THE UNIVERSITY OF CHICAGO

METHODS AND APPLICATIONS FOR THE IDENTIFICATION AND DIRECTED EVOLUTION OF BIOCATALYSTS FOR CHEMICAL SYNTHESIS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

HARRISON MCGRATH SNODGRASS

CHICAGO, ILLINOIS

JUNE 2023

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

APPENDIX I

APPENDIX II

APPENDIX III

APPENDIX IV

APPENDIX V

LIST OF SCHEMES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

ACKNOWLEDGEMENTS

Like all students, I owe my success to my mentors who came before me and helped me grow as a scientist, and even a little as a person. Prof. Mary Andorfer introduced me to flavin-dependent halogenases and helped me to understand the potential of this amazing enzyme class, and also inspired me with her compassion and drive. Dr. Krysten Jones mentored me in molecular biology and without her I would have likely been forced to use restriction cloning. Considering this alternative, my debt can never be paid. Dr. Brian Fisher would spend much of his time in the lab teaching me experimental design, HPLC method development, and some basic chemistry; then leave and spend the rest of his evening keeping my character alive in our D&D campaign. Dr. Rui Huang helped me to perfect my primer design and was a good source of pragmatic wisdom. Dr. Dibyendu Mondal assisted with synthesis and also cooked some amazing foods for special occasions.

Lastly, I want to thank my friends and family. Dan, Mike, Parth, and Anthony made the first months of the Covid-19 pandemic bearable by offering up way too much of their time to hang out in Verdansk and helped make me feel closer to my loved ones during my last year in Bloomington. I also need to thank Tyler, John, and Andrew, who have known me for fifteen years and, despite this, remain my friends. My sister Antoinette, her husband Sean, and their family help inspire me by demonstrating hard work and perseverance. My brother Mitchel encouraged me to keep an open mind and to explore new possibilities which has helped shape my relationship with science. My parents Michael (Duke) and Joyce (Mom) are some of the funniest people I know and have loved and supported me even if they thought I didn't like them that much and wasn't going anywhere. Mousumy, Yogendra, and Aditya have accepted me into their family like a son and brother and I am blessed to have them as my in-laws. Atreyi Bhattacharya has been by my side since I started graduate school, and I am so fortunate to be able to share in this success with her.

ABSTRACT

This dissertation describes both new applications and methodology for the engineering and identification of biocatalysts. Significant portions of this thesis also discuss specifically the characterization, identification, and improvement of flavin-dependent halogenases. This family of enzymes is capable of site-selective halogenation of aryl C-H bonds using benign halide sources and molecular oxygen as the terminal oxidant. Prior to the work in this dissertation, these enzymes had been demonstrated as competent catalysts for regio- and enantio- selective transformations and, importantly, very amenable for improvement by directed evolution. These efforts from our lab had focused largely on the single halogenase RebH as a template for evolution in large part as a result of the dearth of explored enzymes.

Chapter 1 introduces biocatalysis and the role of directed evolution in enzyme development. It also serves as an overview of flavin-dependent halogenases, biosensors in native and synthetic context, continuous evolutions, and selection strategies, as well as the application of genome mining for enzyme identification. This section of the dissertation examples accomplishments of directed evolution while acknowledging the limitations of how it has been applied to biocatalysis development. It also introduces the previous work conducted in our lab for the directed evolution of flavin-dependent halogenases.

Chapter 2 discusses the results of an effort to identify new flavin-dependent halogenases by genome mining. In collaboration with the Joint Genome Institute, over 120 genes encoding putative FDHs identified by primary amino acid sequence were synthesized and cloned into *E. coli* for expression. Solubly expressed enzymes were then screened for function and over 40 new halogenases were identified. The enzymes characterized in this study exhibited novel brominase activity and regiodivergent halogenation on a variety of substrates. Clusters of highly active enzymes identified by this effort were then further mined and used to identify several enzymes with interesting biochemical properties and novel activity.

Chapter 3 summarizes the directed evolution of a RebH variant for atroposelective halogenation of a medicinally relevant 3-Aryl-4(3 H)-quinazolinones via both traditional and dynamic kinetic resolutions. This effort combines the site-selective catalysis enabled by halogenases with highly enantioselective functionalization of a rapidly racemizing substrate to generate valuable atropisomers in high yield and excellent e.r.. The RebH variant used as a template in this study was identified by screening nearly 150 evolved and wild-type enzymes and improved by three rounds of directed evolution. Modeling and docking was used to explain the improvements to site selectivity throughout the evolution. Preparative halogenation enabled the derivatization of enzyme products and proceeded without the loss of enantioselectivity.

Chapter 4 highlights a collaborative effort between our lab and the Dickinson lab to develop a continuous evolution and screen using phage-assisted continuous evolution or PACE. After initial evaluation of a simple biosensor system utilizing IPTG as the ligand for driving gene expression, we eventually identified an engineered abscisic acid system as being more amenable to esterase driven PACE catalyzed by the hydrolysis of ABA esters. A continuous selection (PACS) was found to greatly improve the activity of variants identified. Kinetic characterization and analysis of substrate scope showed that variants identified from this study had improved activity and selectivity.

Chapter 5 describes the utilization of engineered biosensors for the directed evolution of enzymes by increasing screening throughput. This chapter details the use of an abscisic acid biosensor applied to FACS for the screening of an esterase library, resulting in the identification of two variants with greatly improved activity for hydrolysis of the model substrate. Efforts to enable a dual-sort using two orthogonal biosensor systems are also briefly described. On-going efforts to apply an engineered biosensor for halogenated tryptophan derivatives to the screening of FDH libraries is also briefly described.

Chapter 6 details the biocatalytic potential of a newly described halogenase, AetF. This enzyme is a unique bifunctional flavin reductase/flavin-dependent halogenase first described in 2019. Through evaluation of the substrate scope, we found this enzyme to be a competent biocatalyst for halogenation of a variety of substrates, including relatively electron deficient arenes and several enantioselective transformations. This enzyme is also uniquely capable of highly regioselective iodination and enantioselective cycloiodoetherification. This reactivity is rationalized by modeling of the enzyme active site and the model accuracy was supported by mutagenesis studies. A collaboration with Pfizer designed to investigate scalable halogenation with this enzyme is also briefly described.

Chapter 7 discusses an investigation of an active RebH variant with the active site lysine residue mutated to arginine. Though a seemingly small change, this exact mutation had been introduced into a variety of halogenase scaffolds and resulted in inactive variants. Through site-directed mutagenesis, we found that activity of the RebH arginine variants was not limited to the scaffold where it was originally identified. In fact, upon introduction of the mutation K79R into three RebH variants we observed each variant maintained activity for the substrates investigated and often had different selectivity. In collaboration with the Houk group, computational investigation shows that the arginine residue is competent for halogenation via activation of the putative HOX halogenating species, consistent with the proposed mechanism.

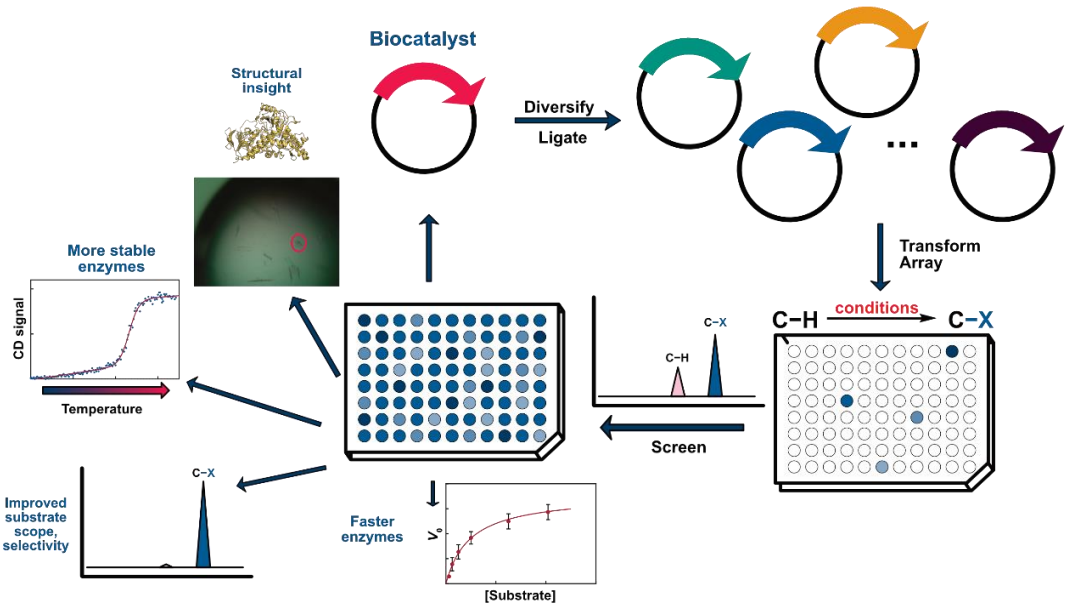
CHAPTER 1: AN INTRODUCTION TO THE METHODOLOGIES, METHODS, AND APPLICATIONS OF DIRECTED EVOLUTION WITH A FOCUS ON FLAVIN-DEPENDENT HALOGENASES.

1.1 Biocatalysis and directed evolution

The high activity and specificity of biocatalysts has long garnered interest among chemists[1]. Hydrolases found extensive use early in the adoption of enzymes to chemistry, with lipases being used for chemo-, regio-, and stereo-specific transformations of a variety of scaffolds[2]. Beyond hydrolysis or esterification however, biocatalysis was initially viewed as a field of limited practical applications. It was noted that wild-type enzymes may be useful for a given transformation but lacked the broader substrate scope that chemical catalysts often enjoyed. On the other hand, enzymes with a broader substrate scope tended to suffer from lower selectivity or stability.[3] Biocatalysis was therefore largely limited to production of primary metabolites such as amino acids[4] or sugars[5].

In the 1990s[6], researchers began to implement changes to the gene sequences encoding these enzymes and subsequently screening these enzyme libraries for improved or modified activity. Thus began the field of directed evolution, in which improvements to enzymatic activity are accumulated through multiple rounds of gene diversification and screening to ultimately achieve the desired activity and selectivity (**Figure 1.1**). Much of the initial work in directed evolution was done on hydrolases[7] due to the convenient screening associated with the hydrolysis of a pro-fluorogenic substrate, but such efforts now include transformations from C-H activations[8,9] to reductive aminations[10] with sophisticated screening[11] and selection[12] strategies. In the time since the invention and implementation of directed evolution, it has become a staple in both academic and industrial fields, and evolved biocatalysts are often the preferred method of synthesis for certain transformations on scales from milligrams to metric tons[13].

Figure 1. 1 Scheme for directed evolution



General scheme for diversification and screening of a biocatalyst for improved function. Genetic diversity is introduced into the starting template gene to construct an enzyme library, which is subsequently transformed into an appropriate expression vector and arrayed into 96-well plates. These plates are then screened for improved function, and the process is iterated until the desired function is achieved.

One of the earliest benefits of biocatalysis recognized from an industrial standpoint was the merging of often opposing goals: reducing long-term catalyst costs and cleaner reaction outcomes[14]. Biocatalysis has allowed the pharmaceutical industry to reduce its reliance on precious metals, which exact an environmental toll from extraction and purification, and have volatile market pricing.[15] Perhaps even more important to produce large-scale quantities of drug material is the ability to modify the catalyst for the desired process environment[16]. Unlike chemical catalysts which may require given conditions to function at optimal rates, biocatalysts can be molded through directed evolution to function at different pH values, temperatures, and cosolvent loadings. Even in cases where no enzyme is available for a desired transformation, techniques such as substrate walking[17] can be implemented to gradually gain function. In light of these and other reasons, the importance of directed evolution was

recognized in 2018 when Prof. Francis Arnold[18] was jointly awarded the Nobel Prize in chemistry for her work in applying directed evolution to enzyme catalysis.

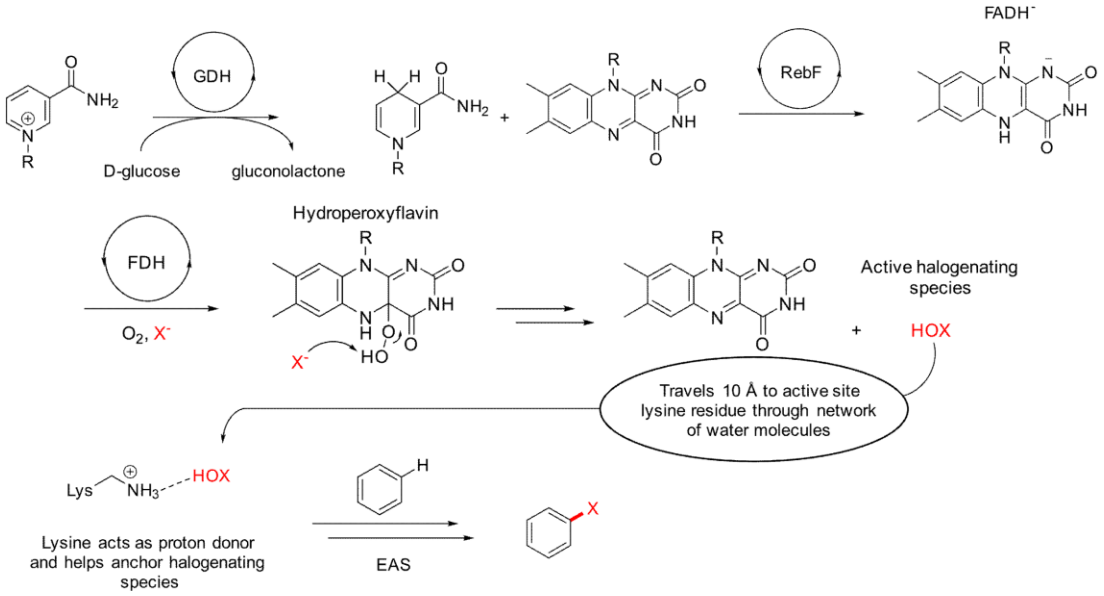
1.2 An introduction to flavin dependent halogenases

Prior work in our lab has focused on applying directed evolution engineering of flavin-dependent halogenases (FDHs) for organic synthesis[19]. These enzymes natively catalyze the site-selective chlorination or bromination of C-H bonds in electron rich aromatic[20] compounds and enolates[21]. Flavin-dependent halogenases utilize reduced flavin ($FADH_2$) typically produced from a flavin reductase partner protein[22]. The proposed mechanism for this class of enzyme (**Figure 1.2**) involves the formation of an HOX species after attack of oxygen to the reduced flavin[23]. This results in the formation of a hydroperoxyflavin which is subsequently attacked by X^- (X = Cl, Br). The resulting HOX then travels to the active site lysine residue which stabilizes this species within the active site via hydrogen bonding[24]. The enzyme then positions a suitably reactive C-H bond towards the HOX species, and a Wheland intermediate forms in the rate limiting step. This species is then deprotonated by a general base, suspected to be a glutamate residue in many FDHs.

A prototypical example of a flavin-dependent halogenase is the L-tryptophan 7-halogenase RebH, which chlorinates free-standing tryptophan at the 7-position of the indole ring. Our lab has previously demonstrated that in addition to the surprisingly broad substrate scope of wild-type RebH[25], this enzyme is amenable to directed evolution, including for improved substrate scope[26], thermostability[27], and site selectivity[28]. Evolved variants of RebH and other halogenases have also been extensively profiled for their activity, which has helped to establish thresholds for electronic limitations and informed about substrate scope[29]. More recently, we have also shown how these enzymes can be used for enantioselective catalysis on a variety of scaffolds, including the halocyclization of olefins with pendant nucleophiles[30,31] and the desymmetrization of methylenedianilines[32]. In the former case, RebH

variants capable of highly enantioselective halocyclization were available from campaigns directed at improving substrate scope, and in the latter, homology models were used to improve activity and selectivity.

Figure 1. 2: Mechanism for flavin-dependent halogenase catalysis



General mechanism for aromatic halogenation using flavin dependent halogenases. In two-component systems, reduced nicotinamide cofactor is used to produce reduced flavin ($FADH^-$). Reduced flavin then travels to the flavin-binding site of an FDH where it is used to activate molecular oxygen forming a hydroperoxyflavin intermediate. This intermediate is then attacked by a molecule of halide (e.g. Cl^-, Br^-) to form HOX, the active halogenating species. HOX then travels to the active site lysine residue where the HOX species is activated by hydrogen bonding from the lysine residue.

While studies of the catalytic flexibility of RebH have shown the utility of this enzyme in synthetic applications, it suffers from the same limitations as all explored FDHs. Primarily, low catalytic rate prohibits these enzymes from being more widely exploited for synthesis. This is especially true for evolved RebH variants, for which the highest catalytic efficiency recorded is one tenth[33] that of wild-type RebH. A potential reason for this problem is the diffusion of the reactive HOX species from the active[34] site, which is known to occur in FDH variants lacking the catalytic lysine mutation. More striking, however, is the distance between the flavin binding site of the enzyme and the catalytic lysine. For the HOX species to travel from the FAD binding site to the active site, it must travel through a poorly defined

10 Å "tunnel."[35] A possible consequence of the need to travel this tunnel can be observed when comparing the rates of FDH catalysis to flavin-dependent monooxygenases which function by positioning the reactive flavin species directly next to the substrate. For example, cyclohexanone monooxygenases have k_{cat} values of roughly 10 s^{-1},[36] while the flavin-dependent halogenase RebH exhibits a k_{cat} for the native tryptophan chlorination of 1.4 min^{-1}.[22]

In 2021, the flavin-dependent halogenase AetF was identified during efforts to characterize the biosynthetic gene cluster responsible[37] for producing a toxin resulting in avian vacuolar myelinopathy (AVM) in bald eagles. The native activity of AetF was later explored by the Moore group,[38] which identified it as a flavin-dependent halogenase that sequentially brominates the C5 then C7 position of tryptophan. Interestingly, AetF exists as a single component combination halogenase/reductase and does not require an additional partner protein to reduce flavin. From the perspective of applied biocatalysis, manipulation of a single protein is desirable over the incorporation of partner proteins as it reduces the time spent expressing and purifying protein for cofactor regeneration. Construction of a sequence-similarity network used to visualize the relationships between protein sequences confirmed that AetF is unique among known flavin-dependent halogenases. Unfortunately, in both the original paper describing AetF and the follow up characterizing the individual enzymes in the BGC, analysis of AetF activity on substrates other than the native tryptophan substrate was not explored.

1.3 Biosensor driven directed evolution of biocatalysts

While the results of directed evolution have been profound, the methods for directed evolution have largely remained the same since its invention. A typical round of evolution involves the *in vitro* generation of an enzyme library, cloning to transform or transfect the host with the enzyme library, expression of the library in well plates to facilitate recovery of improved variants, some form of lysis to release enzyme from the expression vector, screening of the library against the target substrate, and

analysis of the reactions to identify improved variants. Much effort has been made to reduce the labor-intensive nature of these campaigns, and the application of liquid-handling robots and other forms of automation has resulted in a substantial improvement to the number of library plates that can be processed in a day[39]. On-chip gene synthesis has also made exploring large amounts of the sequence space more accessible as lower costs for saturation per site allow for the construction of high-quality NNK libraries to facilitate deep-mutagenic scanning of a template sequence.[40]

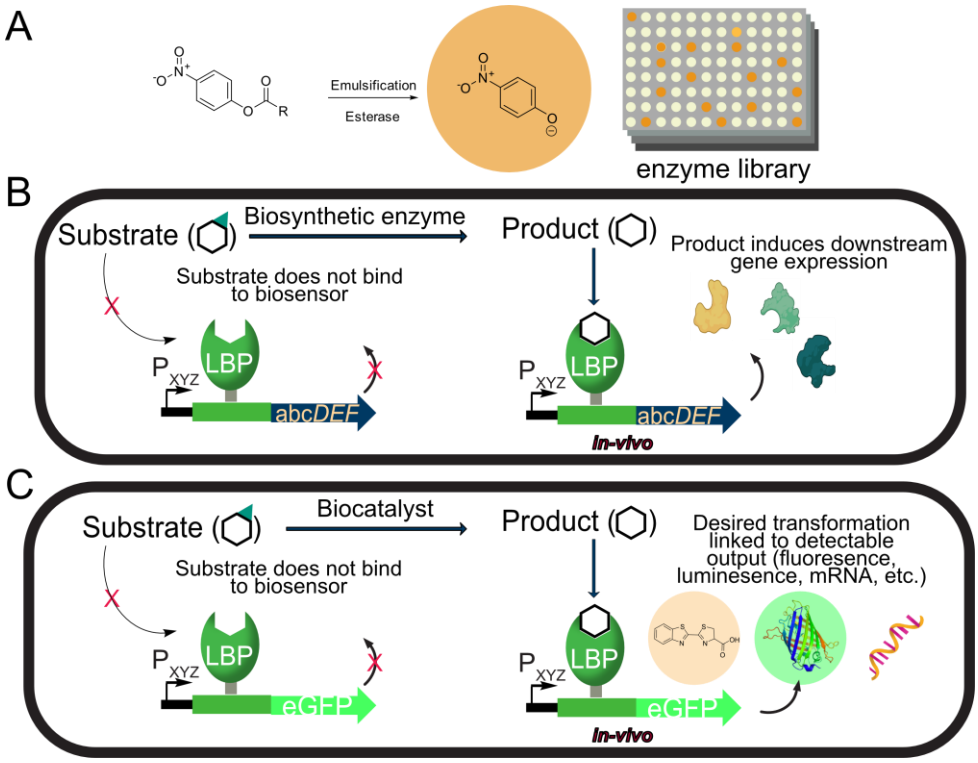
Despite these improvements to the 'front-end' of protein engineering, the lack of a generalizable assay for small molecules has precluded the same sort of increase to throughput in the directed evolution of biocatalysts. Without a generalizable colorimetric screen for the presence of a target product, more direct methods of analysis such as UHPLC or mass-spectrometry are required. For most labs this equipment and the associated consumables can be prohibitively expensive.[41] Even with well-maintained equipment and extensive method optimization, UHPLC-MS typically limits library sizes to 10^3-10^4 with a traditional 1-2 minute method.[42] Unlike in the case for colorimetric assays, this bottleneck caused by the limited throughput for reaction analysis severely restricts the sequence space that can be explored in a single round of evolution. Ultimately, this means only a small portion of the possible residue mutations are screened in each round of evolution and potentially vital and hard-to-predict mutations can easily be missed.[43] As the ability to deeply sequence and barcode enzyme libraries increases, the bottleneck that screening places on the genetic diversity explored in each round grows tighter and more high-throughput methods are required.

Table 1. 1: Methods of directed evolution and throughput[42]

Method	Library size	Advantages	Disadvantages	Examples
Selection	10^9	Only active variants obtained	Must connect organism fitness to enzyme activity	[44], [45,46], [47]
Agar Plate	10^5	Operationally simple	Limited analytical methods, low dynamic range	[48]
Microtiter plate	10^4	Many analytical strategies, target agnostic	Relatively small library sizes	[12,26,49,50]
Cell sorting	10^6	Large libraries	Requires connection between enzyme activity and fluorescent signal	[51]
Compartmentalization	10^9	Large libraries	Output must be connected to thermostable polymerase activity	[52]

One approach to increasing the throughput of screening techniques is to utilize methods that allow for more rapid screening than the ~0.25-2 min/sample typically allowed with UHPLC-MS approaches (**Table 1.1**).[53] One of the longest standing technologies for rapid screening is the generation of a fluorescent output.[54] Fluorogenic reactions have been used for the analysis of biocatalytic reactions both as a direct readout resulting from the formation of a fluorescent product and by the manipulation of side-product generation[55] or cofactor consumption.[56] These methods are useful for rapidly screening libraries that can reach 10^5 variants using automated liquid handler/plate reader systems, but each suffers from drawbacks inherent to the screening method. For example, fluorescent methods that rely on detecting cofactor consumption or turn-over are not linked to product formation, and often the enzyme will evolve to decouple cofactor consumption from substrate turnover.[57] Fluorogenic substrates (**Figure 1.3 A**) are useful for evaluating a large library for basic function or proper enzyme folding, but often the 'probe' substrates in such experiments are structurally distinct from the target substrate and the activity on one substrate may not transfer to the other. Rather than rely on the inherent properties of a given substrate to generate a detectable output, an alternative would be to tune the detector to recognize the desired product.

Figure 1. 3: Methods for fluorescent product production.



A) Fluorogenic substrates can be used to rapidly evaluate enzyme libraries but rely on inherent properties of substrates for detection. B) Transcription factor regulated gene expression can be used to control the expression of downstream genes only in the presence of biosynthetic intermediates. C) Engineered gene circuits can exploit the regulation of transcription factors to correlate the activity of a biocatalyst to a measurable output.

In biological systems, small molecule recognition is typically accomplished by proteins.[58] Often, especially in the context of metabolism and biosynthetic gene clusters, these proteins are ligand-dependent transcription factors (TFs) (**Figure 1.3 B**).[59] In these systems, the accumulation of a significant amount of some primary or secondary metabolite triggers the transcription of genes which may ultimately further modify the product into the next intermediate or encode transporter genes to release the final product into the local environment. This level of regulation allows organisms to carefully control the intracellular small molecule concentration and to thus avoid toxicity or metabolic bottlenecks. Once the potential of TFs for *in vivo* control of gene expression was recognized, these TFs began to be utilized in non-native contexts as well (**Figure 1.3 C**).[60] Desirable properties of TFs can be

improved through directed evolution, as previously described for biocatalysts, and a variety of positive and negative selection strategies have been developed to tune ligand specificity and dynamic range. [61] This adaptability has allowed transcription factors to become ubiquitous tools in synthetic biology. Entire engineered biosynthetic gene clusters have been developed using modified TFs for the high titer *in vivo* production of medicinally relevant products such as artemisinin[62] and cannabinoids[63] which otherwise require extraction from plants, the dependence on which can be vulnerable to political and environmental conditions.

Despite the application of biosensors to synthetic biology, they are rarely used in assays for the directed evolution of specific enzymes. This omission is in part due to the issue of small molecule diffusion, which often results in 'crosstalk' between cells where an active variant produces a significant amount of the product, and an inactive variant does not. Crosstalk caused by diffusion can be remedied with an oil-water-droplet system[64] in which cells expressing the biosensor and the enzyme of interest are encapsulated within a water droplet to facilitate the function of the genetic system required for transcription of biosensor genes. The water droplet is then suspended within an oil medium to reduce ligand diffusion. Frequently, to facilitate the rapid analysis of these samples, the emulsion is then submitted for analysis by fluorescence activated cell sorting (FACS), which can allow for the sampling of up to 10^6 variants per library, expanding the throughput over traditional HPLC based screens by orders of magnitude.

While FACS and related fluorescence-based screens can greatly increase the number of enzymes evaluated within a directed evolution campaign, these methods for screening larger libraries still suffer from some of the typical drawbacks of directed evolution efforts. The need for researcher intervention to construct libraries and identify improved variants introduces additional time and labor, therefore reducing the throughput of evolution campaigns. *In vivo* continuous evolution has the potential to

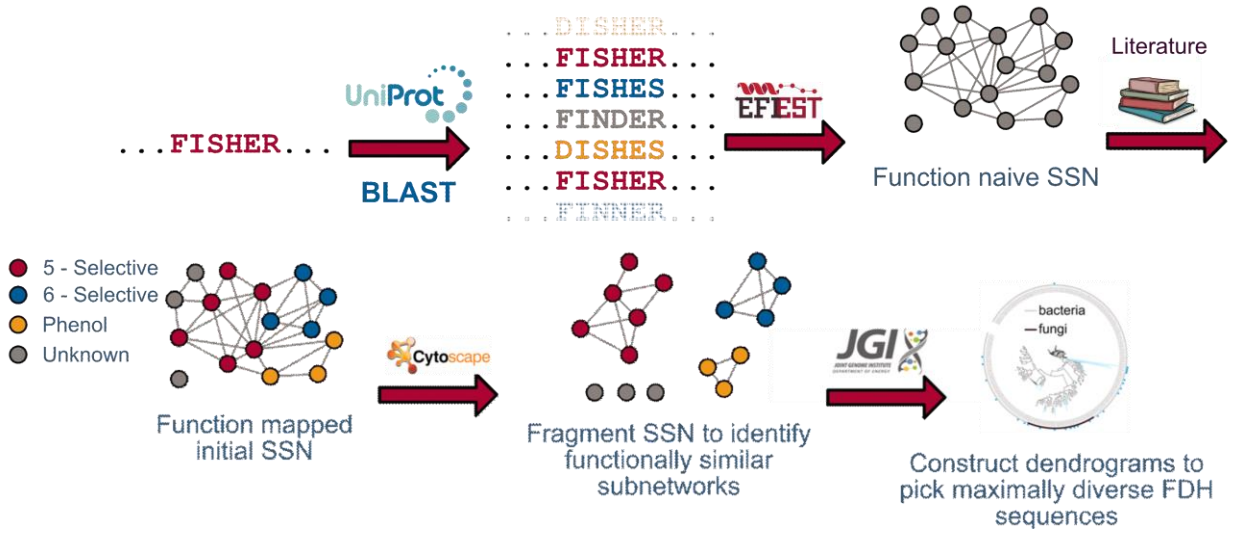
greatly accelerate this process by applying selection pressure based on biocatalyst activity, which eliminates the need for manual cycles of diversification, transformation, and screening required by traditional directed evolution approaches.[65] For continuous evolution to be implemented *in vivo*, however, the activity of the enzyme must be connected to the fitness of the organism via a biosensor. Different systems for continuous evolution, such as OrthoRep,[47] have been adapted for enzyme evolution in eukaryotic hosts, but in prokaryotic organisms, phage assisted continuous evolution (PACE) has found extensive use for the development of enzymes including proteases,[66] DNA-binding proteins,[67] and RNA polymerases.[68] A key feature of PACE that differentiates it from conventional directed evolution approaches is that by exploring such a breadth of the sequence space for a given protein, one evolution campaign can produce multiple improved variants with diverse sequences. The unbiased selection method can also provide mutations that would be difficult to arrive at rationally, such as the premature stop codon observed during tRNA synthetase evolution resulting in a split protein found to improve unnatural amino acid incorporation.[69] Application of this technology to the evolution of biocatalysts may provide a method for more rapid development of enzyme libraries enriched in active variants with diverse function.

1.4 Genome mining: searching for function from sequence

The previously mentioned FDH RebH was discovered through the characterization of the rebeccamycin biosynthetic pathway. Enzyme identification through BGC elucidation has remained a common method for identifying new enzymes. The decreasing cost associated with whole-genome sequencing and the advent of metagenomics for sampling DNA obtained from the environment has resulted in an enormous wealth of potential enzymes to be explored for biocatalysis.[69] Many practically useful enzymes have been identified from such efforts, but the complexity of BGCs [70] and metagenomic sampling makes functional assignment difficult. One such difficulty is that genes encoding proteins

involved in the biosynthetic pathway may not be localized with most other genes in the BGC, which results in incomplete recapitulation when the BGC is reconstituted in a heterologous host.[71] Additionally, while putative gene products obtained from genome sequencing or metagenomic sampling can have assigned annotations meant to inform about protein function, these are often misannotated due to lack of experimental characterization.[72]

Figure 1. 4: Generic genome mining scheme for the identification of enzymes from primary sequence



Starting from an input protein sequence, a BLAST search is used to identify related homologs. Using the enzyme function initiative's enzyme similarity tools, this list of sequences can be visualized as a sequence similarity network. Mapping of known function and increasing cluster stringency results in enzymes with similar function clustering together and informs about unknown sequences, which can then be sampled for functional analysis.

To simplify the assignment of putative protein functions, gene sequences are often translated to the corresponding protein sequence and scanned for commonly utilized motifs such as cofactor or metal binding residues that may hint at the native activity of the enzyme.[73] Putative enzyme assignments based on the presence of such motifs is often used to compare closely related BGC's to each other to identify enzymes responsible for differences in natural product formation. Using these structural motifs to identify potential proteins of a given function is one approach to so-called "genome mining".[74] As gene synthesis costs shrink and become more accessible, genome mining has been used in biocatalysis

to supplement directed evolution efforts by providing an advanced multiple potential enzymes to screen for a desired transformation. Furthermore, like directed evolution, genome mining can be viewed as an iterative process where multiple sequence clusters are explored to identify particularly active regions of the SSN which are then further mined for enzymes with improved activity.

1.5 Conclusions

The impressive impact of biocatalysis in chemical synthesis has brought new recognition to the field. Through the evaluation of enzyme libraries, new applications and modes of catalysis can be identified to expand the set and scope of transformations beyond those identified in nature. New approaches to enzyme identification may provide a complementary approach to directed evolution and facilitate identification of new enzyme activity. Likewise, traditional directed evolution approaches stand to benefit from the increased understanding of biological systems. By coupling enzyme function to gene circuits driving either fluorescent protein expression or proteins governing organism fitness, directed evolution campaigns can be shortened and made less operationally intensive.

1.6 References

(1) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nature* **2012**, *485* (7397), 185–194. https://doi.org/10.1038/nature11117.

(2) Hills, G. Industrial Use of Lipases to Produce Fatty Acid Esters. In *European Journal of Lipid Science and Technology*; 2003; Vol. 105, pp 601–607. https://doi.org/10.1002/ejlt.200300853.

(3) Khersonsky, O.; Tawfik, D. S. Enzyme Promiscuity - Evolutionary and Mechanistic Aspects. In *Comprehensive Natural Products III*; 2020; pp 705–734. https://doi.org/10.1016/B978-008045382-8.00155-6.

(4) May, O.; Verseck, S.; Bommarius, A.; Drauz, K. Development of Dynamic Kinetic Resolution Processes for Biocatalytic Production of Natural and Nonnatural L-Amino Acids. *Org Process Res Dev* **2002**, *6* (4), 452–457. https://doi.org/10.1021/op020009g.

(5) Beerens, K.; Desmet, T.; Soetaert, W. Enzymes for the Biocatalytic Production of Rare Sugars. *J Ind Microbiol Biotechnol* **2012**, *39* (6), 823–834. https://doi.org/10.1007/s10295-012-1089-x.

(6) Arnold, F. H. Design by Directed Evolution. *Acc Chem Res* **1998**, *31* (3), 125–131. https://doi.org/10.1021/ar960017f.

(7) Moore, J. C.; Arnold, F. H. Directed Evolution of a Para-Nitrobenzyl Esterase for Aqueous-Organic Solvents. *Nat Biotechnol* **1996**, *14* (4), 458–467. https://doi.org/10.1038/nbt0496-458.

(8) Hanefeld, U.; Hollmann, F.; Paul, C. E. Biocatalysis Making Waves in Organic Chemistry. *Chemical Society Reviews*. 2022, pp 594–627. https://doi.org/10.1039/d1cs00100k.

(9) Dong, J. J.; Fernández-Fueyo, E.; Hollmann, F.; Paul, C. E.; Pesic, M.; Schmidt, S.; Wang, Y.; Younes, S.; Zhang, W. Biocatalytic Oxidation Reactions: A Chemist's Perspective. *Angewandte Chemie - International Edition*. Wiley-VCH Verlag July 20, 2018, pp 9238–9261. https://doi.org/10.1002/anie.201800343.

(10) Aleku, G. A.; France, S. P.; Man, H.; Mangas-Sanchez, J.; Montgomery, S. L.; Sharma, M.; Leipold, F.; Hussain, S.; Grogan, G.; Turner, N. J. A Reductive Aminase from Aspergillus Oryzae. *Nat Chem* **2017**, *9* (10), 961–969. https://doi.org/10.1038/nchem.2782.

(11) Marshall, J. R.; Yao, P.; Montgomery, S. L.; Finnigan, J. D.; Thorpe, T. W.; Palmer, R. B.; Mangas-Sanchez, J.; Duncan, R. A. M.; Heath, R. S.; Graham, K. M.; Cook, D. J.; Charnock, S. J.; Turner, N. J. Screening and Characterization of a Diverse Panel of Metagenomic Imine Reductases for Biocatalytic Reductive Amination. *Nat Chem* **2021**, *13* (2), 140–148. https://doi.org/10.1038/s41557-020-00606-w.

(12) Jones, K. A.; Snodgrass, H. M.; Belsare, K.; Dickinson, B. C.; Lewis, J. C. Phage-Assisted Continuous Evolution and Selection of Enzymes for Chemical Synthesis. *ACS Cent Sci* **2021**, *7* (9), 1581–1590. https://doi.org/10.1021/acscentsci.1c00811.

(13) Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P.; Hayes, M. A.; Hepworth, L. J.; Lovelock, S. L.; Niikura, H.; Osuna, S.; Romero, E.; Ryan, K. S.; Turner, N. J.; Flitsch, S. L. Biocatalysis. *Nature Reviews Methods Primers* **2021**, *1* (1), 1–21. https://doi.org/10.1038/s43586-021-00044-z.

(14) Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med Chem Lett* **2017**, *8* (5), 476–480. https://doi.org/10.1021/acsmedchemlett.7b00114.

(15) Ludwig, J. R.; Schindler, C. S. Catalyst: Sustainable Catalysis. *Chem*. Elsevier Inc March 9, 2017, pp 313–316. https://doi.org/10.1016/j.chempr.2017.02.014.

(16) Ma, S. K.; Gruber, J.; Davis, C.; Newman, L.; Gray, D.; Wang, A.; Grate, J.; Huisman, G. W.; Sheldon, R. A. A Green-by-Design Biocatalytic Process for Atorvastatin Intermediate. *Green Chemistry* **2010**, *12* (1), 81–86. https://doi.org/10.1039/b919115c.

(17) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science (1979)* **2010**, *329* (5989), 305–309. https://doi.org/10.1126/science.1188934.

(18) Fasan, R.; Jennifer Kan, S. B.; Zhao, H. A Continuing Career in Biocatalysis: Frances H. Arnold. *ACS Catal* **2019**, *9* (11), 9775–9788. https://doi.org/10.1021/acscatal.9b02737.

(19) Andorfer, M. C.; Lewis, J. C. Understanding and Improving the Activity of Flavin-Dependent Halogenases via Random and Targeted Mutagenesis. *Annual Review of Biochemistry*. 2018, pp 159–185. https://doi.org/10.1146/annurev-biochem-062917-012042.

(20) Büchler, J.; Papadopoulou, A.; Buller, R. Recent Advances in Flavin-Dependent Halogenase Biocatalysis: Sourcing, Engineering, and Application. *Catalysts*. 2019, pp 1–20. https://doi.org/10.3390/catal9121030.

(21) Liu, M.; Ohashi, M.; Hung, Y. S.; Scherlach, K.; Watanabe, K.; Hertweck, C.; Tang, Y. AoiQ Catalyzes Geminal Dichlorination of 1,3-Diketone Natural Products. *J Am Chem Soc* **2021**, *143* (19), 7267–7271. https://doi.org/10.1021/jacs.1c02868.

(22) Yeh, E.; Garneau, S.; Walsh, C. T. Robust in Vitro Activity of RebF and RebH, a Two-Component Reductase/Halogenase, Generating 7-Chlorotryptophan during Rebeccamycin Biosynthesis. *Proc Natl Acad Sci U S A* **2005**, *102* (11), 3960–3965. https://doi.org/10.1073/pnas.0500755102.

(23) Yeh, E.; Blasiak, L. C.; Koglin, A.; Drennan, C. L.; Walsh, C. T. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases. *Biochemistry* **2007**, *46* (5), 1284–1292. https://doi.org/10.1021/bi0621213.

(24) Andorfer, M. C.; Evans, D.; Yang, S.; He, C. Q.; Girlich, A. M.; Vergara-Coll, J.; Sukumar, N.; Houk, K. N.; Lewis, J. C. Analysis of Laboratory-Evolved Flavin-Dependent Halogenases Affords a Computational Model for Predicting Halogenase Site Selectivity. *Chem Catalysis* **2022**. https://doi.org/10.1016/j.checat.2022.07.003.

(25) Andorfer, M. C.; Park, H. J.; Vergara-Coll, J.; Lewis, J. C. Directed Evolution of RebH for Catalyst-Controlled Halogenation of Indole C-H Bonds. *Chem Sci* **2016**, *7* (6), 3720–3729. https://doi.org/10.1039/c5sc04680g.

(26) Payne, J. T.; Poor, C. B.; Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angewandte Chemie* **2015**, *127* (14), 4300–4304. https://doi.org/10.1002/ange.201411901.

(27) Poor, C. B.; Andorfer, M. C.; Lewis, J. C. Improving the Stability and Catalyst Lifetime of the Halogenase RebH by Directed Evolution. *ChemBioChem* **2014**, *15* (9), 1286–1289. https://doi.org/10.1002/cbic.201300780.

(28) Andorfer, M. C.; Park, H. J.; Vergara-Coll, J.; Lewis, J. C. Directed Evolution of RebH for Catalyst-Controlled Halogenation of Indole C-H Bonds. *Chem Sci* **2016**, *7* (6), 3720–3729. https://doi.org/10.1039/c5sc04680g.

(29) Andorfer, M. C.; Grob, J. E.; Hajdin, C. E.; Chael, J. R.; Siuti, P.; Lilly, J.; Tan, K. L.; Lewis, J. C. Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal* **2017**, *7* (3), 1897–1904. https://doi.org/10.1021/acscatal.6b02707.

(30) Jiang, Y.; Mondal, D.; Lewis, J. C. Expanding the Reactivity of Flavin Dependent Halogenases Toward Olefins via Enantioselective Intramolecular Haloetherification and Chemoenzymatic Oxidative Rearrangements Yuhua. *ChemRxiv* **2022**, No. July, 1–6.

(31) Mondal, D.; Fisher, B. F.; Jiang, Y.; Lewis, J. C. Flavin-Dependent Halogenases Catalyze Enantioselective Olefin Halocyclization. *Nat Commun* **2021**, *12* (1). https://doi.org/10.1038/s41467-021-23503-3.

(32) Payne, J. T.; Butkovich, P. H.; Gu, Y.; Kunze, K. N.; Park, H. J.; Wang, D. S.; Lewis, J. C. Enantioselective Desymmetrization of Methylenedianilines via Enzyme-Catalyzed Remote Halogenation. *J Am Chem Soc* **2018**, *140* (2), 546–549. https://doi.org/10.1021/jacs.7b09573.

(33) Snodgrass, H. M.; Mondal, D.; Lewis, J. C. Directed Evolution of Flavin-Dependent Halogenases for Site- and Atroposelective Halogenation of 3-Aryl-4(3 H)-Quinazolinones via Kinetic or Dynamic Kinetic Resolution . *J Am Chem Soc* **2022**, *4*. https://doi.org/10.1021/jacs.2c07422.

(34) Phintha, A.; Prakinee, K.; Jaruwat, A.; Lawan, N.; Visitsatthawong, S.; Kantiwiriyawanitch, C.; Songsungthong, W.; Trisrivirat, D.; Chenprakhon, P.; Mulholland, A.; van Pée, K. H.; Chitnumsub, P.; Chaiyen, P. Dissecting the Low Catalytic Capability of Flavin-Dependent Halogenases. *Journal of Biological Chemistry* **2021**, *296*, 100068. https://doi.org/10.1074/jbc.RA120.016004.

(35) Prakinee, K.; Phintha, A.; Visitsatthawong, S.; Lawan, N.; Sucharitakul, J.; Kantiwiriyawanitch, C.; Damborsky, J.; Chitnumsub, P.; van Pée, K. H.; Chaiyen, P. Mechanism-Guided Tunnel Engineering to Increase the Efficiency of a Flavin-Dependent Halogenase. *Nat Catal* **2022**, *5* (6), 534–544. https://doi.org/10.1038/s41929-022-00800-8.

(36) Fürst, M. J. L. J.; Boonstra, M.; Bandstra, S.; Fraaije, M. W. Stabilization of Cyclohexanone Monooxygenase by Computational and Experimental Library Design. *Biotechnol Bioeng* **2019**, *116* (9), 2167–2177. https://doi.org/10.1002/bit.27022.

(37) Breinlinger, S.; Phillips, T. J.; Haram, B. N.; Mareš, J.; Martínez Yerena, J. A.; Hrouzek, P.; Sobotka, R.; Henderson, W. M.; Schmieder, P.; Williams, S. M.; Lauderdale, J. D.; Wilde, H. D.; Gerrin, W.; Kust, A.; Washington, J. W.; Wagner, C.; Geier, B.; Liebeke, M.; Enke, H.; Niedermeyer, T. H. J.; Wilde, S. B. Hunting the Eagle Killer: A Cyanobacterial Neurotoxin Causes Vacuolar Myelinopathy. *Science (1979)* **2021**, *371* (6536). https://doi.org/10.1126/science.aax9050.

(38) Adak, S.; Lukowski, A. L.; Schäfer, R. J. B.; Moore, B. S. From Tryptophan to Toxin: Nature's Convergent Biosynthetic Strategy to Aetokthonotoxin. *J. Am. Chem. Soc* **2022**, *144*, 2022. https://doi.org/10.1021/jacs.1c12778.

(39) Yi, D.; Bayer, T.; Badenhorst, C. P. S.; Wu, S.; Doerr, M.; Höhne, M.; Bornscheuer, U. T. Recent Trends in Biocatalysis. *Chemical Society Reviews*. 2021, pp 8003–8049. https://doi.org/10.1039/d0cs01575j.

(40) Kuhn, P.; Wagner, K.; Heil, K.; Liss, M.; Netuschil, N. Next Generation Gene Synthesis: From Microarrays to Genomes. *Engineering in Life Sciences*. Wiley-VCH Verlag January 1, 2017, pp 6–13. https://doi.org/10.1002/elsc.201600121.

(41) Welch, C. J.; Wu, N.; Biba, M.; Hartman, R.; Brkovic, T.; Gong, X.; Helmy, R.; Schafer, W.; Cuff, J.; Pirzada, Z.; Zhou, L. Greening Analytical Chromatography. *TrAC - Trends in Analytical Chemistry*. Elsevier July 1, 2010, pp 667–680. https://doi.org/10.1016/j.trac.2010.03.008.

(42) Leemhuis, H.; Kelly, R. M.; Dijkhuizen, L. Directed Evolution of Enzymes: Library Screening Strategies. *IUBMB Life*. 2009, pp 222–228. https://doi.org/10.1002/iub.165.

(43) Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nature Reviews Molecular Cell Biology 2009 10:12* **2009**, *10* (12), 866–876. https://doi.org/10.1038/nrm2805.

(44) Palmer, A. C.; Kishony, R. Understanding, Predicting and Manipulating the Genotypic Evolution of Antibiotic Resistance. *Nat Rev Genet* **2013**, *14* (4), 243–248. https://doi.org/10.1038/nrg3351.

(45) Dickinson, B. C.; Packer, M. S.; Badran, A. H.; Liu, D. R. A System for the Continuous Directed Evolution of Proteases Rapidly Reveals Drug-Resistance Mutations. *Nat Commun* **2014**, *5*. https://doi.org/10.1038/ncomms6352.

(46) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **2011**, *472* (7344), 499–503. https://doi.org/10.1038/nature09929.

(47) Rix, G.; Watkins-Dulaney, E. J.; Almhjell, P. J.; Boville, C. E.; Arnold, F. H.; Liu, C. C. Scalable Continuous Evolution for the Generation of Diverse Enzyme Variants Encompassing Promiscuous Activities. *Nat Commun* **2020**, *11* (1), 1–11. https://doi.org/10.1038/s41467-020-19539-6.

(48) Huang, R.; Chen, H.; Upp, D. M.; Lewis, J. C.; Zhang, Y.-H. H. P. J. A High-Throughput Method for Directed Evolution of NAD(P)+-Dependent Dehydrogenases for the Reduction of Biomimetic Nicotinamide Analogues. *ACS Catal* **2019**, *9* (12), 11709–11719. https://doi.org/10.1021/acscatal.9b03840.

(49) Fisher, B. F.; Snodgrass, H. M.; Jones, K. A.; Andorfer, M. C.; Lewis, J. C. Site-Selective C-H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. *ACS Cent Sci* **2019**, *5* (11), 1844–1856. https://doi.org/10.1021/acscentsci.9b00835.

(50) Gomez, C. A.; Mondal, D.; Du, Q.; Chan, N.; Lewis, J. C. Directed Evolution of a Fe(II)-and a-Ketoglutarate-Dependent Dioxygenase for Site-Selective Azidation of Unactivated Aliphatic C-H Bonds. *Chem RXIV* **2022**. https://doi.org/10.26434/CHEMRXIV-2022-DDFZP.

(51) Cheng, F.; Kardashliev, T.; Pitzler, C.; Shehzad, A.; Lue, H.; Bernhagen, J.; Zhu, L.; Schwaneberg, U. A Competitive Flow Cytometry Screening System for Directed Evolution of Therapeutic Enzyme. *ACS Synth Biol* **2015**, *4* (7), 768–775. https://doi.org/10.1021/sb500343g.

(52) Ellefson, J. W.; Ledbetter, M. P.; Ellington, A. D. Directed Evolution of a Synthetic Phylogeny of Programmable Trp Repressors. *Nat Chem Biol* **2018**, *14* (4), 361–367. https://doi.org/10.1038/s41589-018-0006-7.

(53) Kempa, E. E.; Galman, J. L.; Parmeggiani, F.; Marshall, J. R.; Malassis, J.; Fontenelle, C. Q.; Vendeville, J.-B.; Linclau, B.; Charnock, S. J.; Flitsch, S. L.; Turner, N. J.; Barran, P. E. Rapid Screening of Diverse Biotransformations for Enzyme Evolution. *JACS Au* **2021**, *1* (4), 508–516. https://doi.org/10.1021/JACSAU.1C00027.

(54) Wahler, D.; Reymond, J. L. Novel Methods for Biocatalyst Screening. *Current Opinion in Chemical Biology*. Current Opinion in Chemical Biology 2001, pp 152–158. https://doi.org/10.1016/S1367-5931(00)00184-8.

(55) Reisky, L.; Büchsenschütz, H. C.; Engel, J.; Song, T.; Schweder, T.; Hehemann, J.-H.; Bornscheuer, U. T. Oxidative Demethylation of Algal Carbohydrates by Cytochrome P450 Monooxygenases. *Nat Chem Biol* **2018**. https://doi.org/10.1038/s41589-018-0005-8.

(56) Schnepel, C.; Minges, H.; Frese, M.; Sewald, N. A High-Throughput Fluorescence Assay to Determine the Activity of Tryptophan Halogenases. *Angewandte Chemie - International Edition* **2016**, *55* (45), 14159–14163. https://doi.org/10.1002/anie.201605635.

(57) Morlock, L. K.; Böttcher, D.; Bornscheuer, U. T. Simultaneous Detection of NADPH Consumption and H 2 O 2 Production Using the Ampliflu^{TM} Red Assay for Screening of P450 Activities and Uncoupling. *Appl Microbiol Biotechnol* **2018**, *102* (2), 985–994. https://doi.org/10.1007/s00253-017-8636-3.

(58) Henley, M. J.; Koehler, A. N. Advances in Targeting 'Undruggable' Transcription Factors with Small Molecules. https://doi.org/10.1038/s41573-021-00199-0.

(59) Gupta, A.; Bedre, R.; Sudarshan, ‖; Thapa, S.; Sabrin, A.; Wang, G.; Dassanayake, M.; Grove, A. Global Awakening of Cryptic Biosynthetic Gene Clusters in Burkholderia Thailandensis. **2017**. https://doi.org/10.1021/acschembio.7b00681.

(60) Ding, N.; Zhou, S.; Deng, Y. Transcription-Factor-Based Biosensor Engineering for Applications in Synthetic Biology. **2021**, *15*, 7. https://doi.org/10.1021/acssynbio.0c00252.

(61) Xie, V. C.; Styles, M. J.; Dickinson, B. C. Methods for the Directed Evolution of Biomolecular Interactions. *Trends Biochem Sci* **2022**, *47* (5), 403–416. https://doi.org/10.1016/j.tibs.2022.01.001.

(62) Arsenault, R. P.; Wobbe, K. K.; Weathers, J. P. Recent Advances in Artemisinin Production Through Heterologous Expression. *Current Medicinal Chemistry*. 2008, pp 2886–2896. https://doi.org/http://dx.doi.org/10.2174/092986708786242813.

(63) Luo, X.; Reiter, M. A.; D'Espaux, L.; Wong, J.; Denby, C. M.; Lechner, A.; Zhang, Y.; Grzybowski, A. T.; Harth, S.; Lin, W.; Lee, H.; Yu, C.; Shin, J.; Deng, K.; Benites, V. T.; Wang, G.; Baidoo, edward E. K.; Chen, Y.; Dev, I.; Petzold, C. J.; Keasling, J. D. Complete Biosynthesis of Cannabinoids and Their Unnatural Analogues in Yeast. *Nature* **2019**, *567* (7746), 123–126. https://doi.org/10.1038/s41586-019-0978-9.

(64) Flachbart, L. K.; Sokolowsky, S.; Marienhagen, J. Displaced by Deceivers: Prevention of Biosensor Cross-Talk Is Pivotal for Successful Biosensor-Based High-Throughput Screening Campaigns. *ACS Synth Biol* **2019**, *8* (8), 1847–1857. https://doi.org/10.1021/acssynbio.9b00149.

(65) Morrison, M. S.; Podracky, C. J.; Liu, D. R. The Developing Toolkit of Continuous Directed Evolution. *Nat Chem Biol* **2020**, *16* (6), 610–619. https://doi.org/10.1038/s41589-020-0532-y.

(66) Dickinson, B. C.; Packer, M. S.; Badran, A. H.; Liu, D. R. A System for the Continuous Directed Evolution of Proteases Rapidly Reveals Drug-Resistance Mutations. *Nat Commun* **2014**, *5*. https://doi.org/10.1038/ncomms6352.

(67) Hubbard, B. P.; Badran, A. H.; Zuris, J. A.; Guilinger, J. P.; Davis, K. M.; Chen, L.; Tsai, S. Q.; Sander, J. D.; Joung, J. K.; Liu, D. R. Continuous Directed Evolution of DNA-Binding Proteins to Improve TALEN Specificity. *Nat Methods* **2015**, *12* (10), 939–942. https://doi.org/10.1038/nmeth.3515.

(68) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **2011**, *472* (7344), 499–503. https://doi.org/10.1038/nature09929.

(69) Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Söll, D.; Liu, D. R. Continuous Directed Evolution of Aminoacyl-TRNA Synthetases. *Nat Chem Biol* **2017**, *13* (12), 1253–1260. https://doi.org/10.1038/nchembio.2474.

(70) Itoh, N. Chapter 18 - Metagenomics for Improved Biocatalysis; Matsuda, T. B. T.-F. D. in B. (Second E., Ed.; Elsevier: Amsterdam, 2017; pp 375–384. https://doi.org/https://doi.org/10.1016/B978-0-444-63743-7.00018-4.

(71) Hertweck, C. Hidden Biosynthetic Treasures Brought to Light. *Nature Chemical Biology*. 2009, pp 450–452. https://doi.org/10.1038/nchembio0709-450.

(72) Xu, W.; Klumbys, E.; Ang, E. L.; Zhao, H. Emerging Molecular Biology Tools and Strategies for Engineering Natural Product Biosynthesis. *Metabolic Engineering Communications*. Elsevier B.V. June 1, 2020. https://doi.org/10.1016/j.mec.2019.e00108.

(73) Schnoes, A. M.; Brown, S. D.; Dodevski, I.; Babbitt, P. C. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput Biol* **2009**, *5* (12), 1000605. https://doi.org/10.1371/journal.pcbi.1000605.

(74) Höhne, M.; Schätzle, S.; Jochens, H.; Robins, K.; Bornscheuer, U. T. Rational Assignment of Key Motifs for Function Guides in Silico Enzyme Identification. *Nature Chemical Biology 2010 6:11* **2010**, *6* (11), 807–813. https://doi.org/10.1038/nchembio.447.

(75) Zaparucha, A.; de Berardinis, V.; Vaxelaire-Vergne, C. Chapter 1: Genome Mining for Enzyme Discovery. In *RSC Catalysis Series*; 2018; Vol. 2018-Janua, pp 3–27. https://doi.org/10.1039/9781788010450-00001.

CHAPTER 2: EVALUATING THE SUBSTRATE SCOPE AND REGIOSELECTIVITY OF FLAVIN-DEPENDENT HALOGENASES OBTAINED BY GENOME MINING

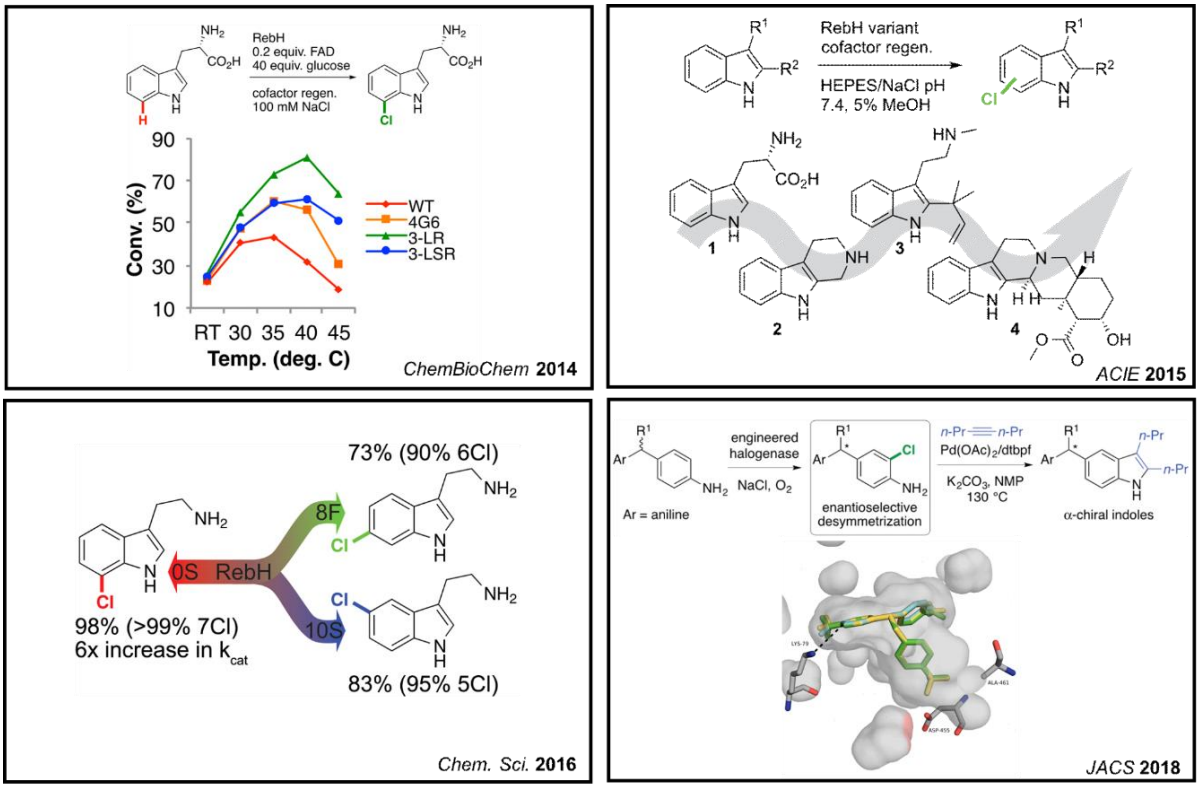
2.1 Introduction

Enzymes have found growing use in industry and academia as complements[1] or alternatives[2] for chemical synthesis. One drawback to directed evolution is that if an enzyme with the desired activity or selectivity is not available, substrate walking, a process in which intermediate substrates that are progressively more 'target-like' are screened, must be employed to develop activity on the desired subsrate.[3] This approach is time consuming and such intermediate substrates are not always readily available. Our group used substrate walking to evolve flavin-dependent halogenases (see chapter 1 for more details) for halogenation of biologically active compounds by challenging the enzyme libraries with progressively larger indole substrates.[4] While successful in developing activity on biologically active compounds, this evolution required several rounds on intermediate substrates to gain access to the target compounds, which slows the progress on improving activity on compounds of interest. Expanded access to halogenases with diverse activities could expedite these campaigns by providing advanced starting points for evolution of different substrates. The biocatalytic potential that could be enabled by a broader set of active enzymes highlights a clear need in the biocatalysis community to identify new biocatalysts that may have enhanced biocatalytic properties, such as broader substrate scope, enhanced stability, or higher activity.

Identifying and characterizing enzymes for biocatalysis has largely depended on the identification of natural products and the subsequent elucidation of the associated biosynthetic gene clusters (BGCs). [5] In addition to providing new biocatalysts, these efforts have led to approved drug compounds[6] and insights into transcriptional regulation[5], making contributions from this field invaluable. As high-throughput sequencing[7] and metagenomic technology[8] continues to improve and become cheaper, thousands of whole BGCs as well as individual gene sequences that potentially encode

synthetically useful biocatalysts have been catalogued and are recorded in readily accessible databases like Uniprot[9]. While many of these sequences remain uncharacterized, conserved sequence motifs provide insight to the native function of these enzymes.[10] Furthermore, sequence similarity networks (SSN's) can be used to cluster these genes together based on their relative identity. Supplementing these SSNs with experimental data from the characterized sequences can provide some insight into the potential function of related but uncharacterized genes.[11]

Figure 2. 1: Representative Results Of Directed Evolution Of FDHs



Summary of previous work for the biocatalytic application of flavin-dependent halogenases. In each case, the lack of available wild-type halogenases necessitated the evolution from RebH or other evolved RebH variants.

Genome mining for new biocatalysts has emerged as a powerful tool for biocatalysis. By directly screening for desired activity and characterizing deposited gene sequences, researchers have identified enzymes that provide distinct selectivity on substrates of interest,[12] that can generate novel natural products,[13] and that allow for predictions regarding uncharacterized enzyme activity,[14] all without

characterization of the gene sequence in the native context. In the absence of a general platform for evaluating varied enzymatic transformations, many of these studies were focused on a relatively small subset of enzymes or substrates. Profiling of enzyme activity using many different substrates is comparatively rare. Additionally, none of these previous studies involved C-H activation reactions, a highly sought-after synthetic transformation. In the work described herein, we sought to demonstrate the utility of genome mining by using it to greatly expand the number of experimentally characterized FDH's, as well as to gain insights to the family of FDH enzymes by developing activity-sequence relationships that may help to further inform subsequent mining experiments to improve the rate of active halogenase discovery.

Authorship

This work was performed with Dr. Brian Fisher, Dr. Krysten Jones, and Dr. Mary Andorfer. Prof. Jared Lewis developed the SSN, Dr. Brian Fisher performed the high-throughput method development for each substrate, and all authors were involved in the determination of soluble expression for the different enzymes described herein. These results have been included to provide context.

2.2 Results and Discussion

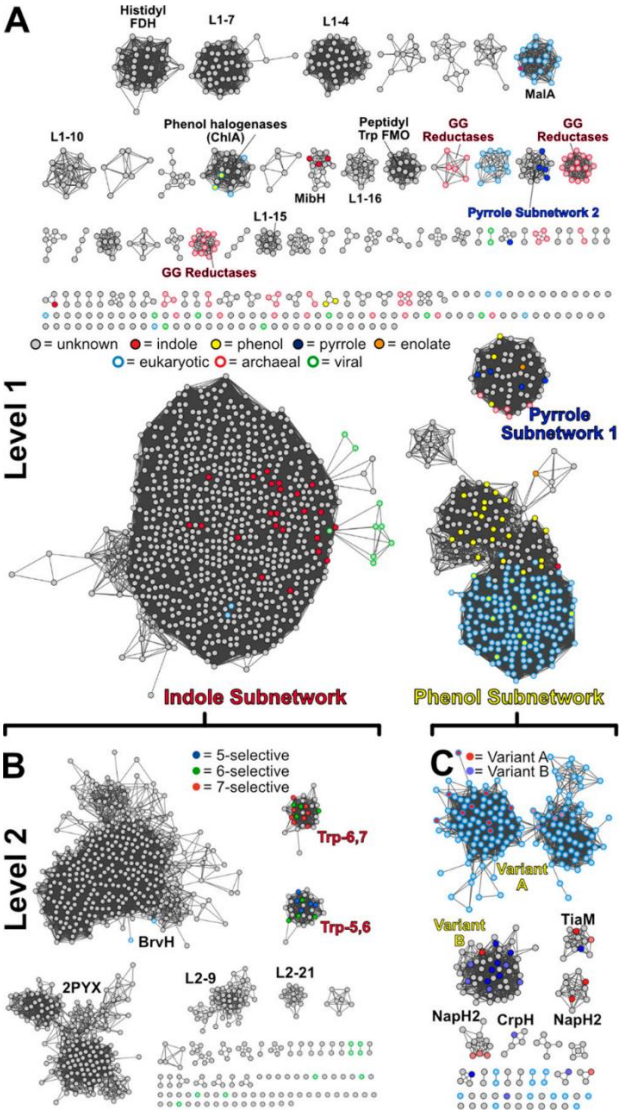
2.2.1 Construction of the SSN and expression of representative FDH family members

The Enzyme Similarity Tool (EST) developed by the Enzyme Function Initiative was used to conduct a BLAST search of the UniProt sequence database using RebH as the query sequence with an *E*-value threshold of 10^{-5}. This process returned 4380 unique gene sequences encoding potential halogenases. To cull the data set to more accurately represent previously characterized halogenase genes, sequences corresponding to less than 285 residues and greater than 1038 residues were removed, and sequences corresponding to FDHs discovered after the initial BLAST search were included to arrive at a final number of 3975 putative FDH genes. Importantly, nearly all previously identified (>90%) FDH genes were represented in this set from a diverse host organism pool with viral, eukaryotic,

prokaryotic, and archael genes. The GxGxxG FAD-binding motif common to flavin-dependent monoxygenases and halogenases[15] was present in 92% of the sequences and the WxWxI[R,P,G] motif characteristic of previously characterized halogenases was found in 78% of the sequences.[16] The presence of both motifs indicates that the majority of these sequences are likely FDHs (**Figure 2.2**).

A representative sequence similarity network (SSN)[17] was generated using an alignment score of 70 to facilitate visual inspection of the putative FDH sequences. Of the 3975 sequences included in the SSN, 129 were known enzymes for which the experimental data was mapped onto the SSN to provide some context for the activity of the unknown sequences. To our gratification, the previously characterized enzymes tended to cluster together based on the known or native substrates halogenated by the known FDHs. This finding allowed us to define the various subnetworks by substrate class. For example, the largest subnetwork comprised 2270 sequences and contained within it all the known tryptophan halogenases, including the 5-, 6-, and 7- halogenases PyrH, SttH, and RebH. Additionally, several halogenases identified from metagenomic studies that are found in this network were found to halogenase small indoles, despite the native substrate being unknown. Similar clustering of known activity can be found in the second and third largest subnetworks, one which contains 438 sequences and includes most of the known phenol halogenases and a second which includes 212 sequences that includes PrnC, which natively halogenates a pyrrole-containing compound.

Figure 2. 2: Level 1 And 2 SSNs Constructed form FDH Sequences



(A) Sequence-similarity network for flavin-dependent halogenases. Each circle is a representative node, grouping protein sequences with >50% sequence identity as determined by CD-HIT.65 Edge detection threshold set at alignment score of 70 (≈30% sequence identity). Nodes are filled according to native substrate functional group of at least one sequence in the representative node; colored stroke indicates domain (thin black stroke = bacterial). Subnetworks with ≥15 sequences but without any characterized protein are labeled numerically. Level 2 subnetworks formed from the Indole (B) and Phenol (C) subnetwork using a stricter alignment score cutoff of 140 (≈40% sequence identity). Level 2 subnetworks are labeled based on known sequences in the subnetwork. For Indole Subnetwork sequences, nodes containing known tryptophan halogenases are filled according to their regioselectivity, and subnetworks with ≥15 sequences are labeled numerically. For Phenol Subnetwork sequences, nodes are filled according to the halogenase variant type (A = free small molecule native substrate, B = ACP-tethered native substrate). SOURCE: ACS Cent. Sci. 2019, 5, 1844−1856

Flavin-dependent halogenases natively halogenate either free small molecule substrates (so-called “variant A” halogenases) or protein-bound substrates (“variant B” halogenases), and this difference was also captured by the SSN. Examining the SSN with increasing stringency (alignment score 140) resulted in generation of a Level 2 SSN, which further segments the phenol subnetwork into additional subnetworks. This increased stringency separated nearly all previously characterized Variant A from Variant B halogenases. The isolation of variant A and B FDH’s further demonstrates that even with the sparse population of characterized FDH’s in the SSN, useful information could still be drawn from it.

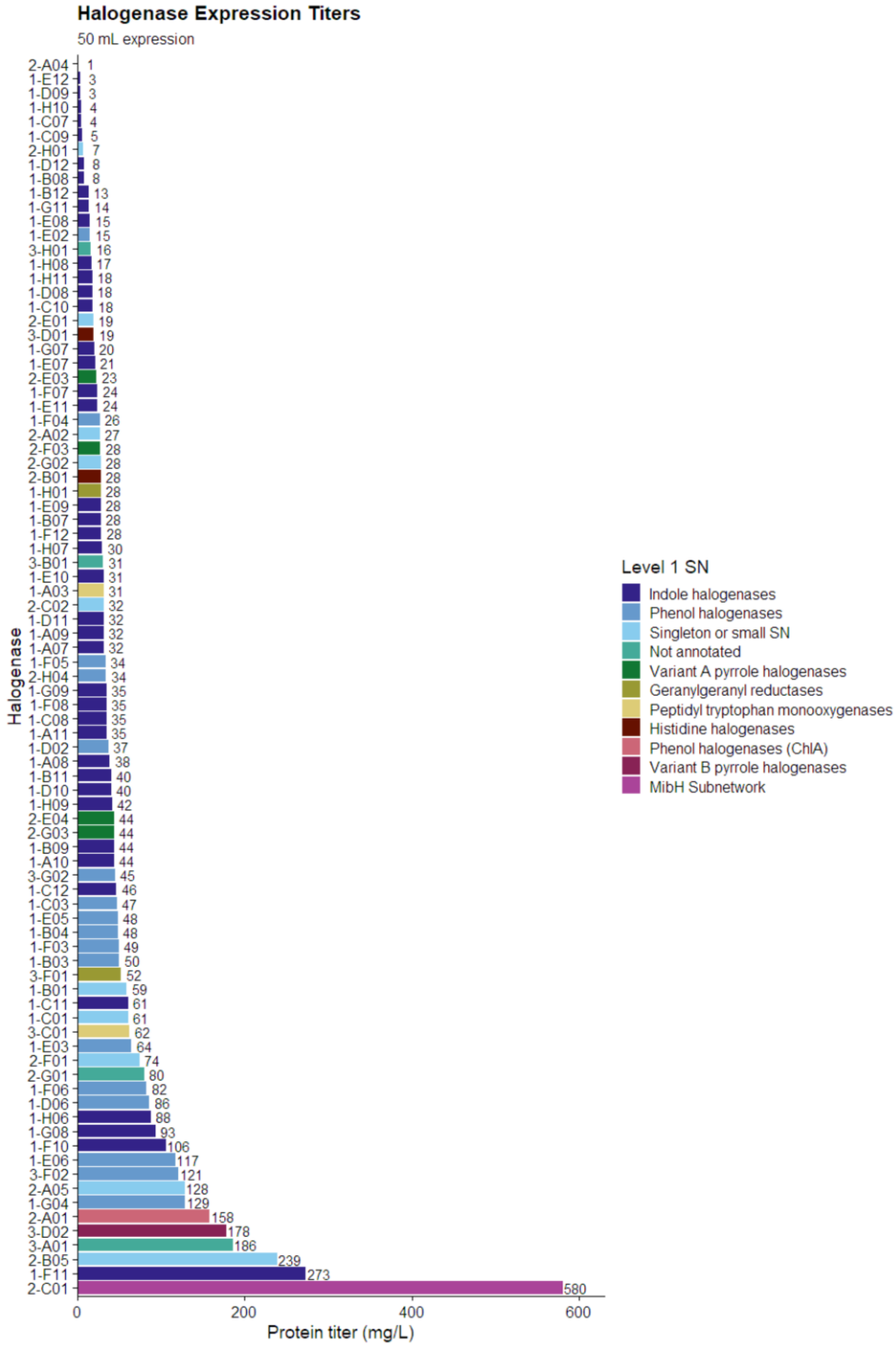
CD-HIT, a clustering algorithm designed to reduce redundancy within reference databases,[18] was used to select sequences from each of the level 1 SSN subnetworks for synthesis and expression. This analysis allowed us to broadly sample the subnetworks while removing sequences with high identity to one another, increasing the chances of obtaining enzymes with differing substrate scope. In collaboration with Dr. Krysten Jones and Dr. Brian Fisher, I cloned 120 putative FDH genes into pET28b, and the resulting plasmids were individually co-transformed into BL21(DE3) *Escherichia coli* with pGro7 chaperone plasmid. Small-scale expression cultures (3mL) were prepared, and the cells were lysed. SDS-PAGE analysis was used to determine whether sufficient quantities of each FDH could be obtained to warrant large-scale expression. Of the 128 putative halogenase genes synthesized, 87 new proteins (68% of the entire set) were obtained in at least single digit milligram per liter quantity upon large-scale expression, suitable for biocatalytic characterization. Phenol halogenases (49% overall, 17 total) were not affected by host domain (45% soluble eukaryotic, 50% soluble prokaryotic) but were overall less soluble than indole halogenases (91%, 42 total). A fair amount of soluble pyrrole subnetwork halogenases were obtained (71%, 5 total), whereas the smaller subnetworks tended to produce fewer soluble halogenases under the expression conditions. Taken together, these results demonstrate the robustness of the general halogenase expression conditions system used here.

2.2.2 Purification, high-throughput screening, and validation

Methodology for screening soluble FDHs on different substrates was next developed. An important decision in this regard was whether to screen the halogenase genes in lysate or as purified enzymes. Evaluation of enzymes in lysate obviates the need for purification and can result in a cheaper and quicker workflow, but the earlier evaluation of the solubility of our genome mined FDH panel showed a high level of variability in expression levels from 100's of mg/L to single digit mg/L titers (**Figure 2.3**). To avoid biasing our study solely to high expressing enzymes we opted for individual purification. Purified enzyme after IMAC was then normalized to a concentration of 25 µM in storage buffer (HEPES, pH 7.4, 25 mM with 10% glycerol) and stored at -20 °C in a PCR plate until use.

The 87 soluble putative halogenase proteins were then subjected to a high-throughput activity screen using a panel of 12 representative halogenase substrates. These 12 substrates were chosen due to their good detection limits by LC-MS, previously established compatibility of known FDHs, and high calculated halenium affinity values,[19] which collectively improved the probability of observing activity on at least one substrate even if the substrate binding was poor or conversion was low. Both chlorination and bromination of the 12 probe substrates with all 87 putative FDHs totaled 2,088 independent experiments, prompting Dr. Brian Fisher to develop an MS-based method analogous to the MISER (Multiple Injections in a Single Experimental Run) approach to expedite screening.[20] Separation of the buffer front and the substrate and product was achieved on a reverse-phase guard column, and the sample was injected directly into the MS to quantitate the substrate and product ions. We considered the possibility of multiplexing the halogenation step by including multiple substrates in a single reaction, but a preliminary experiment using the active halogenase 1-F11 showed that when two known substrates were included in one reaction yield for one substrate dropped to zero showing the potential for lost activity for reasons that were not further evaluated (**Scheme 2.1A**). Instead, to increase throughput, we investigated pooling samples together after biocatalysis. To determine if this approach

Figure 2. 3: Expression titers for genome mined FDHs

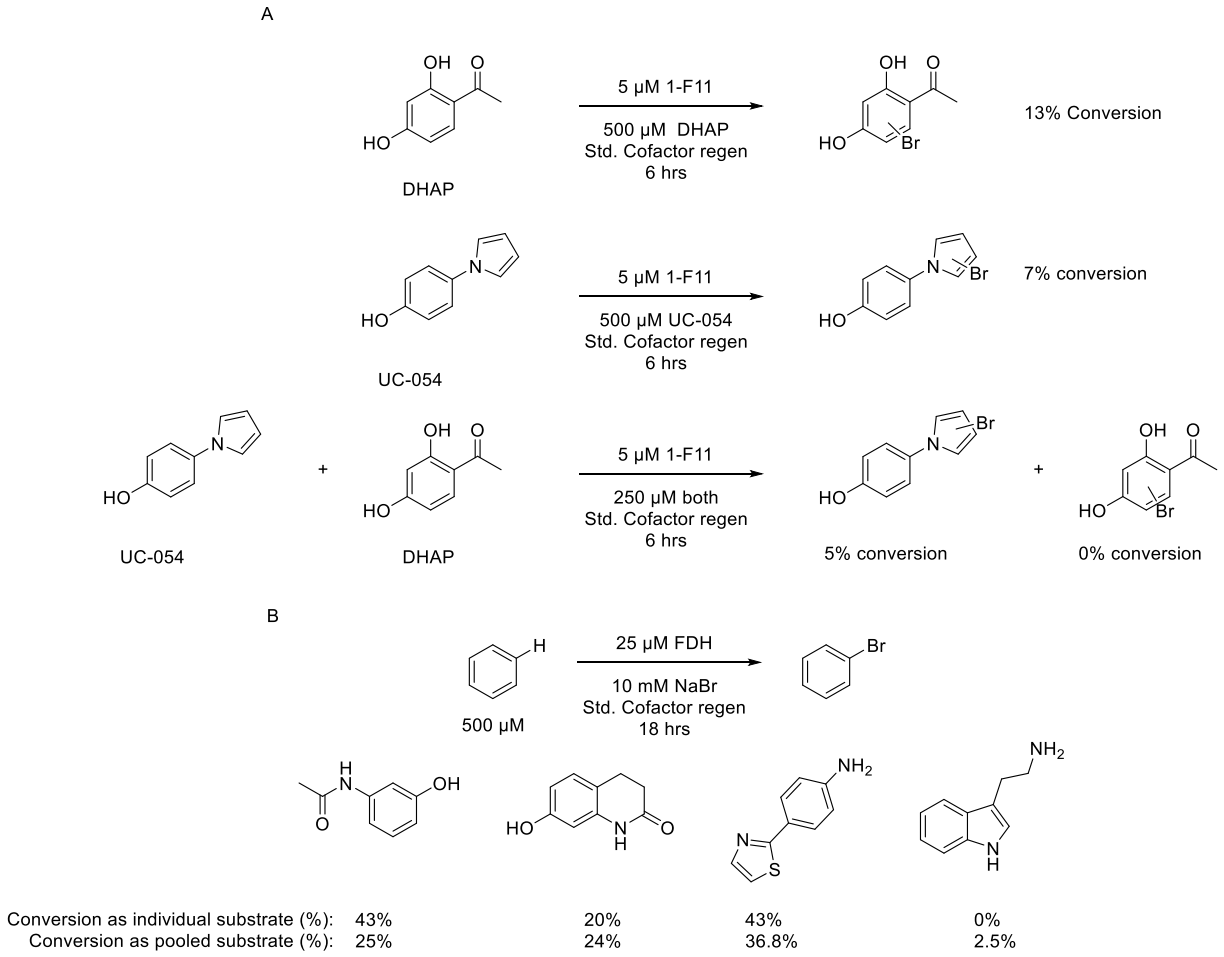


Expression titers correspond to the values obtained from coexpression of halogenase + chaperone plasmid pGro7 as determined by the concentration of halogenase purified from 50 mL culture normalized to 1L expression assuming linear correlation. SOURCE: ACS Cent. Sci. 2019, 5, 1844−1856

would result in false negatives, a preliminary experiment was conducted using the same enzyme, 1-F11, and bromination of a panel of compounds that we suspected would ionize well in positive ion mode due to the presence of basic amines. After overnight incubation, an aliquot of each reaction was pooled, and the conversion was evaluated by MS both for the individual samples and pooled reactions (**Scheme 2.1B**). This showed that the overall conversion values trended well whether the samples were pooled together or analyzed individually, and importantly no false negatives were observed which may lead us to incorrectly mark an enzyme as inactive. Given these results, we decided to pursue the latter approach, and for the substrate scope of our study we pooled substrates after quenching the bioconversion. Depending on the ionizability of the substrate, either 4 or 2 substrates were analyzed in a single injection to ultimately provide a sampling time of 10 seconds per reaction. To preserve precious enzyme material, the reactions were scaled down from the 96-well plates traditionally used for FDH reactions to a 384-well plate, reducing the volume from 75 µL to 10 µL, allowing for even poorly expressing enzymes to be tested in dozens of experiments.

Ultimately, we achieved an analysis time of ≈ 11s per sample and screened ≈ 20,000 halogenase reactions. Of the 87 soluble proteins screened for activity, 39 new halogenases were identified from this probe substrate screen. At this point, a subset of the FDH-substrate pairs was chosen for low-throughput validation by UPLC. Enzyme-substrate pairs chosen for validation were picked to sample both high and low activity enzymes to examine the potential for false negatives or positives and were meant to test the accuracy of the HTS for both positively and negatively ionizing substrates. Comparing the activity values for the UPLC data and the LCMS data resulted in a good correlation, showing the HTS to be reliable for interpreting activity (**Figure 2.4**). One caveat to this analysis is that conversion for substrates that were detected in negative ion mode tended to be overestimated due the low signal to noise ratio due to ion suppression by TFA in the mass spectrometer.

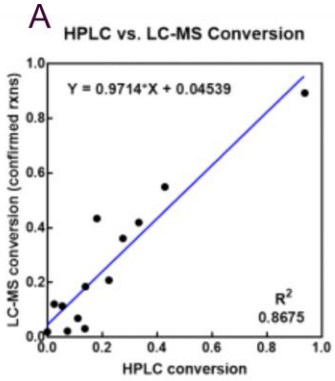
Scheme 2. 1: Conversion With Individual And Pooled Substrates



A) Results of pooling samples during biocatalysis. DHAP, previously converted to 13%, shows no activity demonstrating the potential for ablation of activity for low-activity enzymes. B) Results of pooling samples after biocatalysis, showing correlation between activity when analyzed as an individual sample and activity when measured as a pool of samples.

Figure 2. 4: Correlation Between Low And High Throughput Screen For Subset Of Enzyme-Substrate Pairs

A



B

Enzyme	Substrate	Reason
BSA	L-Trp	Negative control (non-FDH)
1-F07	Pindolol	Negative control (non-active enzyme/substrate pair)
RebH	UC-066	Positive control (negative mode ionization)
1-F08	Pindolol	Highly active enzyme-substrate pair
1-F10	Pindolol	active enzyme/substrate pair
1-F11	Pindolol	active enzyme/substrate pair
1-B12	Pindolol	active enzyme/substrate pair
1-H11	Pindolol	Low activity enzyme-substrate pair

A) Relationship between high-throughput LCMS data and low throughput HPLC data. A correlation coefficient of 0.87 corresponds to an excellent positive relationship and validates the high-throughput screening for discovery of new FDHs. For HPLC screening, conversion was measured as the total area of all products over starting material and all products measured at 220 nm, whereas for the high-throughput screen EIC values were used for quantification. B) Table showing the identity of the enzyme-susbtrate pairs chosen for validation and the reasons for selection. SOURCE: ACS Cent. Sci. 2019, 5, 1844−1856

A prevalent trend that emerged from this screening effort was the higher bromination activity when compared to chlorination. Compared to 45% for bromination, only 16% of the enzyme set exhibited any chlorinase activity, in stark contrast to much of the existing FDH literature. Of the host domains, only archael proteins were completely inactive; 48% of bacterial, 56% of eukaryotic, and one of three viral proteins having some detectable activity on the probe substrates. Dr. Brian Fisher constructed a heatmap and observed clustering of the halogenases based on activity that roughly mirrored activity assignments in the original SSN subnetworks, showing that our SSN can be used in a predictive manner to identify halogenases that act on a particular substrate class.

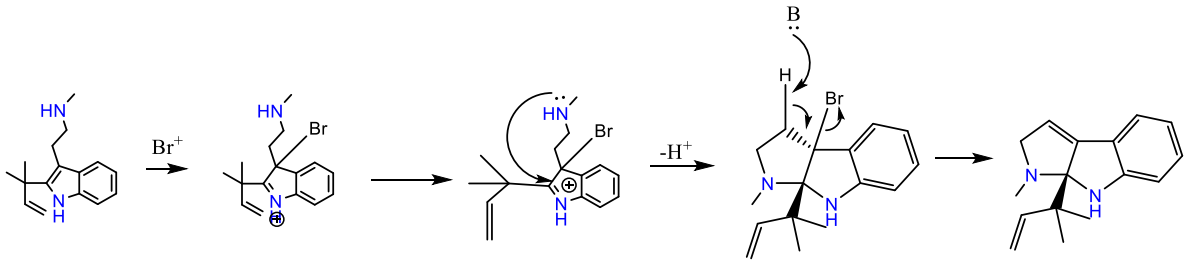
Based on the number of active halogenases identified using probe substrates, we sought to expand the set of substrates to encompass more three-dimensionally complex compounds, many of which are biologically active compounds that either required directed evolution to access previously or are compounds for which FDH activity has never been tested. 50 of these substrates were chosen, and 48% of these were successfully halogenated by at least one enzyme in the HTS. Some reactions also led

to the formation of multiple monobrominated products, prompting us to investigate the reactions on preparative scale to identify if the observed masses were a result of regio-complementary halogenations. In addition to the enzyme purification and reaction/assay optimization noted above, this avenue of research was my primary contribution to this project.

2.2.3 Preparative halogenation reactions and determination of regioselectivity

While the high-throughput mass-spectrometer based screening proved reliable for determining activity of the various enzymes included in the screen, subsequent low-throughput UPLC-MS screens were required to identify enzyme-substrate pairs that had differing regioselectivity. To evaluate FDH site selectivity, several compounds that were halogenated by multiple enzymes, as well as some that could not be included in the high-throughput screen due to poor ionizability, were included in a secondary screen. While most of the halogenases tended to form only one major monobrominated product, a few substrates provided notable examples of regiocomplementarity. In addition to the examples that will be highlighted below, 2-C01-catalyzed bromination of 2-(1,1-dimethyl-2-propen-1-yl)-*N*-methyl-1*H*-Indole-3-ethanamine was found to result in formation of a product with a mass corresponding to starting material -2H. We suspect a transient halogenation analogous to the rearrangement affected by NBS[21] as detailed in **Scheme 2.2** but an attempt to isolate the product failed.

Scheme 2. 2: Proposed Cyclization Of 2-(1,1-Dimethyl-2-Propen-1-Yl)-N-Methyl-1H-Indole-3-Ethanamine



SOURCE: ACS Cent. Sci. 2019, 5, 1844−1856

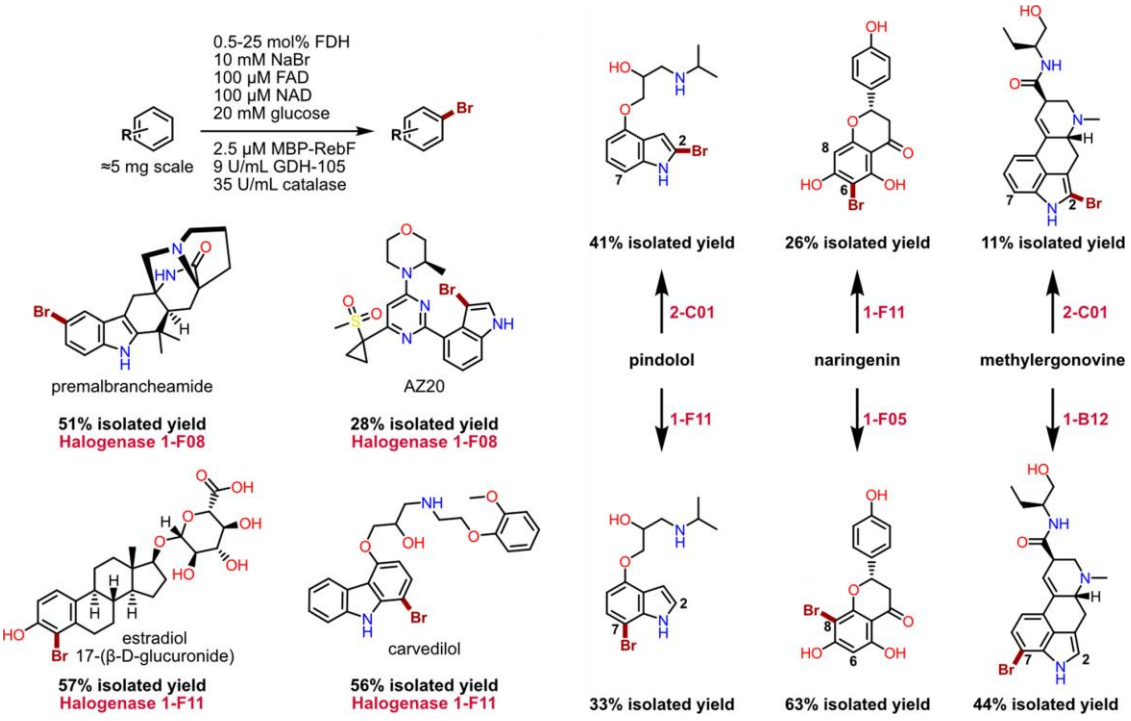
With LCMS traces indicating multiple mono-brominated products for naringenin, pindolol, and cabergoline with different enzyme-substrate pairs, we then investigated conditions for scaling up the

enzymatic reactions to maximize the isolated yield. The general workflow for isolating material from biocatalytic reactions was to conduct halogenase reactions on 3-5 mL scale in 20 mL scintillation vials. Reactions were incubated via shaking in an incubator or plate shaker to provide agitation in the absence of a stir bar. Multiple reactions were then pooled together and quenched via the addition of 1 M HCl to pH ~2 to denature proteins. The precipitated proteins were then filtered through celite, and if the compound was an alkaloid, the reaction was basified to pH > 10 and extracted into DCM. The organic layer was then washed with brine, dried with magnesium sulfate, filtered, and concentrated via rotary evaporation. The resulting crude mixture was either adsorbed onto celite and purified via biotage or dissolved in DMSO and purified via reverse phase HPLC.

Some substrates proved problematic to the standard workflow and required individualized optimization to facilitate isolation of material for characterization. For example, cabergoline was found to be problematic during the isolation process, with an apparent degradation product emerging that made isolation of a substantial amount of the desired brominated products impossible. Fortunately, switching to the structurally similar ergot-alkaloid methylergonovine maintained a very similar activity profile for the enzymes tested and halogenated products were successfully isolated. Initial HPLC purification of the 2-C01 reaction with pindolol resulted in oxidation of the 2-bromopindolol product into the corresponding oxindole under the conditions of heat + 0.1% TFA in water during the removal of solvent by rotary evaporation. In the absence of TFA as an additive during purification, the peak became so broad no separation was observed and purification was not possible. Switching to formic acid in the mobile phase instead circumvented this problem and it was possible to use the same method for isolating both the 7- and 2-bromopindolol. For the 1-F05 bromination of Naringenin low activity of enzyme 1-F05 initially prohibited isolation of the major product. After optimization of temperature, reaction at 4 °C significantly improved the yield of desired product, and lowering the substrate concentration to 500 µM of the substrate gave the highest conversion to the monobrominated material.

It was also found during the workup of both this reaction and the 1-F11 reaction that high heat could not be used during the concentration step as it resulted in non-specific polybromination of the highly electron rich flavanone ring.

Figure 2. 5: Scaled Up FDH Reactions

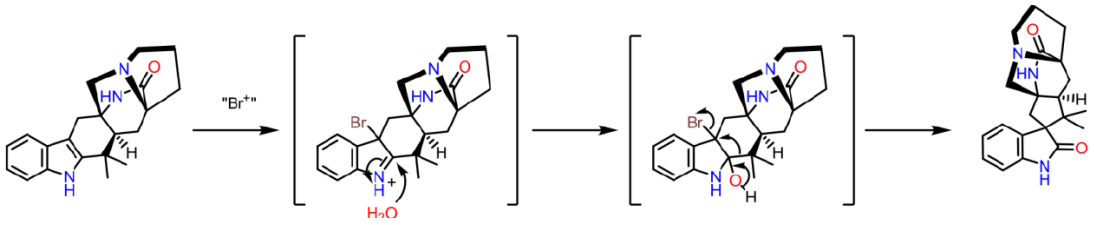


Regioselectivity was determined via 2D-NOESY. SOURCE: ACS Cent. Sci. 2019, 5, 1844−1856

Other halogenase reactions were of interest for the complex or biologically active substrates they halogenated (**Figure 2.5**). For the 1-F11-catalyzed reaction with 17β-estradiolglucoronide, acidification of the enzymatic reaction with 0.5 M HCl prior to extraction with DCM resulted in hydrolysis of the glucose moiety, and careful titration of the reaction with 10% citric acid to a pH just below the pKa of the glucuronic acid (pH=3) was required to maximize the yield during isolation. While the primary product in the 1-F08 catalyzed bromination of premalbrancheamide was stable to the typical isolation conditions, a secondary monobrominated material was also observed by LCMS but could not be isolated without decomposition. Based on work from the Sarpong group,[22] we suspect a transient bromination at C3 which degrades under our conditions to a spirocyclic product (**Scheme 2.3**), but this possibility was

not investigated. The hydration of the C2-brominated AZ20 to the corresponding oxindole obtained from the 1-F08 bioconversion could not be mitigated and was formed spontaneously upon reaction prior to any workup.

Scheme 2. 3: Proposed Oxindole Formation Observed During Workup Of 1-F08 Catalyzed Halogenation Of Premalbrancheamide

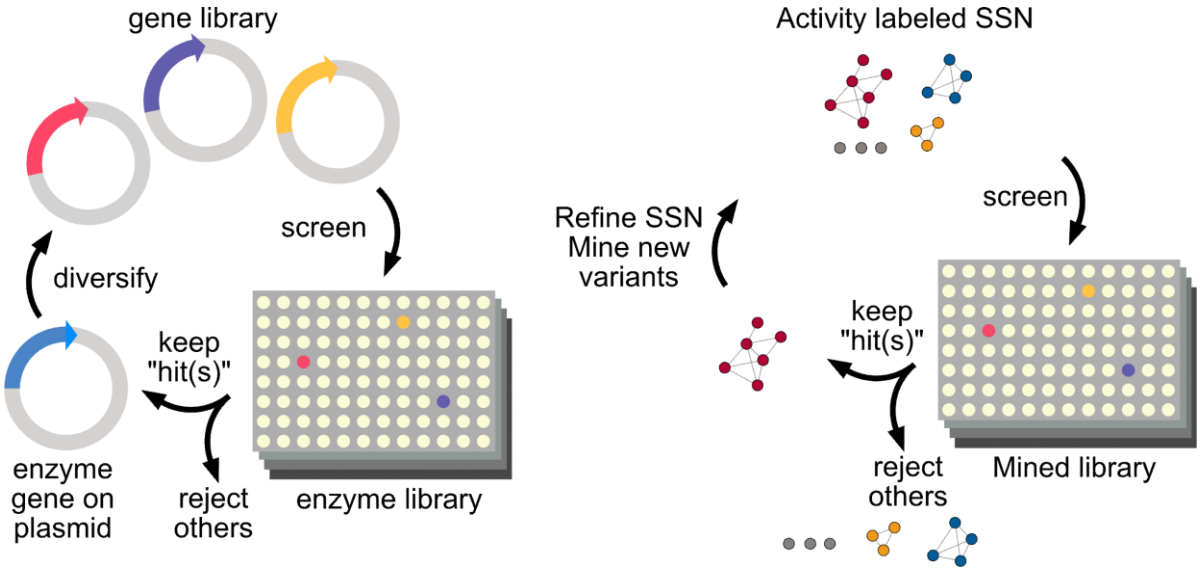


SOURCE: ACS Cent. Sci. 2019, 5, 1844−1856

2.2.4 Follow-up genome mining of active gene clusters

From the high-throughput and preparative analysis outlined above, we identified two halogenase clusters in the level 2 SSN (**Figure 2.2B**) that were particularly enriched in active halogenases. These clusters were most exemplified by halogenases 1-F11 and 2-C01, which were both highly active enzymes with favorable expression and stability profiles. These enzymes also halogenated large, structurally complex substrates, several of which had previously required directed evolution to halogenate with RebH variants. Encouraged by the activity of these enzymes, we decided to further mine these regions of the SSN by having additional genes synthesized (**Figure 2.6**). The process for gene selection largely mirrored the initial process, except only the sub networks containing 2-C01 and 1-F11 were analyzed by CD-HIT to suggest genes for synthesis.

Figure 2. 6: Iterative genome mining for identification of new activity

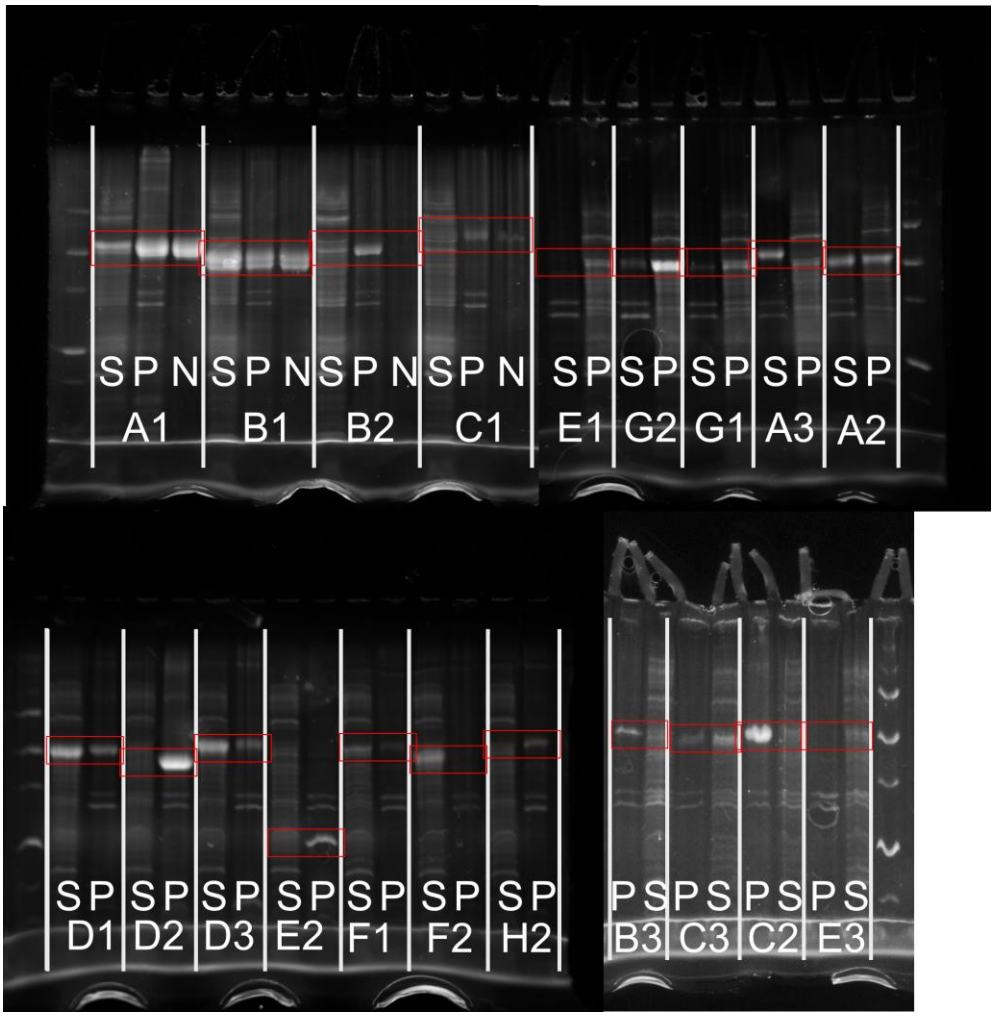


Comparison of traditional directed evolution and our approach to iterative genome mining. A) Directed evolution can be used to improve existing activity B) Iterative genome mining can be used to identify clusters enriched in active enzymes which may exhibit new activity.

Once the genes had been selected and synthesized, the obtained gene blocks were then ligated into pET28 expression vectors via Gibson assembly and used to transform *E. coli* for expression. SDS page gel analysis of the second-round genome mined enzymes showed that of the 21 genes obtained, 12 led to soluble protein under typical halogenase expression conditions (For gels, see **Figure 2.7**, for a summary, see **Table 2.1**). A set of probe substrates was chosen for screening purified enzymes, and several notable examples of activity were observed. Of particular interest was the identification of the halocyclase activity using enzymes R2A1 and R2F2, the latter of which also generated two distinct products matching the m/z of the monobrominated material, perhaps indicating regiodivergent halogenation (**Figure 2.8**). More generally, of the 12 proteins screened for halogenation activity obtained in the iterative mining effort, each enzyme was found to be active on at least one substrate, demonstrating a much higher rate of active enzyme discovery from this cluster than was observed in the initial study. As with the parent enzymes 1-F11 and 2-C01 used to identify D3, no chlorinase activity was observed. During the purification of these proteins for bioconversions, SDS PAGE analysis indicated that

several proteins showed significant degradation had occurred over the course of the purification process, likely resulting from proteolysis (**Figure 2.9**). While many techniques have been developed for slowing or ablating protein proteolysis,[23] activity profiling of the stable FDHs was prioritized to validate the iterative genome mining approach.

Figure 2. 7: SDS Page gels



Gels of halogenases from the iterative genome mining effort. P = Pellet or insoluble fraction, S = Soluble fraction of lysate, N = Purified fraction after Ni-NTA.

Table 2. 1: Information for genome mined enzymes from follow up study

Well Location	Name	Soluble	Organism	Notes
A1	A0A0H4PDU7_9BAC	yes	Cyclobacterium amurskyense	related to an organism that produces halogenated macrocycles
A2	A0A0S1YN77_9ALT	yes	Alteromonas stellipolaris LMG 21856	arctic organism
A3	K6YJ88_9ALTE	no	Paraglaciecola arctica	arctic organism
B1	A0A0U5LQG0_STRR	yes	Streptomyces reticuli	associated with netted scab
B2	A0A0T0PYD6_9SPH	no	Sphingomonas sp. Leaf412	unable to find data
B3	K6ZH11_9ALTE	trace	Glaciecola pallidula	arctic organism
C1	K7YTV9_BDEBC	toxic protein?	Bdellovibrio bacteriovorus	parasitic organism
C2	A0A0U4V0Z7_9GAM	no	Rheinheimera sp. F8	organism produces a blue-pigment labeled glaukothalin
C3	Q21DG9_SACD2	yes	Saccharophagus degradans	organism is known to degrade cellulose
D1	A0A010PBW3_9ALT	yes	Alteromonas sp. ALT199	organism produces 2-n-pentyl-4-quinolinol
D2	A0A0W1G664_9SPH	no	Sphingopyxis sp. H071	organism degrades ?-hexachlorocyclohexane
D3	Q21N77_SACD2	yes	Saccharophagus degradans	organism is known to degrade cellulose
E1	A0A099K8J3_COLP	unclear	Colwellia psychrerythraea	arctic organism
E2	A0A0X3U5P9_9GAM	yes	Microbulbifer flavimaris	comes from 'modestly halophilic' organism
E3	T1W5E6_9ZZZZ	unclear	uncultured organism	unable to find data
F1	A0A099K9Z7_COLP	yes	Colwellia psychrerythraea (Vibrio psychroerythus)	arctic organism
F2	A0A101M6D7_9GAM	yes	Rheinheimera sp. EpRS3	organism produces a blue-pigment labeled glaukothalin
G1	A0A0G9MRG1_9SPH	yes	Aurantiacibacter gangjinensis	basically uncharacterized organism
G2	I3I495_9GAMM	yes	Cellvibrio sp. BR	xylanase producing organism?
H1	A0A0P9GE54_9GAM	not expressed	Pseudoalteromonas sp. P1-9	unable to find data
H2	K6XK41_9ALTE	largely no	Paraglaciecola arctica	arctic organism

Figure 2. 8: Halocyclization with substrate 4-phenylpent-4-enoic acid

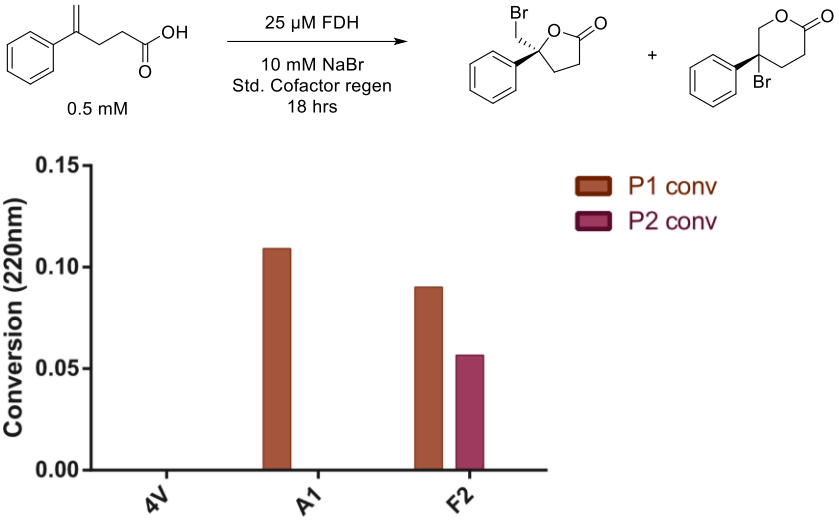
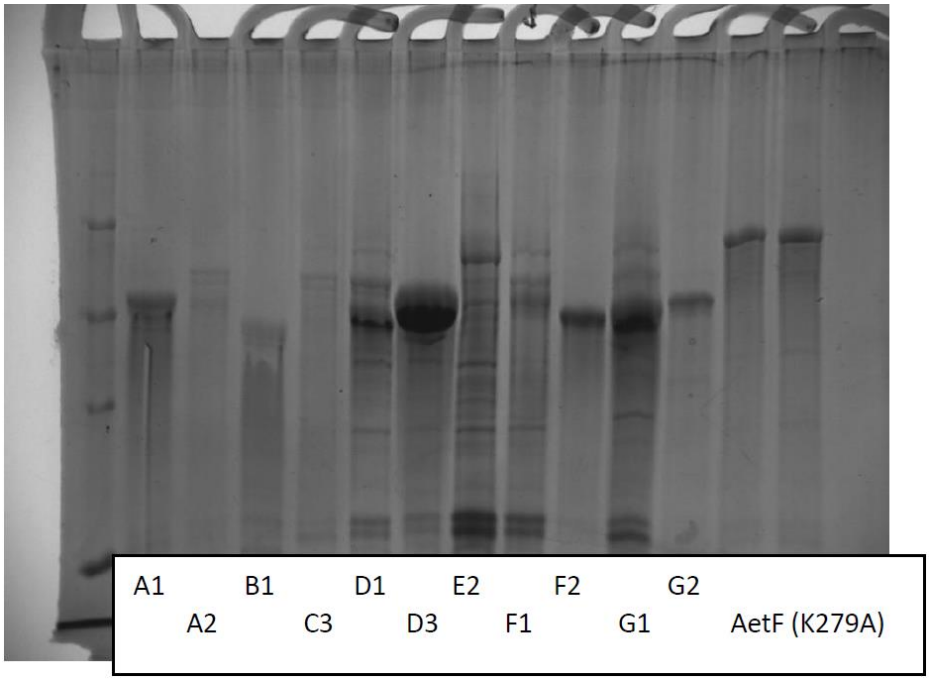


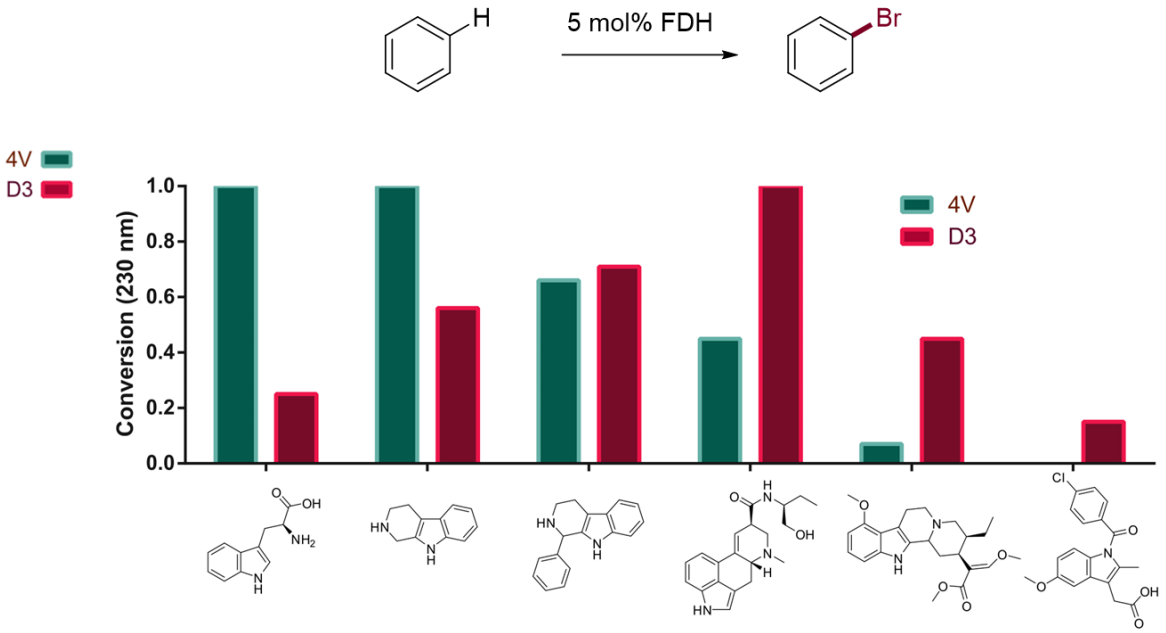
Figure 2. 9: SDS page gels of purified enzymes showing degradation



Despite the issues with stability exhibited by some enzymes in this follow up set of FDHs, we began to explore some of the proteins that could be purified in reasonable purity. One standout from this subset was the enzyme located at well D3 (Uniprot ID: Q21N77_SACD2), which halogenated several substrates from the initial probe substrate panel. Among this substrate set were three indole containing substrates of varying steric bulk: L-tryptophan, tryptoline, and 1-phenylcarbazole. Compared to RebH

variant 4-V, evolved for the halogenation of bulky biologically active substrates, D3 was superior at halogenating the largest substrate 1-phenylcarbazole, but had the lowest activity on L-tryptophan. The observed preference for bromination of sterically demanding molecules led us to believe that perhaps D3 favored the halogenation of large, indole containing substrates.

Figure 2. 10: R2D3 and 4V substrate scope

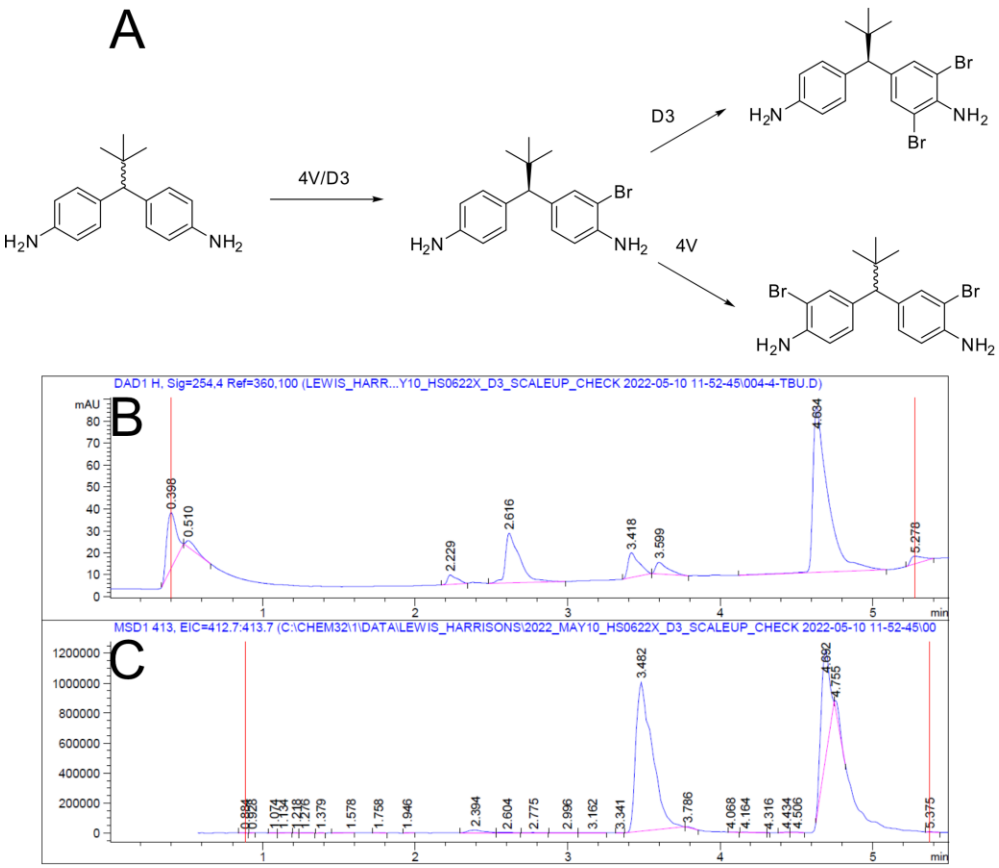


Substrate structures are shown below the bar graphs corresponding to conversion. Only bromination was screened.

To test this hypothesis, we assembled a panel of 11 large or structurally complex, biologically active indole compounds and compared the activity of D3 and 4V (For the full substrate set, see **Table 2.2**, a selection can be seen in **Figure 2.10**). Of the 11 substrates examined, D3 had activity on 7. Among this set was indomethacin, a compound with a highly-substituted pyrrole ring. While the evolved halogenase 4V had no detectable activity, D3 generated a single product in 15% yield. Of the remaining 6 substrates, D3 gave higher conversion for 4. Only 3-((1H-indol-3-yl)methyl)-1-methylpiperazine-2,5-dione and bopindolol were brominated by 4V in higher yield than D3, perhaps due to the relatively exposed native C7 position in those compounds. In addition to the 11 compounds evaluated in this screen, researchers from Prof. David Sherman's group at the University of Michigan became interested

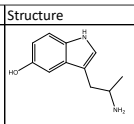
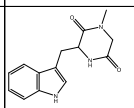
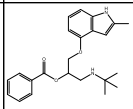
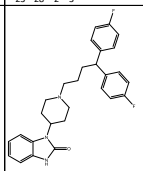
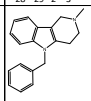
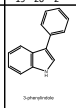
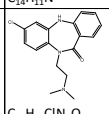
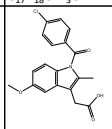
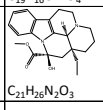
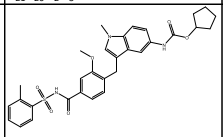
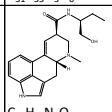
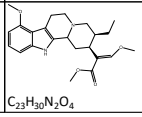
in the pain-relieving effects of the opioid agonist mitragynine[24] and sought to identify a halogenase capable of biohalogenation of this compound. While RebH variant 4V successfully chlorinated this compound to a single product in 25% conversion by UHPLC-MS, it exhibited poor activity and selectivity for bromination. In contrast, halogenase D3 gave a single brominated product in 45% conversion under the same reaction conditions. Upon increasing enzyme loading to 5% and scaling the reaction up to 30 mL, we were successfully able to isolate and verify the formation of 12-Br-Mitragynine in high enough yields that material could be shipped to our collaborators for activity assays. Additonally, when R2D3 was evaluated for the bromination of a prochiral methylenedianiline compound previously found to be desymmeterized with high e.r. by RebH variant 4V,[25] it was found that R2D3 formed a unique dibrominated product (**Figure 2.11 A**) with identical m/z to the racemic dibrominated material, leading us to suspect that R2D3 was exhibiting sequential bromination of the same aniline ring (**Figure 2.11 B & C**). In total, these results, though preliminary, demonstrate that the newly discovered enzyme R2D3 exhibits catalytic activity extending beyond what was previously accessible using previously described FDHs.

Figure 2. 11: Unique dibrominated material formed by R2D3 for a methylenedianiline



A) Putative reaction scheme for the observed new dibrominated product B) UV trace of the reaction measured at 254 nm C) EIC for the m/z of the major dibrominated ion (m/z = 413) showing two distinct dibrominated products.

Table 2. 2: Full substrate scope of D3 and 4V

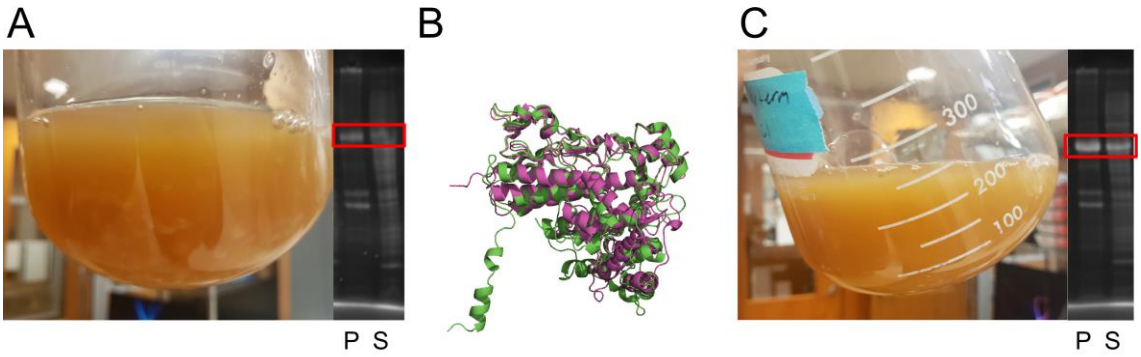
Compound name	Structure	4V Activity	D3 Activity	Notes
α-methylserotonin		5.60E+05	1.40E+06	starting material eluted in buffer front
3-((1H-indol-3-yl)methyl)-1-methylpiperazine-2,5-dione	 $C_{14}H_{15}N_3O_2$	24%	11%	
bopindolol	 $C_{23}H_{28}N_2O_3$	100%	54%	
pimozide	 $C_{28}H_{29}F_2N_3O$	0%	0%	
mebhydrolin	 $C_{19}H_{20}N_2$	24%	68%	
3-phenylindole	 $C_{14}H_{11}N$	40%	100%	
7-chloro-10-(2-(dimethylamino)ethyl)-5,10-dihydro-11H-dibenzo[b,e][1,4]diazepin-11-one	 $C_{17}H_{18}ClN_3O$	0%	0%	
indomethacin	 $C_{19}H_{16}ClNO_4$	0%	15%	
vincamine	 $C_{21}H_{26}N_2O_3$	0%	0%	
zafirlukast	 $C_{31}H_{33}N_3O_6S$	0%	0%	
methylergonovine	 $C_{20}H_{25}N_3O_2$	45%	100%	
mitragynine	 $C_{23}H_{30}N_2O_4$	7%	45%	

Conversion data for the substrate scope evaluated for 4V and R2D3. Conversion compares starting material and product at 230 nm.

While evaluating the expression of variants obtained in this follow up study, it was noted that expression of the enzyme R2C1 (Uniprot ID: K7YTV9_BDEBC) resulted in a reproducible phenotype of *E. coli* cell death distinct in the way the *E. coli* cells began to form clusters of dead cells. Despite the resulting low O.D. cultures and poorly separating pellet, this enzyme was obtained in low yield after standard IMAC purification, and it was determined to be an active halogenase, generating the halocyclized product of an activated styrene with pendant carboxylic acid nucleophile and halogenating phenylpiperazine with low conversion. Analysis of the primary sequence found that this protein was significantly longer than RebH closely related halogenases from the same subnetwork. Most of this additional length came from a ~20 amino acid sequence at the N-terminus of the protein and a ~10 residue sequence at the C-terminus. To determine if these additional residues could be impacting expression of the enzyme and the observed *E. coli* cell death, an Alphafold model of R2C1 was generated and overlayed with a similarly constructed model of the closely related halogenase R2D3. Evident from this model was the addition of an N-terminal helix which we began to suspect was a signal peptide. While this enzyme like other genome mined FDHs has not been characterized within its native biosynthetic gene cluster, Uniprot designates this enzyme as originating in *Bdellovibrio bacteriovorus*, a parasitic organism.[26] *B. bacteriovorus* functions by first binding to its bacterial prey and enters the periplasm of the host through the outer membrane. Once localized within the periplasm of the organism, the *B. bacteriovorus* then modifies the peptidoglycan linkages of the bacterial membrane such that the host organism accommodates growth of *B. bacteriovorus* in a bdelloplast, though much of the mechanism through which the bacteriovorus enters and manipulates the bacterial host is still unknown. To both validate the Alphafold model of R2C1 and boost expression yields of the active halogenase, we removed the putative N-terminal portion of the protein via Quikchange PCR and substituted it for an N-terminal 6xHis tag embedded in a flexible linker. To our delight, not only did this result in a substantial improvement to *E. coli* cell health but the overall expression and ratio of soluble/insoluble expression

for R2C1 improved (**Figure 2.12**). While no activity data was obtained for the truncated R2C1 enzyme, the improved expression provides a platform for evaluating this putative FDH, and perhaps more importantly this validates the Alphafold model guided approach for investigating these newly mined enzymes.

Figure 2. 12: Expression of R2C1



A) Image of *E. coli* culture expression holo-R2C1 and resulting SDS page gel showing improperly separated soluble fraction (S) and most of the enzyme in the pellet (P) after lysis. B) Alphafold model of R2C1 overlaid with the Alphafold model of the closely related enzyme R2D3, showing the existence of an additional N-terminal tag on R2-C1. C) Image of E. coli cell culture of ΔN-R2C1 and resulting SDS page gel showing improved soluble protein expression and loss of clumping phenotype.

2.3 Conclusions

Through this work we characterized 39 new active halogenase genes from our initial screen and an additional 12 from our iterative second study. These enzymes were identified from diverse organisms and functionalized a variety of substrates. We also gained insight into the halide selectivity of FDHs beyond the scope of those identified based on natural product studies, identified several FDHs with complementary regioselectivity on biologically active compounds, and demonstrated how the clustering of enzymes in an SSN can be predictive of the activity of those enzymes on probe substrates. My contribution to this project was essential for the identification of key substrate-enzyme pairs for regiodivergent halogenation and the determination of regioselectivity for functionalization of biologically active compounds. Furthermore, my latest contributions to this project have demonstrated how the enzymes we initially characterized could be used to inform a subsequent genome mining study

currently underway to further explore the FDH sequence space. By focusing on subnetworks of the SSN that were previously unexplored but contained functional enzymes demonstrating promising catalytic properties, we found that the rate of active enzyme identification was improved substantially compared to our first study. Although preliminary, early results from this iterative approach show that both improved and novel activity can be rapidly obtained. Furthermore, by coupling the activity information from related sequences to structural models generated from Alphafold using primary sequence alone we found that successful protein engineering efforts to improve enzyme properties could be undertaken without the need for laborious structure determination. This ability to reasonably predict enzyme function and structure from sequence alone promises to be widely exploited for biocatalyst identification and optimization in the future. More narrowly, the groundwork we laid in this in-depth study of these halogenase sequences will certainly provide future researchers with the ability to rapidly identify more functional members of this enzyme class.

2.4 Experimental

2.4.1 General experimental procedures

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

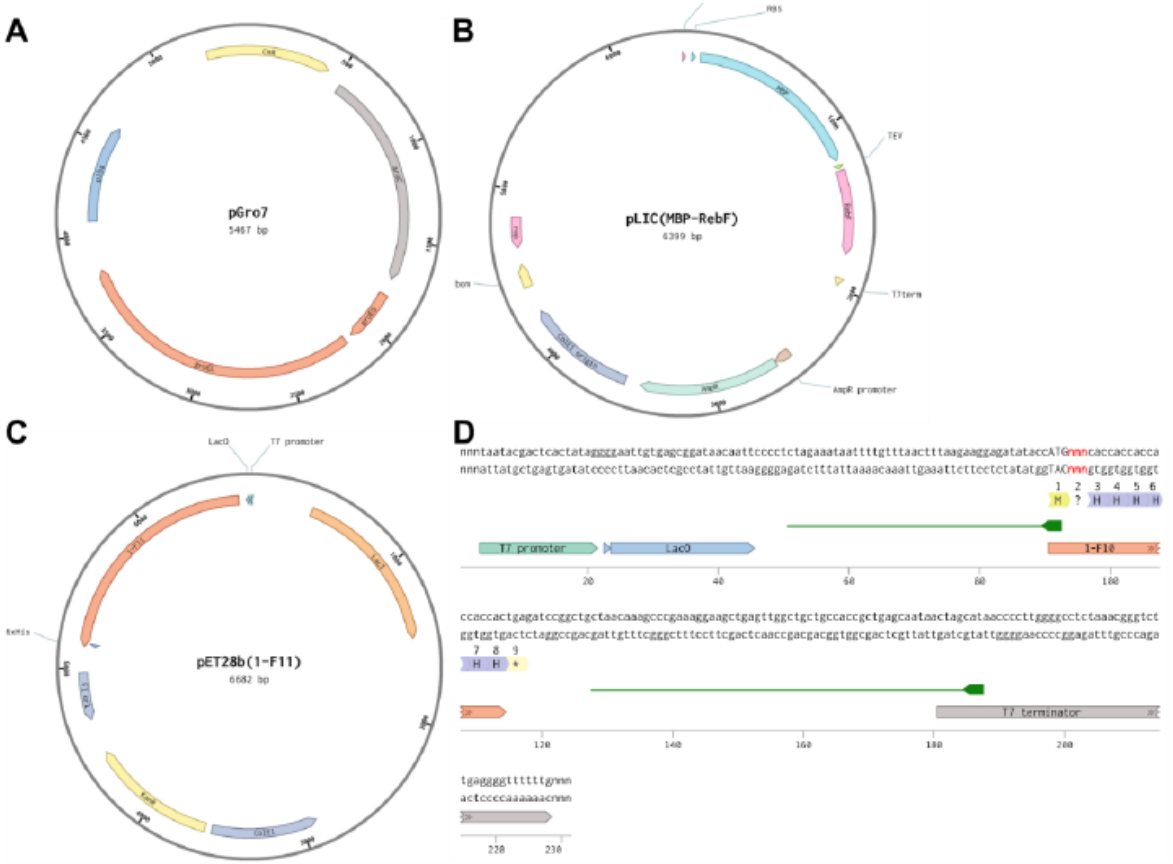
NAD, FAD, and antibiotics were purchased from Chem-Impex International Inc. (Wood Dale, IL). Antibiotics were prepared as 1000x stock solutions: 1000x chloramphenicol was prepared at 25 mg/mL in EtOH, and 1000x kanamycin was prepared at 50 mg/mL. Substrates were purchased from Sigma-

Aldrich, Toronto Research Chemicals, Chem-Impex, or Santa Cruz Biotechnologies. Tryptamine was recrystallized from hot 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). DH5α and BL21(DE3) E. coli were purchased from Invitrogen (Carlsbad, CA). Taq DNA polymerase and Phusion HF polymerase were purchased from New England Biolabs (Ipswitch, MA). Luria broth (LB) and Terrific broth (TB) media were purchased from Research Products International (Mt. Prospect, IL). Qiagen Miniprep Kits were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. Protein ladder (Blue Prestained Protein Standard, Broad Range (11-190 kDa); product number P7706) was purchased from New England Biolabs (Ipswitch, MA).

FDH genes were synthesized by the Joint Genome Institute (Department of Energy, USA), integrated into pET28b with a C-terminal 6x His-tag, and transformed into E. coli Top10. For each fungal 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 2. 13: Example Plasmid Maps Used In This Study



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. SOURCE: ACS Cent. Sci. 2019, 5, 1844−1856

Stock solutions of 10 mM NAD and 10 mM FAD were prepared in 25 mM HEPES pH 7.4 buffer (reaction buffer) and stored at -20 °C until use. Stock solutions of 1.5 M NaCl, 1.5 M NaBr, and 1 M glucose were prepared in reaction buffer and stored at 4 °C until use. Stock solutions of substrate were prepared at 30 mM in either water or DMSO; for the high-throughput screen, substrate stocks were then diluted to 1.67 mM in reaction buffer, manually arrayed into 96-well plates as described later and stored at -20 °C until use. RebF was expressed in E. coli BL21(DE3) as an MBP fusion from pLIC-MBP as described previously[27] 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.

2.4.2 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.

2.4.3 FDH Expression

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

Transformation

Chemically-competent BL21(DE3) *E. coli* cells containing the pGro7 chaperone plasmid were transformed with pET28b containing FDH insert. Aliquots of competent cells were transferred to 5 mL polypropylene culture tubes on ice, to which approximately 10 ng of plasmid was added. Competent cells were incubated with plasmid on ice for 30 min, and the cells were heat shocked at 42 °C in water

bath for 45 seconds. The culture tubes were afterward incubated on ice for 2 minutes, then 350 µL SOC medium was added. The culture tubes were transferred to an incubator at 37 °C shaking at 250 rpm for 1 hr to recover. After recovery, 100 µL cells were added to agar plates (with 50 µg/mL kanamycin and 25 µg/mL chloramphenicol) and spread using 3 mm glass beads. After drying, the agar plates were transferred to a 37 °C incubator and grown overnight. Single colonies were picked and grown in 300 µL LB in a 1 mL deep well 96-well plate with kanamycin and chloramphenicol overnight. To each well of this 96-well plate was then added 200 µL autoclaved 50% glycerol, and this glycerol stock plate was kept at -80 °C.

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 (≈15 min) protein expression was induced with 2 mg/mL L-arabinose and 10 µM IPTG. Protein expression proceeded for 20 hr, at which point the cells were pelleted by centrifugation at 3600 rpm, 4 °C for 15 min. The supernatant was discarded, and the cell pellets were resuspended in 200 µl lysis buffer (0.75 mg/mL lysozyme, 25 mM HEPES, pH 7.4). Cells were incubated in lysis buffer at 37 °C, 250 rpm for 30 min. After lysis, the suspensions were frozen by immersing the 24-well plates in liquid N_2, thawed at room temperature for 15 min, then transferred to a warm water bath. Once thawing was complete, 20 µL DNAse buffer (1 mg/mL DNAse, 25 mM HEPES, pH

7.4) was added, and the 24-well plates were incubated at 37 °C, 250 rpm for 15 min. The insoluble fraction of the cell lysate was pelleted by centrifugation at 3600 rpm, 4 °C for 15 min. The supernatant was isolated, and the insoluble fraction of the cell lysate was washed by resuspending in 300 µL 25 mM HEPES, pH 7.4, then centrifugation at 3600 rpm, 4 °C for 15 min. This supernatant wash was discarded.

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 ≈0.6-0.8, at which point the incubator was cooled to 25 °C. After the cultures were cooled sufficiently (≈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.

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 ≈0.6-0.8, at 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.

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 OD600 ≈0.6-0.8, at which point the incubator was cooled to 15 °C. Once the liquid cultures were cool (about 15 min), protein expression was induced with 2 mg/mL L-arabinose and 100 µM IPTG, and the expression cultures were incubated for 20 hr. Once protein expression was complete, the expression cultures were transferred to 750 mL centrifuge bottles and centrifuged at 3600 rpm, 4 °C for 15 min. The supernatant media was discarded, and the cell pellets were resuspended in 30 mL 25 mM HEPES, pH 7.4. Cell suspensions were sonicated on ice using a QSonica S-4000 with a 0.5" horn at 40W using 1 min on/1 min off cycles for 5 min total cycle time. Cell lysates were clarified by centrifuging at 15,000 rpm in a high-speed fixed-angle rotor for 40 min at 4 °C. The soluble fraction of the lysate was decanted into a new 50 mL centrifuge tube, then transferred to 10 mL polypropylene frit-bottomed spin columns capped at the bottom and containing 3 mL Ni-NTA resin pre-equilibrated with equilibration buffer (20 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4). The columns were capped on top, inverted a few times to mix evenly, and transferred back to the centrifuge tubes which were then capped. Protein was bound to resin by gentle mechanical inversion of these centrifuge tubes for 1 hr at 4 °C. After the binding step, the Ni-NTA suspensions were transferred back to uncapped spin columns and allowed to drain by gravity into a waste basin. 5 mL of wash buffer (20 mM phosphate, 300 mM NaCl, 25 mM imidazole, pH 7.4) was added to the columns, which were allowed to drain by gravity into a waste basin. The spin columns were nested within new 50 mL centrifuge tubes, and 5 mL elution buffer (20 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4) was added and allowed to drain into the centrifuge tubes by gravity. The eluted protein solutions were transferred to 15 mL Amicon spin filters Ultra 30K MWCO spin filters

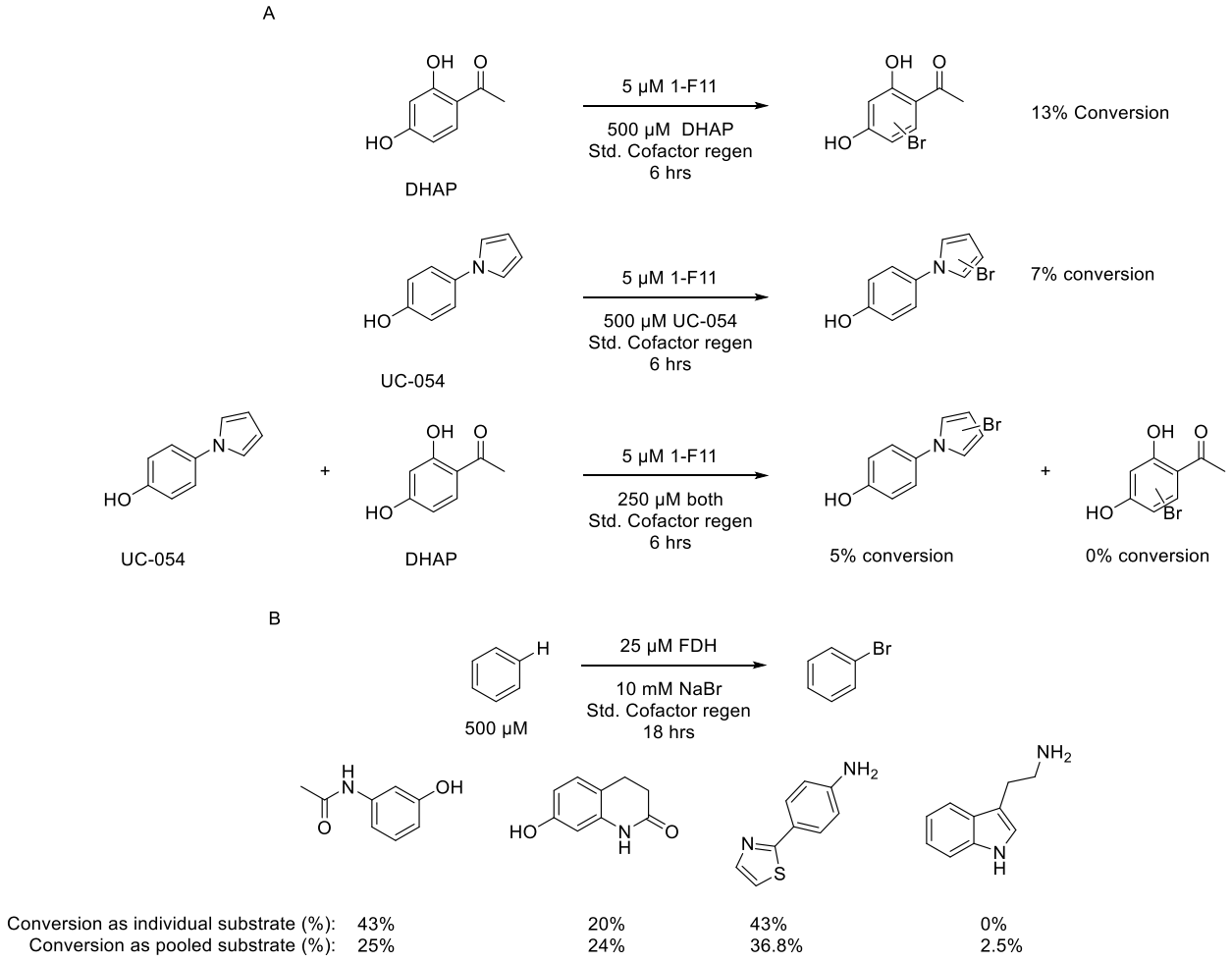
and either buffer exchanged as described previously or concentrated prior to overnight dialysis at 4 °C against 25 mM HEPES, pH 7.4.

SDS-PAGE

SDS-PAGE gels, 1 mm thick, were cast with 10% or 12% acrylamide using standard procedures, but including 0.5% trichloroethanol in the separating portion of the gel to enable fluorescent detection of Trp-containing proteins.[28] 8 µL of the soluble or insoluble fraction of the cell lysate, 16 µL water, and 12 µL 4x SDS-PAGE loading buffer (200 mM Tris-Cl (pH 6.8), 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol) were combined in Eppendorf tubes, which were heated at 95 °C in a dry incubator for 10 min. Samples were cooled to room temperature, then centrifuged for a few seconds in a mini centrifuge to collect liquid at the bottom. SDS-PAGE gels were loaded into the gel chamber, and 8 µL of each sample was loaded onto the gels. SDS-PAGE gels were run at 130 V for 90 minutes, after which they were removed from the chambers and glass slides, washed three times with deionized water, and placed onto the UV stage of a gel imager. Gels were irradiated at 302 nm for 2 min to develop the photochemically modified, fluorescent Trp residues, and images were taken. Serendipitously, the chaperone groEL, one of the chaperones expressed by pGro7 that unfortunately has a molecular weight close to that of many halogenases, has zero Trp residues, and is not stained by this procedure, allowing for easy identification of the expressed FDHs. Gels were then stained using Coomassie Brilliant Blue R-250 staining solution 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.

2.4.4 Bioconversions

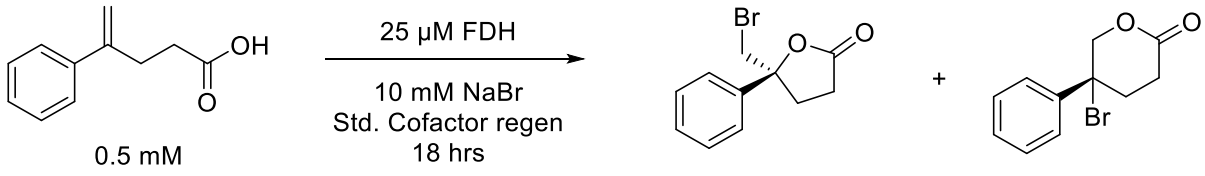
Analytical bioconversions for pooled substrate screening



1-F11 was expressed and purified according to the general procedure. Stock solutions of 50 µM purified FDH were prepared in storage buffer (HEPES, pH 7.4 25 mM, 10% glycerol). An FDH stock was prepared by diluting 7.5 µL of 50 µM 1-F11 with 42.5 µL of reaction buffer (HEPES, pH 7.4 25 mM) in a microtiter plate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 20 mM D-glucose, 10 mM NaBr, and 2.5 mM of substrate from 30 mM stock solutions in DMSO, were prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the FDH stock in the microtiter plate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the

enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours in the case of substrate pooling after bioconversions (**Scheme 2.1B**), and 6 hours in the case of substrate pooling in the same bioconversion (**Scheme 2.1A**). Reactions were quenched by the addition of 1 volume of methanol when the reactions were finished. For pooling substrates before analysis, 20 µL of the 1:1 methanol:water mixtures were combined into one well of a filter plate and filtered via centrifugation using a 0.2 µm filter plate. Samples were analyzed via UHPLC-MS via the MISER method described in the general section. Conversion was calculated as the AUC of product over AUC of starting material and product with integrations measured using the EIC of the major product ion(s) and the starting material ion.

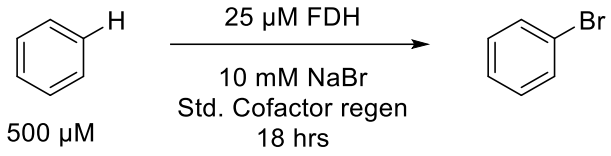
Analytical bioconversions for halocyclization of 4-phenylpent-4-enoic acid



4V, R2F2 and R2A1 were expressed and purified according to the general procedure. Stock solutions of 100 µM purified FDH were prepared in storage buffer (HEPES, pH 7.4 25 mM, 10% glycerol). An FDH stock was prepared by diluting 18.75 µL of 100 µM FDH with 31.25 µL of reaction buffer (HEPES, pH 7.4 25 mM, final concentration 25 µM) in a microtiter plate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 20 mM D-glucose, 10 mM NaBr, and 2.5 mM of substrate from 30 mM stock solutions in DMSO, were prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the FDH stock in the microtiter plate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. Reactions were quenched by

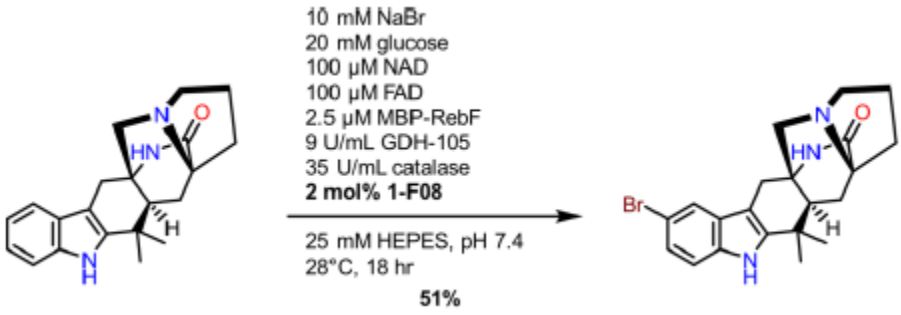
the addition of 1 volume of methanol when the reactions were finished. Protein was precipitated by centrifugation at 3600 RPM in a Sorvall centrifuge for 15 minutes. 50 µL of this sample was then added to a filter plate with 150 µL of water and samples were filtered via centrifugation at 3600 RPM in a Sorvall centrifuge for 15 minutes. Samples were analyzed via UHPLC-MS. Conversion was calculated as the AUC of product over AUC of starting material and product with integrations measured using the TIC of the product ions and the starting material ion. The relevant results can be seen in **Figure 2.8**.

Analytical bioconversions with 4V and R2D3 for substrate screening.



R2D3 and 4V were expressed and purified according to the general procedure. Stock solutions of 100 µM purified FDH were prepared in storage buffer. An FDH stock solution was prepared by diluting 18.75 µL of 100 µM FDH with 31.25 µL of reaction buffer, and 50 µL of this stock solution was added to a microtiter plate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 2.5 mM of substrate the various substrates, were prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the clarified lysate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. After 16 hours, the reactions were quenched and processed as previously described above. Aliquots were taken and analyzed and the results can be seen in **Table 2.2.** Reactions were analyzed by UHPLC-MS with integrations taken from the 230 nm chromatogram.

Premalbrancheamide bromination with 1-F08

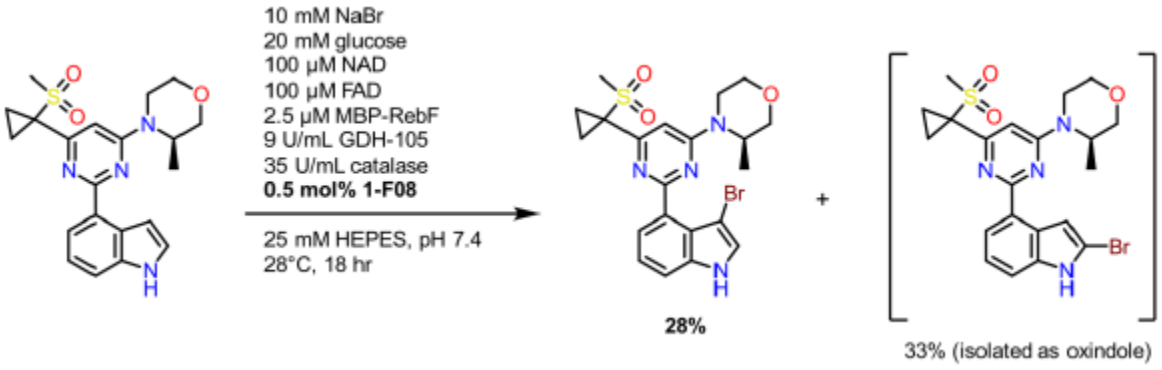


Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: premalbrancheamide (30 mM DMSO stock, 500 µM final concentration 12 µmol, 4.0 mg), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-F08 (2 mol%, 10 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and finally GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated in a VWR 1585 incubator and agitated at 250 rpm at 28 °C. After 18 hours, the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over MgSO4, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (Supelco Discovery C18, 3.0mL/min, 20%B 0-5 min, 20-35%B 5.01-23 min, 35%B 23.01-24 min, 95%B 24.01-29

min, 20%B 28.01-30 min). The fractions found to contain product via LC-MS were collected and concentrated by rotary evaporation to produce product in 51% yield.

^{1}H NMR: (600 MHz, Methylene Chloride-d_2) δ 8.03 (s, 1H), 7.49 (d, J = 1.6 Hz, 1H), 7.30 (d, J = 8.3 Hz, 1H), 7.19 (dd, J = 8.4, 1.7 Hz, 1H), 6.40 (s, 1H), 3.42 – 3.35 (m, 1H), 3.35 – 3.28 (m, 1H), 2.95 (d, J = 17.1 Hz, 1H), 2.88 (d, J = 17.1 Hz, 1H), 2.78 (d, J = 10.8 Hz, 1H), 2.65 – 2.59 (m, 1H), 2.51 (q, J = 9.0 Hz, 1H), 2.39 – 2.31 (m, 1H), 2.28 (dd, J = 10.0, 3.6 Hz, 1H), 2.01 (d, J = 4.5 Hz, 2H), 1.96 (dd, J = 13.7, 3.6 Hz, 1H), 1.65 – 1.55 (m, 1H), 1.29 (s, 3H), 1.19 (s, 3H). ^{13}C NMR: (126 MHz, DMSO-d6) δ 173.07, 143.63, 137.74, 126.28, 121.32, 119.74, 113.69, 113.45, 103.74, 64.72, 62.45, 55.00, 53.22, 46.09, 34.48, 31.54, 27.94, 27.51, 27.37, 24.17, 23.01. HRMS (ESI-MS): calc. for $[C_{21}H_{26}BrN_3O]^+$ ([M+H]$^+$): 414.1176 and 416.1157, found 414.1179 and 416.1159.

AZ20 bromination with 1-F08



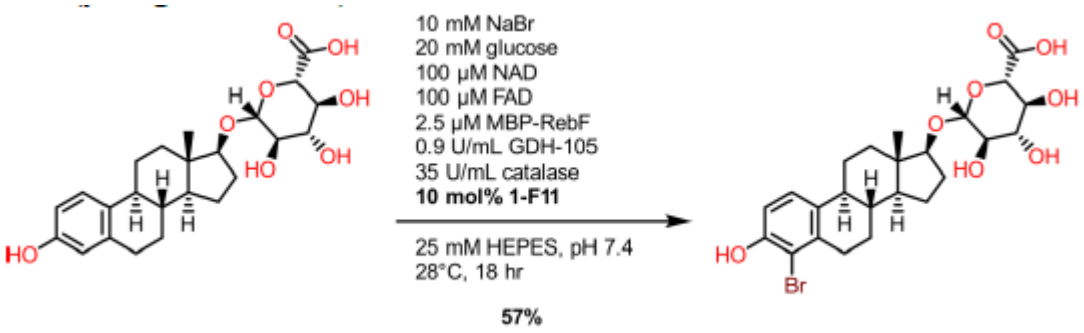
Ten 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: AZ20 (30 mM DMSO stock, 500 µM final concentration 15 µmol, 6.2 mg), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration), and 200 µL DMSO (8% cosolvent). 1-F08 (0.5 mol%, 2.5 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were

incubated in a VWR 1585 incubator and agitated at 250 rpm at 28 °C. After 18 hours, the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This solution was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (Supelco Discovery C18 0-5 min 32%B, 5.01-20 min 32-38%B, 20-25 min 38%B, 25.01-28 min 95%B, 28.01-30 min 32%B). Two non-starting material peaks were found in the semiprep HPLC chromatogram, and each was collected separately, and solvent removed by rotary evaporation. One fraction contained the 3-brominated AZ20 product, isolated in 28% yield, and the other contained an oxindole likely formed from hydrolysis of the 2-brominated AZ20 product, isolated in 33% yield.

3-Br-AZ20:
^{1}H NMR (500 MHz, Methylene Chloride-*d*2) δ 8.62 (s, 1H), 7.50 (dd, *J* = 5.8, 3.4 Hz, 1H), 7.32 – 7.31 (m, 1H), 7.29 (s, 1H), 7.29 – 7.28 (m, 1H), 6.88 (s, 1H), 4.51 – 4.41 (m, 1H), 4.11 (d, *J* = 13.5 Hz, 1H), 4.00 – 3.92 (m, 1H), 3.76 (d, *J* = 11.5 Hz, 1H), 3.71 (dd, *J* = 11.6, 3.2 Hz, 1H), 3.56 (td, *J* = 11.9, 3.1 Hz, 1H), 3.30 (td, *J* = 12.9, 4.0 Hz, 1H), 3.02 (s, 3H), 1.76 (dd, 7.9 Hz, 4.7 Hz, 2H), 1.57 (dd, 7.9 Hz, 4.7 Hz, 2H), 1.32 (d, *J* = 6.8 Hz, 3H). ^{13}C NMR**:** (126 MHz, DMSO-*d*6) δ 164.98, 161.57, 161.22, 136.71, 132.88, 127.33, 122.89, 121.82, 121.60, 113.22, 101.34, 88.20, 70.68, 66.56, 46.62, 46.44, 40.81, 13.92, 12.65. One peak at ~39ppm is covered by the DMSO-*d*6 peak. HRMS (ESI-MS): calc. for $[C_{21}H_{24}BrN_4O_3S]^+$ ($[M+H]^+$): 491.0747 and 493.0728, found 491.0745 and 493.0723.

Oxindole-AZ20:
^{1}H NMR: (400 MHz, Methylene Chloride-*d*2) δ 8.04 (d, *J* = 8.0 Hz, 1H), 7.64 (s, 1H), 7.32 (t, *J* = 7.9 Hz, 1H), 6.96 (d, *J* = 7.7 Hz, 1H), 6.84 (s, 1H), 4.52 – 4.42 (m, 1H), 4.10 – 4.00 (m, 2H), 3.97 (s, 2H), 3.82 (d, *J* = 11.6 Hz, 1H), 3.73 (dd, *J* = 11.6, 3.2 Hz, 1H), 3.59 (td, *J* = 11.9, 3.1 Hz, 1H), 3.34 (td, *J* = 12.8, 4.0 Hz, 1H), 3.03 (s, 3H), 1.85 – 1.78 (m, 2H), 1.33 (d, *J* = 6.8 Hz, 3H). Methylene hydrogens concealed under residual H2O peak. ^{13}C NMR**:** (126 MHz, DMSO-*d*6) δ 177.09, 163.18, 162.33, 161.85, 145.07, 134.44, 127.99, 125.93, 121.99, 111.26, 102.16, 70.67, 66.47, 46.95, 46.56, 40.58, 38.83, 13.89, 12.85. HRMS (ESI-MS): calc. for $[C_{21}H_{25}N_4O_4S]^+$ ($[M+H]^+$): 429.1591, found: 429.1588.

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

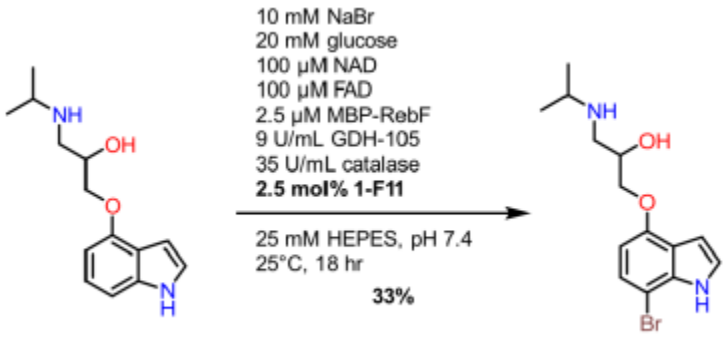


Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: β-Estradiol 17-(β-D-glucuronide) sodium salt (30 mM DMSO stock, 500 µM final concentration 12 µmol, 5.6 mg sodium salt), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-F11 (10 mol%, 50 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (0.9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated in a VWR 1585 incubator and agitated at 250 rpm and 28 °C. After 18 hours, the reactions were collected into an Erlenmeyer flask and acidified to pH 3 using 10% citric acid. The compound was then extracted into 10 mL of a 3:1 $CHCl_3$/^iPrOH mixture five times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC. (Supelco Discovery C18; 3.0mL/min, 35%B 0-5 min, 35-45%B 5.01-23 min, 45%B 23.01- 25 min, 95%B 25.01-32 min, 35%B

32.01-35 min). Purified fractions were collected and solvent was removed by rotary evaporation to produce product in 57% yield.

^{1}H NMR: (400 MHz, Methanol-*d*4) δ 7.14 (d, *J* = 8.3 Hz, 1H), 6.72 (d, *J* = 8.4 Hz, 1H), 4.59 (s, 2H), 4.42 (d, *J* = 7.8 Hz, 1H), 3.91 (d, *J* = 8.8 Hz, 1H), 3.58 (s, 1H), 3.24 (d, *J* = 8.8 Hz, 1H), 2.93 (s, 1H), 2.32 (s, 1H), 2.26 – 2.09 (m, 2H), 1.99 (s, 1H), 1.71 (s, 2H), 1.46 (s, 2H), 1.33 (d, *J* = 13.8 Hz, 3H), 0.90 (s, 3H). ^{13}C NMR: (126 MHz, DMSO-d6) δ 172.68, 152.54, 136.65, 132.74, 125.35, 113.70, 113.00, 103.17, 86.65, 77.48, 74.26, 74.15, 72.72, 49.82, 43.96, 43.27, 37.83, 37.50, 31.23, 28.80, 27.34, 26.60, 22.87, 11.79. HRMS (ESI-MS): calc. for $[C_{24}H_{30}BrO_8]^-$ $[M-H]^-$: 525.1130 and 527.1113, found 525.1122 and 527.1101.

Pindolol bromination with 1-F11: 7-bromination

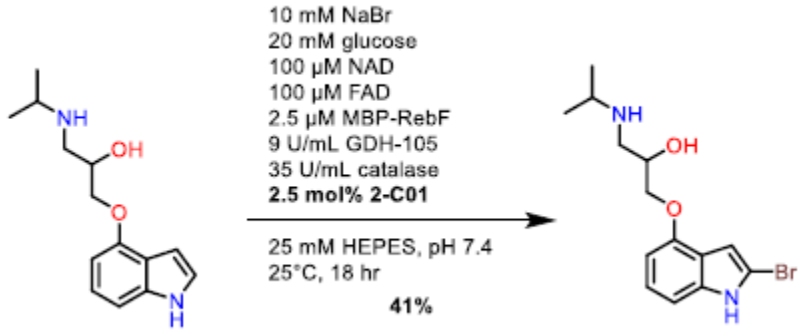


Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 5 mL of buffered solution was as follows: Pindolol (30 mM from DMSO stock, 1 mM final concentration, 40 µmol, 9.9 mg), and solutions in reaction buffer of NaBr (10 eq., final concentration 10 mM), glucose (20 eq., final concentration 20 mM), NAD (0.1 eq., 100 µM), FAD (0.1 eq., 100 µM final concentration). 1-F11 (2.5 mol%, 25 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated at room temperature (25 °C) and agitated at 250 rpm in a VWR 1585 incubator. After 18 hours the reactions were collected into an Erlenmeyer flask and

acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated onto Celite. This material was loaded into a biotage samplet and purified via reverse-phase chromatography (Biotage) as previously reported.[4] The fractions were collected and concentrated, then basified with 1M NaOH to precipitate trifluoroacetate, which was removed by filtration. Product distribution: 1:15.8 2-brominated:7-brominated product. Solvent from the filtrate was removed by rotary evaporation to produce product in 33% isolated yield.

^{1}H NMR: (400 MHz, Methanol-*d*4) δ 7.21 (d, *J* = 3.1 Hz, 1H), 7.17 (d, *J* = 8.3 Hz, 1H), 6.65 (d, *J* = 3.2 Hz, 1H), 6.49 (d, *J* = 8.3 Hz, 1H), 4.31 (d, *J* = 7.3 Hz, 1H), 4.20 (dd, *J* = 10.0, 5.0 Hz, 1H), 4.12 (dd, *J* = 10.0, 5.8 Hz, 1H), 3.51 – 3.41 (m, 1H), 3.35 (dd, *J* = 12.7, 3.0 Hz, 0H), 3.20 (dd, *J* = 12.7, 9.5 Hz, 1H), 1.38 (d, *J* = 6.4 Hz, 6H). ^{13}C NMR**:** (126 MHz, Methanol-*d*4) δ 151.78, 135.72, 123.67, 120.09, 101.33, 99.63, 96.10, 70.62, 67.98, 49.30, 48.84, 23.84, 20.54. HRMS (ESI-MS): calc. for $[C_{14}H_{20}BrN_2O_2]^+$ ($[M+H]^+$): 327.0703 and 329.0683, found 327.0708 and 329.0687.

Pindolol bromination with 2-C01: 2-bromination

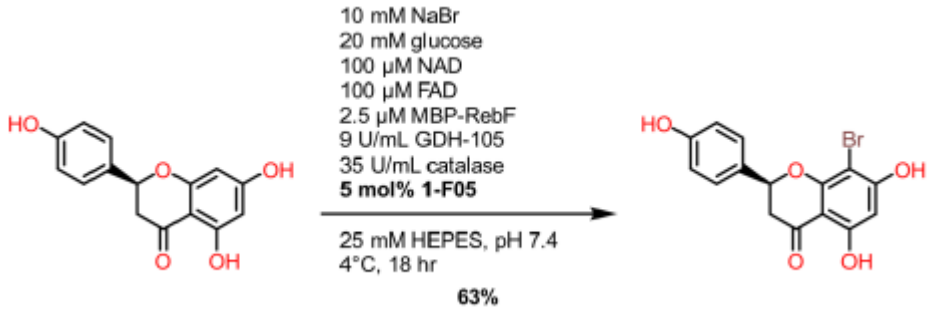


Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 5 mL of buffered solution was as follows: Pindolol (30 mM DMSO stock, 1 mM final concentration, 40 µmol, 9.9 mg), NaBr (10 eq. in reaction buffer, final concentration 10 mM), glucose (20 eq. in reaction buffer, final concentration 20 mM), NAD (0.1 eq. in reaction buffer, 100 µM), FAD (0.1 eq. in reaction buffer, 100 µM final concentration). 2-C01 (2.5 mol%, 25 µM final

concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated at room temperature (25 °C) agitated at 250 rpm in a VWR 1585 incubator. After 18 hours, the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This solution was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (Supelco Discovery C18; 3 mL/min, 19%B 0-5 min, 19-45%B 5.01-25 min, 45%B 25-27 min, 95%B 27.01 min-32 min, 19%B 32.01 min-35 min). Product distribution: 15.7:1.4:1 2-brominated:7-brominated:dibrominated product. The fractions found to contain product via LC-MS were collected and concentrated by rotary evaporation to produce product in 41% isolated yield.

^{1}H NMR: (500 MHz, Methanol-*d*4) δ 7.13 (s, 1H), 7.04 (t, *J* = 7.9 Hz, 1H), 7.00 (d, *J* = 8.2 Hz, 1H), 6.53 (d, *J* = 7.5 Hz, 1H), 4.19 (m, 1H), 4.14 (dd, *J* = 9.4, 4.6 Hz, 1H), 3.99 (dd, *J* = 9.3, 6.5 Hz, 1H), 3.20 (dd, *J* = 12.2, 3.4 Hz, 1H), 2.96 (hept, *J* = 6.5 Hz, 1H), 2.86 (dd, *J* = 12.2, 8.7 Hz, 1H), 1.15 (d, *J* = 6.2 Hz, 3H), 1.14 (d, *J* = 6.2 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-*d*4) δ 154.47, 139.92, 125.48, 124.97, 118.07, 107.33, 102.84, 88.74, 72.40, 70.13, 51.88, 50.95, 22.91. HRMS (ESI-MS): calc. for $[C_{14}H_{20}BrN_2O_2]^+$ ($[M+H]^+$): 327.0703 and 329.0683, found 327.0708 and 329.0688.

Naringenin bromination with 1-F05: 8-bromination

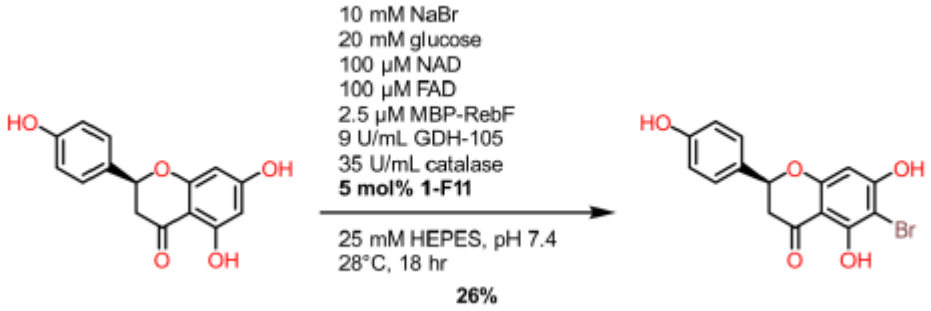


Eight 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3mL of buffered solution was as follows: Naringenin (30 mM DMSO stock, 500 µM final concentration 12 µmol, 3.3 mg), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-F05 (5 mol%, 25 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were placed into a crystallization dish, this crystallization dish was taped to the top of an Eppendorf Thermomixer R, and the reactions were incubated in a 4 °C cold room held and agitated at 300 rpm using the Thermomixer. After 18 hours, the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite. The filtrate was saturated with NaCl, then the aqueous layer was extracted into 10 mL 3:1 $CHCl_3$/$_i$PrOH three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated by rotary evaporation. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semi-prep HPLC (Supelco Discovery C18; 3.0 mL/min, 45%B 0-5 min, 45-53%B 5.01-23 min, 95%B 23.01-28 min, 45%B 28.01 min-30 min). Product distribution: 27.8:1:4.8 8-brominated:6-brominated:dibrominated product. Fractions with product were collected, and solvent was removed by rotary evaporation to produce product, which matched the $_1$H-NMR for 8-brominated naringenin,[29] in 63% isolated yield.

^{1}H NMR (500 MHz, DMSO-*d*6) δ 12.15 (s, 1H), 9.62 (bs, 1H), 7.34 (d, *J* = 8.3 Hz, 2H), 6.82 (d, *J* = 8.6 Hz, 2H), 6.13 (s, 1H), 5.59 (dd, *J* = 12.4, 3.1 Hz, 1H), 3.32 (dd, *J* = 17.2, 12.4 Hz, 1H), 2.82 (dd, *J* = 17.2, 3.2 Hz, 1H). Note: one phenol hydroxyl peak appears to be broadened into baseline. ^{13}C NMR: (126 MHz,

DMSO-*d*6) δ 196.83, 163.53, 162.41, 159.46, 158.22, 128.96, 128.62, 115.70, 102.86, 96.48, 88.74, 79.43, 41.90. HRMS (ESI-MS): calc. for $[C_{15}H_{10}BrO_5]-$ ([M−H]−): 348.9717; found 348.9711.

Naringenin bromination with 1-F11: 6-bromination

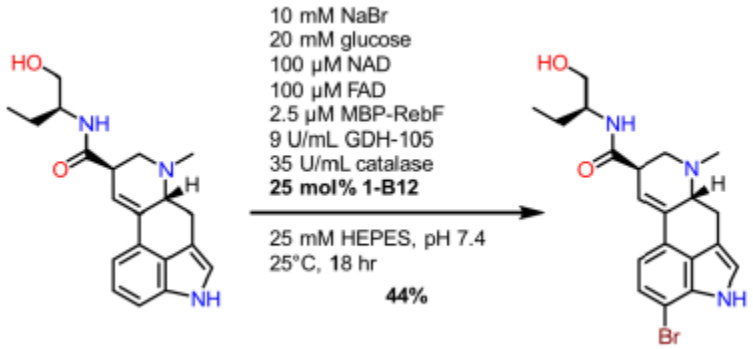


Ten 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stocks such that the final concentration of each component in 3 mL of buffered solution was as follows: Naringenin (30 mM DMSO stock, 500 µM final concentration 15 µmol, 4.1 mg), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-F11 (5 mol%, 25 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stocks of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated in VWR 1585 incubator and agitated at 250 rpm at 28 °C. After 18 hours the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite. This was saturated with NaCl, then the aqueous layer was extracted into 10 mL 3:1 $CHCl_3$/iPrOH eight times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 300 µL DMSO, then purified via via reverse-phase semiprep HPLC (Supelco Discovery C18, 3.0mL/min, 45%B 0-5 min, 45-53%B 5.01-23 min, 95%B 23.01-28 min, 45%B 28.01 min-30 min). Product distribution: 1:2.4:1.3 8-brominated:6-brominated:dibrominated product.

Purified fractions were collected, and solvent was removed by rotary evaporation to produce product in 26% isolated yield.

^{1}H NMR: (500 MHz, DMSO-*d*6) δ 8.16 (s, 1H), 7.31 (d, *J* = 8.6 Hz, 2H), 6.79 (d, *J* = 8.5 Hz, 2H), 5.98 (s, 1H), 5.41 (dd, *J* = 12.7, 3.0 Hz, 1H), 3.24 (dd, *J* = 17.1, 12.7 Hz, 1H), 2.68 (dd, *J* = 17.1, 3.1 Hz, 1H). Note: two phenolic resonances are broadened into baseline. ^{13}C NMR: (126 MHz, DMSO-*d*6) δ 195.62, 163.81, 161.79, 160.34, 158.19, 129.31, 128.77, 115.62, 101.47, 96.21, 90.73, 78.83, 42.02. HRMS (ESI-MS): calc. for $[C_{15}H_{10}BrO_5]-$ ([M−H]−): 348.9717 and 350.9698; found 348.9712 and 350.9690.

Methylergonovine bromination with 1-B12: 7-bromination

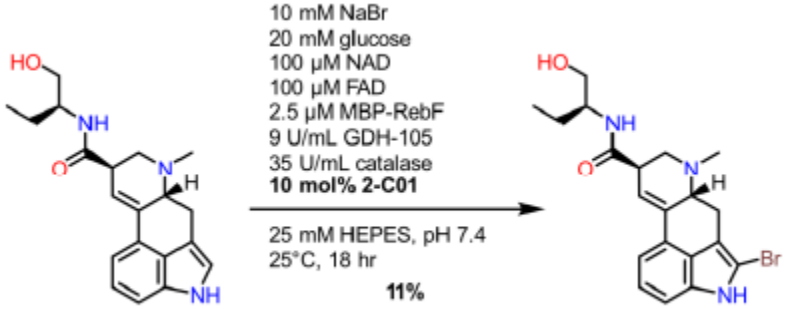


Four 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stock solutions such that the final concentration of each component in 3 mL of buffered solution was as follows: Methylergonovine maleate salt (30 mM DMSO stock solution, 500 µM final concentration 6 µmol, 2.7 mg of maleate salt), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer, 100 µM final concentration). 1-B12 (25 mol%, 100 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stock solution of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated at room temperature (25 °C) agitated at 250 rpm in a VWR 1585 incubator. After 18 hours, the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by

addition of concentrated NaOH. This solution was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (SUPELCO Discovery C18; 3.0 mL/min, 10%B 0-5 min, 10-38%B 5.01-23 min, 38%B 23.01-24 min, 95%B 24.0-28 min, 19%B 28.01-30 min). Product distribution: 1 : 9.4 7-brominated to 2-brominated product. Purified fractions were collected, and solvent was removed by rotary evaporation to produce product in 44% isolated yield.

^{1}H NMR: (600 MHz, Methanol-*d*4) δ 8.34 (bs, 1H), 7.24 (dd, *J* = 7.8, 2.2 Hz, 1H), 7.11 (dd, *J* = 7.8, 2.2 Hz, 1H), 7.07 (s, 1H), 6.53 (s, 1H), 3.83 (dd, *J* = 9.5, 5.1 Hz, 1H), 3.71 – 3.52 (m, 5H), 3.38 – 3.32 (m, 1H), 3.04 (t, *J* = 11.1 Hz, 1H), 2.79 (s, 3H), 2.75 (t, *J* = 12.8 Hz, 1H), 1.67 (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. ^{13}C NMR (126 MHz, DMSO-*d*6) δ 171.91, 134.61, 132.89, 127.54, 125.19, 121.97, 121.07, 113.42, 110.83, 102.05, 63.48, 62.72, 56.03, 52.63, 49.06, 43.81, 43.34, 27.11, 24.16, 10.94.27.11, 24.16, 10.94. HRMS (ESI-MS): calc. for $[C_{20}H_{25}BrN_3O_2]^+$ ($[M+H]^+$): 418.1125 and 420.1107, found 418.1126 and 420.1105.

Methylergonovine bromination with 2-C01: 2-bromination



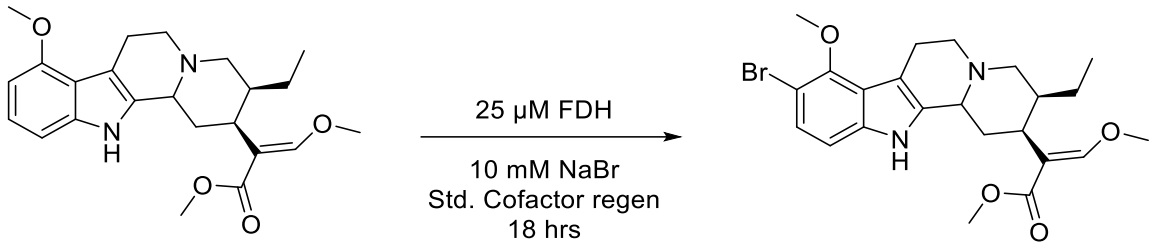
S42

Seven 20 mL scintillation vials consisting of reaction buffer (25 mM HEPES, pH 7.4) were charged with reaction components from concentrated stock solution such that the final concentration of each component in 3 mL of buffered solution was as follows: Methylergonovine maleate salt (30 mM DMSO stock solution, 500 µM final concentration 10.5 µmol, 4.8 mg maleate salt), NaBr (20 eq. in reaction buffer, final concentration 10 mM), glucose (40 eq. in reaction buffer, final concentration 20 mM), NAD (0.2 eq. in reaction buffer, 100 µM), FAD (0.2 eq. in reaction buffer,

100 µM final concentration). 2-C01 (10 mol%, 50 µM final concentration) and MBP-RebF (2.5 µM final concentration) were added next, followed by freshly prepared stock solutions of catalase (35 U/mL final concentration) and GDH (9 U/mL final concentration). The scintillation vials were covered with a non-breathable adhesive film, covered in aluminum foil, then holes were poked into the top of the film. The vials were incubated at room temperature (25 °C) and agitated at 250 rpm in a VWR 1585 incubator. After 18 hours, the reactions were collected into an Erlenmeyer flask and acidified to pH < 2 using 6 M HCl. The resulting solution was filtered through Celite, then basified to pH ≈ 12 by addition of concentrated NaOH. This solution was saturated with NaCl, then the aqueous layer was extracted into 10 mL DCM three times. The organic layer was then washed with brine, dried over $MgSO_4$, and concentrated. The residue was dissolved in 500 µL DMSO, then purified via reverse-phase semiprep HPLC (Supelco Discovery C18; 3.0mL/minute, 10%B 0-5 min, 10-38%B 5.01-23 min, 38%B 23.01-24 min, 95%B 24.01-28 min, 19%B 28.01-30 min). Product distribution: 2.1:1 2-brominated to 7-brominated product. The fractions found to contain product via LC-MS were collected and concentrated by rotary evaporation to produce product in 11% isolated yield.

