

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

I. Materials and Instruments.....	2
A) Materials.....	2
B) Instruments	3
C) Software	4
II. Bioinformatics	8
A) Sequence Lists.....	8
A) Sequence Alignments	9
B) Sequence Similarity Network	9
C) SSN Sequence Analysis	13
III. FDH Expression	24
A) Cloning	24
B) Protein Expression.....	24
IV. High-Throughput LC-MS Bioconversion Screen	29
A) Reaction Setup.....	29
B) High-Throughput Screen Reaction Data Analysis	31
C) Heatmap/Clustering Analysis.....	33
V. Substrate Synthesis	34
VI. Preparative Scale Bioconversions.....	34
VII. Analytical Scale Bioconversions	43
A) Yohimbine Bioconversion.....	43
B) HTS Validation Study	44
C) Halide Selectivity Experiments.....	46
D) Lysate Bioconversions	52
VIII. Halenium Affinity Calculations	59
A) Halenium Affinity Calculation Setup.....	59
B) Halenium Affinity Data.....	60
IX. SDS-PAGE.....	69
A) Exploration of Alternative Expression Conditions	69
B) Example SDS-PAGE Gels	70
X. Circular Dichroism	71
A) CD Data Collection and Analysis	71
B) CD Data.....	72
XI. NMR	73
A) ^1H -NMR	73
B) ^{13}C -NMR	85
C) 2D-NMR Spectra.....	97
XII. SI References	99

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 μ m 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 Et₂O 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). DH5 α and BL21(DE3) *E. coli* were purchased from Invitrogen (Carlsbad, CA). Taq DNA polymerase and Phusion HF polymerase were purchased from New England Biolabs (Ipswich, 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 (Ipswich, 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). For each gene model, the JGI Mycocosm database was used to provide quality scores based on how well the gene model is supported by RNAseq data and how well it is supported by protein homology, including alignment coverage between gene model and homolog from protein databases. We prioritized all complete halogenase genes in Mycocosm, starting from genes of highest RNAseq support, such as genes based on RNAseq assembled transcripts and gene models constructed based on mapped RNAseq reads, and followed by genes of decreasing protein similarity support.

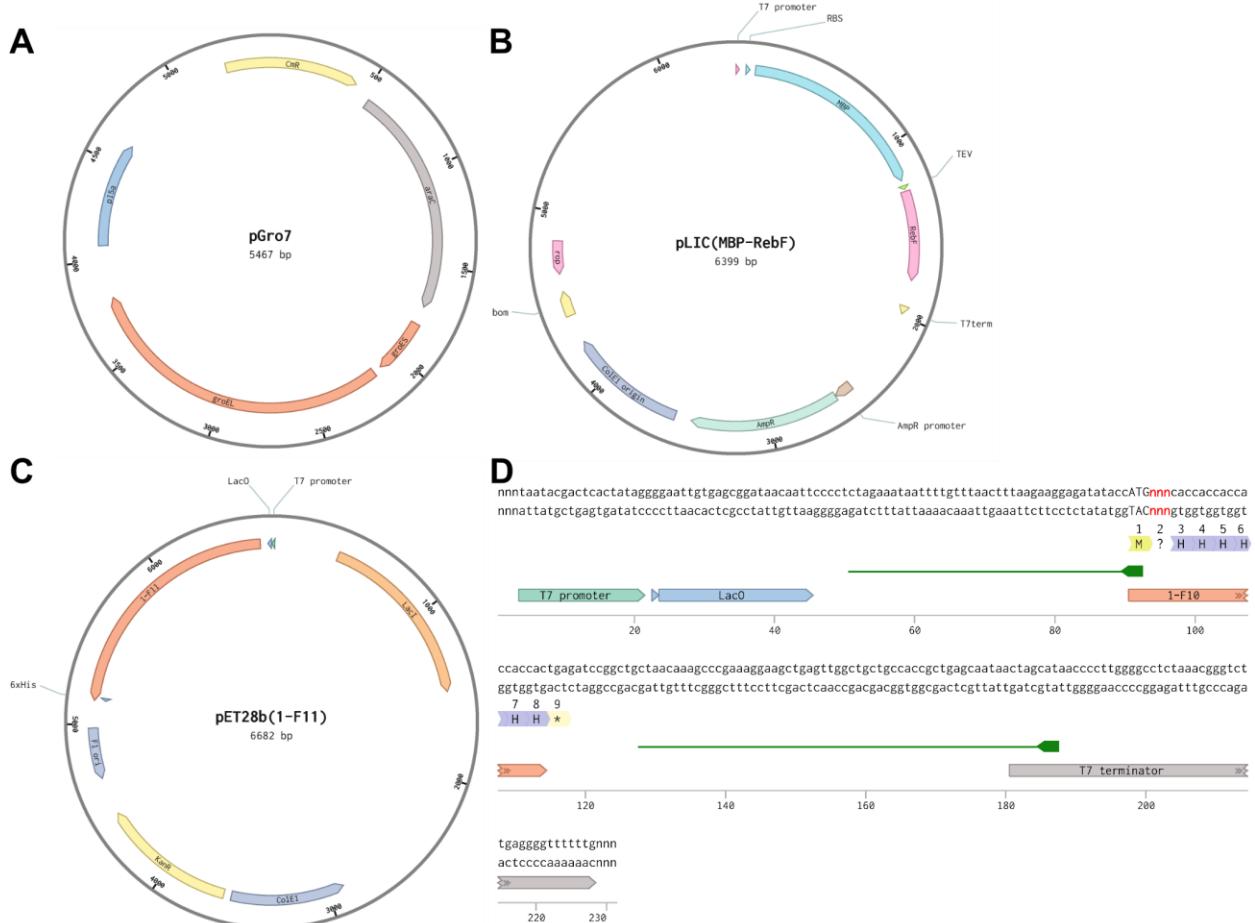


Figure S1. A) Plasmid map of pGro7 chaperone plasmid. B) Plasmid map of pLIC(MBP-RebF). C) Plasmid map of pET28b(1-F11), the expression vector for genome-mined halogenase 1-F11. D) Linear sequence of the insert region into pET28 of the genome-mined halogenases. The red “nnn” segment represents the position 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 previously¹ 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 H₂O 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 H₂O 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)
munspell	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)
tidyverse	* 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)
xm12	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 5W GU 5WGV 5WG 5WG X 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 Bioconductor²) 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” (**Equation S1**).

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

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 ($\approx 30\%$ 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 ($\approx 40\%$ 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-HIT⁴) 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.

Level 1 SSNs

Annotated by genome-mined vs. previously characterized

- = Previously characterized
- = Genome-mined

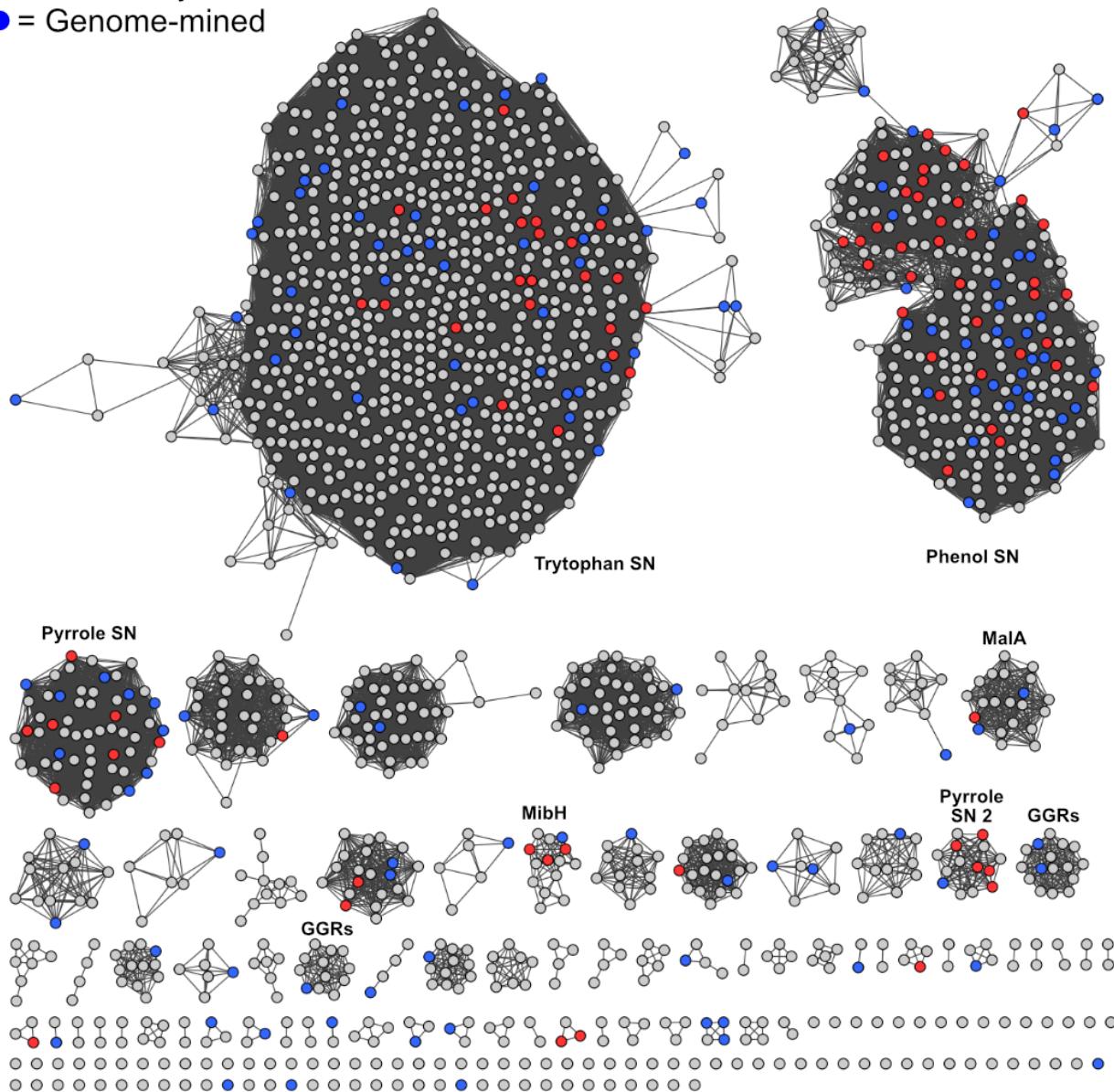


Figure S2. Level 1 SSN annotated with sequences that were characterized in this study in blue and FDHs that have been previously characterized in red. SSN drawn as 50% representative node network in which sequences sharing $\geq 50\%$ sequence similarity as determined by CD-HIT implemented in EFI-EST. Only one node, in the phenol subnetwork, includes more than one genome-mined sequence (FDHs 1-B04 and 1-G04).

Annotated by solubility

- = Soluble
- = Not soluble
- = No expression

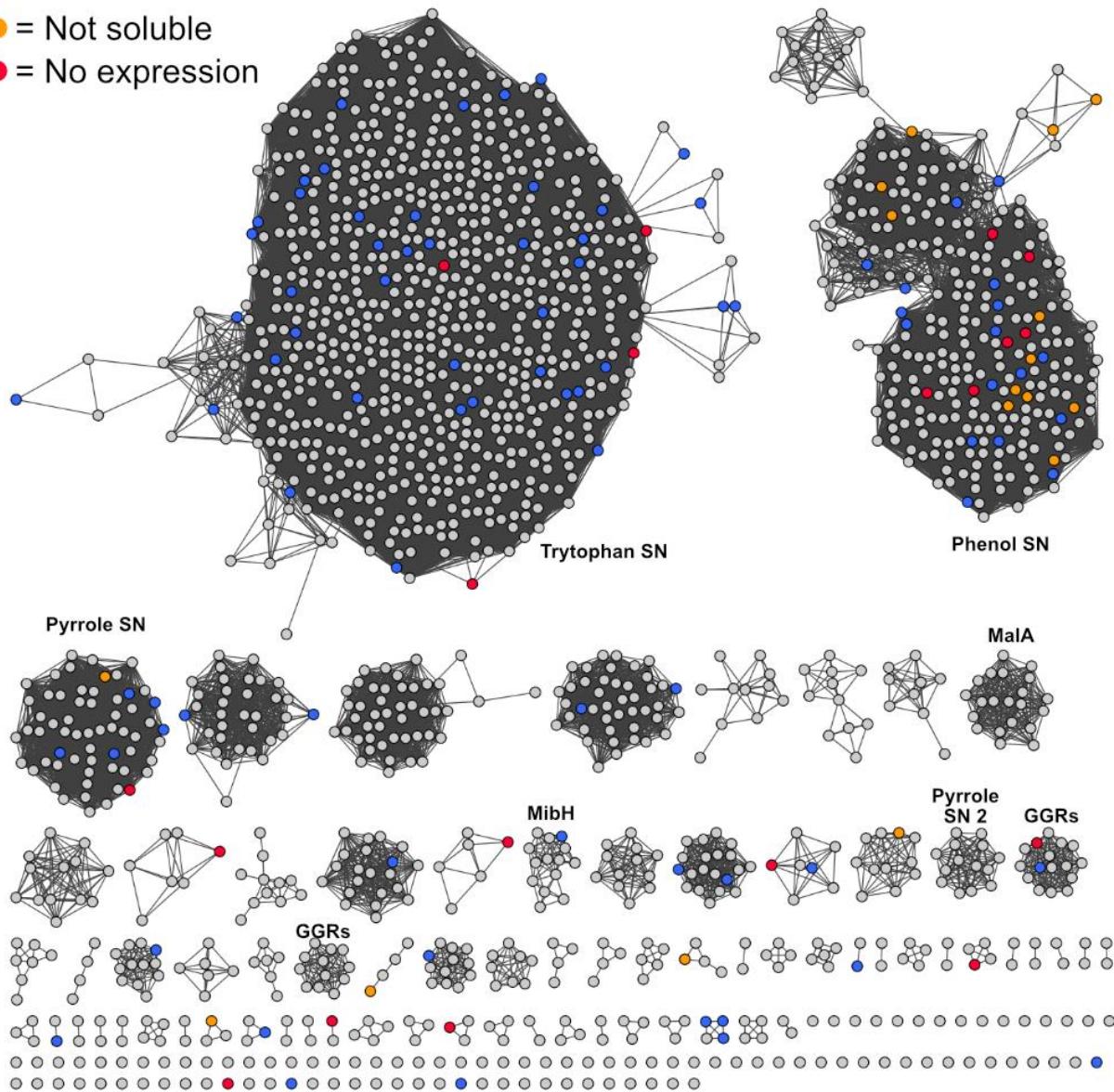


Figure S3. Level 1 SSN annotated with FDH solubility. SSN drawn as 50% representative node network in which sequences sharing $\geq 50\%$ sequence similarity as determined by CD-HIT implemented in EFI-EST. Nodes colored blue contain a protein that was determined to be soluble from small-scale expression 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

● = Structure determined

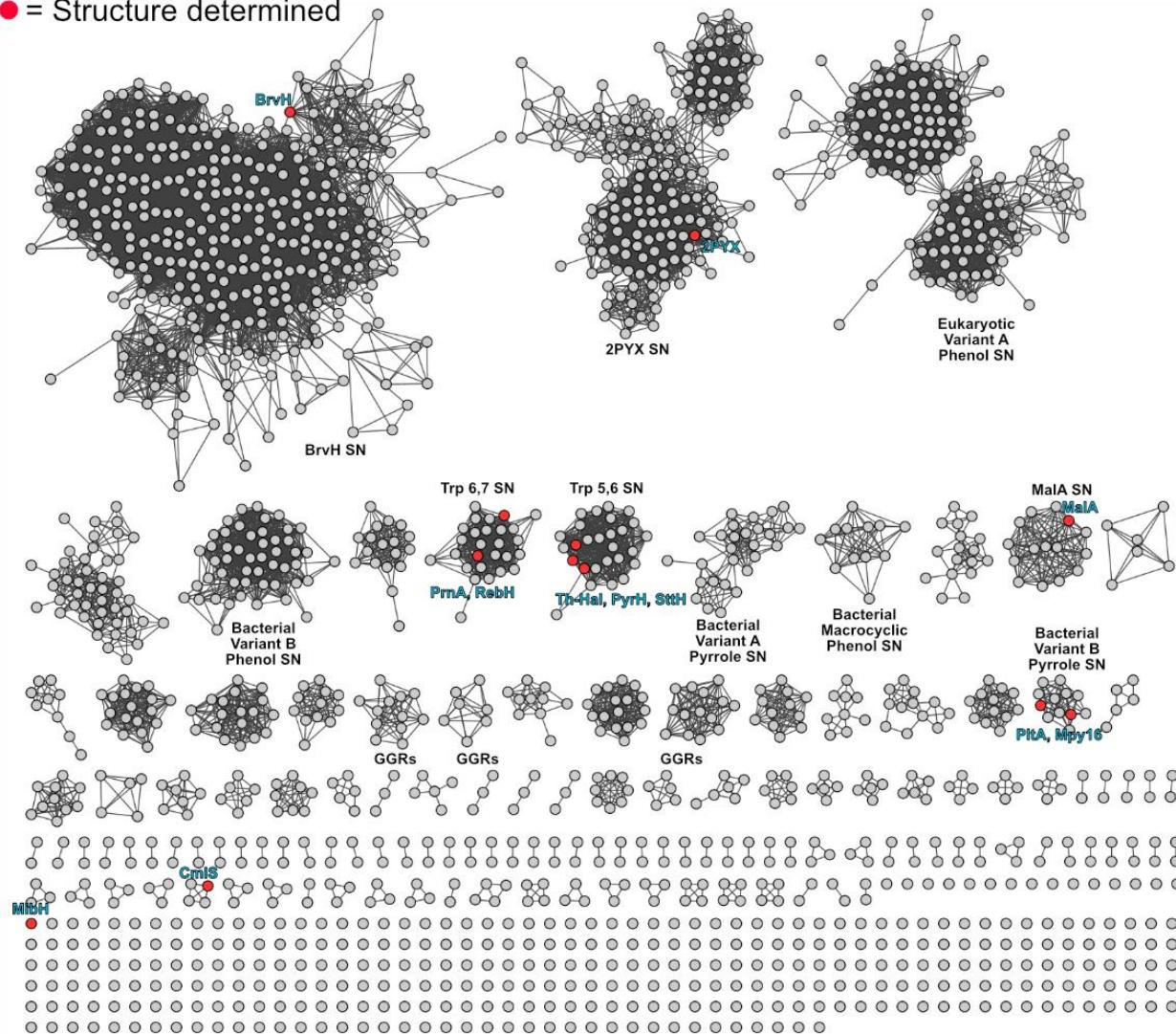


Figure S4. Level 2 SSN annotated with PDB-characterized FDHs. SSN drawn as 50% representative node network in which sequences sharing $\geq 50\%$ sequence similarity as determined by CD-HIT implemented in EFI-EST. Nodes colored red have at least one sequence with a structure reported in the PDB.

C) SSN Sequence Analysis

SSN Statistics

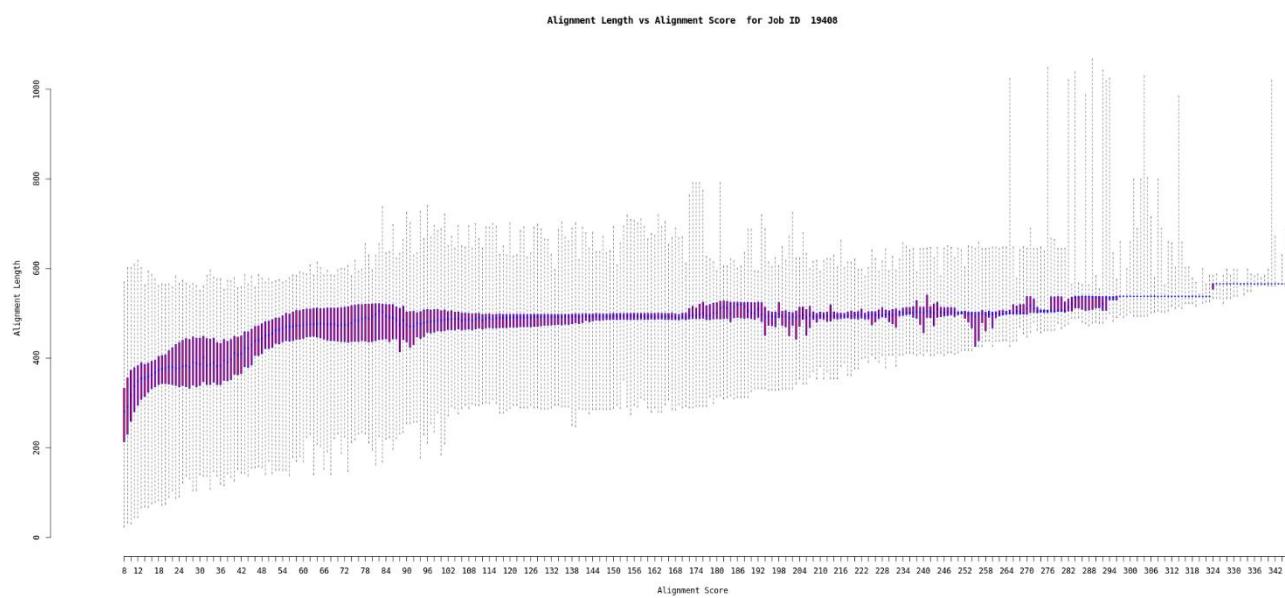


Figure S5. Protein alignment length vs. alignment score.

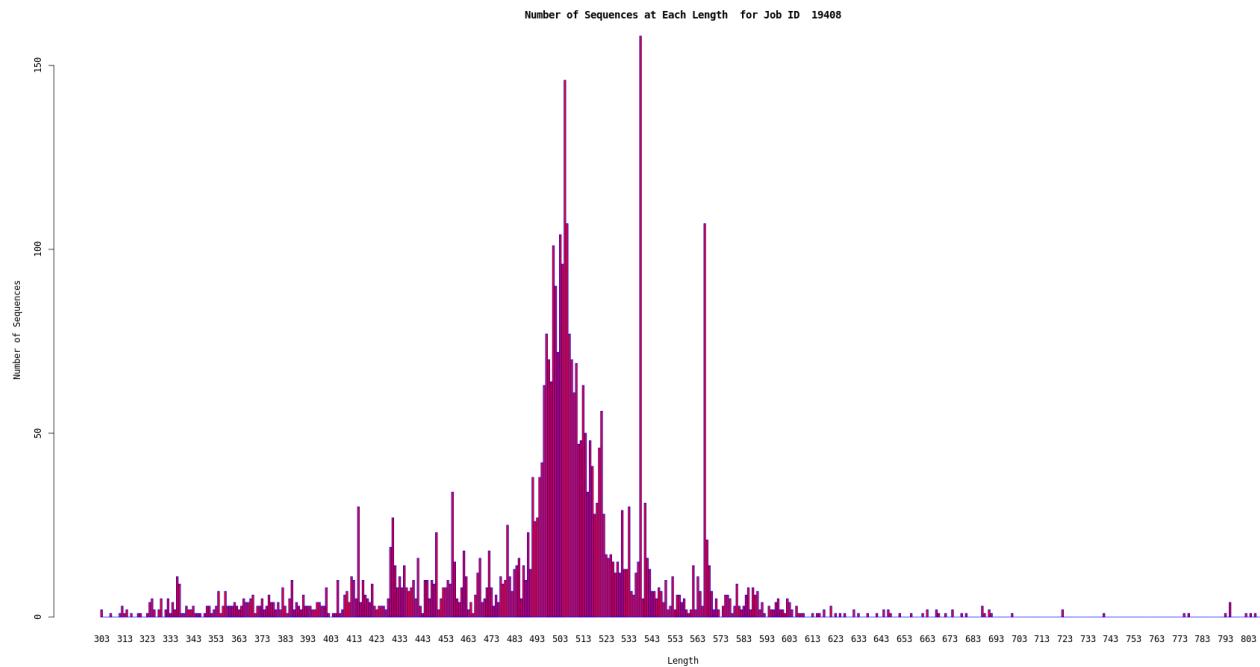


Figure S6. Histogram of number of sequences as a function of protein sequence length.

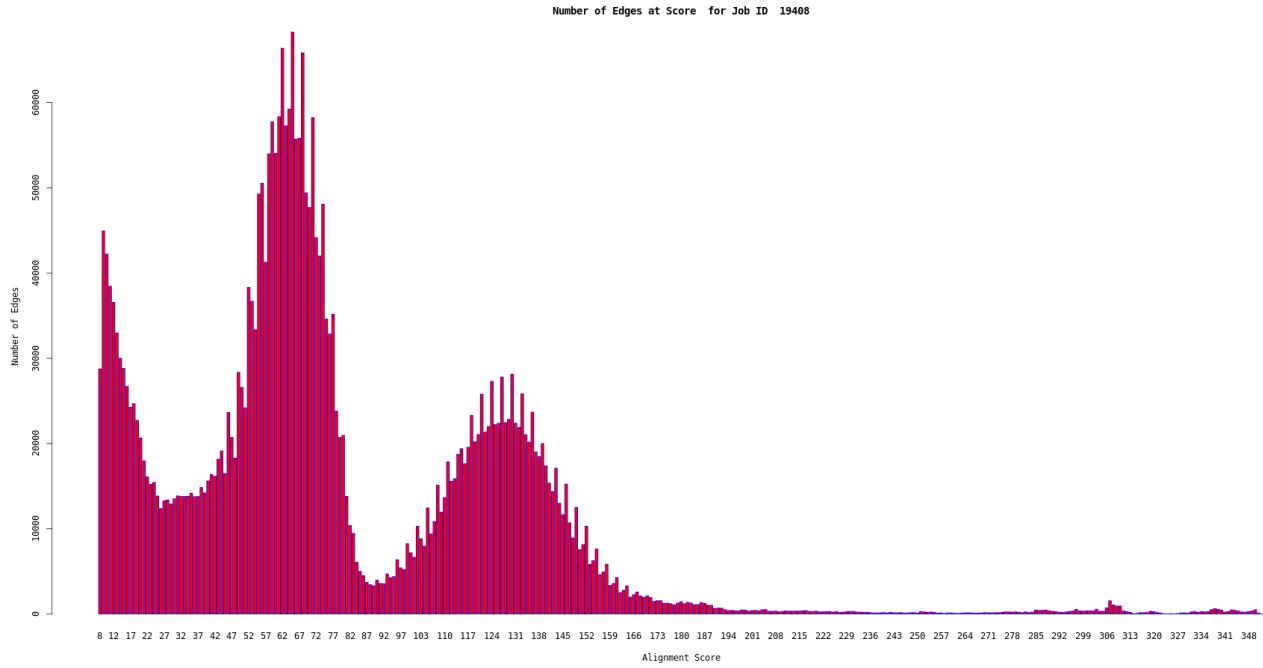


Figure S7. Histogram of number of network edges as a function of alignment score.

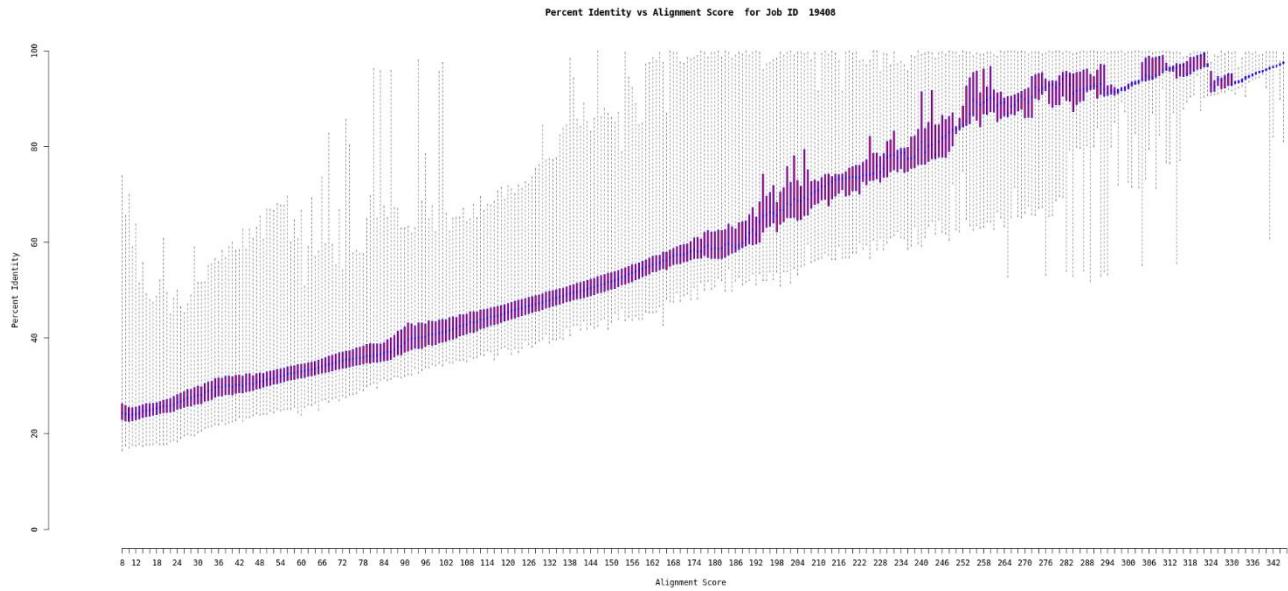


Figure S8. Percent identity as a function of alignment score.

Basic Sequence Statistics

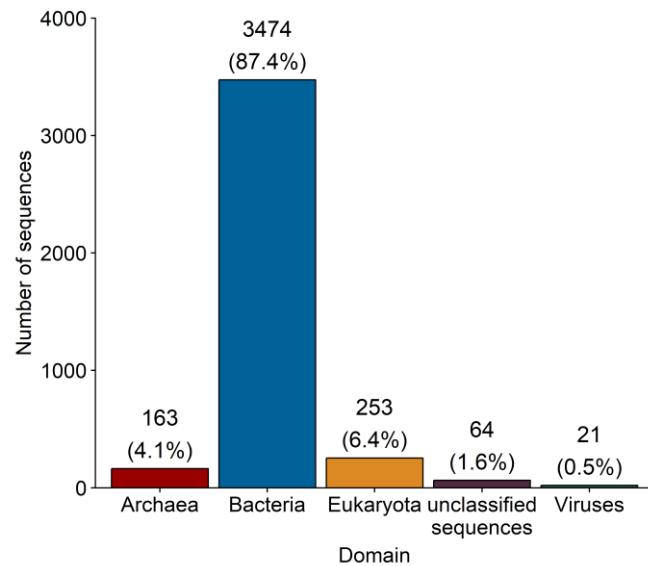


Figure S9. Distribution of host domains among sequences in the FDH sequence similarity network. Numbers above bars are the number of sequences and the percentage of the SSN in parentheses.

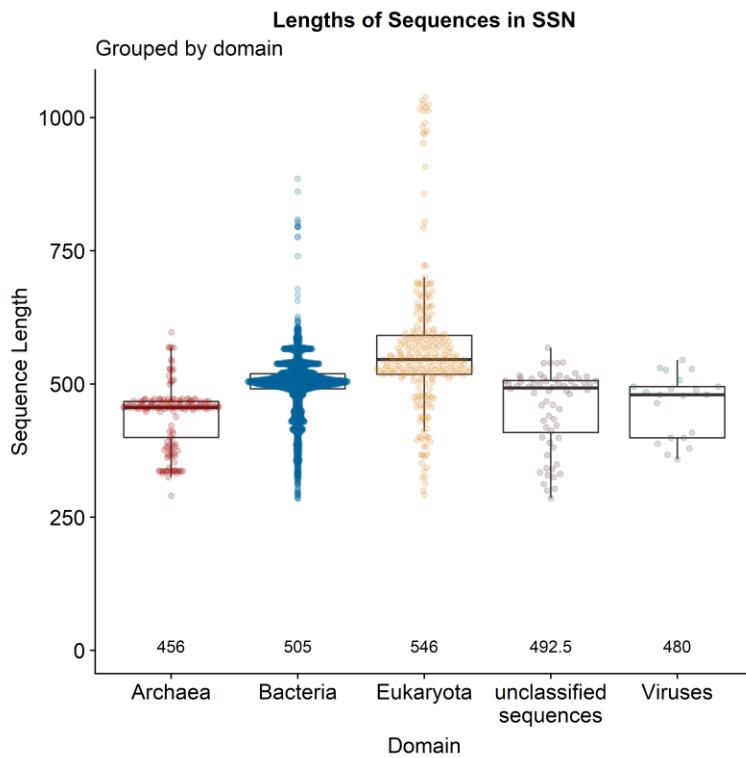


Figure S10. Distribution of sequence lengths among sequences in the FDH sequence similarity network, separated by host domain. Number at bottom of plot is the mean sequence length.

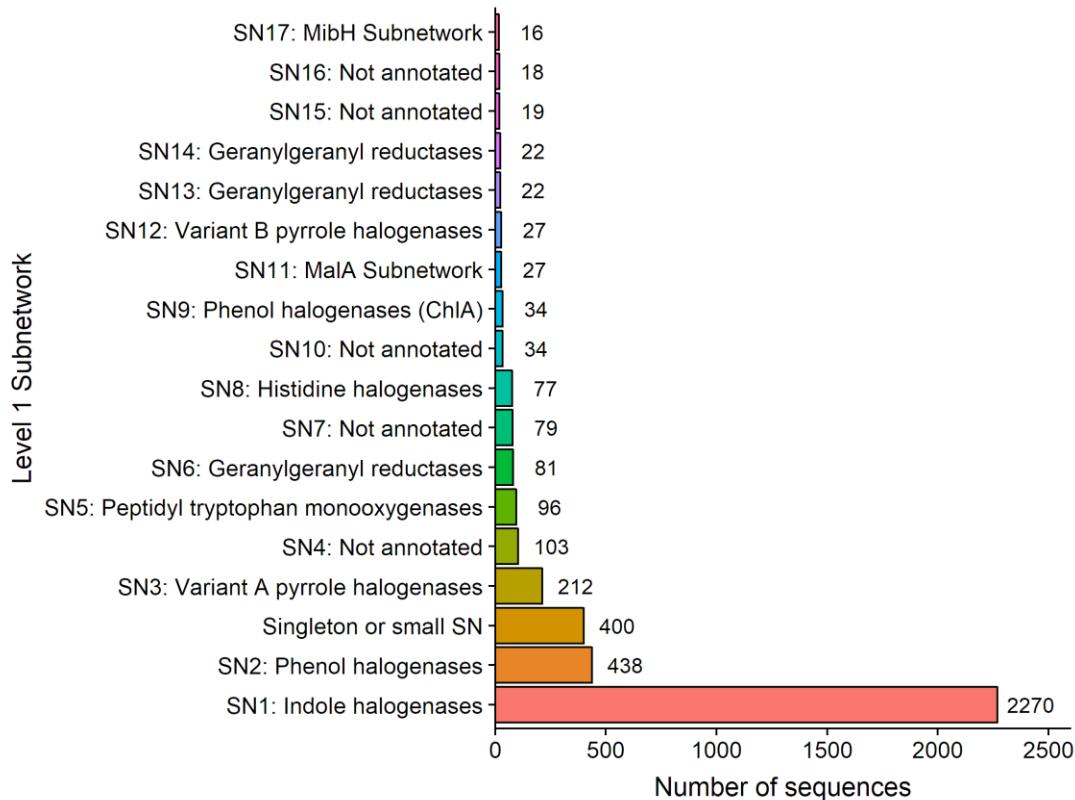


Figure S11. Distribution of sequences in different Level 1 subnetworks of the SSN. Only subnetworks with >15 members are shown.

