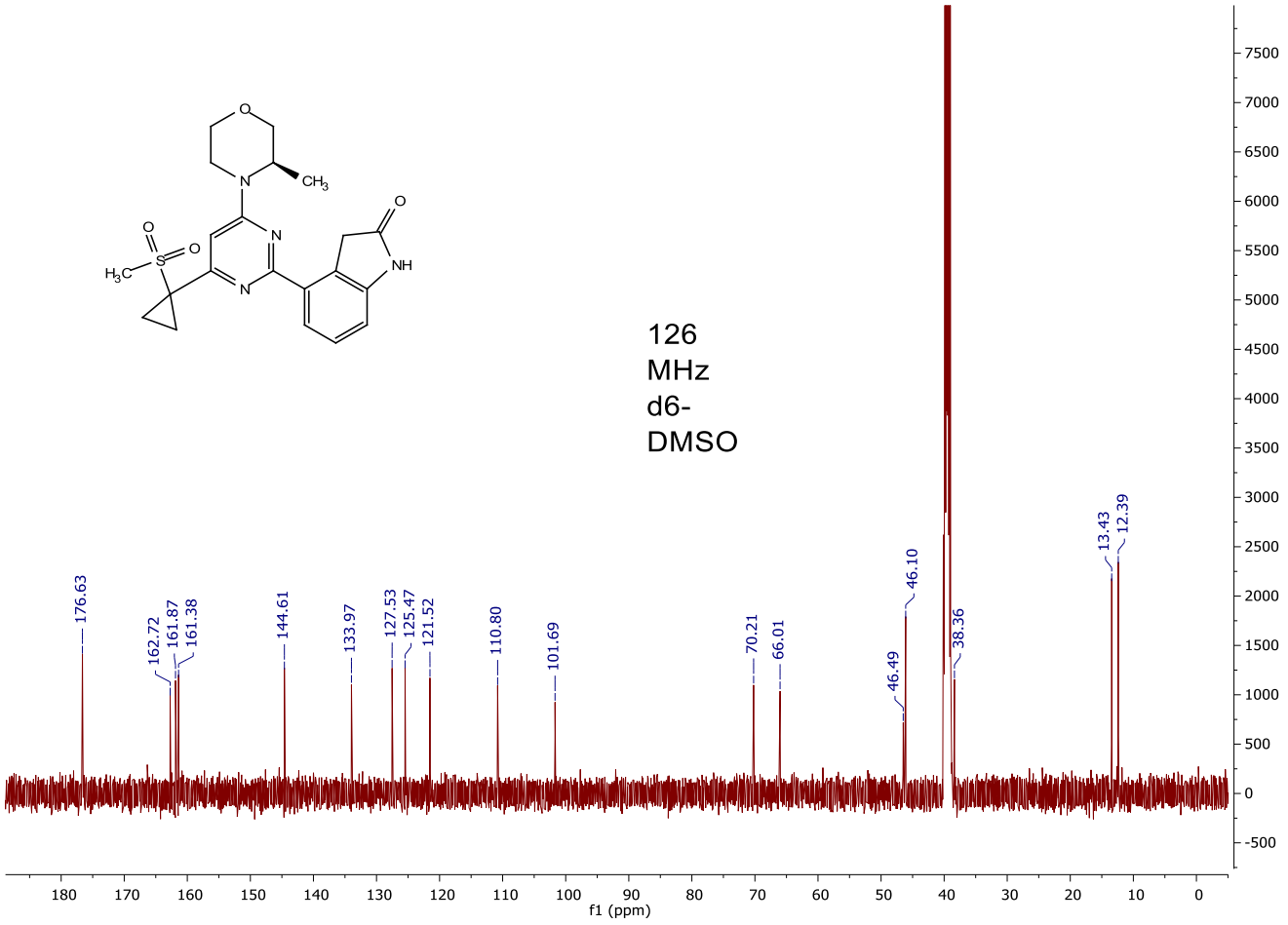
Methylergonovine – 2-C01



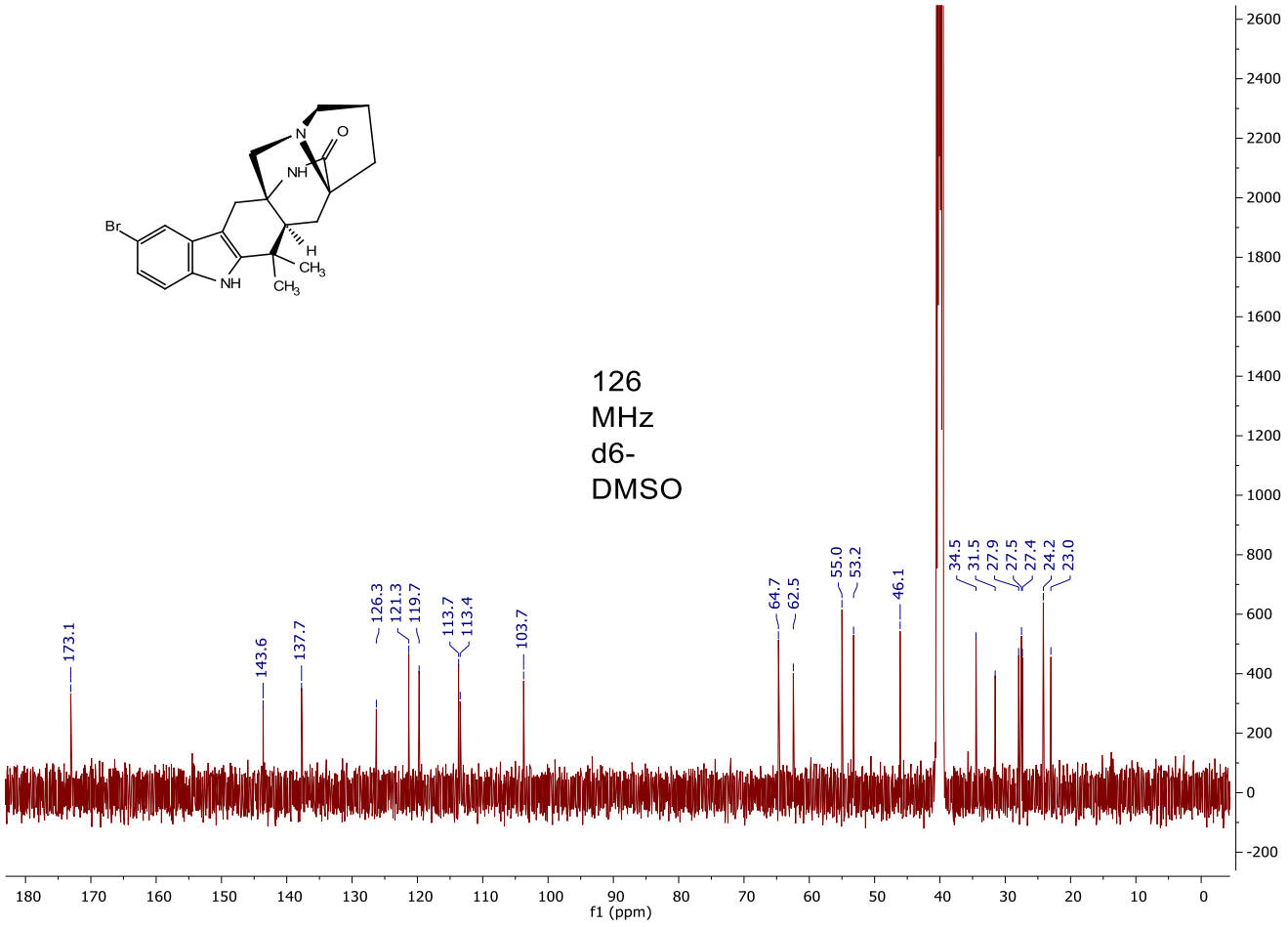
AZ20 – 1-F08


126 MHz
d6-
DMSO
164.6
161.2
160.9
136.4
132.5
127.0
122.5
121.5
121.3
112.9
101.0
87.9
70.3
66.2
46.3
46.1
40.5
38.9
13.6
12.5
12.3
f1 (ppm)