^{1}H NMR: (500 MHz, DMSO-*d*6) δ 11.48 (s, 1H), 8.14 (s, 2H), 7.73 (d, *J* = 8.5 Hz, 1H), 7.13 (dd, *J* = 8.5 Hz, 4.1 Hz, 1H), 7.09 (d, *J* = 5.1 Hz, 1H), 6.40 (s, 1H), 3.66 (tt, *J* = 10.4, 5.2 Hz, 1H), 3.52 – 3.45 (m, 1H), 3.40 (dd, *J* = 10.7, 5.3 Hz, 1H), 3.34 (dd, *J* = 10.7, 6.0 Hz, 1H), 3.29 (dd, *J* = 14.8, 5.8 Hz, 1H), 3.19 – 3.13 (m, 1H), 3.11 (dd, *J* = 11.2, 5.3 Hz, 1H), 2.64 (t, *J* = 11.0 Hz, 1H), 2.53 (s, 3H), 2.42 (dd, *J* = 14.7, 11.2 Hz, 1H), 1.65 – 1.53 (m, 1H), 1.33 (ddt, *J* = 16.1, 14.5, 7.4 Hz, 1H), 0.84 (t, *J* = 7.4 Hz, 3H).^{13}C NMR: (126 MHz, DMSO-*d*6) δ 171.65, 163.62, 134.68, 134.12, 126.80, 126.17, 123.33, 121.75, 112.48, 109.71, 109.50, 104.28, 63.46, 62.17, 55.70, 52.69, 43.50, 42.99, 26.28, 24.16, 10.94. HRMS (ESI-MS): calc. for $[C_{20}H_{23}BrN_3O_2]^-$ ($[M-H]^-$): 416.0979 and 418.0959, found 416.0973 and 418.0952.

Mitragynine bromination with R2D3:



1 mM

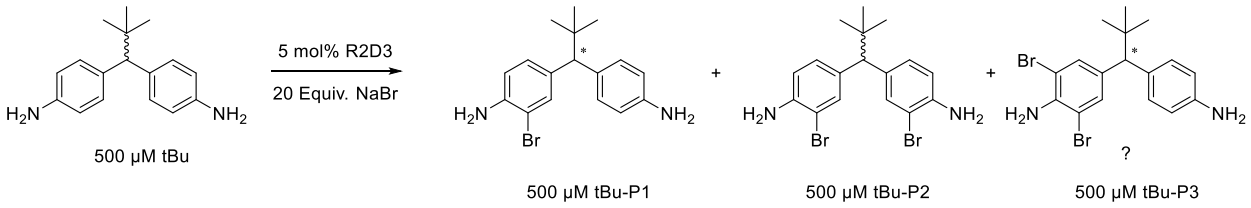
30 mL bioconversions were conducted in 300 mL plastic beakers (Fisherbrand™ Polypropylene Disposable Beaker, catalog No. FB012915) covered with a breathable plate seal. Mitragynine was prepared as a 20 mM stock solution in DMSO. The beakers were charged with 25 mM HEPES, pH 7.4, followed by NaBr (10 mM in reaction buffer), NAD (100 µM final concentration, in reaction buffer), FAD (100 µM final concentration in reaction buffer), glucose (20 mM final concentration, prepared in reaction buffer), glutathione (1 mM final concentration, prepared in reaction buffer), catalase (35 U/mL final concentration), MBP-RebF (2.5 µM final concentration), substrate (1 mM final concentration), D3 (25 µM final concentration), and finally GDH (9 U/mL final concentration). The breathable plate seal was placed over the beaker, which was promptly placed in a Thermo MaxQ 8000 incubator using a 250 mL Erlenmeyer pedestal. The beaker was set to incubate at 150 rpm, 25 °C for 16 hours.

Protein was denatured by the addition of 10% v/v of 1 M HCl. Sodium chloride was added to saturation, 50% v/v DCM was added, and the mixture was stirred vigorously for 10 minutes. The stir bar was removed, and the mixture was filtered through celite. The pH of the mixture was adjusted to pH >10 with sat. sodium bicarbonate, and the mixture was extracted 3x with 50% v/v DCM. The organic layers were combined and dried over magnesium sulfate. Once dry, the solution was filtered and concentrated by rotary evaporation. The resulting green oil was then dissolved onto celite and purified by reverse phage biotage with mobile phase A being water with 0.1% TFA and mobile phase B being ACN with 0.1% TFA according to the following method:

Start (%B)	End (%B)	Length
30	30	1 CV
30	60	30 CV
60	95	1 CV
95	95	2.5 CV

Fractions containing the major peak were collected and the solvent was removed under rotary evaporation. A second round of purification using 1:3 ethyl acetate:hexanes with 2% triethylamine was used to further purify the product. Proton NMR of the product confirmed 12-Br-Mitragynine[30] (See NMR section). 3.1 mg of material was isolated after pentane wash to remove excess grease, corresponding to 22% isolated yield.

4,4'-(2,2-dimethylpropane-1,1-diyl)dianiline bromination with R2D3:



30 mL bioconversions was conducted in 300 mL plastic beakers (Fisherbrand™ Polypropylene Disposable Beaker, catalog No. FB012915) covered with a breathable plate seal. Mitragynine was prepared as a 20 mM stock solution in DMSO. The beakers were charged with 25 mM HEPES, pH 7.4, followed by NaBr (10 mM in reaction buffer), NAD (100 µM final concentration, in reaction buffer), FAD (100 µM final concentration in reaction buffer), glucose (20 mM final concentration, prepared in reaction buffer), glutathione (1 mM final concentration, prepared in reaction buffer), catalase (35 U/mL final concentration), MBP-RebF (2.5 µM final concentration), substrate (500 µM final concentration), D3 (25 µM final concentration), and finally GDH (9 U/mL final concentration). The breathable plate seal was

placed over the beaker, which was promptly placed in a Thermo MaxQ 8000 incubator using a 250 mL Erlenmeyer pedestal. The beaker was set to incubate at 150 rpm, 25 °C for 16 hours.

Protein was denatured by the addition of 10% v/v of 1 M HCl. Sodium chloride was added to saturation, 50% v/v DCM was added, and the mixture was stirred vigorously for 10 minutes. The stir bar was removed, and the mixture was filtered through celite. The pH of the mixture was adjusted to pH >10 with sat. sodium bicarbonate, and the mixture was extracted 3x with 50% v/v DCM. The organic layers were combined and dried over magnesium sulfate. Once dry, the solution was filtered and concentrated by rotary evaporation. The resulting oil was then adsorbed onto celite and purified via reverse-phase biotage according to the following procedure:

%A	%B	Column Volumes
75	25	1 (equilibration)
75	25	1
30	70	25
5	95	2
95	95	3

Clean proton NMRs were only obtained for tBu-P1 and tBu-P2, the NMR spectrum for tBu-P3 is preliminary. Isolated yields were not recorded. As these results were not published, only proton spectrum were obtained for the products.

tBuP1: ^{1}H NMR (600 MHz, cdcl$_3$) δ 7.41 (d, *J* = 2.1 Hz, 1H), 7.14 (dd, *J* = 8.3, 2.0 Hz, 3H), 6.66 (d, *J* = 8.2 Hz, 1H), 6.63 – 6.57 (m, 2H), 3.45 (s, 1H), 0.97 (s, 9H).

tBuP2: ^{1}H NMR (600 MHz, cdcl$_3$) δ 7.38 (d, *J* = 2.1 Hz, 2H), 7.12 (dd, *J* = 8.3, 2.1 Hz, 2H), 6.68 (d, *J* = 8.2 Hz, 2H), 3.41 (s, 1H), 0.97 (s, 9H).

tBuP3: ^{1}H NMR (600 MHz, cdcl$_3$) δ 7.39 (s, 2H), 7.19 (d, *J* = 7.6 Hz, 2H), 6.80 (s, 2H), 4.40 (s, 2H), 3.45 (s, 1H), 0.97 (s, 9H).

2.5 References

(1) Lowell, A. N.; Demars, M. D.; Slocum, S. T.; Yu, F.; Anand, K.; Chemler, J. A.; Korakavi, N.; Priessnitz, J. K.; Park, S. R.; Koch, A. A.; Schultz, P. J.; Sherman, D. H. Chemoenzymatic Total Synthesis and Structural Diversification of Tylactone-Based Macrolide Antibiotics through Late-Stage Polyketide Assembly,

Tailoring, and C-H Functionalization. *J Am Chem Soc* **2017**, *139* (23), 7913–7920. https://doi.org/10.1021/jacs.7b02875.

(2) Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang, H. Design of an in Vitro Biocatalytic Cascade for the Manufacture of Islatravir - SI. *Science (1979)* **2019**, *366* (6470), 1255–1259. https://doi.org/10.1126/science.aay8484.

(3) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic Asymmetric Synthesis of Sitagliptin Manufacture. *Science (1979)* **2010**, *329* (July), 305–310. https://doi.org/10.1126/science.1188934.

(4) Payne, J. T.; Poor, C. B.; Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angewandte Chemie* **2015**, *127* (14), 4300–4304. https://doi.org/10.1002/ange.201411901.

(5) Fridman, E.; Pichersky, E. Metabolomics, Genomics, Proteomics, and the Identification of Enzymes and Their Substrates and Products. *Curr Opin Plant Biol* **2005**, *8* (3 SPEC. ISS.), 242–248. https://doi.org/10.1016/j.pbi.2005.03.004.

(6) Rodrigues, T.; Reker, D.; Schneider, P.; Schneider, G. Counting on Natural Products for Drug Design. *Nat Chem* **2016**, *8* (6), 531–541. https://doi.org/10.1038/nchem.2479.

(7) Reuter, J. A.; Spacek, D. V.; Snyder, M. P. High-Throughput Sequencing Technologies. *Mol Cell* **2015**, *58* (4), 586–597. https://doi.org/10.1016/j.molcel.2015.05.004.

(8) Itoh, N. *Metagenomics for Improved Biocatalysis*; Elsevier B.V., 2017. https://doi.org/10.1016/B978-0-444-63743-7.00018-4.

(9) Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Apweiler, R.; Alpi, E.; Antunes, R.; Arganiska, J.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R.; Bursteinas, B.; Chavali, G.; Cibrian-Uhalte, E.; Da Silva, A.; De Giorgi, M.; Dogan, T.; Fazzini, F.; Gane, P.; Castro, L. G.; Garmiri, P.; Hatton-Ellis, E.; Hieta, R.; Huntley, R.; Legge, D.; Liu, W.; Luo, J.; Macdougall, A.; Mutowo, P.; Nightingale, A.; Orchard, S.; Pichler, K.; Poggioli, D.; Pundir, S.; Pureza, L.; Qi, G.; Rosanoff, S.; Saidi, R.; Sawford, T.; Shypitsyna, A.; Turner, E.; Volynkin, V.; Wardell, T.; Watkins, X.; Zellner, H.; Cowley, A.; Figueira, L.; Li, W.; McWilliam, H.; Lopez, R.; Xenarios, I.; Bougueleret, L.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Blatter, M. C.; Boeckmann, B.; Bolleman, J.; Boutet, E.; Breuza, L.; Casal-Casas, C.; De Castro, E.; Coudert, E.; Cuche, B.; Doche, M.; Dornevil, D.; Duvaud, S.; Estreicher, A.; Famiglietti, L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; Jungo, F.; Keller, G.; Lara, V.; Lemercier, P.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T.; Nouspikel, N.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pozzato, M.; Pruess, M.; Rivoire, C.; Roechert, B.; Schneider, M.; Sigrist, C.; Sonesson, K.; Staehli, S.; Stutz, A.; Sundaram, S.; Tognolli, M.; Verbregue, L.; Veuthey, A. L.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Garavelli, J. S.; Huang, H.; Laiho, K.; McGarvey, P.; Natale, D. A.; Suzek, B. E.; Vinayaka, C. R.; Wang, Q.; Wang, Y.; Yeh, L.

S.; Yerramalla, M. S.; Zhang, J. UniProt: A Hub for Protein Information. *Nucleic Acids Res* **2015**, *43* (D1), D204–D212. https://doi.org/10.1093/nar/gku989.

(10) Höhne, M.; Schätzle, S.; Jochens, H.; Robins, K.; Bornscheuer, U. T. Rational Assignment of Key Motifs for Function Guides in Silico Enzyme Identification. *Nat Chem Biol* **2010**, *6* (11), 807–813. https://doi.org/10.1038/nchembio.447.

(11) Akiva, E.; Brown, S.; Almonacid, D. E.; Barber, A. E.; Custer, A. F.; Hicks, M. A.; Huang, C. C.; Lauck, F.; Mashiyama, S. T.; Meng, E. C.; Mischel, D.; Morris, J. H.; Ojha, S.; Schnoes, A. M.; Stryke, D.; Yunes, J. M.; Ferrin, T. E.; Holliday, G. L.; Babbitt, P. C. The Structure-Function Linkage Database. *Nucleic Acids Res* **2014**, *42* (D1), 521–530. https://doi.org/10.1093/nar/gkt1130.

(12) Gavin, D. P.; Reen, F. J.; Rocha-Martin, J.; Abreu-Castilla, I.; Woods, D. F.; Foley, A. M.; Sánchez-Murcia, P. A.; Schwarz, M.; O'Neill, P.; Maguire, A. R.; O'Gara, F. Genome Mining and Characterisation of a Novel Transaminase with Remote Stereoselectivity. *Sci Rep* **2019**, *9* (1), 1–15. https://doi.org/10.1038/s41598-019-56612-7.

(13) Tan, D.; Jamieson, C. S.; Ohashi, M.; Tang, M. C.; Houk, K. N.; Tang, Y. Genome-Mined Diels-Alderase Catalyzes Formation of the Cis-Octahydrodecalins of Varicidin A and B. *J Am Chem Soc* **2019**, 4–8. https://doi.org/10.1021/jacs.8b12010.

(14) Yang, M.; Fehl, C.; Lees, K. V.; Lim, E. K.; Offen, W. A.; Davies, G. J.; Bowles, D. J.; Davidson, M. G.; Roberts, S. J.; Davis, B. G. Functional and Informatics Analysis Enables Glycosyltransferase Activity Prediction. *Nat Chem Biol* **2018**, *14* (12), 1109–1117. https://doi.org/10.1038/s41589-018-0154-9.

(15) Podzelinska, K.; Latimer, R.; Bhattacharya, A.; Vining, L. C.; Zechel, D. L.; Jia, Z. Chloramphenicol Biosynthesis: The Structure of CmlS, a Flavin-Dependent Halogenase Showing a Covalent Flavin-Aspartate Bond. *J Mol Biol* **2010**, *397* (1), 316–331. https://doi.org/10.1016/j.jmb.2010.01.020.

(16) 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* (34), 12060–12068. https://doi.org/10.1021/jacs.7b06773.

(17) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using Sequence Similarity Networks for Visualization of Relationships across Diverse Protein Superfamilies. *PLoS One* **2009**, *4* (2). https://doi.org/10.1371/journal.pone.0004345.

(18) Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28* (23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

(19) Ashtekar, K. D.; Marzijarani, N. S.; 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* (38), 13355–13362. https://doi.org/10.1021/ja506889c.

(20) 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* (13–14), 1674–1681. https://doi.org/10.1016/j.tetasy.2010.05.029.

(21) Adla, S. K.; Golz, G.; Jones, P. G.; Lindel, T. Study on the NBS-Induced Rearrangement of 2-Tert-Prenyltryptamines. *Synthesis (Stuttg)* **2010**, *2010* (13), 2161–2170. https://doi.org/10.1055/S-0029-1218811.

(22) Mukai, K.; Sant'Ana, D. P. De; Hirooka, Y.; Mercado-Marin, E. V.; Stephens, D. E.; Kou, K. G. M.; Richter, S. C.; Kelley, N.; Sarpong, R. Bioinspired Chemical Synthesis of Monomeric and Dimeric Stephacidin A Congeners. *Nat Chem* **2018**, *10* (1), 38–44. https://doi.org/10.1038/NCHEM.2862.

(23) Sørensen, H. P.; Mortensen, K. K. Advanced Genetic Strategies for Recombinant Protein Expression in Escherichia Coli. *J Biotechnol* **2005**, *115* (2), 113–128. https://doi.org/10.1016/j.jbiotec.2004.08.004.

(24) Takayama, H.; Ishikawa, H.; Kurihara, M.; Kitajima, M.; Aimi, N.; Ponglux, D.; Koyama, F.; Matsumoto, K.; Moriyama, T.; Yamamoto, L. T.; Watanabe, K.; Murayama, T.; Horie, S. Studies on the Synthesis and Opioid Agonistic Activities of Mitragynine-Related Indole Alkaloids: Discovery of Opioid Agonists Structurally Different from Other Opioid Ligands. *J Med Chem* **2002**, *45* (9), 1949–1956. https://doi.org/10.1021/jm010576e.

(25) Payne, J. T.; Butkovich, P. H.; Gu, Y.; Kunze, K. N.; Park, H. J.; Wang, D. S.; Lewis, J. C. Enantioselective Desymmetrization of Methylenedianilines via Enzyme-Catalyzed Remote Halogenation. *J Am Chem Soc* **2018**, *140* (2), 546–549. https://doi.org/10.1021/jacs.7b09573.

(26) Sockett, R. E. Predatory Lifestyle of Bdellovibrio Bacteriovorus. *Annual Review of Microbiology*. Annual Reviews September 8, 2009, pp 523–539. https://doi.org/10.1146/annurev.micro.091208.073346.

(27) Payne, J. T.; Andorfer, M. C.; Lewis, J. C. Regioselective Arene Halogenation Using the FAD-Dependent Halogenase RebH. *Angewandte Chemie - International Edition* **2013**, *52* (20), 5271–5274. https://doi.org/10.1002/ANIE.201300762.

(28) Casas-Finet, J. R.; Wilson, S. H.; Karpel, R. L. Selective Photochemical Modification by Trichloroethanol of Tryptophan Residues in Proteins with a High Tyrosine-to-Tryptophan Ratio. *Anal Biochem* **1992**, *205* (1), 27–35. https://doi.org/10.1016/0003-2697(92)90574-Q.

(29) Yaipakdee, P.; Robertson, L. W. Enzymatic Halogenation of Flavanones and Flavones. *Phytochemistry* **2001**, *57* (3), 341–347. https://doi.org/10.1016/S0031-9422(01)00075-9.

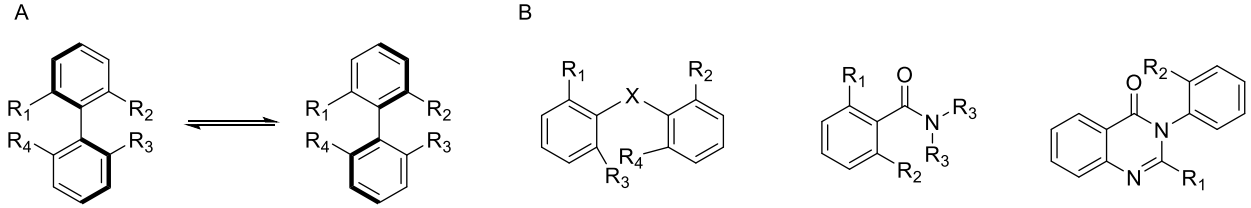
(30) Chakraborty, S.; DiBerto, J. F.; Faouzi, A.; Bernhard, S. M.; Gutridge, A. M.; Ramsey, S.; Zhou, Y.; Provasi, D.; Nuthikattu, N.; Jilakara, R.; Nelson, M. N. F.; Asher, W. B.; Eans, S. O.; Wilson, L. L.; Chintala, S. M.; Filizola, M.; van Rijn, R. M.; Margolis, E. B.; Roth, B. L.; McLaughlin, J. P.; Che, T.; Sames, D.; Javitch, J. A.; Majumdar, S. A Novel Mitragynine Analog with Low-Efficacy Mu Opioid Receptor Agonism Displays Antinociception with Attenuated Adverse Effects. *J Med Chem* **2021**, *64* (18), 13873–13892. https://doi.org/10.1021/acs.jmedchem.1c01273.

CHAPTER 3: IDENTIFICATION AND DIRECTED EVOLUTION OF A REBH VARIANT CAPABLE OF HIGHLY ATROPOSELECTIVE HALOGENATION OF 3-ARYLQUINAZOLIN-4(3*H*)-ONES ON PREPARATIVE SCALE

3.1 Introduction to atropisomerism and atroposelective halogenation

Atropisomerism is a type of axial chirality that results from restricted rotation around a σ-bond, typically resulting from steric hindrance preventing free rotation at a given temperature (**Figure 3.1A**).[1] Like point chirality, specific conformations of atropisomers can interact with other molecules in a stereoselective way. This form of stereo discrimination has been exploited in chemical catalysis to great effect in the last decade (e.g. BINAP, BINOL, etc.). Locking freely rotating drug compounds with sterically constricting substituents can also improve target specificity to reduce side effects that result from off-target protein binding. The combined need for atropisomers in both chemical synthesis[2] and drug development[1] has sparked renewed interest in the stereoselective synthesis of this class of enantiomer.

Figure 3. 1: Origins and scaffold diversity observed in atropisomers

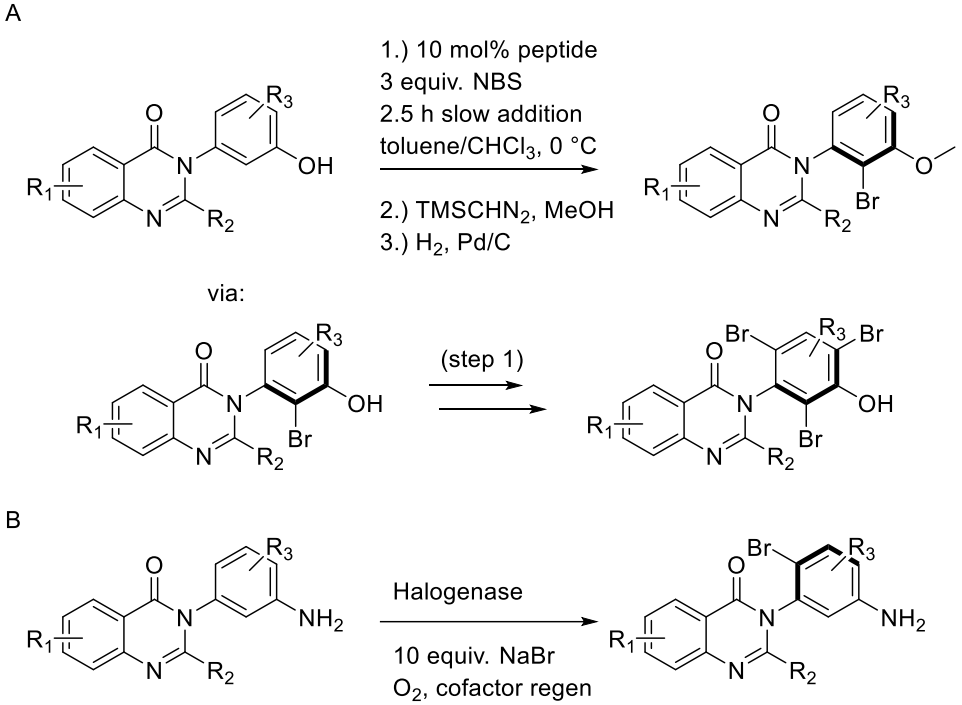


A) When the barrier of rotation around a biaryl axis is larger than 30 kcal/mol, stable atropisomers can be formed. B) A variety of scaffolds can generate stable atropisomers, including diarylamines/ethers, benzamides, and heterocyclic compounds.

Many synthetic approaches for the formation of stable atropisomers have been developed, including both chemical[3,4] and enzymatic methods.[5–7] Of particular relevance to my studies are the methods used to construct axially chiral biaryl compounds (**Figure 3.1B**). For such biaryl compounds, when the barrier to rotation is low, the interconverting atropisomers can be resolved via a dynamic kinetic resolution to a single enantiomer. In several pioneering examples, the Miller group has demonstrated the catalytic atroposelective electrophilic bromination of several scaffolds including

benzamides,[8] simple biaryls bearing directing groups,[9] and 3-arylquinazolinones.[10] In these studies, a peptide catalyst is used to preferentially bind one conformation of the rapidly rotating substrate and deliver the electrophilic bromine to one face of the compound replacing a C-H bond with the bulkier C-Br moiety and locking rotation around the biaryl axis.[11] Notably for the 3-arylquinazolinones, only one halogenation event is required to set the axis of chirality. Despite the requirement for functionalization only ortho to the axis of rotation, the poor site-selectivity of the peptide catalyst results in poly-bromination of the compound (**Figure 3.2A**). The undesired bromine substituents must either be removed[10] or selectively functionalized through additional steps[12]. Site-selective, atroposelective halogenations could streamline this process to increase the efficiency of accessing stable, mono-substituted atropisomers (**Figure 3.2B**).

Figure 3. 2: Atroposelective halogenation of quinazolinones



A) Peptide catalyzed atroposelective tribromination of 3-aryl-4(3*H*)-quinazolinones. B) FDH-catalyzed site- and atroposelective monobromination of 3-aryl-4(3*H*)-quinazolinones. SOURCE: J. Am. Chem. Soc. 2022, 144, 36, 16676–16682

In the last decade, our lab[13,14] and others[15–17] have demonstrated that flavin-dependent halogenases can be engineered to site-selectively halogenate a variety of scaffolds both similar to and distinct from those naturally functionalized by this enzyme class. Especially of note is their non-native application to enantioselective catalysis, both for the desymmetrization of methylenedianilines[18] and halocyclization of olefins.[19,20] FDH's are believed to achieve this site- and enantioselectivity by positioning a specific C-H bond proximal to HOX that is non-covalently bound within the enzyme active site.[21] Given the mechanistic similarity between the electrophilic bromination catalyzed by Miller's peptide catalysts and FDH's, we envisioned that an FDH could be used to selectively functionalize rapidly racemizing compounds, resulting in site-selective atropisomer formation. Compared to the existing literature for biocatalytic formation of point chirality, enzymatic formation of atropisomers is underdeveloped. Existing literature is dominated by redox reactions and hydrolysis mostly focusing on the kinetic resolution of biaryl compounds. Atroposelective lipase catalyzed macrocyclization[6] and oxidative biaryl coupling by P450's has been established,[7] but these require in the former case elevated temperatures and organic solvents, and in the latter suffer from poor yields and selectivity for the non-native examples. Given these limitations, we envisioned that a suitably engineered FDH variant might be capable of not only remedying a long-standing challenge in atroposelective chemical catalysis, but also expand the repertoire of enzymatic reactions as well. A paper was published summarizing a portion of these findings.[22]

Authorship

Benzamide and 3-(3'-hydroxy)quinazolinone synthesis and screening of purified FDH variants to identify hit 1-F08 was done by Dr. Dibyendu Mondal.

3.2 Results and discussion

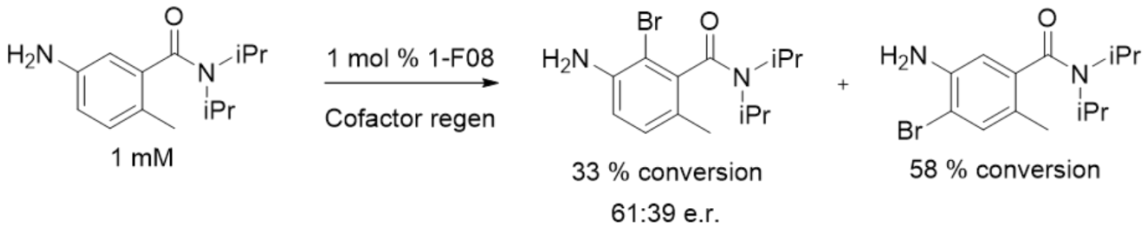
3.2.1 – Identification and attempted directed evolution of 1-F08 for the atroposelective halogenation of benzamide scaffolds.

For initial efforts aimed at identifying a scaffold for atroposelective halogenation, we began with screening a variety of readily accessible compounds that were reported to form stable atropisomers upon appropriate substitution. Particularly appealing in this regard was the benzamide scaffold, of which a variety of compounds can be prepared from readily available benzoic acid and secondary amines. Most atroposelective electrophilic halogenation literature utilizes phenol derivatives. In contrast, we opted to synthesize the aniline derivatives of these compounds, drawing on FDH literature showing improved activity for anilines relative to phenols.[23,24] This relative preference exhibited by FDH's for anilines over phenols likely results from a combination of improved binding to the enzyme and the favorable electronics of anilines for electrophilic halogenation reactions.

Due to the wide variation in expression levels we have previously observed with the genome mined enzymes we opted for screening these compounds against purified enzymes. The large number of both RebH mutants and WT genome mined enzymes in house precluded purification of the entire panel and necessitated a selection of enzymes with the highest activity to be prioritized. One enzyme-substrate pair from this subset, 1-F08 with diisopropyl aniline compound **B1**, demonstrated the formation of two distinct products when analyzed by LC-MS. While the major product was determined to be the para-substituted racemate, the minor product formed in roughly 15% conversion was the desired ortho-substituted atropisomer generated with an e.r. of roughly 39:61 (**Figure 3.3A**).

Figure 3. 3: Identification and attempted evolution of 1-F08 for dynamic kinetic resolution of benzamides.

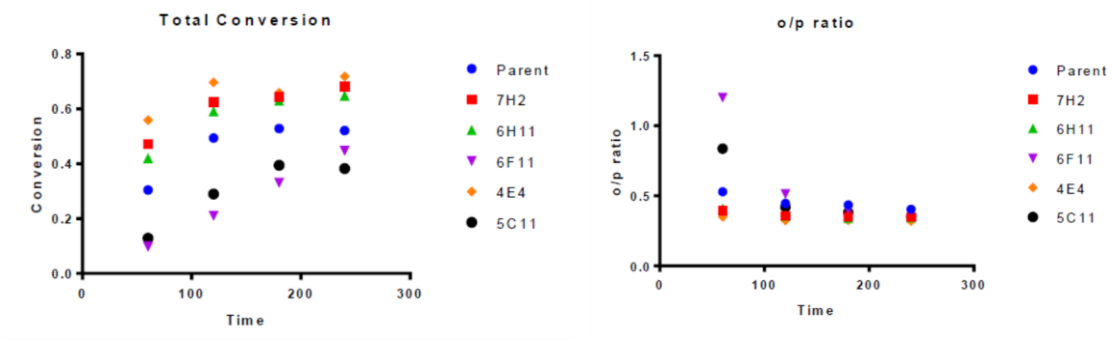
A)



B)

Name	O/P (lysate)	Activity (lysate)	Mutations	Non-silent
6-H-11	5.82	2.09	A160A, E382V	E382V
6-F-11	4.17	2.28	D204D, I259I, L438P	L438P
7-H-02	1.83	0.58	V52A, H231H, G456G	V52A
5-C-11	1.71	1.23	G300G, F432S	F432S
4-E-04	1.17	9.32	E169K, A440V, D480D, R490R	E169K, A440V

C)



A) Scheme for the reaction of 1-F08 with diisopropyl benzamide substrate B1. The e.r. of product B1a was measured by UHPLC with a chiral stationary phase after isolation from preparative halogenation. B) Location, sequence, ortho/para ratio, and fold-improvement to activity over parent 1-F08 as measured via UHPLC-MS C) Conversion and selectivity data for top variants obtained from directed evolution of 1-F08. Despite seeing over 10x improved selectivity for product B1a/B1b in UHPLC-MS screening in lysate, no improvements to selectivity were observed after purification of the top hits.

Given the poor selectivity of 1-F08, we sought to improve this enzyme via directed evolution. Unfortunately, 1-F08 could only be obtained in low yields and has limited activity when screening in lysate. These characteristics necessitated screening by UHPLC-MS with product detection by single-ion monitoring. The alkyl amine additive required for reasonable peak shape of the atropisomers on the reverse-phase stationary phase used for chiral UHPLC was incompatible with our UHPLC-MS system.

Therefore, we envisioned a two-tiered screening approach where initial improvements to activity and regioselectivity were characterized by achiral chromatography, and purified enzymes were then used for determination of improvements to enantioselectivity. Lack of an available crystal structure or acceptable homology model of 1-F08 led us to choose error-prone PCR to randomly introduce mutations across the protein scaffold with 1-2 residue mutations.

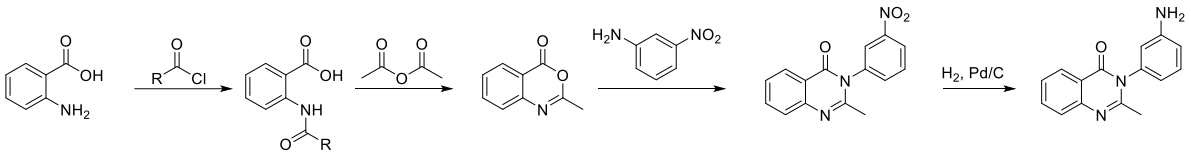
In total, 800 variants were screened (for the scheme, see **Figure 3.3 A,** for more details see the experimental section and **Figure 3.9**). Several improved variants, including 6-H11, 6-F11, 7-H02, 5-C11, and 4-E04, were obtained (sequences, ortho/para ratios, and fold improvement over parent in **Figure 3.3 B**). These variants were sequenced, expressed, and purified, and their activity was compared to WT 1-F08. Much to our surprise, despite the great improvements observed in lysate, none of these variants showed a different result from 1-F08 after overnight reaction (**Figure 3.3 C**). Examination of the time course for these variants showed an odd reaction profile, where formation of the ortho-isomer was initially favored compared to the undesired para product, but this preference was quickly inverted as the reaction proceeded leading to undesired product formation greatly outpacing desired product. We were unable to remedy this issue through reaction optimization, and while further attempting to probe this issue, the Covid-19 pandemic resulted in a lab shutdown preventing further analysis. After shifting to focus on the quinazolinones described in subsequent portions of this chapter, discovery of adventitious HOX release and subsequent non-specific halogenation with FDHs was reported by our lab and others. Given reaction optimization experiments that showed lower activity at higher pH's (**Figure 3.11**), it's possible that some of this reaction profile is due to free HOBr formation by 1-F08,[19,25] which could potentially react with substrate or modify the enzyme. Glutathione and other reducing agents are known to quench free HOX species, which suggests the reduced activity and higher selectivity observed when screening reactions in lysate results from free HOBr being quenched by the relatively reducing

environment of cell lysate. This perhaps offers a route to fix 1-F08 for atroposelective halogenation of benzamides but has not been explored.

3.2.2 – Quinazolinone synthesis and compilation plate screening to identify 6-TLP as a starting point for directed evolution of an atroposelective halogenase.

Given the lack of improvement observed with 1-F08 catalyzed halogenation of benzamides, we decided to reevaluate the literature to identify other potential substrates for atroposelective halogenation. Primarily, we drew potential substrates from the genome mining project[23] outlined in Chapter 2 and our lab's past efforts to profile RebH variant activity.[24] This in-depth examination of past substrate scope revealed that the benzamide scaffold did not bear high resemblance to substrates commonly accepted by characterized FDHs. In addition to the low activity typically observed for small aryl compounds, the sterically demanding diispropyl group was also likely resulting in inefficient binding to the FDH active site, which typically halogenate flat aromatic compounds. A second issue with the benzamide scaffold is the cis-trans isomerization observed with unsymmetrical disubstituted amides, which either limits the scope of the atroposelective halogenation or adds further complexity to the dynamic kinetic resolution by introducing an equilibrium event that can erode initial e.r..[26]

Scheme 3. 1: General quinazolinone synthesis



In contrast to bulky amide compounds, quinolines of various substitution patterns were halogenated by RebH and other genome mined enzymes in high yields. We envisioned that the quinoline portion of the compound could be used to anchor the substrate within the active site, and that an enzyme could be able to stabilize one conformation of the rapidly racemizing aryl ring. To this end, we set out to synthesize a panel of quinoline compounds that could be readily generated with high

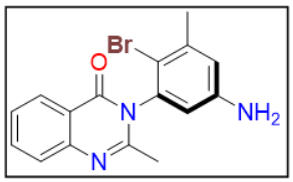
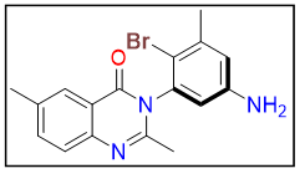
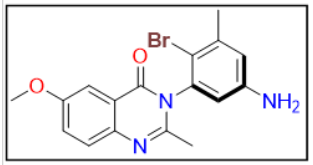
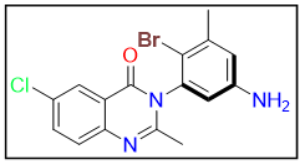
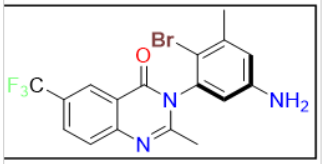
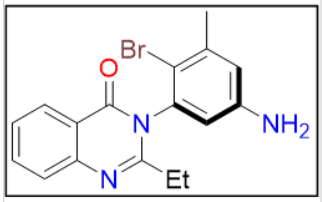
yields in four or fewer steps (**Scheme 3.1**). The condensation reaction with anthranilic acid and acetic anhydride is facile and goes to completion without the need for purification, facilitating large scale synthesis of the starting benzoxazole. This intermediate could then be condensed with widely available nitroanilines according to literature procedures to afford the 3-(3'-nitro)quinazolinones, which are readily reduced to the anilines via standard heterogenous reduction conditions. The resulting anilines often co-elute with an impurity whose m/z value matches the hydroxylamine resulting from incomplete reduction of the nitro compound, so purification via normal phase column chromatography was required after an initial reverse phase chromatography step. Authentic products were prepared for each substrate via bromination with NBS, and compilation plates comprising 47 WT FDHs and 95 evolved RebH variants were screened in lysate for bromination. In general, of the three substrates screened, quinazolinone **1** was halogenated by the most enzymes, though most gave only the undesired para isomer. Three RebH variants, 6-TLP, 3-S, and 3-SS showed detectable conversion to product **1a**, and upon isolation of the product from 6-TLP and chiral analysis via SFC, we were delighted to observe that **1a** was formed in near perfect enantioselectivity with only one atropisomer detectable (**Figure 3.13**). Despite the remarkable enantioselectivity of this enzyme, the regioselectivity was only favorable at low temperature, and the overall activity of the enzyme was low, requiring 5 mol% catalyst loading to adequately turn over 500 μM of **1**.

3.2.3 – Directed evolution of 6-TLP for activity and substrate scope

To improve the activity of 6-TLP and enhance the selectivity for **1a**/**1b**, we decided to begin directed evolution with our initial aim to improve regioselectivity. While crystal structures of RebH variants closely related to 6-TLP were available, the flexible loop thought to be involved in gating access of substrate to the active site was poorly defined in available structures. We therefore decided on error-prone PCR as our method of diversification for our first round of evolution. As with 1-F08, 3-4 base pair mutations/gene were targeted to give 1-2 residue changes per protein. In total, 800 variants were

screened for improved activity and regioselectivity as determined by UHPLC (**Figure 3.14**). One of these variants demonstrated 21-fold improvement in regioselectivity and 7-fold improvement to activity for substrate **1**. This variant contained one mutation, F402L (**Figure 3.15**), in the secondary sphere of the enzyme and was therefore designated as 1-L. Preparative reactions with 1-L revealed that in addition to the improvements in yield and site-selectivity, the enantioselectivity was perfectly maintained as well, validating the two-tiered screening approach for the development of an atroposelective halogenase.

Figure 3. 4: Substrate scope of 1-L

Product	Conv	e.r.
	72%	>99.5:0.5
	43%	98:2
	24%	97.5:2.5
	48%	96:4
	>5%	N.D.
	48%	92.5:7.5

Evaluation of the substrate scope of evolved RebH variant 1-L. Conversion refers to the NMR conversion of the product aniline relative to residual starting material and undesired products. E.r. determined by chiral SFC. SOURCE: J. Am. Chem. Soc. 2022, 144, 36, 16676–16682

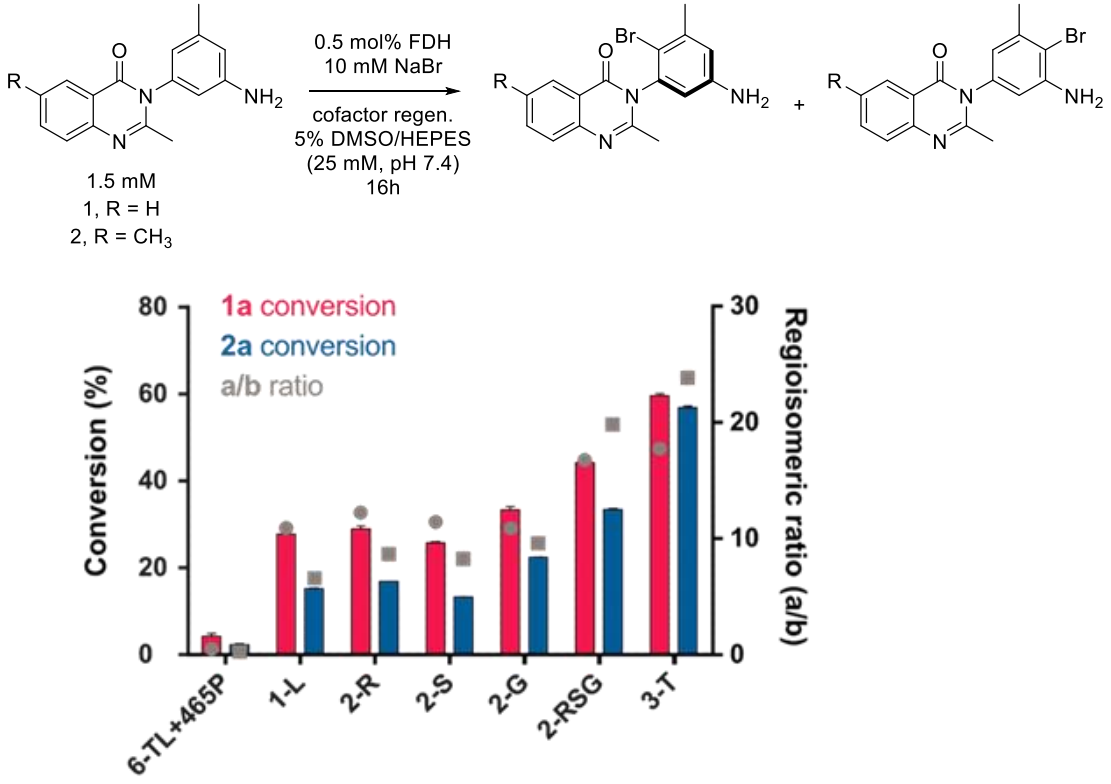
With the activity and selectivity of 1-L verified using purified enzyme, we then sought to characterize an expanded substrate scope for the enzyme as well (**Figure 3.4**). Quinazolinones of varying electronic properties and steric bulk were successfully halogenated with 1-L, and e.r. values ranging

from 85:15 to >99.5:0.5 were obtained using this variant. Deviation from the simple core of substrate **1**, however, resulted in losses to both yield and enantioselectivity. For example, while bromination of substrate **1** provided **1a** with 72% conversion and >99:1 e.r., bromination of 6-methyl substituted compound **2** provided **2a** with only 43% conversion and 98:2 e.r., and an analogous 6-chloro substituted product **12a** was likewise produced with reduced conversion and e.r., 46% and 96:4 respectively. Further examination of substitution at the 6-position revealed that 1-L provided only 24% conversion of 6-methoxy-substituted **11** to **11a** and had almost no activity on 6-trifluoromethyl substrate **13**. The lowest enantioselectivity was observed for product **8a** which bears an ethyl group in the 2-position of the quinazolinone core. While conversion as determined by NMR spectroscopy was 48%, the measured e.r. obtained from the crude reaction mixture was only 92.5:7.5. Taken together, these data indicated that while 1-L had reasonable flexibility in the scope of substituted quinazolinones it accepted, other substrates were halogenated in much lower yields and selectivity. This finding indicated that further directed evolution was required for biocatalyst optimization. We also evaluated a panel of 3-hydroxyquinazolinones (substrates **16**, **17**, **18**, data not shown) but low activity and no selectivity was observed so we instead continued our focus on 3-aminoquinazolinone substrates.

While 1-L showed high activity and selectivity for **1**, the significantly reduced activity and modestly reduced e.r. for compounds with steric bulk at position 6 of the quinazolinone ring led us to believe further evolution using substrate **1** would potentially lead to an enzyme with narrow substrate scope. To ensure a more broadly active final variant from our evolution campaign, we opted to take a substrate walking approach, switching to the 6-methylquinazolinone substrate **2** for screening a second round of error-prone PCR. Evaluating 1000 variants via UHPLC revealed only variants with modest improvements compared to 1-L (**Figure 3.16**). Nonetheless, a compilation plate of the best variants was re-screened, which revealed three mutations resulting in small improvements to activity and regioselectivity, W455R, N467S, and S469G (**Figures 3.17** and **3.18**). Analysis of a predicted structure for

1-L generated using the Alphafold[27] colab server revealed that these mutations were clustered on a flexible portion of the RebH active site that is proposed to 'gate' access of substrates to the enzyme active site. Through Quikchange PCR,[28] these variants were combined into all possible double mutants and the triple mutant combining all three mutations. Gratifyingly, while the double mutants demonstrated modest improvements to yield and selectivity, 2-RSG was nearly 2-fold improved for conversion of substrate **2** and exhibited similar improvements to regioselectivity. Importantly, comparable improvement for conversion of substrate **1** were also observed, demonstrating the improvements on the substrate used for evolution may extend to others with 2-RSG as well. After workup, SFC analysis of crude reactions demonstrated 2-RSG had either maintained or improved atroposelectivity for the substrates evaluated. This enzyme was also amenable to large-scale preparative reactions with higher substrate loadings, allowing for isolation of enough material to obtain the absolute configuration of the product via XRD. This showed 2-RSG and the other variants from this lineage formed the M-atropisomer of substrate **1**, which was presumed to match the configuration of other products as well.

Figure 3. 5: Directed evolution of variant 6-TLP for atroposelective halogenation.



Lineage of 6-TLP. Conversion and selectivity data for the directed evolution of 6-TL + F465P. Conversion is reported as the fraction of the area under the curve for the desired product (1a or 2a) over the total area (SM + 1/2a + 1/2b). Site selectivity is reported as the relative ratio of product 1/2a over b. Reactions were conducted in duplicate and are plotted, showing the standard deviation from UPLC analysis. Variants are named using the generation number, 1-3, followed by the single letter residue abbreviation for the mutation(s) introduced in each round: L = F402L, R = W455R, S = N476S, G = S469G, and T = S467T. SOURCE: J. Am. Chem. Soc. 2022, 144, 36, 16676–16682

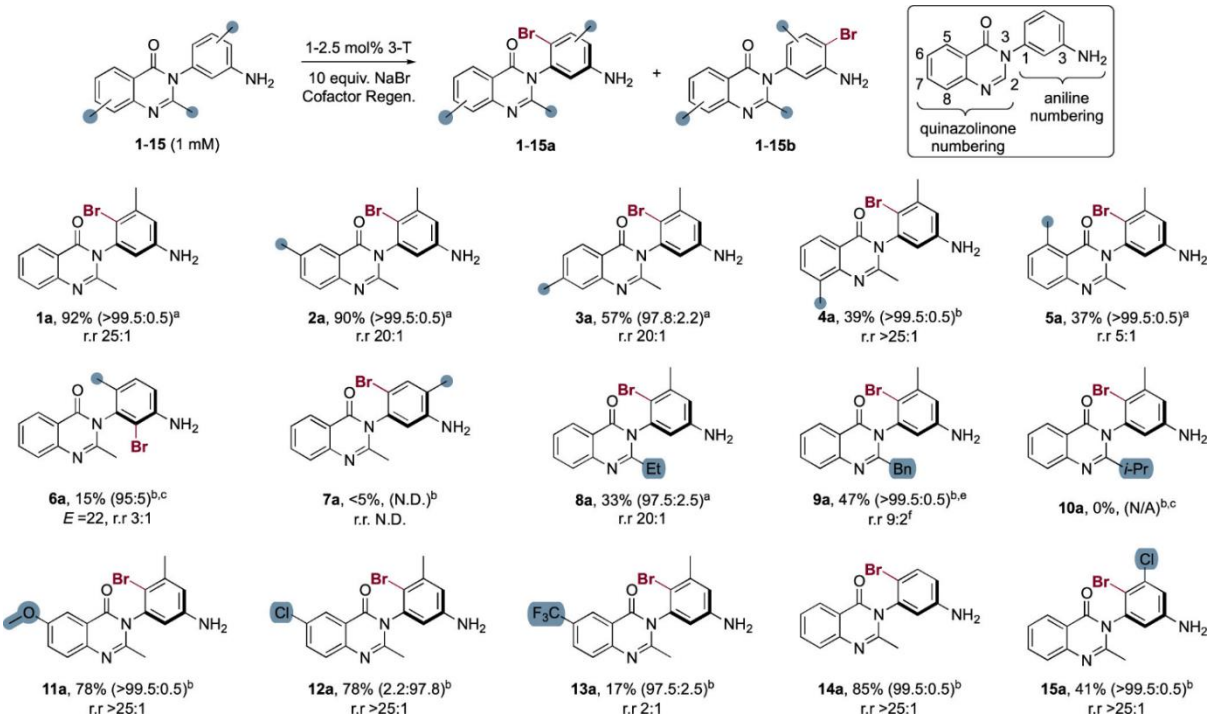
As a final optimization of 2-RSG, site saturation mutagenesis was performed on residues 455, 467, and 469 using degenerate NNK codons, and enough variants were screened to ensure 95% coverage of mutations at each site. Sequencing the top hits from the 455 and 469 libraries showed that the top variants were the parent sequence (**Figure 3.19**). Analysis of the top hits from the 467 library showed improvements resulting from both S467R (3R) and S467T (3T), which both provided roughly 2-fold improved conversion for substrate **2** in the assay compared to the 2-RSG enzyme used as the template in this round. Comparing 3-R and 3-T to the lineage at 0.5% equivalents of enzyme showed that for the halogenation of substrate **2**, 3-T was 91-fold improved for regioselectivity compared to

parent 6-TLP, and 25-fold improved for activity (**Figure 3.5**). For both substrates, the e.r. of 3-T was either maintained or improved when compared to that observed in the parent enzyme.

3.2.3 – Evaluation of the substrate scope of 3-T and kinetic characterization

We next sought to evaluate the substrate scope of variant 3-T. Sixteen quinazolinone derivatives were synthesized and evaluated with 3-T on 30 mL scale at 1 mM substrate loading, roughly corresponding to about 10 mg of expected product in each case. For all compounds where assay yield was over 5%, we obtained isolated yield for products. Compared to 1-L for the same substrates, we found in all cases that the activity, site-selectivity, and enantioselectivity was improved using 3-T. In many cases, 3-T was able to efficiently brominate in over 90% isolated yield with catalyst loading as low as 0.5 mol % relative to substrate at 1 mM substrate loading. Beyond the quinazolinone core, we also attempted the kinetic dynamic resolution of a 3-(6-methylquinolin-5-yl)aniline and 3-(2,5-dimethyl-1*H*-pyrrol-1-yl)aniline. Unfortunately, no variant from the lineage could successfully provide high yield of the halogenated quinoline, and the pyrrole was found to be unstable to conditions used in FDH catalysis.

Figure 3. 6: Substrate scope of final variant 3-T



Substrate scope of 3-T. Reactions were conducted on a 30 mL scale corresponding to roughly 10 mg of theoretical yield of the product. See the SI for details. Relevant compound numbering is provided in the inset. Product yields are isolated % yields, enantiomeric ratio (e.r.) was determined by chiral SFC analysis, and regioisomeric ratio (r.r.) was obtained from the ratio of integrals for products 1-15a:b obtained from UPLC analysis (r.r. > 25:1 denotes the second isomer not detected; N.D. indicates that the value was not determined). [a]1 mol % 3-T. [b]2.5 mol % 3-T. [c]Starting material exists as stable atropisomers. [d]Number refers to conversion as determined by UPLC. [e]20 equiv NaBr, 5 mol % 3-T. [f]See product characterization in the SI for details. SOURCE: J. Am. Chem. Soc. 2022, 144, 36, 16676–16682

Content with the improvement of activity and selectivity, we next evaluated the substrate scope of 3-T (**Figure 3.6**) for selective bromination of a panel of 3-(3-aminophenyl)-2-alkylquinazolin-4-(3*H*)-ones. To demonstrate the utility of this enzyme we chose to obtain isolated yields from preparative reactions for all substrates that provided >5% UPLC assay yield of the desired product. The isolated yields for these compounds ranged from 17-92%, high enantioselectivity (≥ 95:5 e.r.) was observed for all of the isolated products (for chiral SFC traces, see **Figures 3.20-3.35**). Previous studies using FDHs for enantioselective desymmetrization showed a minor improvement to e.r. from a secondary kinetic resolution of the monochlorinated species at the expense of yield for the desired product. To investigate whether this was also occurring for the developed atroposelective bromination we investigated the

UHPLC-MS trace looking at the major extracted ion for the di-brominated material of both a compound exhibiting high conversion and site-selectivity (**1**) and a second compound with low-site selectivity (**5**) (see section **3.4.5, Figures 3.38 and 3.39**). Only trace dibromination was observed, indicating that any improvement to e.r. resulting from a secondary resolution was insignificant compared to the initial chirality-setting halogenation event. For the substrates in **Figure 3.6**, only the 4'-methylaniline substrate **7** and the R2 = 2-isopropyl substrate **10** provided yields below 5%, illustrating how subtle differences in substrate structure can dramatically alter reaction outcomes in this system. Using variant 2-RSG to isolate compound **1a** we established that the (*M*) atropisomer of **1a** was formed, and this selectivity was presumed for the remaining products.

Given that methyl substitution at the of the quinazolinone ring lowered yield and e.r. for variant 1-L from the lineage, the activity of 3-T on substrates with differential methyl substitution on both the quinazolinone core and the 3-aminophenyl group was examined. The methyl substitution of the 6- (**2a**) and 7-positions (**3a**) of the quinazolinone was well tolerated, and only a slight loss in site selectivity was observed. The substitution of the 8-position led to decreased yield (**4a**), but high site selectivity and enantioselectivity were obtained with slightly increased catalyst loading. Reduced site selectivity (5:1) was obtained with a methyl group at the 5-position, though enantioselectivity of the isolated product **5a** remained high. The methyl substitution of the 4-position of the 3-aminophenyl group in substrate **7** was not tolerated. The analogous substitution of the 6-position of this group led to stable atropisomers in substrate 6, providing the opportunity to establish whether 3-T could be used to resolve these atropisomers via classical kinetic resolution. Even though the 6-methyl group occupies the position typically halogenated by 3-T, **6a** was formed in 15% yield with an enantiomeric ratio of 95:5, constituting an E value of 22. Notably, while **6a** was the major product for 3-T and therefore was the only isomer for which e.r. was determined, **6b** was the major product when using NBS. This again shows how FDHs can be used for regioselective brominations overriding substrate electronics.

We also evaluated the effect of steric bulk at the quinazolinone 2-position, which is crucial for restricting rotation of the biaryl axis, and found that it was well tolerated. Both the ethyl- and benzyl-substituted compounds **8** and **9** were highly selectively halogenated with e.r. values of 97.5:2.5 and >99:1 respectively, although for the benzyl compound a third regioisomer was observed after purification of **9a** resulting in reduced site-selectivity. The corresponding isopropyl compound **10**, however, was not halogenated by any enzyme in the lineage. In most cases, 3-T tolerated various substituents at the 6-position including the electron donating methyl (**2a**), methoxy (**11a**), and chloro (**12a**) groups, but a strongly electron withdrawing triflouoromethyl group significantly reduced yield and site-selectivity while maintaining high e.r. (**13a**). In following with the electrophilic halogenation mechanism, yield for the desired product trended with more electron donating substituents at the 5' position of the aniline ring. The 5'-methyl substituted product **1a** was produced in 92% yield with 1 mol% catalyst loading, whereas the 5'-H (**14a**) and 5'-Cl (**15a**) substituted compounds required 2.5 mol% catalyst loading and were isolated in 85% and 41% yield respectively.

For all substrates evaluated except 3-amino-4-methylphenyl substituted substrate **7** and 2-isopropyl substituted substrate **10**, we observed substantial improvements in yield of the desired product from 1-L to 3-T. In addition to the improved yields, 3-T either improved or maintained both site-selectivity and atroposelectivity when compared to 1-L for each product (see **Figure 3.4** and **3.6**). While site-selectivity for most reactions exceeded 10:1, regioisomeric ratios for substrates 5 and 13 were relatively low (5:1 and 2:1, respectively). Traditional resolution substrate 6 and 2-benzyl-substituted 9 were also halogenated with reduced site selectivity (3:1 and 9:2, respectively), perhaps relating to the more complex structures of these substrates.

Substrate solubility was observed to be an important consideration for activity as demonstrated by the 8-methyl substituted compound substrate **4a**. When this substrate was added to the small

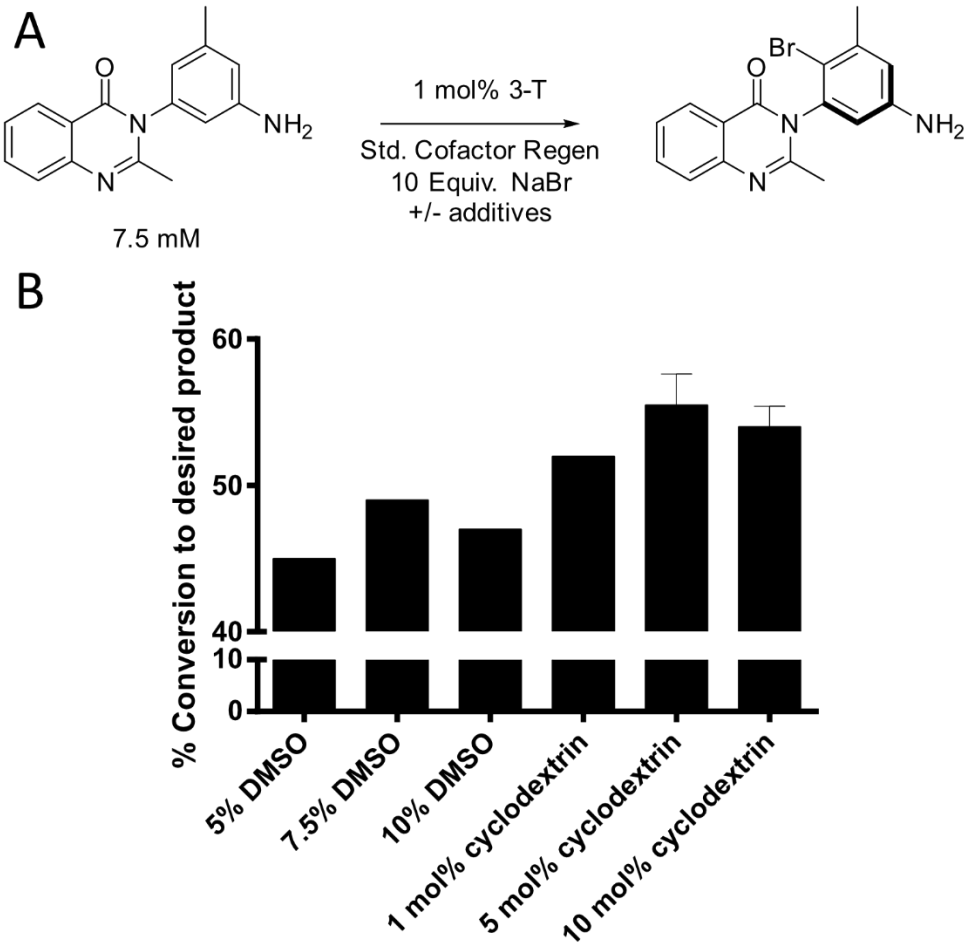
molecule mix as a concentrated stock solution, it precipitated immediately and would not resolubilize over the course of the reaction. The product assay yield in this case was limited to ~15% compared to when the substrate was added directly to the reaction which resulted in substantially improved yields. While this effect was also noted for the 2-benzyl substituted compound, which was remedied by lowering the substrate concentration, it was most prominent for an additional substrate not covered in the publication, 2'-methyl substituted quinazolinone **19** (see experimental). This substrate exists as two stable atropisomers and was synthesized to determine the ability of 3-T for kinetic resolution. This compound was uniquely insoluble in organic solvents, including DMSO, DMF, and alcoholic solvents, leading us to abandon this compound in favor of the more soluble substrate **6**. A quinazolinone with a pyridyl aniline (**20**, see experimental) was also evaluated for halogenation by 3-T and other variants in the lineage, but no activity was detected, likely a result of the relative electron deficiency of heterocyclic arenes.

After determining the substrate scope of 3-T, we sought to evaluate the steady-state kinetics of the halogenation of substrate **1**. A calibration curve was constructed, and the initial rate of **1a** formation was measured by UHPLC. The K_M was determined to be 2.0 μM and the k_{cat} was found to be 0.03 min^{-1} (**Figures 36** and **37,** see experimental). While the k_{cat} is low compared to WT-RebH with the native substrate L-tryptophan, this result compares favorably to previously evolved halogenase variants. For example, among evolved RebH variants evaluated in our lab the previous highest reported k_{cat}/K_M observed with a non-native substrate is the halogenation of tryptamine by evolved RebH variant 0S (k_{cat}/K_M = 0.013 min^{-1} μM^{-1}).[14] Considering the conformational complexity associated with substrate **1** compared to the simple indole scaffold in tryptamine this result is encouraging for the application of FDHs to atroposelective halogenation.

3.2.4 – Scale up halogenation of and subsequent cross-coupling

While the capability of FDHs to halogenate arenes site-selectively has been exploited for preparative halogenation reactions, the low catalytic rates and suspected deactivation of FDHs often prevents ready adaptation to preparative scale reactions. Additionally, when applying FDHs to desymmetrization of methylenedianilines and the halocyclization of electron rich olefins, previous efforts at scaling up FDH reactions to larger volumes has resulted in reduced yield and enantioselectivity (data not shown). Encouraged by the relatively favorable kinetics observed for the halogenation of substrate **1**, we sought to determine if 3-T could successfully be applied to preparative scale. Several different enzyme-substrate concentrations and ratios were evaluated to maximize the yield of product while preserving precious enzyme. While the activity of 3-T toward substrate **1** was the highest observed in the substrate scope study, concerns about subsequent derivatization of the sterically encumbered aryl bromide **1a** drove us to also evaluate if halogenation of the less encumbered aniline substrate **14** could be accomplished at similarly high substrate loading. On analytical scale, 2 mol% of 3-T generated product **14a** in 76% conversion at 2.5 mM substrate loading. If these results were translated to a 40 mL reaction, they would provide roughly 25 mg of product. While this would be acceptable, the high enzyme loading could require many expression cultures and complicate product extraction due to high levels of emulsification. We chose instead to attempt derivatization of the sterically challenging substrate **1a** with **14a** as an alternate should the cross-coupling fail.

Figure 3. 7: Screening of additional DMSO and cyclodextrin for improved substrate solubility

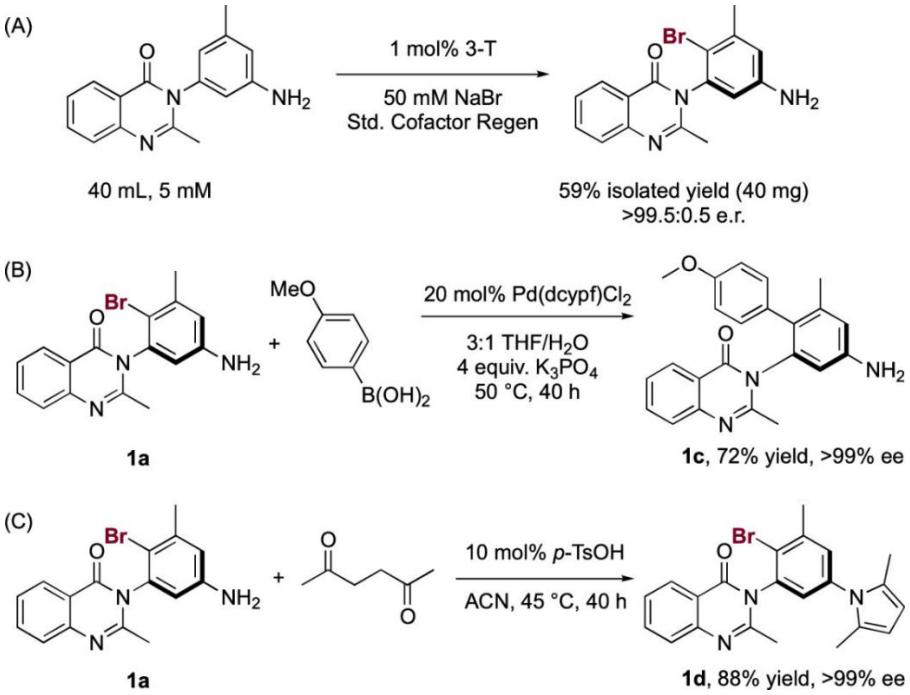


A) Scheme for the halogenation for screening additives to improve solubility. Reactions were conducted in duplicate and the average of the values with standard deviation is shown. Mol % of cyclodextrin is relative to substrate.

Additives that are believed to sequester hydrophobic molecules within a cavity have also been used to increase substrate loadings in aqueous reactions. The effects of DMSO and α-cyclodextrin (a common amphiphilic additive) concentration on the halogenation of substrate 1 by 3-T halogenation was therefore examined. On analytical scale, we observed only a small impact of the % DMSO in the reaction corresponding to less than 5% assay yield compared to the standard reaction conditions. A more pronounced effect was observed on the addition of α-cyclodextrin, with 5 mol% of the additive relative to substrate improving yield by over 10% when conducted in duplicate (**Figure 3.7**).

Unfortunately, this effect did not initially seem to translate to the scaled-up reaction when conducted on 4 mL scale; a significant reduction in yield was obtained after workup compared to the standard reaction. In the last stages of this project, it was recognized that product **1a** was substantially less soluble than the starting material **1**. At high substrate loadings, precipitated product was lost during the workup on the celite cake used to remove precipitated protein from the reaction. It was found that washing this celite cake with methanol to release precipitated product from the celite significantly improved isolated yield. Considering this information, perhaps the loss in observed yield after workup with inclusion of α-cyclodextrin resulted from increased product concentration, and subsequent increased precipitation from the reaction. Unfortunately, due to time constraints this was not reinvestigated prior to the completion of this work, and ultimately the preparative halogenation was conducted without additives.

Scheme 3. 2: 5 mM halogenation of substrate 1 and subsequent derivatization



Preparative-scale reaction of substrate 1. See the SI for more details, including the reaction setup. (B) Suzuki coupling of (M)-1a to (M)-1c. (C) Paal-Knorr reaction of (M)-1a to (M)-1d. SOURCE: J. Am. Chem. Soc. 2022, 144, 36, 16676–16682

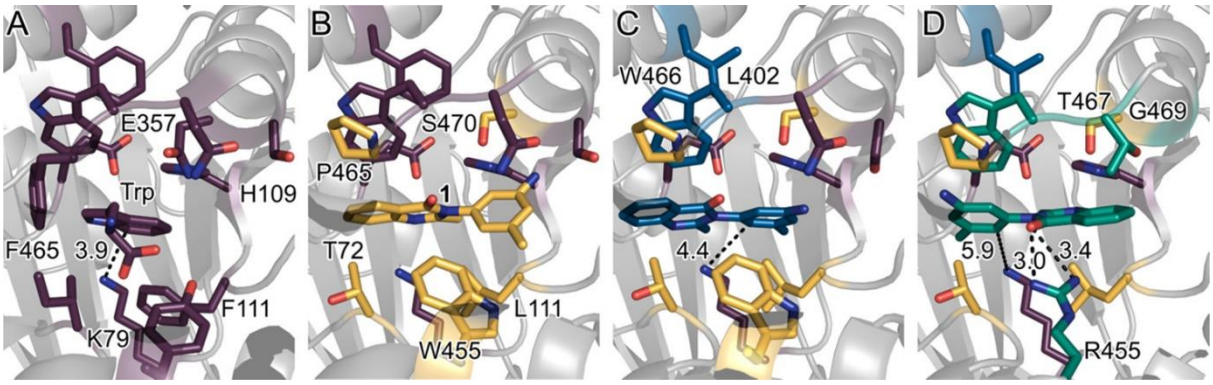
To our delight upon scale up of the 3-T halogenation with **1** using the conditions outlined in **Scheme 3.2** we observed 70% conversion to **1a** by UHPLC and obtained **1a** in 60% isolated yield, corresponding to 39.6 mg of product after purification (**Scheme 3.2A**). This product titer (0.99 mg/mL) is among the highest reported for a non-native preparative reaction catalyzed by an FDH without immobilization or enzyme recycling. Crucially, unlike in the previous enantioselective examples of FDH catalysis,[18] the enantioselectivity of the reaction was perfectly maintained, with the minor enantiomer not detectable. Isolation of this quantity of material allowed us to derivatize this compound in two ways from a single biocatalytic reaction. Despite the steric constriction around the aryl bromide, **1a** readily underwent Suzuki coupling with *p*-methoxyboronic acid to afford the triaryl product **1c** in 72% yield without erosion of e.r. (**Scheme 3.2B**). Pyrroles are privileged biological moieties, and to demonstrate the benefits of functionalizing unprotected anilines, the pyrrole compound **1d** was synthesized in 88% yield under standard Paal-Knorr conditions again maintaining the e.r. (**Scheme 3.2C**). These reactions combined with the observed substrate scope provide examples of how 3-T could be used to build complexity onto simple quinazolinone scaffolds.

3.2.5 – Structural analysis and rationalization of improved selectivity by modeling

After analysis of the biocatalytic potential of 3-T, we wanted to probe how structural changes in the 6-TLP scaffold may be influencing the improved activity and selectivity observed throughout the evolution. Our lab has successfully crystallized both the wild-type RebH and RebH mutants derived from various directed evolution projects leading us initially to attempt resolving the structure through X-ray crystallography. Unfortunately, these initial attempts at this did not provide crystals of sufficient length for analysis by XRD. Additionally, while RebH mutants have been successfully crystallized, a flexible loop comprising residues 446 to 456 is frequently not well ordered enough for diffraction. Given the importance of residue R455 and the proximity of S467 and G469 to this loop, the possibility remained that even upon obtaining a structure of 3-T it may not provide the insight we desired. Given these

limitations, we decided to shift our focus to modeling the protein using Alphafold, a recently developed AI-based protein folding algorithm. Using the publicly available Alphafold colab web server, models were built of 6-TLP, 1-L, and 3-T. Substrate **1** was docked into the enzyme active sites using the Chimera software suite from UCSF.[29]

Figure 3. 8: Docking studies of variants from 6-TLP lineage



Comparison of substrate binding in (A) the crystal structure of the RebH-tryptophan-FAD-Cl complex (PDB 2OA1) (39) with calculated docking poses of 1 in computational models of RebH variants (B) 6TL F465P, (C) 1-L, and (D) 3-T. Tryptophan and key residues in the RebH active site are shown in purple; the colors of substrate 1 and key residues/mutations then change to yellow, blue, and green as they are introduced in 6TL F465P, 1-L, and 3-T.

Previous work from our group and others has indicated that in many cases the site of halogenation can be predicted or rationalized by docking of substrate within the active site of models or crystal structures. By identifying the sufficiently reactive C-H bond positioned proximate to the catalytic lysine residue these models are generally predictive of halogenase selectivity as seen for tryptophan in the active site of RebH (**Figure 3.8 A**). Comparing the active sites of wild-type RebH and 6-TLP shows that mutations accumulated in the lineage of 6-TLP increased the volume of the FDH active site potentially allowing the typically narrow active site of RebH to accommodate the larger quinazolinone. Docking of substrate **1** into 6-TLP (**Figure 3.8 B**) failed to produce any structures where lysine K79 was positioned proximate to C6 of the aniline ring, consistent with the low activity of this enzyme for production of **1a**. In contrast, variant 1-L shows the shortest distance between this position and the catalytic lysine (**Figure**

3.8 C). This shortened distance is potentially enabled by F402L which is predicted to reposition W466 towards **1** bound in the active site. This rotates the core of **1** such that C6 of the aniline ring is only 4.4 Å away from the catalytic lysine. In the docked model of substrate **1** in the active site of 3-T, the substrate adopts a pro-M conformation with C6 of the aniline ring 5.9 Å from the catalytic lysine. This pose is consistent with the observed site- and enantioselectivity of the reaction. In variant 3-T, mutation W455R significantly changes the shape and charge of the active site and makes up a part of the loop gating access of substrate to the active site. This mutation is predicted to flip the orientation of **1** in the active site by forming a bidentate hydrogen bond with the carbonyl oxygen of **1** (**Figure 3.8D**). The importance of this residue for high selectivity is potentially highlighted by the observation that 5-methyl substituted compound **5** exhibits substantially lower regioselectivity than the other methyl substituted quinazolinones. This docking model also helps to explain the lack of activity for substrates **7** and **10** as well. The 3-methyl substitution on the aniline ring of **7** would clash with Y362 and the isopropyl group of **10** is predicted to clash with I82, explaining the lack of robust halogenation observed with these substrates. Our model does not predict any direct influence of residues T467 and G469 on the positioning of **1** within the active site, but these residues are near S470. This serine residue has been implicated in a persistent hydrogen bonding network with residues E357 and H109 that lines the substrate binding pocket of the enzyme.[21] The importance of T467 and G469 in improving activity and selectivity could implicate modulation of the hydrogen bonding network to help accommodate the quinazolinone substrates.

3.3 – Conclusions

In conclusion, this work demonstrates the engineering of a RebH variant for the kinetic and dynamic kinetic resolution of 3-(3-aminophenyl)-4(3H)-quinazolinones. This effort not only contributes to the growing field of enantioselective FDH catalysis but also more broadly to methods in atroposelective synthesis. The demonstrated preparative reactions conducted with 3-T show that this

enzyme can be used for the isolation of enough enantiopure material for subsequent derivatization. In addition to the published efforts for scalable halogenation, additive screening showed that inclusion of reagents meant to sequester hydrophobic compounds can improve the activity of halogenases for substrates with low aqueous solubility. This work also shows how in addition to the desymmetrization of meso-methylenedianilines and the halocyclization of styrenes with pendant nucleophiles, new modes of asymmetric FDH catalysis continue to be discovered. The variety of enantiopure scaffolds accessed via electrophilic atroposelective halogenations suggests that FDHs could find further use for site- and atroposelective reactions.

3.4 – Experimental

3.4.1 – General procedures

Materials

All chemical reagents were purchased from commercial suppliers and used without further purification unless otherwise noted. Primers used in this study were purchased from SigmaAldrich. NAD, FAD, and antibiotics were purchased from Chem-Impex International Inc. GDH-105 (hereafter, GDH) was purchased from Codexis. Catalase from bovine liver was obtained from Millipore Sigma (2000-5000 U/mg, product number C9322). 2-methyl-4H-benzo[d][1,3]oxazin-4-one (2-methylbenzoxazone) and other benzoxazone derivatives as well as N-acylanthranilic acids,[10] 1,3-dianiline, and 5-methylbenzene-1,3-diamine[30] were synthesized as previously reported and used without further purification. 3-(3'-hydroxy)quinazolinones were synthesized and purified as previously reported. Antibiotics were prepared as 1000x stock solutions of 50 mg/mL kanamycin and 100 mg/mL ampicillin in water and 25 mg/mL chloramphenicol in 95% ethanol. The pGro7 plasmid encoding the chaperone proteins groEL and groES chaperone set was purchased from Takara (Otsu, Shiga, Japan). RebH variant[14,31,32] and genome mined[23] compilation plates were obtained from prior studies. Reaction buffer is constant throughout the study and is defined as 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) at a concentration of 25

mM titrated to pH 7.4 using 1 M NaOH. This buffer was prepared using Milli-Q water to ensure no chloride sources were present in solution. Data visualization and kinetic analysis were conducted using Microsoft Excel and GraphPad Prism.

Semi prep, Biotage, UPLC, SFC column, method info

Ultra-high pressure liquid chromatography (UHPLC) was performed using an Agilent 1200 series system with a 1290 Infinity II high-speed pump, a 1260 Infinity II diode array detector, and a 1290 Infinity II multisampler with single-needle configuration. All achiral UPLC analysis was performed using an Eclipse Plus C18 2.1x5 mm guard column with a 1.8 µm pore size (part number 821725- 901) connected to a ZORBAX rapid resolution C18 column (part number 959757-902) with 0.1% TFA in water as mobile phase A and 0.1% TFA in acetonitrile as mobile phase B. The UHPLCMS instrument used in this study was the same, except this was connected to a 6135X single quadrupole mass spectrometer with an Agilent Jet Stream ESI source. Mobile phase A for UHPLC-MS analysis was water with 0.1% FA, mobile phase B was acetonitrile with 0.1% FA. Reverse phase purification was performed using either or both: 1.) a Biotage solera One with 12 g SNAP-KP-C18-HS columns using 0.1% FA (formic acid) in water as mobile phase A and 0.1% FA in methanol as mobile phase B or 2.) 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% FA in water as mobile phase A and 0.1% FA in acetonitrile as mobile phase B. Most chiral chromatography was performed using an Agilent 1260 Infinity SFC equipped with a CHIRALPAK®IC-3 column with supercritical CO2 as mobile phase A and methanol with 0.1% diethylamine (DEA) as mobile phase B. The general method for SFC chromatography was an 8- minute isocratic method with 30% mobile phase B and 2.5 mL/min flow rate. For one product, **13a**, the enantiomers could not be separated using this method. Instead, this product was analyzed on an Agilent 1100 HPLC

equipped with a Phenomenex Lux 3u Cellulose-1 150 x 4.6 mm column using normal phase chiral chromatography method 1 as detailed in the analysis section of this SI.

Chromatography/Purification methods

UHPLC Method A (used for the general analysis of quinazolinone bioconversions and reactions with **B1**)

Method A		
Time (min)	%A	%B
0	82	18
0.5	82	18
4	30	70
4.01	5	95
4.5	5	95
4.51	82	18
0.5 min post run	82	18

UHPLC method B (used for kinetic analysis of 3-T with substrate **1**, library screening for substrate **2**, and preparation of the calibration curve)

Method B		
Time (min)	%A	%B
0	74	26
0.5	74	26
1.25	68	32
1.3	55	45
2.5	48	52
2.75	5	95
2.8	74	26

Normal phase chiral chromatography 1 (used to determine the enantioselectivity of **13a** from the reaction with 3-T)

Time (min)	%A (hexanes)	%B (IPA)
0	88	12
30	80	20

Purification methods

Semiprep method 1 (used to purify product **1a** and **1b** from bioconversions using 6-TLP as the catalyst using the HPLC system described in the general methods)

Time (min)	%A	%B
5	57	43
20	57	43
22	57	43
22.01	5	95
28	5	95
28.01	57	43

Semi prep method 2 (used to purify products **2a** and **11a** from bioconversions with 1-L)

Time (min)	%A	%B
5	55	45
20	45	55
22	45	55
22.01	5	95
28	5	95
28.01	55	45

Semi prep method 3 (used to purify products **8a** and **12a** from bioconversions with 1-L)

Time (min)	%A	%B
5	45	55
20	30	70
22	30	70
22.01	5	95
28	5	95
28.01	45	55

Semi prep method 4 (used to purify product 6a from bioconversions with 3-T and NBS reactions)

Time	%A	%B
0	51	49
5	51	49
13	48	52
14	5	95
19	5	95
21	51	49

Biotage method 1 (used to purify most of the starting materials either from the condensation or reduction reactions)

%A	%B	Column Volumes
90	10	1 (equilibration)
90	10	1
55	45	25
5	95	2
95	95	3

Biotage method 2 (used to purify most products after NBS or bioconversions, as well as crosscoupled product **1d**)

%A	%B	Column Volumes
75	25	1 (equilibration)
75	25	1
30	70	25
5	95	2
95	95	3

UHPLC-MS method 1 (used to verify the lack of significant dibromination in reactions with 3-T and substrates 1 and 5).

Method 1		
Time (min)	%A	%B
0	80	20
0.5	80	20
4	20	80
4.5	5	95
4.51	80	20

3.4.2 – Cloning and Protein Production

Cloning procedures

Standard molecular biology protocols were followed.[33] Oligonucleotides were ordered from Sigma and prepared according to the manufacturer's recommendations. DNA sequences were visualized using the Benchling website. All FDH expression plasmids were cloned such that the start codon for the gene was in line with the Nco1 cut site such that halogenases had an Nterminal 6xHis tag and constructed using ibson assembly[34] or QuikChange[28] from PCR products. All primer sequences and descriptions are included in a table below. Error prone PCR was performed using Taq DNA polymerase. Vector and QuikChange PCR fragments were generated using PrimeSTAR® Max DNA Polymerase. After PCR, the resulting linear DNA fragments were purified via gel extraction from agarose using the Qiagen Gel Extraction kits, purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. The linear DNA fragments were eluted with 30 µL water. 3 µL of rCutSmart buffer (NEB#B6004S) was then added to the linear DNA, followed by 1 µL of DPN1 (NEB#R0176L), and the reaction was incubated at 37 °C for 1 hour. The reaction was then purified using the Zymo Research Clean and Concentrate kits (Cat#D4004 or D4013) according to the manufacturers protocol and both insert and vector were eluted with 10 µL water. Assembly reactions were performed using the NEBuilder Hifi DNA Assembly Master Mix using 0.1 pmol of the insert and 0.05 pmol of the vector. 2 µL of the resulting assembly was used to transform E. coli BL-21(λDE3) electrocompetent cells containing the pGro7 plasmid10 via electroporation and plated with the appropriate selection antibiotics. Point

mutations were introduced using QuikChange PCR. After QuikChange PCR, the resulting linear DNA fragments were purified via gel extraction and treated with DPN1 as described above. The reaction was then purified using the Zymo Research Clean and Concentrate kits according to the manufacturers protocol and eluted with 10 µL molecular-grade water. 5 µL of these linear DNA fragments were then used to transform E. coli as described above. Electroporation was carried out on a Bio-Rad MicroPulser using method Ec2. Genes were confirmed by sequencing from whole plasmids by Quintara Biosciences (625 Mt Auburn, St STE 105, Cambridge, MA).

Error prone PCR of 1-F08 to prepare the insert:

Reagent	Stock Concentration	Volume (µL)	Final Concentration
Forward Primer (HS150)	10 µM	2	200 nM
Reverse Primer (HS151)	10 µM	2	200 nM
dNTP	10 mM	2	200 µM
Taq polymerase	5000 U/mL	0.5	100 U
Taq 10x buffer	10 x	10	1 x
$MnCl_2$	6 mM	2.5	150 µM
Template (G3)	50 ng/µL	1	0.5 ng/µL
Mg H_2O		80	

Vector PCR of 1-F08 to prepare the vector:

Reagent	Stock Concentration	Volume (µL)	Final Concentration
Forward Primer (1-F08 vector FP)	10 µM	2.5	500 nM
Reverse Primer (1-F08 vector RP)	10 µM	2.5	500 nM
2x Primestar	2x	25	1x

Template (1-F08 in pET28)	50 ng/µL	1	5 ng/ µL
Water		19	

Error prone PCR of RebH variant 6-TLP to prepare the insert for round 1 of evolution:

PCR mix:

Reagent	Stock Concentration	Volume (µL)	Final Concentration
Forward Primer (HS209)	10 µM	2	200 nM
Reverse Primer (HS210)	10 µM	2	200 nM
dNTP	10 mM	2	200 µM
Taq polymerase	5000 U/mL	0.5	100 U
Taq 10x buffer	10 x	10	1 x
$MnCl_2$	5 mM	2.5	125 µM
Template (G3)	50 ng/µL	1	0.5 ng/µL
Mg H_2O		80	

Error prone PCR cycle:

Temperature (°C)	Length of cycle (min:sec)	Repetitions
95	5	1
95	0:30	22
58	0:30	
68	1:45	
68	10:00	1
4	hold	1

Vector PCR:

Reagent	Stock concentration	Volume (µL)	Final concentration
Primestar 2x Mix	2x	25	1x
Forward primer (HS211)	10 µM	2.5	500 nM
Reverse primer (HS208)	10 µM	2.5	500 nM
Template (pET28b)	50 ng/µL	0.5	0.5 ng/µL
Water		19.5	

Error prone PCR of evolved variant 1-L:

Insert:

Reagent	Stock Concentration	Volume (µL)	Final Concentration
Forward Primer (HS209)	10 µM	2	200 nm
Reverse Primer (HS210)	10 µM	2	200 nm
dNTP	10 mM	2	200 µM
Taq polymerase	5000 U/mL	0.5	100 U
Taq 10x buffer	10 x	10	1 x
$MnCl_2$	5 mM	2.5	125 µM
Template (1-L)	50 ng/µL	1	0.5 ng/µL
Mg H_2O		80	

Error prone PCR cycle:

Temperature (°C)	Length of cycle (min:sec)	Repetitions
95	5	1
95	0:30	22
58	0:30	
68	1:45	
68	10:00	1
4	hold	1

Vector PCR:

Reagent	Stock concentration	Volume (µL)	Final concentration
Primestar 2x Mix	2x	25	1x
Forward primer (HS211)	10 µM	2.5	500 nM
Reverse primer (HS208)	10 µM	2.5	500 nM
Template (pET28)	50 ng/µL	0.5	0.5 ng/µL
Water		19.5	

QuikChange PCR to generate 2-RS

Reagent	Stock	Volume (µL)	Final Concentration
Forward primer (HS243)	10 µM	2.5	500 nM
Reverse primer (HS244)	10 µM	2.5	500 nM
Template (2-S)	50 ng/µL	1	1 ng

Water		19	
Primestar 2x mix	2x	25	1x

QuikChange PCR to generate 2-SG

Reagent	Stock	Volume (µL)	Final Concentration
Forward primer (HS237)	10 µM	2.5	500 nM
Reverse primer (HS238)	10 µM	2.5	500 nM
Template (2-G)	50 ng/µL	1	1 ng
Water		19	
Primestar 2x mix	2x	25	1x

QuikChange PCR to generate 2-RG

Reagent	Stock	Volume (µL)	Final Concentration
Forward primer (HS243)	10 µM	2.5	500 nM
Reverse primer (HS244)	10 µM	2.5	500 nM
Template (2-G)	50 ng/µL	1	1 ng
Water		19	
Primestar 2x mix	2x	25	1x

QuikChange PCR to generate 2-RSG

Reagent	Stock	Volume (µL)	Final Concentration

Forward primer (HS237)	10 µM	2.5	500 nM
Reverse primer (HS238)	10 µM	2.5	500 nM
Template (2-R)	50 ng/µL	1	1 ng
Water		19	
Primestar 2x mix	2x	25	1x

NNK PCRs of 2-RSG:

455 NNK PCR:

Reagent	Stock concentration	Volume (µL)	Final concentration
Primestar 2x Mix	2x	25	1x
Forward primer (HS230)	10 µM	2.5	500 nM
Reverse primer (HS229)	10 µM	2.5	500 nM
Template (2-RSG)	50 ng/µL	1	0.5 ng/µL
Water		19	

467 NNK PCR:

Reagent	Stock concentration	Volume (µL)	Final concentration
Primestar 2x Mix	2x	25	1x
Forward primer (HS239)	10 µM	2.5	500 nM
Reverse primer (HS240)	10 µM	2.5	500 nM
Template (2-RSG)	50 ng/µL	1	0.5 ng/µL
Water		19	

469 NNK PCR:

Reagent	Stock concentration	Volume (µL)	Final concentration
Primestar 2x Mix	2x	25	1x
Forward primer (HS241)	10 µM	2.5	500 nM
Reverse primer (HS242)	10 µM	2.5	500 nM
Template (2-RSG)	50 ng/µL	1	0.5 ng/µL
Water		19	

List of primers used in this study:

Primer name	**Sequence**	**Use**
1-F08 vector FP	caccaccaccaccaccactgagatccggctgctaacaaagcTTgaaaggaag	1-F08 vector PCR
1-F08 vector RP	CTTTTCACCATggtatatctccttcttaaagttaaacaaaattatttctag	1-F08 vector PCR
HS150	tttgttagcagccggatctcagtggtgg	1-F08 ePCR
HS151	tttaagaaggagatataccATGGTGAAAAGTATCGTGGTG	1-F08 ePCR
HS208	CTGCTGCCCATGGtatatctccttcttaaagttaaac	Error prone PCR - insert
HS209	taagaaggagatataCCATGGGCAGCAGCCATCATCATCA	Error prone PCR - insert
HS210	CGAGTGCGGCATCAAGCTTTCAGCGGCCGTGCTGTTG	Error prone PCR - vector
HS211	ACAGCACGGCCGCTGAAAGCTTGATGCCGCACTCG	Error prone PCR - vector
HS229	GAACTCCTCCTCGAAGTTGCCMNNGTAGAGCTGGGCGTCG	W455 NNK
HS230	GGCAACTTCGAGGAGGAGTTCCGCAAC	W455 NNK
HS237	GAGTTCCGCAACCCTTGGAGCAACGGCAGCTACTACTGCGTGCT GGC	Quickchange 2H5 triple mutant
HS238	CCAAGGGTTGCGGAACTCCTCCTCGAAGTT	Quickchange 2H5 triple mutant
HS239	GAGTTCCGCAACCCTTGGnnkAACGGCAGCTACTACTGCGTGCTG	N467 NNK
HS240	CCAAGGGTTGCGGAACTCCTCCTCGAAGTTG	N467 NNK
HS241	CGCAACCCTTGGAGCAACnnkAGCTACTACTGCGTGCTG	S469 NNK - 2H5 RSG
HS242	GTTGCTCCAAGGGTTGCGGAACTCCTCCTC	S469 NNK - 2H5 RSG
HS243	ACGACGCCCAGCTCTACAGGGGCAACTTCGAGGAGGAGTTCCG	W455R for S469G and N467S
HS244	GTAGAGCTGGGCGTCGTCGGGCGCGGGCGCGTTGATC	W455R for S469G and N467S
T7 'long'	GTAATACGACTCACTATAGGGC	Sequencing
T7 reverse	GCTAGTTATTGCTCAGCGG	Sequencing

Halogenase library preparation in culture plates

Method 1:

This method was used when starting from the RebH and genome mined compilation plates. Glycerol stocks containing BL-21(λDE3) *E. coli* cells transformed with pGro7 and pET28b harboring the RebH variants or genome mined enzymes were stamped into autoclaved 1 mL 96-well deep well plates containing 300 µL of LB supplemented with the appropriate antibiotics. These inoculated plates were sealed with an AeraSeal adhesive film and incubated overnight at 37 °C, 225 rpm overnight. 30 µL of these overnight cultures was used to inoculate 1 mL of TB supplemented with the appropriate antibiotics in 2 mL 96-well deep well plates. Inoculated expression cultures were sealed with an AeraSeal adhesive film and incubated at 37 °C, 225 rpm until OD_{600} ≈ 0.6-0.8, at which point the incubator was allowed to cool to 30 °C. Gene expression was induced with 100 µM IPTG, 0.2 mg/mL L-arabinose and proceeded for 20 hours, 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 100 µL lysis buffer (0.75 mg/mL lysozyme, 25 mM HEPES, pH 7.4). Cells were incubated in lysis buffer for 45 min at 37 °C, 225 rpm. After 45 minutes had elapsed, the cells were removed from the incubator and flash frozen in liquid nitrogen until all liquid had frozen. The 96-well plate was then removed from liquid nitrogen and allowed to warm for 5 minutes at ambient temperature. After 5 minutes, the 96-well plate was placed into a lukewarm water bath to ensure uniform thawing of the lysed cell pellets. When all wells were thawed, 10 µL of DNAse buffer (0.1 mg/mL DNAse1 in 25 mM HEPES, pH 7.4) was added to each well. The plates were then centrifuged at 500 rpm for 30 seconds to ensure mixing of DNAse buffer and lysed cell pellets. The 96-well plate was then incubated for 15 minutes at 37 °C, 225 rpm. The insoluble fraction of the cell lysate was then removed by centrifugation at 3600 rpm at 4 °C for 20 minutes, and the clarified lysate was transferred to a microtiter plate for bioconversions.

Method 2:

This method was used for libraries after plating *E. coli* onto LB agar. Single colonies were picked using a Norgen systems colony-picking robot from rectangular agar plates. *E. coli* cells transformed with pGro7 and pET28b harboring RebH variants were picked into 1 mL 96-well deep well plates containing 300 µL of LB supplemented with the appropriate antibiotics. After overnight incubation of the *E. coli* cultures, secondary cultures were inoculated and gene expression induced as described in method 1.

Large-scale halogenase preparation

For large scale halogenase production, 14 mL culture tubes containing 10 mL LB with kanamycin and chloramphenicol were inoculated with BL-21(λDE3) *E. coli* cells that had been transformed with pGro7 and pET28b containing the appropriate flavin-dependent halogenase (FDH) gene (*vide infra*) as previously described. Construction and description of the halogenase expression plasmids is included in the cloning section of this SI. The cultures were incubated overnight at 37 °C, 250 rpm. The next day, 750 mL TB with the appropriate antibiotics was inoculated with the entire overnight culture. The inoculated expression cultures were incubated at 37 °C, 225 rpm until OD_{600} was between 0.6 and 0.8. The incubator was allowed to cool to 30 °C, and gene expression was induced with 2 mg/mL arabinose and 100 µM IPTG, and the expression culture was incubated for 20 hours. Cells were harvested by centrifugation at 3600 rpm for 20 minutes in a Sorvall XT centrifuge, then either frozen and kept at -20 °C or immediately resuspended in 25 mM HEPES buffer, pH 7.4. Cells were lysed using a QSonica S-4000 with a 0.5" horn at 40 W and a total processing time of 5 minutes with 1 minute on/off cycles. Cell lysate was clarified at 15,000 rpm for 30 minutes, and the resulting clarified lysate was transferred to a fresh 50 mL centrifuge tube and added to pre-equilibrated Ni-NTA (equilibration buffer: 20 mM phosphate, 300 mM NaCl, 10 mM imidazole pH 7.4). Clarified lysate was incubated with resin for approximately 30 minutes at which point it was transferred to uncapped spin columns and the lysate was allowed to flow through. The resin was washed with at least 5 CV wash buffer (20 mM phosphate, 300 mM NaCl, 25 mM imidazole pH 7.4), at which point the spin column was transferred to a new centrifuge tube and the resin was washed with elution buffer (20 mM

phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4). Eluted protein was concentrated via diafiltration using Amicon spin filters Ultra 30K MWCO spin filters and the buffer was exchanged for storage buffer (25 mM HEPES and 10% glycerol, pH 7.4). For long term storage, proteins were immediately frozen in liquid nitrogen and stored at – 80 °C until use. When stored in this way, enzyme activity was stable for at least two months. Halogenase concentration was determined by A_{280} measurements taken on a Tecan Infinite M200 pro microplate reader using the extinction coefficient calculated by the amino acid sequence of the enzyme. MBP-RebF concentrations were determined using the Pierce BCA Protein Assay Kit.

Protein sequences of major variants discussed:

Bold red letters are used to denote mutations from Wt-RebH. All variants in this study (6-TLP and mutants) included an N-terminal his tag in a flexible GS linker with the sequence MGSSHHHHHHSSGLVPRGSH**M**… where **M** is the start of the halogenase gene. The mutations of 1-F08 are included in **Figure 3.3**.

* = Stop

1-F08

MVKSIVVVGGGSAGWITANLLNAHCRRAGRQTQITLVESPDIPTIGVGEATVPSIRRTLSDIGLPEQELLISAEATFKTLIR
YKNWNLGDSYDHPFDRRRRPLTDGAVRRWTAQNNLPFDRSFSALSQLADNHISPKSPRQPQYQGAFPYAYHLDAIKL
ANRLASFGRERGIQHKLCKVTDVKVSPEGLISAVHTDQGEALTADLYVDCTGFRAALLGGGLGVKMKSFAKHLLCNRAV
TMRIPYEVYKPERIVTYTQAIARDHGWQWDINLQTRRGIGYVYSSDFLSEDAAETALRQNEGPHSDGLEARHIRFTSGK
RLVSWKGNCVAVGLADGFLEPLESSGLYLVEFAGRAIATMLDDFATAPQATARYYNRIVDELYDEILGFLNLHYVTSKRR
DTPFWQAVTADDAILDDLRDRLELWKRRSPTEFDFPGADRLFAQDSYEFILHGMKYVDGPVSGAPAVPDMSALISKSR
SELPDHEAILALVQRAAAARAHHHHHH*

Wt-RebH (his tag not included)

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATIPNLQTAFFDFLGIPEDEWMRECNASYK
VAIKFINWRTAGEGTSEARELDGGPDHFYHSFGLLKYHEQIPLSHYWFDRSYRGKTVEPFDYACYKEPVILDANRSPRRL
DGSKVTNYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTATGRVFDADLFVDCSGFRGLLINK
AMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYSSRFATEDEAVREFC
EMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQLVKHFPDKSLNPVLTARFNREIE
TMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPASDDAQLYYGNFEEEFRNFWNNS
NYYCVLAGLGLVPDAPSPRLAHMPQATESVDEVFGAVKDRQRNLLETLPSLHEFLRQQHGR*

6-TLP (parent):

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEAT**T**PNLQTAFFDFLGIPEDEWMRECNASY
KVAIKFINWRTAGEGTSEARELDGGPDHFYH**PL**GLLKYHEQIPLSHYWFDR**L**YRGKTVEPFDYACYKEPVILDANRSPRR

LDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTATGRVFDADLFVDCSGFRGLLIN
KAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYSSRFATEDEAVREF
CEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQLVKHFPDKSFNPVLTARFNREI
ETMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAPDDAQLYWGNFEEEFRNPWN
NSSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLRQQHGR*

1-L

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATTPNLQTAFFDFLGIPEDEWMRECNASY
KVAIKFINWRTAGEGTSEARELDGGPDHFYHPLGLLKYHEQIPLSHYWFDRLYRGKTVEPFDYACYKEPVILDANRSPRR
LDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTATGRVFDADLFVDCSGFRGLLIN
KAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYSSRFATEDEAVREF
CEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQLVKHFPDKSFNPVLTARFNREI
ETMFDDTRDLIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAPDDAQLYWGNFEEEFRNPWN
NSSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLRQQHGR*

2-RSG

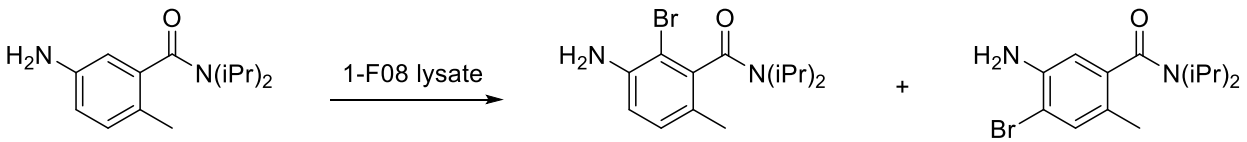
MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATTPNLQTAFFDFLGIPEDEWMRECNASY
KVAIKFINWRTAGEGTSEARELDGGPDHFYHPLGLLKYHEQIPLSHYWFDRLYRGKTVEPFDYACYKEPVILDANRSPRR
LDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTATGRVFDADLFVDCSGFRGLLIN
KAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYSSRFATEDEAVREF
CEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQLVKHFPDKSFNPVLTARFNREI
ETMFDDTRDLIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAPDDAQLYRGNFEEEFRNPWSN
GSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLRQQHGR*

3-T

MGSSHHHHHHSSGLVPRGSHMSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATTPNLQT
AFFDFLGIPEDEWMRECNASYKVAIKFINWRTAGEGTSEARELDGGPDHFYHPLGLLKYHEQIPLSHYWFDRLYRGKTV
EPFDYACYKEPVILDANRSPRRLDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTAT
GRVFDADLFVDCSGFRGLLINKAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGR
FGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQ
LVKHFPDKSFNPVLTARFNREIETMFDDTRDLIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAP
DDAQLYRGNFEEEFRNPWTNGSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLRQ
QHGR*

3.4.3 – Biocatalysis

Plate screening of 1-F08 enzyme libraries



B1 B1a B1b

Clarified lysate was obtained as described in culture plate method 1. 50 µL of the clarified lysate was added via a multichannel pipette to a 96 well flat-bottom poly-styrene microtiter plate (Fisherbrand, Cat. No. 12565501). A solution of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 2.5 mM of substrate **B1** was prepared in reaction buffer (small molecule stock solution). A solution of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer (enzyme stock solution). 15 µL of the small molecule stock solution was added to the 96 well microtiter plates containing clarified lysate. 10 µL of the enzyme stock solution was then added to give a total reaction volume of 75 µL. The plate was sealed using an aluminum heat seal and incubated at 25 °C in for 20 hours using a Ika MTS 2/4 digital plate shaker set to 600 rpm.

The reactions were quenched by adding 75 µL methanol to the microtiter plates, which were then further incubated for 10 minutes to allow for denaturation of protein. The plates were centrifuged at 3600 rpm for 15 minutes to pellet the precipitated protein, and 50 µL of the supernatant was added to 75 µL of water in a 0.22 µm plate filter (Agilent, part number 203980-100) and filtered into a new microtiter plate via centrifugation at 3600rpm for 15 minutes. The microtiter plate containing the filtered reaction mixtures was heat sealed as before and analyzed via UHPLC using method A with integrations taken from the 230 nm chromatogram. These data are summarized in **Figure 3.9**, and the full mutation list from WT 1-F08 for the top 5 hits can be found in **Figure 3.3**. Note that only the first 500 variants of the 800 member library were screened before the lack of correlation between purified and lysate data was observed.

Figure 3. 9: Results of 1-F08 library screening

P4 o/p

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.858471	0	0	0.791454	0	0	0.95532	0	0	0	0	0
B	1.341071	0	0	0	0	0	1.293705	0	0	0	0	0
C	0	0	0.731666	0	0	1.244221	1.005277	0	0	0	0	0
D	0.891152	0	0	0	0	0	0	0	0	0	0	0
E	0	1.208498	0	1.171733	0	0	1.354152	1.048291	0	0	0	0.959192
F	0	0	0	0.974629	0.845643	0	0.909365	0	0	0	0	0
G	0.760313	0	0	1.190871	0	0	0	0	0	0	0	0
H	0.652469	0	0	0.873262	0	0	0	1.000833	0	0.781967	0	1.111393

P4 activity

	1	2	3	4	5	6	7	8	9	10	11	12
A	2.248372	0.149942	0	3.48682	0	0	1.719989	0	0	0	0	0
B	0.833744	0	0	0	0	0	0.313401	0	0	0.336254	0	0
C	0	0	0.573892	0	0	0.311675	2.306334	0	0	0.056396	0.255286	0
D	0.989704	0.250467	0	0.094761	0	0	0	0	0	0	0	0
E	0	0.596032	0	11.10616	0	0	0.334247	0.280304	0	0	0	1.170693
F	0	0	0	2.112707	2.48684	0	2.394246	0.134776	0	0	0	0
G	0.623218	0	0	0.298858	0	0	0	0	0	0	0.174371	0
H	0.391025	0.081884	0.09233	2.377094	0	0	0	1.376005	0	2.612078	0	0.78905

P5 o//p

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.897458	1.187433	0	0.919306	0	0.991262	0	0	0	0	0	0
B	1.224767	0	0	0	1.231612	0	1.009717	1.2916	0	1.526937	0	0
C	0	0	0.930705	0	0	0	0	0	0	0	1.714	0
D	0	1.375801	0	0	0.99268	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	1.059579	1.381819	0	0	0.97684	0.90521	0	0	0
G	0	0	0	0	0.863384	0	1.084454	0	0.982055	0	0	0
H	1.242038	0	0	0	0	0	0	1.243697	0	0	0	0

P5 activity

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.909797	1.144699	0	1.325243	0	1.014809	0	0	0	0	0	0
B	0.318115	0	0	0	0.805784	0	1.036399	0.721713	0	0.227296	0	0
C	0	0.083468	1.068068	0	0	0	0.104228	0	0	0	1.228522	0
D	0	0.757919	0	0	1.322235	0	0	0	0	0	0	0
E	0	0	0	0	0.142323	0	0	0	0	0	0	0.119677
F	0	0	0	0.895534	0.665063	0	0.095103	1.296769	1.191728	0	0	0
G	0	0	0	0	1.151557	0	1.010112	0	1.265877	0	0	0
H	0.895228	0	0	0	0	0	0	0.744774	0	0	0	0

P6 o/p

	1	2	3	4	5	6	7	8	9	10	11	12
A	0	0	0	0	0	0	0	0	0	0	1.269254	0
B	0	0	0	0.794317	0	0	0.98335	0.867269	0	0.754715	0	0
C	1.019927	0	0.994137	0	0	0	0	0	2.896031	0	0	1.210179
D	0	0.944862	0	1.412608	0	0	0	0	0	0	0	0
E	0	0	1.275316	1.314439	0	0	0	1.021609	0	0	0	0
F	0	0	0	1.022512	0	0	2.289282	0.942057	0	0	4.166036	0
G	0	0	0	0	0	0	1.141276	1.082126	0	1.043659	0	0
H	0	0	0	0	0	0	0.989247	0	0	0	5.823049	0

P6 activity

	1	2	3	4	5	6	7	8	9	10	11	12
A	0	0	0	0	0	0.478567	0	0	0	0	0.536005	0
B	0	0	0.22837	0.838794	0	0	0.945439	1.030092	0	1.047187	0	0
C	0.922855	0	1.056372	0	0	0	0.114187	0	3.352892	0	0	0.514611
D	0	1.048686	0	0.428675	0	0	0	0	0	0	0	0
E	0	0	1.282671	0.372775	0	0	0	1.685575	0	0	0	0
F	0	0.281602	0	0.998189	0.094557	0.101713	2.224065	1.959507	0.101023	0.075704	2.279602	0
G	0.136524	0	0	0	0	0	1.373092	0.424328	0	1.461194	0	0
H	0	0	0	0	0	0	1.037501	0	0	0	2.086467	0

P7 o/p

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.909645	0	0.934918	0	0	0	0	1.11607	0	0.993605	0	0
B	0	0	0	0	0	0	1.058861	0	0	0	1.769199	0
C	0	0	0.980981	0	0	0	0	0	0	1.110567	0	0
D	0	0	0	0.981428	0	0	0	1.128532	0.869957	0	0	0.945092
E	0	0	0	1.298011	0	0	0	0	0	0	0	0
F	0	0	0	0.960158	0	0	0	0	0	0	0	0
G	0.839327	0	0	0	0.833475	1.001523	0	0	0	0	0	0
H	0	1.833195	0.98911	0	0	0	0.898524	0	0.939956	0	0.84493	0

P7 activity

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.87778	0	0.695395	0	0	0	0	0.119379	0	0.660396	0	0
B	0.028446	0	0	0	0	0	0.882425	0	0	0	0.12194	0
C	0	0	1.147483	0	0	0	0	0	0	0.119168	0	0
D	0	0	0	0.864579	0	0	0	0.241058	1.171176	0	0	0.782232
E	0	0	0	0.40757	0	0	0	0	0	0	0	0
F	0	0	0	0.970092	0	0.023233	0	0	0	0	0	0.047788
G	1.115697	0	0	0	1.53752	0.917289	0	0	0	0.026148	0	0
H	0	0.579592	0.784761	0	0.03497	0	0.870144	0	0.538622	0	0.296007	0

P8 o/p

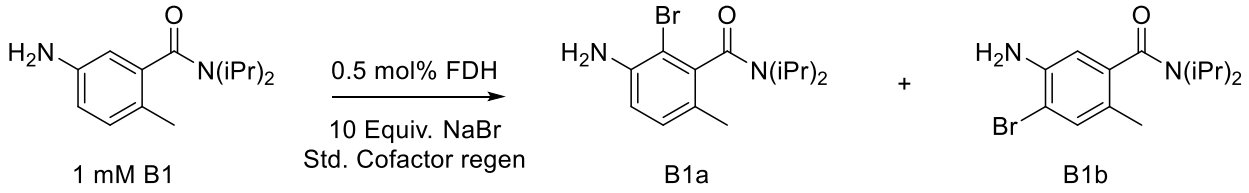
	1	2	3	4	5	6	7	8	9	10	11	12
A	0	0	0	0	0	0	1.109582	0	1.21921	0	0	0
B	0	0	0	0	0	0	0.756372	0	0	0	0	0
C	0	1.039806	1.085627	0	0.788138	0	0	0.895282	0	0.810614	0	0
D	0	0	0	0	0	0	0.729453	0	0.840803	0	0	0
E	0	0.861001	0	0	0	0	0	0	0	0	0	0
F	0	0	0	1.158001	0	0	1.042322	0	0	0	0	0
G	0	0	0	0	0	0	0	1.063635	0	0.994063	0	1.095273
H	0	1.210308	0	0	0	0	1.973742	0.748558	0	0	0	0.862918

P8 activity

	1	2	3	4	5	6	7	8	9	10	11	12
A	0	0	0	0	0	0	0.44057	0	0.763289	0	0	0
B	0	0	0	0	0	0	1.01584	0	0	0	0	0
C	0	0.742506	0.998906	0.239164	1.081758	0	0	1.135188	0	1.24745	0	0
D	0.293036	0.638826	0	0.384689	0	0	1.200952	0	1.09047	0	0	0
E	0	0.799323	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0.985255	0	0	0.663828	0	0	0	0	0
G	0	0	0	0	0	0	0	0.990557	0	0.977765	0	0.123891
H	0	0.118801	0	0	0	0	0.403865	0.514396	0	0	0	0.605995

Values in the 96-well plates are normalized for activity (defined as total conversion to product **B1a** using the TIC for quantification) and selectivity (defined as product **B1a/B1b**) compared to the intra-plate values for Wt-1-F08. WT-1-F08 is located at wells B7, C3, and F4 on each plate (wells are highlighted bold). Wells are shaded with the highest values shaded darker green to make hit identification easier.

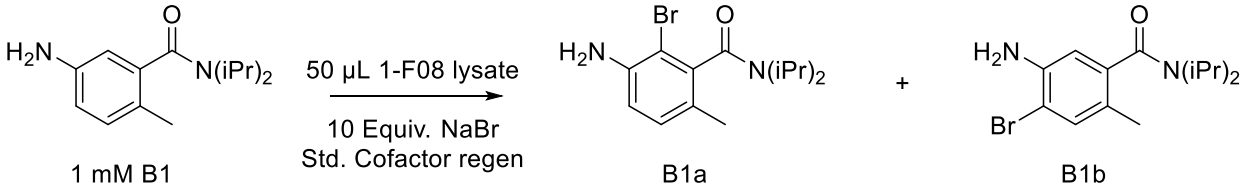
Screening of purified hits from 1-F08 lineage



1 mM B1 B1a B1b

1-F08 variants were expressed and purified according to the general procedure. Stock solutions of 100 µM purified FDH were prepared in storage buffer. An FDH stock solution was prepared by diluting 18.75 µL of 100 µM 1-F08 variant with 31.25 µL of reaction buffer, and 50 µL of this stock solution was added to a microtiter plate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 5 mM of substrate **1**, was prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the clarified lysate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. After 16 hours, the reactions were quenched and processed as previously described above. Aliquots were taken and analyzed and the results can be seen in **Figure 3.3C.** Reactions were analyzed by UHPLC-MS using method A with integrations taken from the 230 nm chromatogram.

Buffer optimization for 1-F08 biocatalysis



Clarified lysate was obtained as described in culture plate method 1, except several solutions of lysozyme were prepared in different buffers at different pHs, as seen in the figure below:

Figure 3. 10: Plate layout and pKas of buffers used in buffer optimization of 1-F08

Identity	Buffer range	pka
HEPES	6.8-8.2	7.45
Phosphate	various	7.21
MES	5.5-6.7	6.1
Tris	7.9-9.0	8.08
Tricine	7.4-8.8	8.15
PIPES	6.1-7.5	6.76
MOPS	6.5-7.9	7.2
CAPS	9.7-11.1	10.4

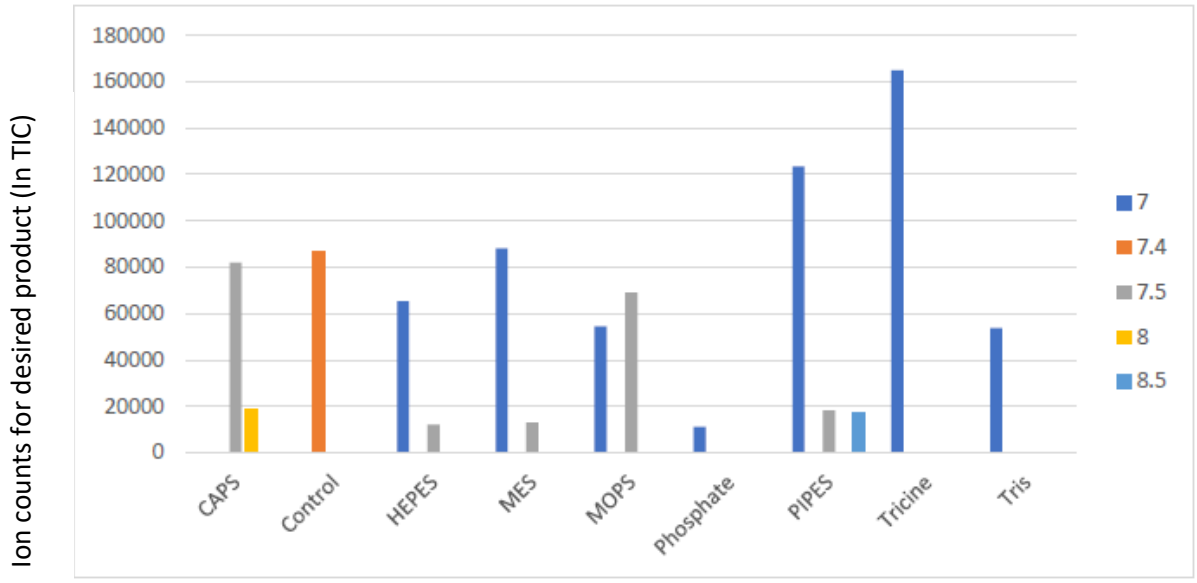
Plate	1	2	3	4
HEPES	7	7.5	8	8.5
Phosphate	7	7.5	8	8.5
MES	7	7.5	8	8.5
Tris	7	7.5	8	8.5
Tricine	7	7.5	8	8.5
PIPES	7	7.5	8	8.5
MOPS	7	7.5	8	8.5
CAPS	Positive control (25 mM HEPES)		8	8.5

50 µL of the clarified lysate was added via a multichannel pipette to a 96 well flat-bottom poly-styrene microtiter plate (Fisherbrand, Cat. No. 12565501). A solution of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 2.5 mM of substrate **B1** was prepared in reaction buffer (small molecule stock solution). A solution of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer (enzyme stock solution). 15 µL of the small molecule stock solution was added to the 96 well microtiter plates containing clarified lysate. 10 µL of the enzyme stock solution was then added to give a total reaction volume of 75 µL. The plate was sealed using an aluminum heat seal and incubated at 25 °C in for 20 hours using a Ika MTS 2/4 digital plate shaker set to 600 rpm.

The120eactionns were quenched by adding 75 µL methanol to the microtiter plates, which were then further incubated for 10 minutes to allow for denaturation of protein. The plates were centrifuged at 3600 rpm for 15 minutes to pellet the precipitated protein, and 50 µL of the supernatant was added to 75 µL of water in a 0.22 µm plate filter (Agilent, part number 203980-100) and filtered into a new microtiter plate via centrifugation at 3600rpm for 15 minutes. The microtiter plate containing the filtered reaction

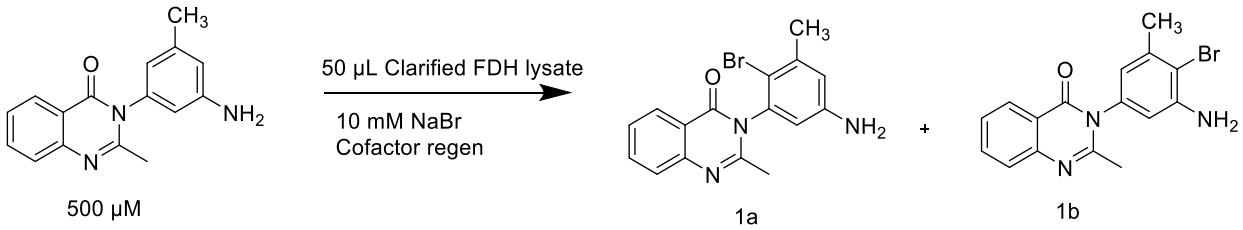
mixtures was heat sealed as before and analyzed via UHPLC using method A with integrations taken from the 230 nm chromatogram. These data are summarized in **Figure 3.11**.

Figure 3. 11: Results of buffer optimization study with enzyme 1-F08.



Interpretation: The enzyme is more active in more acidic buffers.

Plate screening of the RebH and genome mining compilation plates



Clarified lysate was obtained as described in culture plate method 1. 50 µL of the clarified lysate was added via a multichannel pipette to a 96 well flat-bottom poly-styrene microtiter plate (Fisherbrand, Cat. No. 12565501). A solution of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 2.5 mM of substrate **1** was prepared in reaction buffer (small molecule stock solution). A solution of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer (enzyme stock solution). 15 µL of the small molecule stock solution was added to the 96 well microtiter plates containing clarified lysate. 10 µL of the enzyme stock solution was then added to give a total reaction volume of 75 µL. The plate was sealed using an aluminum heat seal and incubated at 4 °C in a constant temperature room for 48 hours using a Ika MTS 2/4 digital plate shaker set to 600 rpm.

The reactions were quenched by adding 75 µL methanol to the microtiter plates, which were then further incubated for 10 minutes to allow for denaturation of protein. The plates were centrifuged at

3600 rpm for 15 minutes to pellet the precipitated protein, and 50 µL of the supernatant was added to 75 µL of water in a 0.22 µm plate filter (Agilent, part number 203980-100) and filtered into a new microtiter plate via centrifugation at 3600rpm for 15 minutes. The microtiter plate containing the filtered reaction mixtures was heat sealed as before and analyzed via UHPLC using method A with integrations taken from the 230 nm chromatogram. These data are summarized in **Figure 3.9**, and the full mutation list from WT RebH for the top 5 hits can be found in **Figure 3.10**.

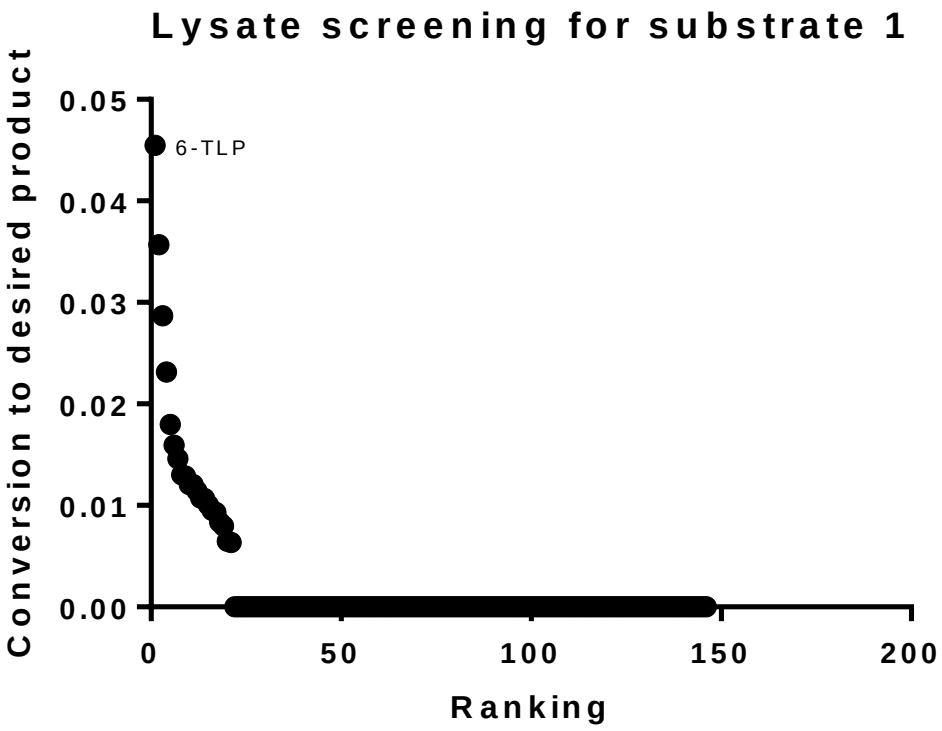
Example reaction set up:

Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	10.25	25 mM
NaBr	1.5 M	0.5	10 mM
Glucose	1 M	1.5	20 mM
NAD	10 mM	0.75	0.1 mM
FAD	10 mM	0.75	0.1 mM
Substrate	30 mM	1.25	0.5 mM
	Total Volume SM mix per reaction	15	

Enzyme mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	3.7	25 mM
GDH	180 U/mL	3.75	9 U/mL
Catalase	2000 U/mL	1.31	35 U/mL
MBPF	150 µM	1.25	2.5 µM
	Total volume enzyme mix per reaction	10	

Total volume clarified lysate per reaction (µL)	Volume reaction buffer added per reaction (µL)
50	0

Figure 3. 12: Lysate screening of the RebH and genome mined compilation plates for the halogenation of substrate 1

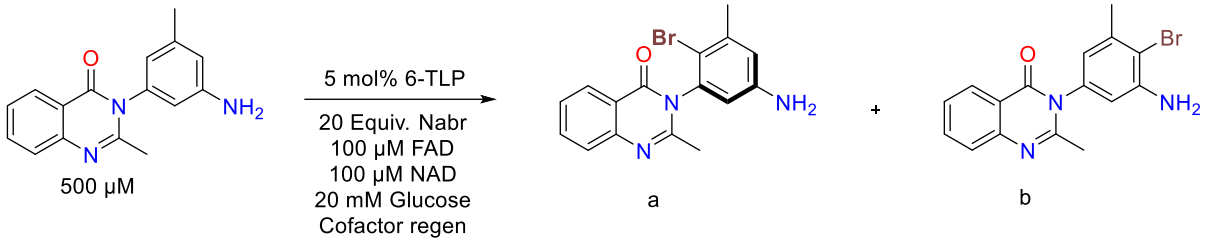


Data callout for the most active variant, 6-TLP, is included.

Table 3. 1: Table of top 5 active variants with sequences from wt-RebH

Activity rank	RebH variant name	Mutations from WT RebH
1	6-TL + L465P (6-TLP)	I52T, S110P, F111L, S130L, N166S, L380F, S448P, Y455W, L465P, N470S, Q494R, R509Q
2	3-SS	S2P, M71V, G112S, K145M, N467T, N470S
3	3-S	S2P, M71V, K145M, N467T, N470S
4	2RFQ-F111S	F111S, L380F, S448P, N470S, Q494R, R509Q
5	6-TL + L465W	I52T, S110P, F111L, S130L, N166S, L380F, S448P, Y455W, L465W, N470S, Q494R, R509Q

Preparative scale bioconversion of 6-TLP with substrate 1 to determine enantioselectivity

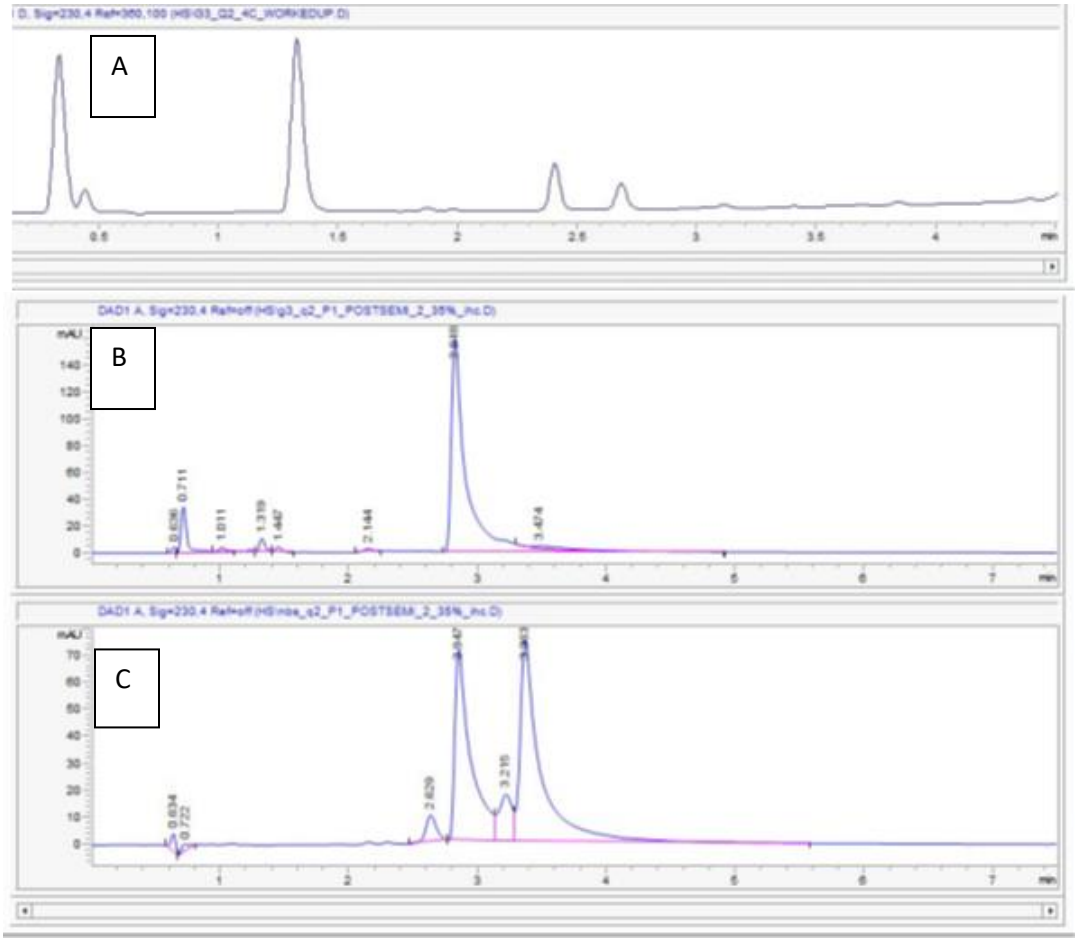


FDH variant 6-TLP was expressed and purified using the general procedure provided above. Reactions were conducted in 20 mL plastic scintillation vials secured to an Ika MTS 2/4 digital plate shaker. Unless otherwise noted, all reagents were prepared in reaction buffer (25 mM HEPES, pH 7.4). Each scintillation vial was charged with 3.21 mL of reaction buffer, followed by 26.6 µL of a 1.5 M solution of NaBr (final concentration 10 mM, 20 equiv.), 80 µL of a 1 M solution of glucose (final concentration 20 mM, 40 equiv.), 40 µL of both NAD and FAD from a 10 mM stock solution, and 66.6 µL of a 30 mM DMSO stock solution of substrate **1** (500 µM final concentration). After all small molecule components were added, 200 µL of a 500 µM stock solution of 6-TLP (final concentration 25 µM, 5 mol% relative to **1**), 70 µL of a 2000 U/mL stock solution of catalase (final concentration 35 U/mL), and 66.6 µL of a 150 µM stock solution of MBP-RebF (final concentration 2.5 µM) were added. Reactions were initiated by the addition of 200 µL of a 180 U/mL stock solution of GDH (final concentration 9 U/mL) to give a final volume of 4 mL in each scintillation vial. Ten reactions were assembled in this way, and the vials were quickly covered with an AeraSeal breathable plate seal. The reactions were shaken at 600 rpm with temperature maintained at 4 °C in a constant temperature room for 48 hours. The contents of the vials were pooled together into an Erlenmeyer flask, saturated with NaCl, and the pH of the mixture was adjusted to ~1 using 1 M HCl. 20 mL DCM and a stir bar were added to the flask, and the mixture was mixed vigorously to precipitate remaining protein. The resulting mixture was then filtered through celite to remove precipitated protein, and the pH of the filtrate was adjusted to ~10 using 1 M NaOH. The aqueous layer was extracted 3x with 20 mL DCM, and the organic layers were combined, washed with brine, and then dried over magnesium sulfate. The solvent was then removed by rotary evaporation under reduced pressure. The resulting oil was then dissolved in DMSO and filtered using a 0.22 micron filter (WHEATON MicroLiter Syr Filter, 13 mm, 0.2 um nylon, part number F13-2020) into a reduced-volume HPLC vial (Thermo Scientific™ part number C4011-16). Product **1a** was purified by semi-preparative HPLC using Semiprep method 1 listed in the general section. Achiral and chiral chromatograms can be seen in **Figure 3.11** below.

Example reaction:

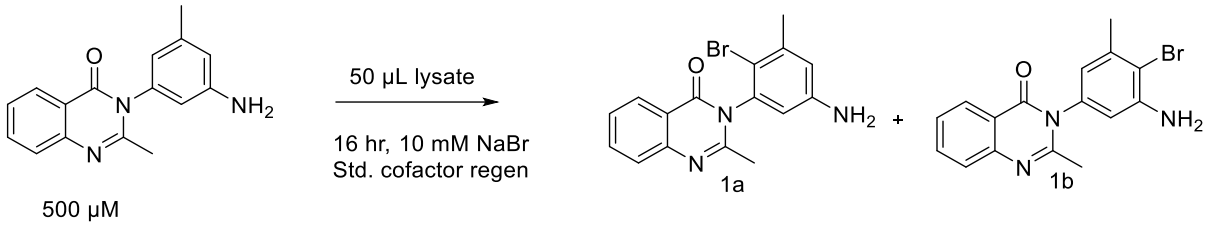
Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	3210	25 mM
NaBr	1.5 M	26.6	10 mM
Glucose	1 M	80	20 mM
NAD	10 mM	40	0.1 mM
FAD	10 mM	40	0.1 mM
Substrate	30 mM	66.6	0.5 mM
MBPF	150 µM	66.6	2.5 µM
Catalase	2000 U/mL	70	35 U/mL
6-TLP	500 µM	200	25 µM
GDH	180 U/mL	200	9 U/mL
	Total volume	4000	

Figure 3. 13: UHPLC and SFC analysis of 6-TLP bioconversions



A.) Crude reaction mixture after 48-hour incubation of 6-TLP with substrate **1** as described above. The peak with retention time 2.4 min corresponds to product **1a**, and retention time 2.7 min was identified as **1b**. B.) Chiral analysis of product **1a** generated from (Top): 6-TLP after purification using semiprep method 1 ©: **1a** generated via NBS reaction purified using semiprep method 1.

Evaluation of the error-prone PCR library of 6-TLP



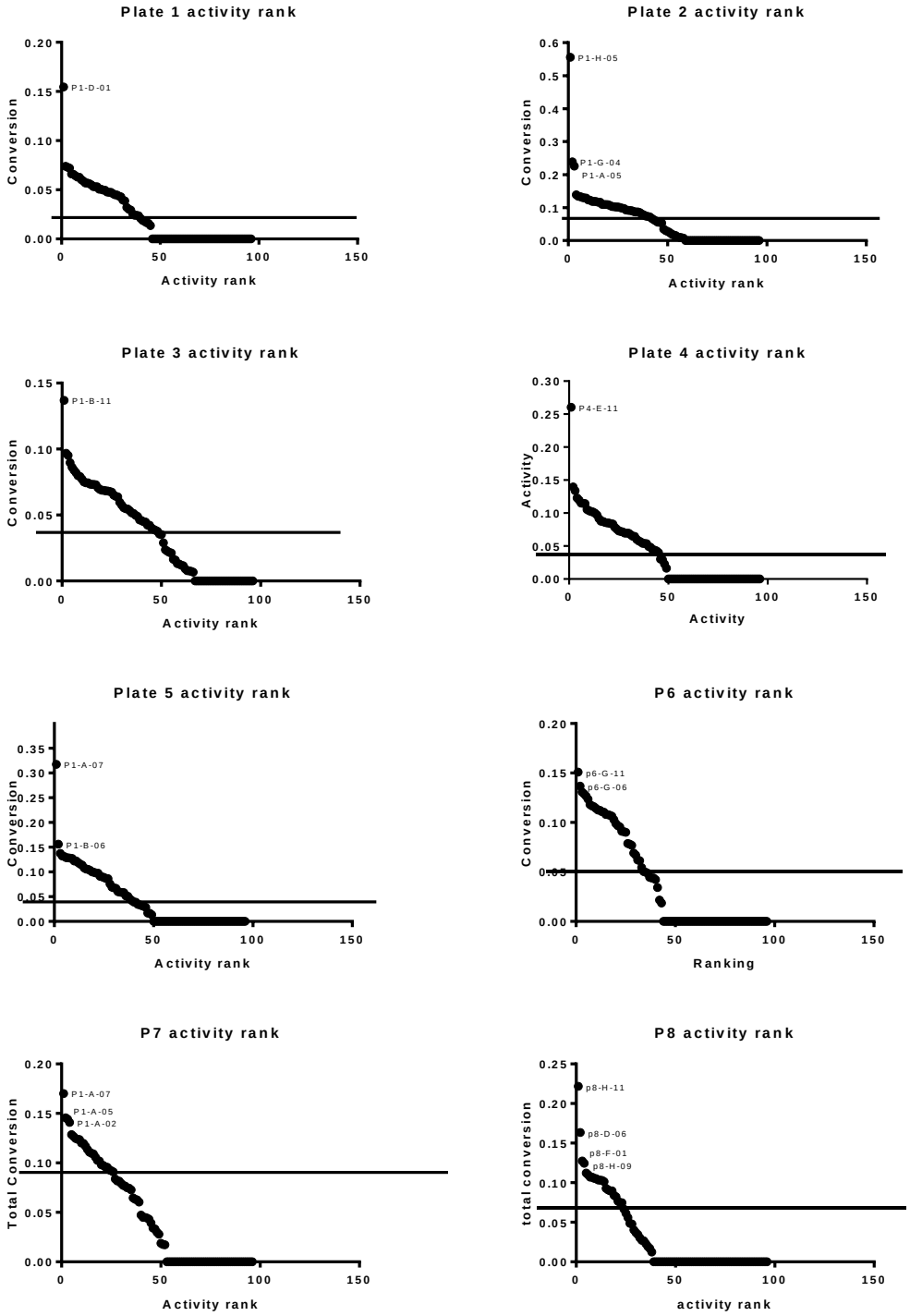
Library preparation including PCR sequences, primer identities, and the conditions used are included in the cloning section of this SI. Clarified lysate was obtained using culture plate method 2. 50 µL of clarified lysate was added to a microtiter plate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 2.5 mM of substrate **1** was prepared in reaction buffer (25 mM HEPES, pH 7.4). An enzyme stock solution consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the clarified lysate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours using a Heidolph Tetramax 1000 set to rotate at 700 rpm. After the reaction, the reactions were quenched and filtered using the same process as previously described. The plates were analyzed by UHPLC using method A with integrations taken from the 230 nm chromatogram. These data are summarized in **Figure 3.12** below.

Example reaction:

Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	10.25	25 mM
NaBr	1.5 M	0.5	10 mM
Glucose	1 M	1.5	20 mM
NAD	10 mM	0.75	0.1 mM
FAD	10 mM	0.75	0.1 mM
Substrate	30 mM	1.25	0.5 mM
	Total Volume SM mix per reaction	15	
Enzyme mix	**Stock concentration**	**Volume from concentrated stock (µL)**	**Final concentration in reaction**
25 mM HEPES, pH 7.4	25 mM	3.7	25 mM
GDH	180 U/mL	3.75	9 U/mL
Catalase	2000 U/mL	1.31	35 U/mL
MBPF	150 µM	1.25	2.5 µM
	Total volume enzyme mix per reaction	10	

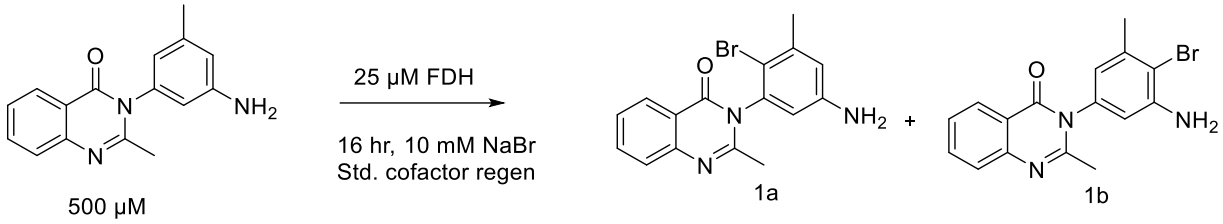
Total volume clarified lysate per reaction (µL)	Volume reaction buffer added per reaction (µL)
50	0

Figure 3. 14: Screening of an error prone library of 6-TLP for improved activity on substrate 1



Black bar denotes the average of 6-TLP.

Screening of purified enzymes from the error-prone library of 6-TLP

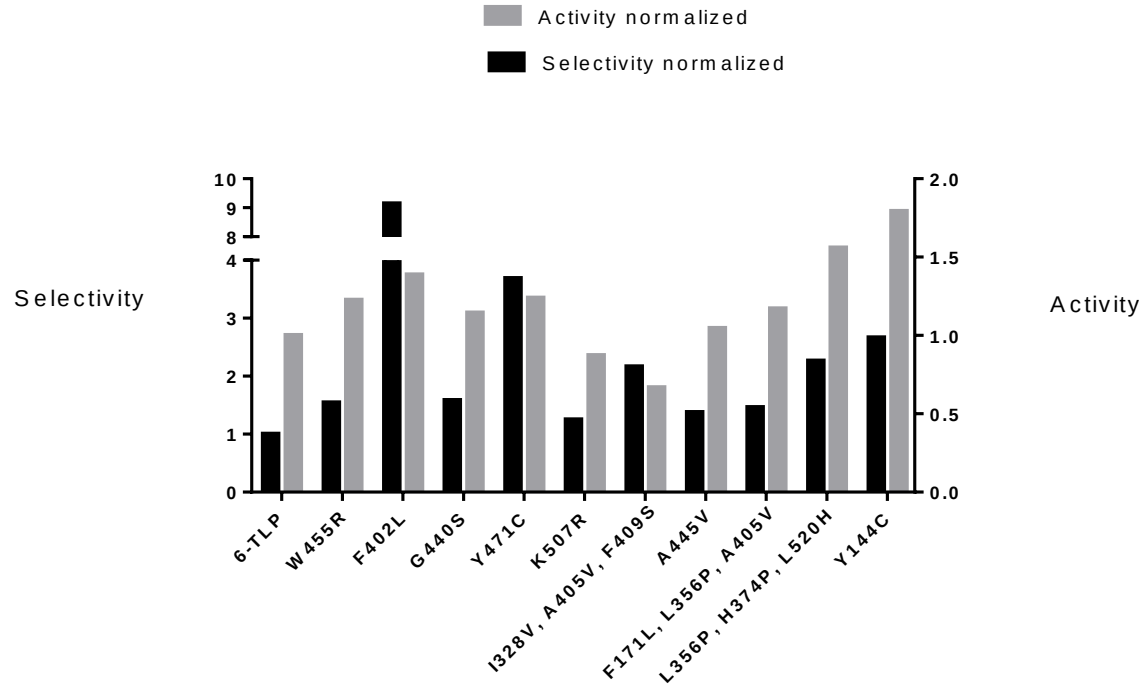


6-TLP variants were expressed and purified according to the general procedure. Stock solutions of 100 µM purified FDH were prepared in storage buffer. An FDH stock solution was prepared by diluting 18.75 µL of 100 µM 6-TLP variant with 31.25 µL of reaction buffer, and 50 µL of this stock solution was added to a microtiter plate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 2.5 mM of substrate **1**, was prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the clarified lysate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. After 16 hours, the reactions were quenched and processed as previously described above. Reactions were analyzed by UHPLC using method A with integrations taken from the 230 nm chromatogram. These data are summarized in **Figure 3.13** below.

Example reaction set up (one 75 µL reaction):

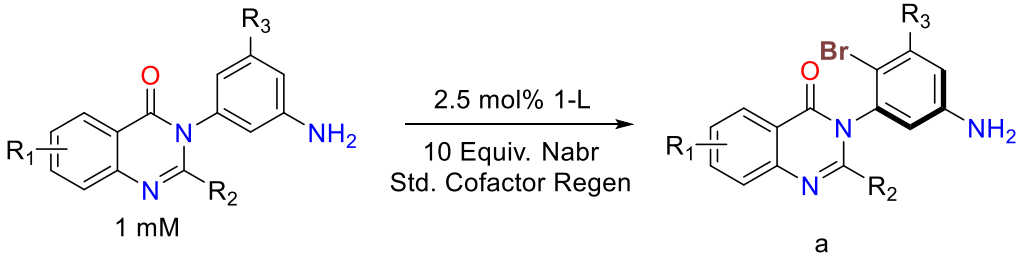
Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	10.25	25 mM
NaBr	1.5 M	0.5	10 mM
Glucose	1 M	1.5	20 mM
NAD	10 mM	0.75	0.1 mM
FAD	10 mM	0.75	0.1 mM
Substrate	30 mM	1.25	0.5 mM
	Total Volume SM mix per reaction	15	
Enzyme mix	**Stock concentration**	**Volume from concentrated stock (µL)**	**Final concentration in reaction**
25 mM HEPES, pH 7.4	25 mM	3.7	25 mM
GDH	180 U/mL	3.75	9 U/mL
Catalase	2000 U/mL	1.31	35 U/mL
MBPF	150 µM	1.25	2.5 µM
	Total volume enzyme mix per reaction	10	
FDH Mix	**Stock concentration**	**Volume from concentrated stock (µL)**	**Final concentration in reaction**
25 mM HEPES, pH 7.4	25 mM	31.25	25 mM
6-TLP var.	100 µM	18.75	25 µM
	Total volume FDH mix per reaction	50	

Figure 3. 15: Evaluation of purified hits from 6-TLP error prone library.



F402L (1-L) was chosen as the parent for the next round of evolution due to increased activity and selectivity. Selectivity and activity were normalized to 6-TLP. Activity was the total conversion of substrate **1** into products **1a** and **1b** as determined by AUC measured at 230 nm. Selectivity is given by the ratio of **1a**/**1b** as determined by the AUC measured at 230 nm.

Evaluation of 1-L substrate scope



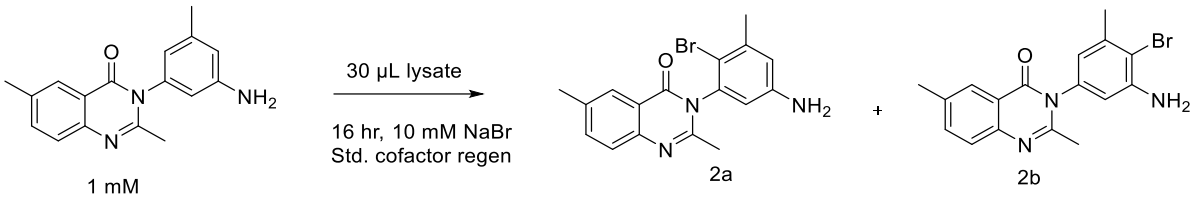
1-L was expressed and purified according to the general procedure. Reactions were conducted in 20 mL plastic scintillation vials. Unless otherwise noted, all reagents were prepared in reaction buffer (25 mM HEPES, pH 7.4). Each scintillation vial was charged with 3.103 mL of reaction buffer, followed by 26.6 µL of a 1.5 M solution of NaBr (final concentration 10 mM, 10 equiv.), 80 µL of a 1 M solution of glucose

(final concentration 20 mM, 20 equiv.), 40 µL of both NAD and FAD from a 10 mM stock solution (100 µM final concentration, 0.1 Equiv.), 40 µL of L-glutathione from a 100 mM stock solution (1 mM final concentration, 1 Equiv.) and 133.2 µL of a 30 mM DMSO stock solution of substrate (1 mM final concentration). After addition of small molecules, 200 µL of a 500 µM stock solution of 1-L was added (final concentration 25 µM, 2.5 mol% relative to substrate), followed by 70 µL of a 2000 U/mL stock solution of catalase (final concentration 35 U/mL), then 66.6 µL of a 150 µM stock solution of MBP-RebF (final concentration 2.5 µM). Reactions were initiated by the addition of 200 µL of a 180 U/mL stock solution of GDH (final concentration 9 U/mL) to a final volume of 4 mL in each scintillation vial. 2 vials were set up for each substrate in this way and were quickly covered with a breathable plate seal. The vials were shaken in a VWR 1585 shaking incubator at 200 rpm and 25 °C as measured by a thermometer for 20 hours. After this time had passed, the contents of the vials were pooled together into an Erlenmeyer flask, saturated with NaCl, and the pH of the mixture was adjusted to ~1 using 1 M HCl. 10 mL DCM was added to the reaction along with a stir bar, and the samples were mixed vigorously to precipitate remaining protein. The resulting mixture was then filtered through celite to remove precipitated protein, and the pH of the filtrate was adjusted to ~10 using 1 M NaOH. The aqueous layer was extracted 3x with 20 mL DCM, and the combined organic extracts were washed with brine and dried over magnesium sulfate. The solvent was then removed by rotary evaporation under reduced pressure. Conversion was determined by NMR-integrations of starting material compared to product in the crude ^{1}H-NMR spectrum. After conversion determination via crude NMR, the reaction was dissolved in DMSO and processed as previously described for purification via semi-preparative HPLC. HPLC methods used to purify the products are listed in the general section. Chiral chromatography was performed using the general SFC method using methanol as mobile phase B with 0.1% DEA as an additive. These data are summarized in **Figure S6** below.

Example reaction set up:

Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	3103	25 mM
NaBr	1.5 M	26.6	10 mM
Glucose	1 M	80	20 mM
NAD	10 mM	40	0.1 mM
FAD	10 mM	40	0.1 mM
Substrate	30 mM	133.33	1 mM
Glutathione	100 mM	40	1 mM
MBPF	150 µM	66.6	2.5 µM
Catalase	2000 U/mL	70	35 U/mL
1-L	500 µM	200	25 µM
GDH	180 U/mL	200	9 U/mL
	Total volume	4000	

Plate screening of the 1-L error prone library



All cloning details including PCR components and the PCR cycles used are included in the cloning section. . Clarified lysate was obtained using culture plate method 2.. 30 µL of clarified lysate was added to a microtiter plate followed by 20 µL of reaction buffer. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 5 mM of substrate **2** from a 30 mM stock solution in DMSO was prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the clarified lysate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. The plates were quenched and processed as previously described and analyzed by UHPLC using method B with integrations done using 230 nm. Activity was the total conversion of substrate **2** into product **2a** as determined by AUC measured at 230 nm. These data are summarized in **Figure 3.14** below.

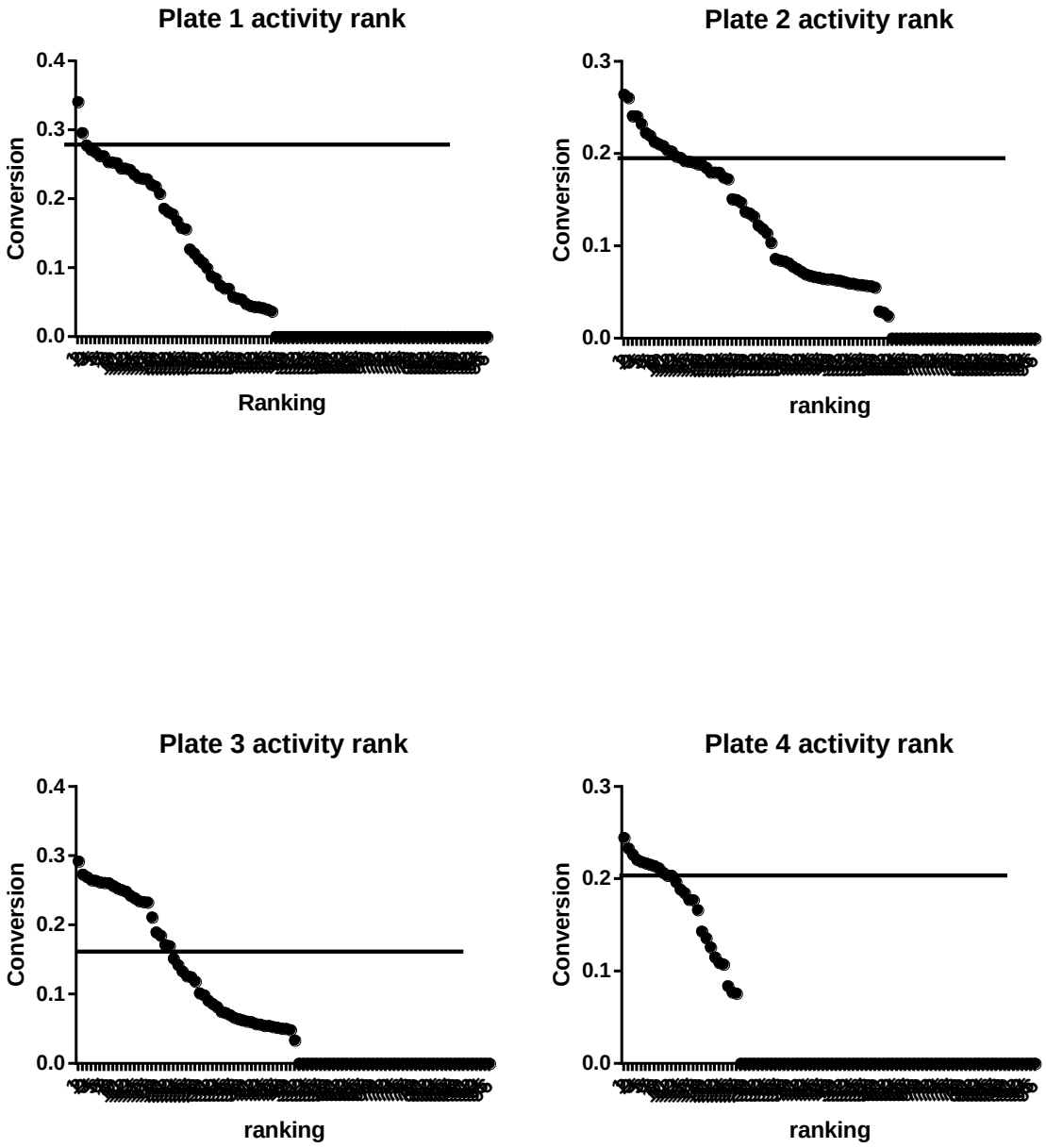
Example reaction set up:

Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	9	25 mM
NaBr	1.5 M	0.5	10 mM
Glucose	1 M	1.5	20 mM
NAD	10 mM	0.75	0.1 mM
FAD	10 mM	0.75	0.1 mM
Substrate	30 mM	2.5	1 mM
	Total Volume SM mix per reaction	15	

Enzyme mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	3.7	25 mM
GDH	180 U/mL	3.75	9 U/mL
Catalase	2000 U/mL	1.31	35 U/mL
MBPF	150 µM	1.25	2.5 µM
	Total volume enzyme mix per reaction	10	

Total volume clarified lysate per reaction (µL)	Volume reaction buffer added per reaction (µL)
30	20

Figure 3. 16: Results from error-prone library of 1-L



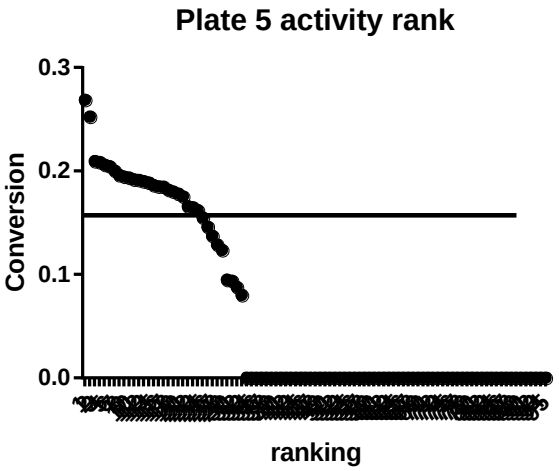


Plate 5 activity rank
Conversion
0.3
0.2
0.1
0.0
ranking

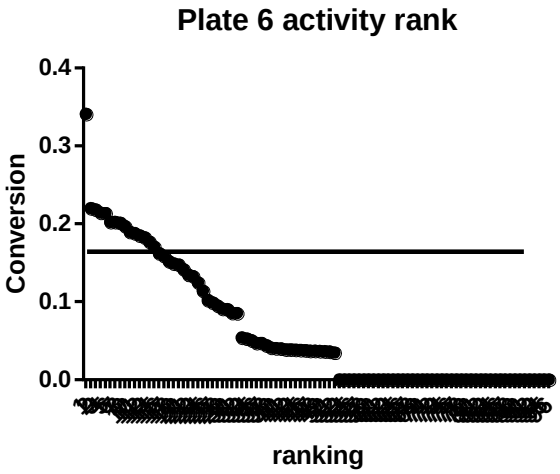


Plate 6 activity rank
Conversion
0.4
0.3
0.2
0.1
0.0
ranking

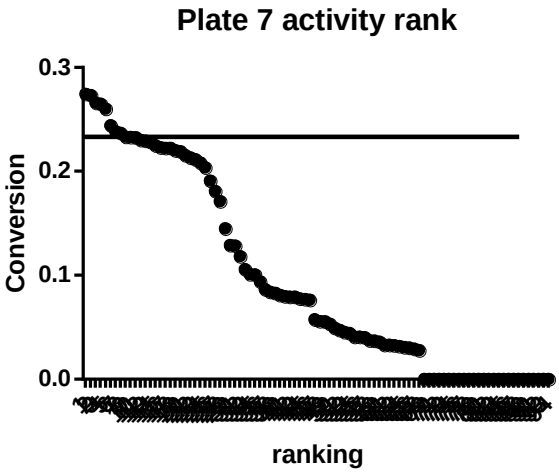


Plate 7 activity rank
Conversion
0.3
0.2
0.1
0.0
ranking

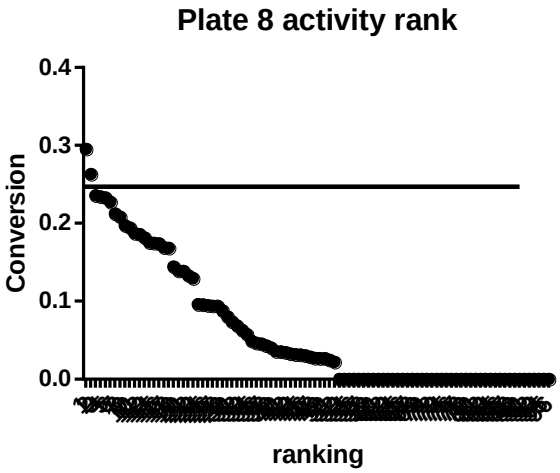
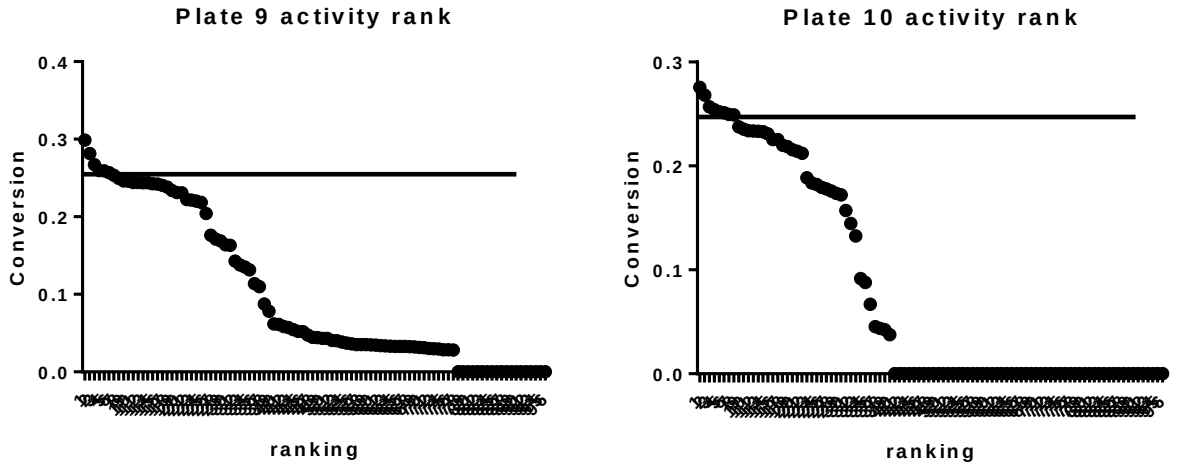
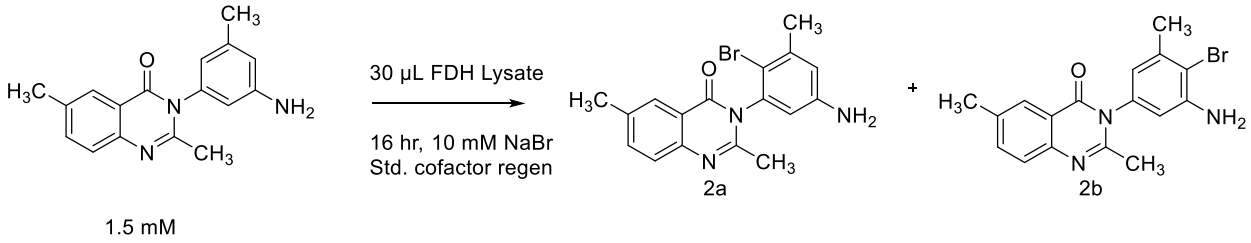


Plate 8 activity rank
Conversion
0.4
0.3
0.2
0.1
0.0
ranking



Conversion is measured by the ratio of product 2a/(2+2a) with integrations measured at 230 nm. For most variants, product 2b was only detected in trace quantities during the lysate screening, and therefore was not considered for determining conversion. The black bar denotes the average of the activity of 1-L.

Screening of 1-L compilation plate



Cultures of improved variants identified from the UHPLC screen were retrieved from refrigerated culture plates and used to inoculate individual wells of a 1 mL 96-well culture plate containing 300 µL of LB media supplemented with the appropriate antibiotics. The plate was incubated with shaking at 250 rpm for 16 hours at 37 °C. The next day, 50 µL of this overnight culture was used to inoculate two 2 mL deep well plates containing 1 mL of TB media with the appropriate antibiotics. Protein expression and lysis was performed as previously described. After lysis, 30 µL of lysate was added to a microtiter plate followed by 20 µL of reaction buffer. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 7.5 mM of substrate **2** from a 30 mM solution in DMSO was prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the

clarified lysate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours as previously described. After 16 hours had passed, the reactions were quenched and processed as previously described. Conversion was calculated as the AUC of product over AUC of starting material and product with integrations measured at 230 nm. The normalized conversion values are summarized in **Figure 3.15**, and the mutations from 1-L identified in the top 10 sequences are included in **Figure 3.16** found below.

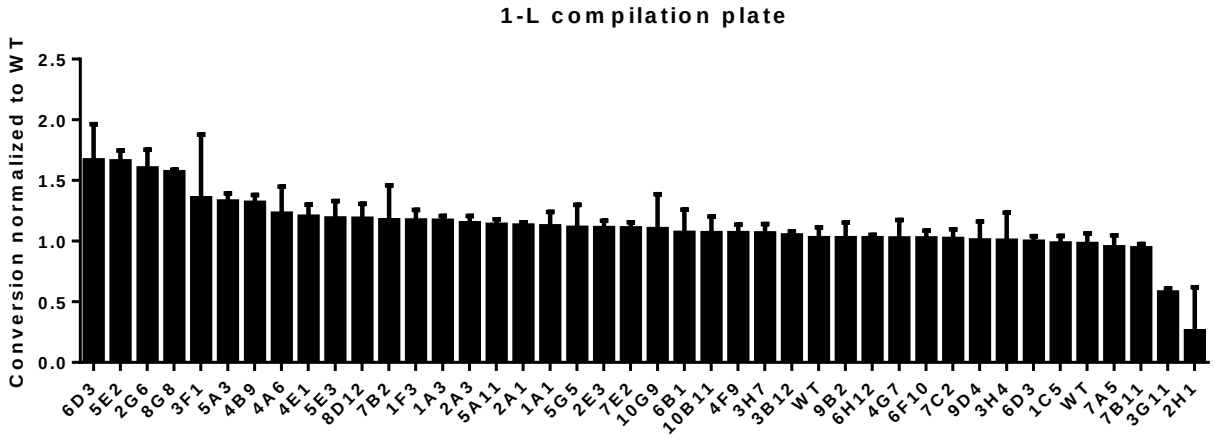
Reaction set up:

Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	7	25 mM
NaBr	1.5 M	0.5	10 mM
Glucose	1 M	1.5	20 mM
NAD	10 mM	0.75	0.1 mM
FAD	10 mM	0.75	0.1 mM
Substrate	30 mM	3.75	1.5 mM
Glutathione	100 mM	0.75	1 mM
	Total Volume SM mix per reaction	15	

Enzyme mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	3.7	25 mM
GDH	180 U/mL	3.75	9 U/mL
Catalase	2000 U/mL	1.31	35 U/mL
MBPF	150 µM	1.25	2.5 µM
	Total volume enzyme mix per reaction	10	

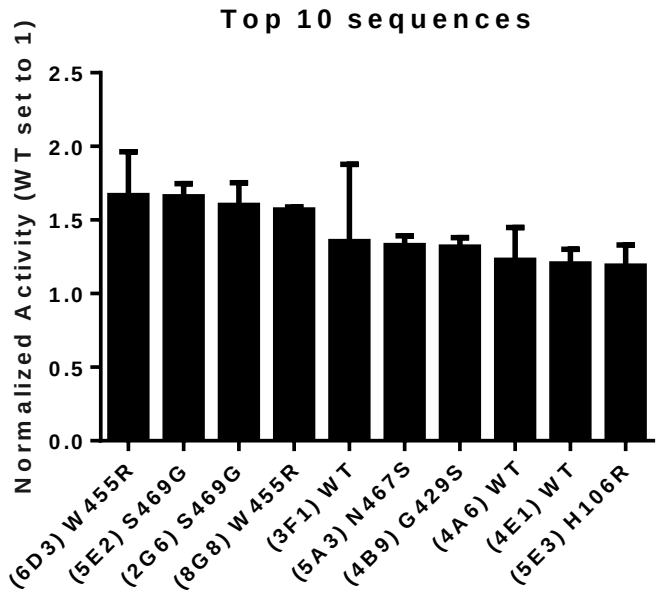
Total volume clarified lysate per reaction (µL)	Volume reaction buffer added per reaction (µL)
30	20

Figure 3. 17: Normalized data from screening the compilation plate of 1-L



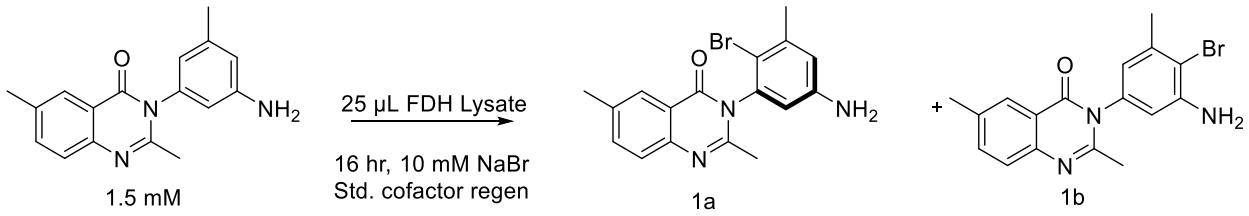
Average of two replicates for each variant is included, normalized to the value of 1-L (not shown).

Figure 3. 18: Top 10 Variants from the compilation plate of the 1-L library with sequences



Top 10 variants from **Figure 3.13** with the sequences shown.

Screening of 2-RSG NNK libraries



All cloning details used to generate the NNK libraries including primers, PCR components, and PCR cycles are included in the cloning portion of this SI. Clarified lysate was obtained using culture plate method 2. After lysis, 30 µL of lysate was added to a microtiter plate followed by 20 µL of reaction buffer. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 7.5 mM of substrate **6** from a stock solution of 30 mM in DMSO was prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the clarified lysate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours as previously described. After this time, the reactions were quenched and processed as previously described, and the reactions were analyzed by UHPLC using method B. Conversion was calculated as the

AUC of product over AUC of starting material and product with integrations measured at 230 nm. The results from this experiment are depicted in **Figure 3.17**.

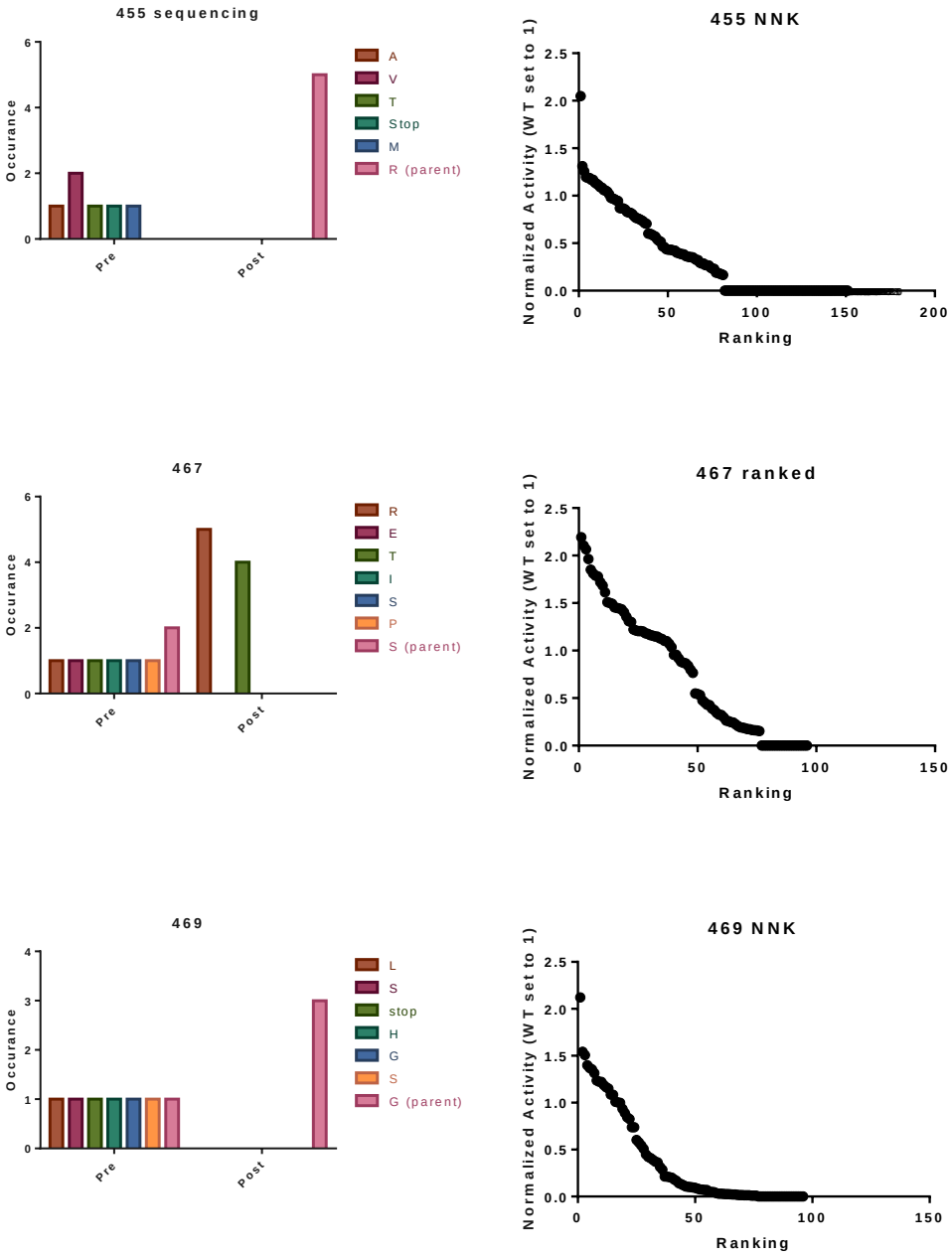
Example reaction:

Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	7	25 mM
NaBr	1.5 M	0.5	10 mM
Glucose	1 M	1.5	20 mM
NAD	10 mM	0.75	0.1 mM
FAD	10 mM	0.75	0.1 mM
Substrate	30 mM	3.75	1.5 mM
Glutathione	100 mM	0.75	1 mM
	Total Volume SM mix per reaction	15	

Enzyme mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	3.7	25 mM
GDH	180 U/mL	3.75	9 U/mL
Catalase	2000 U/mL	1.31	35 U/mL
MBPF	150 µM	1.25	2.5 µM
	Total volume enzyme mix per reaction	10	

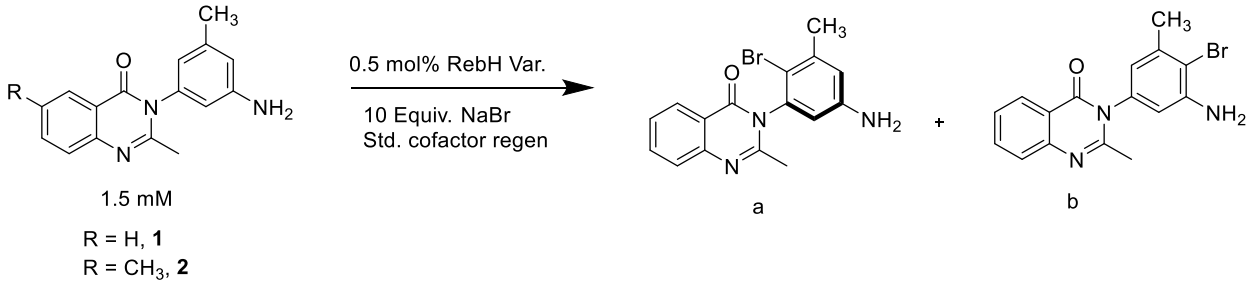
Total volume clarified lysate per reaction (µL)	Volume reaction buffer added per reaction (µL)
25	25

Figure 3. 19: Evaluation of the NNK library of 2-RSG for the halogenation of substrate 2



A.) Sequencing of random variants from R455NNK library and results of sequencing top hits after screening. Due to the high G-C content of the sequence around this region, we observed occasional cloning artifacts and screened 150 variants to account for this. B.) Normalized activity of R455NNK variants compared to 2-RSG. C.) Sequencing of random variants from S467NNK library and results of top hits after screening. D.) Normalized activity of S467NNK variants compared to 2-RSG. E.) Sequencing of random variants from the 469 NNK library and sequencing of the top hits after screening. F.) Normalized activity of the G469NNK library compared to 2-RSG.

Analytical-scale reactions for the 6-TLP lineage on substrates **1** and **2**



6-TLP and variants were expressed and purified according to the general procedure. Stock solutions of 75 µM purified FDH were prepared in storage buffer. An FDH stock was prepared by diluting 7.5 µL of 75 µM 6-TLP variant with 42.5 µL of reaction buffer in a microtiter plate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 7.5 mM of substrates **1** or **2** from 30 mM stock solutions in DMSO, were prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the FDH stock in the microtiter plate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. Conversion was calculated as the AUC of product over AUC of starting material and product with integrations measured at 230 nm.

Example reaction set up:

Small molecule mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	7	25 mM
NaBr	1.5 M	0.5	10 mM
Glucose	1 M	1.5	20 mM
NAD	10 mM	0.75	0.1 mM
FAD	10 mM	0.75	0.1 mM
Substrate	30 mM	3.75	1.5 mM
Glutathione	100 mM	0.75	1 mM
	Total Volume SM mix per reaction	15	

Enzyme mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	3.7	25 mM
GDH	180 U/mL	3.75	9 U/mL
Catalase	2000 U/mL	1.31	35 U/mL
MBPF	150 µM	1.25	2.5 µM
	Total volume enzyme mix per reaction	10	

FDH Mix	Stock concentration	Volume from concentrated stock (µL)	Final concentration in reaction
25 mM HEPES, pH 7.4	25 mM	42.5	25 mM
6-TLP var.	75 µM	7.5	7.5 µM
	Total volume FDH mix per reaction	50	

General procedure for prep-scale bioconversions with 3-T

30 mL and 40 mL bioconversions were conducted in 300 mL plastic beakers (Fisherbrand™ Polypropylene Disposable Beaker, catalog No. FB012915) covered with a breathable plate seal.

Substrates were prepared as either 30 mM (for 1 mM reactions) or 100 mM (for the 5 mM reaction of substrate **1**) stock solutions in DMSO. Substrate **12** was prepared as a 15 mM stock solution in DMSO. The beakers were charged with 25 mM HEPES, pH 7.4, followed by NaBr (10 mM in reaction buffer), NAD (100 µM final concentration, in reaction buffer), FAD (100 µM final concentration in reaction buffer), glucose (20 mM final concentration, prepared in reaction buffer), glutathione (1 mM final concentration, prepared in reaction buffer), catalase (35 U/mL final concentration), MBP-RebF (2.5 µM final concentration), substrate (1 mM final concentration, 500 µM for substrate **12**), 3-T (10 or 25 µM final concentration), and finally GDH (9 U/mL final concentration). The breathable plate seal was placed over the beaker, which was promptly placed in a Thermo MaxQ 8000 incubator using a 250 mL Erlenmeyer pedestal. The beaker was set to incubate at 150 rpm, 25 °C for 6 hours.