Domains in SSN

Grouped by Level 1 Subnetwork

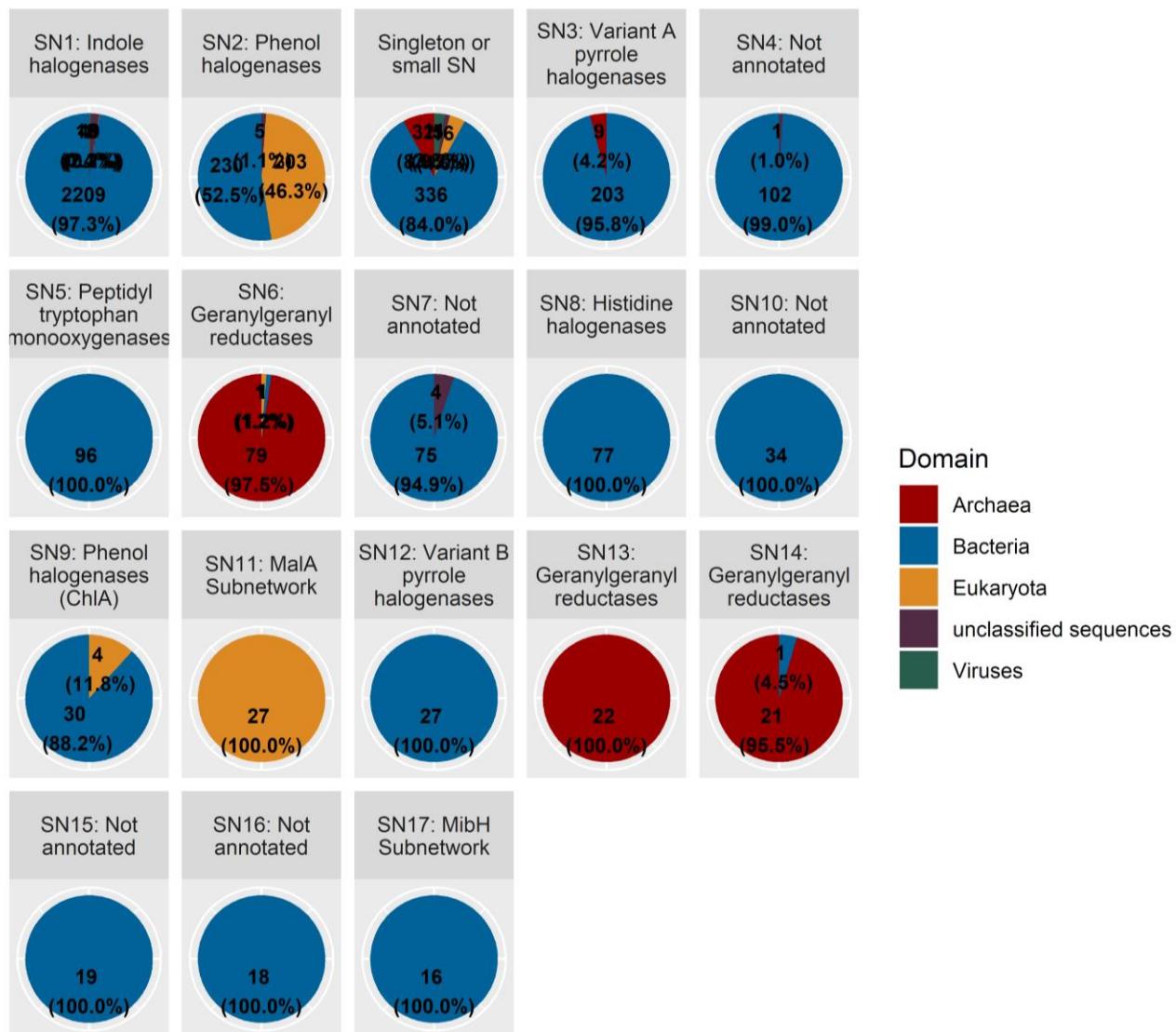


Figure S12. Distribution of sequence host domains across the FDH sequence similarity network, separated by Level 1 subnetworks. Note the high fraction of archaeal sequences in the subnetworks assigned as geranylgeranyl reductase subnetworks, the eukaryotic abundance of the MalA subnetwork, and the roughly even split between bacterial and eukaryotic hosts in the Phenol Subnetwork.

Pairwise Alignment Analysis

Max %ID of SSN Sequences to Nearest-Neighbor Genome-Mined Sequence

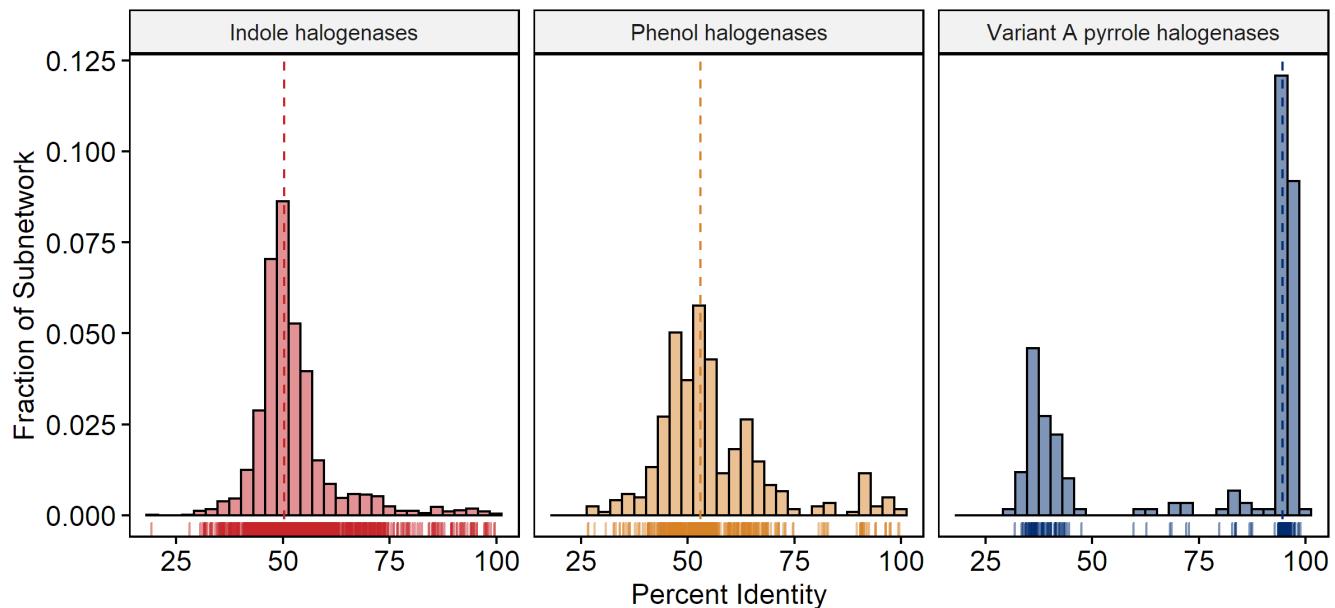


Figure S13. Percent identity of each sequence in the SSN to the most-similar non-identical sequence in the set of 128 genome-mined set of sequences evaluated experimentally. Data for only the three largest Level 1 subnetworks are shown. Each sequence in the SSN was pairwise-aligned using the BLOSUM62 substitution matrix and the PID1 option of the pid function in the Biostrings R package. Colored vertical lines below histograms indicate each individual sequence. Dashed lines on histograms indicate median %ID in the subnetwork. Note the high similarity of sequences in the Pyrrole Subnetwork.

Percent ID to Known Flavin-Dependent Halogenases

Max %ID for each sequence in SSN to manually-curated list of FDHs using pairwise alignments

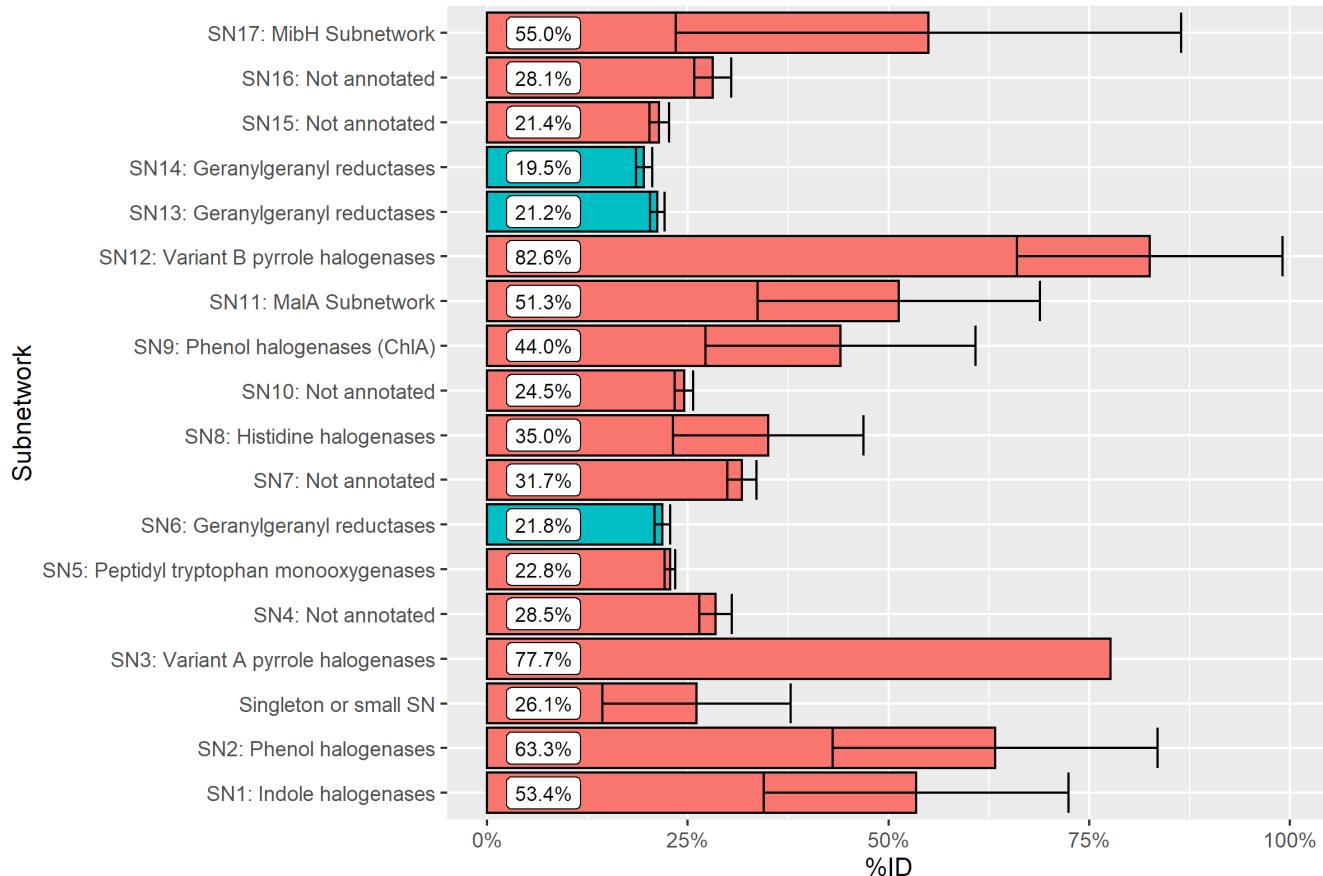


Figure S14. Percent identity of sequences in the sequence similarity network to known FDHs. Each sequence in the SSN was pairwise-aligned to each sequence within a manually curated list of known FDHs using the BLOSUM62 substitution matrix and the PID1 option of the pid function in the Biostrings R package. For each SSN sequence, the highest %ID to any known FDH was found, and the average of these maximum %ID's for all sequences within a subnetwork are illustrated in the bar chart above. Subnetworks that are assigned as geranylgeranyl reductase (GGR) subnetworks are highlighted in blue. Note the low similarity to known FDHs for the GGR subnetworks.

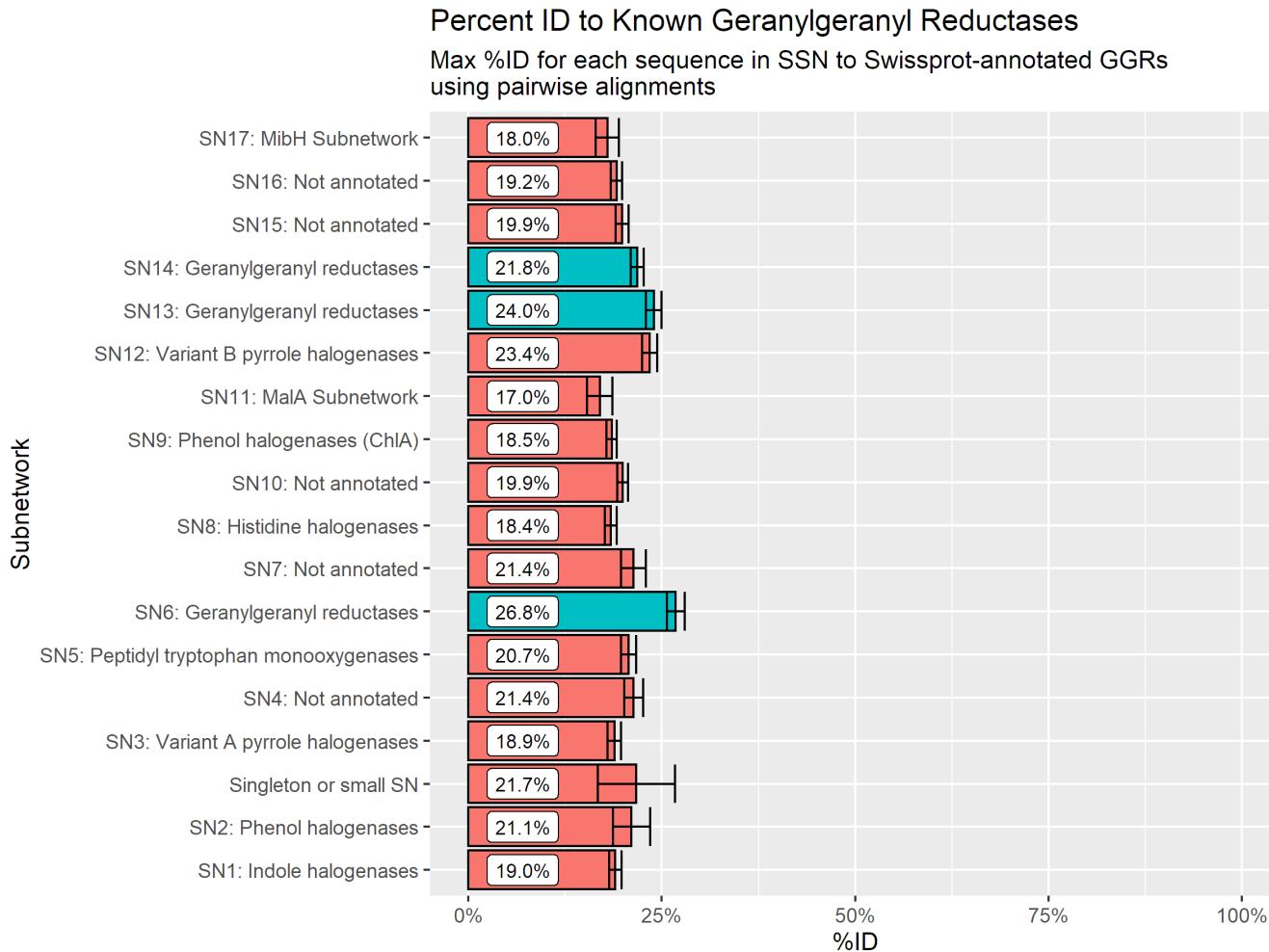
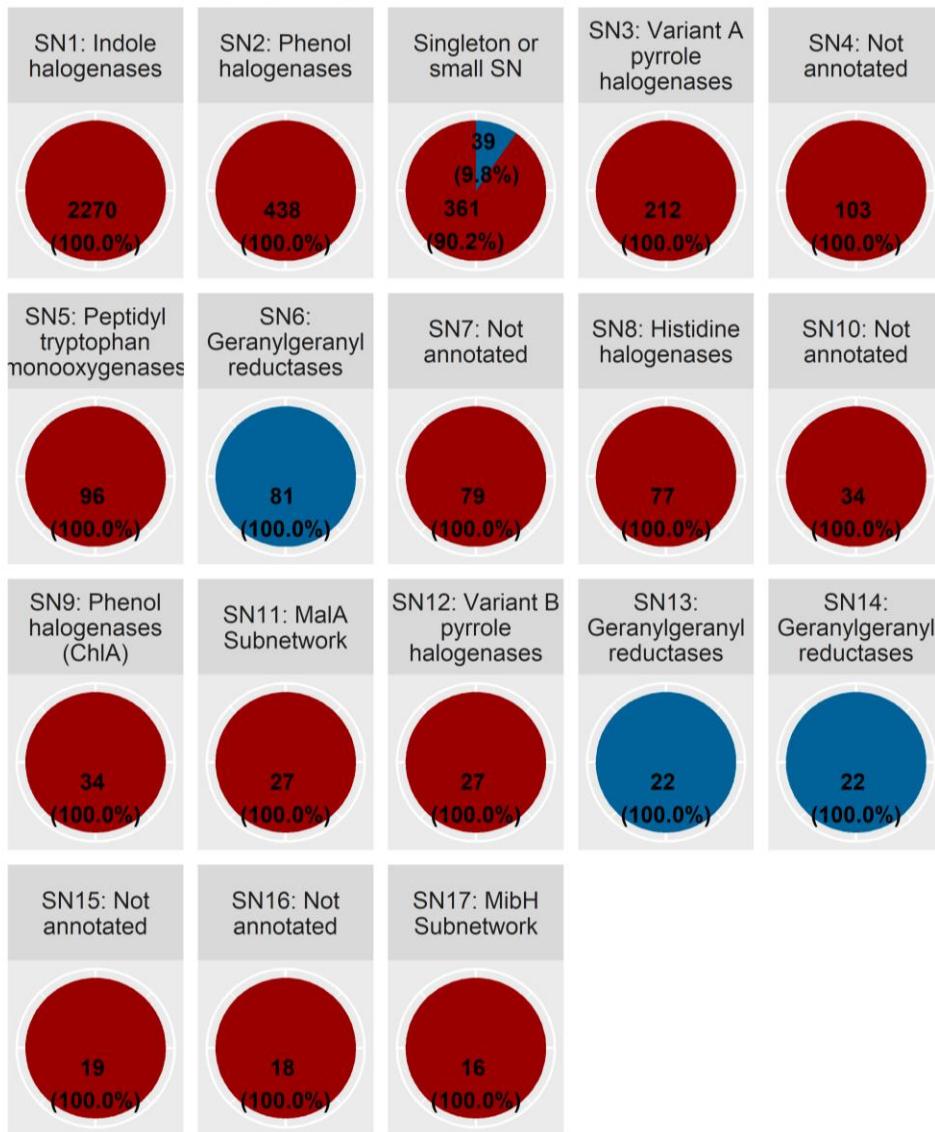


Figure S15. Percent identity of sequences in the sequence similarity network to geranylgeranyl reductases (GGRs). Each sequence in the SSN was pairwise-aligned to each reviewed sequence in the geranylgeranyl reductase family in InterPro as of 5/31/2019 (IPR011777; 36 sequences) using the BLOSUM62 substitution matrix and the PID1 option of the pid function in the Biostrings R package. For each SSN sequence, the highest %ID to any GGR was found, and the average of these maximum %ID's for all sequences within a subnetwork are illustrated in the bar chart above. Subnetworks that are assigned as geranylgeranyl reductase subnetworks are highlighted in blue. Note the comparatively high %ID for the subnetworks assigned as GGR subnetworks.

Motif Analysis

GGR Motifs in the SSN

YxWxFP motif; grouped by Level 1 Subnetwork



YxWxFP Motif Present?

- █ FALSE
- █ TRUE

Figure S16. Distribution of the YxWxFP motif across the SSN sequences, separated by Level 1 subnetwork. This motif has been found to be characteristic of GGRs.⁵

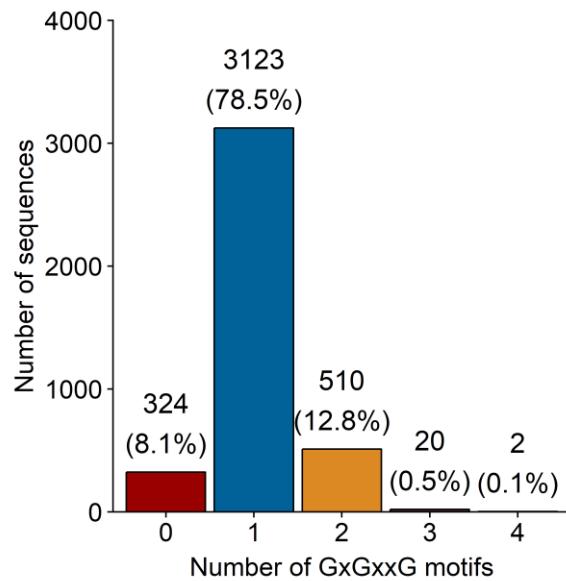


Figure S17. Distribution of the number of GxGxxG motifs in the SSN sequences. X axis separates sequences with the given number of GxGxxG motifs in the sequence. Y axis denotes number of sequences with the given number of GxGxxG motifs.

Phylogenetic Analysis



Figure S18. Phylogenetic tree for sequences in the Level 1 Pyrrole subnetwork. All 212 sequences were first subjected to clustering using CD-HIT⁴ with threshold of 90% sequence identity to reduce the redundancy of the sequence set, resulting in a set of 55 representative sequences. RebH and Rdc2 were added to this set to serve as outgroups, and multiple sequence alignment was performed using MUSCLE implemented in MEGA X.⁶ The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The tree with the highest log likelihood (-48388.01) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a

JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Uniparc IDs for the representative sequences are used as labels; labels are annotated with either the protein name for known halogenases or with the FDH ID if they are genome-mined sequences. Colored dots are drawn next to sequences with known native substrates, and the color indicates compound class. Orange = phenolic, blue = pyrrole, red = enolate.

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 OD₆₀₀ 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 Ca²⁺/glycerol buffer (60 mM CaCl₂, 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 Ca²⁺/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 Ca²⁺/glycerol buffer. The cell suspension was dispensed in 100 μL aliquots into chilled PCR tubes, snap frozen in liquid N₂, 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 TB media in 24-well deep well plates. Inoculated expression cultures were sealed with an AeraSeal adhesive film and incubated at 37 °C, 235 rpm until OD₆₀₀ ≈0.6-0.8, at which point the incubator was cooled to 15

°C. After the cultures were cooled sufficiently (\approx 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 N₂, 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.

Small-scale evaluation of expression conditions

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 overnight cultures (20 µL) were used to inoculate either 2 ml antibiotic-containing TB media or 2 mL antibiotic-containing autoinduction media (see below for composition) in 24-well deep well plates. Inoculated expression cultures were sealed with an AeraSeal adhesive film and incubated at 37 °C, 235 rpm until OD₆₀₀ \approx 0.6-0.8, at which point the incubator was cooled to 25 °C. After the cultures were cooled sufficiently (\approx 15 min) protein expression in the TB wells 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. Cells were lysed as described above.

Autoinduction Media Base

3.9 g Na₂HPO₄ (anhydrous)

3 g KH₂PO₄ (anhydrous)

20 g tryptone

5 g yeast extract

5 g NaCl

Autoclaved.

Autoinduction Media Sugar Mix

12.5 g glucose

50 g lactose

150 mL glycerol

Sterile filtered.

Autoinduction media was prepared by adding 40 mL of the sugar mix per 1 L of the autoinduction media base.

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 OD₆₀₀ \approx 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 concentration⁷ 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 S19).

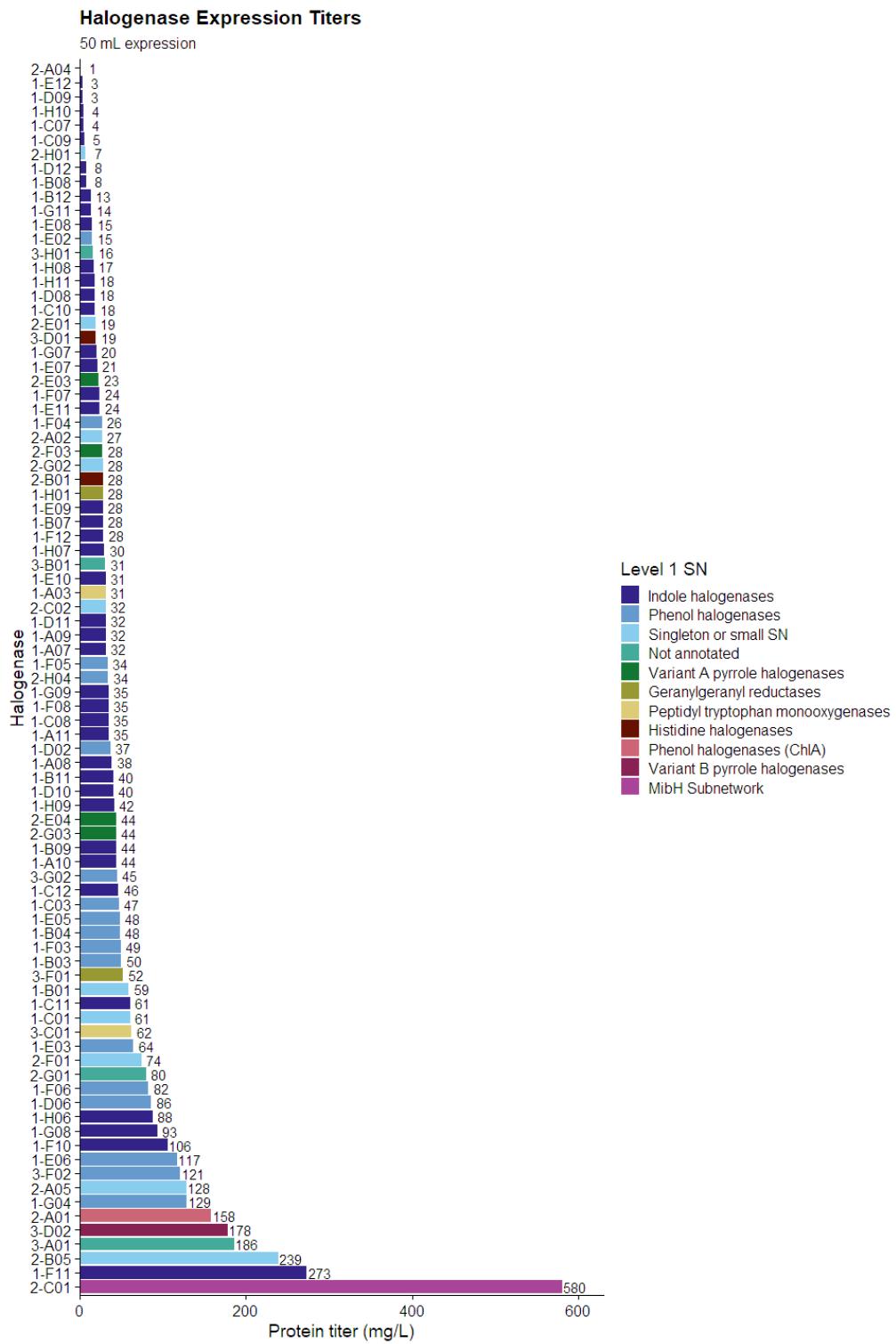


Figure S19. Protein expression titers determined by A_{280} from the 50 mL expressions of genome-mined FDHs. Expressions done as single experiments.

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 $OD_{600} \approx 0.6\text{-}0.8$, 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 μ M 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,⁸ but including 0.5% trichloroethanol in the separating portion of the gel to enable fluorescent detection of Trp-containing proteins.⁹ 8 μ L of the soluble or insoluble fraction of the cell lysate, 16 μ L water, and 12 μ L 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 μ L 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 solution¹⁰ 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 solution¹¹ 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.¹² 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 S21). 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 $[^{79}\text{Br}-\text{tryptamine}+\text{H}]^+$) 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.

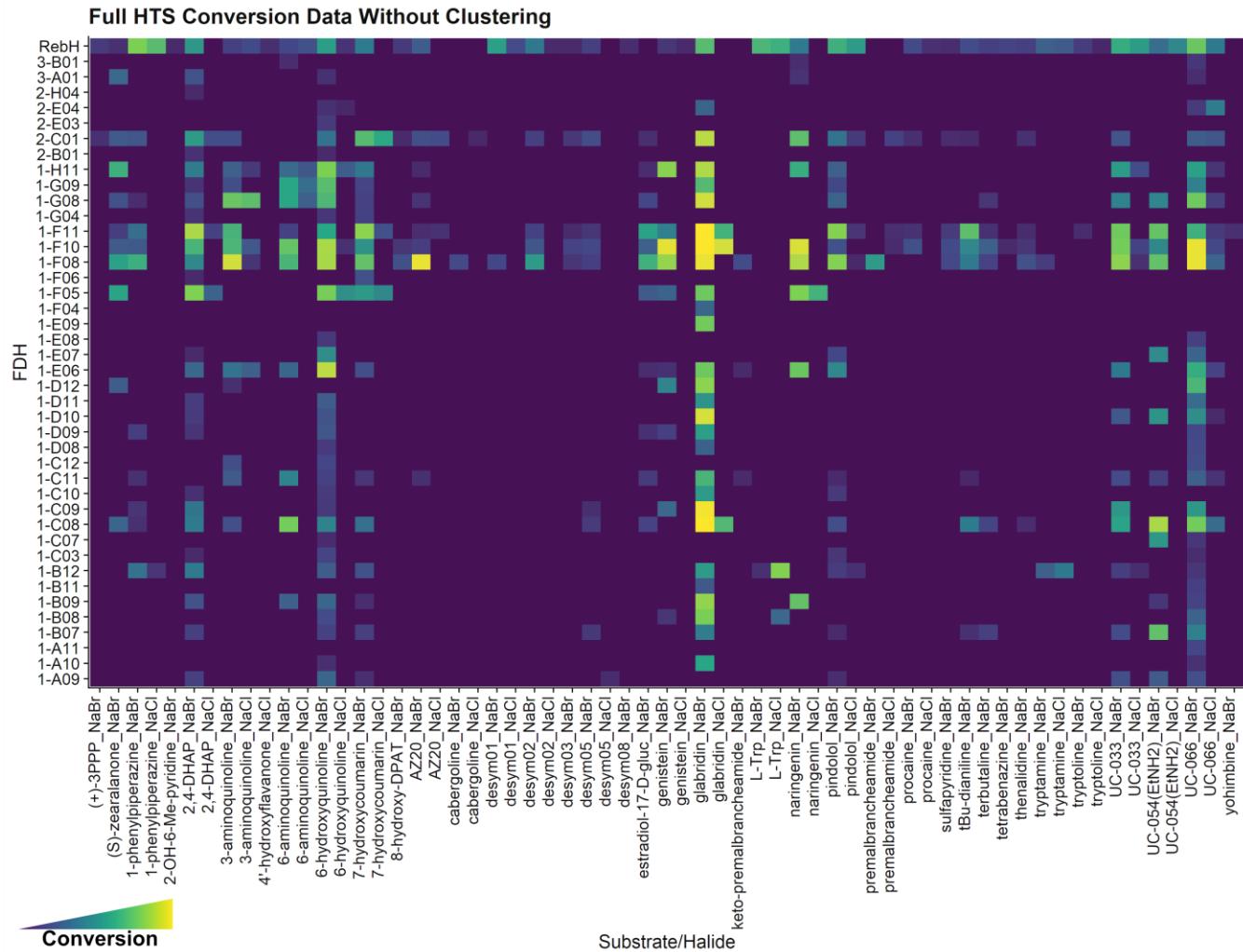


Figure S20. Heatmap depicting high-throughput screening results. Rows depict individual FDHs, and columns depict substrate/halide pairs. Intensity of yellow color represents conversion as determined in the high-throughput LCMS screen. No hierarchical clustering was applied to the dataset.

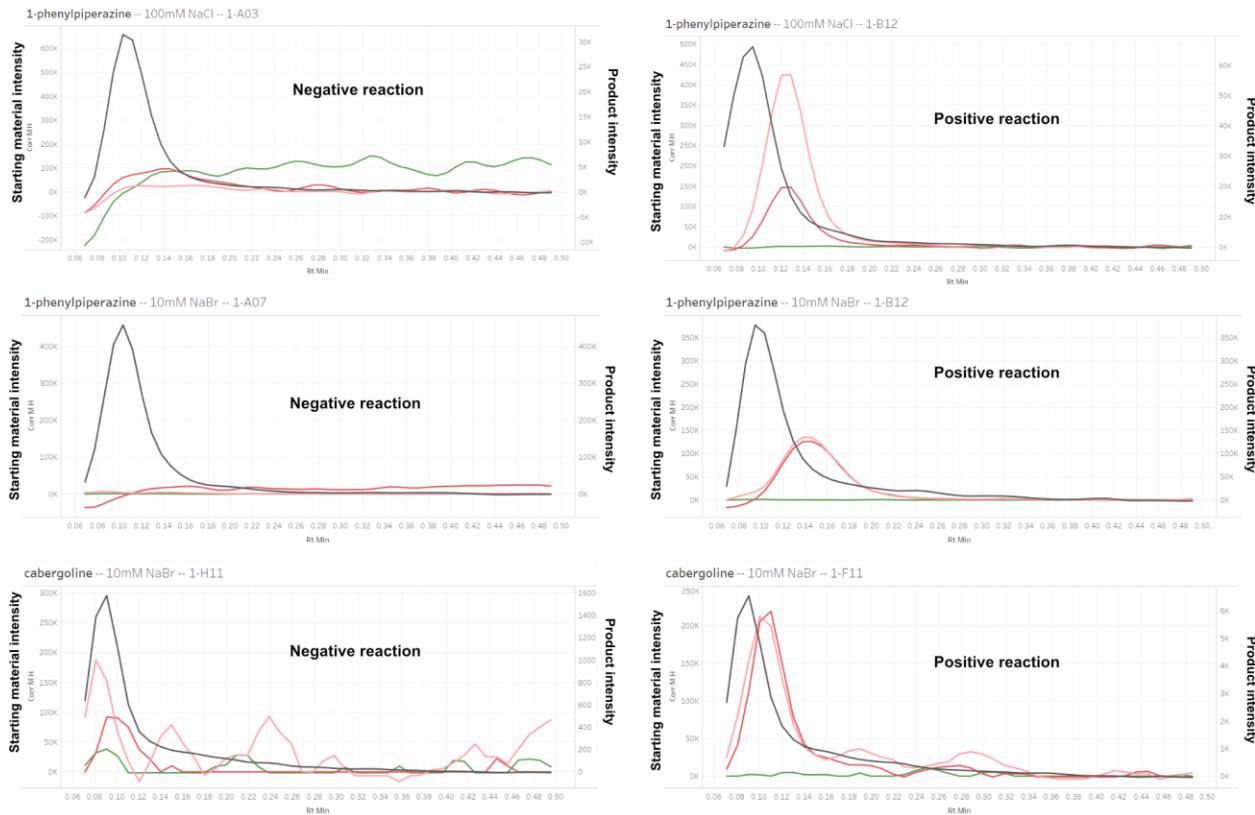
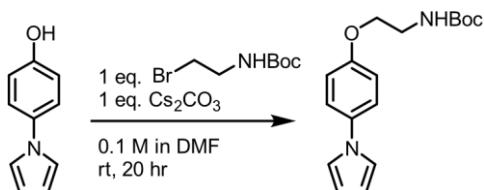


Figure S21. Example EIC chromatograms for select reactions. Dark gray curve is the EIC chromatogram for starting material; pink is the EIC chromatogram for the monohalogenated product from the lighter halogen isotope (^{35}Cl or ^{79}Br); red is the EIC chromatogram for the monohalogenated product from the heavier halogen isotope (^{37}Cl or ^{81}Br); green is the EIC chromatogram for doubly-halogenated product (the most abundant isotopomer).

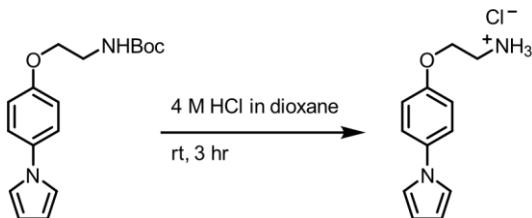
C) Heatmap/Clustering Analysis

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 package¹³ 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



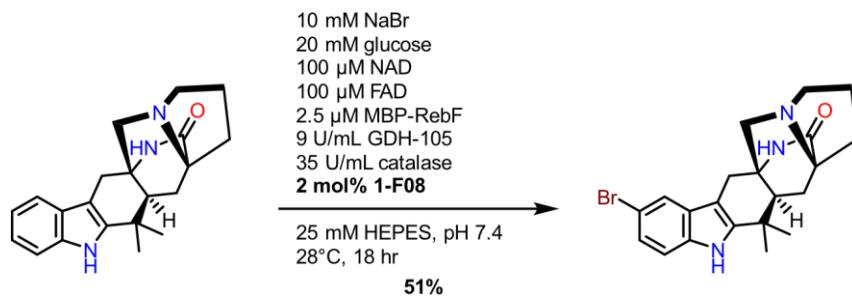
Starting material **S1** (4-(1H-pyrrol-1-yl)phenol; 1 eq., 100 mg, 0.63 mmol) was dissolved in DMF in a 100-mL round-bottom flask charged with a stirbar, after which Cs_2CO_3 (1 eq., 205 mg, 0.63 mmol) was added. 2-(Boc-amino)ethyl bromide (1 eq., 141 mg, 0.63 mmol) was added at room temperature, and 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 powder (57% yield). **$^1\text{H NMR}$** : (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). **$^{13}\text{C NMR}$** : (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 $[\text{C}_{17}\text{H}_{22}\text{O}_3\text{N}_2\text{Na}]^+$ ($[\text{M} + \text{Na}]^+$): 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 $[\text{C}_{12}\text{H}_{15}\text{ON}_2]^+$ ($[\text{M} + \text{H}]^+$): 203.1179; found 203.1178.