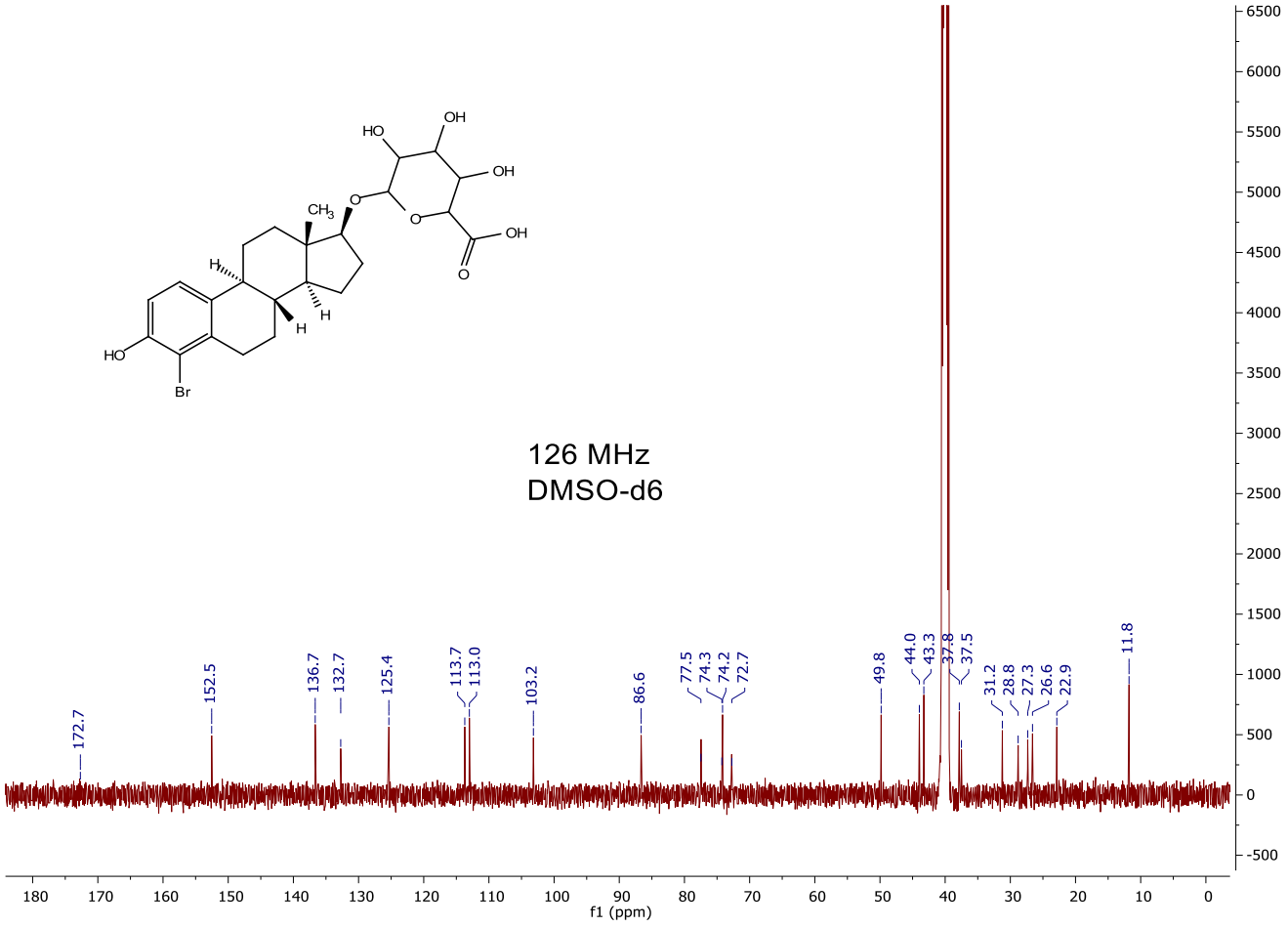
AZ20 – 1-F08 (oxindole product from C2 bromination)