Example reaction set up:

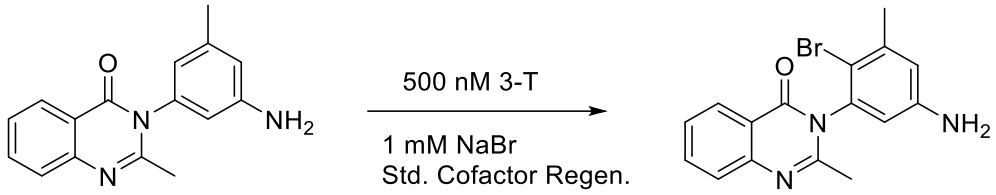
Component	Stock	Final	Volume (mL)
HEPES, pH 7.4	25 mM	25 mM	24
NaBr	1.5 M	10 mM	0.2
glucose	1 M	20 mM	0.6
NAD	10 mM	100 µM	0.3
FAD	10 mM	100 µM	0.3
substrate	30 mM	1 mM	1
GDH	180 U/mL	9 U/mL	1.5
catalase	2000 U/mL	35 U/mL	0.525
MBPF	150 µM	2.5 µM	0.5
FDH	500 µM	10 µM	0.6
Glutathione	100 mM	1 mM	0.4
		Total	30

Protein was denatured by the addition of 10% v/v of 1 M HCl. Sodium chloride was added to saturation, 50% v/v DCM was added, and the mixture was stirred vigorously for 10 minutes. The stir bar was removed, and the mixture was filtered through celite. The pH of the mixture was adjusted to pH >10 with 1 M NaOH, and the mixture was extracted 3x with 50% v/v DCM. The organic layers were combined and dried over magnesium sulfate. Once dry, the solution was filtered and concentrated by rotary evaporation. The resulting solids were then dissolved in DCM, adsorbed onto celite, and purified by reverse phase using the methods listed in the product characterization section of this SI. For products where the observed regioselectivity was less than 20:1, Biotage was insufficient for adequate purification and a subsequent silica column was required. When necessary, normal phase purification was performed using 3:1 ethyl acetate:hexanes with 1% TEA as an additive. The general procedure for normal phase purification is as follows: After Biotage purification, solvent was removed by rotary evaporation under reduced pressure. If the NMR of the product mixture after Biotage was less than 20:1, the product mixture was dissolved in DCM, then adsorbed onto silica which was dried thoroughly by rotary evaporation. This silica was then dry loaded onto a packed silica column, and product elution was monitored by TLC with UV detection. Fractions containing the major component of the reaction were collected and solvent was removed by rotary evaporation. Isolated yields and e.r. obtained from chiral chromatography are available in **Figure 3.6** in the main text. Chiral chromatograms are available in

the product characterization portion of this SI.

Calibration curve generation and kinetics performed with 3-T

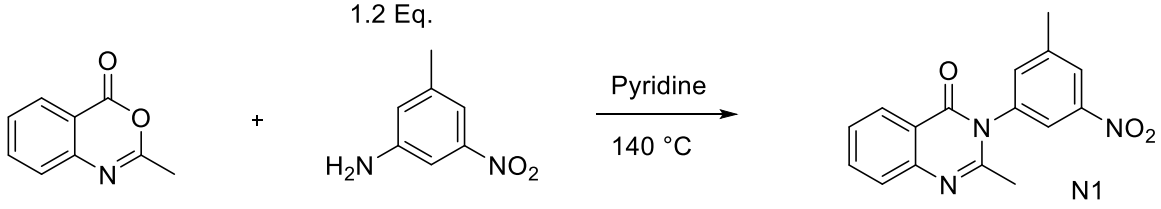
Enzyme aliquots of halogenase 3-T used for kinetics were always freshly thawed and then subsequently marked so a previously thawed sample was not used for kinetic calculations. Indole-3-acetic acid (IAA) was used as the internal standard for determining the yield of **1a** by generating a calibration curve from 50 nM to 25 µM **1a** with integrations taken from the 230 nm chromatogram (See Figure S28 for calibration curve). Both the calibration curve and kinetic measurements were taken in triplicate and are reported as the average measurement with standard deviation.



For kinetics, small molecule mixes including substrate **1** at 0.5, 1, 2.5, 5, 10, and 25 µM were prepared according to the previously described procedure. 3-T (50 µL, 500 nM final concentration) was added to a microtiter plate, followed by 15 µL small molecule mix, and the reactions were initiated by the addition of 10 µL of the enzyme mix as previously described and immediately set to incubate on a plate shaker as previously described. To stop the bioconversions for timepoint analysis, one volume (75 µL) of a quench solution consisting of 100 µM IAA (50 µM final concentration internal standard) in methanol was added to the microtiter plate to halt biocatalytic halogenation. Once all the timepoints had been quenched, the microtiter plates were spun down and and the reactions filtered as previously described. The reactions were processed using UHPLC method 2 with integrations taken from the 230 nm chromatogram. Yield of **1a** was calculated from the calibration curve in **Figure S28.** Kinetic data was obtained from the Michaelis-Menten plot in **Figure S29**.

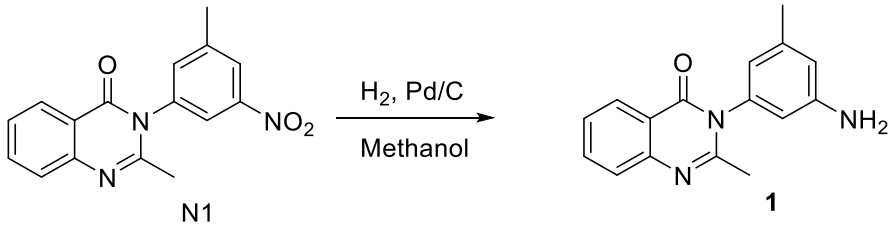
3.4.4 - Synthetic procedures

Scheme 3. 3: General quinazolinone synthesis A



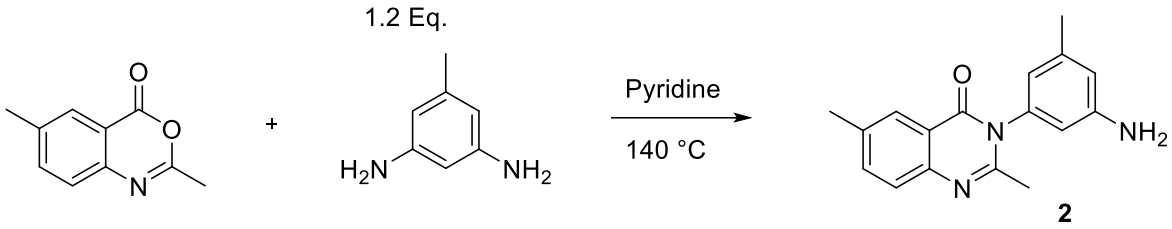
Solid 2-methylbenzoxazone (1 mmol, 161 mg) and 3-nitro-5-methylaniline (1.2 mmol, 184 mg) and a stir bar were added to a thick walled vial (Kimble Kimax 60700-2). 2 mL of pyridine was added, and the vial was tightly capped (Kontes Article No. 410119-2015 open cap with Kimble Article No. 749110-0022 valve) and left to stir at 140 °C for at least 24 hours. Over time, the reaction became deep orange, and the solids dissolved. After 24 hours, the reaction was removed from the oil bath and allowed to cool. The yellow liquid was transferred to a round bottom flask, and the vial was washed with copious amounts of toluene until no color was observed after washing the vial. The resulting liquid was concentrated by rotary evaporation, and the resulting oil was dissolved in DCM. Celite was added to the round bottom, and the product was adsorbed onto the celite by rotary evaporation. The desired compound was purified by reverse phase chromatography using a Biotage Isolera 1 system equipped with a Biotage® SNAP KP-C18-HS 12 g cartridge (part number FSL0-1118-0012) using Biotage method 2. Compound N**1** (0.24 mmol, 70.8 mg) was isolated in 24% yield as a pale yellow solid.

Scheme 3. 4: General procedure for reduction of 3-(3'-nitro)-quinazolinones to 3-(3'-amino)-quinazolinones



Compound N1 (50 mg, 0.17 mmol) was added to a round bottomed flask containing a stir bar and dissolved in methanol. 10% w/w Pd/C was added to the solution, and the flask was then tightly sealed with a septum. This mixture was sparged with a nitrogen balloon for at least 20 minutes. After saturating the methanol with nitrogen, a hydrogen balloon was sparged through the solution for at least 30 minutes. After this, the needle was removed from solution and the mixture was left to stir under hydrogen atmosphere overnight. The next day, the mixture was filtered through celite and the solvent was removed by rotary evaporation under reduced pressure. The resulting oil was dissolved in DCM, and silica gel was added to the solution, which was again concentrated by rotary evaporation under reduced pressure until dryness. The compound was purified via normal phase chromatography (3:1 EtOAc:Hexanes with 1% TEA). Substrate **1** was obtained as a white solid (0.11 mmol, 30 mg) in 67% yield.

Scheme 3. 5: General quinazolinone synthesis B

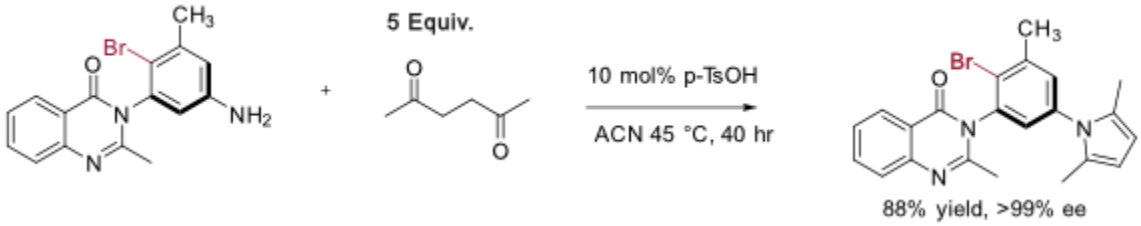


Solid 2,6-dimethylbenzoxazone (300 mg, 1.71 mmol) was added to a thick walled vial (Kimble Kimax 60700-2) and then the oil 5-methylbenzene-1,3-diamine (250 mg, 2.1 mmol) was added via syringe (measured gravimetrically). Finally, a stir bar was added followed by 2 mL of pyridine and the vial was tightly capped (Kontes Article No. 410119-2015 open cap with Kimble Article No. 749110-0022 valve) and left to stir at 140 °C in an oil bath for at least 24 hours. Over time, the reaction became deep red and the solids dissolved. After 24 hours, the reaction was removed from the oil bath and allowed to cool. The red liquid was transferred to a round bottom flask and the vial was washed with copious amounts of toluene (about 50 mL). The resulted red liquid was concentrated by rotary evaporation under reduced pressure, and the resulting oil was taken up in DCM. This liquid was washed with 1 M NaOH, at which point the red color changed to yellow. The aqueous phase was extracted 2x with DCM, then the organic layers were concentrated and washed with brine, then dried over magnesium sulfate. The resulting aniline compound was purified by reverse phase chromatography following the same procedure as compound N1, followed by normal phase (3:1 EtOAc:Hexanes with 1% TEA). Substrate 2 was obtained as a white solid in 41% yield (190 mg, 0.7 mmol).

General NBS reaction and purification

Prior to use, NBS was stored at 4 °C in a dark vial. To a dried round bottom flask was added a stir bar and 25 mg of aniline starting material (~0.09 mmol) followed by 10 mL of dry DCM. The solution was cooled to 0 °C using an ice bath. 0.9 equivalents of NBS (14 mg) was dissolved in 1 mL acetone and added dropwise to the aniline over the course of 5 minutes. After the addition was complete, the solution was allowed to stir for an additional hour, at which point the stir bar was removed and celite was added directly to the solution, which was concentrated under reduced pressure and immediately loaded onto the Biotage and purified using Biotage method 2 unless otherwise mentioned in the product characterization portion of this SI. If further purification was required, the resulting monobrominated material was collected from the Biotage and purified by either semi-preparative HPLC, or normal phase chromatography as mentioned in the product characterization portion of this SI.

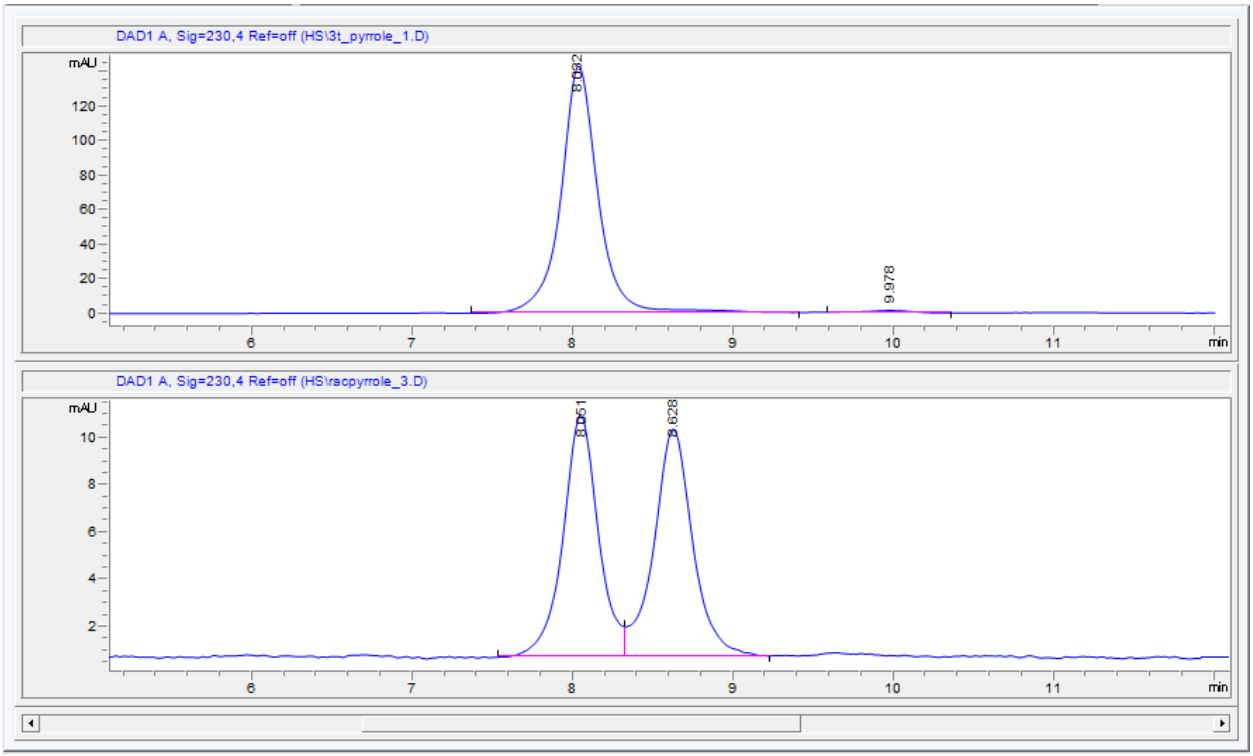
Scheme 3. 6: Synthesis of pyrrole 1c



To a thick-walled vial (Kimble Kimax 60700-2) containing a stir bar were added **1a** (20 mg, 0.058 mmol), *p*-toluenesulfonic acid monohydrate (1 mg, 0.006 mmol), and 2 mL acetonitrile. 2,5-hexanedione (5 equiv., 0.3 mmol) was added to the stirring solution via syringe, and the vial was tightly capped (Kontes Article No. 410119-2015 open cap with Kimble Article No. 749110-0022 valve) and heated to 45 °C for 40 hours. After the starting material could not be identified by TLC, the vial was removed from the oil bath and allowed to cool to room temperature. Silica was then added to the solution, and the solvent was removed by rotary evaporation under reduced pressure. Compound **1c** was purified via normal phase chromatography with 1:3 EtOAc:hexanes as the mobile phase. Compound **1c** was obtained as a white solid (21.5 mg, 0.051 mmol) in 88%yield.

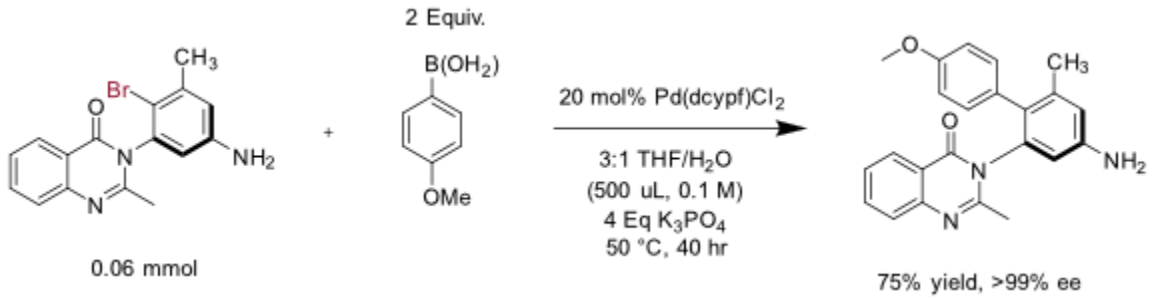
^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.29 (dd, *J* = 8.0, 1.5 Hz, 1H), 7.79 (ddd, *J* = 8.5, 7.2, 1.6 Hz, 1H), 7.72 – 7.69 (m, 1H), 7.52 – 7.47 (m, 1H), 7.27 (d, *J* = 2.5 Hz, 1H), 7.07 (d, *J* = 2.4 Hz, 1H), 5.90 (s, 2H), 2.57 (s, 3H), 2.28 (s, 3H), 2.10 (s, 6H). ^{13}C NMR (126 MHz, Chloroform-*d*) δ 161.29, 153.19, 147.50, 141.56, 138.96, 137.89, 134.86, 131.14, 128.55, 127.18, 127.00, 126.94, 126.83, 124.44, 120.64, 106.66, 23.80, 23.67, 13.16. **HRMS** (ESI-MS) Calc. for $[C_{22}H_{20}BrN_3OH]^+$ ($[M+H]$)$^+$: 422.0863, found 422.0864.

Figure 3. 20: Chiral SFC trace of pyrrole product 1c



Top: Chemoenzymatic product with **1a** obtained from biconversions with 3-T. Bottom: Product obtained with **1a** obtained with NBS reaction from **1**.

Scheme 3. 7: Suzuki coupling reaction to form product 1d

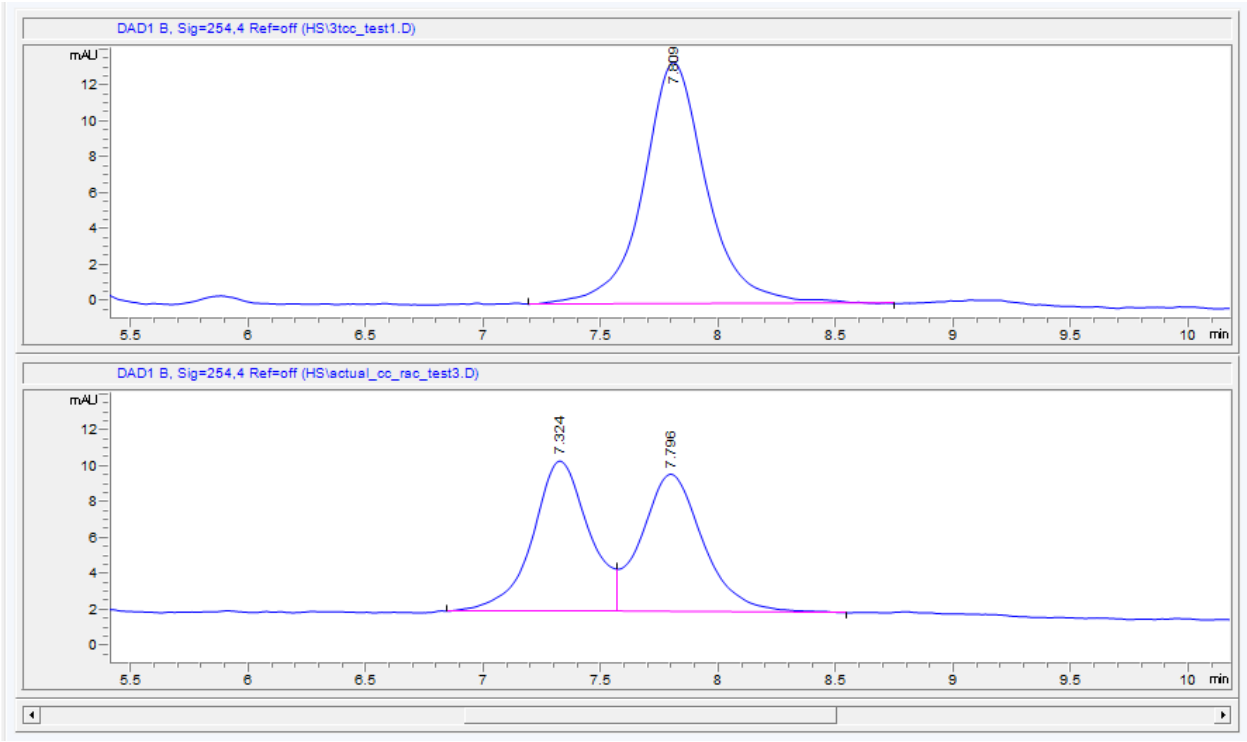


To a thick-walled vial (Kimble Kimax 60700-2) containing a stir bar was added **1a** (20 mg, 0.058 mmol), K_3PO_4 (4 equiv., 49 mg, 0.232 mmol), and *p*-methoxyboronic acid (2 equiv., 0.116 mmol, 17.6 mg). The vial was then equipped with a septum and evacuated and purged 3x with nitrogen. This vial was then imported to a drybox where 20 mol% of [1,1'-Bis(di-cyclohexylphosphino)ferrocene]dichloropalladium(II) (8.8 mg, 0.011 mmol) was added. The septum was replaced on the vial, and it was exported from the dry box and imported to a wet box, where it was subsequently suspended in 0.5 mL of a 3:1 mixture of THF/H_2O. The vial was then capped tightly (Kontes Article No. 410119-2015 open cap with Kimble Article No. 749110-0022 valve) and removed from the wet box. The vial was then placed in an oil bath set

to 50 °C and left to stir vigorously. After 40 hours, the vial was removed from the oil bath and allowed to cool to room temperature. The contents of the vial were diluted with DCM and transferred to a separatory funnel, where the organic layer was washed with 5 mL of an aqueous solution of 1 M NaOH. The organic layer was then washed with brine and dried over magnesium sulfate. The solution was then filtered into a round bottom flask and silica was added, and the solvent removed by rotary evaporation under reduced pressure. The compound was then first purified by normal phase chromatography (3:1 DCM:ACN with 1% TEA), followed by reverse phase purification using Biotage method 2. Product **1d** was obtained as a white film in 75% yield (17 mg, 0.046 mmol).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.23 (d, *J* = 7.9 Hz, 1H), 7.67 (t, *J* = 7.7 Hz, 1H), 7.50 (d, *J* = 8.2 Hz, 1H), 7.39 (t, *J* = 7.6 Hz, 1H), 7.12 (d, *J* = 8.5 Hz, 1H), 7.02 (d, *J* = 8.4 Hz, 1H), 6.78 (d, *J* = 8.3 Hz, 1H), 6.74 (d, *J* = 2.3 Hz, 1H), 6.62 (d, *J* = 8.5 Hz, 1H), 6.43 (d, *J* = 2.4 Hz, 1H), 3.80 (s, 2H), 3.68 (s, 3H), 2.22 (s, 3H), 2.08 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.63, 158.59, 154.30, 147.41, 146.51, 139.97, 137.18, 134.22, 130.36, 130.05, 129.83, 128.90, 127.03, 126.54, 126.20, 120.51, 118.05, 113.78, 113.47, 111.86, 55.03, 23.95, 21.06. **HRMS** (ESI-MS) Calc. for $[C_{23}H_{21}N_3O_2H]^+$ ([M+H])$^+$: 372.1707, found 372.1707.

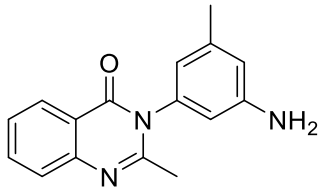
Figure 3. 21: Chiral SFC trace of cross coupled product 1d



Top: Chemoenzymatic product with **1d** obtained from biconversions with 3-T with substrate **1**. Bottom: Product obtained with **1a** obtained with NBS reaction from **1**.

Substrate Synthesis

Substrate **1**



Nitro compound N1 and substrate **1** were synthesized using general method A, and detailed synthetic information is included above. Nitro compound N1 was isolated in 24% yield. After the reduction, substrate **1** was isolated in 67% yield as a fine white powder.

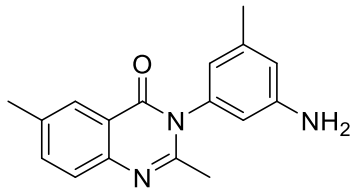
Nitro compound N1:

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.28 – 8.18 (m, 2H), 7.96 (s, 1H), 7.85 (d, *J* = 5.8 Hz, 2H), 7.57 (t, *J* = 6.9 Hz, 1H), 7.42 (s, 1H), 2.53 (s, 3H), 2.41 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.11, 152.85, 149.15, 147.33, 142.05, 138.59, 135.09, 135.04, 127.09, 127.06, 127.05, 124.92, 120.94, 120.44, 24.43, 21.39.

Substrate **1**:

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.31 – 8.24 (m, 1H), 7.76 (t, *J* = 7.6 Hz, 1H), 7.68 (d, *J* = 8.3 Hz, 1H), 7.46 (t, *J* = 7.6 Hz, 1H), 6.62 (s, 1H), 6.45 (s, 1H), 6.37 (s, 1H), 3.77 (s, 2H), 2.35 – 2.30 (m, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.29, 154.60, 147.90, 147.52, 141.03, 138.60, 134.46, 127.08, 126.70, 126.50, 120.84, 118.44, 116.67, 111.48, 24.10, 21.42. HRMS (ESI-MS) Calc. for $[C_{16}H_{15}N_3OH]^+$ ($[M+H]$)$^+$: 266.1288, found 266.1286.

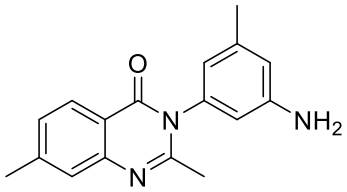
Substrate **2**



Substrate **2** was synthesized according to general method B, and detailed synthetic information is provided above. Substrate **2** was obtained in 41% yield as a white powder after normal phase chromatography.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.05 (s, 1H), 7.63 – 7.48 (m, 2H), 6.59 (s, 1H), 6.43 (s, 1H), 6.35 (s, 1H), 3.78 (s, 2H), 2.47 (s, 3H), 2.31 (s, 3H), 2.29 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.42, 153.82, 148.00, 145.62, 141.10, 138.86, 136.74, 136.05, 126.62, 126.57, 120.68, 118.62, 116.75, 111.67, 24.14, 21.55, 21.42. HRMS (ESI-MS) Calc. for $[C_{17}H_{17}N_3OH]^+$ ($[M+H]$)$^+$: 280.1444, found 280.1443.

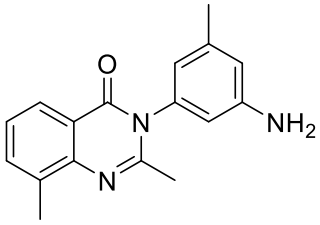
Substrate **3**



Substrate **3** was synthesized using general procedure B starting from 2,7-dimethylbenzoxazone (100 mg, 0.57 mmol) and 5-methylbenzene-1,3-diamine (83 mg, 0.68 mmol). Substrate **3** was obtained in 45% yield (71 mg, 0.26 mmol) as a pale orange powder following normal phase chromatography (3:1 EtOAc:Hexanes 1% TEA).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.14 (d, *J* = 8.1 Hz, 1H), 7.47 (s, 1H), 7.27 (d, *J* = 7.9 Hz, 1H), 6.60 (s, 1H), 6.44 (s, 1H), 6.35 (d, *J* = 2.1 Hz, 1H), 3.79 (s, 2H), 2.51 (s, 3H), 2.31 (s, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.18, 147.87, 145.49, 140.98, 138.63, 128.11, 126.92, 126.37, 118.48, 118.37, 116.63, 111.53, 24.04, 21.97, 21.42. HRMS (ESI-MS) Calc. for $[C_{17}H_{17}N_3OH]^+$ ($[M+H]$)$^+$: 280.1444, found 280.1444.

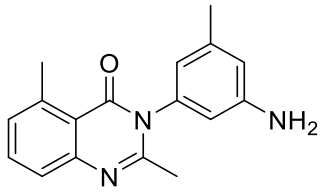
Substrate **4**



Substrate **4** was synthesized using general procedure B starting from 2,8-dimethylbenzoxazone (100 mg, 0.57 mmol) and 5-methylbenzene-1,3-diamine (83 mg, 0.68 mmol). Substrate **4** was obtained in 34% yield (54mg, 0.19 mmol) as a pale-red powder following normal phase chromatography (3:1 EtOAc:Hexanes 1% TEA).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.12 (d, *J* = 7.9 Hz, 1H), 7.59 (d, *J* = 7.3 Hz, 1H), 7.33 (t, *J* = 7.6 Hz, 1H), 6.60 (s, 1H), 6.44 (s, 1H), 6.36 (d, *J* = 2.0 Hz, 1H), 3.77 (s, 2H), 2.63 (s, 3H), 2.33 (s, 3H), 2.31 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.75, 153.11, 147.86, 146.16, 140.93, 138.86, 135.26, 134.98, 125.94, 124.71, 120.80, 118.49, 116.57, 111.55, 24.31, 21.42, 17.32, 14.21. HRMS (ESI-MS) Calc. for $[C_{17}H_{17}N_3OH]^+$ ($[M+H]$)$^+$: 280.1444, found 280.1444.

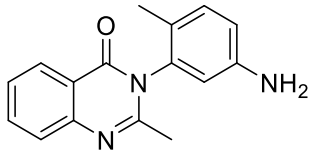
Substrate **5**



Substrate **5** was synthesized using general procedure B starting from 2,5-dimethylbenzoxazone (100 mg, 0.57 mmol) and 5-methylbenzene-1,3-diamine (83 mg, 0.68 mmol). Substrate **5** was obtained in 27% yield (0.15 mmol, 43 mg) as a white powder following reverse phase purification using Biotage method 1.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.58 (t, *J* = 7.7 Hz, 1H), 7.50 (d, *J* = 8.1 Hz, 1H), 7.20 (d, *J* = 7.3 Hz, 1H), 6.60 (s, 1H), 6.44 (s, 1H), 6.36 (d, *J* = 2.1 Hz, 1H), 3.79 (s, 2H), 2.83 (s, 3H), 2.32 (s, 3H), 2.27 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.94, 154.22, 149.11, 147.91, 141.53, 141.05, 138.86, 133.61, 129.21, 124.92, 119.26, 118.52, 116.54, 111.56, 23.93, 23.04, 21.45. HRMS (ESI-MS) Calc. for $[C_{17}H_{17}N_3OH]^+$ ($[M+H]$)$^+$: 280.1444, found 280.1445.

Substrate **6**



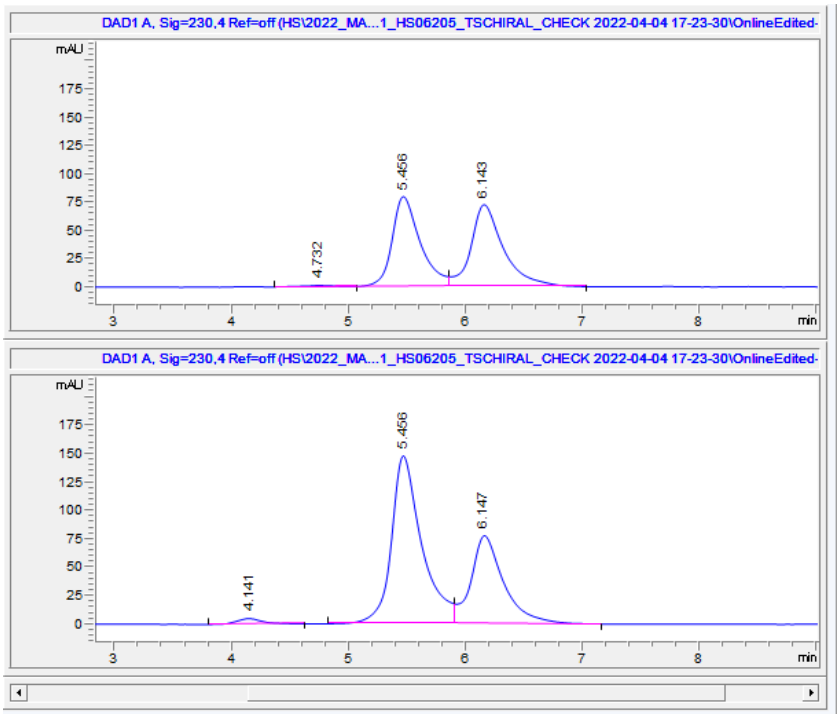
Substrate **6** was synthesized using general procedure A starting from 2-methylbenzoxazone (161 mg, 1 mmol) and 2-methyl-5-nitroaniline (182 mg, 1.2 mmol). Nitro compound N6 was obtained in 35% yield (103 mg, 0.35 mmol) as a pale-yellow powder using Biotage method 2. N6 (100 mg, 0.34 mmol) was reduced in the same way as substrate 1 using 10 mg of 10% Pd/C. After reduction, substrate **6** was isolated as a white powder in 95% yield (86 mg, 0.32 mmol).

N6 characterization: ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.29 (td, *J* = 8.2, 7.7, 2.0 Hz, 2H), 8.11 (d, *J* = 2.3 Hz, 1H), 7.87 – 7.78 (m, 1H), 7.73 (d, *J* = 8.1 Hz, 1H), 7.61 (d, *J* = 8.5 Hz, 1H), 7.52 (t, *J* = 7.6 Hz, 1H), 2.26 (s, 3H), 2.22 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.43, 152.81, 147.51, 147.34, 143.91, 137.71, 135.12, 132.33, 127.13, 127.12, 124.49, 123.85, 120.40, 23.92, 17.86. HRMS (ESI-MS) Calc. for $[C_{16}H_{13}N_3O_3H]^+$ ($[M+H]$)$^+$: 296.1030, found 296.1030.

Substrate **6** characterization: ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.29 (dd, *J* = 7.9, 1.4 Hz, 1H), 7.77 (td, *J* = 7.5, 6.9, 1.4 Hz, 1H), 7.68 (d, *J* = 8.2 Hz, 1H), 7.47 (t, *J* = 7.5 Hz, 1H), 7.16 (d, *J* = 8.2 Hz, 1H), 6.73 (dd, *J* = 8.3, 2.4 Hz, 1H), 6.50 (d, *J* = 2.4 Hz, 1H), 3.70 (s, 3H), 2.24 (s, 3H), 1.99 (s, 3H). ^{13}C NMR (126 MHz,

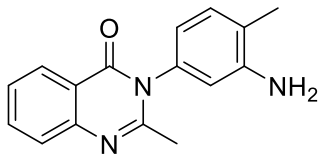
Chloroform-*d*) δ 161.63, 154.54, 147.64, 145.99, 137.32, 134.51, 132.17, 127.13, 126.73, 126.51, 124.53, 120.79, 116.61, 114.18, 23.69, 16.40. HRMS (ESI-MS) Calc. for $[C_{16}H_{15}N_3OH]^+$ $([M+H])^+$: 266.1288, found 266.1287.

Figure 3. 22: SFC chromatograms for substrate 6.



Top: purified substrate **6**. Bottom: recovered substrate 6 after reaction with 3-T (62% recovered, e.r. 65:35).

Substrate **7**

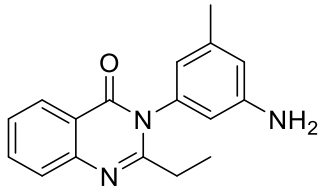


Substrate **7** was synthesized using general procedure A starting from 2-methylbenzoxazone (161 mg, 1 mmol) and 4-methyl-3-nitroaniline (182 mg, 1.2 mmol). Nitro compound N7 was obtained in 28% yield (83 mg, 0.28 mmol) as a pale-yellow powder after reverse phase purification using Biotage method 2. Nitro compound N7 (50 mg, 0.17 mmol) was reduced to substrate **7** using the same method as described for substrate 1 with 5 mg 10% Pd/C and was purified by reverse phase chromatography using Biotage method 1. Substrate **7** was obtained in 78% yield (35 mg, 0.13 mmol) after the reduction.

N7 characterization: ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.30 (d, *J* = 7.9 Hz, 1H), 8.00 (s, 1H), 7.91 (q, *J* = 8.5, 8.0 Hz, 2H), 7.62 (d, *J* = 7.0 Hz, 2H), 7.48 (s, 1H), 2.75 (s, 3H), 2.48 (s, 3H).^{13}C NMR (126 MHz, $CDCl_3$) δ 162.12, 153.02, 149.73, 147.32, 136.36, 135.28, 135.02, 134.44, 132.87, 127.09, 127.07, 127.04, 124.85, 120.46, 24.45, 20.52. HRMS (ESI-MS) Calc. for $[C_{16}H_{13}N_3O_3H]^+$ ($[M+H]$)$^+$: 296.1030, found 296.1031.

Substrate **7**: ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.27 (dd, *J* = 7.9, 1.6 Hz, 1H), 7.75 (ddd, *J* = 8.5, 7.1, 1.6 Hz, 1H), 7.66 (d, *J* = 8.1 Hz, 1H), 7.48 – 7.41 (m, 1H), 7.19 (d, *J* = 7.7 Hz, 1H), 6.59 – 6.52 (m, 2H), 3.76 (s, 2H), 2.30 (s, 3H), 2.22 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.34, 154.78, 147.47, 146.08, 136.46, 134.46, 131.73, 127.12, 126.67, 126.49, 123.45, 120.86, 117.45, 113.78, 24.16, 17.22. HRMS (ESI-MS) Calc. for $[C_{16}H_{15}N_3OH]^+$ ($[M+H]$)$^+$: 266.1288, found 266.1286.

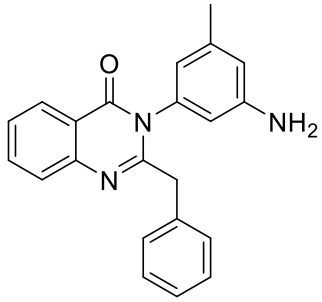
Substrate **8**



Substrate **8** was synthesized according to general procedure B starting from 2-ethylbenzoxazone (100 mg, 0.57 mmol) and 5-methylbenzene-1,3-diamine (84 mg, 0.69 mmol). After reverse phase purification of the crude condensation reaction using Biotage method 2, a closely eluting yellow impurity was removed by silica chromatography using 3:1 DCM:ACN with 1% TEA as an additive. Substrate **8** was isolated in 16% yield (25 mg, 0.09 mmol) as a white powder.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.30 – 8.24 (m, 1H), 7.79 – 7.67 (m, 2H), 7.45 (t, *J* = 7.4 Hz, 1H), 6.61 (s, 1H), 6.44 (s, 1H), 6.36 (d, *J* = 2.3 Hz, 1H), 3.77 (s, 2H), 2.54 (q, *J* = 7.4 Hz, 2H), 2.32 (s, 3H), 1.25 (t, *J* = 7.4 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.48, 158.13, 147.74, 147.6, 140.89, 138.19, 134.33, 127.03, 126.98, 126.40, 120.83, 118.81, 116.62, 111.79, 28.98, 21.43, 11.42. HRMS (ESI-MS) Calc. for $[C_{17}H_{17}N_3OH]^+$ ($[M+H]$)$^+$: 280.1444, found 280.1444.

Substrate **9**



Substrate **9** was synthesized using general procedure A starting from 2-benzylbenzoxazone (100 mg, 0.36 mmol) and 3-nitro-5-methylaniline (67 mg, 0.44 mmol). Nitro compound N9 was purified by reverse phase Biotage using the same column and system as previously reported, but using the following method:

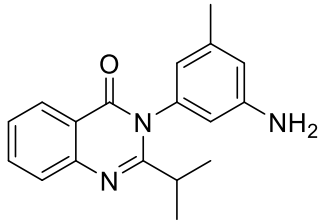
%A (Water, 0.1% Formic acid)	%B (MeOH, 0.1% Formic acid)	CV (column volumes)
60	40	1 (Equilibration)
60	40	1
10	90	25
5	95	1
5	95	3

Nitro compound N9 was obtained in 33% yield (44 mg, 0.12 mmol) as a pale yellow solid. Reduction of the N9 to substrate **9** was performed the same way as substrate **1** using 40 mg of N9 (10.8 mmol) and 4 mg of 10% Pd/C. After the reduction of N9, substrate **9** was purified by silica chromatography using 1:1 EtOAc:Hexanes with 1% TEA as an additive. After purification, substrate 9 was obtained in 75% yield (0.08 mmol, 27 mg) as a white powder.

Nitro compound N9 characterization: ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.27 (dt, *J* = 8.0, 1.1 Hz, 1H), 8.11 – 8.06 (m, 1H), 7.90 – 7.80 (m, 2H), 7.63 (d, *J* = 2.2 Hz, 1H), 7.54 (ddd, *J* = 8.2, 6.4, 2.0 Hz, 1H), 7.22 – 7.12 (m, 3H), 6.92 (q, *J* = 1.4 Hz, 1H), 6.84 – 6.76 (m, 2H), 3.93 (q, *J* = 15.0 Hz, 2H), 2.34 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.40, 154.09, 148.49, 147.24, 141.12, 137.44, 135.97, 135.05, 134.85, 128.64, 128.31, 127.58, 127.40, 127.39, 127.07, 124.53, 121.55, 120.66, 43.15, 21.17. HRMS (ESI-MS) Calc. for $[C_{22}H_{17}N_3O_3H]^+$ $([M+H])^+$: 372.1343, found 372.1342.

Aniline substrate **9** characterization: ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.20 (d, *J* = 8.0 Hz, 1H), 7.70 (d, *J* = 5.5 Hz, 2H), 7.43 – 7.37 (m, 1H), 7.13 (s, 3H), 6.93 – 6.88 (m, 2H), 6.48 (s, 1H), 6.02 (s, 1H), 5.97 (s, 1H), 3.85 (s, 2H), 3.57 (s, 2H), 2.10 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.53, 155.53, 147.48, 147.40, 140.36, 137.46, 136.02, 134.43, 128.76, 128.19, 127.25, 127.08, 126.79, 126.74, 121.00, 119.38, 116.55, 112.28, 42.51, 21.31. HRMS (ESI-MS) Calc. for $[C_{22}H_{19}N_3OH]^+$ $([M+H])^+$: 342.1601, found 342.1601.

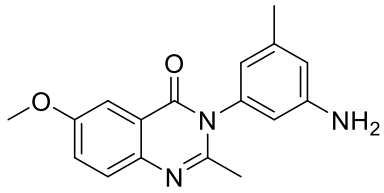
Substrate **10**



Substrate **10** was synthesized using general procedure B starting from 2-isopropylbenzoxazone (75 mg, 0.4 mmol) and 5-methylbenzene-1,3-diamine (58 mg, 0.48 mmol) according to the procedure for substrate 2 and was isolated as a white powder in 22% yield (25 mg, 0.09 mmol) after silica purification in 3:1 EtOAc:Hexanes with 1% TEA as an additive.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.26 (d, *J* = 7.9 Hz, 1H), 7.78 – 7.67 (m, 2H), 7.43 (t, *J* = 7.4 Hz, 1H), 6.62 (s, 1H), 6.44 (s, 1H), 6.36 (s, 1H), 2.84 (s, 1H), 2.32 (s, 3H), 1.24 (d, *J* = 6.7, 1.4 Hz, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.68, 161.91, 147.83, 147.67, 140.81, 138.31, 134.24, 127.10, 126.97, 126.26, 120.80, 118.89, 116.59, 111.87, 32.19, 21.58, 21.55, 21.44. HRMS (ESI-MS) Calc. for $[C_{18}H_{19}N_3OH]^+$ $([M+H])^+$: 294.1601, found 294.1605.

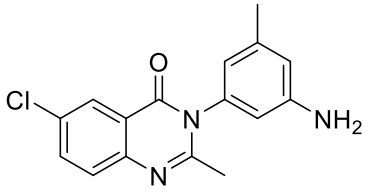
Substrate **11**



Substrate 11 was synthesized according to general procedure B starting from 6-methoxy-2-methylbenzoxazone (150 mg, 0.78 mmol) and 5-methylbenzene-1,3-diamine (115 mg, 0.94 mmol). After reverse phase purification, a closely eluting colored impurity was removed via normal phase chromatography in 3:1 DCM:ACN with 1% TEA as an additive. Substrate 11 was isolated as a slightly brown powder in 10% yield.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.66 – 7.55 (m, 2H), 7.34 (d, *J* = 8.9, 3.0 Hz, 1H), 6.60 (s, 1H), 6.44 (s, 1H), 6.35 (s, 1H), 3.90 (s, 3H), 3.77 (s, 2H), 2.31 (s, 3H), 2.28 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.32, 158.28, 152.36, 148.02, 142.27, 141.09, 138.90, 128.41, 124.73, 121.67, 118.56, 116.74, 111.62, 106.63, 55.93, 23.97, 21.54. HRMS (ESI-MS) Calc. for $[C_{17}H_{17}N_3O_2H]^+$ $([M+H])^+$: 296.1394, found 296.1395.

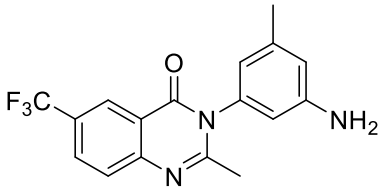
Substrate **12**



Substrate **12** was initially synthesized according to general procedure A starting from 6-chloro-2-methylbenzoxazone and 3-methyl-5-nitroaniline, however during the reduction step we observed hydrogenative dehalogenation. Synthesis of substrate 12 was accomplished using the same protocol as substrate **2** starting from 6-chloro-2-methylbenzoxazone (150 mg, 0.77 mmol) and 5-methylbenzene-1,3-diamine (112 mg, 0.92 mmol). After reaction, substrate 12 was purified by normal phase chromatography with 1:1 EtOAc:Hexanes with 1% TEA as an additive. Substrate 12 was obtained in 57% yield (131 mg, 0.44 mmol) as a white powder.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.22 (d, *J* = 2.5 Hz, 1H), 7.68 (dd, *J* = 8.7, 2.4 Hz, 1H), 7.60 (d, *J* = 8.7 Hz, 1H), 6.62 (s, 1H), 6.43 (s, 1H), 6.35 (d, *J* = 2.3 Hz, 1H), 3.82 (s, 2H), 2.32 (s, 3H), 2.30 (s, 3H).^{13}C NMR (126 MHz, $CDCl_3$) δ 161.40, 155.10, 148.08, 146.18, 141.31, 138.44, 135.01, 132.36, 128.59, 126.50, 122.05, 118.42, 116.95, 111.42, 24.22, 21.57. HRMS (ESI-MS) Calc. for $[C_{16}H_{14}ClN_3OH]^+$ ($[M+H]$)$^+$: 300.0898, found 300.0899.

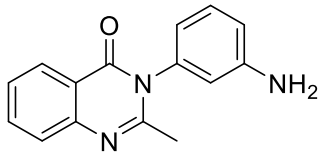
Substrate **13**



Substrate **13** was synthesized using general procedure B starting from 2-methyl-6-trifluoromethylbenzoxazone (100 mg, 0.44 mmol) and 5-methylbenzene-1,3-diamine (65 mg, 0.53 mmol). After the reaction, the compound was purified by normal phase chromatography using 2:1 EtOAc:Hexanes with 1% TEA as an additive. This material was further purified by reverse phase chromatography using Biotage method 2. Substrate **13** was obtained in 57% (83 mg, 0.25 mmol) yield as a brown powder.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.54 (s, 1H), 7.96 (d, *J* = 7.7 Hz, 1H), 7.81 (d, *J* = 8.6 Hz, 1H), 6.64 (s, 1H), 6.44 (s, 1H), 6.35 (s, 1H), 2.37 (s, 3H), 2.33 (s, 3H). ^{19}F NMR (376 MHz, $CDCl_3$) δ -62.34. ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.52, 157.02, 149.63, 148.24, 141.15, 138.01, 130.62, 130.59, 130.57, 130.54, 128.77, 128.51, 128.24, 127.98, 127.74, 126.96, 125.00, 124.97, 124.93, 124.90, 124.79, 122.63, 120.71, 120.46, 117.89, 116.83, 111.15, 24.18, 21.38. HRMS (ESI-MS) Calc. for $[C_{17}H_{14}F_3N_3OH]^+$ ($[M+H]$)$^+$: 334.1162, found 334.1161.

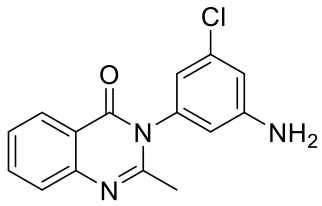
Substrate **14**



Substrate **14** was synthesized according to general procedure B starting from 2-methylbenzoxazone (161 mg, 1 mmol) and 1,3-dianiline (129 mg, 1.2 mmol). Substrate 14 was purified by reverse phase chromatography using Biotage method 1. Substrate **14** was obtained in 40% yield (100 mg, 0.40 mmol) as a white powder.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.28 (d, 1H), 7.82 – 7.71 (m, 2H), 7.48 (t, 1H), 7.31 (t, *J* = 7.9 Hz, 1H), 6.79 (d, 1H), 6.62 (d, *J* = 7.7, 1.4 Hz, 1H), 6.55 (s, 1H), 2.34 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.63, 154.88, 148.63, 147.89, 139.08, 134.88, 131.07, 127.43, 127.10, 126.92, 121.19, 117.85, 116.24, 114.61, 111.59, 107.47, 24.48. HRMS (ESI-MS) Calc. for $[C_{15}H_{13}N_3OH]^+$ ([M+H])$^+$: 252.1131, found 252.1131.

Substrate **15**

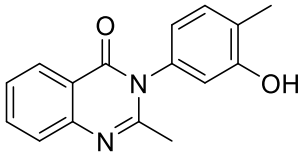


Substrate **15** was synthesized using general procedure A starting from 2-methylbenzoxazone (161 mg, 1 mmol) and 3-chloro-5-nitroaniline (206 mg, 1.2 mmol). Nitroaniline N15 was purified by reverse phase chromatography using Biotage method 2 and obtained in 16% yield (50 mg, 0.16 mmol) as a yellow powder. N15 was reduced to substrate **15** using the same protocol as for substrate 1. After the reduction, substrate **15** was purified by reverse phase chromatography using Biotage method 1. Substrate **15** was obtained in 81% yield (37 mg, 0.13 mmol) after the reduction.

Nitro compound N15: ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.39 (t, *J* = 2.0 Hz, 1H), 8.25 (dd, *J* = 8.0, 1.5 Hz, 1H), 8.10 (t, *J* = 1.9 Hz, 1H), 7.82 (ddd, *J* = 8.5, 7.2, 1.6 Hz, 1H), 7.70 (d, *J* = 8.1 Hz, 1H), 7.66 (t, *J* = 1.9 Hz, 1H), 7.55 – 7.48 (m, 1H), 2.29 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.88, 152.05, 149.50, 147.19, 139.66, 136.85, 135.28, 134.85, 127.33, 127.19, 127.06, 124.87, 122.32, 120.23, 24.43. HRMS (ESI-MS) Calc. for $[C_{15}H_{10}ClN_3O_3H]^+$ ([M+H])$^+$: 316.0483, found 316.0484.

Aniline substrate **15**: ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.26 (d, *J* = 7.9, 1.5 Hz, 1H), 7.77 (td, *J* = 7.7, 7.1, 1.6 Hz, 1H), 7.67 (d, *J* = 8.1 Hz, 1H), 7.47 (t, *J* = 7.7 Hz, 1H), 6.78 (t, *J* = 2.0 Hz, 1H), 6.64 (t, *J* = 1.8 Hz, 1H), 6.44 (t, *J* = 1.9 Hz, 1H), 3.94 (s, 2H), 2.33 (s, 3H). HRMS (ESI-MS) Calc. for $[C_{15}H_{13}ClN_3OH]^+$ ([M+H])$^+$: 286.0742, found 286.0742.

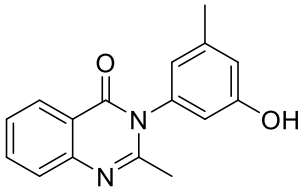
Substrate **16**



Substrate **16** was synthesized according a previous literature report[1].

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.31 (d, *J* = 8.0 Hz, 1H), 7.97 (t, *J* = 7.6 Hz, 1H), 7.90 (d, *J* = 8.2 Hz, 1H), 7.71 (t, *J* = 7.6 Hz, 1H), 7.35 (d, *J* = 7.7 Hz, 1H), 6.77 – 6.71 (m, 2H), 2.62 (s, 3H), 2.31 (s, 3H).

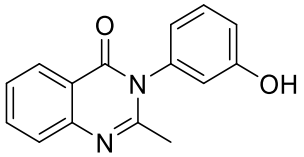
Substrate **17**



Substrate **17** was synthesized according a previous literature report[1].

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.29 (d, *J* = 8.0 Hz, 1H), 7.93 (d, *J* = 7.9 Hz, 2H), 7.66 (t, *J* = 7.4 Hz, 1H), 6.83 (s, 1H), 6.57 (s, 1H), 6.48 (s, 1H), 2.53 (s, 3H), 2.32 (s, 3H).

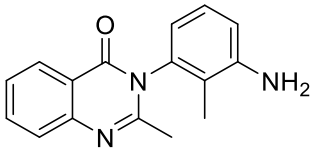
Substrate **18**



Substrate **18** was synthesized according a previous literature report[1].

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.29 (d, *J* = 7.8 Hz, 1H), 7.79 (t, *J* = 7.7 Hz, 1H), 7.70 (d, *J* = 8.5 Hz, 1H), 7.49 (t, *J* = 7.6 Hz, 1H), 7.38 (t, *J* = 7.9 Hz, 1H), 6.89 (d, *J* = 8.5 Hz, 1H), 6.76 (d, *J* = 7.7 Hz, 1H), 6.65 (s, 1H), 2.29 (s, 3H).

Substrate **19**

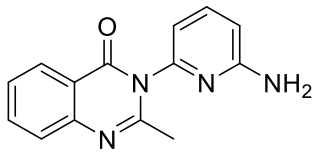


Note that because substrate **19** was not reported in the final publication, only the proton NMRs for **N19** and **19** were acquired.

Nitrosubstrate **N19**: ^{1}H NMR (400 MHz, cdcl$_3$) δ 8.29 (td, *J* = 8.2, 7.7, 2.0 Hz, 2H), 8.11 (d, *J* = 2.3 Hz, 1H), 7.87 – 7.78 (m, 1H), 7.73 (d, *J* = 8.1 Hz, 1H), 7.61 (d, *J* = 8.5 Hz, 1H), 7.52 (t, *J* = 7.6 Hz, 1H), 2.26 (s, 3H), 2.22 (s, 3H).

Aniline substrate **19:** 1H NMR (400 MHz, cdcl3) δ 8.29 (d, J = 8.0 Hz, 1H), 7.80 (s, 2H), 7.51 (s, 1H), 7.19 (t, J = 7.9 Hz, 1H), 6.82 (d, J = 8.1 Hz, 1H), 6.60 (d, J = 7.8 Hz, 1H), 2.30 (s, 2H), 2.04 (s, 6H).

Substrate **20**

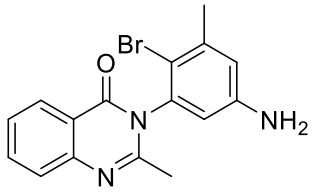


Note that because substrate **20** was not reported in the final publication, only the proton NMR was acquired.

^{1}H NMR (400 MHz, cdcl$_3$) δ 8.28 (d, *J* = 7.9 Hz, 1H), 7.82 – 7.62 (m, 3H), 7.46 (t, *J* = 7.5 Hz, 1H), 6.70 (d, *J* = 7.4 Hz, 1H), 6.62 (d, *J* = 8.2 Hz, 1H), 4.62 (s, 2H), 2.34 (s, 3H).

Product characterization

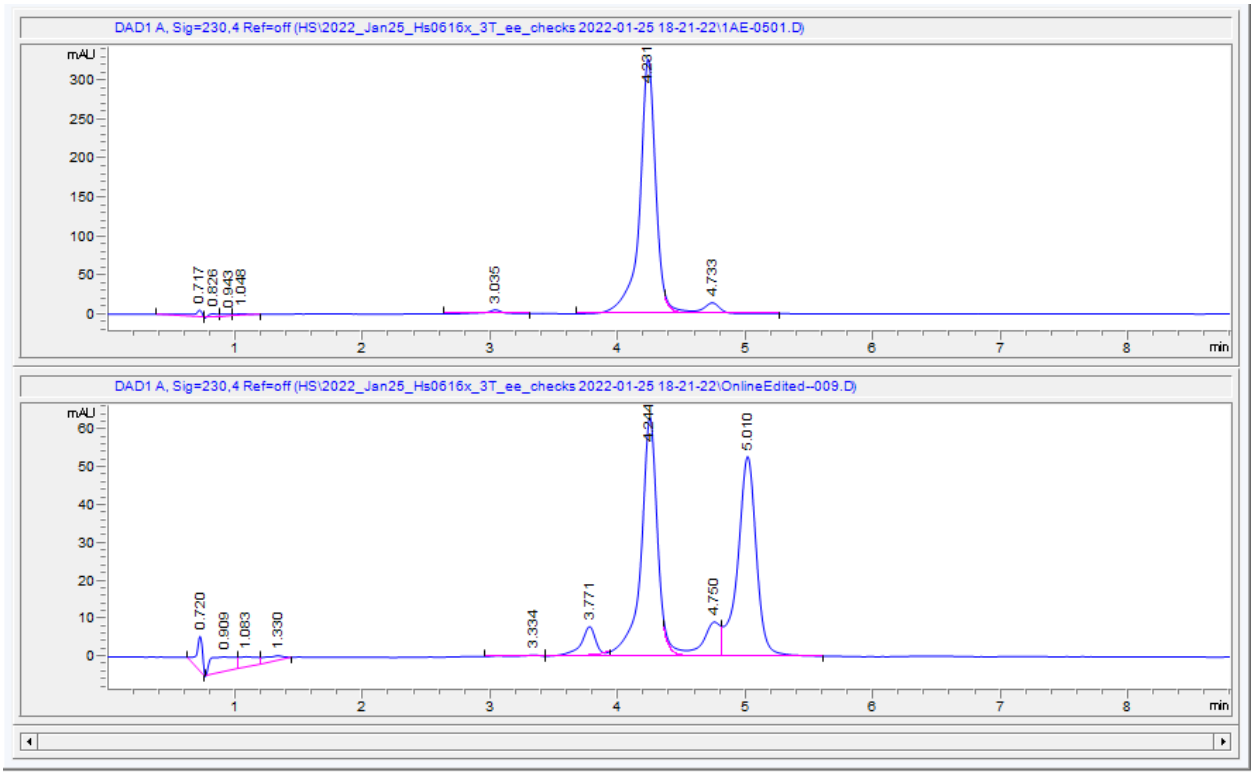
Product **1a**



Product **1a** was obtained after purification from Biotage using method 2 following bioconversions with 3-T. With earlier variants and NBS reactions when further purification was required, 1a was further purified via normal phase purification using 3:1 ethyl acetate:hexanes with 1% TEA. From bioconversions with 3-T, 1a was obtained in 92% yield (9.5 mg) as a white powder.

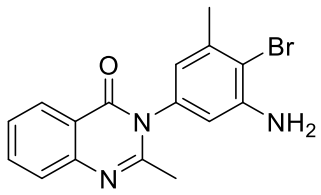
^{1}H NMR (500 MHz, CDCl$_3$) δ 8.18 (dd, *J* = 7.9, 1.6 Hz, 1H), 7.71 (ddd, *J* = 8.6, 7.1, 1.6 Hz, 1H), 7.61 (d, 1H), 7.41 (ddd, *J* = 8.2, 7.1, 1.2 Hz, 1H), 6.68 (d, 1H), 6.45 (d, *J* = 2.8 Hz, 1H), 3.83 (s, 2H), 2.35 (s, 3H), 2.19 (s, 3H). ^{13}C NMR (126 MHz, Chloroform-*d*) δ 162.27, 154.58, 147.88, 147.50, 141.02, 138.58, 134.45, 127.07, 126.69, 126.48, 120.83, 118.43, 116.65, 111.46, 24.09, 21.42. HRMS (ESI-MS) Calc. for $[C_{16}H_{14}BrN_3OH]^+$ ([M+H])$^+$: 344.0397, found 344.0393.

Figure 3. 23: Chiral SFC trace of product 1a



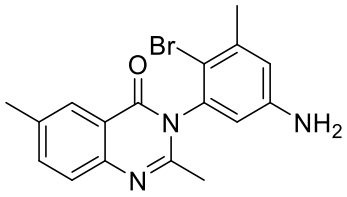
Top: **1a** obtained from biconversions with 3-T. Bottom: **1a** obtained with NBS reaction.

Product **1b**



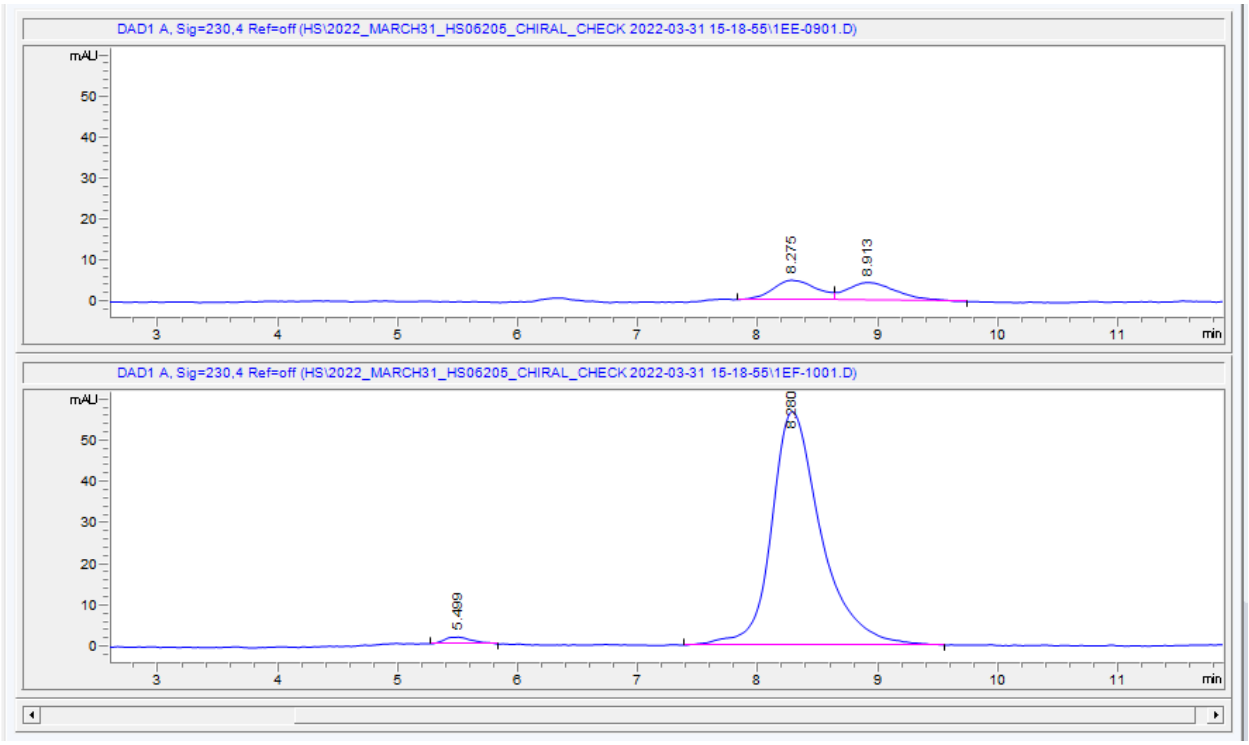
Product **1b** was obtained from bioconversions using 6-TLP and NBS reactions using semiprep method 1 as a white film. From bioconversions with 6-TLP, 1b was obtained in 11% yield (0.6 mg, 0.02 mmol). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.27 (d, 1H), 7.83 – 7.70 (m, 2H), 7.48 (t, 1H), 6.56 – 6.48 (m, 2H), 2.43 (s, 3H), 2.37 (s, 3H). HRMS (ESI-MS) Calc. for $[C_{16}H_{14}BrN_3OH]^+$ ($[M+H]$)$^+$: 344.0393, found 344.0392.

Product **2a**



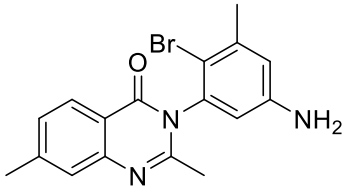
Product **2a** was obtained after purification from Biotage using method 2 with 3-T. With earlier variants and NBS reactions when further purification was required, **2a** was further purified via normal phase purification using 3:1 ethyl acetate:hexanes with 1% TEA. From bioconversions with 3-T, product **2a** was obtained in 90% yield (9.6 mg) as a white powder. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.07 (s, 1H), 7.58 (d, *J* = 1.3 Hz, 2H), 6.70 (d, *J* = 2.7 Hz, 1H), 6.49 (d, *J* = 2.7 Hz, 1H), 2.48 (s, 3H), 2.40 (s, 3H), 2.26 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.52, 153.15, 146.72, 145.56, 140.89, 137.82, 136.70, 136.07, 126.63, 126.56, 120.43, 118.18, 113.30, 112.42, 23.59, 23.45, 21.28. HRMS (ESI-MS) Calc. for $[C_{17}H_{16}BrN_3OH]^+$ ($[M+H]$)$^+$: 358.0550, found 358.0554.

Figure 3. 24: Chiral SFC trace of product 2a



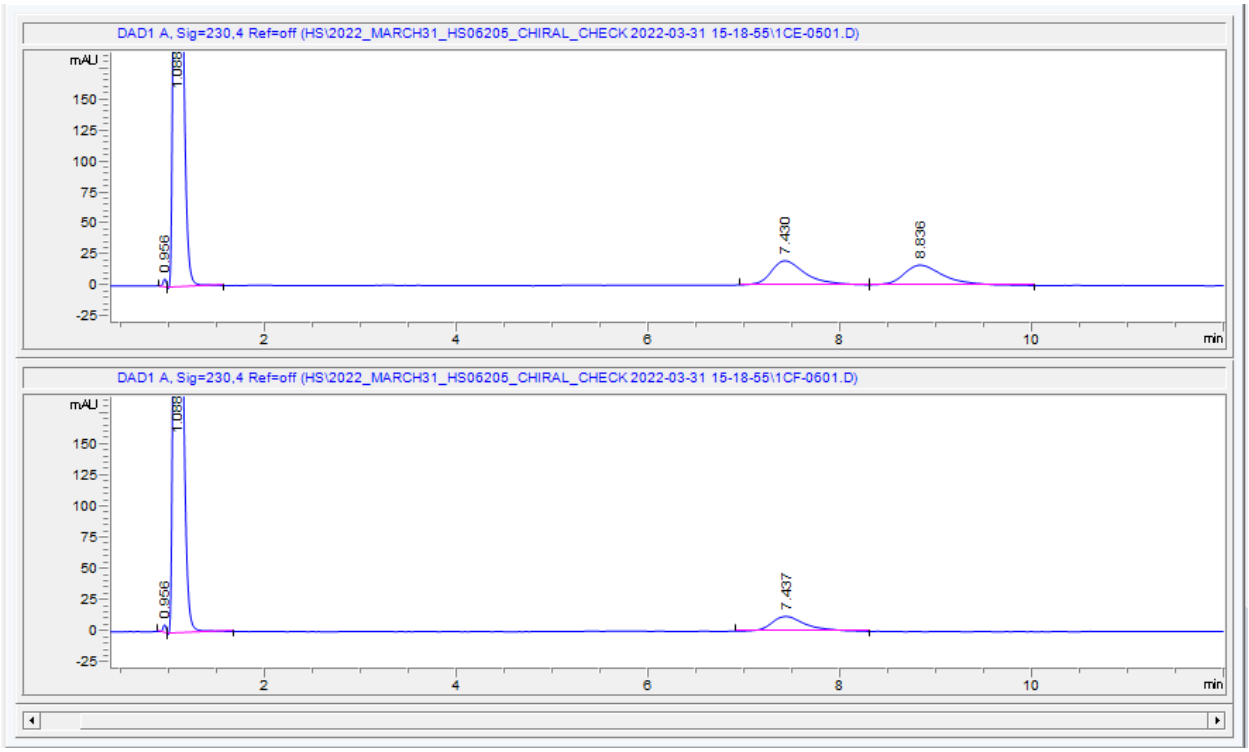
Top: **2a** obtained from NBS reaction. Bottom: 2**a** obtained with 3-T bioconversion.

Product **3a**



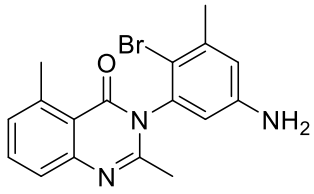
Product **3a** was initially purified from Biotage using method 2 following bioconversions with 3-T. **3a** was further purified via normal phase purification using 3:1 ethyl acetate:hexanes with 1% TEA. From bioconversions with 3-T, product **3a** was obtained in 57% yield (6.1 mg) as a white powder. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.17 (d, *J* = 8.1 Hz, 1H), 7.47 (s, 1H), 7.28 (d, *J* = 9.1 Hz, 1H), 6.70 (d, *J* = 2.7 Hz, 1H), 6.50 (d, *J* = 2.7 Hz, 1H), 3.78 (s, 2H), 2.51 (s, 3H), 2.40 (s, 3H), 2.26 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.52, 153.15, 146.72, 145.56, 140.89, 137.82, 136.70, 136.07, 126.63, 126.56, 120.43, 118.18, 113.30, 112.42, 23.59, 23.45, 21.28. HRMS (ESI-MS) Calc. for $[C_{17}H_{16}BrN_3OH]^+$ ($[M+H]$)$^+$: 358.0550, found 358.0553.

Figure 3. 25: Chiral SFC trace of product 3a.



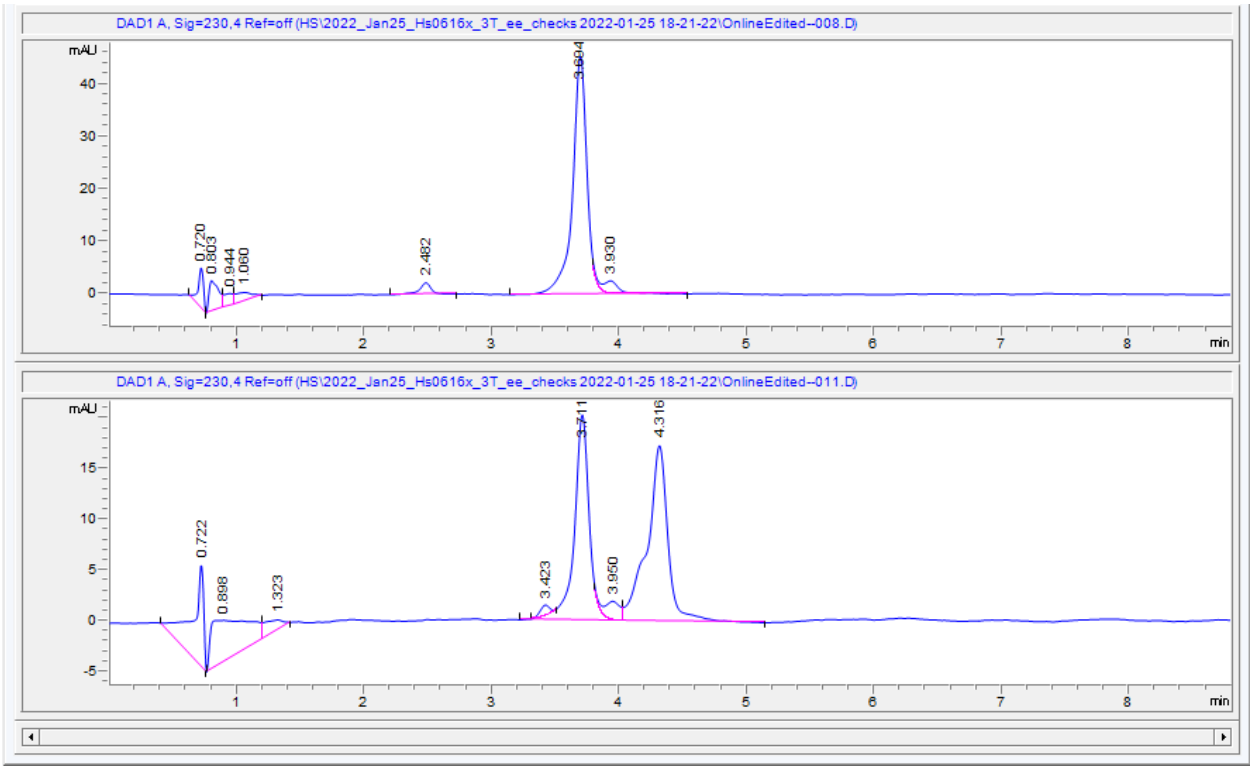
Top: **3a** obtained from NBS reaction. Bottom: **3a** obtained with 3-T bioconversion.

Product **4a**



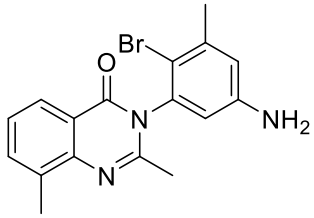
Product **4a** was initially purified from Biotage using method 2. Additional normal phase purification was performed using 3:1 ethyl acetate:hexanes with 1% TEA. From bioconversions with 3-T, product **4a** was obtained as a white film in 39% yield (4.1 mg). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.60 (t, *J* = 7.8 Hz, 1H), 7.53 (d, *J* = 8.1 Hz, 1H), 7.22 (d, *J* = 7.3 Hz, 1H), 6.70 (d, *J* = 2.7 Hz, 1H), 6.50 (d, *J* = 2.7 Hz, 1H), 2.84 (s, 3H), 2.40 (s, 3H), 2.25 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 162.02, 153.88, 149.00, 146.71, 141.68, 140.97, 137.90, 133.79, 129.34, 124.89, 119.15, 118.11, 113.31, 112.50, 23.59, 23.28, 23.01. HRMS (ESI-MS) Calc. for $[C_{17}H_{16}BrN_3OH]^+$ ($[M+H]$)$^+$: 358.0550, found 358.0552.

Figure 3. 26: Chiral SFC trace of product 4a



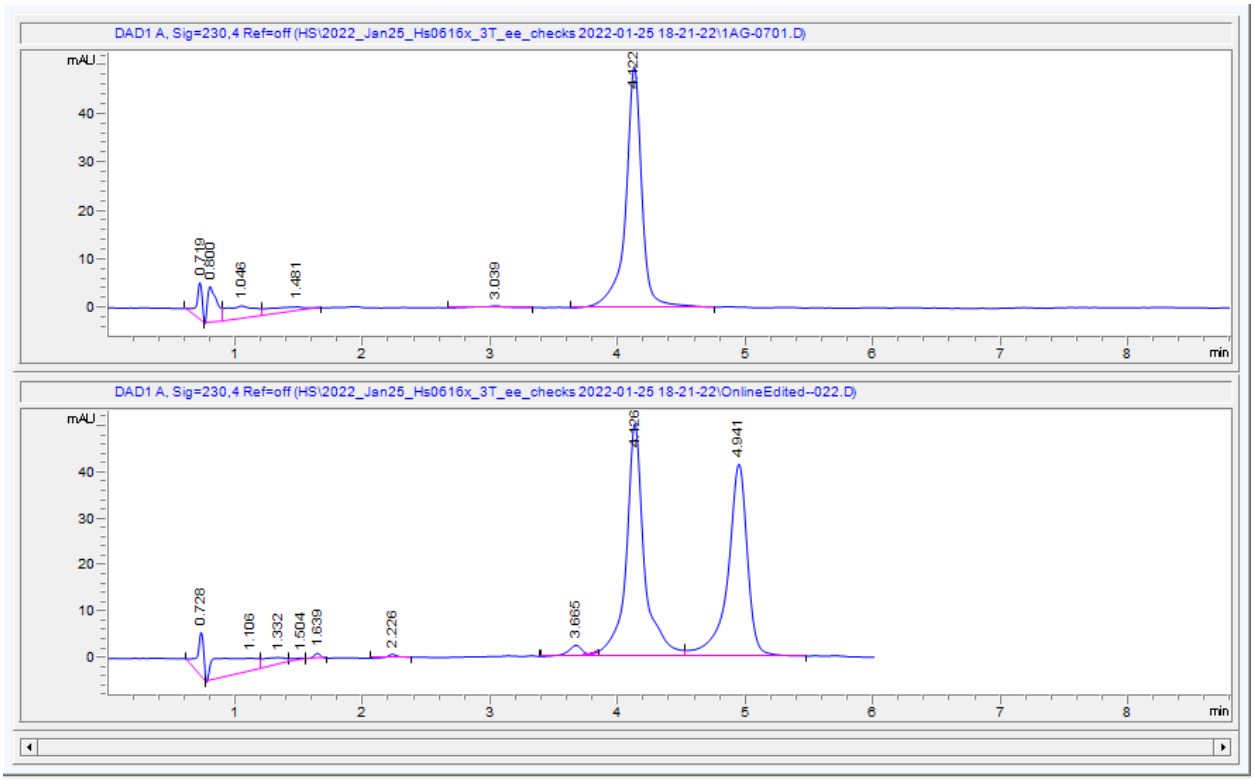
Top: **4a** obtained from NBS reaction. Bottom: 4**a** obtained with 3-T bioconversion.

Product **5a**



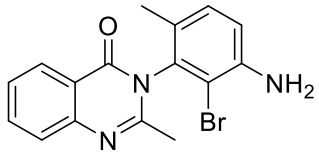
Product **5a** was initially purified from Biotage using method 2. Additional normal phase purification was performed using 3:1 ethyl acetate:hexanes with 1% TEA. From bioconversions with 3-T, product **5a** was obtained as a flaky white solid in 37% yield (3.9 mg). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.14 (d, 1H), 7.61 (d, *J* = 7.3 Hz, 1H), 7.34 (t, *J* = 7.6 Hz, 1H), 6.70 (d, *J* = 2.7 Hz, 1H), 6.49 (d, *J* = 2.7 Hz, 1H), 2.65 (s, 3H), 2.40 (s, 3H), 2.29 (s, 3H). 13C NMR (126 MHz, $CDCl_3$) δ 162.26, 152.80, 146.98, 146.10, 140.78, 137.53, 135.42, 135.34, 126.15, 124.58, 120.40, 118.15, 113.10, 111.91, 23.40, 23.38, 17.26. HRMS (ESI-MS) Calc. for $[C_{17}H_{16}BrN_3OH]^+$ ($[M+H]$)$^+$: 358.0550, found 3358.0552.

Figure 3. 27: Chiral SFC trace of product 5a



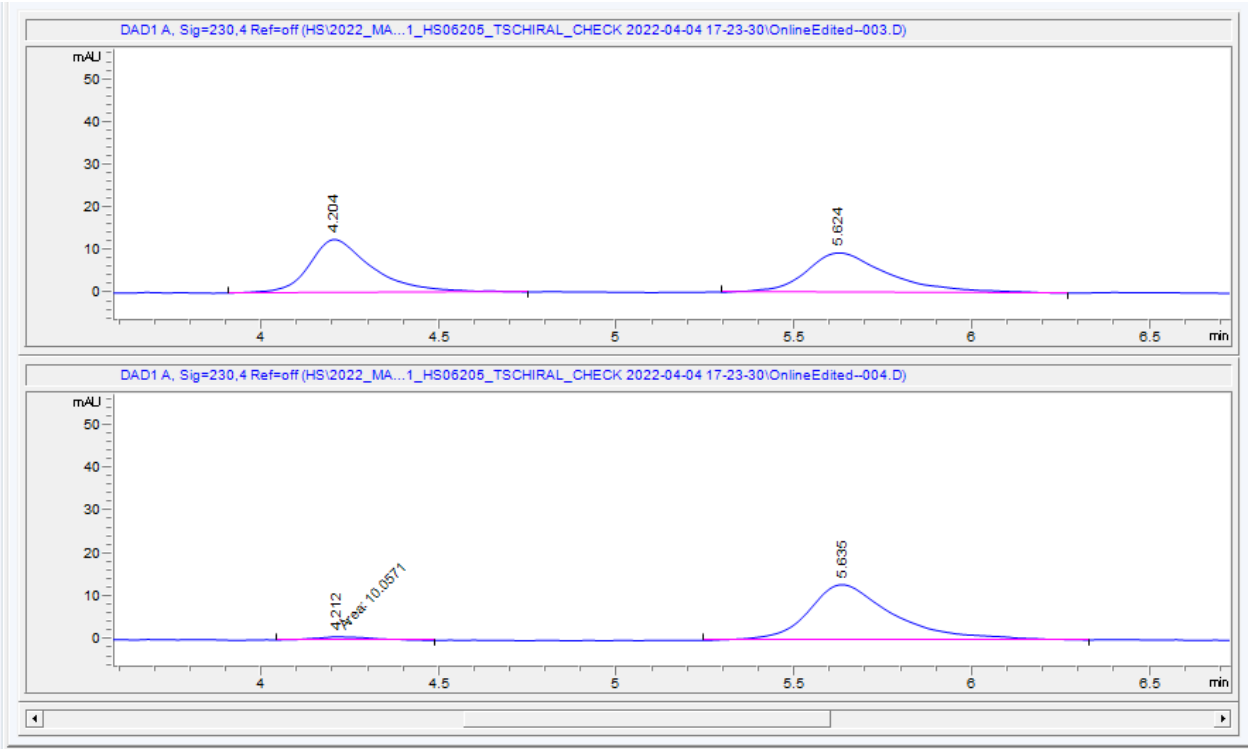
Top: **5a** obtained from 3-T bioconversion. Bottom: **5a** obtained from NBS reaction.

Product **6a**



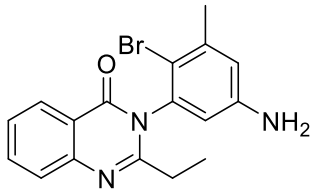
After bioconversions the reaction was processed as previously reported and purified by Biotage method 2, which allowed for the recovery of enantioenriched substrate **6** and products **6a** and **6b**. Product **6a** was further purified using semiprep method 4. From bioconversions with 3-T, 6a was obtained in 15% yield (1.5 mg) as a white film. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.29 (d, *J* = 7.8 Hz, 1H), 7.76 (t, *J* = 7.3 Hz, 1H), 7.69 (d, *J* = 8.2 Hz, 1H), 7.46 (t, *J* = 7.5 Hz, 1H), 7.12 (d, *J* = 8.3 Hz, 1H), 6.82 (d, *J* = 8.3 Hz, 1H), 2.19 (s, 3H), 2.04 (s, 3H). 13C NMR (126 MHz, $CDCl_3$) δ 160.80, 154.01, 147.75, 144.12, 136.22, 134.68, 130.42, 127.28, 126.89, 126.57, 126.08, 120.75, 116.52, 109.36, 23.00, 17.61. HRMS (ESI-MS) Calc. for $[C_{16}H_{14}BrN_3OH]^+$ ($[M+H]$)$^+$: 344.0393, found 344.0395.

Figure 3. 28: Chiral SFC trace of product 6a



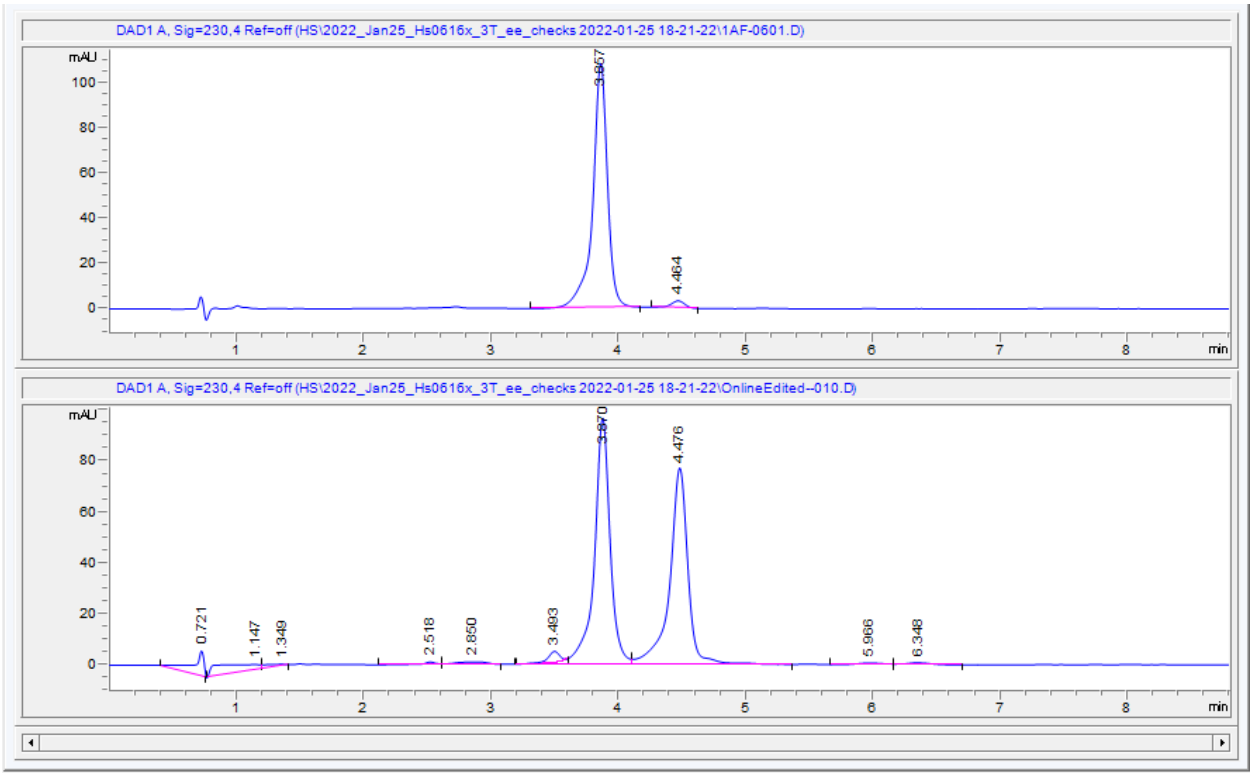
Top: **6a** obtained from NBS reaction. Bottom: **6a** obtained from bioconversion with 3-T.

Product **8a**



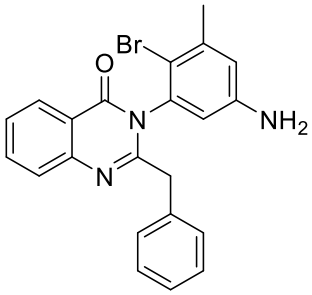
Product **8a** was purified by Biotage method 2. From bioconversions with 3-T, 8a was obtained as a white powder in 33% yield (3.5 mg). ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.28 (d, *J* = 8.0 Hz, 1H), 7.79 – 7.70 (m, 2H), 7.46 (d, *J* = 7.9 Hz, 1H), 6.69 (s, 1H), 6.47 (s, 1H), 2.46 (d, *J* = 7.9 Hz, 2H), 2.39 (s, 3H), 1.27 (t, *J* = 8.2 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.66, 157.43, 147.66, 146.51, 140.89, 137.39, 134.48, 127.14, 126.45, 120.79, 118.17, 113.64, 112.83, 28.52, 23.62, 10.90. HRMS (ESI-MS) Calc. for $[C_{17}H_{16}BrN_3OH]^+$ ($[M+H]$)$^+$: 358.0550, found 358.0551.

Figure 3. 29: Chiral SFC trace of product 8a



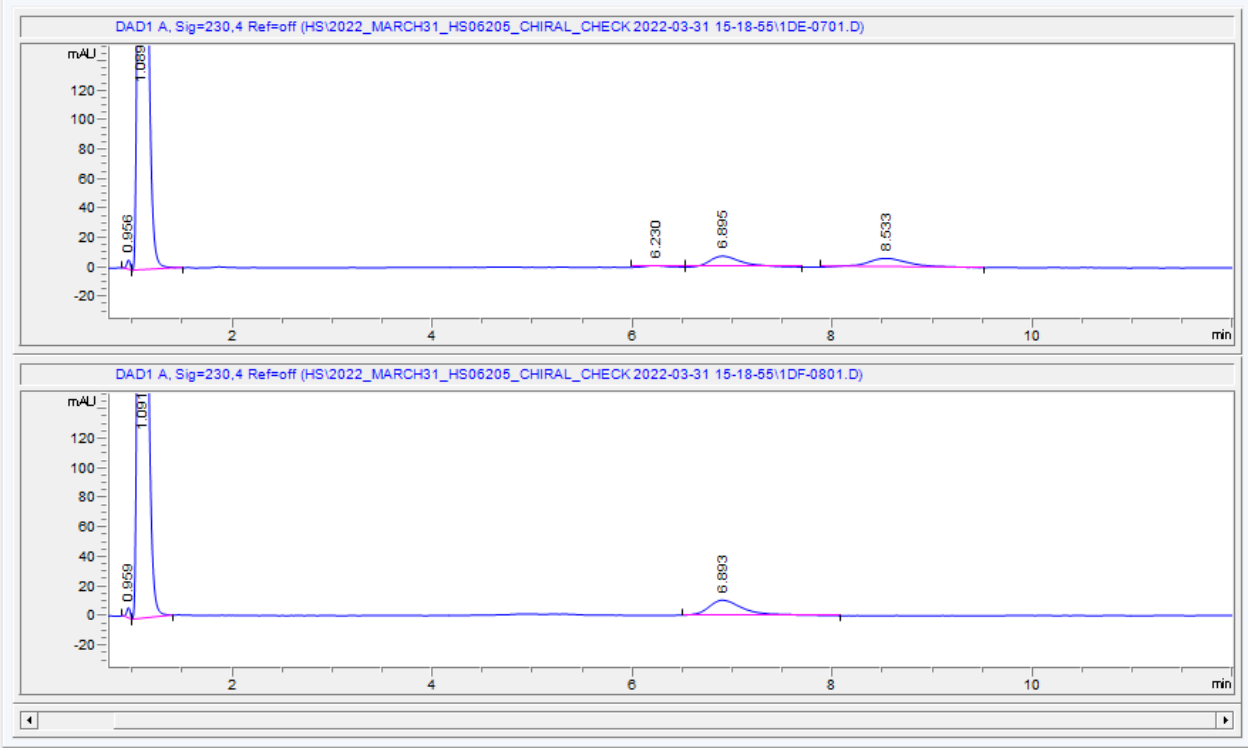
Top: **8a** obtained from 3-T bioconversion. Bottom: **8a** obtained from NBS reaction.

Product **9a**



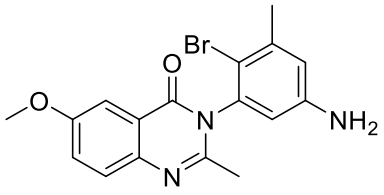
Product **9a** was purified by Biotage using the same method as for substrate **9**. Bioconversions with 3-T generated a third monobrominated product that could not be separated from the desired product 9a, suspected to be 2'-brominated material based on the proton NMR. This additional product could not be isolated for characterization. From bioconversions with 3-T, product **9a** was obtained as a white film in 47% yield (2.9 mg).^{1}H NMR (600 MHz, $CDCl_3$) δ 8.29 (d, *J* = 7.9 Hz, 1H), 7.82 – 7.76 (m, 2H), 7.49 (ddd, *J* = 8.2, 5.8, 2.3 Hz, 1H), 7.24 – 7.14 (m, 3H), 6.99 (d, 2H), 6.66 (s, 1H), 5.97 (s, 1H), 4.03 (d, *J* = 14.8 Hz, 1H), 3.73 (d, *J* = 14.9 Hz, 1H), 2.37 (s, 3H). ^{13}C NMR (126 MHz, Chloroform-*d*) δ 161.69, 154.97, 147.50, 146.03, 140.46, 136.62, 135.36, 134.57, 129.03, 128.20, 127.37, 127.18, 126.88, 126.79, 120.92, 118.04, 114.87, 112.89, 42.45, 23.58. HRMS (ESI-MS) Calc. for $[C_{22}H_{18}BrN_3OH]^+$ ($[M+H]$)$^+$: 420.0706, found 420.0708.

Figure 3. 30: Chiral SFC trace of product 9a



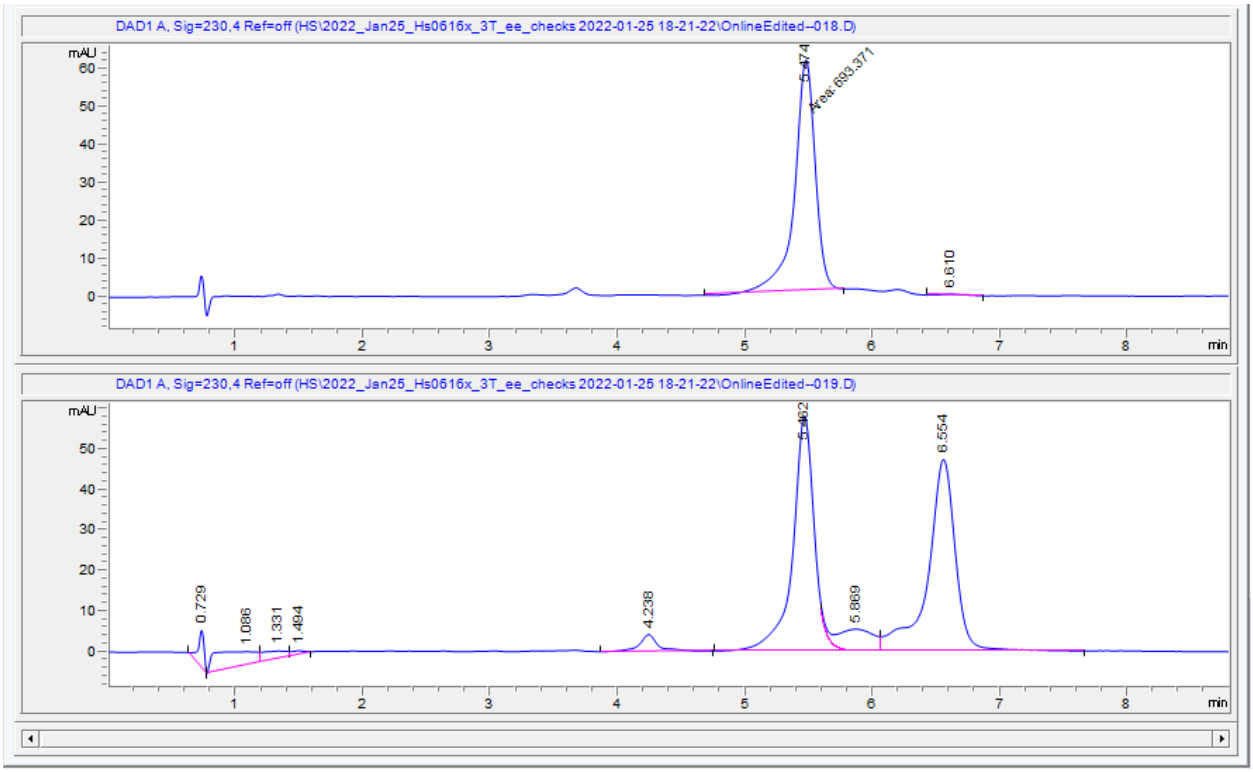
Top: **9a** obtained from NBS reaction. Bottom: **9a** obtained from 3-T bioconversion.

Product **11a**



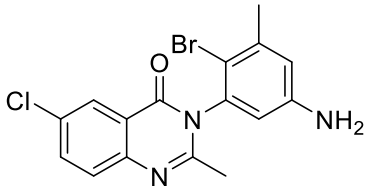
Product **11a** was purified by reverse phase chromatography using Biotage method 2. From bioconversions with 3-T, 11a was obtained as a white solid in 78% yield (8.7 mg). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.68 – 7.59 (m, 2H), 7.37 (dd, *J* = 8.9, 3.0 Hz, 1H), 6.71 (d, *J* = 2.7 Hz, 1H), 6.50 (d, *J* = 2.8 Hz, 1H), 3.91 (s, 3H), 2.41 (s, 3H), 2.26 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.46, 158.19, 151.68, 146.74, 142.22, 140.93, 137.88, 128.42, 124.79, 121.43, 118.20, 113.24, 112.40, 106.63, 55.81, 23.60, 23.30. HRMS (ESI-MS) Calc. for $[C_{17}H_{17}BrN_3O_2H]^+$ ([M+H])$^+$: 374.0499, found 374.0498.

Figure 3. 31: Chiral SFC trace of product 11a



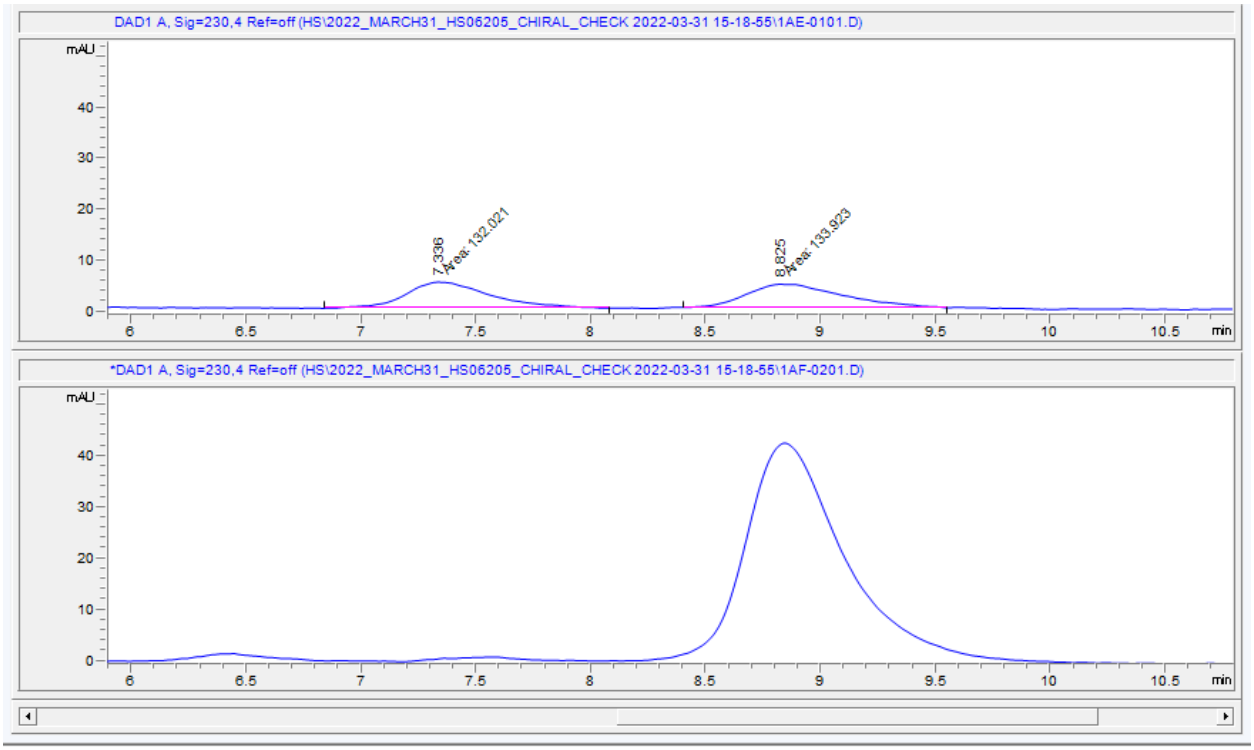
Top: **11a** obtained from 3-T bioconversion. Bottom: **11a** obtained from NBS reaction.

Product **12a**



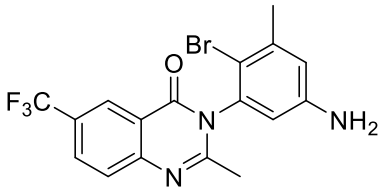
Product **12a** was purified by reverse phase chromatography using Biotage method 2. From bioconversions with 3-T, 12a was obtained as a slightly gray film in 78% yield (8.8 mg). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.22 (d, *J* = 2.4 Hz, 1H), 7.68 (dd, *J* = 8.7, 2.4 Hz, 1H), 7.61 (d, *J* = 8.6 Hz, 1H), 6.70 (d, 1H), 6.47 (d, *J* = 2.7 Hz, 1H), 2.39 (s, 3H), 2.25 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 160.50, 154.46, 146.79, 146.11, 141.10, 137.40, 135.03, 132.32, 128.57, 126.48, 121.81, 118.36, 113.07, 112.17, 23.58, 23.52. HRMS (ESI-MS) Calc. for $[C_{16}H_{13}BrClN_3OH]^+$ ($[M+H]$)$^+$: 379.9982, found 379.9983.

Figure 3. 32: Chiral SFC trace of product 12a



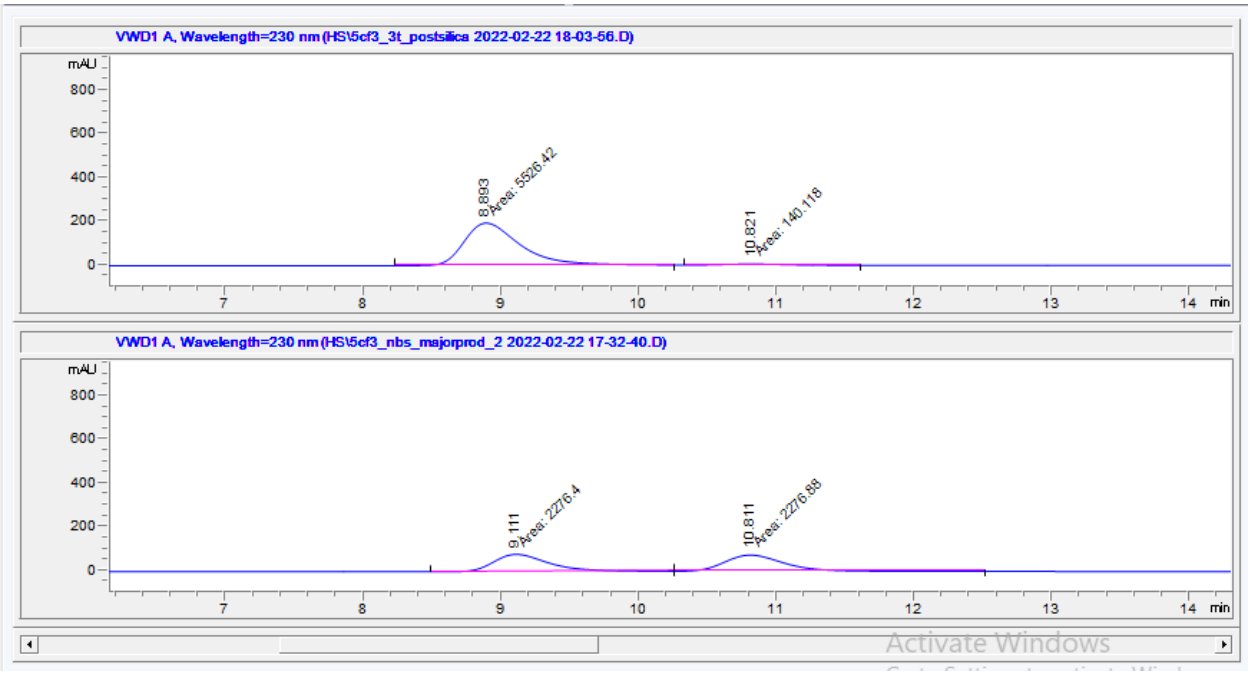
Top: **12a** obtained from NBS reaction. Bottom: **12a** obtained from bioconversion with 3-T.

Product **13a**



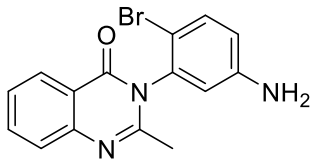
Product **13a** was initially purified by reverse phase chromatography using Biotage method 2, which provided a mixture of products. Further purification of **13a** was performed by normal phase chromatography with 2:1 ethyl acetate:hexanes with 1% TEA as the mobile phase. From bioconversions with 3-T, **13a** was obtained as a white powder in 17% yield (2.1 mg). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 8.56 (s, 1H), 7.95 (dd, *J* = 8.4, 2.1 Hz, 1H), 7.78 (d, *J* = 8.5 Hz, 1H), 6.71 (d, *J* = 2.7 Hz, 1H), 6.48 (d, *J* = 2.7 Hz, 1H), 3.84 (s, 2H), 2.40 (s, 3H), 2.30 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 160.72, 156.49, 149.74, 146.90, 141.14, 137.15, 130.82, 130.80, 130.77, 128.90, 128.63, 128.37, 128.10, 127.89, 126.94, 125.14, 125.11, 125.07, 125.04, 124.78, 122.61, 120.67, 118.43, 112.93, 111.94, 23.69, 23.57. HRMS (ESI-MS) Calc. for $[C_{17}H_{13}BrF_3N_3OH]^+$ ([M+H])$^+$: 412.0267, found 412.0269.

Figure 3. 33: Chiral HPLC trace of product 13a



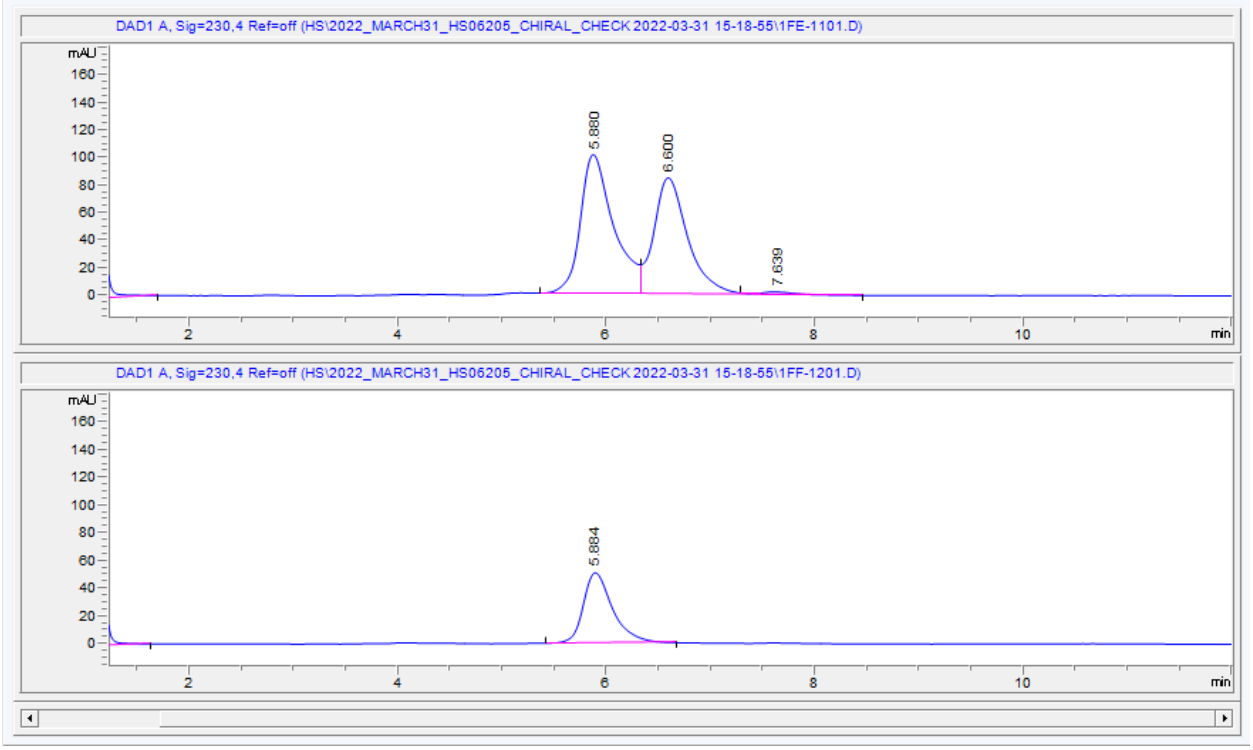
Top: **13a** obtained from bioconversion with 3-T. Bottom: **13a** obtained from NBS reaction

Product **14a**



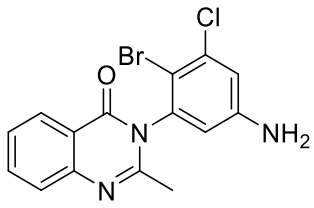
Product **14a** was purified from bioconversions by reverse phase chromatography using Biotage method 2. From bioconversions with 3-T, 14a was obtained as white powder in 85% yield (8.4 mg). ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.18 (dd, *J* = 7.9, 1.5 Hz, 1H), 7.67 (ddd, *J* = 8.5, 7.1, 1.6 Hz, 1H), 7.61 – 7.56 (m, 1H), 7.37 (dd, *J* = 8.4, 6.6 Hz, 2H), 6.57 (dd, *J* = 8.6, 2.7 Hz, 1H), 6.53 (d, *J* = 2.7 Hz, 1H), 3.84 (s, 2H), 2.18 (s, 3H). ^{13}C NMR (126 MHz, Chloroform-*d*) δ 161.46, 153.89, 147.55, 137.49, 134.69, 134.12, 127.17, 126.86, 126.63, 120.67, 117.56, 115.63, 109.62, 23.57. HRMS (ESI-MS) Calc. for $[C_{15}H_{12}BrN_3OH]^+$ ($[M+H]$)$^+$: 330.0237, found 330.0239.

Figure 3. 34: Chiral SFC trace of product 14a



Top: **14a** obtained from NBS reaction**.** Bottom: **14a** obtained from bioconversion with 3-T.

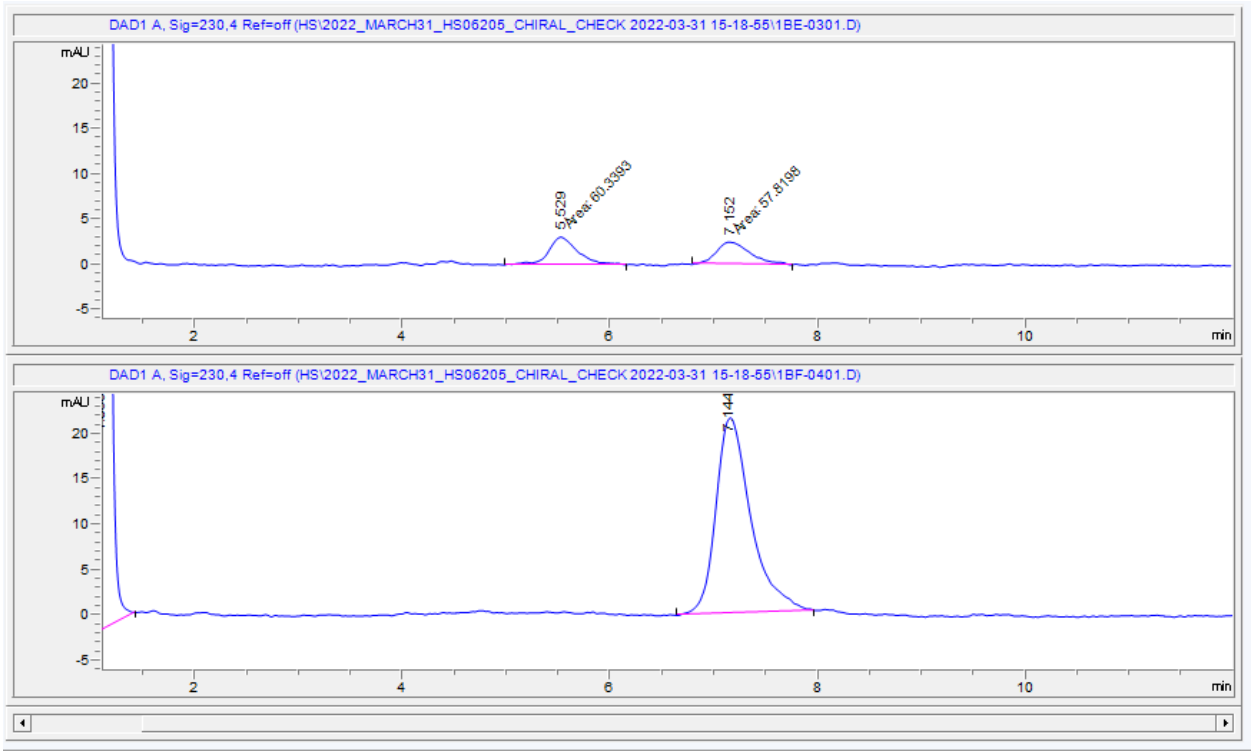
Product **15a**



Product **15a** was purified from bioconversions by reverse phase chromatography using Biotage method 2. From bioconversions with 3-T, 15a was obtained in 41% yield (4.4 mg) as white film.

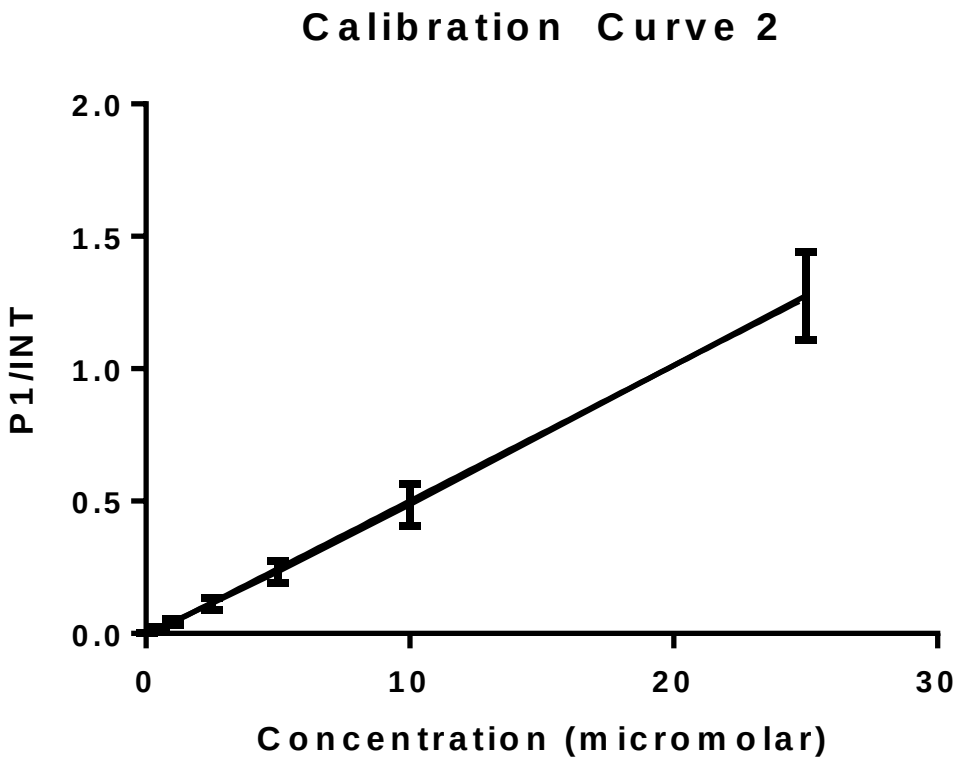
^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.28 (dd, *J* = 8.0, 1.5 Hz, 1H), 7.78 (ddd, *J* = 8.5, 7.1, 1.6 Hz, 1H), 7.69 (d, *J* = 8.1 Hz, 1H), 7.53 – 7.39 (m, 1H), 6.92 (d, *J* = 2.7 Hz, 1H), 6.56 (d, *J* = 2.6 Hz, 1H), 3.98 (s, 2H), 2.29 (s, 3H).^{13}C NMR (126 MHz, Chloroform-*d*) δ 161.31, 153.43, 147.48, 147.41, 139.11, 136.94, 134.84, 127.17, 126.93, 126.77, 120.57, 117.24, 114.31, 110.43, 23.50. HRMS (ESI-MS) Calc. for $[C_{15}H_{11}BrClN_3OH]^+$ $([M+H])^+$: 365.9825, found 365.9827.

Figure 3. 35: Chiral SFC trace of product 15a



Top: **15a** obtained from NBS reaction**.** Bottom: **15a** obtained from bioconversion with 3-T.

Figure 3. 36: Calibration curve for 1a relative to internal standard (indole-3-acetic acid).

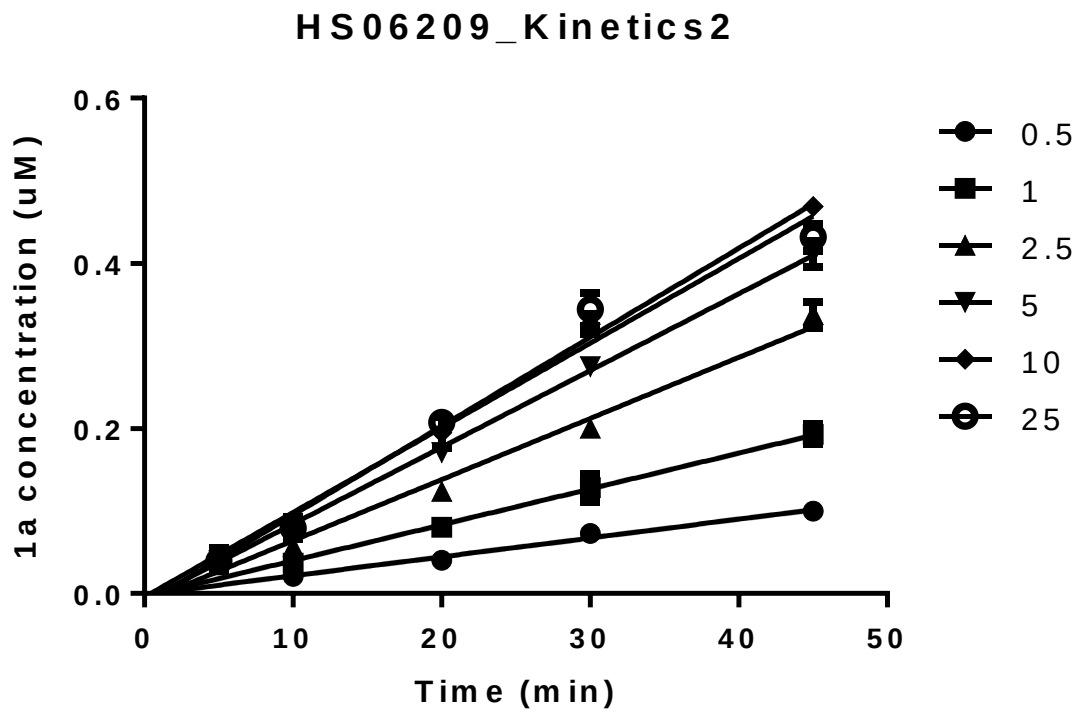


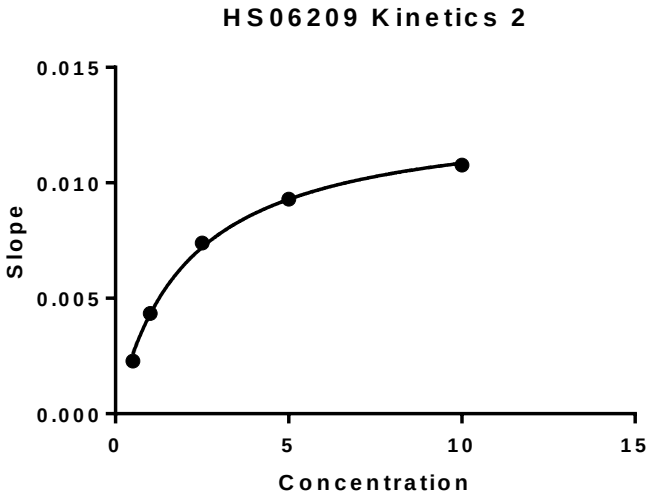
Best-fit values	
Slope	0.05113 ± 0.001662
Y-intercept when X=0.0	-0.01080 ± 0.01729
X-intercept when Y=0.0	0.2113

Goodness of Fit	
R square	0.9803

Integrations were taken from the 230 nm chromatogram. Measurements were taken in triplicate and are represented with the standard deviation.

Figure 3. 37: Initial rate analysis and Michaelis-Menten kinetics for 3-T with substrate 1

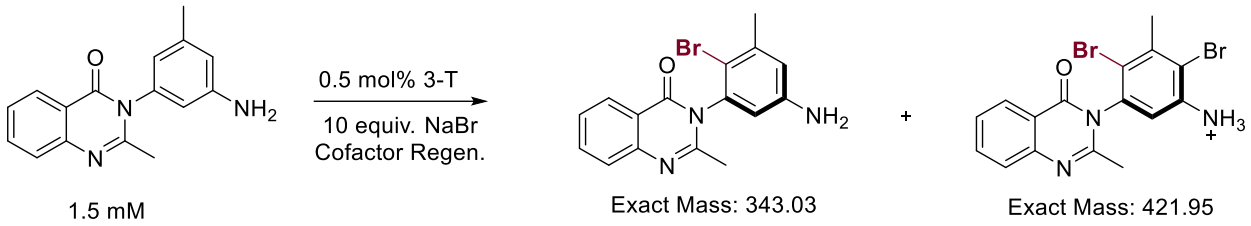




kcat	
Best-fit values	
Et	= 0.5000
kcat	0.02611
Km	2.032
Vmax	= 0.01306
Std. Error	
kcat	0.0004563
Km	0.1027

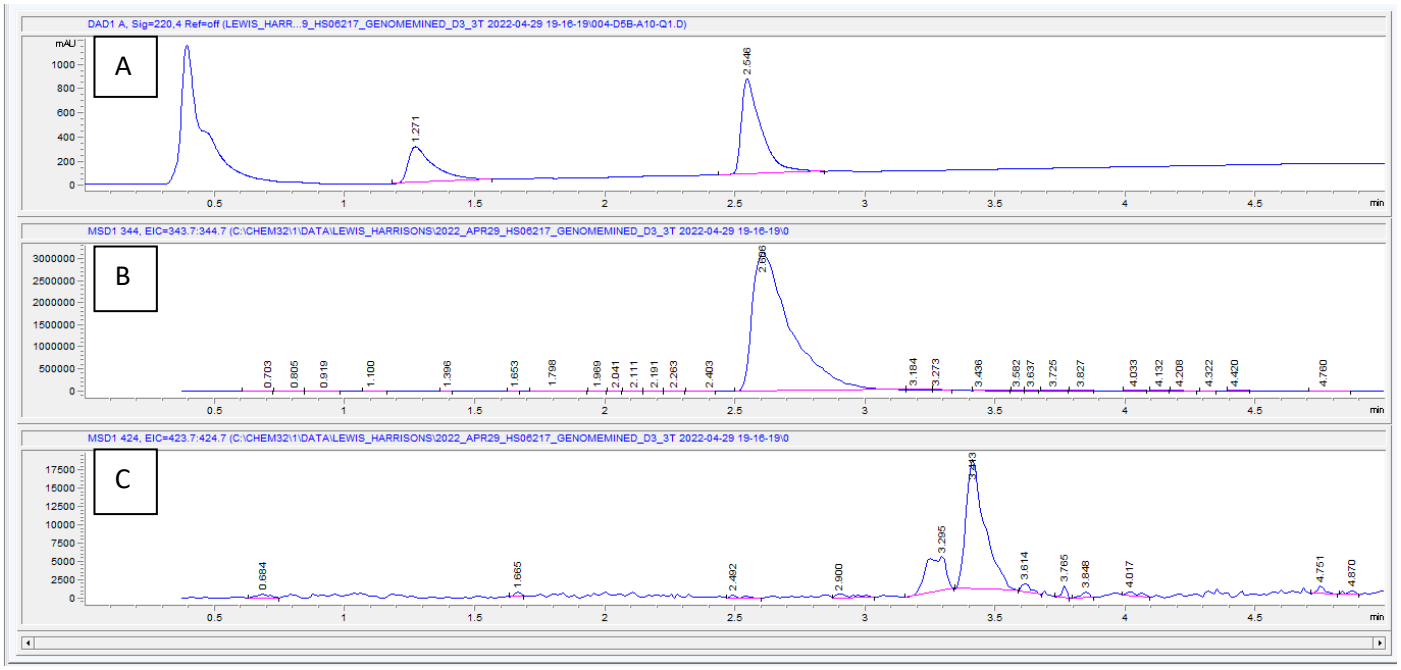
Values were plotted in Graphpad, which was also used to calculate initial rates for each substrate concentration, KM, and kcat. Values are plotted with the standard deviation of three measurements each.

3.4.5 – LCMS to confirm trace dibromination

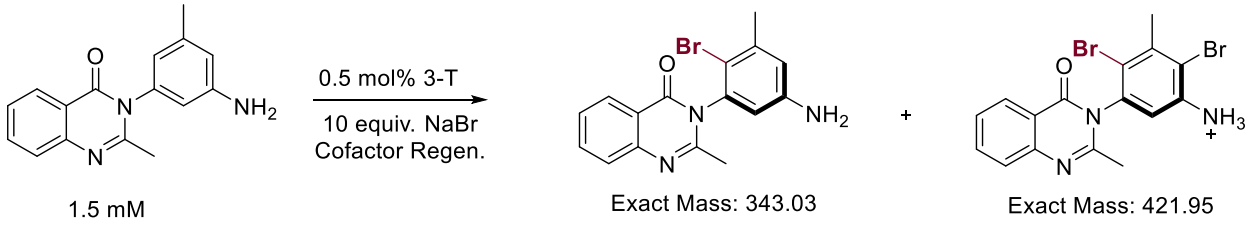


To verify the lack of di-bromination, a reaction of 3-T with substrate **1** was set up according to the same protocol and procedure as for the lineage reactions discussed in earlier in this manuscript. After quenching the reactions and filtering them as previously described, these reactions were analyzed via UHPLC-MS using UHPLC-MS method 1.

Figure 3. 38: UHPLC-MS analysis of the 3-T halogenation of substrate 1.

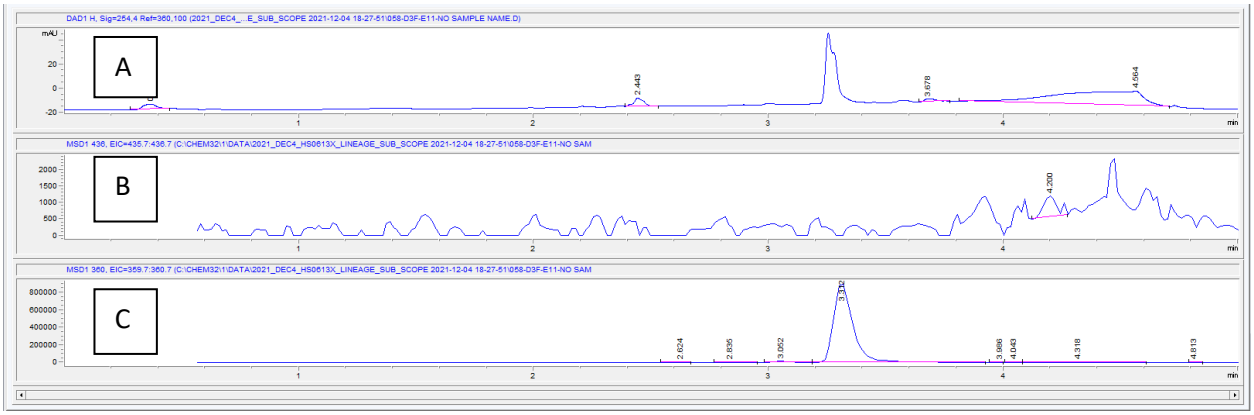


A) UV trace measured at 220 nm. B) EIC for the major ion of the monobrominated product. Note, due to the high conversion to **1a** and the inefficient peak sharpening from the formic acid additive, **1a** and **1b** overlap. C) EIC for the major ion of the di-brominated products, showing that although these products are detectable they are formed in trace yield (extracted ion intensity two orders of magnitude lower than monobrominated product).



To verify the lack of di-bromination, a reaction of 3-T with substrate **5** was set up according to the same protocol and procedure as for the lineage reactions discussed in earlier in this manuscript. After quenching the reactions and filtering them as previously described, these reactions were analyzed via UHPLC-MS using UHPLC-MS method 1.

Figure 3. 39: UHPLC-MS analysis of the 3-T halogenation of substrate 5



A) UV trace measured at 254 nm. B) EIC for the major ion of the di-brominated products, showing that despite the low regioselectivity for the initial halogenation event, no subsequent di-halogenation of the product is observed. C) EIC for the major ion of the monobrominated product. Note, due to the high conversion of substrate **5** to **5a** and **5b** and the inefficient peak sharpening from the formic acid additive, **5a** and **5b** overlap.

3.5 References

(1) Laplante, S. R.; Fader, L. D.; Fandrick, K. R.; Fandrick, D. R.; Hucke, O.; Kemper, R.; Miller, S. P. F.; Edwards, P. J. Assessing Atropisomer Axial Chirality in Drug Discovery and Development. *J Med Chem* **2011**, *54* (20), 7005–7022. https://doi.org/10.1021/jm200584g.

(2) Bringmann, G.; Gulder, T.; Gulder, T. A. M.; Breuning, M. Atroposelective Total Synthesis of Axially Chiral Biaryl Natural Products. *Chem Rev* **2011**, *111* (2), 563–639. https://doi.org/10.1021/cr100155e.

(3) Wencel-Delord, J.; Panossian, A.; Leroux, F. R.; Colobert, F. Recent Advances and New Concepts for the Synthesis of Axially Stereoenriched Biaryls. *Chem Soc Rev* **2015**, *44* (11), 3418–3430. https://doi.org/10.1039/c5cs00012b.

(4) Carmona, J. A.; Rodríguez-Franco, C.; Fernández, R.; Hornillos, V.; Lassaletta, J. M. Atroposelective Transformation of Axially Chiral (Hetero)Biaryls. From Desymmetrization to Modern Resolution Strategies. *Chemical Society Reviews*. 2021, pp 2968–2983. https://doi.org/10.1039/d0cs00870b.

(5) Mazzaferro, L. S.; Hüttel, W.; Fries, A.; Müller, M. Cytochrome P450-Catalyzed Regio- and Stereoselective Phenol Coupling of Fungal Natural Products. *J Am Chem Soc* **2015**, *137* (38), 12289–12295. https://doi.org/10.1021/jacs.5b06776.

(6) Gagnon, C.; Godin, É.; Minozzi, C.; Sosoe, J.; Pochet, C.; Collins, S. K. Biocatalytic Synthesis of Planar Chiral Macrocycles. *Science (1979)* **2020**, *367* (6480), 917–921. https://doi.org/10.1126/science.aaz7381.

(7) Zetzsche, L. E.; Yazarians, J. A.; Chakrabarty, S.; Hinze, M. E.; Murray, L. A. M.; Lukowski, A. L.; Joyce, L. A.; Narayan, A. R. H. Biocatalytic Oxidative Cross-Coupling Reactions for Biaryl Bond Formation. *Nature* **2022**, *603* (7899), 79–85. https://doi.org/10.1038/s41586-021-04365-7.

(8) Barrett, K. T.; Miller, S. J. Enantioselective Synthesis of Atropisomeric Benzamides through Peptide-Catalyzed Bromination. *J Am Chem Soc* **2013**, *135* (8), 2963–2966. https://doi.org/10.1021/ja400082x.

(9) Gustafson, J. L.; Lim, D.; Mille, S. J. Dynamic Kinetic Resolution of Biaryl Atropisomers via Peptide-Catalyzed Asymmetrie Bromination. *Science (1979)* **2010**, *328* (5983), 1251–1255. https://doi.org/10.1126/science.1188403.

(10) Diener, M. E.; Metrano, A. J.; Kusano, S.; Miller, S. J. Enantioselective Synthesis of 3-Arylquinazolin-4(3H)-Ones via Peptide-Catalyzed Atroposelective Bromination. *J Am Chem Soc* **2015**, *137* (38), 12369–12377. https://doi.org/10.1021/jacs.5b07726.

(11) Metrano, A. J.; Miller, S. J. Peptide-Based Catalysts Reach the Outer Sphere through Remote Desymmetrization and Atroposelectivity. *Acc Chem Res* **2019**, *52* (1), 199–215. https://doi.org/10.1021/acs.accounts.8b00473.

(12) Gustafson, J. L.; Lim, D.; Barrett, K. T.; Miller, S. J. Synthesis of Atropisomerically Defined, Highly Substituted Biaryl Scaffolds through Catalytic Enantioselective Bromination and Regioselective Cross-Coupling. *Angewandte Chemie* **2011**, *123* (22), 5231–5235. https://doi.org/10.1002/ange.201101147.

(13) Payne, J. T.; Andorfer, M. C.; Lewis, J. C. Regioselective Arene Halogenation Using the FAD-Dependent Halogenase RebH. *Angewandte Chemie - International Edition* **2013**, *52* (20), 5271–5274. https://doi.org/10.1002/ANIE.201300762.

(14) Andorfer, M. C.; Park, H. J.; Vergara-Coll, J.; Lewis, J. C. Directed Evolution of RebH for Catalyst-Controlled Halogenation of Indole C-H Bonds. *Chem Sci* **2016**, *7* (6), 3720–3729. https://doi.org/10.1039/c5sc04680g.

(15) Shepherd, S. A.; Karthikeyan, C.; Latham, J.; Struck, A. W.; Thompson, M. L.; Menon, B. R. K.; Styles, M. Q.; Levy, C.; Leys, D.; Micklefield, J. Extending the Biocatalytic Scope of Regiocomplementary Flavin-Dependent Halogenase Enzymes. *Chem Sci* **2015**, *6* (6), 3454–3460. https://doi.org/10.1039/c5sc00913h.

(16) Glenn, W. S.; Nims, E.; O'Connor, S. E. Reengineering a Tryptophan Halogenase to Preferentially Chlorinate a Direct Alkaloid Precursor. *J Am Chem Soc* **2011**, *133* (48), 19346–19349. https://doi.org/10.1021/ja2089348.

(17) Shepherd, S. A.; Menon, B. R. K.; Fisk, H.; Struck, A. W.; Levy, C.; Leys, D.; Micklefield, J. A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase. *ChemBioChem* **2016**, *17* (9), 821–824. https://doi.org/10.1002/cbic.201600051.

(18) Payne, J. T.; Butkovich, P. H.; Gu, Y.; Kunze, K. N.; Park, H. J.; Wang, D. S.; Lewis, J. C. Enantioselective Desymmetrization of Methylenedianilines via Enzyme-Catalyzed Remote Halogenation. *J Am Chem Soc* **2018**, *140* (2), 546–549. https://doi.org/10.1021/jacs.7b09573.

(19) Mondal, D.; Fisher, B. F.; Jiang, Y.; Lewis, J. C. Flavin-Dependent Halogenases Catalyze Enantioselective Olefin Halocyclization. *Nat Commun* **2021**, *12* (1), 1–7. https://doi.org/10.1038/s41467-021-23503-3.

(20) Jiang, Y.; Mondal, D.; Lewis, J. C. Expanding the Reactivity of Flavin Dependent Halogenases Toward Olefins via Enantioselective Intramolecular Haloetherification and Chemoenzymatic Oxidative Rearrangements Yuhua. *ChemRxiv* **2022**, *17* (July), 1–6. https://doi.org/10.1021/ACSCATAL.2C03383.

(21) Andorfer, M. C.; Evans, D.; Yang, S.; He, C. Q.; Girlich, A. M.; Vergara-Coll, J.; Sukumar, N.; Houk, K. N.; Lewis, J. C. Analysis of Laboratory-Evolved Flavin-Dependent Halogenases Affords a Computational Model for Predicting Halogenase Site Selectivity. *Chem Catalysis* **2022**. https://doi.org/10.1016/j.checat.2022.07.003.

(22) Snodgrass, H. M.; Mondal, D.; Lewis, J. C. Directed Evolution of Flavin-Dependent Halogenases for Site- and Atroposelective Halogenation of 3-Aryl-4(3 H)-Quinazolinones via Kinetic or Dynamic Kinetic Resolution . *J Am Chem Soc* **2022**, *4*. https://doi.org/10.1021/jacs.2c07422.

(23) Fisher, B. F.; Snodgrass, H. M.; Jones, K. A.; Andorfer, M. C.; Lewis, J. C. Site-Selective C-H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. *ACS Cent Sci* **2019**, *5* (11), 1844–1856. https://doi.org/10.1021/acscentsci.9b00835.

(24) Andorfer, M. C.; Grob, J. E.; Hajdin, C. E.; Chael, J. R.; Siuti, P.; Lilly, J.; Tan, K. L.; Lewis, J. C. Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal* **2017**, *7* (3), 1897–1904. https://doi.org/10.1021/acscatal.6b02707.

(25) Phintha, A.; Prakinee, K.; Jaruwat, A.; Lawan, N.; Visitsatthawong, S.; Kantiwiriyawanitch, C.; Songsungthong, W.; Trisrivirat, D.; Chenprakhon, P.; Mulholland, A.; van Pée, K. H.; Chitnumsub, P.; Chaiyen, P. Dissecting the Low Catalytic Capability of Flavin-Dependent Halogenases. *Journal of Biological Chemistry* **2021**, *296*, 100068. https://doi.org/10.1074/jbc.RA120.016004.

(26) Barrett, K. T.; Metrano, A. J.; Rablen, P. R.; Miller, S. J. Spontaneous Transfer of Chirality in an Atropisomerically Enriched Two-Axis System. *Nature* **2014**, *508* (7498), 71–75. https://doi.org/10.1038/nature13189.

(27) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(28) Hogrefe, H. H.; Cline, J.; Youngblood, G. L.; Allen, R. M. Creating Randomized Amino Acid Libraries with the Quikchange® Multi Site-Directed Mutagenesis Kit. *Biotechniques* **2002**, *33* (5), 1158–1165. https://doi.org/10.2144/02335pf01.

(29) Huang, C. C.; Meng, E. C.; Morris, J. H.; Pettersen, E. F.; Ferrin, T. E. Enhancing UCSF Chimera through Web Services. *Nucleic Acids Res* **2014**, *42* (W1), 478–484. https://doi.org/10.1093/nar/gku377.

(30) Cournoyer; Richard Leo; Keitz; Paul Francis; O'Yang; Counde; Yasuda; Dennis Mitsugu. 2-Imidazoline, 2-Oxazoline, 2-Thiazoline, and 4-Imidazole Derivatives of Methylphenyl, Methoxyphenyl, and Aminophenyl Alkylsulfonamides and Ureas and Their Use, 1999.

(31) Poor, C. B.; Andorfer, M. C.; Lewis, J. C. Improving the Stability and Catalyst Lifetime of the Halogenase RebH by Directed Evolution. *ChemBioChem* **2014**, *15* (9), 1286–1289. https://doi.org/10.1002/cbic.201300780.

(32) Payne, J. T.; Poor, C. B.; Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angewandte Chemie* **2015**, *127* (14), 4300–4304. https://doi.org/10.1002/ange.201411901.

(33) Sambrook, J. *Molecular Cloning. A Laboratory Manual*; Third edition. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, [2001] ©2001, 1983; Vol. 11. https://doi.org/10.1016/0307-4412(83)90068-7.

(34) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Methods* **2009**, *6* (5), 343–345. https://doi.org/10.1038/nmeth.1318.

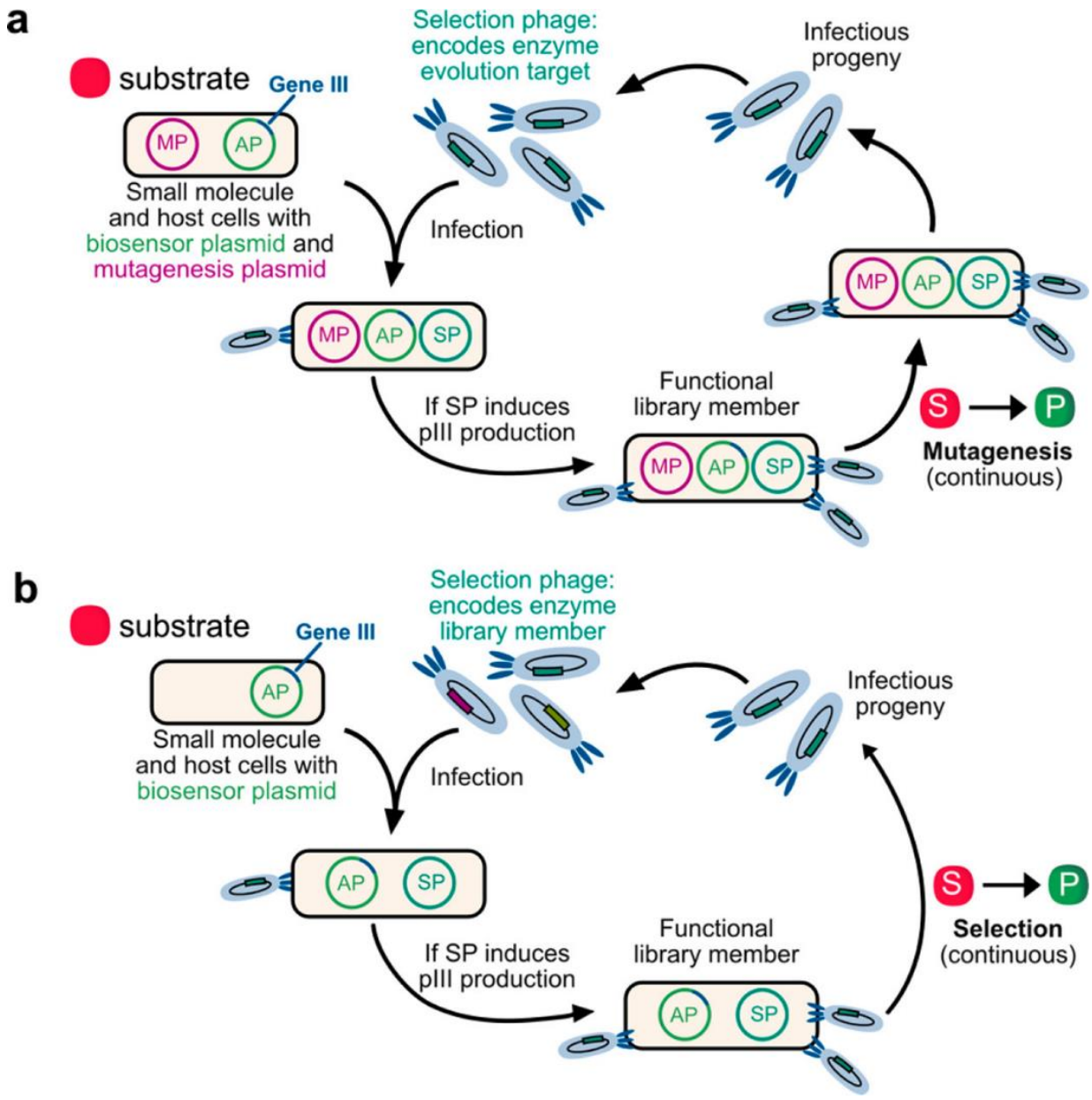
CHAPTER 4: PHAGE-ASSISTED CONTINUOUS EVOLUTION AND SELECTION OF AN ESTERASE

4.1: Introduction

The synthetic value of enzymes to accomplish challenging chemical reactions under environmentally benign conditions has driven the development of biocatalysts in chemistry.[1] Hydrolases in particular have found wide use for site- and enantio-specific ester bond hydrolysis and have been valuable for both early and late-stage modifications of a variety of compounds. While in some cases naturally occurring enzymes exhibit properties that allow them to be readily applied to synthesis, most biocatalysts require optimization.[2] Through multiple rounds of *in vitro* mutagenesis, protein expression, and bioconversion screening (typically via HPLC) these enzymes can be improved to match target process conditions[3] or can enable new chemistries unknown in nature.[4]

Despite the great promise directed evolution has for biocatalyst development, the slow pace of directed evolution campaigns limits the impact biocatalysis can have on chemical synthesis.[5] Significant efforts have been made to reduce this time burden. For example, the increasing implementation and availability of tools to aid directed evolution efforts by providing bioinformatic analysis[6] and structural prediction[7] has accelerated protein engineering efforts by expanding the set of enzymes available for initial screening. Automation methods including liquid handling robots can reduce the amount of researcher intervention required for screening libraries, but such systems are costly to acquire and difficult to maintain. In many cases, the originally identified enzyme exhibits a small substrate scope and techniques such as substrate walking must be implemented to broaden activity towards a desired substrate.[8] While effective, substrate walking introduces additional rounds of evolution and may prolong already lengthy evolution efforts.

Figure 4. 1: Biocatalyst driven PACE



Envisioned method for continuous evolution and selection of biocatalysts (a) Overview of small-molecule-dependent PACE for the evolution of selective biocatalysts. (b) Overview of small-molecule-dependent PACS for the selection of selective biocatalysts. SOURCE: ACS Central Science (2021) 7(9) 1581-1590

The most direct way to shorten a directed evolution campaign is to decrease the time burden required to evaluate each variant in a library. UHPLC and automation is limited to ~10^4 variants per round of mutagenesis with a sample time of about 1-2 minutes per sample, which vastly under samples

the potential variants generated from error-prone mutagenesis of even very small proteins of interest. In comparison, *in vivo* continuous evolution techniques such as PACE (phage-assisted continuous evolution) have the potential to vastly accelerate protein engineering efforts by sampling library sizes that surpass 10^7.[9] Beyond sampling larger segments of the available sequence space, continuous evolution eliminates the *in vitro* steps of library diversification, expression, and screening. By utilizing *E. coli* as 'microreactors', each step of the directed evolution campaign occurs within the *E. coli* cell using native cellular machinery. PACE had previously been used to evolve many protein molecules that directly regulate gene expression, such as RNA polymerases[10], DNA-binding proteins,[11] and invertases.[9]

PACE functions by connecting the activity of a protein of interest (POI) to the production of an essential phage gene, gIII. By encoding the POI on the phage genome, one can ensure that only phage harboring active protein variants can successfully propagate through the system. This selection for protein activity can be transformed into a continuous evolution strategy by the incorporation of a mutagenic plasmid on the *E. coli* host. This plasmid expresses proteins that increase the error rate of DNA replication within the *E. coli* cell to effectively drive POI mutation. By exploiting the rapid lifetime of phage relative to *E. coli*, a continuous flow of fresh bacterial hosts can be provided at a rate that localizes mutations to the phage genome. A key feature of this system is that the connection between biosensor and POI activity is not simply binary, which is to say that increasing activity of the mutated POI results in an increased level of gIII expression and consequently more infectious phage. This sort of positive selection ultimately leads to the enrichment of a library with active variants. More recent efforts have demonstrated that careful design of gene circuits allows for the continuous evolution of proteins which do not natively function on DNA or elements regulating gene expression,[12] expanding the set of proteins amenable to *in vivo* evolution with PACE. Due to the way PACE enriches in active proteins based on function, it often results in multiple variants with distinct gene sequences, showing that many different mutations can result in improved performance.

We envisioned that with an appropriate selection of a biosensor-ligand pair, a continuous evolution strategy such as PACE could be applied to the evolution of a biocatalyst (**Figure 4.1A**). Prior to the work described herein, only one such biocatalytic PACE campaign had been conducted. This work evolved a bacterially expressible methanol dehydrogenase (MDH) utilizing the irreversible reaction between the transcription factor FmrR and acetaldehyde to drive gIII expression.[13] Due to the low phage propagation rates from the wild-type protein, a non-continuous system described as PANCE (phage assisted non-continuous evolution) was initially used to evolve variants of MDH that could survive the more stringent selection conditions observed during PACE (For an example of PANCE, see **Figure 4.1B**). Variants evolved from this non-continuous evolution platform were then successfully able to propagate in a normal PACE system due to their enhanced performance over the wild type. Even with improved variants that could successfully propagate, the presence of 'cheaters' who developed non-productive ways of escaping selection required more complex controls to regulate gIII expression. We envisioned that in addition to being a novel method of biocatalyst evolution, this project could prove valuable for understanding PACE as a platform for future research. This work has resulted in a publication.[14]

Authorship

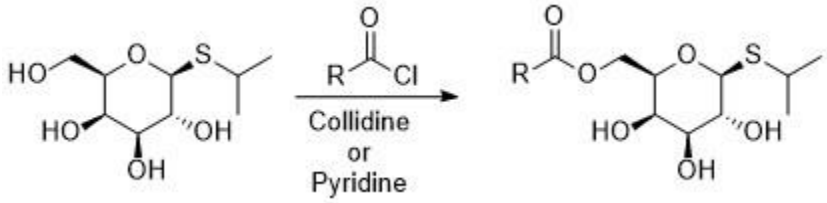
My responsibilities on this project included IPTG biosensor cloning and evaluation, enzyme identification, ligand synthesis for all biosensors except the estradiol biosensor, MISER method development, library screening, and enzyme kinetics. The estradiol biosensor was designed and evaluated by Dr. Ketaki Belsare, and Dr. Krysten Jones designed the ABA biosensor and the PACE and PACS experiments used for continuous evolution. Unless otherwise mentioned, all biosensor optimization and cloning was performed by Dr. Krysten Jones or Dr. Ketatki Belsare.

4.2 Results and discussion

4.2.1 Initial biosensor design and ligand synthesis

Inspired by the RNA-templated release of IPTG via a proximity induced Staudinger reaction used to identify complementary RNA,[15] we sought to use the IPTG inducible LacI repressor as the biosensor in our ligand-biosensor pair for initial attempts at PACE. To probe whether IPTG could be used as a masked ligand for the LacI biosensor, IPTG esters of different R-group bulkiness were synthesized to test whether observable background hydrolysis occurred in *E. coli* lysate (**Scheme 4.1**). An earlier reported synthesis detailed straightforward acylation conditions using the bulky pyridyl base collidine, which could be readily purified via silica chromatography. This synthesis could be adapted for other acid chlorides to readily synthesize the four 'masked' IPTG based ligands used to assay for background hydrolysis.

Scheme 4. 1: Synthesis of IPTG esters used as ligands for initial studies



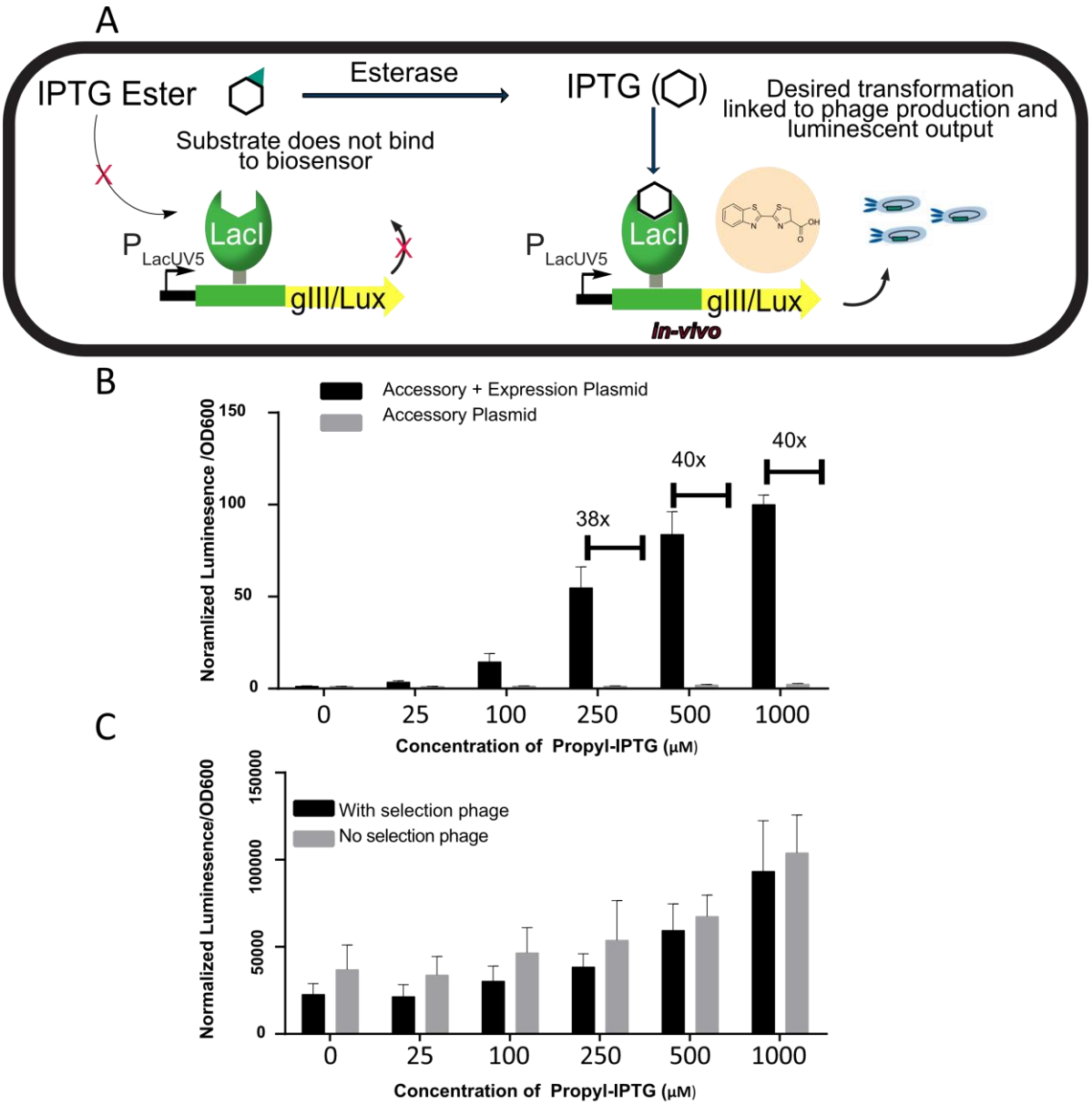
General scheme for synthesis of IPTG esters. For more details, see **Section 4.4.2 – Ligand synthesis**.

To detect background hydrolysis of the synthesized IPTG esters, we developed an assay using the P_{UVLac5} promoter to control expression of the *luxAB* gene. The assay was based on a luminescent readout using the expression of the gIII-luxAB gene cassette. This cassette is used as a readout in PACE screens as the luciferase protein produced from the luxAB portion of the gene cassette provides an easily measured luminescent signal. To further simplify the assay, it is conducted in cells that natively produce myristyl aldehyde, the native substrate of luciferase, from a chromosomally encoded luxCED cassette to obviate the need for addition of octanal or other external substrates. While GFP based assays are more common due to the absence of the required substrate for the observed fluorescence,

the luminescent reaction catalyzed by the protein products of the *luxAB* cassette provides a quicker readout due to the nature of the enzymatic reaction compared to the timeline required for protein expression and fluorophore maturation. This assay can be used both with the phage ultimately utilized in PACE or with co-transformed plasmids expressed in *E. coli* which are typically used to probe initial gene circuit designs. Initial assays conducted in aldehyde producing cells showed that no significant turn-on was observed with any of the esterified IPTG molecules after incubation in *E. coli* lysate in the absence of any co-expressed esterase (see additional information, **Figure 4.10**).

Encouraged by the lack of observed background hydrolysis, we next sought to identify a potential esterase to act as our biocatalyst. Porcine-liver esterase was initially investigated due to its broad substrate scope and application in biotechnology, but poor expression in *E. coli* prohibited its use in PACE. The failure to adequately express this esterase led us to consider expression as a primary consideration for future enzyme choice. This ultimately led to the investigation of the *B. subtilis* esterase BS2 which has been reported to express well in typical *E. coli* expression vectors.[16] To test whether BS2 might be a suitable enzyme for *in vivo* IPTG ester hydrolysis, *E. coli* were transformed with BS2 harbored within an expression plasmid meant to mimic the expression of genes from phage vectors (For biosensor schematic, see **Figure 4.2A**). When cells co-expressing the BS2 esterase and the *luxAB* gene were provided with either the acetyl or n-propyl IPTG esters we observed 32 and 79 fold turn on over the signal for background (see the experimental section, **Figure 4.11**). Optimization of substrate concentration showed that using 500 μM of the n-propyl IPTG ester resulted in 40-fold luminescent signal over background.

Figure 4. 2: IPTG driven gIII production



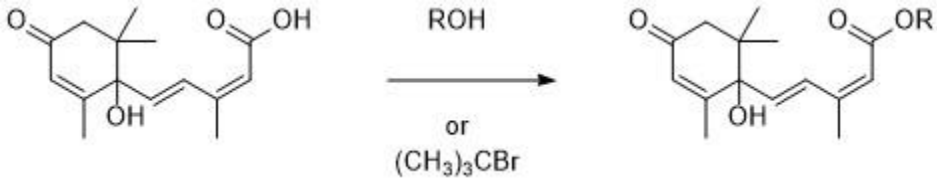
A) Schematic for the design of the IPTG biosensor. B) Screening of the n-propyl IPTG ester **1b** for in-vivo hydrolysis from BS2 produced by E. coli C) When BS2 is expressed from selection phage, no significant turn on over background is observed.

Despite this optimization, after using Gibson assembly[17] to clone the BS2 esterase into the selection phage, no luminescent signal was observed in plate-reader based hydrolysis assays (**Figure 4.2C**) even after further optimization of the accessory plasmid (Data of optimized plasmids in the experimental, for results see **Figure 4.12** and **4.13**). Additionally, no plaques were observed from plaque

assays used to measure phage propagation levels. A further analysis of phage display literature made clear the importance of low basal gene expression from the biosensor implemented in PACE. When pIII, the protein product of gIII, is produced in the absence of other phage proteins this free pIII embeds itself in the *E. coli* membrane, occupying sites for phage to bind and precluding the phage from propagating.[18] LacUV5 and other IPTG promoted gene expression systems are known to be "leaky"[19] and have relatively high levels of basal gene expression due to a weak binding interaction between the LacI repressor protein and its cognate DNA operator sequence.[20] The tight gene regulation required in PACE likely prohibits these 'leaky' repressors from being implemented without substantial efforts, and this avenue for PACE driven biocatalyst design was concluded.

4.2.2 Abscisic acid based biosensor designs prove amenable for PACE and PACS

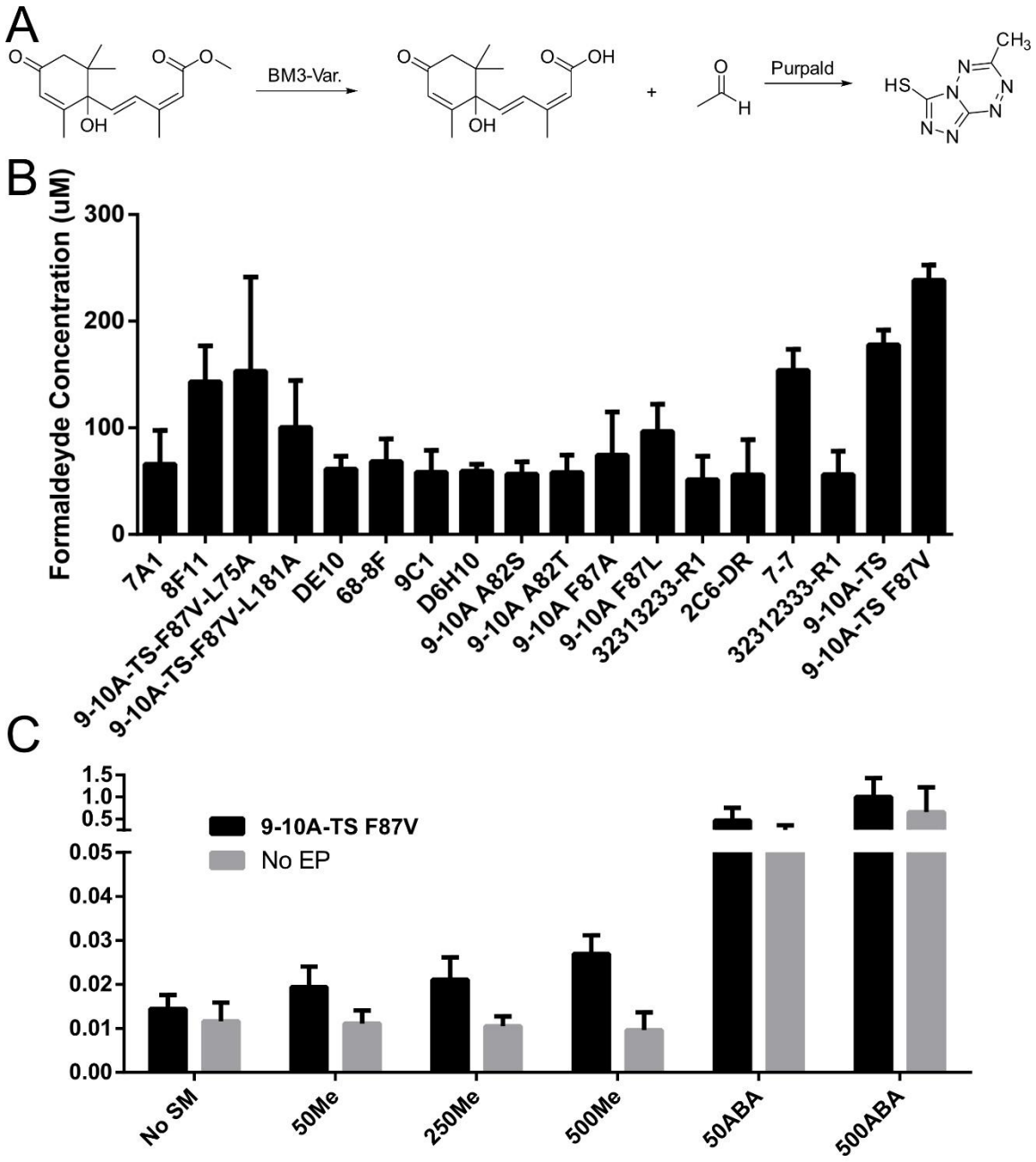
Scheme 4. 2: Synthesis of ABA esters



Given that the LacI biosensor was incompatible with PACE, we next focused on a split sensor that had been previously used. After evaluating the literature, we found the ABI-PYL based abscisic acid binding system appealing.[21] This biosensor is based on a split T7 RNA polymerase, each half of which is genetically fused to either the PYL or ABI proteins. These proteins natively engage in a protein-protein interaction only in the presence of abscisic acid (ABA). In the engineered system, when this ABA dependent PPI occurs, it brings together the two halves of the T7 RNAP and recovers function, allowing gene expression from the T7 promoter. In addition to being very specific for the presence of ABA, this also serves as an amplification of signal detection, as the functionally active T7 RNAP rapidly generates multiple mRNA copies.

After Dr. Krysten Jones optimized the biosensor for maximal turn-on in the presence of ABA, I next investigated the masking of ABA to confirm that the biosensor did not have activity on potential enzyme substrates that could release ABA. The synthesis of the methyl, ethyl, and benzyl esters of ABA were accomplished using simple methods adapted from similar ester syntheses (**Scheme 4.2**). When the DMAP catalyzed synthesis of the isopropyl ester was attempted however, NMR spectroscopy revealed the formation of a new impurity with similar chemical shifts. This impurity could not be adequately separated from the desired product and as such a new synthesis was required. A Mitsunobu reaction[22] using triphenylphosphine and diethyl azodicarboxylate as the acylating agents gave the isopropyl product in low yield but with no significant impurity. Optimization of this with the more reactive tri-*n*-butylphosphine and tetramethylazodicarboxamide[23] increased yield to from 31% to 76% and allowed for the ready preparation of this compound. As expected for the bulky *t*-butyl ester, none of these methods were successful. Instead, a reference for the effective *t*-butylation of esters proved useful and substituting bicarbonate for cesium carbonate gave the *t*-butyl ester in high yield. In a luminescent assay with coexpression of the biosensor components and a 'dummy' phage harboring the HRV protease none of the masked ligands had significant turn-on over background, suggesting suitability for PACE.

Figure 4. 3: P450 driven release of ABA from masked ligand

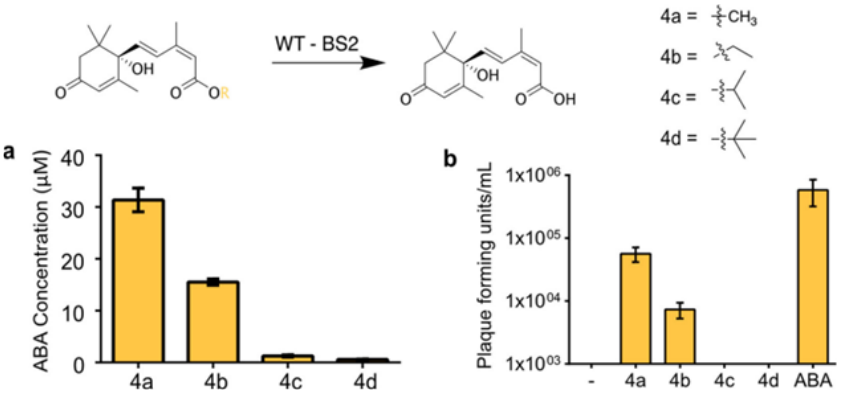


P450 catalyzed gene expression. A) Scheme for the purpald assay and B) The resulting formation of purpald from P450 compilation screen. Reactions were conducted in duplicate and purpald formation was determined from a calibration curve made fresh the day of the analysis. C) Results of luminescent assay from P450 co-expressed with gIII-luxAB from accessory plasmid in E. coli based assasy. EP = Expression Plasmid

For initial studies, we looked at two methods to deprotect the 'masked' ABA ester ligands. The first is a similar esterase catalyzed hydrolysis approach originally undertaken for the IPTG project

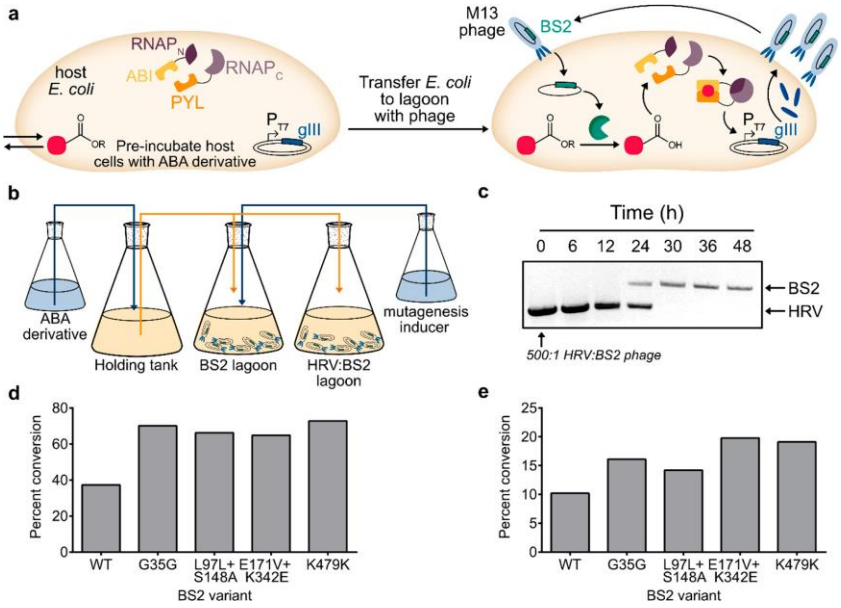
outlined above. We also investigated the well-studied cytochrome $P450_{BM3}$. This enzyme is a heme-dependent monooxygenase that contains a fused reductase domain capable of using NADPH to reduce the iron center of the heme, circumventing the need for a secondary protein for catalytic activity. BM3-catalyzed hydroxylation alpha to the ester oxygen atom leads to carboxylic acid deprotection with concomitant aldehyde formation. Because methyl ester substrate **4a** would lead to formation of formaldehyde, a compilation plate of BM3 variants was screened using a colorimetric assay for the release of formaldehyde[24] from the corresponding ABA ester (**Figure 4.3A**). Hits from this screen were then further analyzed via HPLC to confirm ester cleavage. We also observed roughly two-fold turn on over background by expressing the BM3 variants in an *E. coli* expression vector as previously described for the BS2 esterase (**Figure 4.3B**). Unfortunately, repeated attempts to clone this gene onto the phage genome failed. Truncation of the BM3 gene within the phage genome led us to believe that the large size of the BM3 gene (3.15 kb) is incompatible with the relatively small phage genome (~4 kb without an enzyme-gIII cassette). While this may prove viable for other experiments utilizing the ABA biosensor, the inability to render phage harboring the BM3 gene led us to prioritize the esterase catalyzed hydrolysis.

Figure 4. 4: Esterase-catalyzed hydrolysis of abscisic acid (ABA) can support phage replication



(a) BS2 activity on ABA esters. ABA esters (750 μM) were incubated with 5 μM BS2 over 60 min, and percent conversion to ABA was determined. (b) BS2 phage replication with ABA derivatives. Phage carrying BS2 were incubated with E. coli expressing ABA biosensor in the absence or presence of ABA esters or ABA for 6 h. Phage cultures were then collected and analyzed for phage replication. Error bars are the standard error for n = 3 replicates. SOURCE: ACS Central Science (2021) 7(9) 1581-1590

Figure 4. 5: PACE set up and results of enrichment and continuous evolution



Evolution of esterase variants with activity on ABA esters. (a) Schematic of PACE for biocatalyst evolution. ABA derivative is preincubated with E. coli expressing ABA biosensor 3 h prior to transfer to a lagoon containing phage carrying BS2. Only BS2 phage that can hydrolyze the derivative to ABA will be able to replicate on the host cells and produce gIII to generate infectious progeny. (b) Biocatalyst PACE setup with a holding tank for host cell preincubation with ABA derivatives. (c) Phage competition in biocatalyst PACE. ABA ester 4a was preincubated with host cells, and phage carrying BS2 or HRV were mixed (1:500) as shown in part b. Phage samples were collected 0–48 h after addition to the lagoon and then analyzed by PCR. Bioconversions of BS2 and PACE BS2 variants on (d) ABA ester 4a or © ABA ester 4b. SOURCE: ACS Central Science (2021) 7(9) 1581-1590

Having already established the expression and activity of BS2 in the *E. coli* co-expression system, we tested this esterase *in vitro* against the synthesized ABA esters. We found that BS2 hydrolyzed the methyl ester efficiently with about 5 % conversion by UHPLC (**Figure 4.4A**). As expected, increased steric bulk correlated with lower activity, and while BS2 hydrolyzed the ethyl ester with lower activity only trace activity could be observed for the isopropyl and *t*-butyl esters. Unlike with the earlier IPTG driven biosensor, this observed activity also translated to the phage propagation assay with the BS2 containing phage able to propagate when provided the masked ABA ligands in plaque assays (**Figure 4.4B**). Notably, the ability of phage to form plaques correlated to the observed conversions from *in vitro* bioconversions with only a small amount of plaque formation observed using the ethyl ester. An additional functional assay showed the enrichment of BS2-containing phage relative to HRV-protease harboring phage variants, suggesting that active esterases should enrich over inactive variants in a selection (**Figure 4.5C**). Even with the extensive optimization on the biosensor system, the phage titers observed with the BS2 catalyzed release of ABA from methyl-ABA were low compared to other PACE systems that utilize systems functioning directly on DNA regulating elements. Despite the low titers, the successful propagation of phage in this system led us to believe biocatalyst-driven PACE could be accomplished with this system (for biosensor scheme, see **Figure 4.5A**).