VI. Preparative Scale Bioconversions

Premalbrancheamide bromination with 1-F08



concentration 12 μ mol, 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 μ M), FAD (0.2 eq. in reaction buffer, 100 μ M final concentration). 1-F08 (2 mol%, 10 μ M final concentration) and MBP-RebF (2.5 μ M 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 \approx 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 MgSO₄, and concentrated. The residue was dissolved in 500 μ L 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. **¹H 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). **¹³C 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 [C₂₁H₂₆BrN₃O]⁺ ([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 S22**),^{14, 15} but we were unable to produce enough material to confirm this hypothesis.

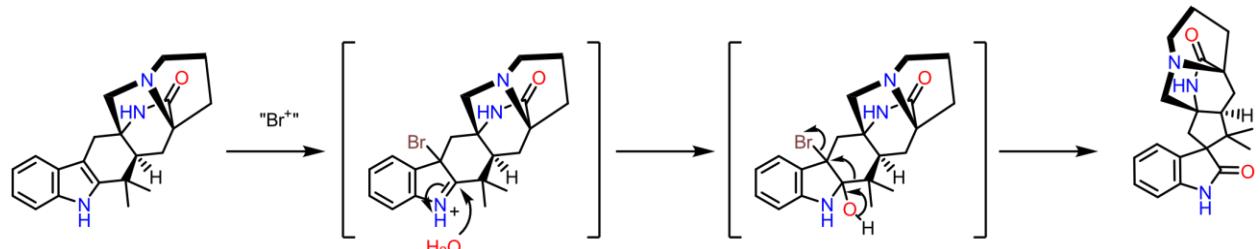
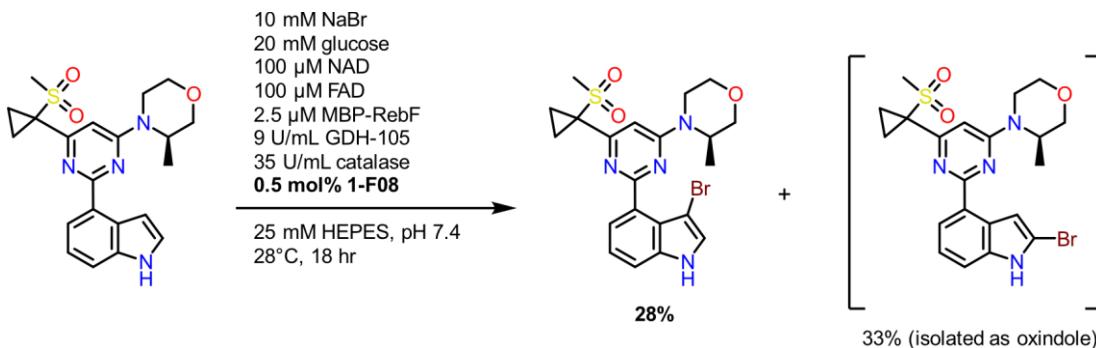


Figure S22: Proposed mechanism formation of side product spiro-oxindole from 1-F08 bromination of premalbrancheamide.

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 μ M final concentration 15 μ mol, 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 μ M), FAD (0.2 eq. in reaction buffer, 100 μ M final concentration), and 200 μ L DMSO (8% cosolvent). 1-F08 (0.5 mol%, 2.5 μ M final concentration) and MBP-RebF (2.5 μ M 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 \approx 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 MgSO₄, and concentrated. The residue was dissolved in 500 μ L 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:

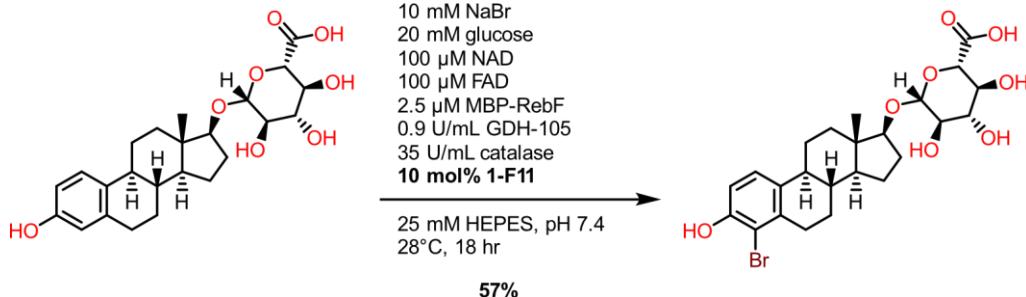
¹H NMR (500 MHz, Methylene Chloride-*d*₂) δ 8.62 (s, 1H), 7.50 (dd, *J* = 5.8, 3.4 Hz, 1H), 7.32 – 7.31 (m, 1H), 7.29 (s, 1H), 7.29 – 7.28 (m, 1H), 6.88 (s, 1H), 4.51 – 4.41 (m, 1H), 4.11 (d, *J* = 13.5 Hz, 1H), 4.00 – 3.92 (m, 1H), 3.76 (d, *J* = 11.5 Hz, 1H), 3.71 (dd, *J* = 11.6, 3.2 Hz, 1H), 3.56 (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). **¹³C 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 \sim 39 ppm is covered by the DMSO-*d*₆ peak. **HRMS (ESI-MS)**: calc. for [C₂₁H₂₄BrN₄O₃S]⁺ ([M+H]⁺): 491.0747 and 493.0728, found 491.0745 and 493.0723.

Oxindole-AZ20:

¹H 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,

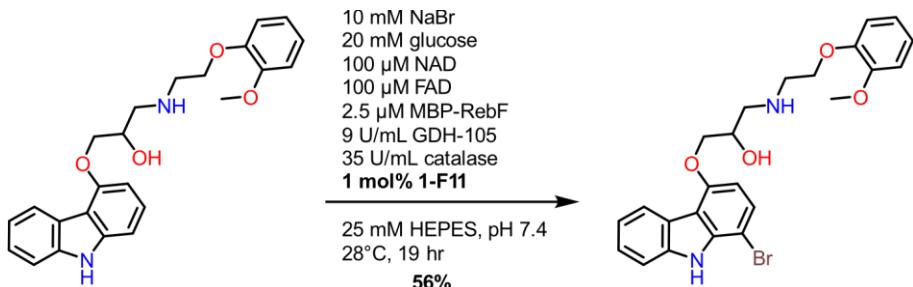
4.0 Hz, 1H), 3.03 (s, 3H), 1.85 – 1.78 (m, 2H), 1.33 (d, J = 6.8 Hz, 3H). Methylene hydrogens concealed under residual H₂O peak. **¹³C NMR:** (126 MHz, DMSO-*d*₆) δ 177.09, 163.18, 162.33, 161.85, 145.07, 134.44, 127.99, 125.93, 121.99, 111.26, 102.16, 70.67, 66.47, 46.95, 46.56, 40.58, 38.83, 13.89, 12.85. **HRMS (ESI-MS):** calc. for [C₂₁H₂₅N₄O₄S]⁺ ([M+H]⁺): 429.1591, found: 429.1588.

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



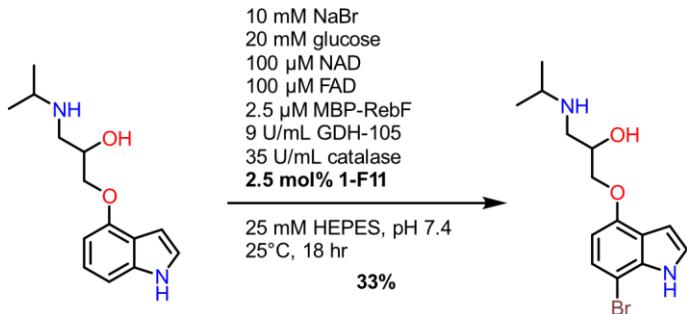
Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: β -Estradiol 17-(β -D-glucuronide) sodium salt (30 mM DMSO stock, 500 μ M final concentration 12 μ mol, 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 μ M), FAD (0.2 eq. in reaction buffer, 100 μ M final concentration). 1-F11 (10 mol%, 50 μ M final concentration) and MBP-RebF (2.5 μ M 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 CHCl₃/PrOH mixture five times. The organic layer was then washed with brine, dried over MgSO₄, and concentrated. The residue was dissolved in 500 μ L 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. **¹H 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). **¹³C NMR:** (126 MHz, DMSO-*d*₆) 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 [C₂₄H₃₀BrO₈]⁻ [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 μ mol) 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 μ M final concentration), and FAD (0.1 eq., 100 μ M 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 μ M final concentration) and 1-F11 (1 mol%, 10 μ M 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 μ L 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 NaHCO₃ 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). **¹H 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). **¹³C 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 [C₂₄H₂₆BrN₂O₄]⁺ ([M+H]⁺ with ⁷⁹Br) 485.1070, found 485.1073.

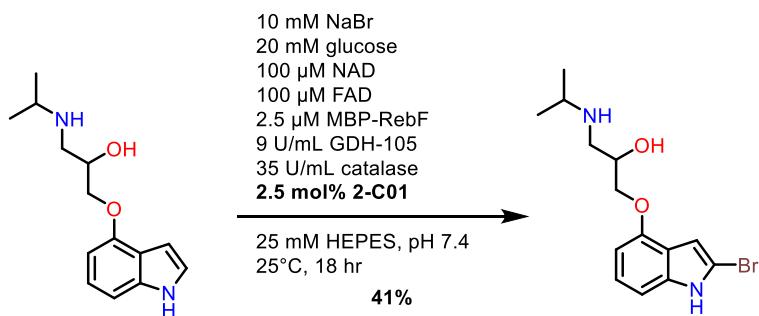
Pindolol bromination with 1-F11: 7-bromination



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 μ mol, 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 μ M), FAD (0.1 eq., 100 μ M final concentration). 1-F11 (2.5 mol%, 25 μ M final concentration) and MBP-RebF (2.5 μ M 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 \approx 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 MgSO₄, and concentrated onto Celite. This was then purified via reverse-phase chromatography (Biotage) as previously reported.¹⁶ 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. **¹H 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). **¹³C 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 [C₁₄H₂₀BrN₂O₂]⁺ ([M+H]⁺): 327.0703 and 329.0683, found 327.0708 and 329.0687.

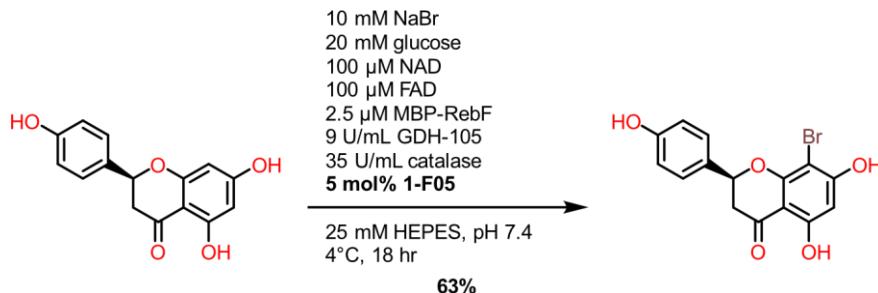
Pindolol bromination with 2-C01: 2-bromination



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 μ mol, 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 μ M), FAD (0.1 eq. in reaction buffer, 100 μ M final concentration). 2-C01 (2.5 mol%, 25 μ M final concentration) and MBP-RebF (2.5 μ M 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 $\text{pH} < 2$ using 6 M HCl. The resulting solution was filtered through Celite, then basified to $\text{pH} \approx 12$ 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. **$^1\text{H NMR}$** : (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). **$^{13}\text{C NMR}$** (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 $[\text{C}_{14}\text{H}_{20}\text{BrN}_2\text{O}_2]^+$ ($[\text{M}+\text{H}]^+$): 327.0703 and 329.0683, found 327.0708 and 329.0688.

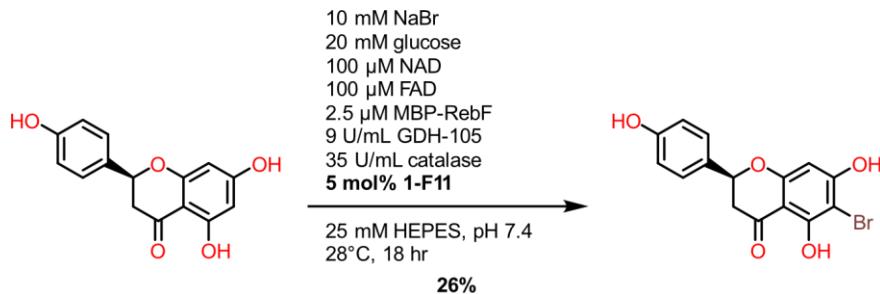
Naringenin bromination with 1-F05: 8-bromination



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 $\text{pH} < 2$ 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 $\text{CHCl}_3/\text{iPrOH}$ 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 $^1\text{H-NMR}$ for 8-brominated naringenin,¹⁷ in 63% isolated yield. **$^1\text{H NMR}$** (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, $J = 12.4, 3.1$

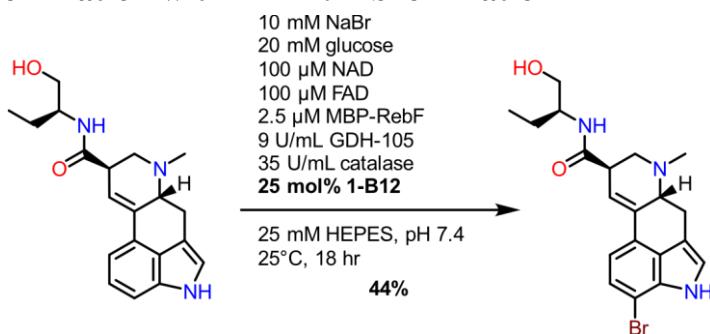
Hz, 1H), 3.32 (dd, J = 17.2, 12.4 Hz, 1H), 2.82 (dd, J = 17.2, 3.2 Hz, 1H). Note: one phenol hydroxyl peak appears to be broadened into baseline. **^{13}C NMR:** (126 MHz, DMSO- d_6) δ 196.83, 163.53, 162.41, 159.46, 158.22, 128.96, 128.62, 115.70, 102.86, 96.48, 88.74, 79.43, 41.90. **HRMS (ESI-MS):** calc. for $[\text{C}_{15}\text{H}_{10}\text{BrO}_5]^-$ ($[\text{M}-\text{H}]^-$): 348.9717; found 348.9711.

Naringenin bromination with 1-F11: 6-bromination



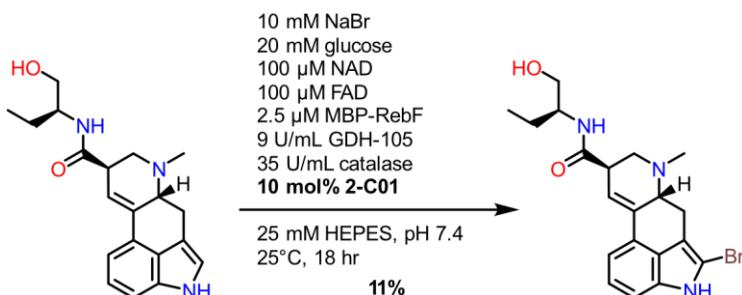
Ten 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: Naringenin (30 mM DMSO stock, 500 μM final concentration 15 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 CHCl₃/ⁱPrOH eight times. The organic layer was then washed with brine, dried over MgSO₄, and concentrated. The residue was dissolved in 300 μL DMSO, then purified 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. **^1H 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 $[\text{C}_{15}\text{H}_{10}\text{BrO}_5]^-$ ($[\text{M}-\text{H}]^-$): 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 μ M final concentration 6 μ mol, 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 μ M), FAD (0.2 eq. in reaction buffer, 100 μ M final concentration). 1-B12 (25 mol%, 100 μ M final concentration) and MBP-RebF (2.5 μ M 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 \approx 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 MgSO₄, and concentrated. The residue was dissolved in 500 μ L 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. **¹H 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 (dq, *J* = 13.8, 7.3 Hz, 1H), 1.47 (dq, *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. **¹³C 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. **HRMS (ESI-MS):** calc. for [C₂₀H₂₅BrN₃O₂]⁺ ([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 μ M final concentration 10.5 μ mol, 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 μ M), FAD (0.2 eq. in reaction buffer, 100 μ M final concentration). 2-C01 (10 mol%, 50 μ M final concentration) and MBP-RebF (2.5 μ M 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 \approx 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 MgSO₄, and concentrated. The residue was dissolved in 500 μ L 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. **¹H 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). **¹³C 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 [C₂₀H₂₃BrN₃O₂]⁻ ([M-H]⁻): 416.0979 and 418.0959, found 416.0973 and 418.0952.

VII. Analytical Scale Bioconversions

A) Yohimbine Bioconversion

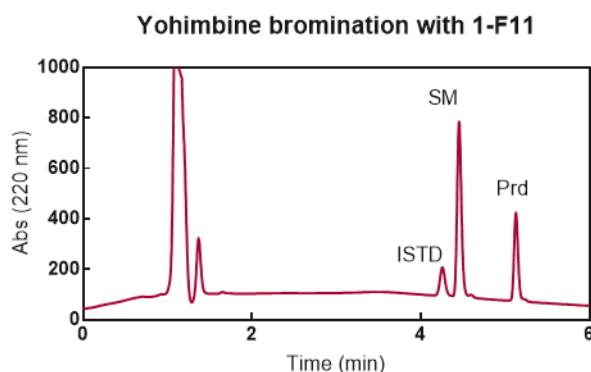
Analytical-scale bioconversion of 1-F11 bromination of yohimbine for LC analysis were prepared using substrate/small molecule/FDH/cofactor regen mixtures similarly to the high-throughput screen reaction setup.

NaBr	10 mM
glucose	20 mM
NAD	100 μ M
FAD	100 μ M
GDH	9 U/mL
MBP-RebF	2.5 μ M
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,¹⁹⁻²¹ 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

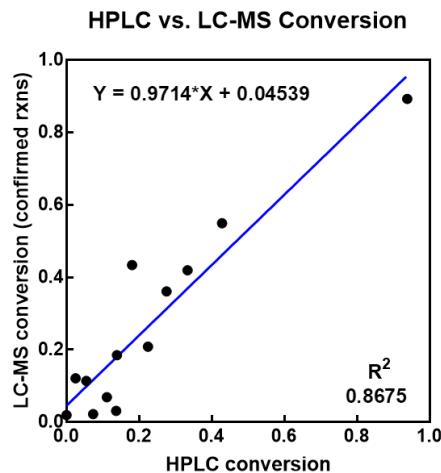


Figure S23. Comparison of conversions obtained from integrating UV chromatogram peaks by HPLC (220 nm) and conversions obtained from the high-throughput LC-MS screen.

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 μ M
FAD	100 μ M
GDH	9 U/mL
MBP-RebF	2.5 μ M
Catalase	35 U/mL
FDH	25 μ M
substrate	1 mM
Rxn Volume	45 μ L

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 μ L MeOH. Precipitated protein was pelleted by centrifugation at 3600 rpm at 4°C for 10 min. Supernatant was transferred to a 0.2 μ m 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 μ m particle size, 2.1 x 5.0 mm) with 0.1% formic acid in H₂O 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

C) Halide Selectivity Experiments

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 [M³⁵Cl+H]⁺ and [M³⁷Cl+H]⁺ signal added; for monochlorinated 2,4-dihydroxyacetophenone, the EIC chromatograms consist of the [M³⁵Cl-H]⁻ and [M³⁷Cl-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 [M⁷⁹Br+H]⁺ and [M⁸¹Br+H]⁺ signal added; for monobrominated 2,4-dihydroxyacetophenone, the EIC chromatograms consist of the [M⁷⁹Br-H]⁻ and [M⁸¹Br-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 [M³⁵Cl³⁵Cl+H]⁺ and [M³⁵Cl³⁷Cl+H]⁺ signals added; for 2,4-dihydroxyacetophenone, the EIC chromatograms consist of the [M³⁵Cl³⁵Cl-H]⁻ and [M³⁵Cl³⁷Cl-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

FDH Halide Preferences

FDH: RebH; substrate: L-Trp; conversions from EIC area

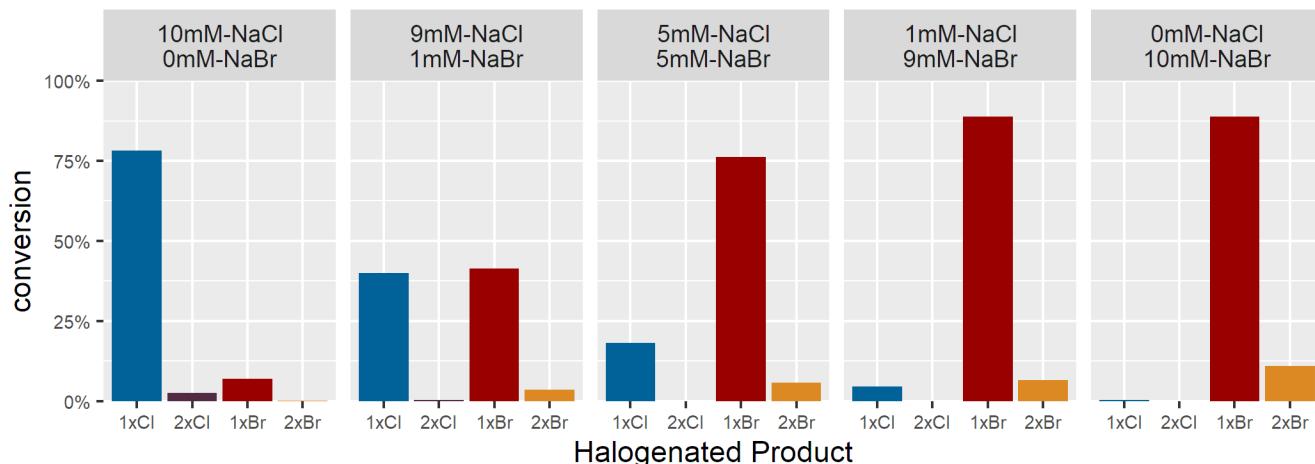


Figure S24. Halide selectivity of RebH in L-Trp halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

FDH Halide Preferences

FDH: RebH; substrate: 1-phenylpiperazine; conversions from EIC area

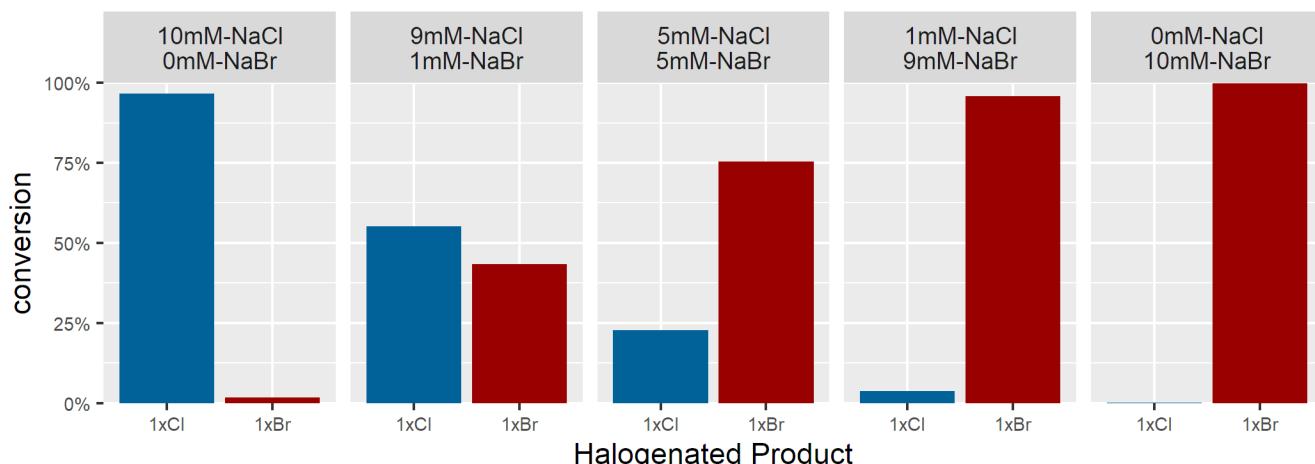


Figure S25. Halide selectivity of RebH in 1-phenylpiperazine halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

FDH Halide Preferences

FDH: RebH; substrate: pindolol; conversions from EIC area

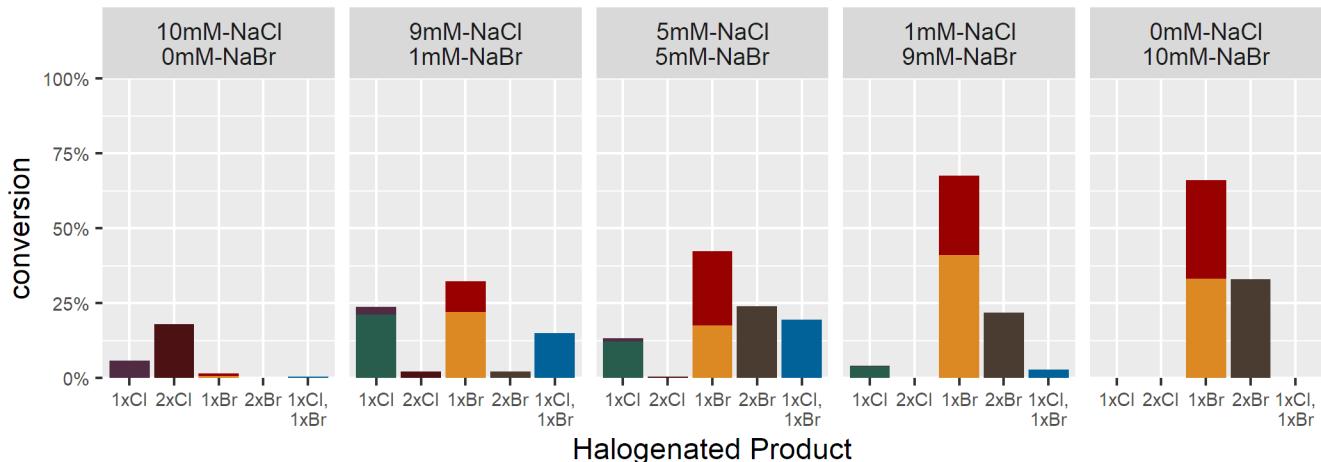


Figure S26. Halide selectivity of RebH in pindolol halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product m/z).

FDH Halide Preferences

FDH: RebH; substrate: 2,4-DHAP; conversions from EIC area

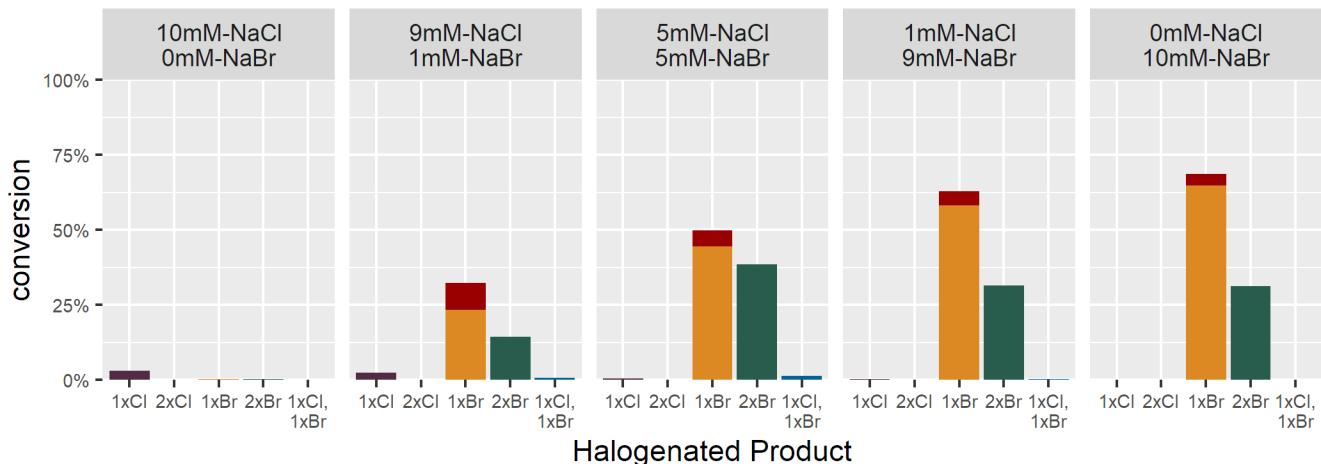


Figure S27. Halide selectivity of RebH in 2,4-dihydroxyacetophenone halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product m/z).

1-B12

FDH Halide Preferences

FDH: 1-B12; substrate: L-Trp; conversions from EIC area

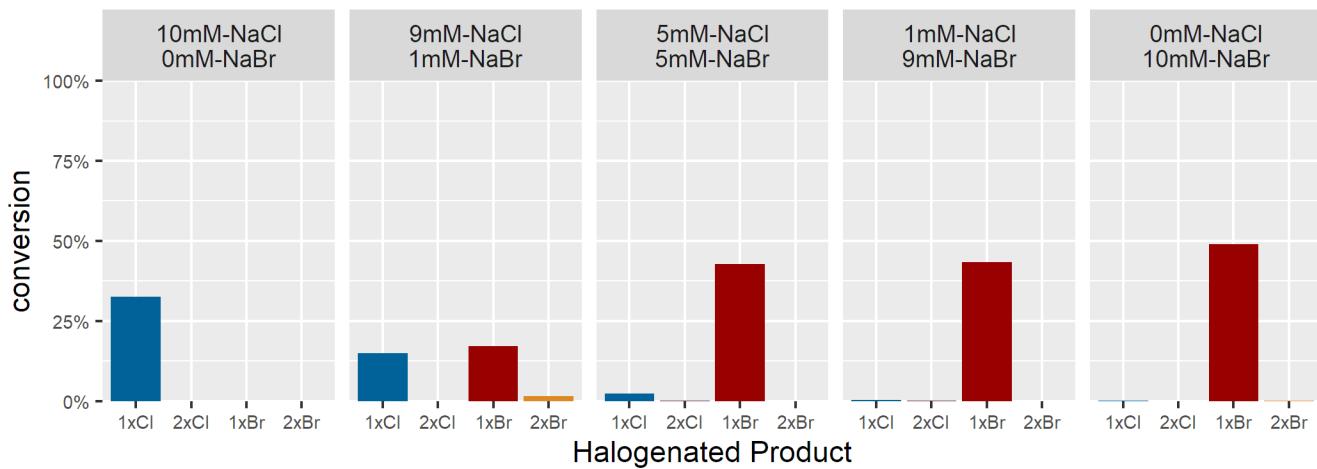


Figure S28. Halide selectivity of 1-B12 in pindolol halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

FDH Halide Preferences

FDH: 1-B12; substrate: 1-phenylpiperazine; conversions from EIC area

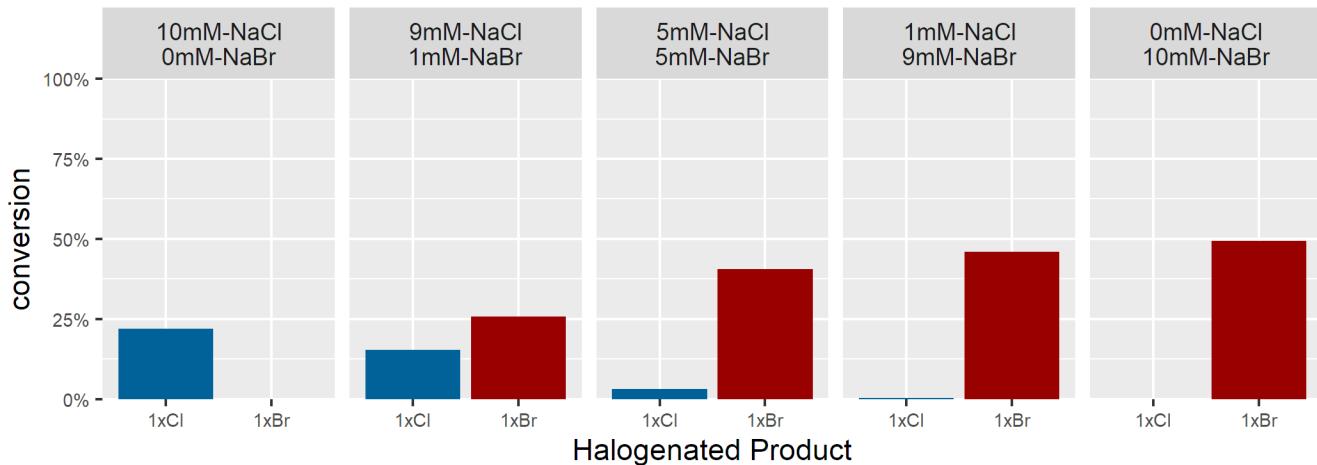


Figure S29. Halide selectivity of 1-B12 in 1-phenylpiperazine halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

FDH Halide Preferences

FDH: 1-B12; substrate: pindolol; conversions from EIC area

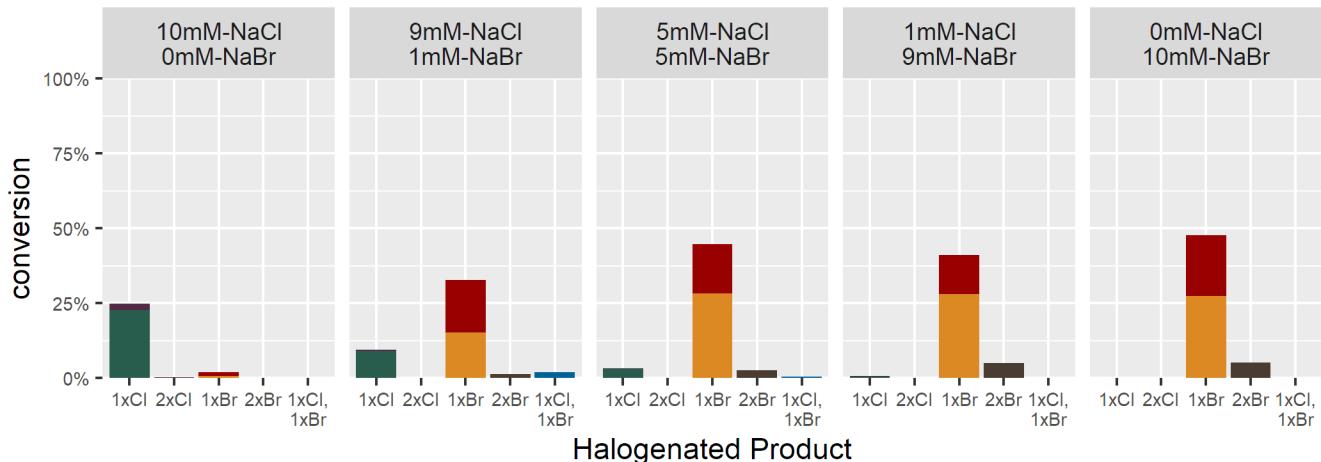


Figure S30. Halide selectivity of 1-B12 in pindolol halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product m/z).

FDH Halide Preferences

FDH: 1-B12; substrate: 2,4-DHAP; conversions from EIC area

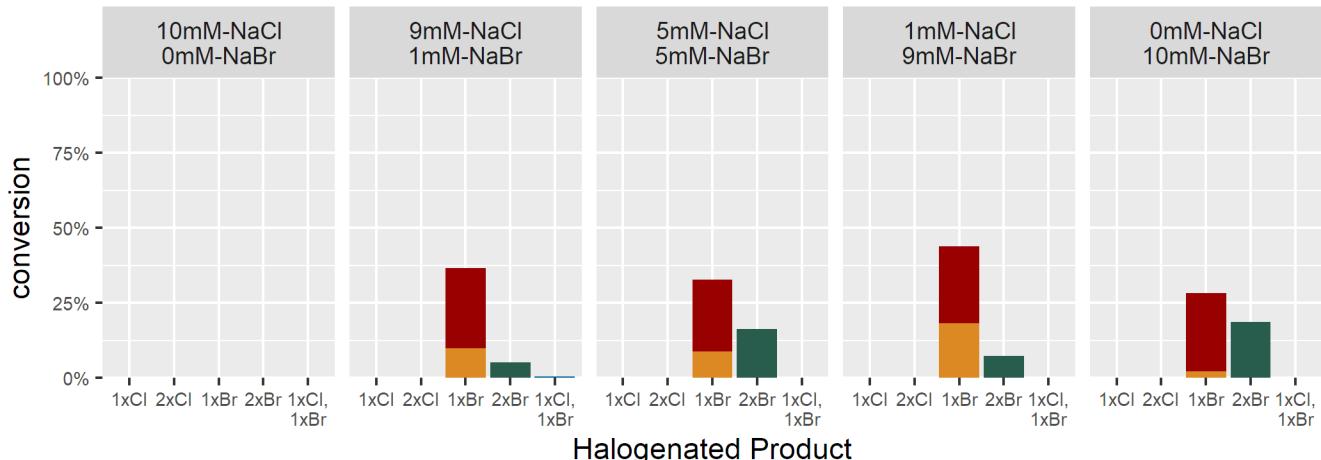


Figure S31. Halide selectivity of 1-B12 in 2,4-dihydroxyacetophenone halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product m/z).