Premalbrancheamide – 1-F08

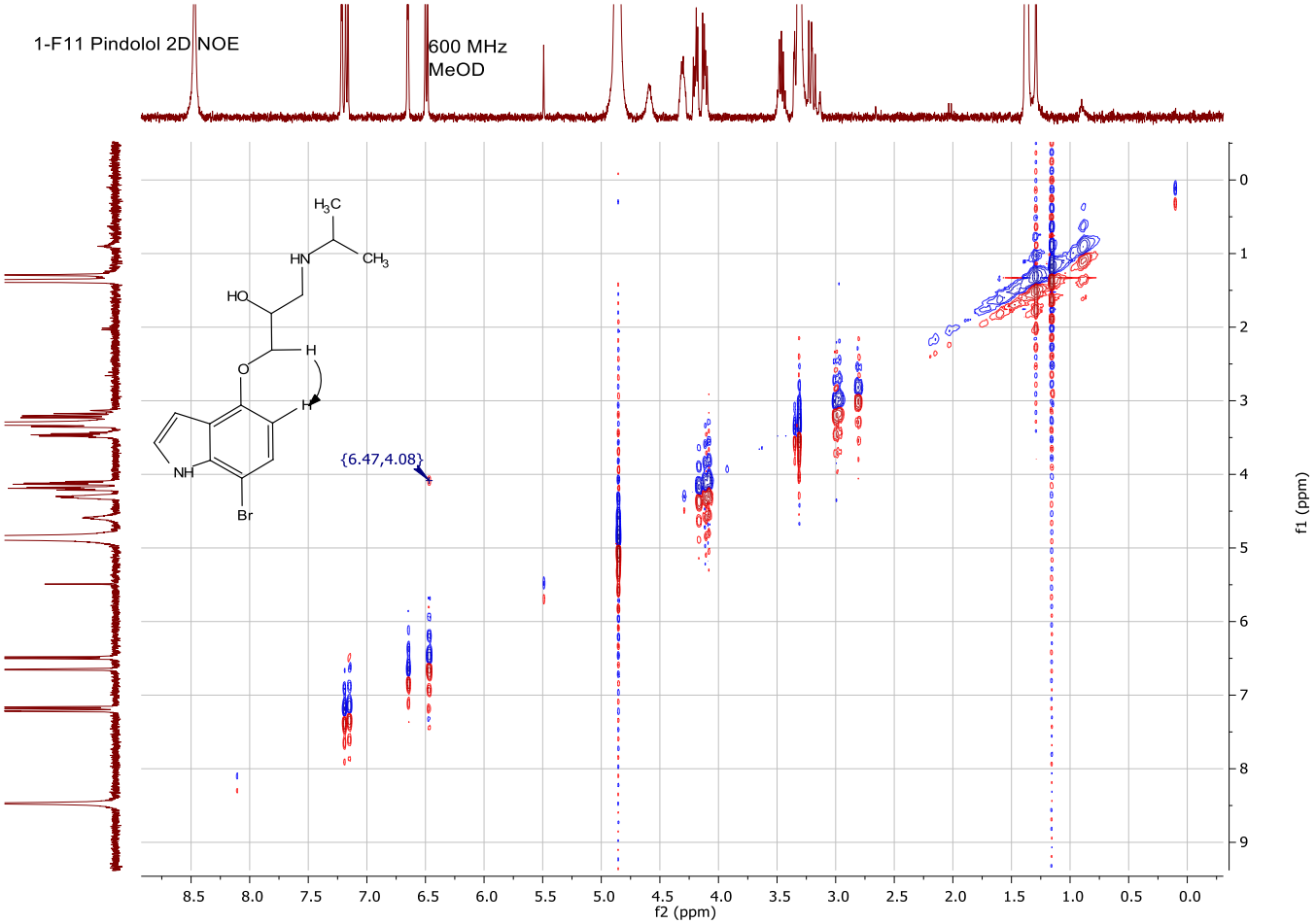


Estradiol 17β-D-glucuronide – 1-F11