As ABA has previously been noted to require incubation with *E. coli* to allow for diffusion across the membrane, an additional "holding tank" was incorporated into the flow path which we suspected would allow for cells to become saturated with esterified ABA before exposing the *E. coli* to selection phage (**Figure 4.5B**). For the first 48 hours of the PACE experiment, methyl ester **4a** and ABA were both administered to the both lagoon to ensure adequate phage replication and provide a period of neutral drift to accumulate mutations before applying stronger selection pressure. Over this initial 48-hour period the ratio of **4a**/ABA was increased until only **4a** was pumped into the lagoon and PACE was continued for an additional 48 hours, at which point the esterase variants were sub-cloned into an

arabinose inducible vector for expression in *E. coli*. Analysis of the genotype of the resulting BS2 variants from the PACE experiment showed many of the mutations were silent, and only a few coding mutations were observed. When these variants were evaluated against the methyl ester in lysate improvements up to 2-fold were observed over WT BS2 (**Figure 4.5D and 4.5E**), but the purified enzymes were largely identical to the wild-type. This led us to suspect that the selection conditions applied during even the lowest stringency tests led to phage populations too low to meaningfully sample the sequence space available, so most of the improvements we observed resulted from improved expression and not activity. Additionally, an attempt to evolve variants by selection on ethyl substrate **4b** did not yield any variants with further mutations, indicating that the current system for PACE was insufficient for further evolution.

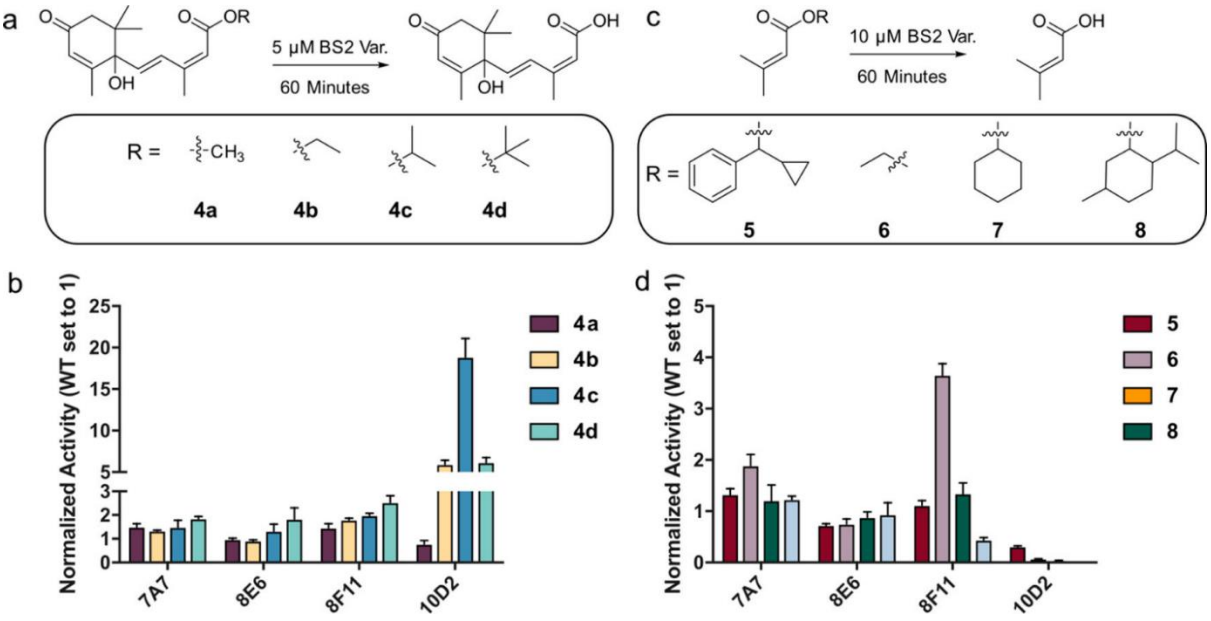
Considering the low rate of mutation in PACE, we instead sought to utilize the phage for an analogous, non-continuous selection. Rather than rely on the mutagenic plasmid for library generation, this approach would instead use phage-assisted continuous selection of an error-prone library of BS2 to enrich in active variants. This library was constructed with a high error-rate(average of 5+ residue mutations/gene, see **Figure 4.14**) that would result in mostly inactive variants for a typical screen. Initial attempts at cloning this library using Gibson assembly resulted in mostly truncated or fragmented BS2 gene constructs within the phage. We believe this is likely because of the impurities present in the primers used for Gibson assembly, which may result in smaller or non-expressing gene products which would have less of a burden on phage replication. Restriction cloning proved effective for generation of the BS2 library harbored within the selection phage. The PACS experiment was identical to that used in the previously described PACE set up, except the mutagenesis plasmid typically used to induce mutation in the target gene was removed (**Figure 4.1B**). After 24 hours circulating in the lagoon, the *in vitro* constructed BS2 phage library was able to replicate using the ethyl ABA ligand whereas the control phage consisting of WT-BS2 showed phage extinction. This confirmation of the improved catalytic

activity of the phage circulating in the library signaled a successful selection. Just as PACE can evolve a target gene for improved function with multiple unique mutations, we envisioned that PACS would likely result in a pool of variants with improved activity. To verify this, the phage samples obtained post-selection were subsequently sub-cloned into the pET28 vector for expression in BL21(λDE3).

4.2.3 Evaluation of esterase variants produced from PACS.

In order to sample the variants generated through the PACS experiment, we sought to implement a mass-spec based high-throughput screening approach. The multiple injections in a single experimental run (MISER) method uses only a guard-column to achieve separation between reaction components (e.g. buffer and substrate/product) to facilitate rapid analysis of large libraries. With this method, separation between the buffer, free ABA, and the ethyl ester could be achieved with a total sample processing time of 42 seconds, which represents a 4-fold decrease in processing time per sample compared to the conventional UHPLC method that we typically used to analyze individual reactions. After verifying a good correlation between MISER activity and UHPLC activity for a subset of the library (**Figure 4.9** in the experimental section), this method was used to process 900 samples plus 60 control reactions qualitatively. The top 90 hits were then pooled into a single 96-well microtiter plate and re-screened in lysate against the ethyl ester. This analysis revealed 10 variants with at least 2-fold improved activity over WT for the hydrolysis of the ethyl ABA ester, including two variants later identified to have identical protein sequences. This variant 10D5 contained six new coding mutations representing a large mutational burden unlikely to emerge from a typical round of directed evolution (See **Table 4.1** for the list of mutations).

Figure 4. 6: Evaluation of substrate scope for selected hits from a PACS screen



(a) Reaction scheme for ABA esters evaluated. 750 µM substrate was incubated with 5 µM BS2 variant for 1 h. (b) Conversion of each reaction normalized to the WT reaction. Each bar represents the average of three reactions. Yield of ABA obtained from a calibration curve using 2-acetamidophenol as an internal standard (calibration curve Figure S17). (c) Reaction scheme for the hydrolysis of "truncated" ABA fragment, 3,3-dimethyl acrylate (DMA) esters. 5 mM substrate was incubated with 10 µM BS2 variant for 1 h. (d) Conversion of each reaction normalized to the WT reaction. Each bar represents the average of three reactions. Yield of 3,3-DMA obtained from a calibration curve using 5-bromoindole as the internal standard (calibration curve Figure S18). SOURCE: ACS Central Science (2021) 7(9) 1581-1590

After identifying a panel of significantly improved enzymes (for enzyme sequences see **Table 4.1**), we next evaluated the substrate scope of the variants obtained via PACS (**Figure 4.6A**). For all variants except 10D5, the observed activity decreased with increasing steric bulk of the ester fragment (Me > Et >> *i*-Pr >> *t*-Bu) (**Figure 4.6B**). In contrast, 10D5 exhibited the highest activity for the ethyl ester followed by the isopropyl compound and exhibited a surprising decrease in activity compared to the wild-type for the methyl substrate. Steady state kinetic analysis revealed that the likely cause of this inverted preference was the dramatic increase in K_M from 111 µM for the wild-type to approximately 5.615 mM in the evolved variant (for kinetic characterization, see the **Additional Data Section 4.5** and **Figures 4.16-4.19**). The K_M for the ethyl substrate and 10D5 also increased to 1265 µM compared to 143

μM in the wild-type, but a compensatory 50-fold increase in Kcat results in a more efficient reaction despite the less effective binding.

The top hits from this library were also evaluated against a panel of substrates protected with a 3,3-dimethyl acrylate group (**Figure 4.6C**), which we envisioned would act as a synthetically tractable protecting group that could mimic the ABA fragment to facilitate binding to evolved variants. A series of DMA esters were synthesized and screened against the BS2 panel. Interestingly, 10D5 was found to have minimal activity on any of the other DMA substrates evaluated. Instead, the two mutants found to be broadest in substrate scope were 7A7 and 8F11, which contain 4 (P90T, I130V, N269S, L339P) and 3 (P76R, T220A, D320G) unique, new coding mutations respectively. In the latter case of 8F11, all mutations are predicted to be on the surface of the BS2 esterase far from the active site. This finding demonstrates how non-targeted *in vitro* libraries can be functionally enriched through PACS to generate a population of variants that would be difficult to arrive at by rational mutagenesis and with mutational loads higher than is typically permitted in traditional directed evolution. Although the purified variants examined generally showed an increase in activity towards the substrate panel compared to wild-type BS2, none of the evaluated variants had significantly improved activity towards the α-cyclopropylbenzyl substrate. Encouraged by the range of activities from our narrow panel of enzymes, we broadened the panel of enzymes by rescreening the compilation plate in lysate. One variant from this screen, 1E12, contained 3 additional unique mutations and was found to be 1.8-fold improved over wild-type for hydrolysis of this compound after IMAC purification, outperforming any enzymes from the previously examined set (see **Figure 4.21** in the additional data section).

4.3 Conclusions

The work described within this chapter helped to establish methods and systems that can be used to link gIII expression to the modification of a small molecule by a biocatalyst. In addition to expanding the set of biosensor designs that have been successfully linked to gIII production, we

demonstrated the potential of biocatalyst dependent PACE by evolving an esterase using a previously designed ABA biosensor. This biosensor was also shown to be amenable for the evaluation of P450 enzymes in the same context, but these could not be successfully incorporated onto the phage genome. After modifying the PACE system to evaluate a library constructed *in vitro* we eventually identified an esterase with 48-fold improved Kcat for the hydrolysis of probe substrate **4b** and a change in preference from the smaller methyl ester **4a** to the bulkier isopropyl (**4c**) and *t*-butyl (**4d**) ABA esters. Many of the mutations accumulated throughout these evolutions occurred in regions outside of the active site and may have been difficult to predict or rationalize. Despite these successes, we also found some limitations of biocatalyst driven PACE that will require further understanding of the system before it can be widely adopted for biocatalyst development.

4.4 Experimental

4.4.1 Materials, Instruments and methods

Greiner Bio-One polypropylene 96-well plates (product number 651201) or Fisherbrand 96-well deepwell polypropylene plates (product number 12-566-121) were purchased from Fisher. Corning-costar 96-well black wall, clear bottom plates (product number 3631) were purchased from Fisher. Antibiotics were purchased from Gold Biotechnology and were prepared as 1000x stock solutions. Sodium acrylate was purchased from Combi-Blocks (product number QC-1489-005). IPTG was purchased from Gold Biotechnology (product number I2481). Estradiol was purchased from Combi-Blocks (product number QA-4119). Abscisic acid was purchased from Biosynth International (product number A-0120). Stock solutions (50 mM) of ligand were prepared at 50 mM in either DMSO (estradiol and abscisic acid) or water (sodium acrylate and IPTG) and stored at -20 °C until use. Acrylate ester substrates and ethyl 3,3-dimethylacrylate were purchased from Sigma-Aldrich and used without further purification. Estradiol ethers were synthesized as previously reported (PMID: 11141094).

Phusion polymerase was expressed, purified, and titrated prior to use[1]. Taq DNA polymerase and Q5 High-Fidelity DNA polymerase was purchased from New England Biolabs. LBD and SRC RID3 genes were synthesized by Twist Bioscience. The genes for ABI and PYL were gifts from Professor Fu-Sen Liang. The genes for AcuR was a gift from George Church (pJKR-L-acuR, Addgene plasmid #62568; http://n2t.net/addgene:62568; RRID:Addgene_62568). The above genes were integrated into a plasmid containing *gIII* translationally coupled to the bacterial lux operon.

Luria broth (LB) media and Super Optimal broth with Catabolite repression (SOC) media were purchased from Research Products International and Davis rich media for PACE or PACS was prepared as previously described[2]. *E. coli* 10β were purchased from New England Biolabs and *E. coli* S1030 cells[2] were courtesy of the Liu lab, Harvard University. Restriction enzymes, Antarctic Phosphatase, and T4 DNA ligase were purchased from New England Biolabs. DNA clean & concentrator kits were purchased from Zymo Research (product number D4013).

Luciferase and optical density measurements were performed using a Synergy Neo2 Hybrid Multi-Mode Reader (BioTek). For UHPLC-MS analysis, 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 was used for both low-throughput analysis and the qualitative MISER screen. For UHPLC analysis, a similar Agilent system was used, except the autosampler used was single-needle configuration. For both UHPLC-MS and UHPLC, low-throughput analysis was performed on an Eclipse Plus C18 2.1x5mm guard column with a 1.8 µm pore size (part number 821725-901) connected to a ZORBAX rapid resolution C18 column (part number 959757-902). MISER analysis was done using only the guard column.

4.4.2 Cloning and molecular biology

Cloning and sequencing

All plasmids were constructed by Gibson assembly from PCR products generated using Phusion Polymerase or Q5 High-Fidelity DNA Polymerase (NEB). PCR reactions and thermocycling conditions were performed according to NEB's recommended protocols. Annealing temperatures were determined with the NEB T_m Calculator. Phage were cloned by Gibson assembly of BS2 into a previously optimized[3] SP phage backbone and transformed into previously developed 1059 cells[2], which supply *gIII* in an activity-independent manner. After overnight growth of phage in media, the supernatant was isolated and plaque assays were performed on 1059 cells. Single plaques were grown overnight and sequenced to identify phage samples with the correct insert. All plasmids and phage were sequenced by the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility.

Biosensor screening in *E. coli* for small molecule detection.

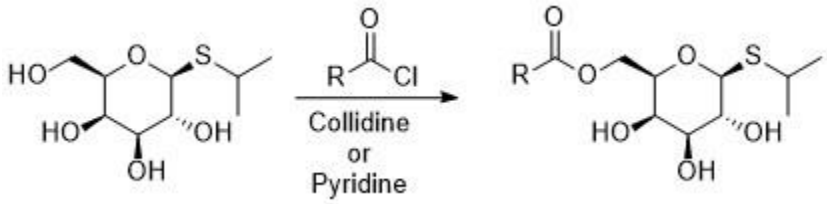
10β cells were transformed by heat shock with either AcuR, IPTG, estradiol, or ABA biosensors. Single colonies were grown to saturation overnight at 37 °C and then each well of a 96-well deep well plate containing 0.185 mL of LB with antibiotics and 500 µM control small molecule substrate or ester substrate was inoculated with 20 µL of the overnight culture. After growth with shaking for 6 h at 37 °C, 150 µL of each culture was transferred to a 96-well black wall, clear bottom plate (Costar), and luminescence and OD_{600} were measured on a Synergy Neo2 Hybrid Multi-Mode Reader (BioTek). The data were analyzed by dividing the luminescence values by the background-corrected OD_{600} value. All values were then normalized to the emission of cells expressing each biosensor without small molecule, which was assigned an arbitrary value of 1 to allow for values from each luminescence plot to be compared to each other. All experiments were performed in four technical replicates.

Phage growth assays.

S1030 cells were transformed by heat shock (42 °C, 45 sec) with either AcuR, IPTG, estradiol, or ABA biosensors. Single colonies were grown to saturation overnight at 37 °C and then each well of a 24-well deep well plate containing 0.375 mL of LB with antibiotics and 1.5 mM small molecule ester substrates or 500 µM control substrates was inoculated with 125 µL of the overnight culture. After growth with shaking for 2 h at 37 °C, 10 µL of BS2 phage (1 x 10^5 PFU/µL) was added to each well. Cultures were then grown for an additional 5-6 h at 37 °C with shaking and supernatant was then isolated for plaque assays on 1059 cells to calculate phage titers.

4.4.3 Synthesis

Scheme 4. 3: General procedure for the synthesis of IPTG esters 1a-1c.



To an oven-dried round bottomed flask under nitrogen atmosphere was added IPTG (500 mg, 2.09 mmol). The flask was evacuated and backfilled with nitrogen 3x before 3 mL 2,4,6-collidine was added to the flask. This solution was cooled to –35 °C in a 40:60 methanol:water bath containing dry ice. The corresponding acid chloride was added dropwise over 20 minutes. Upon addition of the acid chloride, the solution solidified into a white gummy solid. After 30 minutes, product was confirmed by TLC and the solution was removed from the dry ice bath and allowed to warm to room temperature. The resulting gummy solid was dissolved in DCM and purified by silica chromatography using a gradient of 1 column volume each of 4:1 DCM:Acetone, 3:1 DCM:Acetone, then 2:1 DCM:Acetone. Product was identified by staining TLC plates in $KmnO_4$, then collected and concentrated by rotary evaporation. Residual collidine was removed under high-vacuum.

6-O-Acetyl-IPTG (**1a**): Obtained 491 mg, 1.75 mmol. 84% Yield

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.43 – 4.35 (m, 2H), 4.30 (dd, *J* = 11.6, 6.9 Hz, 1H), 3.97 (q, *J* = 1.3 Hz, 1H), 3.74 (ddd, *J* = 7.0, 5.9, 1.2 Hz, 1H), 3.69 – 3.62 (m, 2H), 3.23 (hept, *J* = 6.8 Hz, 1H), 2.10 (s, 3H), 1.37 (dd, *J* = 9.9, 6.8 Hz, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.05, 86.01, 75.90, 74.46, 70.52, 68.48, 62.93, 36.05, 24.23, 23.95, 20.83. HRMS: calc. for $[C_{12}H_{22}O_6Sna]^+$ ([M+Na]$^+$: 303.0873, found 303.0873.

6-O-propionyl-IPTG (**1b**): Obtained 369 mg, 1.254 mmol. 60% Yield

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.42 – 4.32 (m, 2H), 4.29 – 4.20 (m, 1H), 3.92 (s, 1H), 3.69 (t, *J* = 6.5 Hz, 1H), 3.65 – 3.57 (m, 2H), 3.18 (hept, *J* = 6.8 Hz, 1H), 2.33 (q, *J* = 7.6 Hz, 2H), 1.32 (dd, *J* = 8.2, 6.8 Hz, 6H), 1.13 (t, *J* = 7.8 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 174.50, 85.95, 75.99, 74.49, 70.51, 68.54, 62.81, 36.00, 27.47, 24.22, 23.90, 9.03. HRMS: calc. for $[C_{12}H_{22}O_6Sna]^+$ ([M+Na]$^+$: 317.1029, found 317.1030.

6-O-isobutyl-IPTG (**1c**): Obtained 290 mg, 0.94 mmol. 45 % Yield.

The procedure for **1c** is the same as **1a** and **1b**, but 0.1 equiv. 4-dimethylaminopyridine was added to the reaction before collidine.

^{1}H NMR (500 MHz, Chloroform-*d*) δ 4.45 – 4.36 (m, 2H), 4.28 (dd, *J* = 11.5, 6.9 Hz, 1H), 3.95 (q, *J* = 1.2 Hz, 1H), 3.73 (ddd, *J* = 7.0, 5.9, 1.2 Hz, 1H), 3.70 – 3.61 (m, 2H), 3.23 (hept, *J* = 6.7 Hz, 1H), 2.59 (hept, *J* = 7.0 Hz, 1H), 1.36 (dd, *J* = 11.1, 6.8 Hz, 6H), 1.19 (d, *J* = 7.0 Hz, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 177.20, 85.92, 76.03, 74.49, 70.55, 68.52, 62.75, 35.96, 33.94, 24.24, 23.88, 18.93, 18.89. HRMS: calc. for $[C_{13}H_{24}O_6Sna]^+$ ([M+Na]$^+$: 331.1186, found 331.1187

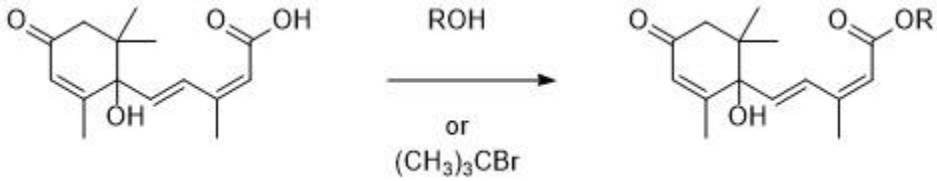
6-O-trimethylacetyl-IPTG (**1d**): To an oven-dried round bottomed flask under nitrogen atmosphere was added IPTG (500 mg, 2.09 mmol), and 0.1 equiv. 4-dimethylaminopyridine (25 mg, 0.21 mmol). This was evacuated under vacuum and purged with nitrogen three times, after which 3 mL pyridine was added to the flask. The solution was cooled to 0 °C with an ice bath, at which point trimethylacetyl chloride (255 µL, 2.09 mmol) was added dropwise over 20 minutes. Upon addition of trimethylacetyl chloride the solution became a white gummy solid. After addition of trimethylacetyl chloride, the solution was allowed to stir for a half hour at which point the ice bath was removed and the gummy solid dissolved in DCM and purified via silica chromatography using a gradient of 1 column volume each of 4:1 DCM:Acetone, 3:1

DCM:Acetone, then 2:1 DCM:Acetone. Product was identified by staining TLC plates in KmnO4, then collected and concentrated under reduced pressure. Residual collidine was removed under high-vacuum until white crystals crashed out of solution.

Obtained 225 mg (1.46 mmol) 70 % yield.

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.36 – 4.27 (m, 2H), 4.17 (dd, *J* = 11.6, 7.0 Hz, 1H), 3.85 (s, 1H), 3.64 (t, *J* = 6.5 Hz, 1H), 3.57 (d, *J* = 8.3 Hz, 2H), 3.14 (p, *J* = 6.8 Hz, 1H), 1.27 (dd, *J* = 12.7, 6.7 Hz, 6H), 1.13 (s, 9H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 178.68, 85.86, 76.09, 74.51, 70.56, 68.58, 62.92, 38.80, 35.92, 27.14, 24.26, 23.83. HRMS: calc. for $[C_{14}H_{26}O_6Sna]^+$ ($[M+Na]^+$: 345.1342, found 345.1344.

Scheme 4. 4: Synthesis of abscisic acid esters 4a-4d.



4a – methyl abscisic acid: To a stirring solution of ABA (250 mg, 0.94 mmol) in 5 mL of a 9:1 mixture of toluene to methanol was added a 2M solution of TMS-diazomethane in diethyl ether. The TMS-diazomethane solution was added dropwise until a yellow color persisted, at which point the solution was neutralized with glacial acetic acid until colorless. The solution was then concentrated under rotovap and washed with toluene until white crystals were obtained, and the compound was used without further purification. 261 mg obtained, quantitative yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.81 (d, *J* = 16.1 Hz, 1H), 6.08 (d, *J* = 16.1 Hz, 1H), 5.88 (s, 1H), 5.69 (s, 1H), 3.64 (s, 3H), 2.47 – 2.17 (m, 2H), 1.94 (d, *J* = 1.3 Hz, 3H), 1.86 (d, *J* = 1.4 Hz, 3H), 1.05 (s, 3H), 0.95 (s, 3H). ^{13}C NMR (126 MHz, Methylene Chloride-d_2) δ 196.54, 165.54, 161.44, 148.88, 135.72, 127.24, 126.21, 117.15, 78.94, 50.24, 48.90, 40.65, 23.26, 22.06, 20.12, 17.90. HRMS: calc. for $[C_{18}H_{26}O_4Na]^+$ ($[M+Na]^+$: 301.1410, found 301.1412.

4b – ethyl abscisic acid: To a stirring solution of ABA (250 mg, 0.94 mmol) in 5 mL DCM was added 1 equiv. of DBU (0.94 mmol, 140 µL, followed by 1.2 equiv. Of bromoethane (1.13 mmol, 90 µL). This solution was allowed to stir overnight at room temperature. The next day, the solution was transferred to a separatory funnel and washed with saturated sodium bicarbonate, then brine and dried over magnesium sulfate.

After concentrating the solution by rotary evaporation the resulting oil was taken up in DCM and the compound was purified via silica chromatography using a 6:1 hexanes:ethyl acetate solvent mixture. After removing solvent under reduced pressure, white crystals were obtained in 94% yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 (d, *J* = 16.1 Hz, 1H), 6.08 (d, *J* = 16.0 Hz, 1H), 5.87 (s, 1H), 5.68 (s, 1H), 4.10 (q, *J* = 7.1 Hz, 2H), 2.41 (d, *J* = 17.1 Hz, 1H), 2.28 – 2.17 (m, 1H), 1.94 (s, 3H), 1.85 (s, 3H), 1.22 (t, *J* = 7.1 Hz, 3H), 1.04 (s, 3H), 0.95 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 197.64, 166.02, 162.31, 148.89, 136.05, 128.19, 127.08, 118.89, 79.66, 59.98, 49.78, 41.52, 24.32, 23.05, 21.15, 18.88, 14.30. HRMS: calc. for $[C_{17}H_{24}O_4Na]^+$ ($[M+Na]^+$: 315.1567, found 315.1568

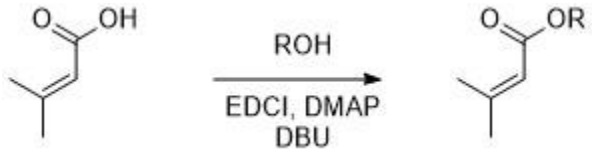
4c – isopropyl abscisic acid: ABA (250 mg, 0.94 mmol) was added to an oven-dried round bottomed flask that was evacuated and purged with nitrogen 3x. 1.5 equiv. N,N,N',N'-tetramethylazodicarboxamide (TMAD) (243 mg, 1.41 mmol) was added, followed by 2.5 mL anhydrous benzene. The solution was allowed to cool to 0 °C under nitrogen 200tmosphere in an ice bath and 1.5 equiv. tri-n-butylphosphine (350 µL, 1.41 mmol) was added, followed by 3 equiv. Isopropanol (215 µL, 2.82 mmol). Upon addition of the alcohol, the solution solidified immediately. The ice bath was replaced with an oil bath and the solution heated to 60 °C at which point the mixture gradually went back into solution. The reaction was monitored by TLC until no further product formation was obtained, at which point the solution was concentrated under reduced pressure and the resulting residue taken up in DCM. A work up consisting of a 0.5 M HCl wash, followed by a brine wash was done and the solution was dried over magnesium sulfate. This was filtered away and the solution was dried under reduced pressure. After removing solvent, the resulting oil was purified via silica chromatography using a 6:1 hexanes:ethyl acetate solvent mixture and concentrated under rotary evaporation until a white solid crashed out of solution. Obtained 201 mg (0.66 mmol), 70% yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 (d, *J* = 16.0 Hz, 1H), 6.06 (d, *J* = 16.1 Hz, 1H), 5.87 (s, 1H), 5.66 (s, 1H), 5.05 – 4.93 (m, 1H), 2.41 (d, *J* = 17.0 Hz, 1H), 2.22 (d, *J* = 17.2 Hz, 1H), 1.93 (d, *J* = 1.3 Hz, 3H), 1.86 (d, *J* = 1.4 Hz, 3H), 1.19 (dd, *J* = 6.2, 1.2 Hz, 6H), 1.04 (s, 3H), 0.95 (s, 3H). ^{13}C NMR (101 MHz, Methylene Chloride-d_2) δ 197.10, 165.35, 161.96, 148.64, 135.97, 128.04, 126.94, 119.08, 79.64, 67.06, 49.62, 41.34, 23.96, 22.74, 21.62, 20.80, 18.57. HRMS: calc. for $[C_{18}H_{26}O_4Na]^+$ ($[M+Na]^+$: 329.1723, found 329.1725.

4d – t-butyl ABA: To a stirring solution of ABA (100 mg, 0.378 mmol) in N,N-dimethylacetamide (3 mL) was added benzyltriethylammonium chloride (86 mg, 0.378 mmol), followed by 26 equiv. Cs_2CO_3 (3.205 g, 9.86 mmol) and 48 equiv. 2-bromo-2-methylpropane (2.037 mL, 18.14 mmol). The mixture was placed into an oil bath and heated to 55 °C for 24 hours, at which point starting material could not be observed on TLC. The solution was diluted with DI water, and the product was extracted from the aqueous phase 3x with EtOAc. This solution was then washed with 0.5 M HCl, 5% LiCl, then brine. The organic layer was dried over magnesium sulfate and the solvent removed under rotary evaporation. The resulting residue was taken up in DCM and the product was purified via silica chromatography using a 3:1 hexanes:ethyl acetate solvent mixture. A clear oil was obtained, which after rinsing with diethyl ether and concentrating by rotary evaporation crashed out of solution as white crystals. Obtained 107.8 mg (0.34 mmol), 89% yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (d, *J* = 16.1 Hz, 1H), 6.03 (d, *J* = 16.1 Hz, 1H), 5.85 (s, 1H), 5.61 (s, 1H), 2.42 (d, *J* = 17.1 Hz, 1H), 2.23 (d, *J* = 17.1 Hz, 1H), 1.90 (d, *J* = 1.3 Hz, 3H), 1.86 (d, *J* = 1.4 Hz, 3H), 1.42 (s, 9H), 1.04 (s, 3H), 0.94 (s, 3H). ^{13}C NMR (126 MHz, Methylene Chloride-d_2) δ 196.56, 164.71, 161.59, 146.91, 134.96, 127.17, 126.09, 119.71, 79.30, 78.86, 48.94, 40.61, 27.21, 23.31, 22.07, 20.06, 17.87. HRMS: calc. for $[C_{19}H_{28}O_4Na]^+$ ($[M+Na]^+$): 343.1880, found 343.1880.

Scheme 4. 5: General procedure for the synthesis of 3,3-dimethylacrylate esters (5-8).



To a stirring solution of 3,3-dimethylacrylic acid (500 mg, 5 mmol) in DCM was added 1.2 Equiv. DBU (6 mmol, 895 μL) and 1 equiv. Of EDCI•HCl (5 mmol, 960 mg), followed by 1 equiv. DMAP (5 mmol, 610 mg). 1.5 equiv. Of the corresponding alcohol was then added to the reaction mixture, and the solution was allowed to stir overnight. The next day, the solution was worked up by washing the organic layer with 0.5

M HCl, followed by a solution of saturated bicarbonate and brine. The organic layer was then dried over Na_2SO_4, which was removed by filtering and the resulting organic layer was concentrated under reduced pressure. The resulting oil was then dissolved in a small amount of DCM and the product was purified by silica chromatography using 24:1 hexanes:ethyl actetate.

Characterization of compound **6**, cyclopropyl(phenyl)methyl 3-methylbut-2-enoate:

Yield: 483 mg (2.09 mmol), 43%, obtained as clear oil.
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.46 – 7.28 (m, 5H), 5.81 (s, 1H), 5.30 (d, *J* = 8.6 Hz, 1H), 2.19 (s, 3H), 1.93 (s, 3H), 1.43 – 1.27 (m, 1H), 0.69 – 0.53 (m, 3H), 0.48 – 0.39 (m, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 165.99, 157.01, 140.97, 128.42, 127.75, 126.65, 116.33, 78.64, 27.49, 20.32, 16.81, 4.14, 3.20. HRMS: calc. for $[C_{15}H_{18}O_2]^+$ ($[M]^+$): 230.1307, found 230.1306.

Characterization of compound **7**, cyclohexyl 3-methylbut-2-enoate:

Yield: 554 mg, (3.04 mmol), 61%, obtained as clear oil.
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.66 (p, *J* = 1.4 Hz, 1H), 4.77 (td, *J* = 8.9, 4.2 Hz, 1H), 2.16 (d, *J* = 1.3 Hz, 3H), 1.80 – 1.68 (m, 3H), 1.66 – 1.49 (m, 3H), 1.49 – 1.28 (m, 5H), 1.27 (s, 2H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 166.13, 155.68, 116.62, 71.52, 31.71, 27.24, 25.37, 23.75, 20.04. HRMS: calc. for $[C_{11}H_{18}O_2]^+$ ($[M]^+$): 182.1307, found 182.1305.

Characterization of compound **8**, 2-isopropyl-5-methylcyclohexyl 3-methylbut-2-enoate (menthylbutenoate):

Yield: 536 mg (2.25 mmol), 52%, obtained as pale yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.64 (p, *J* = 1.4 Hz, 1H), 4.68 (td, *J* = 10.8, 4.4 Hz, 1H), 2.14 (d, *J* = 1.3 Hz, 3H), 2.04 – 1.95 (m, 1H), 1.91 – 1.81 (m, 4H), 1.65 (dp, *J* = 13.2, 3.1 Hz, 2H), 1.48 (dtq, *J* = 8.7, 6.2, 3.1 Hz, 1H), 1.36 (ddt, *J* = 14.1, 10.9, 3.1 Hz, 1H), 1.12 – 0.99 (m, 1H), 0.99 – 0.90 (m, 1H), 0.87 (dd, *J* = 6.8, 4.2 Hz, 7H), 0.74 (d, *J* = 7.0 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 166.35, 156.05, 116.56, 73.08, 47.16, 41.19, 34.36, 31.44, 27.38, 26.30, 23.57, 22.05, 20.78, 20.16, 16.44. HRMS: calc. for $[C_{15}H_{26}O_2]^+$ ($[M]^+$): 238.1933, found 238.1932.

4.4.4 Biocatalysis

General procedure for library lysis.

After subcloning BS2 variants from the phage vectors into the pET28 expression vector, the resulting library was transformed into *E. coli* and plated onto LB agar containing kanamycin. After 16 hours of growth at 37 °C, single colonies from the plate along with wild-type BS2 and empty pET28 vector were then picked into ten 1 mL 96 well plates charged with 300 µL LB media supplemented with kanamycin. The plates were allowed to grow with incubation at 37 °C at 250 rpm for 18 hours, at which point the cultures were removed from the incubator and 30 µL was used to inoculate 1 mL of TB media supplemented with kanamycin. Glycerol plates containing 50 µL of 50% glycerol and 50 µL of the cell cultures were prepared. The 2 mL deep well plates were grown at 37 °C at 250 rpm until the cells harboring the WT BS2 esterase reached 0.6-0.8 at which point gene expression was induced with 10 µL of an IPTG stock solution (10 mM stock concentration, 0.1 mM final). Gene expression was allowed to continue at 30 °C at 250 rpm for 20 hours, at which point the plates were removed from the incubator and centrifuged at 3600 rpm for 15 minutes in a Sorvall XT centrifuge. The supernatant was removed and the plates were frozen at -80 °C until needed.

The frozen 96-well plates were removed from the –80 °C freezer and allowed to warm for 10 minutes at room temperature. To each well in the 96 well plate, 100 µL of lysis buffer (0.75 mg/mL lysozyme, 50 mM phosphate, pH 7.5) was added, and the cells were resuspended by vortexing until homogenous. A non-breathable plate seal was affixed to each 96-well plate, and the cells were lysed at 37 °C, 250 rpm for 45 minutes. Once the lysis was complete, the samples were then flash frozen with liquid nitrogen and then gradually allowed to warm to room temperature. Once thawed, 10 µL of DNAse buffer (0.1 mg/mL DNAse I in 50 mM phosphate, pH 7.4) was added to each well and the samples were incubated at 37 °C for an additional 15 minutes. After DNAse treatment was finished, the cell lysate was centrifuged at 3600 rpm, 4 °C for 15 minutes to pellet cell debris. Cell lysate was then used immediately for bioconversions.

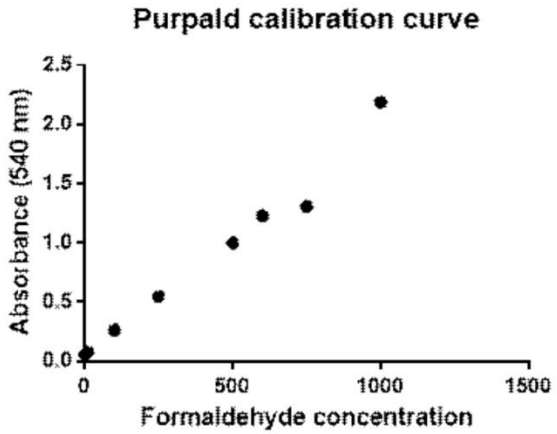
Purpald assay for P450 activity

A compilation plate of P450 enzymes in LB media + 25% glycerol was stamped into a 1 mL 96-well culture plate containing 300 µL sterile LB media supplemented with ampicillin. This plate was incubated at 37 °C in a horizontal incubator for 250 rpm for at least 16 hours. The next day, 120 µL aliquots of overnight cultures were transferred into 2 mL, deep well plates containing 1 mL TB containing ampicillin, trace metals mix, and 1 mM aminolevulinic acid. The cultures were incubated at 37 °C for 4h at 250 rpm, and 30 min after reducing the incubation temperature to 25 °C. 50 µLof IPTG (4.5 mM stock solution) was used to induce gene expression and the cultures were allowed to incubate for an additional 24 h at 25 °C. The next day, cells were pelleted via centrifugation at 3600 rpm in a sorvall XT centrifuge held to 4 °C for 15 minutes. The cell pellets were stored at – 80 °C until needed.

Cell lysis was accomplished according to the general procedure, except 300 µL of lysis buffer was added to each well of the plate containing the P450 variants. Once cell lysis was accomplished, 75 µL of clarified lysate was added to a microtiter plate, followed by 10 µL of phosphate buffer (50 mM, pH 7.4). To this lysate was added 5 µL NADP (10 mM stock, 500 µM final concentration), 3 µL **4a** (30 mM stock, 1 mM final concentration), 2 µL glucose (1 M stock, 20 mM final concentration), and reactions were initiated by 5 µL GDH-109 (180 U/mL stock, 9 U/mL final concentration). The plate was then covered with an aluminum plate seal and the reactions were then left to incubate for 3 hours at room temperature on a plate shaker set to 600 rpm. The reactions were then quenched via the addition of purpald in 2 M NaOH (20 µL of a 160 mM stock solution) along with a calibration curve consisting of known concentrations of benzaldehyde. For the calibration curve, see **Figure 4.7** below. The plate was incubated at room temperature for 45 minutes and the absorbance of each well was measured using a plate reader

at 540 nm. The results for active variants can be seen in **Figure 4.3A**.

Figure 4. 7: Purpald calibration curve for determination of formaldehyde concentration

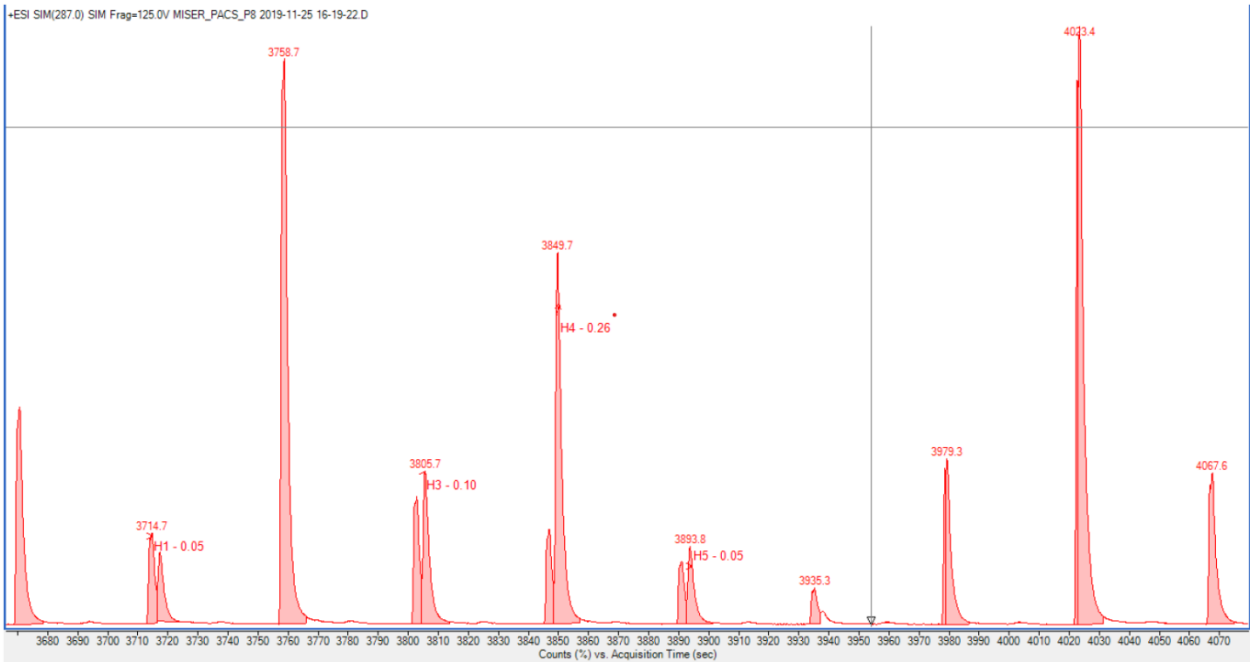


Each point represents a single measurement. Formaldehyde concentrations used were 1mM, 0.75 mM, 0.6 mM, 0.5 mM, 0.25 mM, 0.1 mM, and 0.01 mM.

MISER (Multiple Injections in a Single Experimental Run)[25] screening.

For all LC-MS usage, mobile phase A was water with 0.1% formic Acid, and mobile phase B was acetonitrile with 0.1% formic acid. MISER screening was performed on 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. A guard column was used to separate buffer from the analyte of interest, and the flow was directly analyzed by the mass-spec, bypassing the UV-Vis detector in the LC-MS system. A flow rate of 1 mL/min with an isocratic method of 32%B with a 42 second delay between injections was used for all MISER screening. The method was setup as an FIA method using the Chemstation software, and the peak height of the resulting abscisic acid product was used as a qualitative measure of conversion. MassHunter software was used to visually inspect and label the chromatograms. To monitor the product ion, SIM was used for the sodium adduct of the product peak (M+Na)+ corresponding to an m/z of 287. An example of the output from the MISERgram is included below.

Figure 4. 8: MISERgram cut-out



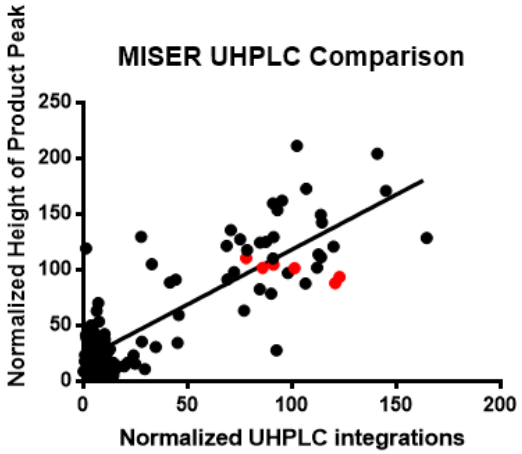
The presence of the second peak indicates product. SOURCE: ACS Central Science (2021) 7(9) 1581-1590

High-throughput MISER screen and comparison to low-throughput UHPLC.

To compare the reliability of the MISER screen to a low-throughput UHPLC screen, the first two plates of post-PACS samples were grown and lysed as described above. 5 µL of a 11.25 mM stock of **4b** in DMSO was delivered to a 96-well microtiter plate, followed by 20 µL of phosphate buffer (50 mM, pH 7.4). Reactions were initiated by the addition of 50 µL of lysate, and then the microtiter plate was covered by a non-breathable plate seal and the reactions incubated for 90 minutes without stirring. After 90 minutes time, the reactions were quenched via the addition of one volume of methanol. The resulting precipitated protein was separated via centrifugation at 3600 rpm, 4 °C for 15 minutes. The samples were then further diluted via addition of 150 µL water to 30 µL of quenched bioconversion and filtered through a 0.22 µm filter plate (Agilent product number 201276-100) into a polystyrene microtiter plate. These plates were then heat-sealed and submitted for analysis by the MISER screen described above. After the high-throughput analysis by LC-MS, the same plates were then removed from the autosampler and re-filtered through the 0.22 µm filter plates and analyzed by low-throughput UHPLC-MS using method A. To provide

a qualitative measurement of activity in the high-throughput MISER approach, the peak height of the abscisic acid product peak was used. For the low-throughput UHPLC data, the qualitative measure was the area under the curve of the product peak over the total area under the curve of the product and starting material peaks. To normalize the data between the two different analyses, the values for each well were normalized to the intra-plate average of the WT values. Microsoft Excel was used for data processing, and Graphpad Prism used for data visualization.

Figure 4. 9: MISER/Low throughput UHPLC comparison



Low-throughput UHPLC compared to high-throughput MISER screen, with WT conversion normalized to 100 for both. WT values are highlighted in red. Pearson correlation coefficient was calculated to be 0.7234. SOURCE: ACS Central Science (2021) 7(9) 1581-1590

Creation and screening of the PACS compilation plate.

After screening the 10 post-PACS plates via MISER, active variants were pooled into a single 96-well plate. This plate was stamped out twice, and secondary plates were inoculated and induced as described per the general library preparation. For the bioconversions with **4b**, 5 µL of a 11.25 mM stock solution of **4b** in DMSO was added to a microtiter plate and diluted with 25 µL of phosphate buffer (50 mM, pH 7.5). The reactions were initiated with the addition of 45 µL of clarified lysate, and these reactions were allowed to incubate for 90 minutes at room temperature without shaking. After 90 minutes, the reactions were quenched via addition of one volume of methanol, and the precipitated protein was sedimented via

centrifugation at 3600 rpm for 15 minutes at 4 °C. The reactions were then diluted with 30 µL of lysate and 150 µL of water, then analyzed by LC-MS. UV-absorbance at 254 nm was used to provide a qualitative measure of conversion. Data processing was done using the ChemStation software and Microsoft Excel, and the data was visualized using GraphPad Prism.

To screen the compilation plate for substrate **5**, the same procedure was followed, except 5 µL of a 100 mM stock solution of substrate **5** was diluted in 45µL phosphate buffer, followed by the addition of 50 µL clarified lysate. These reactions were then incubated for 60 minutes while shaking on a Heidolph Titramax 1000 plate shaker at 750 rpm. They were processed the same as the reactions above, and analyzed by UHPLC using method C. The plate was screened in triplicate, and the top five mutants expressed for purified reactions with substrate **5**.

Purification and characterization of PACS hits with abscisic acid esters 4a-4d and 3,3-dimethylacrylate esters 5, 6, 7, 8.

The top 8 variants for which good sequencing could be obtained were grown as overnights in 10 mL LB with 50 ng/µL kannamycin. 5 mL of this overnight culture was used to innoculate 250 mL of TB + 50 ng/µL kannamycin and the cells were grown at 37 °C shaking at 250 rpm until the OD600 reached 0.8, at which point the temperature was reduced to 30 °C and protein expression was induced with 25 µL of a 1 M stock solution of IPTG for a final concentration of 100 µM. These cultures were allowed to continue incubating for 20 hours at 250 rpm, at which point the cells were transferred to cell buckets and centrifuged at 3600 rpm, 4 °C for 20 minutes. The supernatent was removed and the cell pellets were stored at –20 °C until ready for purification. To begin purifying the protein, each cell bucket was resuspended in 40 mL of phopshate buffer (50 mM, pH 7.5), and the resuspended cell pellets were transferred to a 50 mL centrifuge tube. These were immediately placed on ice, and the cells were lysed by sonication 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. The cells were

then centrifuged in a high-speed rotor at 15,000 rpm for 30 minutes at 4 °C. The clarified lysate was then purified via IMAC using Nickel-NTA resin. Buffer exchanging was done by dialyzing the eluted protein sample in phosphate buffer (50 mM, pH 7.5) overnight. The next day, the samples were concentrated in a Amicon ultracentrifugation column (10 kDa MW cutoff) and the buffer was diluted with 10 % glycerol to act as a cryoprotectant. Protein concentration was determined by the absorbance at 280 nm with the extinction coefficient calculated from the protein sequence using the Expasy Protein Parameters website. Purified enzymes were stored at –80 °C until used.

To begin the bioconversions with **4a-4d**, 100 µM stock solutions of each enzyme were prepared. 1.88 µL of 30 mM 4a-4d were added to a microtiter plate, followed by 69.4 µL of phosphate buffer (50 mM, pH 7.5). Reactions were initiated by the addition of 3.75 µL purified enzyme to the reactions, at which point the plates were sealed with a non-breathable plate seal and the reactions were left at room temperature with no shaking for one hour. After this point, one volume of quench solution (200 µM 2-Acetamidophenol in methanol) was added and the reactions were processed as described for the lysate reactions above. Reactions were analyzed by UHPLC monitored at 280 nm using method B. Product concentration was determined using a calibration curve.

Bioconversions with substrates **5-8** were performed at room temperature. Due to inconsistent mixing in the absence of stirring as a result of substrate insolubility, bioconversion plates were shaken at 750 rpm using a Heidolph Titramax 1000 plate shaker. 100 µM stock solutions were prepared for each enzyme. Substrate stock solutions were prepared as 100 mM solutions in DMSO, prepared fresh prior to use. From the 100 mM stocks, 5 µL of each substrate was added to a polypropylene microtiter plate, and these were diluted with 85 µL of phosphate buffer (pH 7.5, 50 mM). To initiate the reactions, 10 µL of the 100 µM stock solution of each enzyme was added to the microtiter plate. Reactions were run in triplicate and

allowed to react for one hour. To end the reactions, each reaction was quenched using one reaction volume of quench solution (200 µM 5-bromoindole as internal standard in methanol). These reactions were then further diluted 1:4 in water and processed the same as the general lysate bioconversions. Reactions were analyzed by UHPLC, using method C and the amount of product was quantified by using a calibration curve at 230 nm.

Kinetic characterization of WT-BS2 and 10D2 with 4a and 4b.

In general for kinetic characterization, 3x stock solutions of the enzyme were prepared (i.e. for 2.5 µM final concentration, stocks of 7.5 µM enzyme were prepared). Substrates were prepared as 1.5x stocks. 50 µL of the substrate stock solutions were added to a 96-well microtiter plate in triplicate, and the reactions were initiated by the addition of 25 µL of the enzyme stock solutions. Reactions were left at room temperature without any stirring until the timepoints were finished, at which point one volume of quench solution (200 µM internal standard in methanol, 100 µM final concentration) was added to the reactions. After the last timepoint was quenched, the reactions were processed as described for the general lysate bioconversions. Reactions were analyzed using UHPLC using method B. Abscisic acid was quantified using a calibration curve with the interal standard measured at 280 nm.

Table 4. 1: Sequences of top 8 variants from PACS compilation plate screen

Identity	Plate	Mutations	Normalized lysate conversion	New Mutations	Non-silent mutations
P10-D-02	P10	E18G, M193T, M221T, K342E, E352G, T462A, N473S	456	E18G, M193T, M221T, E352G, T462, N473S	E18G, M193T, M221T, E352G, T462, N473S
P05-F-05	P05	L68P, K342E, E424G	356	L68P, E424G	L68P
P01-B-08	P01	N240S, A291A, K342E	251	N240S, A291A	N240S
P02-G-04	P02	F39L, N240S, A291A, K342E	233	F39L, N240S, A291A	F39L, N240S
P08-F-11	P08	L67L, P76R, E171V, T220A, D260D, D320G, K342E	215	L67L, P76R, T220A, D260D, D320G	P76R, T220A, D320G
P02-E-06	P02	L334F, M357L, R365C	203	L334F, M357L, R365C	L334F, M357L, R365C
P08-E-06	P08	K14R, Q36Q, R263W, F405L	199	K14R, Q36Q, R263W, F405L	K14R, R263W, F405L
P07-A-07	P07	P90T, I130V, A203A, A227A, N269S, L339P, K479K	181	P90T, I130V, A203A, A227A, N269S, L339P	P90T, I130V, N269S, L339P

Table 4. 2: Kinetic parameters for WT-BS2 and evolved variant 10D2 on substrates 4a & 4b

	4a			4b		
Variant	K_{cat} (min^{-1})	K_M (μM)	K_{cat}/K_M (min^{-1} μM^{-1})	K_{cat} (min^{-1})	K_M (μM)	K_{cat}/K_M (min^{-1} μM^{-1})
WT	0.092	111	0.000829	0.06784	143.2	0.000474
10D2	1.583	5615	0.00282	3.244	1265	0.00256

Table 4. 3: Methods for UHPLC and UHPLC-MS analysis

Method A:

UHPLC-MS ABA	Time (min)	Starting %B	Ending %B
Mobile Phase A = Water + 0.1 % FA	0-0.5	30	30
Mobile Phase B = Acetonitrile + 0.1 % FA	0.51-1.25	30	50
	1.26-1.3	50	80
	1.31-1.75	80	95
	1.76-2.25	95	95
	2.26-2.5	30	30
	0.5 post run	30	30

Method B:

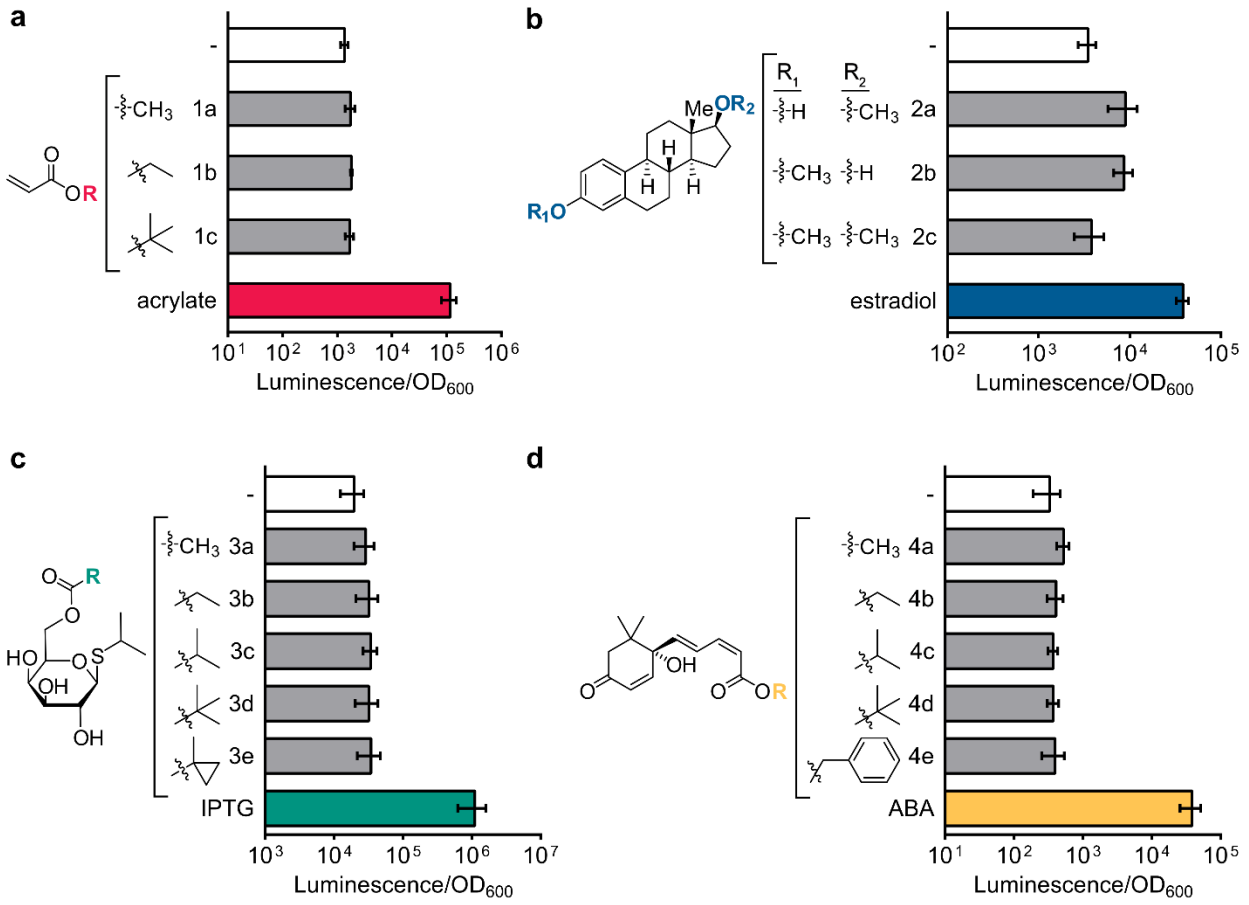
UHPLC-ABA	Time (min)	Starting %B	Ending %B
Mobile Phase A = Water + 0.1 % TFA	0-0.5	25	25
Mobile Phase B = Acetonitrile + 0.1 % TFA	0.51-1.2	25	45
	1.21-2	60	68
	2.01-2.5	68	95
	2.51-2.8	95	95
	2.81-3	25	25
	0.5 post run	25	25

Method C:

UHPLC – DMA Esters	Time (min)	Starting %B	Ending %B
Mobile Phase A = Water + 0.1 % TFA	0-0.5	25	25
Mobile Phase B = Acetonitrile + 0.1 % TFA	0.51-2	25	50
	2.01-3.5	50	95
	3.51-4.5	95	95
	4.51-5	25	25
	1.0 post run	25	25

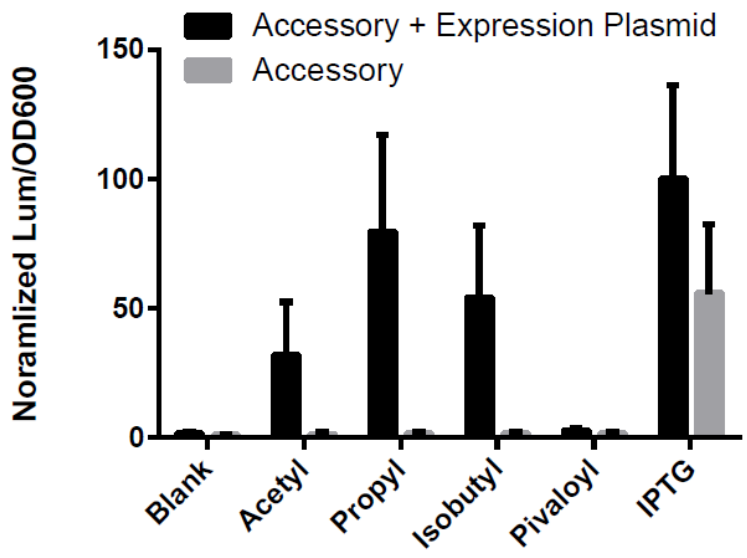
4.5 Additional Data

Figure 4. 10: Small molecule biosensors can be used to detect selective biocatalysis



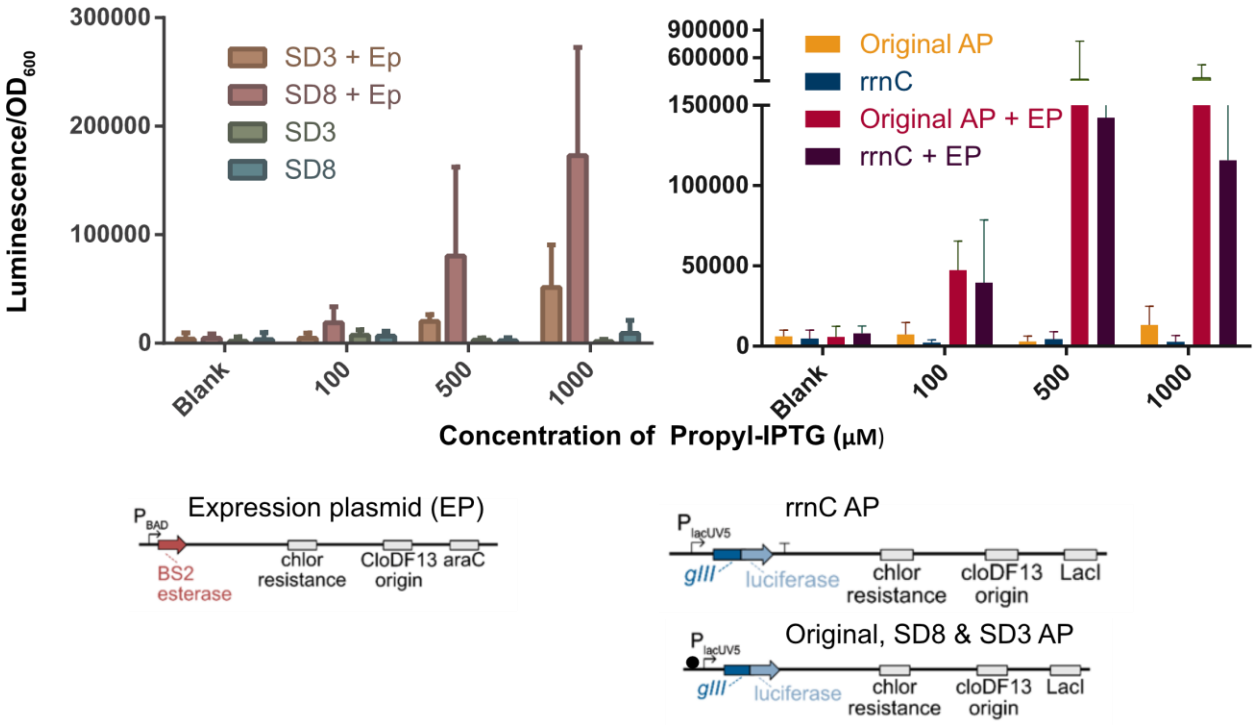
(a) *E. coli* expressing AcuR biosensor were incubated in the absence or presence of acrylate esters or acrylate for 3 h and then analyzed for luminescence. (b) *E. coli* expressing estradiol biosensor were incubated in the absence or presence of methylated estradiol or estradiol for 3 h and then analyzed for luminescence. (c) *E. coli* expressing IPTG biosensor were incubated in the absence or presence of IPTG esters or IPTG for 3 h and then analyzed for luminescence. (d) *E. coli* expressing ABA biosensor were incubated in the absence or presence of ABA esters or ABA for 3 h and then analyzed for luminescence. Error bars are the standard deviation for n = 4 replicates. SOURCE: ACS Central Science (2021) 7(9) 1581-1590

Figure 4. 11: In vivo IPTG ester hydrolysis by BS2 esterase



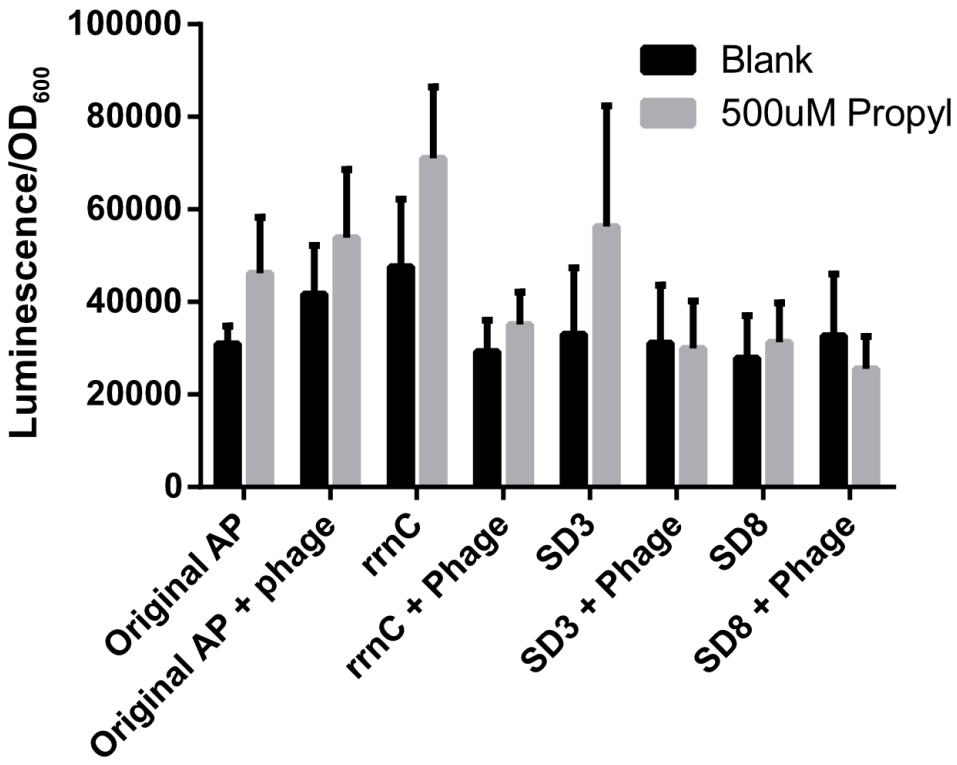
E. coli expressing either just the IPTG biosensor (gray, "accessory plasmid") or both the biosensor and expression plasmid (black, 'Accessory + Expression plasmid) were incubated in the absence or presence of IPTG esters or IPTG for 3 h and then analyzed for luminescence.

Figure 4. 12: Optimization of accessory plasmids for IPTG biosensor



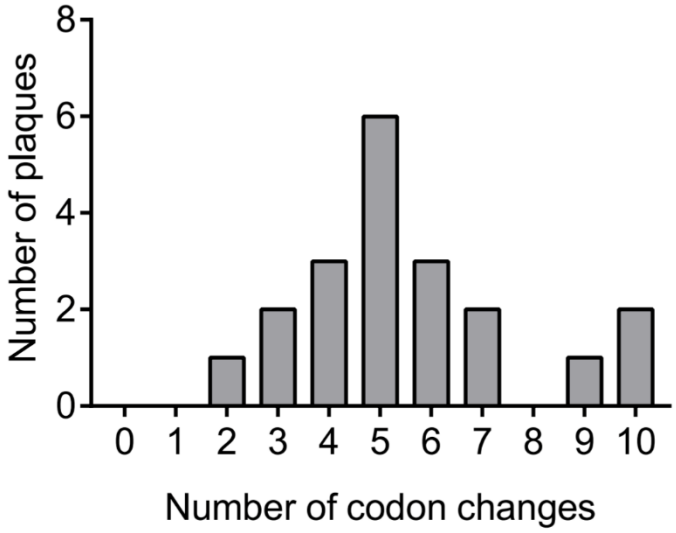
Accessory plasmids containing modified RBS sequences (SD8 = strong RBS, SD3 = weak RBS) and an additional terminator sequence (rrnC) were compared to the original plasmid. In *E. coli*, the rrnC containing accessory plasmid shows lower background signal in the absence of BS2 expression while maintaining high turn-on when esterase is expressed. Black circle represents the location of RBS inclusion, terminator is represented by the "T" on the plasmid map.

Figure 4. 13: Optimized plasmids show no difference from original accessory plasmid when esterase is expressed from selection phage.



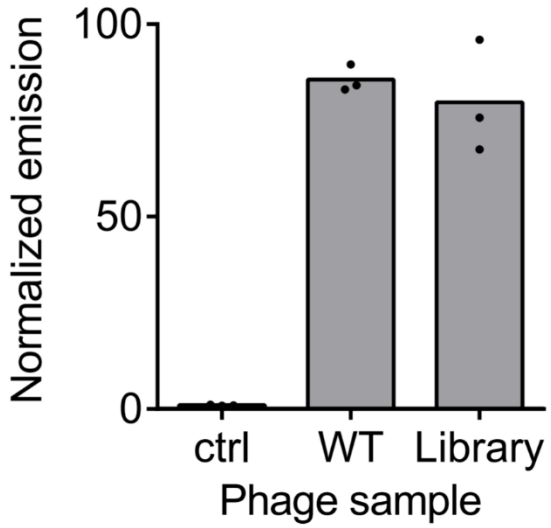
In each example, no increase in luminescent signal is observed when the selection phage is included, indicating no BS2 catalyzed hydrolysis is occurring.

Figure 4. 14: Library of BS2 phage variants used in ABA PANCE experiment generated via error prone PCR.



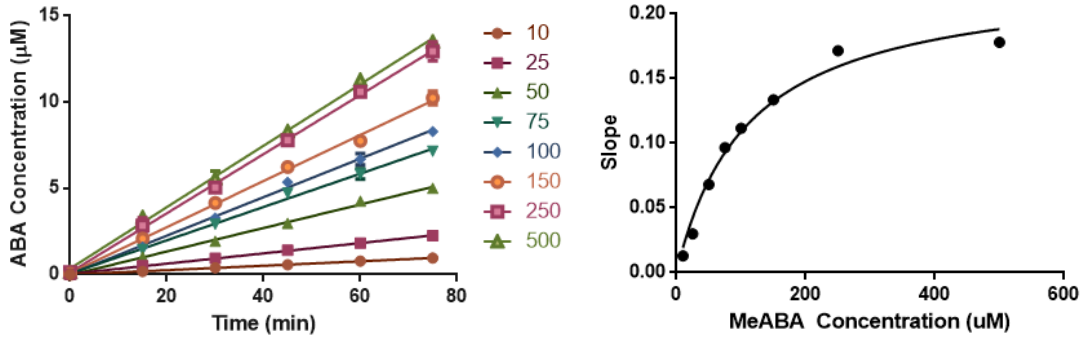
Codon changes per gene for 20 randomly selected plaques from BS2 phage library. SOURCE: ACS Central Science (2021) 7(9) 1581-1590.

Figure 4. 15: ePCR generated library of BS2 variants contains active variants



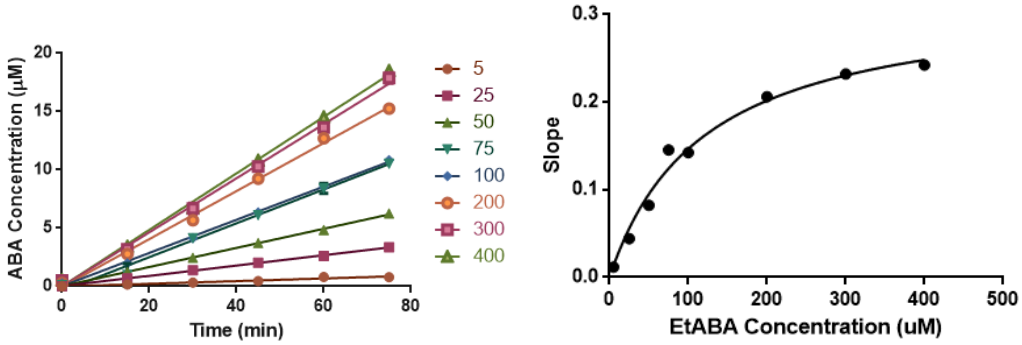
Control phage (HRV), BS2 WT, or BS2 phage library (1 µL) was incubated with a fluorescein α-cyclopropyl ester (fluorescein-CM_2) fluorogenic molecule and 1059 cells overnight. Cultures were then analyzed for emission. Three samples of bulk phage (1 x 10^{12} PFU/mL) were analyzed. SOURCE: ACS Central Science (2021) 7(9) 1581-1590.

Figure 4. 16: Kinetics for wild-type BS2 with methyl ABA ester 4a



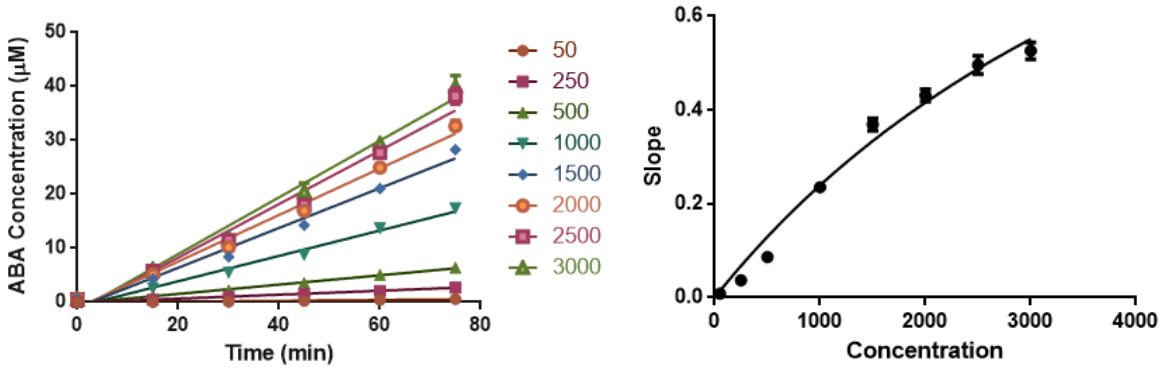
Final enzyme concentration was 2.5 µM, with timepoints taken every 10 minutes. (a) Conversion vs time (b) Michaelis-Menten plot used to calculate K_M, K_{cat}. Slope calculations and kinetic parameters were calculated using Graphpad Prism. SOURCE: ACS Central Science (2021) 7(9) 1581-1590.

Figure 4. 17: Wild-type BS2 kinetics with ethyl ABA ester 4b



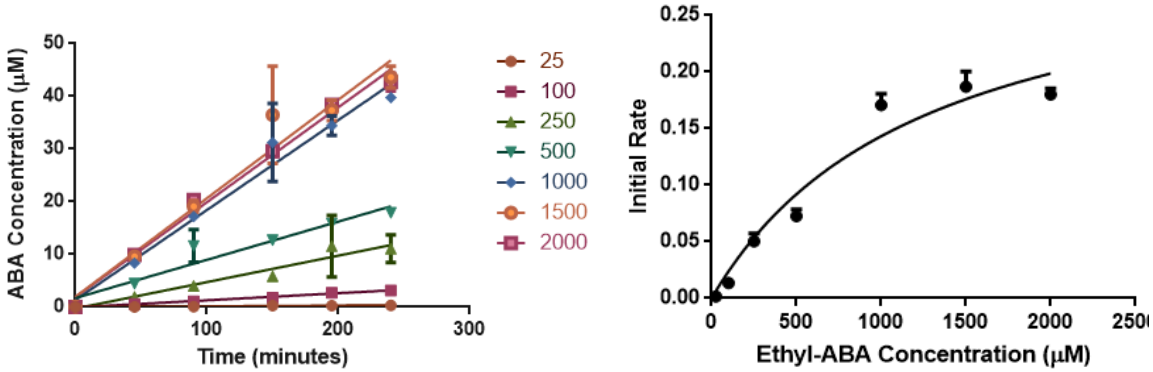
Final enzyme concentration was 5 µM, with timepoints taken every 10 minutes. (a) Conversion vs time (b) Michaelis-Menten plot used to calculate KM, Kcat. Slope calculations and kinetic parameters were calculated using Graphpad Prism. SOURCE: ACS Central Science (2021) 7(9) 1581-1590.

Figure 4. 18: 10D2 BS2 variant kinetics with methyl ABA substrate 4a



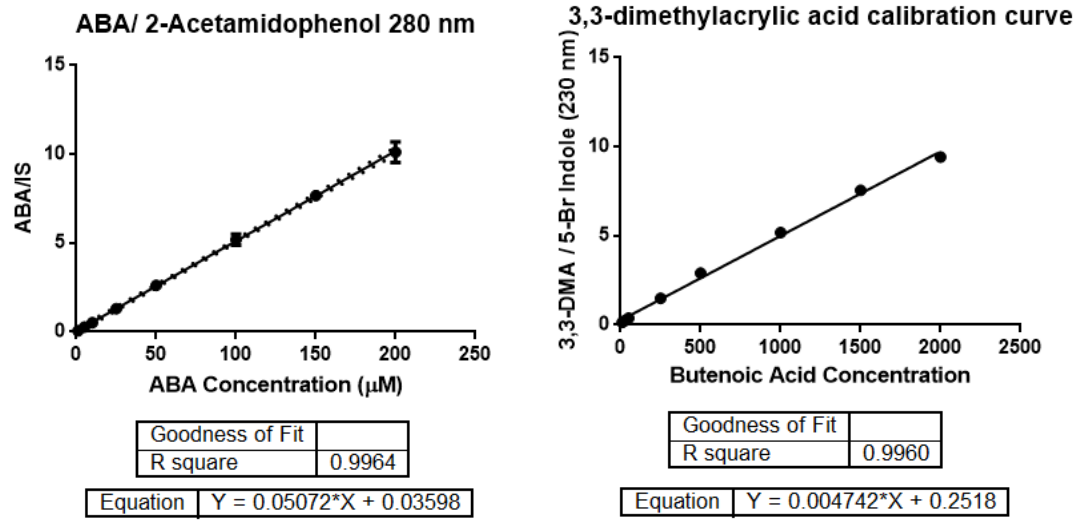
Final enzyme concentration was 1 µM, with timepoints taken every 10 minutes. (a) Conversion vs time (b) Michaelis-Menten plot used to calculate KM, Kcat. Slope calculations and kinetic parameters were calculated using Graphpad Prism. SOURCE: ACS Central Science (2021) 7(9) 1581-1590.

Figure 4. 19: 10D2 kinetics with ethyl ABA substrate 4b



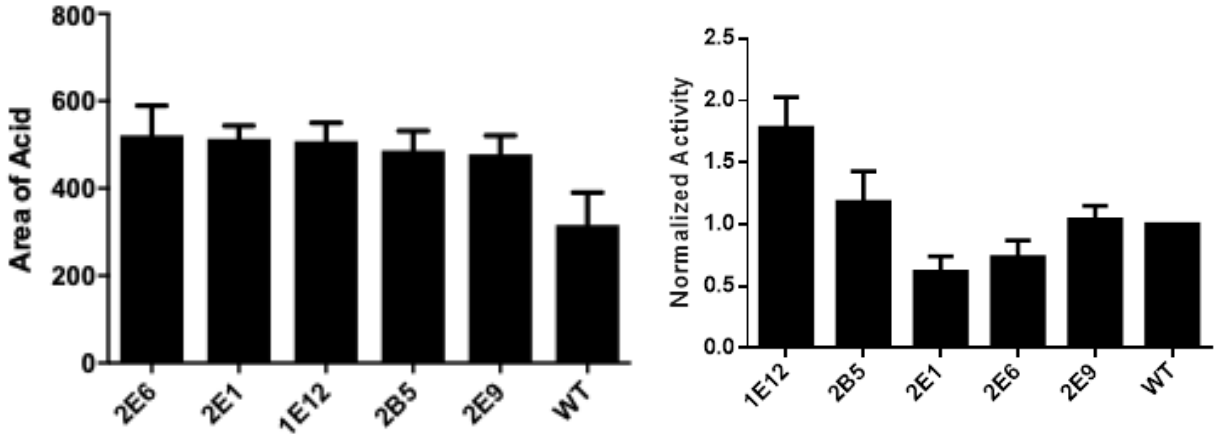
Final enzyme concentration was 100 nM, with timepoints taken every 10 minutes. (a) Conversion vs time (b) Michaelis-Menten plot used to calculate KM, Kcat. Slope calculations and kinetic parameters were calculated using Graphpad Prism. SOURCE: ACS Central Science (2021) 7(9) 1581-1590.

Figure 4. 20: Calibration curves used to quantitate analytes in this study



Left: Calibration curve for abscisic acid with 2-acetamidophenol as the internal standard. Right: A. Calibration curve for 3,3-dimethylacrylic acid with 5-bromo indole as the internal standard. SOURCE: ACS Central Science (2021) 7(9) 1581-1590.

Figure 4. 21: Evaluation of BS2 variants obtained from PACS for the hydrolysis of substrate 5



Left: Lysate screening to identify variants from the compilation plate improved for hydrolysis of **5**. Right: Activity of purified variants with substrate **5**.

4.6 References

(1) Hughes, G.; Lewis, J. C. Introduction: Biocatalysis in Industry. *Chemical Reviews*. 2018, pp 1–3. https://doi.org/10.1021/acs.chemrev.7b00741.

(2) Burton, S. G.; Cowan, D. A.; Woodley, J. M. The Search for the Ideal Biocatalyst. *Nature Biotechnology*. 2002, pp 37–45. https://doi.org/10.1038/nbt0102-37.

(3) Ma, S. K.; Gruber, J.; Davis, C.; Newman, L.; Gray, D.; Wang, A.; Grate, J.; Huisman, G. W.; Sheldon, R. A. A Green-by-Design Biocatalytic Process for Atorvastatin Intermediate. *Green Chemistry* **2010**, *12* (1), 81–86. https://doi.org/10.1039/b919115c.

(4) Liu, Z.; Arnold, F. H. New-to-Nature Chemistry from Old Protein Machinery: Carbene and Nitrene Transferases. *Current Opinion in Biotechnology*. 2021, pp 43–51. https://doi.org/10.1016/j.copbio.2020.12.005.

(5) Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med Chem Lett* **2017**, *8* (5), 476–480. https://doi.org/10.1021/acsmedchemlett.7b00114.

(6) Fisher, B. F.; Snodgrass, H. M.; Jones, K. A.; Andorfer, M. C.; Lewis, J. C. Site-Selective C-H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. *ACS Cent Sci* **2019**, *5* (11), 1844–1856. https://doi.org/10.1021/acscentsci.9b00835.

(7) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(8) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science (1979)* **2010**, *329* (5989), 305–309. https://doi.org/10.1126/science.1188934.

(9) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **2011**, *472* (7344), 499–503. https://doi.org/10.1038/nature09929.

(10) Dickinson, B. C.; Leconte, A. M.; Allen, B.; Esvelt, K. M.; Liu, D. R. Experimental Interrogation of the Path Dependence and Stochasticity of Protein Evolution Using Phage-Assisted Continuous Evolution. *Proc Natl Acad Sci U S A* **2013**, *110* (22), 9007–9012. https://doi.org/10.1073/pnas.1220670110.

(11) Hubbard, B. P.; Badran, A. H.; Zuris, J. A.; Guilinger, J. P.; Davis, K. M.; Chen, L.; Tsai, S. Q.; Sander, J. D.; Joung, J. K.; Liu, D. R. Continuous Directed Evolution of DNA-Binding Proteins to Improve TALEN Specificity. *Nat Methods* **2015**, *12* (10), 939–942. https://doi.org/10.1038/nmeth.3515.

(12) Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Söll, D.; Liu, D. R. Continuous Directed Evolution of Aminoacyl-TRNA Synthetases. *Nat Chem Biol* **2017**, *13* (12), 1253–1260. https://doi.org/10.1038/nchembio.2474.

(13) Roth, T. B.; Woolston, B. M.; Stephanopoulos, G.; Liu, D. R. Phage-Assisted Evolution of Bacillus Methanolicus Methanol Dehydrogenase 2. *ACS Synth Biol* **2019**, *8* (4), 796–806. https://doi.org/10.1021/acssynbio.8b00481.

(14) Jones, K. A.; Snodgrass, H. M.; Belsare, K.; Dickinson, B. C.; Lewis, J. C. Phage-Assisted Continuous Evolution and Selection of Enzymes for Chemical Synthesis. *ACS Cent Sci* **2021**, *7* (9), 1581–1590. https://doi.org/10.1021/acscentsci.1c00811.

(15) Shibata, A.; Ito, Y.; Abe, H. RNA-Templated Molecule Release Induced Protein Expression in Bacterial Cells. *Chemical Communications* **2013**, *49* (3), 270–272. https://doi.org/10.1039/c2cc37826d.

(16) Schmidt, M.; Henke, E.; Heinze, B.; Kourist, R.; Hidalgo, A.; Bornscheuer, U. T. A Versatile Esterase from Bacillus Subtilis: Cloning, Expression Characterization, and Its Application in Biocatalysis. *Biotechnol J* **2007**, *2* (2), 249–253. https://doi.org/10.1002/biot.200600174.

(17) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Methods* **2009**, *6* (5), 343–345. https://doi.org/10.1038/nmeth.1318.

(18) Boeke, J. D.; Model, P.; Zinder, N. D. Effects of Bacteriophage F1 Gene III Protein on the Host Cell Membrane. *MGG Molecular & General Genetics* **1982**, *186* (2), 185–192. https://doi.org/10.1007/BF00331849.

(19) Chao, Y. P.; Chiang, C. J.; Hung, W. bin. Stringent Regulation and High-Level Expression of Heterologous Genes in Escherichia Coli Using T7 System Controllable by the AraBAD Promoter. *Biotechnol Prog* **2002**, *18* (2), 394–400. https://doi.org/10.1021/BP0101785.

(20) Penumetcha, P.; Lau, K.; Zhu, X.; Davis, K.; Eckdahl, T. T.; Campbell, A. M. Improving the Lac System for Synthetic Biology. *Bios* **2010**, *81* (1), 7–15. https://doi.org/10.1893/011.081.0104.

(21) Pu, J.; Kentala, K.; Dickinson, B. C. Multidimensional Control of Cas9 by Evolved RNA Polymerase-Based Biosensors. *ACS Chem Biol* **2018**, *13* (2), 431–437. https://doi.org/10.1021/acschembio.7b00532.

(22) Hughes, D. L. The Mitsunobu Reaction. *Organic reactions* **1992**, 335–656. https://doi.org/10.1002/chin.200331284.

(23) Tsunoda, T.; Otsuka, J.; Yamamiya, Y.; Itô, S. N , N , N ', N '-Tetramethylazodicarboxamide (TMAD), A New Versatile Reagent for Mitsunobu Reaction. Its Application to Synthesis of Secondary Amines. *Chem Lett* **1994**, *23* (3), 539–542. https://doi.org/10.1246/cl.1994.539.

(24) Reisky, L.; Büchsenschütz, H. C.; Engel, J.; Song, T.; Schweder, T.; Hehemann, J.-H.; Bornscheuer, U. T. Oxidative Demethylation of Algal Carbohydrates by Cytochrome P450 Monooxygenases. *Nat Chem Biol* **2018**. https://doi.org/10.1038/s41589-018-0005-8.

(25) 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* (13), 1674–1681. https://doi.org/https://doi.org/10.1016/j.tetasy.2010.05.029.

CHAPTER 5: BIOSENSOR DRIVEN SCREENING OF BIOCATALYSTS

5.1: Introduction

The implementation of directed evolution for the improvement of biocatalysts has undoubtedly transformed chemistry in the last two decades.[1–3] Despite this potential and promise, the elongated time frame required for directed evolution campaigns is largely the result of the limited throughput of screening technology.[4] Automation and the use of liquid handling robots can provide improvements to sample handling capacity, but the bottleneck of directed evolution is typically screening capacity. Even with advancements in target agnostic methods such as mass spectrometry, most methods are limited to the evaluation of roughly 1000 variants per round of evolution.[5] One of the longest standing methods for screening large libraries quickly is to link the desired effect to a fluorescent output. Fluorogenic reactions have been used since the beginning of directed evolution, with some of the earliest evolution campaigns utilizing esterase catalyzed hydrolysis of *p*-nitrobenzoate esters.[6]

The advantage to this screening method is the ability to rapidly assay entire 96- or 384-well plates of variants with assay frequency of less than 5 seconds per sample. Coupling these fluorescent plate assays to automated liquid handling methods and on-line sequencers could provide structure-activity information for thousands of variants with minimal researcher intervention and greatly facilitate directed evolution (**Figure 5.1**). The disadvantage to utilizing fluorogenic substrates for screening large libraries is the reliance on properties inherent to the substrate itself. Most common fluorogenic substrates consist of large conjugated aromatic systems to provide a signal in the visible region. Often, the required uncaging of the fluorophore in these substrates may not be amenable for screening the desired biotransformation. Furthermore, even when the uncaging can be linked to the desired biotransformation, the substrate is often not representative of viable targets. This structural difference between the probe and target substrate means there is no guarantee the observed uncaging of the fluorogenic probe substrate correlates with the desired reaction.

Figure 5. 1: Methods for screening and advantages and disadvantages

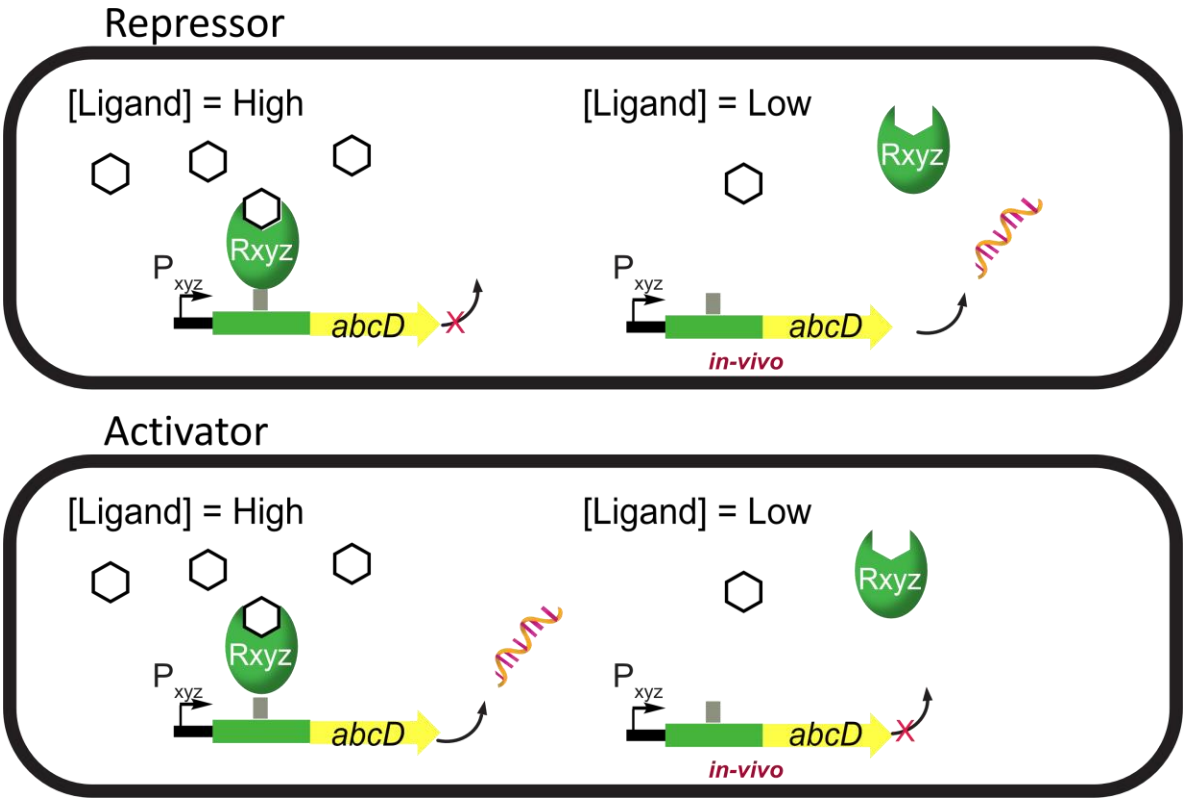
Advantages	Target agnostic	Least expensive option, cells are sequesteered in well plates, library size can expand up to 10^5	Library size can expand to 10^6
Disadvantages	Expensive, library size limited to 10^3	Typically relies on substrate or cofactor dependent detection	The potential for substrate diffusion must be mitigated, typically relies on substrate or cofactor dependent detection, expensive to operate and maintain.

Rather than relying on the substrate properties, an alternative approach would be to develop a system for detecting transformations dependent on the enzyme class. Methods to monitor specific enzyme reactions such as imine reductases,[7] P450 hydroxylases,[8] and other redox chemistries[9] show the promise of these approaches. These methods most commonly rely on the production of a reactive byproduct such as formaldehyde[8] or the detection of a specific form of a cofactor.[10] Byproduct coupled fluorescent outputs have been successfully used for screening large enzyme libraries but suffer from drawbacks such as the toxicity of reactive byproducts or the potential for uncoupling of cofactor consumption and enzyme activity. By comparison, methods that directly connect the fluorescent output to biocatalyst activity are rare.

In nature, proteins that modulate expression of genes in response to small molecule ligands are called transcription factors. Transcription factors are commonly found upstream of gene clusters and act to regulate the expression of metabolic genes in response to the concentration of a specific metabolite. This regulation of gene clusters can prevent the production of cofactor consuming proteins, redirect the ATP used in protein and mRNA synthesis for essential cell function when the gene products are unnecessary, or control the concentration of toxic secondary metabolites. Ligand dependent transcription factors respond to the presence of ligand in two ways. Transcription factors can act as activators to upregulate gene expression or repressors to decrease expression of downstream genes

(**Figure 5.2**). In addition to being able to detect the presence of a small molecule ligand, transcription factors commonly exhibit a dose-response factor.[11] This means the more a ligand is present to bind to the transcription factor, the more the transcription factor can influence downstream regulation of gene expression. Ligand responsive transcription factors have not gone unnoticed in the synthetic biology community. Protein based TFs are commonly used to improve titers of products produced by heterologously expressed biosynthetic gene clusters. The dose-dependent regulation of gene expression allows for the prevention of toxic metabolite buildup or prevents bottlenecks from occurring by carefully regulating the levels of intermediates.[12]

Figure 5. 2: Transcription factors as activators and repressors



When ligand-inducible transcription factors act as repressors, high concentration of the effector ligand results in binding of the TF to an operator region of DNA preventing transcription of downstream genes. When the concentration of ligand is low, the TF undergoes a conformational change that releases it from the operator sequence allowing transcription. For activators the inverse is true, and high concentration of ligand results in downstream gene expression.

Transcription factors can be evolved for improved dynamic range,[11] sensitivity,[13] or substrate specificity,[14] making them ideal for product-dependent assays to detect biocatalyst activity. Rather than regulating expression of downstream metabolic genes, these ligand-responsive transcription factors can also be used to regulate expression of a reporter protein, commonly a fluorescent protein. In this context, the transcription factor allows for gene expression and the eventual maturation of fluorescent protein only in cell lines that produce sufficient quantities of the effector ligand. Due to the fluorescent protein being localized within the cell, it is possible to use these biosensors in conjunction with cell sorters to allow for the rapid evaluation of up to hundreds of thousands of variants if the effector ligand does not diffuse out from the cell to turn on gene expression in non-producing cells.[15] In cases where the effector ligand is prone to diffusion, these assays can often be amended to allow for screening of samples in microtiter plates. While the throughput of microtiter plate analysis is lower than for cell sorting (**Figure 5.1**), it still exceeds limitations of UHPLC screening by an order of magnitude.

Despite the ability to develop transcription factors as product dependent biosensors they are rarely implemented to improve biocatalytic workflows. With the aim to bridge this gap, we envisioned that by demonstrating several gene circuits coupling the production of products from biocatalytic reactions to fluorescent gene production, we could show the potential of transcription factors to screen enzyme libraries. To this end, we explore both the *in vivo* utilization of biosensors for cell sorting, as well as the *in vitro* use of biosensor expressing cells to provide a fluorescent output in response to biocatalyst activity. Those efforts are described herein.

Authorship

My primary responsibilities on this project included ligand synthesis, library screening after completion of the ABA sort, and lead of the initial studies towards implementing the tryptophan biosensor for halogenase screening. Unless otherwise mentioned, all biosensor design, evaluation, and other cloning

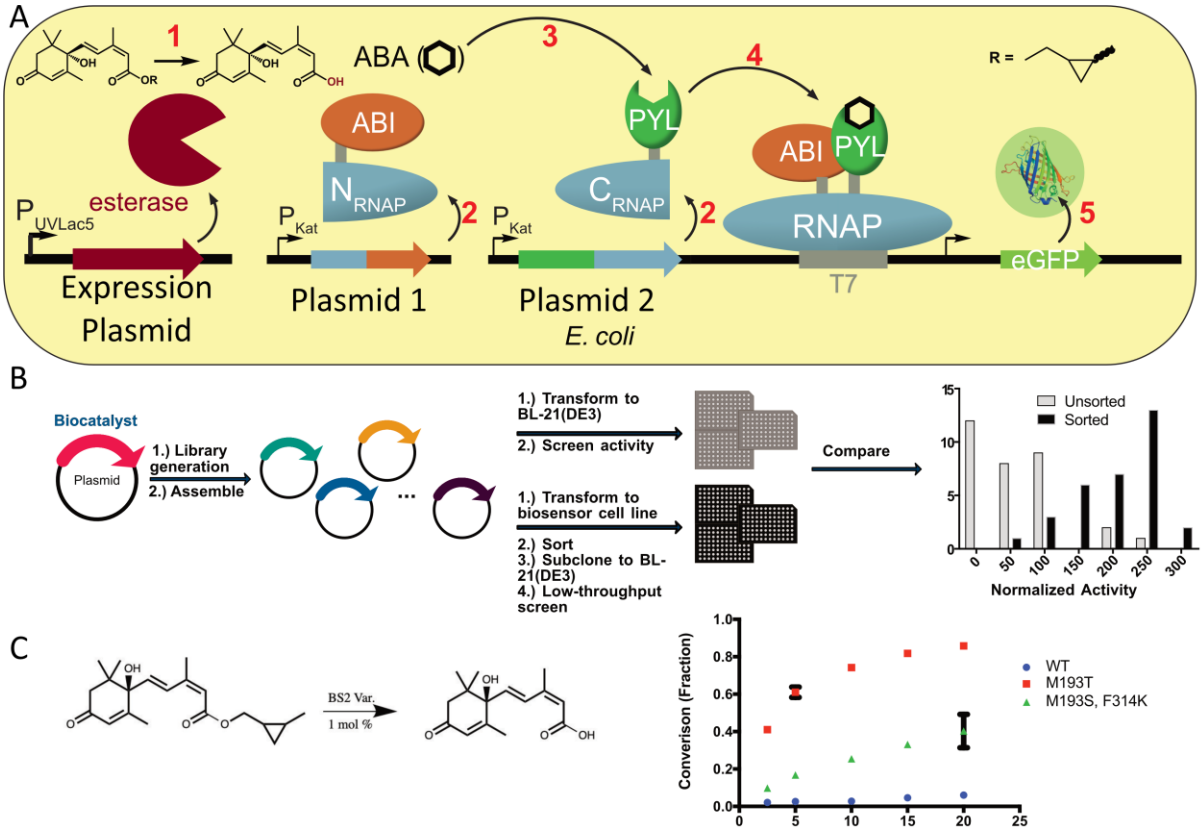
was conducted by Dr. Krysten Jones. Studies for the *in vitro* evaluation of the 6-halogenase Thal with the engineered tryptophan biosensor were conducted by Caitlin Roof.

5.2 Results & Discussion

5.2.1 Evaluation of BS2 libraries using abscisic acid driven GFP expression.

While developing the biosensor system required to link esterase activity to gIII production in PACE, we wondered if a similar system would be amenable to production of a fluorescent output (**Figure 5.3A**). The discovery that a holding tank was required to enable diffusion of the esterified ABA esters into the cells led us to believe that perhaps libraries of the BS2 esterase could be evaluated using fluorescence activated cell sorting (FACS). Given that negatively charged ligands are known to cross the cell membrane more slowly than neutral hydrophobic compounds or positively charged hydrophilic ones, the free ABA acid may potentially linger in the cell long enough to result in GFP turn-on without substantial 'cross-talk' to non-ABA producing cells. After optimization of the biosensor for GFP expression, a library of BS2 variants constructed from an NNK scan of the active site were co-transformed into cell lines compatible with the production of GFP using the split T7-RNAP described in Chapter 4. A cell sort performed after incubating cells expressing the BS2 esterase supplemented with substrate **ABA-1** and the top collecting the top 0.1% of variants from the screen were collected. These variants were then sub-cloned back into pET28 for expression and screening in lysate. Simultaneously, a library was constructed in the same way but in the absence of an activity enrichment sort to see if our FACS screen resulted in a meaningful increase of active variants.

Figure 5. 3: Cell sorting for improved esterase activity with the ABA ester



A) Gene circuit for GFP in response to ABA. 1) Esterase catalyzed hydrolysis of ABA releases unmasked ligand 2) Split RNA polymerases are constitutively expressed 3) Unmasked ABA binds to the PYL-C_{RNAP} protein construct 4) The ABA-PYL complex recruits the ABI-NRNAP complex and reconstitutes the T7-RNAP 5) The reconstituted T7-RNAP binds to the T7 promoter and drives transcription of *eGFP*. B) Workflow for FACS sort of BS2 library and results of comparison between sorted (Black) and unsorted library (Grey). The bar graph compares the activity of the top 30 variants in lysate for the hydrolysis of **ABA-1** for the unsorted and sorted libraries, values normalized to wild-type BS2. C) Purified variants obtained from sort show improved activity compared to the wild-type BS2 variant.

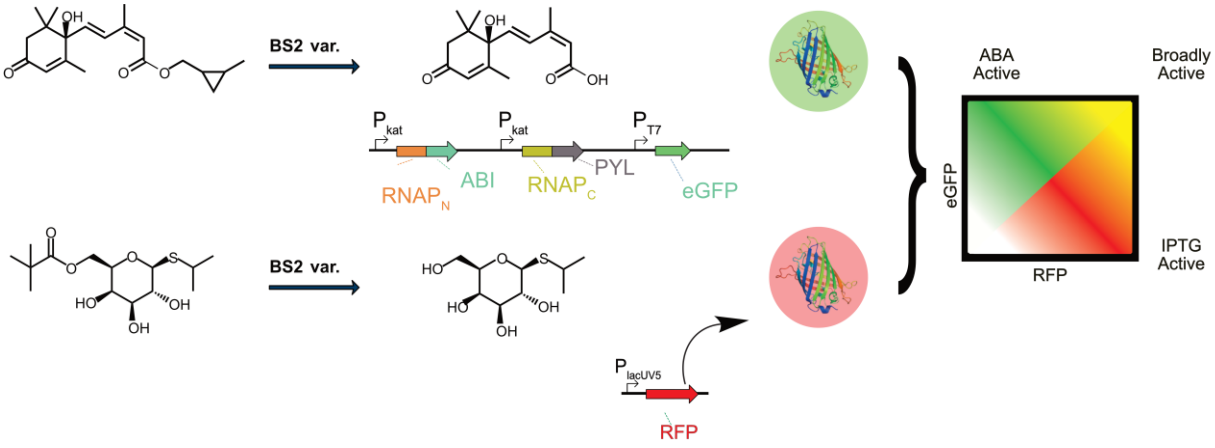
After the cell sort, 600 variants were screened both from the sorted and unsorted library. From these, 30 variants from both the sorted and unsorted plates were selected for screening in duplicate to verify the activity. From this analysis, an enrichment in active variants from the samples obtained post cell sort was clear, with the vast majority of the top 30 variants from the unsorted library exhibiting roughly equivalent or lower activity than the wild-type enzyme. In contrast, the sorted library was significantly shifted towards variants with higher activity, with nearly 90% of the top 30 variants exhibiting 1.5-fold improved activity and several up to 3-fold improved (**Figure 5.3B**). After purification

of these enzymes, we found that two variants, BS2-T and BS2-SK, exhibited drastically improved rates of hydrolysis for the sorted substrate (**Figure 5.3C**). This activity also extended to other ABA esters as well (see the experimental section and **Figure 5.9** for more details).

5.2.2 Evaluation of secondary biosensors towards enabling a dual sort

Having established that the cell sort successfully enriched the BS2 library in active variants over an unsorted library, we began to be interested in developing a second biosensor meant to recognize an orthogonal ligand. We decided to evaluate several other biosensors that could be linked to esterase activity. Including the IPTG biosensor mentioned in Chapter 4, we also decided to evaluate the *ttgR* which responds to flavanoid compounds, and the *padR* which responds to coumaric acid. Synthesis of esterified versions of these compounds was accomplished easily using well established chemistry (for more details, see section 5.4.3), and we verified that each of these masked ligands did not result in significant turn on over the background signal in plate reader assays.

Figure 5. 4: Envisioned dual sort and results of IPTG diffusion study and coumarate ester toxicity

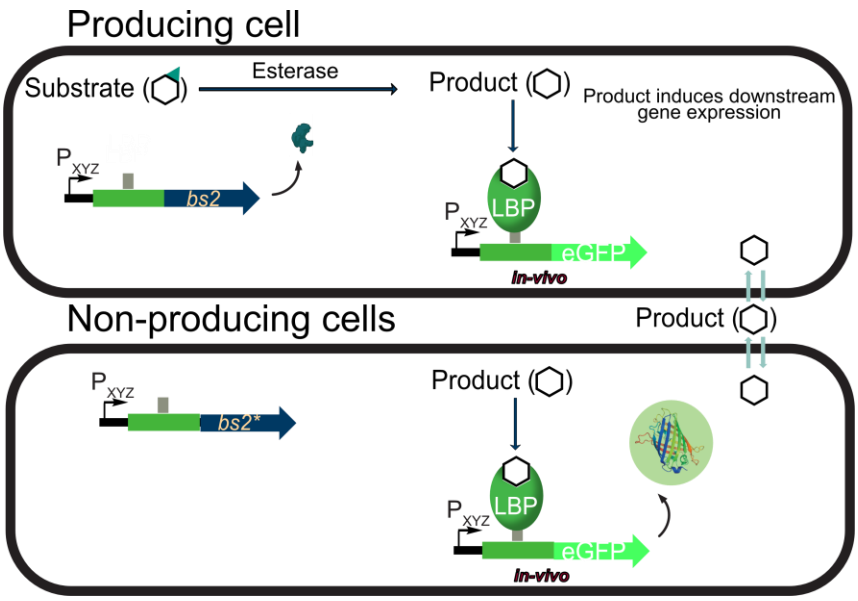


By connecting the hydrolysis of two esters to production of independent fluorescent proteins, a single sort could enable the identification of selective and broadly active esterase variants.

Unfortunately, for each of these systems we found problems that were not easily surmountable. A cell sort of BS2 variants using esterified versions of the IPTG ester found that after collecting variants from the top 0.1% of the library these improvements did not translate to *in vitro* screening. An

experiment designed to probe diffusion of unmasked IPTG from BS2 producers to non-producers showed that released IPTG was rapidly diffusing out of the cells and activating *egfp* (**Figure 5.5**). Further literature analysis showed that this diffusion occurs even at low concentrations of IPTG which would likely prevent IPTG from being used in a cell sort without manipulation of the medium such as an oil-water emulsion. For the *ttgR* biosensor the esterified flavonoid compounds were largely insoluble in the aqueous medium which could complicate dosing of the cell sort or analysis of the cells after reaction. Additionally, literature analysis shows that these flavonoid compounds rapidly diffuse through the *E. coli* cell membrane due to their small, hydrophobic nature, which would likely prevent cell sorting with this biosensor. While we had hoped the *padR* ligands would behave similar to ABA as negatively charged species and self-sequester BS2 catalyzed hydrolysis, the higher concentrations of esterified *p*-coumarate esters necessary to turn on the biosensor resulted in lower OD_{600} values for cultured *E. coli* cells, likely a result of the esters acting as (toxic) Michael acceptors. In total, the work done in this study to optimize the biosensors may enable these systems to be used for *in vitro* studies as will be described in the next section but are unlikely to be easily utilized in FACS.

Figure 5. 5: Diffusion of the small molecule results in fluorescent signal from non-producing strains



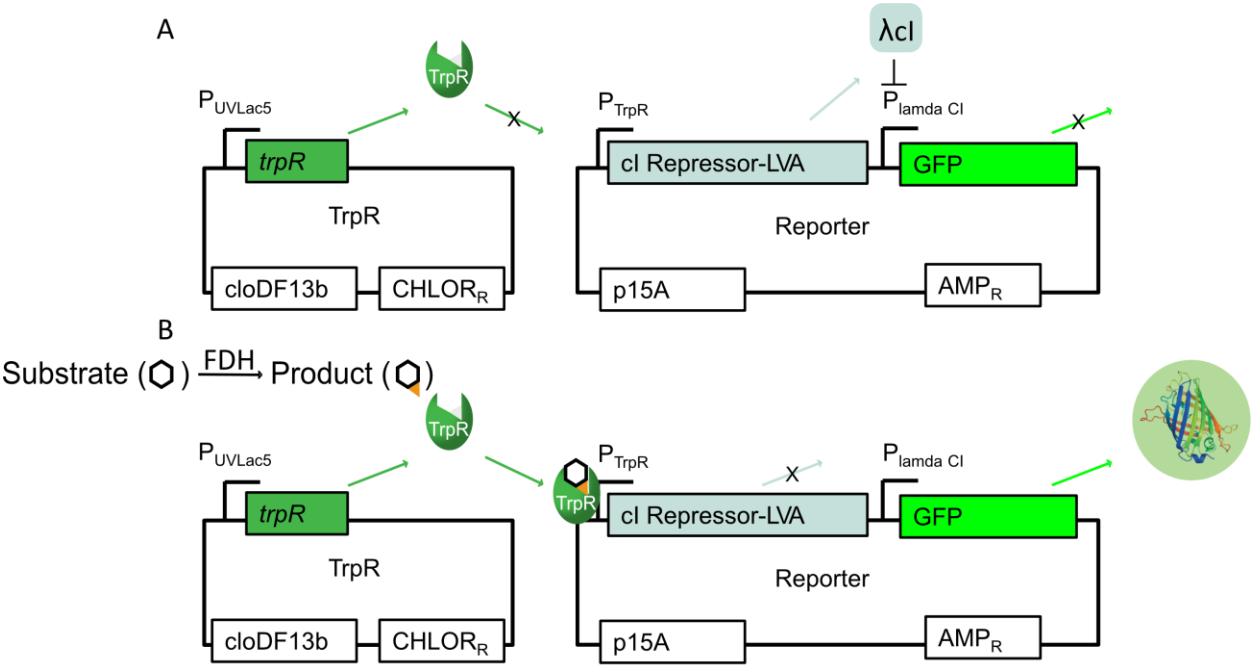
For molecules like IPTG and flavanone derivatives, rapid diffusion through the membrane of *E. coli* results in fluorescent signal in variants that are non-producers, resulting in mostly false positives from attempts at cell sorting.

5.2.3 Evaluation and optimization of a biosensor for brominated tryptophan compounds

Around the time we began to explore cell sorting as a potential method for evaluating large enzyme libraries, work from Prof. Ellington's group described the use of an intricate selection for the reprogramming of the ligand specificity for the tryptophan repressor encoded by the gene *trpR*. By utilizing alternating rounds of positive and negative selection selection based on the thermostability of an RNA polymerase, trpR was engineered to specifically recognize the 5- and 6-brominated analogs with no significant response to the native ligand L-tryptophan. By using the TrpR protein to regulate expression of a secondary repressor controlling GFP expression, the native repressive activity of the TrpR transcription factor was inverted to produce a gene circuit that resulted in turn-on in response to effector ligand rather than the native turn-off (**Figure 5.6**). In plate reader assays, they reported 30-fold fluorescent turn on when cells expressing the biosensor were supplied with the proper effector ligand and little to no cross-reactivity with off-target brominated isomers or the non-halogenated compound was observed. We envisioned that by co-transforming a flavin-dependent halogenase into the cell line

used for the development of the *trpR* biosensor would allow this system to be used for cell sorting like the ABA system.

Figure 5. 6: Depiction of tryptophan biosensor



A) In situations with low levels of 6-bromotryptophan, the TrpR repressor protein is not able to bind to the target operator sequence, and expression of the λcI repressor protein inhibits transcription of *egfp*. B) With high levels of 6-bromotryptophan, the TrpR repressor protein binds to the operator sequence and prevents expression of the λcI repressor protein. This derepresses the expression of *egfp* and allows for production of the fluorescent protein.

One drawback to utilizing the FDHs for *in vivo* biocatalysis is the requirement for the flavin reductase partner protein. While we considered the possibility of co-expressing the reductase protein independently, we held concerns that the burden of protein expression for the *E. coli* cell would hamper GFP expression. Instead, we envisioned utilizing a fusion halogenase protein, previously explored in our lab and noted for improvement of *in vivo* biocatalysis. We decided to generate fusions of the 5-tryptophan halogenase PyrH and the 6-halogenase Thal. Both of these enzymes have been used as biocatalysts in previous studies, they are capable of both bromination and chlorination, and they can be solubly expressed in *E. coli*. Fusions of these genes were ligated to the flavin reductase *rebF* via Gibson

assembly, and a flexible G4S linker was included between the halogenase and reductase since this was shown to improve activity over the previously designed linker for the RebH construct we originally evaluated.[16] *In vivo* assays with the fusion enzymes expressed in *E. coli* showed that the PyrH fusion protein exhibited no detectable activity via UHPLC and the Thal-RebF enzyme showed low turnover to the 6-brominated product (see **Figure 5.11** in the additional figures section). We decided to pursue this enzyme for the *in vivo* experiments necessary for cell sorting with the engineered tryptophan biosensor.

A key feature of a cell sort with the engineered tryptophan biosensor that must be taken into consideration is the requirement for the deletion of the chromosomal copy of the tryptophan repressor protein. If the native repressor is present, it outcompetes the engineered repressor for the native operator sequence. While alternative operator sequences could be considered, the developed orthogonal sequences in the Ellington paper resulted in lower turn on than utilization of the native trpR operator sequence. Given the low activity expected from our initial variants, our concerns that further loss of signal from a sub-optimal operator sequence would provide insufficient signal for reliable detection led us to use cells containing the Δ*trpR* genotype (cell line JW4356 purchased from the Coli Genetic Stock Center). A second consideration was the construction of the plasmid constructs used for the cell sort. The Δ*trpR* cell line was modified to include a kanamycin selection marker in the place of the *trpR* gene. Examination of the biosensor plasmid showed that the chloramphenicol resistant plasmid bearing the PBR322 origin showed the highest overall signal when supplied the 6-bromotryptophan effector ligand, leading us to select the spectinomycin resistant ColE1 origin for the plasmid bearing the Thal-RebF construct. A final ampicillin resistant plasmid was also required by the system for supplying the inverter circuit to drive *egfp* expression from the TrpR repressor protein. This three-plasmid system utilized the four commonly applied antibiotics and prevented us from including the pGro7 chaperone plasmid which was required for high levels of halogenase expression and *in vivo* halogenase activity in BL21 cells. A combined construct including the halogenase and the *trpR* gene was also evaluated to

provide a free antibiotic for the chaperone plasmid, but it showed no signal over background when supplied with the effector ligand (see experimental section **Figure 5.14**).

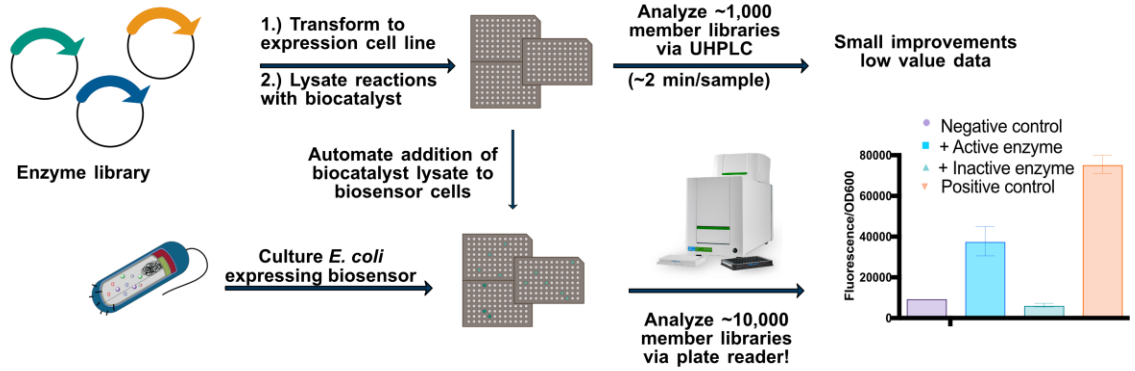
In addition to requiring modification of the biosensor system for use in *in vivo* cell sorting, the expression of the Thal-RebF construct also required significant modification and optimization. Our typical systems are designed for expression in the BL21(λDE3) cell line, which includes a chromosomal copy of the T7 RNA polymerase for expressing proteins from the T7 promoter. This promoter is orthogonal to *E. coli* function as this polymerase is not natively found in *E. coli*. Unfortunately, this polymerase was not present in the Δ*trpR* cell line, and while we had successfully cloned a copy of the T7 RNAP into the Δ*trpR* cell line, these cells exhibited significantly lower signal/background ratio when supplied with the 6-bromotryptophan effector ligand (see additional figures, **Figure 5.12**). Instead, we focused on modifying the promoters used for expressing the Thal-RebF construct. Three promoters were considered: the T5 and tac promoters derived from phage and the LacUV5 promoter, all of which are regulated by the LacI repressor protein and inducible with IPTG.[17] Of these, we selected the T5 promoter due to known leaky expression from the LacUV5 and tac promoters, which we envisioned would be detrimental to cell health. Evaluating the expression and activity of the Thal-RebF fusion revealed that the T5 promoter showed activity for *in vivo* halogenation of L-tryptophan when expressed from BL21(λDE3) cells, but only when coexpressed with the chaperone plasmid (see additional figures, **Figure 5.13**) which was incompatible with the other plasmids in our system due to both overlapping antibiotic resistances and plasmid origins. Given the lack of meaningful progress made in these efforts for establishing an *in vivo* system to enable cell sorting, we decided to transition utilizing this system for *in vitro* studies instead.

5.2.4 Early efforts to establishing an *in vitro* system for evaluating enzyme libraries

While the cell sort of the BS2 library on the ABA ester demonstrated how connecting a biosensor system to enzyme catalysis *in vivo* can dramatically improve the ability to screen large

libraries, the difficulties we encountered in generalizing this system to other biosensors led us to consider alternative approaches by which biosensors could simplify high throughput screening. While plate reader analysis lacks the ability to sort through a library as large as 10^6, fluorescent readouts in a plate reader format have been used to evaluate libraries as large as 10^5 and can be integrated into automation platforms to greatly reduce the amount of time required for researcher intervention.[18] Additionally, this approach obviates the need for the biocatalyst of interest to have a generalizable expression system. By expressing the biocatalyst in an optimized cell line or with plasmids to aid expression, we envisioned this approach could extend beyond well-expressed enzymes.

Figure 5. 7: Early efforts of using the L-tryptophan biosensor for halogenase analysis



Workflow for analyzing Thal halogenase libraries showing results of preliminary screening of Thal (blue, + active enzyme), no biosensor turn on (purple, negative control), 2-C01 (green, + Inactive enzyme), and with the addition of 6-bromotryptophan (orange, positive control).

To demonstrate this approach, we decided to pursue the *in vitro* analysis of the 6-tryptophan halogenase Thal. This enzyme was chosen due to the high signal obtained for the 6-bromotryptophan effector ligand in the plate screening assay and because this enzyme was found to be better expressed than PyrH in our previous expression and activity studies. We envisioned that by running these halogenase reactions at high substrate loadings and appropriately diluting these samples into the cell line expressing the biosensor, reactions could be screened well beyond the saturation point of the biosensor to screen for halogenases that operate a high substrate concentrations. To this end, we

screened reactions of the Thal halogenase using 10 mM of substrate and added the completed reactions to cells expressing a biosensor for 6-bromotryptophan. We then compared the activity of Thal to that of the halogenase 2-C01, which is known to not functionalize tryptophan, and to a control reaction with no halogenase expression. From this initial study, we observed that only the cells expressing the Thal halogenase exhibited significant turn on, and that the total turn on was roughly ~45% of the positive control where the cells were provided with the 6-bromotryptophan effector ligand (**Figure 5.7**). This result suggests that active variants can be picked out from inactive samples and that variants with up to 2-fold higher activity relative to the wild-type enzyme could be identified.

We envision that with the drop in cost associated with construction of well-defined NNK libraries and emerging methods for DNA barcoding or library sequencing approaches to make sequencing large libraries affordable, such a biosensor system may enable the characterization of the enzyme-sequence relationship for Thal. This ability to screen libraries over 10x larger than previous efforts for an enzyme class with untapped industry potential that cannot be directly correlated to a fluorescent product would show the utility of biosensor driven enzyme evolution. Coupled to the automation system available to us, this may provide the ability to screen multiple thousands of variants in each round of evolution. While the limitation of detecting only 5- or 6-bromotryptophan may not provide enzymes with incredibly broad scope, the insight gained to halogenase function from such a deep mutational scan may provide information that extends to other halogenases.

Beyond analyzing halogenase libraries, we also considered the potential for the plate reader approach to evaluate other enzymatic reactions as well. While esterase-catalyzed turn on of the *ttgR* biosensor system was not amenable for *in vivo* screening, the high activity we observed with this system in the plate reader assays means it could be used for similar *in vitro* efforts as described for the Thal halogenase. One noteworthy property of the *ttgR* biosensor is the broad range of ligands that it

recognizes. We found that various mono, di, and tri-hydroxylated flavanone compounds, for example, provided significant turn-on over background. We envisioned that this biosensor could be used to identify or develop a P450 enzyme for aromatic hydroxylation of the flavanone compound. We identified two P450 enzymes through bioconversions in lysate that produced the same hydroxylated product as determined by retention times on UHPLC-MS. Unfortunately however, the product of these reactions was identified as the aliphatic hydroxylation product after isolating material from bioconversions (see additional data **Figure 5.10**). While this product possibly turns on the biosensor this has not been explored. Future efforts utilizing this biosensor could evaluate an NNK library of the active site of these P450s to perhaps improve this hydroxylation activity if it does respond to the biosensor and reposition the substrate in the active site of these enzymes to enable aromatic hydroxylation. The broad specificity of the ttgR biosensor means that any hydroxylation of the aromatic rings is likely to be captured by the biosensor and could open the door to variants capable of producing new regioisomers.

5.3 Conclusions

Biosensor driven enzyme evolution is still a field that requires further development. The progress made on several projects through my efforts here have shown how these biosensors can be used directly for the enrichment of active enzymes. Even in cases where the direct application of a biosensor for screening enzyme activity is not possible due to the limitations of the substrate, native and engineered biosensors can be amended to plate reader assays that may allow for deep mutational scanning.

5.4 Experimental

5.4.1 Materials, instruments, and methods

Greiner Bio-One polypropylene 96-well plates (product number 651201) or Fisherbrand 96-well deepwell polypropylene plates (product number 12-566-121) were purchased from Fisher. Corning-costar 96-well black wall, clear bottom plates (product number 3631) were purchased from Fisher. Antibiotics were purchased from Gold Biotechnology and were prepared as 1000x stock solutions. Sodium acrylate was

purchased from Combi-Blocks (product number QC-1489-005). IPTG was purchased from Gold Biotechnology (product number I2481). Abscisic acid was purchased from Biosynth International (product number A-0120). Flavanone, p-coumaric acid and other reagents used for chemical synthesis were purchased from Sigma Aldrich and used without further purification. Stock solutions (30 mM) of ligand were prepared at 50 mM DMSO and stored at -20 °C until use. The JW4356 cell line contains the *kanR* gene in place of the *trpR* gene and is referred to as the Δ*trpR* cell line in the text. Taq DNA polymerase was purchased from New England Biolabs and PRIMESTAR from Takara. The genes for ABI and PYL were gifts from Professor Fu-Sen Liang. *In vivo* biocatalysis with Thal and PyrH fusion proteins was conducted according to previous literature.

Luria broth (LB) media and Super Optimal broth with Catabolite repression (SOC) media were purchased from Research Products International and Davis rich media for PACE or PACS was prepared as previously described[2]. *E. coli* 10β were purchased from New England Biolabs and *E. coli* S1030 cells[2] were courtesy of the Liu lab, Harvard University. Restriction enzymes, Antarctic Phosphatase, and T4 DNA ligase were purchased from New England Biolabs. DNA clean & concentrator kits were purchased from Zymo Research (product number D4013).

Fluorescence and optical density measurements were performed using a Synergy Neo2 Hybrid Multi-Mode Reader (BioTek). For UHPLC-MS analysis, 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 was used for both low-throughput analysis and the qualitative MISER screen. For UHPLC analysis, a similar Agilent system was used, except the autosampler used was single-needle configuration. For both UHPLC-MS and UHPLC, low-throughput analysis was performed on an Eclipse Plus C18 2.1x5mm guard column

with a 1.8 μm pore size (part number 821725-901) connected to a ZORBAX rapid resolution C18 column (part number 959757-902). MISER analysis was done using only the guard column.

5.4.2 Cloning, molecular biology, and enzyme preparation

Cloning and sequencing.

All plasmids were constructed by Gibson assembly from PCR products generated using Primestar 2x Mix. PCR reactions and thermocycling conditions were performed according to NEB's recommended protocols. Annealing temperatures were determined with the NEB T_m Calculator. Lysogenization of the Δ*trpR* cell line with the T7 RNAP was performed using the λDE3 lysogenization kit according to the manufacturers protocol.

Plate reader assays for optimization of the *trpR* biosensor

Cells containing the *trpR* biosensor and the relevant components of the biosensor were re-streaked onto LB Agar containing the appropriate antibiotics fresh each day before beginning the assay process. From the freshly re-streaked plates, single colonies were grown to saturation overnight at 37 °C in a 96-well deep well plate containing 0.3 mL of LB with antibiotics. This overnight culture was then used to inoculate 1 mL of 2-XYT media containing the appropriate antibiotics and the cells were allowed to grow for 3 hours at 37 °C. After growth with shaking for 3 h at 37 °C, gene expression was induced with IPTG (1 mM final concentration) and either 5- or 6-bromotryptophan (1 mM final concentration). After induction, the plate was replaced into the incubator and fluorophore expression and maturation was allowed to continue for either 6 or 20 hours. Following this period, a 100 μL aliquot of the samples were taken from the deep well plate and added to a black-walled microtiter plate, at which point fluorescence and OD_{600} were measured on a Synergy Neo2 Hybrid Multi-Mode Reader (BioTek). The data were analyzed by dividing the fluorescence values by the background-corrected OD_{600} value. All values were then normalized to the emission of cells expressing each biosensor without small molecule, which was assigned an arbitrary value

of 1 to allow for values from each fluorescence plot to be compared to each other. All experiments were performed in three replicates.

Enzyme expression and purification

For esterase variants, overnight cultures were prepared in 10 mL LB with 50 ng/µL kannamycin. 5 mL of this overnight culture was used to innoculate 250 mL of TB + 50 ng/µL kannamycin and the cells were grown at 37 °C shaking at 250 rpm until the OD600 reached 0.8, at which point the temperature was reduced to 30 °C and protein expression was induced with 25 µL of a 1 M stock solution of IPTG for a final concentration of 100 µM. These cultures were allowed to continue incubating for 20 hours at 250 rpm, at which point the cells were transferred to cell buckets and centrifuged at 3600 rpm, 4 °C for 20 minutes. The supernatent was removed and the cell pellets were stored at –20 °C until ready for purification. To begin purifying the protein, each cell bucket was resuspended in 40 mL of phopshate buffer (50 mM, pH 7.5), and the resuspended cell pellets were transferred to a 50 mL centrifuge tube. These were immediately placed on ice, and the cells were lysed by sonication 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. The cells were then centrifuged in a high-speed rotor at 15,000 rpm for 30 minutes at 4 °C. The clarified lysate was then purified via IMAC using Nickel-NTA resin. Buffer exchanging was done by dialyzing the eluted protein sample in phosphate buffer (50 mM, pH 7.5) overnight. The next day, the samples were concentrated in a Amicon ultracentrifugation column (10 kDa MW cutoff) and the buffer was diluted with 10 % glycerol to act as a cryoprotectant. Protein concentration was determined by the absorbance at 280 nm with the extinction coefficient calculated from the protein sequence using the Expasy Protein Parameters website. Purified enzymes were stored at –80 °C until used.

The process for the purification of P450 variants was largely identical up to the purification process, except ampicillin was used as the selection antibiotic for all cell cultures, and gene expression was induced with

250 µL of a 1 M stock solution of IPTG for a final concentration of 1 mM, as well as 2.5 mL stock of 50 mM 5-ALA for a final concentration of 500 µM. After cell lysis performed identically to previously described, the enzyme was purified by anion exchange according to literature protocol. As only a preliminary evaluation of the enzyme product was required, no concentration of the purified P450 sample was measured prior to use in bioconversions.

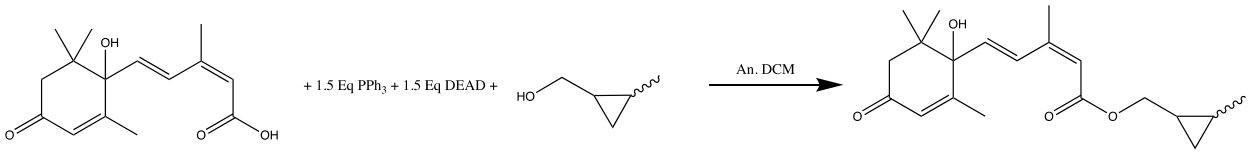
Table 5. 1: List of Primers used by me for this study:

Primer name	**Sequence**	**Use**	**Direction**
HS039	CGCACCAAGTCGGGA	Thal middle portion sequencing primer	F
HS040	GCCTGGTGCCGCGCGGCAGCCATATGGACAATCGAATCAAGACAGTCG	Thal fusion insert primer	F
HS041	CGACTGTCTTGATTCGATTGTCCATATGGCTGCCGCG	Thal fusion vector primer	R
HS042	AGACCCACCGCCACCTGAGCCACCACCGCCCGACGCACCGTGCAA	thal fusion insert primer	R
HS043	GACCTGCTGCGTCAGTTGCACGGTGCGTCGGGCGGTGGTGGCTCA	thal fusion vector primer	F
HS165	ATGATGGCTGCTGCCCATagttaatttctcctctttaatgaattc	vector primer for Thal-F into PT5	R
HS166	GGACACCGGAGGGATAAagcttaattaatgactgagcttgg	vector primer for thal-f into Pt5	F
HS167	ccaagctcagtcattaattaagctTTATCCCTCCGGTGTCC	insert primer for thal-f into pt5	R
HS168	tcattaaagaggagaaattaactATGGGCAGCAGCCATCATCATC	insert primer for thal-f into pt5	F
HS189	gccacctgacgtctaaga	PT5 sequencing primer	F

5.4.3 - Synthesis

For the synthesis of all IPTG esters, see chapter 4 section 4.4.3. All ABA esters discussed in this section except the 2-methylcyclopropyl substrate were reported in chapter 4. Esters of p-coumarate were either purchased and used without purification or were synthesized as previously reported.[19]

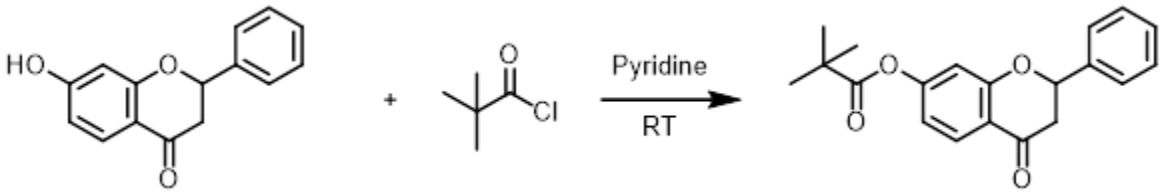
Scheme 5. 1: 2-methylcyclopropyl-ABA (ABA-1):



To a stirring solution of ABA (150 mg, 0.568 mmol) in 3mL anhydrous DCM under nitrogen, triphenylphosphine (223 mg, 0.852 mmol) was added as a solid, followed by the alcohol (170 µL, 1.758 mmol). DEAD (334 µL of 40% solution in toluene, 0.852 mmol) was added dropwise over five minutes, and the solution stirred overnight. The next day, the solution was concentrated under vacuum, then taken into EtOAc, and washed with 0.5M HCl, brine, then dried over $MgSO_4$. This was then concentrated under vacuum again, and the solution was purified via silica chromatography (4:1 Hex:EtOAc), final yield 91%. ^{1}H NMR (400 MHz, Methylene Chloride-d_2) δ 7.84 (d, *J* = 16.1 Hz, 1H), 6.16 (d, *J* = 16.2 Hz, 1H), 5.88 (s, 1H), 5.76 (s, 1H), 5.31 (s, 1H), 4.13 – 3.71 (m, 2H), 2.47 (d, *J* = 17.0 Hz, 1H), 2.22 (d, *J* = 17.0 Hz, 1H), 2.00 (s, 3H), 1.89 (s, 3H), 1.08 (s, 3H), 1.04 (dd, *J* = 6.1, 1.8 Hz, 3H), 1.00 (s, 3H), 0.84 (dd, *J* = 8.0, 3.2 Hz, 1H), 0.69 (t, *J* = 7.0 Hz, 1H), 0.48 – 0.41 (m, 1H), 0.30 (dd, *J* = 5.6, 3.0 Hz, 1H). ^{13}C NMR (126 MHz, Methylene Chloride-d_2) δ 196.69, 165.33, 161.70, 148.51, 135.63, 127.25, 126.10, 117.71, 78.89, 67.49, 48.91, 40.66, 23.28, 22.08, 20.13, 17.92, 17.49, 17.30, 10.81, 10.56.

Flavanone preparation:

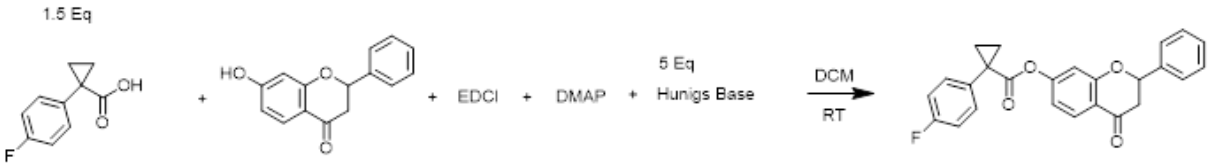
Scheme 5. 2: Pivoyl-Flavanone:



To a stirring solution of 7-hydroxyflavanone (200 mg, 0.832 mmol) in 5 mL of pyridine was added trimethylacetyl chloride (125 µL, 1 mmol) dropwise. This solution was allowed to stir overnight at room temperature. The following day, the solution was diluted into DCM and washed 3x with 10% $CuSO_4$, 3x with saturated EDTA, once with brine, and then dried over $MgSO_4$ and concentrated. The resulting oil was dissolved in DCM and then purified by silica chromatography, 7:1 hexanes : ethyl acetate. Obtained clear oil, 257.1 mg, 95% yield.

^{1}H NMR (400 MHz, Methylene Chloride-d_2) δ 7.84 (d, *J* = 8.5 Hz, 1H), 7.49 – 7.27 (m, 5H), 6.81 – 6.61 (m, 2H), 5.45 (dd, *J* = 13.2, 2.9 Hz, 1H), 3.01 (dd, *J* = 16.9, 13.2 Hz, 1H), 2.80 (dd, *J* = 16.9, 3.0 Hz, 1H), 1.26 (s, 9H).

Scheme 5. 3: 1-(4-fluorophenyl)flavanone ester:

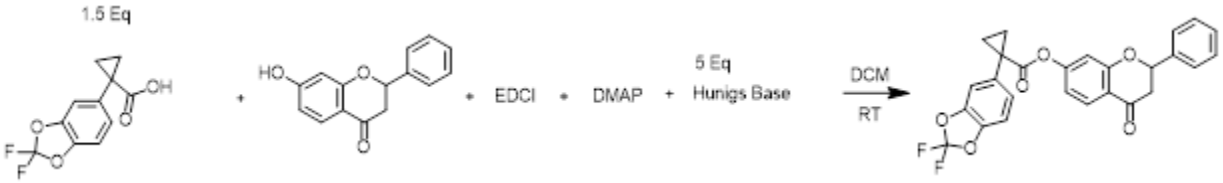


To a stirring solution of 1-(4-fluorophenyl)cyclopropanecarboxylic acid (1.248 mmol, 225 mg) in 2 mL DCM was added Hunigs base (4.16 mmol, 742 µL), followed by EDCI (HCl salt) (0.832 mmol, 160 mg), then DMAP (0.832 mmol, 101 mg). This was allowed to stir for ten minutes prior to the addition of 7-hydroxyflavanone (0.832 mmol, 200 mg). This solution stirred overnight. The following day, the solution was diluted with DCM, then washed once with 10% citric acid, once with saturated bicarbonate, once

with brine then dried over $MgSO_4$ and concentrated under vacuum. The resulting oil was dissolved in DCM then purified by silica chromatography, 7:1 hexane : ethyl acetate. Obtained white solid, 180.2 mg, 45% yield.

^{1}H NMR (400 MHz, Methylene Chloride-d_2) δ 7.80 (d, *J* = 8.6 Hz, 1H), 7.49 – 7.22 (m, 7H), 6.96 (t, *J* = 8.7 Hz, 2H), 6.72 – 6.61 (m, 2H), 5.43 (dd, *J* = 13.1, 3.0 Hz, 1H), 2.98 (dd, *J* = 16.9, 13.2 Hz, 1H), 2.78 (dd, *J* = 16.9, 3.0 Hz, 1H), 1.70 (q, *J* = 4.1 Hz, 2H), 1.28 (t, *J* = 3.6 Hz, 2H). ^{13}C NMR (101 MHz, Methylene Chloride-d_2) δ 190.41, 172.20, 163.22, 162.27, 160.78, 156.83, 138.58, 134.66 (d, *J* = 3.3 Hz), 132.33, 132.25, 128.71, 128.06, 126.14, 118.75, 115.41, 115.11, 114.90, 110.91, 79.97, 44.22, 28.40, 17.52.

Scheme 5. 4: 1-(2,2-difluorobenzodioxyl)flavanone ester:



To a stirring solution 1-(2,2-difluorobenzo[d][1,3]dioxol-5-yl)cyclopropane-1-carboxylic acid (1.25 mmol, 201 mg) in 2 mL DCM was added Hunigs base (4.16 mmol, 742 µL), followed by EDCI (HCl salt) (0.832 mmol, 160 mg), then DMAP (0.832 mmol, 101 mg). This was allowed to stir for ten minutes prior to the addition of 7-hydroxyflavanone (0.832 mmol, 200 mg). This solution stirred overnight. The following day, the solution was diluted with DCM, then washed once with 10% citric acid, once with saturated bicarbonate, once with brine then dried over $MgSO_4$ and concentrated under vacuum. The resulting oil was dissolved in DCM then purified by silica chromatography, 10:1 hexane : ethyl acetate. Obtained white solid, 102 mg, 26% yield.

^{1}H NMR (400 MHz, Methylene Chloride-d_2) δ 7.81 (d, *J* = 8.5 Hz, 1H), 7.43 – 7.27 (m, 5H), 7.16 – 7.07 (m, 2H), 6.98 (d, *J* = 8.1 Hz, 1H), 6.72 – 6.63 (m, 2H), 5.43 (dd, *J* = 13.2, 3.0 Hz, 1H), 2.99 (dd, *J* = 16.9, 13.2 Hz, 1H), 2.78 (dd, *J* = 16.9, 3.0 Hz, 1H), 1.72 (q, *J* = 4.2 Hz, 2H), 1.29 (q, *J* = 4.2 Hz, 2H). ^{13}C NMR (126 MHz, Methylene Chloride-d_2) δ 190.40, 171.87, 162.30, 156.73, 143.46, 142.99, 138.59, 135.05, 128.74, 128.11, 126.16, 126.03, 118.85, 115.38, 112.15, 110.92, 109.08, 80.02, 44.25, 28.99, 17.77.

5.4.4 Biocatalysis

Esterase library expression and screening:

BS2 and the BS2 libraries were expressed from BL21(DE3) from pET28(a). Libraries were picked by an automated colony picker and transferred to 96-well 1 mL deep well plates with 300 µL of LB supplemented with 50 µg/mL kanamycin. These cultures were incubated for 12-16 hours shaking at 250 rpm at 37 °C. 30 µL of overnight culture was then used to inoculate 1 mL of TB supplemented with 50 µg/mL kanamycin in a 96-well 2 mL deep well plate. These cultures were grown while shaking at 250 rpm and heated to 37 °C until the OD_{600} = 0.6-0.8, at which point the cultures were induced by 10 µL of a 10 mM stock solution of IPTG, for a final induction concentration of 100 µM. The temperature of the incubator was reduced to 30 °C, at which point the cultures were incubated for 20 hours at 250 rpm, after which the cultures were harvested by centrifugation at 3600 rpm for 15 minutes and stored in a -80 °C freezer until use.

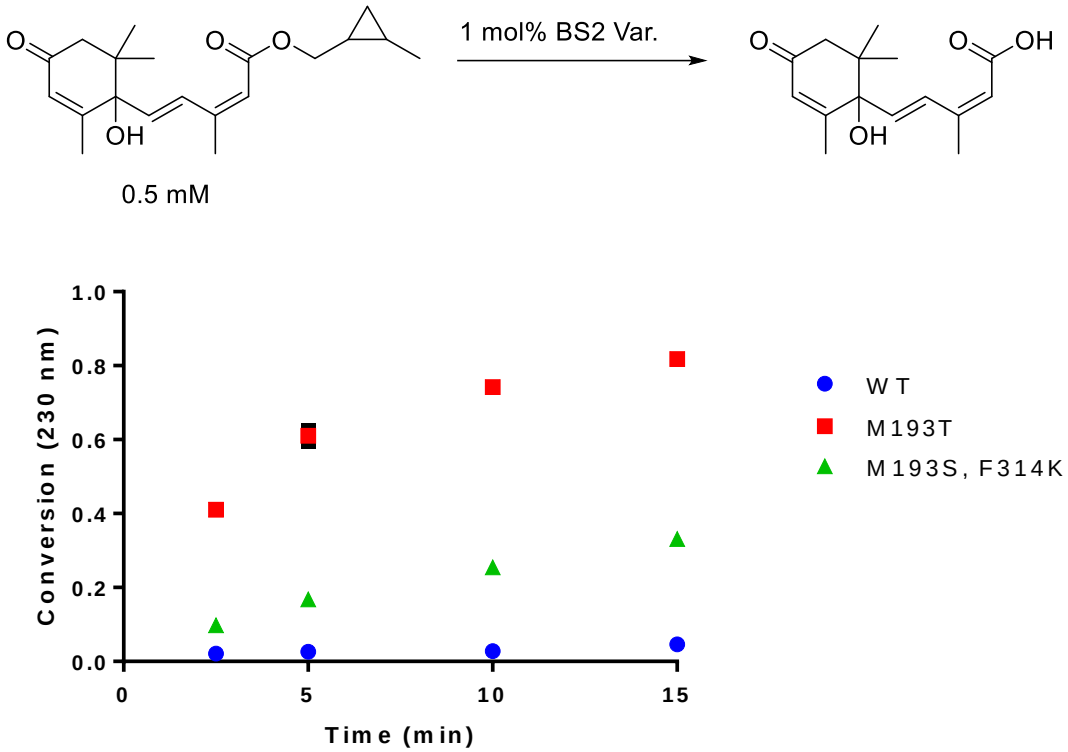
For screening of libraries, cells were thawed and suspended in 100 µL HEPES buffer (25mM, pH 7.4) containing 0.75 mg/mL lysozyme. After resuspension by gentle vortexing, the cells incubated in the lysis buffer for 45 minutes at 37 °C and 250 rpm. The 96-well plate was then flash frozen in liquid nitrogen, and then allowed to thaw, and 10 µL of a 1 mg/mL solution of DNAse in HEPES buffer (25mM, pH 7.4) was added to the wells. This was allowed to incubate for 15 minutes at 37 °C and 250 rpm, following which the lysate was clarified by centrifugation at 3600 rpm for 15 minutes, and the lysate was transferred to a microtiter plate for screening. For initial library screening, 70 µL of lysate was added to the microtiter plate, followed by 5 µL of substrate (11.25 mM stock solution, 0.75 mM final concentration) and for follow up screening of hits from the library, 25 µL of lysate was transferred followed by 45 µL of buffer and 5 µL of substrate (11.25 mM stock solution, 0.75 mM final concentration). After 90 minutes, the reactions were quenched via addition of one volume of methanol. Precipitated protein was sedimented via centrifugation at 3600 rpm for 15 minutes at 4 °C. The

reactions were then diluted with 30 µL of lysate and 150 µL of water, then analyzed by LC-MS. UV-absorbance at 230 nm was used to provide a qualitative measure of conversion. Data processing was done using the ChemStation software and Microsoft Excel, and the data was visualized using GraphPad Prism. Results for the top 30 variants normalized to wild-type BS2 are available in **Figure 5.3**.

Evaluation of purified variants from sorted library

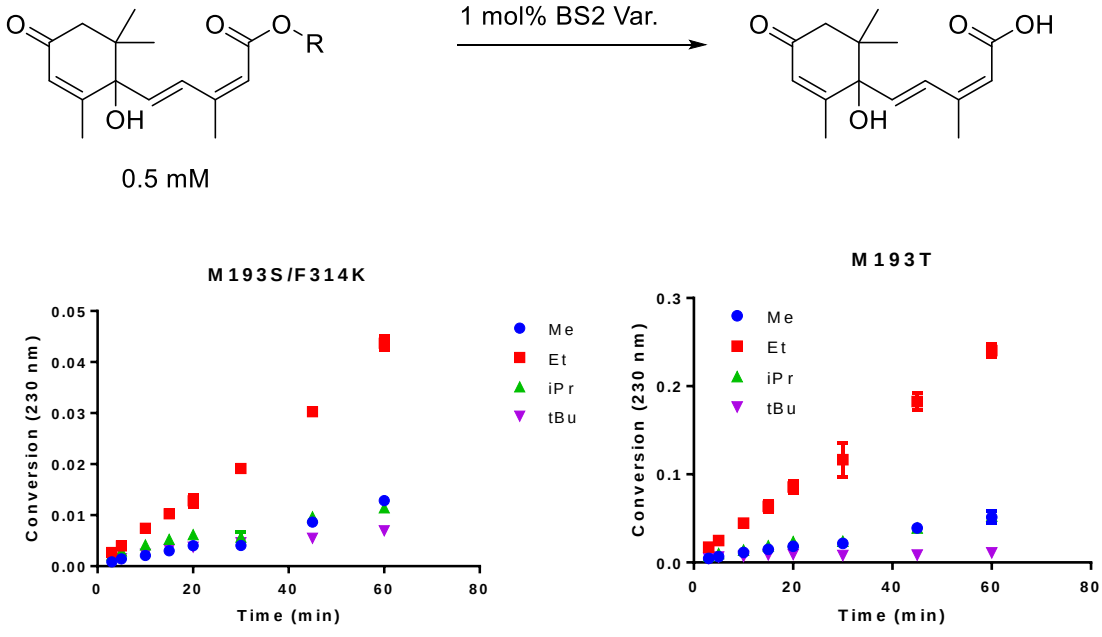
Wild-type BS2 and the M193T and M193S-F314K variants were expressed and purified according to the general protocol. Stock solutions of the BS2 esterase variants at 100 µM were prepared and 5 µL of this solution was added to microtiter plate, followed by 90 µL of phosphate buffer (pH 7.5, 50 mM), and reactions were initiated by the addition of 5 µL substrate (20 mM stock concentration, 500 µM final concentration). Reactions were allowed to continue without shaking for 1 hour, and time points were taken at the appropriate intervals by the addition of 1 volume of methanol to the microtiter plate. After the final timepoint was quenched, precipitated protein was sedimented via centrifugation at 3600 rpm for 15 minutes at 4 °C. The reactions were then diluted with 30 µL of lysate and 150 µL of water, then analyzed by LC-MS. UV-absorbance at 230 nm was used to provide a qualitative measure of conversion. Data processing was done using the ChemStation software and Microsoft Excel, and the data was visualized using GraphPad Prism.

Figure 5. 8: Bioconversions with esterases identified from cell sorting



Each point represents the average of 3 replicates. Only the first 4 timepoints are shown.

Figure 5. 9: Variants BS2-T and BS2-SK show improved activity for other ABA esters

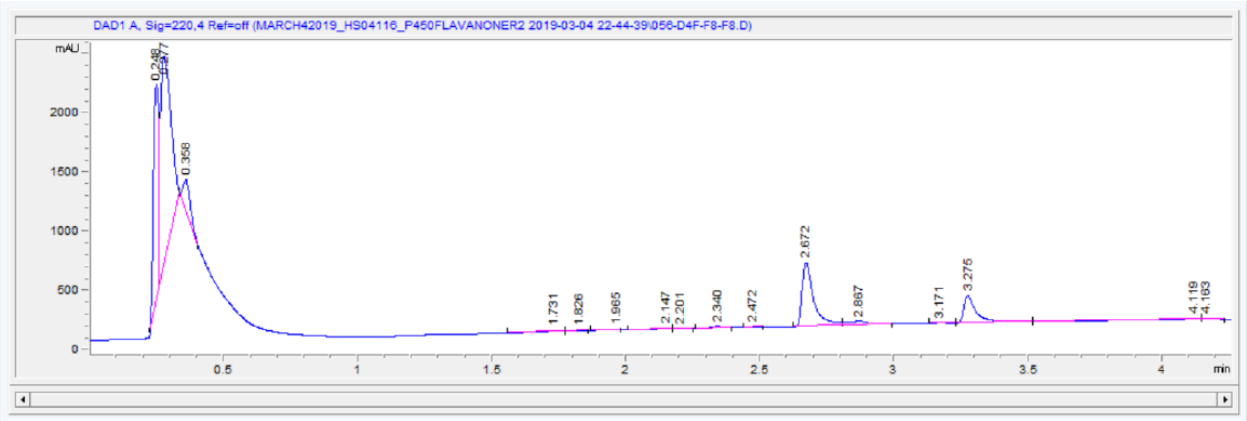


P450 library expression and screening:

The P450 libraries were obtained from the Francis Arnold lab at Caltech as glycerol stocks in BL21(DE3) from pCWori, and contained mutants of the *B. megaterium* P450 BM3. The library was grown by stamping the glycerol plate into 300 µL of LB supplemented with 100 µg/mL of ampicillin. These cultures were incubated for 12-16 hours shaking at 250 rpm at 37 ^{o}C. 30 µL of overnight culture was then used to inoculate 1 mL of TB supplemented with 100 µg/mL of ampicillin, 1 µL/mL of trace metal mix, and 20 mg/L of 1-aminolevulinic acid. These cultures were incubated four hours at 37 ^{o}C, then the cultures were removed from the incubator and the temperature was reduced to 25 ^{o}C. P450 expression was induced by addition of 50 µL of a 4.5 mM stock solution of IPTG and the cultures were allowed to express for 24 hours at 25 ^{o}C, 250 rpm, at which point the cultures were pelleted by centrifugation at 3600 rpm for 15 minutes. The cells were placed in a -80 ^{o}C freezer until use.

For cell lysis, plates were allowed to thaw at room temperature followed by the addition of 275 µL lysis buffer (0.1 M, pH = 8.0 phosphate buffer, 0.75 mg/mL lysozyme, 0.1 mg/mL DNAse I). The cells were resuspended by gentle vortexing, then incubated for one hour at 37 ^{o}C at 250 rpm. After this the lysate was separated by centrifugation at 3600 rpm for 15 minutes. For screening, 75 µL of clarified lysate was added to a microtiter plate, followed by 10 µL of phosphate buffer (50 mM, pH 7.4). To this lysate was added 5 µL NADP (10 mM stock, 500 µM final concentration), 5µL flavanone substrate (100 mM stock, 5mM final concentration), 2 µL glucose (1 M stock, 20 mM final concentration), and reactions were initiated by 5 µL GDH-109 (180 U/mL stock, 9 U/mL final concentration). The plate was then covered with an aluminum plate seal and the reactions were then left to incubate for 24 hours at room temperature on a plate shaker set to 600 rpm. The reactions were then quenched via the addition of 1 volume of methanol and the reactions were analyzed via UHPLC.

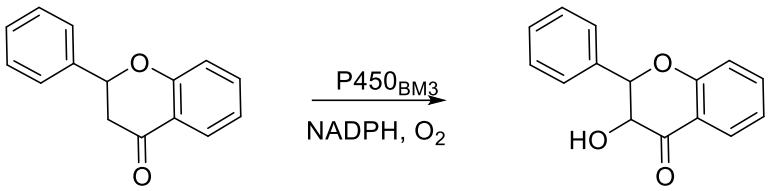
Figure 5. 10: UHPLC trace of WT F87A catalyzed hydroxylation of flavanone



Retention time 2.672 represents the hydroxylated product, RT 3.276 is the starting material.

Expression and purification P450BM3 variants for identification of the hydroxylated product:

Scheme 5. 5: P450 catalyzed flavanone hydroxylation

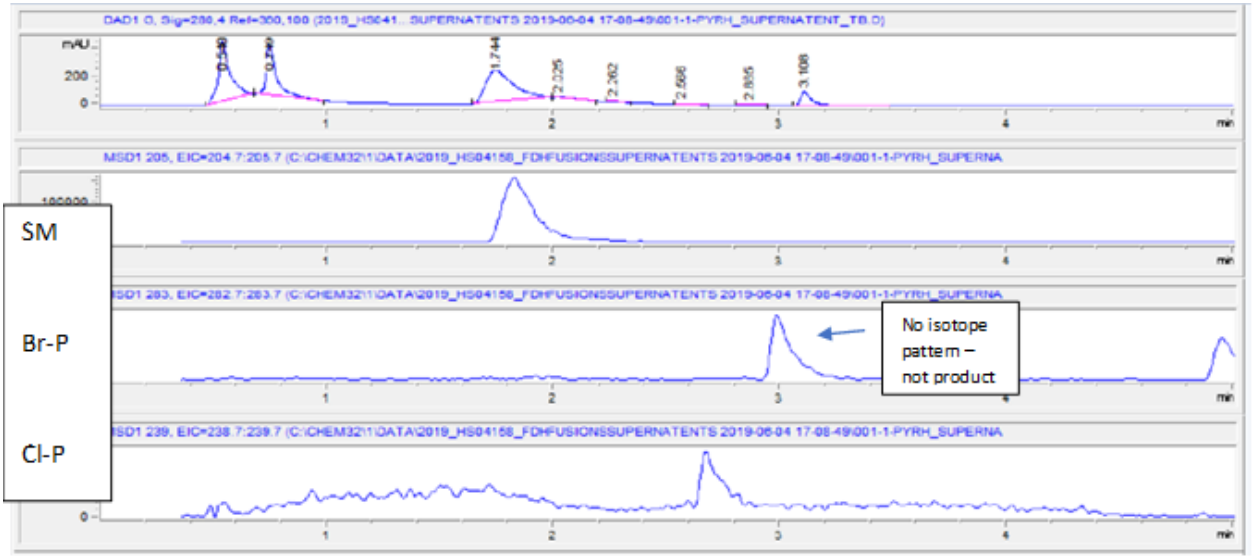


$P450_{BM3}$ variants WT-F87A and 9-10A F87L were grown and purified according to the general procedure. To a 20 mL glass scintillation vial was added 2.19 mL phosphate buffer (50 mM, pH 7.4), 0.15 mL GDH-109 (180 U/mL stock, 9 U/mL final concentration), 0.3 mL substrate (50 mM stock concentration, 1 mM final), 60 µL glucose (1 M stock, 20 mM final concentration), and 300 µL of the P450 variant (unknown concentration). Reactions were initiated by the addition of 10 mg NADP (disodium salt, ~5 mM) and the reactions were sealed with a breathable plate seal and allowed to incubate for 20 hours in a horizontal incubator with gentle shaking. The next day aliquots of the reactions were analyzed by UHPLC and it was found that the reaction with 9-10A F87L showed no conversion, whereas the WT-F87A reaction mimicked the lysate reaction. The WT-F87A reaction was removed from the incubator and diluted with 5 volumes ethyl acetate, the other reaction was discarded. The aqueous layer was washed 3x with additional volumes of ethyl acetate and the combined organic layers were dried with brine followed by

magnesium sulfate. Solvent was removed by rotary evaporation and the resulting oil was dissolved in DMSO and purified via semi preparative HPLC.

5.5 Additional Data

Figure 5. 11: UHPLC-MS traces from in vivo activity of Thal-(G4S)2-RebF and PyrH-(G4S)2-RebF fusions



Thal:

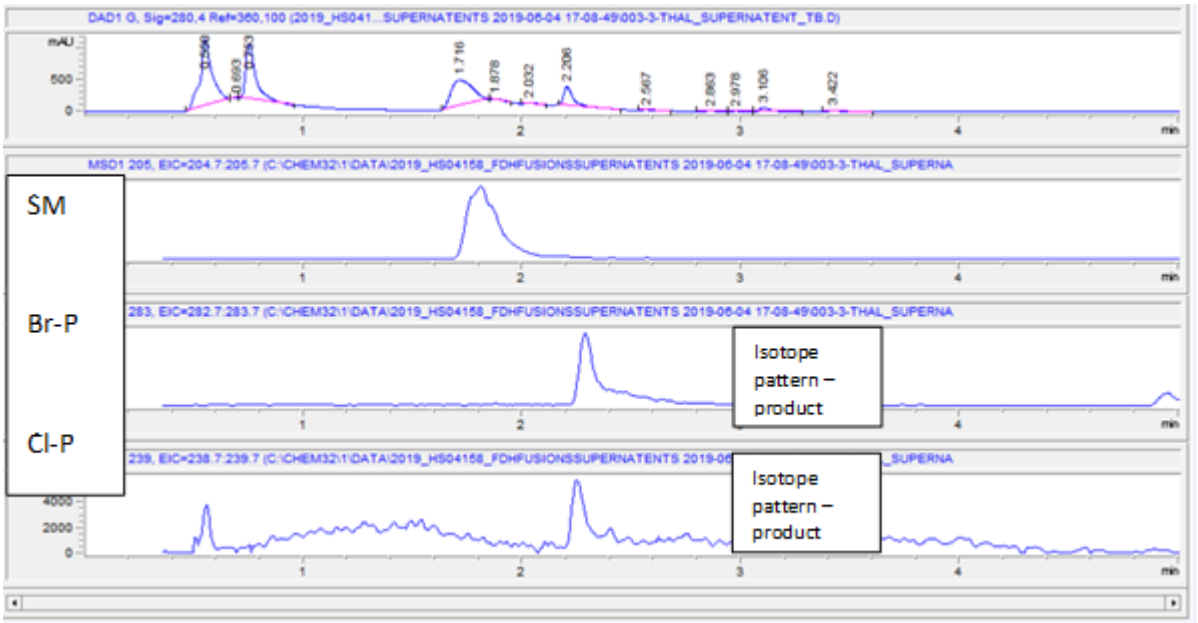
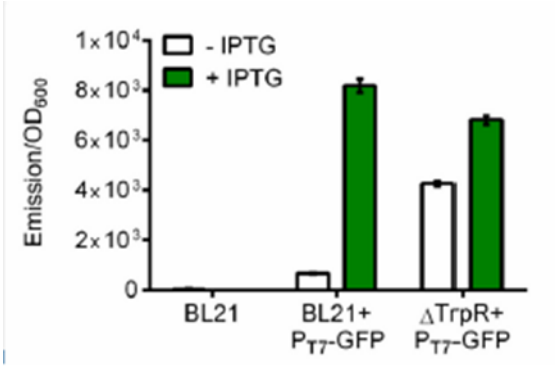
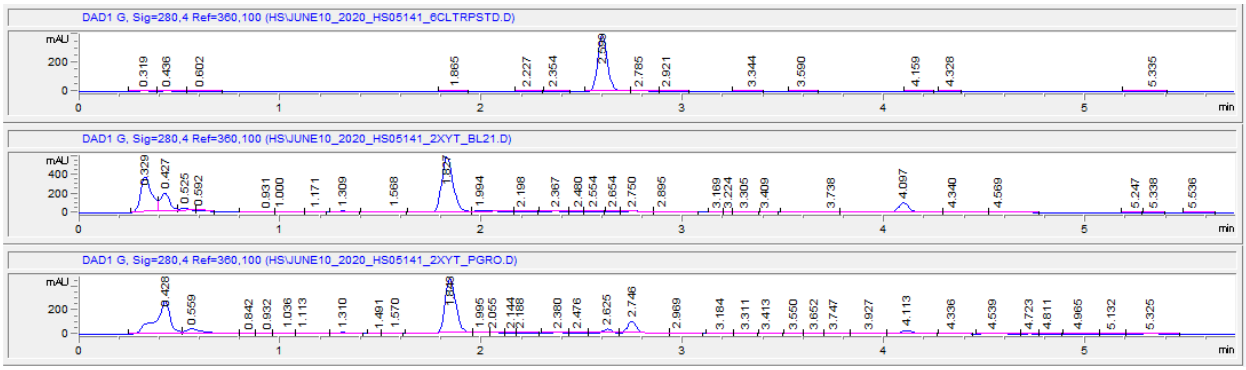


Figure 5. 12: Fluorescence assay with λDE3 lysogenized ΔtrpR cell line



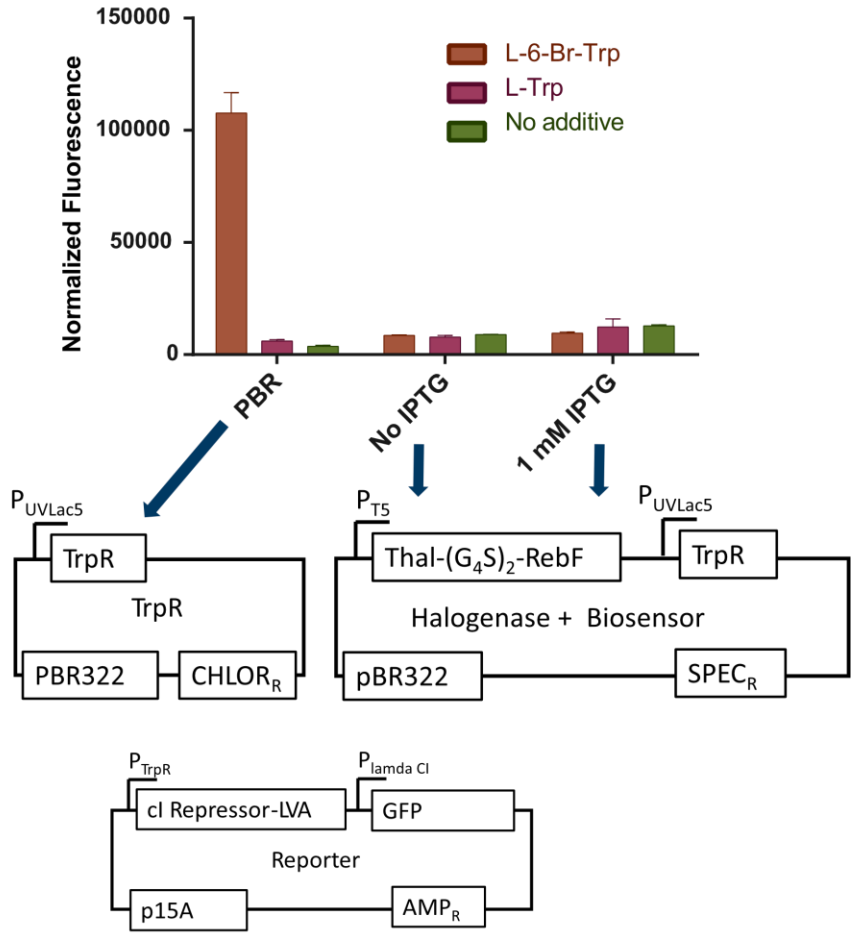
Interpretation: The lysogenized cell line has a very low signal/noise ratio for GFP expression, indicating very poor repression of T7 RNAP expression.

Figure 5. 13: UHPLC trace of in-vivo Thal-RebF fusion activity with the halogenase expressed from T5 promotor in BL21 λDE3 cells



Top: UPLC trace of 6-chlorotryptophan, used as a surrogate for the authentic of 6-bromotryptophan which was not available at the time. Middle: Trace of Thal-RebF fusion expressed from the T5 promoter in the absence of the pGro7 chaperone plasmid. Bottom: Trace of Thal-RebF fusion expressed from T5 promoter with the co-expression of the pGro7 chaperone plasmid. Interpretation: With the T5 promoter, the pGro7 plasmid is required for activity.

Figure 5. 14: Fluorescence data for Thal-RebF fusion and biosensor single plasmid system



Interpretation: The plasmid containing both the halogenase and the biosensor shows significantly lower fluorescent signal production than the original biosensor system and would likely not be viable for cell sorting.

5.6 References

(1) Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P.; Hayes, M. A.; Hepworth, L. J.; Lovelock, S. L.; Niikura, H.; Osuna, S.; Romero, E.; Ryan, K. S.; Turner, N. J.; Flitsch, S. L. Biocatalysis. *Nature Reviews Methods Primers* **2021**, *1* (1), 1–21. https://doi.org/10.1038/s43586-021-00044-z.

(2) Hughes, G.; Lewis, J. C. Introduction: Biocatalysis in Industry. *Chemical Reviews*. 2018, pp 1–3. https://doi.org/10.1021/acs.chemrev.7b00741.

(3) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nature* **2012**, *485* (7397), 185–194. https://doi.org/10.1038/nature11117.

(4) Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med Chem Lett* **2017**, *8* (5), 476–480. https://doi.org/10.1021/acsmedchemlett.7b00114.

(5) Leemhuis, H.; Kelly, R. M.; Dijkhuizen, L. Directed Evolution of Enzymes: Library Screening Strategies. *IUBMB Life*. 2009, pp 222–228. https://doi.org/10.1002/iub.165.

(6) Moore, J. C.; Arnold, F. H. Directed Evolution of a Para-Nitrobenzyl Esterase for Aqueous-Organic Solvents. *Nat Biotechnol* **1996**, *14* (4), 458–467. https://doi.org/10.1038/nbt0496-458.

(7) Marshall, J. R.; Yao, P.; Montgomery, S. L.; Finnigan, J. D.; Thorpe, T. W.; Palmer, R. B.; Mangas-Sanchez, J.; Duncan, R. A. M.; Heath, R. S.; Graham, K. M.; Cook, D. J.; Charnock, S. J.; Turner, N. J. Screening and Characterization of a Diverse Panel of Metagenomic Imine Reductases for Biocatalytic Reductive Amination. *Nat Chem* **2021**, *13* (2), 140–148. https://doi.org/10.1038/s41557-020-00606-w.

(8) Reisky, L.; Büchsenschütz, H. C.; Engel, J.; Song, T.; Schweder, T.; Hehemann, J. H.; Bornscheuer, U. T. Oxidative Demethylation of Algal Carbohydrates by Cytochrome P450 Monooxygenases Brief-Communication. *Nat Chem Biol* **2018**, *14* (4), 342–344. https://doi.org/10.1038/s41589-018-0005-8.

(9) Morlock, L. K.; Böttcher, D.; Bornscheuer, U. T. Simultaneous Detection of NADPH Consumption and H 2 O 2 Production Using the Ampliflu™ Red Assay for Screening of P450 Activities and Uncoupling. *Appl Microbiol Biotechnol* **2018**, *102* (2), 985–994. https://doi.org/10.1007/s00253-017-8636-3.

(10) Huang, R.; Chen, H.; Upp, D. M.; Lewis, J. C.; Zhang, Y.-H. H. P. J. A High-Throughput Method for Directed Evolution of NAD(P)+-Dependent Dehydrogenases for the Reduction of Biomimetic Nicotinamide Analogues. *ACS Catal* **2019**, *9* (12), 11709–11719. https://doi.org/10.1021/acscatal.9b03840.

(11) Snoek, T.; Chaberski, E. K.; Ambri, F.; Kol, S.; Bjørn, S. P.; Pang, B.; Barajas, J. F.; Welner, D. H.; Jensen, M. K.; Keasling, J. D. Evolution-Guided Engineering of Small-Molecule Biosensors. *Nucleic Acids Res* **2020**, *48* (1). https://doi.org/10.1093/nar/gkz954.

(12) Ding, N.; Zhou, S.; Deng, Y. Transcription-Factor-Based Biosensor Engineering for Applications in Synthetic Biology. **2021**, *15*, 7. https://doi.org/10.1021/acssynbio.0c00252.

(13) Javanpour, A. A.; Liu, C. C. Evolving Small-Molecule Biosensors with Improved Performance and Reprogrammed Ligand Preference Using OrthoRep. *ACS Synth Biol* **2021**, *10* (10), 2705–2714. https://doi.org/10.1021/acssynbio.1c00316.

(14) Taylor, N. D.; Garruss, A. S.; Moretti, R.; Chan, S.; Arbing, M. A.; Cascio, D.; Rogers, J. K.; Isaacs, F. J.; Kosuri, S.; Baker, D.; Fields, S.; Church, G. M.; Raman, S. Engineering an Allosteric Transcription Factor to Respond to New Ligands. *Nat Methods* **2016**, *13* (2), 177–183. https://doi.org/10.1038/nmeth.3696.

(15) Cheng, F.; Kardashliev, T.; Pitzler, C.; Shehzad, A.; Lue, H.; Bernhagen, J.; Zhu, L.; Schwaneberg, U. A Competitive Flow Cytometry Screening System for Directed Evolution of Therapeutic Enzyme. *ACS Synth Biol* **2015**, *4* (7), 768–775. https://doi.org/10.1021/sb500343g.

(16) Andorfer, M. C.; Belsare, K. D.; Girlich, A. M.; Lewis, J. C. Aromatic Halogenation by Using Bifunctional Flavin Reductase–Halogenase Fusion Enzymes. *ChemBioChem* **2017**, *18* (21), 2099–2103. https://doi.org/10.1002/cbic.201700391.

(17) Hannig, G.; Makrides, S. C. Strategies for Optimizing Heterologous Protein Expression in Escherichia Coli. *Trends Biotechnol* **1998**, *16* (February), 54–60. https://doi.org/10.3233/978-1-61499-880-8-89.

(18) Janzen, W. P.; Bernasconi, P. *High Throughput Screening. Methods and Protocols, Second Edition.*; 2009; Vol. 565.

(19) Prabhu, R. N.; Ramesh, R. Catalytic Application of Dinuclear Palladium(II) Bis(Thiosemicarbazone) Complex in the Mizoroki-Heck Reaction. *Tetrahedron Lett* **2012**, *53* (44), 5961–5965. https://doi.org/10.1016/j.tetlet.2012.08.120.

CHAPTER 6: EVALUATION OF THE BIFUNCTIONAL HALOGENASE AETF FOR BIOCATALYSIS

6.1 Introduction

The structure and biosynthetic pathway of aetokthonotoxin (AETX) was recently elucidated.[1] This compound had previously eluded detection even after the likely producer of this compound was identified. It wasn't until the mass spectrum of the product was obtained and the researchers recognized the distinctive isotope pattern of a penta-brominated compound that the organism was cultured in media supplemented with bromide to produce AETX, which is responsible for bald eagle die-offs. This toxin possesses a number of unique structural features, including a nitrile group (rare in natural products), an unsymmetrical indole dimer core, and five bromine substituents. In initial efforts to characterize the biosynthesis of AETX, it was found that two different halogenases were involved. AetA installs three bromide substituents on one indole core, and AetF brominates tryptophan at its 5- and 7-positions. It was also noted that when supplied with NADPH, AetF did not require a reductase partner for robust halogenation activity, meaning that AetF functioned both as a flavin reductase and flavin-dependent halogenase. Later work from the Moore group[2] found AetF is unique relative to other previously characterized flavin dependent halogenases (FDHs). The bifunctional catalytic activity of AetF is reminiscent of some flavin monooxygenases, such as the BVMO cyclohexanone monooxygenase used for esomepreazole oxidation.[3]

Figure 6. 1: Retrosynthesis and enzymatic construction of aetokthonotoxin

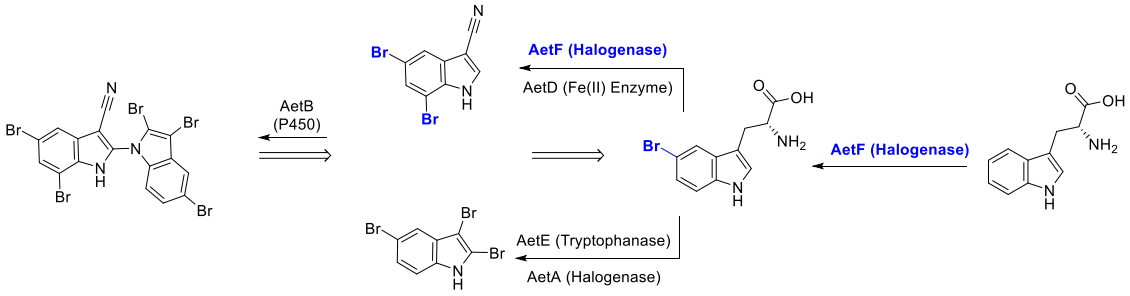


Figure adapted from Journal of the American Chemical Society (2022) 144(7) 2861-2866

From a biocatalysis standpoint, single component enzymes have many benefits over two-component systems.[4] Reductase enzymes required for catalysis contribute to protein waste and often need to be produced in-house due to limited commercial availability. The presence of extraneous enzymes also complicates reaction scalability, as the high amounts of cosolvent required for substrate solubilization may deactivate the reductase enzyme even if the primary biocatalyst is functional. In cases where bifunctional enzymes are desired, but none are available from genomic sources, this bifunctionality can be engineered by genetic fusion of two enzymes connected by a chain of linker residues. For example, previous efforts from our group found that by genetically fusing a flavin-dependent halogenase to a flavin reductase, product titers from *in vivo* halogenation reactions were significantly improved.[5] For *in vitro* reactions however, the fusion enzyme expressed poorly and steady state kinetic analysis of this fusion enzyme showed a significantly reduced k_{cat}/K_M compared to the wild-type RebH.