2-C01

FDH Halide Preferences

FDH: 2-C01; substrate: L-Trp; conversions from EIC area

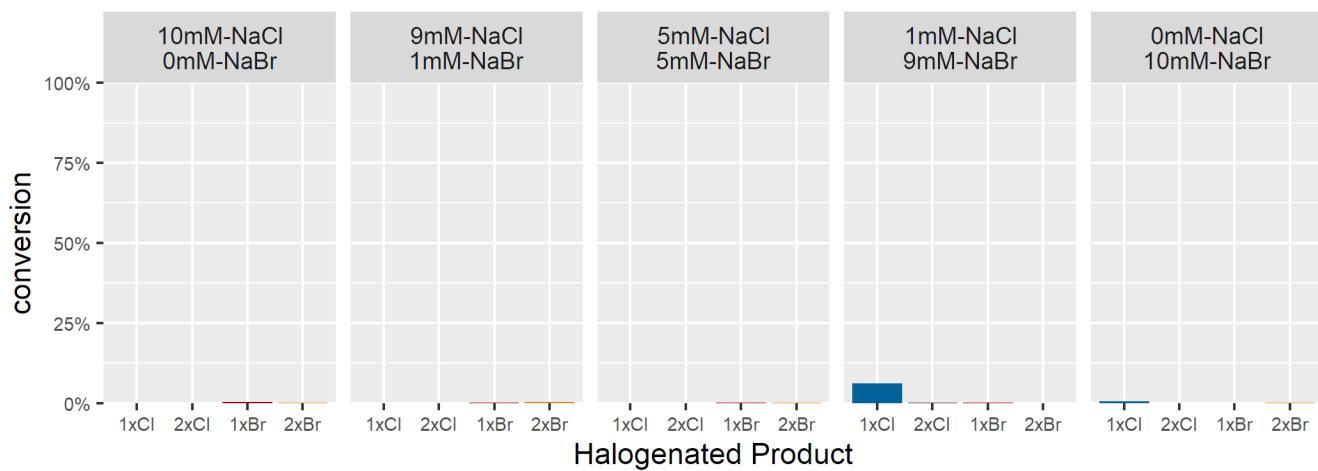


Figure S32. Halide selectivity of 2-C01 in L-Trp halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

FDH Halide Preferences

FDH: 2-C01; substrate: 1-phenylpiperazine; conversions from EIC area

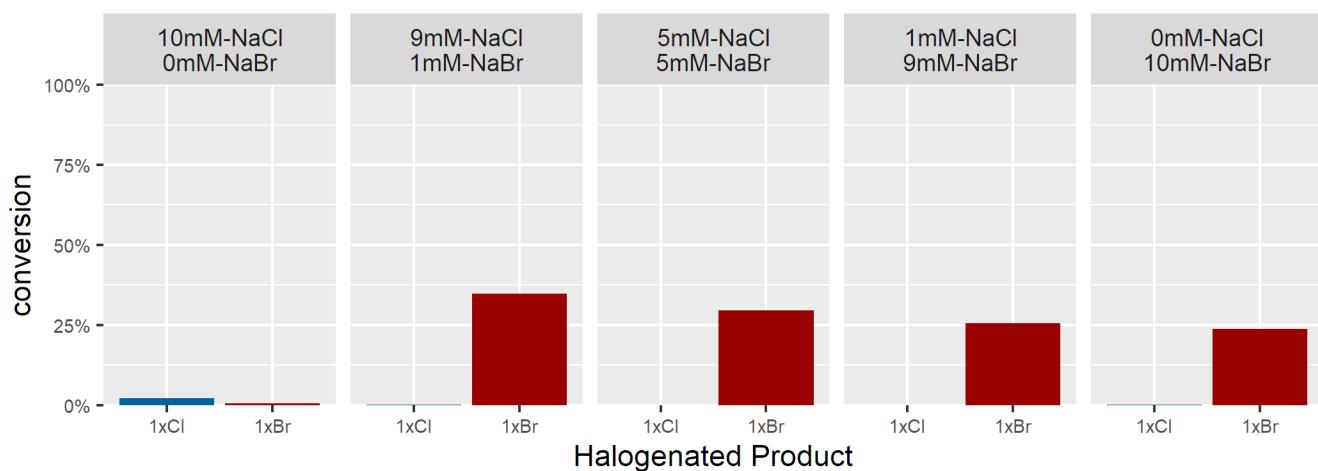


Figure S33. Halide selectivity of 2-C01 in 1-phenylpiperazine halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products.

FDH Halide Preferences

FDH: 2-C01; substrate: pindolol; conversions from EIC area

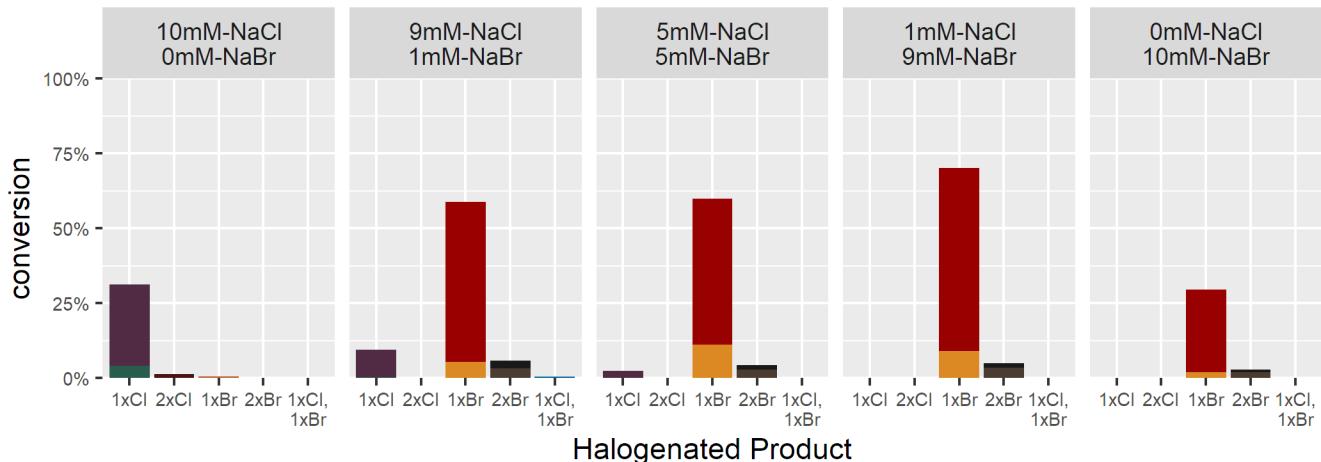


Figure S34. Halide selectivity of 2-C01 in pindolol halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product m/z).

FDH Halide Preferences

FDH: 2-C01; substrate: 2,4-DHAP; conversions from EIC area

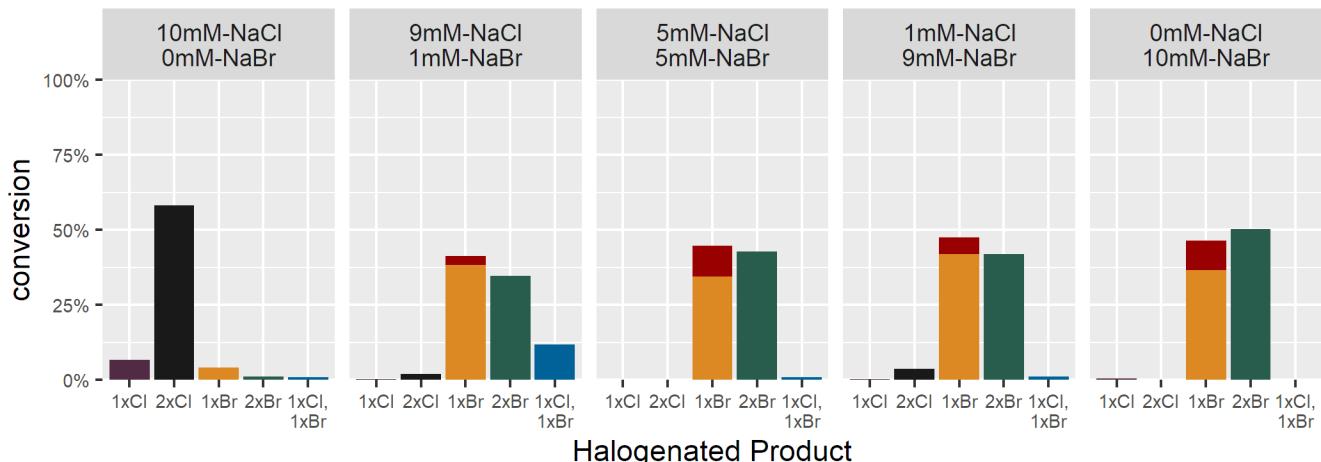


Figure S35. Halide selectivity of 2-C01 in 2,4-dihydroxyacetophenone halogenation using conversion from EIC integrals. Concentration of NaCl and NaBr are given above each plot. Different bars represent conversion to different halogenated products. Different colored portions of each bar represent different halogenated regioisomers (i.e., different peaks in the EIC chromatogram for a particular halogenated product m/z).

D) Lysate Bioconversions

Lysate bioconversion procedure

1 mL TB in 2 mL 96-well deep well plate were inoculated using 50 μ L from overnight cultures of BL21(DE3) + pGro7 + pET28b(FDH). The inoculated 2 mL 96-well deep well plate was incubated at 37

$^{\circ}\text{C}$ and shaken at 225 rpm until OD \approx 0.6-0.8, at which point the incubator was cooled to 18 $^{\circ}\text{C}$. After cooling for 30 min, protein expression was induced with 2 mg/mL L-arabinose and 100 μM IPTG. The induced plate was shaken for 18 hr, at which point the cells were pelleted by centrifugation at 3600 rpm, 4 $^{\circ}\text{C}$ for 15 min. Media supernatant was discarded, and cells were resuspended in 100 μL of lysis buffer (0.75 mg/mL lysozyme, 0.5 mg/mL DNase I in 25 mM HEPES, pH 7.4). Cells were lysed at 37 $^{\circ}\text{C}$ shaking at 225 rpm for 45 min, after which point the lysate was snap-frozen by floating the deep well plate on liquid N₂. The frozen plate was allowed to thaw at room temperature for 10 min, then thawed further by floating in a 37 $^{\circ}\text{C}$ water bath. Lysate was clarified by centrifuging at 3600 rpm, 4 $^{\circ}\text{C}$ for 15 min, and 15 μL of the soluble fraction of the lysate was added to each reaction such that the final volume of each reaction was 45 μL and the concentrations of all components of the reaction were as follows in Table S3.

Table S3. Reaction conditions for testing lysate bioconversions.

Component	Concentration
NaBr	10 mM
glucose	20 mM
NAD	100 μM
FAD	100 μM
GDH	9.0 U/mL
catalase	35 U/mL
MBP-RebF	2.5 μM
substrate	500 μM

Lysate bioconversion tests: poorly-soluble FDHs

Early experiments suggested that some enzymes had poor activity in lysate but were active in purified form. Because of this, and because we also did not want activity data to be overly biased toward enzymes that expressed well, we sought to obtain purified enzymes and evaluate all enzymes at standard concentrations. It is possible, however, that in prioritizing more soluble enzymes for further study, highly active but perceptibly insoluble FDHs might have been overlooked. Indeed, Goss reports that for an unusual and poorly soluble multidomain FDH, KrmI, activity was more easily observed in lysate.²² We therefore attempted bioconversion using the lysate of *E. coli* overexpressing a selection of insoluble and poorly-expressed FDHs. Some soluble FDHs were also included in this study, including active enzymes such as 1-F08 and enzymes for which no activity was observed in the high-throughput screen including 1-G07 (Table S4). Substrates for the lysate bioconversion tests were mostly drawn from the most easily halogenated compounds in the probe substrate screen (2,4-dihydroxyacetophenone, [2,4-DHAP], 6-hydroxyquinoline, and pindolol). One expanded screen substrate, UC-054(EtNH₂), was easily halogenated and was also included. Premalbrancheamide was included specifically to test for halogenation by 3-B02, an insoluble enzyme closely related to the FDH MalA, which natively dichlorinates premalbrancheamide.²³

Table S4. FDHs selected for lysate bioconversion tests.

FDH	Solubility	Activity
1-F08	Soluble	Active
1-F05	Soluble	Active
1-C09	Soluble	Active
1-G07	Soluble	Inactive
1-B04	Soluble	Inactive
2-G02	Soluble	Inactive
1-C06	Not expressed	-
3-H02	Not expressed	-
1-E01	Not expressed	-
2-C03	Not expressed	-
3-B02	Insoluble	-
1-D04	Insoluble	-
3-A03	Insoluble	-
1-H05	Insoluble	-
2-D02	Insoluble	-
2-F02	Insoluble	-

Only two enzymes that were not assigned as active and soluble, 1-D04 and 3-H02, exhibited any conversion in the lysate bioconversion screens (Table S5, Figure S36). Both brominated pindolol, but only to 1.3% and 1.5% conversion, respectively. No bromination of the other probe substrates by these two enzymes was observed, suggesting that the substrate scope of these enzymes is likely to be narrow. Overall, the experiments suggested that extensive work outside the scope of the present study would likely be required to observe significant halogenation activity among the more difficult to work with proteins in the genome-mined set.

Table S5. Lysate bioconversion results. Conversion determined by integration of the UV chromatogram at 220 nm.

Substrate	FDH Expression	FDH Activity	FDH	Overall Conversion	
2,4-DHAP	Insoluble	-	1-D04	0.0%	
			1-H05	0.0%	
			2-D02	0.0%	
			2-F02	0.0%	
			3-A03	0.0%	
	Not expressed		3-B02	0.0%	
			1-C06	0.0%	
			1-E01	0.0%	
			2-C03	0.0%	

			3-H02	0.0%
6-hydroxyquinoline	Soluble	Active	1-C09	0.9%
			1-F05	5.4%
			1-F08	8.6%
		Inactive	1-B04	0.0%
			1-G07	0.0%
			2-G02	0.0%
			1-D04	0.0%
pindolol	Insoluble	-	1-H05	0.0%
			2-D02	0.0%
			2-F02	0.0%
			3-A03	0.0%
			3-B02	0.0%
			1-C06	0.0%
			1-E01	0.0%
			2-C03	0.0%
	Not expressed	Active	3-H02	0.0%
			1-C09	10.3%
			1-F05	19.6%
			1-F08	100.0%
	Soluble	Inactive	1-B04	0.0%
			1-G07	0.0%
			2-G02	0.0%
			1-D04	1.3%
	Insoluble	-	1-H05	0.0%
			2-D02	0.0%
			2-F02	0.0%
			3-A03	0.0%
	Not expressed	-	3-B02	0.0%
			1-C06	0.0%
			1-E01	0.0%

premalbrancheamide	Soluble	Active	2-C03	0.0%
			3-H02	1.5%
			1-C09	2.4%
			1-F05	0.0%
			1-F08	88.8%
		Inactive	1-B04	0.0%
			1-G07	0.0%
			2-G02	0.0%
	Insoluble	-	1-D04	0.0%
			1-H05	0.0%
			2-D02	0.0%
			2-F02	0.0%
			3-A03	0.0%
UC-054(EtNH2)	Not expressed	-	3-B02	0.0%
			1-C06	0.0%
			1-E01	0.0%
			2-C03	0.0%
			3-H02	0.0%
	Soluble	Active	1-C09	0.0%
			1-F05	0.0%
			1-F08	33.6%
		Inactive	1-B04	0.0%
			1-G07	0.0%
			2-G02	0.0%
UC-054(EtNH2)	Insoluble	-	1-D04	0.0%
			1-H05	0.0%
			2-D02	0.0%
			2-F02	0.0%
			3-A03	0.0%
			3-B02	0.0%
			1-C06	0.0%

			1-E01	0.0%
			2-C03	0.0%
			3-H02	0.0%
Soluble	Active		1-C09	83.5%
			1-F05	0.0%
	Inactive		1-F08	70.7%
			1-B04	0.0%
	Inactive		1-G07	0.0%
			2-G02	0.0%

Lysate Tests

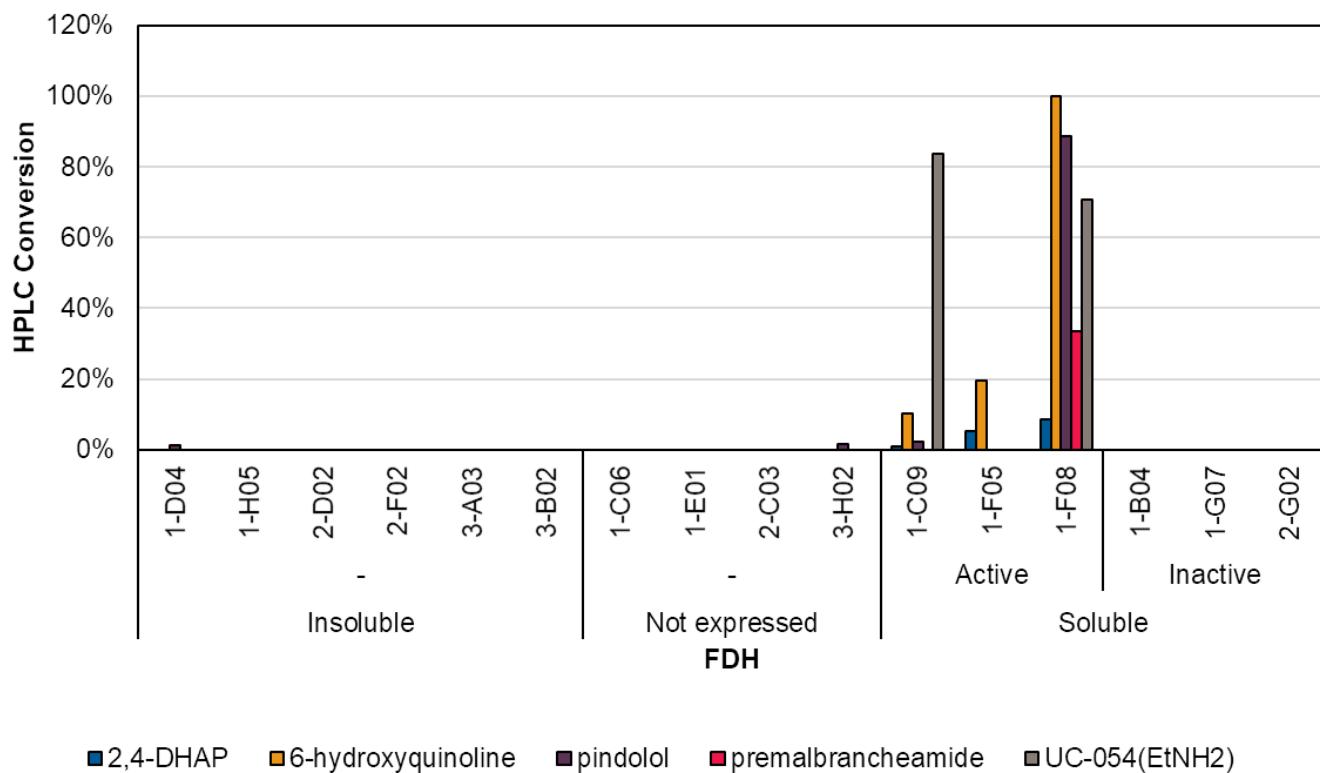


Figure S36. Analysis of lysate bioconversion test reactions. Conversion determined by integration of the UV chromatogram at 220 nm. FDHs are grouped by expression observed in expression tests, and soluble enzymes are further subdivided into enzymes that had activity in the high-throughput screen and those that did not.

Lysate bioconversion tests: 1-F11 vs. RebH

Bromination of several substrates was tested using lysate of 1-F11 and RebH to directly compare their catalytic capabilities under conditions similar to those that would be used for directed evolution. Six substrates were selected, five of which 1-F11 was found to brominate in the high-throughput screen and one of which (carvedilol) was not evaluated in the high-throughput screen but tested later (Figure S37).

RebH (codon-optimized, with C-terminal His-tag, like others in the genome-mined set) was found to brominate pindolol in the HTS, but not the other four compounds; its bromination activity on carvedilol had not been characterized previously.

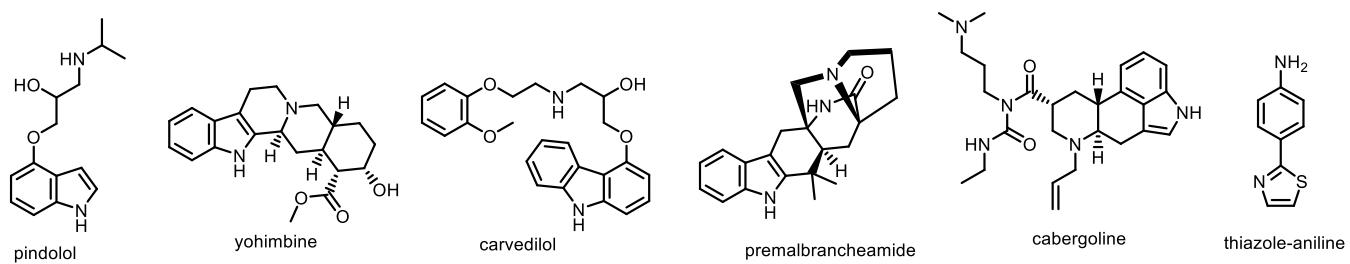


Figure S37. Substrates selected for lysate bioconversion tests to compare RebH and 1-F11 activity.

Bioconversions were conducted as described above in triplicate. Time points were taken at 1 hr and 26 hr. Conversion was calculated using UV chromatograms (for pindolol, yohimbine, and carvedilol: 220 nm; for premalbrancheamide, cabergoline, and thiazole-aniline, 280 nm) by adding integrals of brominated products and dividing by the sum of the integrals of brominated products and starting material.

Table S6. Bioconversion results for lysate bioconversions of RebH and 1-F11.

Time (hr)	Substrate	FDH	Conversion	Conversion std. dev.
1	pindolol	1-F11	9.4%	0.5%
		RebH	4.9%	2.2%
	yohimbine	1-F11	3.6%	0.6%
		RebH	0.0%	
	carvedilol	1-F11	17.2%	0.6%
		RebH	2.1%	1.7%
	premalbrancheamide	1-F11	0.0%	
		RebH	0.0%	
	cabergoline	1-F11	1.1%	0.1%
		RebH	0.0%	
	thiazole-aniline	1-F11	ND ^a	ND ^a
		RebH	0.0%	
26	pindolol	1-F11	71.4%	4.5%
		RebH	21.3%	12.9%
	yohimbine	1-F11	8.8%	0.8%
		RebH	0.0%	
	carvedilol	1-F11	96.7%	2.4%
		RebH	5.9%	2.1%
	premalbrancheamide	1-F11	1.7%	0.2%
		RebH	0.3%	0.2%
	cabergoline	1-F11	2.1%	0.4%
		RebH	0.0%	
	thiazole-aniline	1-F11	6.8%	0.4%
		RebH	0.0%	

^aInjection error resulted in no data collection for this time point.

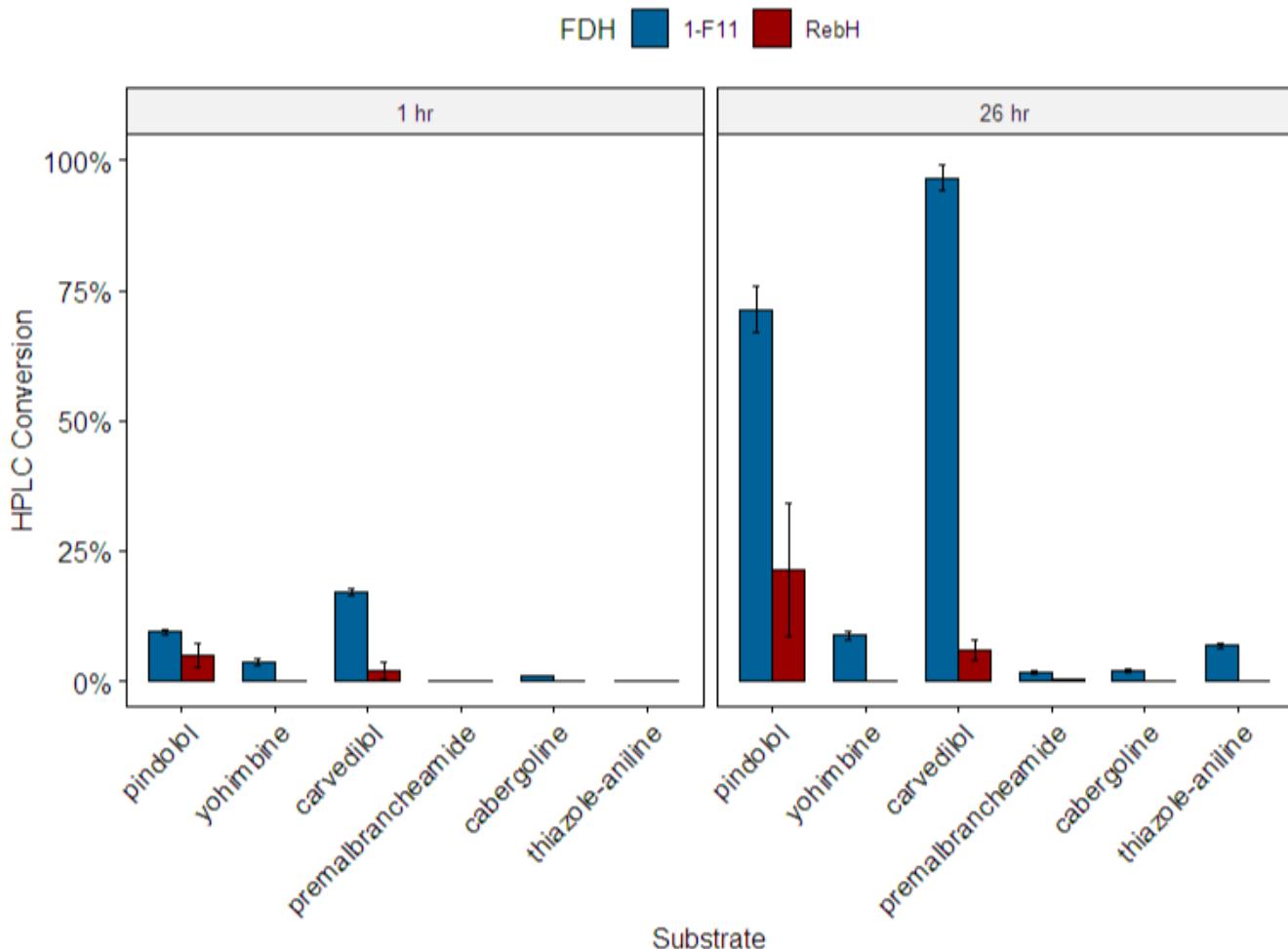


Figure S38. Conversions in reactions using RebH or 1-F11 added as lysate.

VIII. Halenium Affinity Calculations

A) Halenium Affinity Calculation Setup

Procedure for HalA calculations was adapted from Borhan *et al.*¹⁸ 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 **Equation S2**.

$$HalA = -H^\circ_{Sub-X} + H^\circ_{Sub} + H^\circ_{X+} + \frac{5}{2}RT \quad \text{Equation S2}$$

For the sake of speed, only chlorenium affinity was computed for the substrates reported below, but the overall trends for chlorenium and for bromenium affinities of different positions in simpler test compounds were qualitatively identical (data not shown). Geometry optimization for particularly large and flexible compounds did not reach equilibrium within several days of window time. The structures of these compounds were therefore simplified. The carbazole and catechol rings were analyzed separately. Different stereoisomers at the halogenated carbon were tested in cases where the halogenated carbon was judged to be near to a stereocenter.

B) Halenium Affinity Data

Structures used in halenium affinity calculations

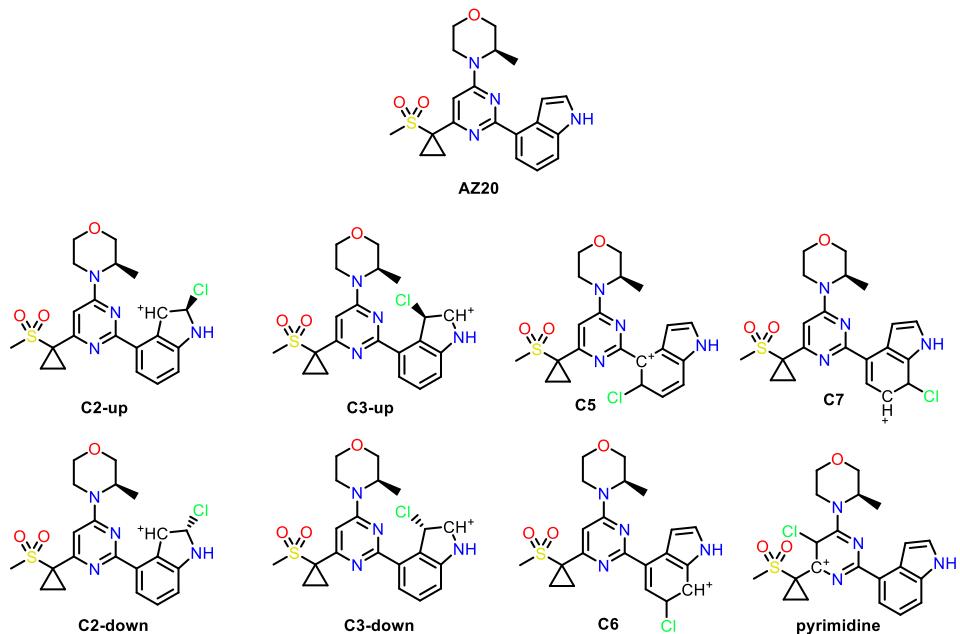


Figure S39. Structures used in halenium affinity calculations of AZ20.

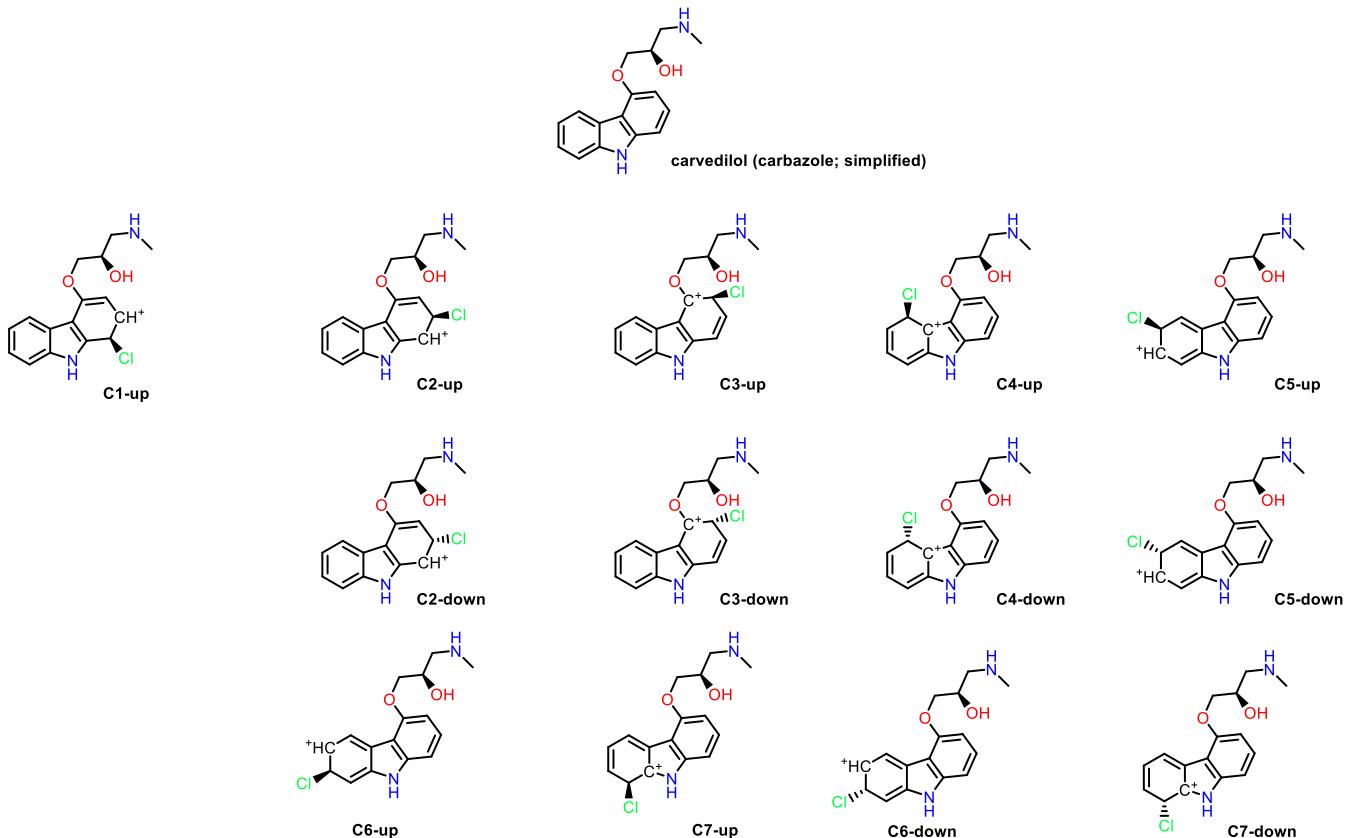


Figure S40. Structures used in halonium affinity calculations of the carbazole ring of carvedilol.

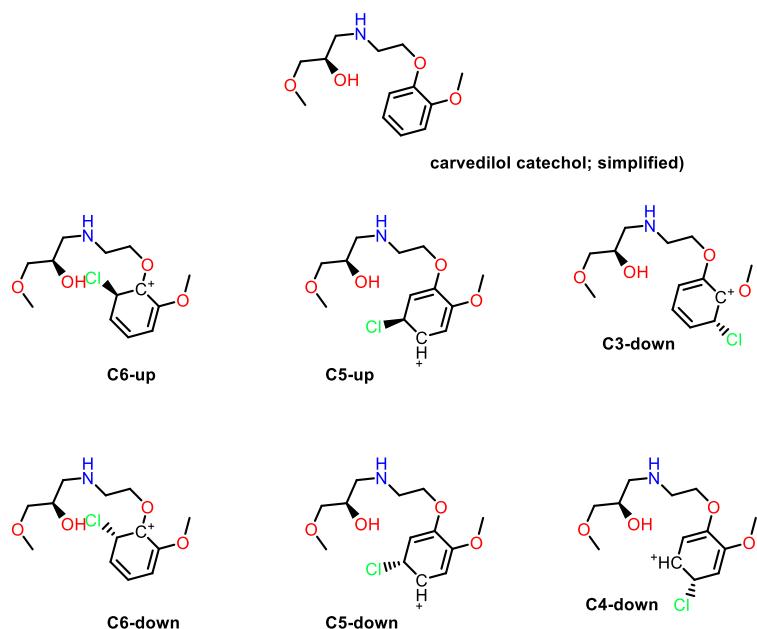


Figure S41. Structures used in halonium affinity calculations of the catechol ring of carvedilol.

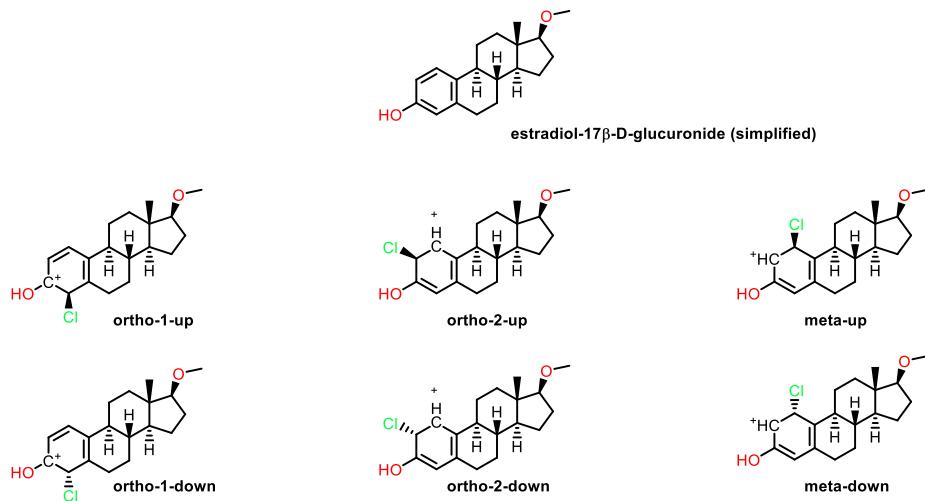


Figure S42. Structures used in halonium affinity calculations of estradiol-17 β -D-glucuronide.