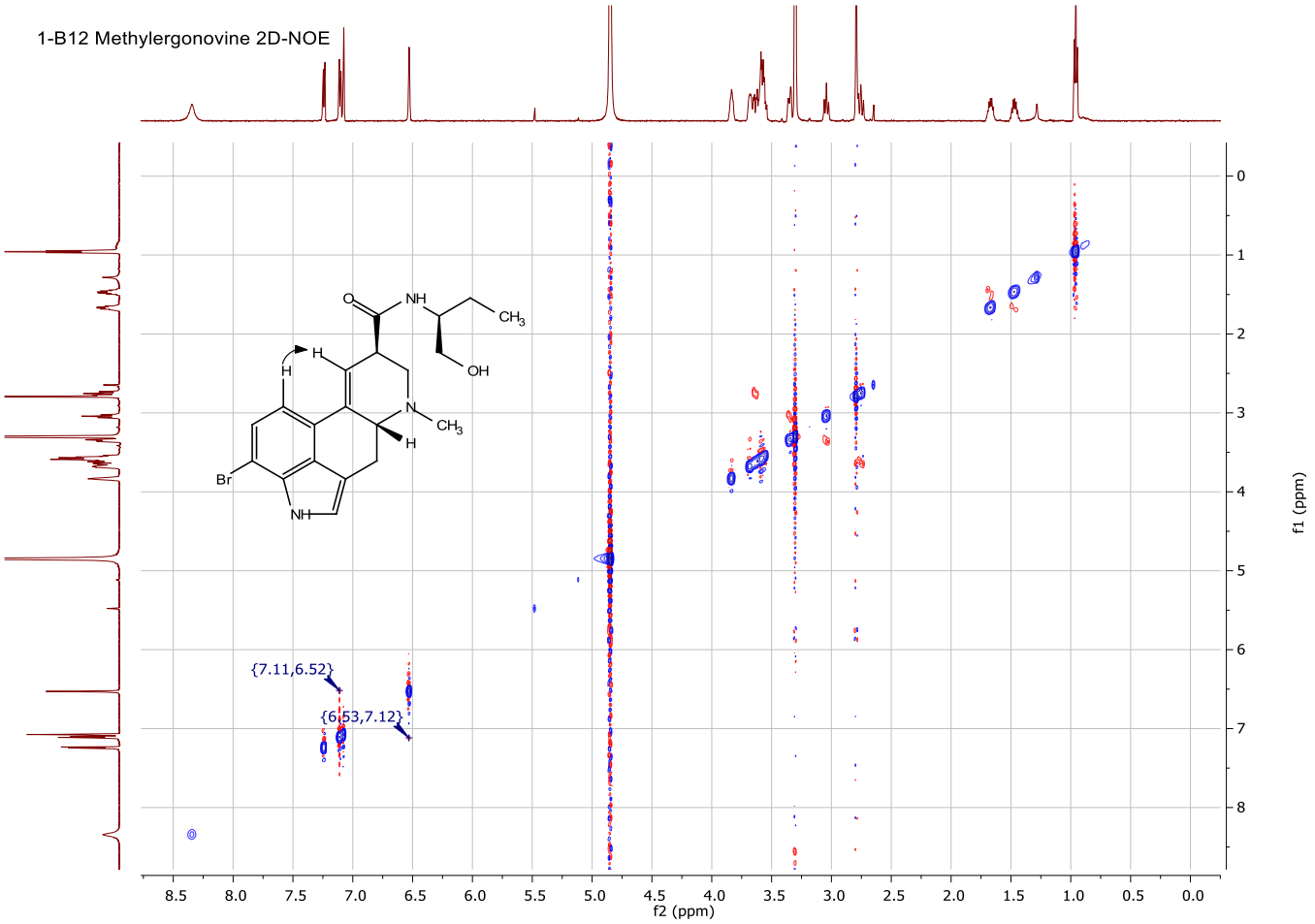
C) 2D-NMR Spectra

Pindolol – 1-F11



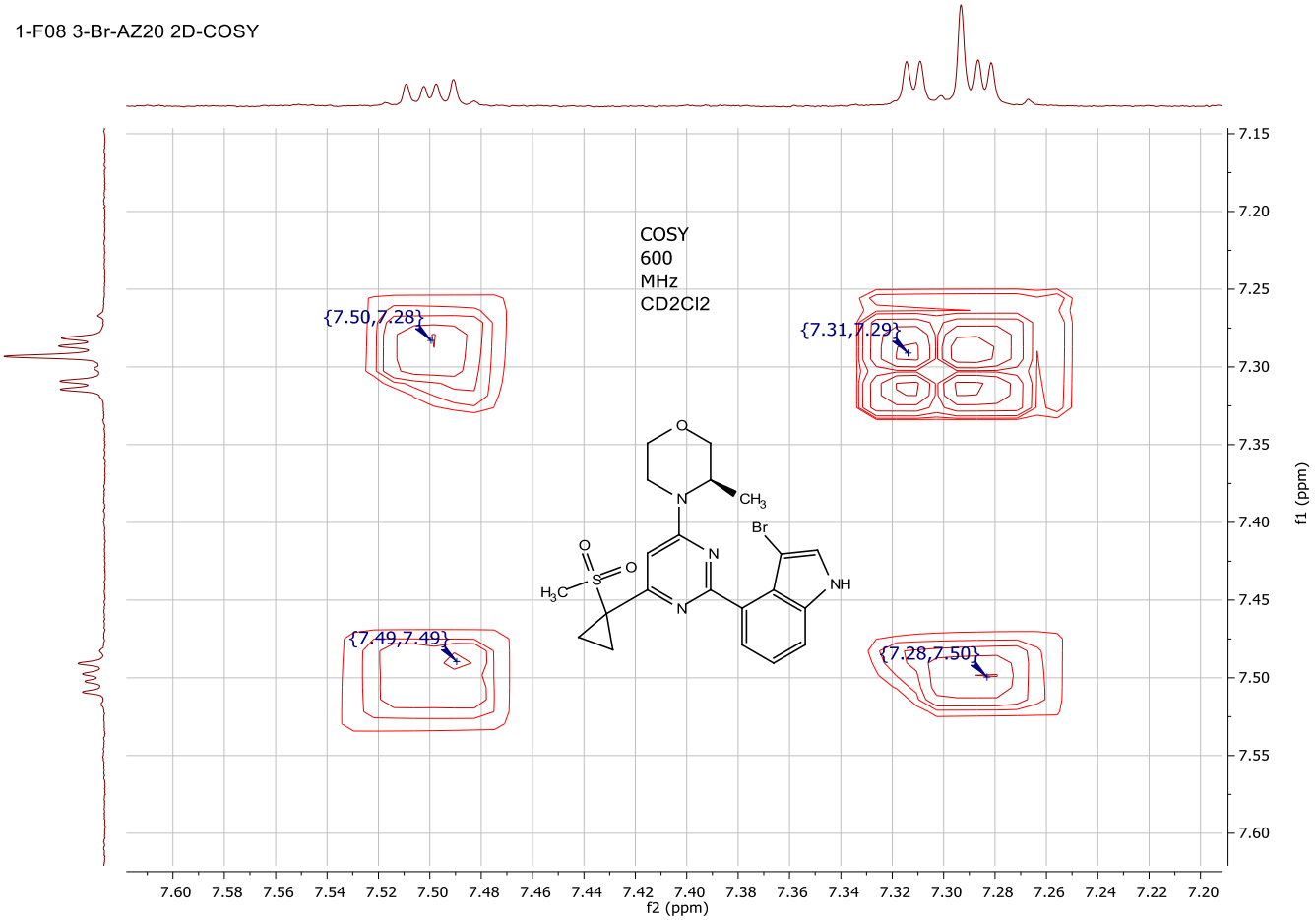
NOE shown above, in combination with aryl multiplet patterns, was used to assign regiochemistry of the bromination reaction.