Among flavin dependent halogenases, single-component flavin reductases/flavin-dependent halogenases are rare. Only one other bifunctional flavin reductase/flavin-dependent halogenase has been described in literature: the phenol halogenase and decarboxylase Bmp5.[6] The only known reaction catalyzed by this enzyme is the di-bromination of an activated 3-hydroxybenzoic acid substrate to generate 2,4-dibromophenol. In contrast, AetF exhibits selective dibromination of the benzene ring of L-tryptophan despite the fact that the C2 position of the pyrrole ring is significantly more electronically activated toward halogenation. Given the selectivity of AetF, we envisioned that the natively bifunctional AetF could simplify biocatalytic halogenation and explored its potential as a biocatalyst. A paper based on some of the experiments described in this chapter has been published.[7]

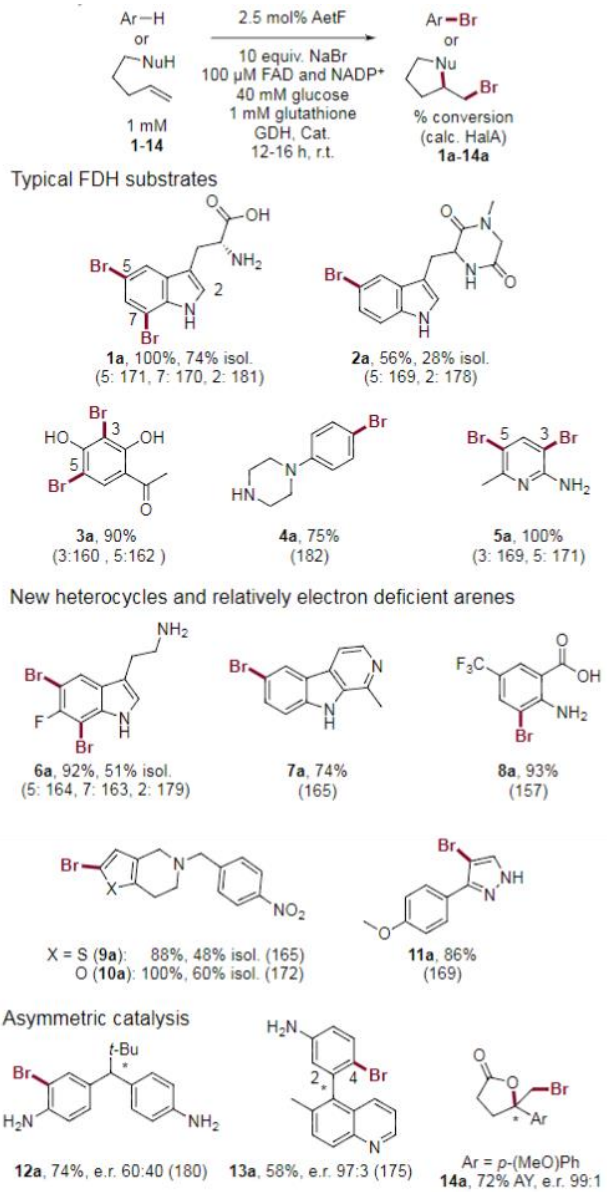
Authorship

Initial characterization of the substrate scope for AetF, the idea to explore iodination, cloning for the K258A variant, reaction optimization, and chiral method development for the bromolactonization substrate was all conducted or led by myself. Preparative product isolation, product characterization, and chiral method development for the cycloiodoetherification reaction was conducted by Yuhua Jiang. Chiral method development reactions for other substrates in this study were either conducted by Ariana Vargas or Dr. Maria Brown.

6.2.1 – Exploration of the substrate scope of AetF and verification of the catalytic lysine

Initial studies of the substrate scope of AetF were focused on recapitulating the known activity on L-tryptophan and testing a panel of substrates known to be accepted by a variety of halogenases.[8] AetF was expressed and purified according to literature protocols[9] and isolated as a bright yellow solution, presumably a result of bound FAD. Expression yields of AetF were lower than the 40 mg/L reported by Moore et al.,[2] and instead were closer to 20 mg/L. It was found that, like for other reported halogenases, the inclusion of the pGro7 chaperone plasmid encoding the proteins GroEL and GroES improved soluble expression of AetF (see **Figure 6.5** in the additional data section). The isolated enzyme was used to set up bioconversions at enzyme loading typical for halogenase reactions, 2.5 mol% of enzyme and 1 mM of substrate (**Figure 6.2**). UHPLC analysis of bioconversions confirmed that L-tryptophan and a variety of other indole containing and small, activated aromatic compounds were successfully brominated by AetF. Analysis of accepted substrates from this initial screen showed that AetF had a general preference for smaller substrates, exhibiting lower activity on 2-phenylcarbazole than for harmane, for example. A screen substituting bromide for 100 mM NaCl showed that AetF had nearly no chlorination activity even for very activated aromatic compounds. Only L-tryptophan was chlorinated about 22% conversion to a single monochlorinated product, contrasting with the observed robust bromination activity for a variety of substrates.

Figure 6. 2: Bromination substrate scope of AetF.

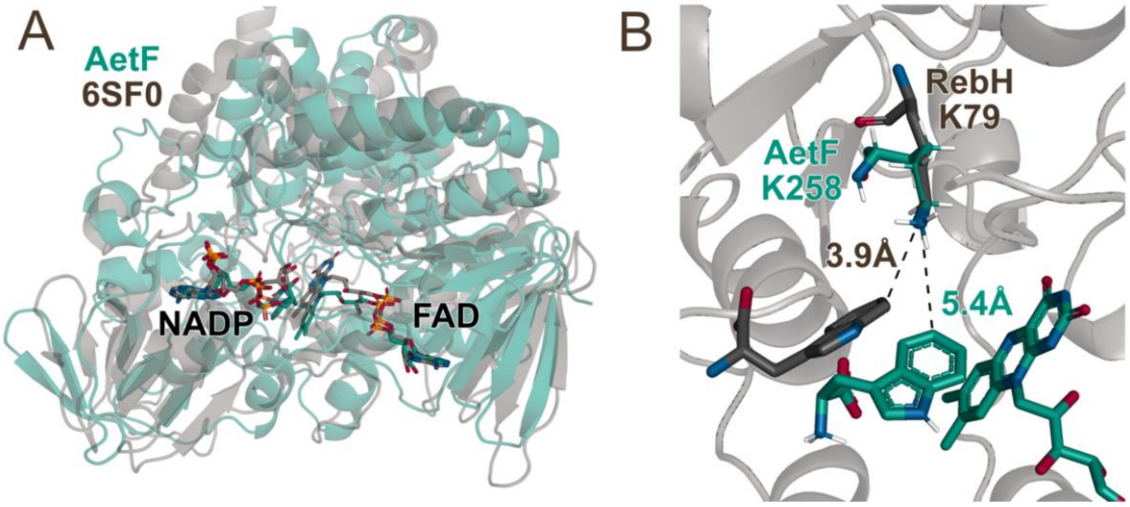


Substrate scope of AetF. %Conversion, %Isol. Yield (HalA value of halogenated position). Reactions were conducted using purified enzyme on analytical scale. When NBS material did not match the retention time of enzymatic product on UHPLC, preparative reactions were conducted to isolate material for characterization. SOURCE: *Angewandte Chemie* 2022.

In many of the chromatograms for these reactions, an additional peak consistent with a product bearing a tri-brominated isotope pattern was observed. Based on its m/z value, this product was assigned to be tribromoimidazole resulting from bromination of imidazole that was present in AetF

samples due to incomplete buffer exchange following IMAC purification. This result was surprising due to the relatively deactivated nature of the imidazole compared to most accepted halogenase substrates [10] as measured by halenium affinity, which is a calculated score that correlates to the ability of a potential substrate to stabilize a halenium ion.[11] A second panel of relatively unactivated arenes was screened to establish the generality of this finding (**Figure 6.2**). Many of these substrates, which had never been accepted by other FDHs, were halogenated by AetF in high yield. Recent reports have shown that RebH and other traditional FDHs can be used for asymmetric catalysis, including desymmetrization,[12] halocyclization,[13,14] and dynamic kinetic resolution of atropisomers.[15] Inspired by this work, we also evaluated AetF for enantioselective catalysis (**Figure 6.2**). AetF exhibited excellent to moderate enantioselectivity for each asymmetric reaction. While AetF did not successfully halogenate the quinazolinone substrate used for the evolution of 3-T (**Chapter 3**),[15] it did generate essentially a single atropisomer of a quinoline compound that no enzyme in the 3-T lineage generated in high yields. The *t*-butyl dianiline compound **12a** was used as a model to sculpt the active site of 4-V for enhanced enantioselectivity,[12] and AetF was able to generate a the desymmetrized product in 74% assay yield with 60:40 e.r. Whereas these reactions took significant optimization with wt-RebH, AetF was capable of each reaction without engineering.

Figure 6. 3: Alphafold model of AetF overlaid generated for structural analysis



A) FMO 6SF0, a human derived FMO generated via ancestral sequence reconstruction. B) Comparison of the active site of AetF and RebH with the lysine residues overlaid showing the analogy between the known catalytic lysine residue of RebH (K79) and the putative catalytic residue of AetF (K258). SOURCE: *Angewandte Chemie* 2022.

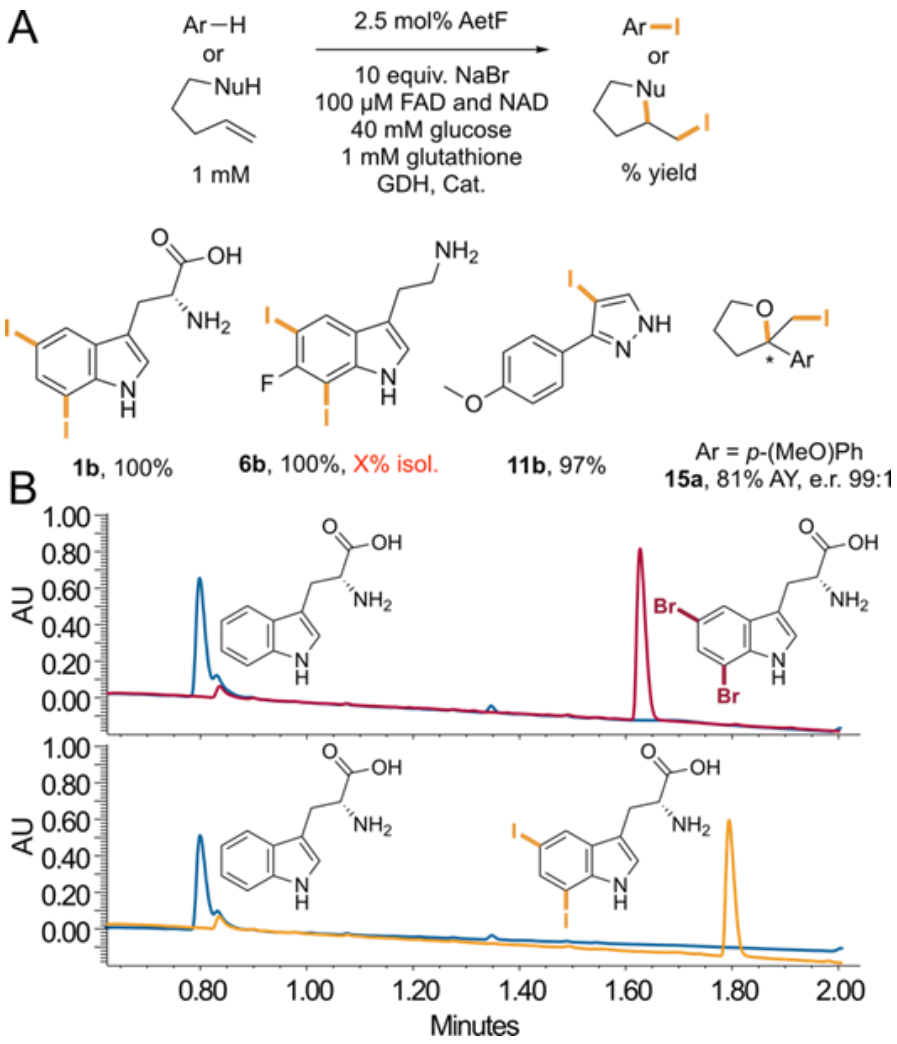
As previously mentioned, AetF is unique relative to other FDHs and has only been recently described in literature. As such, no structural data or homology model could be analyzed for identification. We instead turned to Alphafold[16] for modeling the AetF structure and used this predicted structure as input for a DALI search to identify the nearest structural homologs. This search indicated that the nearest homologs are ancestral reconstructions of human membrane bound bifunctional flavin-reductase/monooxygenases (FMOs). Overlaying the crystal structure of one such FMO with the AetF model showed a resemblance of the active site between the two enzymes much stronger than for AetF with other FDH active sites. Particularly noteworthy was the FAD binding site in 6SF0.[17] When 6SF0 was used as a model to dock the flavin cofactor into the active site of AetF, we found a conspicuous lysine residue, K258, within 5.4 angstroms of the C-5 atom of L-tryptophan docked within the AetF active site. The positioning of this lysine residue bears strong resemblance to the catalytic lysine present in traditional FDHs proposed to activate the reaction HOX species for electrophilic halogenation. We envisioned that if mutagenesis of this residue ablated activity it could confirm the reliability of our

model to facilitate future mutagenesis of the active site. The K258A mutant was cloned via SOE-PCR,[18] and the activity of the resulting enzyme was compared to the wild-type AetF in lysate against the native L-tryptophan substrate. After confirming soluble expression of AetF K258A via SDS PAGE analysis (See **Figure 6.5** in the experimental section) and evaluating halogenation activity via UHPLC, we confirmed that K258A completely lacked the bromination activity of the wild type halogenase (**Figure 6.4**). This result helps to confirm the validity of our model and demonstrates the utility of Alphafold for sequence-based protein engineering efforts which would otherwise be hindered by laborious crystallization efforts.

6.2.2 - Discovery of selective iodination catalyzed by AetF

Iodination remains rare for flavin-dependent halogenases. Despite previous work with flavin-dependent halogenases having demonstrated the ability of iodine to attack the hydroperoxy flavin intermediate and generate HOI with a faster rate than was measured for Br or Cl,[19] it wasn't until 2019 that an FDH was described which preferred bromination and iodination over the corresponding chlorination reactions.[20] It was hypothesized that the relatively open active site observed in this halogenase was sufficient for accommodating the iodine atom and facilitated the enzyme catalyzed iodination. More recently however, a report called this observation into question.[21] Control experiments performed showed that in the absence of an FDH, reduced flavin could halogenate a variety of electronically activated substrates. It was suspected that these reactions were occurring in solution through the reaction of HOI generated via oxidation of I^- by peroxide formed from the reduced flavin, as the inclusion of catalase which converts peroxide into oxygen and water with rapid kinetics. The lack of suck controls being conducted in previous iodination experiments along with the substrate scope consisting of electronically activated compounds halogenated at highly activated positions calls into question the enzyme-controlled selectivity of previously identified iodinases.

Figure 6. 4: Iodination results with AetF and K258A



A) AetF-catalyzed aromatic iodination and cycloiodoetherification. Conversions were determined from relative integration values for starting material and product; selected isolated yields (isol.) and assay yields (AY) relative to authentic internal standards were also obtained. B) Bromination and iodination of tryptophan catalyzed by AetF (red and orange traces, respectively) and AetF K258A (blue traces), showing now conversion to either halogenated product using the latter enzyme. GDH = glucose dehydrogenase, Cat. = catalase. SOURCE: *Angewandte Chemie* 2022.

It has been hypothesized that perhaps the absence of selective iodination activity in many halogenases is due to a constricted halide binding site that cannot accommodate iodide. We envisioned that if our model of the relatively open AetF active site is accurate, HOI could be generated in the active site of AetF. If K258 in the catalytic lysine could activate HOI as it does for HOBr, this might enable iodination of compounds previously reported as being inactive with the earlier described 'iodinases',

such as tryptophan or other less electronically activated indoles. Toward this end, we screened for iodination activity with both the wild-type AetF enzyme and the K258A mutant that had previously showed ablated brominase activity. These enzymes were screened against a reduced panel of substrates including substrates for which AetF gave full conversion and one where AetF had less than 10% activity when screened in lysate. Not only did the wild-type enzyme supplemented with NaI largely recapitulate the bromination results with this substrate panel, but the alanine variant K258A still showed no activity (See **Figure 6.9** in additional information). Given our reactions included catalase to quench peroxide formed from off-cycle oxygen activation and glutathione which rapidly reduces oxidative species, we believe this study proves that AetF is capable of catalyzing iodination activity. Isolation of the brominated and iodinated L-tryptophan and 6-fluorotryptamine material confirmed that AetF maintains the regioselectivity for iodination, over-riding the inherent electronics for both compounds. Additionally, AetF was found to catalyze the iodoetherification of an aromatic styrene with a pendant alcohol nucleophile with 99:1 e.r., strongly suggesting this iodination is occurring within the enzyme active site (**Figure 6.4**). Taken together, this data verifies that AetF catalyzes site-selective iodination of several compounds in high yields and could provide a valuable way to access iodinated arenes to enable new chemistry.

6.2.3 – Reaction optimization

Having verified the impressive activity and selectivity of AetF for electrophilic halogenation we next sought to optimize conditions for improved enzyme activity and increased reaction scale. These optimization efforts were conducted in lysate to facilitate the exploration of conditions more rapidly. The first optimization study was designed to analyze the effects of different buffers, pH values, and substrate concentrations on halogenation yields. HEPES and phosphate buffer were chosen due to their broad buffering ranges and compatibility with FDHs in previous studies. 6-fluorotryptamine was chosen as a substrate due to the high activity of AetF towards this substrate and its electron deficient character,

though the di-halogenation of this substrate complicated analysis (*vide infra*), and a substrate that undergoes monohologenation should be used in future studies. *E. coli* harboring plasmids encoding AetF and GroEL-GroES was cultured in a 96-well plate and lysed using a stock solution of lysozyme at the desired buffer concentration. Conversion to product was evaluated by UHPLC. This study clearly showed that AetF had improved activity at higher pH and buffer concentration, though it was sensitive to high levels of phosphate buffer (**Figure 6.10**). Notably, measuring the pH of a reaction conducted on 25 mL scale with a starting pH of 7.4 (HEPES, 25 mM) showed that the pH dropped significantly course of the reaction. This observation suggested that AetF deactivation was occurring at more acidic pH, which explained why a higher pH was advantageous for biocatalysis (up to the limit of stability of the enzyme).

Based on the acidification of the reaction, a follow up study was conducted to test more basic pH conditions to nullify this effect. This acidification was determined to likely be the result of the conversion of glucose to gluconic acid by GDH, and such a dramatic change indicated that most of the glucose was being consumed in a manner that is decoupled from product formation. In light of this finding, the impact of additional glucose was evaluated as well, and we found that 2x glucose resulted in improved conversion. The inclusion of 1 mM NaCl also improved yields, although the reason why is unclear. Lastly, the observed sensitivity of the reaction to lower pH and the observed acidification process led us to explore an alternate cofactor regeneration system that does not produce an acidic byproduct. Alcohol-dehydrogenase based cofactor regeneration systems reduce NADP to NADPH with concomitant oxidation of isopropanol to acetone and have been applied to industrial synthesis, demonstrating the efficiency of this system. A comparison of the optimized buffer conditions (25 mM Tris-HCl, pH 9.0) and the original buffer conditions (25 mM HEPES, pH 7.4) showed that using this cofactor regeneration system, the buffer optimized for the GDH cofactor regeneration system underperformed the original conditions (**Figure 6.12**). This perhaps shows that the benefits from our optimization potentially came from limiting the acidification from gluconic acid buildup. Further

optimization of the IPA loading showed that 1% v/v IPA gave the highest conversion (**Figure 6.13).** While more studies are required to demonstrate which cofactor regeneration system is superior for reaction scale up, both have advantages and disadvantages and the option to use either could assist future reaction development.

6.2.4 – Efforts towards scaled up halogenation with AetF

Considering the poor expression of AetF in microtiter plates and the relatively high substrate loading used for optimization studies, it was clear that AetF exhibited abnormally high TTN for a flavin-dependent halogenase. The rapid acidification of buffer solutions indicated that significant cofactor turnover was occurring independent of substrate halogenation and suggested uncoupling of NADPH consumption and either FAD reduction or HOBr. These observations suggested that to maximize the efficiency of AetF for preparative halogenations, tighter control and a deeper understanding of reaction parameters would be required. We envisioned that by collaborating with the biocatalytic process chemistry group at Pfizer we could take advantage of the expertise and equipment of industrial chemists to study and improve the reaction dynamics of AetF.

One of our primary interests was to understand the impact of dissolved oxygen content on biocatalysis using AetF. To accomplish this, we used an EasyMax reactor equipped with a dissolved oxygen probe. In addition to measuring the total dissolved oxygen content in a sample, the EasyMax reactor allowed us to control the pH and temperature of the reactions, necessary for maintaining a stable environment in the presence of gluconic acid build-up. The first reaction was set up at 130 mL using clarified lysate. The EasyMax reactor was sealed such that there was no opening of the reaction to atmosphere, limiting airflow into the system. It is important to note that the reaction is not completely isolated from the environment. Tracking the reaction with the EasyMax system, we noted that upon the addition of all the reaction components the dissolved oxygen in the reaction plummeted from 16% to 0.5% (See **Figure 6.14** in the experimental). The dissolved oxygen stayed at 0.5% for roughly 10 hours, at

which point it began to slowly climb until it reached 16% again where it stayed flat. The reactor was disassembled and sampling the reaction via UHPLC revealed 47% total conversion from starting material. Given the very low solubility of oxygen in solution we expected that the further drop in dissolved oxygen may create an oxygen-limited environment and slow enzyme catalysis. We envisioned that increasing air flow in the reaction by bubbling air would increase available oxygen and therefore, conversion. A repeat of the reaction including an air sparge was confirmed and allowed to react in the same way for 24 hours (**Figure 6.15**). To probe whether the reaction format impacted conversion we also took a 1 mL aliquot from the reaction and arrayed 75 µL of this sample into a microtiter plate. Upon sampling of the reaction by UHPLC we instead observed significantly reduced conversion, down from 47% from our first attempt to 10%. In contrast, the aliquoted sample that was placed into a microtiter plate was fully converted to dibrominated material. The observed difference in activity between these scaled up experiments and the unexpected difference observed with and without the air sparge clearly indicates that more study is necessary to scale these reactions up successfully.

Another consideration for scaling reactions with biocatalysts is the extent to which the enzyme needs to be purified to facilitate large scale reactions. While purified enzymes typically exhibit the highest level of specific activity, the cost associated with purification is typically prohibitive for large scale processes. Instead, the preferred method for enzyme formulation is lyophilized clarified lysate or whole cells, which allow for convenient storage. To this end, we wanted to screen various formats of AetF to see if lysate, whole cells, or lyophilized powder could be used for screening preparative reactions. A set of bioconversions conducted with each of these showed that of the three, whole cells seemed to perform the best for bromination of the substrate 6-flourotryptamine upon scale up, generating brominated products in 28% yield compared to only 18% yield from clarified lysate and 18% using lyophilized enzyme (for more information, see the experimental section). During efforts to isolate material from bioconversions with AetF, preparative reactions were conduced using low loadings of

purified enzyme with both NaBr and NaI as the halide sources. Surprisingly, the bromination reaction showed lower conversion than the iodination reaction both at preparative scale and in a microtiter plate (Table 6.2. In fact, for a 30 mL preparative reaction we observed full conversion to 5,7-diiodotryptophan using only 0.1 mol% of purified AetF, a remarkable 2000 TTN assuming no byproduct formation. Attempts to increase substrate concentration beyond this to 2.5 and 5 mM resulted in much lower activity for the preparative reactions. Clearly, much work is still to be done for developing AetF, but this high observed TTN combined with the substrate scope of the enzyme may indicate that with more reaction and enzyme engineering AetF may be a starting point to develop a viable biocatalyst for scalable halogenation.

6.3. Conclusion

The work described in this chapter is the first exploration of a single-component flavin-dependent halogenase for biocatalysis. Exploration of AetF substrate scope revealed that this enzyme brominates a range of compounds, including indoles, anilines, phenols, and heterocycles, some of which contain electron withdrawing groups and are not reactive toward other FDHs. AetF is also a competent catalyst for several enantioselective transformations such as desymmetrization, atroposelective halogenation, and halocyclization, demonstrating surprising flexibility for a wild type enzyme. While this enzyme is poorly active for chlorination, we found that it catalyzes site- and enantioselective iodination reactions not described for other FDHs. An Alphafold model of AetF revealed a putative catalytic lysine residue, and mutation of this residue to alanine resulted in loss of activity, helping to validate the proposed model of the enzyme active site. Optimization of the buffer and cofactor regeneration system shows that AetF functions well with both GDH and ADH driven NADPH production and exhibits sensitivity to acidic conditions. Attempts at reaction scale up using the EasyMax bioreactor showed that for unknown reasons there is a substantial loss of activity, which became even more pronounced when an air stream was applied to the reaction. Taken together, these data suggest that while AetF holds

promise for simplifying halogenase biocatalysis, there remains a significant amount of protein engineering and reaction optimization required to make this enzyme ready for process chemistry.

6.4 Experimental

6.4.1 General methods, materials, and chromatography

Materials

Greiner Bio-One polypropylene 96-well V-bottom plates (product number 651201) were purchased from Fisher Scientific. Agilent 0.2 µm PVDF 96-well filter plates (product number 203980-100) were purchased from Agilent. NADP, FAD, and antibiotics were purchased from Chem-Impex International Inc. (Wood Dale, IL). Substrates were purchased from Sigma-Aldrich, Chem-Impex, AK Scientific, Enamine, and Santa Cruz Biotechnologies. A gene encoding AetF was purchased as an *E. coli* codon-optimized construct in pET28(a) from Twist Biosciences as described below. 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). L-glutathione reduced was obtained from Millipore Sigma. 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 Pre-stained Protein Standard, Broad Range (11-190 kDa); product number P7706) was purchased from New England Biolabs (Ipswitch, MA).

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. Stock solutions of 10 mM NADP and 10 mM FAD were prepared in 25 mM HEPES, pH 7.4 (reaction buffer) and stored at -20 °C until use. Stock solutions of 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 100 mM in DMSO or in IPA. GDH was prepared as 180 U/mL stock solution

in reaction buffer immediately before reaction setup. Catalase and glutathione stock solutions were prepared immediately before reaction setup.

Instruments

Ultra-high pressure liquid chromatography (UHPLC) was performed using one of two systems. Studies of the AetF substrate scope were performed on an Agilent 1200 series system with a 1290 Infinity II high-speed pump, a 1260 Infinity II diode array detector, and a 1290 Infinity II multisampler with single-needle configuration connected to a 6135X single quadrupole mass spectrometer with an Agilent Jet Stream ESI source. All achiral UPLC analysis was performed using an Eclipse Plus C18 2.1x5 mm guard column with a 1.8 µm pore size (part number 821725-901) connected to a ZORBAX rapid resolution C18 column (part number 959757-902) with mobile phase A of water with 0.1% FA and mobile phase B of acetonitrile with 0.1% FA.

The second system was an Acquity Waters UPLC equipped with a Waters BEH C18 column (2.1 × 50 mm, 1.7 µm, part no. 186002350) maintained at 30 °C, using 0.1% trifluoracetic acid in water (mobile phase A) and acetonitrile (mobile phase B) at a flow rate of 0.6 mL min^{-1}. For the iodination experiments, the m/z values of the iodinated products were verified using the same system but equipped with an SQ Detector/Mass spectrometer with the cone voltage set to 30V and the capillary voltage set to 4.5 kV.

Reverse phase purification was performed using either or both: 1.) a Biotage Isolera One with 12 g SNAP-KP-C18-HS columns using 0.1% FA (formic acid) in water as mobile phase A and 0.1% FA in methanol as mobile phase B or 2.) 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% FA in water as mobile phase A and 0.1% FA in acetonitrile as mobile phase B.

Chromatography (UHPLC, UHPLC-MS, SFC)

Biotage method 1: Used to purify quinoline compound **13**:

Column Volumes	%A	%B
1	90	10
15	0	100
20	0	100

UHPLC-MS Method 1: Used to analyze products **1a**, **2a**, **4a**, **6a**, **7a**, **13a**, and **1b**.

Column: Agilent Eclipse Plus C18 column (3.5 µm particle size; 4.6 x 150 mm)
Mobile phase A: Water with 0.1% TFA

Mobile phase B: ACN with 0.1% TFA

Method:

Time (min)	%A	%B
0	95	5
0.5	95	5
8	20	80
8.01	5	95
8.25	5	95
8.5	95	5

UHPLC-MS Method 2: Used to analyze product **3a**, **5a**, **8a**, **9a**, **10a**, **11a**, **12a**, and **11b**.

Column: Agilent Eclipse Plus C18 column (3.5 µm particle size; 4.6 x 150 mm)
Mobile phase A: Water with 0.1% TFA

Mobile phase B: ACN with 0.1% TFA

Method:

Time (min)	%A	%B

0	80	20
0.5	80	20
8	20	80
8.01	5	95
8.25	5	95
8.5	80	20

UHPLC-MS Method 3: Used to analyze products **14a** and **15a** during substrate scope studies.

Column: Agilent Eclipse Plus C18 column (3.5 µm particle size; 4.6 x 150 mm)
Mobile phase A: Water with 0.1% TFA

Mobile phase B: ACN with 0.1% TFA

Method:

Time (min)	%A	%B
0	60	40
0.5	60	40
4	30	70
4.01	5	95
4.25	5	95
4.5	60	40

UHPLC Method 1: Used for AetF optimization studies with 6-fluorotryptamine and L-Tryptophan on UHPLC system #2.

Column: ACQUITY UPLC BEH C18 Column, 130 Å, 1.7 µm, 2.1 x 50 mm (part number 186002350) with a flow rate of 0.6 mL/min.
Mobile phase A: Water with 0.1% TFA
Mobile phase B: Acetonitrile

Gradient:

Time (min)	%A	%B
0	95	5

0.25	95	5
3	50	50
3.01	5	95
3.5	5	95
3.51	95	5

6.4.2 Enzyme preparation and molecular biology

Molecular Cloning

Standard molecular biology protocols were followed.[22] Oligonucleotides were purchased from either Sigma-Aldrich or Integrated DNA technologies as lyophilized powder and prepared according to the manufacturers specification. All primer sequence and their use in this project are included in a tale below. AetF was purchased as an *E. coli* codon-optimized gene construct in pET28(a) from Twist Biosciences with the 6x-his tag placed at the N-terminus of the protein embedded in a flexible linker.

Full protein sequence for AetF used in this study. The methionine that starts translation of AetF is in bold, and the putative catalytic lysine is highlighted red:

MGSSHHHHHHSSGLVPRGSH**M**LEVCIIGFGFSAIPLVRELARTQTEFQIISAESGSVWDRLSESGRLDFSLVSSFQTSFYS
FDLVRDYEKDYYPTAKQFYEMHERWRSVYEEKIIRDFVTKIENFKDYSLISTRSGKTYEAKHVVLATGFDRLMNTFLSNFD
NHVSNKTFVFDTMGDSANLLIAKLIPNNNKIILRTNGFTALDQEVQVLGKPFTLDQLESPNFRYVSSELYDRLMMSPVYP
RTVNPAVSYNQFPLIRRDFSWVDSKSSPPNGLIAIKYWPIDQYYYHFNDDLENYISKGYLLNDIAMWLHTGKVILVPSDT
PINFDKKTITYAGIERSFHQYVKGDAEQPRLPTILINGETPFEYLYRDTFMGVIPQRLNNIYFLGYTRPFTGGLANITEMQS
LFIHKLITQPQFHQKIHQNLSKRITAYNQHYYGAAKPRKHDHTVPFGFYTEDIARLIGIHYQPNECRSVRDLLFYYAFPNN
AFKYRLKGEYAVDGVDELIQKVNDKHDHYAQVFVQALSIRNMNSDEAAEWDHSARRFSFNDMRHKEGYRAFLDTYLK
AYRQVENISVDDTVVDEEWNFMVKEACQVRDKVAPNIEEKTHYSKDEDVNKGIRLILSILDSDISSLPDSNGSRGSGNLK
EGDRLCKFEAQSIEFIRRLLQPKNYELLFIRESTVSPGSHRHGETA

The K258A mutation was introduced using SOE-PCR[23] using the following conditions:

Fragment 1 PCR:

PCR mix:

Component	Volume (uL)	Concentration
Forward primer (HSPF016)	2.5	500 nM
Reverse Primer (HSPF-023)	2.5	500 nM
Template (AetF)	1	50 ng
Water	19	NA
Primestar 2x mix	25	NA

PCR Cycle:

Temperature (°C)	Length of cycle (min:sec)	Repetitions
98	5:00	1
98	0:10	28
60	0:05	
72	0:30	
72	5:00	1
4	hold	1

Fragment 2 PCR:

PCR Mix:

Component	Volume (uL)	Concentration
Forward primer (HSPF022)	2.5	500 nM
Reverse Primer (HSPF-020)	2.5	500 nM
Template (AetF)	1	50 ng
Water	19	NA
Primestar 2x mix	25	NA

PCR cycle:

Temperature (°C)	Length of cycle (min:sec)	Repetitions
98	5:00	1
98	0:10	28
60	0:05	
72	0:30	
72	5:00	1
4	hold	1

Vector PCR:

PCR Mix:

Component	Volume (uL)	Concentration
Forward primer (HSPF021)	2.5	500 nM
Reverse Primer (HSPF-017)	2.5	500 nM
Template (AetF)	1	50 ng
Water	19	NA
Primestar 2x mix	25	NA

PCR cycle:

Temperature (°C)	Length of cycle (min:sec)	Repetitions
98	5:00	1
98	0:10	28
60	0:05	
72	1:30	
72	5:00	1
4	hold	1

After the F1, F2, and vector PCRs, the resulting amplicons were purified via gel extraction using the Qiagen gel extraction kit according to the manufacturers protocol. Samples were eluted off the spin column with 30 µL of molecular-grade water. To each tube, 3 µL of NEB Cutsmart buffer (NEB#B6004S) and 1 µL of DPN1 enzyme (NEB#R0176L) was added to each reaction and allowed to incubate at 37 °C for 1 hour. After digestion of the template DNA, the reaction mixture was purified using the Qiagen PCR purification kit according to the manufacturer's protocol, and eluted with 10 µL of molecular grade water. After purification of the F1 and F2 fragment, the assembly PCR was performed to generate the AetF(K258A) insert.

Assembly PCR:

PCR mix:

Component	Volume (uL)	Concentration
Forward primer (HSPF016)	2.5	500 nM
Reverse Primer (HSPF-020)	2.5	500 nM
Template (1:1 mixture of F1:F2)	1	50 ng (total)
Water	19	NA
Primestar 2x mix	25	NA

PCR cycles:

Temperature (°C)	Length of cycle (min:sec)	Repetitions
98	5:00	1
98	0:10	28
60	0:05	
72	1:00	
72	5:00	1
4	hold	1

The resulting PCR product was purified via gel extraction according to the manufacturer's protocol. Concentrations of the purified DNA products was measured using a Thermo NanoDrop 8000 by measuring the absorbance at 260 nm. Ligation of the assembled AetF(K258A) fragment into pET28(a) was done using Gibson Assembly. The Gibson assembly was performed as follows:

Component	Volume (uL)	Amount (ng, pmol)
Insert (AetF[K258A])	1.5	150 (0.15)
Vector (pET28a)	1	100 (0.05)
NEB 2x Gibson Mix	2.5	n/a

This reaction mixture was allowed to incubate for thirty minutes at 50 °C. After the reaction was complete, 2 µL of the assembly mixture was used directly to transform *E. coli* BL21(λDE3) cells Agilent, part number 230132) via electroporation using the EC1 program. Electroporated cells were recovered using 750 µL (Expression recovery medium, Biosearch Technologies Item ID 80030-1) and transferred to a 12 mL culture tube, then incubated at 37 °C for one hour with 210 rpm in a Kuhner Shaker. After this recovery period, selection was performed by plating the culture onto agar containing kanamycin and the plate was incubated at 37 °C overnight. The next day, 5 colonies were picked into 10 mL of LB media supplemented with kanamycin and allowed to grow overnight at 37 °C with 210 rpm. After allowing the cultures to grow overnight, they were removed from the incubator and plasmid was isolated following the Qiagen miniprep

protocol. The K258A mutation was verified by Sangar sequencing performed by Functional Biosciences (Madison, WI).

List of primers used in this study:

Primer name	Sequence (5'->3')	Function	Dir.
HSPF-007	TACGGCATTAGATCAGGAAG	Sequencing	F
HSPF-016	CCGCGCGGCAGCCATATGTTGGAAGTCTGTATCATCG	SOE Fragment 1	F
HSPF-017	GATACAGACTTCCAACATATGGCTGCCGCGCGGCAC	SOE Vector	R
HSPF-020	CTCGAGTCATTAAGCGGTTTCTCCATGGC	SOE Fragment 2	R
HSPF-021	GCCATGGAGAAACCGCTTAATGACTCGAG	SOE vector	F
HSPF-022	ATGGACTGATCGCTATTGCGTACTGGCCCATTGACCAG	SOE Fragment 2	F
HSPF-023	CTGGTCAATGGGCCAGTACGCAATAGCGATCAGTCCATTAG	SOE Fragment 1	R

Gene Expression

After verification of the correctly assembled sequences, AetF constructs harbored within pET28 were transformed into BL21 Gold (DE3) competent cells (Agilent 230132) with the pGro7 chaperone plasmid (Takara Bio, catalog # 3340) and electroporation and recovery was done as described above. Recovered media was plated onto LB agar plates with kanamycin and chloramphenicol for selection and incubated at 37 °C overnight. The next day, single colonies of *E. coli* were picked into 12 mL culture tubes containing 10 mL of LB media supplemented with the appropriate antibiotics. After incubation for 16 hours at 37 °C in a vertical incubator set to 210 rpm, the cultures were removed from the incubator and 400 µL of each was diluted with 400 µL of 50% glycerol to create a glycerol stock. The remaining culture was used to inoculate 1 L of TB media in a 2.8 L Erlenmeyer flask supplemented with the appropriate antibiotics, at which point the flask was placed into a horizontal incubator and the culture was allowed to incubate at 37 °C for 2.5 hours. At this point, the culture was removed from the incubator and placed into an ice bath

for 15 minutes while the incubator was cooled to 18 °C. Gene expression was induced with 100 µL of a 1 M stock of IPTG (final concentration 0.1 mM) and the addition of 2 g of solid L-arabinose directly into the media (final concentration 2 mg/mL). The culture flask was incubated at 18 °C for 24 hours at 210 rpm. After this point in time, the culture flask was removed from the incubator and pelleted in a centrifuge at 15,000 x g for 45 minutes with the centrifuge held at 4 °C. The supernatant was removed from the cell pellet, and the cell pellet was stored at either –20 °C or –80 °C until further use.

For expression in microtiter plates, colonies from the LB agar plate were picked into a 2 mL plate charged with 500 µL of LB media supplemented with the appropriate antibiotics. These cultures were allowed to grow overnight in a Glas-Col incubator set to 37 °C and 900 rpm. The next day, 20 µL of the overnight culture was used to inoculate 800 µL of TB media charged in a 2 mL deep well plate supplemented with the appropriate antibiotics. After inoculation, these cultures were placed into a Glas-Col incubator set to incubate at 37 °C and 900 rpm for 2.5 hours, at which point the culture plate was removed and the incubator was allowed to cool to 18 °C. Once the incubator cooling was complete, the culture plate was set to incubate for 24 hours at 18 °C and 900 rpm. After incubation was complete, the culture plate was placed into a centrifuge and centrifugation was done at 2000 x g and 20 minutes while held at 4 C. After centrifugation was complete, the supernatant was removed from the culture and the cell pellets were stored at –80 °C until use.

Protein Preparation

For preparation of clarified cell lysate from microtiter plates, the cell pellets were resuspended in 100 µL of lysis buffer (0.75 mg/mL lysozyme, 1 uL/mL Benzonase I in the appropriate buffer for final conditions) in a Glas-col inucbator set to 37 °C and 1200 rpm for 15 minutes. After resuspension was complete, the incubation speed was reduced to 900 rpm and the lysis was allowed to continue for an additional 45 minutes. Once cell lysis was complete, the 2-mL deep well plate was placed into the –80 °C freezer until the cell lysate was thoroughly frozen through, at least 1 hour. After the cell lysate was frozen, it was

removed from the –80 °C freezer and set in ambient temperature for 10 minutes. To ensure even thawing of the entire plate, it was then placed into lukewarm water for an additional 20 minutes. Once the cell lysate was thawed, the lysate was clarified via centrifugation at 2000 x g for 20 minutes. The supernatent was then immediately aliquoted into a microtiter plate for bioconversions.

For cell pellets prepared from 1 L cultures, cell lysis was performed by either microfluidization (Microfluidics LM10) or sonication. For microfluidization, the cell pellet was resuspended in 100 mL of lysis buffer (0.75 mg/mL lysozyme, 1 mg/mL DNAse1) and incubated at 37 °C in a vertical incubator for 15 minutes to resuspend the pellet. Once resuspension was complete, the resuspended cells were subjected to microfluidization at 20,000 PSI in two passes through the system. The lysed cell was then clarified via centrifugation at 15,000 x g for 45 minutes, and the resulting clarified lysate was poured into a fresh tube for use. When whole cells were used, the cell pellet was resuspended in 100 mL of HEPES buffer (pH 7.4, 25 mM) via incubation in a vertical incubator set to 18C and 140 rpm for 30 minutes.

For sonication, cells were lysed using a QSonica S-4000 with a 0.5” horn at 40 W and a total processing time of 5 minutes with 1 minute on/off cycles. Cell lysate was clarified at 15,000 rpm for 30 minutes, and the resulting clarified lysate was transferred to a fresh 50 mL centrifuge tube and added to pre-equilibrated Ni-NTA (equilibration buffer: 20 mM phosphate, 300 mM NaCl, 10 mM imidazole pH 7.4). Clarified lysate was incubated with resin for approximately 30 minutes at which point it was transferred to uncapped spin columns and the lysate was allowed to flow through. The resin was washed with at least 5 CV wash buffer (20 mM phosphate, 300 mM NaCl, 25 mM imidazole pH 7.4), at which point the spin column was transferred to a new centrifuge tube and the resin was washed with elution buffer (20 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4). Eluted protein was concentrated via diafiltration using Amicon spin filters Ultra 30K MWCO spin filters and the buffer was exchanged for storage buffer (25 mM HEPES and 10% glycerol, pH 7.4). For long term storage, proteins were immediately frozen in liquid

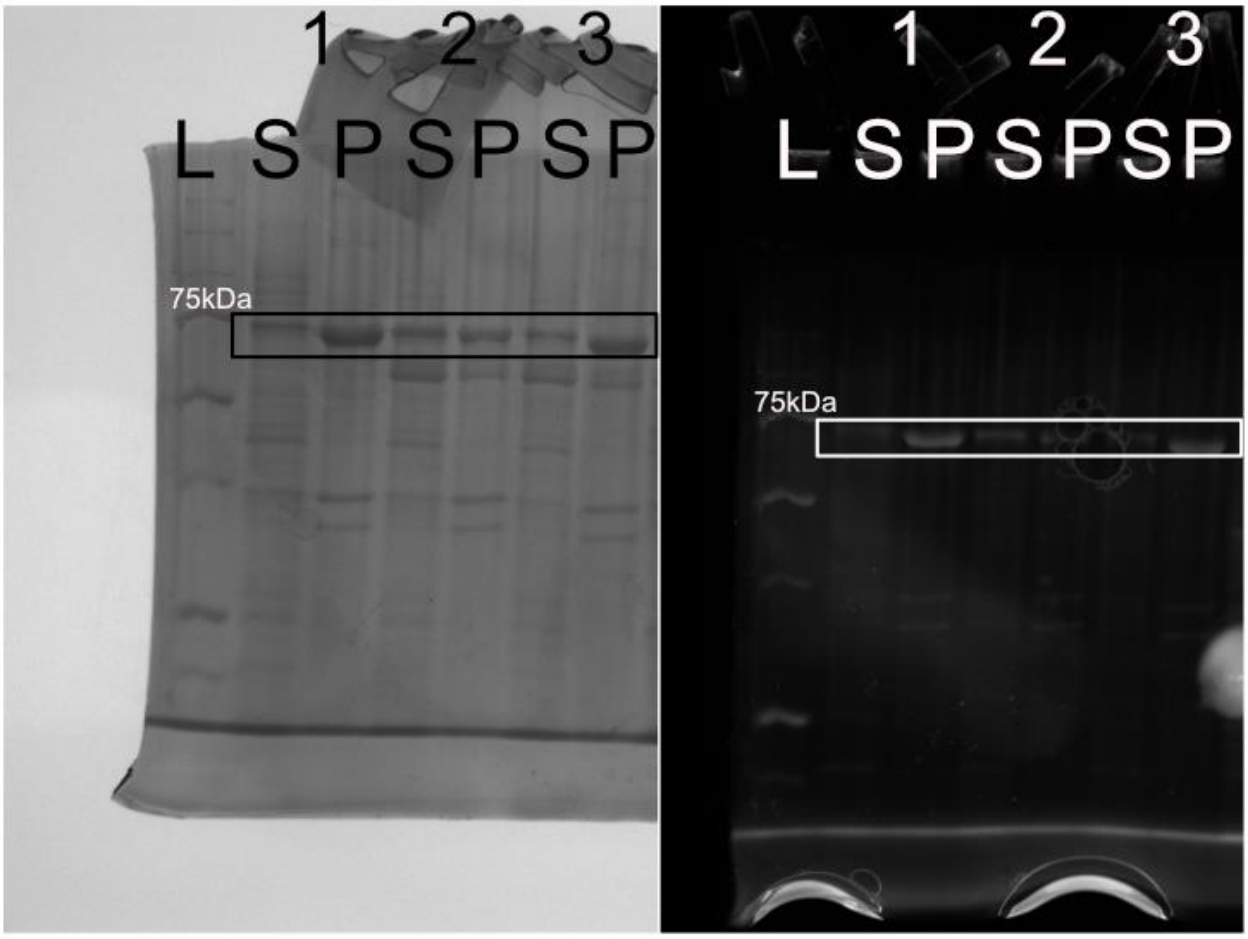
nitrogen and stored at – 80 °C until use. AetF concentration was determined using the Pierce BCA Protein Assay Kit.

For lyophilization, a 1 L culture of AetF was grown and separated into two tared cell buckets in equal weight, about 470 g each. One cell pellet was then lysed using the microfluidizer according to the general protocol, clarified, and diluted to a final volume of 100 mL in HEPES buffer, pH 7.4, 25 mM. This clarified lysate was then poured into 4 x 100 mL recrystallization trays such that the total depth of lysate in the tray was no more than 5 cm. The trays were then fit with a cheese cloth which was wrapped around the tray with a rubber band. The recrystallization trays were then placed into a freezer set to -80 °C until frozen solid, about two hours. The frozen clarified lysate was then placed into a lyophilizer under vacuum and for 72 hours, at which point the water was visibly drained from solution and the vacuum was released. The resulting lyophilized powder was placed into 4 tared 50 mL falcon tubes (16.5 g total lyophilized powder) and stored at – 20 °C until needed.

D) SDS page gel

SDS page gels were taken from cultures expressed in 2 mL microtiter plates according to the analytical scale biocatalysis protocol. Following clarification of cell lysate from centrifugation, the supernatant from the microtiter plate containing soluble enzyme was aliquoted into a 1.5 mL Eppendorf tube. The insoluble fraction was then suspended in the same amount of buffer used for cell lysis (100 µL) and resuspended by shaking vigorously on a plate shaker for 30 minutes (750 rpm, 25 °C). From each sample, 5 µL was taken and added to a PCR strip containing 10 µL water and 5 µL of 4 x Laemmli buffer. The samples were then incubated at 90 °C for 10 minutes and loaded on to an SDS page gel along with protein standard ladder. After electrophoresis at 130 V for 100 minutes, the gel was removed and imaged first using the TCE stain, then Coomassie blue. The gel showing soluble and insoluble fractions of AetF is **Figure 6.5** below

Figure 6. 5: SDS page gels of AetF expression



Left: Coomasie stained image. Right: TCE stained image. L = Ladder, the mark corresponding to 75 kDa is labeled (AetF MW is 74 kDa), S = soluble fraction, P = insoluble pellet. 1 = AetF, 2 = AetF + pGro7, 3 = AetF (K258A) + pGro7. Interpretation: pGro7 boosts soluble expression of AetF, and while AetF K258A expresses less than the wild type enzyme, but is still soluble expressed.

6.4.3 Biocatalysis

General Bioconversion Conditions

Analytical-scale bioconversion and LC analysis were assembled by mixing stock solutions of substrate, small molecule, FDH, and cofactor regeneration system components. Substrate solutions were prepared in DMSO to give a final substrate concentration of 100 mM. A mixture of the small molecule components (NaX (X = Br, I), substrate, glucose, NADP, and FAD) was then prepared in reaction buffer. This small molecule mix was manually arrayed into 96-well plate or 1.5 mL Eppendorf tubes using a multi-channel pipette.

Enzyme mixture: A mixture of the cofactor regeneration components and enzymes of the AetF bioconversions (GDH, AetF, catalase and glutathione) was prepared in reaction buffer. The mix was manually arrayed into 96-well plate containing the small molecule mixture or 1.5 mL Eppendorf tubes using a multi-channel pipette.

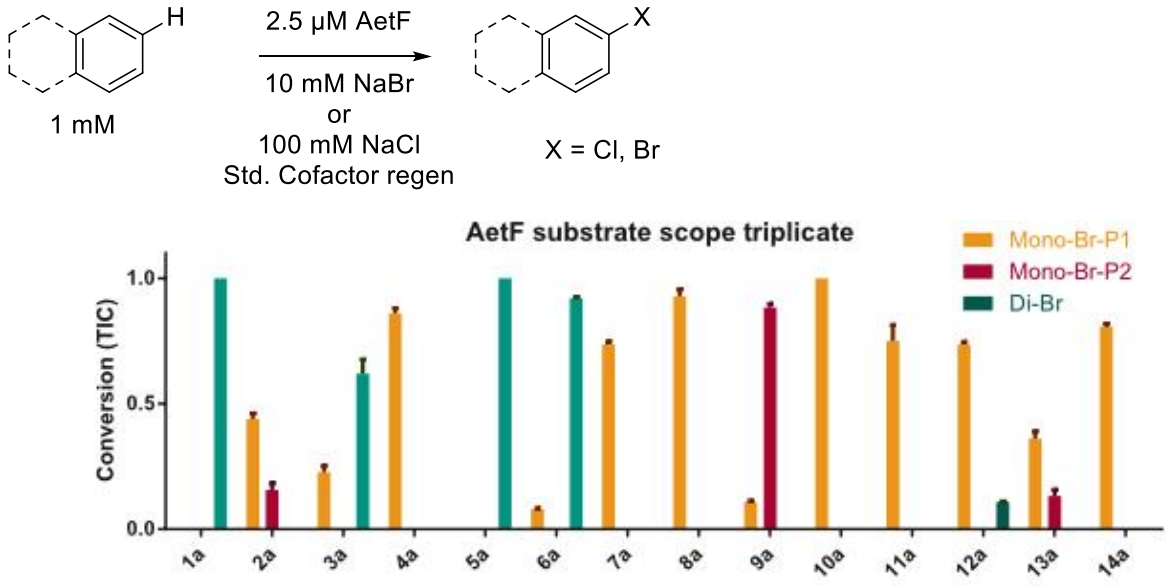
After all reactions had been prepared, the 1.5 mL Eppendorf tubes or 96-well plates sealed with adhesive film were shaken on a plate shaker at 650 rpm for different interval of time. The reactions were quenched and proteins precipitated by adding an equal volume of MeOH premixed with appropriate internal standards to each well of the 96-well reaction plate. Reaction plates were shaken for 5 min, and precipitated material was removed by centrifugation at 3600 rpm, 4 °C for 5 min. A multichannel pipette was then used to transfer 60 µL of from the reaction plates to a new 96-well filter plate retaining the well layout of the previous plate. The filter plate was fitted onto a 96-well V-bottom deep well plates, and the supernatant was filtered via centrifugation at 3600 rpm, 4 °C for 5 min. The collected filtrate was then sealed with aluminum foil and analyzed using LCMS.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). RF values, HPLC traces for the racemic products (needed to identify product enantiomers in bioconversions) and bioconversions are provided in the appropriate section.

Analytical-scale AetF reactions

Purified AetF was used for initial evaluation of the substrate scope and reactions were performed in triplicate. A solution of 125 µM AetF in storage buffer (HEPES pH 7.4, 25 mM with 10% glycerol) was thawed on ice. A stock solution of AetF was prepared at 37.5 uM, and 50 µL of this solution was aliquoted into a microtiter plate. A solution of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr or NaI or 500 mM NaCl, and 5 mM of substrate was prepared in reaction buffer (small molecule stock solution). A solution of 67.5 U/mL GDH and 262.5 U/mL catalase was prepared in reaction buffer (enzyme stock solution). 15 µL of the small molecule stock solution was added to the 96 well microtiter plates containing clarified lysate. 10 µL of the enzyme stock solution was then added to give a total reaction volume of 75 µL and a final AetF concentration of 25 µM. The plate was sealed using an aluminum heat seal and incubated in ambient atmosphere using a plate shaker set to 600 rpm. The next day, reactions were quenched using 1 volume of methanol, and the quenched reactions were further

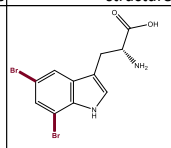
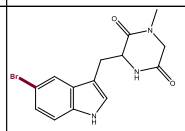
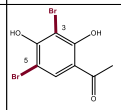
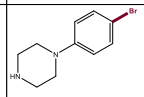
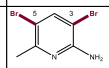
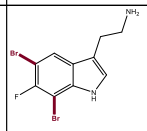
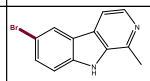
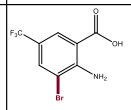
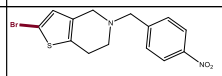
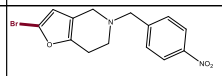
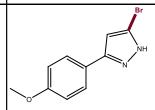
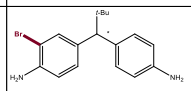
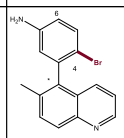
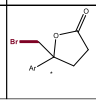
incubated for 15 minutes to facilitate full protein precipitation. Precipitated protein was removed from the solution via centrifugation at 3600 x rpm in a Sorvall XT centrifuge for 15 minutes. From the resulting supernatant, 50 µL was then added to a 0.2 µM filter plate charged with 150 µL of water from a Milli-Q filtration system, and filtration was done gravimetrically at 3600 x rpm in a Sorvall XT centrifuge for 15 minutes into a fresh microtiter plate. The microtiter plate containing filtered reactions was then heat-sealed using an aluminum heat-seal and analyzed via UHPLC-(MS) using the appropriate methods as given in the methods section of this SI. Conversion is defined as % AUC of the product(s) as determined by m/z value compared to the % AUC of the total area of starting material + product(s). Full characterization of all brominated and iodinated products presented in the manuscript is provided below.

Figure 6. 6: Conversion data used to evaluate AetF substrate scope



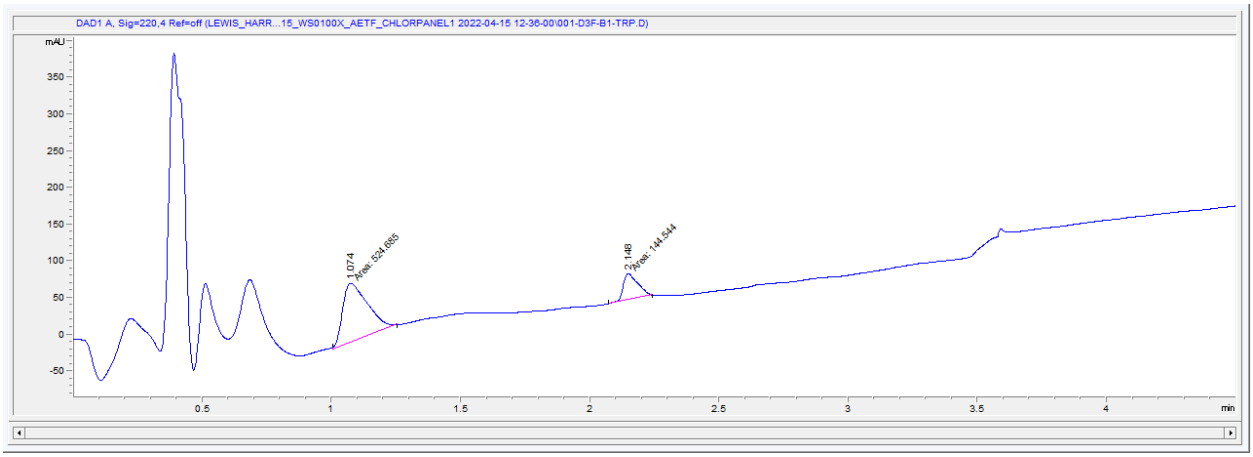
Conversion was evaluated as the AUC of the brominated product(s) on the TIC over the total AUC of substrate and all products. SOURCE: *Angewandte Chemie* 2022.

Table 6. 1: Conversion data from Figure 6.2

Name	structure	Mono-Br-P1			Mono-Br-P2			Di-Br		
1a		0	0	0	0	0	0	1	1	1
2a		0.463458	0.43486	0.41934	0	0	0	0.147059	0.134282	0.18672
3a		0.256637	0.210526	0.212955	0	0	0	0.556718	0.645559	0.658385
4a		0.845746	0.883801	0.850575	0	0	0	0	0	0
5a		0	0	0	0	0	0	1	1	1
6a		0.077953	0.083724	0.07549	0	0	0	0.922047	0.916276	0.92451
7a		0.723711	0.733716	0.74812	0	0	0	0	0	0
89a		0.93266	0.898734	0.953668	0	0	0	0	0	0
9a		0.113823	0.104082	0.106656	0.886177	0.895918	0.868083	0	0	0
10a		1	1	1	0	0	0	0	0	0
11a		0.821711	0.726323	0.704116	0	0	0	0	0	0
12a		0.737542	0.723246	0.744746	0	0	0	0.108154	0.106396	0.110265
13a		0.332467	0.356144	0.392274	0.158618	0.129219	0.109916	0	0	0
14a		0.818509	0.810115	0.796813	0	0	0	0	0	0

Data from each individual replicate is represented. SOURCE: *Angewandte Chemie* 2022.

Figure 6. 7: LC/MS chromatogram for chlorination of tryptophan by AetF



#	Time	Area	Height	Width	Area%	Symmetry
1	1.074	524.7	80.2	0.109	78.401	0.505
2	2.148	144.5	36.6	0.0659	21.599	0.528

Integration was performed at 220 nm and the relative % area is shown. Chlorination was not observed for any other substrates in Table 6.1. SOURCE: *Angewandte Chemie* 2022.

Analysis of the K258A mutation on bromination activity of AetF and discovery of iodination

During cell lysis, the lysis buffer used for verification of the catalytic lysine residue was HEPES, pH 7.4, 25 mM. 50 µL of the clarified lysate containing either wt-AetF or the K258A variant were aliquoted into a microtiter plate in triplicate. A solution of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM of NaBr or NaI and 5 mM of L-tryptophan was prepared in reaction buffer (small molecule stock solution). A solution of 67.5 U/mL GDH and 262.5 U/mL catalase was prepared in reaction buffer (enzyme stock solution). 15 µL of the small molecule stock solution was added to the 96 well microtiter plates containing clarified lysate. 10 µL of the enzyme stock solution was then added to give a total reaction volume of 75 µL. These reactions were incubated at 25 °C in a Thermo Scientific plate mixer (Eppendorf Thermomixer R) at 750 rpm for 20 hours. The next day, reactions were quenched using 1 volume of methanol, and the quenched reactions were further incubated for 15 minutes to facilitate full protein precipitation. 100 µL of this solution was transferred to a 1 mL culture plate (Waters, part # 186002481) and precipitated protein was removed from the solution via centrifugation at 2000 x g rpm in a Sorvall XT centrifuge for 15 minutes. From the resulting supernatant, 50 µL was then added to a 0.2 µM filter plate (OASIS, 186002794) charged with 150 µL of water from a Milli-Q filtration system, and filtration was done gravimetrically at 3600 x rpm in a Sorvall XT centrifuge for 15 minutes into a fresh microtiter plate. The microtiter plate containing filtered reactions was then fitted with a mat (PP cap, part # 186002483) using an aluminum heat-seal and analyzed via UHPLC using method 1. Conversion is defined as % AUC of the product(s) as determined by m/z value compared to the % AUC of the total area of starting material + product(s).

Scheme 6. 1: Reaction to validate the putative catalytic lysine residue as essential for brominase and iodinase activity with AetF as shown in Figures 6.8 and 6.9

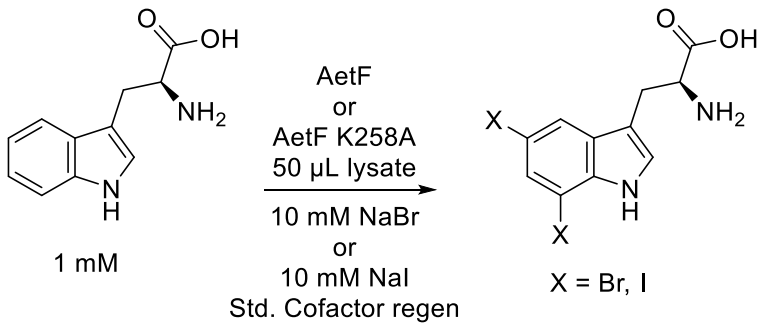
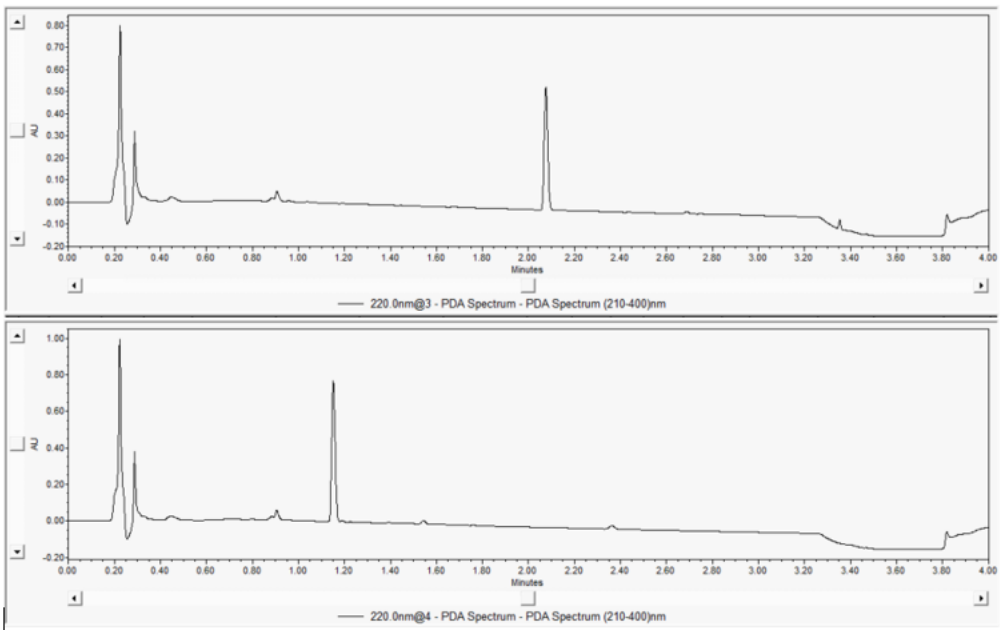
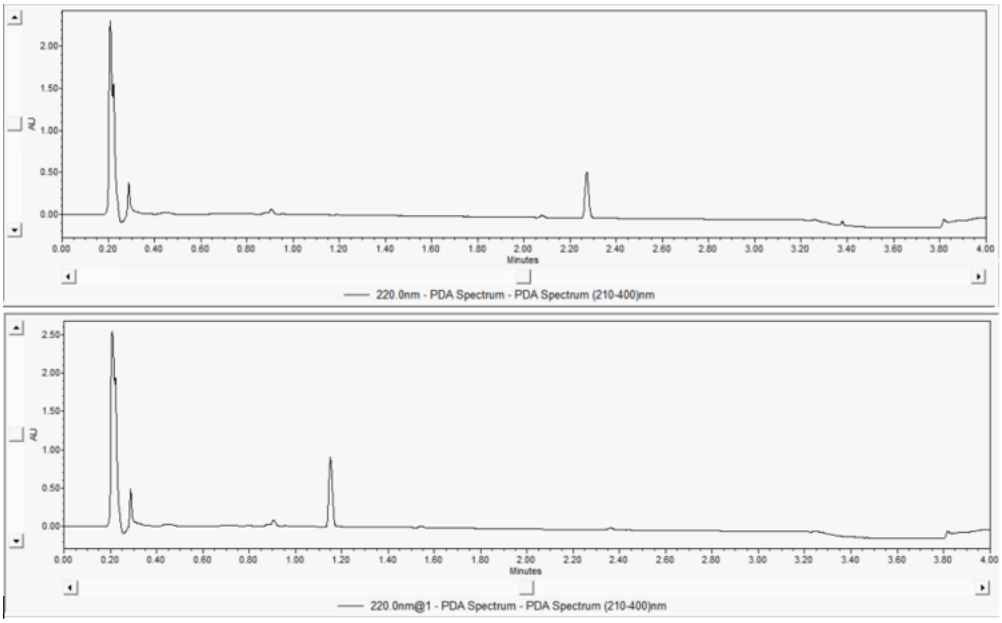


Figure 6. 8: LC/MS chromatogram showing bromination activity using AetF and AetF K258A



Top: Wild-type AetF showing full conversion to dibrominated L-Tryptophan. Bottom: Chromatogram showing no activity with AetF K258A variant. SOURCE: *Angewandte Chemie* 2022.

Figure 6. 9: LC/MS chromatogram showing iodination activity using AetF and AetF K258A



Top: Wild-type AetF showing full conversion to diiodinated L-Tryptophan. Bottom: Chromatogram showing no activity with AetF K258A variant. SOURCE: *Angewandte Chemie* 2022.

Buffer optimization studies to facilitate preparative halogenation

Study 1: This study was meant to identify a buffer concentration and pH that improved results in lysate compared to our standard FDH biocatalysis conditions.

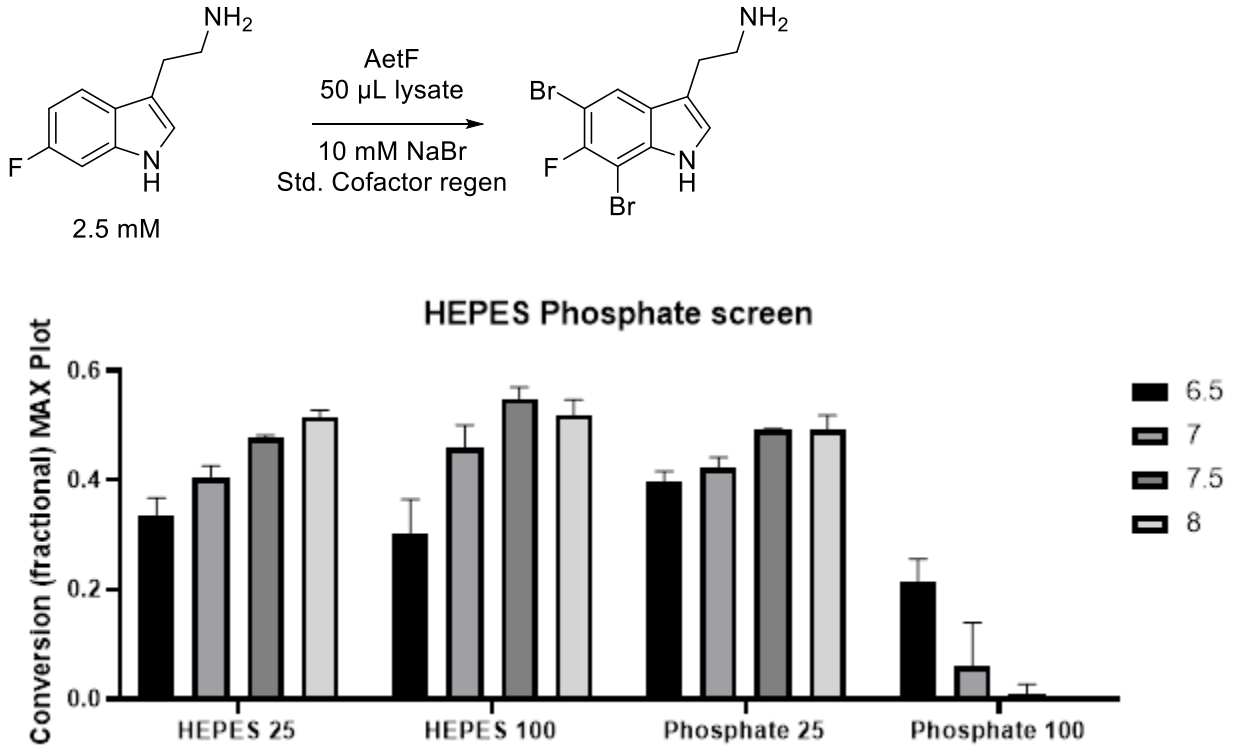
The layout of the plate was as follows:

Buffer	Concentration	Plate	1	2	3	4	5	6	7	8	9	10	11	12
HEPES	25 mM	A	pH 6.5			pH 7.0			pH 7.5			pH 8.0		
HEPES	100 mM	B	pH 6.5			pH 7.0			pH 7.5			pH 8.0		
Phosphate	25 mM	C	pH 6.5			pH 7.0			pH 7.5			pH 8.0		
Phosphate	100 mM	D	pH 6.5			pH 7.0			pH 7.5			pH 8.0		

During cell lysis, lysis buffer was prepared as a 10x stock solution of lysozyme and benzonase I in HEPES pH 7.4, 5 mM (7.5 mg/mL lysozyme, 10 uL/mL Benzonase I). 150 µL of this stock solution was then aliquoted into 1.35 mL of the appropriate pH and concentration of either HEPES or phosphate. From the 100 µL of clarified lysate, 50 µL was aliquoted into a micro titer plate. A solution of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 12.5 mM of 6-fluorotryptamine in DMSO (50 mM stock solution, 5% v/v DMSO final concentration) was prepared in reaction buffer (small molecule stock solution). A solution of 67.5 U/mL GDH and 262.5 U/mL catalase was prepared in reaction buffer (enzyme stock solution). 15 µL of the small molecule stock solution was added to the 96 well microtiter plates containing clarified lysate. 10 µL of the enzyme stock solution was then added to give a total reaction volume of 75 µL and a final AetF concentration of 25 µM. The plate was sealed using an aluminum heat seal and incubated in ambient atmosphere using a plate shaker set (Eppendorf Thermomixer R) to 600 rpm. The next day, reactions were quenched using 1 volume of methanol, and the quenched reactions were further incubated for 15 minutes to facilitate full protein precipitation. Precipitated protein was removed from the solution via centrifugation at 3600 x rpm in a Sorvall XT

centrifuge for 15 minutes. From the resulting supernatent, 50 µL was then added to a 0.2 µm filter plate charged with 150 µL of water from a Milli-Q filtration system, and filtration was done gravimetrically at 3600 x rpm in a Sorvall XT centrifuge for 15 minutes into a fresh microtiter plate. The microtiter plate containing filtered reactions was then heat-sealed using an aluminum heat-seal and analyzed via UHPLC using method 1. Conversion is defined as % AUC of the product(s) as determined by MAX PLOT values compared to the % AUC of the total area of starting material + product(s).

Figure 6. 10: Results of the first optimization study



Interpretation: AetF conversion is highest in HEPES, 100 mM at pH 7.5. SOURCE: *Angewandte Chemie* 2022.

Study 2: This study was meant to build off of the previous report and screen more basic buffer pHs and Tris-HCl which has a higher pKa than HEPES.

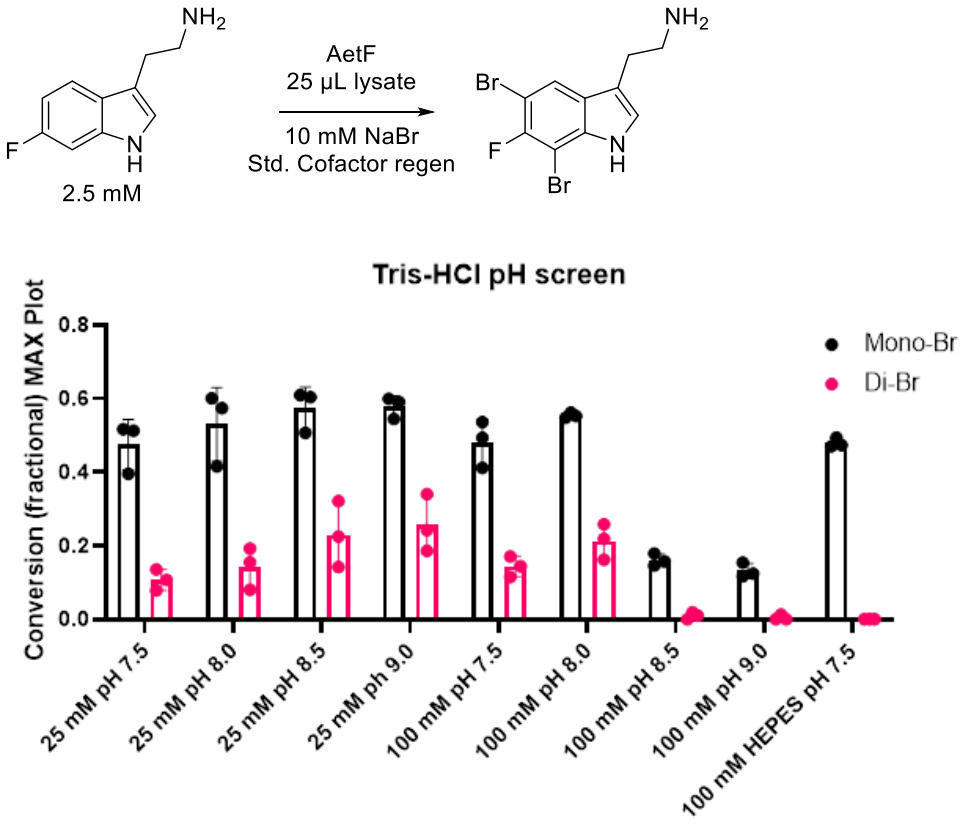
During cell lysis, lysis buffer was prepared as a 10x stock solution of lysozyme and benzonase I in HEPES pH 7.4, 5 mM (7.5 mg/mL lysozyme, 10 uL/mL Benzonase I). 150 µL of this stock solution was then aliquoted into 1.35 mL of the appropriate pH and concentration of either HEPES or phosphate. From the 100 µL of clarified lysate, 25 µL was aliquoted into a micro titer plate followed by 25 µL of the appropriate reaction buffer. Stock solutions of the appropriate composition were comprised as follows:

The final layout of the microtiter plate was as follows. Reactions were conducted in triplicate.

Concentration	Plate	1	2	3	4	5	6	7	8	9	10	11	12
25 mM	A	pH 7.5			pH 8.0			pH 8.5			pH 9.0		
100 mM	B	pH 7.5			pH 8.0			pH 8.5			pH 9.0		
		HEPES, pH 7.5, 100 mM											

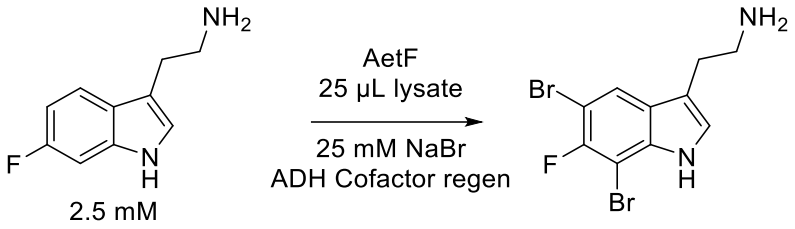
15 µL of the small molecule stock solution was added to the 96 well microtiter plates containing clarified lysate. 10 µL of the enzyme stock solution was then added to give a total reaction volume of 75 µL and a final AetF concentration of 25 µM. The plate was sealed using an aluminum heat seal and incubated in ambient atmosphere using an Eppendorf thermomixer R plate shaker set to 600 rpm. The next day, reactions were quenched using 1 volume of methanol, and the quenched reactions were further incubated for 15 minutes to facilitate full protein precipitation. Precipitated protein was removed from the solution via centrifugation at 3600 x rpm in a Sorvall XT centrifuge for 15 minutes. From the resulting supernatent, 50 µL was then added to a 0.2 µM filter plate charged with 150 µL of water from a Milli-Q filtration system, and filtration was done gravimetrically at 3600 x rpm in a Sorvall XT centrifuge for 15 minutes into a fresh microtiter plate. The microtiter plate containing filtered reactions was then heat-sealed using an aluminum heat-seal and analyzed via UHPLC using method 1. Conversion is defined as % AUC of the product(s) as determined by MAX PLOT value compared to the % AUC of the total area of starting material + product(s).

Figure 6. 11: Results of the second optimization study



Interpretation: AetF has the highest activity in Tris pH 9.0, 25 mM. SOURCE: *Angewandte Chemie* 2022.

Study 3: This study was meant to probe if the improvements observed with more basic pH was a result of enhanced enzyme activity or by preventing acidification of the media. To probe this, the GDH cofactor regeneration system was replaced with one driven by alcohol dehydrogenase (ADH-101) and IPA.



The layout of the microtiter plate was as follows:

ADH (final conc mg/mL)	Row	1	2	3	4	5	6
0.01	A	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3
0.1	B	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3
0.5	C	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3
Buffer		HEPES pH 7.4, 25 mM			Tris-HCl pH 9.0, 25 mM		

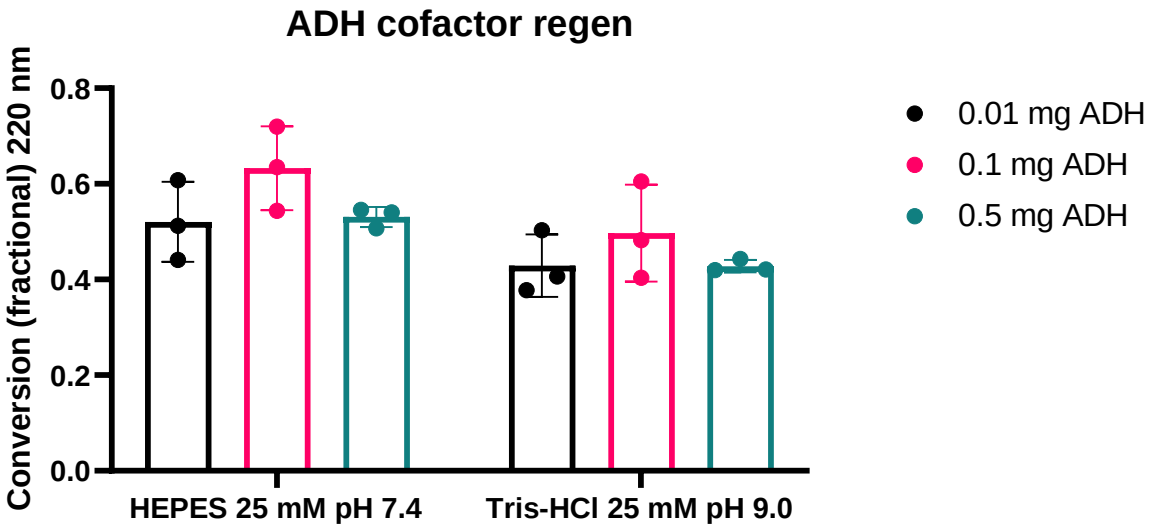
Cell pellets were lysed as described in the general procedure in lysis buffer (either HEPES, pH 7.4, 25 mM or Tris-HCl, pH 9.0, 25 mM with 0.1 mg/mL lysozoyme and 1 µL/mL benzonase I). After cell lysis, 25 µL of lysate was added to a microtiter plate and further diluted with 25 µL of the appropriate buffer according to the layout of the microtiter plate. Two small molecule mixes were made, one with the diluting buffer being HEPES, and the other diluted with Tris-HCl. Otherwise, the small molecule mix components were the same. Three enzyme mixes were prepared, each with a 100x stock solution of ADH. The enzyme mix was prepared in HEPES buffer to simplify reaction set up. The substrate 6-fluorotryptamine was dissolved in IPA as a 25 mM stock solution (final concentration 2.5 mM 6-F tryptamine, 10% v/v IPA). The reaction mixes were constructed as follows:

Small molecule mix			
Component	Stock	Final	Volume (µL)
Buffer	25 mM	25 mM	43.2
NaBr	1.5 mM	25 mM	13.5
FAD	10 mM	0.1 mM	8.1
NADP	10 mM	0.1 mM	8.1
Glutathione	100 mM	1 mM	8.1
Substrate	25 mM	2.5 mM	81

enzyme mix			
Component	Stock	Final	Volume (µL
Buff	25 mM	25 mM	41
ADH	Various	Various	6.8
Cat	2000 U/mL	70 U/mL	24

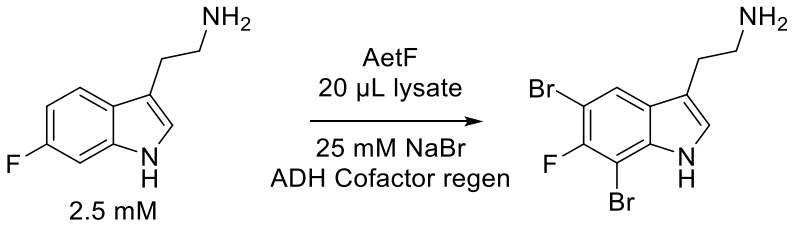
Reactions were initiated by the addition of 15 µL substrate mix followed by 10 µL enzyme mix. The reactions were then covered with a breathable plate seal and allowed to incubate at 650 rpm for 20 hours in a plate mixer set to 25 °C. The following day, reactions were quenched by the addition of 1 volume of methanol and the plate was processed as previously described. The plate was analyzed using UHPLC method 1.

Figure 6. 12: ADH buffer study with AetF lysate



Interpretation: When alcohol dehydrogenase is used as a cofactor regeneration system, more basic conditions to not positively impact AetF activity.

Study 4: This study was meant to optimize the buffer concentration and %v/v IPA for biocatalysis with AetF.



A 96 well plate containing AetF cultures was prepared according to the general protocol for AetF lysate reactions. The plate layout was as follows:

pH	Conc		1	2	3	4	5	6	7	8	9	10	11	12
6.5	25	A												
6.5	100	B												
7	25	C												
7	100	D												
7.5	25	E												
7.5	100	F												
8	25	G												
8	100	H												
		%v/v IPA	1% IPA			2.5 % IPA			5% IPA			10% IPA		

A 20x stock solution of lysozyme and DNAse (15 mg/mL lysozyme, 20 µL/mL benzonase I in 5 mM HEPES, pH 7.4) was prepared and diluted 1/20 into the appropriate buffer such that 8 lysis buffers were prepared (i.e. pH 6.5, 25 mM; pH 6.5, 100 mM; pH 7.0, 25 mM, etc.), and clarified lysate was prepared according to the general protocol. 20 µL of clarified lysate was added to a microtiter plate followed by 28.8 µL of buffer at the appropriate pH and concentration, followed by 1.88 µL of substrate (100 mM

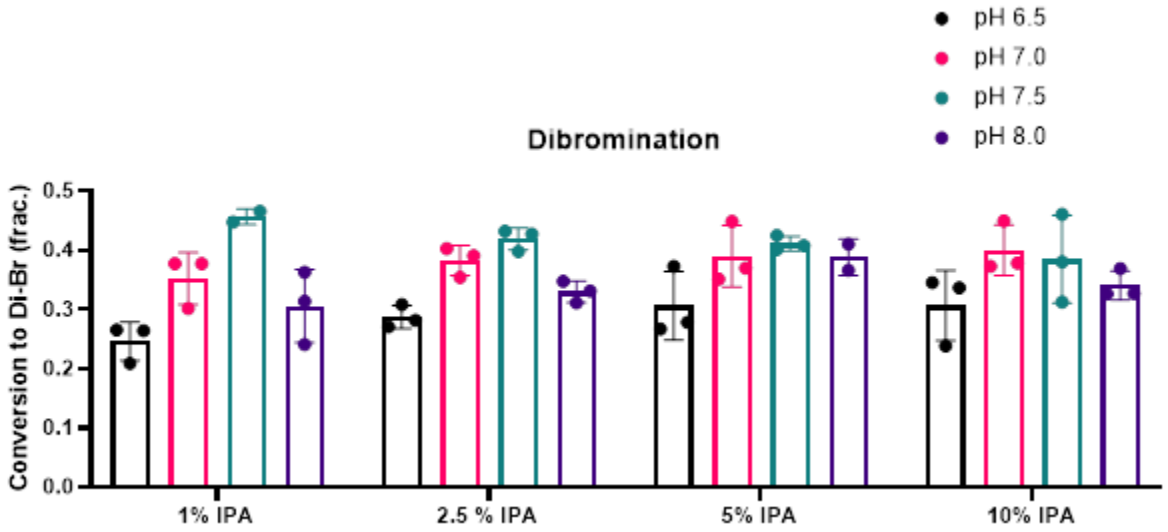
stock concentration, 2.5 mM final concentration, 2.5 % v/v DMSO). Four small molecule stock solutions were prepared using 25 mM of the appropriate pH buffer according to the table below:

Small molecule mix			
Component	Stock	Final	Volume (µL)
HEPES, various pH	25 mM	25 mM	331
NaBr or NaI	1.5 M	10 mM	36
NADP	10 mM	0.1 mM	21.6
FAD	10 mM	0.1 mM	21.6
Glutathione	100 mM	1 mM	21.6

enzyme mix			
Component	Stock	Final	Volume (µL)
HEPES, various pH	25 mM	25 mM	41
ADH	10 mg/mL	0.1 mg/mL	6.8
Cat	2000 U/mL	70 U/mL	24

Reactions were initated by the addition of IPA to the reaction according to the volumes on the table (0.75 µL for 1% IPA, 1.88 µL for 2.5% IPA, 3.75 µL for 5% IPA, 7.5 µL for 10% IPA). Reactions were left to incubate at 25 °C and 650 rpm overnight on a plate shaker. The next day, the reactions were quenched and processed, and analyzed. as previously described. Note that because the overall conversion was higher than expected, only the conversion to the dibrominated material is reported.

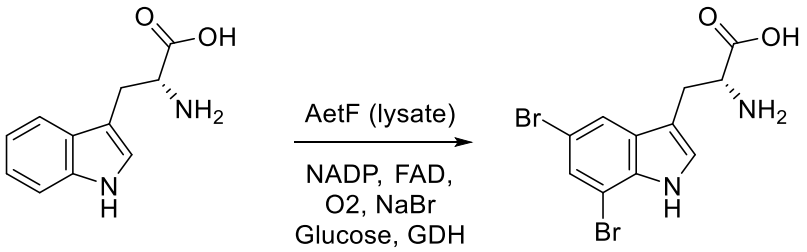
Figure 6. 13: IPA optimization



Interpretation: 1% IPA at pH 7.5 gives the highest conversion, although the % IPA does not seem to have a large impact on the outcome of the reaction.

6.4.4 – Preparative bioconversions

Bioreactor reaction 1: Lysate reaction of AetF with closed system

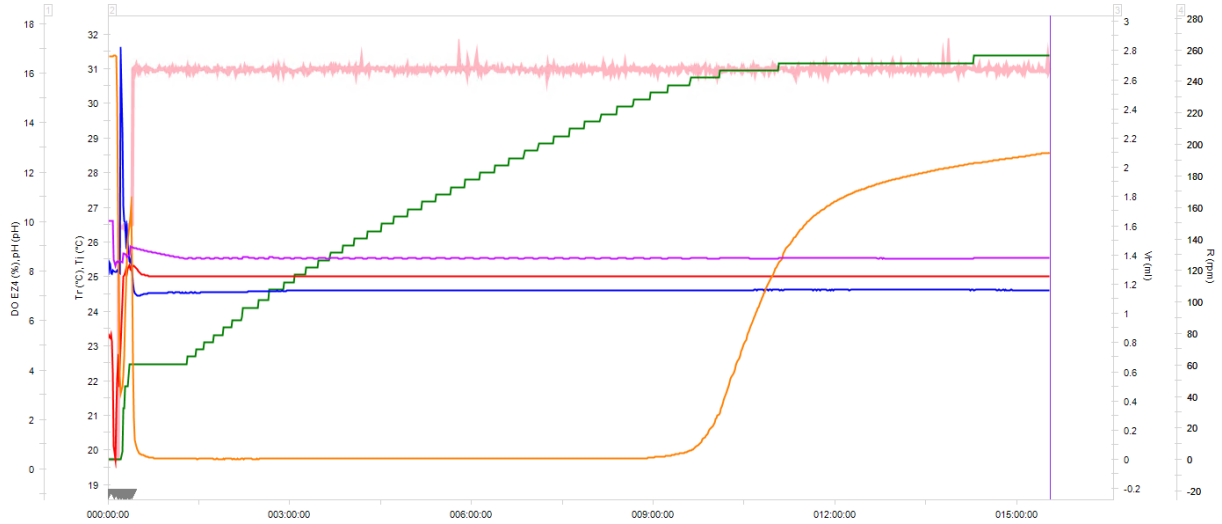


A 1 L culture of AetF was grown and split evenly into two tared centrifuge buckets (462.4 g each of cell culture) and centrifuged, at 15,000 x g for 45 minutes. One cell pellet was stored at – 80 °C, and the other was immediately lysed using the microfluidizer according to the general protocols. The lysate was clarified according to the general protocol and diluted to a final volume of 105 mL using Tris-HCl pH 8.0, 25 mM. This diluted lysate was then added to the EasyMax bioreactor along with the reaction components according to the following table:

Reaction mixture			
Component	Stock	Final	Volume (mL)
Lysate + Buffe	25 mM	25 mM	105
NaBr	1.5 M	25 mM	2.2
Glucose	1M	40 mM	5.2
NADP	10 mM	0.1 mM	1.3
FAD	10 mM	0.1 mM	1.3
Glutathione	100 mM	1 mM	1.3
Substrate	50 mM	2.5 mM	6.5
GDH	180 U/mL	9 U/mL	3.25
Catalase	2000 U/mL	35 U/mL	4.55
		Total volume	130.6

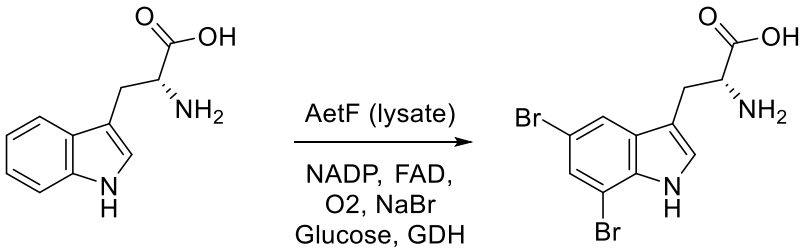
The bioreactor was set up with a pH probe, a line for the addition of base to keep pH constant (pH 8.5), a temperature probe to keep the temperature held to 25 °C consistently, a stirrer held to 250 rpm, and the dissolved oxygen probe. The reaction was initiated by the addition of GDH and allowed to stir overnight. A sixth port in the bioreactor lid was sealed using a glass stopper and the lid was fitted onto the bioreactor vessel using an O-ring creating a pseudo-closed system.

Figure 6. 14: Bioreactor output 1



Pink line: Stir rate, Green line: Addition of 2M NaOH to maintain pH, Purple line: pH, Red line: temperature, Orange line: dissolved oxygen, Blue line: temperature of probe (not used). Upon the addition of the reaction components, a significant drop in the dissolved oxygen content of the solution was observed, from 16% down to 0.5% where it was held until roughly 12 hours into the reaction at which point a gradual climb of the DO was observed. The addition of base to maintain pH resulted in a total base addition of 2 mL (0.0 mol OH-), corresponding to a total glucose consumption of 78%. Base addition slowed upon the saturation of oxygen into the solution, perhaps indicating the enzyme has stopped turning over. Conversion from this reaction as measured by UHPLC was found to be 21% 5-Br-tryptophan and 26% 5,7-diBr-tryptophan.

Bioreactor Reaction 2: Lysate reaction of AetF with air sparge

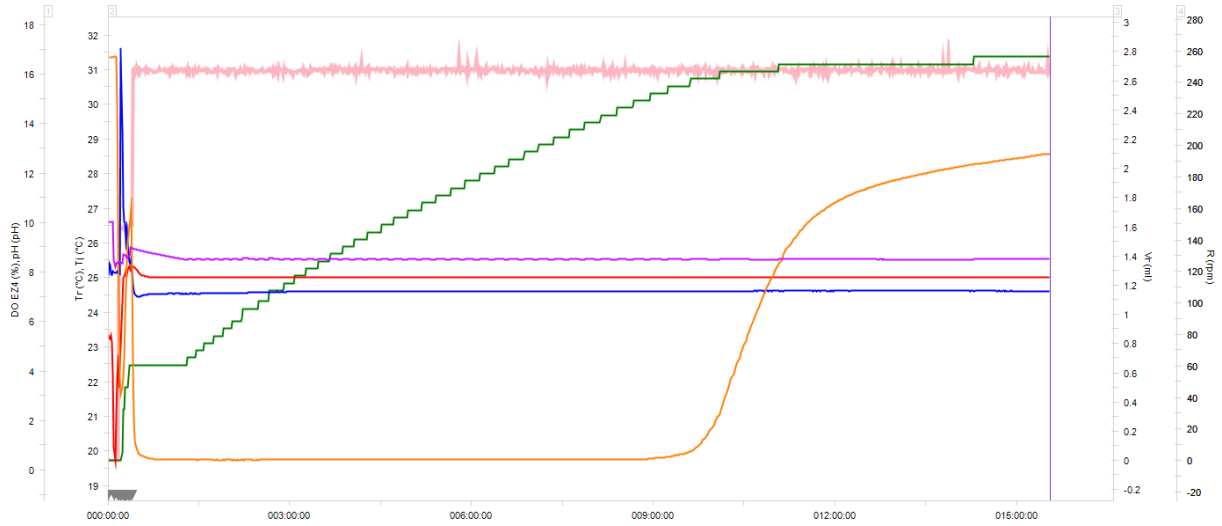


The remaining cell pellet from the previous section (bioreactor reaction 1) was used for this reaction and prepared in the same way. To ensure the additional airflow was fully submerged into solution, the volume of the solution was increased to 150 mL total. The reaction was set up according to the table below:

Reaction mixture			
Component	Stock	Final	Volume (mL)
Lysate + Buffe	25 mM	25 mM	122
NaBr	1.5 M	25 mM	2.5
Glucose	1M	40 mM	6
NADP	10 mM	0.1 mM	1.5
FAD	10 mM	0.1 mM	1.5
Glutathione	100 mM	1 mM	1.5
Substrate	50 mM	2.5 mM	7.5
GDH	180 U/mL	9 U/mL	3.75
Catalase	2000 U/mL	35 U/mL	7.5
		Total volume	153.75

The bioreactor was set up with a pH probe, a line for the addition of base to keep pH constant (pH 8.0), a temperature probe to keep the temperature held to 25 °C consistently, a stirrer held to rpm, and the dissolved oxygen probe. A sixth port in the bioreactor lid was used to accommodate the air sparge which was equipped with a diffuser to increase the surface area of the air bubbling through solution and set to 17.85 sec/min flow rate. Note that compressed air was used, not compressed oxygen for bubbling. The reaction was initiated by the addition of GDH. After the addition of GDH, a 1 mL aliquot of the reaction was stored in an Eppendorf tube, then aliquoted into 3 x 75 µL aliquots in a microtiter plate, which was then fit with a breathable plate seal and allowed to incubate overnight. With all six ports occupied, the lid was fit onto the bioreactor vessel using an O-ring creating a pseudo-closed system, and the reaction was allowed to stir overnight.

Figure 6. 15: Bioreactor output 2

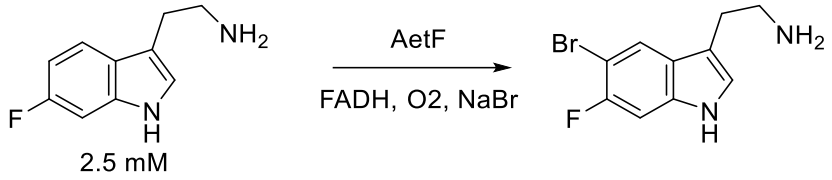


Upon the addition of the reaction components, a significant drop in the dissolved oxygen content of the solution was observed, from 16% down to 0.5% where it was held until roughly 12 hours into the reaction at which point a gradual climb of the DO was observed. The addition of base to maintain pH resulted in a total base addition of X mL (0.0x mol OH-), corresponding to a total glucose consumption of X%. Base addition slowed upon the saturation of oxygen into the solution, perhaps indicating the

enzyme has stopped turning over. Conversion from this reaction as measured by UHPLC was found to be 10%%. The aliquoted samples from this reaction were all observed to be fully converted to the 5,7-dibromotryptophan product, indicating that both in comparison to the closed system (bioreactor reaction 1) and the microtiter plate, an air sparge was detrimental to enzyme activity.

30 mL reaction of AetF with lysate, lyophilized enzyme, and whole cells

A 1 L culture of AetF was grown according to the general procedure and split into 10 x 100 mL cultures in cell buckets and the cells were harvested by centrifugation at 3600 x rpm for 20 minutes, at which point the supernatant was removed. Four cell buckets were resuspended in 10 mL each of HEPES buffer, pH 7.4, 25 mM and combined into 2 x 20 mL cell cultures. One cell culture was lysed using the microfluidizer according to the general protocol, the other was left as intact whole cells. For reactions with lyophilized powder, the powder was weighed into the beaker and diluted with buffer before addition of other reagents. Two concentrations of AetF lyophilized powder was tried, 5 mg/mL and 10 mg/mL corresponding to 150 mg and 300 mg of lyophilized AetF.



The reactions were set up as follows:

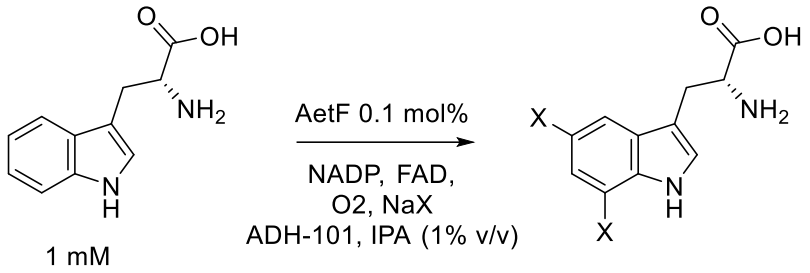
Reaction mixture			
Component	Stock	Final	Volume (mL)
Buffer + Lysate or Whole cells	25 mM	25 mM	26
NaBr	1.5 M	25 mM	0.5
IPA	N/A	1% v/v	0.3
NADP	10 mM	0.1 mM	0.3
FAD	10 mM	0.1 mM	0.3
Glutathione	100 mM	1 mM	0.3
Substrate	50 mM	2.5 mM	0.75
ADH	180 U/mL	9 U/mL	0.3
Catalase	2000 U/mL	35 U/mL	1.05
		Total volume	29.8

Reactions were initiated by the addition of 0.3 mL IPA to begin production of NADPH. After the addition of IPA, the reaction was shaken gently by hand to mix and 0.1 mL of the reaction was aliquoted into an Eppendorf tube. The 30 mL reactions were incubated at 160 rpm in a horizontal incubator held to 25 °C and the Eppendorf tubes were incubated in a plate shaker at 650 rpm held to 25 °C. The next day, 0.1 mL aliquots were taken from the 30 mL reactions and all reactions were quenched via the addition of 1 volume of methanol. Protein was removed from the solution via centrifugation and reactions were analyzed by UHPLC method 1. The results are summarized in **Table 6.2** below.

Table 6. 2: Results of 30 mL reactions with lysate, whole cell, and lyophilized powder

30 mL	0.17	0.28	0.17	0.18
Aliquot	0.2	0.6	DNI	DNI
	5 mg/m	Whole cell	5 mg/mL	10 mg/mL

30 mL reaction of AetF with purified enzyme – iodination and bromination



AetF was purified according to the general procedure. The reactions were set up according to the following table:

Reaction components			
Component	Stock	Final	Volume (mL)
HEPES, pH 7.4	25 mM	25 mM	26
NaBr or NaI	1.5 M	10 mM	0.2
IPA	N/A	1% v/v	0.3
NADP	10 mM	0.1 mM	0.3
FAD	10 mM	0.1 mM	0.3
Glutathione	100 mM	1 mM	0.3
Substrate	30 mM	1 mM	1
ADH	10 mg/mL	0.1 mg/mL	0.3
Catalase	2000 U/mL	70 U/mL	1.1
AetF	32 µM	1 µM	29.8

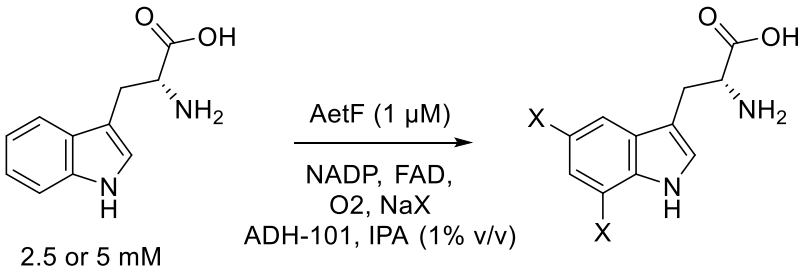
Reactions were initiated by the addition of 0.3 mL IPA to begin production of NADPH. After the addition of IPA, the reaction was shaken gently by hand to mix and 0.1 mL of the reaction was aliquoted into an Eppendorf tube. The 30 mL reactions were incubated at 160 rpm in a horizontal incubator held to 25 °C and the Eppendorf tubes were incubated in a plate shaker at 650 rpm held to 25 °C. The next day, 0.1 mL aliquots were taken from the 30 mL reactions and all reactions were quenched via the addition of 1 volume of methanol. Protein was removed from the solution via centrifugation and reactions were analyzed by UHPLC method 1. The results are summarized in Table 6.3 below.

Table 6. 3: Results of 30 mL reactions with NaBr, NaI

NaBr	17:80	26:45
NaI	0:100	0:100
	Aliquot	30 mL

Results are reported as (Mono-X:Di-X). Interpretation: AetF activity is less inhibited by scaling up when using NaI compared to NaBr. Additionally, the low enzyme loading and high substrate consumption for the aliquoted samples indicates that AetF is possibly capable of turning over more than 1,000 times, atypically high for flavin-dependent halogenases.

30 mL reaction at higher substrate loading - iodination



AetF was purified according to the general procedure. Two stock solutions of L-tryptophan were prepared, one at 50 mM and the other at 100 mM. The reactions were set up according to the following table:

Reaction mixture			
Component	Stock	Final	Volume (mL)
HEPES, pH 7.4	25 mM	25 mM	25
NaI	1.5 M	25 mM	0.5
NADP	10 mM	0.1 mM	0.3
FAD	10 mM	0.1 mM	0.3
Glutathione	100 mM	1 mM	0.3
Substrate	100 or 50 mM	2.5 or 5 mM	1.5
IPA	N/A	1% v/v	0.3
Catalase	2000 U/mL	70 U/mL	1.05
ADH	10 mg/mL	0.1 mg/mL	0.3
AetF	32 µM	1 µM	0.86
		Total volume	30.41

Reactions were initiated by the addition of 0.3 mL IPA to begin production of NADPH. After the addition of IPA, the reaction was shaken gently by hand to mix and 0.1 mL of the reaction was aliquoted into an Eppendorf tube. The 30 mL reactions were incubated at 160 rpm in a horizontal incubator held to 25 °C and the Eppendorf tubes were incubated in a plate shaker at 650 rpm held to 25 °C. The next day, 0.1 mL aliquots were taken from the 30 mL reactions and all reactions were quenched via the addition of 1 volume of methanol. Protein was removed from the solution via centrifugation and reactions were analyzed by UHPLC method 1. The results are summarized in **Table 6.4** below.

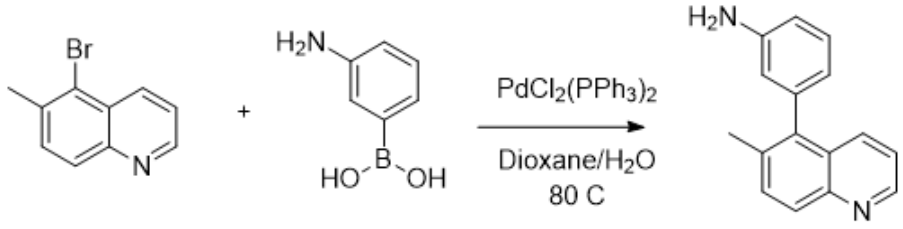
Table 6. 4: Results of reactions with 2.5 and 5 mM L-tryptophan

Substrate concentration	Mono-Br	Di-Br	Mono-Br	Di-Br
2.5	3.7	50	4.4	14
5	4.4	25.2	3.4	6.5
	Aliquot		Preparative	

Interpretation: While AetF still exhibits high TTN for iodination in analytical scale reactions (~54% conversion for 2.5 mM, 30% conversion at 5 mM), this effect largely is not observed for preparative reactions.

6.4.5 – Synthetic procedures

Scheme 6. 2: Quinoline 13:



To an oven-dried 50 mL round-bottom flask containing a stir-bar were added 3-aminophenyl boronic acid (HCl salt, 1.73 mmol, 300 mg, 1.3 equiv), 5-bromo-6-methylquinoline (1.35 mmol, 300 mg, limiting reagent), $PdCl_2(PPh_3)_2$ (0.04 mmol, 32 mg, 3 mol % Pd), and cesium carbonate (3.46 mmol, 1.127 g, 2.5 equiv.). The flask was then evacuated and purged with nitrogen 3 times and fitted with a condenser capped with a rubber septum. 15 mL of a 1:4 mixture of degassed dioxane:water was then added to the round bottom flask and a nitrogen balloon was added to the condenser. The solution was heated to 80 °C with vigorous stirring for 14 hours. The next day, the mixture was diluted with 1 volume of methanol, filtered through celite, and the solvent was removed via rotary evaporation under reduced pressure. The resulting residue was taken up in DCM and washed with 0.1 M NaOH, then brine, and the organic layer was dried over magnesium sulfate and the solvent was removed by rotary evaporation.

To purify the desired compound, silica gel was added to the resulting residue and the mixture was taken up in DCM and the compound was adsorbed onto the silica by removal of solvent. This silica gel was then added to a column presaturated with 1:1 Ethyl acetate: Hexanes with 1% TEA additive. NMR of the major component of the reaction showed an impurity from the Buchwald-Hartwig amination product, which was removed via reverse-phase chromatography using Biotage method 1. The desired product was obtained as a brown solid in 27% yield.

^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.74 (dd, *J* = 4.2, 1.7 Hz, 1H), 7.94 (d, *J* = 8.7 Hz, 1H), 7.76 (dt, *J* = 8.6, 1.2 Hz, 1H), 7.54 (d, *J* = 8.6 Hz, 1H), 7.22 – 7.19 (m, 1H), 7.16 (dd, *J* = 8.5, 4.2 Hz, 1H), 6.71 – 6.66 (m, 1H), 6.55 (dt, *J* = 7.5, 1.3 Hz, 1H), 6.48 (t, *J* = 1.9 Hz, 1H), 3.69 (s, 2H), 2.22 (s, 3H).^{13}C NMR (126 MHz, Chloroform-*d*) δ 149.13, 146.82, 146.49, 139.63, 138.40, 134.63, 133.74, 132.15, 129.46, 128.24, 127.88, 120.77, 120.44, 116.68, 114.13, 20.55.

6.5. References

(1) Breinlinger, S.; Phillips, T. J.; Haram, B. N.; Mareš, J.; Martínez Yerena, J. A.; Hrouzek, P.; Sobotka, R.; Henderson, W. M.; Schmieder, P.; Williams, S. M.; Lauderdale, J. D.; Wilde, H. D.; Gerrin, W.; Kust, A.; Washington, J. W.; Wagner, C.; Geier, B.; Liebeke, M.; Enke, H.; Niedermeyer, T. H. J.; Wilde, S. B. Hunting the Eagle Killer: A Cyanobacterial Neurotoxin Causes Vacuolar Myelinopathy. *Science (1979)* **2021**, *371* (6536). https://doi.org/10.1126/science.aax9050.

(2) Adak, S.; Lukowski, A. L.; Schäfer, R. J. B.; Moore, B. S. From Tryptophan to Toxin: Nature's Convergent Biosynthetic Strategy to Aetokthonotoxin. *J Am Chem Soc* **2022**, *144* (7), 2861–2866. https://doi.org/10.1021/jacs.1c12778.

(3) Bong, Y. K.; Song, S.; Nazor, J.; Vogel, M.; Widegren, M.; Smith, D.; Collier, S. J.; Wilson, R.; Palanivel, S. M.; Narayanaswamy, K.; Mijts, B.; Clay, M. D.; Fong, R.; Colbeck, J.; Appaswami, A.; Muley, S.; Zhu, J.; Zhang, X.; Liang, J.; Entwistle, D. Baeyer-Villiger Monooxygenase-Mediated Synthesis of Esomeprazole As an Alternative for Kagan Sulfoxidation. *Journal of Organic Chemistry* **2018**, *83* (14), 7453–7458. https://doi.org/10.1021/acs.joc.8b00468.

(4) Elleuche, S. Bringing Functions Together with Fusion Enzymes—from Nature's Inventions to Biotechnological Applications. *Appl Microbiol Biotechnol* **2015**, *99* (4), 1545–1556. https://doi.org/10.1007/s00253-014-6315-1.

(5) Andorfer, M. C.; Belsare, K. D.; Girlich, A. M.; Lewis, J. C. Aromatic Halogenation by Using Bifunctional Flavin Reductase–Halogenase Fusion Enzymes. *ChemBioChem* **2017**, *18* (21), 2099–2103. https://doi.org/10.1002/cbic.201700391.

(6) Agarwal, V.; el Gamal, A. A.; Yamanaka, K.; Poth, D.; Kersten, R. D.; Schorn, M.; Allen, E. E.; Moore, B. S. Biosynthesis of Polybrominated Aromatic Organic Compounds by Marine Bacteria. *Nat Chem Biol* **2014**, *10* (8), 640–647. https://doi.org/10.1038/nchembio.1564.

(7) Jiang, Y.; Snodgrass, H. M.; Zubi, Y. S.; Roof, C. v.; Guan, Y.; Mondal, D.; Honeycutt, N. H.; Lee, J. W.; Lewis, R. D.; Martinez, C.; Lewis, J. C. The Single Component Flavin Reductase/Flavin Dependent Halogenase AetF Is a Versatile Catalyst for Selective Bromination and Iodination of Arenes and Olefins. *Angewandte Chemie* **2022**. https://doi.org/10.26434/CHEMRXIV-2022-88XZV.

(8) Fisher, B. F.; Snodgrass, H. M.; Jones, K. A.; Andorfer, M. C.; Lewis, J. C. Site-Selective C-H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. *ACS Cent Sci* **2019**, *5* (11), 1844–1856. https://doi.org/10.1021/acscentsci.9b00835.

(9) Payne, J. T.; Andorfer, M. C.; Lewis, J. C. Regioselective Arene Halogenation Using the FAD-Dependent Halogenase RebH. *Angewandte Chemie - International Edition* **2013**, *52* (20), 5271–5274. https://doi.org/10.1002/anie.201300762.

(10) Andorfer, M. C.; Grob, J. E.; Hajdin, C. E.; Chael, J. R.; Siuti, P.; Lilly, J.; Tan, K. L.; Lewis, J. C. Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal* **2017**, *7* (3), 1897–1904. https://doi.org/10.1021/acscatal.6b02707.

(11) Ashtekar, K. D.; Marzijarani, N. S.; 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* (38), 13355–13362. https://doi.org/10.1021/ja506889c.

(12) Payne, J. T.; Butkovich, P. H.; Gu, Y.; Kunze, K. N.; Park, H. J.; Wang, D. S.; Lewis, J. C. Enantioselective Desymmetrization of Methylenedianilines via Enzyme-Catalyzed Remote Halogenation. *J Am Chem Soc* **2018**, *140* (2), 546–549. https://doi.org/10.1021/jacs.7b09573.

(13) Mondal, D.; Fisher, B. F.; Jiang, Y.; Lewis, J. C. Flavin-Dependent Halogenases Catalyze Enantioselective Olefin Halocyclization. *Nat Commun* **2021**, *12* (1). https://doi.org/10.1038/s41467-021-23503-3.

(14) Jiang, Y.; Mondal, D.; Lewis, J. C. Expanding the Reactivity of Flavin Dependent Halogenases Toward Olefins via Enantioselective Intramolecular Haloetherification and Chemoenzymatic Oxidative Rearrangements Yuhua. *ChemRxiv* **2022**, No. July, 1–6.

(15) Snodgrass, H. M.; Mondal, D.; Lewis, J. C. Directed Evolution of Flavin-Dependent Halogenases for Site- and Atroposelective Halogenation of 3-Aryl-4(3 H)-Quinazolinones via Kinetic or Dynamic Kinetic Resolution . *J Am Chem Soc* **2022**, *4*. https://doi.org/10.1021/jacs.2c07422.

(16) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(17) Nicoll, C. R.; Bailleul, G.; Fiorentini, F.; Mascotti, M. L.; Fraaije, M. W.; Mattevi, A. Ancestral-Sequence Reconstruction Unveils the Structural Basis of Function in Mammalian FMOs. *Nat Struct Mol Biol* **2020**, *27* (1), 14–24. https://doi.org/10.1038/s41594-019-0347-2.

(18) Warrens, A. N.; Jones, M. D.; Lechler, R. I. Splicing by Over-Lap Extension by PCR Using Asymmetric Amplification: An Improved Technique for the Generation of Hybrid Proteins of Immunological Interest. *Gene* **1997**, *186* (1), 29–35. https://doi.org/10.1016/S0378-1119(96)00674-9.

(19) Phintha, A.; Prakinee, K.; Jaruwat, A.; Lawan, N.; Visitsatthawong, S.; Kantiwiriyawanitch, C.; Songsungthong, W.; Trisrivirat, D.; Chenprakhon, P.; Mulholland, A.; van Pée, K. H.; Chitnumsub, P.; Chaiyen, P. Dissecting the Low Catalytic Capability of Flavin-Dependent Halogenases. *Journal of Biological Chemistry* **2021**, *296*, 100068. https://doi.org/10.1074/jbc.RA120.016004.

(20) Gkotsi, D. S.; Ludewig, H.; Sharma, S. v; Connolly, J. A.; Dhaliwal, J.; Wang, Y.; Unsworth, W. P.; Taylor, R. J. K.; McLachlan, M. M. W.; Shanahan, S.; Naismith, J. H.; Goss, R. J. M. A Marine Viral Halogenase That Iodinates Diverse Substrates. *Nat Chem* **2019**, *11* (12), 1091–1097. https://doi.org/10.1038/s41557-019-0349-z.

(21) Zhang, Y.; Chen, L.; Chen, H.; Huang, T.; Shi, Q.; Wang, X.; Wang, Y.; Tang, M.-C.; Zhou, N.-Y.; Lin, S. Aryl C-H Iodination: Are There Actual Flavin-Dependent Iodinases in Nature? *Sci China Chem* **2021**, *64* (10), 1730–1735. https://doi.org/10.1007/s11426-021-1018-0.

(22) Sambrook, J. *Molecular Cloning. A Laboratory Manual*; Third edition. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, [2001] ©2001, 1983; Vol. 11. https://doi.org/10.1016/0307-4412(83)90068-7.

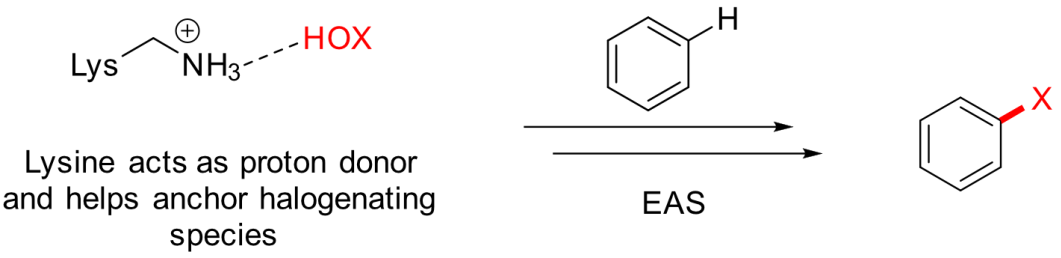
(23) Hogrefe, H. H.; Cline, J.; Youngblood, G. L.; Allen, R. M. Creating Randomized Amino Acid Libraries with the Quikchange® Multi Site-Directed Mutagenesis Kit. *Biotechniques* **2002**, *33* (5), 1158–1165. https://doi.org/10.2144/02335pf01.

CHAPTER 7: EXPLORING THE SURPRISING PLASTICITY OF THE CATALYTIC LYSINE RESIDUE IN THE REBH SCAFFOLD

7.1 Introduction

Halogenated natural products have attracted considerable interest due to their unique biological and medicinal value. Studies on these compounds have revealed that specific placement of halogen substituents is critical to their function.[1] This finding prompted efforts to identify enzymes capable of selective halogenation of diverse scaffolds found in these compounds. Early studies revealed that the heme-dependent enzyme chloroperoxidase, the first halogenating enzyme to be characterized, could oxidize chloride to HOCl using H_2O_2, but this species is believed to diffuse from the active site where it engages in non-selective chlorination.[2] Vanadium dependent haloperoxidases were later found to catalyze stereoselective halocyclization of their native terpene substrates and non-selective halocyclization and aromatic halogenation of non-native substrates. More recently, non-heme Fe(II)- and □-ketoglutarate dependent halogenases (FeDHs)[3] and flavin-dependent halogenases (FDHs)[4] were found to catalyze selective halogenation of sp3 and sp2 C-H bonds, respectively, using oxygen as a stoichiometric oxidant.

Figure 7. 1: Catalytic lysine residue activates HOX in FDH active site



The selectivity of FeDHs and FDHs inspired extensive efforts to establish the origins of their selectivity. In both cases, substrate positioning relative to the active halogenating species plays a key role in selectivity. While a reasonably strong consensus based on extensive mechanistic studies has emerged regarding the nature of this species FeDH catalysis,[3] two related hypotheses are commonly put forth to

rationalize FDH activity. Early studies identified a persistent haloamine species[5] and as such, it was originally assumed that the active halogenating species was a covalently formed chloramine species. More recently, the discovery of HOX release from the active site[6,7] along with computational models probing the feasibility of reactions for known enzyme-substrate pairs suggests that HOX formed from attack by halide on the hydroperoxy flavin intermediate is likely activated by hydrogen bonding of the lysine residue in the FDH active site (**Figure 7.1**). In both cases, this active site lysine is conserved in all known FDHs and likely plays a critical role as either the halogenating agent itself or an H-bond donor that activates HOX. Mutating this residue, typically to alanine, was universally reported to abolish native aromatic halogenase activity. More recently, however, our lab reported that halogenation of electron rich aromatic compounds and the FDH catalyzed halocyclization of 4-(4-methoxyphenyl)pent-4-enoic acid could be affected with alanine variants of RebH with good yields, but lacked the high selectivity of the lysine variant.[6] These results show that while an active site lysine residue is not strictly required for halogenase activity, it uniquely enables reaction of aromatic substrates that do not undergo direct halogenation by HOX and enantioselective halogenation.

Given this precedent, we were surprised to find that an engineered variant of the FDH RebH (6-TLP) exhibited improved activity and altered site selectivity toward 3-(3'-aminophenyl)-4(3*H*)-quinazolinone substrate **1** upon introduction of K79R.[8] This mutation was introduced via error prone PCR along with K279R and was identified based on the improved assay yield of the resulting the double mutant for the undesired product **1b** (**Figure 7.2 A**). After validating the 6-RR mutant by re-sequencing the gene, and verification that mutation of the surface lysine K279R did not significantly influence activity or selectivity (**Figure 7.2 B**) we next sought to explore how general the activity of the lysine to arginine mutation was in the RebH scaffold. At the time of writing, this work is currently in preparation for submission.

Authorship

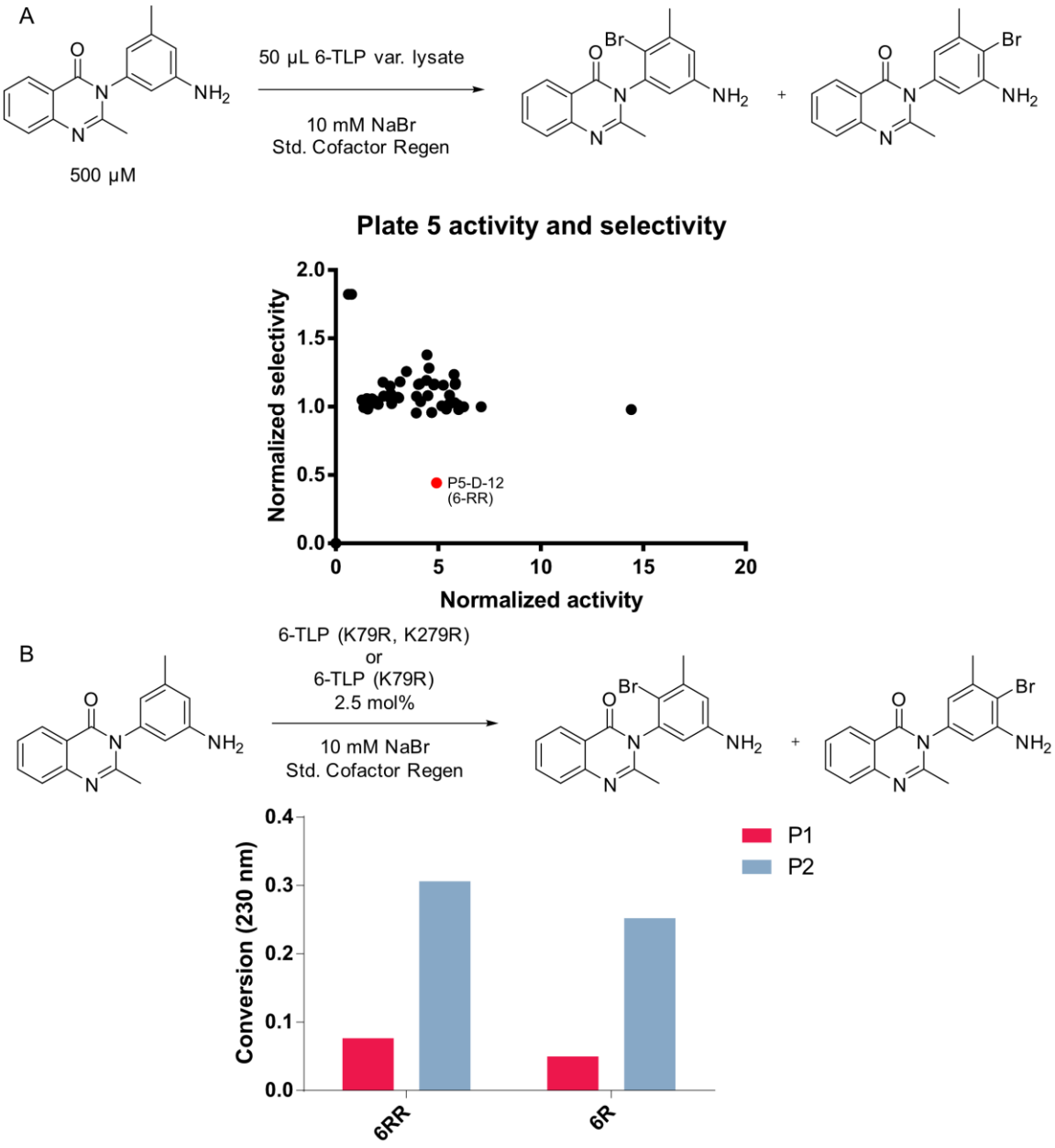
My primary responsibilities on this project were the identification of the 6-RR variant, cloning of the K79R mutation into wt-RebH and 4V, screening and selection of substrates for the evaluation of arginine variants, and evaluation of the genome mined variants. Computational work was done by Declan Evans from the Houk group, and validation of the K79R mutation was done by Caitlin Roof.

7.2 Results and discussion

7.2.1 Evaluation of RebH variants bearing K79R mutation

RebH variant 6-TLP, which contains the K79R mutation, is substantially altered from wild-type RebH, with 12 mutations including 8 first and second-sphere active site residues. Given the lack of activity with analogous mutations to K79R in other FDH scaffolds, we decided that it was essential to verify the observed activity did not result from other mutations in the active site that could influence HOX reactivity. To this end, we cloned the K79R mutation into both the wild-type RebH enzyme and an evolved variant, 4V, which was originally developed for halogenation of large substrates not accepted by wild-type RebH.[9] 4V has several mutations from wild-type RebH that do not overlap with those found in 6-TLP. The Alphafold models[10] for these enzymes also show distinctly different active site conformations, ensuring that if we identified activity from the arginine variant of this halogenase then it was likely other engineered RebH variants would accept this mutation as well.

Figure 7. 2: Initial discovery of 6-RR and comparison of the activity for 6-RR and 6-R



A) Data from the screen where variant 6-RR was originally identified. Data refers to both the activity (total conversion from **1** to **1a** and **1b**) and selectivity (ratio of **1a**/**1b**) normalized to the activity of variant 6-TLP in the same screen. 6-RR is highlighted in red, showing an increased activity but decreased ratio of **1b**/**1a** compared to the parent variant.

The K79R mutation was introduced into both 4V and the wild-type RebH scaffold by QuikChange PCR[11] and verified by Sangar sequencing. With the arginine variants of the enzymes verified, we then

evaluated the activity of each enzyme against a substrate for which the parent enzyme was well characterized. Enzyme-substrate pairs were evaluated both for chlorination and bromination. The release of HOX from the active site of FDHs both with the native lysine residue and altered catalytic residues such as alanine or arginine has been well documented.[6,12] To mitigate any product formation by adventitious HOX, we included the scavenger glutathione in the reaction, as well as the peroxide scavenger catalase.

The arginine variants were active toward each of the substrates evaluated (**Figure 7.3**). In general, the arginine variants tended to prefer bromination over chlorination despite the known native halide for RebH being chloride. This preference was most obvious for RebH variant 4-V, which gave chlorinated product **3-Cl** in 40% conversion, about 4-fold higher than the observed conversion for 4-VR. For bromination of this substrate, however, both enzymes were nearly identical, each generating about 22% conversion to **3-Br**. Halogenation of L-tryptophan also demonstrated this trend, with RebH fully converting substrate **2** to a mixture of **7-Cl-2** and **5,7-di-Cl-2** and RebH-R providing only 25% conversion to **7-Cl-2** with most of the reaction comprising unreacted starting material. Not only did the bromination reaction of RebH-R reveal full conversion of the starting material, but unlike the wild-type RebH bromination reaction which is known to follow the chlorination reaction in terms of selectivity (7-Br-**2** is converted to 5,7-di-Br-**2**), this reaction showed the formation of a new brominated tryptophan product with the same m/z value as the 7-halogenated material.

Figure 7. 3: Results of biocatalysis with RebH-K79R variants

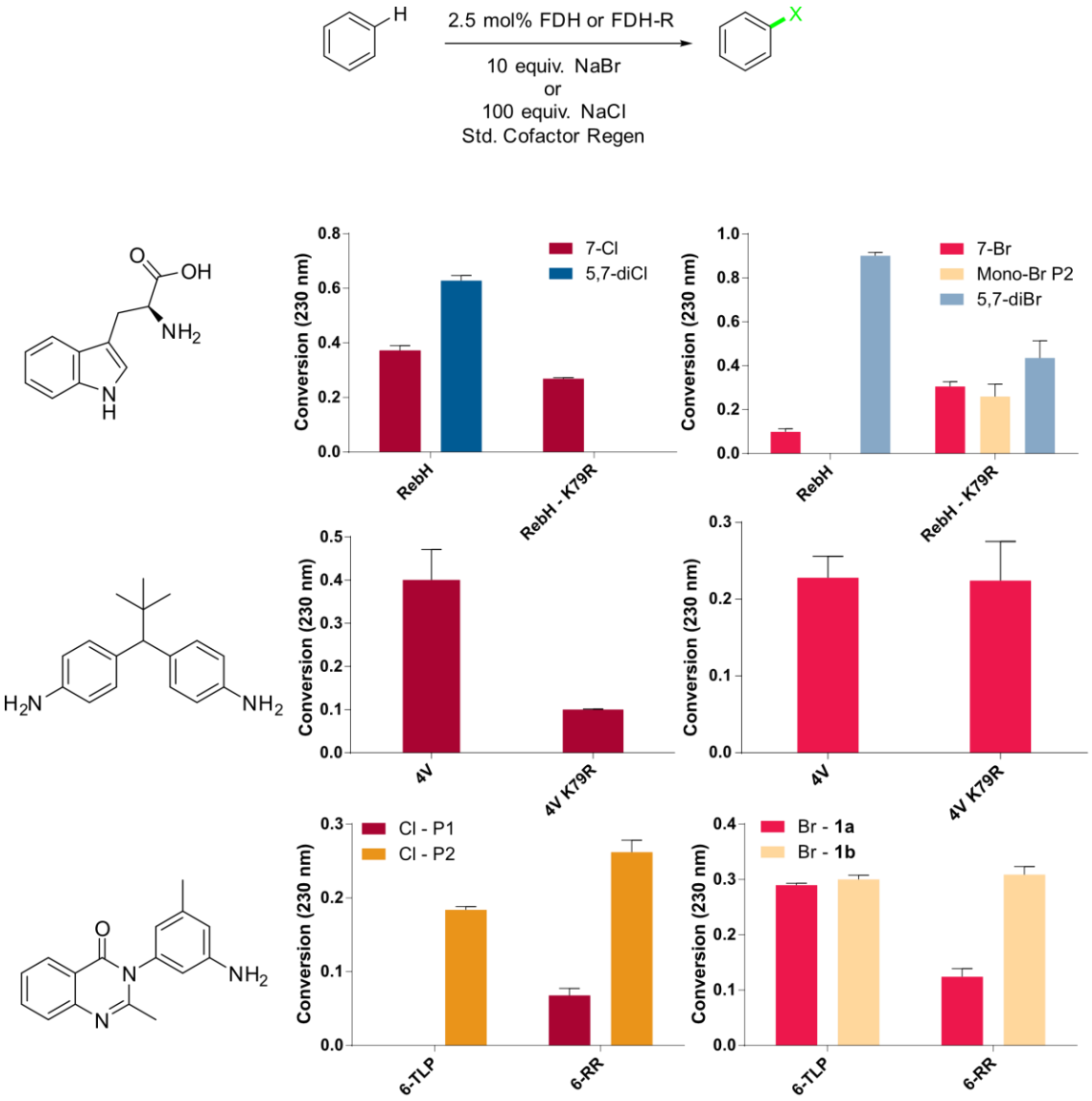


Figure 3: Bioconversions with lysine and arginine derivatives of RebH and variants. Each bar represents the average of three replicates with standard deviation included. Conversion was measured using the integrations of the product and starting material peaks at 230 nm.

Bromination of compound **1** with 6-TLP resulted in a near 1:1 mixture of product **1a**:**1b**, whereas 6-RR has a significantly lower **1a**:**1b** ratio, favoring bromination at the less-sterically hindered 4' position of the aniline ring. In contrast to the arginine variants of RebH and 4-V, 6-RR maintained chlorination activity on substrate **1** and formed a new, earlier eluting monochlorinated product, leading to higher overall TON. Though this new product could not be characterized through isolation due to low yield,

analogy to the retention times of brominated products **1a** and **1b** suggest that this earlier-eluting product may be either the 2' or 6'-chlorinated product. Given the success recently had with directed evolution of 6-TLP for the highly atroposelective bromination of substrate **1** and the relative dearth of developed atroposelective chlorinations, 6-RR could perhaps act as a starting point for directed evolution of the corresponding chlorination reaction.

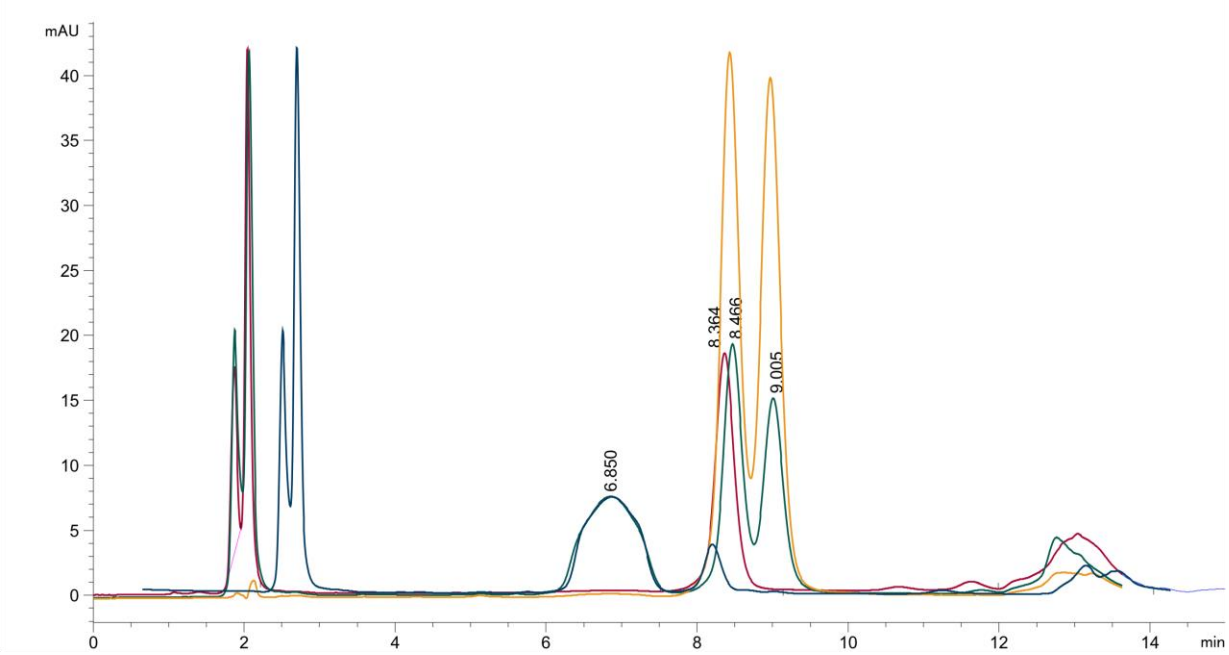
7.2.2: Cloning and evaluation of 3-T Flip (3-T K79E, E354K) and evaluation of putative natural arginine bearing halogenases

Inspired by these results, we decided to probe the extent to which RebH could tolerate other mutations of residues previously considered as catalytically essential . In our recent work, we evolved the RebH variant 6-TLP for highly site- and atroposelective halogenation of 3-arylquinazolinones, a medicinally relevant privileged scaffold. Models of quinazolinone substrate **1** bound within the active site of the final variant 3-T showed the catalytic lysine residue K79 extending towards the 6-position of the aniline ring substituent in a pro-*M* fashion, reflecting the observed site and atroposelectivity observed in the reaction. In this model, substrate **1** is positioned directly between K79 and E354 which has been proposed to act as a general base for the deprotonation of the Wheland intermediate to facilitate substrate turnover into product.

The tight positioning of substrate within the active site in this model lead us to consider that if the relative positioning of K79 and the E354 residues could be flipped without otherwise disturbing substrate binding it may be possible to invert the enantioselectivity of the reaction in a programmable fashion. The double mutant of 3T[8] K79E, E354K (dubbed 3T-flip) was constructed and purified via IMAC chromatography. To our delight, not only was this enzyme solubly expressed from *E. coli*, we also observed activity, albeit both activity and regio-selectivity were significantly reduced from the parent 3-T. Chiral chromatography revealed that the only identifiable enantiomer was the same as generated from the parent 3T, indicating the predicted enantioselectivity flip did not occur (**Figure 7.4**). Despite the

unchanged stereoselectivity, the observed activity of 3T-flip again contradicts long-held assumptions about “catalytically essential” residues in FDHs.

Figure 7. 4: Comparison Of Chiral UHPLC For 3T, 3T-Flip, And NBS Of Quinoline Product 1a



Blue: Crude 3T-Flip reaction. Red: Crude 3T reaction. Gold: Racemic NBS material. Green: Coinjection of racemic material and 3T-Flip reaction. Interpretation: 3T-Flip maintains the enantioselectivity of the parent enzyme 3T. The 3T-Flip reactions were normalized for the starting material peak at 6.85.

In light of the observed activity with variants of RebH containing the active site arginine residue, we began to suspect this residue might exist within other native halogenases. To probe this, we revisited the set of enzymes obtained from our previous work with genome mining of halogenase gene clusters.[13] By first aligning the sequences of the enzymes within this set, then comparing models of the enzymes constructed using Alphafold we identified four enzymes which could be expressed in *E. coli* and contained an arginine residue that overlays with the lysine of RebH. While none of these enzymes were identified as being active enzymes from the high-throughput screen, we expressed and purified them to re-screen them in a lower-throughput format against two electron-rich substrates. Two enzymes from this set, 1-F12 and 2-C02, halogenated the electron rich indole substrate pindolol in trace conversion (0.5%), with yield unaffected by the HOX scavenger glutathione (for example, see **Figure 7.5**). Given the

unknown native substrates of these enzymes, it's possible that these enzymes are native FDHs but functionalize substrates significantly different than what was examined within this study. More work would be needed to confirm there are halogenases bearing catalytic arginine residues.

7.3 Conclusions

The efforts highlighted in this study demonstrate the first examples of active flavin-dependent halogenase proteins with a non-lysine amino acid acting as the catalytic residue. In many cases, modification of the catalytic residue results in an alteration of site-selectivity or halide preference. Given that previous efforts .

7.4 Experimental data

7.4.1 General materials and methods

Materials

All chemical reagents were purchased from commercial suppliers and used without further purification unless otherwise noted. Primers used in this study were purchased from Sigma-Aldrich. NAD, FAD, and antibiotics were purchased from Chem-Impex International Inc. GDH-105 (hereafter, GDH) was purchased from Codexis. Catalase from bovine liver was obtained from Millipore Sigma (2000-5000 U/mg, product number C9322). Antibiotics were prepared as 1000x stock solutions of 50 mg/mL kanamycin and 100 mg/mL ampicillin in water and 25 mg/mL chloramphenicol in 95% ethanol. The pGro7 plasmid encoding the chaperone proteins groEL and groES chaperone set was purchased from Takara (Otsu, Shiga, Japan). Reaction buffer is constant throughout the study and is defined as 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) at a concentration of 25 mM titrated to pH 7.4 using 1 M NaOH. This buffer was prepared using Milli-Q water to ensure no chloride sources were present in solution. Data visualization and kinetic analysis were conducted using Microsoft Excel and GraphPad Prism.

Instruments

Ultra-high pressure liquid chromatography (UHPLC) was performed using an Agilent 1200 series system with a 1290 Infinity II high-speed pump, a 1260 Infinity II diode array detector, and a 1290 Infinity II

multisampler with single-needle configuration. All achiral UPLC analysis was performed using an Eclipse Plus C18 2.1x5 mm guard column with a 1.8 µm pore size (part number 821725-901) connected to a ZORBAX rapid resolution C18 column (part number 959757-902). The UHPLC-MS instrument used in this study was the same, except this was connected to a 6135X single quadrupole mass spectrometer with an Agilent Jet Stream ESI source. Mobile phase A for UHPLC-MS analysis was water with 0.1% FA, mobile phase B was acetonitrile with 0.1% FA. Chiral chromatography was performed using the same UHPLC Agilent 1200 series system as previously described except equipped with a CHIRALPAK®IC-3 column. The mobile phases for chiral chromatography on the UHPLC were water with 20 mM ammonium bicarbonate pH 9.0 for mobile phase A, and acetonitrile without additive for mobile phase B.

UHPLC Methods

UHPLC-MS Method 1: This method was used to analyze the bioconversions of L-Tryptophan

UHPLC-MS Method 2: This method was used to analyze the bioconversions of tBudianiline and the quinazolinone compound.

Chiral UHPLC Method 1: This method was used to determine the enantiomer generated by halogenase 3-T flip

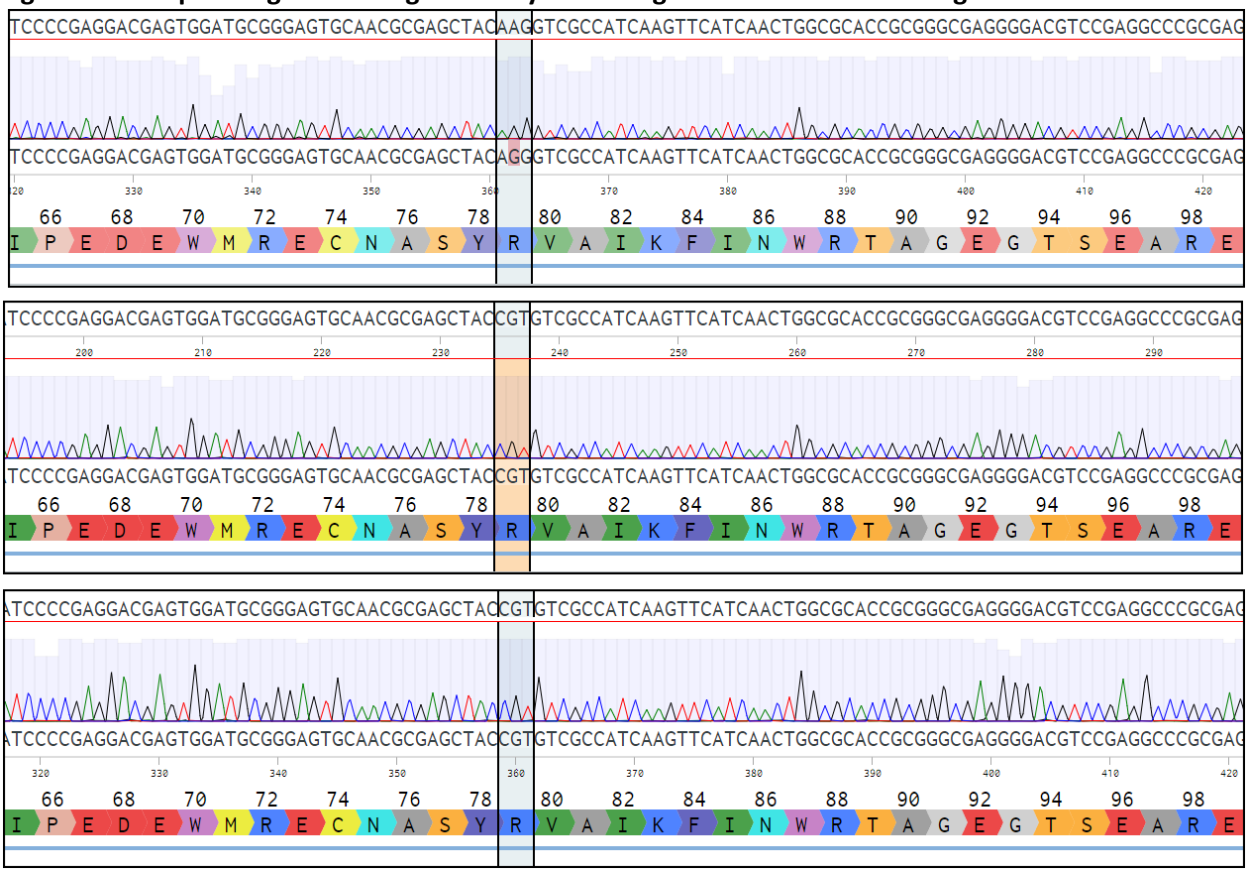
7.4.2 Enzyme cloning and purification

Molecular cloning

Standard molecular biology protocols were followed[7]. Oligonucleotides were ordered from Sigma and prepared according to the manufacturer's recommendations. DNA sequences were visualized using the Benchling website. All FDH expression plasmids were cloned such that the start codon for the gene was in line with the Nco1 cut site such that halogenases had an N-terminal 6xHis tag and constructed using QuikChange[9] from PCR products. All primer sequences and descriptions are included in a table below. QuikChange PCR fragments were generated using PrimeSTAR® Max DNA Polymerase. After PCR, the resulting linear DNA fragments were purified via gel extraction from agarose using the Qiagen Gel Extraction kits, purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. The linear DNA fragments were eluted with 30 µL water. 3 µL of rCutSmart buffer

(NEB#B6004S) was then added to the linear DNA, followed by 1 µL of DPN1 (NEB#R0176L), and the reaction was incubated at 37 °C for 1 hour. The reaction was then purified using the Zymo Research Clean and Concentrate kits (Cat#D4004 or D4013) according to the manufacturers protocol and both insert and vector were eluted with 10 µL water. 2 µL of the resulting ligated product was used to transform *E. coli* BL-21(λDE3) electrocompetent cells containing the pGro7 plasmid[10] via electroporation and plated with the appropriate selection antibiotics. Electroporation was carried out on a Bio-Rad MicroPulser using method Ec2. Genes were confirmed by sequencing from whole plasmids by Quintara Biosciences (625 Mt Auburn, St STE 105, Cambridge, MA). For cloning of 3-T flip, Gibson assembly was used. PCR fragments were generated and prepared as earlier described. For the assembly reaction, a 2:1 molar ratio of insert:vector was used, along with NEB 2x HiFi mixture according to the manufacturers protocol. The assembly reaction was performed for 1 hour at 50 °C in an isocratic method. The following assembly mixture was used without further purification, and 2 µL of the mix was used to transform *E. coli* as previously described. For the conditions of the error-prone library where variant 6-RR was originally identified, please see the appropriate citation for the prior publication. Sequencing chromatograms obtained from Quintara Biosciences for 6-RR, RebH-R, and 4V-R are included in **Figure 7.5** below.

Figure 7. 5: Sequencing chromatograms of lysine to arginine mutations showing successful mutation



Top: 6RR. Middle: RebH-R. Bottom: 4V-R. In each case, the arginine residue corresponding to the catalytic residue is bracketed. The pale bar behind the chromatogram corresponds to sequence 'Quality' as defined by Quintara, and in all cases the quality of the read for the codon encoding the K->R mutation is above 90 indicating high confidence in the arginine residue. The data was visualized in Benchling.

Sequencing chromatograms showing arginine in place of catalytic residues

Table 7. 1: Primer sequences used in this study

HS217	GGAATCCCCGAGGACGAGTGGATGCGGGAG TGCAACGCGAGCTAC	Quickchange K79R	F
HS218	GCGCCAGTTGATGAACTTGATGGCGACACGG TAGCTCGCGTTGC	Quickchange K79R	R
HS219	GACGACGCGAACGGTGTGGAACCGTTCACCT CGGCGATCGCCATG	Quickchange K279R	F
HS220	CATCGGGATCTTCCACGTCCAGCCCGAACGC ATGGCGATCGCCG	Quickchange K279R	R
HS245	GGGAGTGCAACGCGAGCTACGAAGTCGCCA TCAAGTTCATCAAC	3T Flip insert	F

HS246	CGTAGACGAAGTAGATGCCCGTCGATTTCAG TGGTTCCACGAAGC	3T Flip insert	R
HS247	GCTTCGTGGAACCACTGAAATCGACGGGCAT CTACTTCGTCTACG	3T Flip vector	F
HS248	TTGATGAACTTGATGGCGACTTCGTAGCTCG CGTTGCAC	3T Flip vector	R

PCR cycles:

K79R Quikchange PCR:

Used to construct 6-R from 6-TLP, RebH-R from RebH, and 4V-R from 4V

Reagent	Stock concentration	Volume (µL)	Final concentration
Primestar 2x Mix	2x	25	1x
Forward primer (HS217)	10 µM	2.5	500 nM
Reverse primer (HS218)	10 µM	2.5	500 nM
Template (RebH, 4V, or 6-TLP)	50 ng/µL	1	0.5 ng/µL

3-T Flip Insert PCR

PCR Set up:

Reagent	Stock concentration	Volume (µL)	Final concentration
Primestar 2x Mix	2x	25	1x
Forward primer (HS245)	10 µM	2.5	500 nM
Reverse primer (HS246)	10 µM	2.5	500 nM
Template (3-T)	50 ng/µL	0.5	0.5 ng/µL
Water		19.5	

PCR Cycles

Temperature (°C)	Length of cycle (min:sec)	Repetitions
95	5	1
95	0:30	28
60	0:10	
60	0:20	
65	5:00	1
4	hold	1

3-T Vector PCR:

PCR Set up:

Reagent	Stock concentration	Volume (µL)	Final concentration
Primestar 2x Mix	2x	25	1x
Forward primer (HS247)	10 µM	2.5	500 nM
Reverse primer (HS248)	10 µM	2.5	500 nM
Template (3-T)	50 ng/µL	0.5	0.5 ng/µL
Water		19.5	

PCR Cycles

Temperature (°C)	Length of cycle (min:sec)	Repetitions
95	5	1
95	0:30	28
60	0:10	

60	0:50	
65	5:00	1
4	hold	1

Gibson assembly reaction was performed as follows:

	Volume (µL)	Concentration (ng/µL)
Insert	0.5	90
Vector	1	135
NEB 2x HiFi Mix	1.5	N/A

Large-scale halogenase preparation

For large scale halogenase production, 14 mL culture tubes containing 10 mL LB with kanamycin and chloramphenicol were inoculated with BL-21(λDE3) *E. coli* cells that had been transformed with pGro7 and pET28b containing the appropriate flavin-dependent halogenase (FDH) gene (*vide infra*) as previously described. Construction and description of the halogenase expression plasmids is included in the cloning section of this SI. The cultures were incubated overnight at 37 °C, 250 rpm. The next day, 750 mL TB with the appropriate antibiotics was inoculated with the entire overnight culture. The inoculated expression cultures were incubated at 37 °C, 225 rpm until OD_{600} was between 0.6 and 0.8. The incubator was allowed to cool to 30 °C, and gene expression was induced with 2 mg/mL arabinose and 100 µM IPTG, and the expression culture was incubated for 20 hours. Cells were harvested by centrifugation at 3600 rpm for 20 minutes in a Sorvall XT centrifuge, then either frozen and kept at -20 °C or immediately resuspended in 25 mM HEPES buffer, pH 7.4. Cells were lysed using a QSonica S-4000 with a 0.5" horn at 40 W and a total processing time of 5 minutes with 1 minute on/off cycles. Cell lysate was clarified at 15,000 rpm for 30 minutes, and the resulting clarified lysate was transferred to a fresh 50 mL centrifuge tube and added to pre-equilibrated Ni-NTA (equilibration buffer: 20 mM phosphate, 300 mM NaCl, 10 mM imidazole pH 7.4). Clarified lysate was incubated with resin for approximately 30 minutes at which point it was transferred

to uncapped spin columns and the lysate was allowed to flow through. The resin was washed with at least 5 CV wash buffer (20 mM phosphate, 300 mM NaCl, 25 mM imidazole pH 7.4), at which point the spin column was transferred to a new centrifuge tube and the resin was washed with elution buffer (20 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4). Eluted protein was concentrated via diafiltration using Amicon spin filters Ultra 30K MWCO spin filters and the buffer was exchanged for storage buffer (25 mM HEPES and 10% glycerol, pH 7.4). For long term storage, proteins were immediately frozen in liquid nitrogen and stored at – 80 °C until use. Halogenase concentration was determined by A_{280} measurements taken on a Tecan Infinite M200 pro microplate reader using the extinction coefficient calculated by the amino acid sequence of the enzyme. MBP-RebF concentrations were determined using the Pierce BCA Protein Assay Kit.

Protein sequences for enzymes discussed within this study

Note that for all halogenases, the catalytic residue is underlined and bolded red (for the genome mined halogenases, this is the putative catalytic residue). The start codon is underlined and the N-terminal flexible linker is included.

6-TLP

MGSSHHHHHHSSGLVPRGSHMSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATTPNLQT
AFFDFLGIPEDEWMRECNASY**K**VAIKFINWRTAGEGTSEARELDGGPDHFYHPLGLLKYHEQIPLSHYWFDRLYRGKTV
EPFDYACYKEPVILDANRSPRRLDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTAT
GRVFDADLFVDCSGFRGLLINKAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGR
FGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQ
LVKHFPDKSFNPVLTARFNREIETMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAP
DDAQLYWGNFEEEFRNPWNNSSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLR
QQHGR*

6-RR

MGSSHHHHHHSSGLVPRGSHMSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATTPNLQT
AFFDFLGIPEDEWMRECNASY**R**VAIKFINWRTAGEGTSEARELDGGPDHFYHPLGLLKYHEQIPLSHYWFDRLYRGKTV
EPFDYACYKEPVILDANRSPRRLDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTAT
GRVFDADLFVDCSGFRGLLINKAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMRSGWTWKIPMLGR
FGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQ
LVKHFPDKSFNPVLTARFNREIETMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAP

DDAQLYWGNFEEEFRNPWNNSSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLR
QQHGR*

6-R

MGSSHHHHHHSSGLVPRGSHMSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATTPNLQT
AFFDFLGIPEDEWMRECNASYRVAIKFINWRTAGEGTSEARELDGGPDHFYHPLGLLKYHEQIPLSHYWFDRLYRGKTV
EPFDYACYKEPVILDANRSPRRLDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTAT
GRVFDADLFVDCSGFRGLLINKAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGR
FGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQ
LVKHFPDKSFNPVLTARFNREIETMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAP
DDAQLYWGNFEEEFRNPWNNSSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLR
QQHGR*

4V

MGSSHHHHHHSSGLVPRGSHMPGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATIPNLQTA
FFDFLGIPEDEWVRECNASYKVAIKFINWRTAGEGTSEARELDGGPDHFYHSSGLLKYHEQIPLSHYWFDRSYRGKTVEP
FDYACYMEPVILDANRSPRRLDGSKVTNYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTAT
GRVFDADLFVDCSGFRGLLINKAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGR
FGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQ
LVKHFPDKSLNPVLTARFNREIETMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMVINAPAS
DDAQLYYGNFEEEFRNFWTNSSYYCVLAGLGLVPDAPSPRLAHMPQATESVDEVFGAVKDRQRNLLETLPSLHEFLRQ
QHGR*

4V-R

MGSSHHHHHHSSGLVPRGSHMPGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATIPNLQTA
FFDFLGIPEDEWVRECNASYRVAIKFINWRTAGEGTSEARELDGGPDHFYHSSGLLKYHEQIPLSHYWFDRSYRGKTVEP
FDYACYMEPVILDANRSPRRLDGSKVTNYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTAT
GRVFDADLFVDCSGFRGLLINKAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGR
FGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQ
LVKHFPDKSLNPVLTARFNREIETMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMVINAPAS
DDAQLYYGNFEEEFRNFWTNSSYYCVLAGLGLVPDAPSPRLAHMPQATESVDEVFGAVKDRQRNLLETLPSLHEFLRQ
QHGR*

3-T

MGSSHHHHHHSSGLVPRGSHMSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATTPNLQT
AFFDFLGIPEDEWMRECNASYKVAIKFINWRTAGEGTSEARELDGGPDHFYHPLGLLKYHEQIPLSHYWFDRLYRGKTV
EPFDYACYKEPVILDANRSPRRLDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTAT
GRVFDADLFVDCSGFRGLLINKAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGR
FGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLESTGIYFVYAALYQ
LVKHFPDKSFNPVLTARFNREIETMFDDTRDLIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAP
DDAQLYRGNFEEEFRNPWTNGSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLRQ
QHGR*

3-T Flip

MGSSHHHHHHSSGLVPRGSHMSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATTPNLQT
AFFDFLGIPEDEWMRECNASYEVAIKFINWRTAGEGTSEARELDGGPDHFYHPLGLLKYHEQIPLSHYWFDRLYRGKTV
EPFDYACYKEPVILDANRSPRRLDGSKVTSYAWHFDAHLVADFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTAT
GRVFDADLFVDCSGFRGLLINKAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGR
FGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCFVEPLKSTGIYFVYAALYQ
LVKHFPDKSFNPVLTARFNREIETMFDDTRDLIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAINAPAP
DDAQLYRGNFEEEFRNPWTNGSYYCVLAGLGLVPDAPSPRLAHMPRATESVDEVFGAVKDQQRNLLETLPSLHEFLRQ
QHGR*

1-E09 (Uniprot A0A0Q7BW10_9BURK)

MGSSHHHHHHSSGLVPRGSHMSDIQSIVIVGGGTAGWLAACYLQRTLDAVSQRALPITLIEPSQVSSIGADMATVPTL
RNTMQALGLPESTLFTAADATLTNGIRFKGWHRGGDAAHGDAYDHPFDMPLPFSGFAATAHWLNLMQRGLTRQP
MAEACTVQTALFDGHRSPKLMDSPDYQAPVSYGYQLDAVKLAMLLQQTAQQRGVKHVRGQVVRVHRGDSGIESVEL
ADGSRHAASFFIDCSGSQSLLLQQGLGVPWMSYADDLPCDRVATMPLAYAEPHEALRSYTTATAQAAGWTWEMDL
QSRRGTGHVYASRFCSDDEAVHMLQALSGGQRPLAEPRLQRLRIGHAARAWEANCLALGTAAGCIEPLQSTSLYLVEW
MLQLFVDHVAPSGNSDGCRGRVNRVFNDLYEELRDFIVAHYALSQRRDTPFWRVCTEEATLPPRLAELLALWDSKVPT
PTDLDRRLSLFGVANWSYLLAGLHRLPSGGIGLAAHIAPEISLQAMAHVRGIRQQASQQSPTMRDYLRSLAAAAQPAC
AH*

1-F07 (Uniprot I3IAI0_9GAMM)

MGSSHHHHHHSSGLVPRGSHMNDGIKKIVIVGRDLDAWITAFFLKSVLDKSRDSYEVTLVELGTLLTEHDIFAVLPSYKM
LHKTLGANEDKLRQTAKARPFFGQRFTGWNPELPEFFHAYDRLGINFNGVDFFQYWMKAVSNGLKLPLEDFSLGVAA
AKHGRFVASTGQADFSHLAYGYHLSAIEYVNAIARAAMEVGVKRENGNIITINRNNDVIHSLSLDDGMVLKADFFIDAS
GADALLINSLTDNNFESWEHWFLCDRIITASSAPLSPAPAFSQVTAFSSGWCGLYPLNNRTAIQTLYSSRHTDFSGVVAE
MKMRVGVDISQGVERPIKCGMLSRPWIGNCLAVGTAAASMEPLDALQEHSLVISMVMLKQLFPNSNEYQNERDVYN
KKMHSFIENLRDFQIVHYALNSRDELFWTACKNLKLPHILKEKIDIFKCCGYASVREDETFQEENWISVFNGHGLAPEFYS
PLVDNMSDDEMIQNFQKILRMIKERINSSPLV*

1-F12 (Uniprot T1W9P7_9ZZZZ)

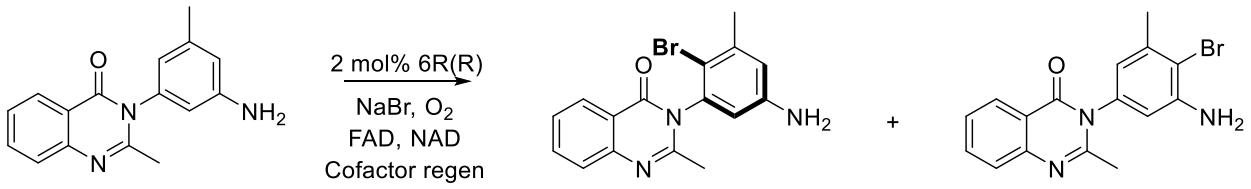
MGSSHHHHHHSSGLVPRGSHMLRERLGHVLVAGGGMTGWCAAAALKRRAPFLDVTILSCAPSPHALADRIACTLPSLI
GFQEDLGIGAADGIVRAGSGYRLGTLFTGWANGVPDYVHAYGRYGAAFGATAFHHHWVRAAQDGSAVPFDRHSPA
AAMARSGRFAAPKPGTPFADHEFGLTIDIARHASMMRAFALHVGVREIPGEIGGVTLDADGAIATIDLTDGTALGADLF
IDATGPEARLRGAFGDDRDDWSAWLPCDRVLIAETNDEPMVLTTITAHEAGWRWASGVQSGFAYASTAISDGKAAR
VLRNTTGTTPDEPIRVRAGTRPQPWRGNCIAIGDAATEIEPLEWCNLHLALSAIDRLIAMLPGRVPVPVEVVEFNRQTLA
EAERVRDFLAMHYHTARRDDPLWRATAATEAPASLAHTLAQFAERGRLPFFEEETFARDSWAAVLIGQGFLPRRVDPL
VSAVPVAATRAAMARQVAAIAAALPHVPTHAAYRADQKRHLTR*

2-C02 (Uniprot A0A0Q4LKP3_9SPHN)

MGSSHHHHHHSSGLVPRGSHMNRRTPPATIAVVGGGQVALLAACALARTLPVARVRLLPTPVPPHAMADRAHGALP
SLARLHDRIGIDEAGLLARAGASHRLATRHDGGWRSDGRDWWIGHGAVADPAAHGWTGRDTTGATGPSGPAVALA
LAERFAPAADDPASPLSDIDHGLRWAPEAYRRHLASLARHLGVVIEAPGDTVPDAVDADLVIDATGRRGDDWIDWSA
ELPIDRVAIAHADPALSLADHVTLDDTALHLISPGRDVTRRVTMRHGGEGTEIASGRRRNPWTGRIVALGDAAALLPPL
GHTNLALAAQAIDLLVELLPGRDIHPLERAEYNRRAGQAADAARDFVAAHLVALPEGHPLRSRPSPLLALRLREFARRAR
LPHVEEDFLPRDLWTQLLTGIGIPPGTPAHLAAADPAVTAAARTAQAQRSTQAVALAEPYPLWLARTLGERG*

7.4.3 Biocatalysis

Comparison of 6R and 6RR

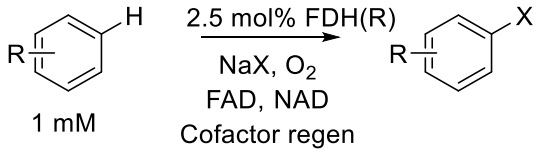


500 µM P1 P2

6R and 6RR were cloned, expressed, and purified as previously described in the general procedure. Due to the low yield of variant 6R, these reactions were performed as single measurements. Stock solutions of 50 µM purified FDH were prepared in storage buffer. An FDH stock was prepared by diluting 15 µL of 50 µM 6-TLP variant with 35 µL of reaction buffer in a microtiter plate (final concentration, 10 µM). A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr, and 2.5 mM of the quinazolinone substrate from 30 mM stock solution in DMSO, were prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the FDH stock in the microtiter plate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. After reactions were finished, protein precipitation was affected by the addition of 1 volume of methanol, and protein was removed from the sample by centrifugation at 3600 RPM in a Sorvall centrifuge held to 4 °C for 15 minutes. From the supernatant, 100 µL was taken and added to a 0.2 µm filter plate and filtered via centrifugation as previously described. Bioconversions were

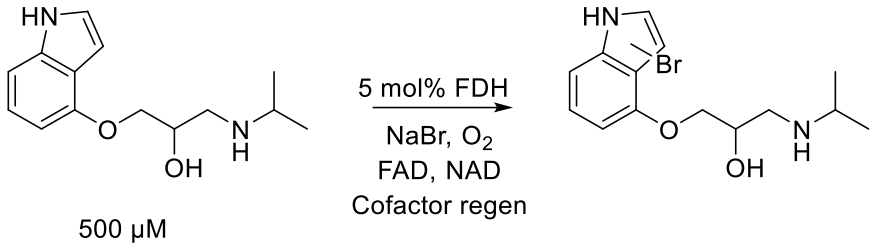
analyzed using UHPLC method 2 as described in the methods section of this SI. Conversion was calculated as the AUC of product over AUC of starting material and product with integrations measured at 230 nm.

Bioconversion protocol for comparison of lysine and arginine variants



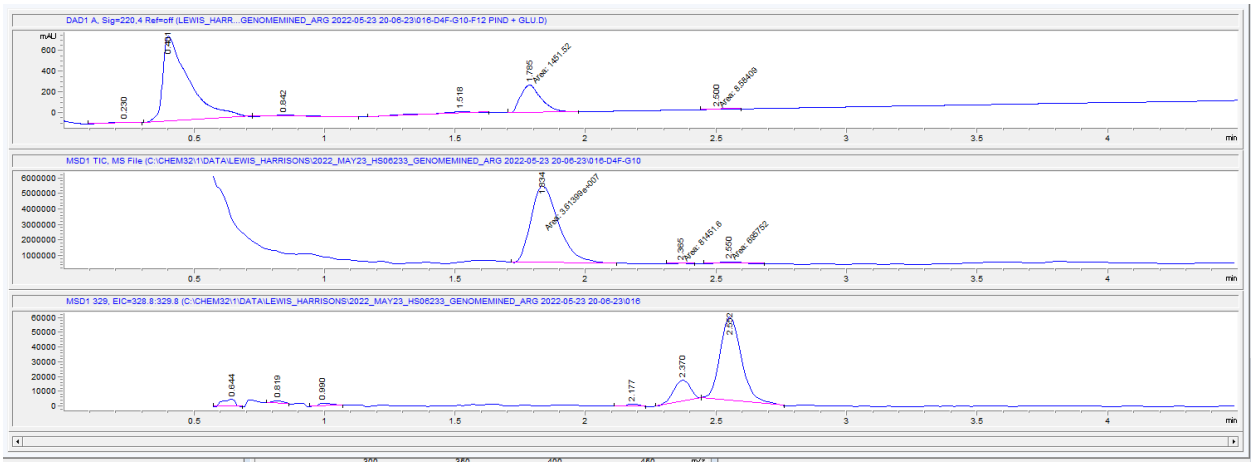
FDH variants were expressed and purified according to the general procedure. Stock solutions of 125 µM purified FDH were prepared in storage buffer. An FDH stock was prepared by diluting 15 µL of 125 µM 6-TLP variant with 35 µL of reaction buffer in a microtiter plate (final concentration, 25 µM). Reactions were performed in triplicate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr or 500 mM NaCl, and 5 mM of the appropriate substrate from 30 mM stock solutions in DMSO, were prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the FDH stock in the microtiter plate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. After reactions were finished, protein precipitation was affected by the addition of 1 volume of methanol, and protein was removed from the sample by centrifugation at 3600 RPM in a Sorvall centrifuge. From the supernatant, 100 µL was taken and added to a 0.2 µm filter plate and filtered via centrifugation as previously described. Bioconversions were analyzed using UHPLC methods appropriate for the compounds as described in the methods section of this SI. Conversion was calculated as the AUC of product over AUC of starting material and product with integrations measured at 230 nm.

Evaluation of genome mined variants.



FDH variants were expressed and purified according to the general procedure. Stock solutions of 125 µM purified FDH were prepared in storage buffer. An FDH stock was prepared by diluting 15 µL of 125 µM 6-TLP variant with 35 µL of reaction buffer in a microtiter plate (final concentration, 25 µM). Reactions were performed in triplicate. A small molecule stock solution consisting of 500 µM FAD, 500 µM NAD, 100 mM D-glucose, 50 mM NaBr or 500 mM NaCl, and 2.5 mM of the appropriate substrate from 30 mM stock solutions in DMSO, were prepared in reaction buffer. An enzyme stock consisting of 67.5 U/mL GDH, 262.5 U/mL catalase, and 18.75 µM of MBP-RebF was prepared in reaction buffer. 15 µL of small molecule mix was added to the FDH stock in the microtiter plate, followed by 10 µL of the enzyme mix for a total reaction volume of 75 µL. After addition of the enzyme mix, the plate was sealed using an aluminum heat seal and set to incubate at room temperature for 16 hours. After reactions were finished, protein precipitation was affected by the addition of 1 volume of methanol, and protein was removed from the sample by centrifugation at 3600 RPM in a Sorvall centrifuge. From the supernatant, 100 µL was taken and added to a 0.2 µm filter plate and filtered via centrifugation as previously described. Bioconversions were analyzed using UHPLC-MS methods appropriate for the compounds as described in the methods section of this SI. Conversion was calculated as the AUC of product over AUC of starting material and product with integrations measured at 230 nm.

Figure 7. 6: UHPLC-MS trace from bioconversions with pindolol and genome mined enzyme 1-F12



All traces are shown from injections including 1 mM glutathione. Reactions were conducted in triplicate, only one representative reaction is shown. Top: UV trace measured at 230 nm. Middle: TIC trace Bottom: EIC for major peak of monobrominated pindolol isomer.

7.5 References

(1) Harris, C. M.; Kannan, R.; Kopecka, H.; Harris, T. M. The Role of the Chlorine Substituents in the Antibiotic Vancomycin: Preparation and Characterization of Mono- and Didechlorovancomycin. *J Am Chem Soc* **1985**, *107* (23), 6652–6658. https://doi.org/10.1021/ja00309a038.

(2) Agarwal, V.; Miles, Z. D.; Winter, J. M.; Eustáquio, A. S.; el Gamal, A. A.; Moore, B. S. Enzymatic Halogenation and Dehalogenation Reactions: Pervasive and Mechanistically Diverse. *Chem Rev* **2017**, *117* (8), 5619–5674. https://doi.org/10.1021/acs.chemrev.6b00571.

(3) Timmins, A.; de Visser, S. P. A Comparative Review on the Catalytic Mechanism of Nonheme Iron Hydroxylases and Halogenases. *Catalysts*. Multidisciplinary Digital Publishing Institute July 31, 2018, p 314. https://doi.org/10.3390/catal8080314.

(4) Andorfer, M. C.; Lewis, J. C. Understanding and Improving the Activity of Flavin-Dependent Halogenases via Random and Targeted Mutagenesis. *Annual Review of Biochemistry*. 2018, pp 159–185. https://doi.org/10.1146/annurev-biochem-062917-012042.

(5) Yeh, E.; Blasiak, L. C.; Koglin, A.; Drennan, C. L.; Walsh, C. T. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases. *Biochemistry* **2007**, *46* (5), 1284–1292. https://doi.org/10.1021/bi0621213.

(6) Mondal, D.; Fisher, B. F.; Jiang, Y.; Lewis, J. C. Flavin-Dependent Halogenases Catalyze Enantioselective Olefin Halocyclization. *Nat Commun* **2021**, *12* (1). https://doi.org/10.1038/s41467-021-23503-3.

(7) Andorfer, M. C.; Evans, D.; Yang, S.; He, C. Q.; Girlich, A. M.; Vergara-Coll, J.; Sukumar, N.; Houk, K. N.; Lewis, J. C. Analysis of Laboratory-Evolved Flavin-Dependent Halogenases Affords a Computational Model for Predicting Halogenase Site Selectivity. *Chem Catalysis* **2022**. https://doi.org/10.1016/j.checat.2022.07.003.

(8) Snodgrass, H. M.; Mondal, D.; Lewis, J. C. Directed Evolution of Flavin-Dependent Halogenases for Site- and Atroposelective Halogenation of 3-Aryl-4(3 H)-Quinazolinones via Kinetic or Dynamic Kinetic Resolution . *J Am Chem Soc* **2022**, *4*. https://doi.org/10.1021/jacs.2c07422.

(9) Payne, J. T.; Poor, C. B.; Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angewandte Chemie* **2015**, *127* (14), 4300–4304. https://doi.org/10.1002/ange.201411901.

(10) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(11) Kunkel, T. A. Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *Proc Natl Acad Sci U S A* **1985**, *82* (2), 488–492. https://doi.org/10.1073/pnas.82.2.488.

(12) Phintha, A.; Prakinee, K.; Jaruwat, A.; Lawan, N.; Visitsatthawong, S.; Kantiwiriyawanitch, C.; Songsungthong, W.; Trisrivirat, D.; Chenprakhon, P.; Mulholland, A.; van Pée, K. H.; Chitnumsub, P.; Chaiyen, P. Dissecting the Low Catalytic Capability of Flavin-Dependent Halogenases. *Journal of Biological Chemistry* **2021**, *296*, 100068. https://doi.org/10.1074/jbc.RA120.016004.