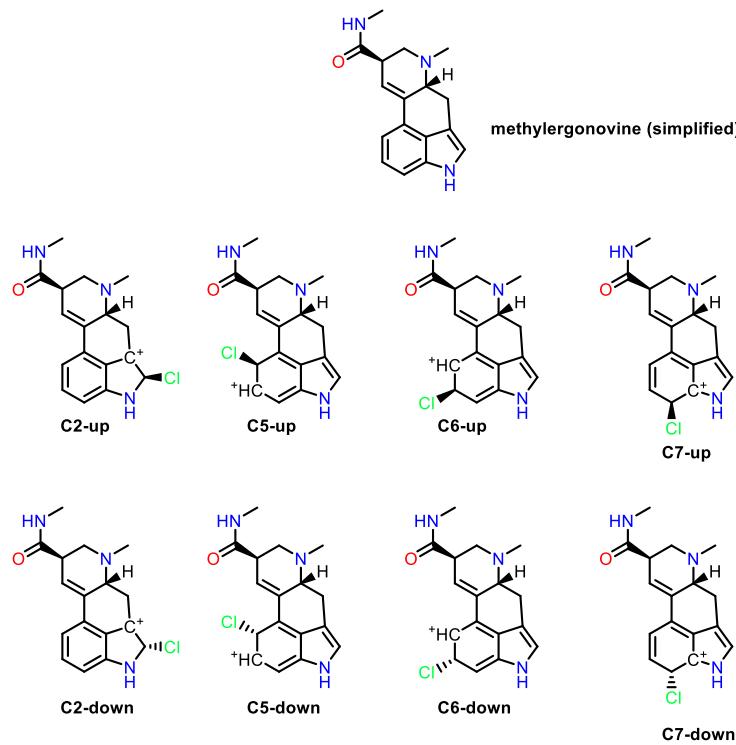


Figure S43. Structures used in halonium affinity calculations of methylergonovine.

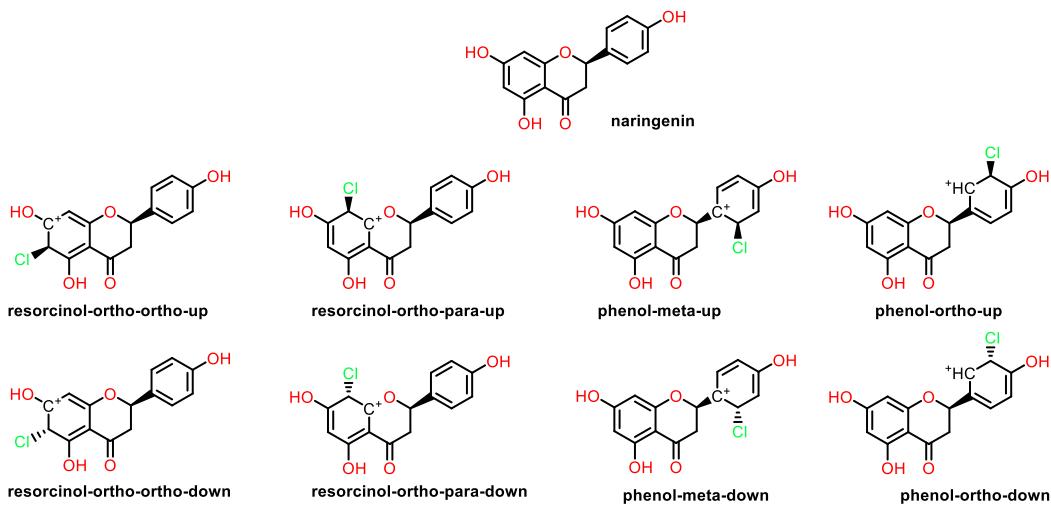


Figure S44. Structures used in halogen affinity calculations of naringenin.

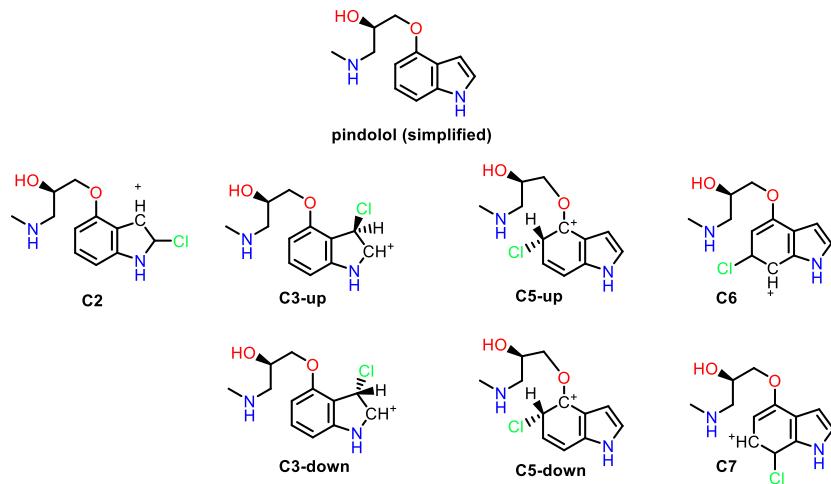


Figure S45. Structures used in halogen affinity calculations of pindolol.

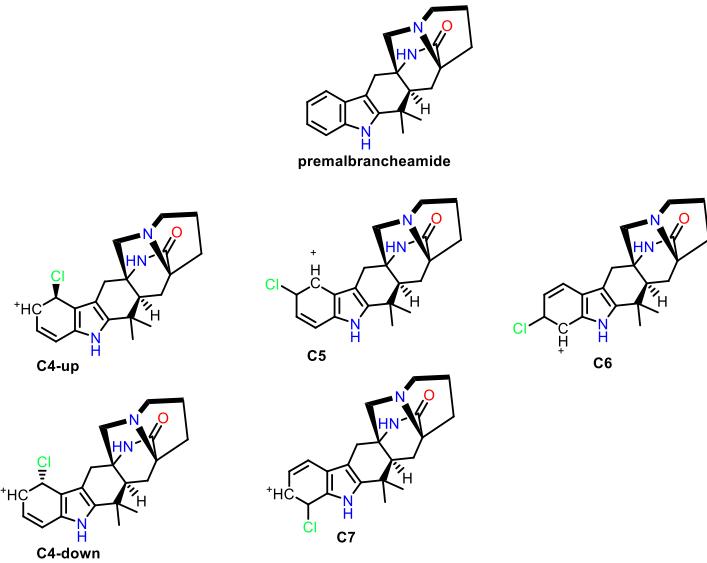


Figure S46. Structures used in halenium affinity calculations of premalbrancheamide.

Halenium affinity values

Table S7. Halenium affinity data for selected substrates.

Compound	Position	Halenium Affinity (kcal/mol)
AZ20	C2-down	177.09
	C2-up	175.06
	C3-down	184.50
	C3-up	184.61
	C5	172.16
	C6	172.54
	C7	172.97
	pyrimidine	184.08
carvedilol	C1-up	188.30
	C2-down	174.13
	C2-up	174.59
	C3-down	183.89
	C3-up	183.53
	C4-up	175.44
	C5-down	174.64
	C5-up	180.28
	C6-down	179.59
	C7-down	177.55
	C7-up	177.54
	C8-down	177.53
	C8-up	177.24
	catechol-C2-up	166.07

	catechol-C3-down	170.26
	catechol-C4-down	175.33
	catechol-C4-up	175.13
	catechol-C5-down	166.80
	catechol-C5-up	171.88
	catechol-C6-down	168.79
	catechol-C6-up	171.00
	meta-down	155.91
	meta-up	152.81
estradiol-17β-D-glucuronide	ortho-1-down	168.49
	ortho-1-up	168.92
	ortho-2-down	169.00
	ortho-2-up	169.00
	C2-down	188.20
	C2-up	188.36
methylergonovine	C5-down	182.86
	C5-up	182.66
	C6-down	175.67
	C6-up	175.55
	C7-down	183.97
	C7-up	184.02
	phenol-meta-down	144.99
	phenol-meta-up	144.53
naringenin	phenol-ortho-down	154.59
	phenol-ortho-up	154.57
	resorcinol-ortho-ortho-down	175.41
	resorcinol-ortho-ortho-up	175.46
	resorcinol-ortho-para-down	176.13
	resorcinol-ortho-para-up	176.22
	C2	187.96
	C3-down	186.37
pindolol	C3-up	185.41
	C5-down	190.55
	C5-up	184.56
	C6	179.16
	C7	190.77

premalbrancheamide	C4-down	177.12
	C4-up	177.02
	C5	173.17
	C6	177.60
	C7-down	170.80

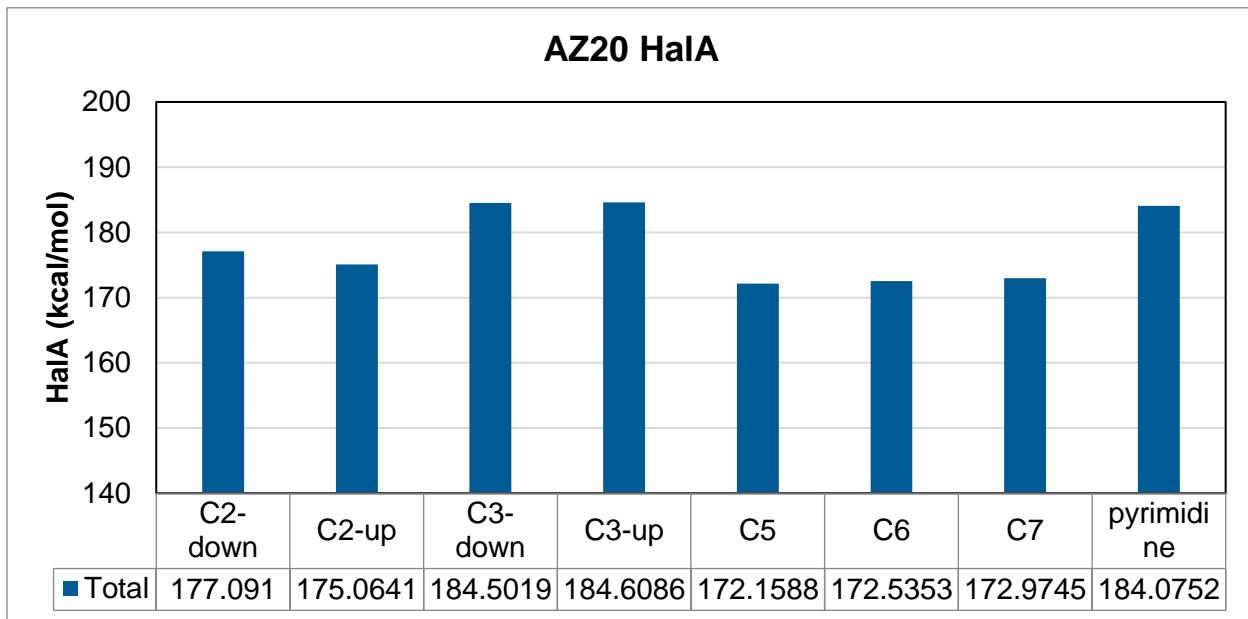


Figure S47. Chlorenium affinity data for AZ20.

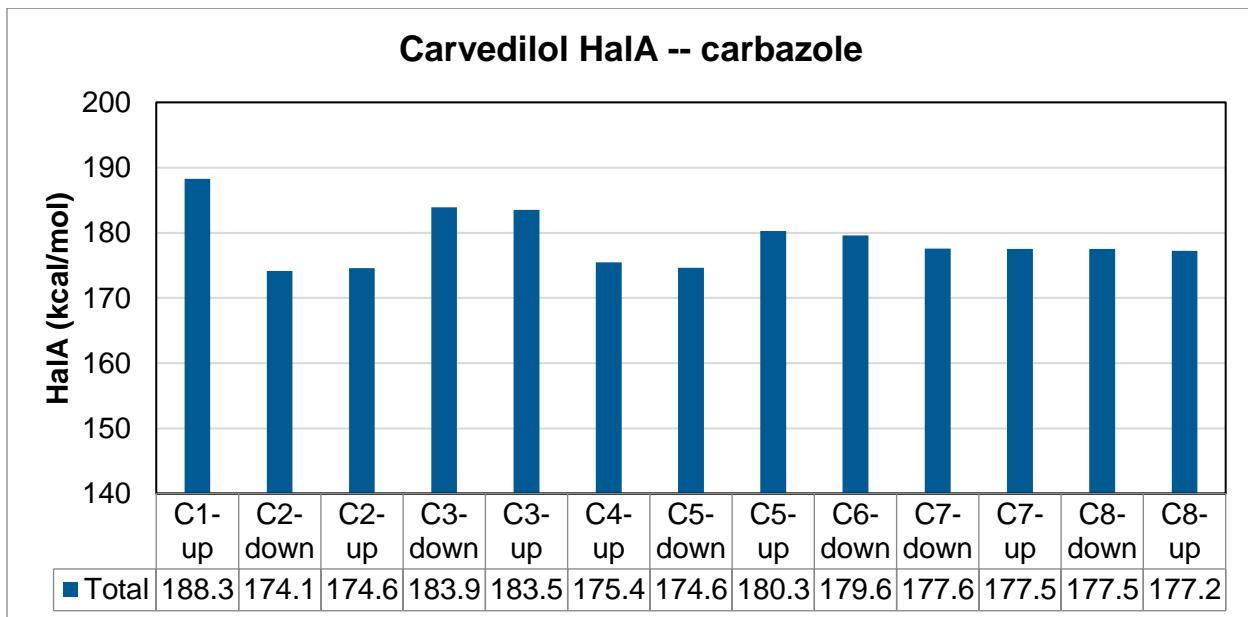


Figure S48. Chlorenium affinity data for the carbazole ring of carvedilol.

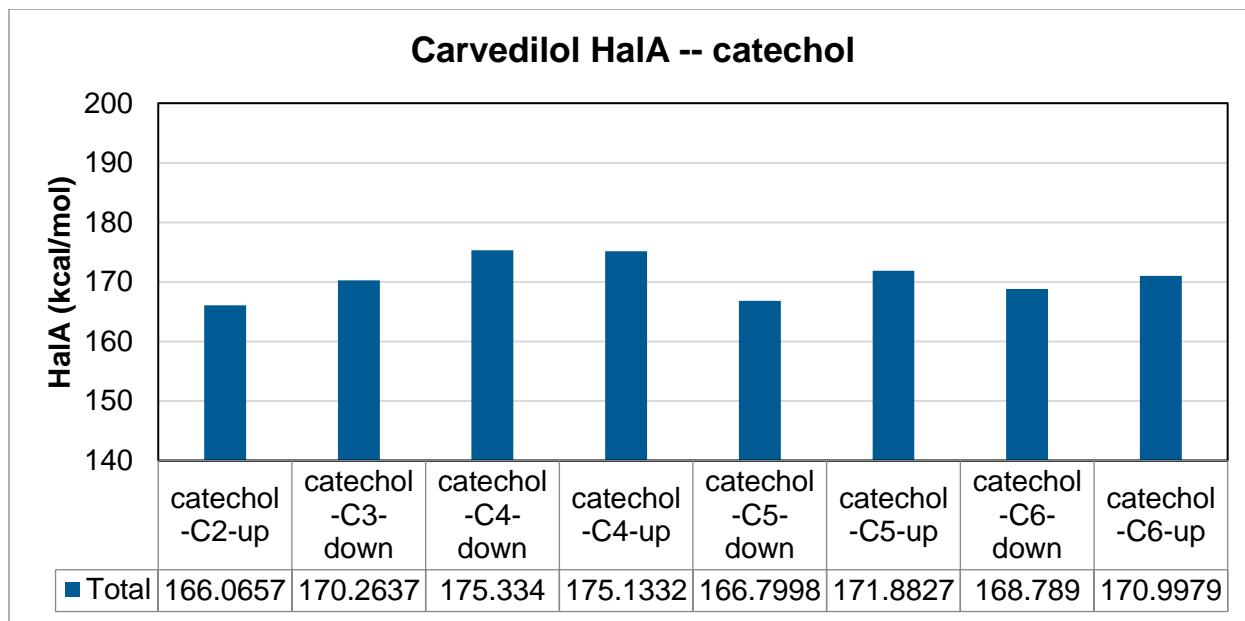


Figure S49. Chlorenium affinity data for the catechol ring of carvedilol.

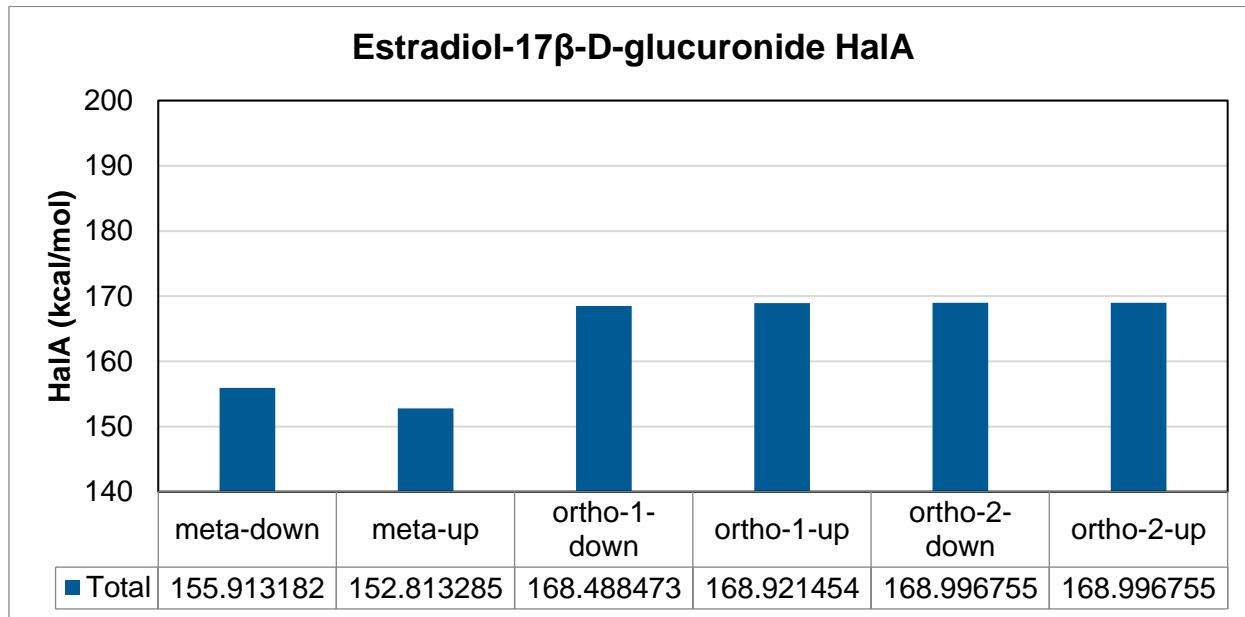


Figure S50. Chlorenium affinity data for estradiol-17 β -D-glucuronide.

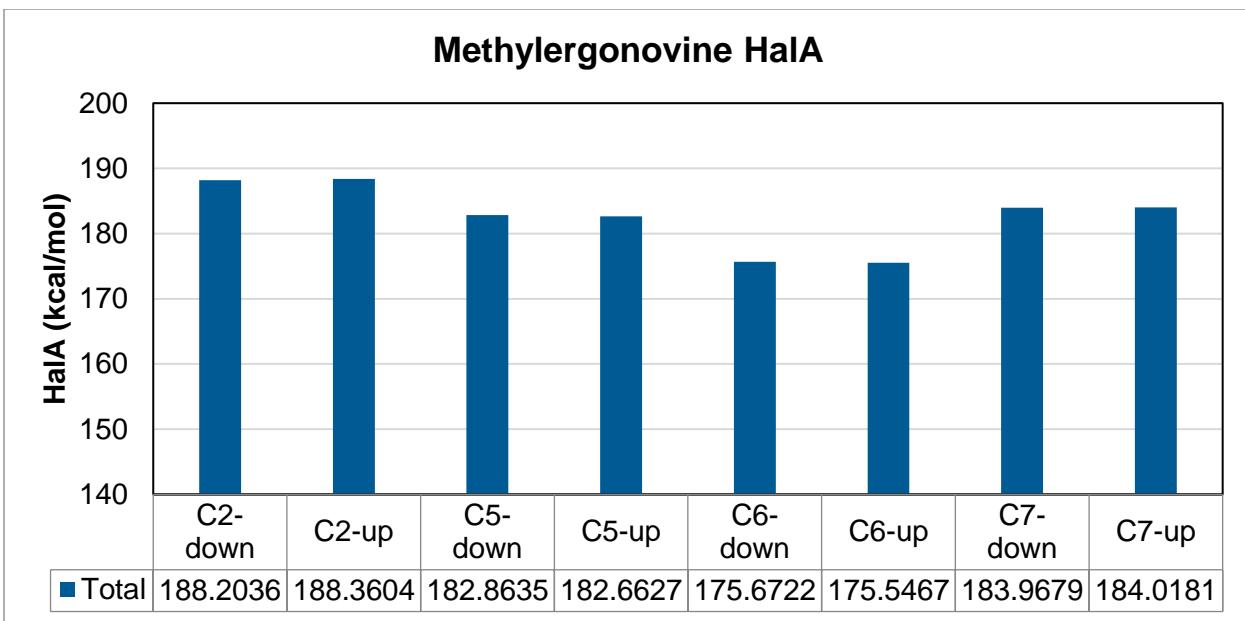


Figure S51. Chlorenium affinity data for methylergonovine.

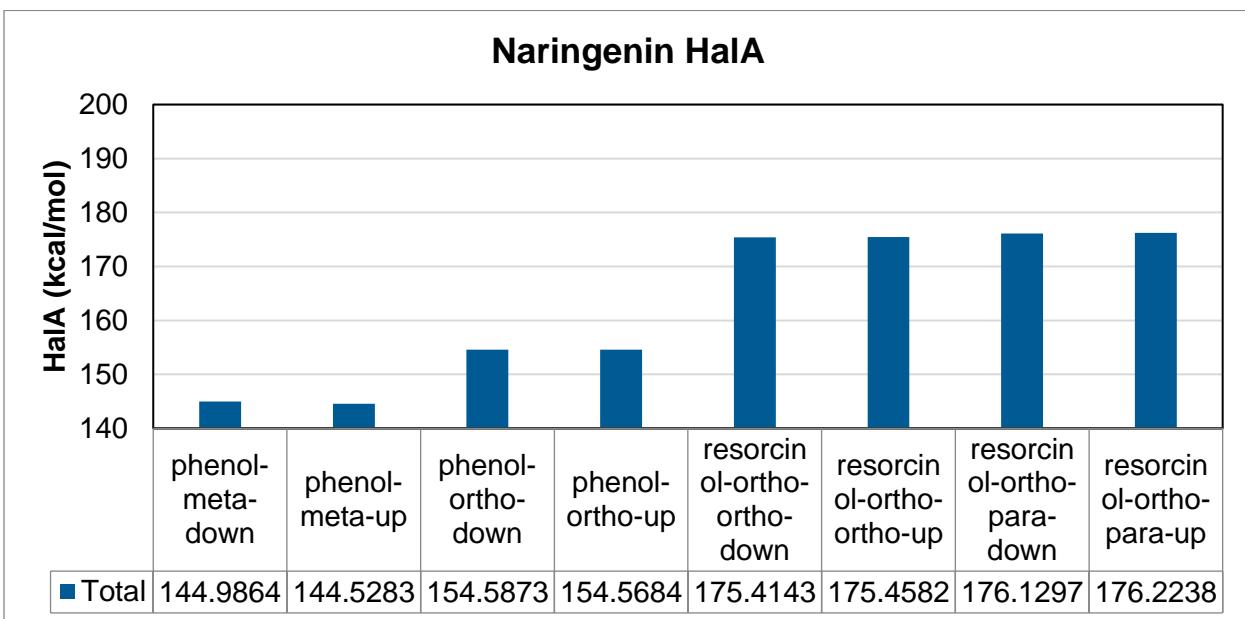


Figure S52. Chlorenium affinity data for naringenin.

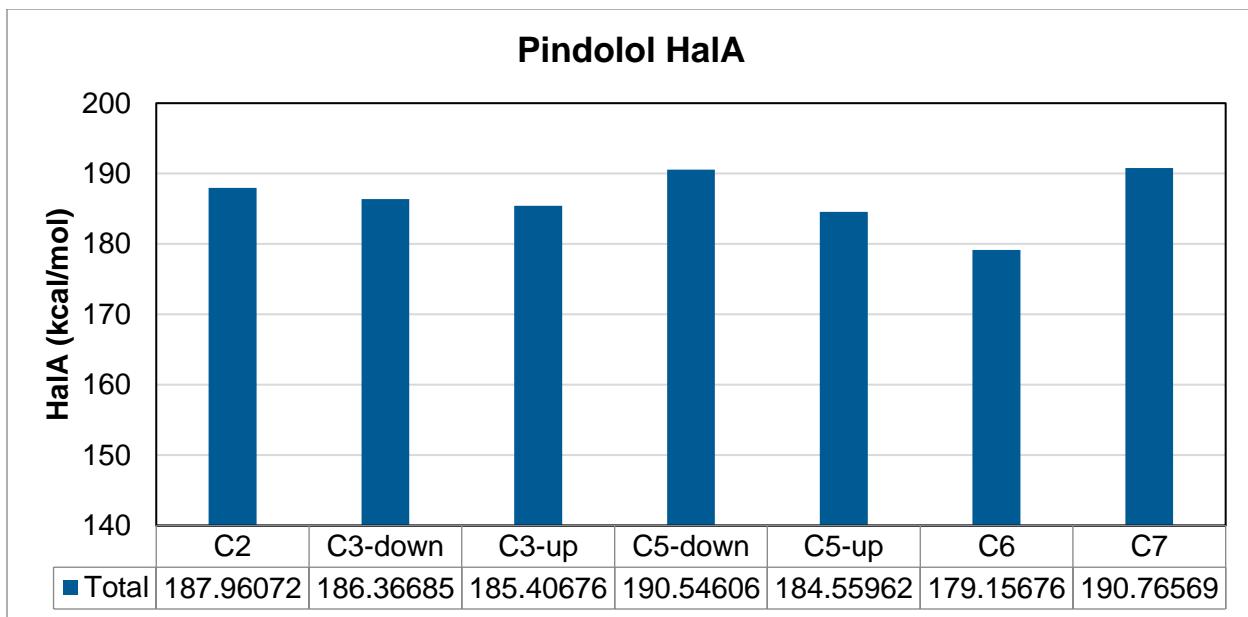


Figure S53. Chlorenium affinity data for pindolol.

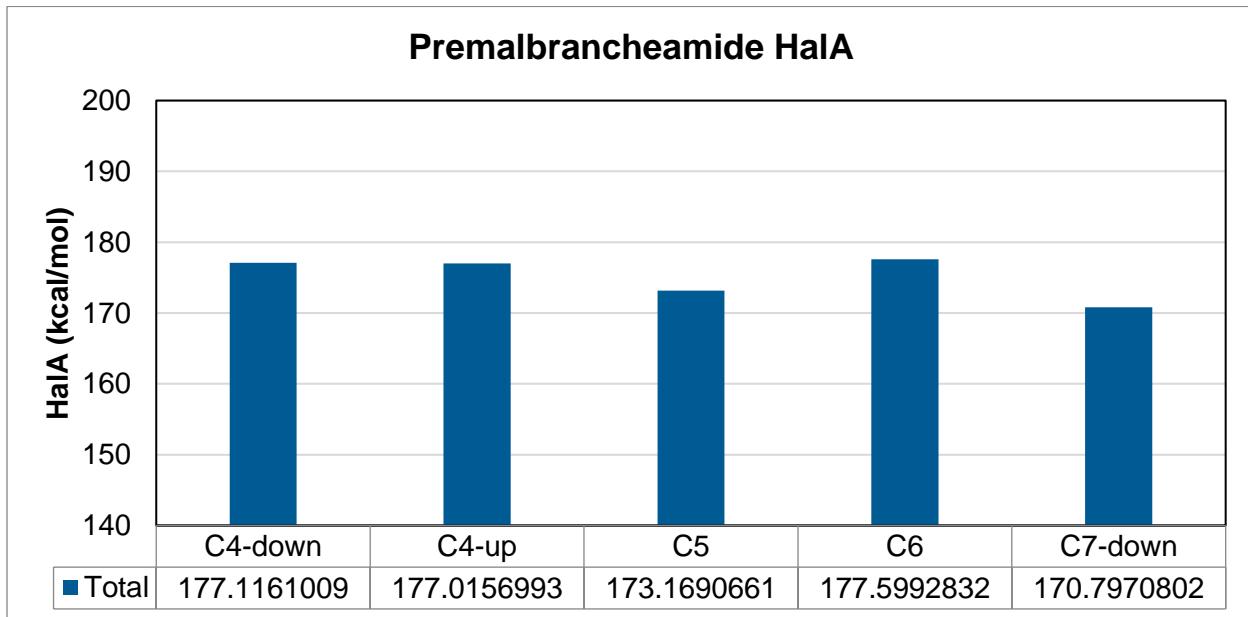


Figure S54. Chlorenium affinity data for premalbrancheamide.

IX. SDS-PAGE

A) Exploration of Alternative Expression Conditions

Autoinduction was also explored for improving soluble expression of genome-mined halogenases. Compared with chemical induction, the overall level of halogenase overexpression decreased without a perceptible change in fraction soluble. We therefore chose to continue expression using chemical induction in TB rather than autoinduction.

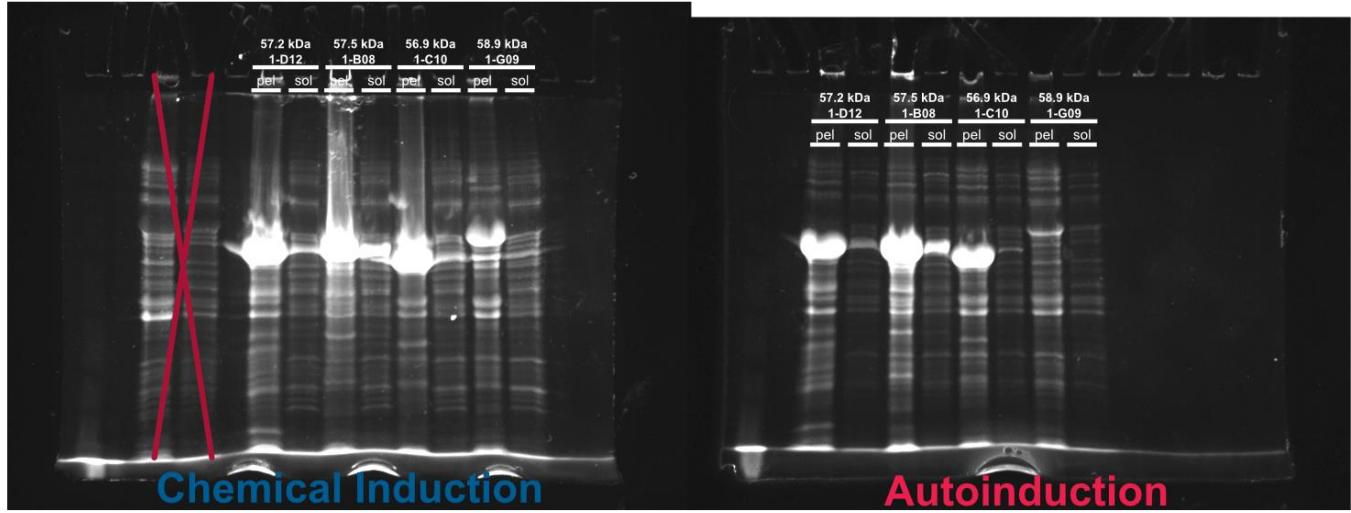
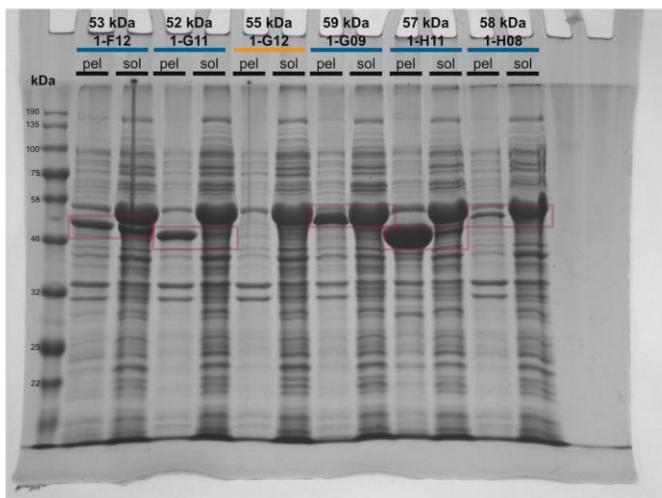
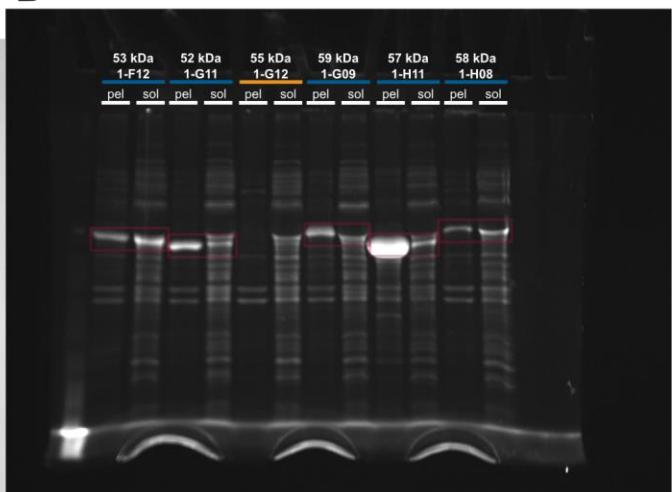
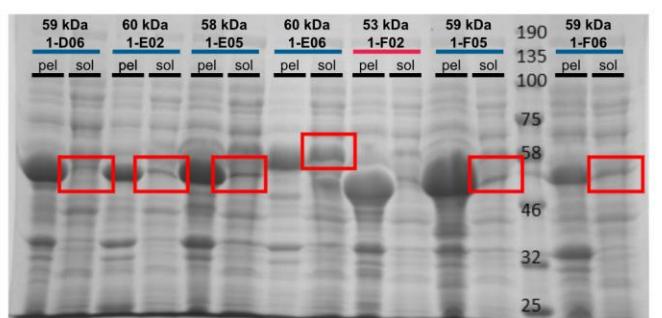
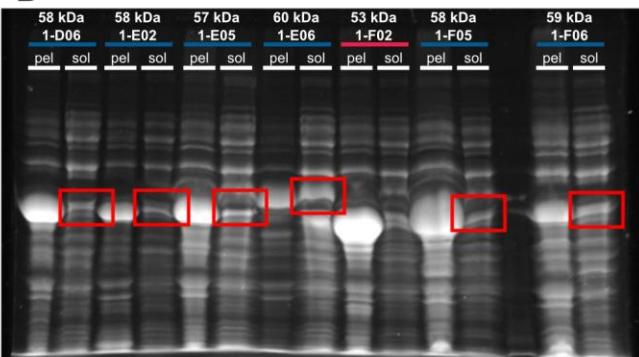


Figure S55. Comparison of chemical induction and autoinduction for expression of a selection of four halogenases. See Section III for expression conditions. Red X over lanes indicates lanes contain samples from an unrelated experiment.

B) Example SDS-PAGE Gels

A selection of SDS-PAGE gels used to assess the solubility of genome-mined enzymes is shown below in Figure S56.

A**B****C****D**

— Not expressed
— Insoluble
— Soluble

Figure S56. 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.