Methylergonovine – 1-B12



The NOESY crosspeak shown above, in combination with multiplet patterns in the ^{1}H-NMR, were used to assign bromination regiochemistry.

AZ20 – 1-F08



There is a doublet of doublets at 7.50 ppm; this is consistent with *ortho* coupling (5.8 Hz) and *meta* coupling (3.4 Hz), meaning that the bromination could not have occurred on the benzo ring of the indole moiety. Also, one aryl resonance at 7.29 ppm has COSY crosspeaks with two other aryl resonances at 7.31 ppm and 7.50 ppm, and although the splitting pattern is obscured by chemical shift overlap, this resonance must be from the aryl hydrogen at the 6-position of the indole.

XIII. References

1. Payne, J. T.; Andorfer, M. C.; Lewis, J. C., *Methods Enzymol.* **2016,** *575*, 93.

2. Gentleman, R. C.; Carey, V. J.; Bates, D. M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A. J.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J. Y.; Zhang, J., *Genome Biology* **2004,** *5*, R80.

3. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L., *BBA – Proteins and Proteomics* **2015,** *1854*, 1019.

4. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W., *Bioinformatics* **2012,** *28*, 3150.

5. Gill, S. C.; Hippel, P. H., *Anal. Biochem.* **1989,** *182*, 319.

6. *Cold Spring Harbor Protocols* **2015,** *2015*, pdb.rec087908.

7. Ladner, C. L.; Yang, J.; Turner, R. J.; Edwards, R. A., *Anal. Biochem.* **2004,** *326*, 13.

8. *Cold Spring Harbor Protocols* **2007,** *2007*, pdb.rec10727.

9. *Cold Spring Harbor Protocols* **2007,** *2007*, pdb.rec10717.

10. Wenig, P.; Odermatt, J., *BMC Bioinformatics* **2010,** *11*, 405.

11. Galili, T.; O'Callaghan, A.; Sidi, J.; Sievert, C., *Bioinformatics* **2018,** *34*, 1600.

12. (a) Williams, R. M.; Glinka, T.; Kwast, E., *Tetrahedron Lett.* **1989,** *30*, 5575; (b) Trost, B. M.; Brennan, M. K., *Synthesis* **2009,** *2009*, 3003.

13. Payne, J. T.; Poor, C. B.; Lewis, J. C., *Angew. Chem. Int. Ed.* **2015,** *54*, 4226.

14. Yaipakdee, P.; Robertson, L. W., *Phytochemistry* **2001,** *57*, 341.

15. Ashtekar, K.; Marzijarani, N.; Jaganathan, A.; Holmes, D.; Jackson, J. E.; Borhan, B., *J. Am. Chem. Soc.* **2014,** *136*, 13355.

16. (a) Welch, C. J.; Gong, X.; Schafer, W.; Pratt, E. C.; Brkovic, T.; Pirzada, Z.; Cuff, J. F.; Kosjek, B., *Tetrahedron: Asymmetry* **2010,** *21*, 1674; (b) Bellomo, A.; Celebi‐Olcum, N.; Bu, X.; Rivera, N.; Ruck, R. T.; Welch, C. J.; Houk, K. N.; Dreher, S. D., *Angew. Chem. Int. Ed.* **2012,** *51*, 6912; (c) Santanilla, A. B.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D., *Science* **2015,** *347*, 49.

17. Niklasson, M.; Andresen, C.; Helander, S.; Roth, M.; Kahlin, A.; Appell, M.; Mårtensson, L. G.; Lundström, P., *Protein Sci.* **2015,** *24*, 2055.