(13) Fisher, B. F.; Snodgrass, H. M.; Jones, K. A.; Andorfer, M. C.; Lewis, J. C. Site-Selective C−H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. **2022**, *10*, 50. https://doi.org/10.1021/acscentsci.9b00835.

APPENDIX I

NMR Spectra for Compounds from Chapter 2

Figure AI. 1: 1HNMR spectra of 7-bromopindolol from 1-F11 bioconversion

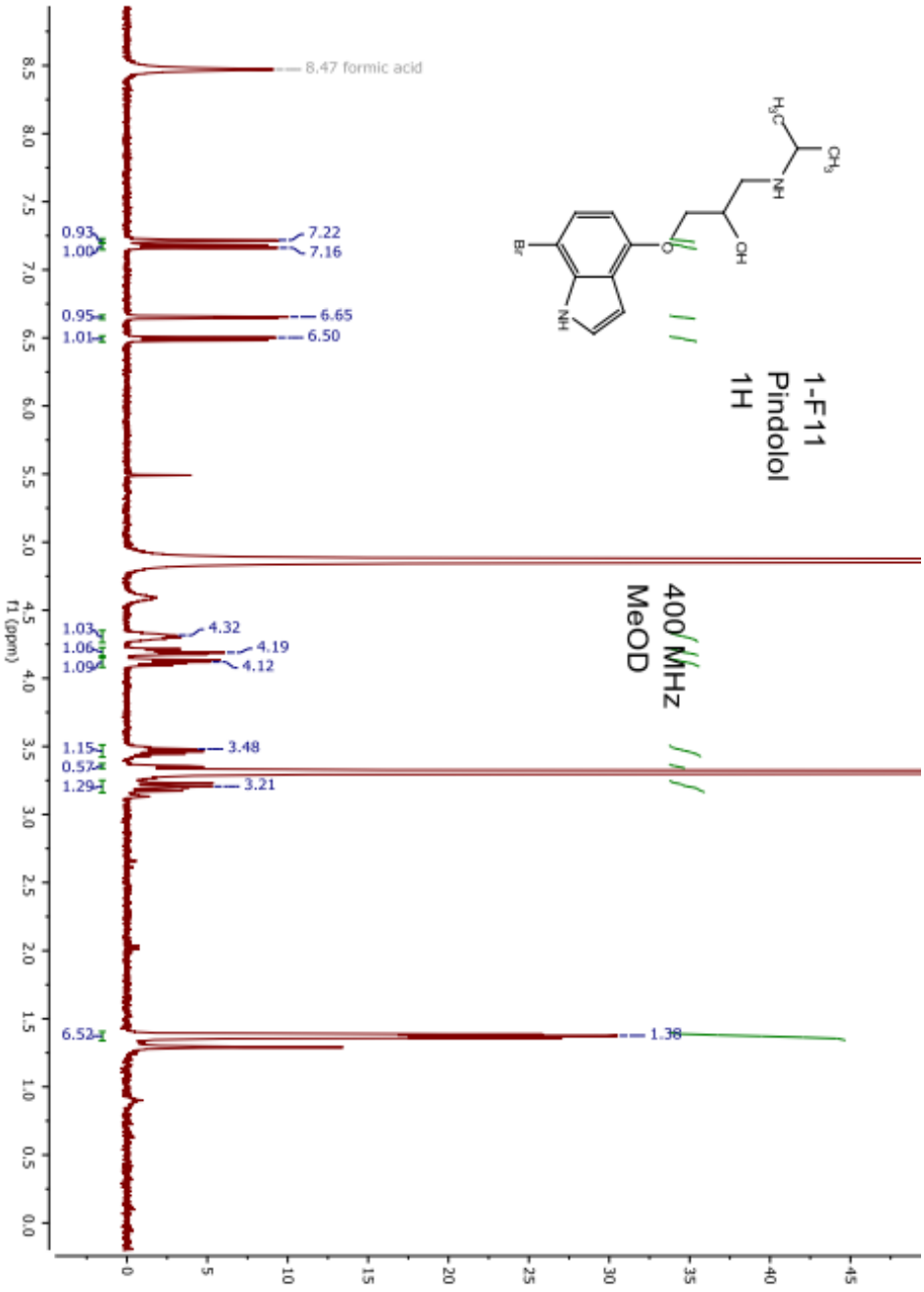


Figure AI. 2: 13CNMR spectra of 7-bromopindolol from 1-F11 bioconversion

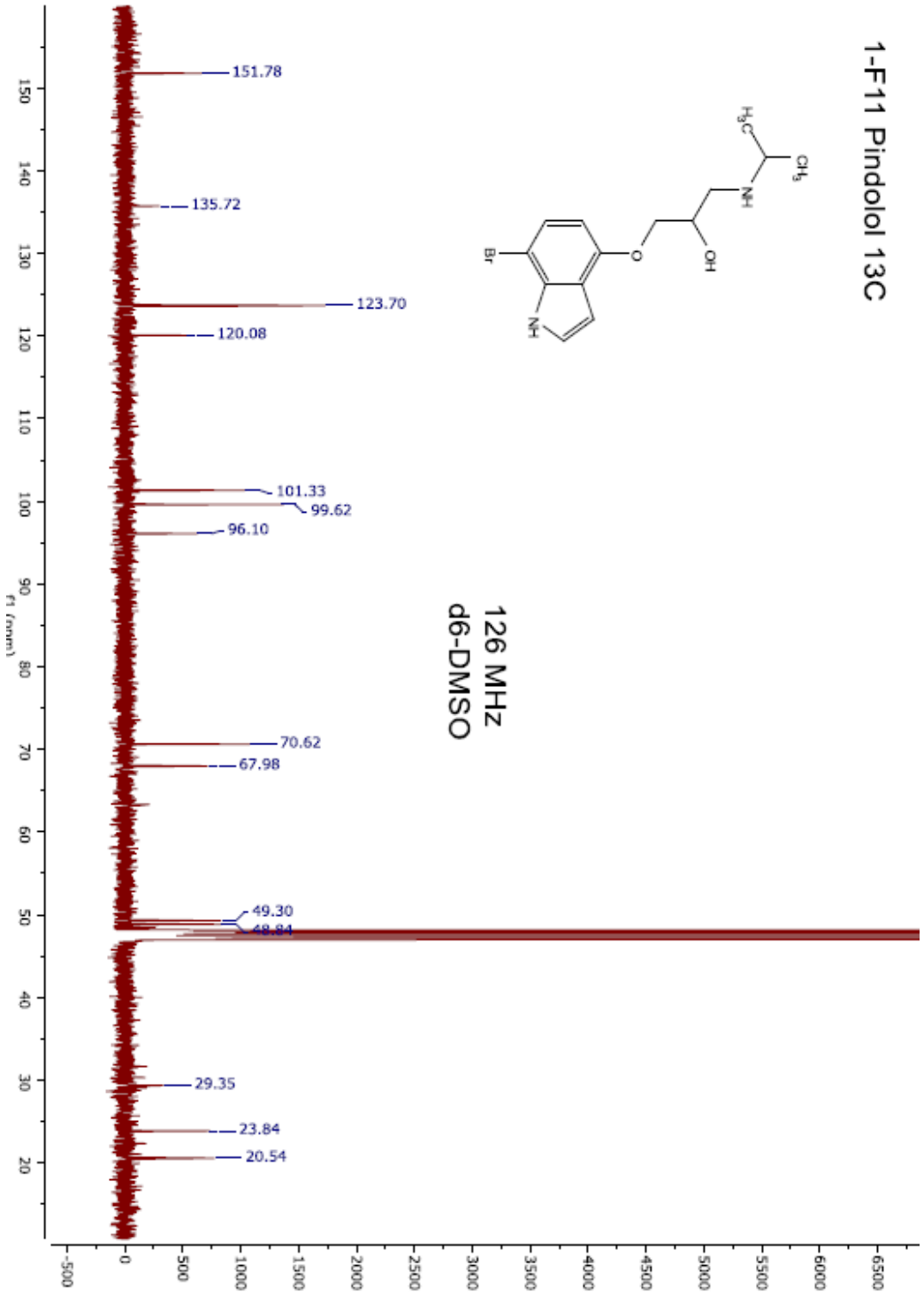


Figure AI. 3: 1HNMR spectra of 2-bromopindolol from 2-C01 bioconversion

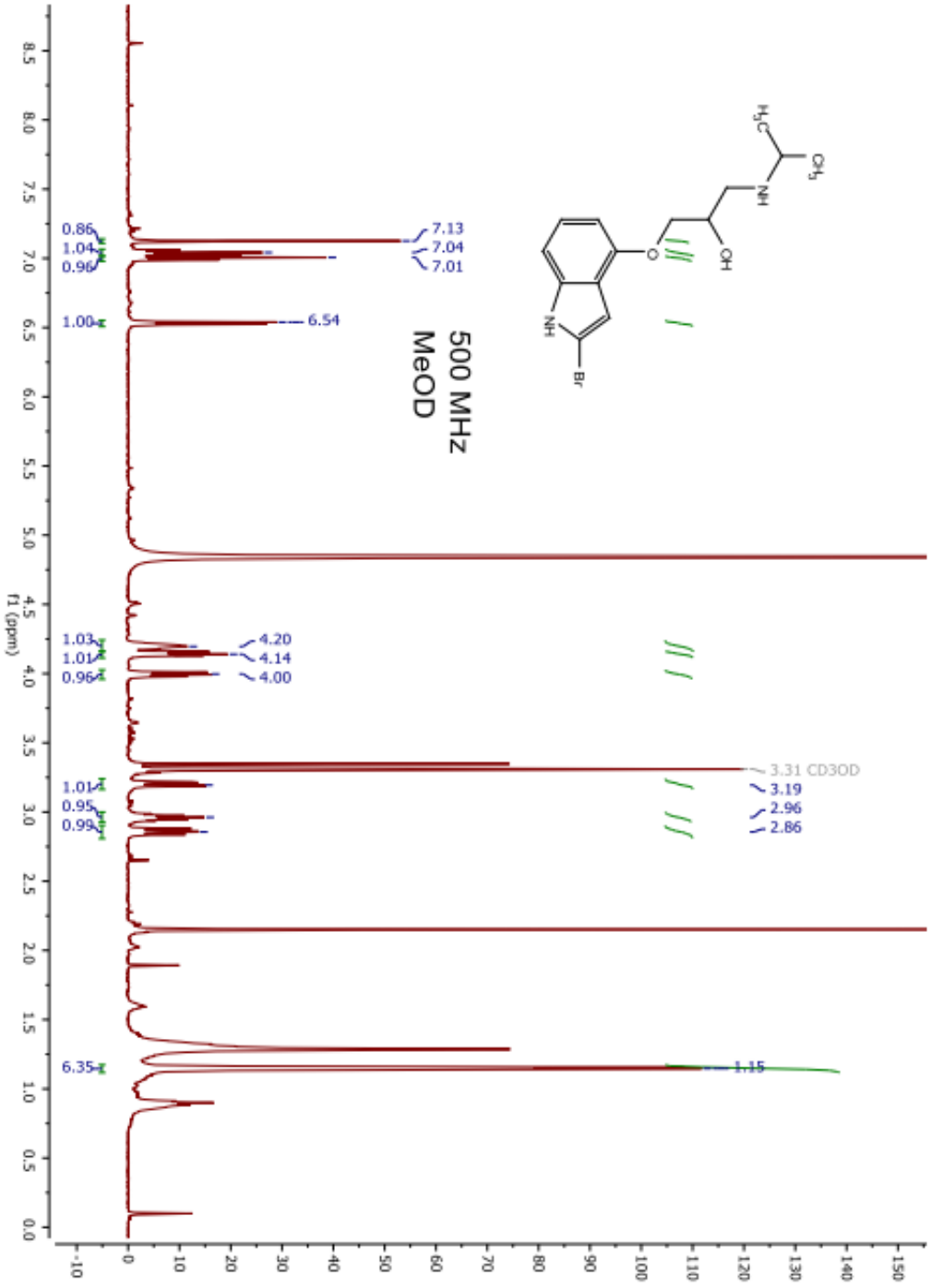


Figure AI. 4: 13CNMR of 2-bromopindolol from 2-C01 bioconversion

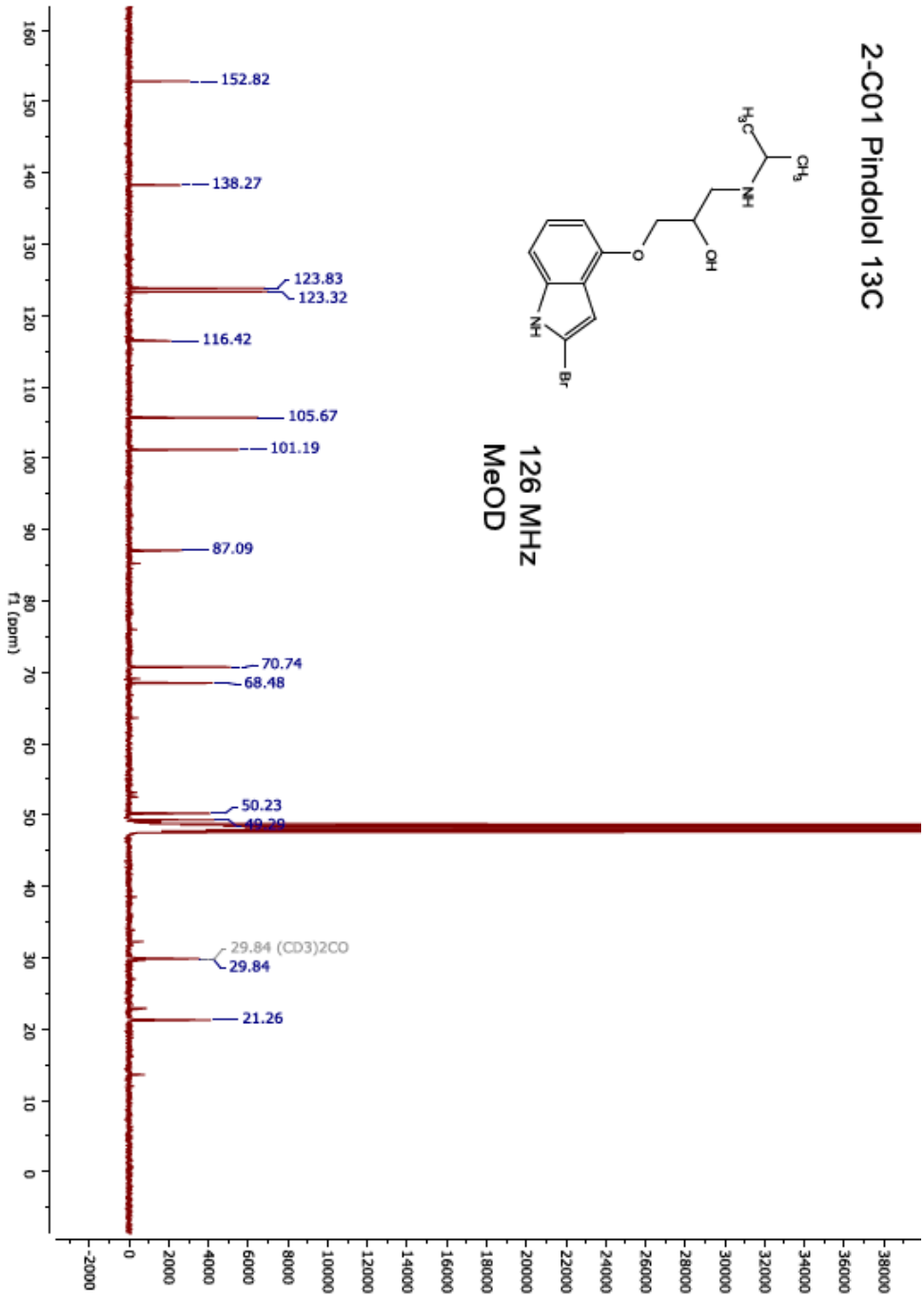


Figure AI. 5: 1H NMR of 8-bromonaringenin from 1-F05 bioconversion

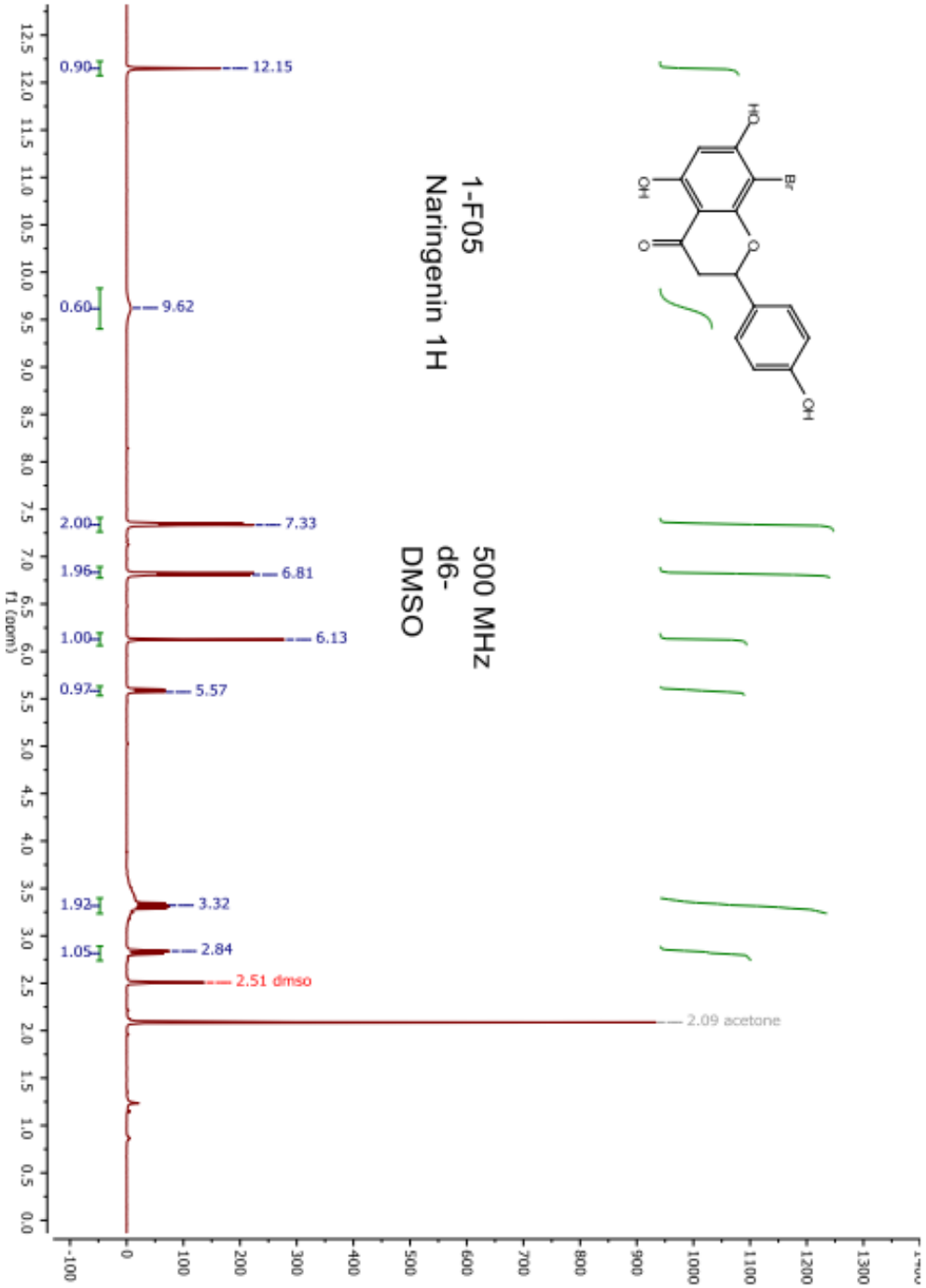


Figure AI. 6: 13CNMR of 8-bromonaringenin from 1-F05 bioconversion

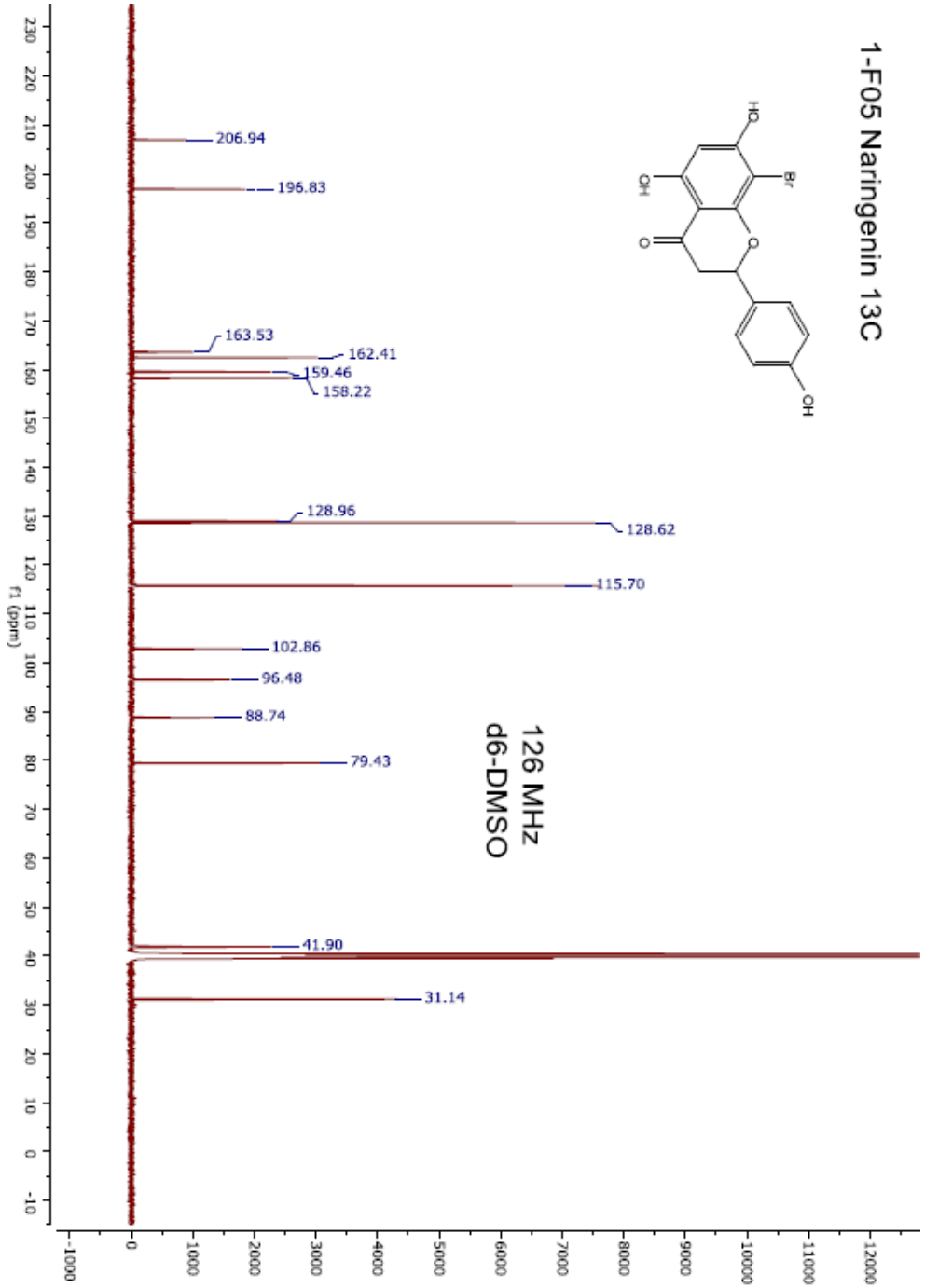


Figure AI. 7: 1HNMR of 6-bromonaringenin from 1-F11 bioconversion

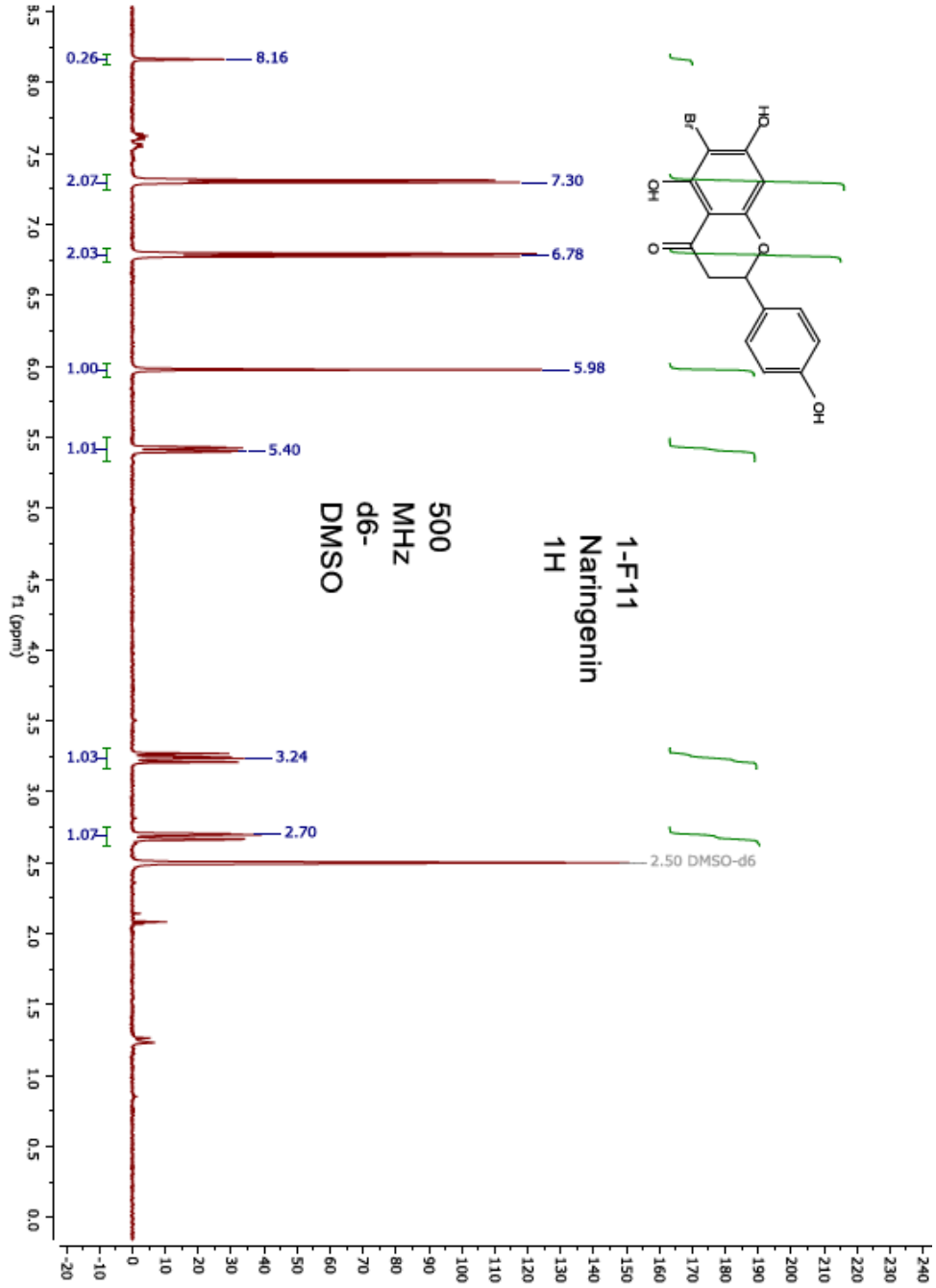


Figure AI. 8: 13CNMR of 6-bromonaringenin from 1-F11 bioconversion

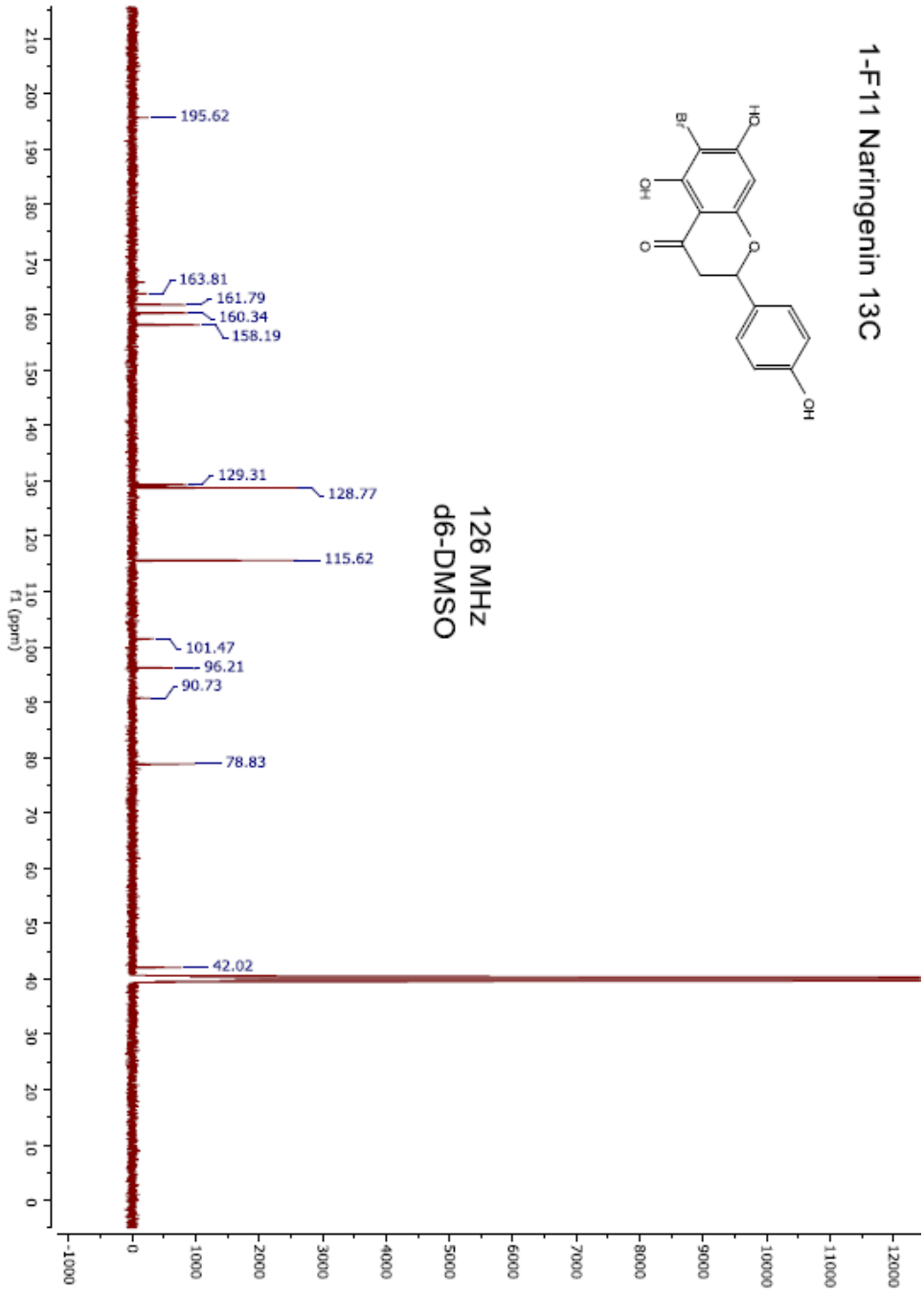


Figure AI. 9: 1HNMR of 7-bromomethylergonovine 1-B12 bioconversion

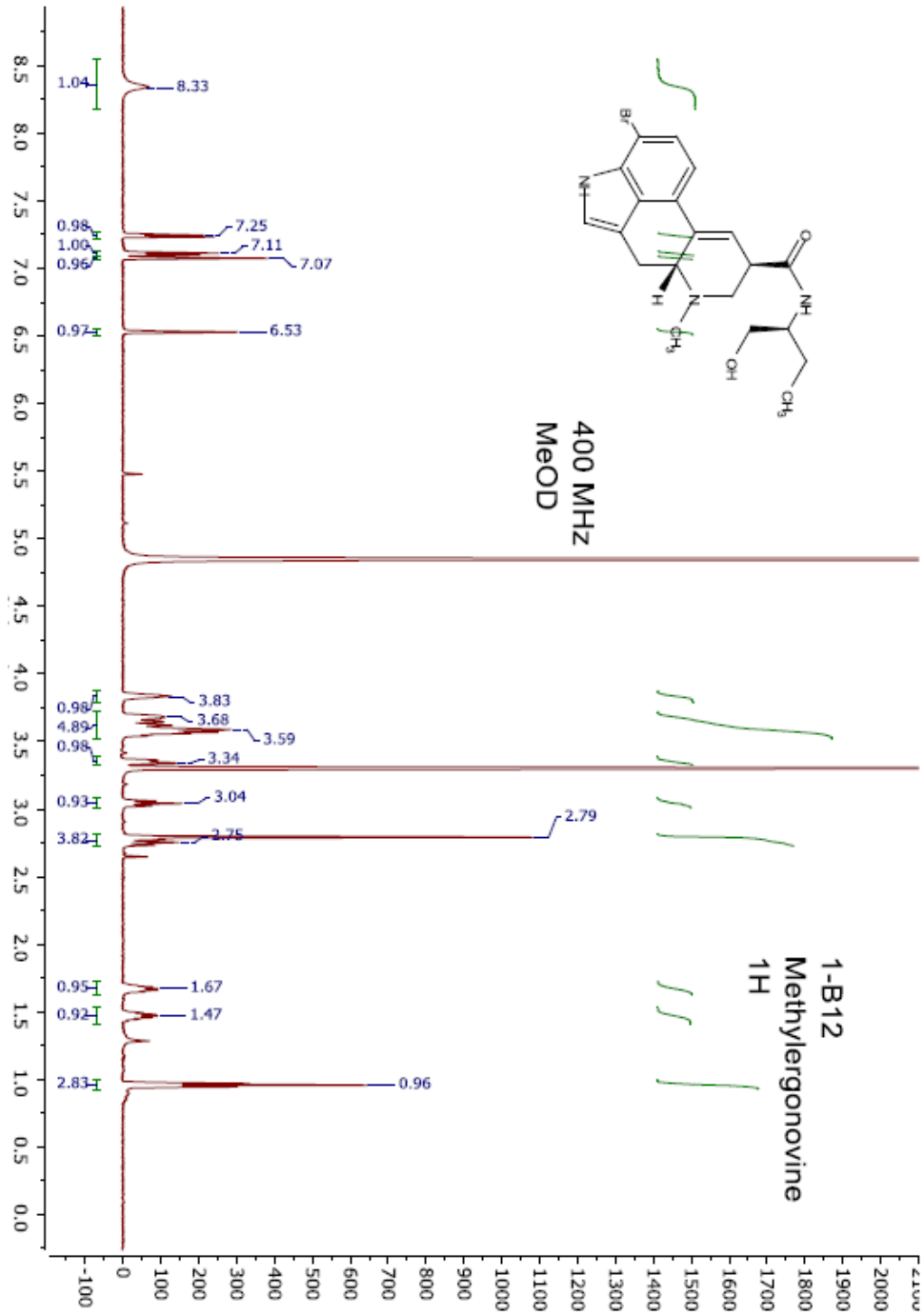


Figure AI. 10: 13CNMR of 7-bromomethylergonovine 1-B12 bioconversion

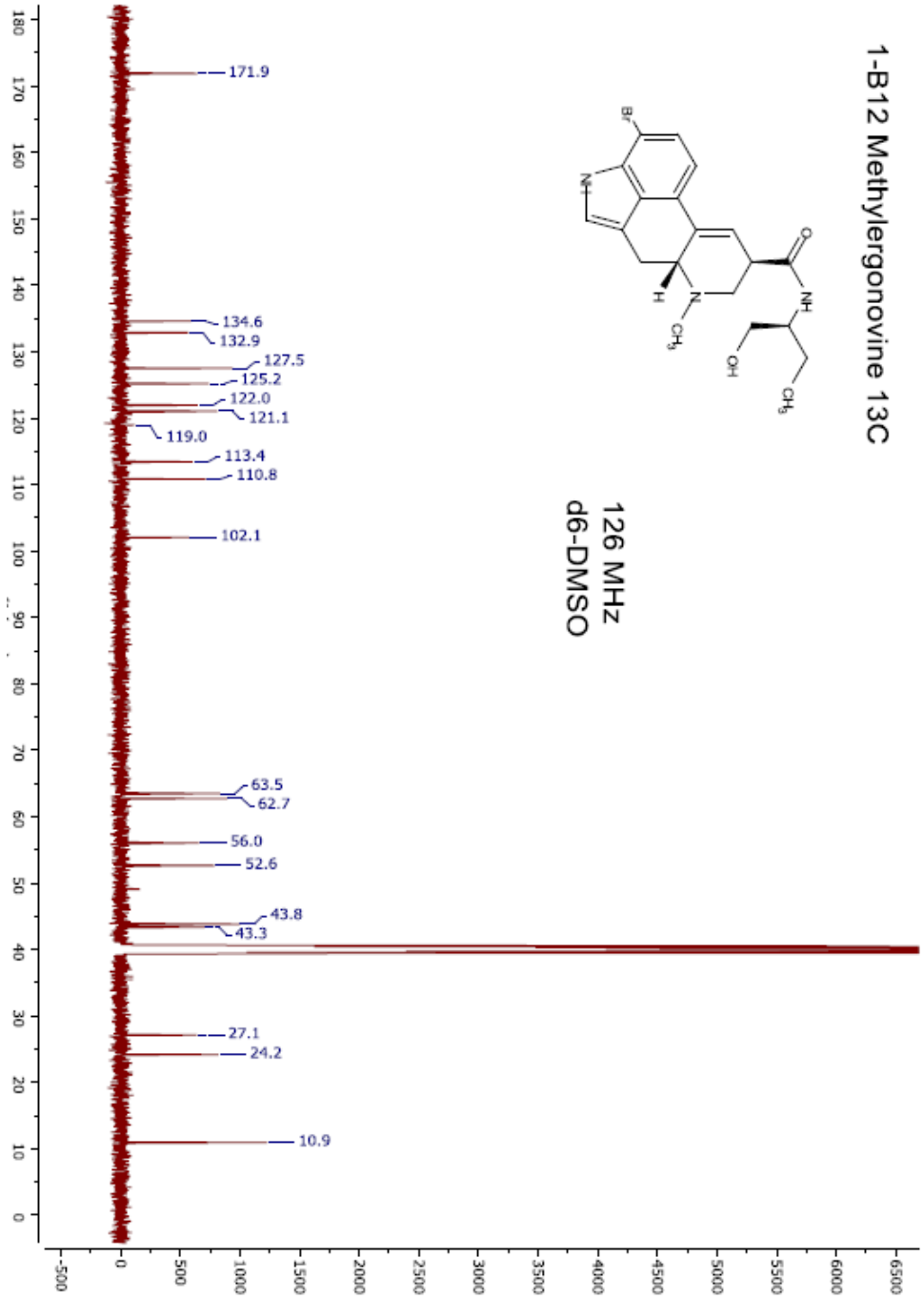


Figure AI. 11: 1HNMR of 2-bromomethylergonovine from 2-C01 bioconversion

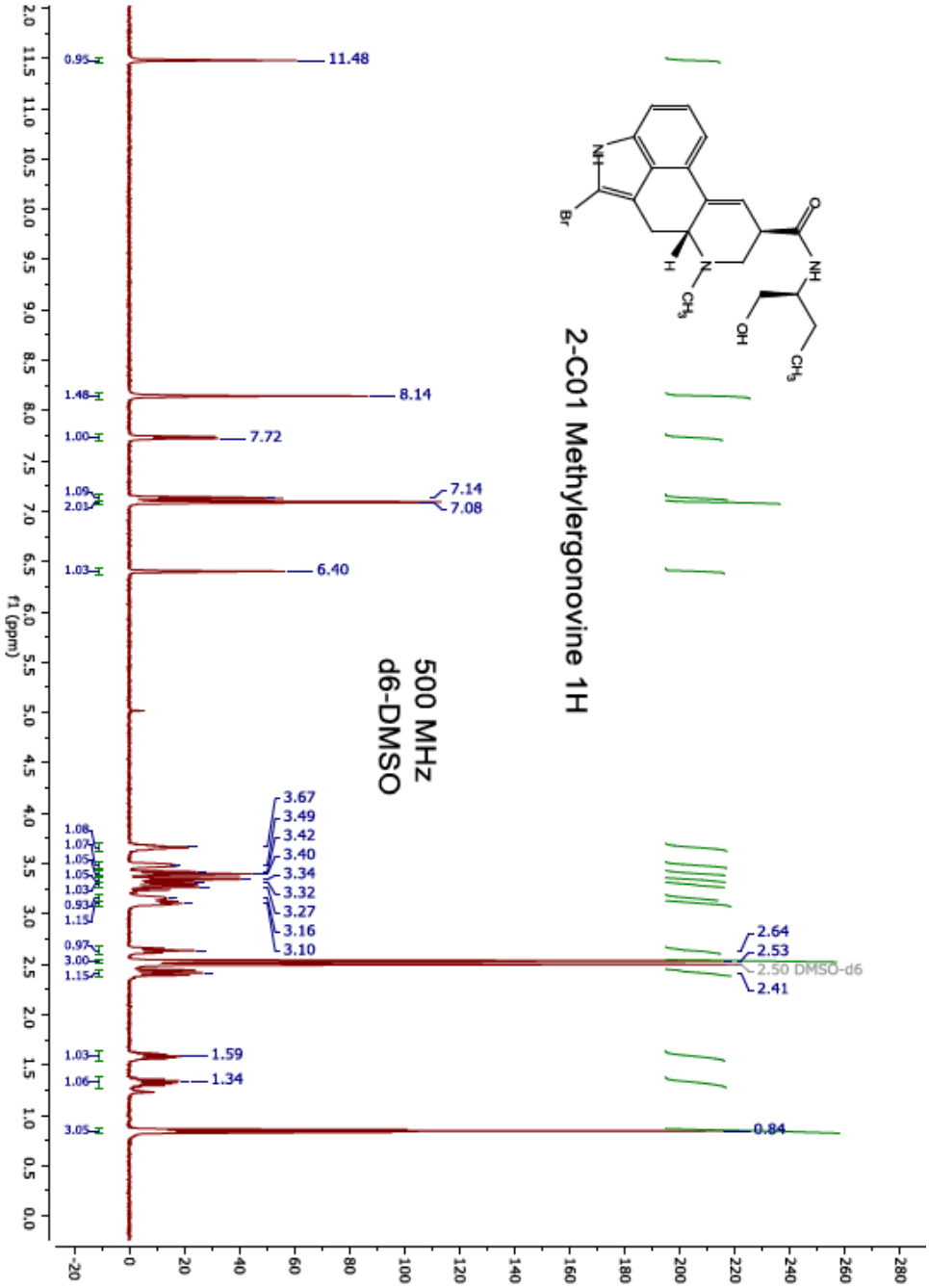


Figure AI. 12: 13CNMR of 2-bromomethylergonovine from 2-C01 bioconversion

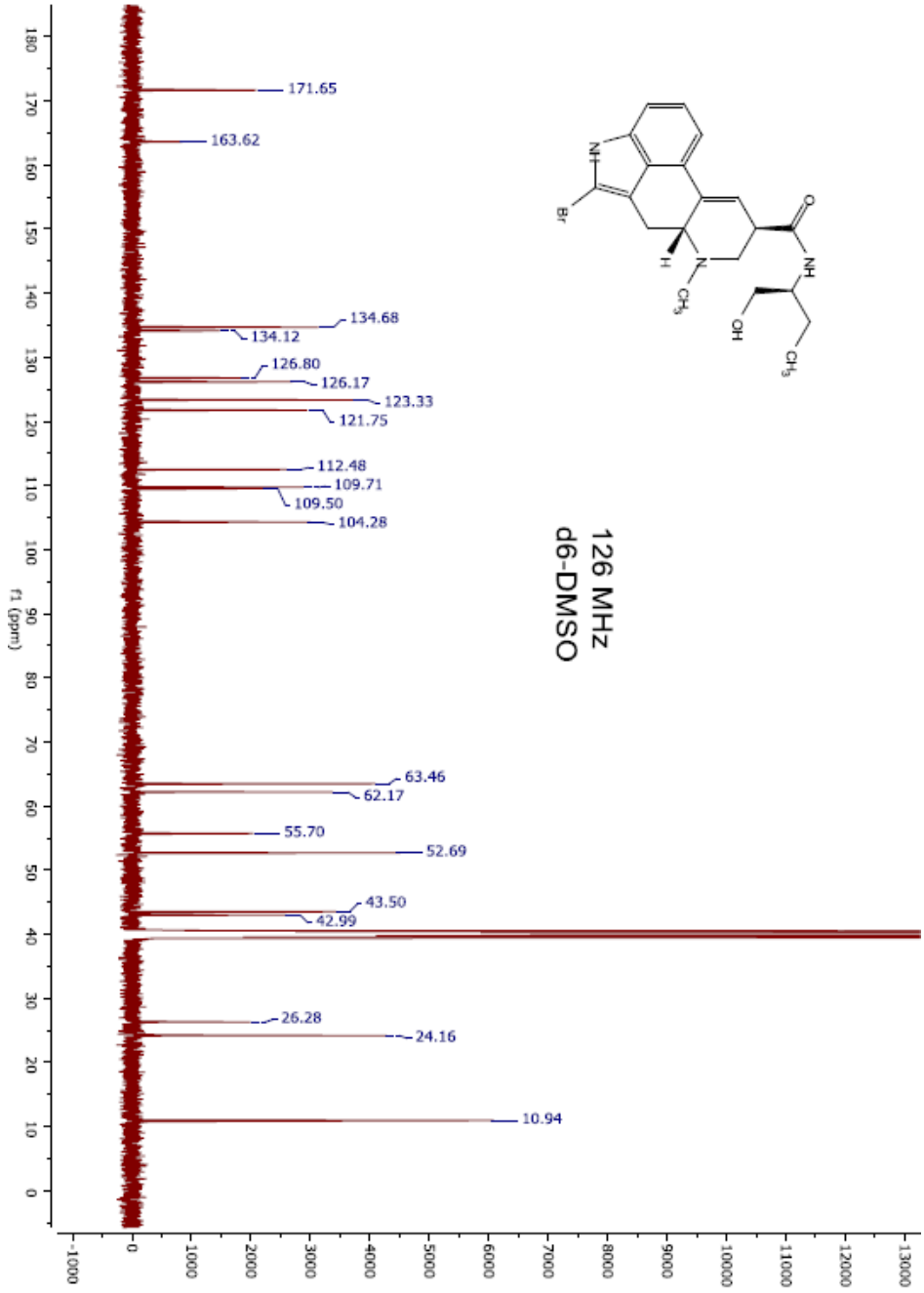


Figure AI. 13: 1HNMR of 3-bromo-AZ20 from 1-F08 bioconversion

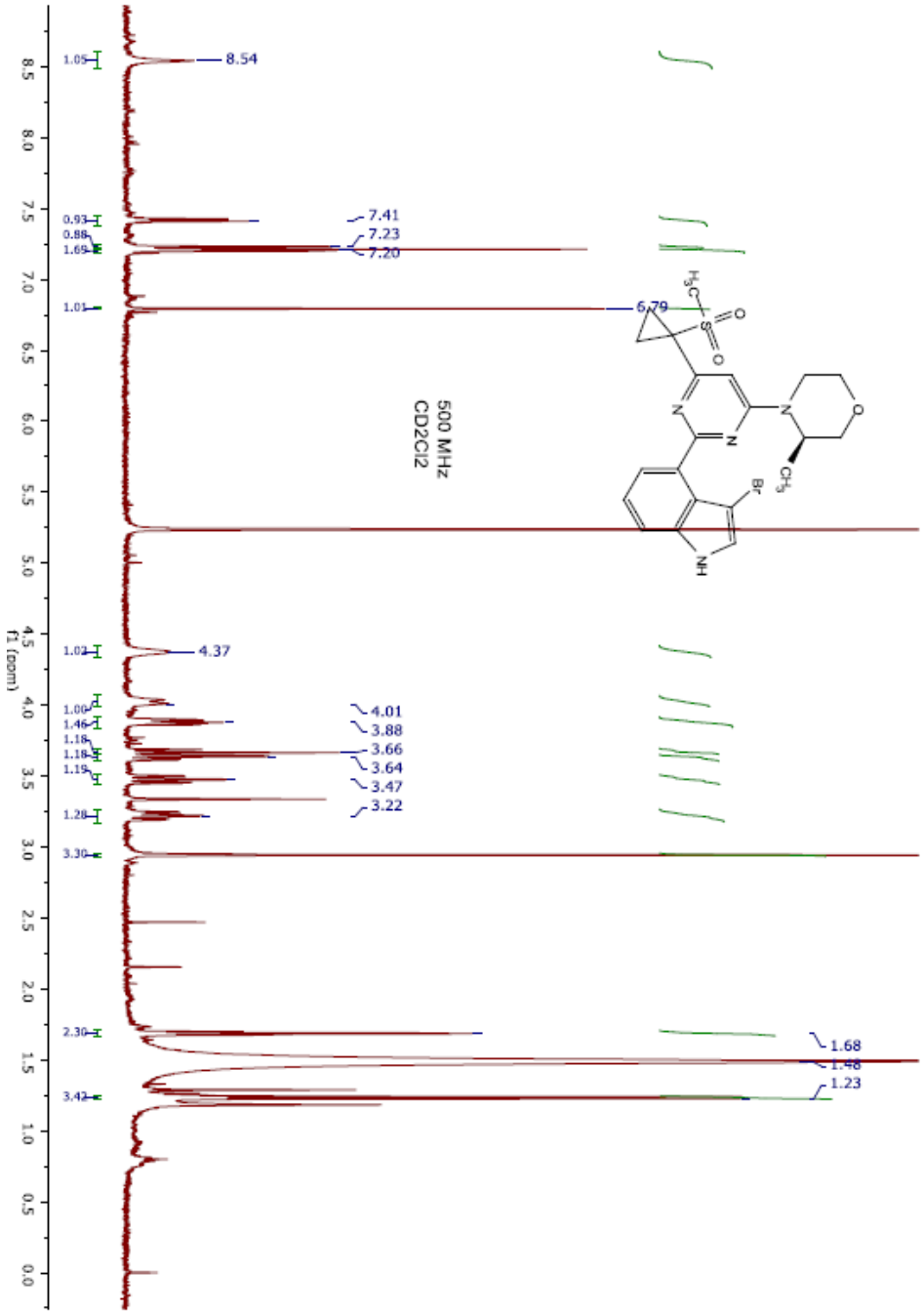


Figure AI. 14: 13CNMR of 3-bromo-AZ20 from 1-F08 bioconversion

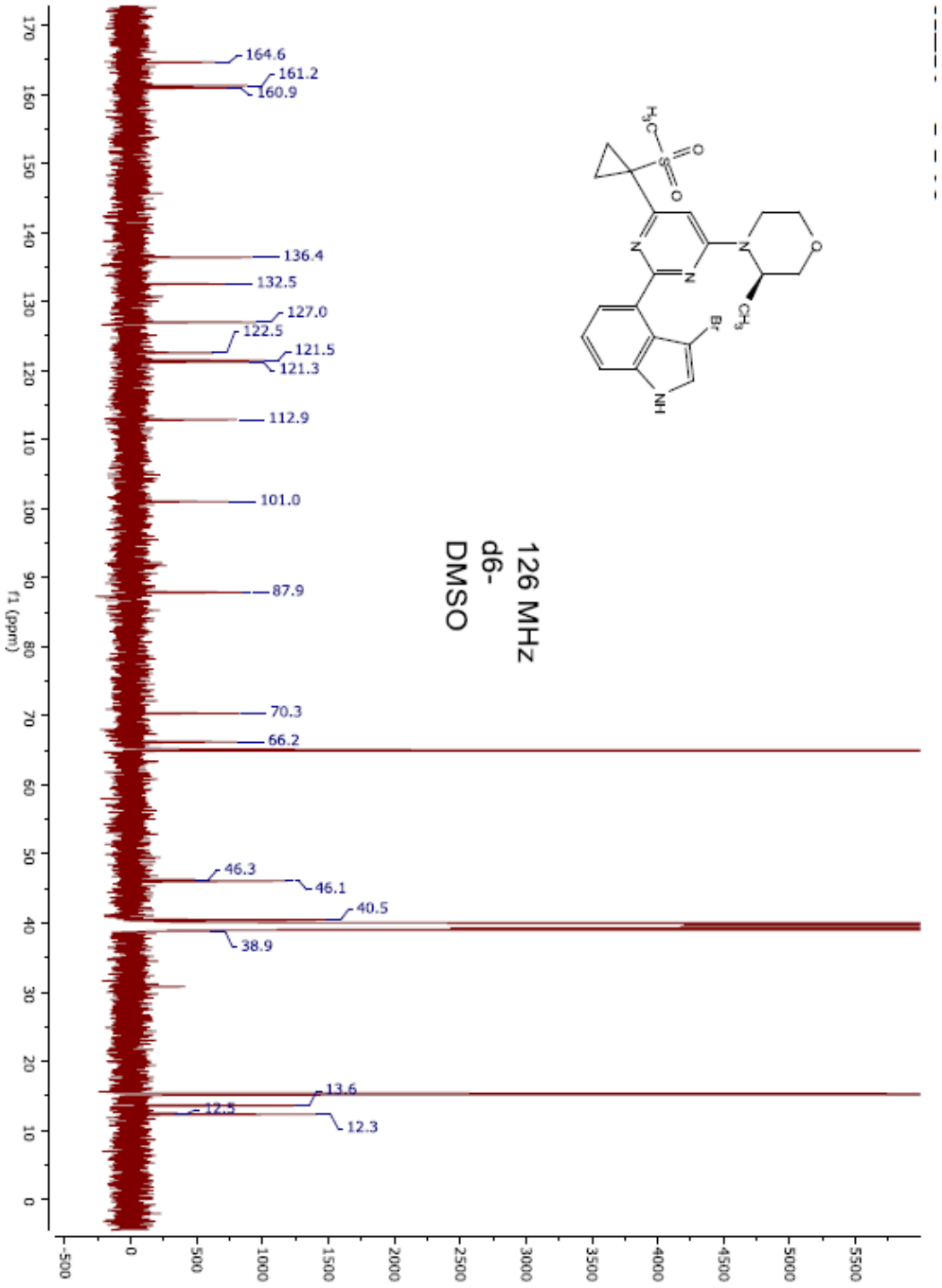


Figure AI. 15: 1HNMR of 5-bromopremalbrancheamide from 1-F08 bioconversion

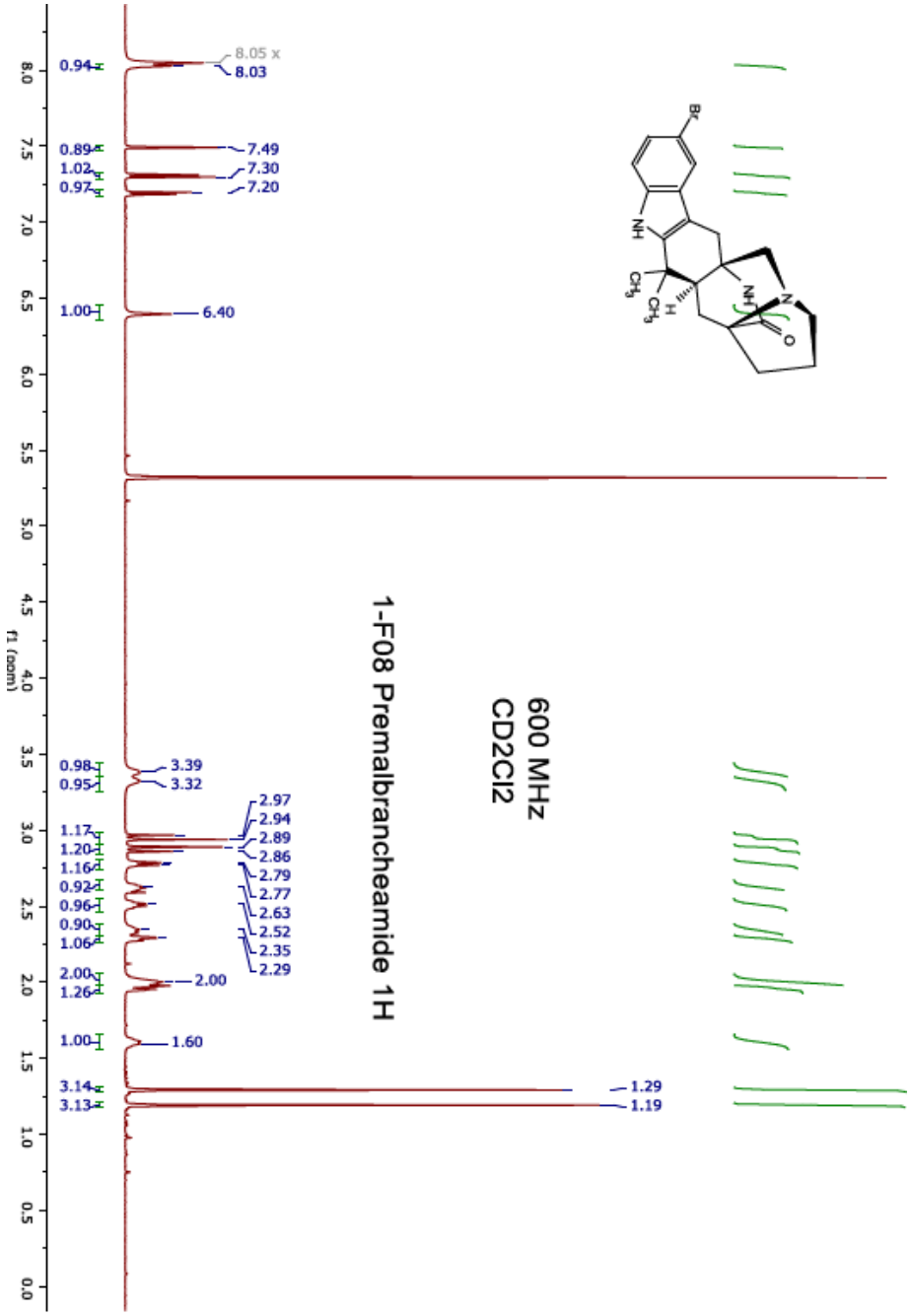


Figure AI. 16: 13CNMR of 5-bromopremalbrancheamide from 1-F08 bioconversion

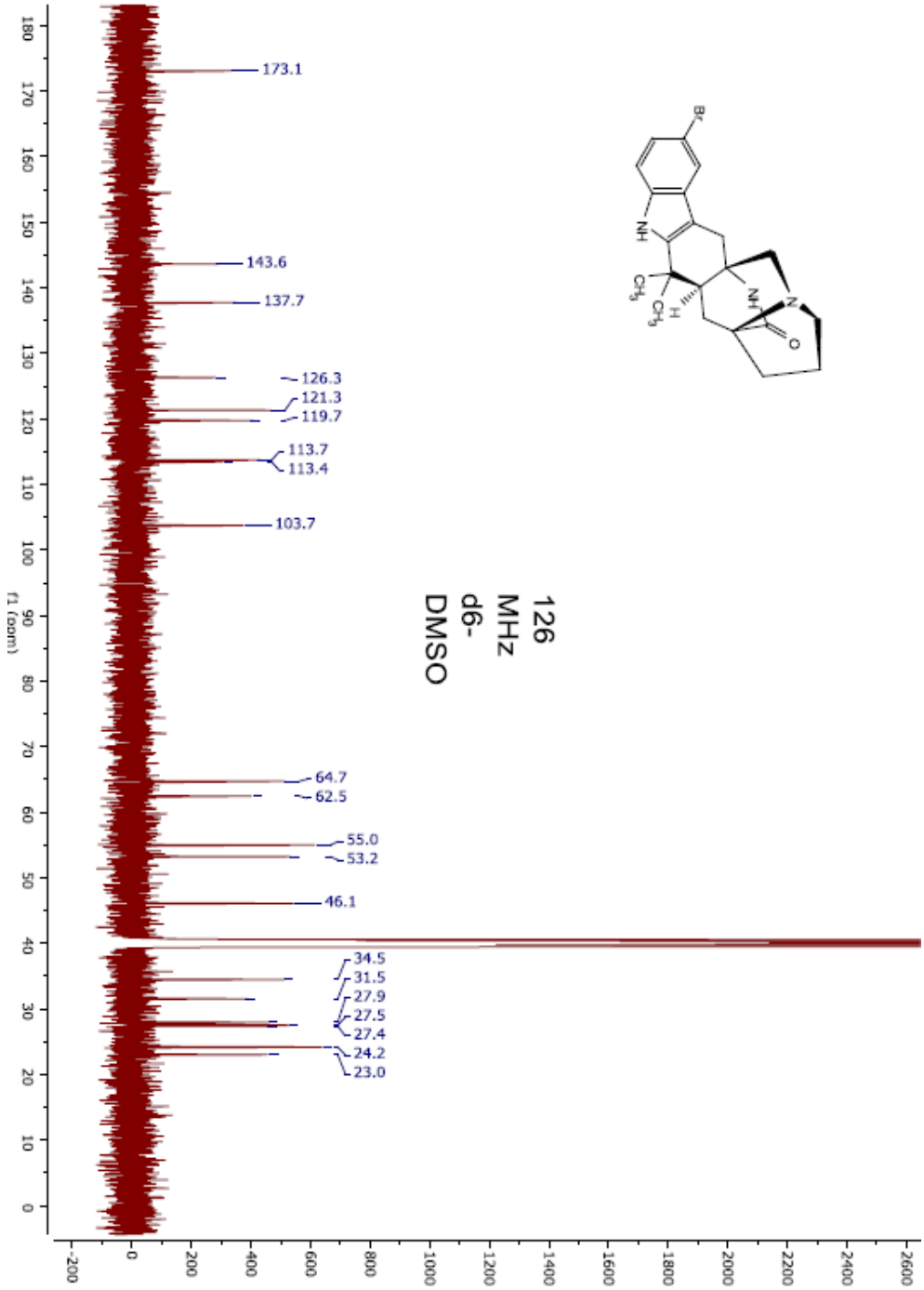


Figure AI. 17: 1HNMR of 4-bromo-17β-estradiol from 1-F11 bioconversion

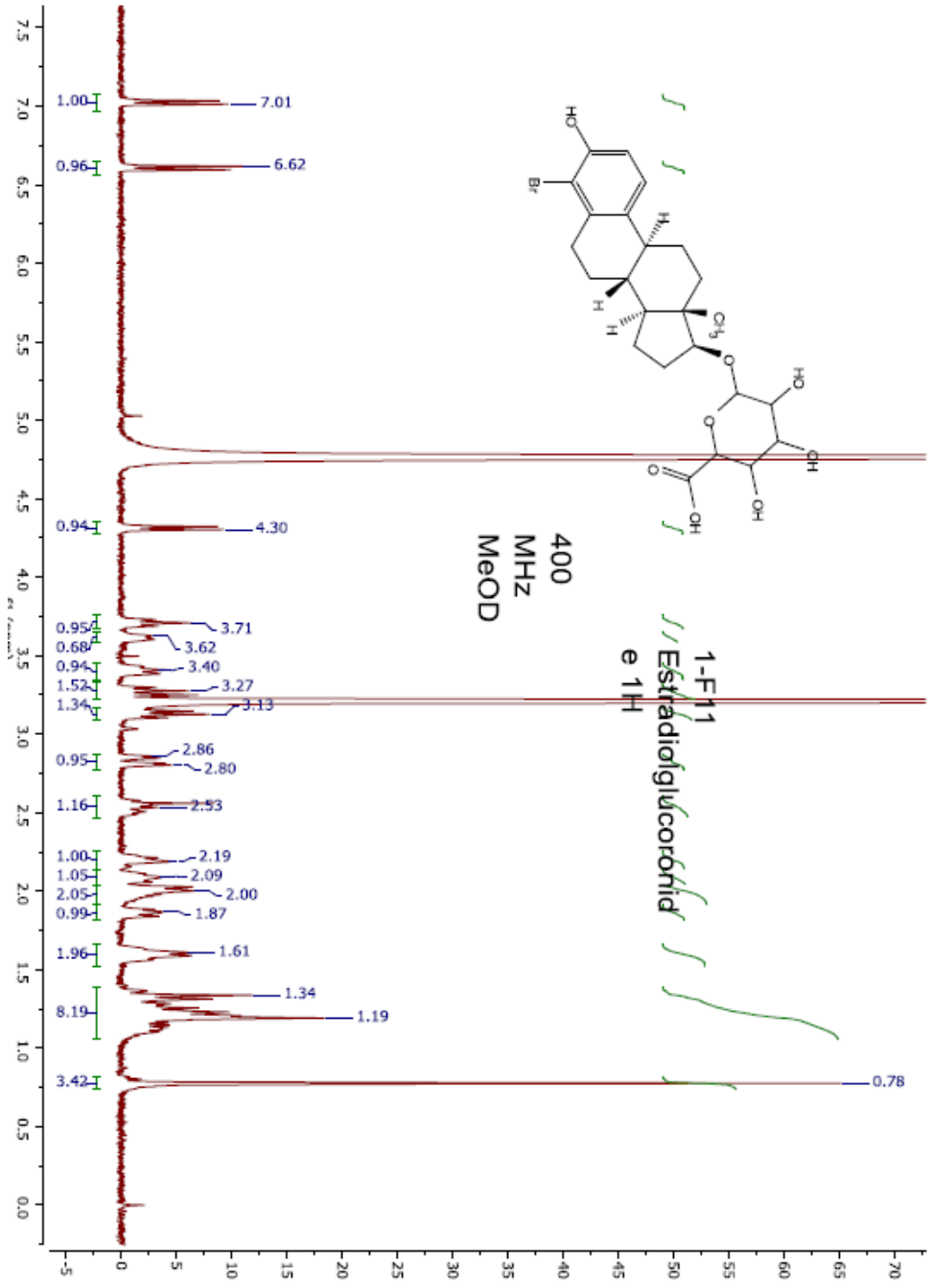


Figure AI. 18: 13CNMR of 4-bromo-17β-estradiol from 1-F11 bioconversion

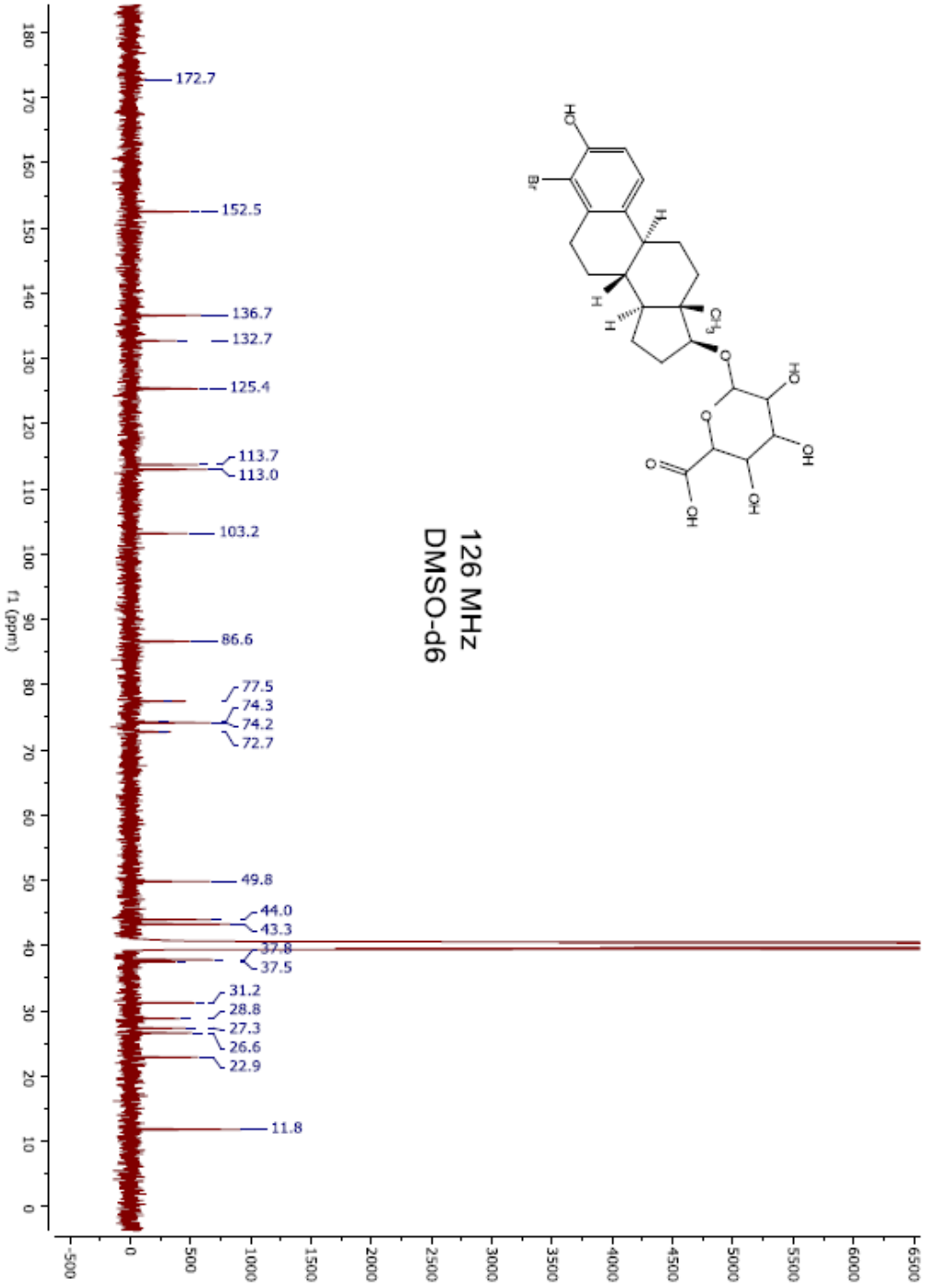


Figure AI. 19: 1HNMR of 4-(1-(4-aminophenyl)-2,2-dimethylpropyl)-2-bromoaniline from R2D3 bioconversion

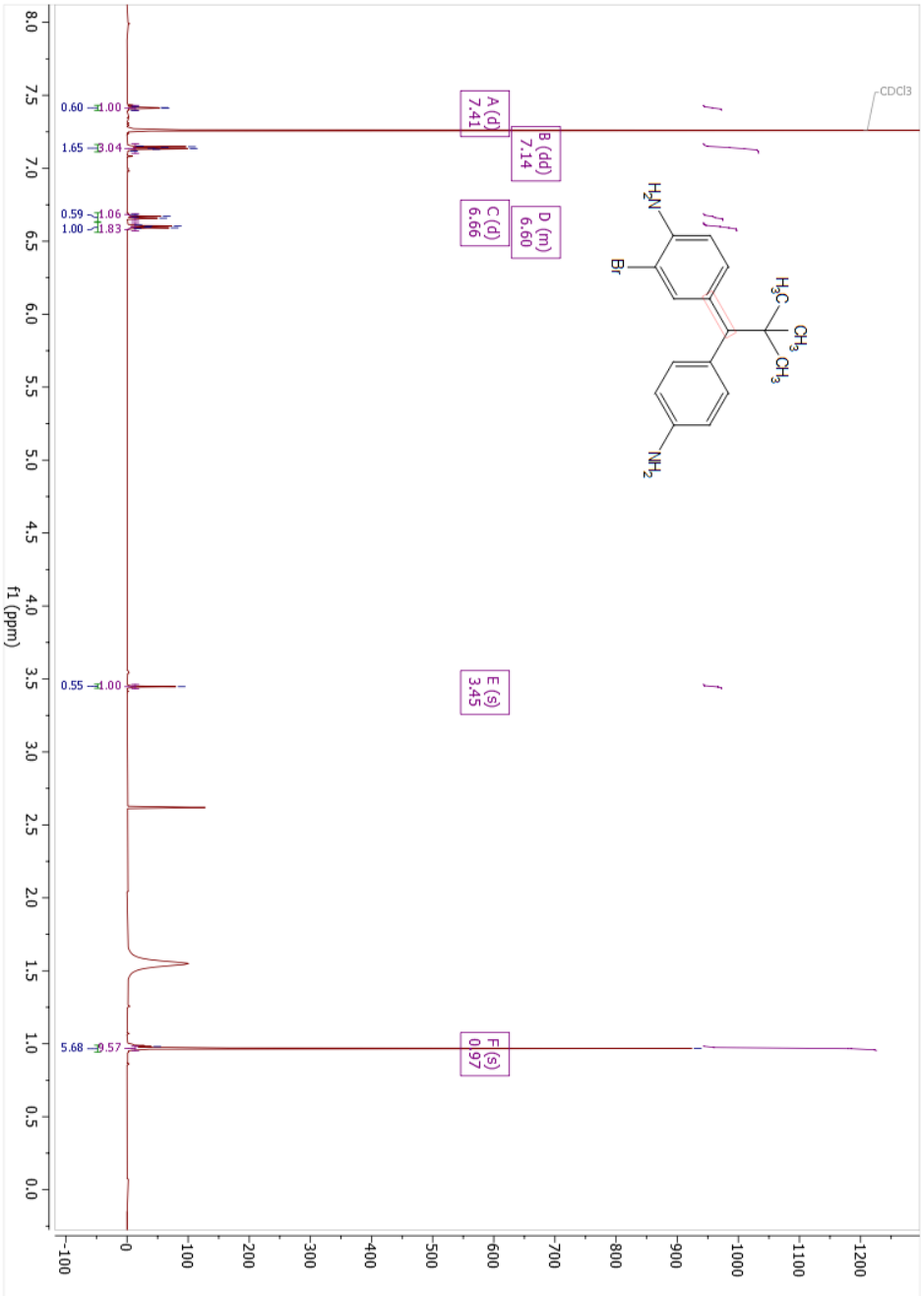


Figure AI. 20: 1HNMR of 4,4'-(2,2-dimethylpropane-1,1-diyl)bis(2-bromoaniline)

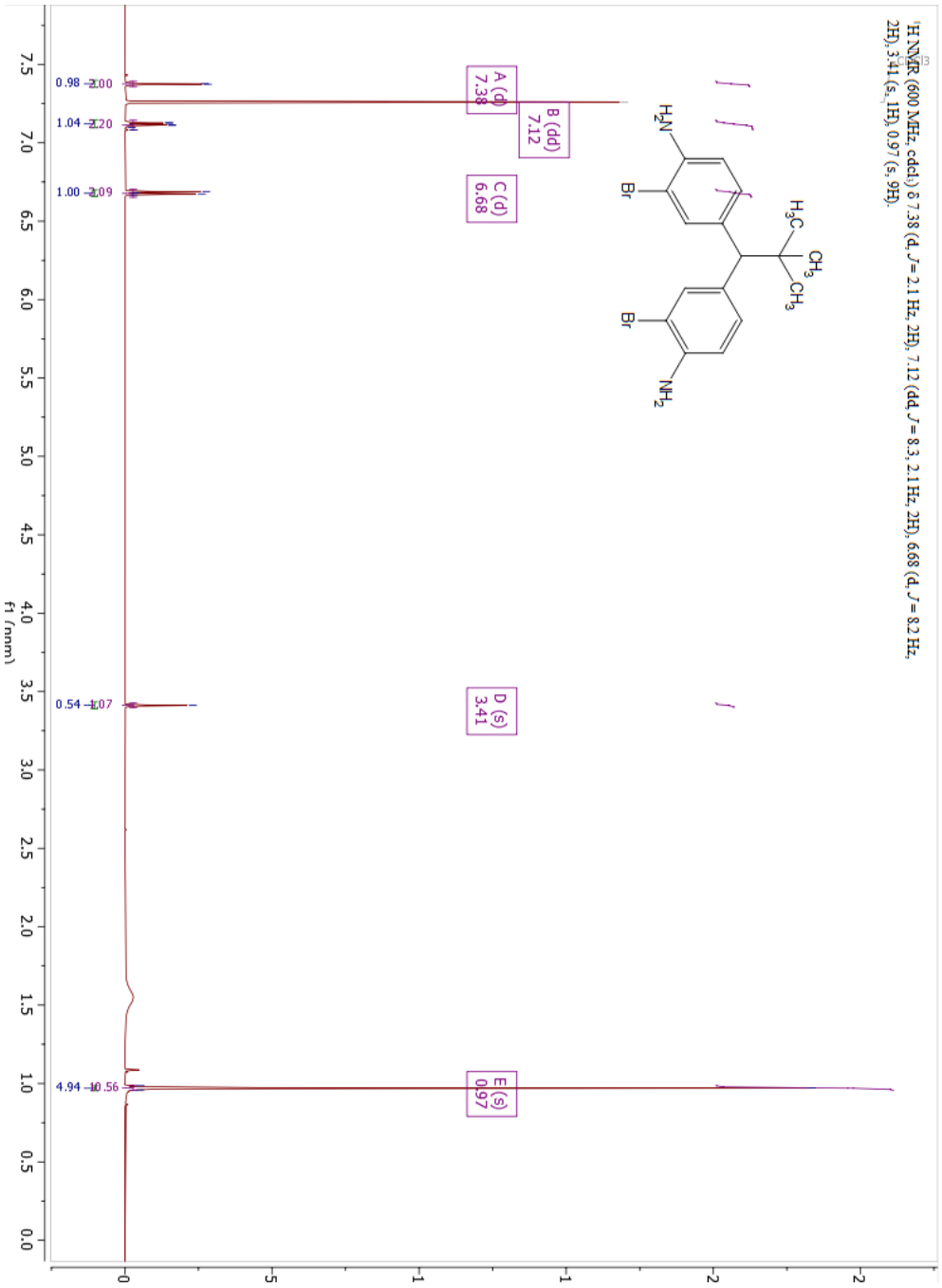
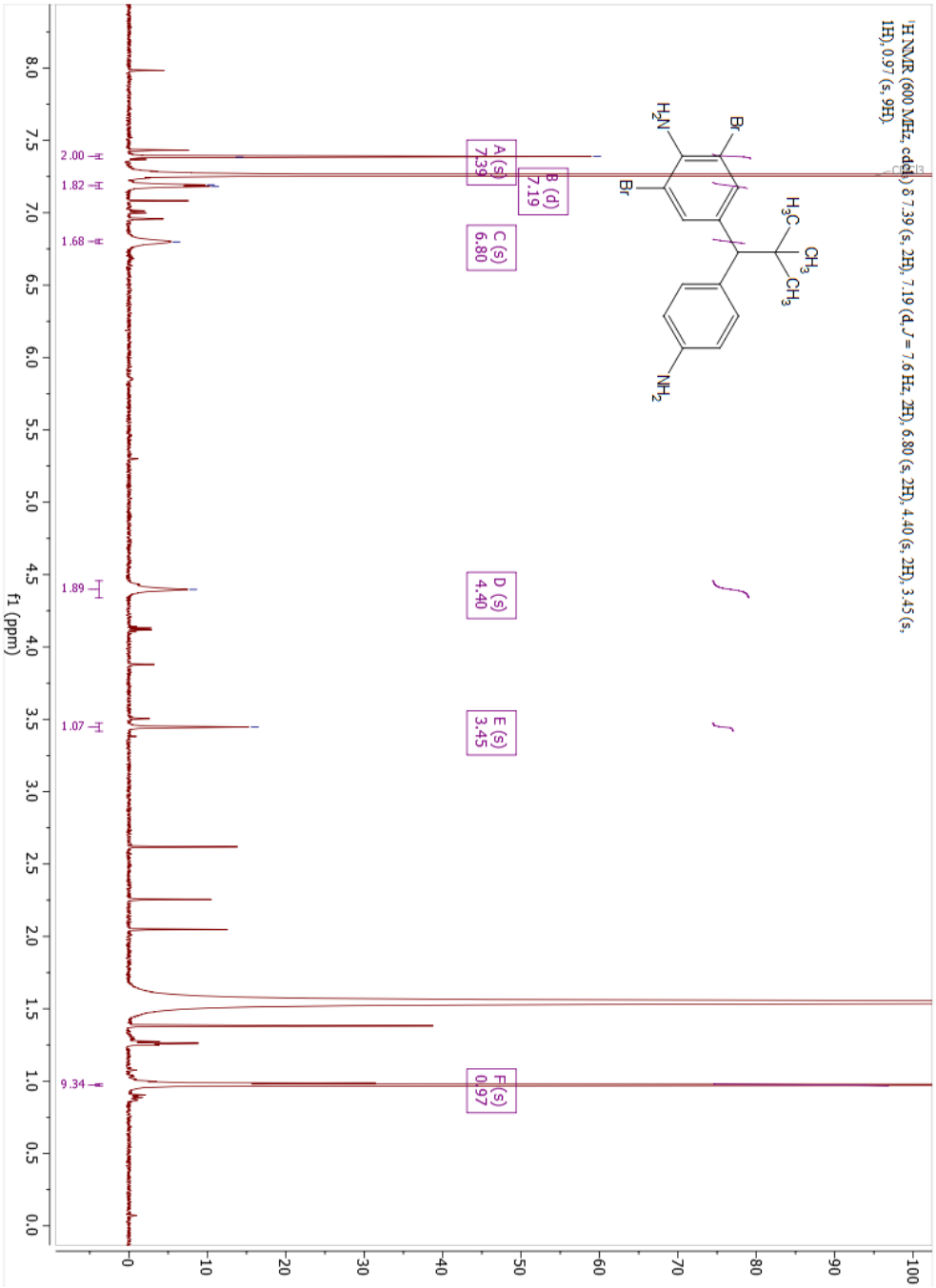


Figure AI. 21: 1HNMR of 4-(1-(4-aminophenyl)-2,2-dimethylpropyl)-2,6-dibromoaniline



Note: Product validation not completed.