X. Circular Dichroism

A) CD Data Collection and Analysis

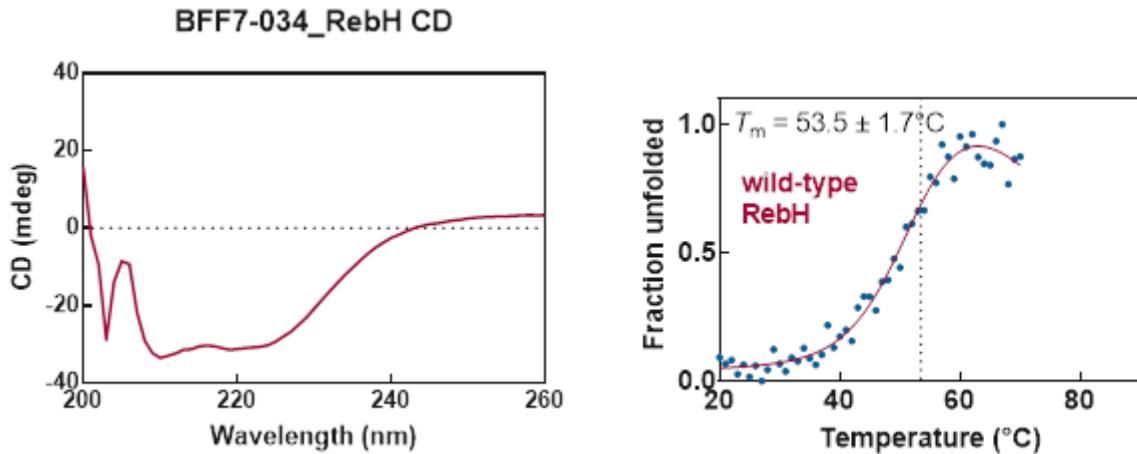
Frozen stocks of enzyme in 10% glycerol, 25 mM HEPES, pH 7.4 stored at -80°C were thawed on ice, then buffer exchanged twice in 4 mL Amicon Ultra 30K MWCO spin filters into 25 mM HEPES, pH 7.4. Enzyme stocks were transferred to ice cold Eppendorf tubes, which were centrifuged at 13.2 krpm, 4°C for 3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

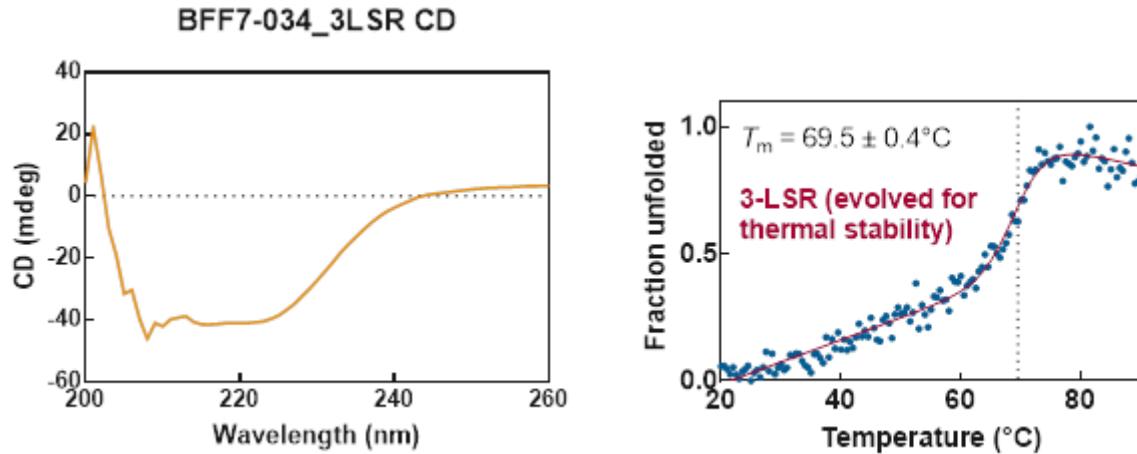
Circular dichroism thermal melts were processed using CDPal using the Autofit All function.²⁴

B) CD Data

RebH

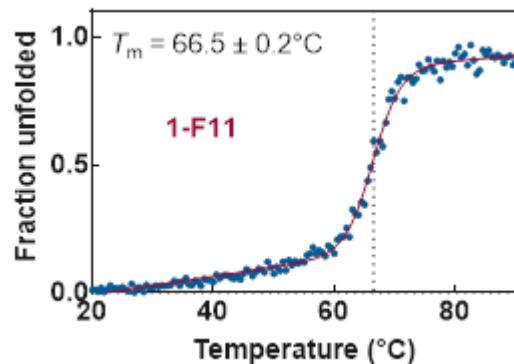
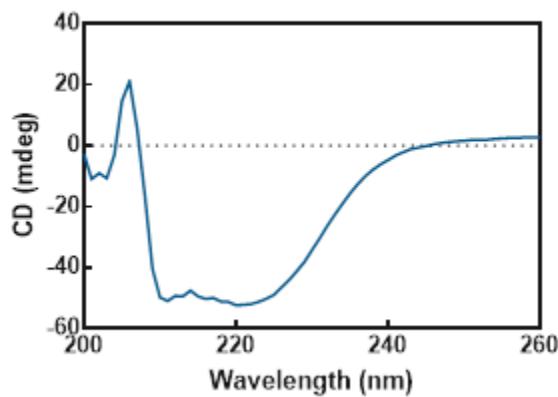


3-LSR



1-F11

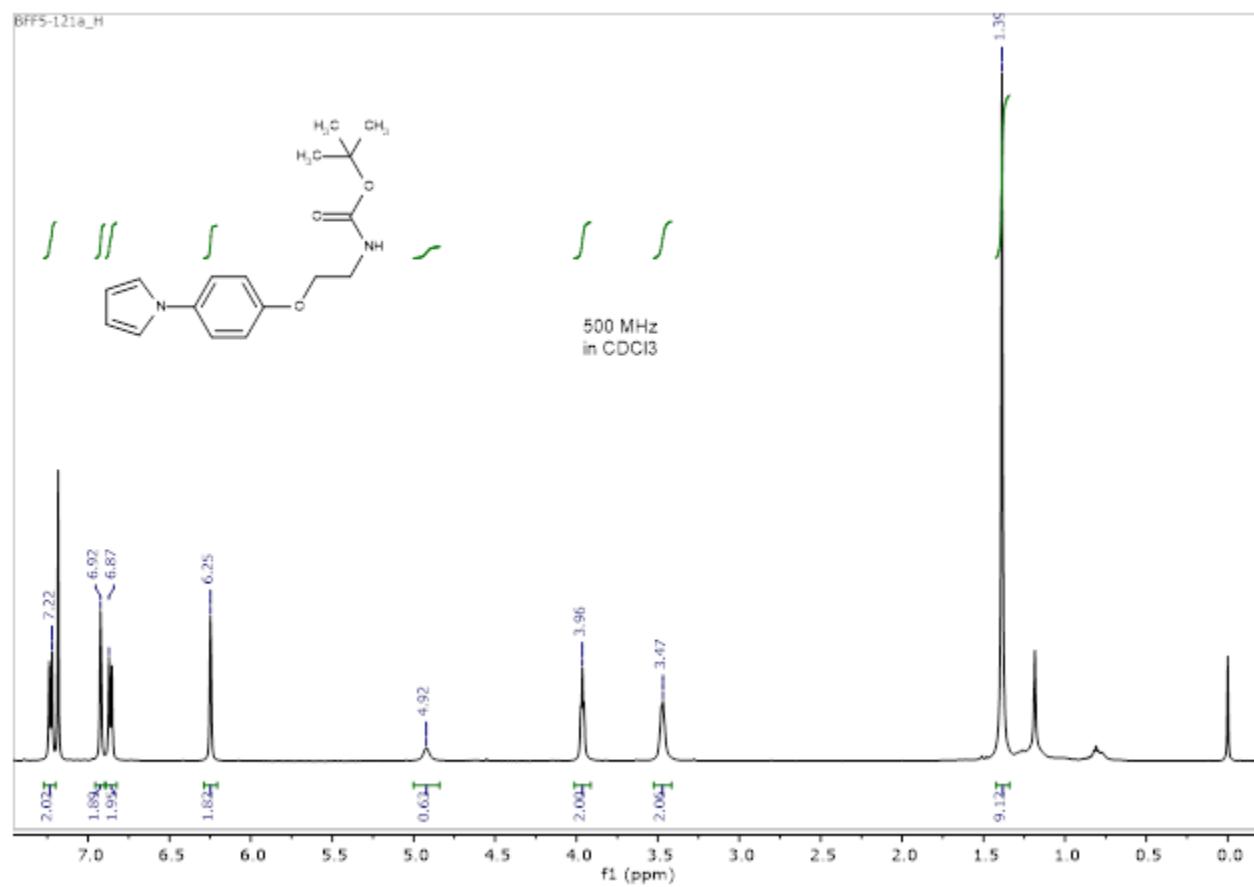
BFF7-034_1F11 CD



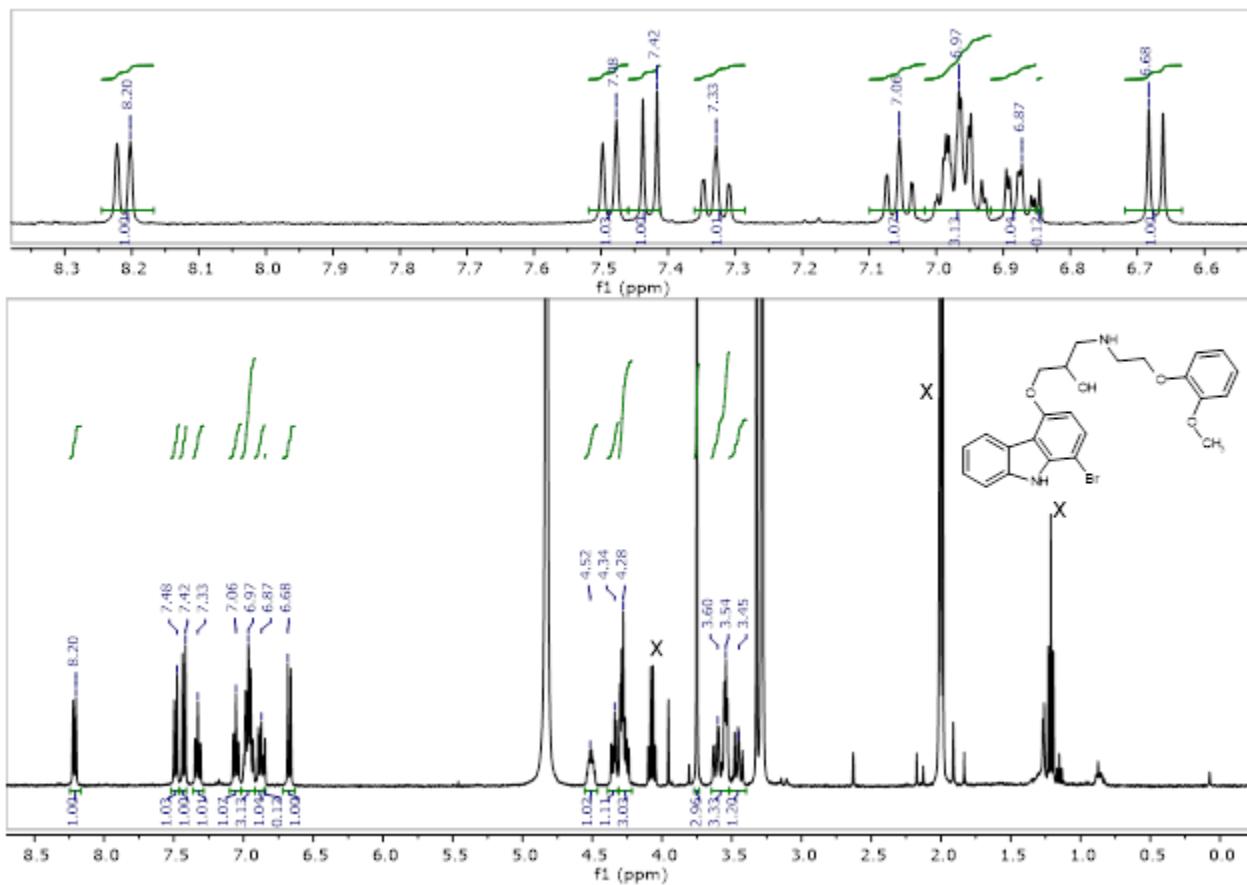
XI. NMR

A) $^1\text{H-NMR}$

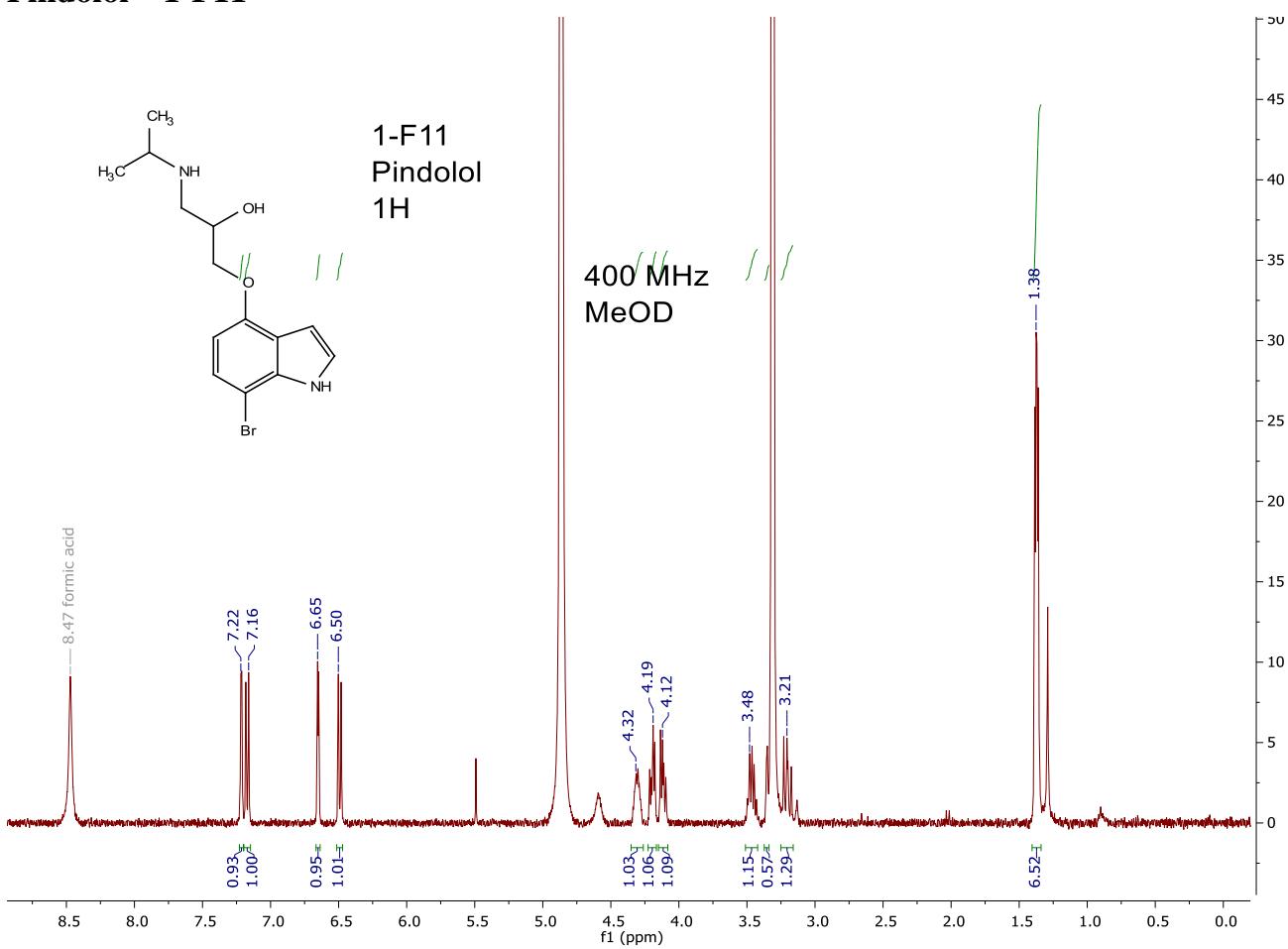
S2



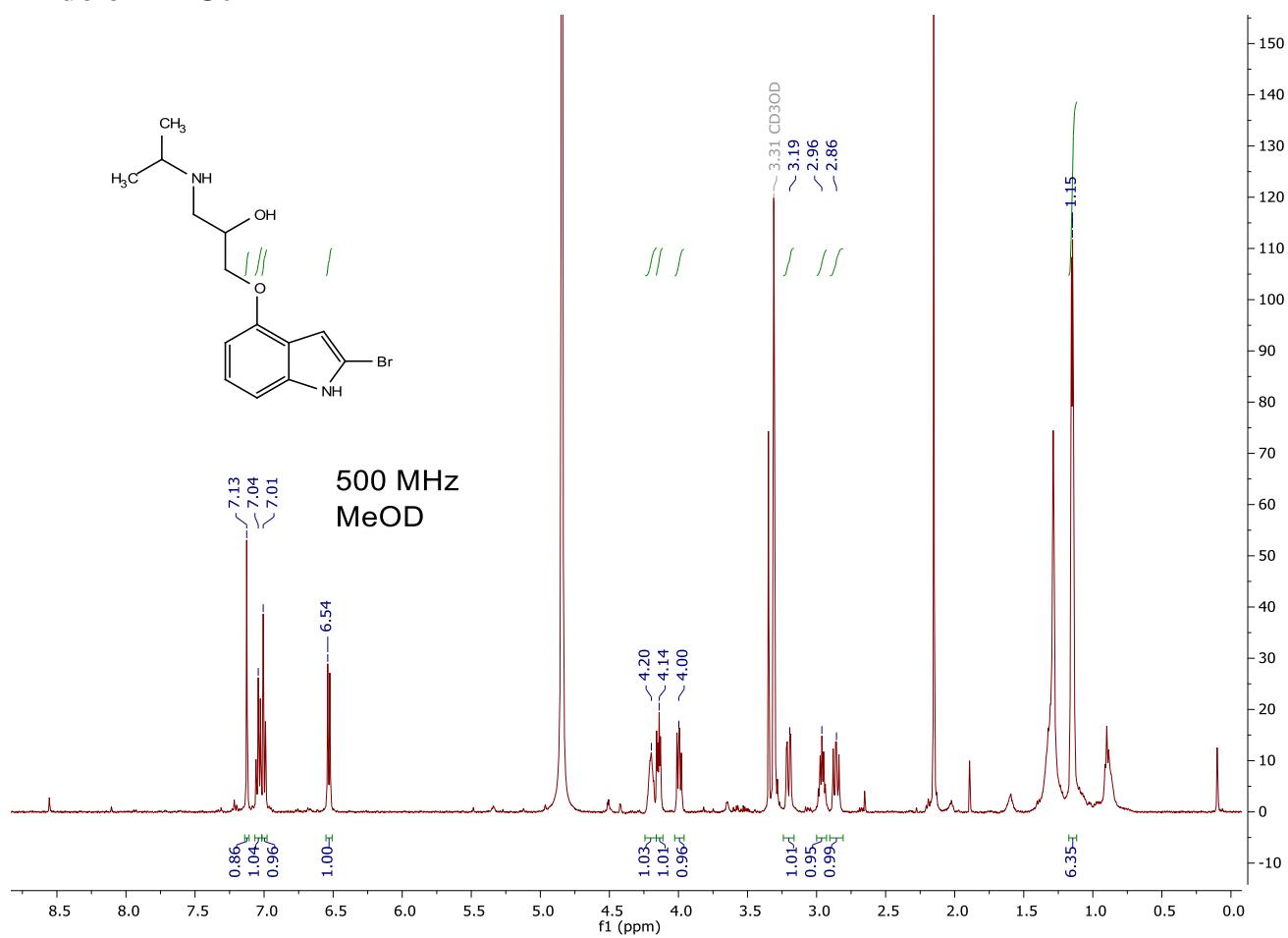
Carvedilol – 1-F11



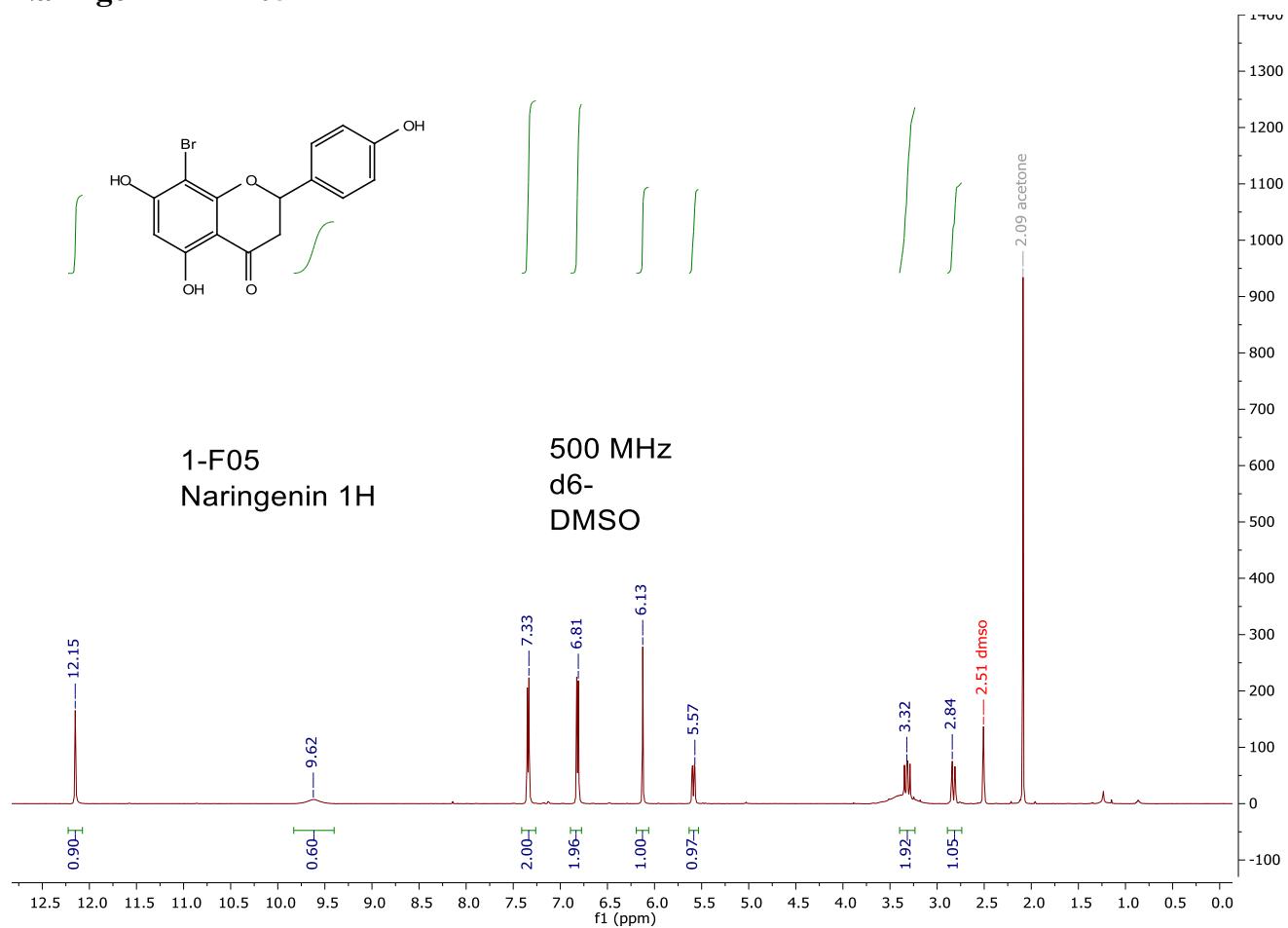
Pindolol – 1-F11



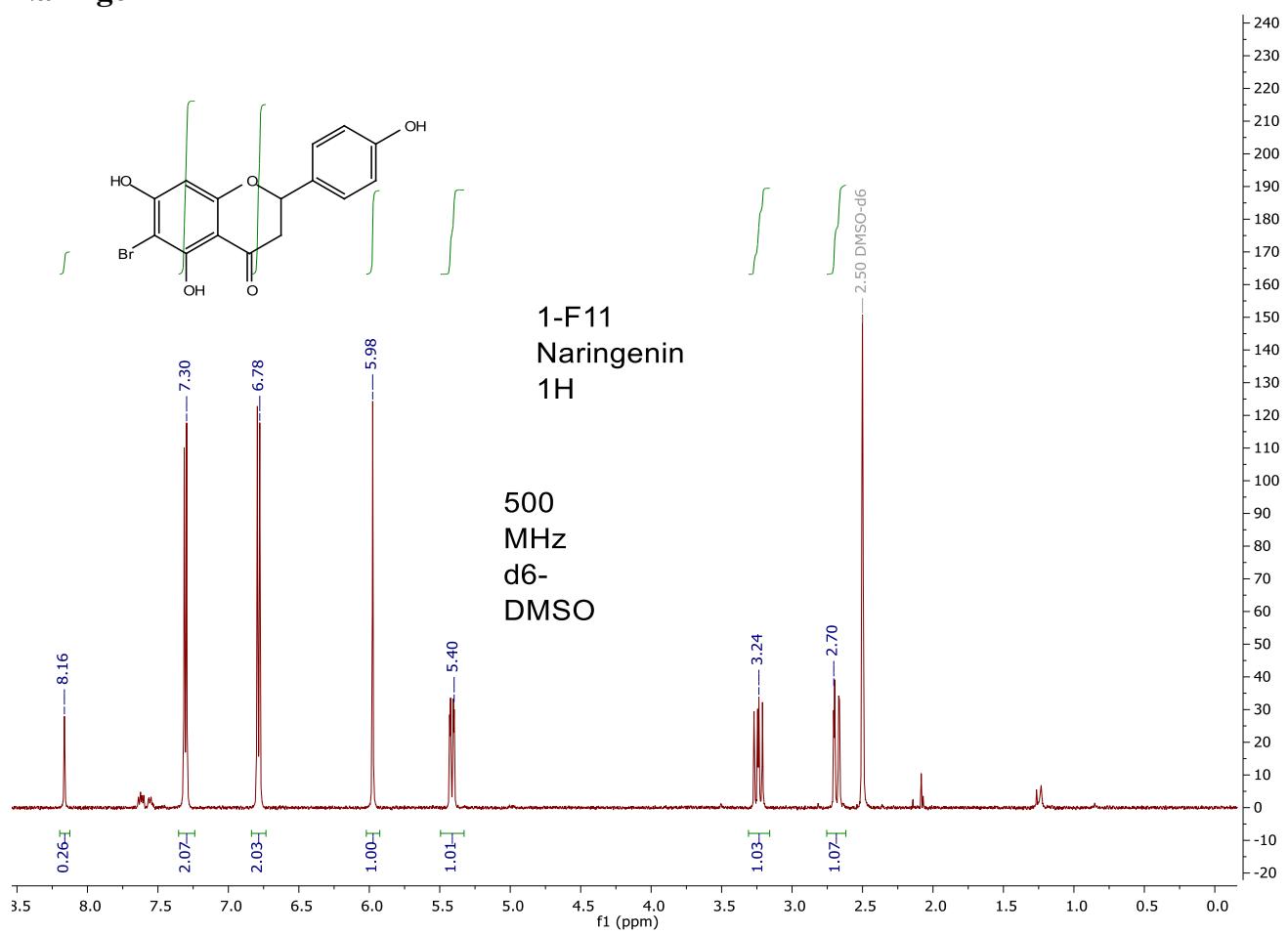
Pindolol – 2-C01



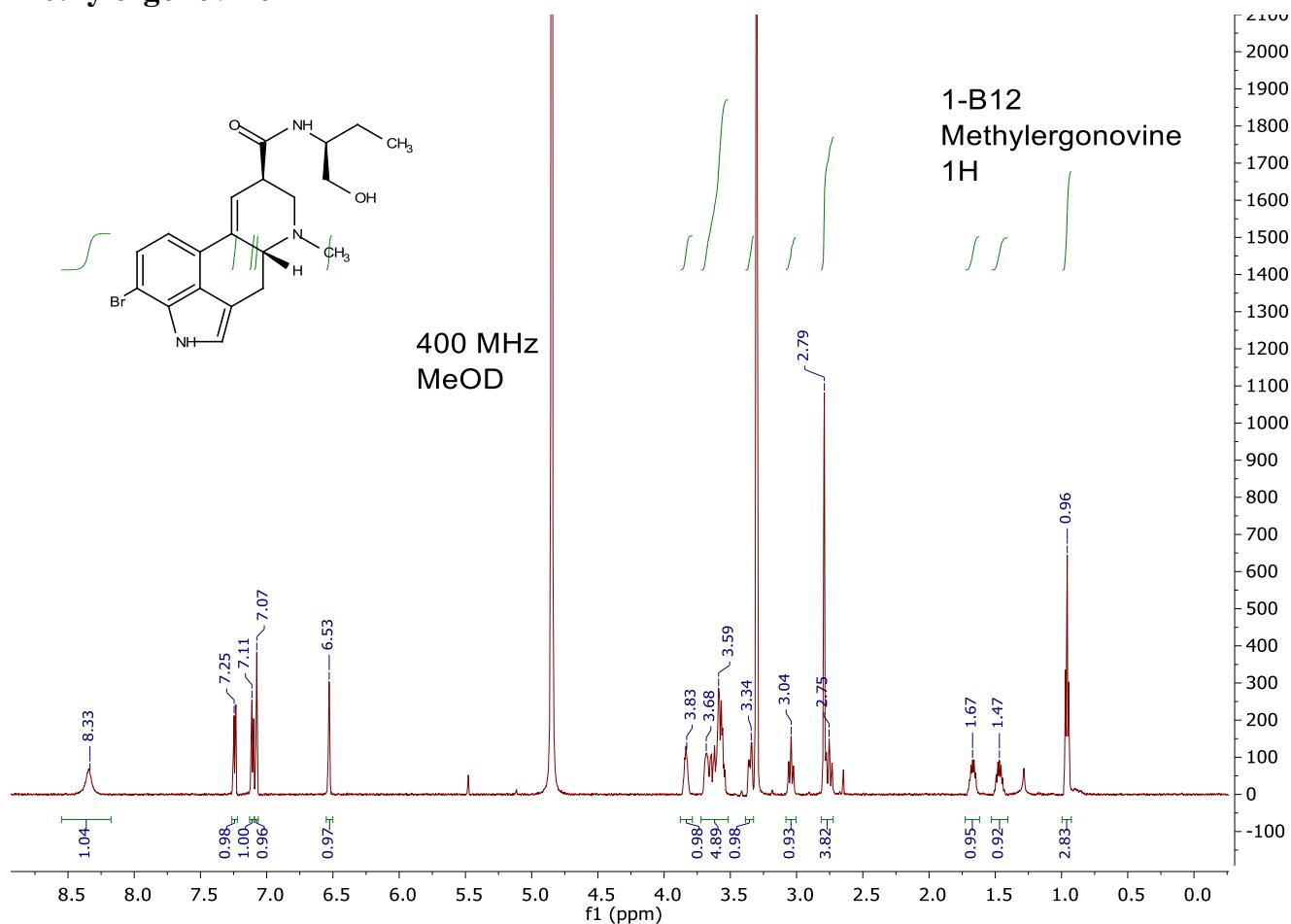
Naringenin – 1-F05



Naringenin – 1-F11

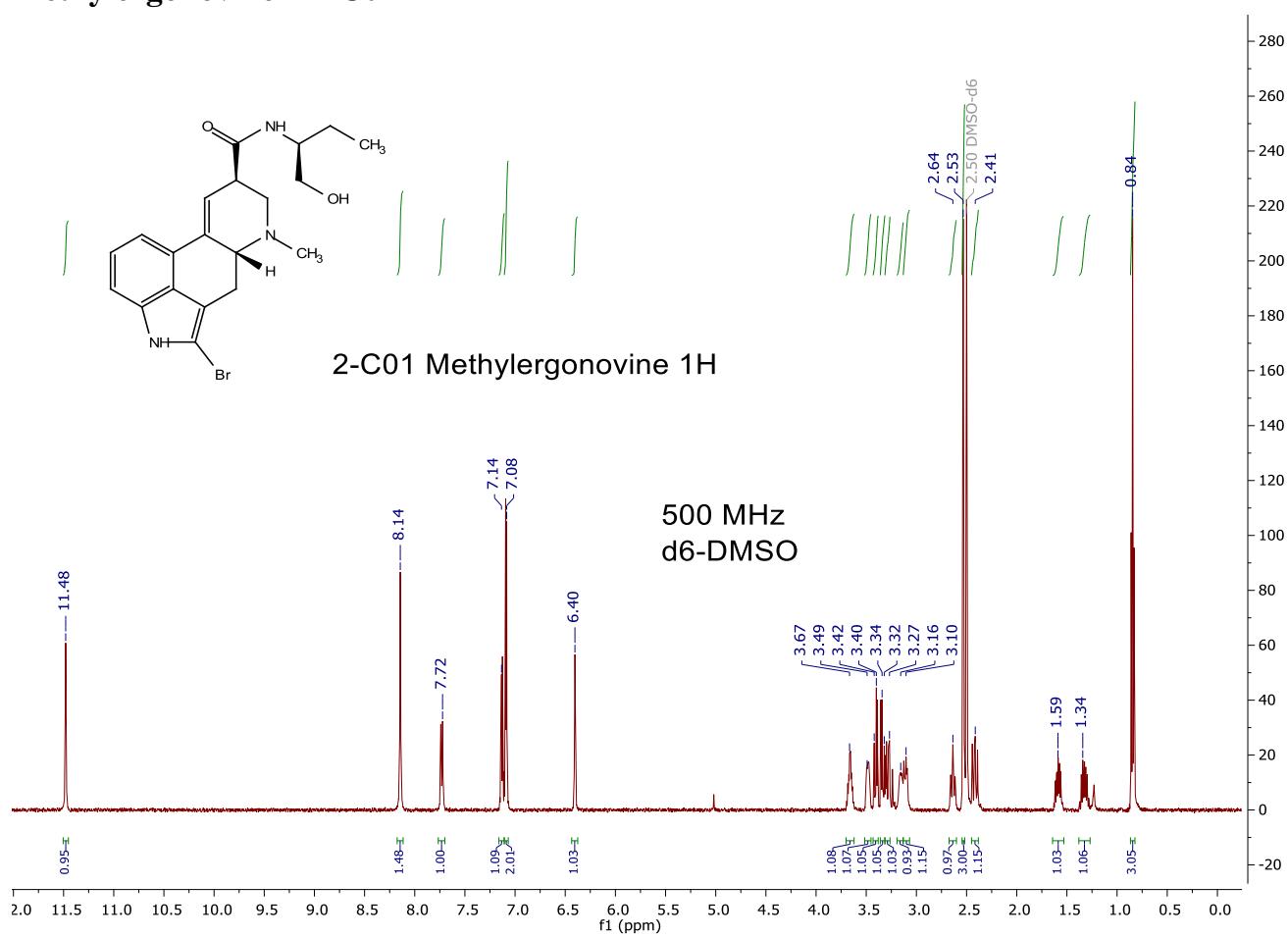


Methylergonovine – 1-B12

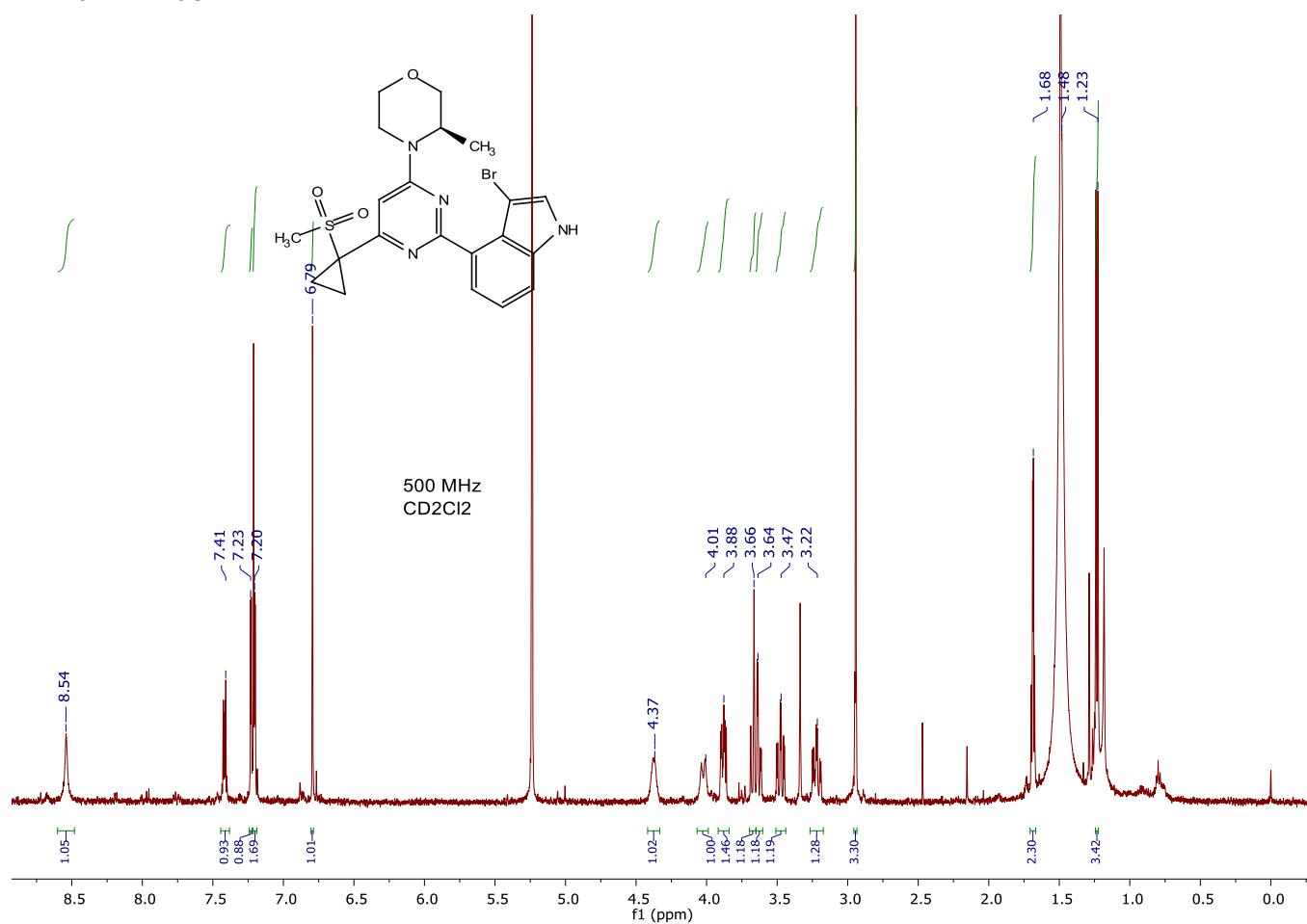


Note: the steric environment of the secondary amide appears to slow down hydrogen-deuterium exchange sufficiently that the N–H resonance appears in the ^1H -NMR spectrum in CD_3OD .