APPENDIX II

NMR Spectra for Compounds from Chapter 3

Figure AII. 1: 1HNMR for Nitro compound N1

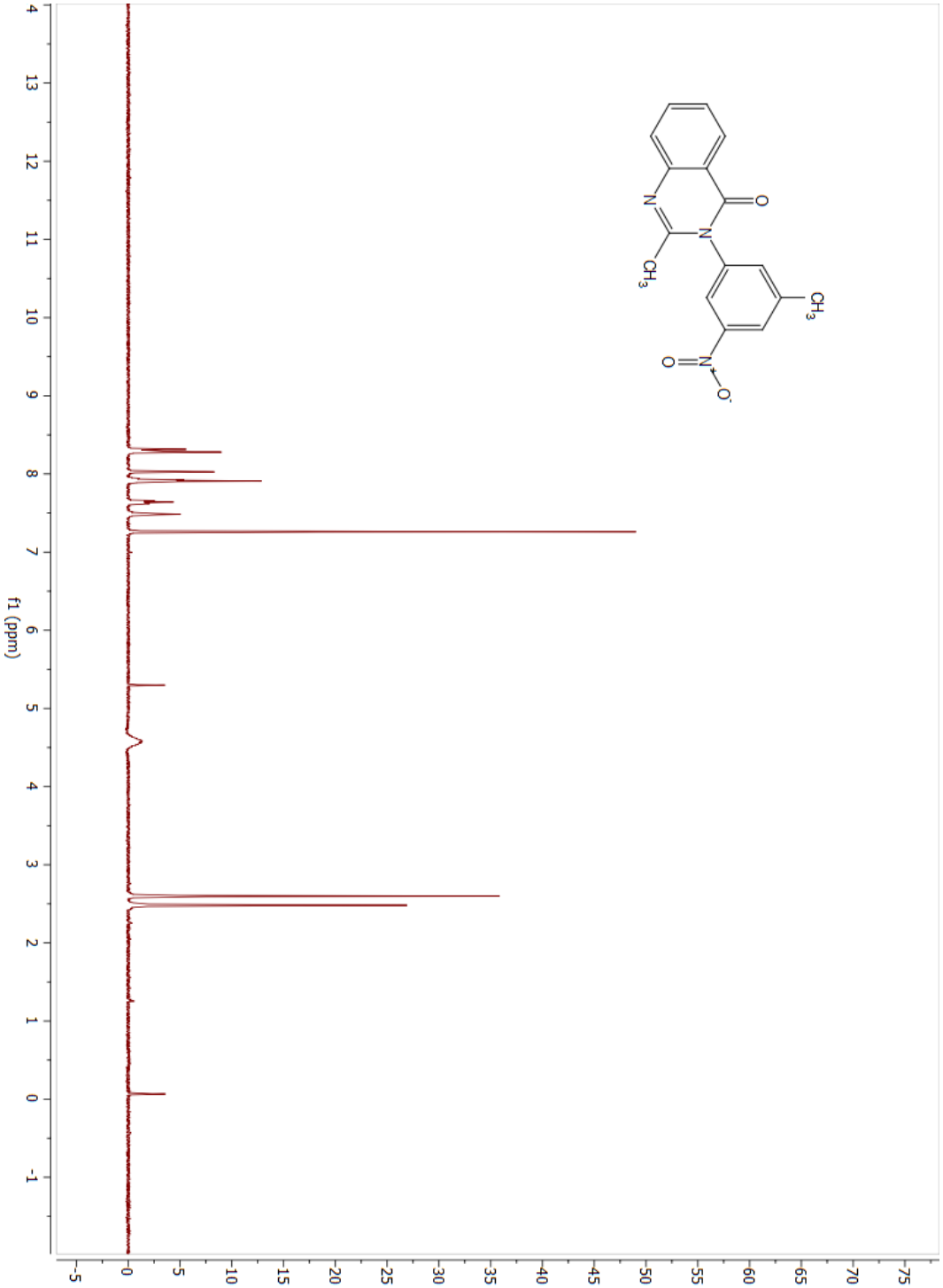


Figure AII. 2: 13CNMR for Nitro compound N1

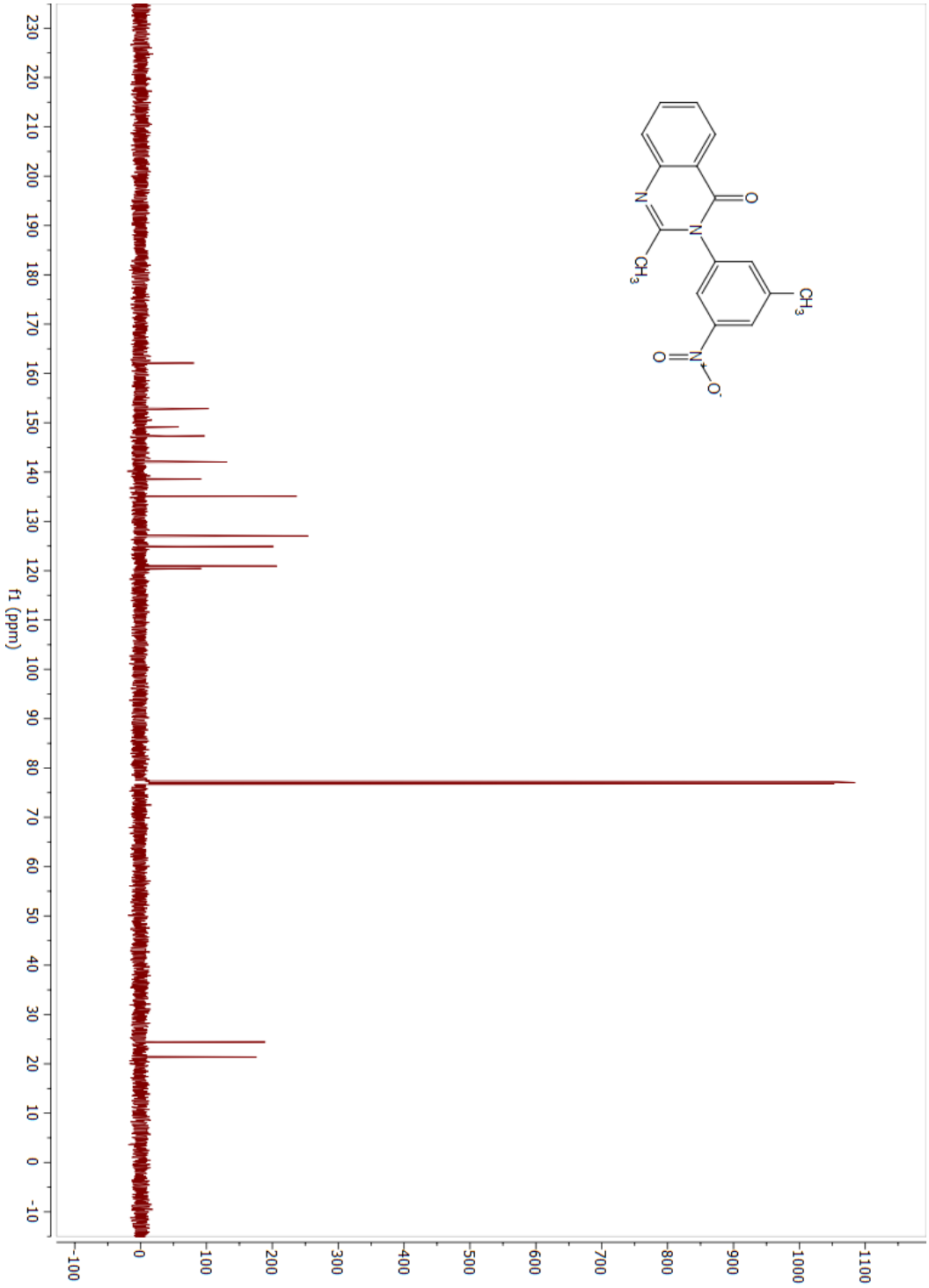


Figure AII. 3: 1HNMR Substrate 1

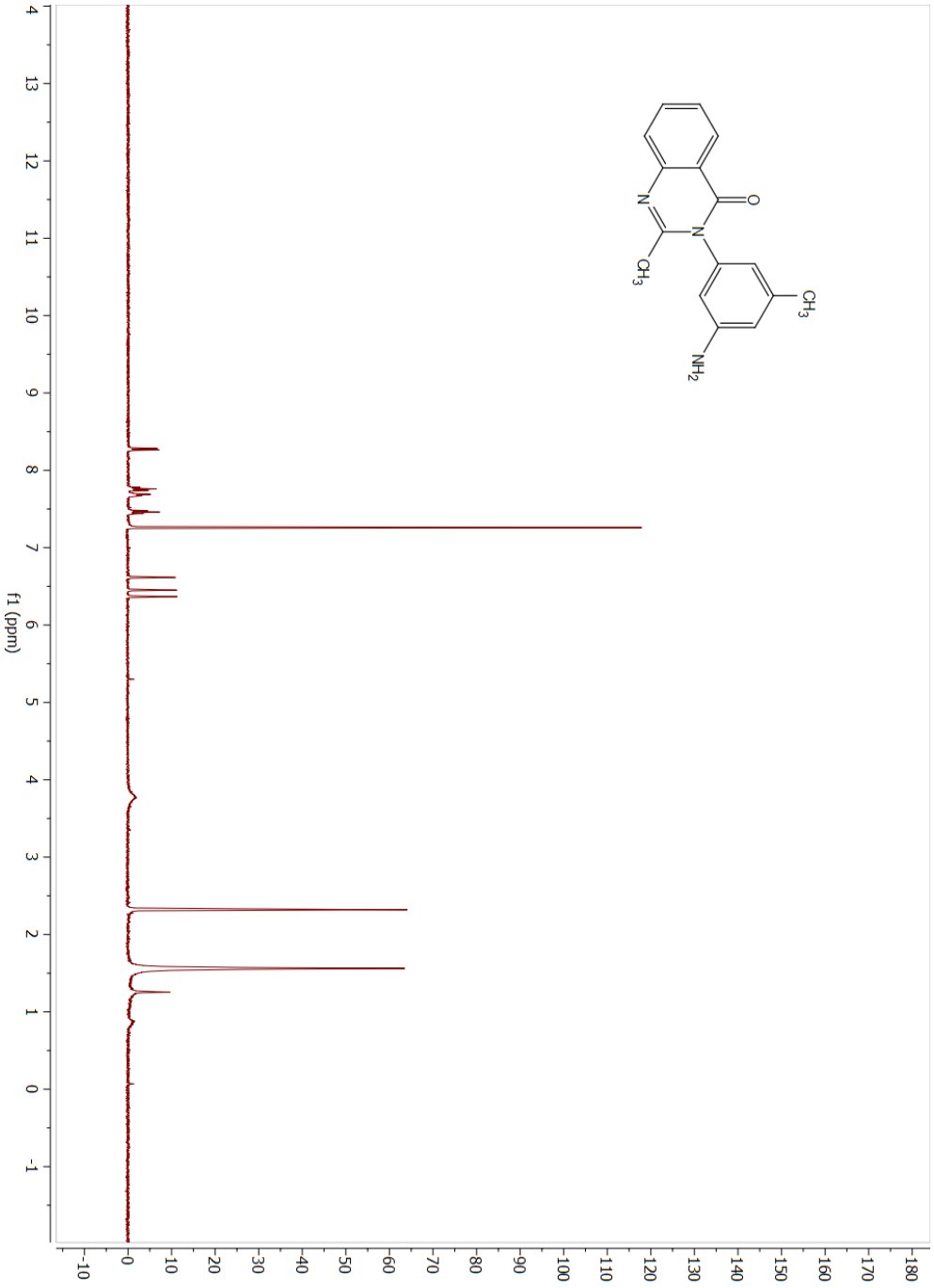


Figure AII. 4: 13CNMR Substrate 1

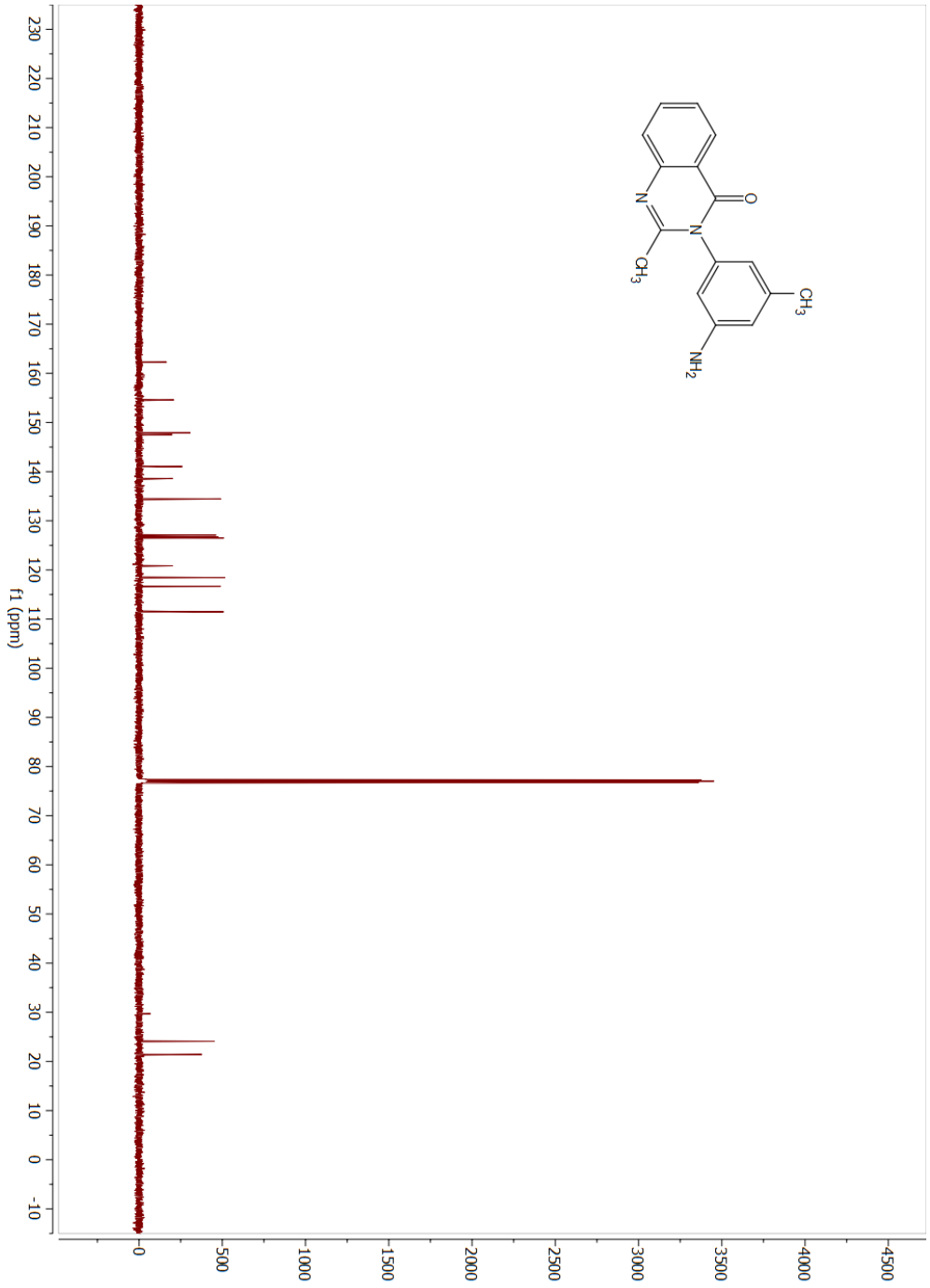


Figure AII. 5: 1HNMR Substrate 2

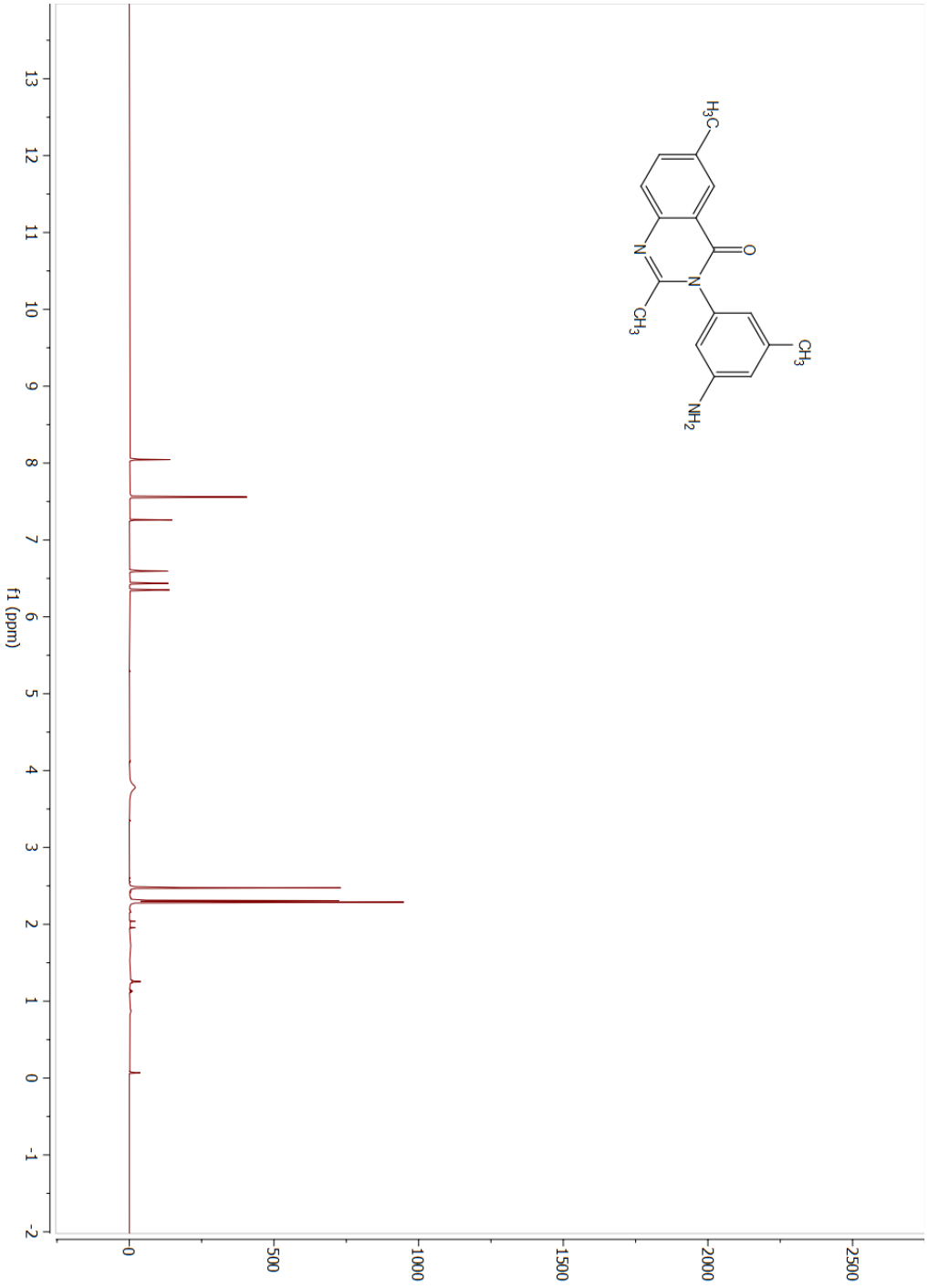


Figure AII. 6: 13CNMR Substrate 2

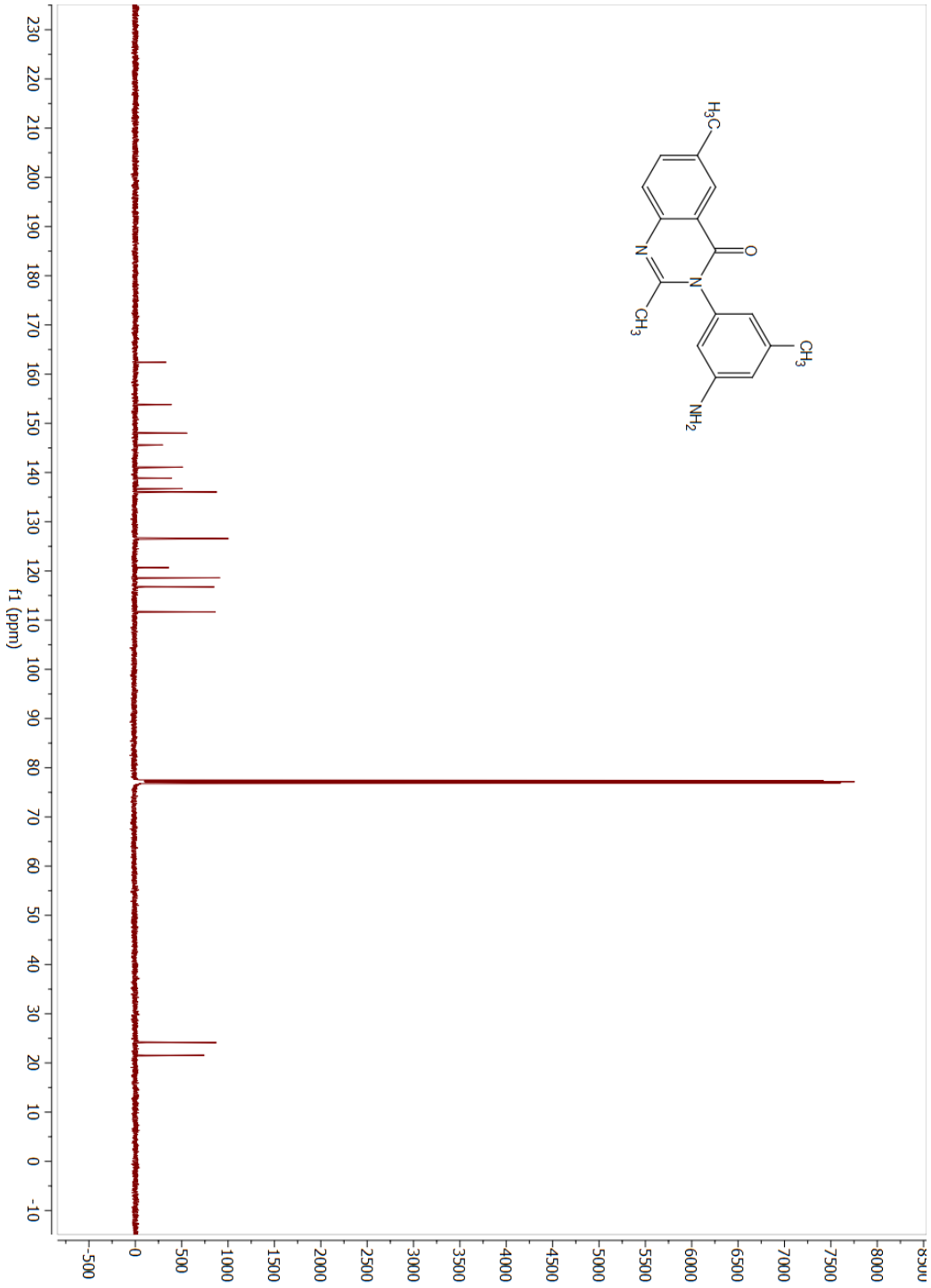


Figure AII. 7: 1HNMR Substrate 3

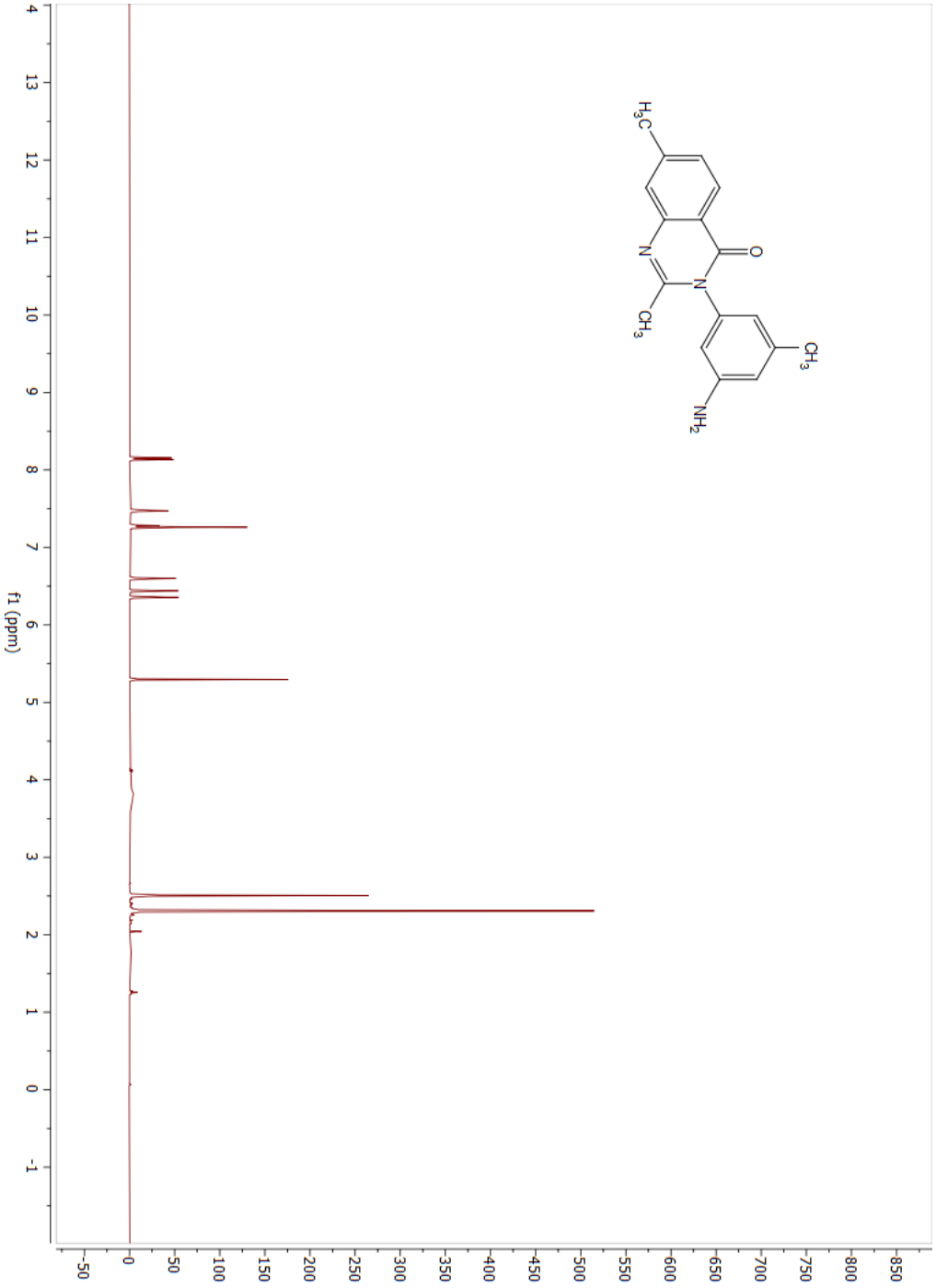


Figure AII. 8: 13CNMR Substrate 3

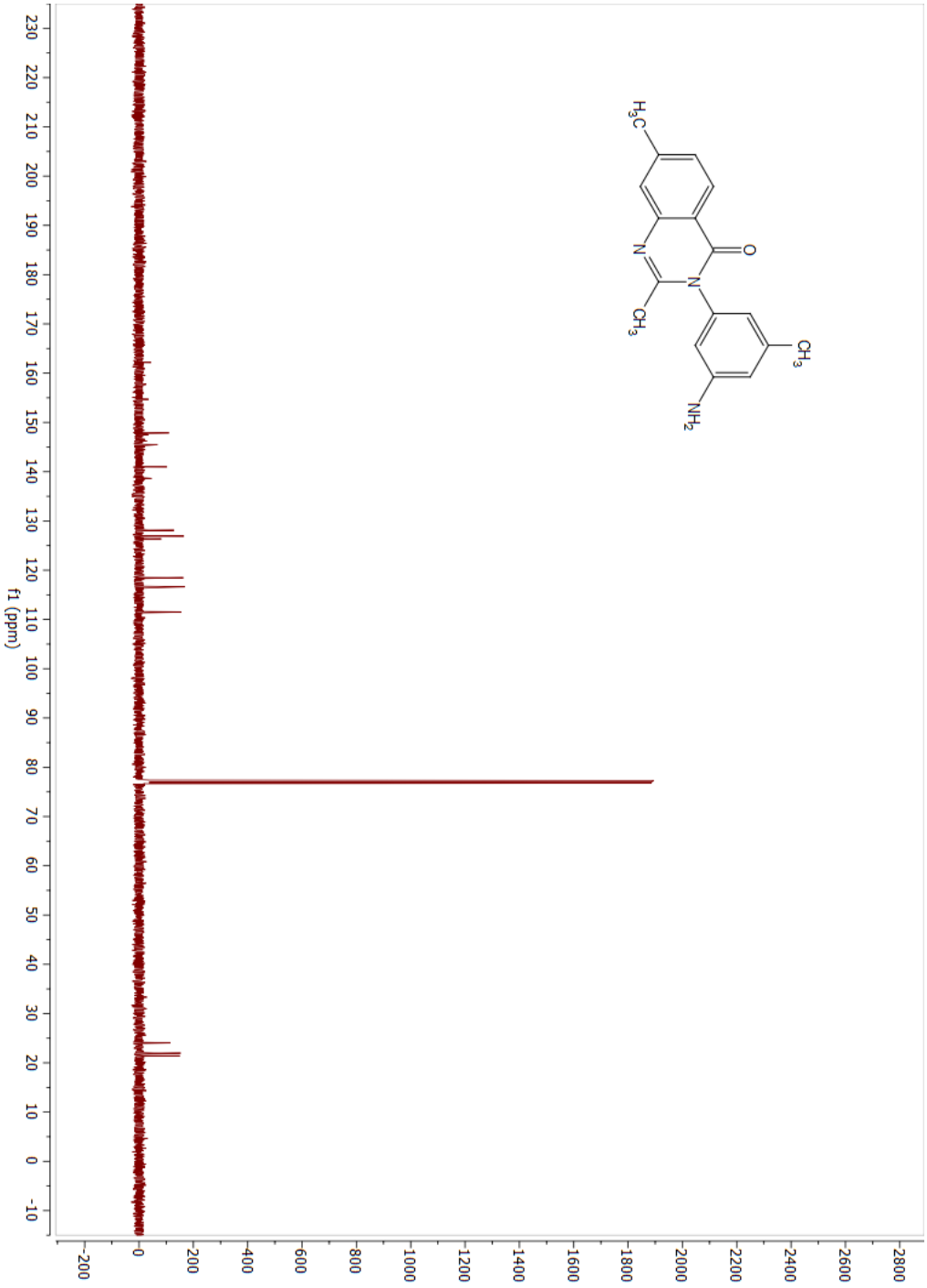


Figure AII. 9: 1HNMR Substrate 4

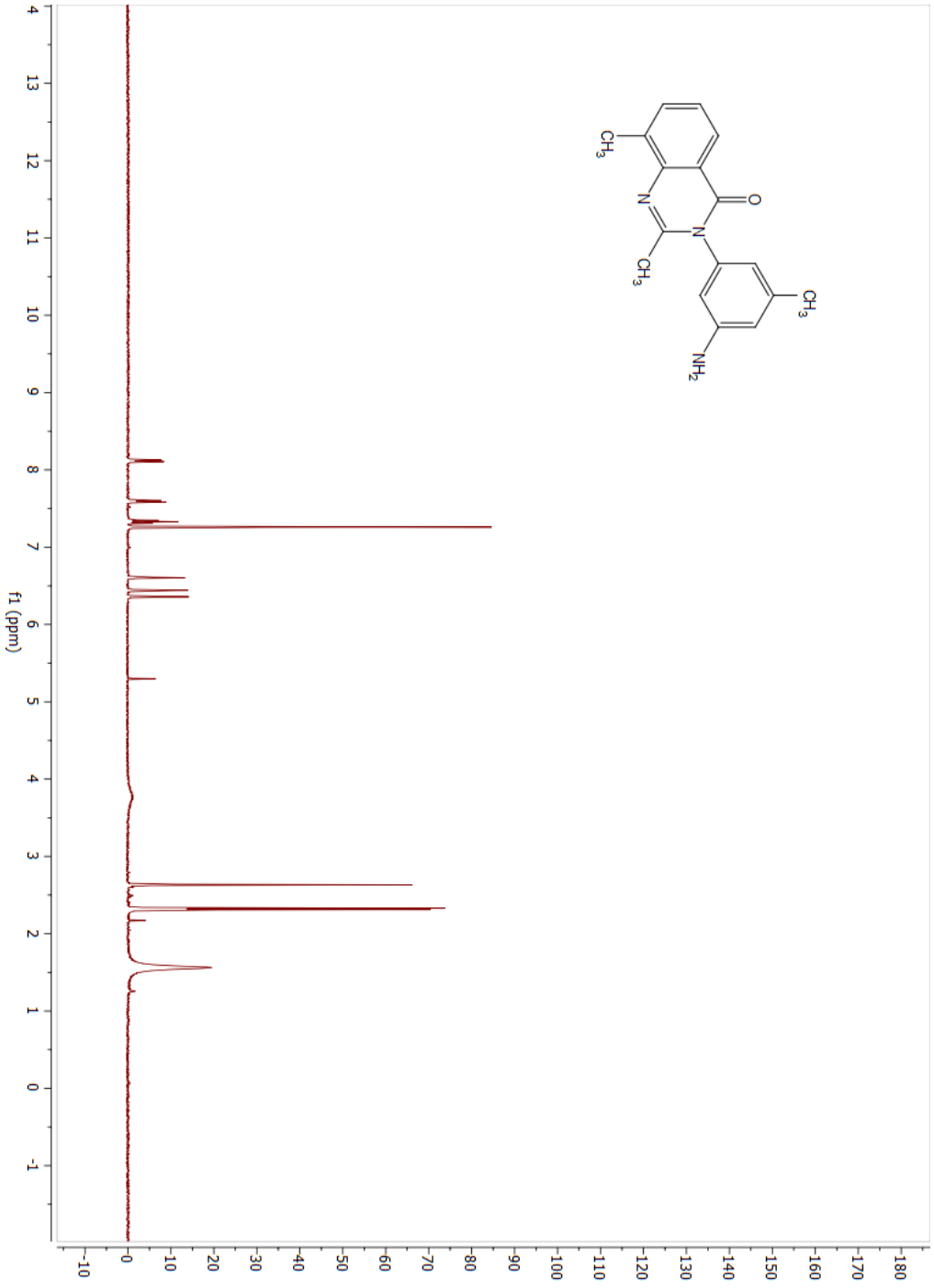


Figure AII. 10: 13CNMR Substrate 4

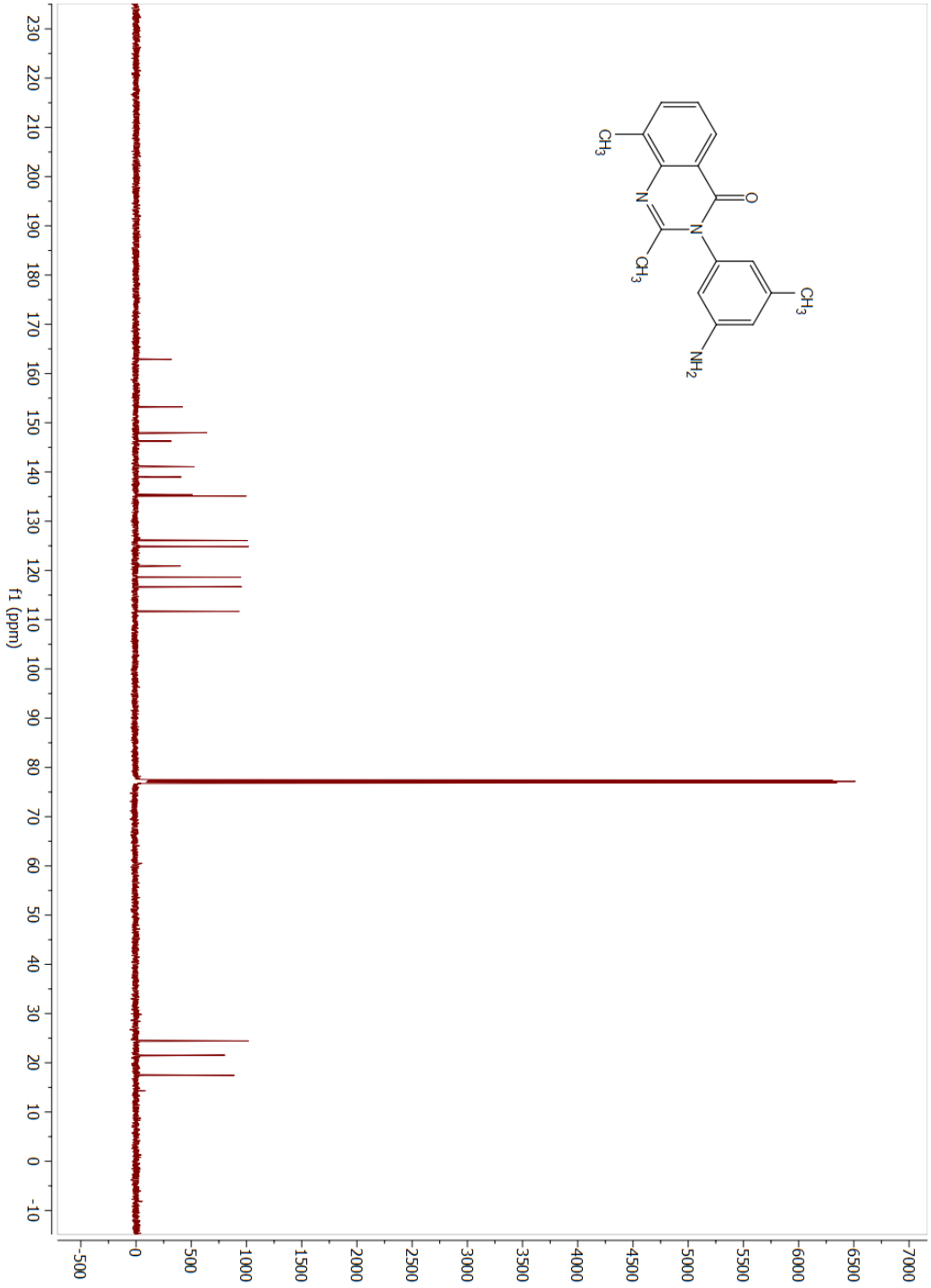


Figure AII. 11: 1HNMR Substrate 5

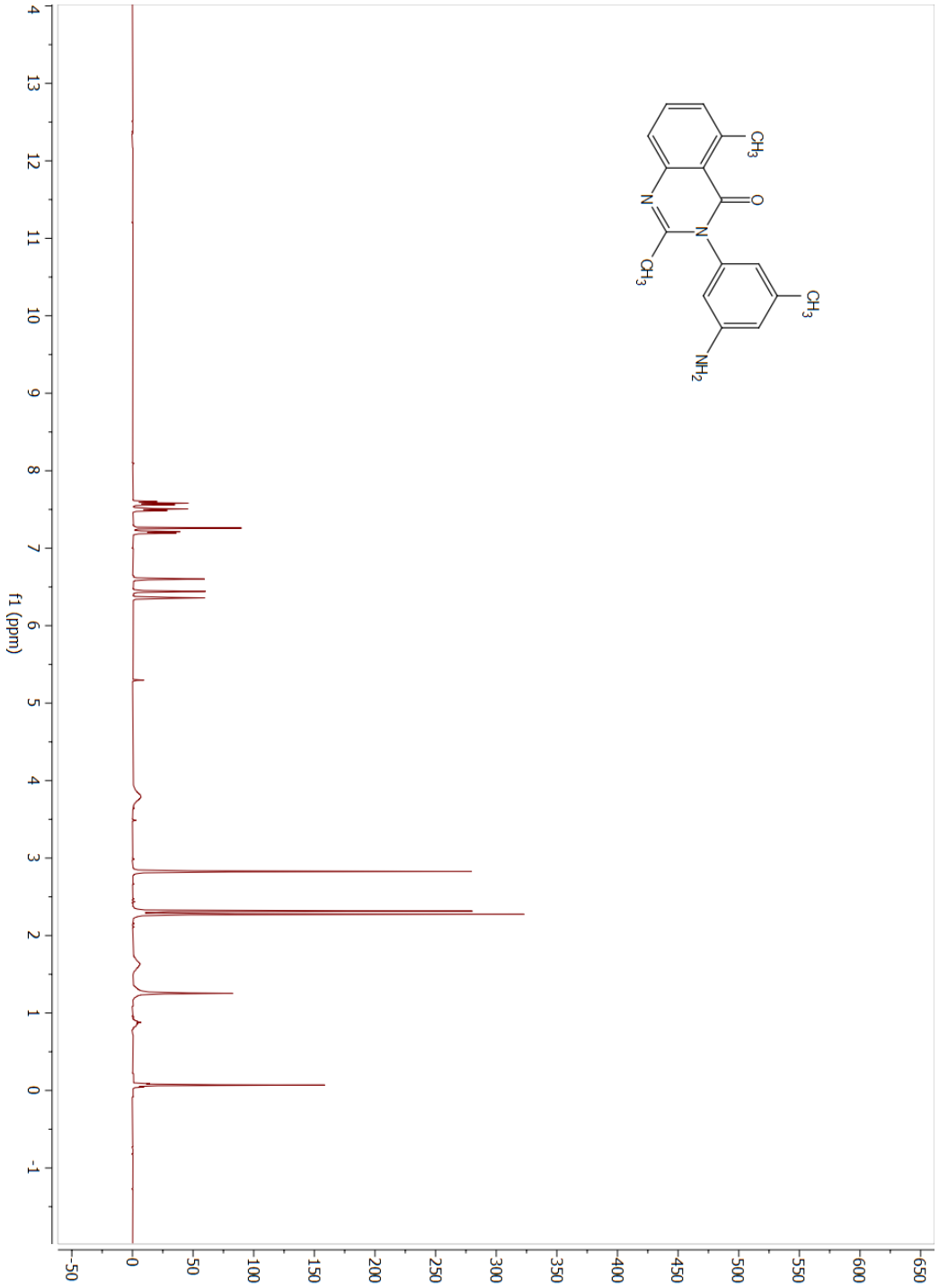


Figure AII. 12: 13CNMR Substrate 5

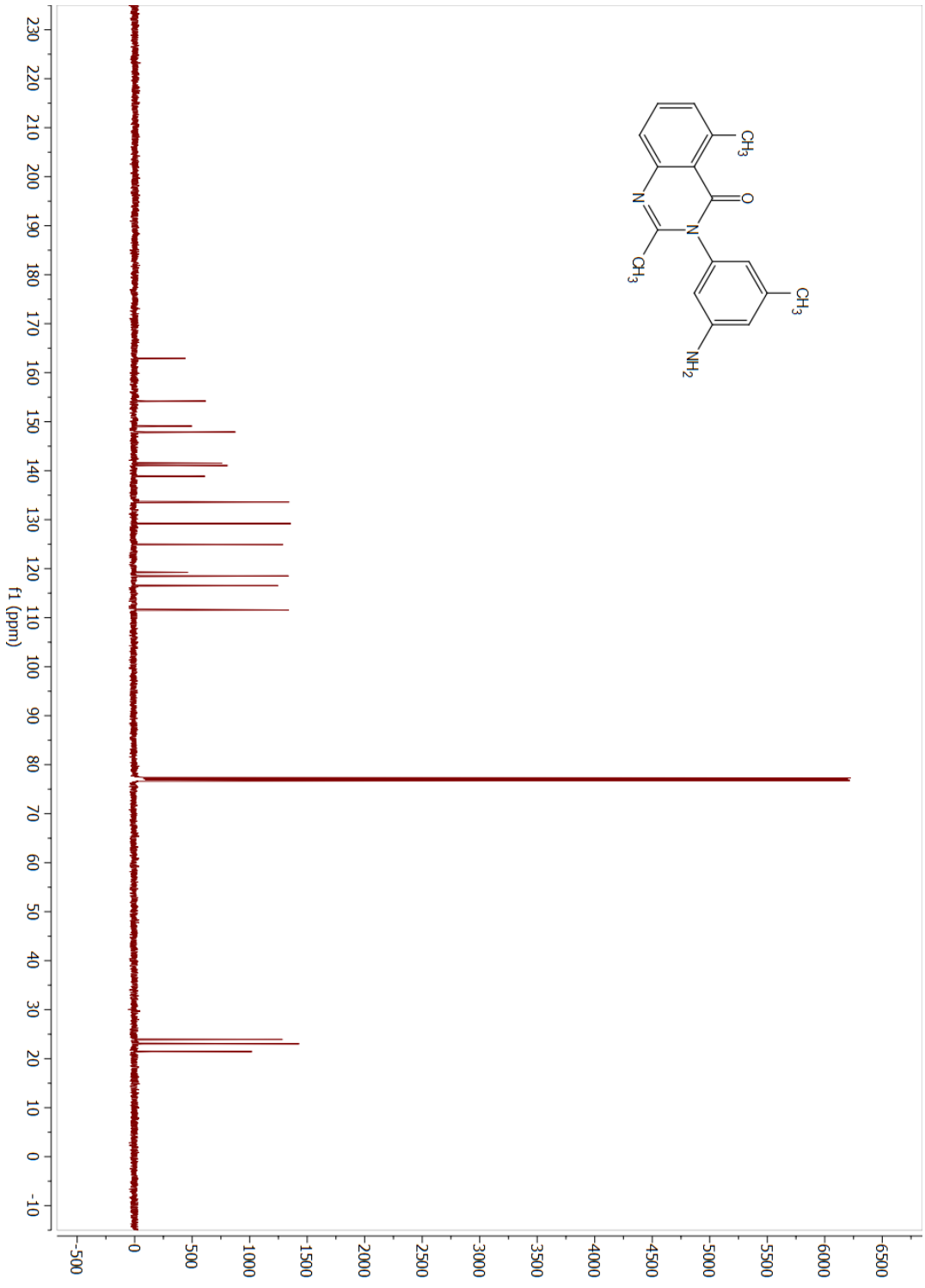


Figure AII. 13: 1HNMR Nitro compound N6

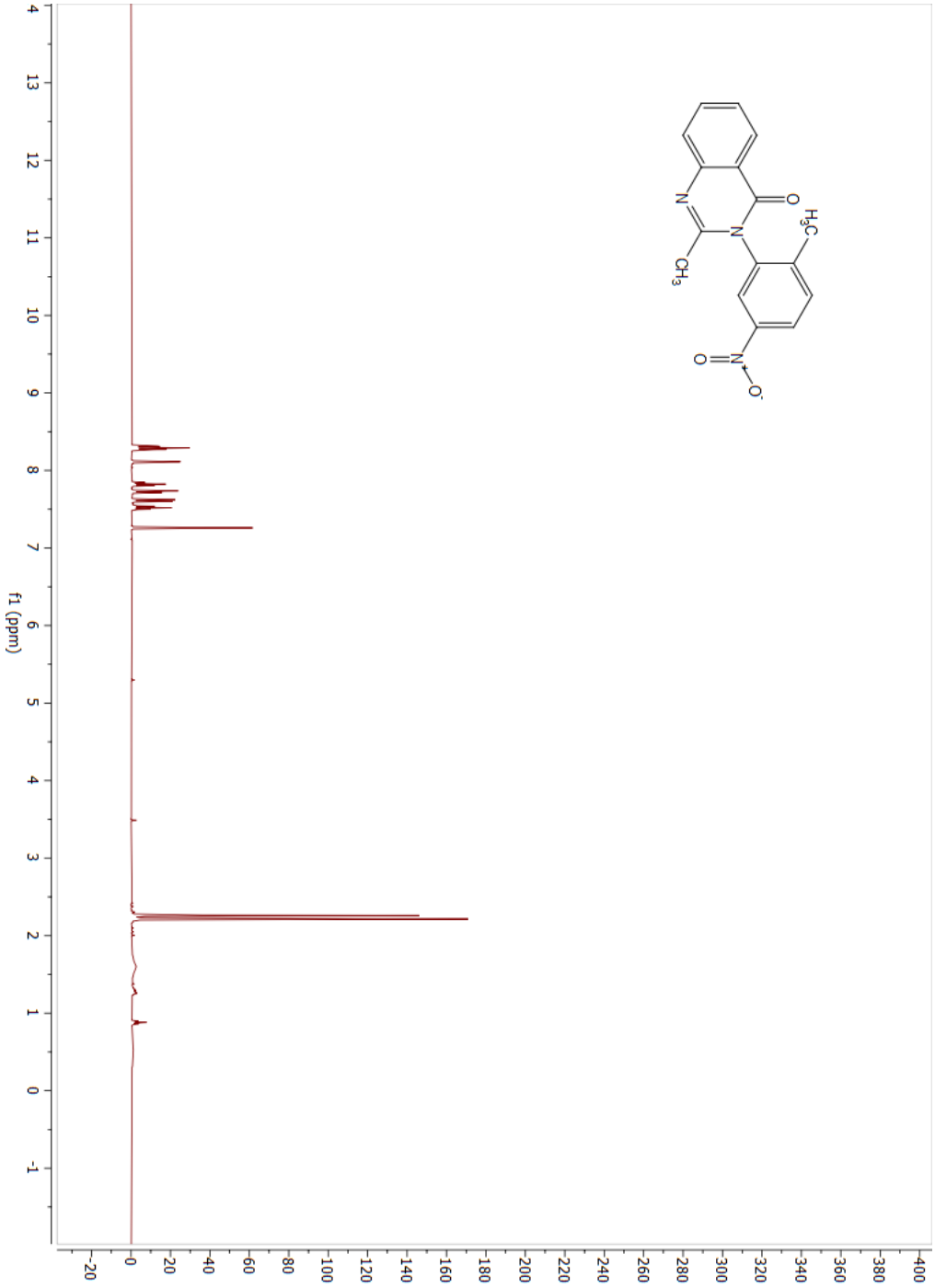


Figure AII. 14: 13CNMR Nitro compound N6

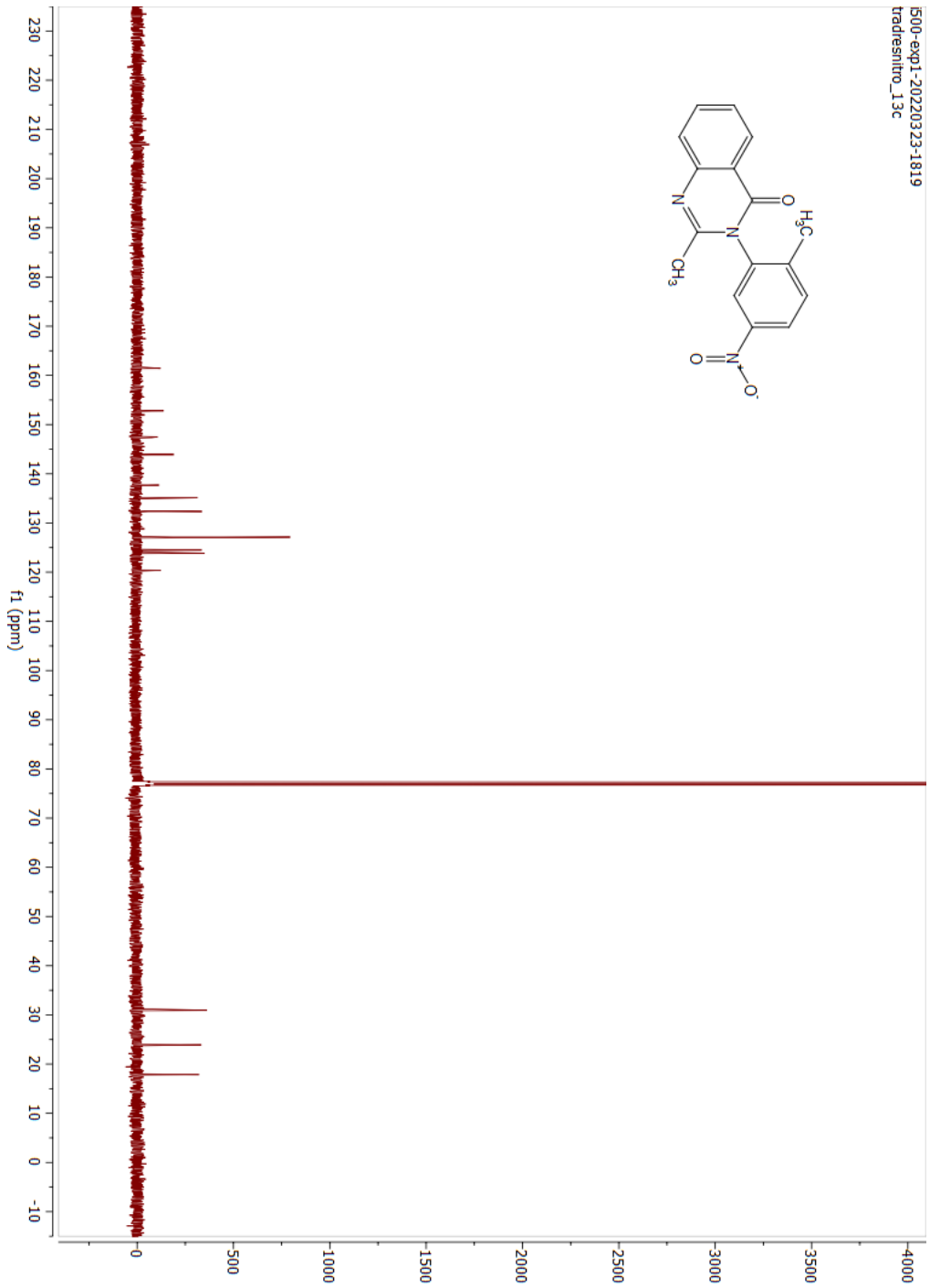


Figure AII. 15: 1HNMR Compound 6

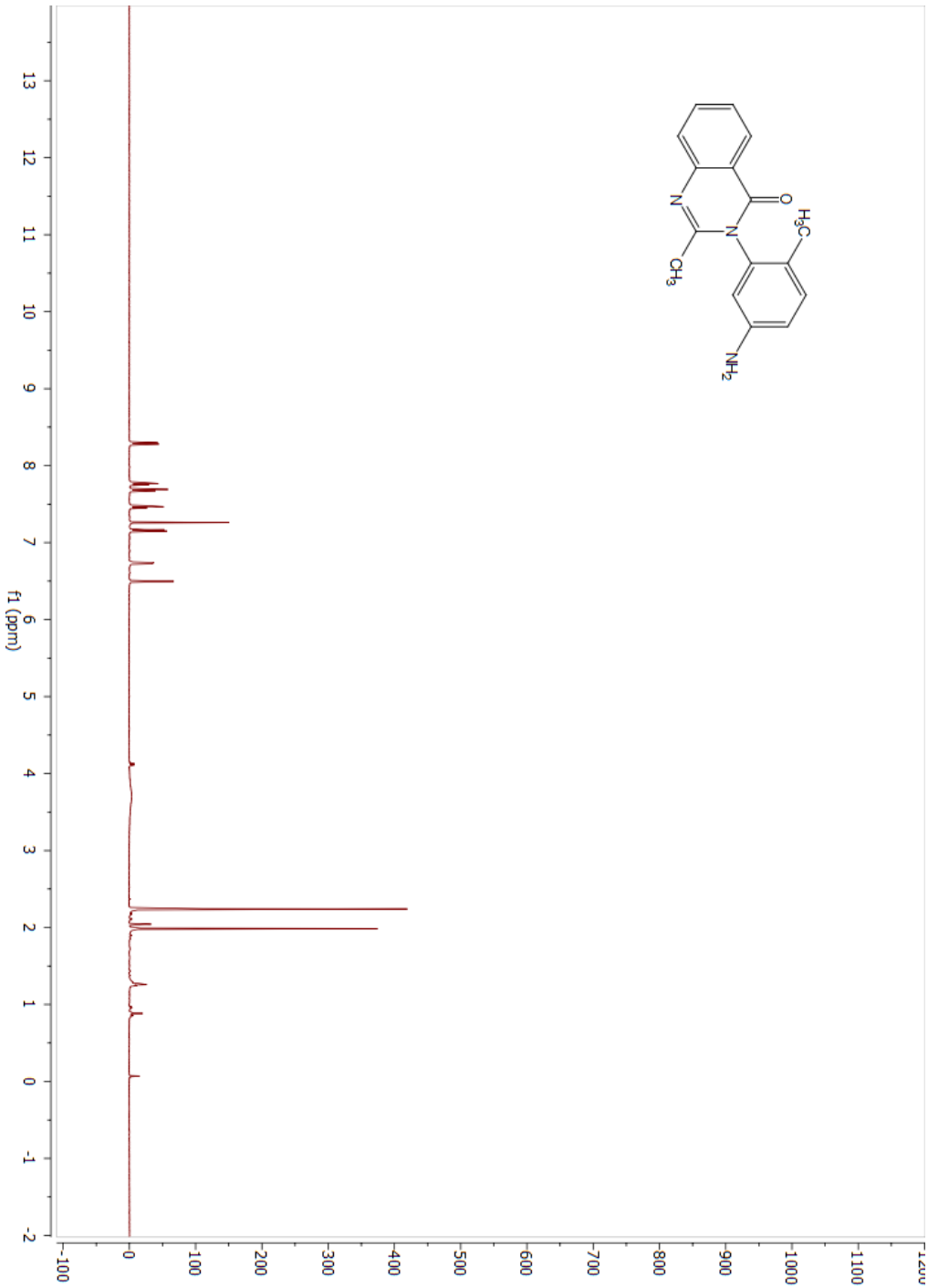


Figure AII. 16: 13CNMR Compound 6

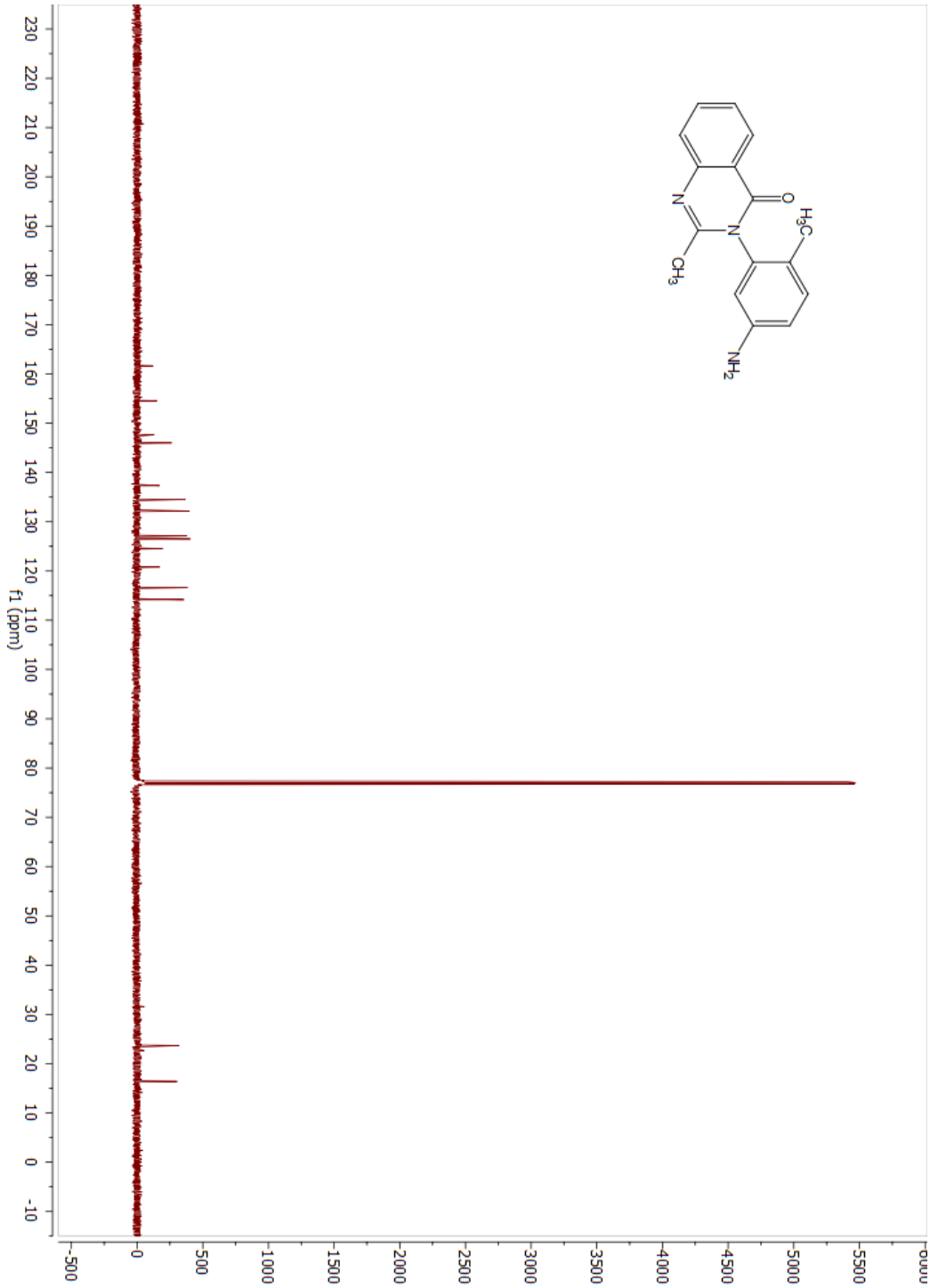


Figure AII. 17: 1HNMR Nitro compound N7

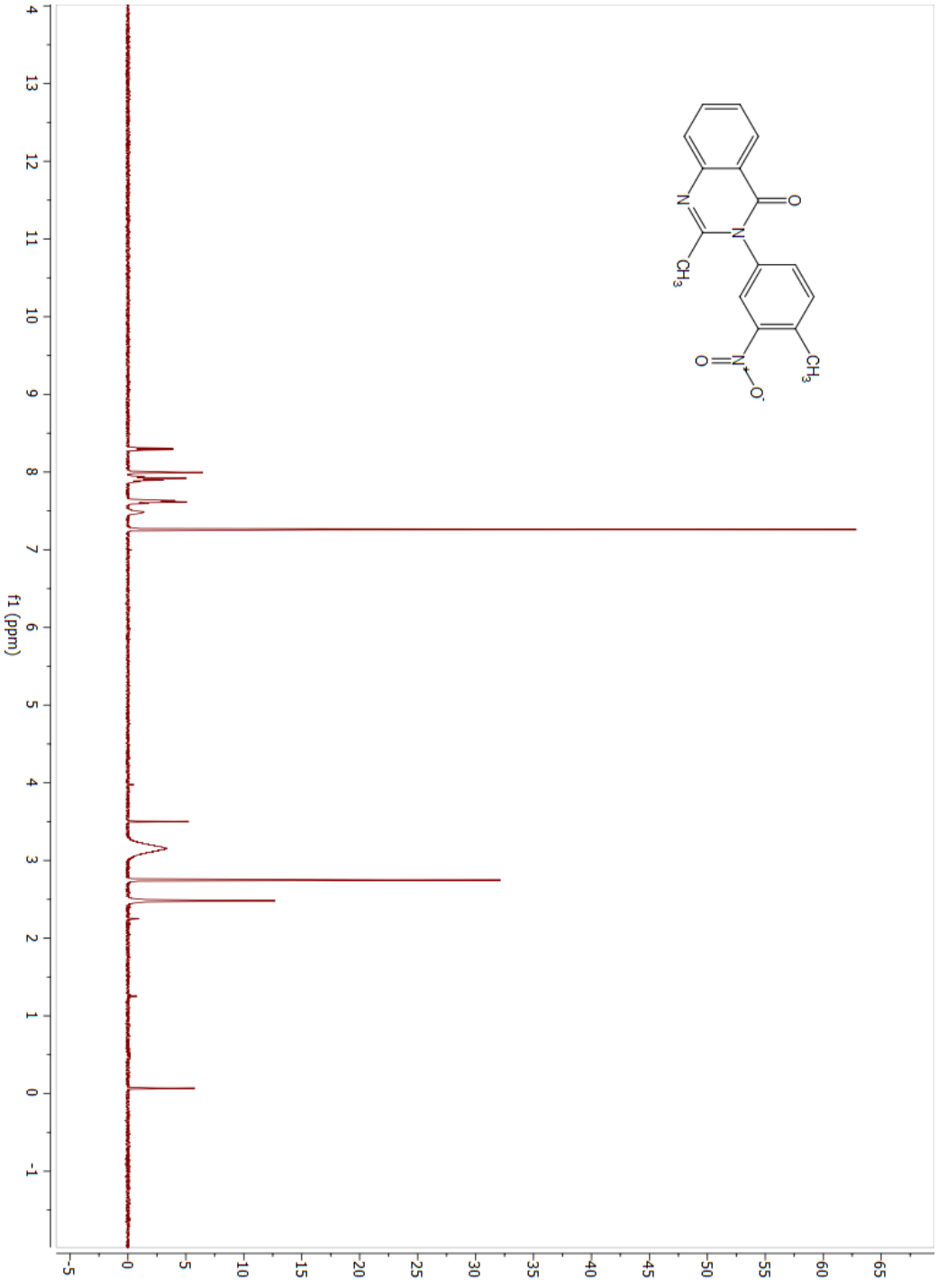


Figure AII. 18: 13CNMR Nitro compound N7

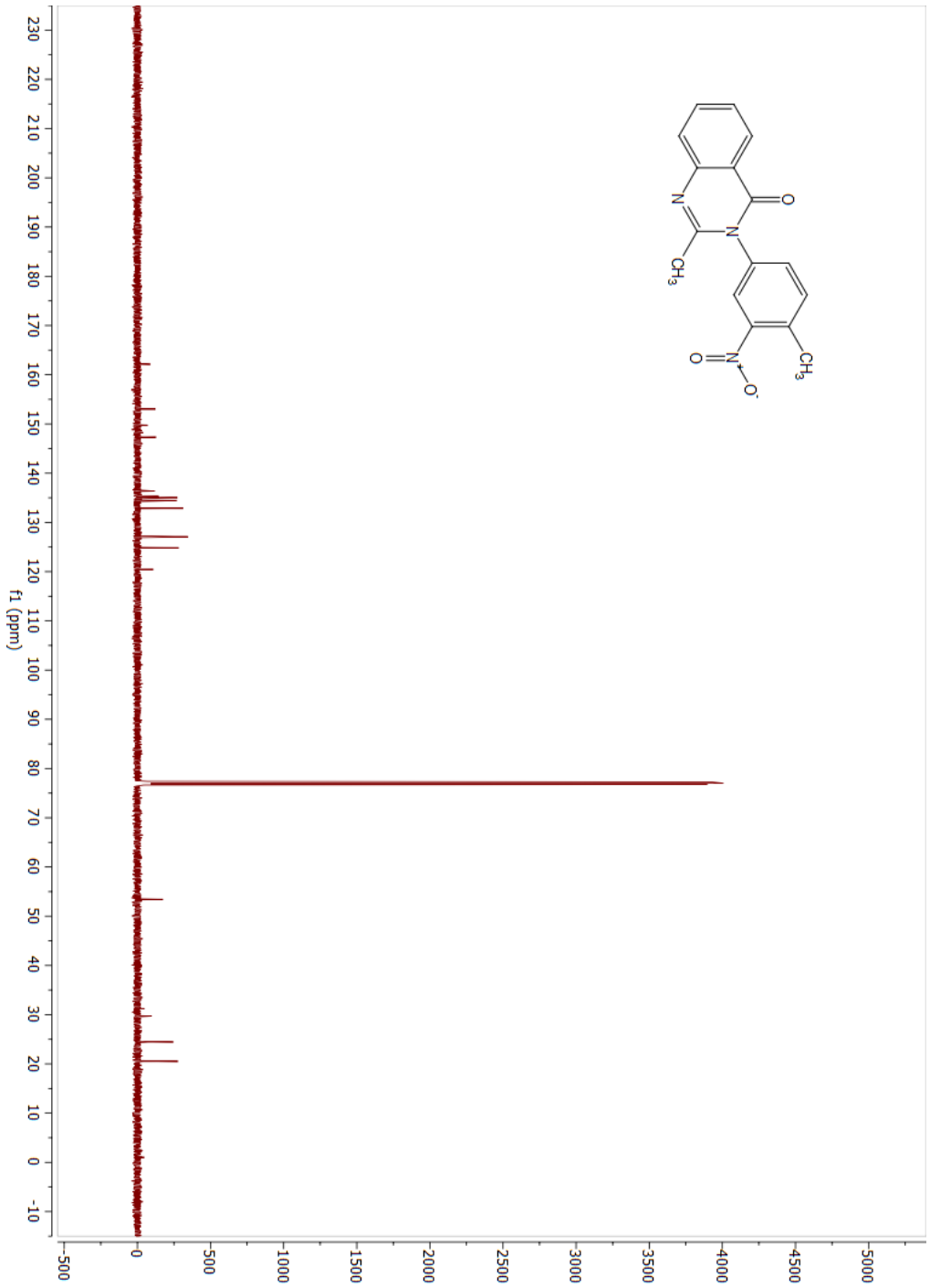


Figure AII. 19: 1HNMR compound 7

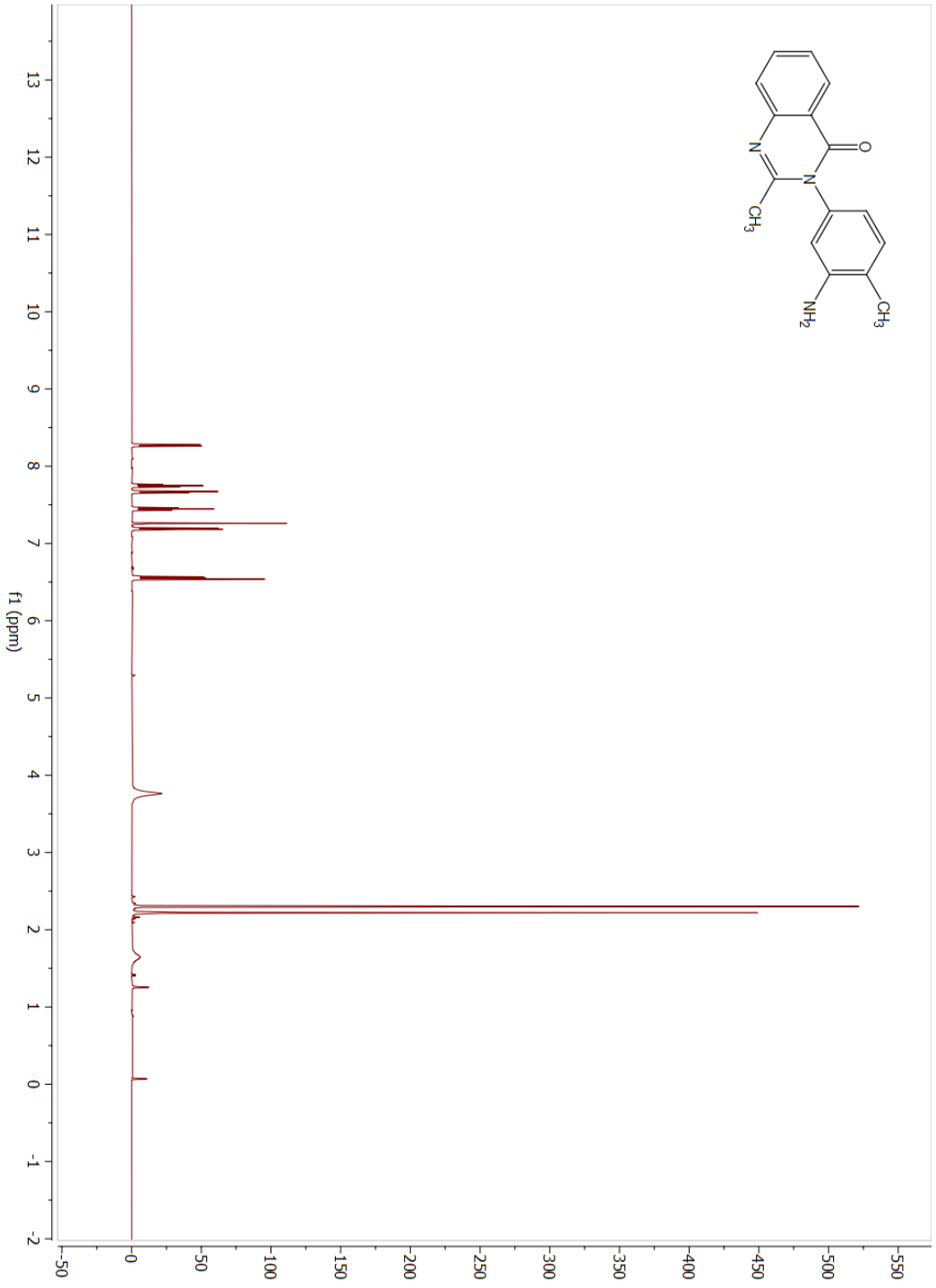


Figure AII. 20: 13CNMR compound 7

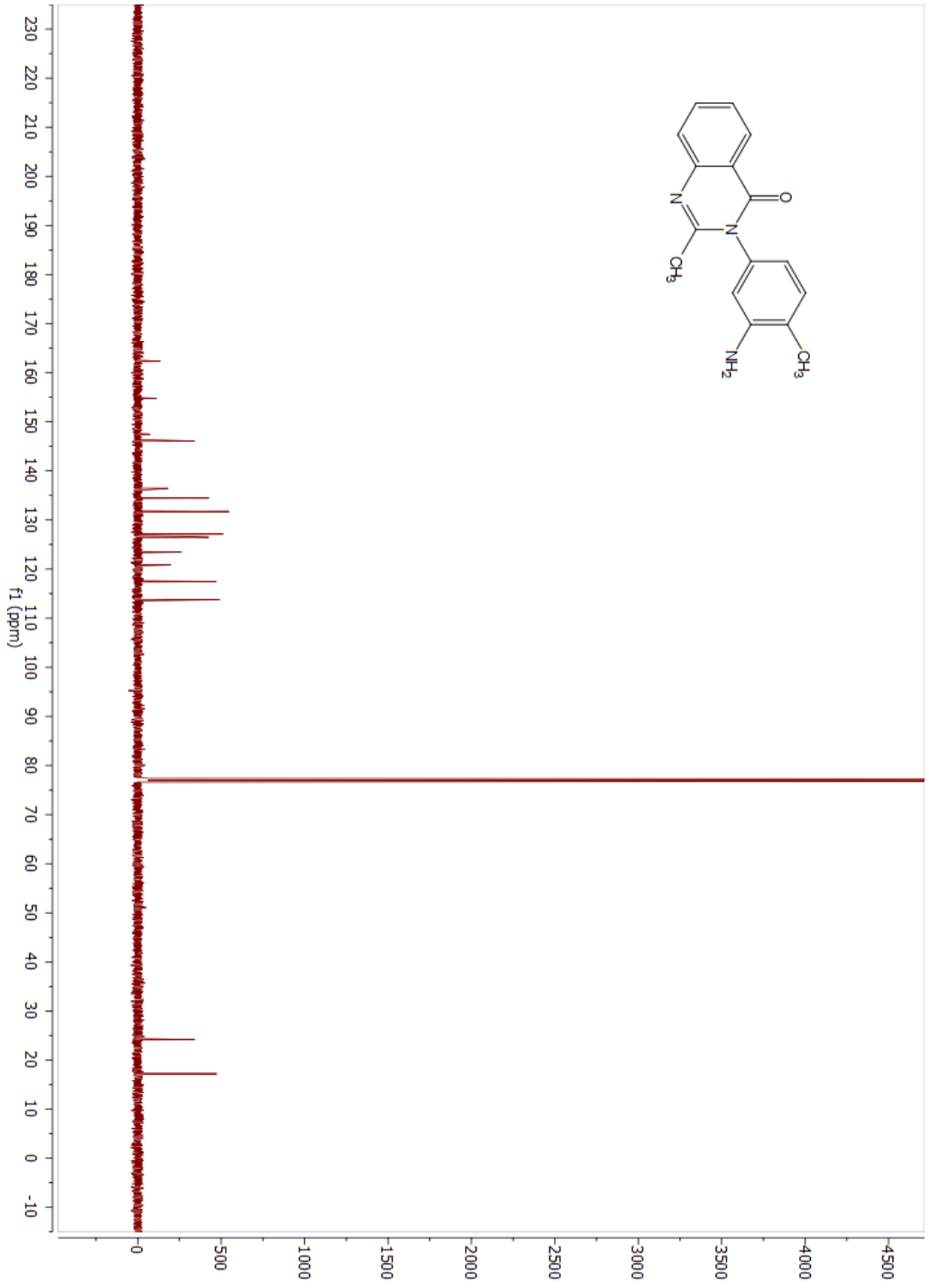


Figure AII. 21: 1HNMR compound 8

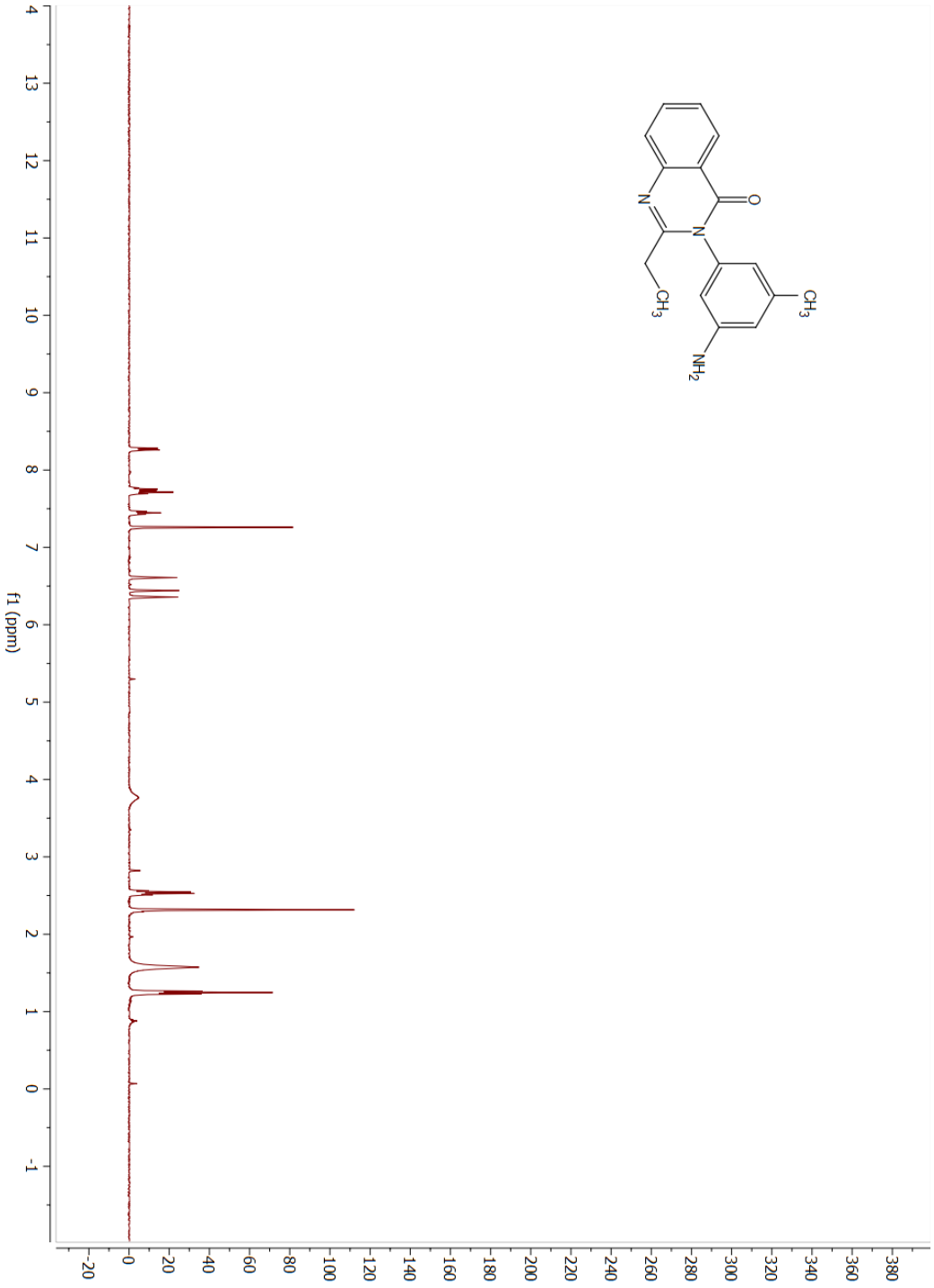


Figure AII. 22: 13CNMR compound 8

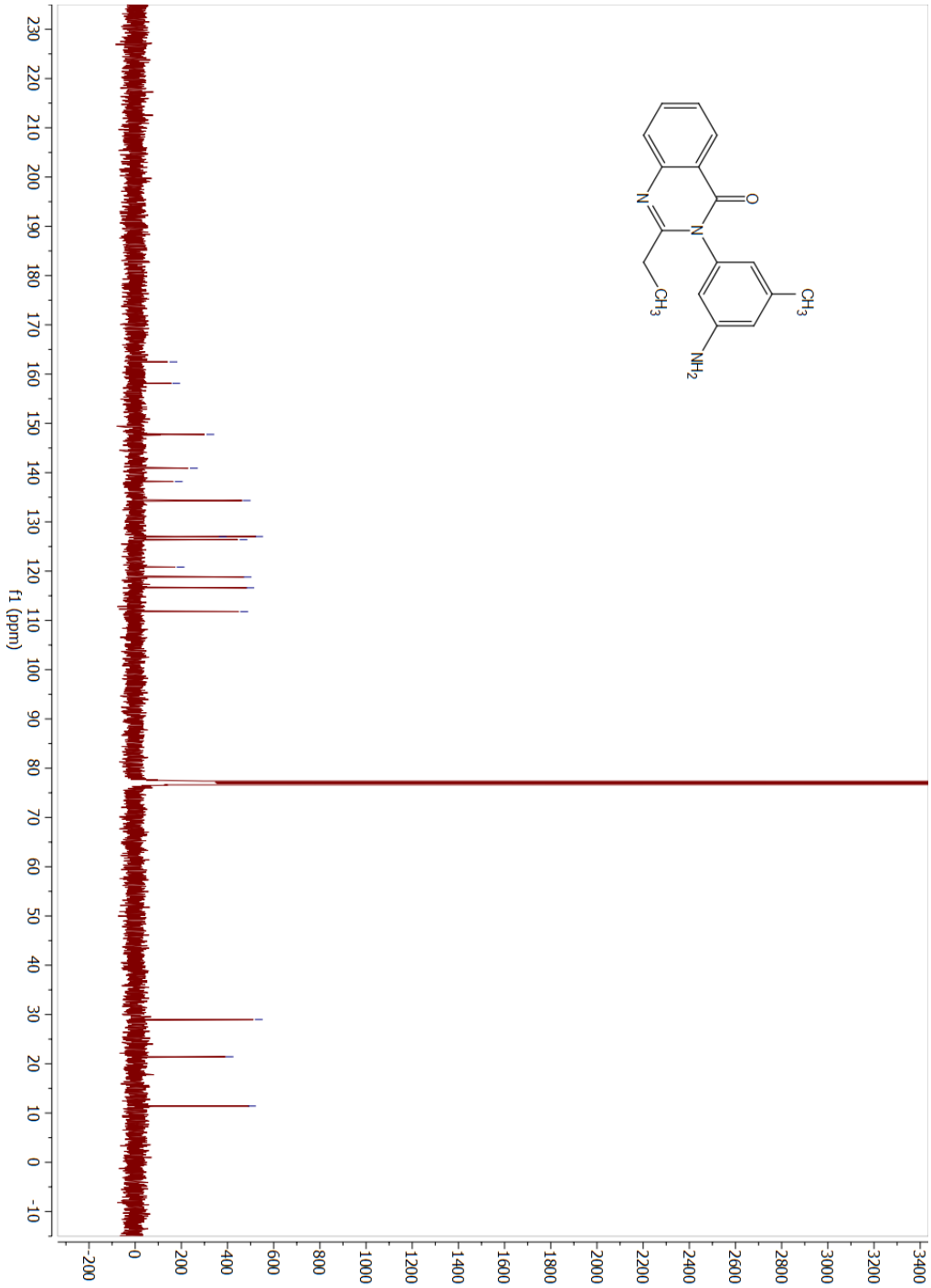


Figure AII. 23: 1HNMR nitro compound 9

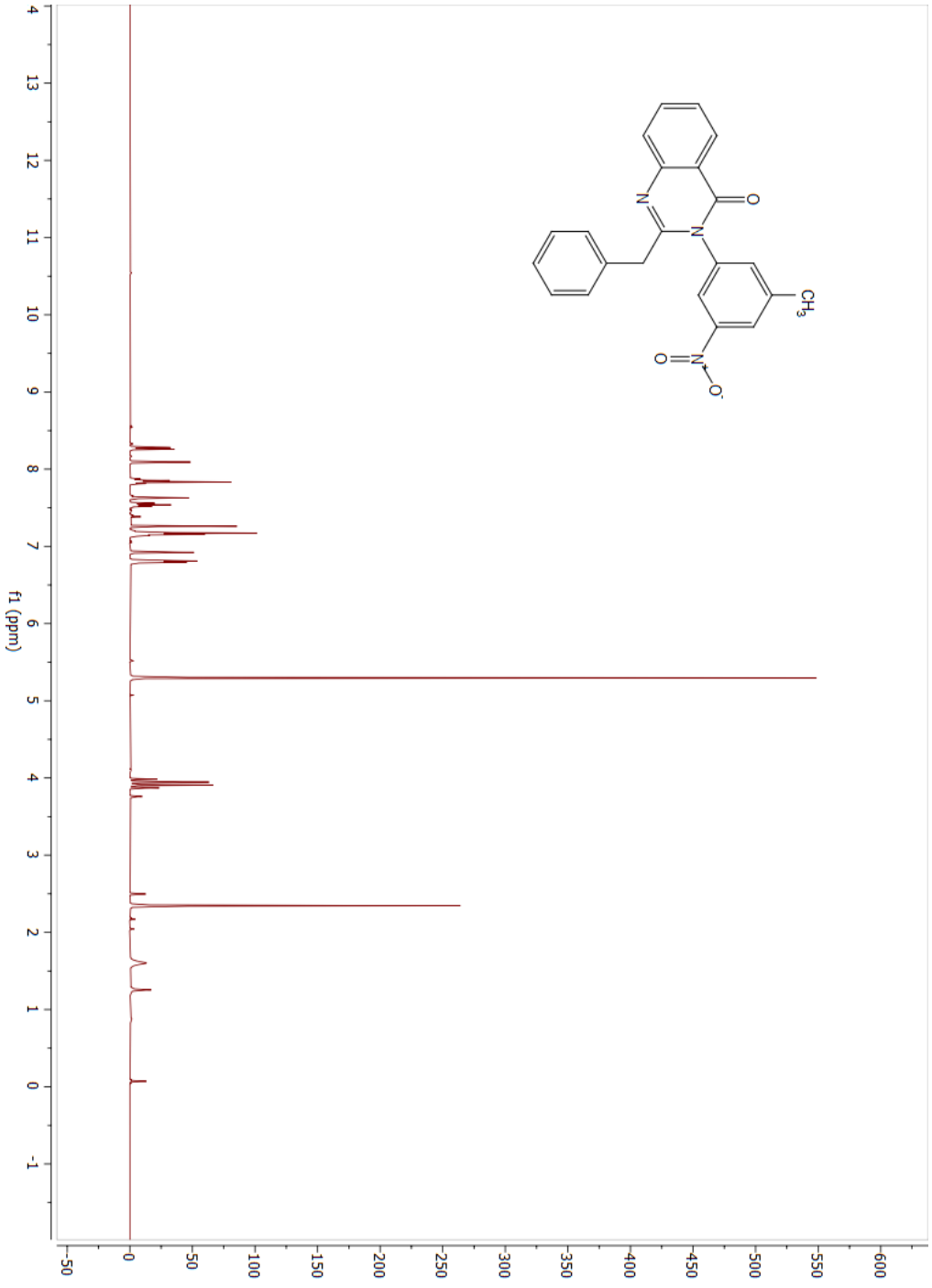


Figure AII. 24: 13CNMR nitro compound 9

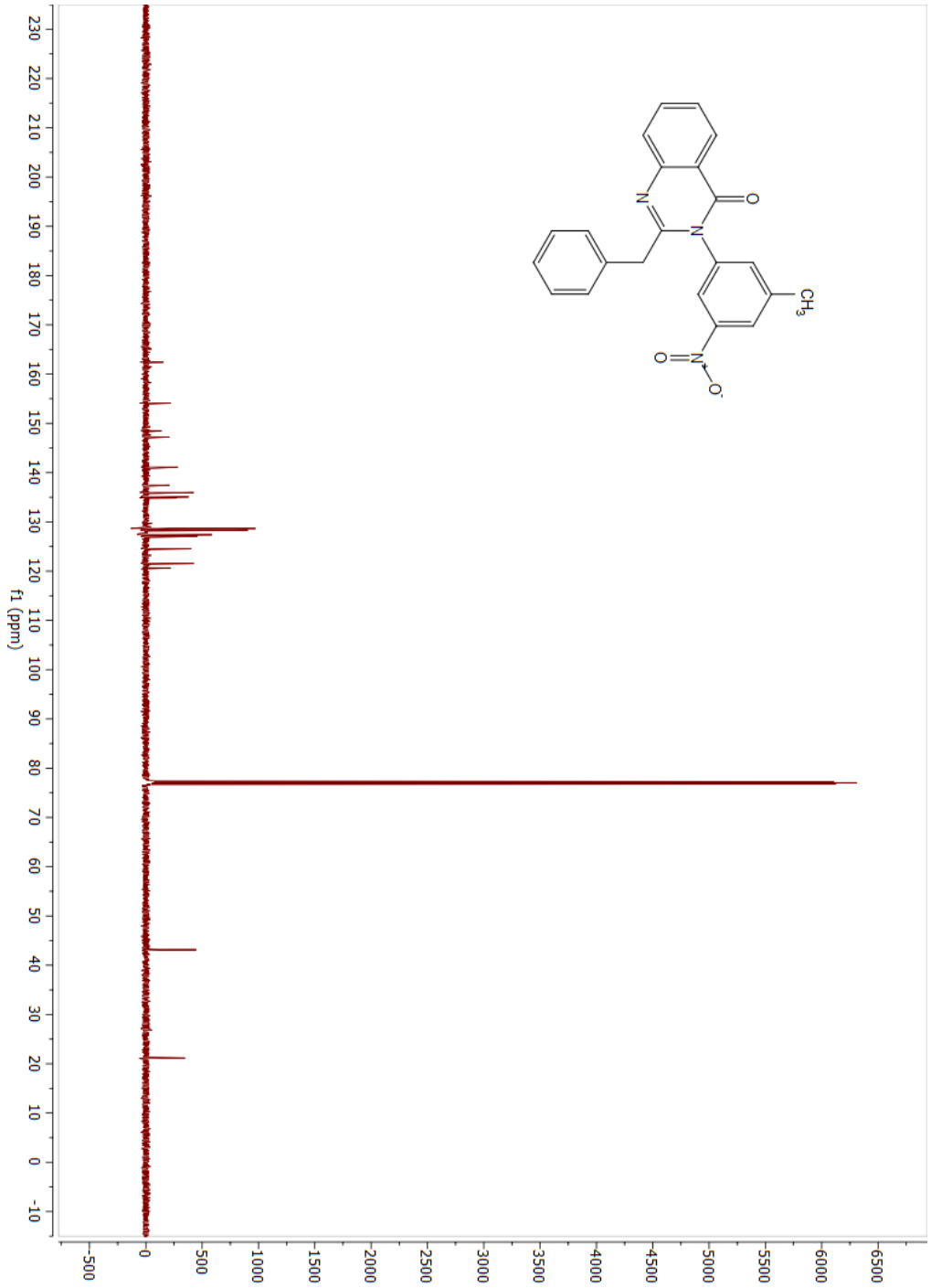


Figure AII. 25: 1HNMR compound 9

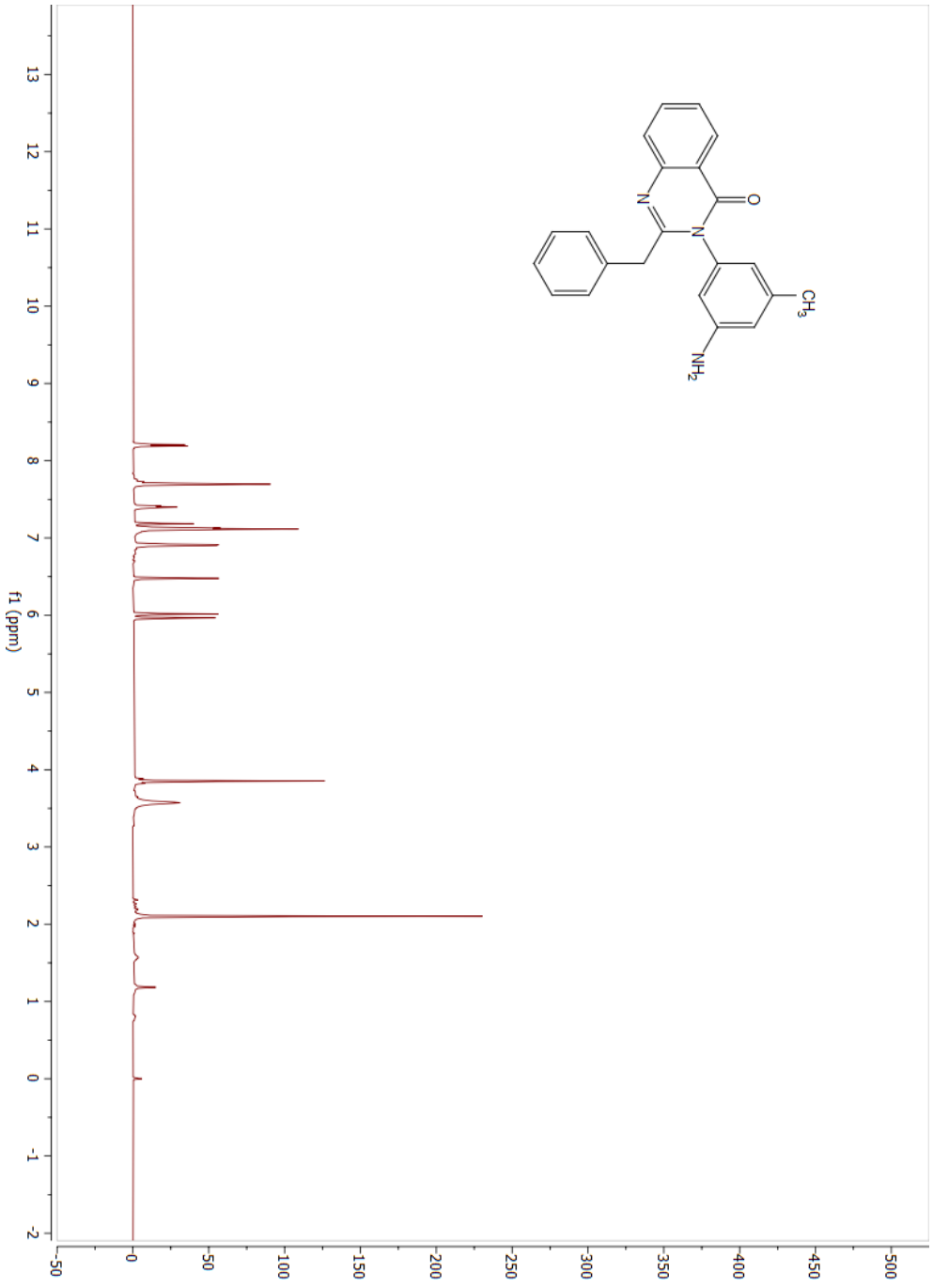


Figure AII. 26: 13CNMR compound 9

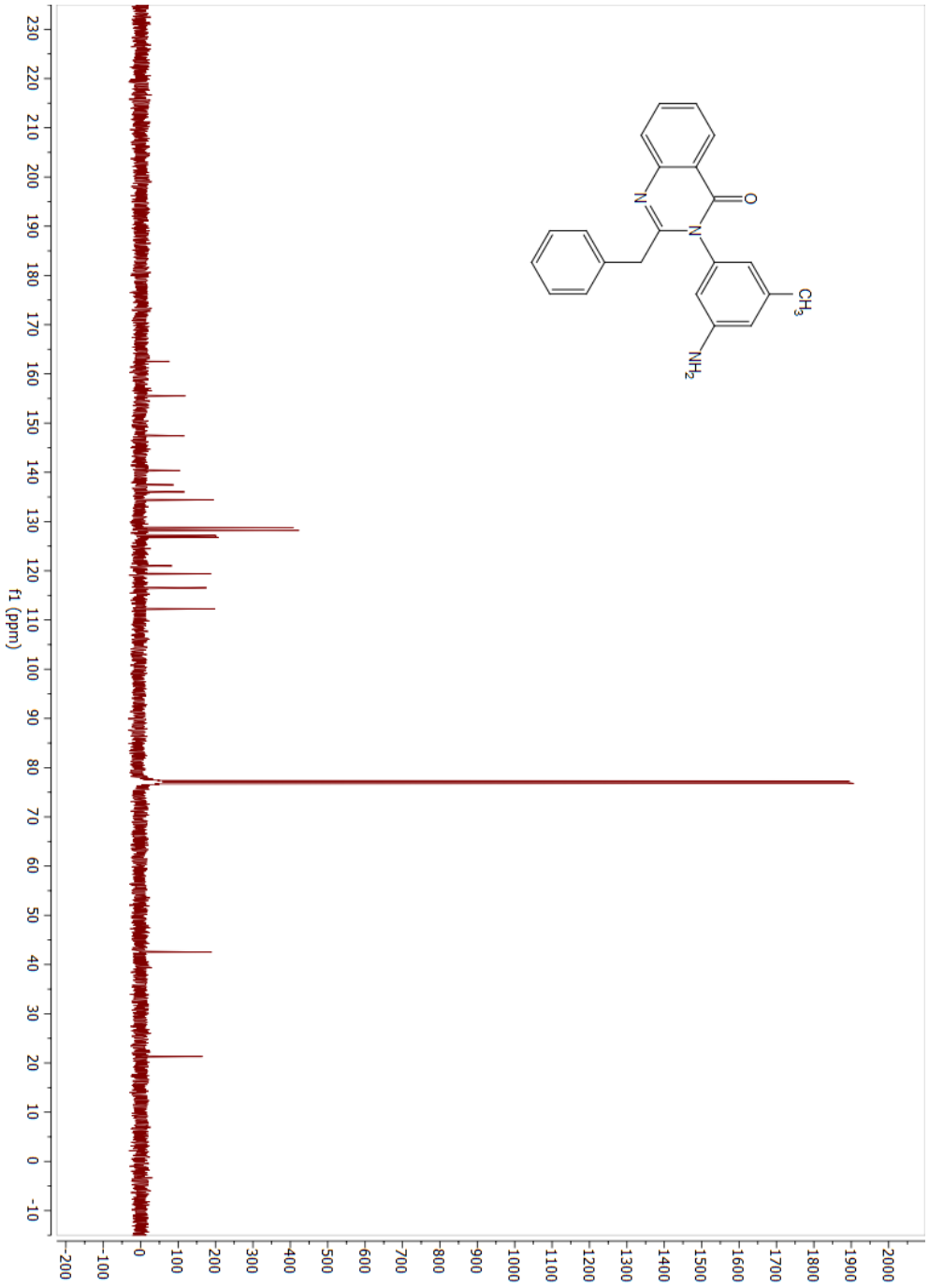


Figure AII. 27: 1HNMR compound 10

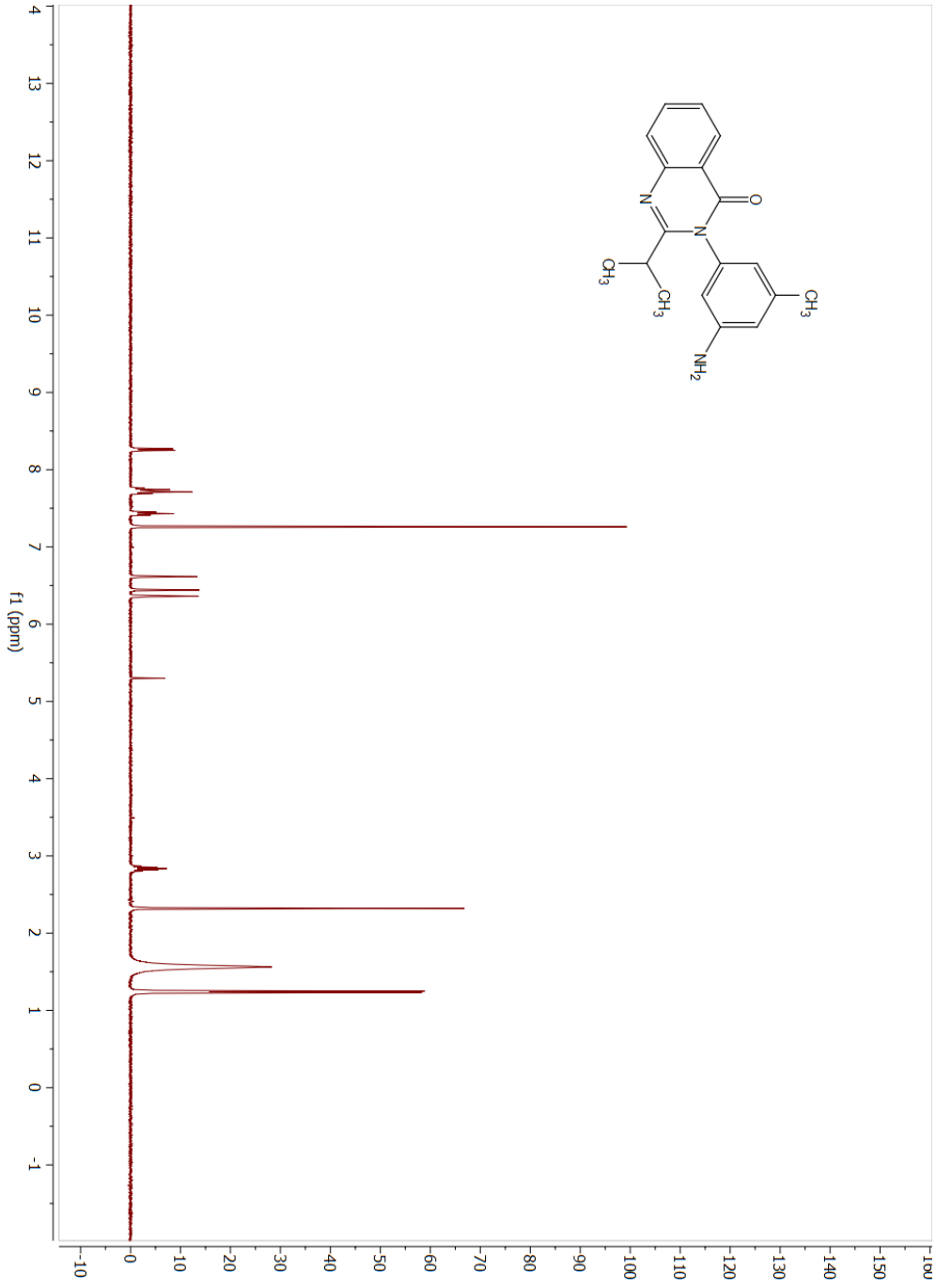


Figure AII. 28: 13CNMR compound 10

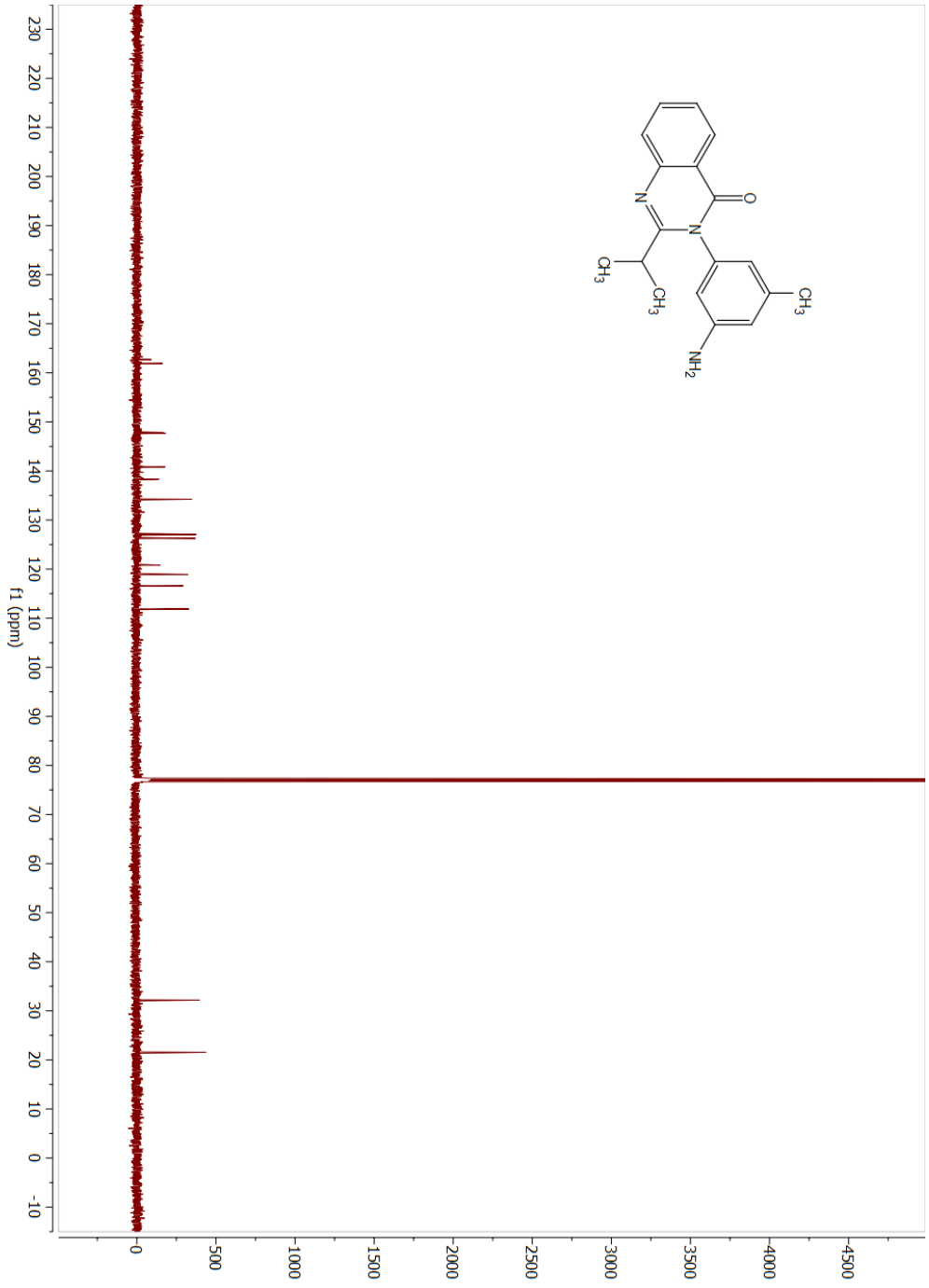


Figure AII. 29: 1HNMR compound 11

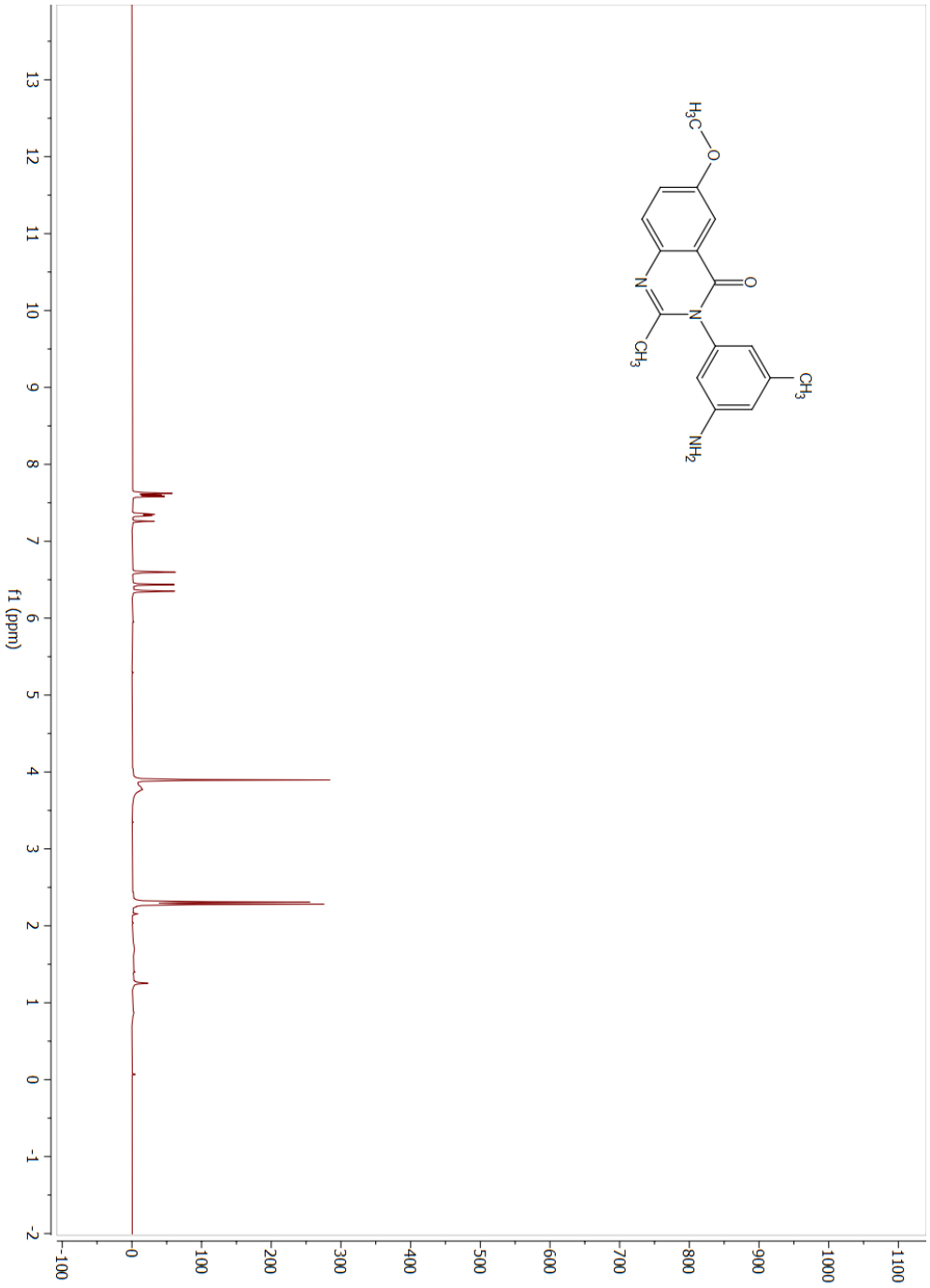


Figure AII. 30: 13CNMR compound 11

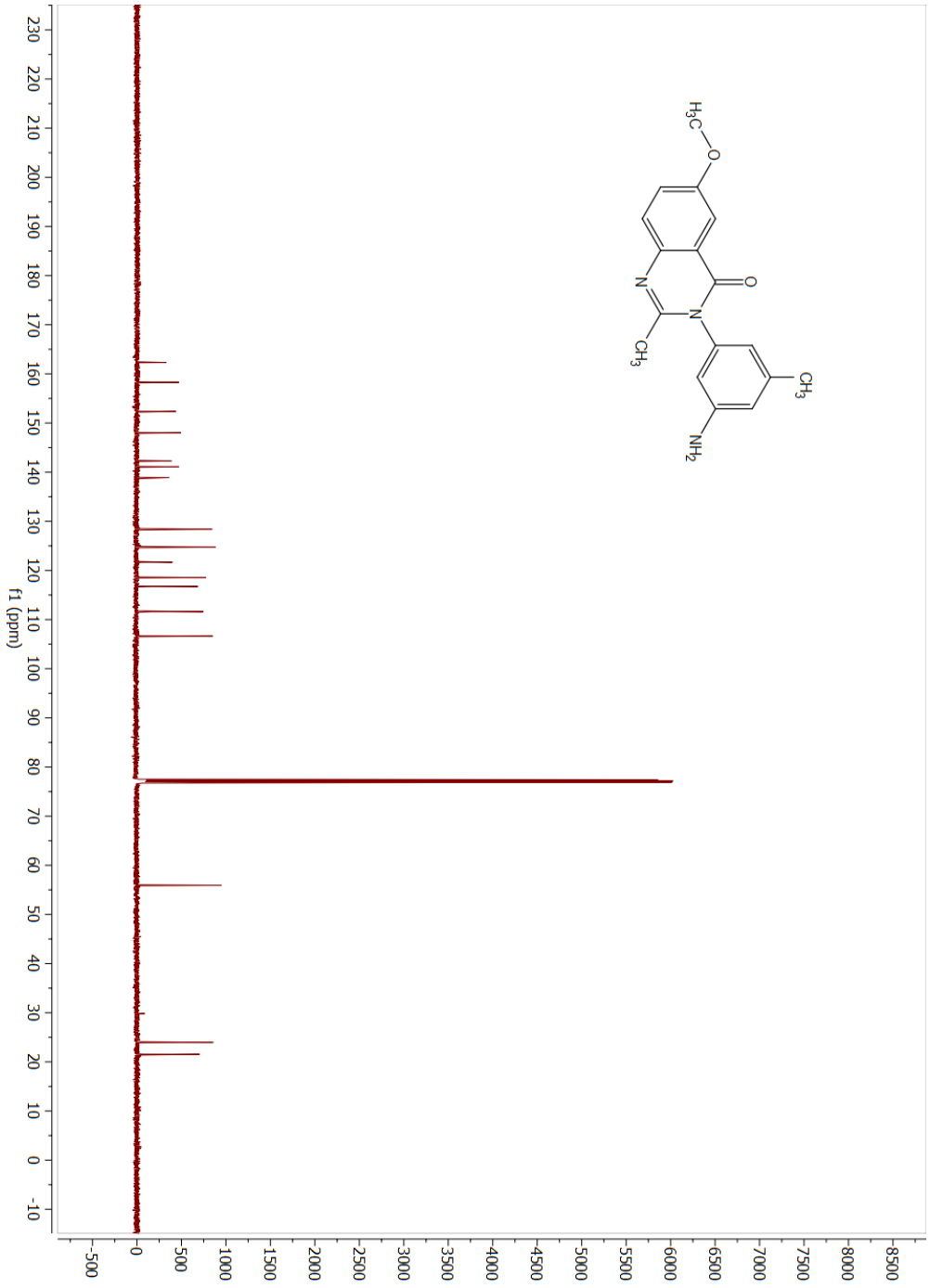


Figure AII. 31: 1HNMR compound 12

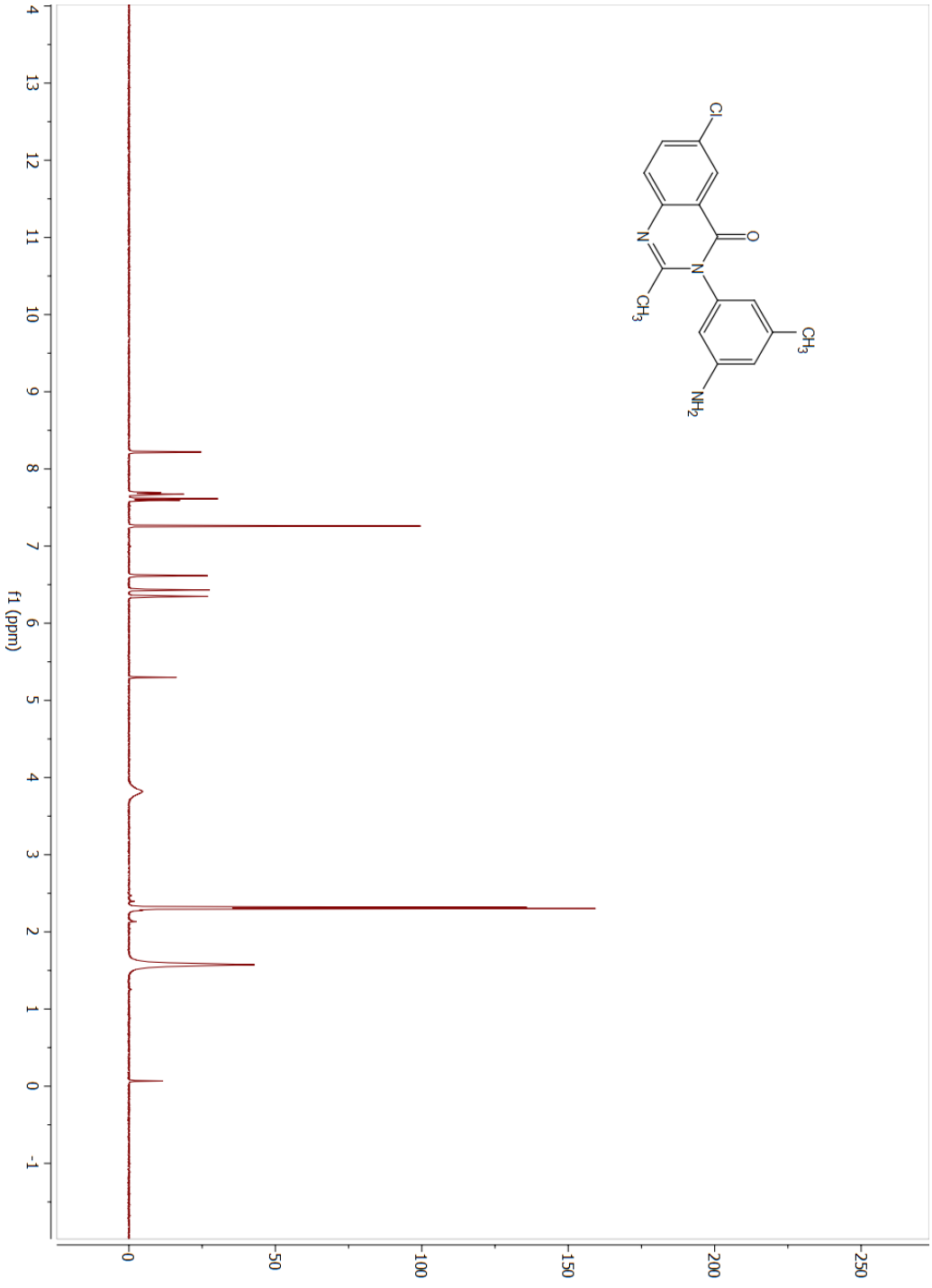


Figure AII. 32: 13CNMR compound 12

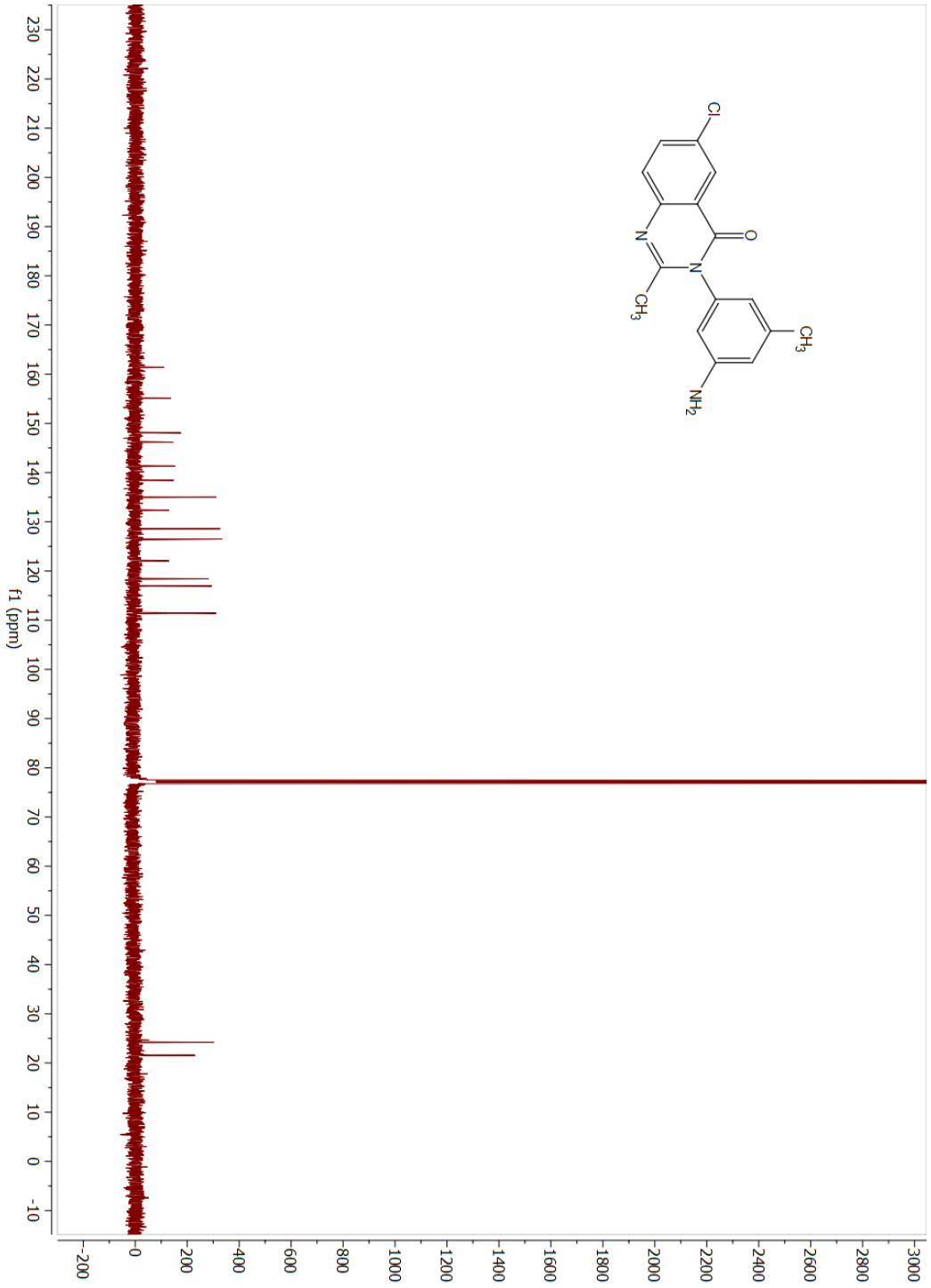


Figure AII. 33: 1HNMR compound 13

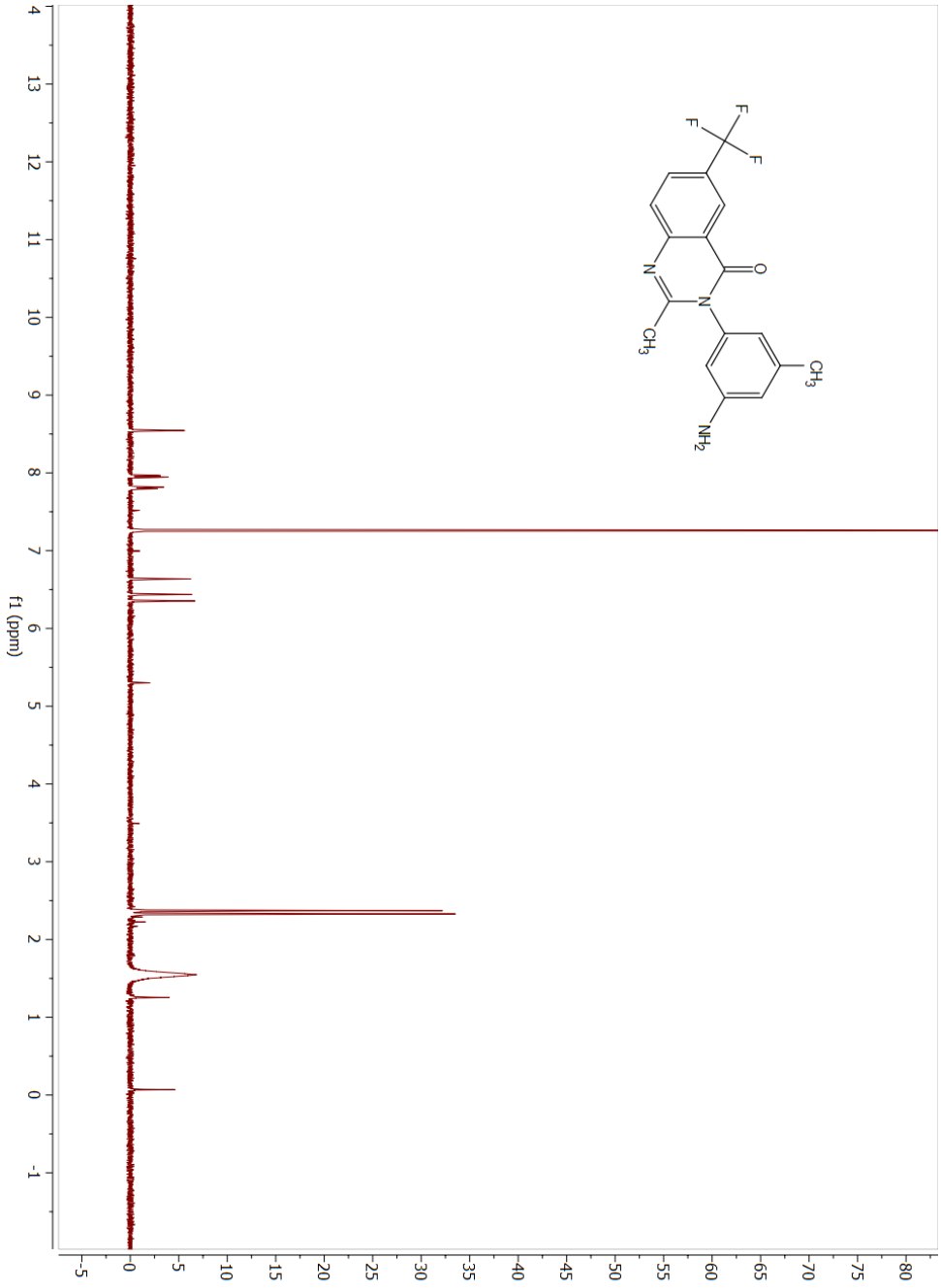


Figure AII. 34: 13CNMR compound 13

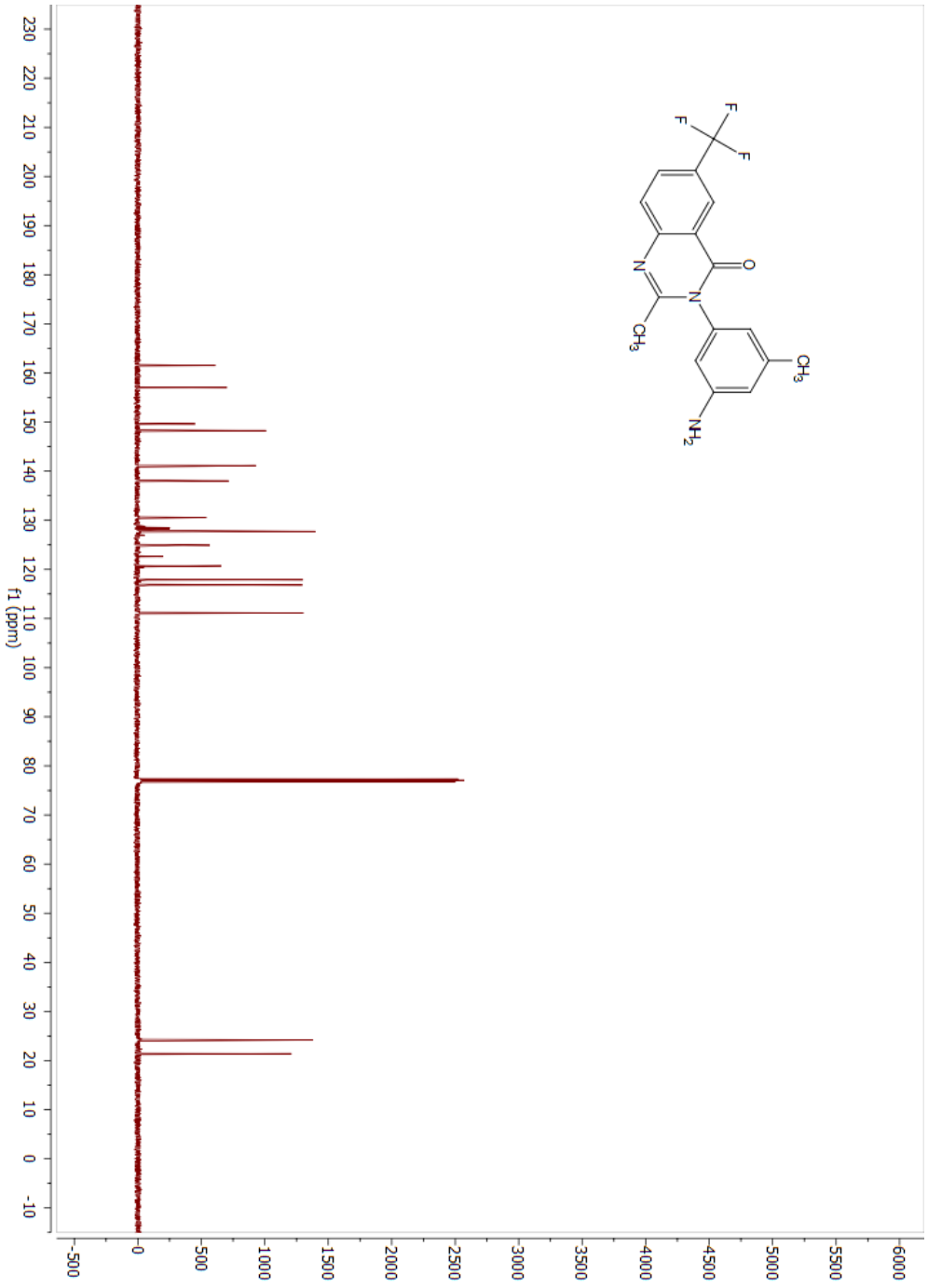


Figure AII. 35: 1HNMR compound 14

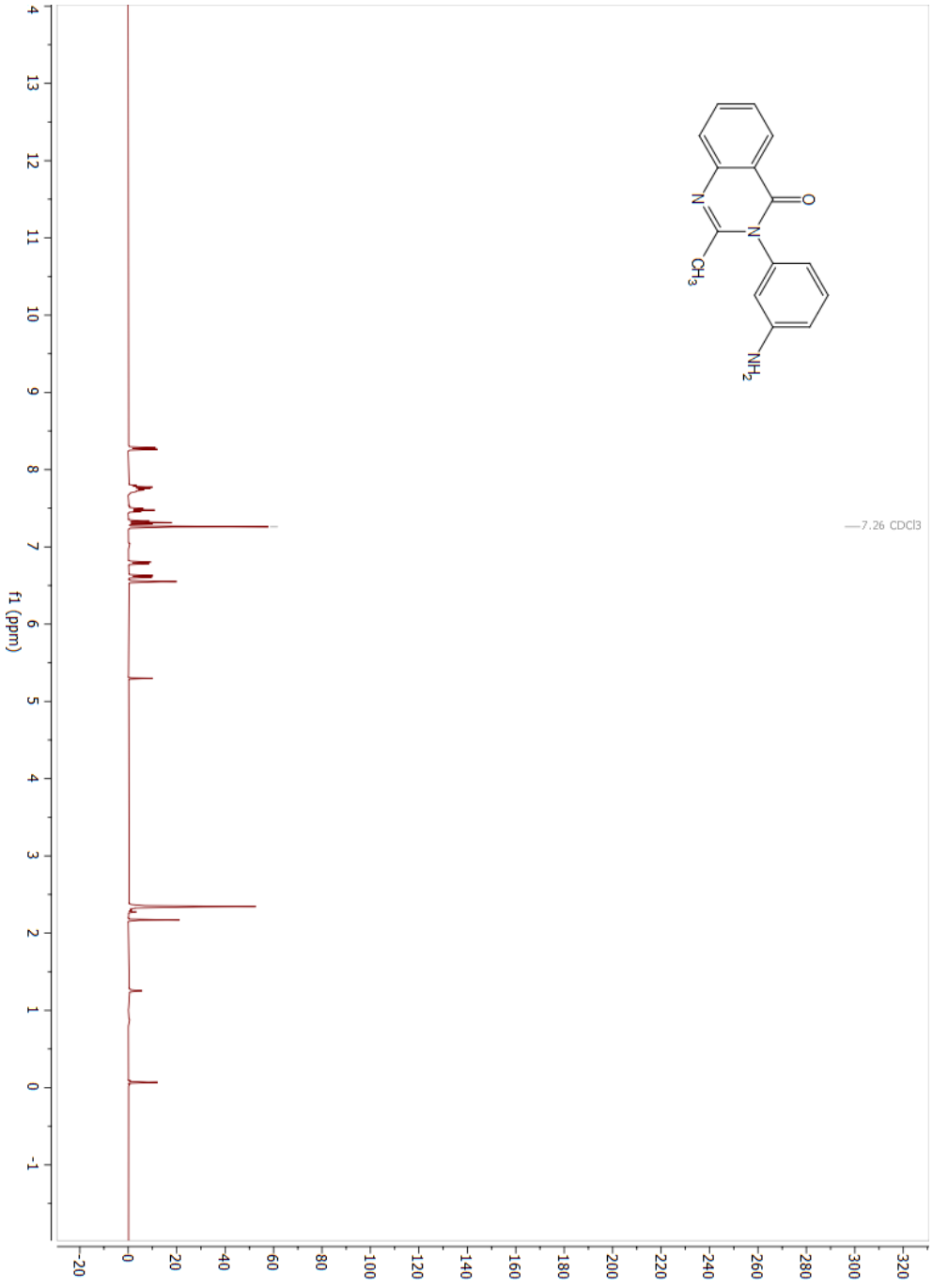


Figure AII. 36: 13CNMR compound 14

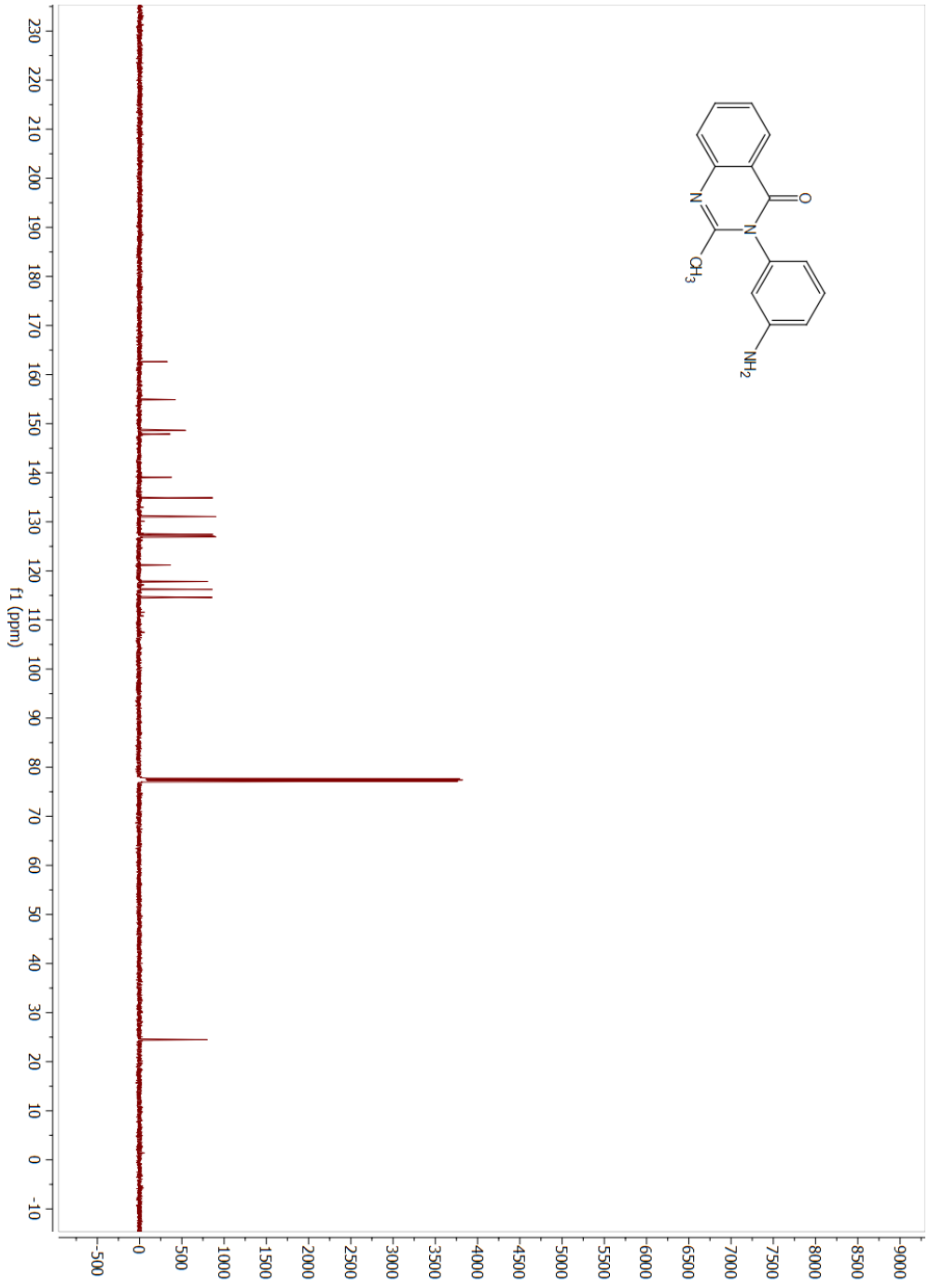


Figure AII. 37: 1HNMR nitro compound N15

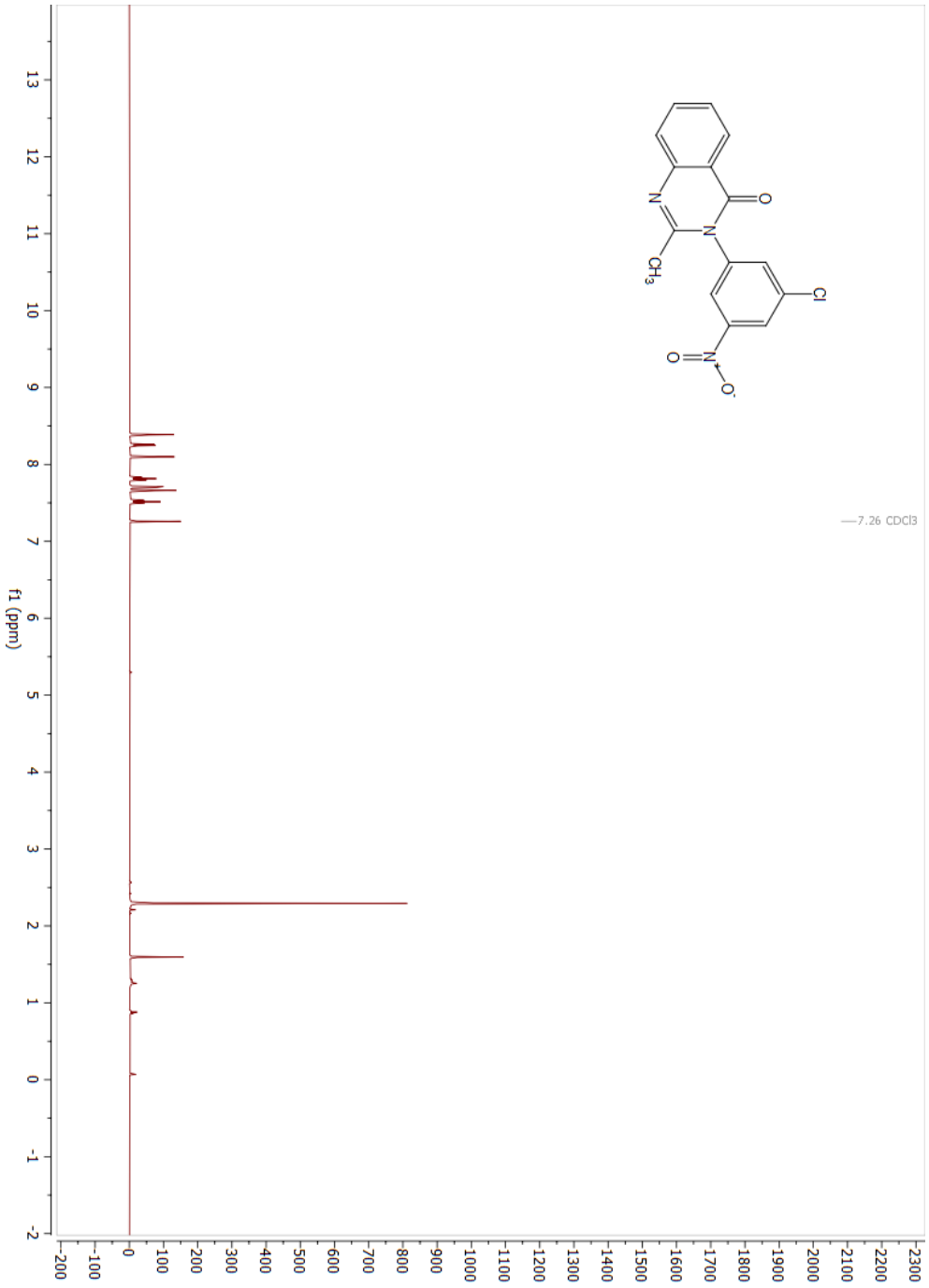


Figure AII. 38: 13CNMR nitro compound N15

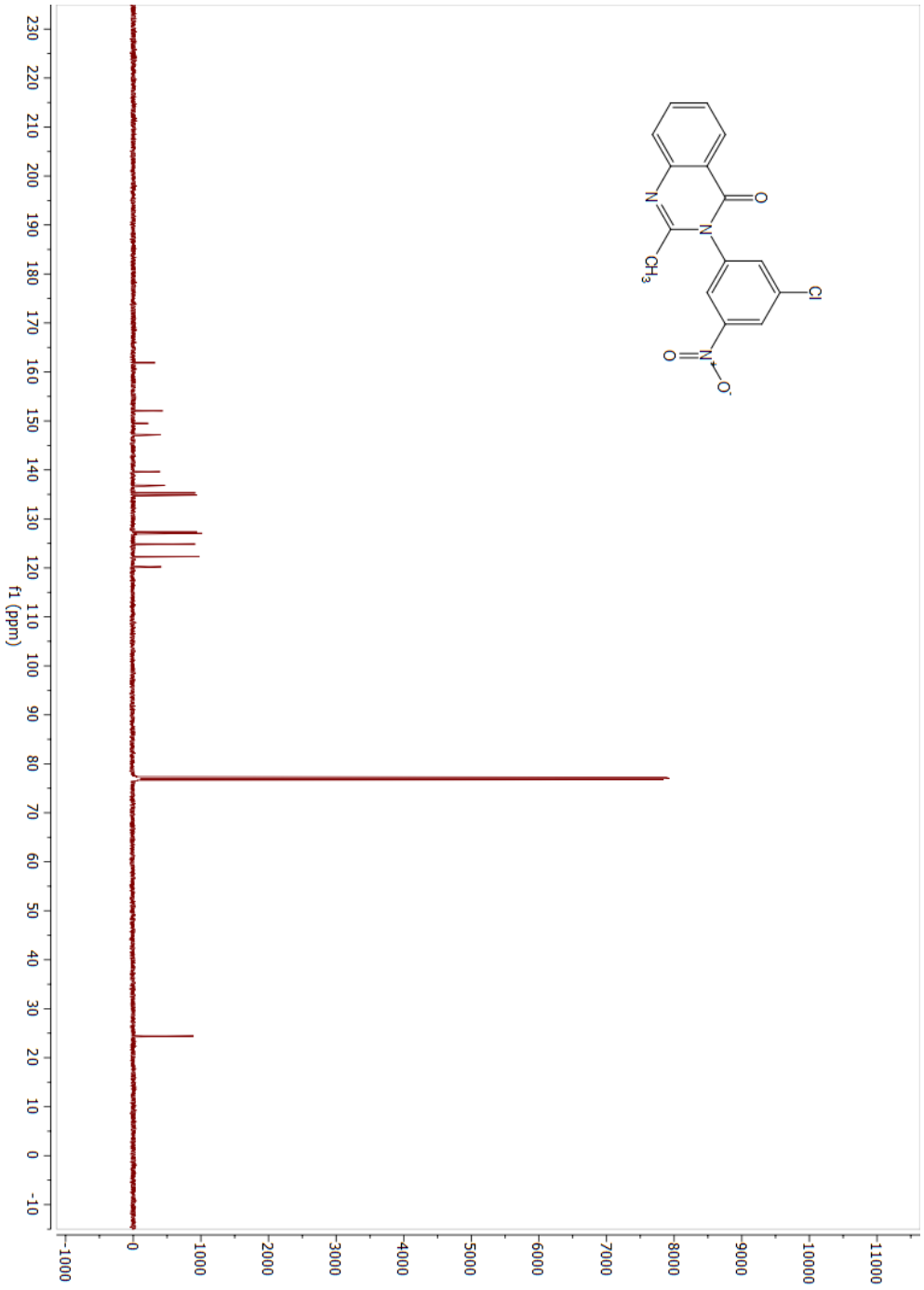


Figure AII. 39: 1HNMR compound 15

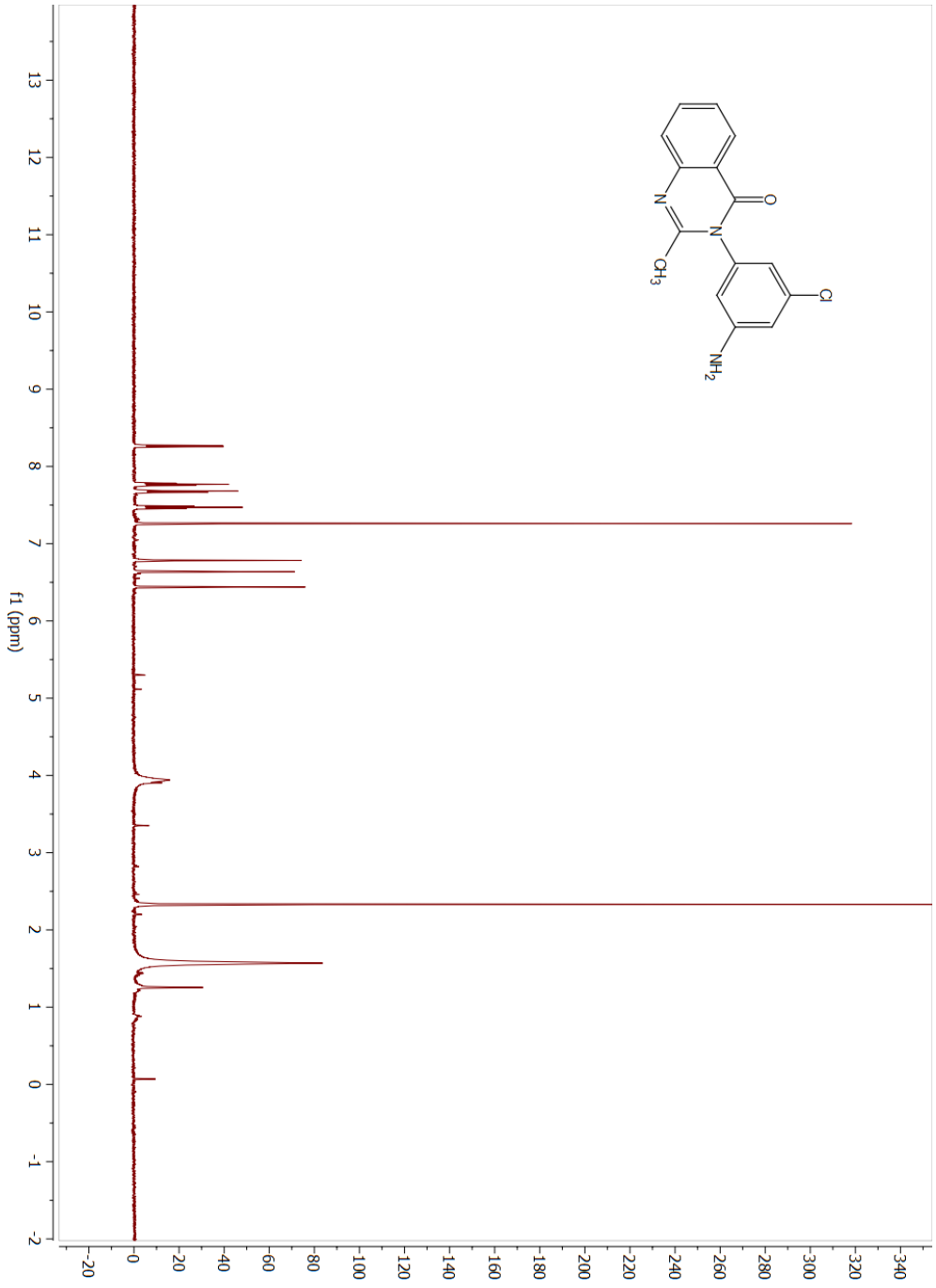


Figure AII. 40: 13CNMR compound 15

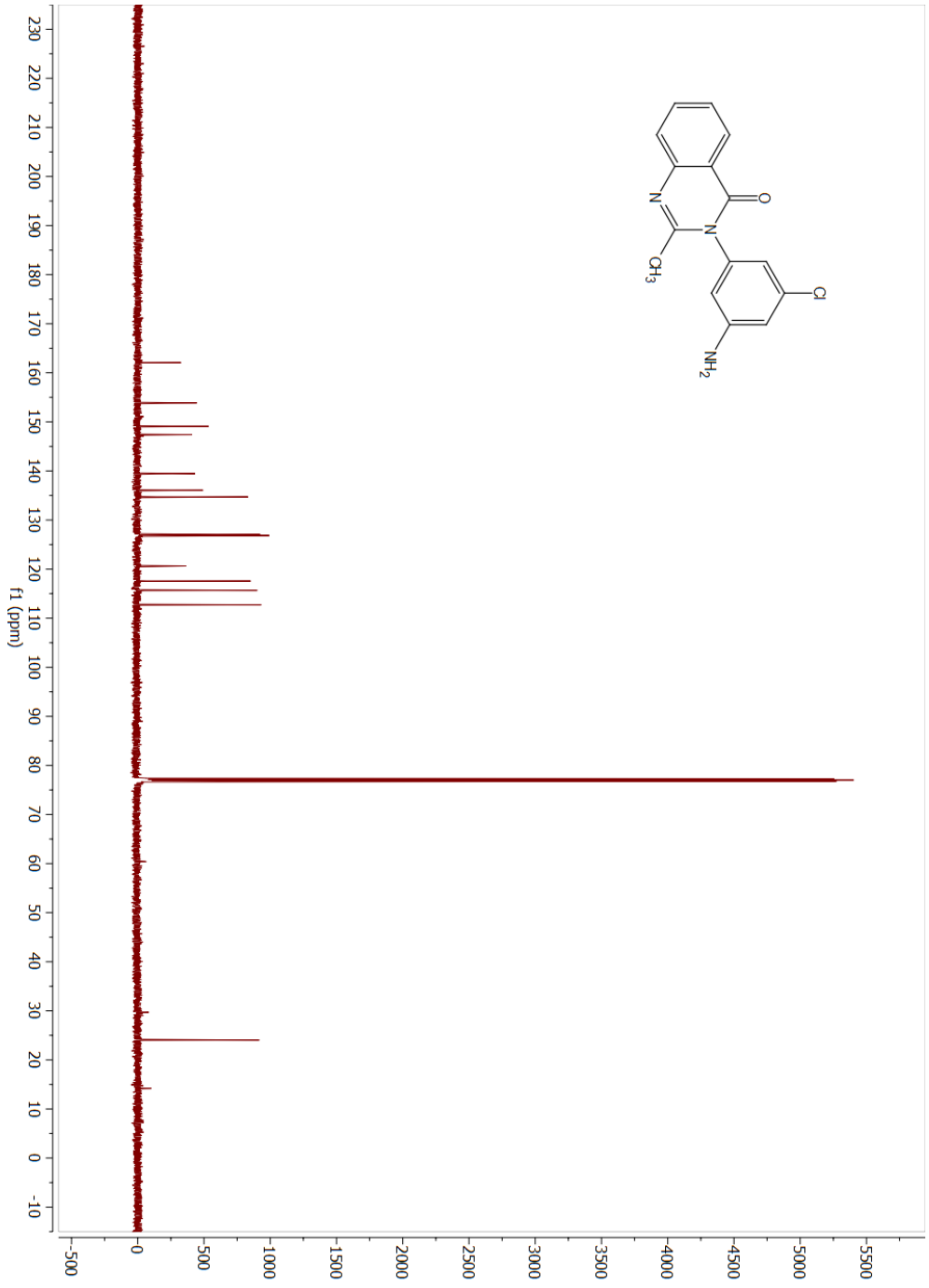


Figure AII. 41: 1HNMR compound 16

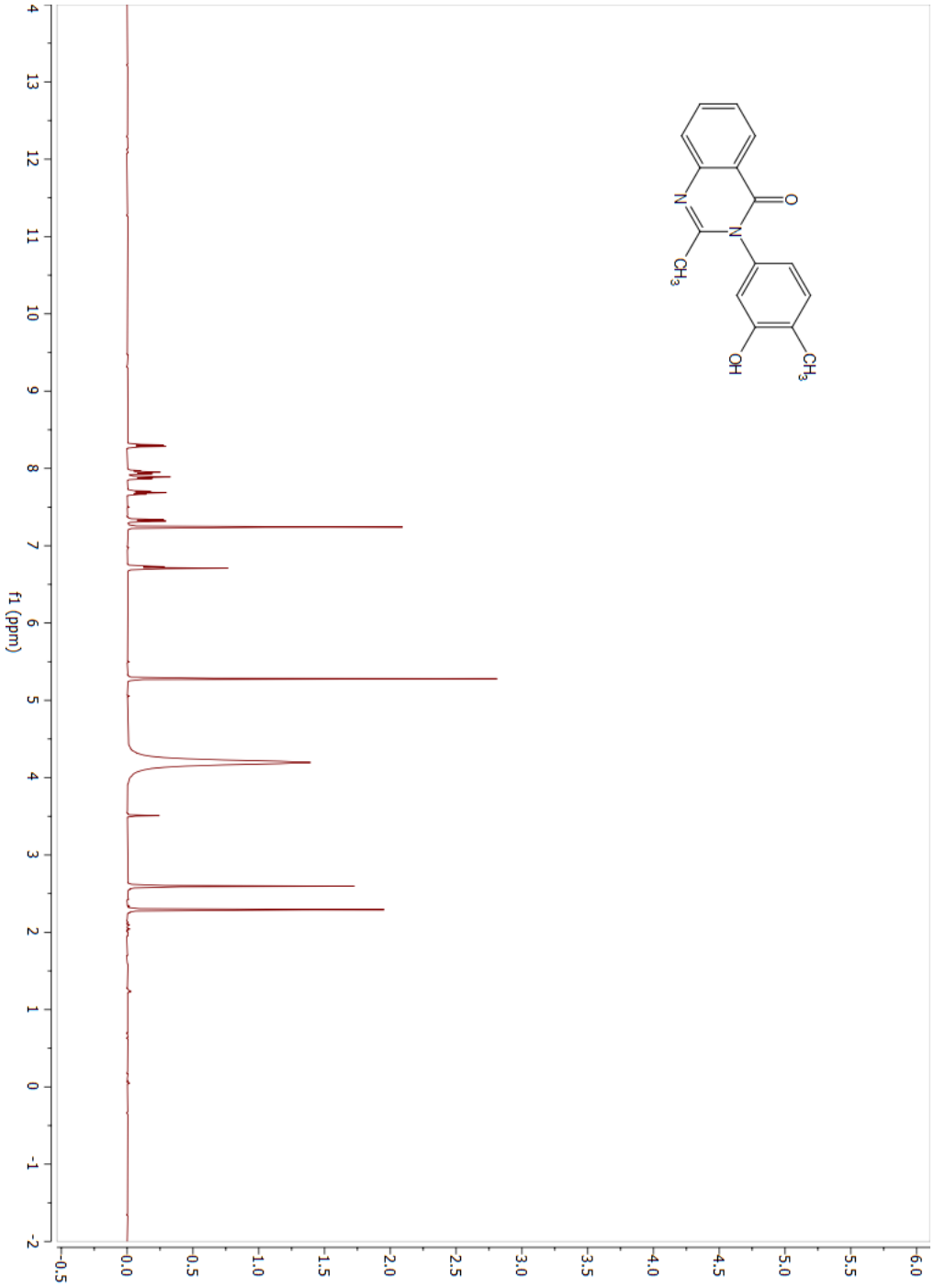


Figure AII. 42: 1HNMR compound 17

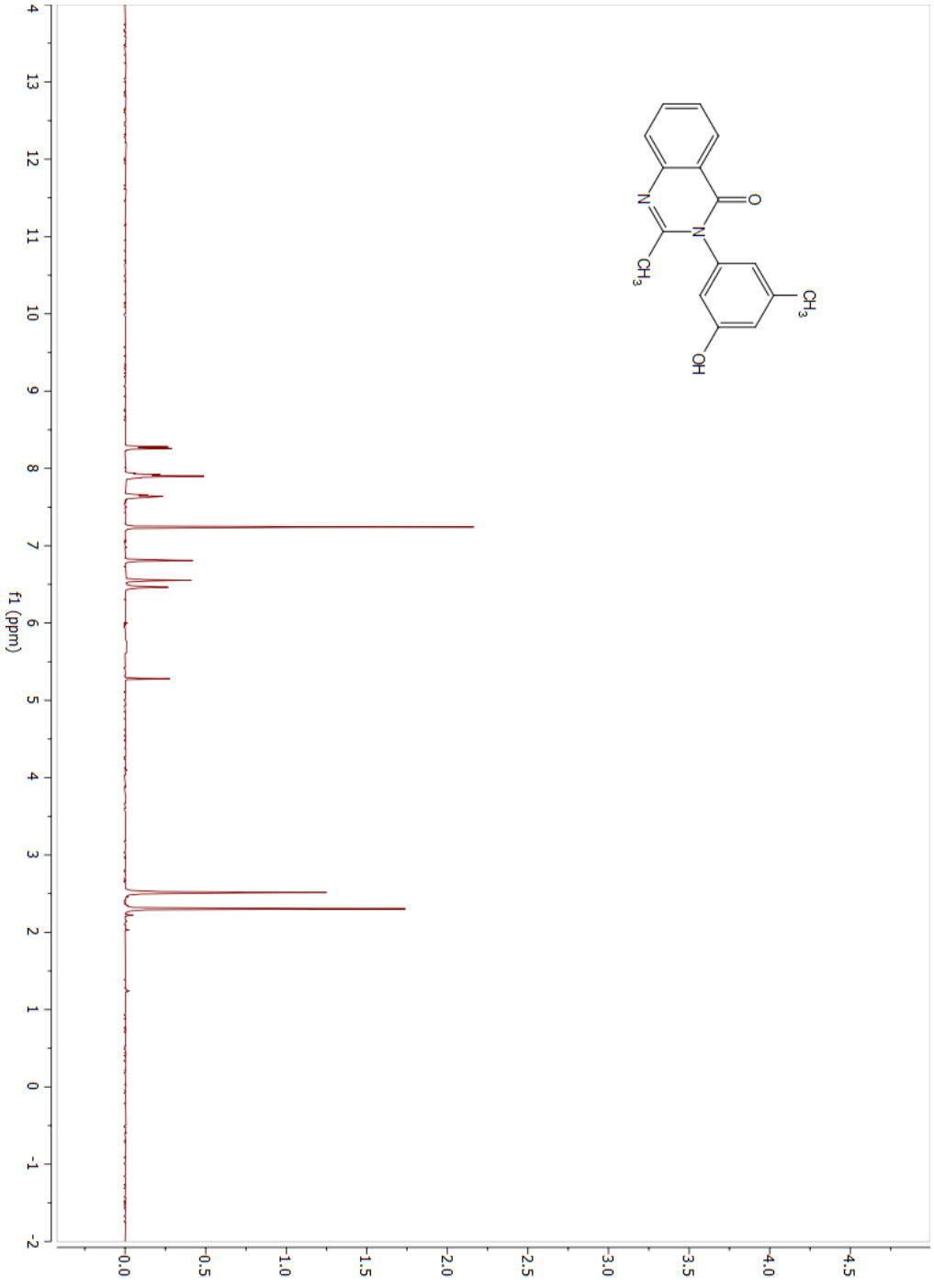


Figure AII. 43: 1HNMR compound 18

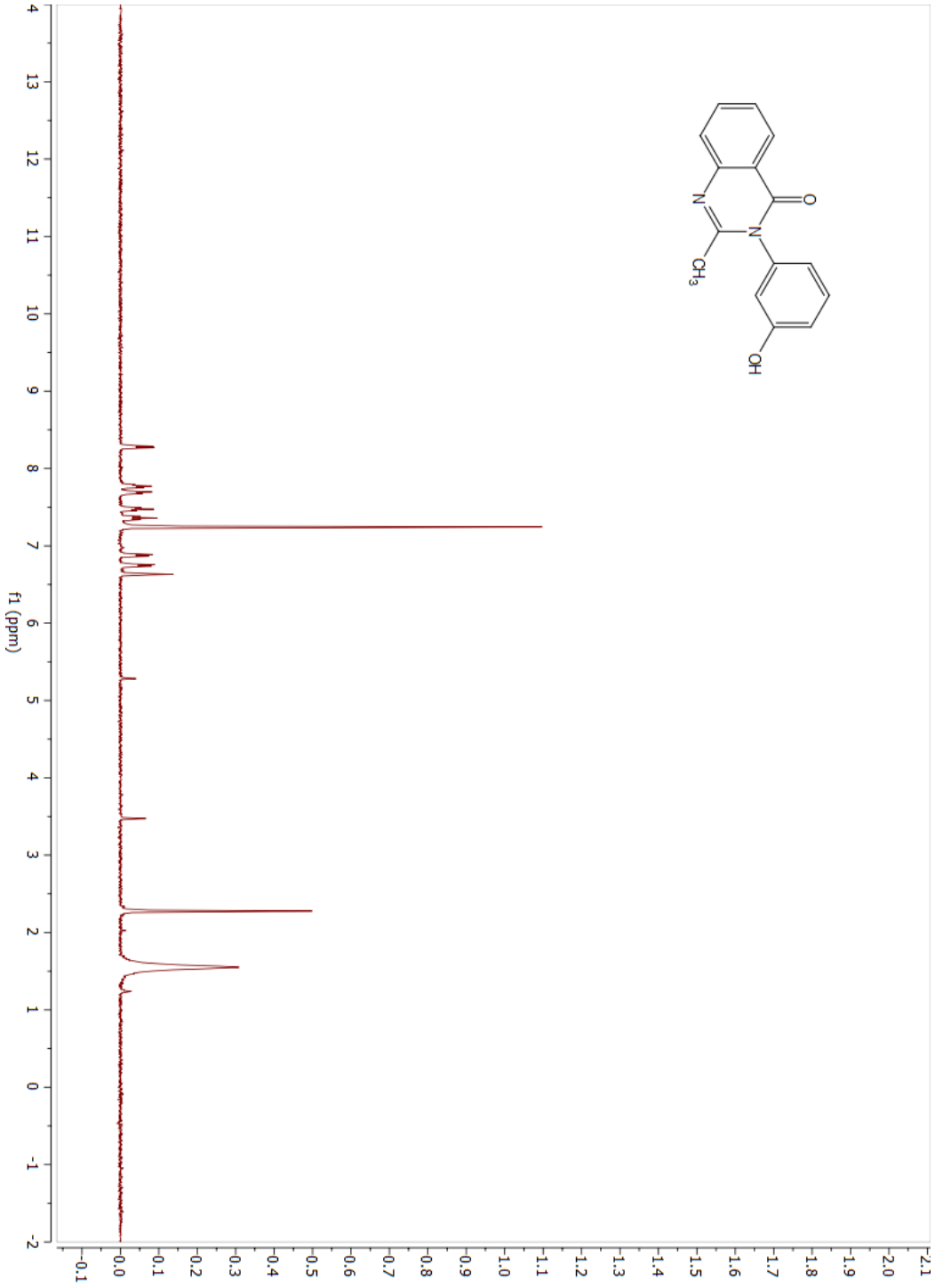


Figure AII. 44: 1HNMR Product 1a

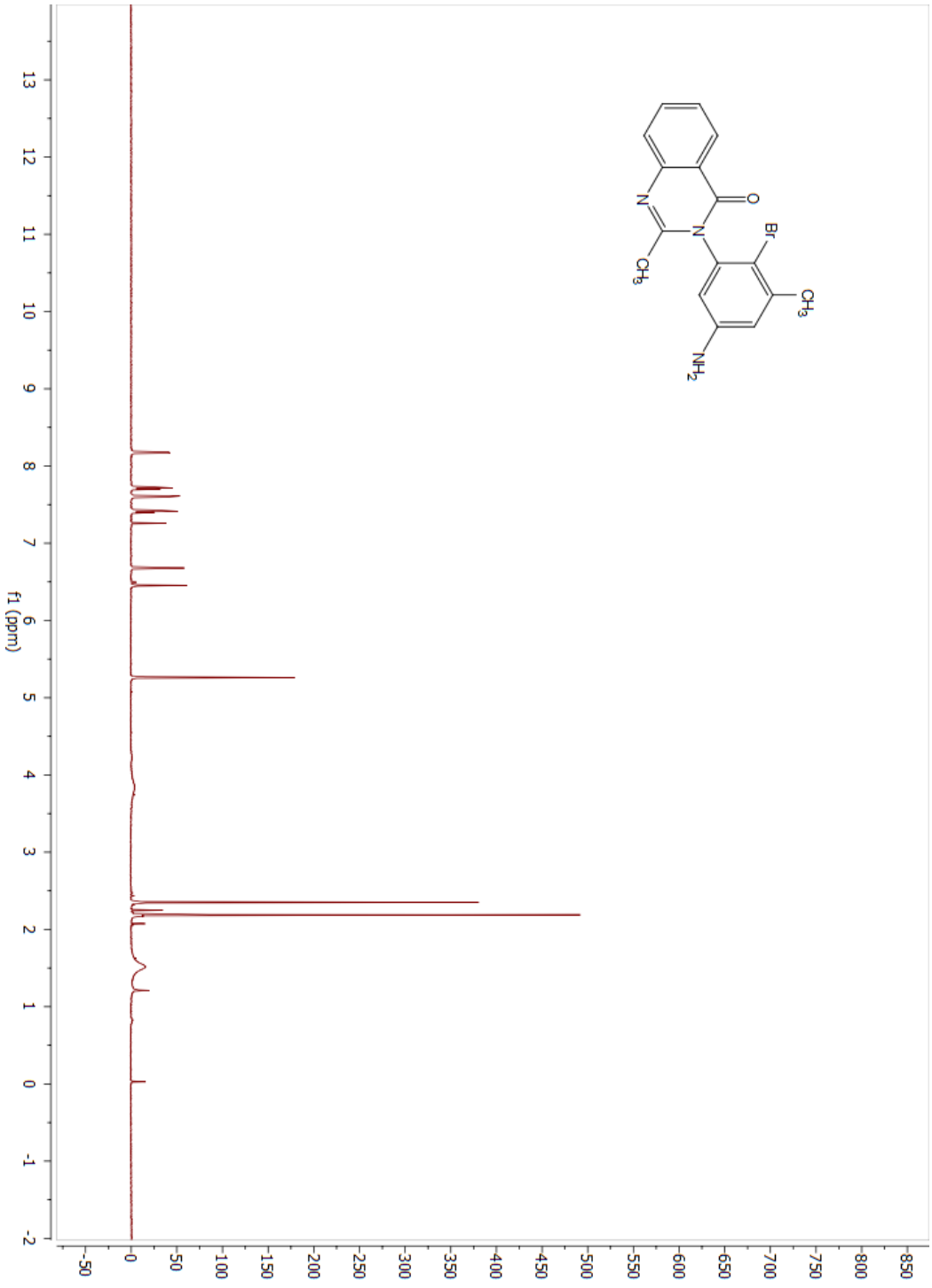


Figure AII. 45: 13CNMR Product 1a

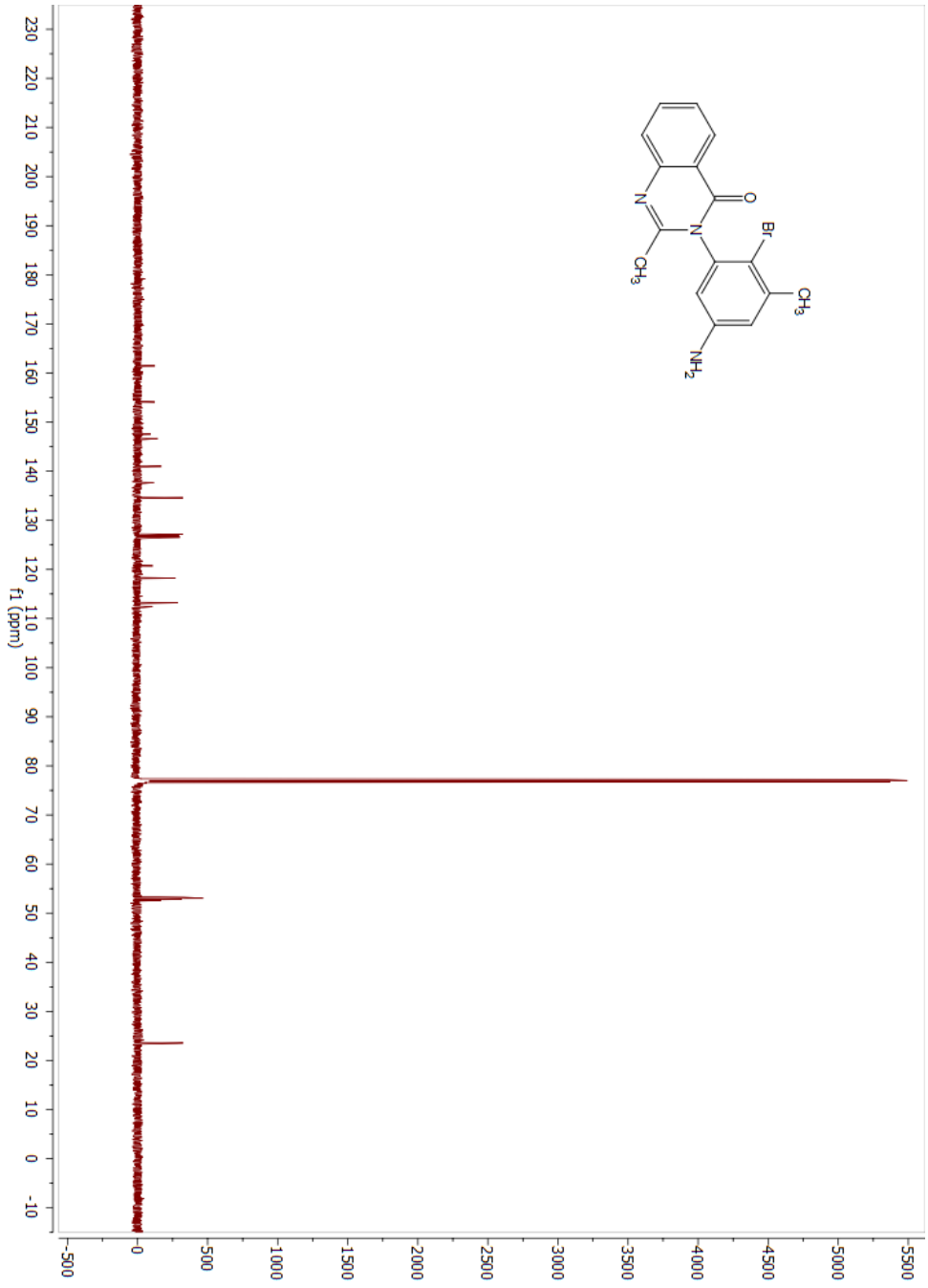


Figure AII. 46: 2D NOESY Product 1a

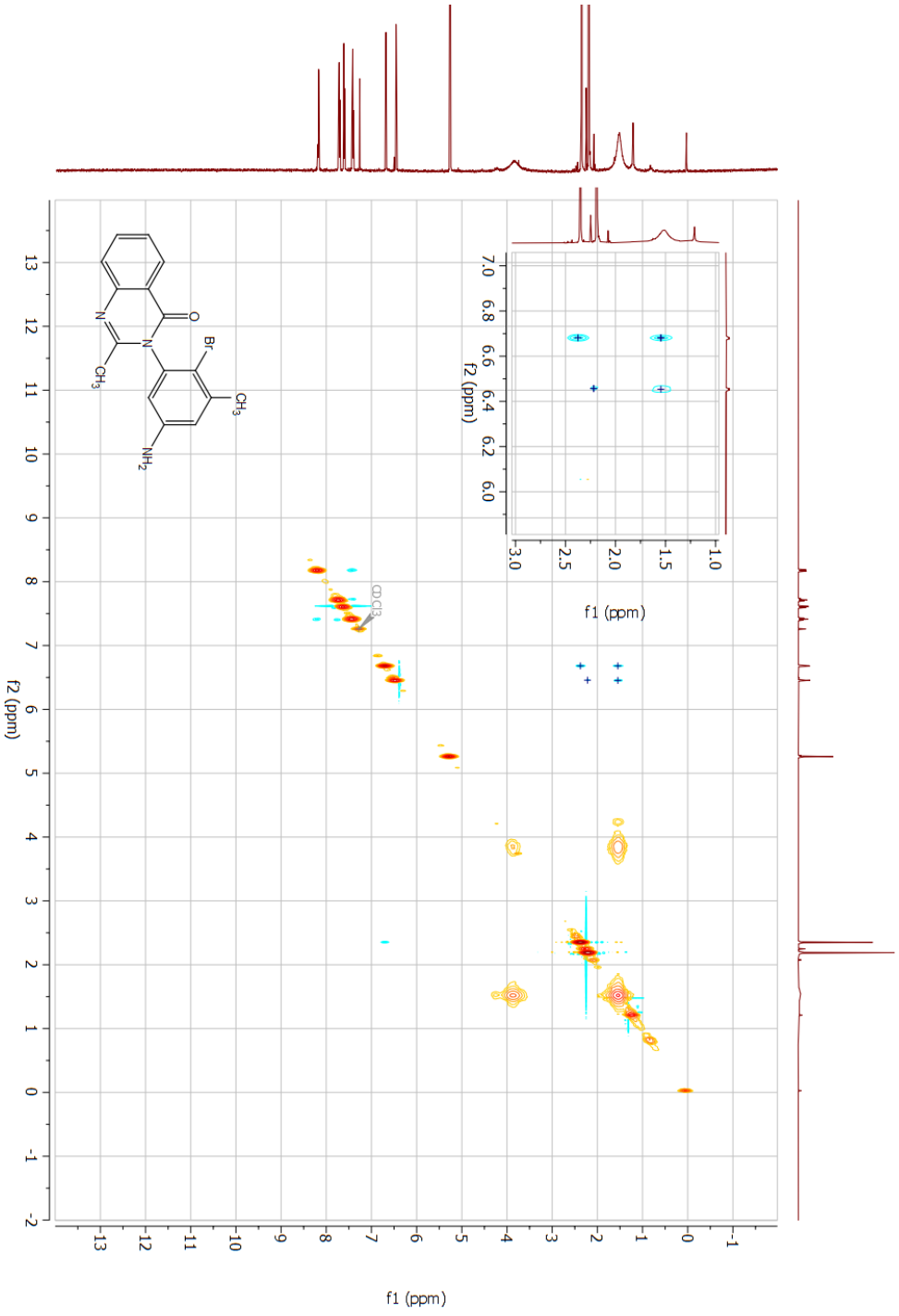


Figure AII. 47: 1H NMR Product 1b

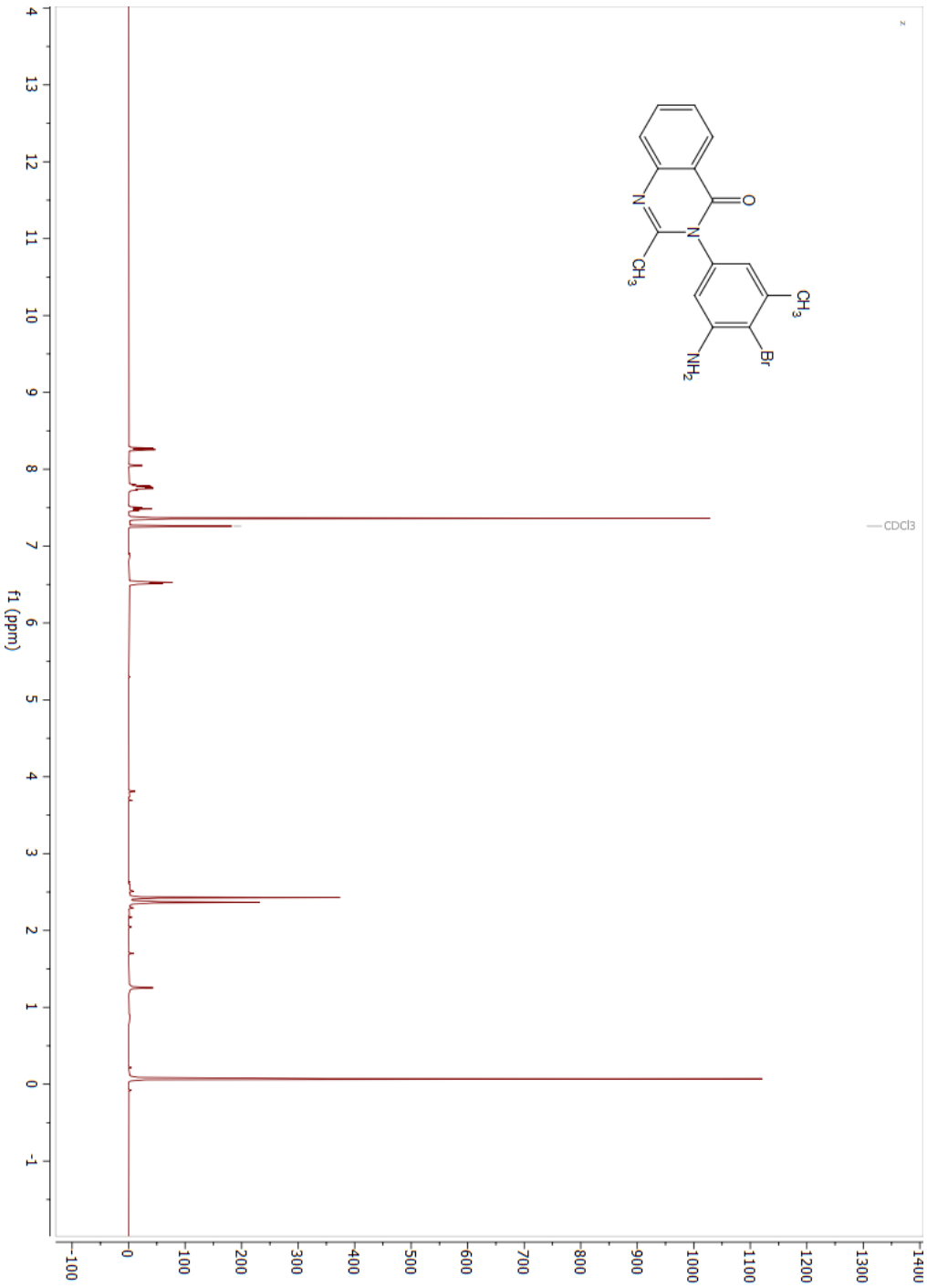


Figure AII. 48: 2D NOESY Product 1b

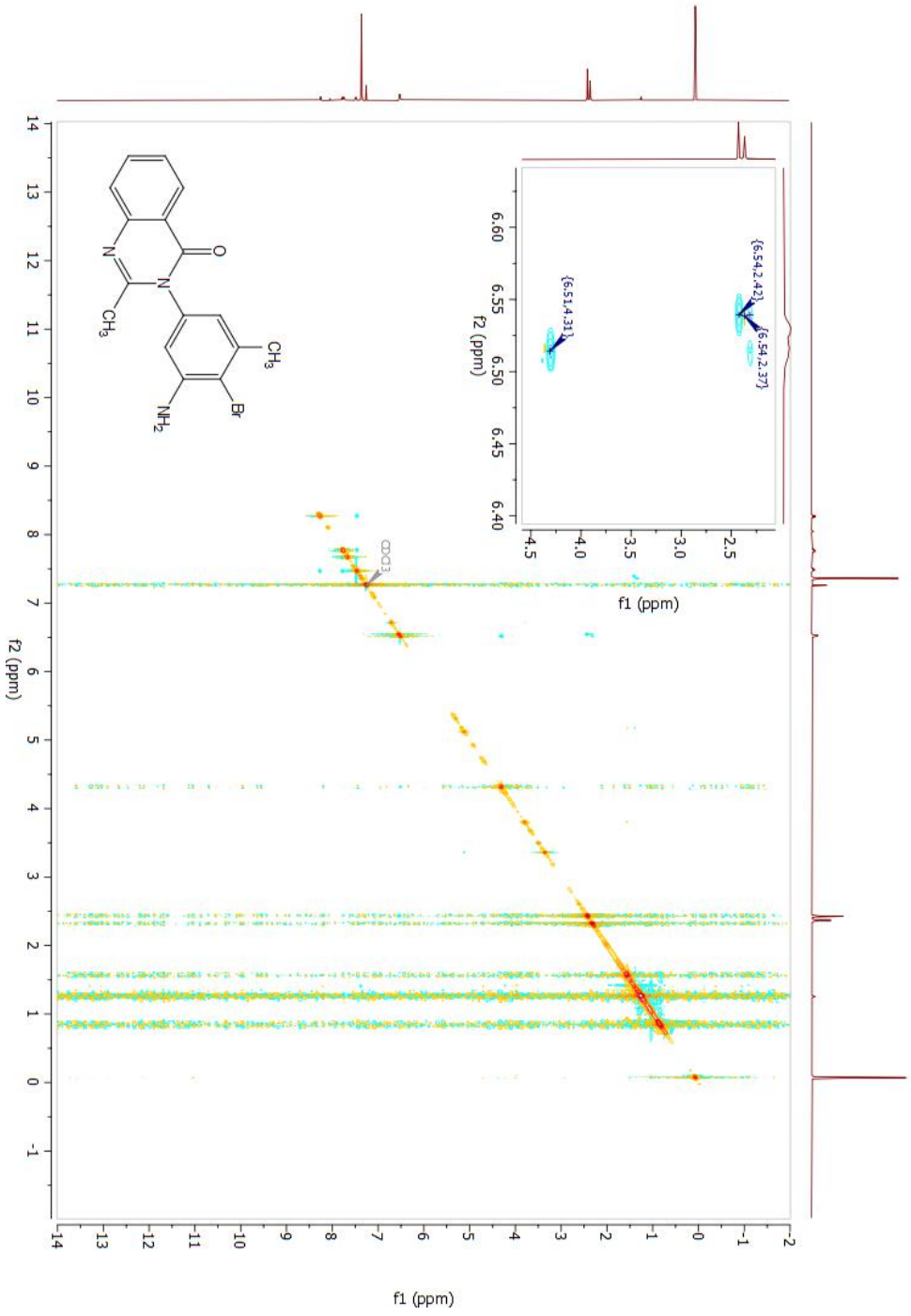


Figure AII. 49: 1H NMR Product 2a

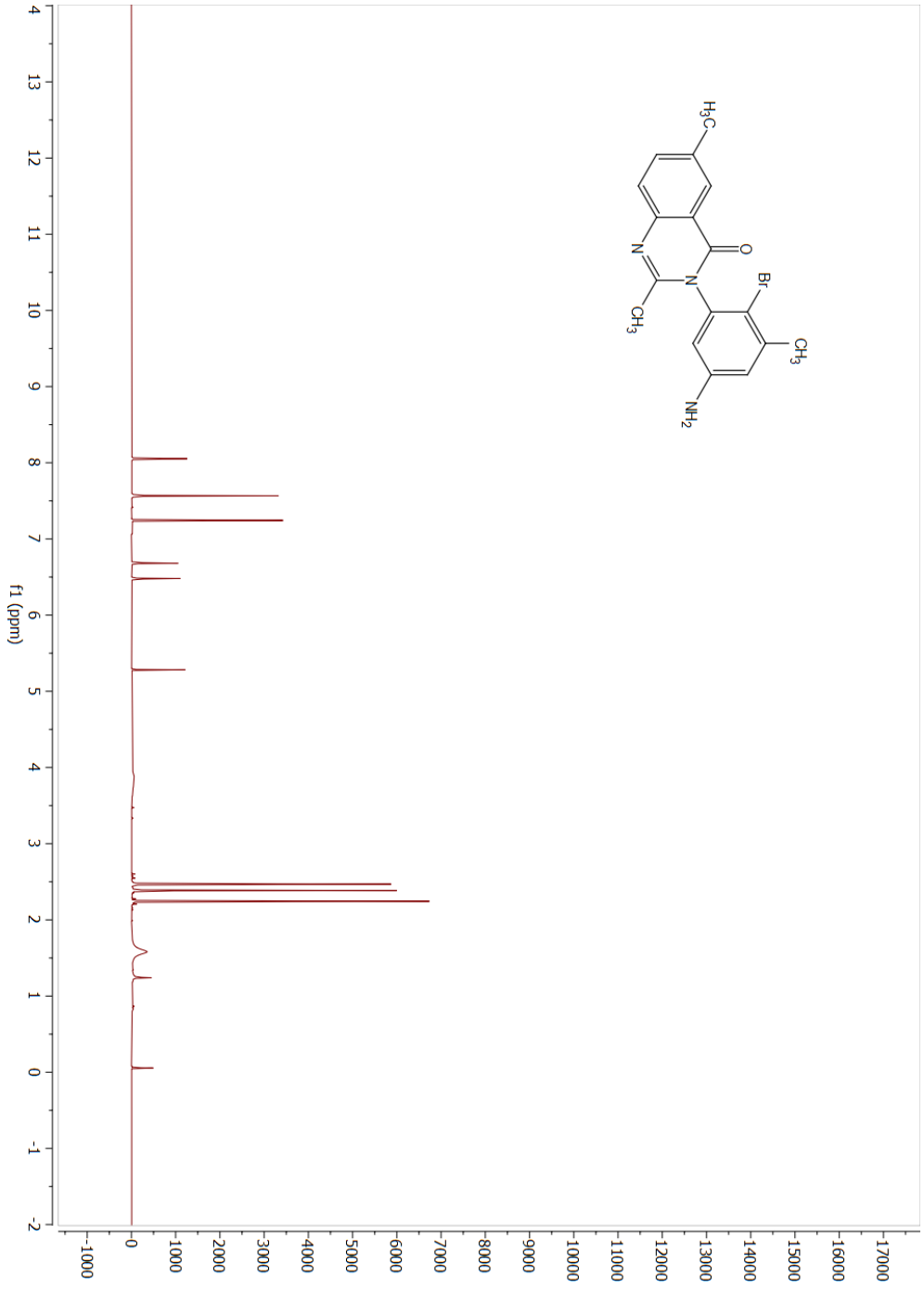


Figure AII. 50: 13C NMR Product 2a

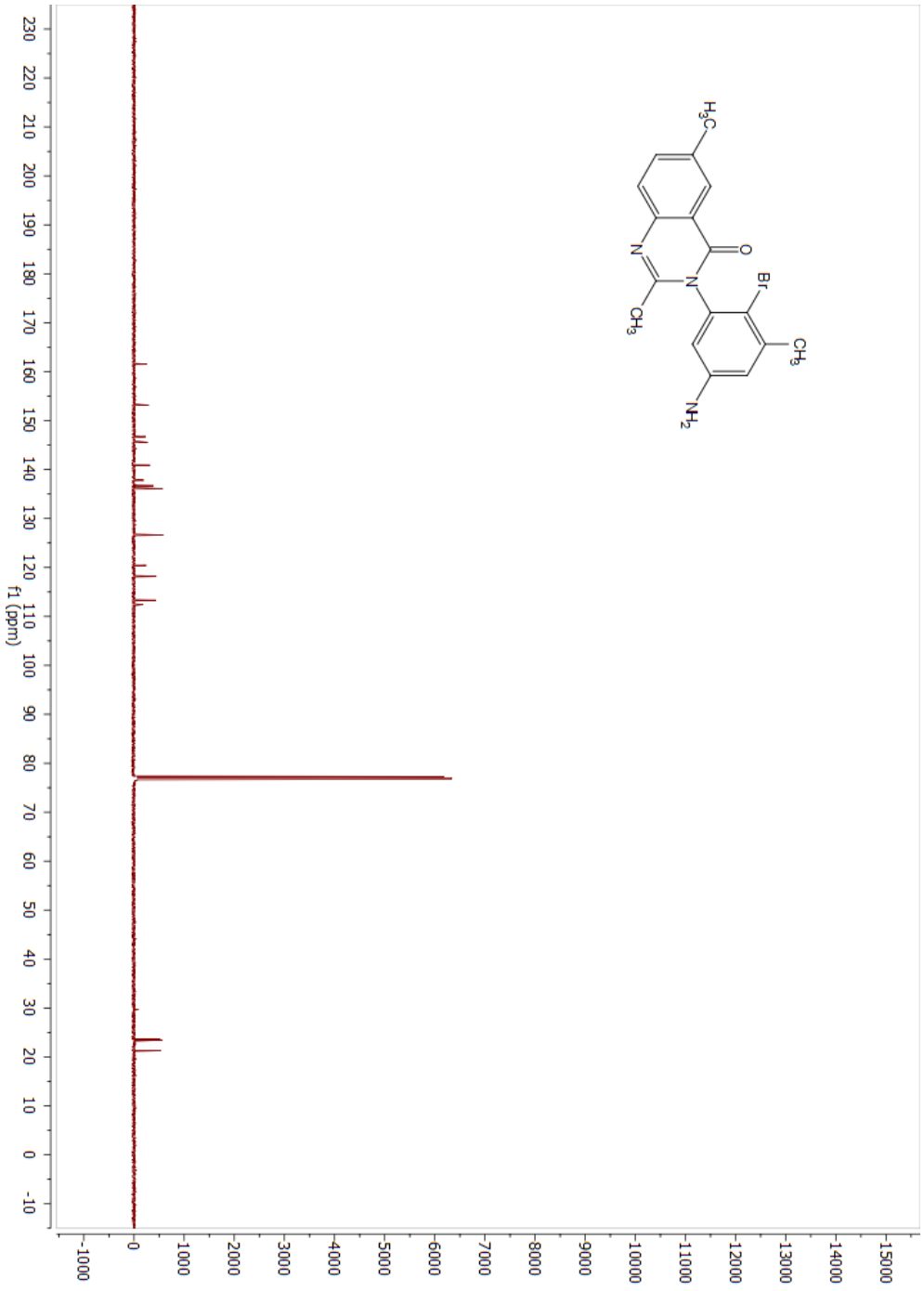


Figure AII. 51: 2D NOESY 2a

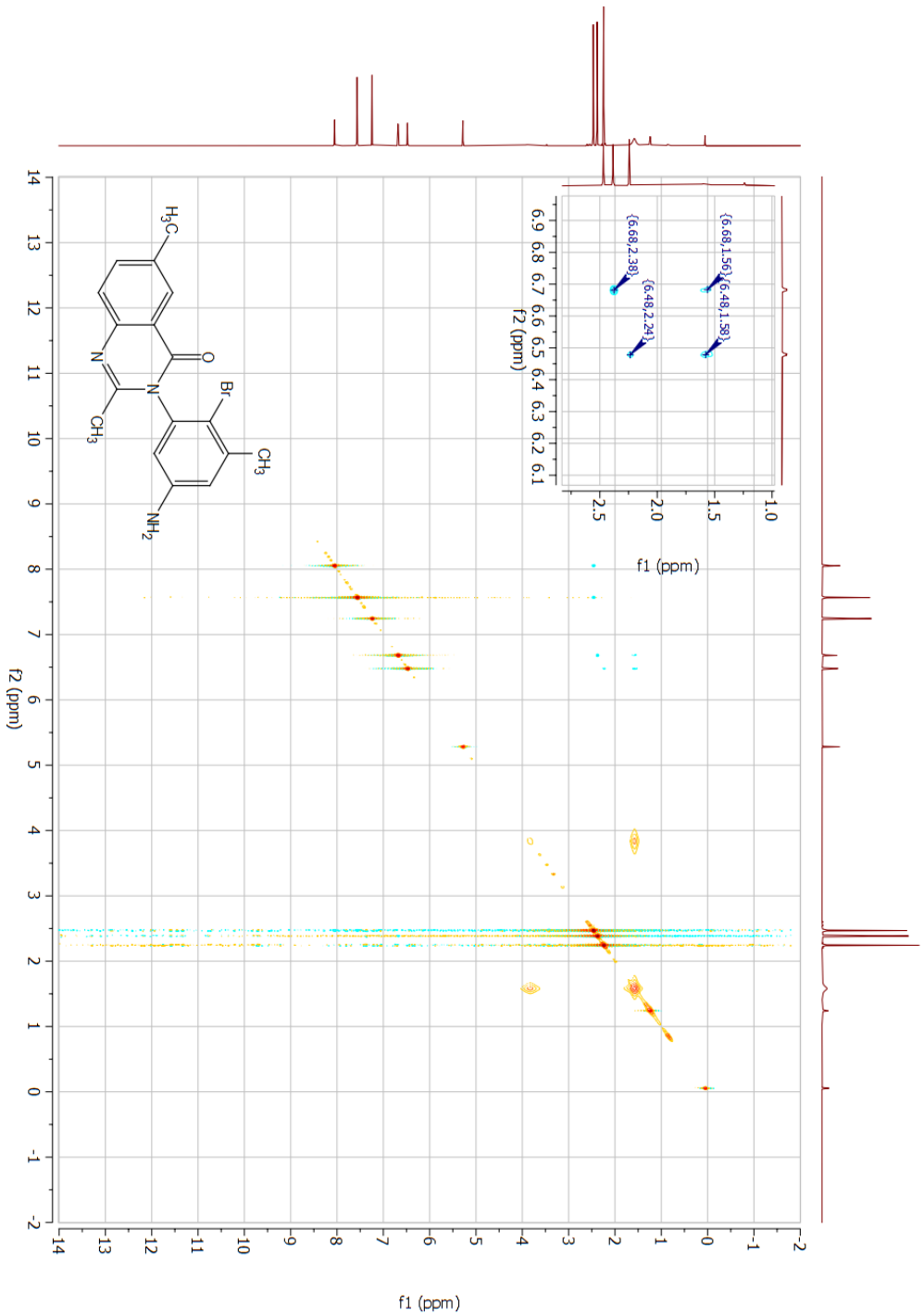


Figure AII. 52: 1H NMR 3a

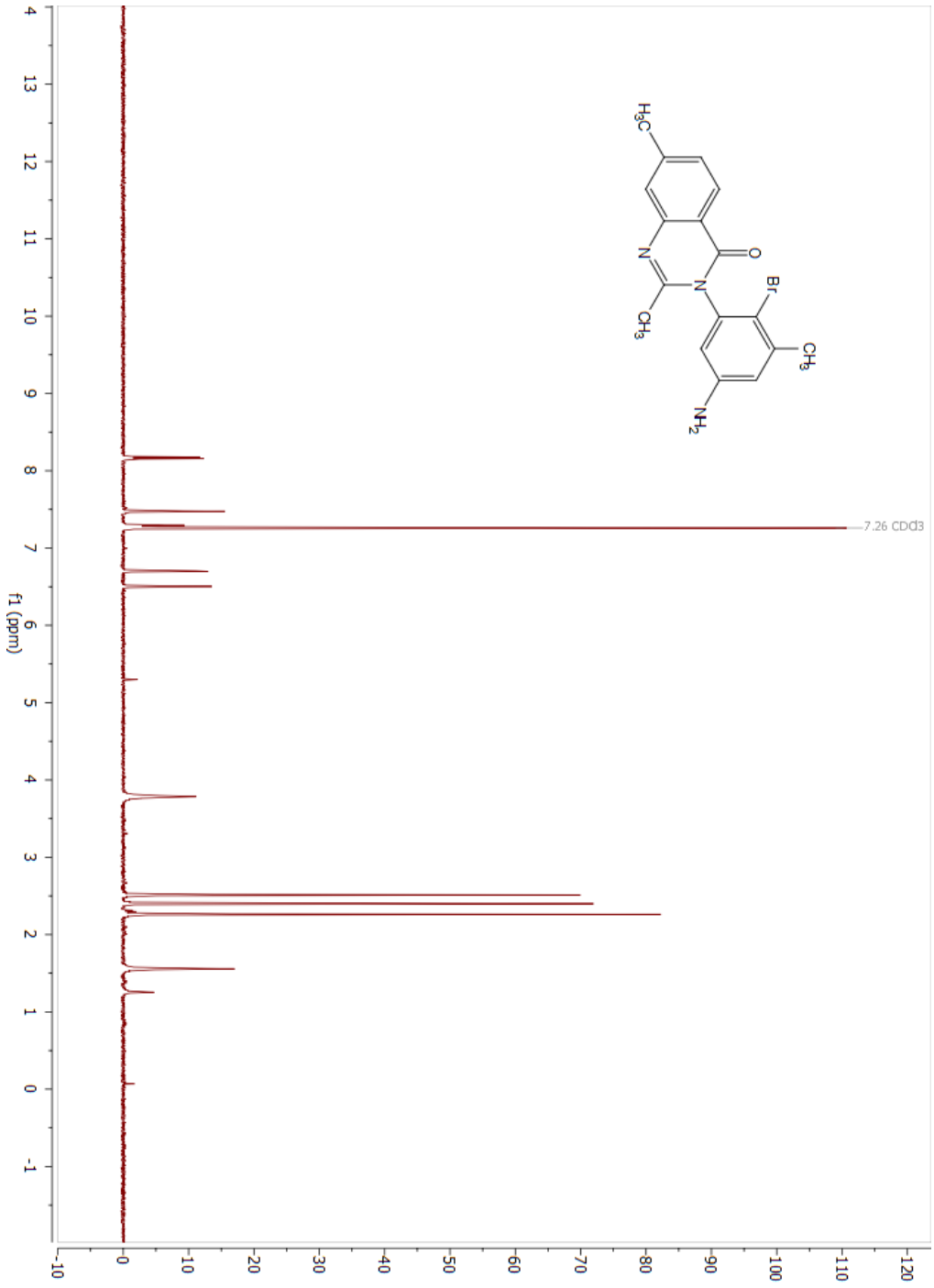


Figure AII. 53: 13CNMR 3a

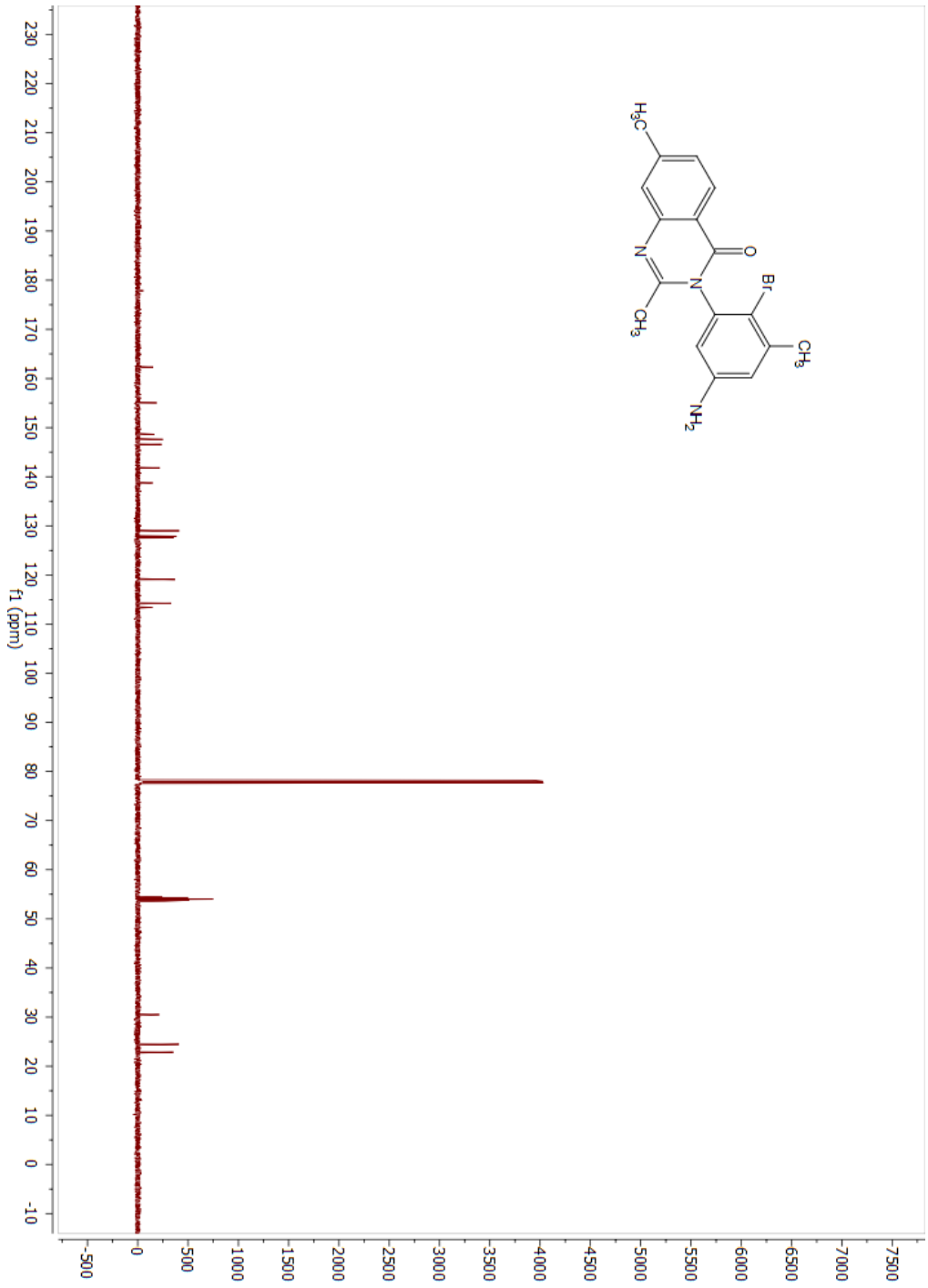


Figure AII. 54: 2D NOESY 3a

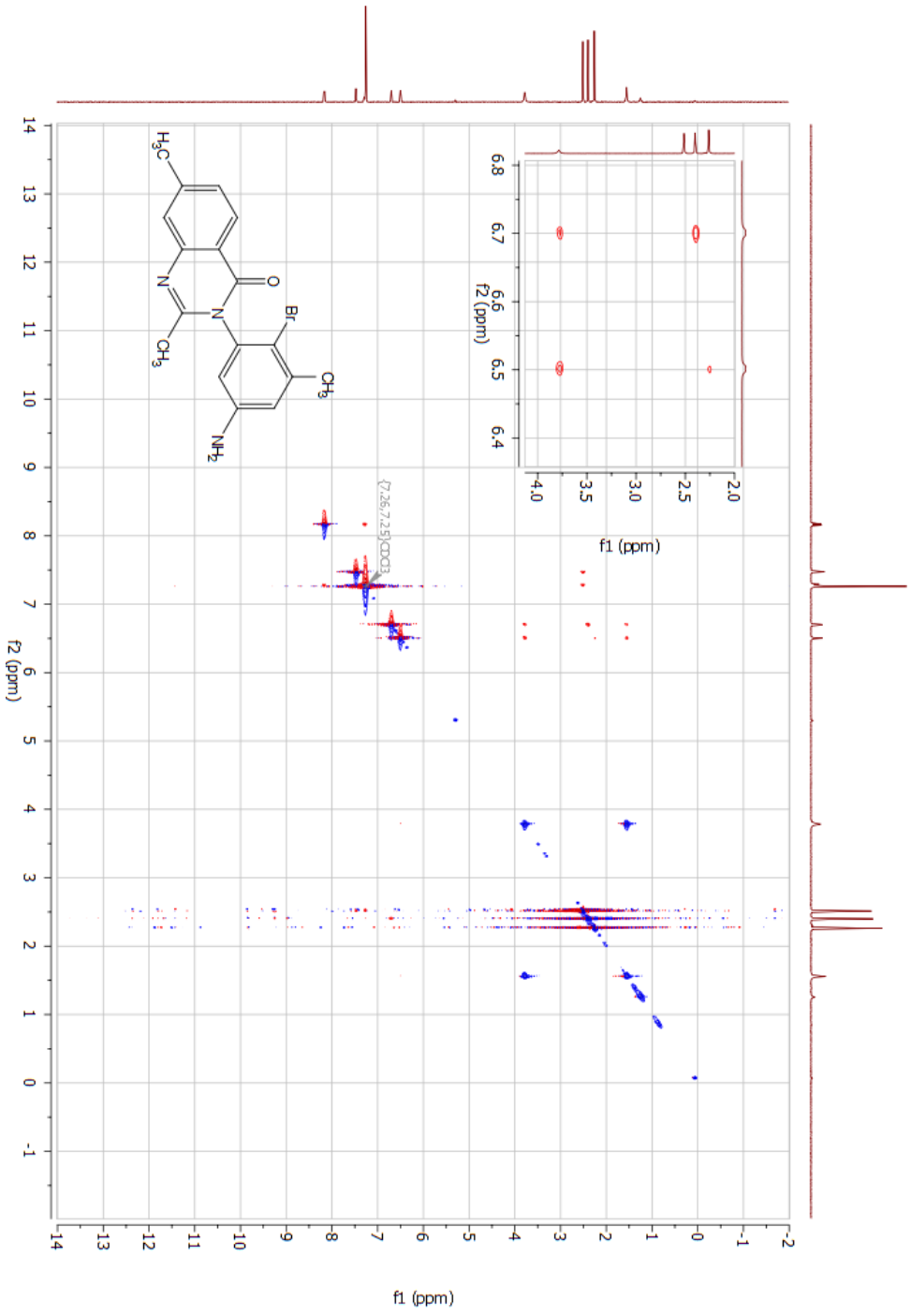


Figure AII. 55: 1HNMR 4a

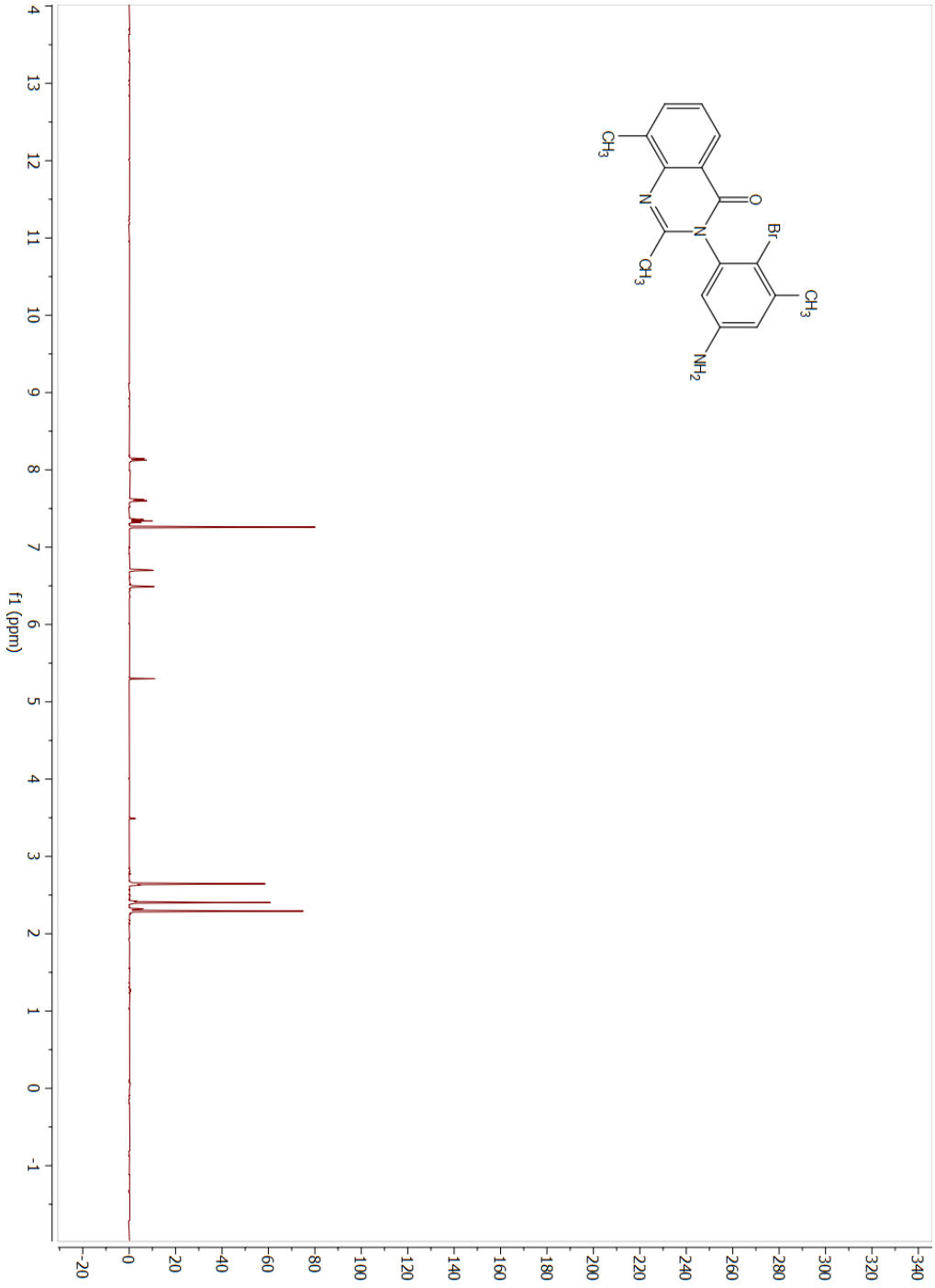


Figure AII. 56: 13CNMR 4a

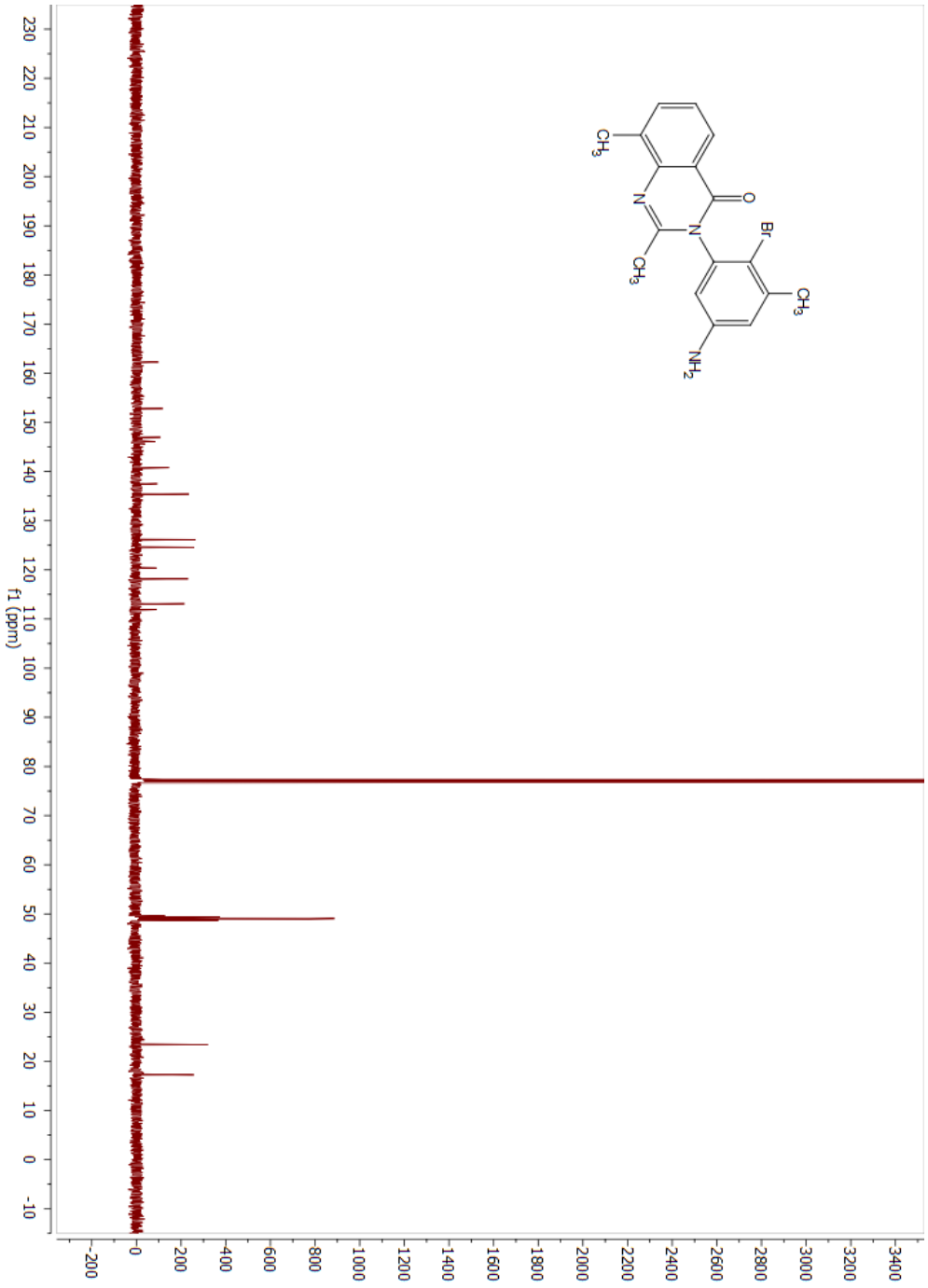