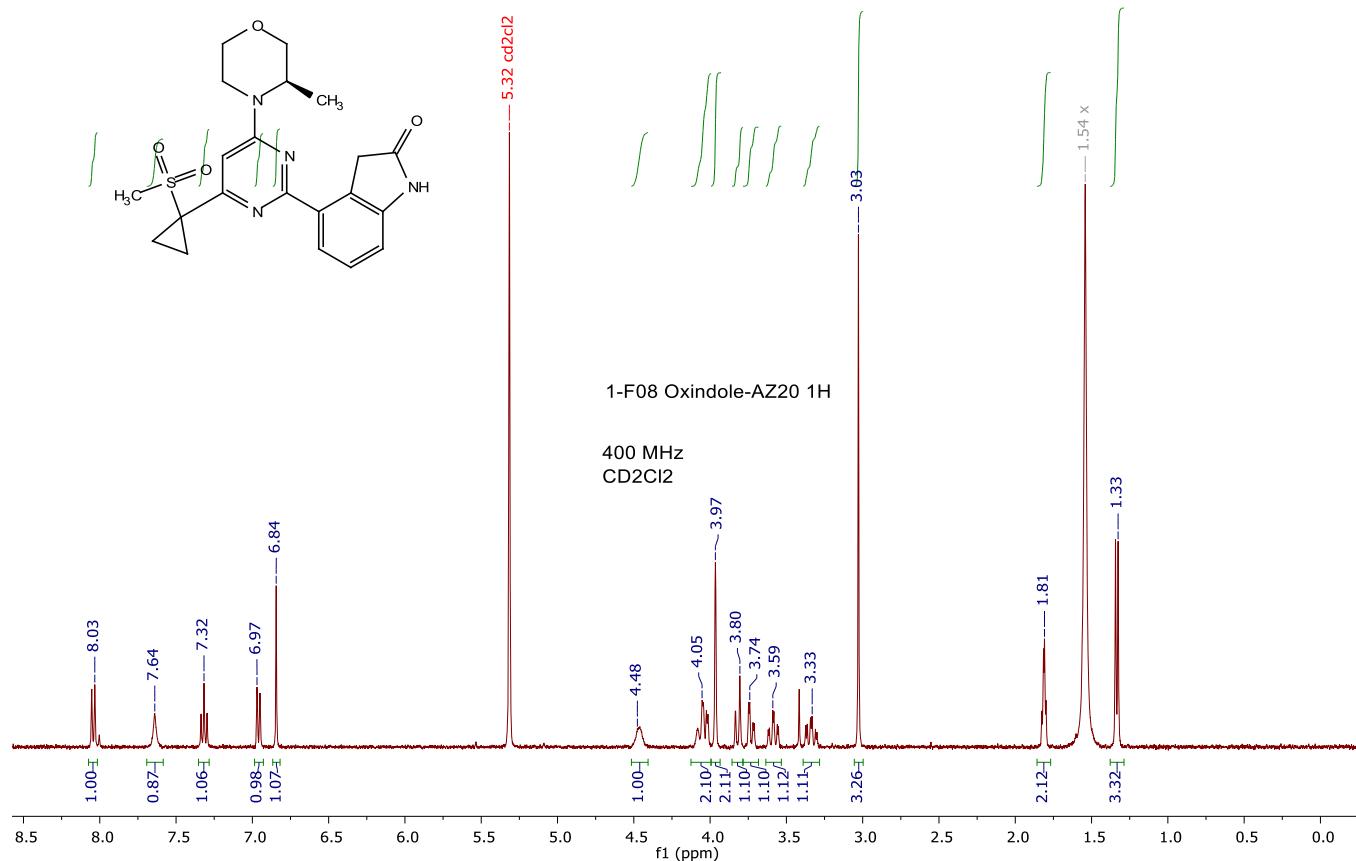
Methylergonovine – 2-C01



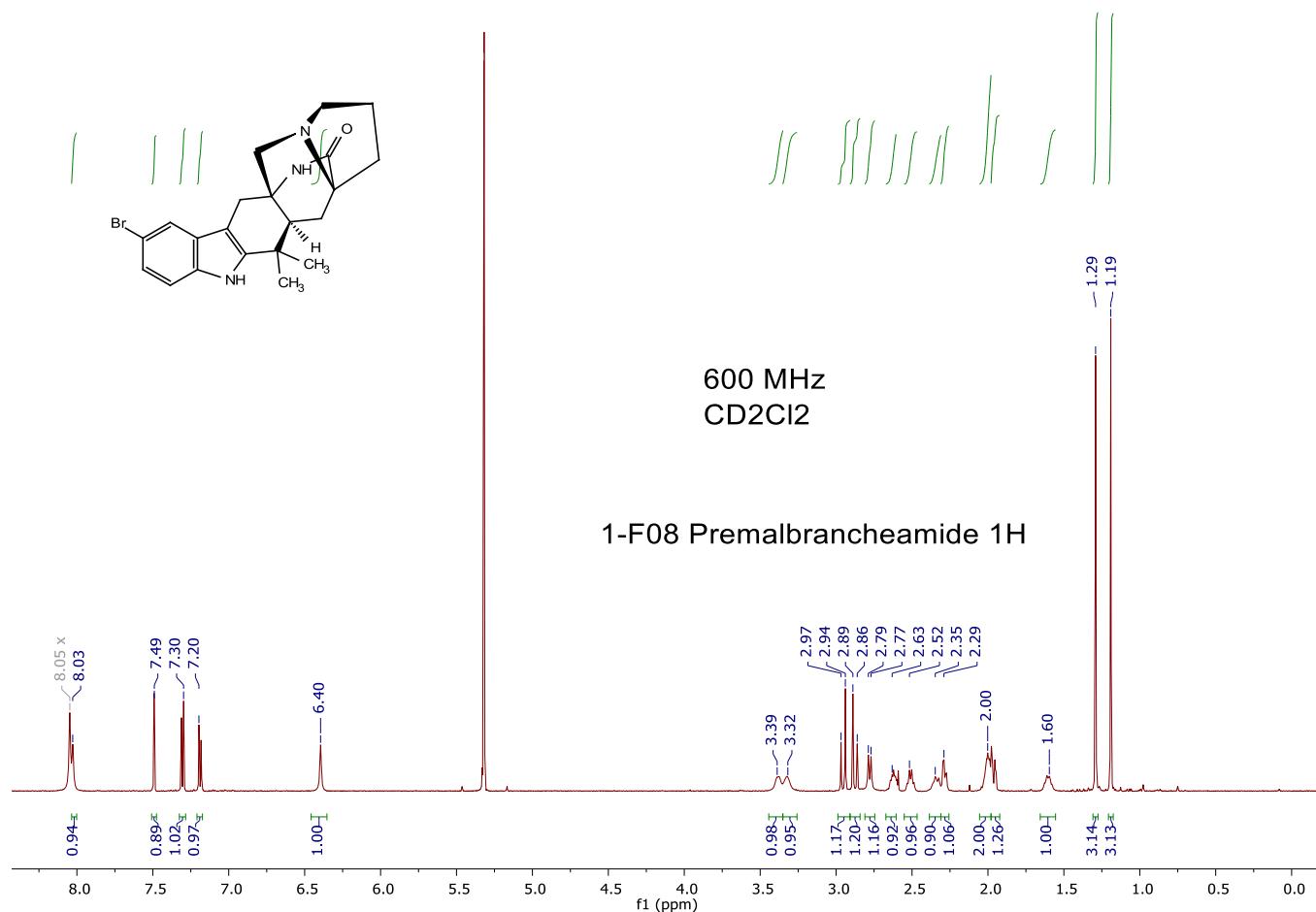
AZ20 – 1-F08



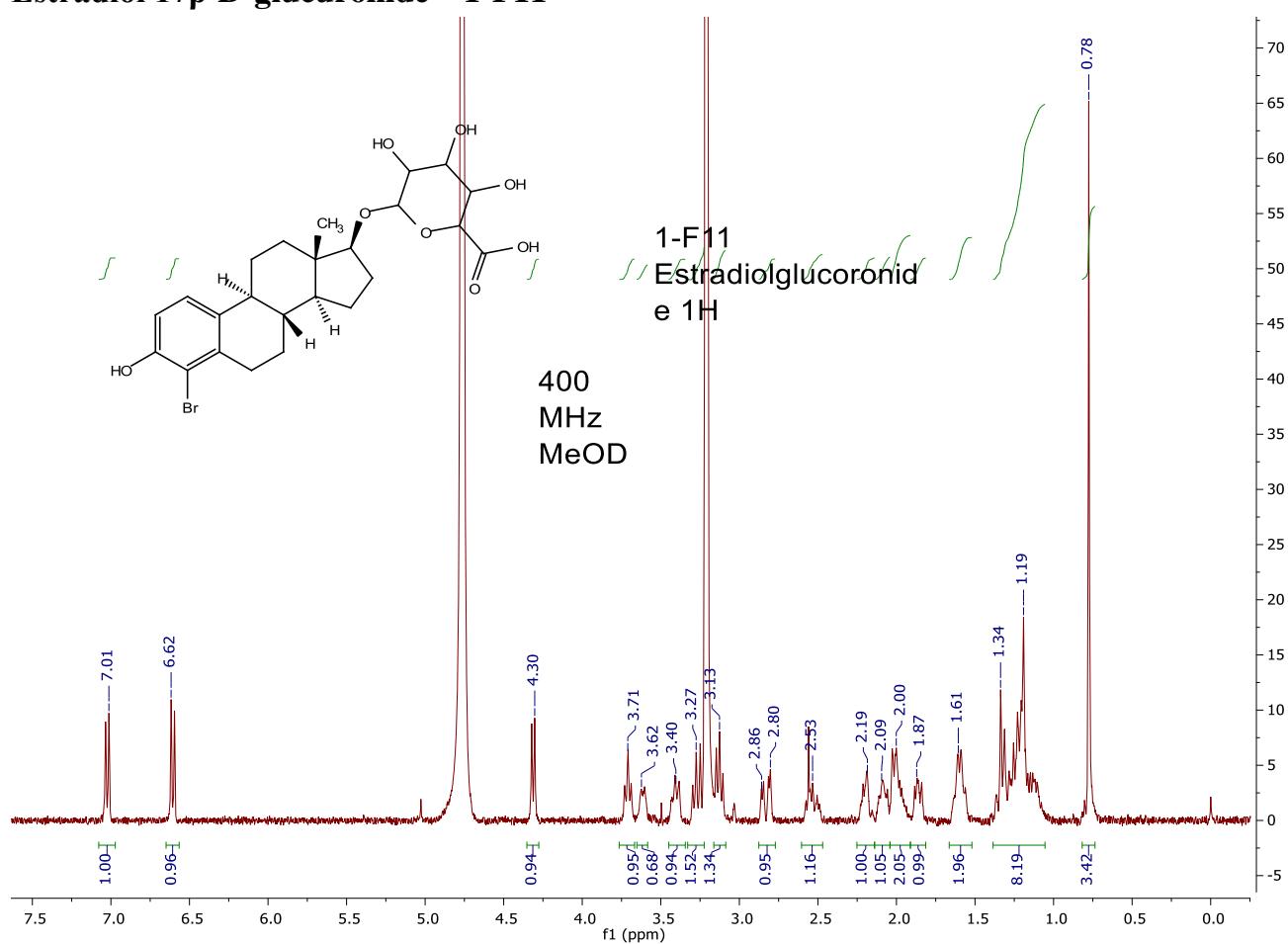
AZ20 – 1-F08 (oxindole product from C2 bromination)



Premalbrancheamide – 1-F08

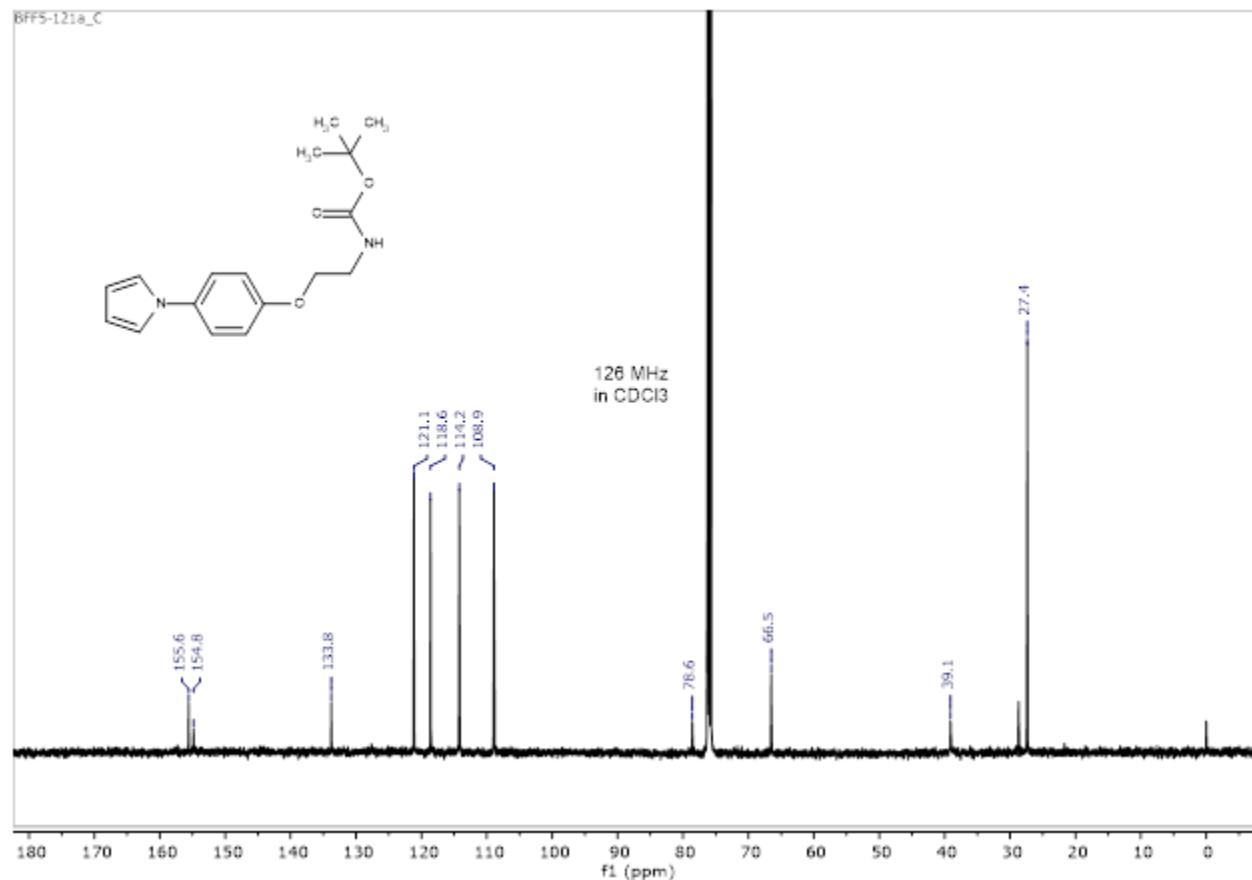


Estradiol 17 β -D-glucuronide – 1-F11

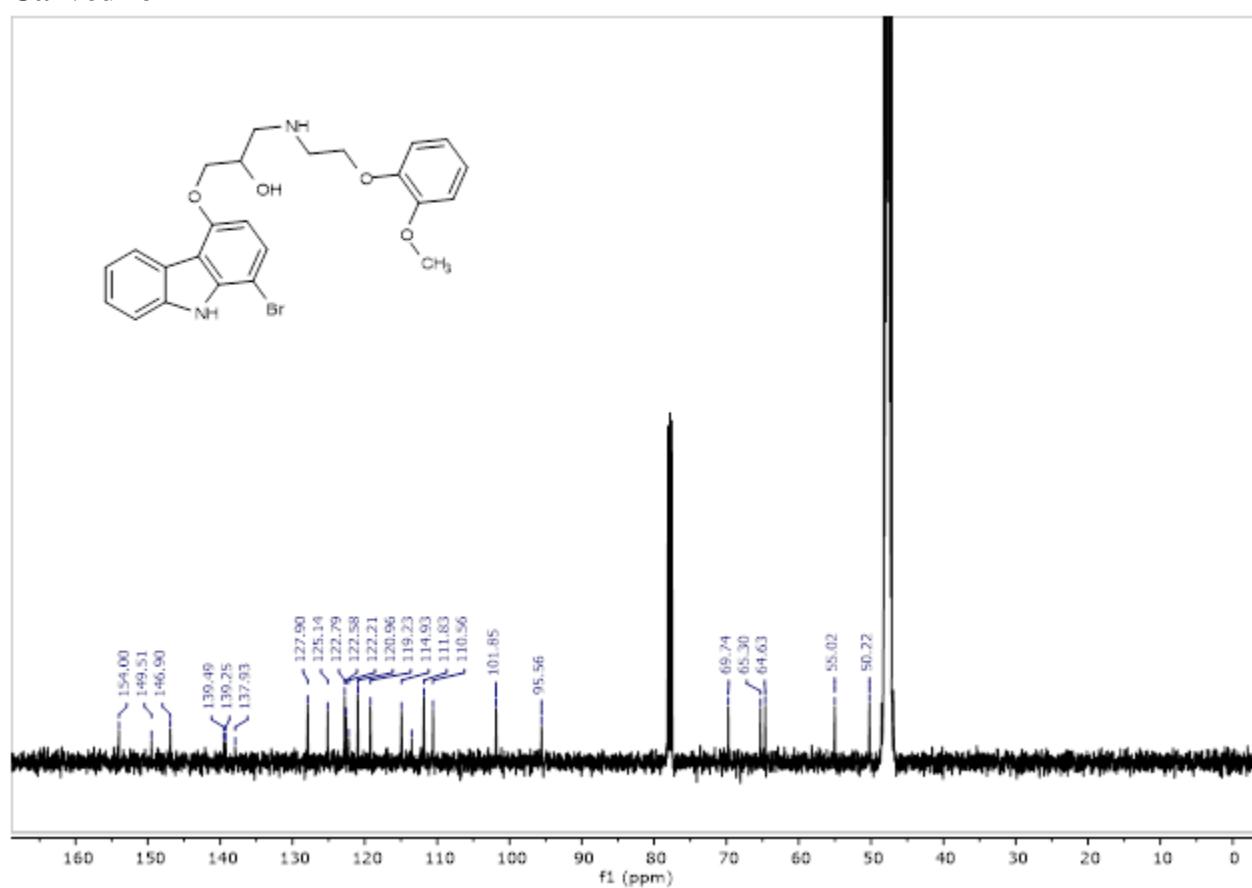


B) ^{13}C -NMR

S2

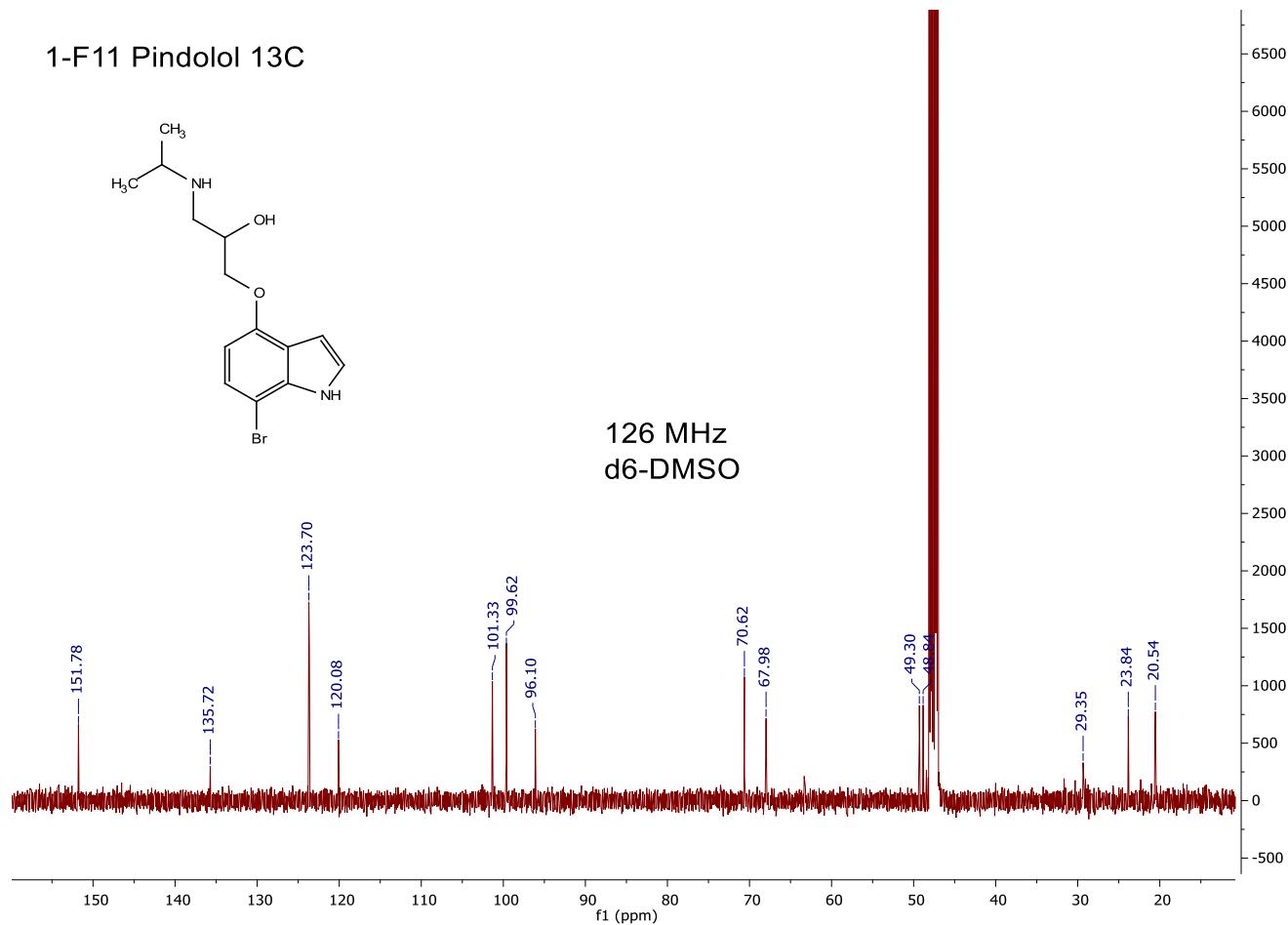


Carvedilol – 1-F11



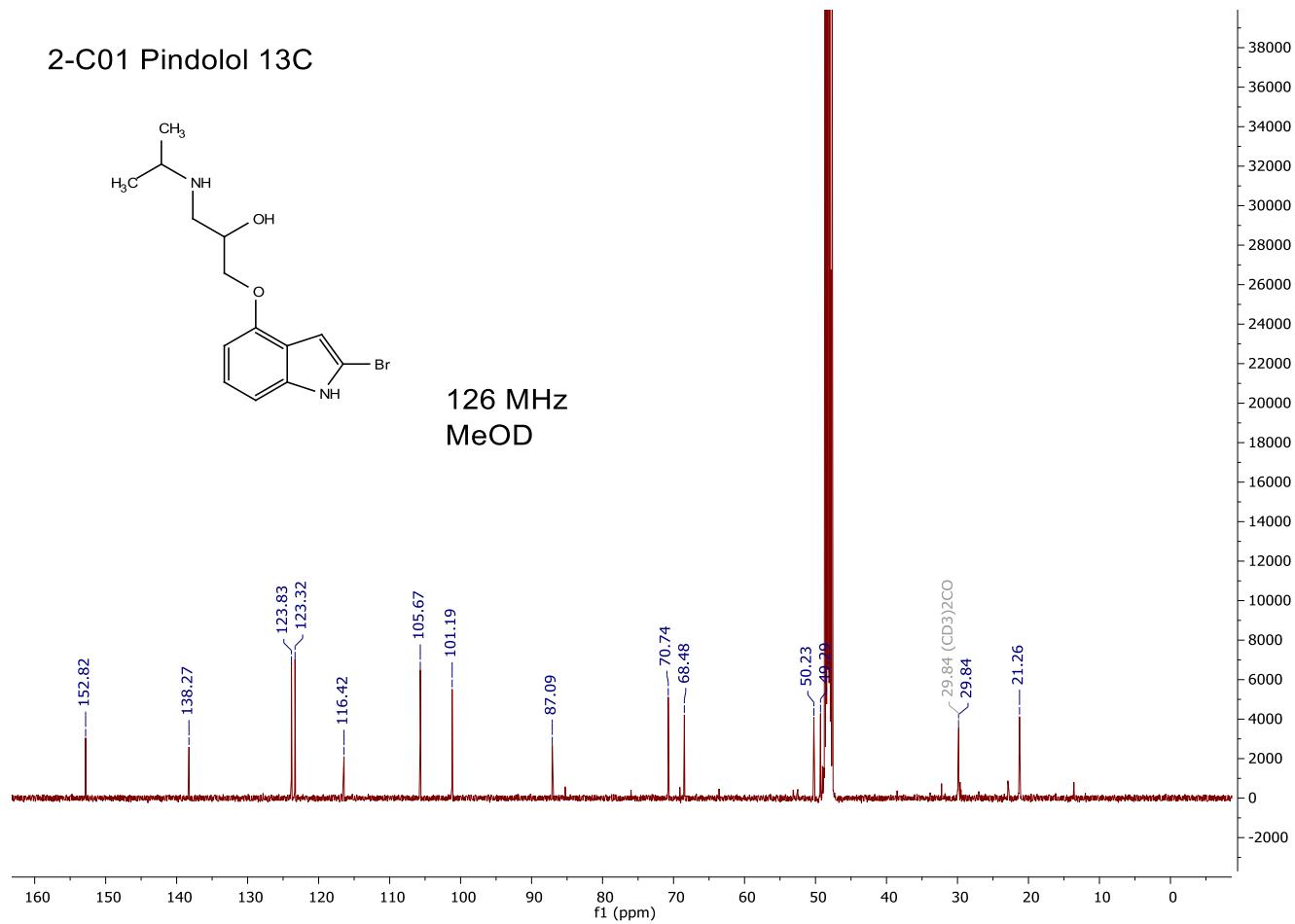
Pindolol – 1-F11

1-F11 Pindolol 13C



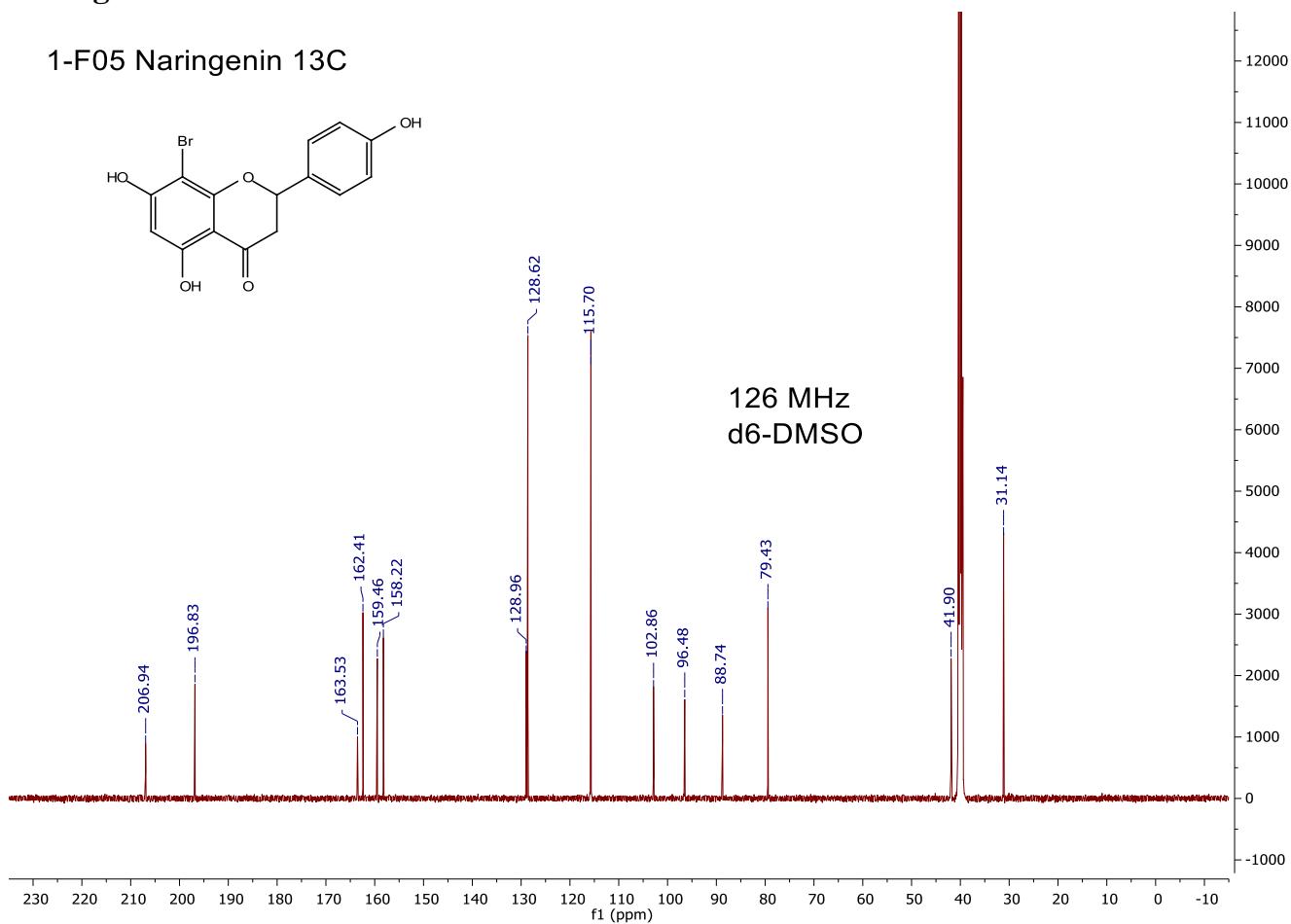
Pindolol – 2-C01

2-C01 Pindolol 13C



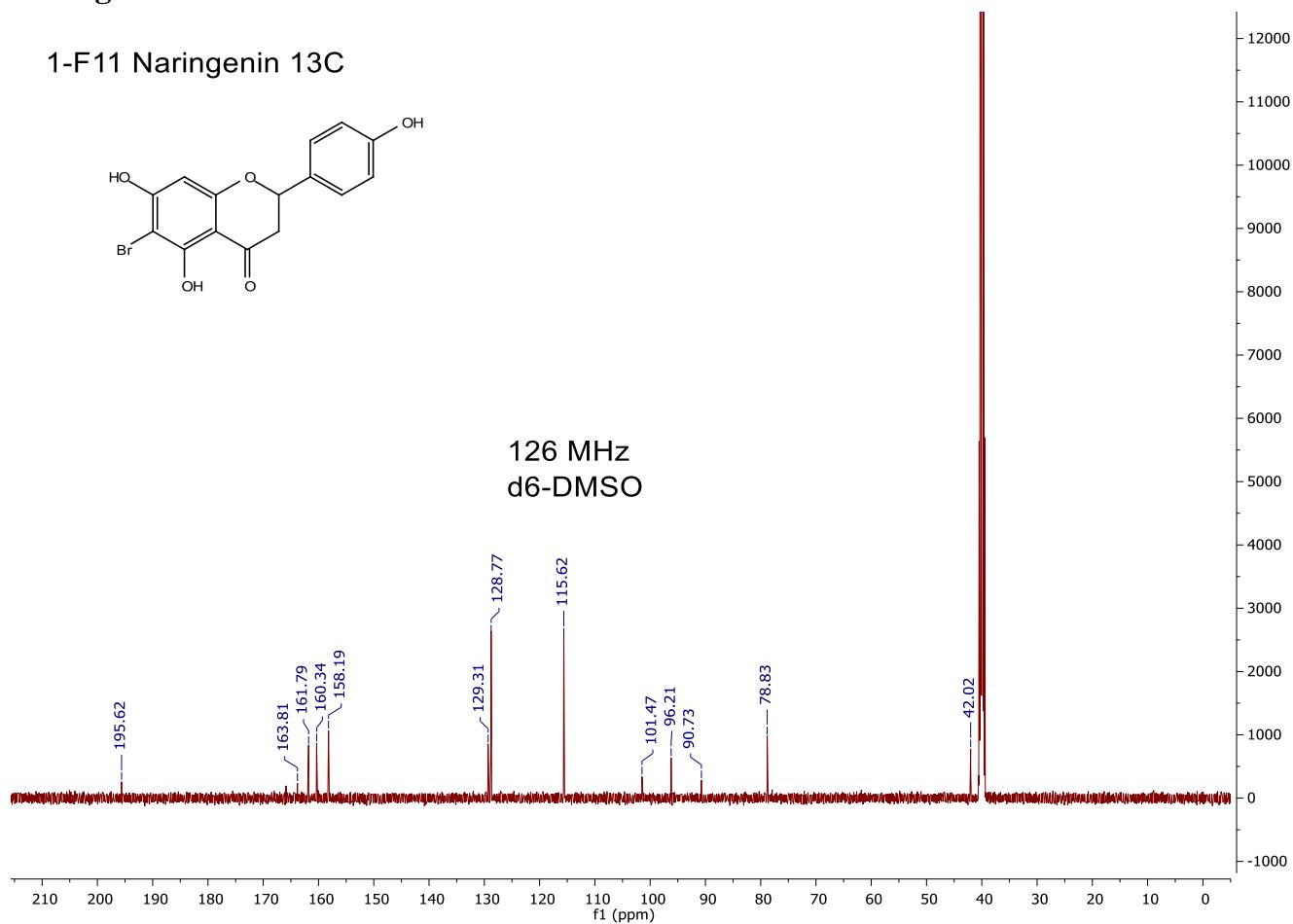
Naringenin – 1-F05

1-F05 Naringenin 13C



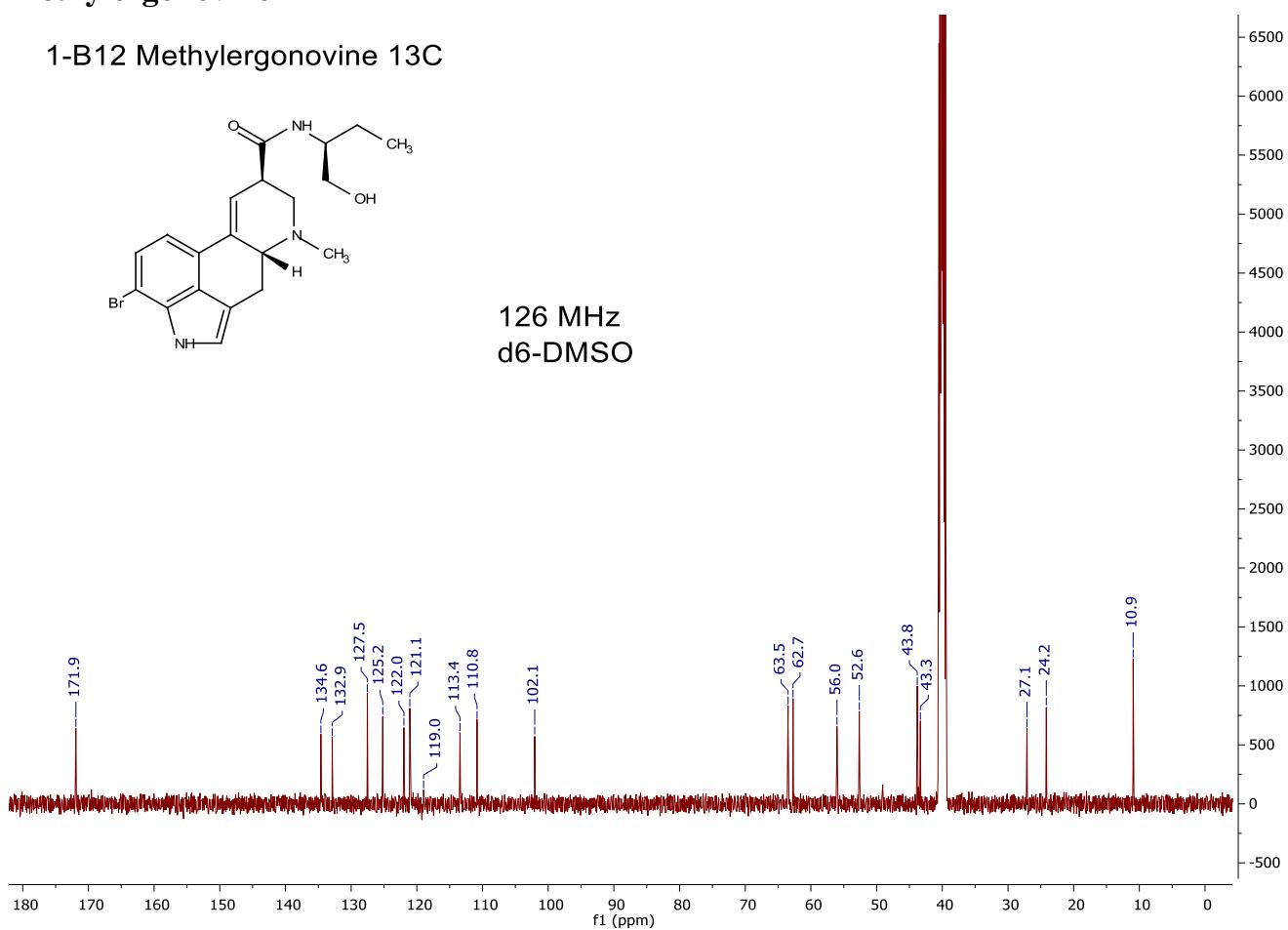
Naringenin – 1-F11

1-F11 Naringenin 13C

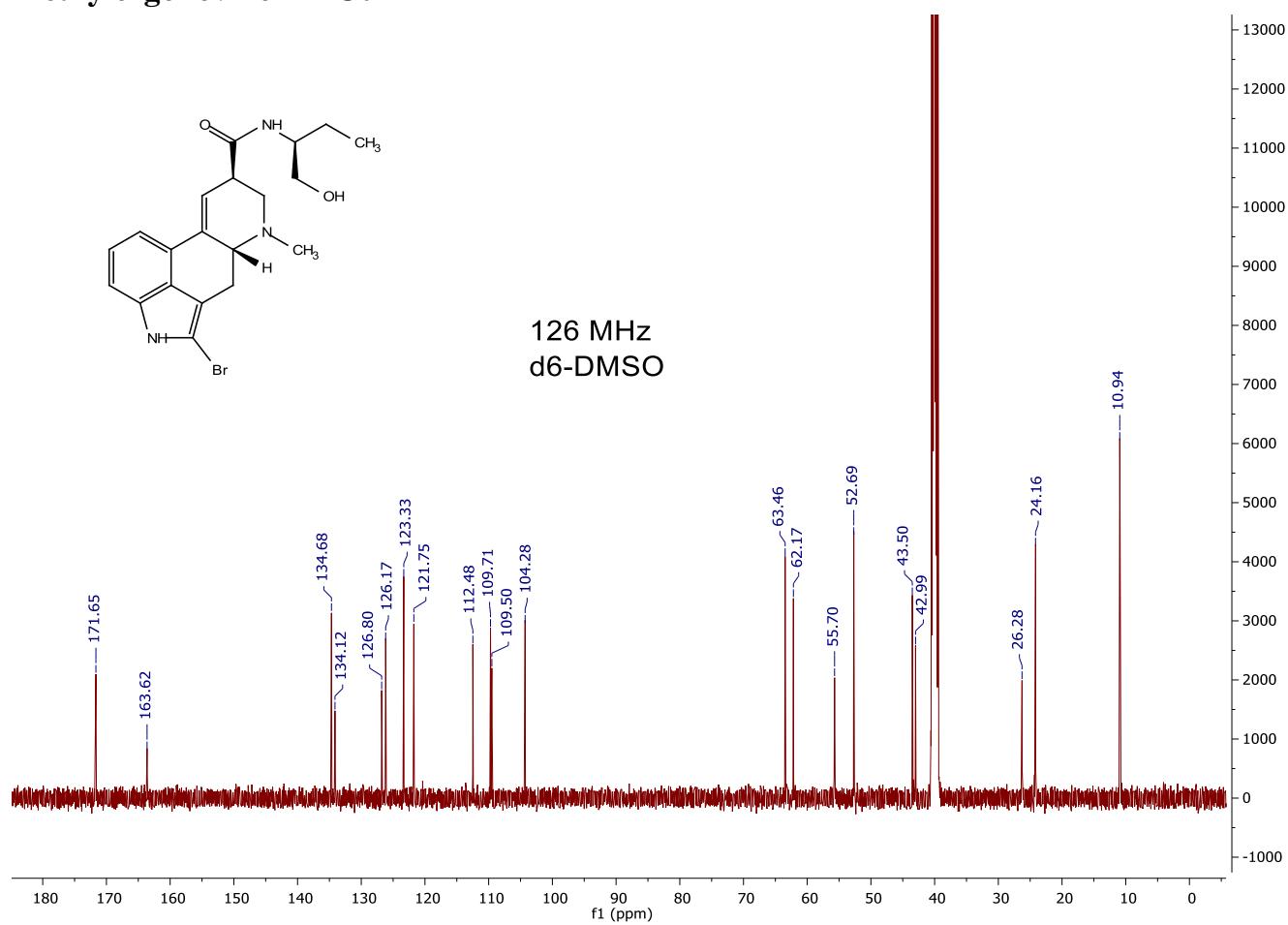


Methylergonovine – 1-B12

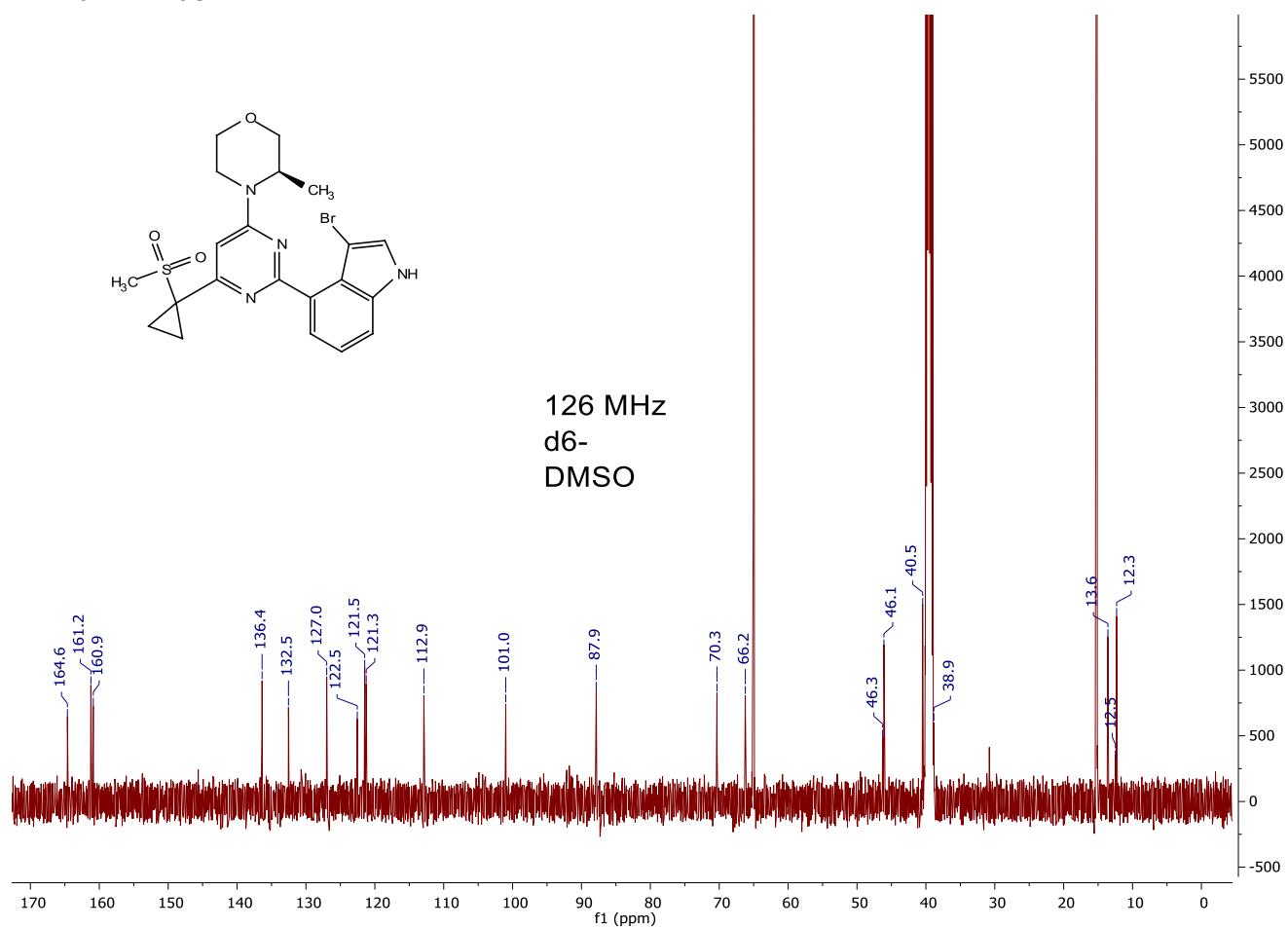
1-B12 Methylergonovine 13C



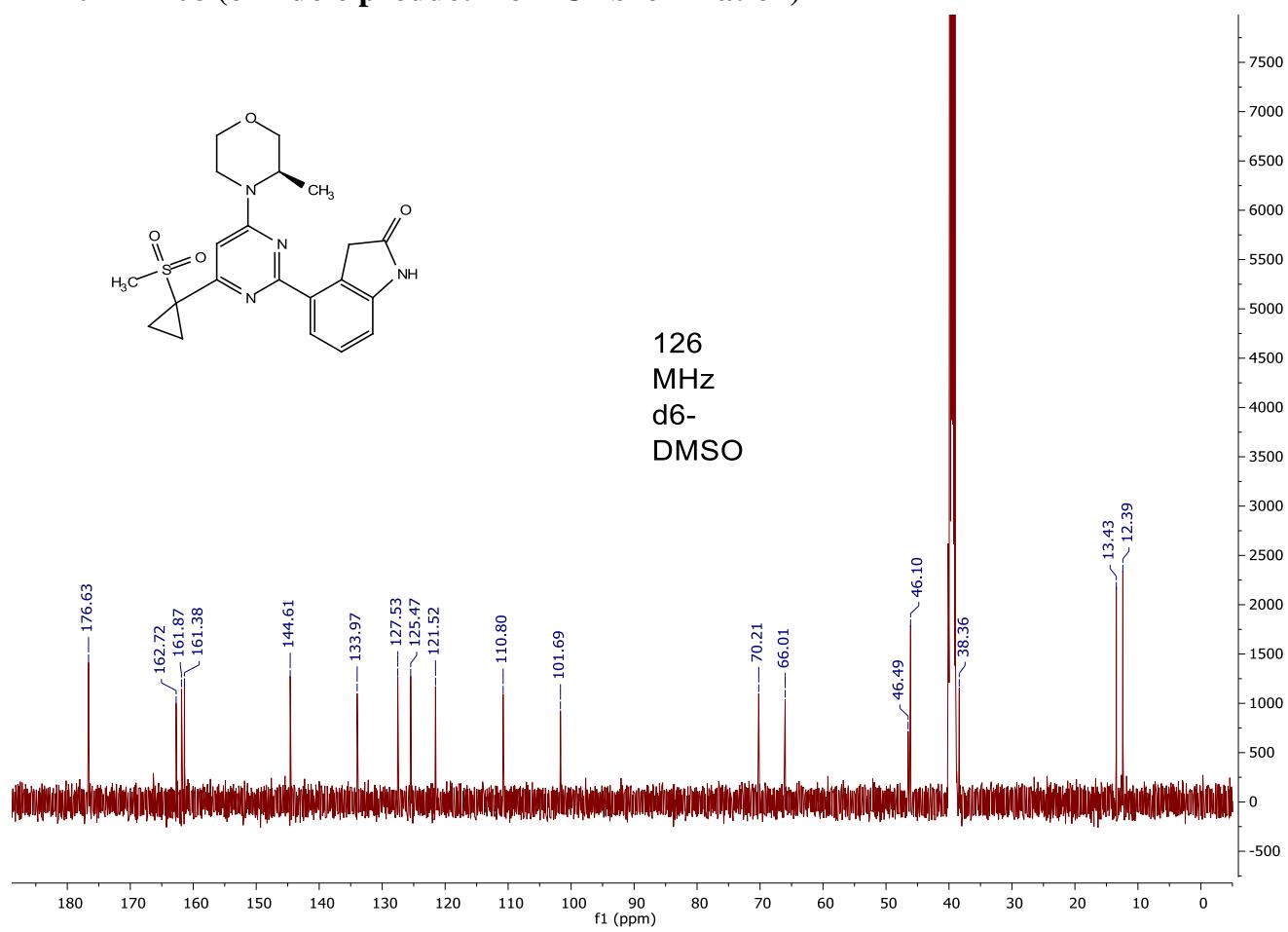
Methylergonovine – 2-C01



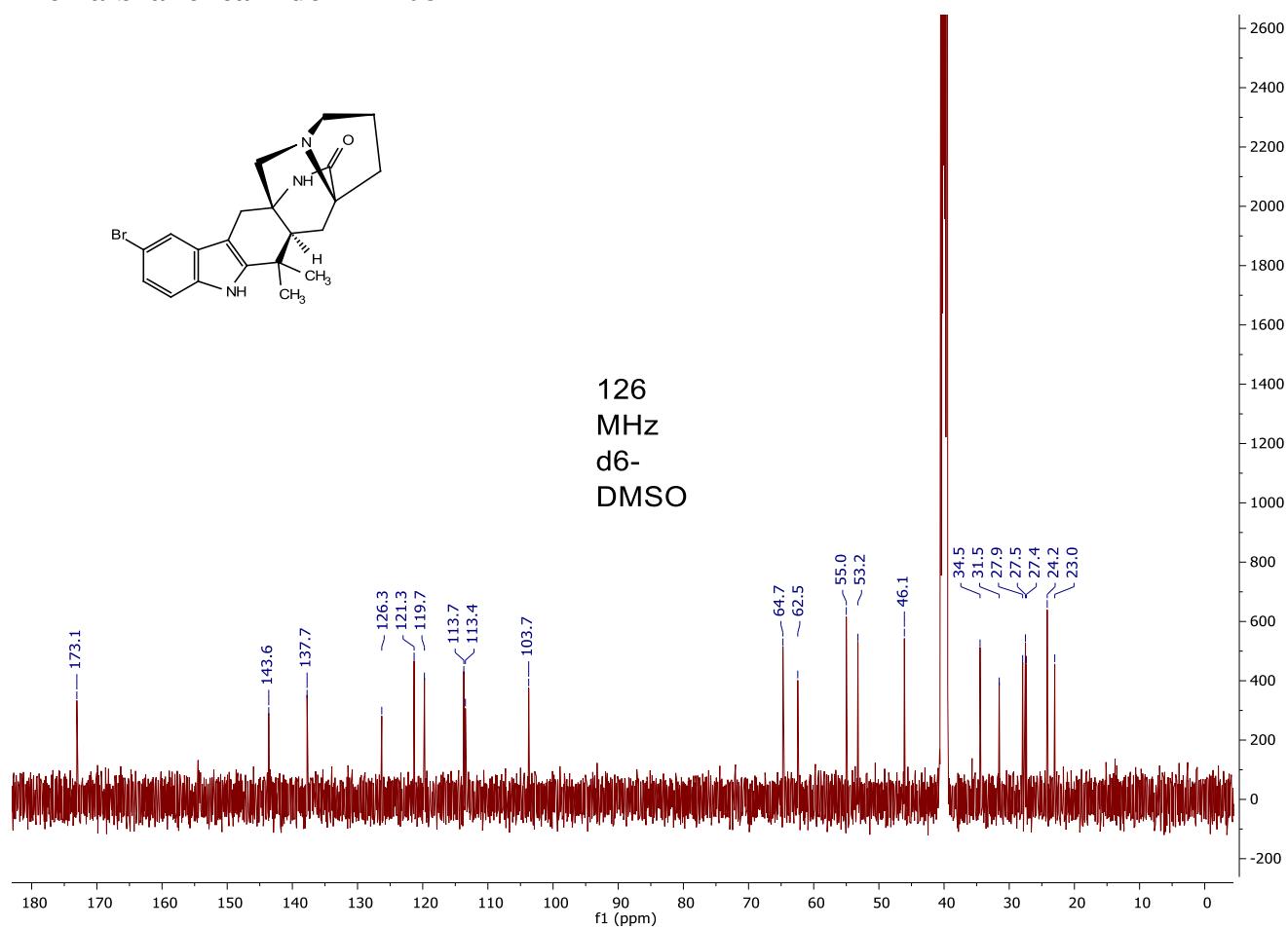
AZ20 – 1-F08



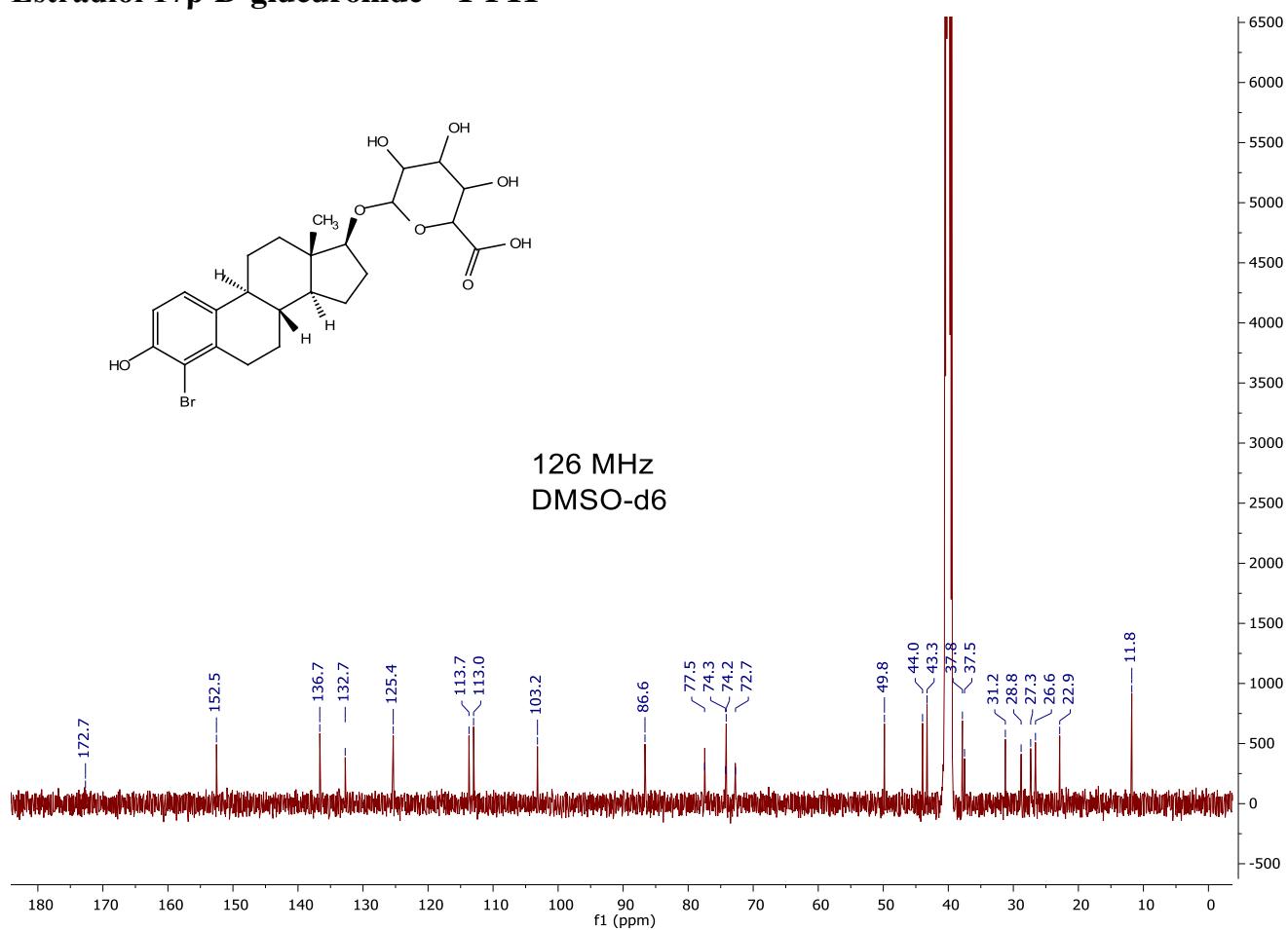
AZ20 – 1-F08 (oxindole product from C2 bromination)



Premalbrancheamide – 1-F08

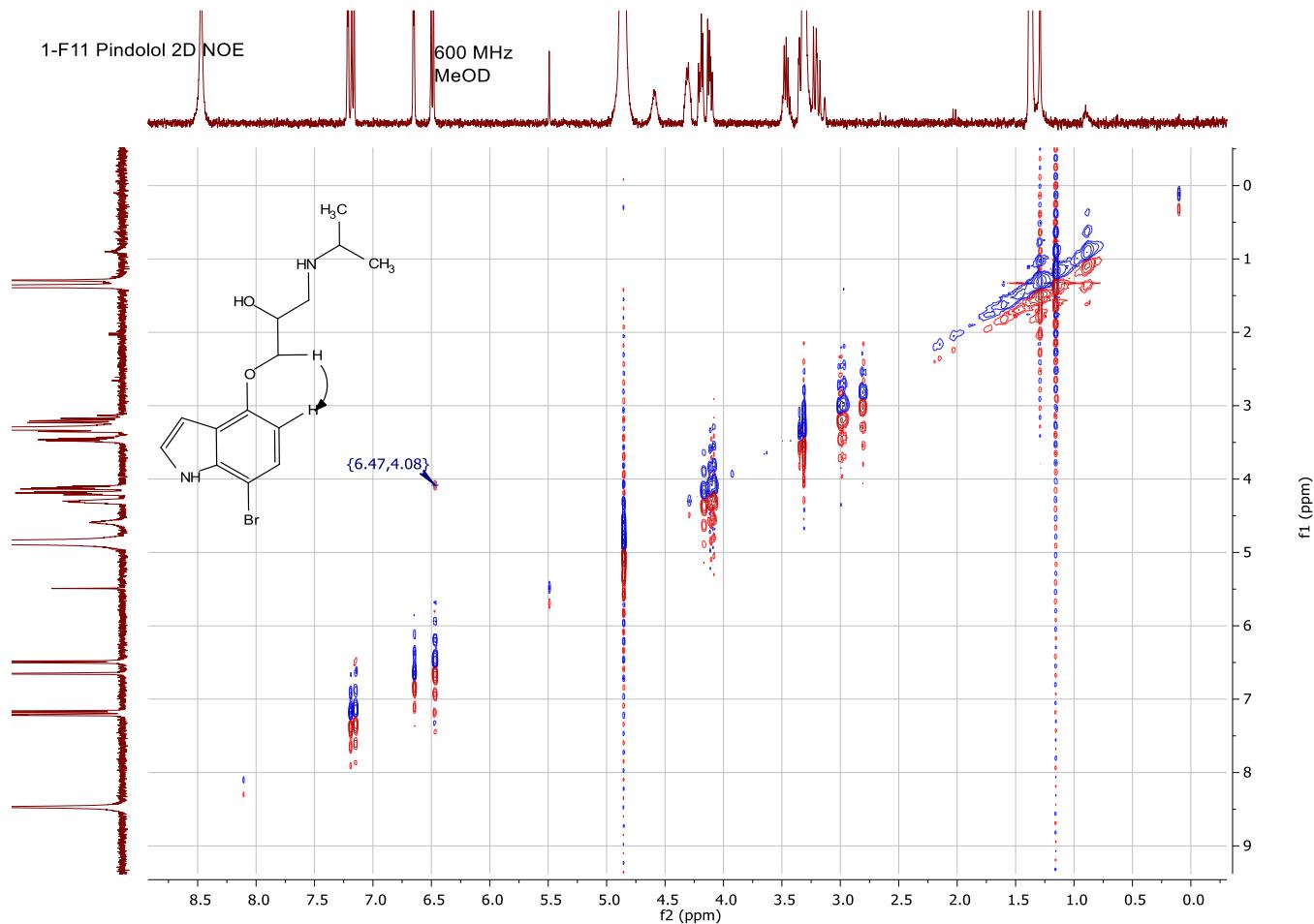


Estradiol 17 β -D-glucuronide – 1-F11



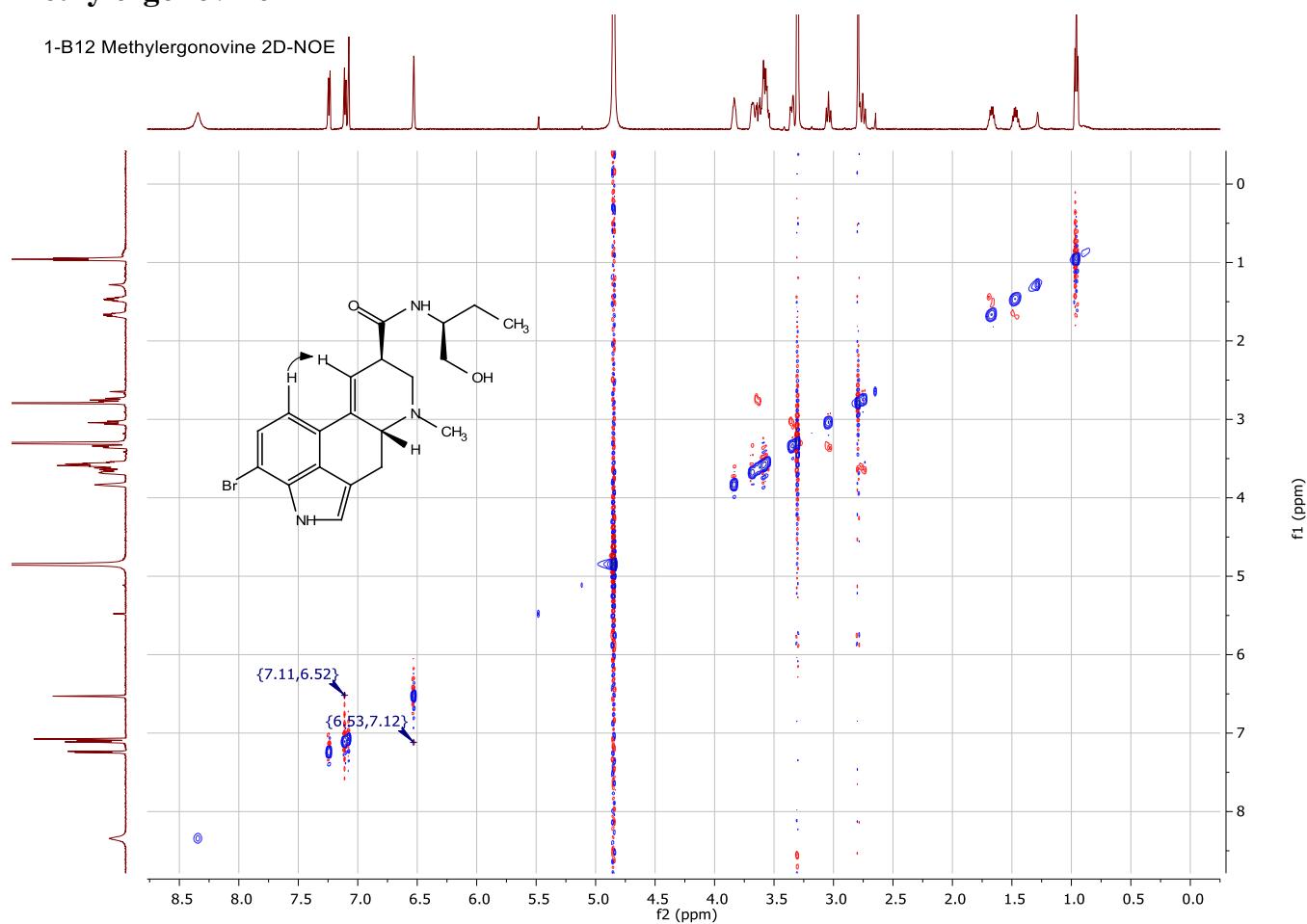
C) 2D-NMR Spectra

Pindolol – 1-F11



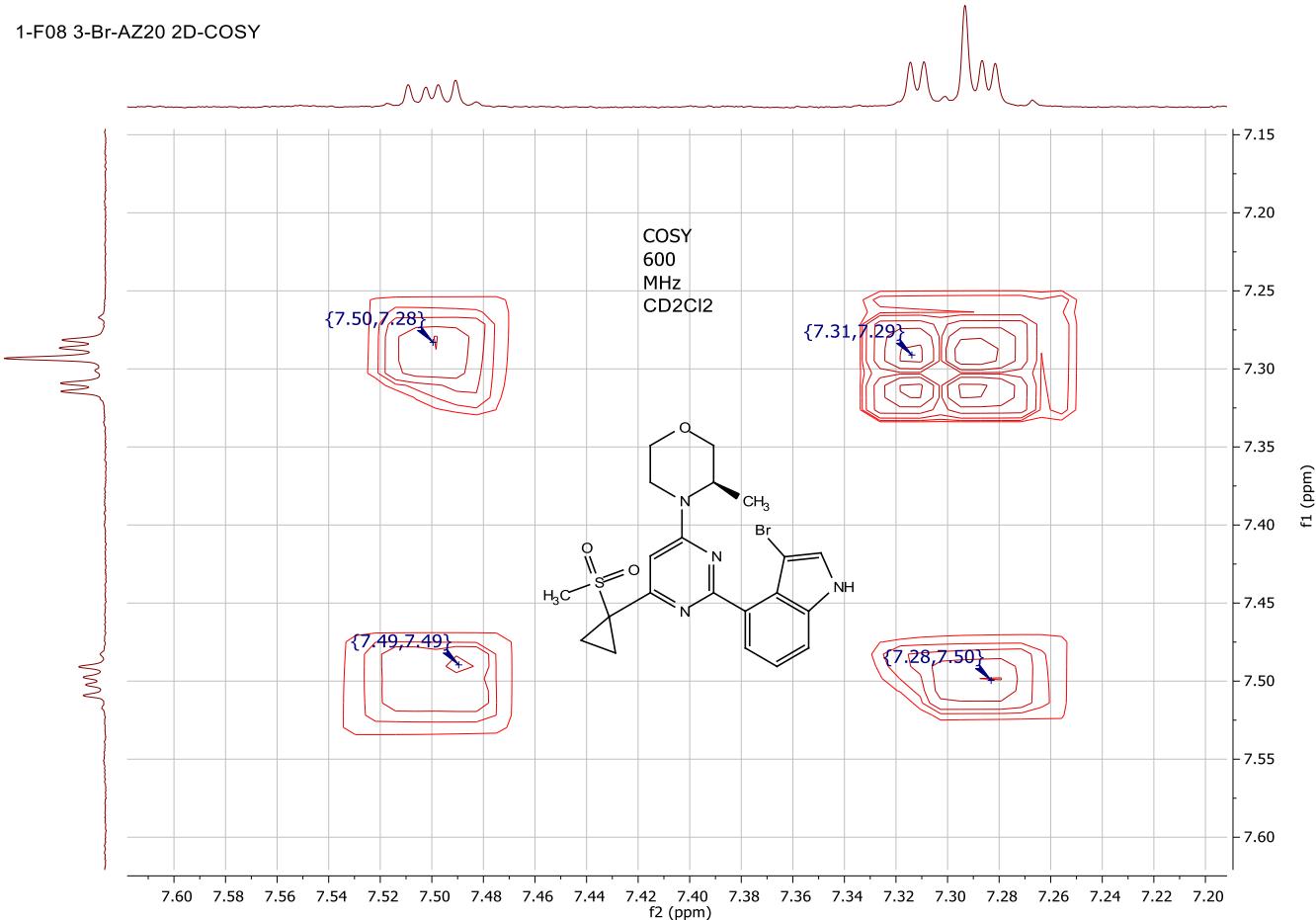
NOE shown above, in combination with aryl multiplet patterns, was used to assign regiochemistry of the bromination reaction.

Methylergonovine – 1-B12



The NOESY crosspeak shown above, in combination with multiplet patterns in the ^1H -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.

XII. SI References

1. Payne, J. T.; Andorfer, M. C.; Lewis, J. C., Engineering Flavin-Dependent Halogenases. *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., Bioconductor: Open Software Development for Computational Biology and Bioinformatics. *Genome Biol.* **2004**, 5, R80.
3. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkin, B.; Slater, D. R.; Whalen, K. L., Enzyme Function Initiative-Enzyme Similarity Tool (Efi-Est): A Web Tool for Generating Protein Sequence Similarity Networks. *Biochim. Biophys. Acta, Proteins Proteomics* **2015**, 1854, 1019.

4. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W., Cd-Hit: Accelerated for Clustering the Next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150.
5. Sasaki, D.; Fujihashi, M.; Iwata, Y.; Murakami, M.; Yoshimura, T.; Hemmi, H.; Miki, K., Structure and Mutation Analysis of Archaeal Geranylgeranyl Reductase. *J. Mol. Biol.* **2011**, *409*, 543.
6. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K., Mega X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547.
7. Gill, S. C.; von Hippel, P. H., Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal. Biochem.* **1989**, *182*, 319.
8. SDS-PAGE Gel. *CSH Protoc.* **2015**, *2015*, pdb.rec087908.
9. Ladner, C. L.; Yang, J.; Turner, R. J.; Edwards, R. A., Visible Fluorescent Detection of Proteins in Polyacrylamide Gels without Staining. *Anal. Biochem.* **2004**, *326*, 13.
10. Coomassie Brilliant Blue Staining Solution. *CSH Protoc.* **2007**, *2007*, pdb.rec10727.
11. Destaining Solution for Coomassie Brilliant Blue R250. *CSH Protoc.* **2007**, *2007*, pdb.rec10717.
12. Wenig, P.; Odermatt, J., OpenChrom: A Cross-Platform Open Source Software for the Mass Spectrometric Analysis of Chromatographic Data. *BMC Bioinformatics* **2010**, *11*, 405.
13. Galili, T.; O'Callaghan, A.; Sidi, J.; Sievert, C., Heatmaply: An R Package for Creating Interactive Cluster Heatmaps for Online Publishing. *Bioinformatics* **2018**, *34*, 1600.
14. Williams, R. M.; Glinka, T.; Kwast, E., Synthetic Studies on Paraherquamide: Regioselectivity of Indole Oxidation. *Tetrahedron Lett.* **1989**, *30*, 5575.
15. Trost, B. M.; Brennan, M. K., Asymmetric Syntheses of Oxindole and Indole Spirocyclic Alkaloid Natural Products. *Synthesis* **2009**, *2009*, 3003.
16. Payne, J. T.; Poor, C. B.; Lewis, J. C., Directed Evolution of Rebh for Site-Selective Halogenation of Large Biologically Active Molecules. *Angew. Chem. Int. Ed.* **2015**, *54*, 4226.
17. Yaipakdee, P.; Robertson, L. W., Enzymatic Halogenation of Flavanones and Flavones. *Phytochemistry* **2001**, *57*, 341.
18. Ashtekar, K.; Marzijarani, N.; Jaganathan, A.; Holmes, D.; Jackson, J. E.; Borhan, B., A New Tool to Guide Halofunctionalization Reactions: The Halenium Affinity (Hala) Scale. *J. Am. Chem. Soc.* **2014**, *136*, 13355.
19. Welch, C. J.; Gong, X.; Schafer, W.; Pratt, E. C.; Brkovic, T.; Pirzada, Z.; Cuff, J. F.; Kosjek, B., Miser Chromatography (Multiple Injections in a Single Experimental Run): The Chromatogram Is the Graph. *Tetrahedron: Asymmetry* **2010**, *21*, 1674.

20. Bellomo, A.; Celebi - Olcum, N.; Bu, X.; Rivera, N.; Ruck, R. T.; Welch, C. J.; Houk, K. N.; Dreher, S. D., Rapid Catalyst Identification for the Synthesis of the Pyrimidinone Core of Hiv Integrase Inhibitors. *Angew. Chem. Int. Ed.* **2012**, *51*, 6912.
21. Santanilla, A. B.; Regaldo, 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., Nanomole-Scale High-Throughput Chemistry for the Synthesis of Complex Molecules. *Science* **2015**, *347*, 49.
22. Smith, D. R. M.; Uria, A. R.; Helfrich, E. J. N.; Milbredt, D.; van Pee, K.-H.; Piel, J. r.; Goss, R. J. M., An Unusual Flavin-Dependent Halogenase from the Metagenome of the Marine Sponge Theonella Swinhoei Wa. *ACS Chem. Biol.* **2017**, *12*, 1281.
23. 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., Function and Structure of Mala/Mala', Iterative Halogenases for Late-Stage C-H Functionalization of Indole Alkaloids. *J. Am. Chem. Soc.* **2017**, *139*, 12060.
24. Niklasson, M.; Andresen, C.; Helander, S.; Roth, M.; Kahlin, A.; Appell, M.; Mårtensson, L. G.; Lundström, P., Robust and Convenient Analysis of Protein Thermal and Chemical Stability. *Protein Sci.* **2015**, *24*, 2055.