Figure AII. 57: 2DNOESY 4a

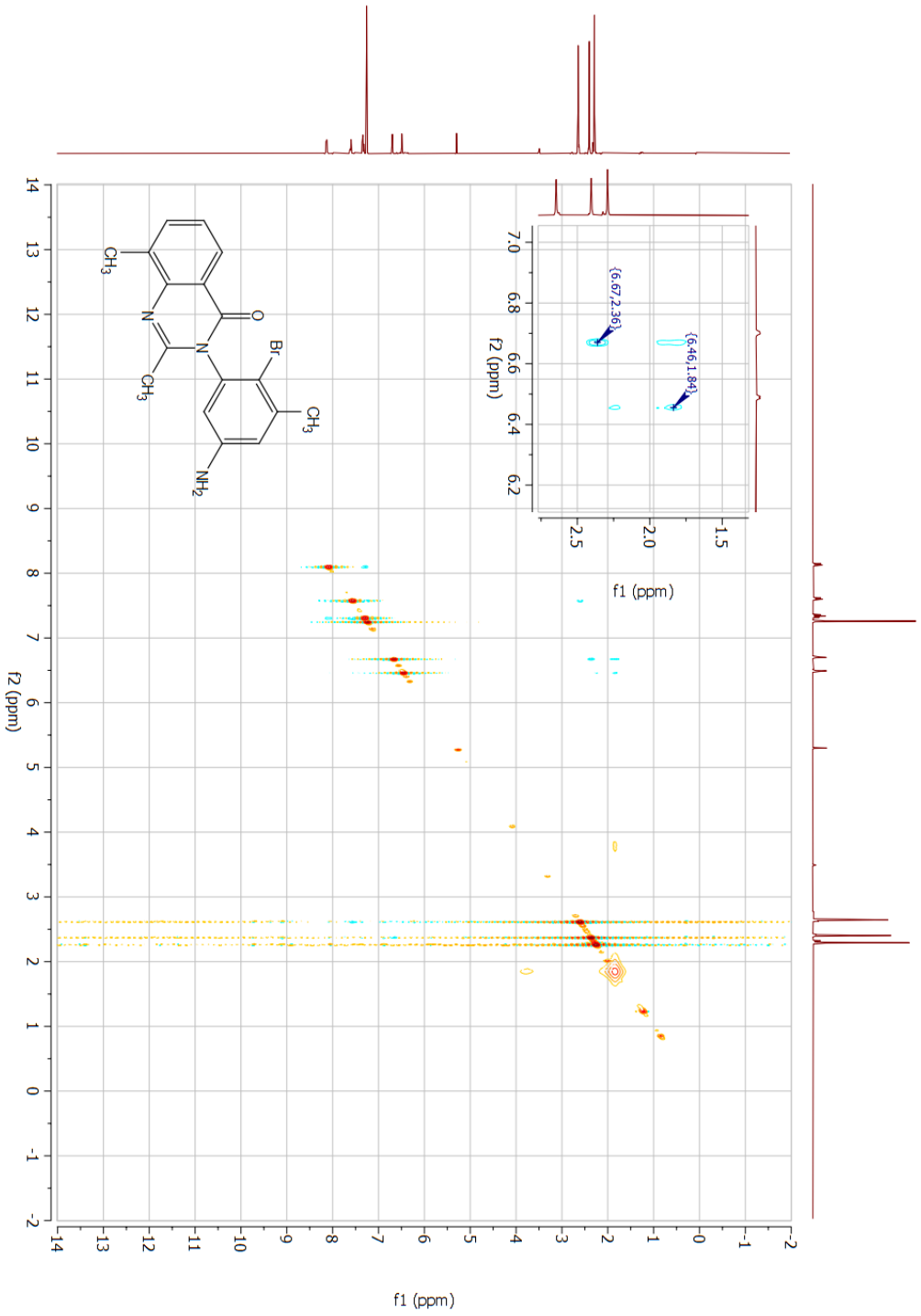


Figure AII. 58: 1HNMR 5a

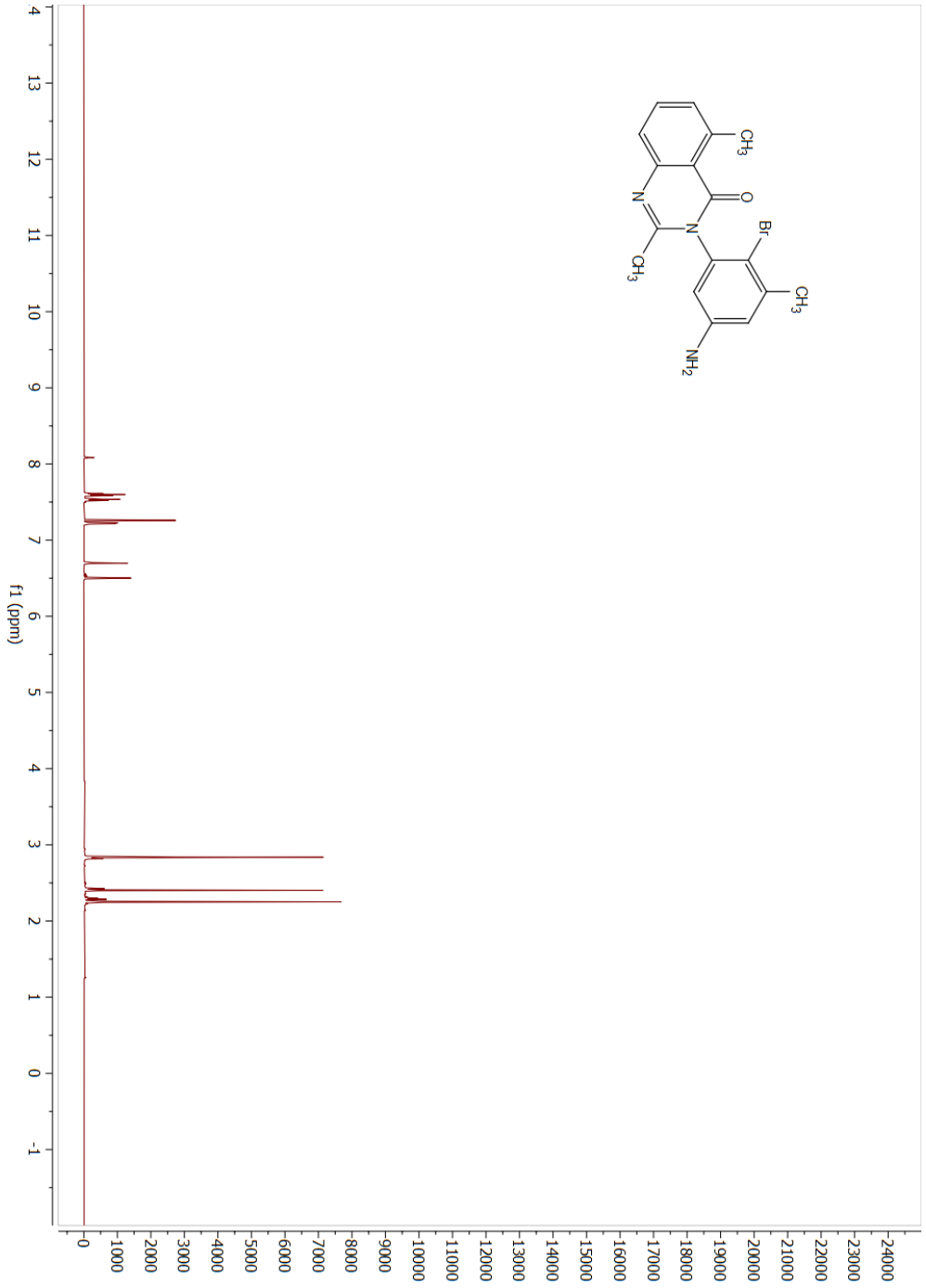


Figure AII. 59: 13CNMR 5a

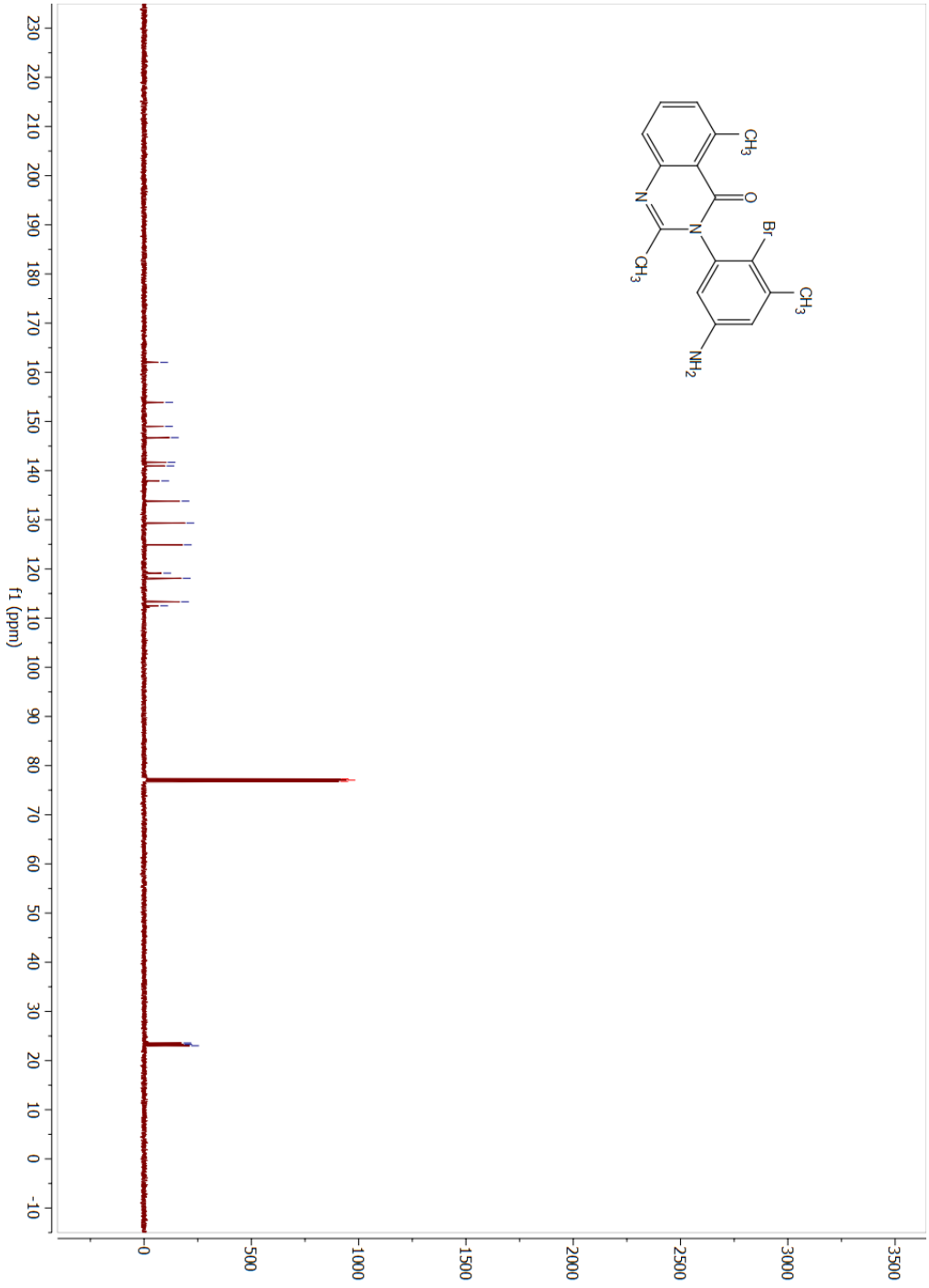


Figure AII. 60: 2DNOESY 5a

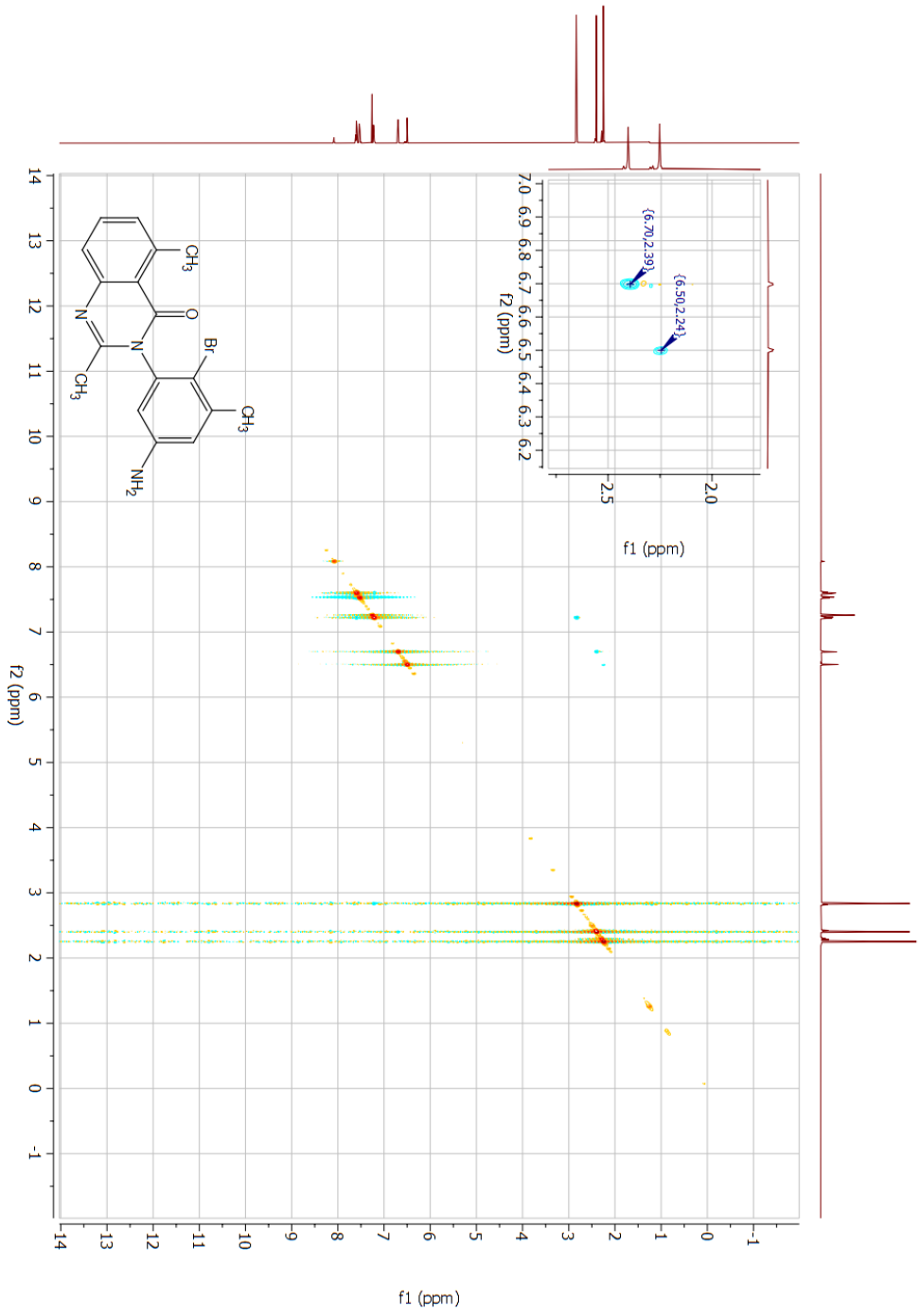


Figure AII. 61: 1HNMR 6a

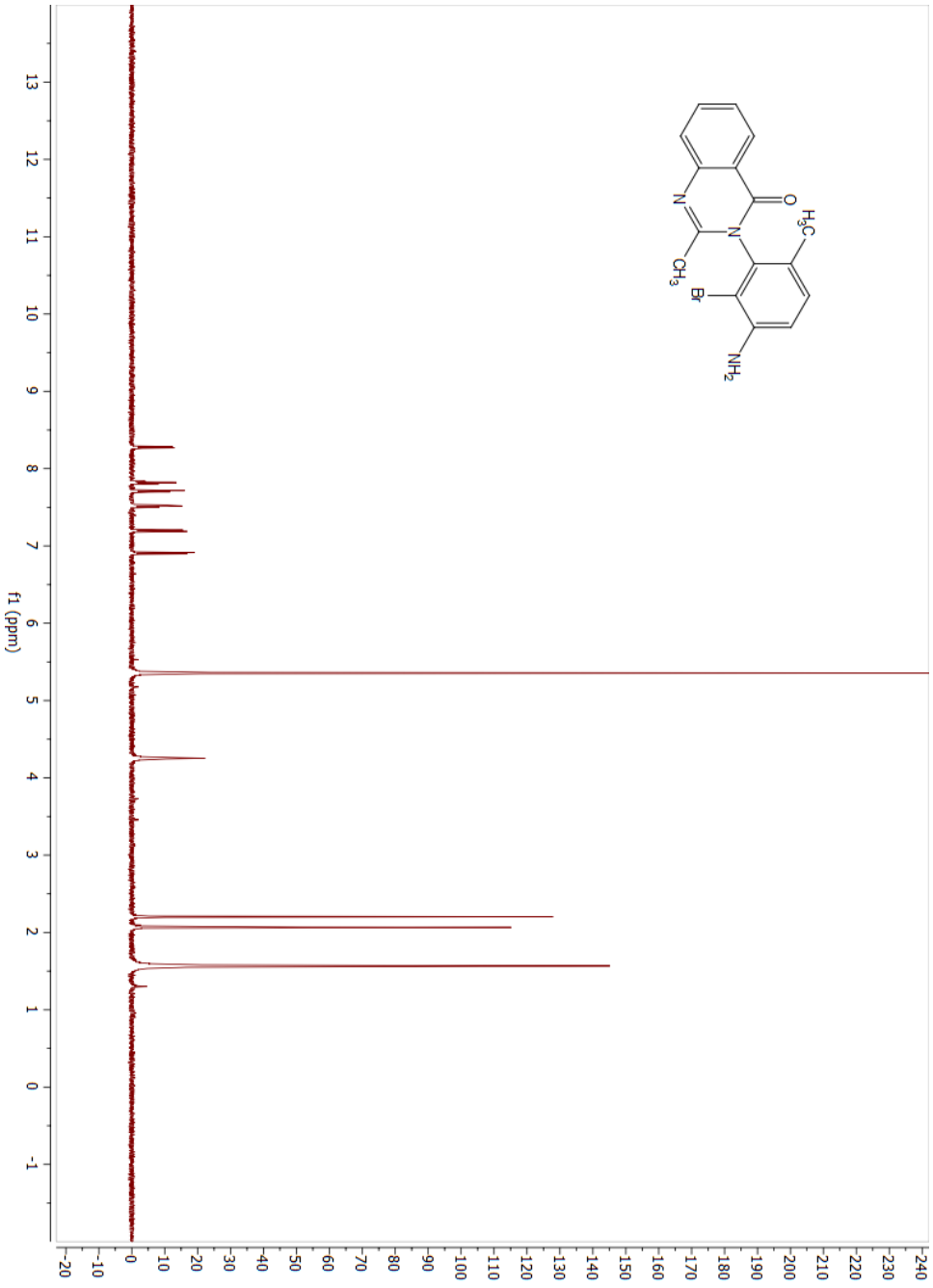


Figure AII. 62: 13CNMR 6a

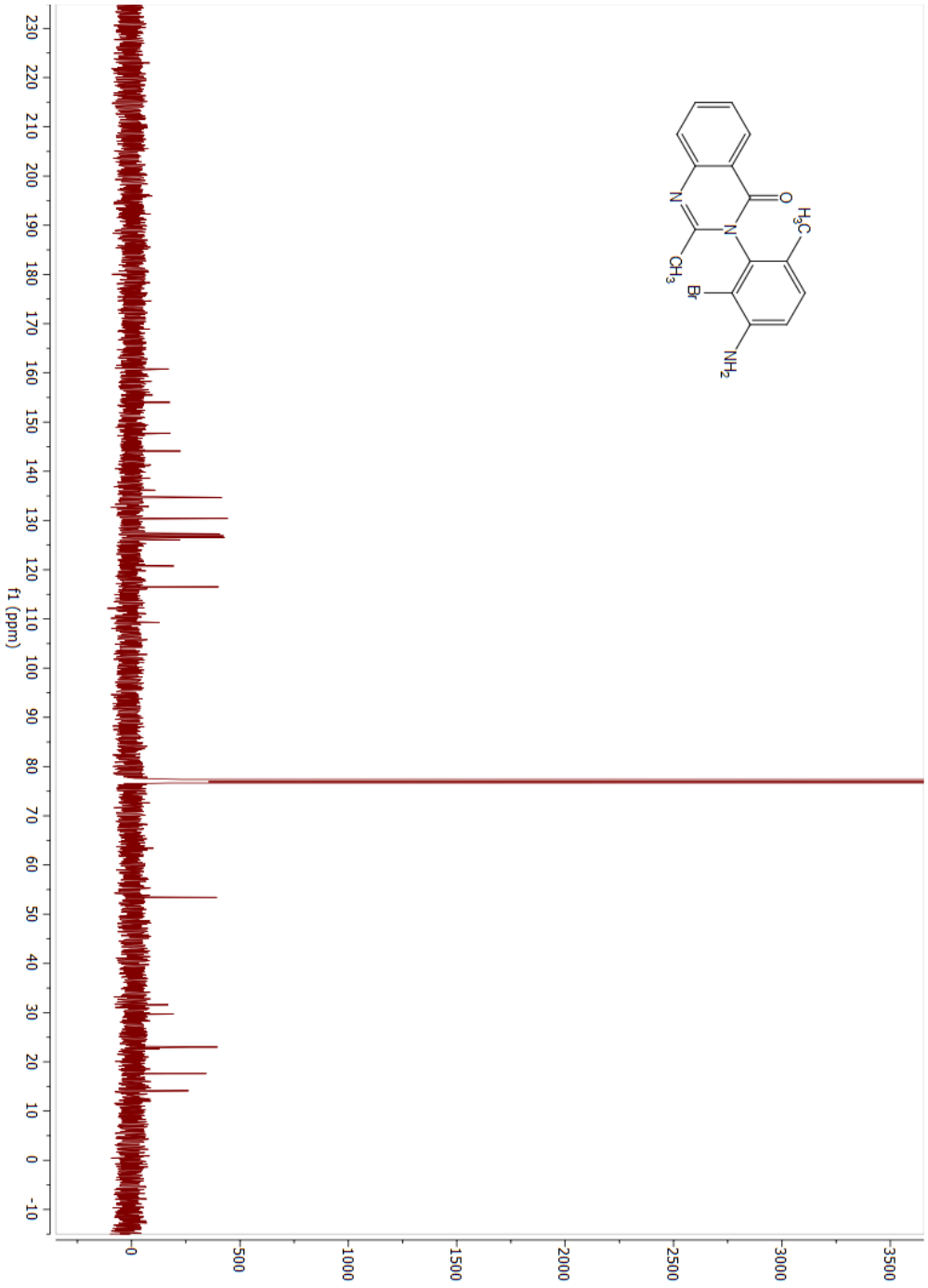


Figure AII. 63: 1HNMR 8a

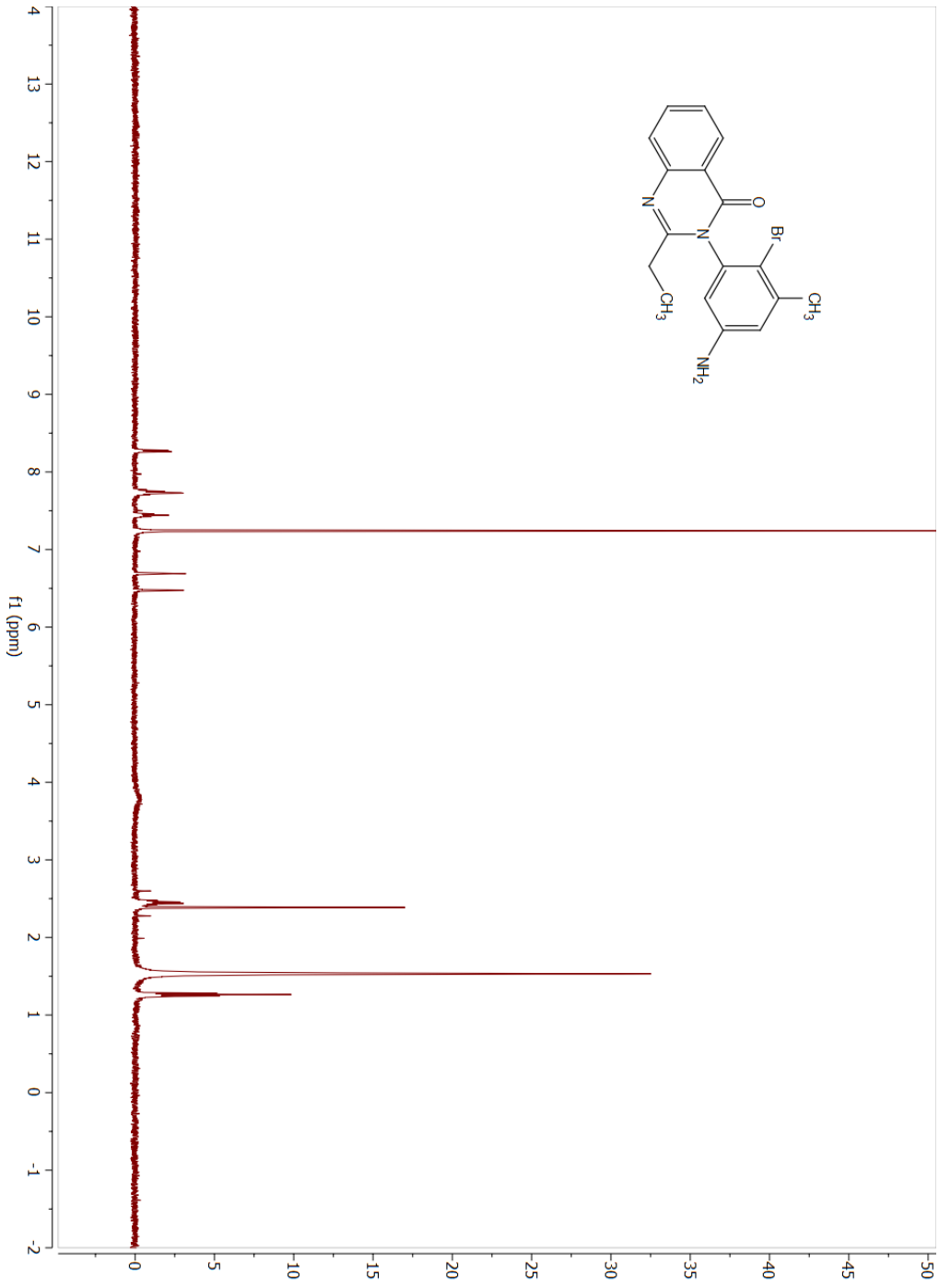


Figure AII. 64: 13CNMR 8a

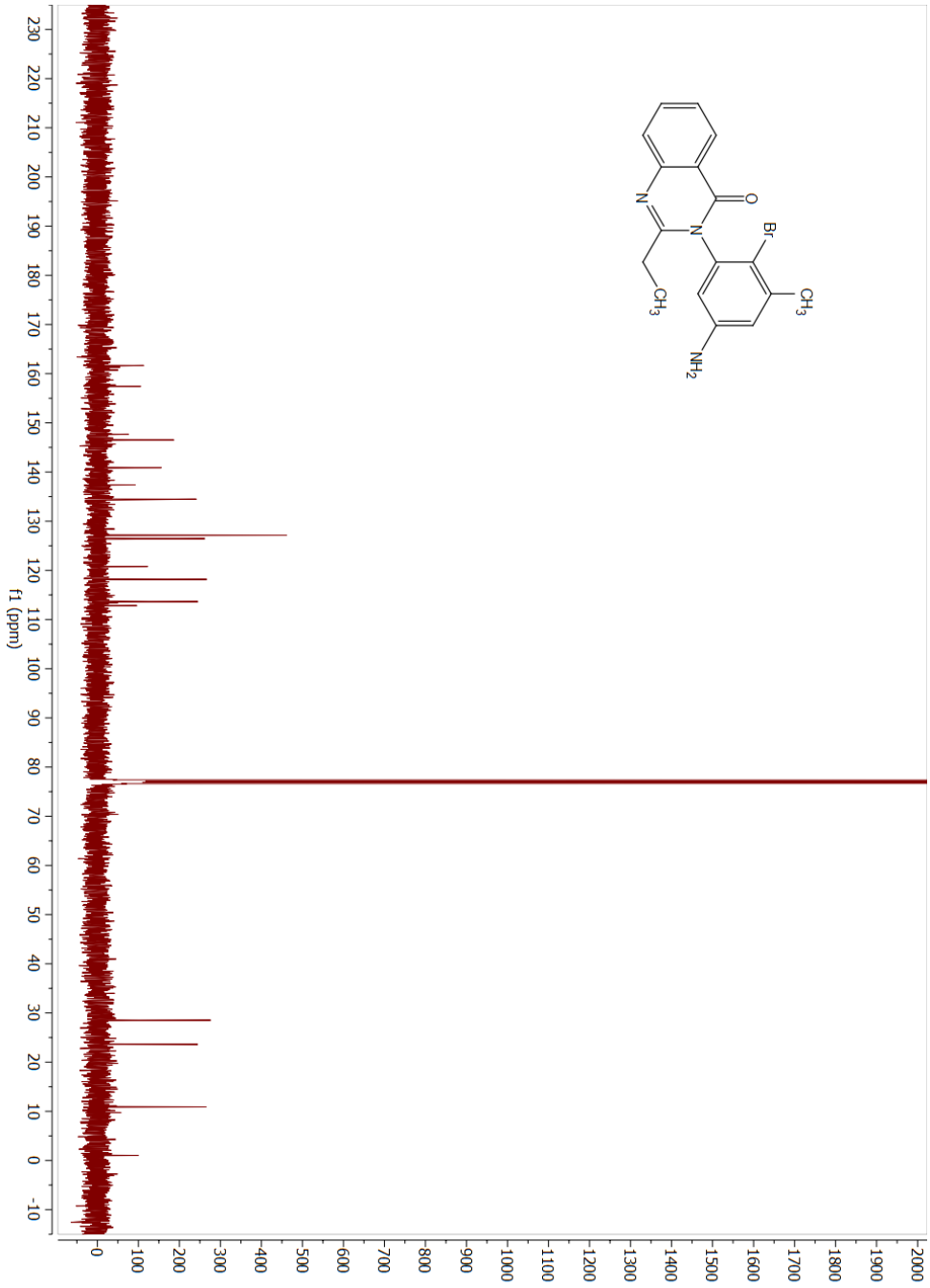


Figure AII. 65: 2DNOESY 8a

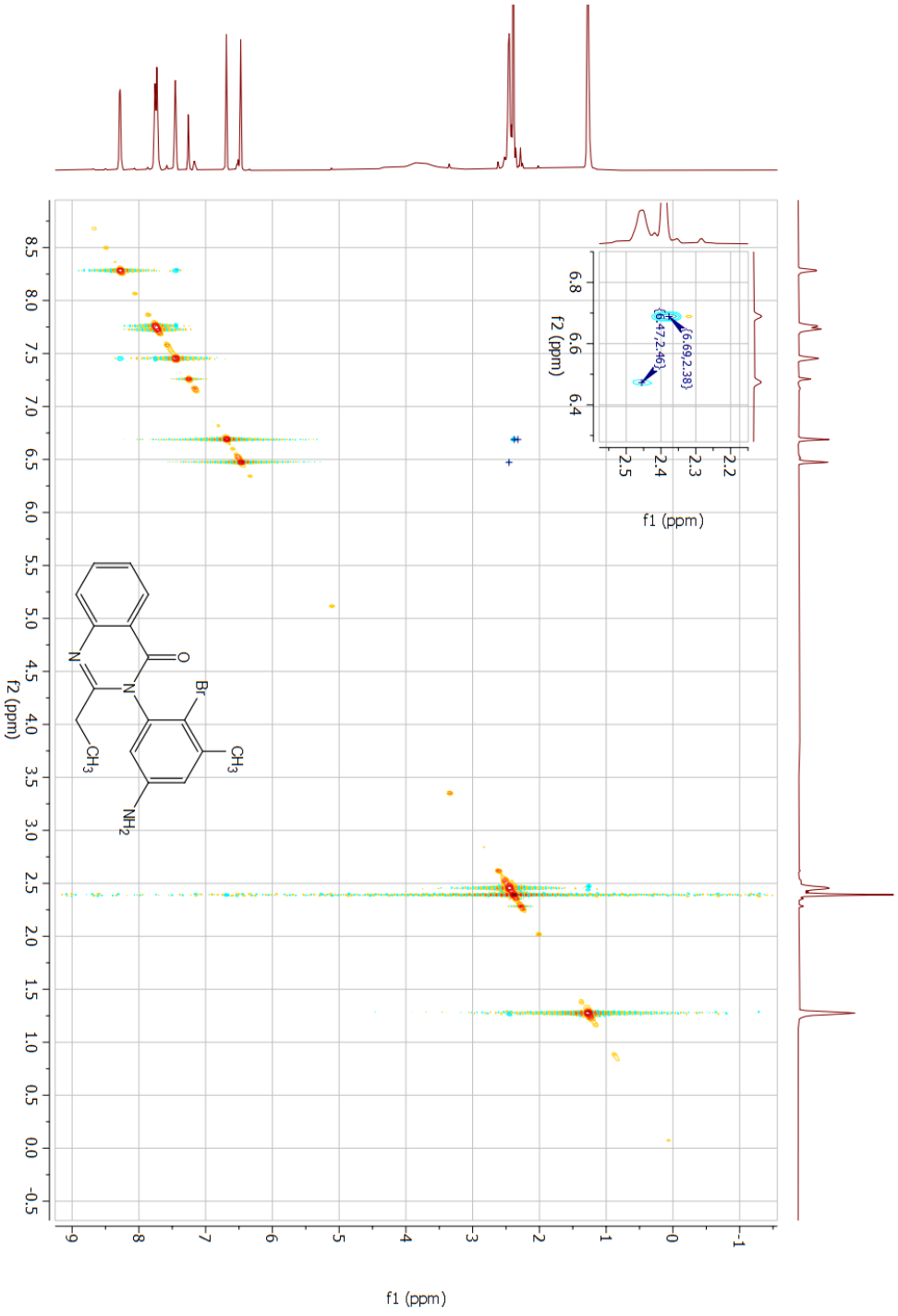


Figure AII. 66: 1HNMR 9a

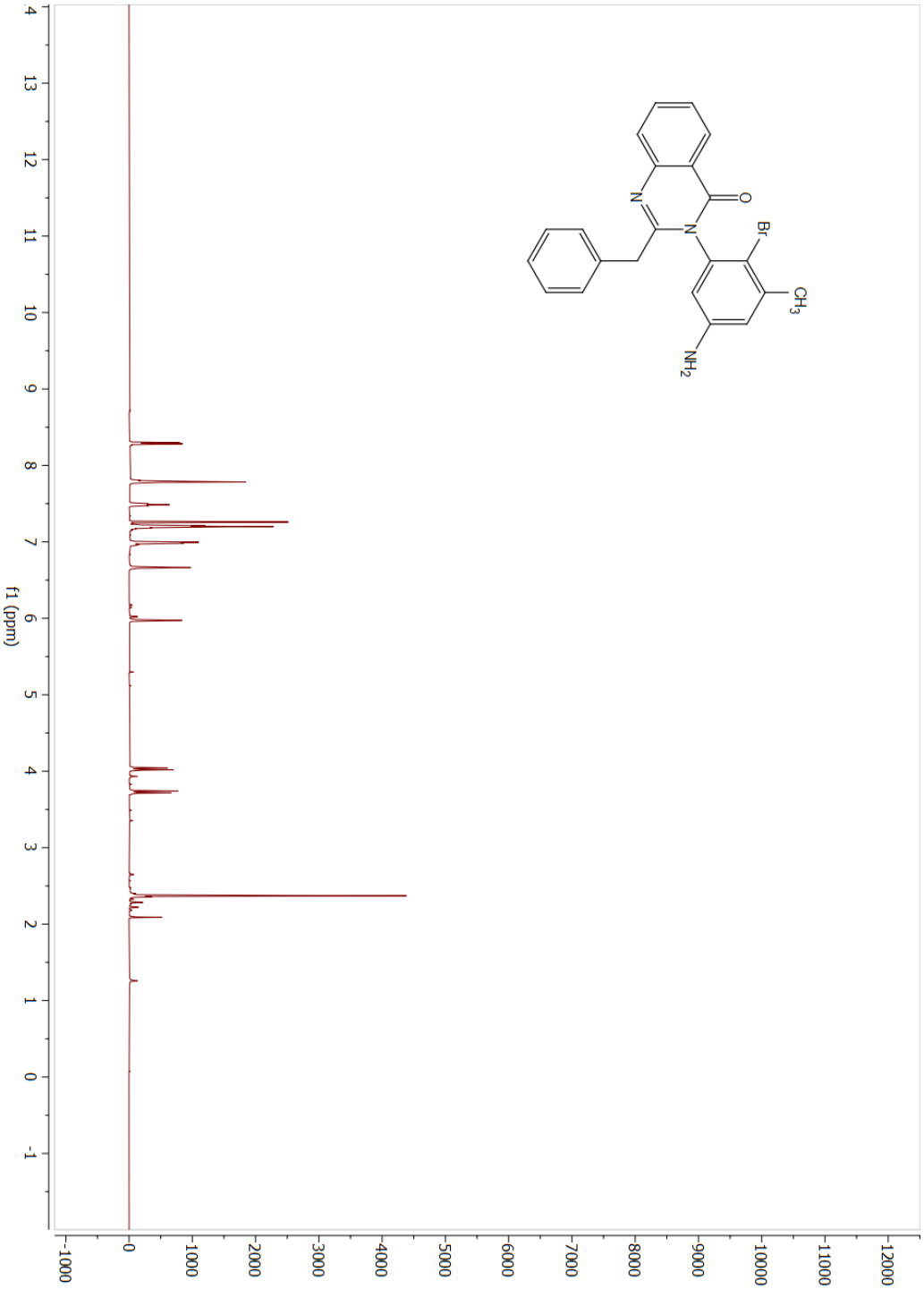


Figure AII. 67: 13CNMR 9a

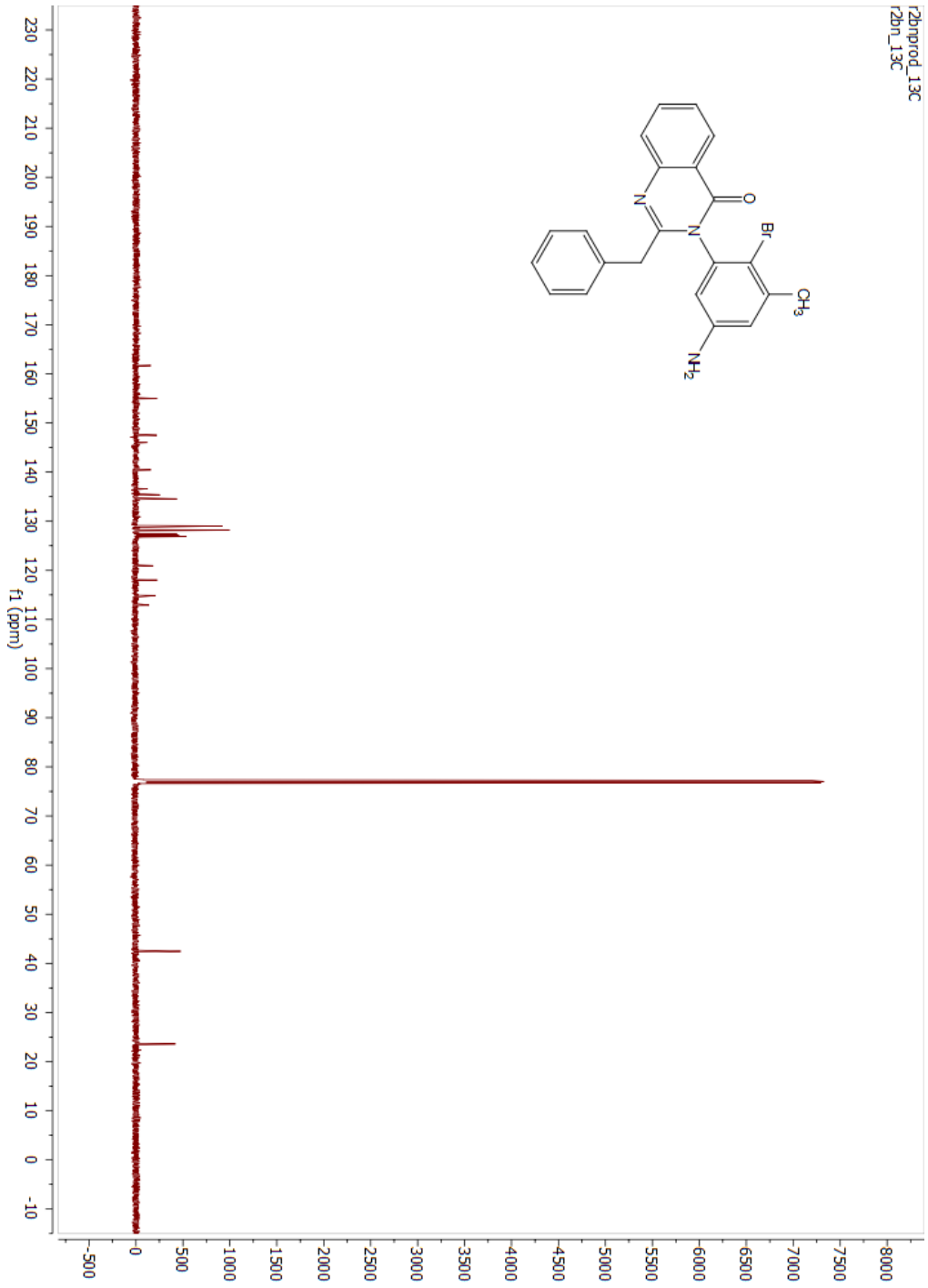


Figure AII. 68: 2DNOESY 9a

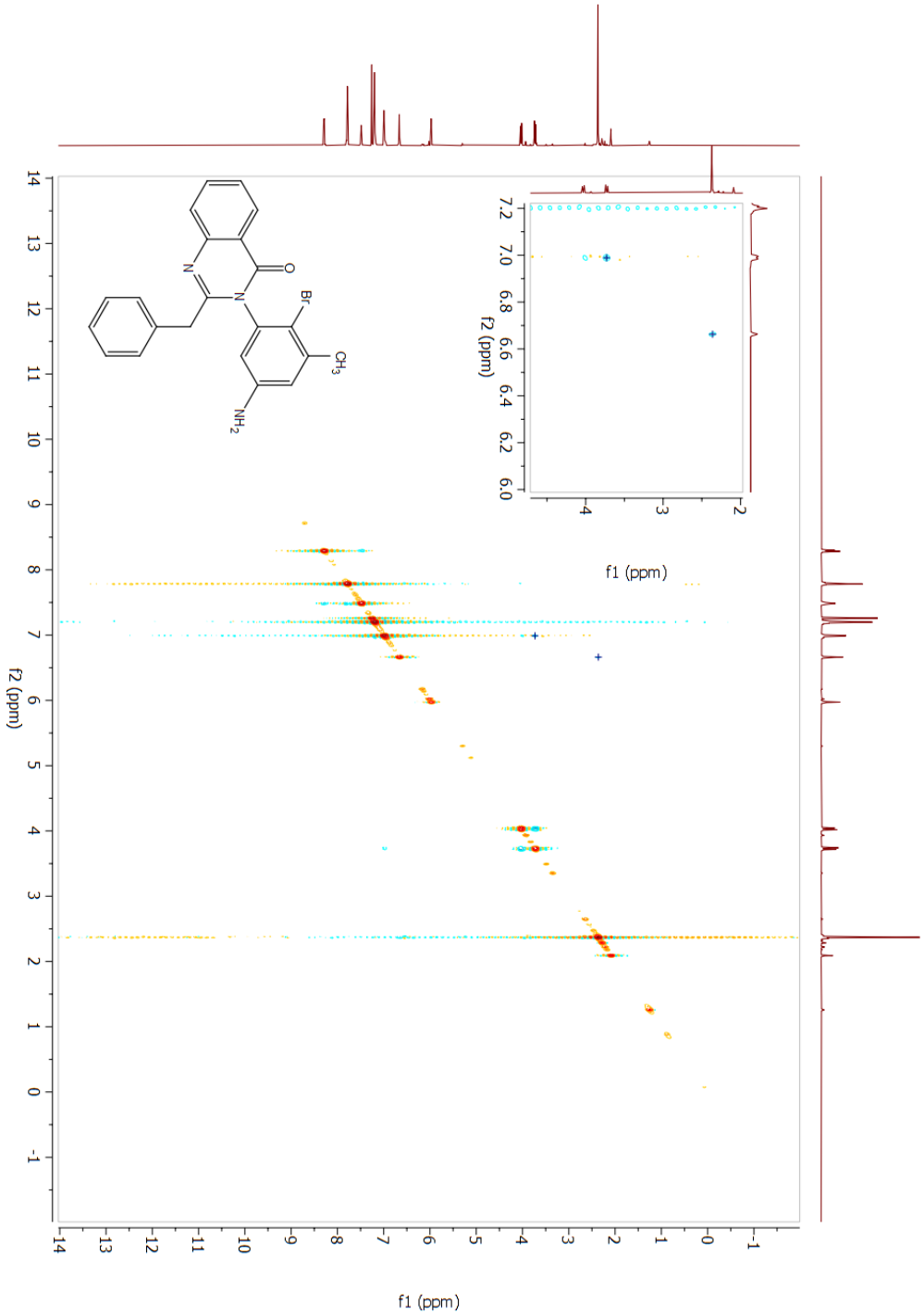


Figure AII. 69: 1HNMR 11a

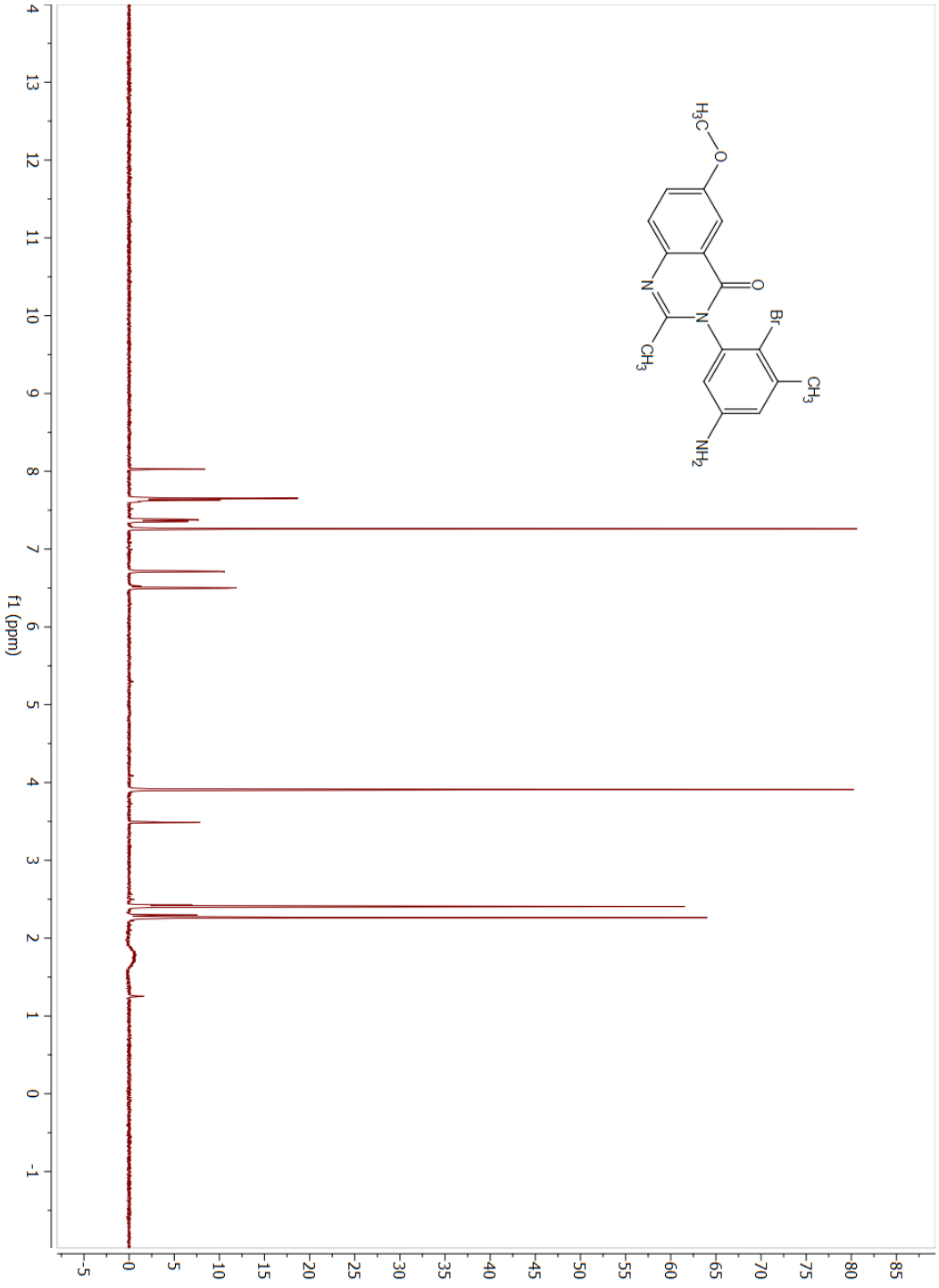


Figure AII. 70: 13CNMR 11a

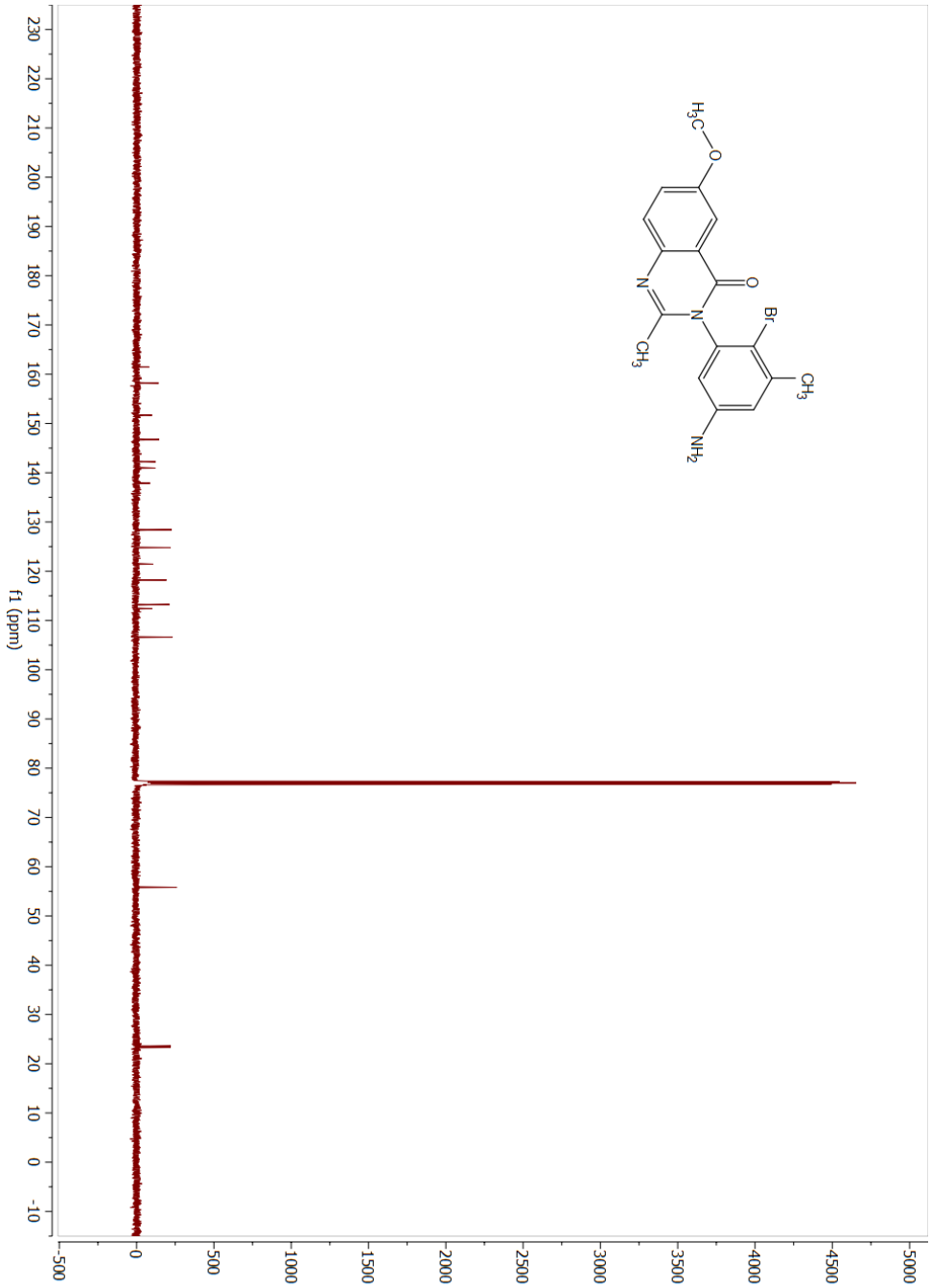


Figure AII. 71: 2DNOESY 11a

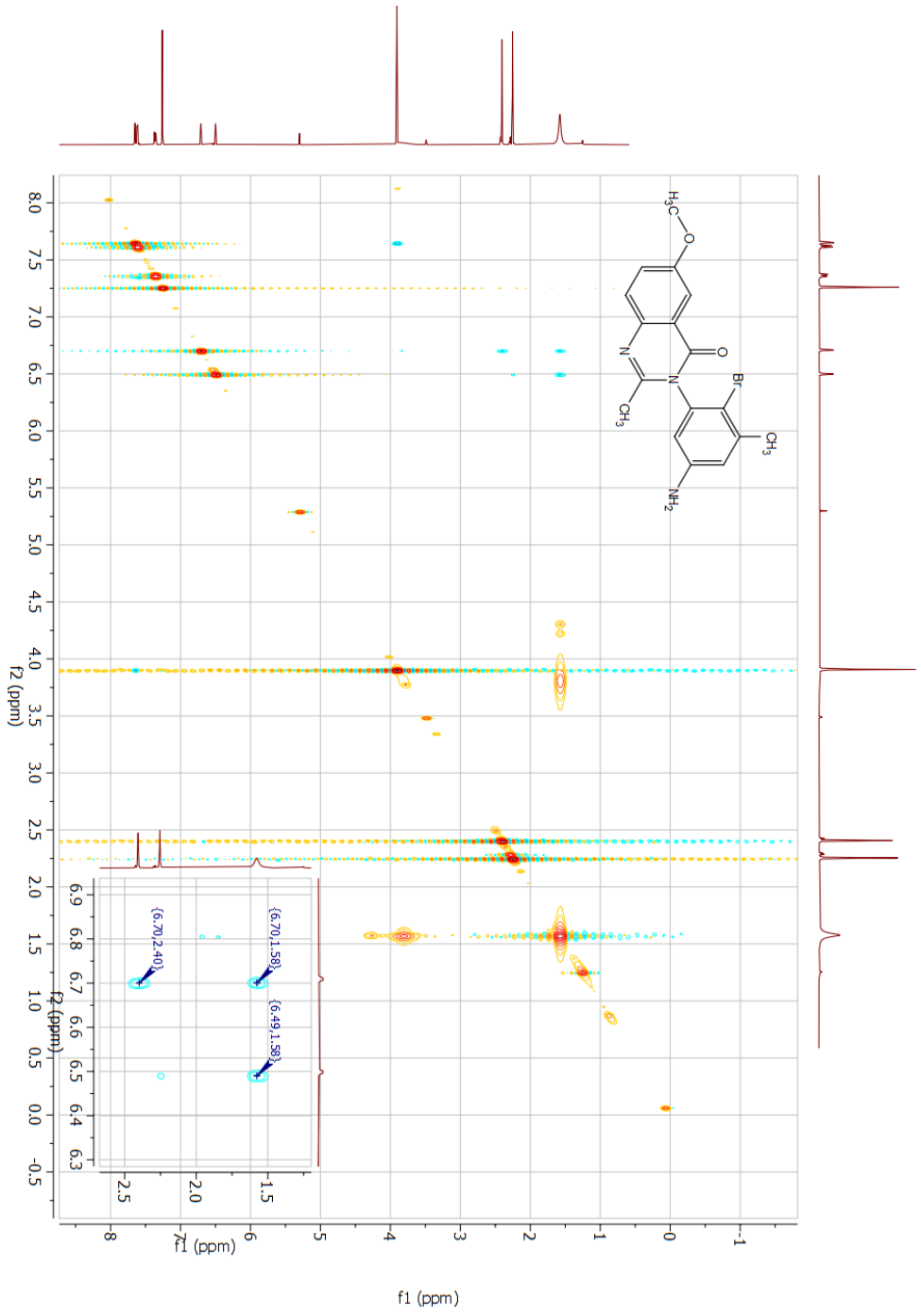


Figure AII. 72: 1HNMR 12a

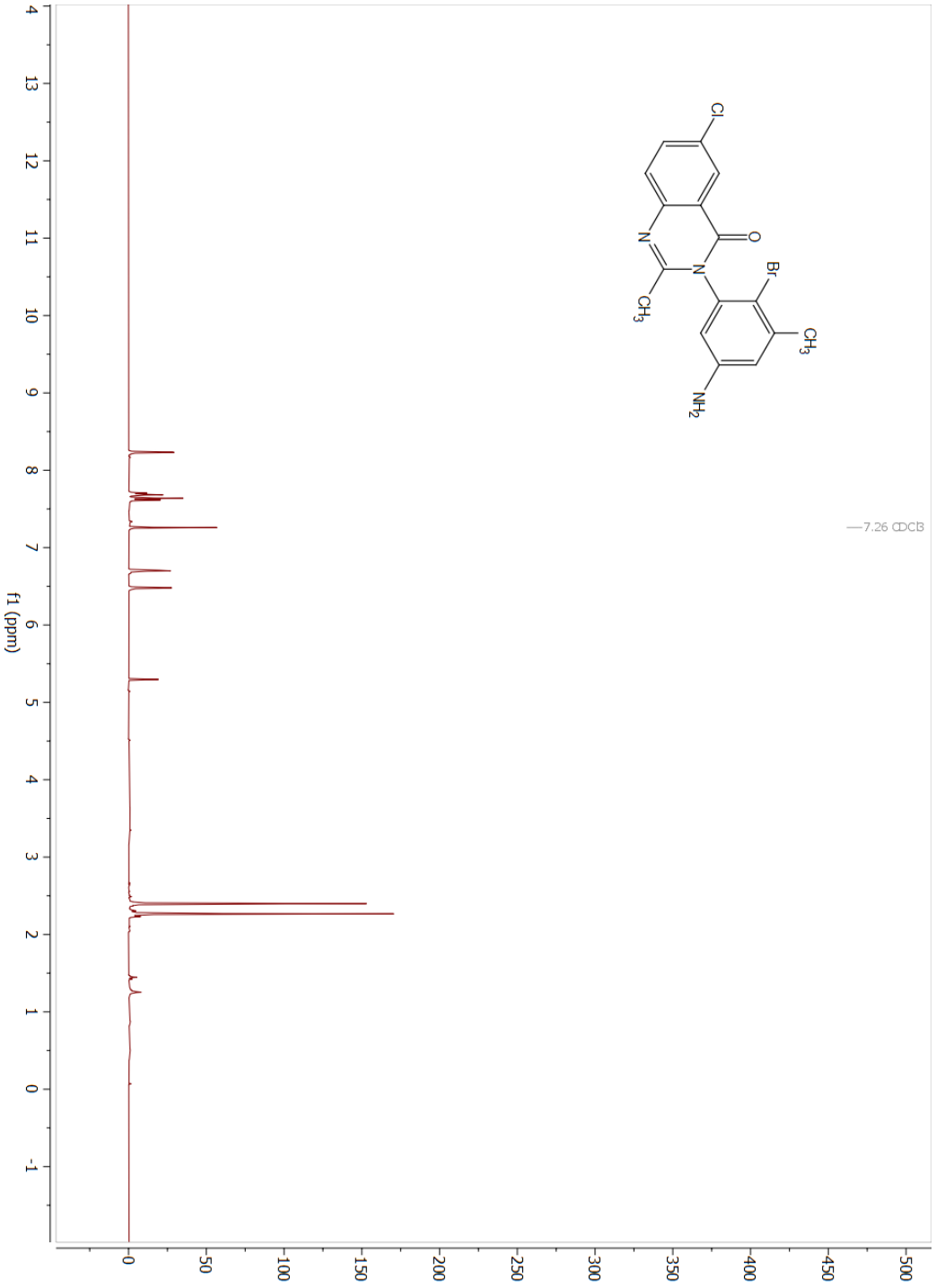


Figure AII. 73: 13CNMR 12a

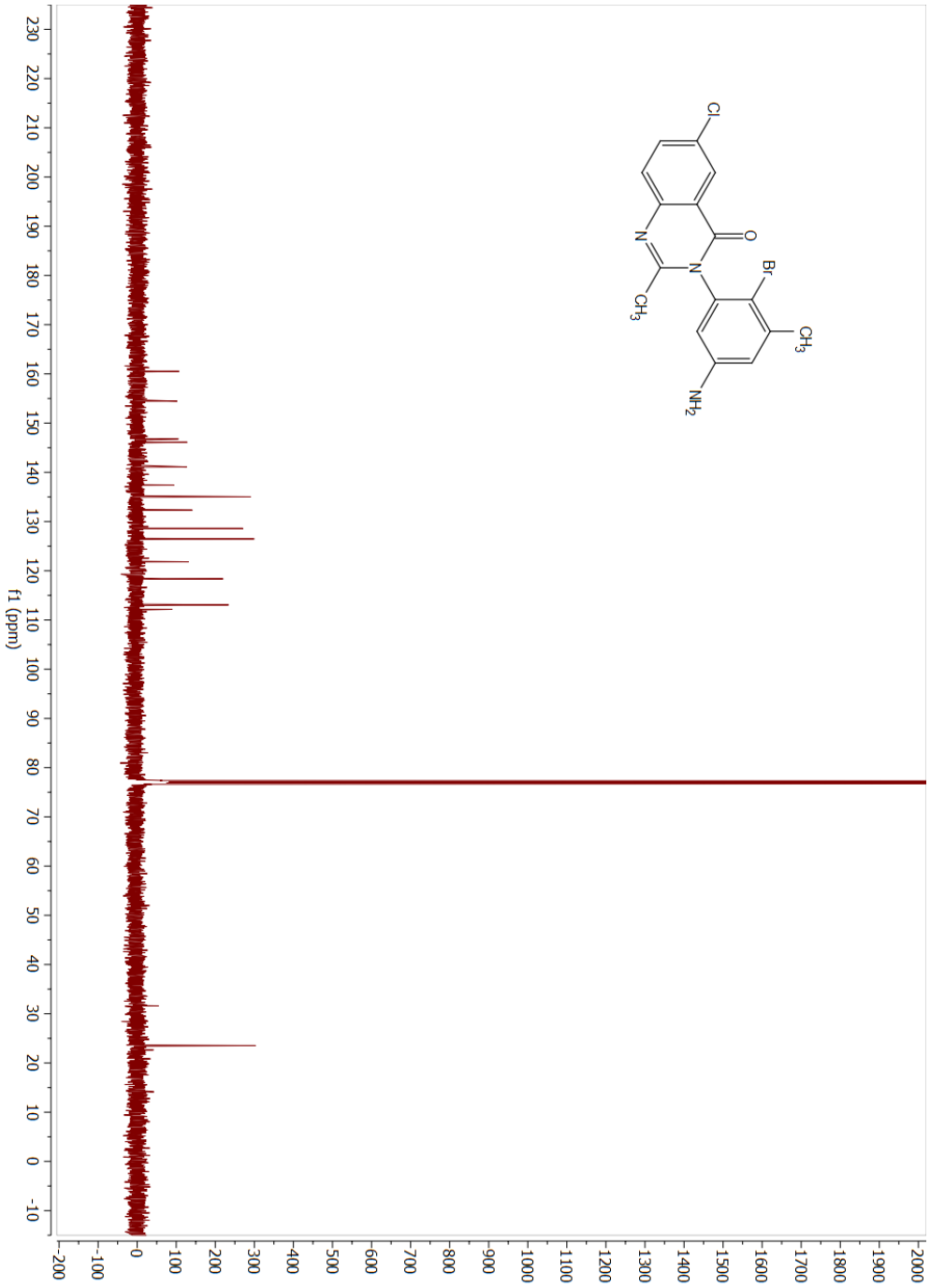


Figure AII. 74: 2DNOESY 12a

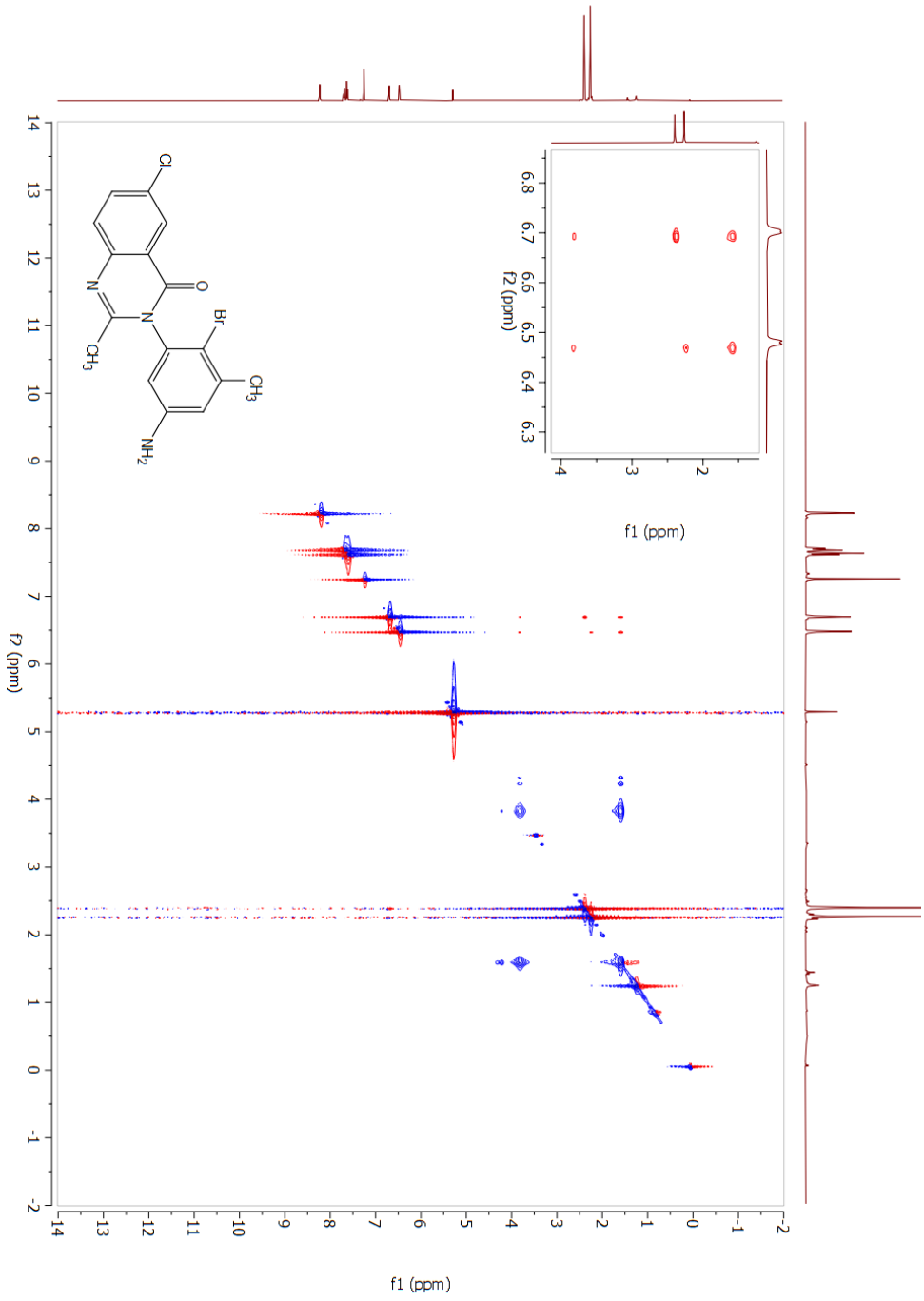


Figure AII. 75: 1HNMR 13a

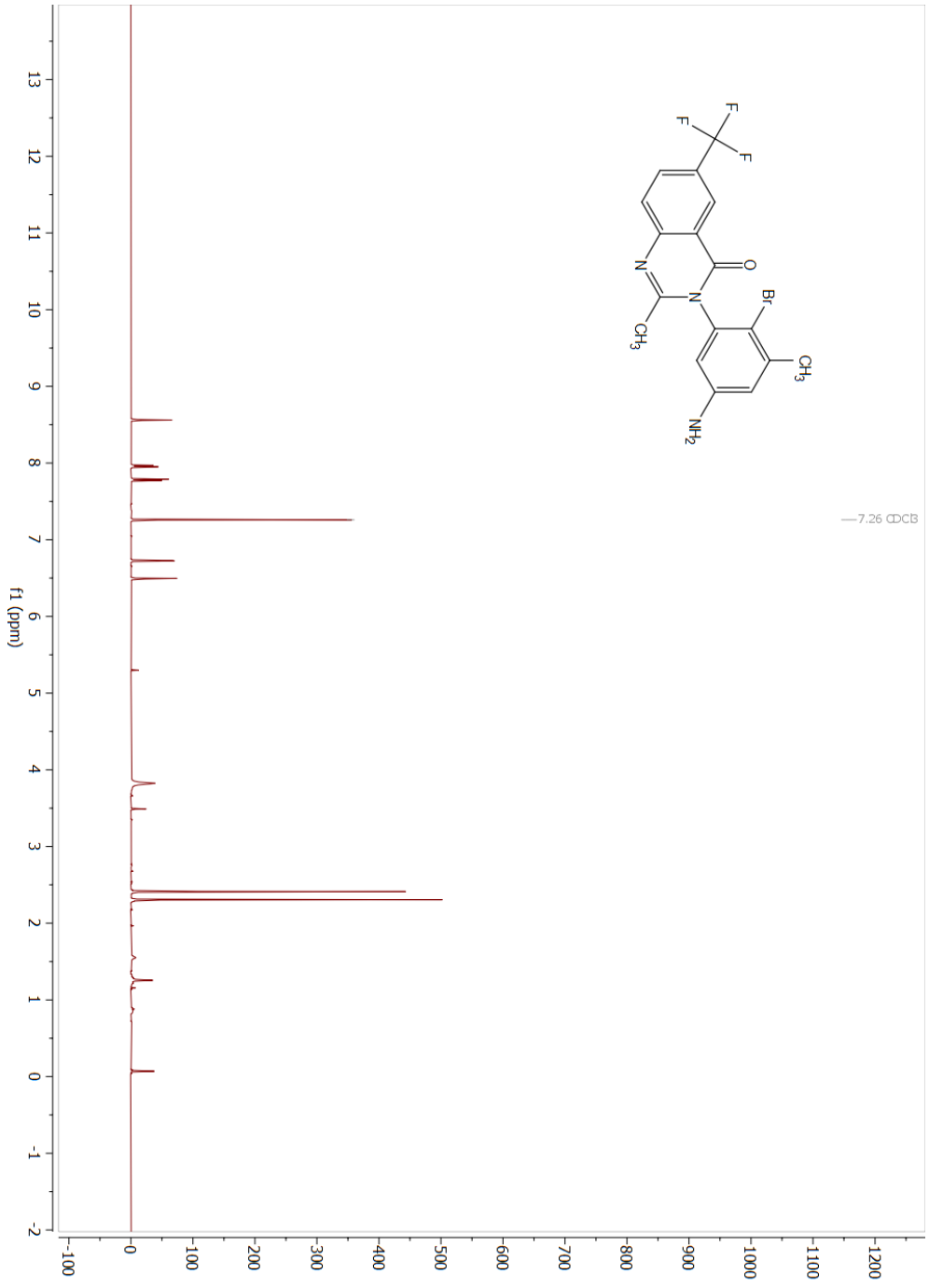


Figure AII. 76: 13CNMR 13a

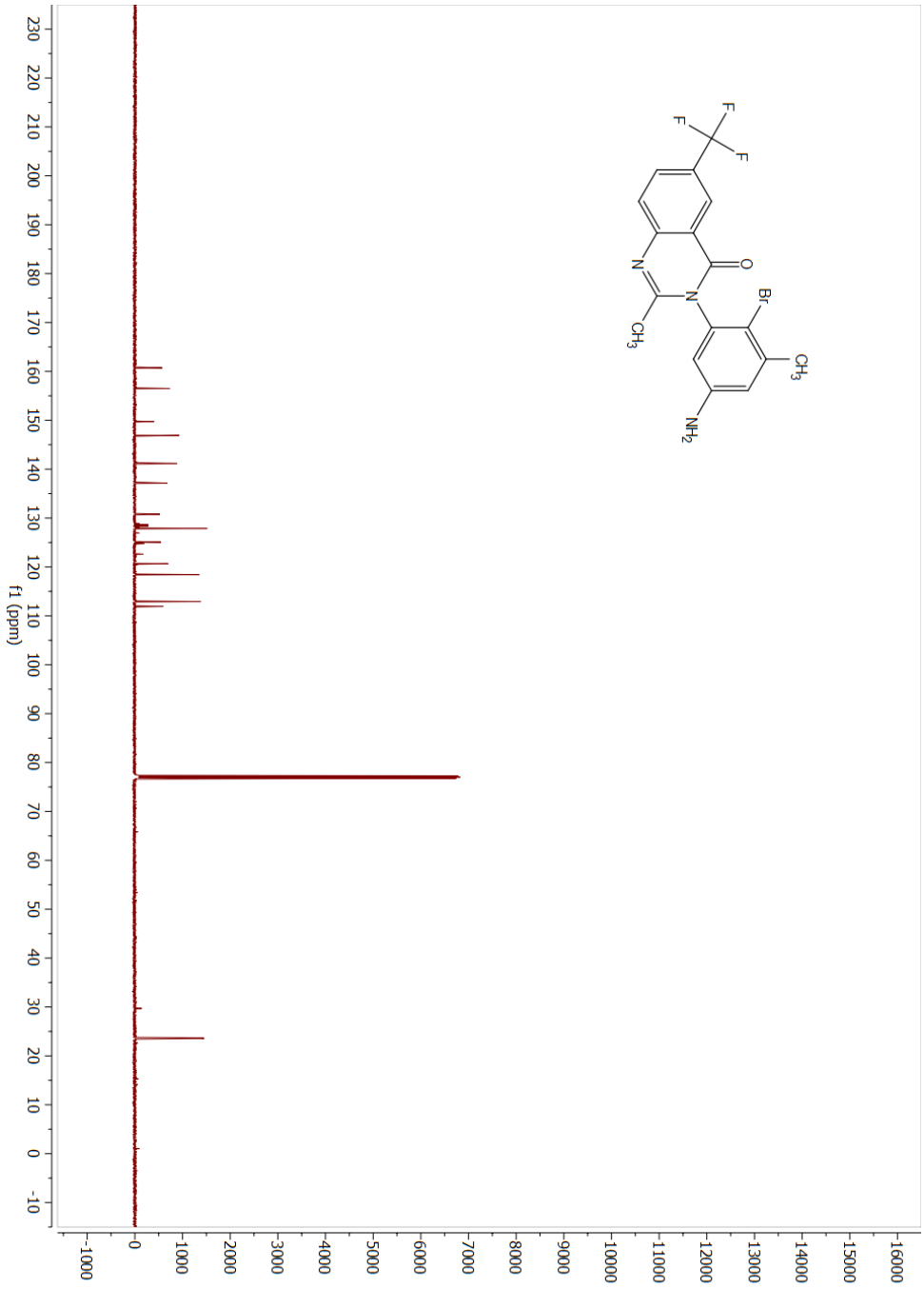


Figure AII. 77: 2DNOESY 13a

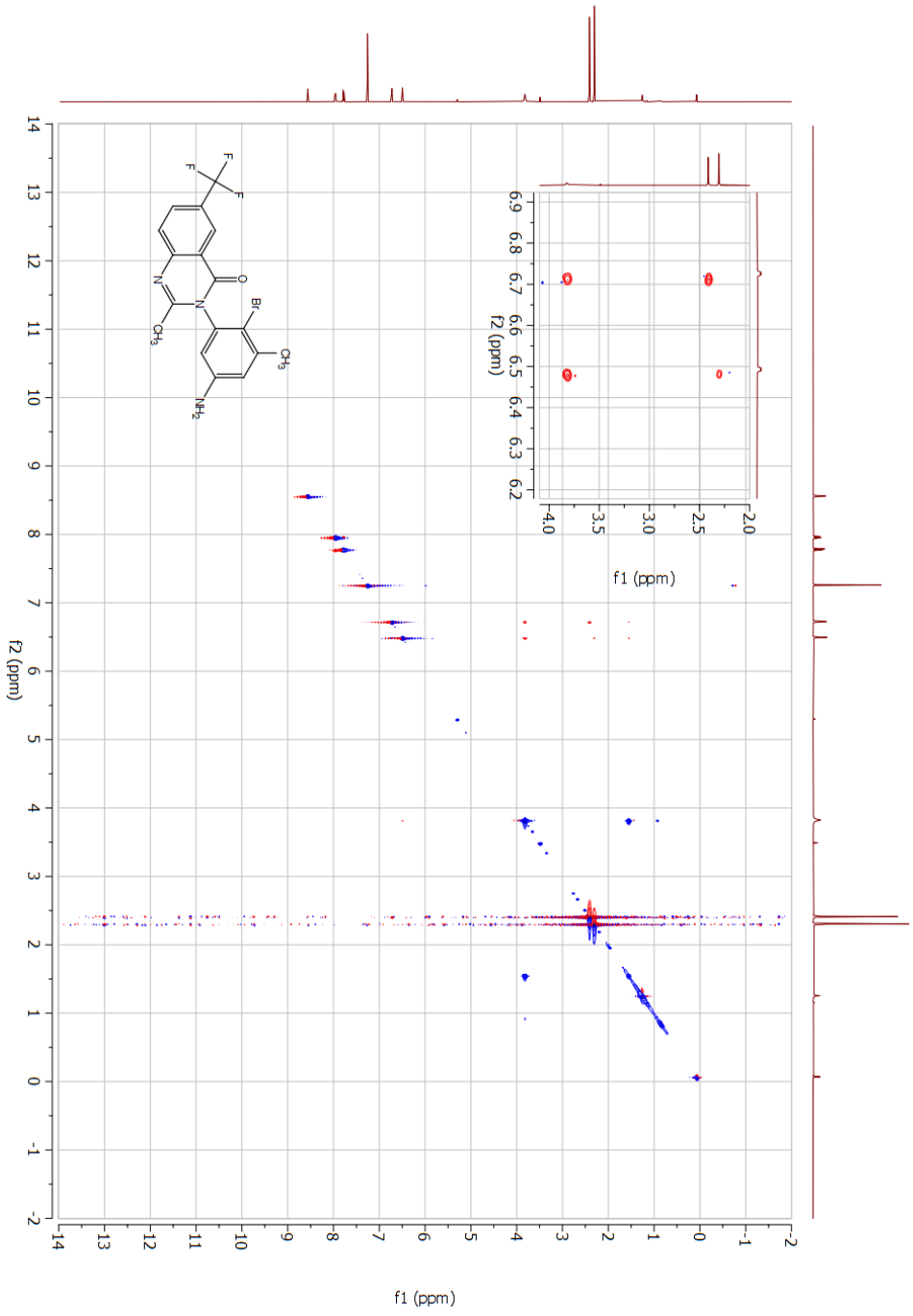


Figure AII. 78: 1HNMR 14a

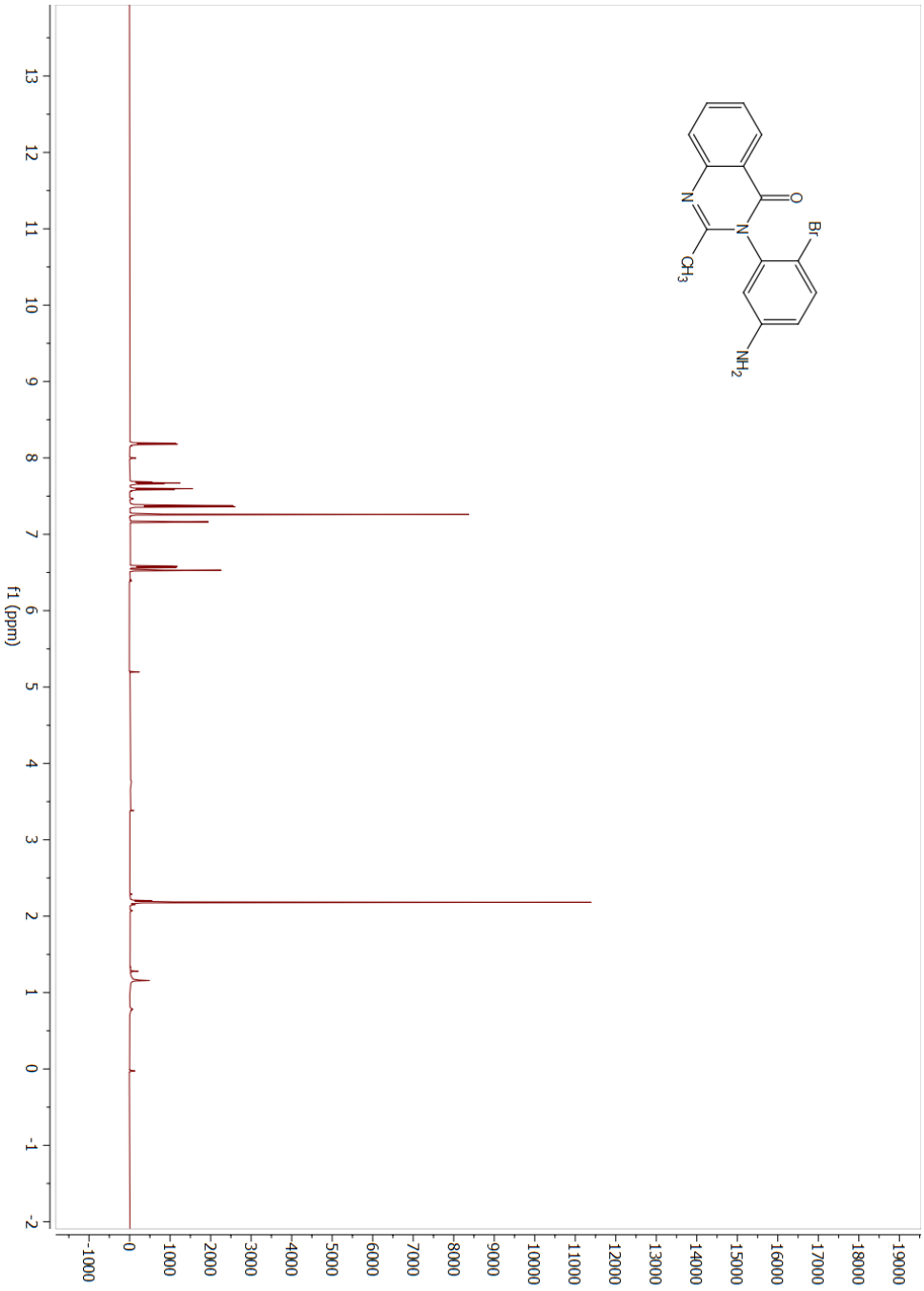


Figure AII. 79: 13CNMR 14a

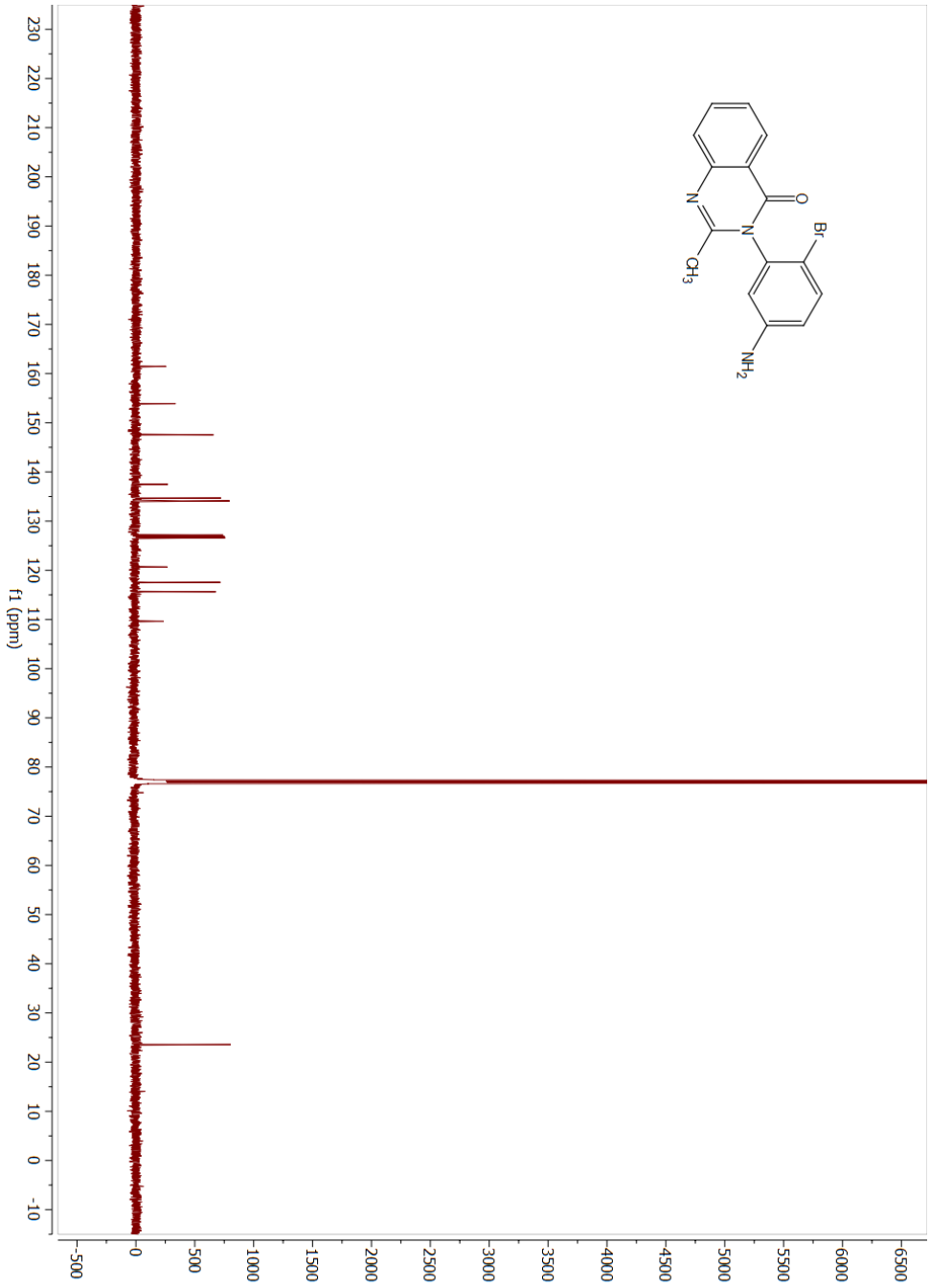


Figure AII. 80: 2DNOESY 14a

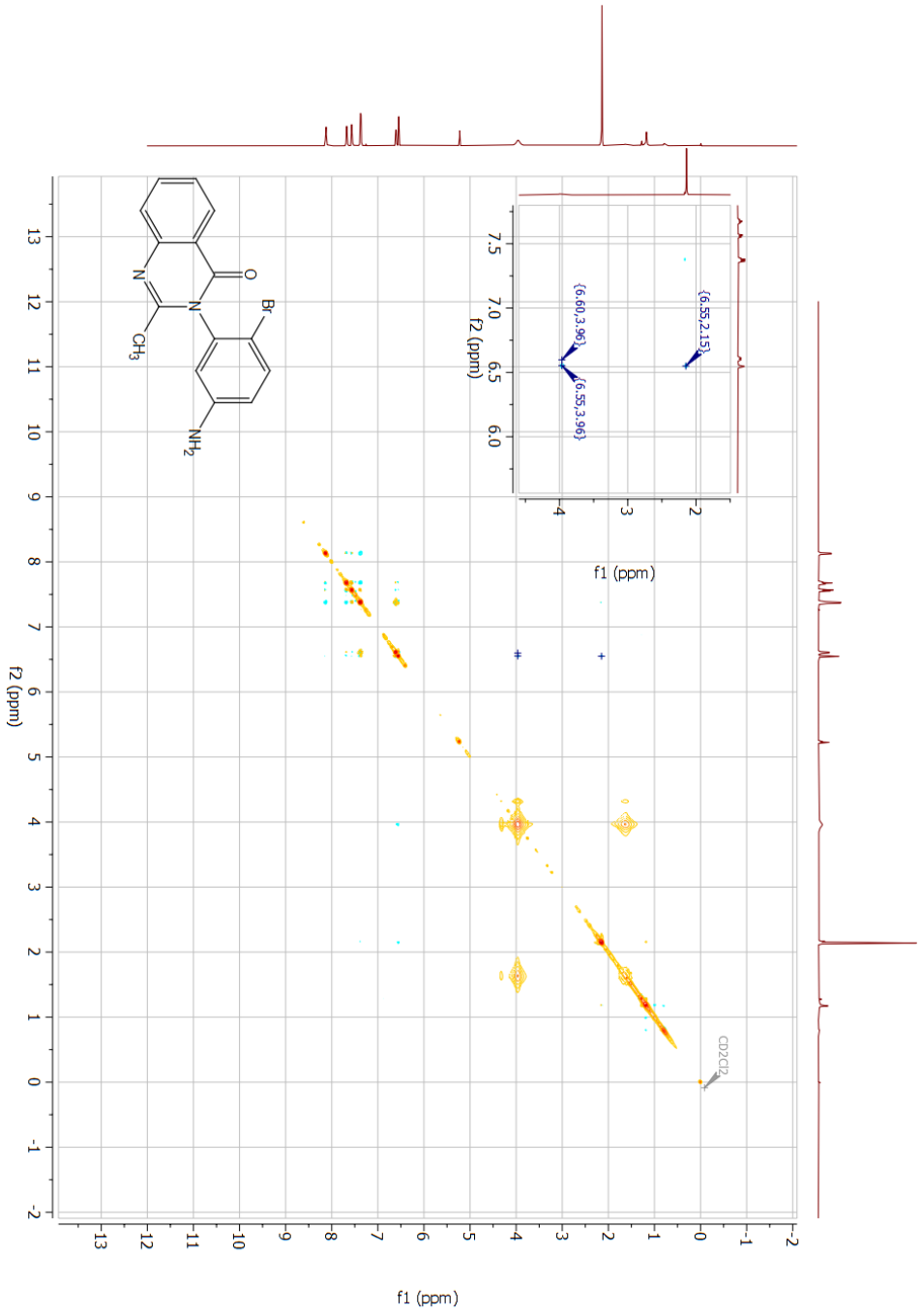


Figure AII. 81: 1HNMR 15a

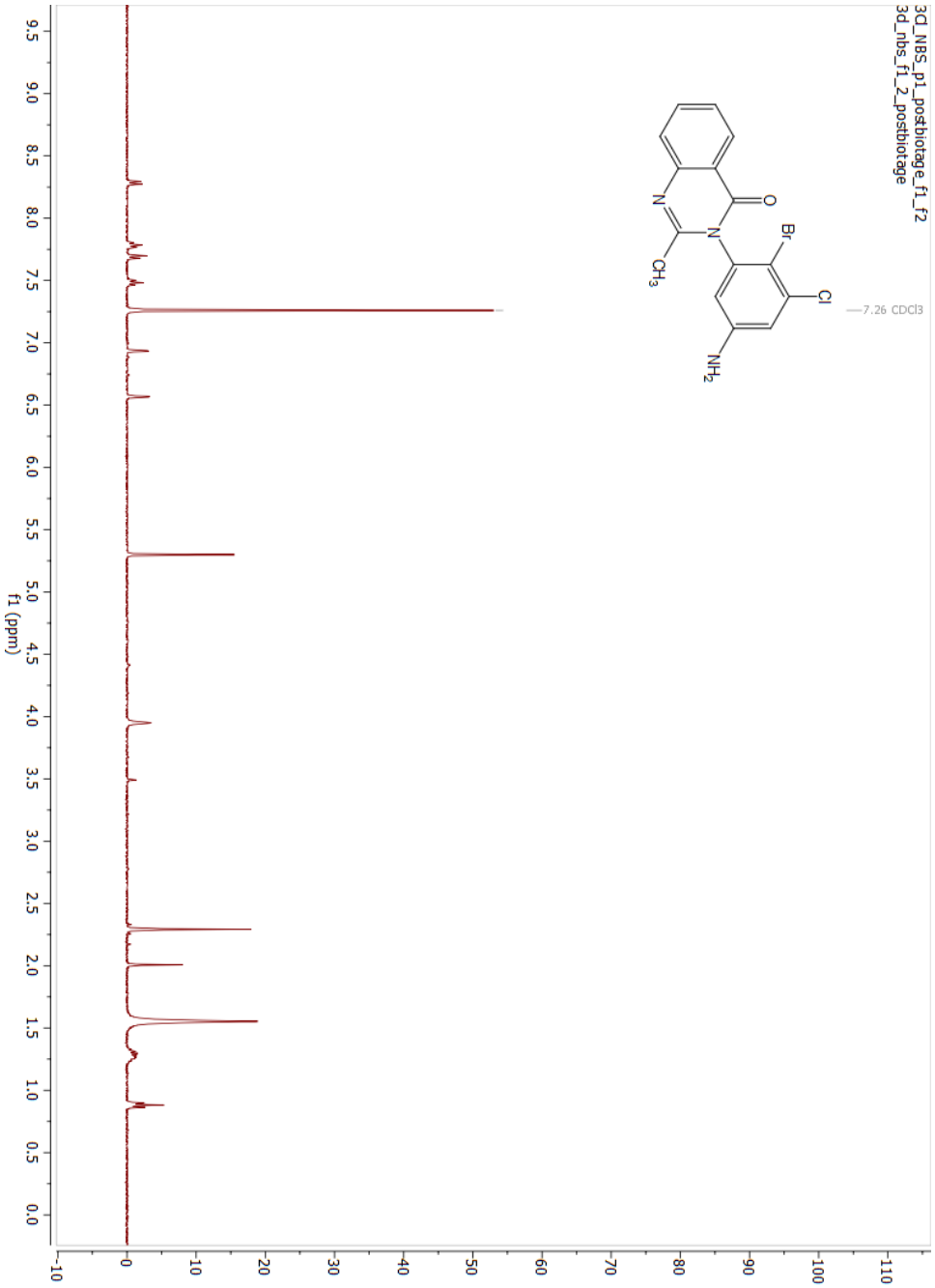


Figure AII. 82: 13CNMR 15a

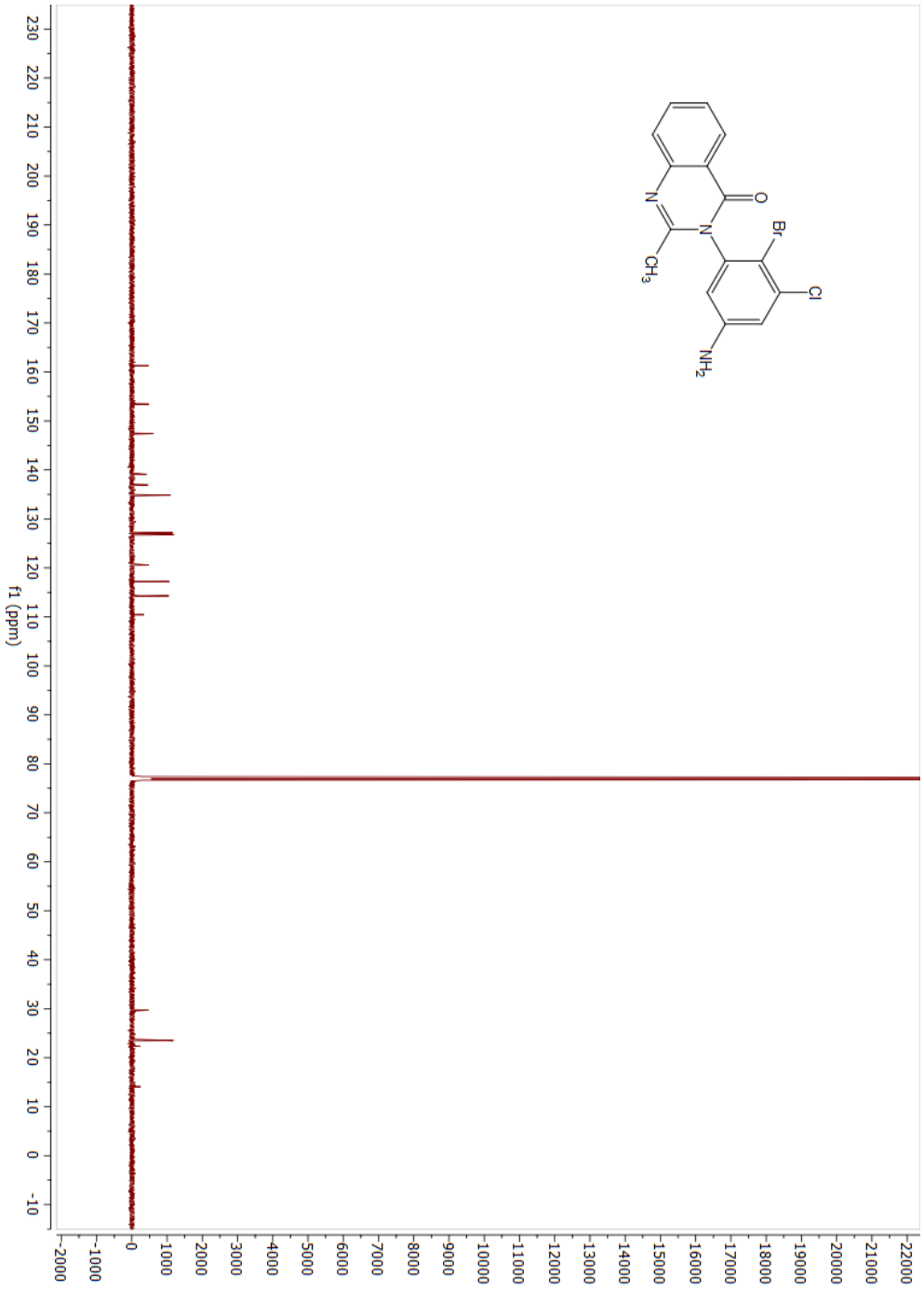


Figure AII. 83: 2DNOESY 15a

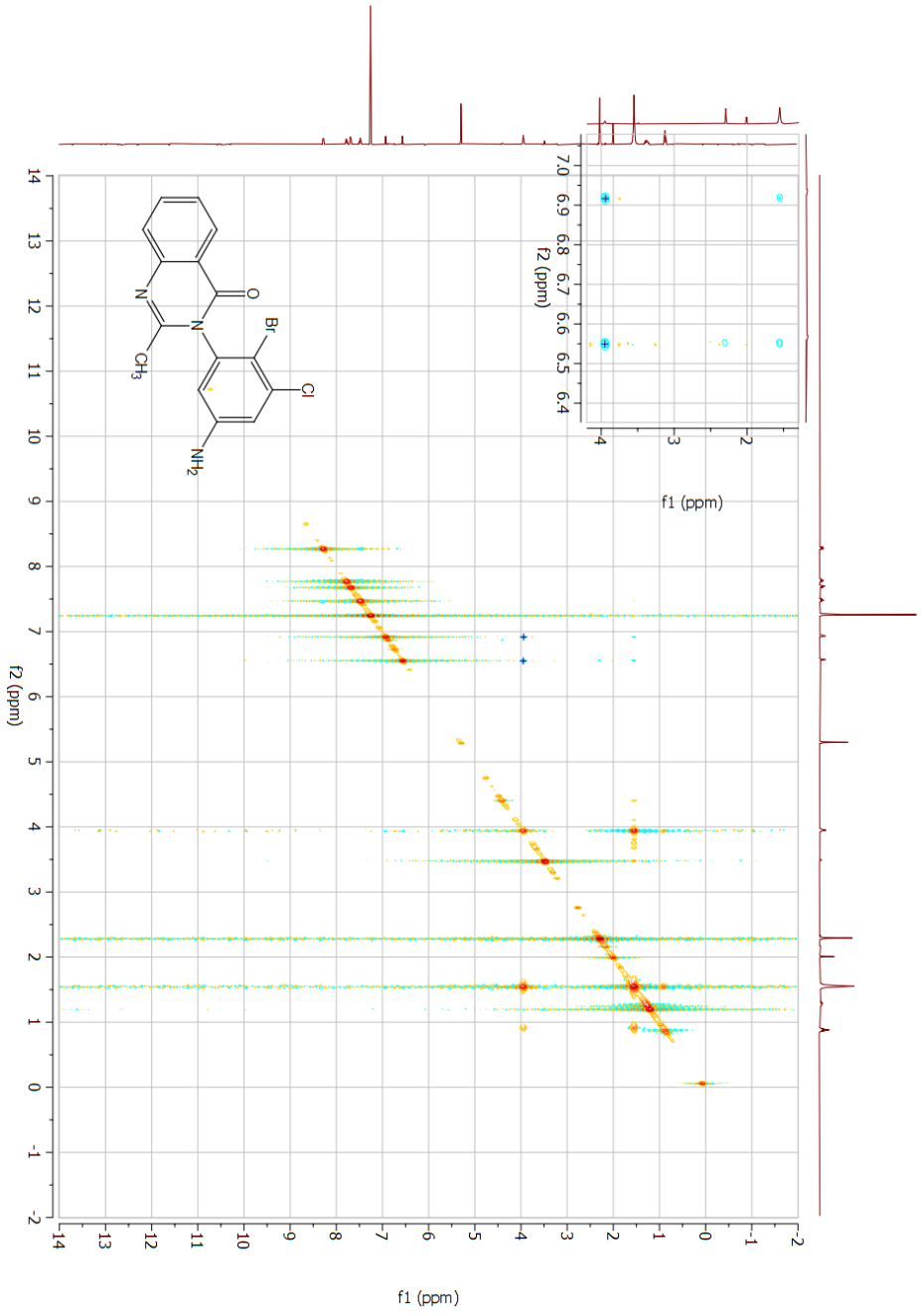


Figure AII. 84: 1HNMR 1c

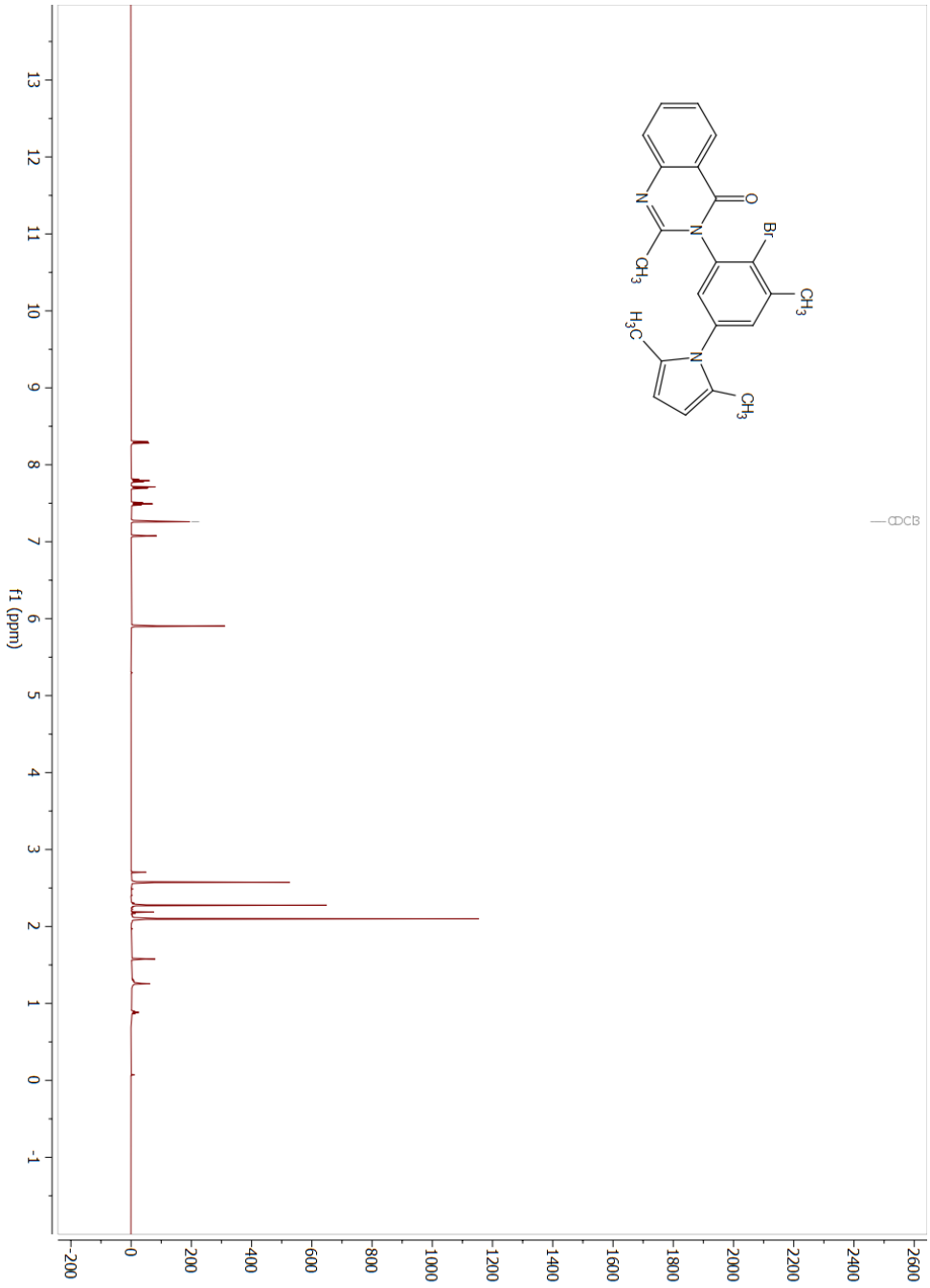


Figure AII. 85: 13CNMR 1c

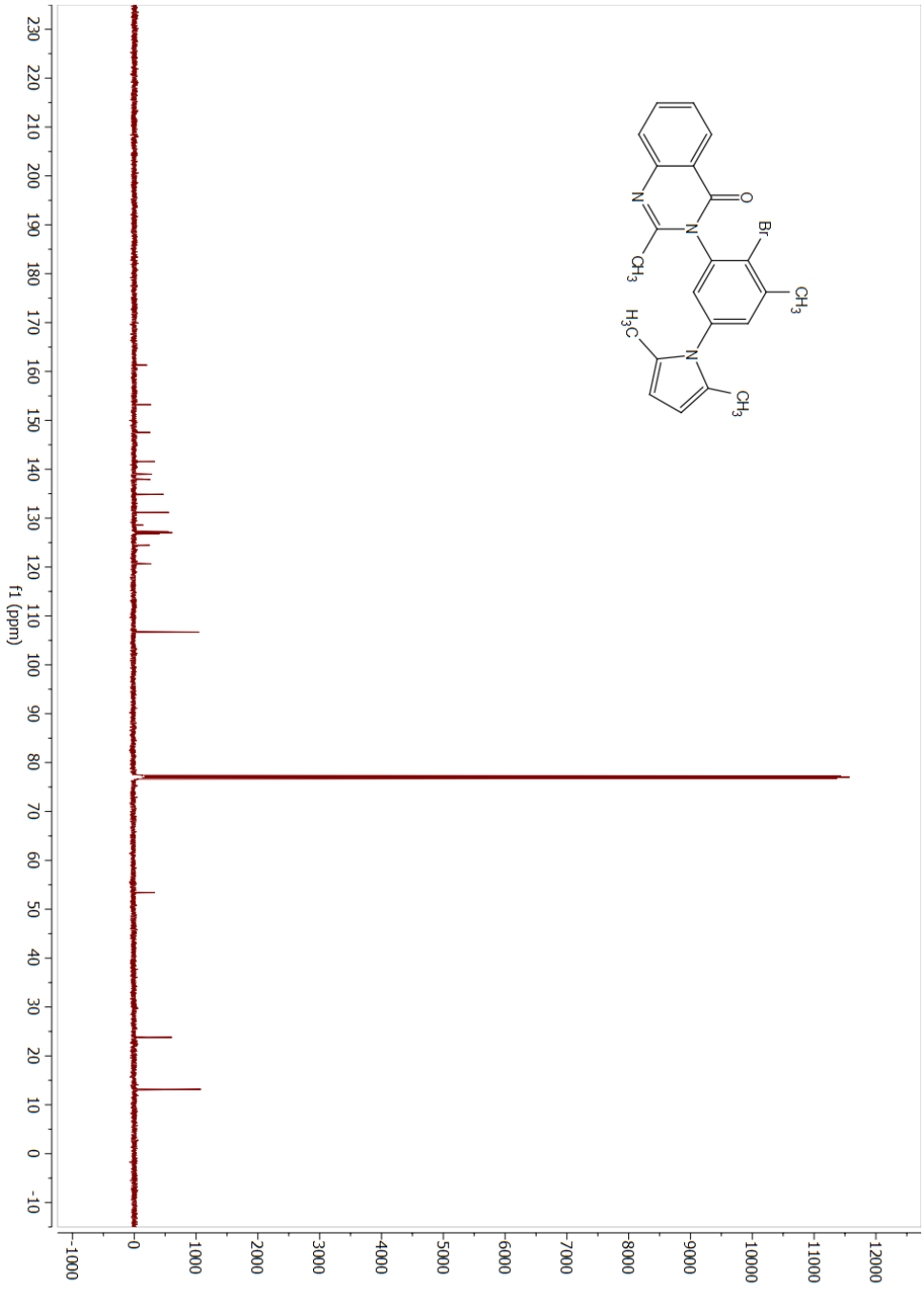


Figure AII. 86: 1HNMR 1d

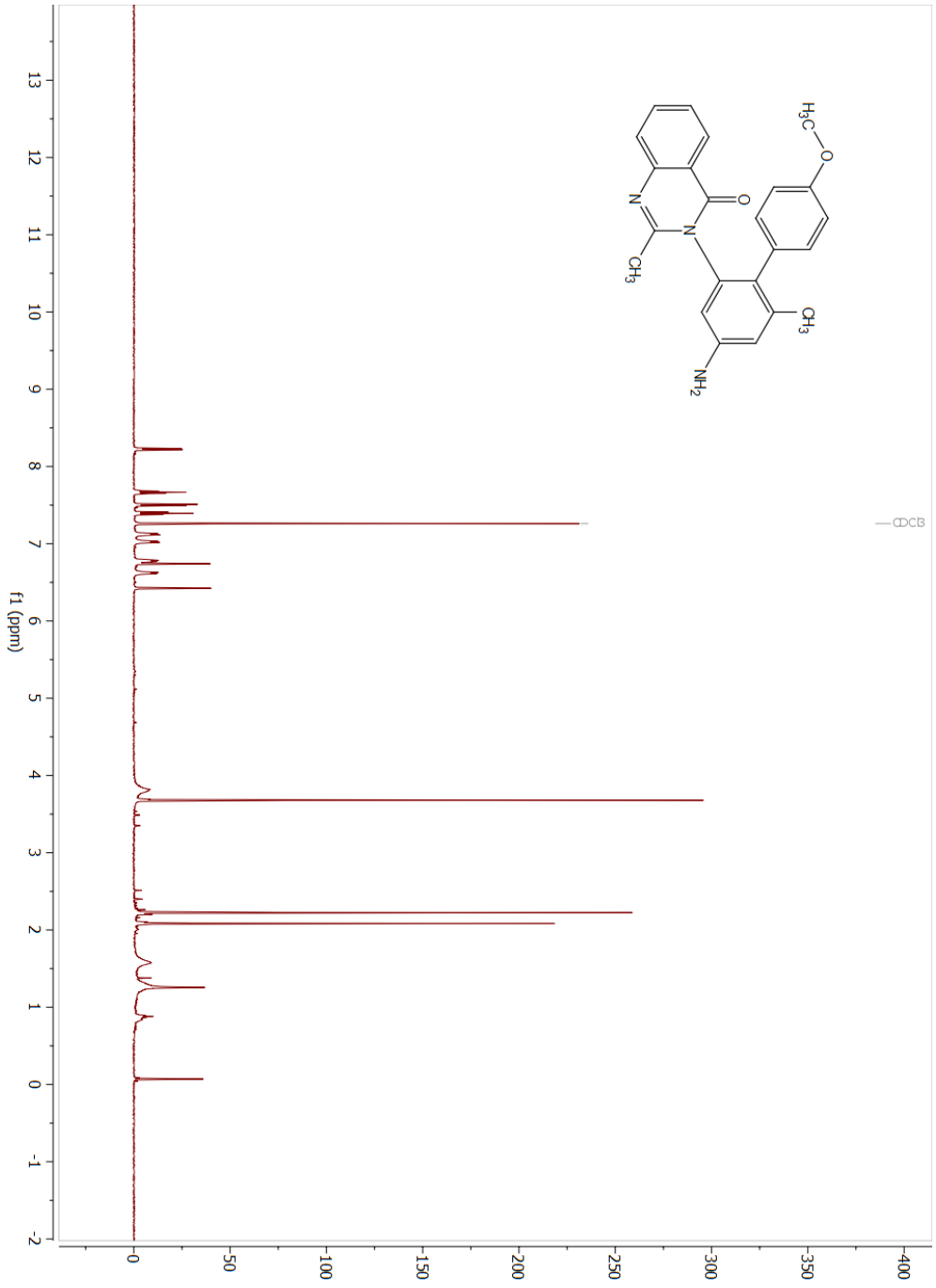
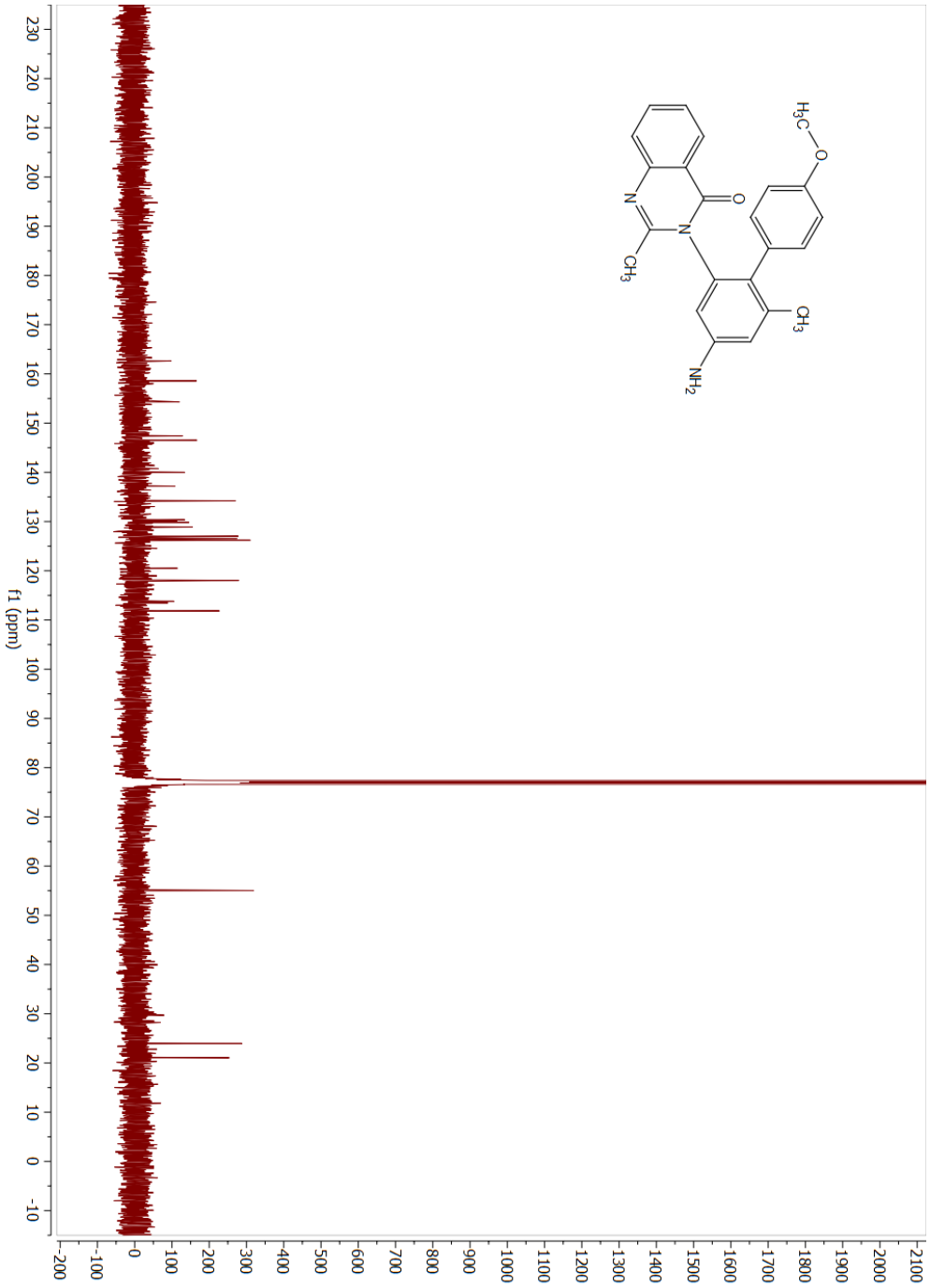


Figure AII. 87: 13CNMR 1d



APPENDIX III

NMR Spectra for Compounds from Chapter 4

Figure AIII. 1: 1HNMR of substrate 1a

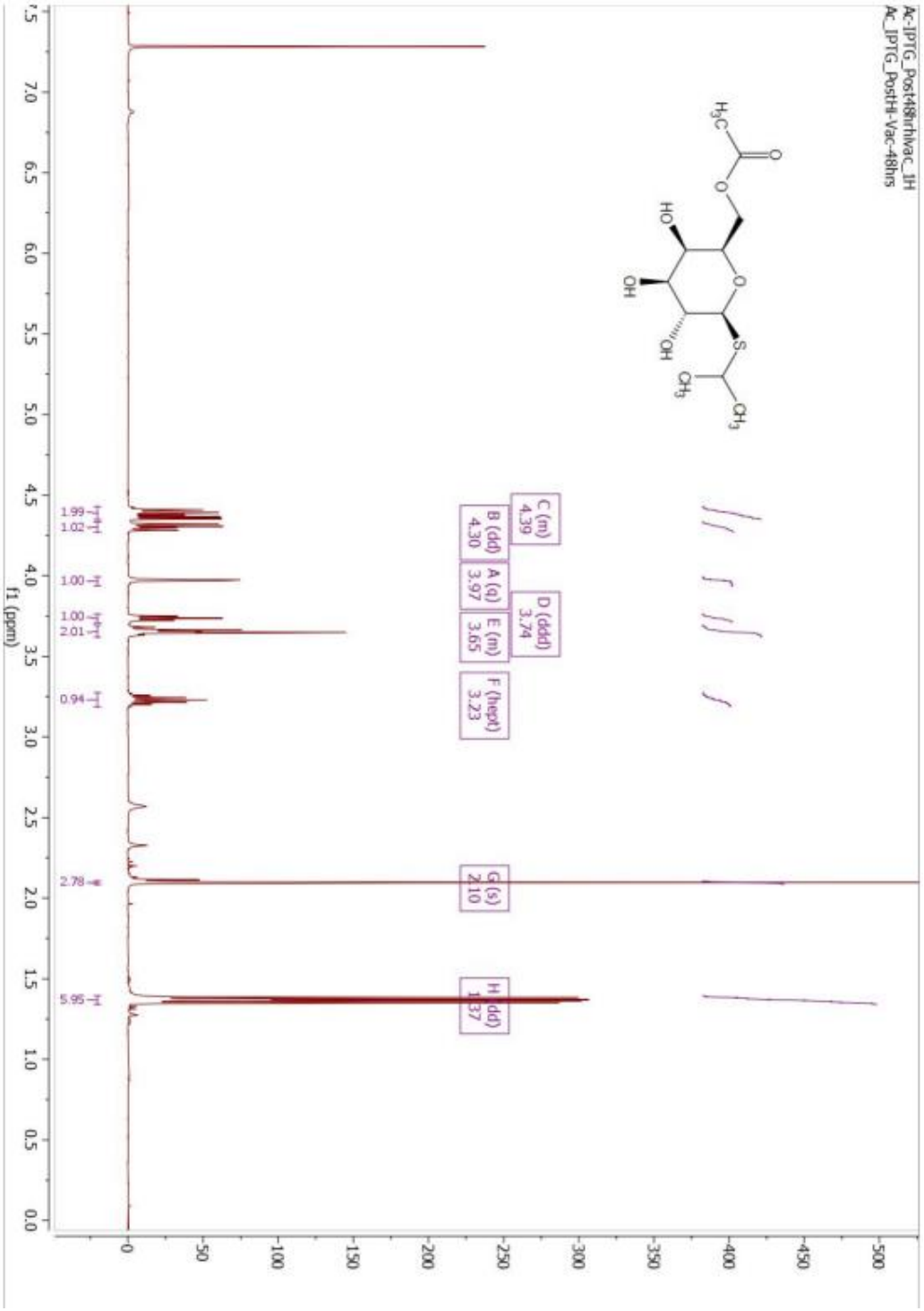


Figure AIII. 2: 13CNMR of substrate 1a

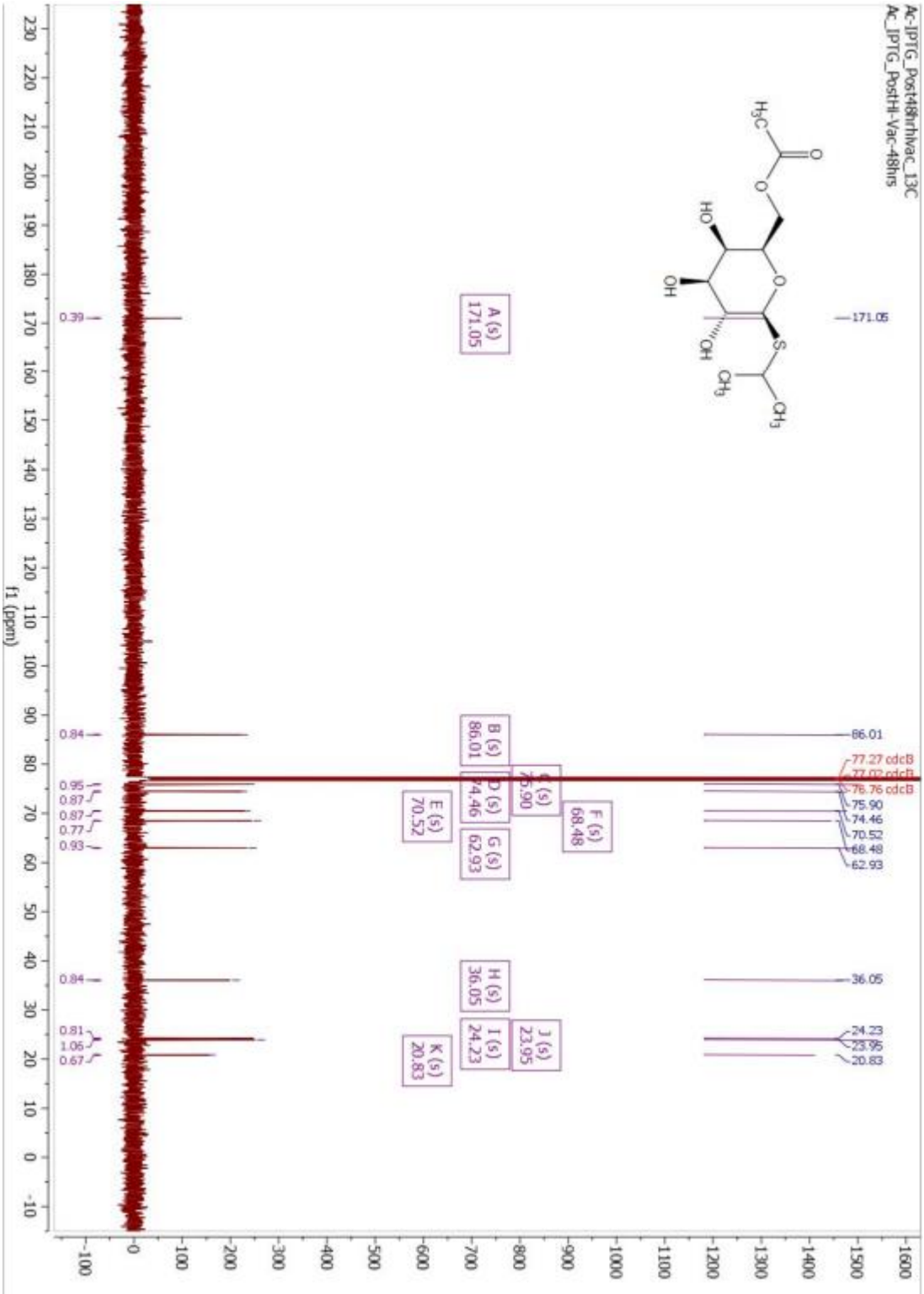


Figure AIII. 3: 1HNMR of substrate 1b

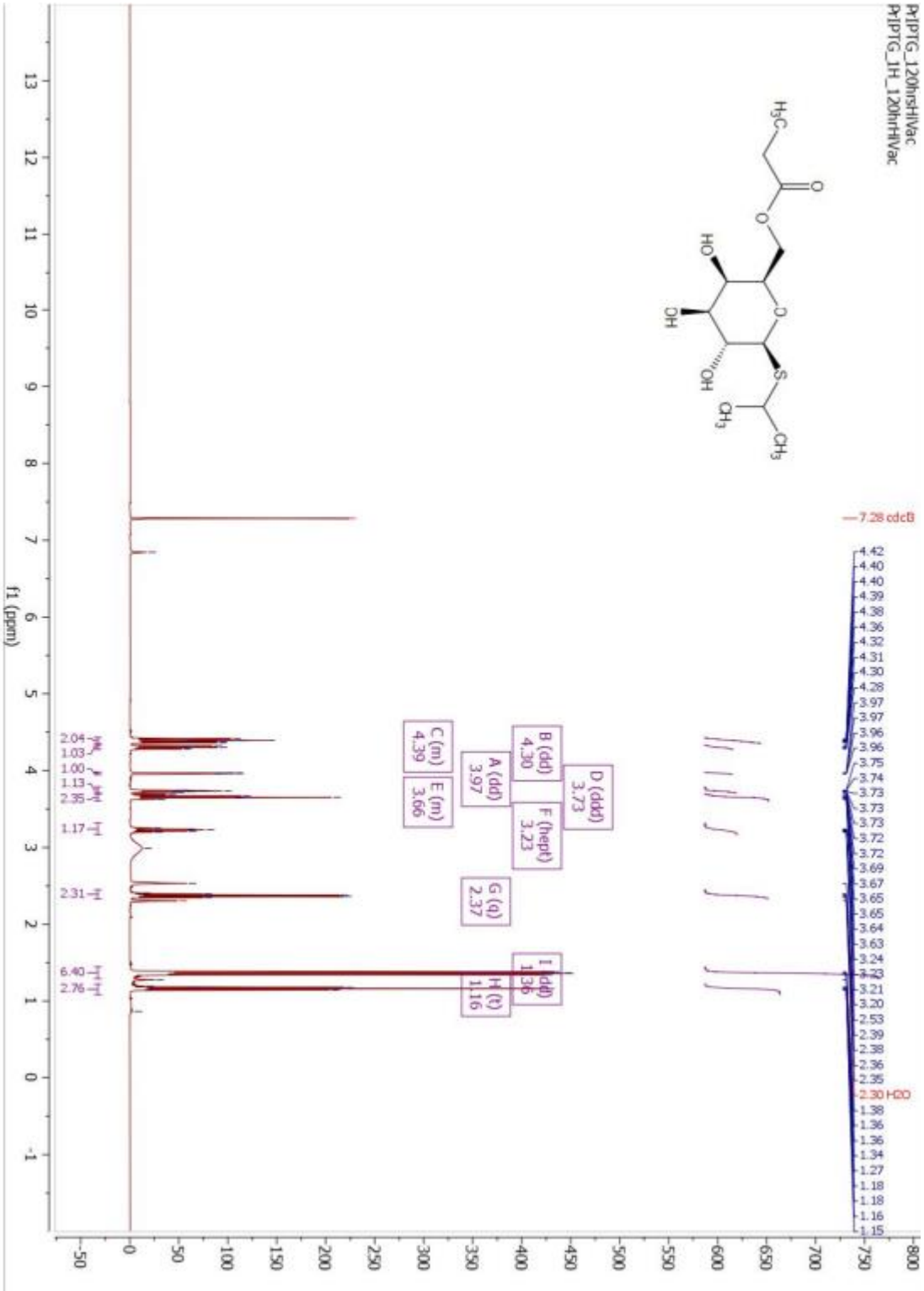


Figure AIII. 4: 13CNMR of substrate 1b

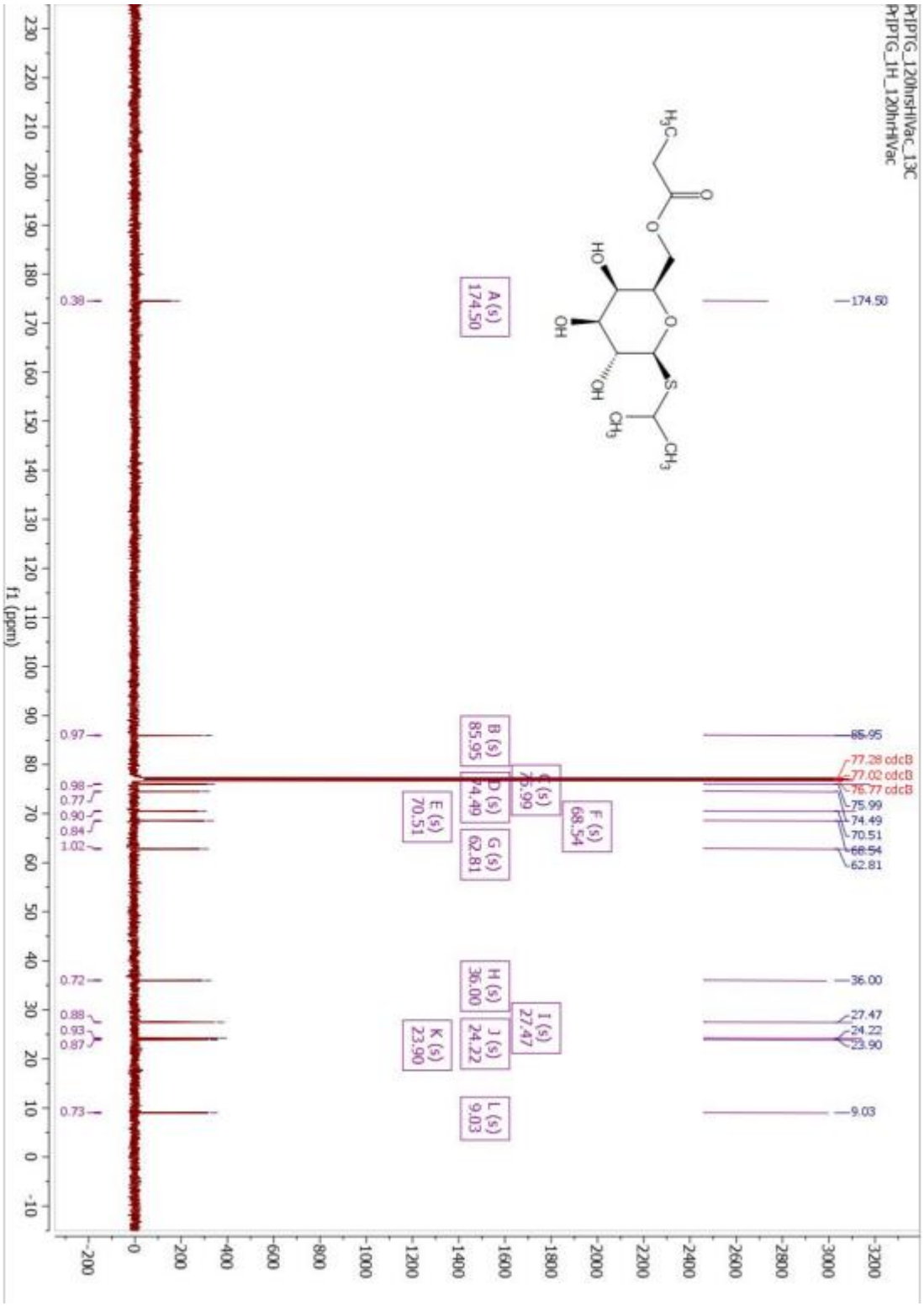


Figure AIII. 5: 1HNMR of substrate 1c

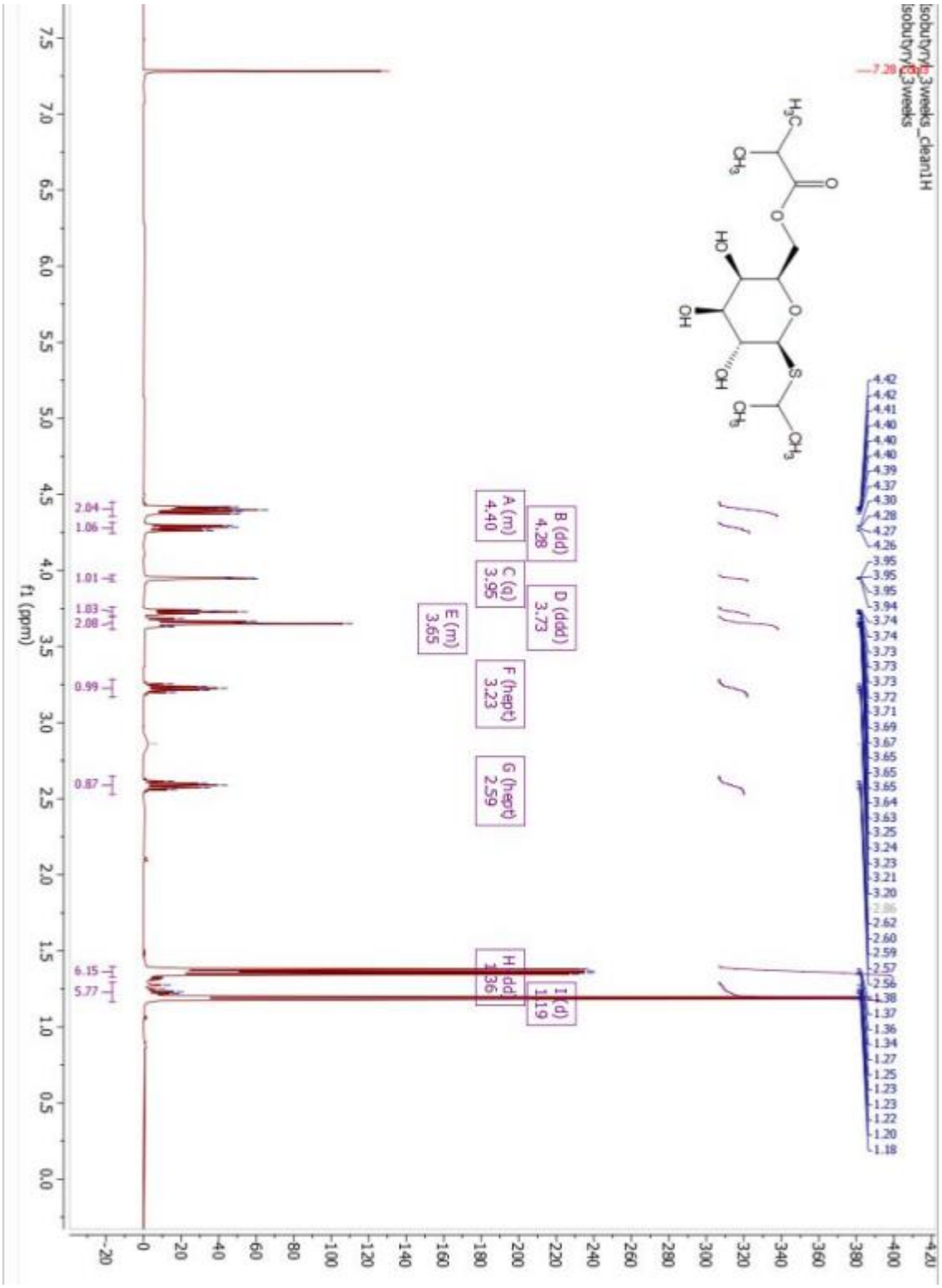


Figure AIII. 6: 13CNMR of substrate 1c

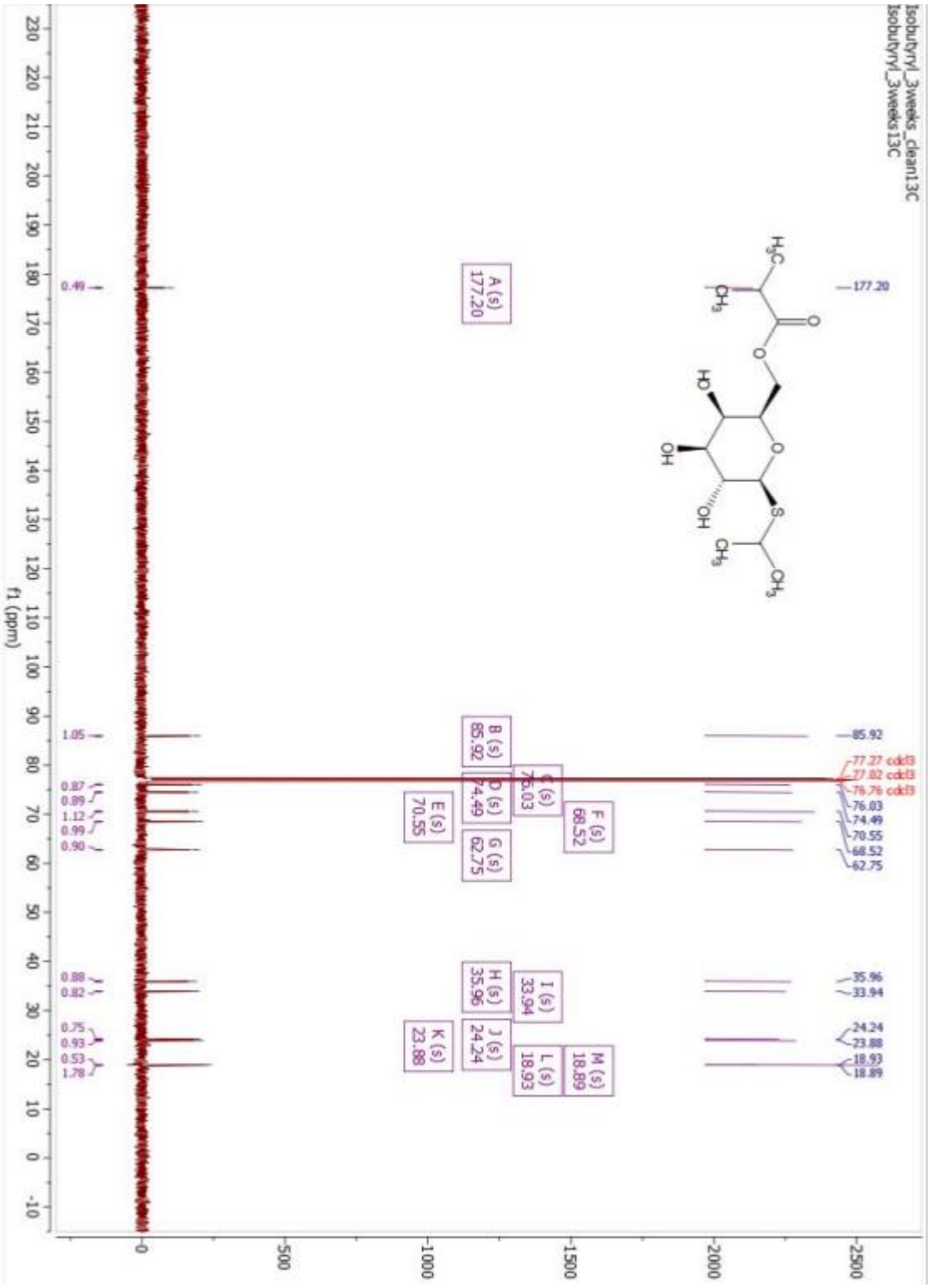


Figure AIII. 7: 1HNMR of substrate 1d

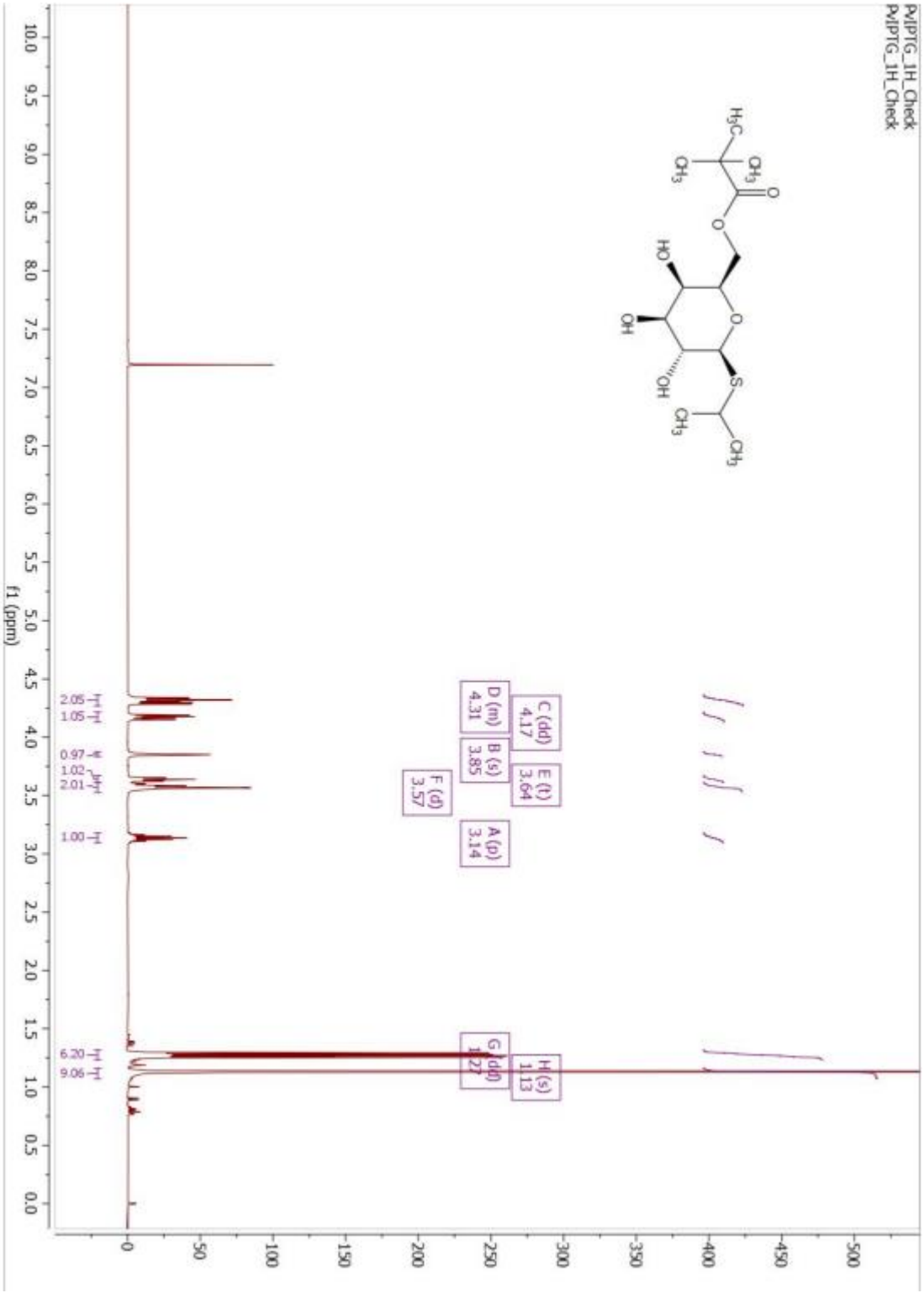


Figure AIII. 8: 13CNMR of substrate 1d

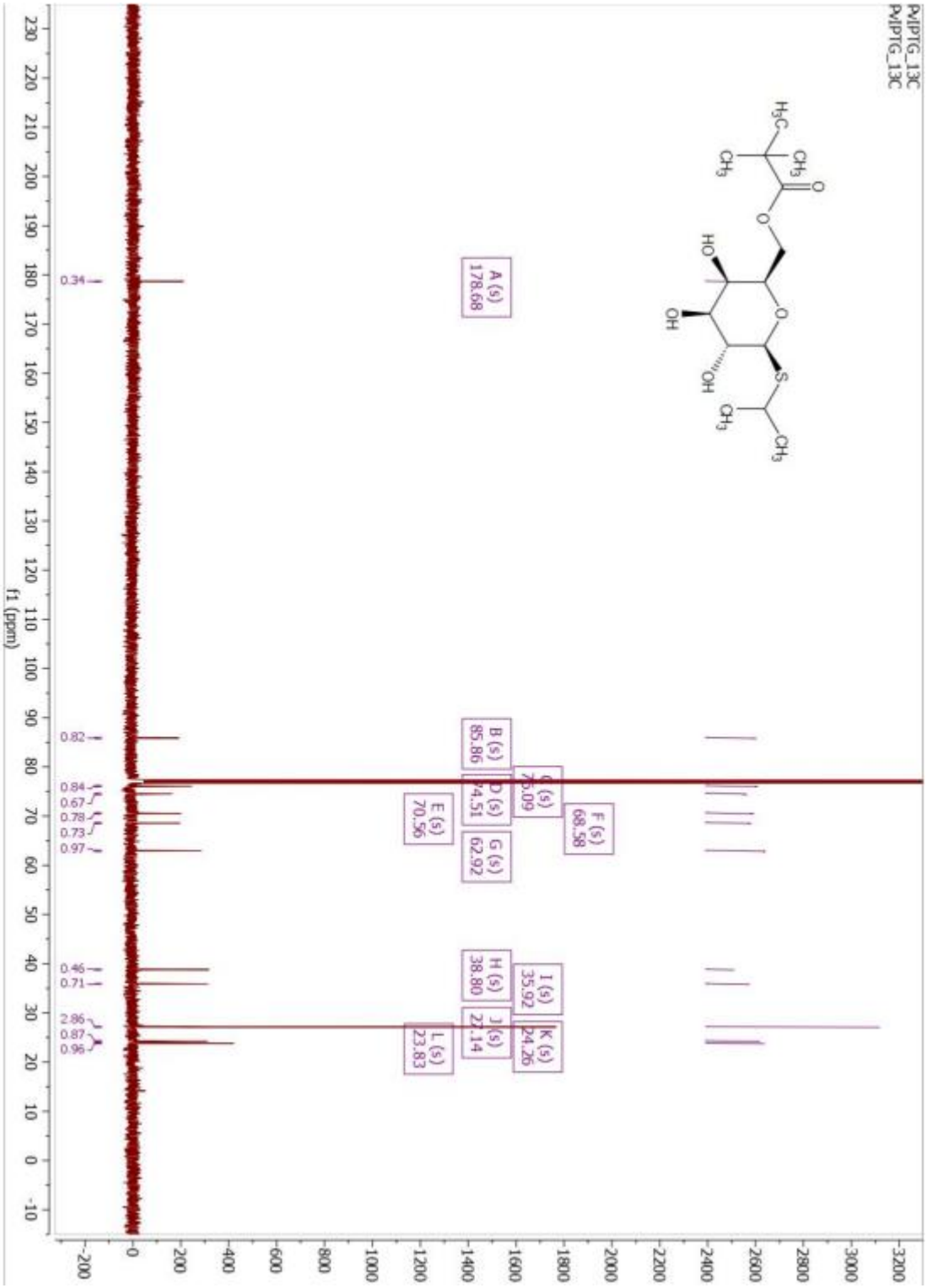


Figure AIII. 9: 1HNMR of substrate 2a

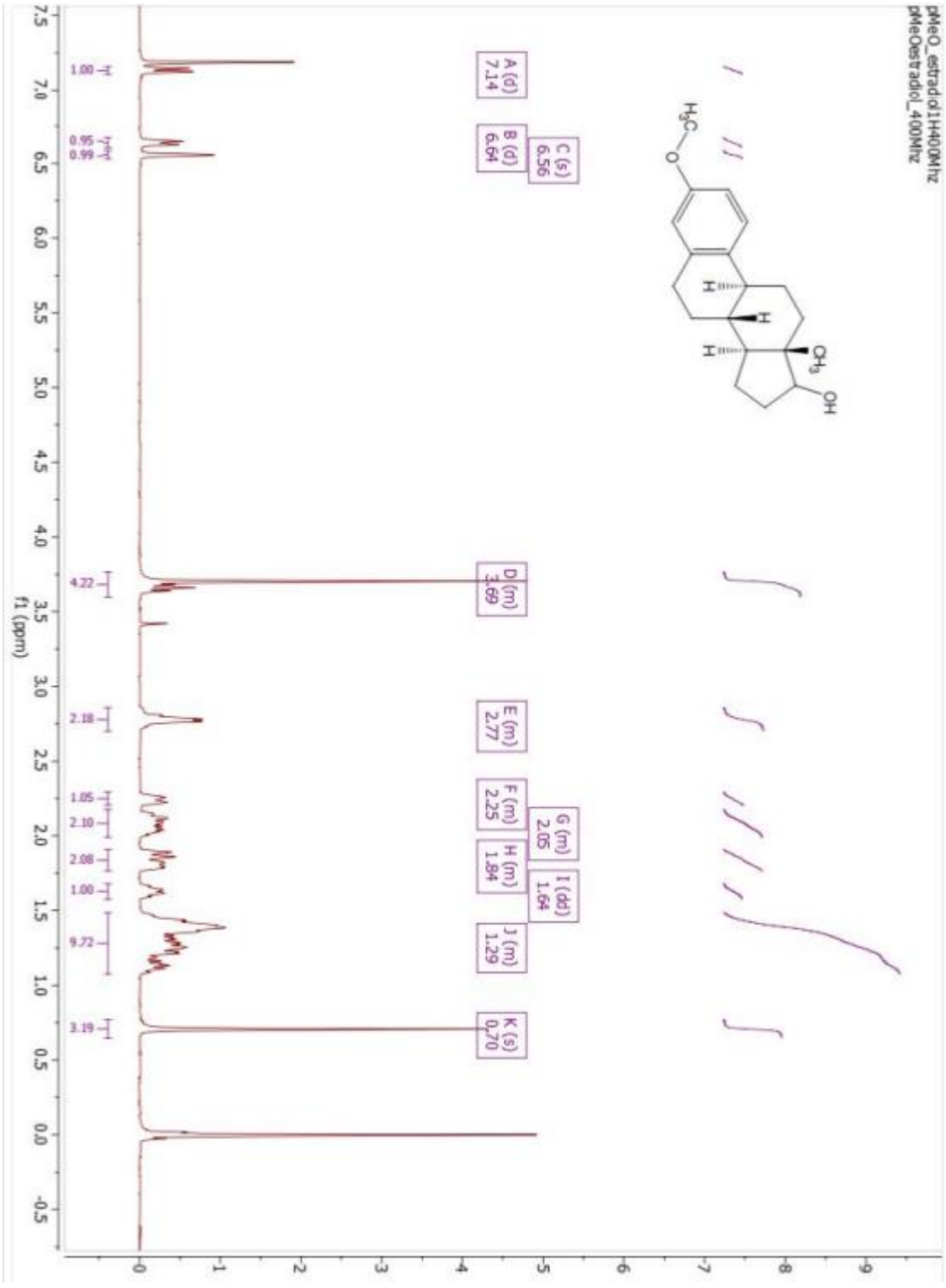


Figure AIII. 10: 13CNMR of substrate 2a

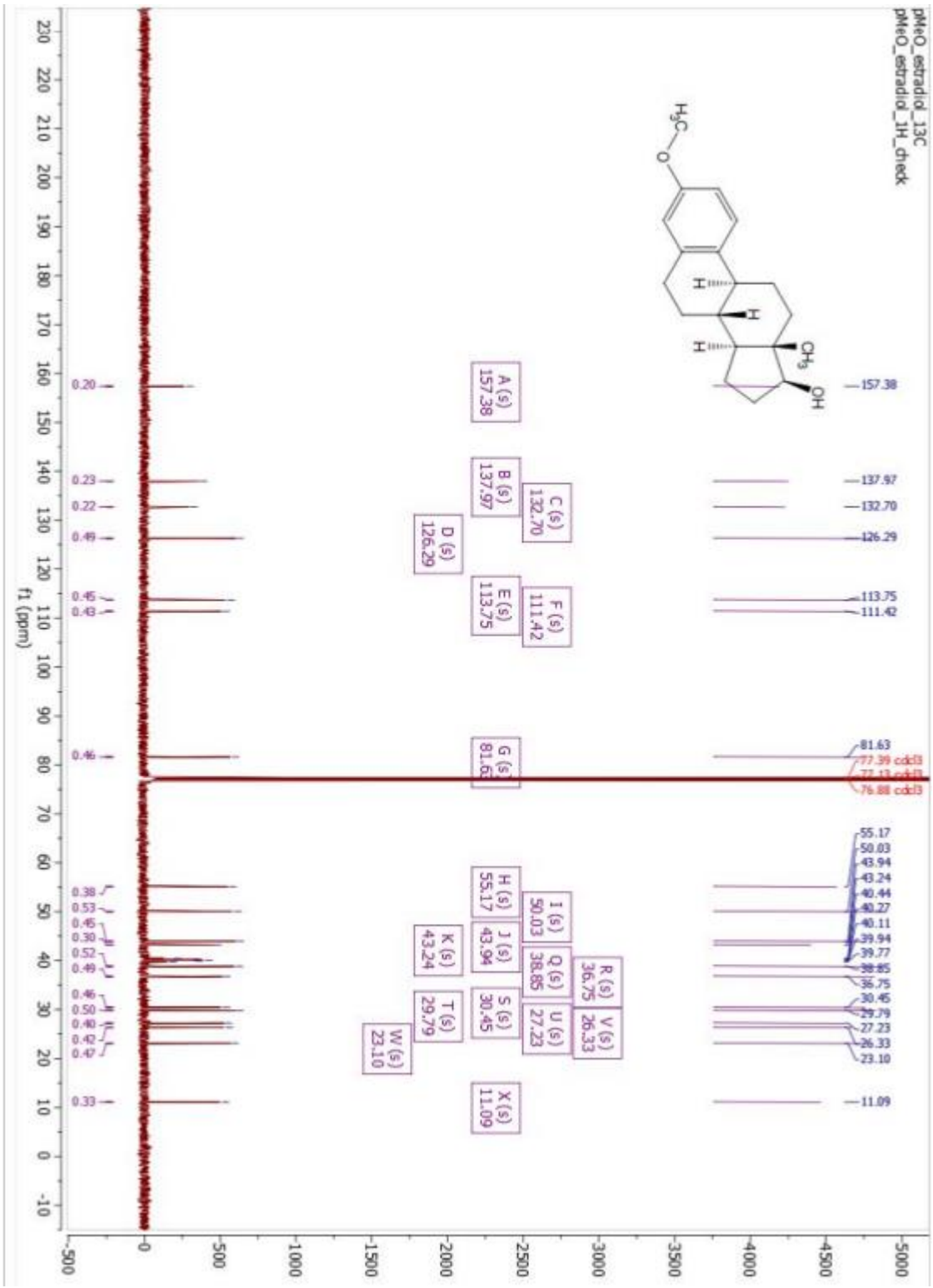


Figure AIII. 11: 1HNMR of substrate 2b

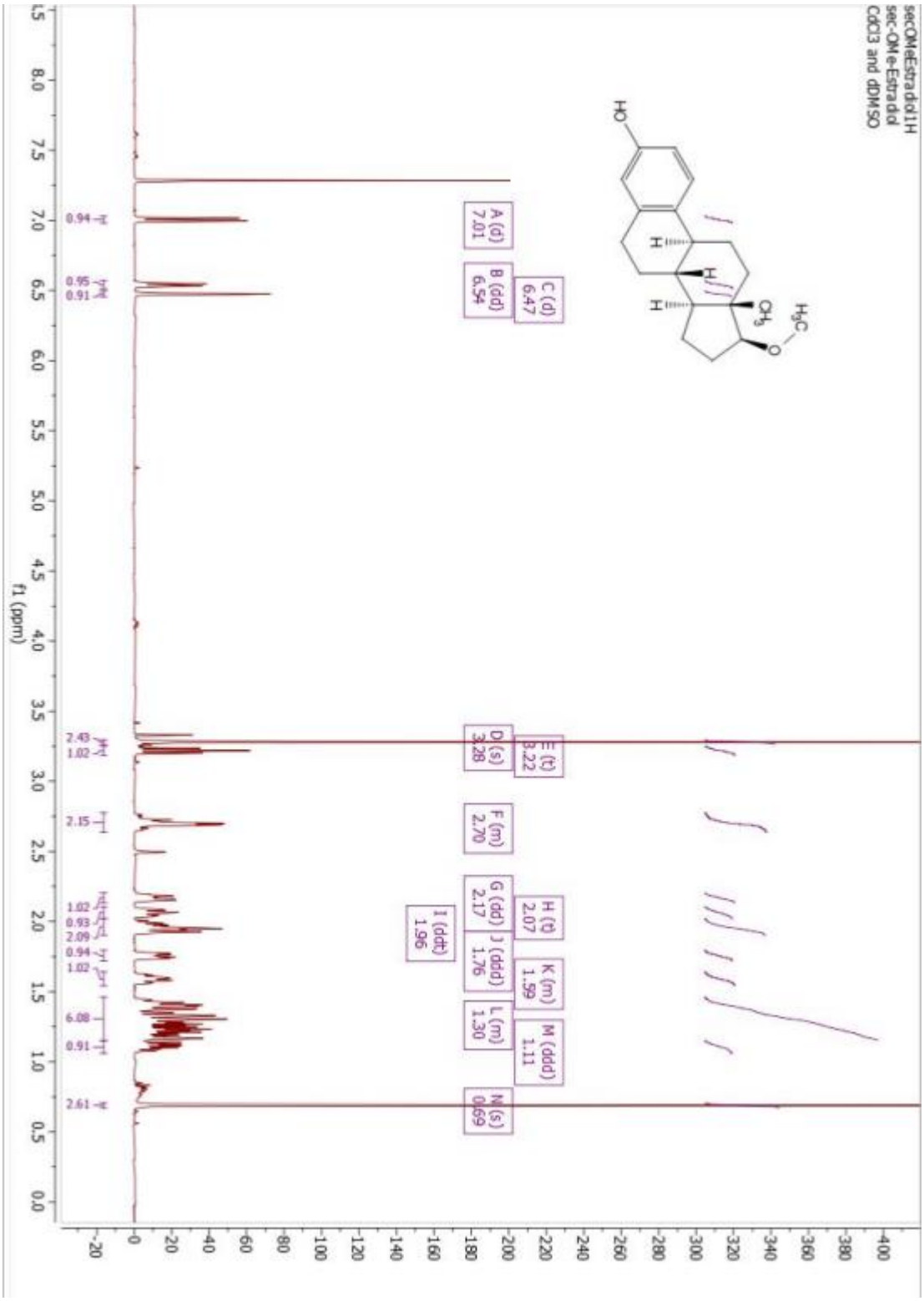


Figure AIII. 12: 13CNMR of substrate 2b

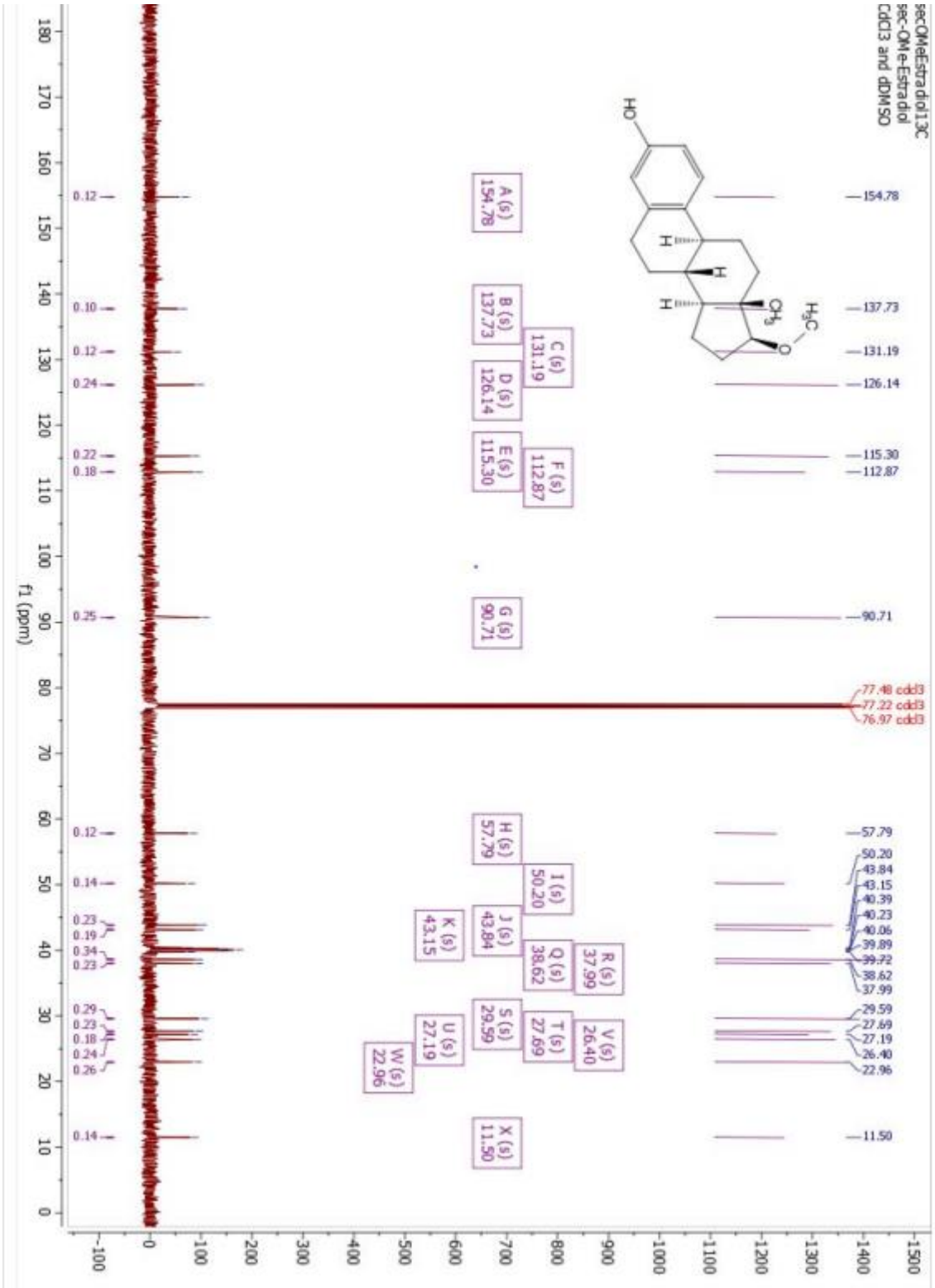


Figure AIII. 13: 1HNMR of substrate 2c

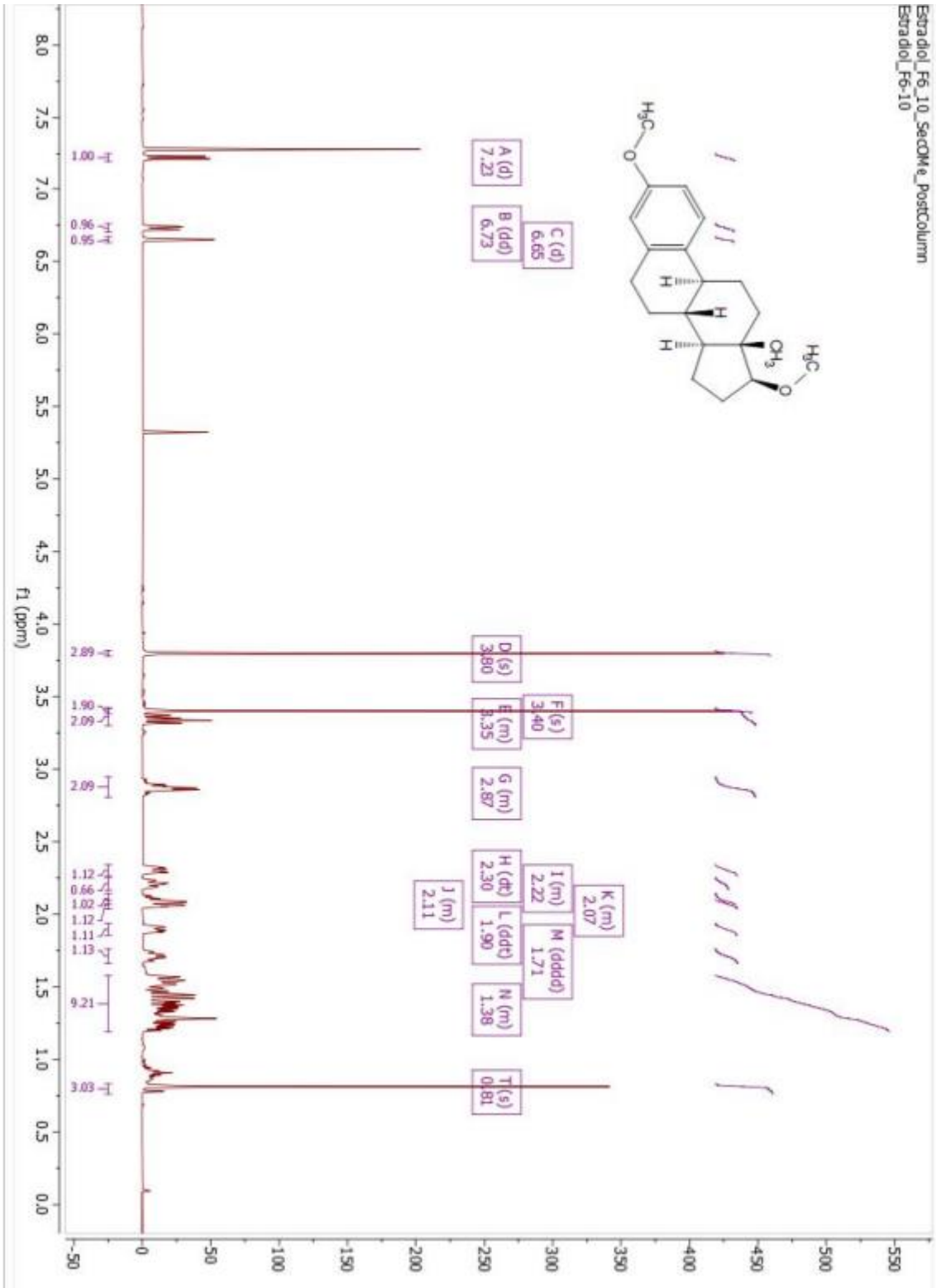


Figure AIII. 14: 13CNMR of substrate 2c

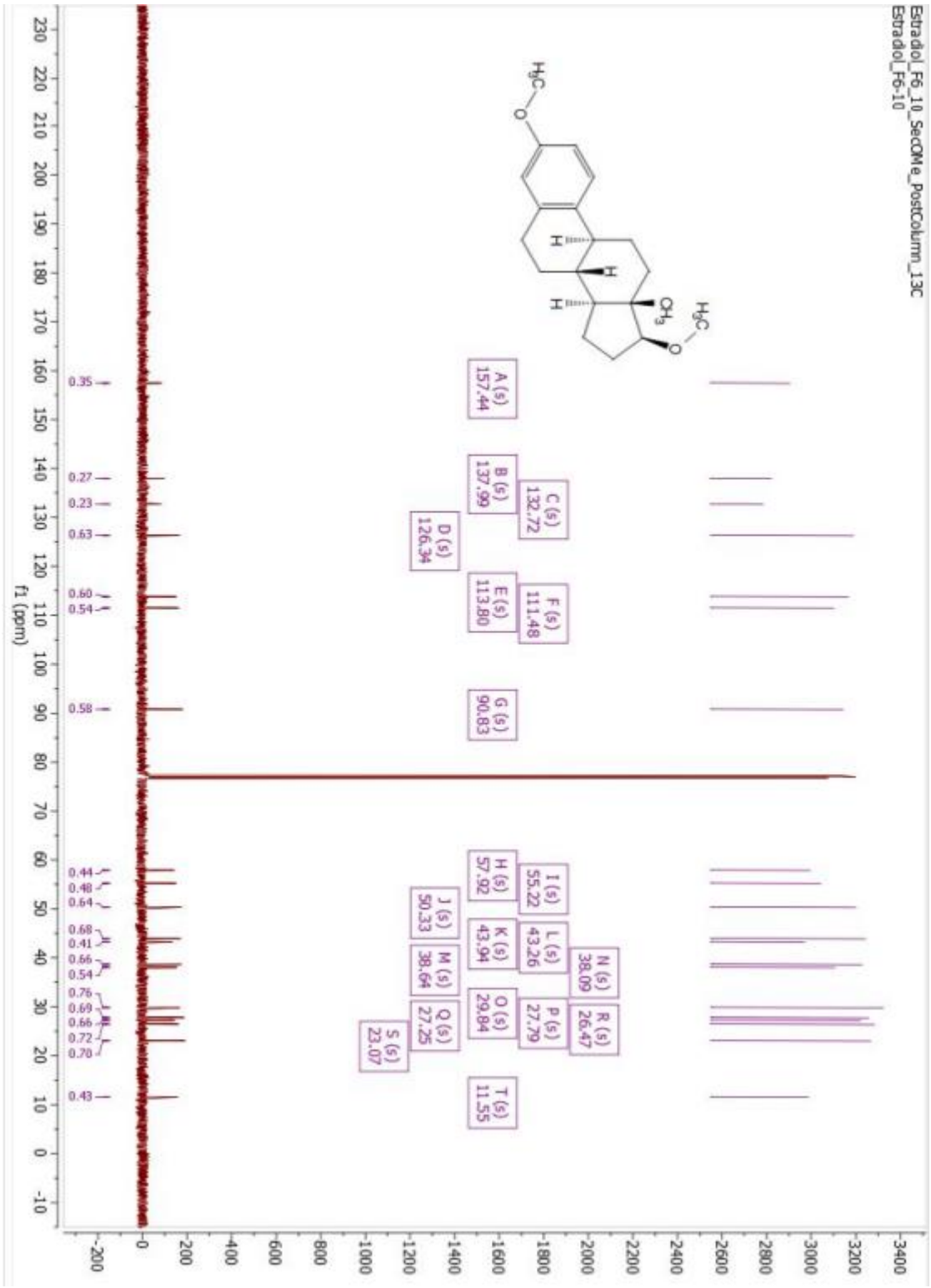


Figure AIII. 15: 1HNMR of substrate 4a

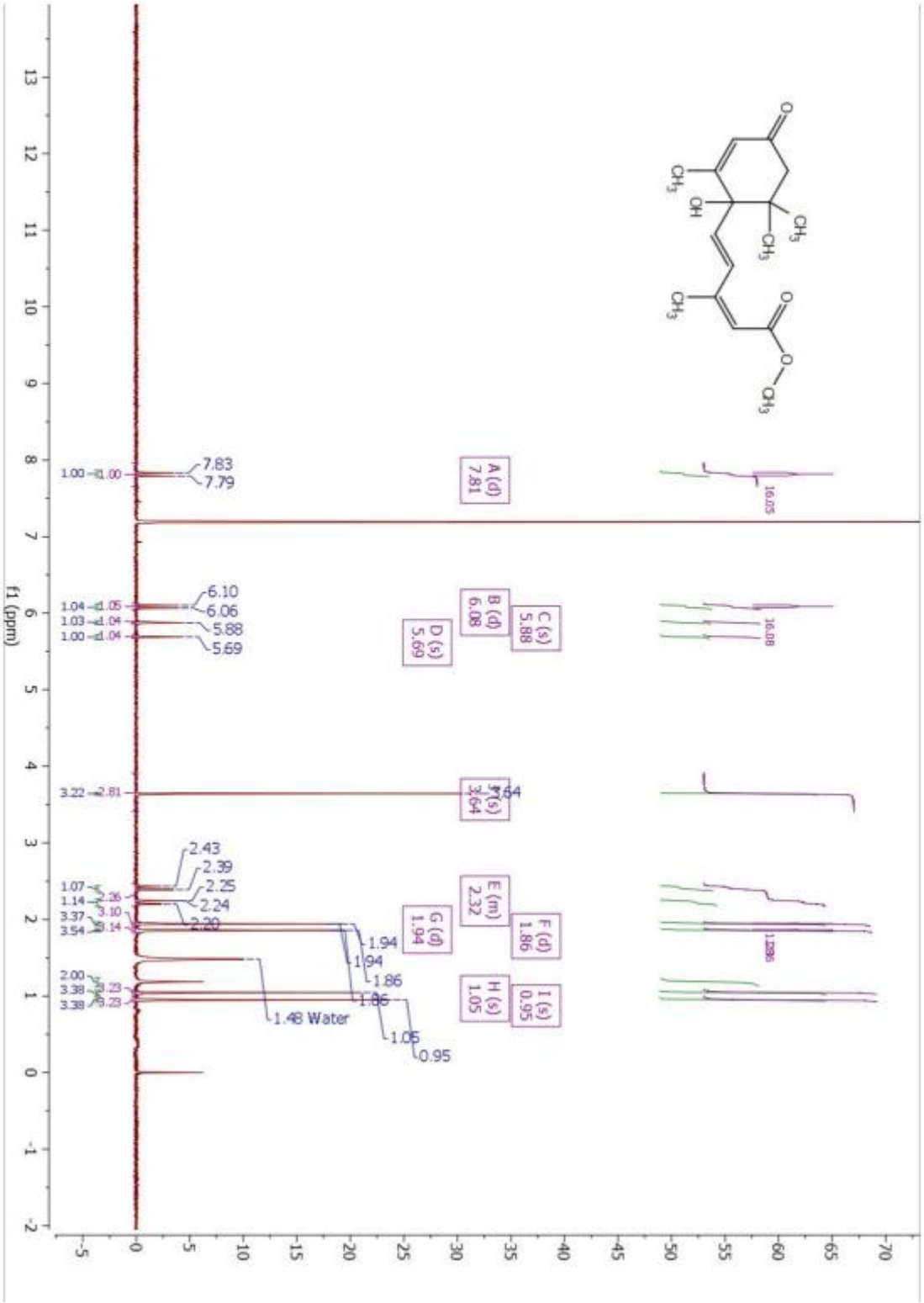


Figure AIII. 16: 13CNMR of substrate 4a

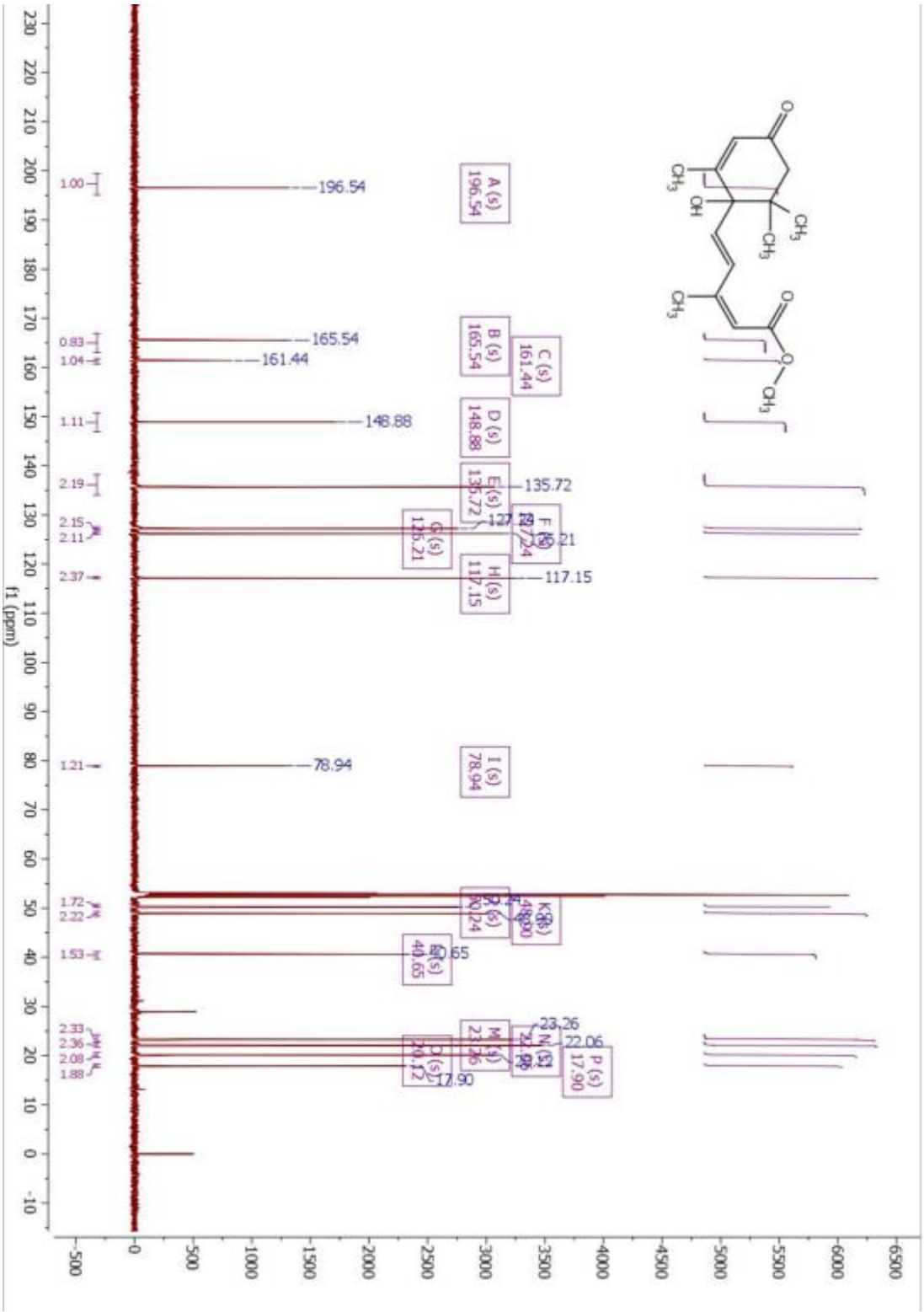


Figure AIII. 17: 1HNMR of substrate 4b

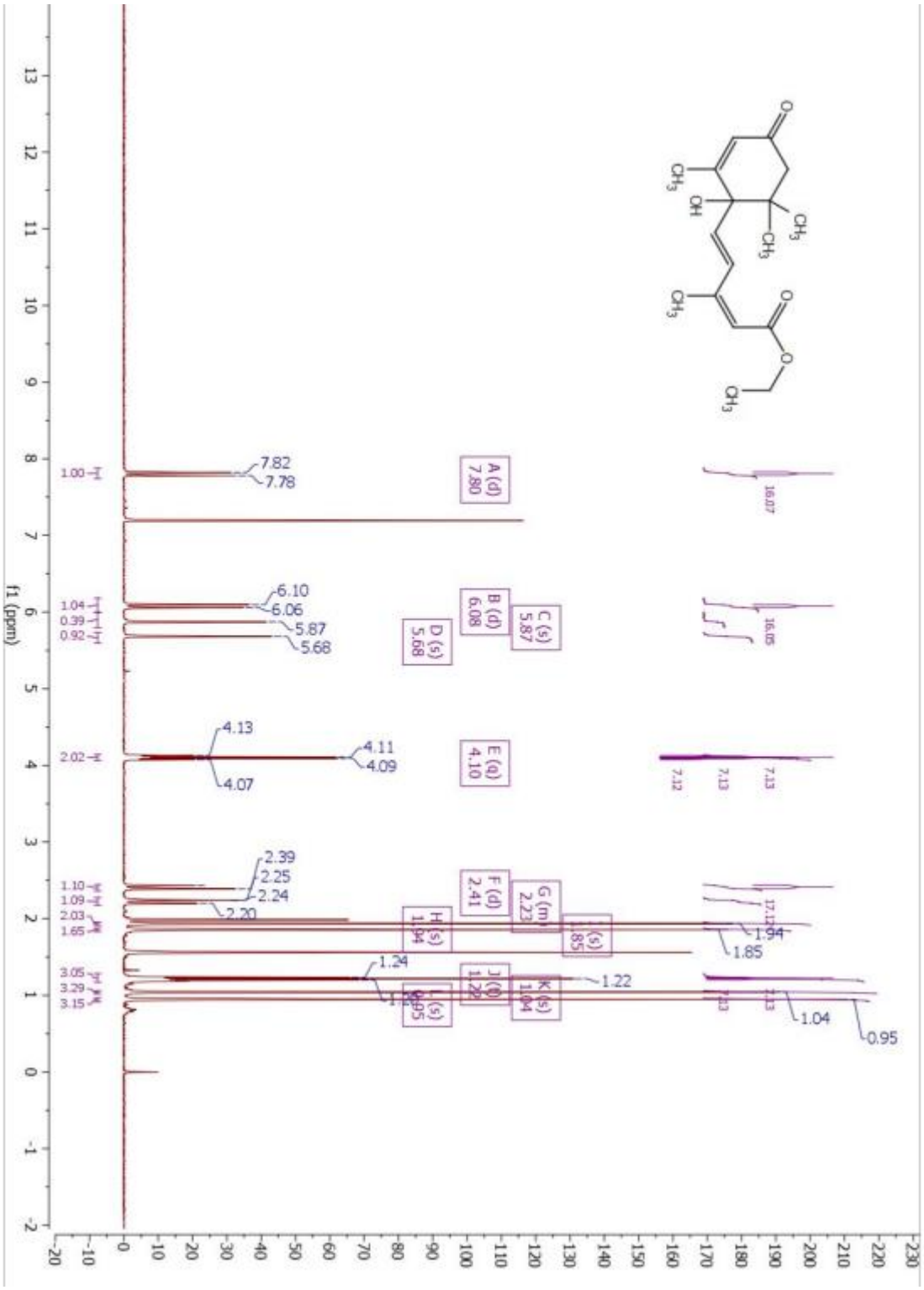


Figure AIII. 18: 13CNMR of substrate 4b

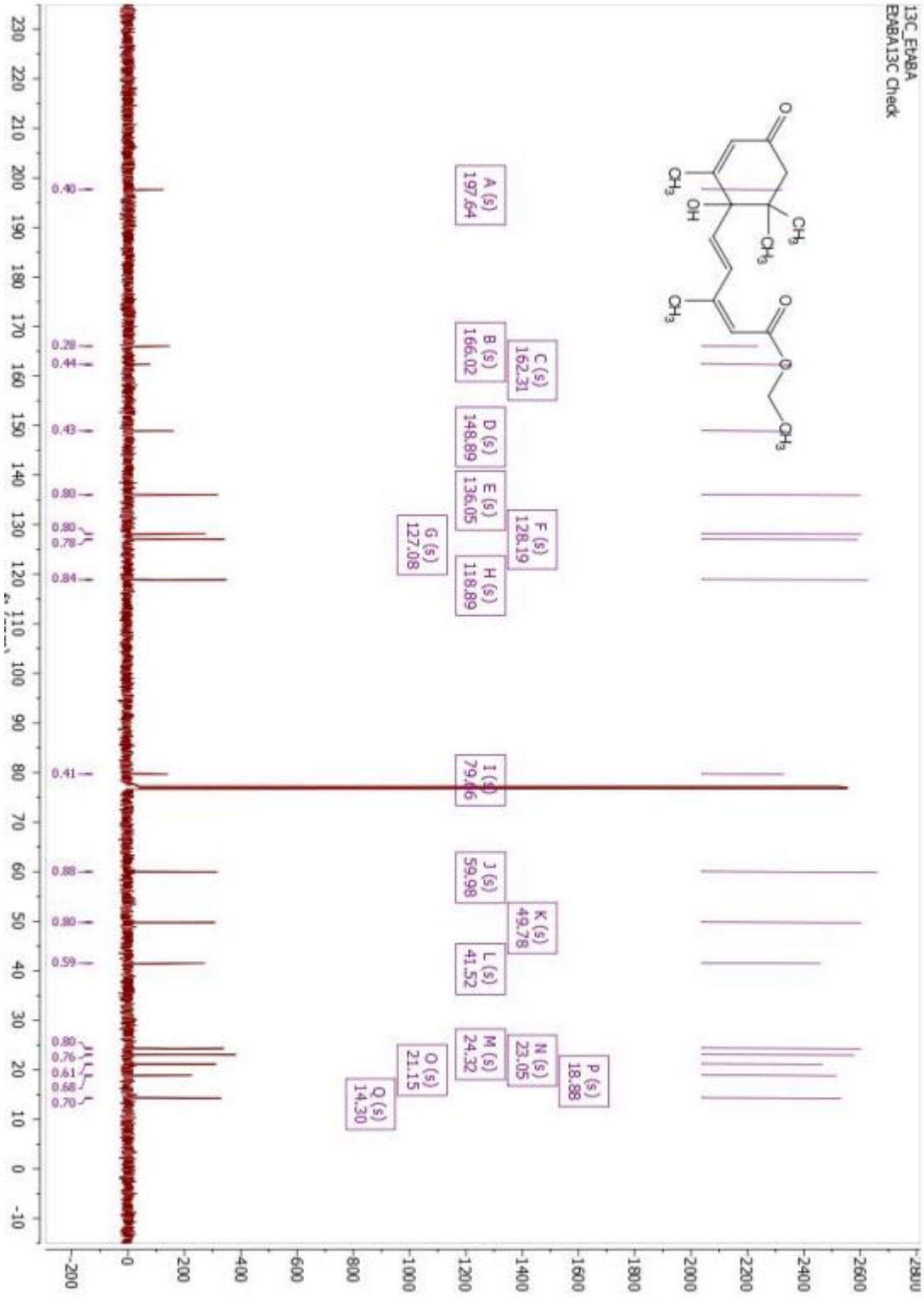


Figure AIII. 19: 1HNMR of substrate 4c

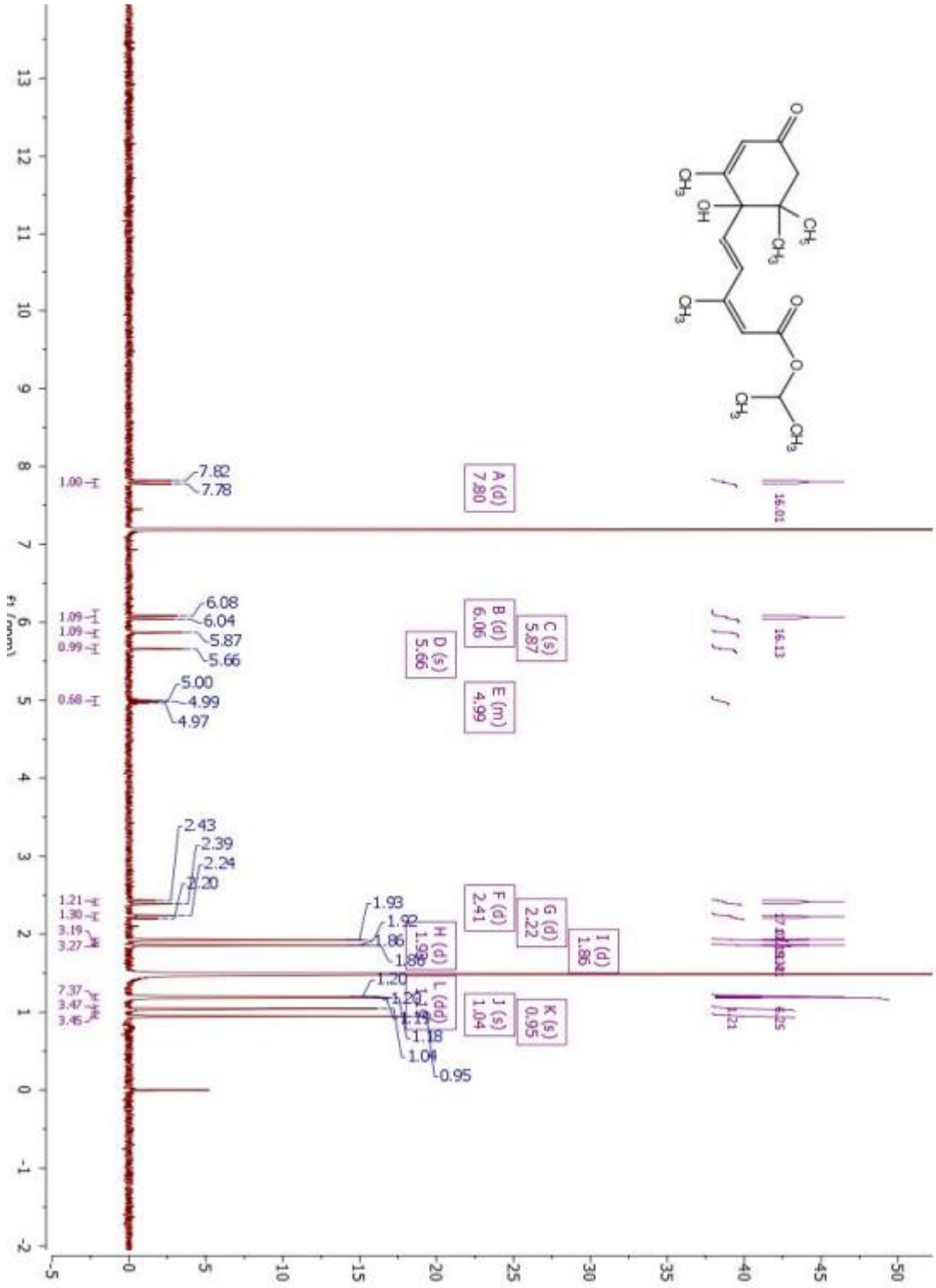


Figure AIII. 20: 13CNMR of substrate 4c

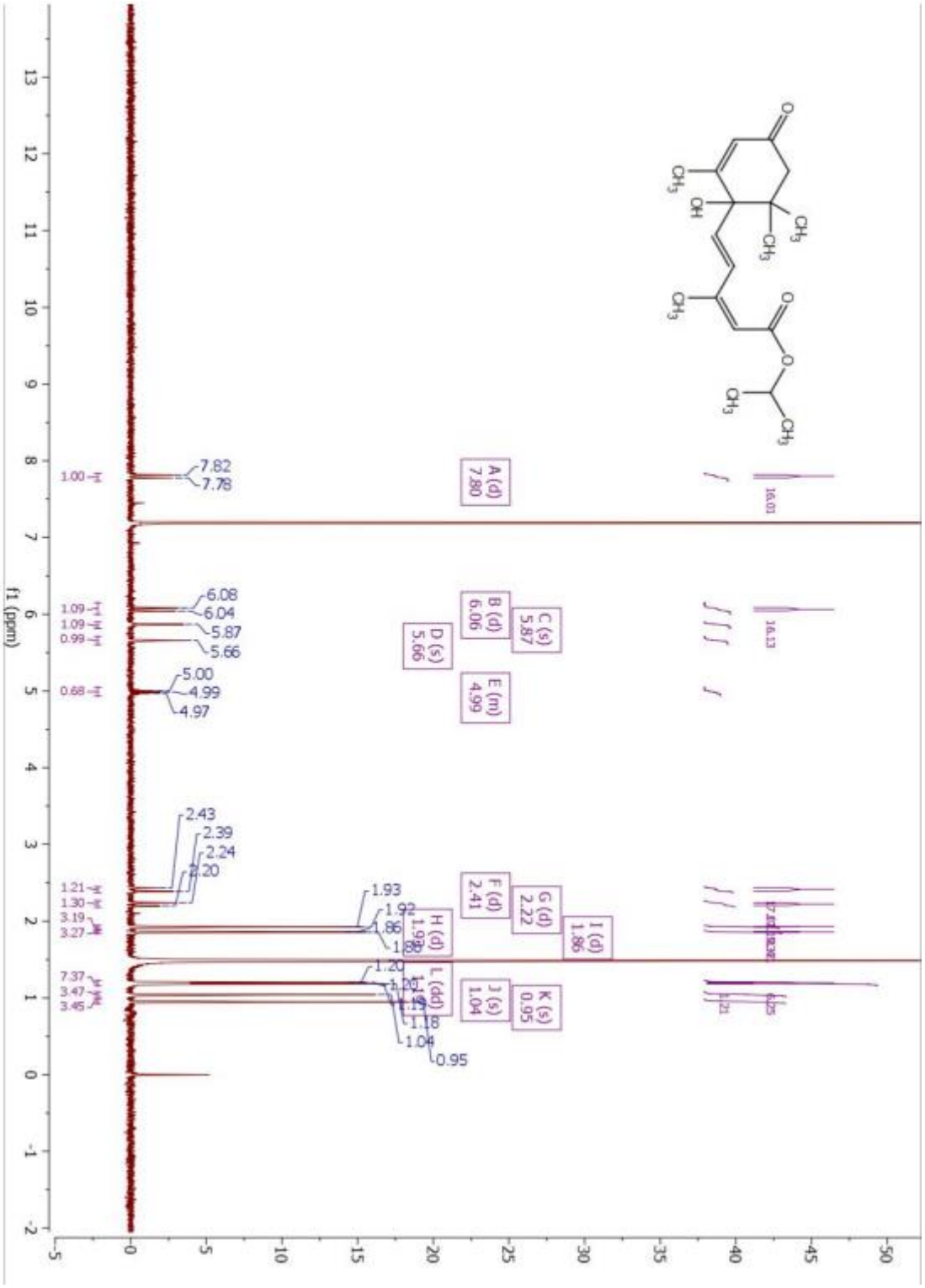


Figure AIII. 21: 1HNMR of substrate 4d

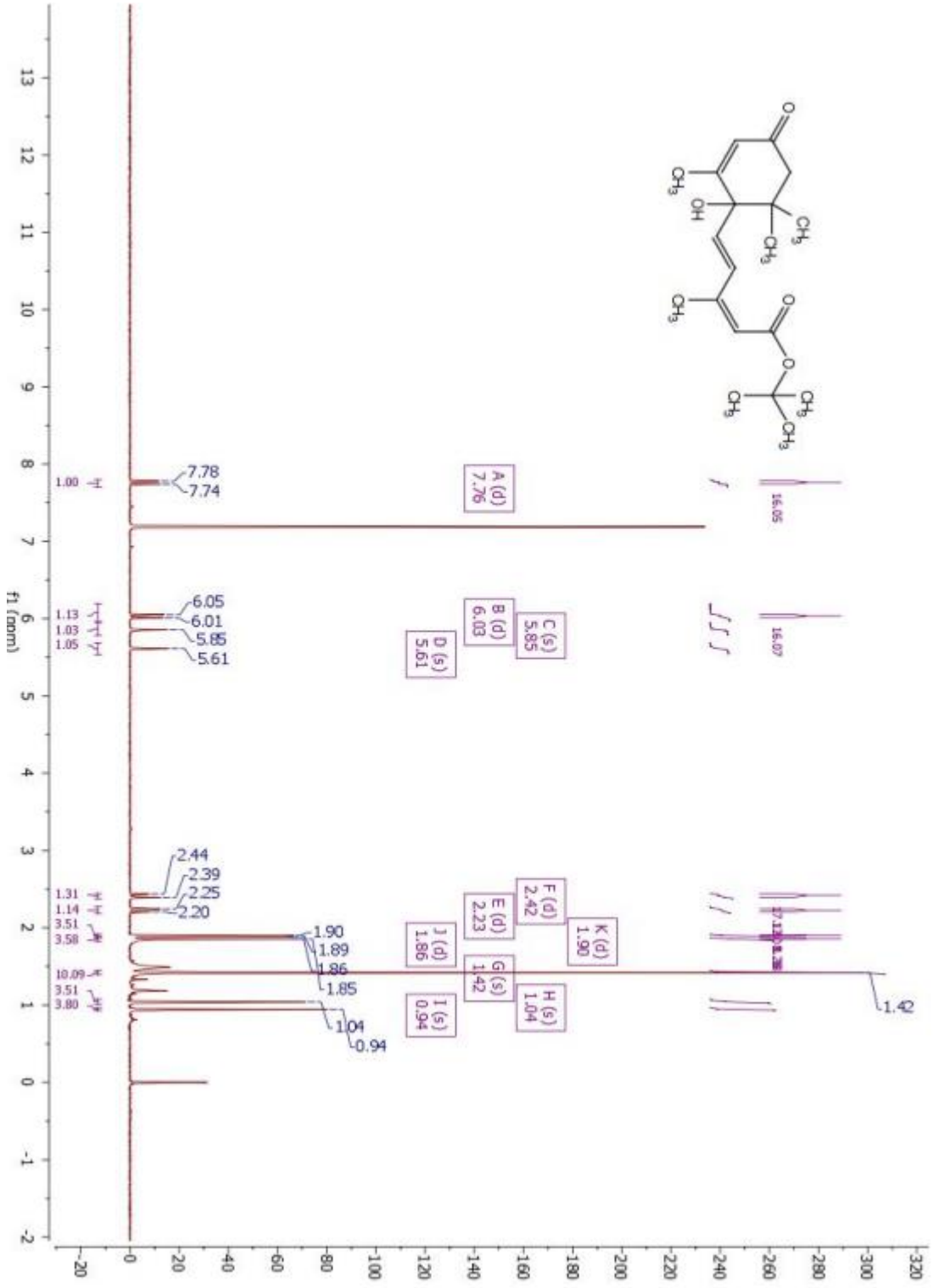


Figure AIII. 22: 13CNMR of substrate 4d

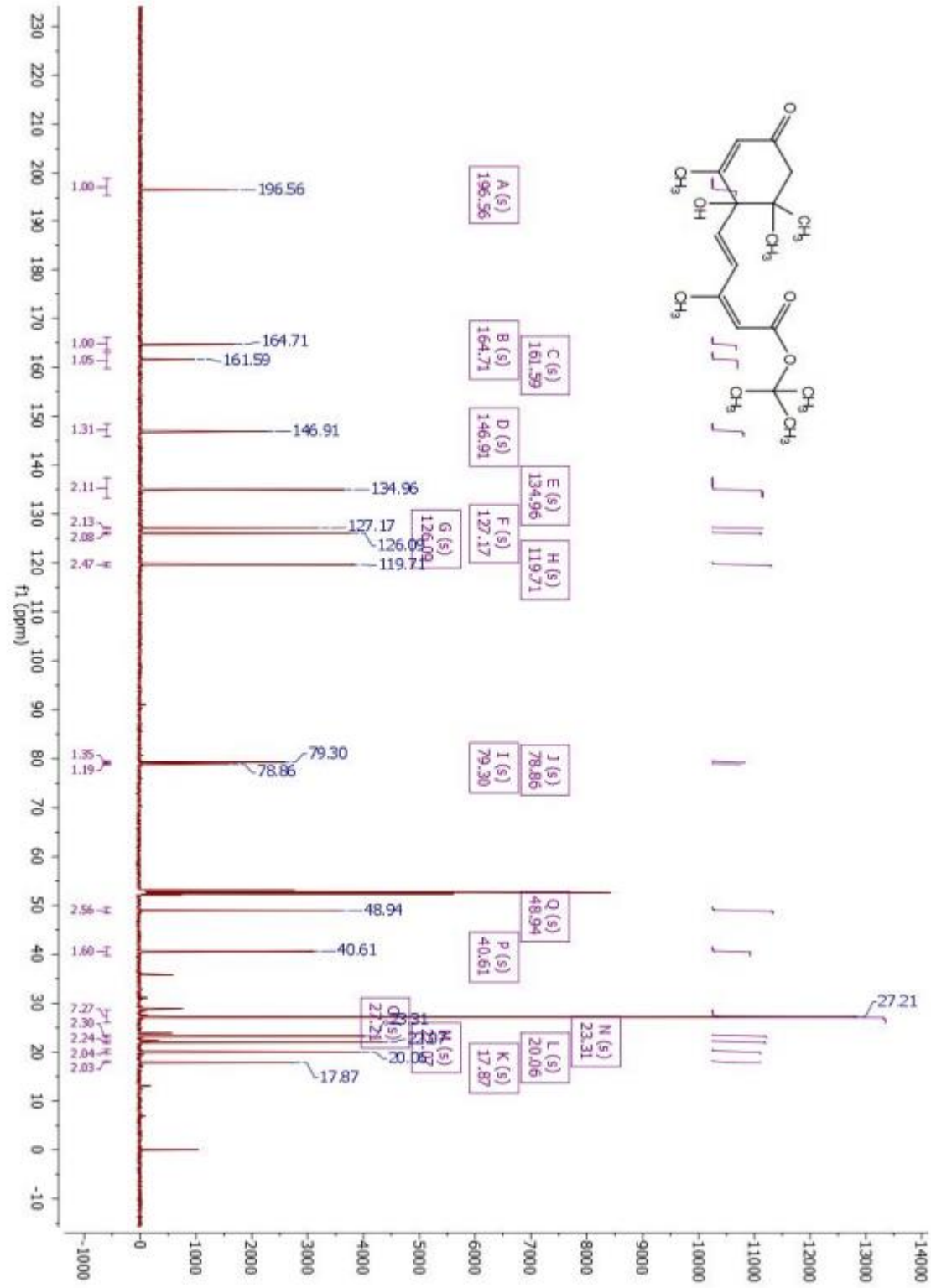


Figure AIII. 23: 1HNMR of substrate 5

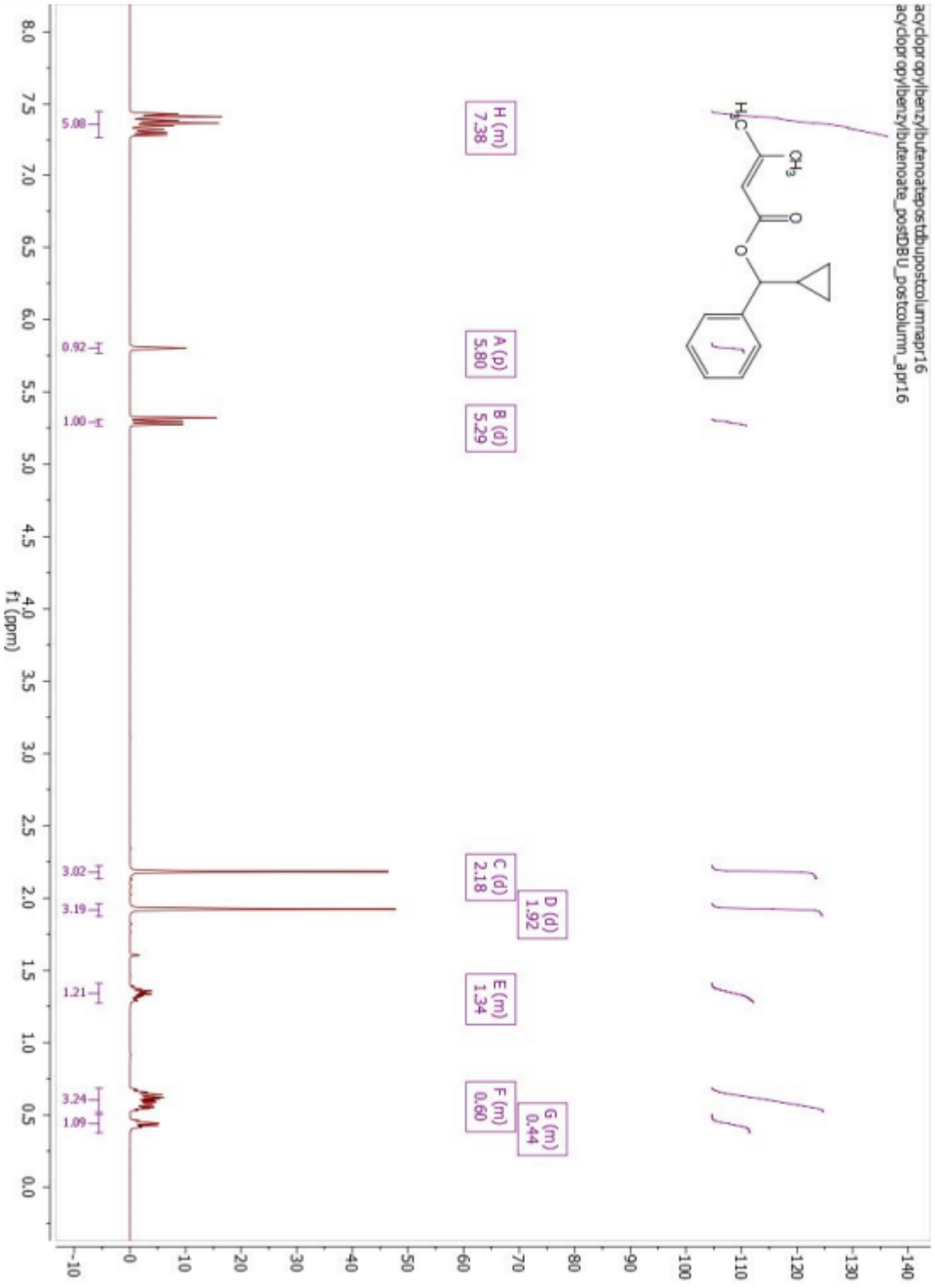


Figure AIII. 24: 13CNMR of substrate 5

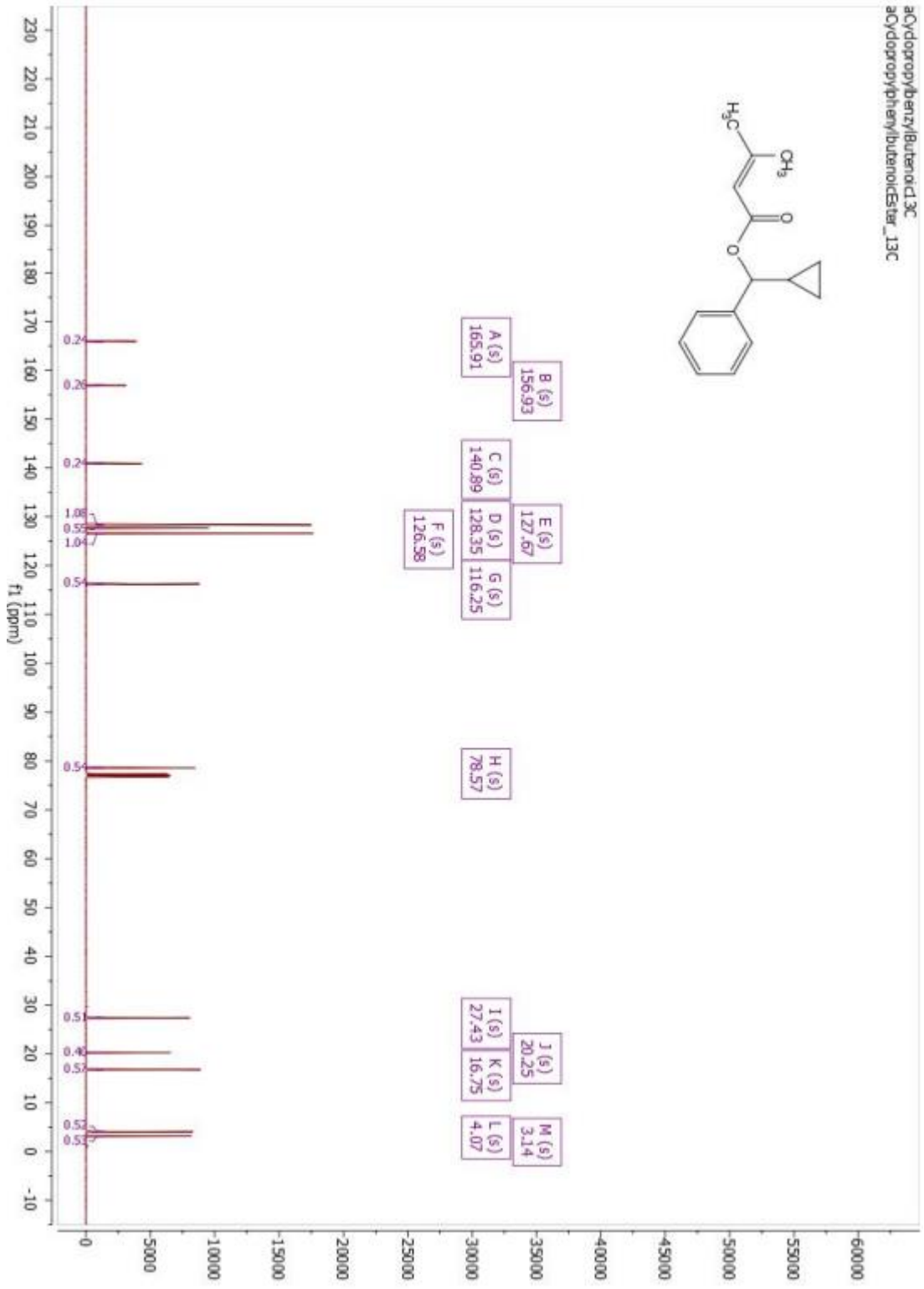


Figure AIII. 25: 1HNMR of substrate 7

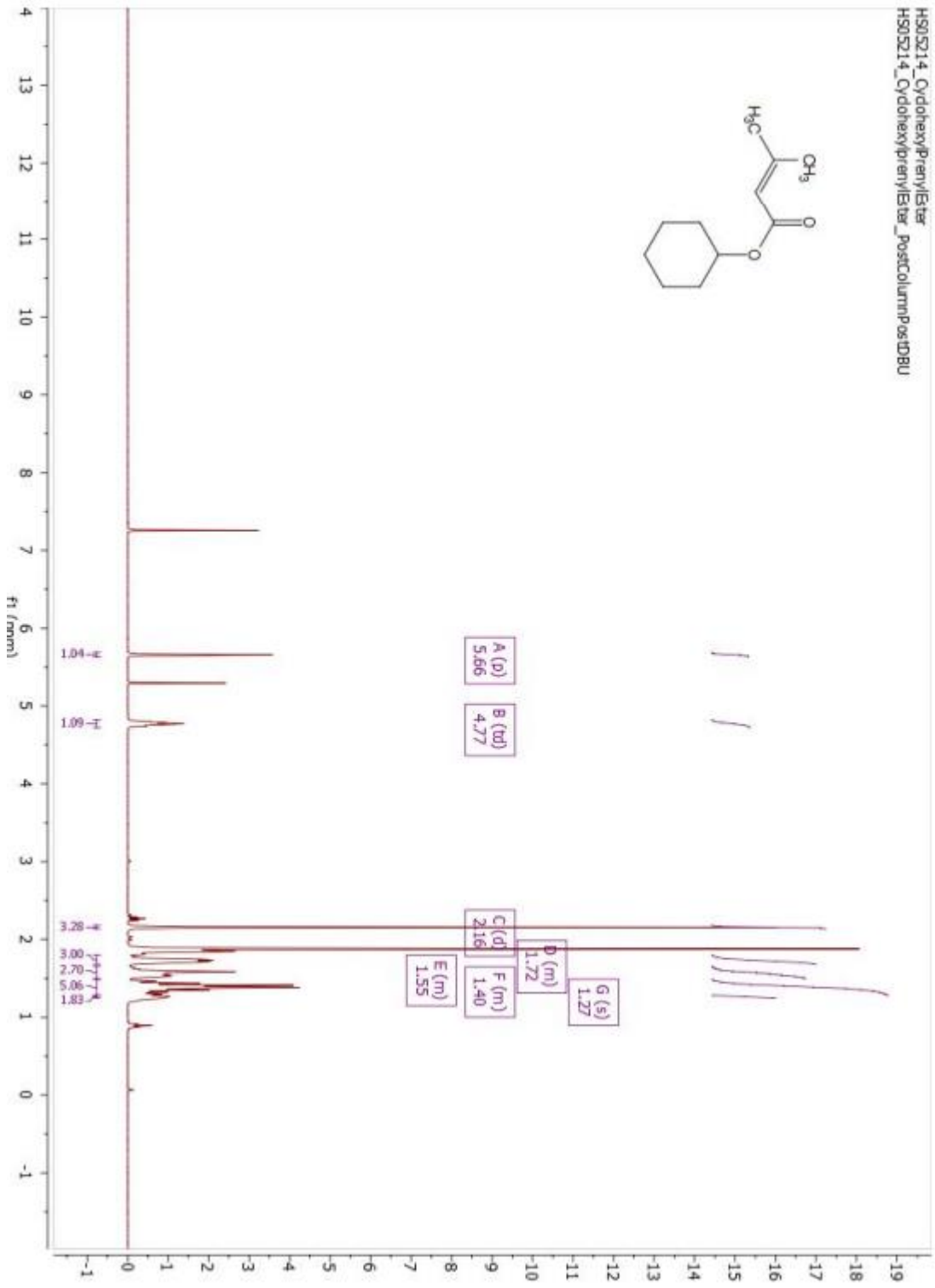


Figure AIII. 26: 13CNMR of substrate 7

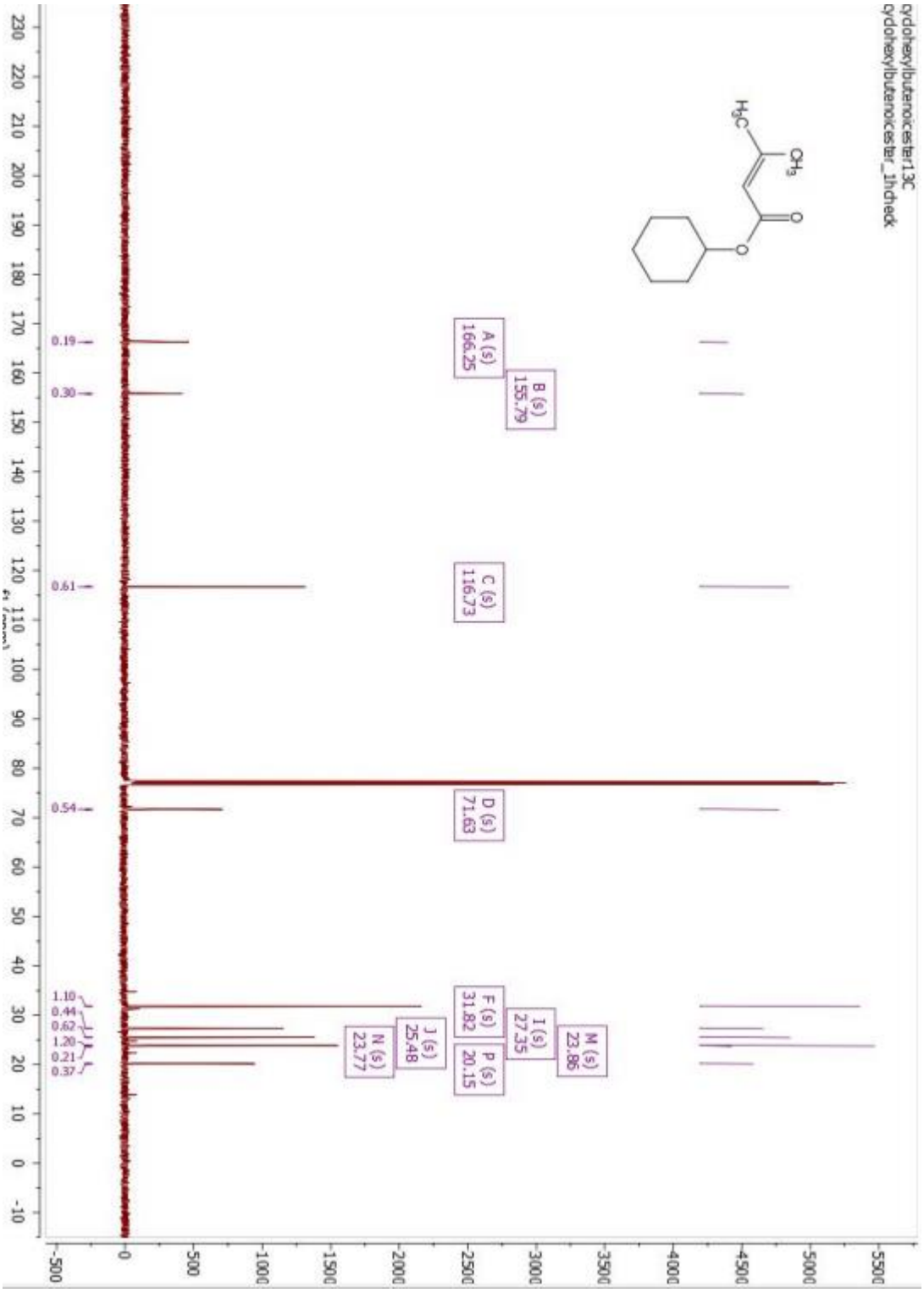


Figure AIII. 27: 1HNMR of substrate 8

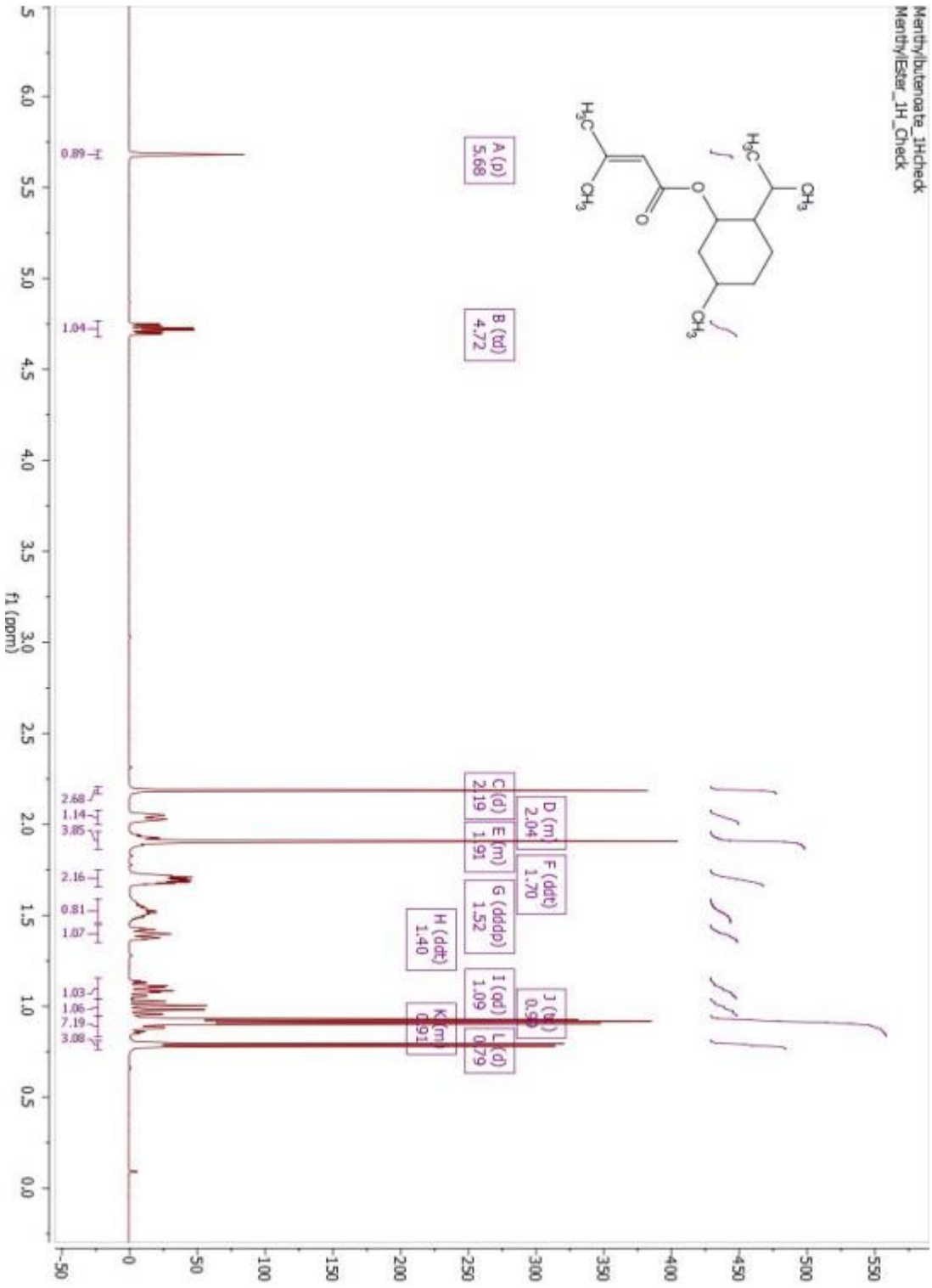
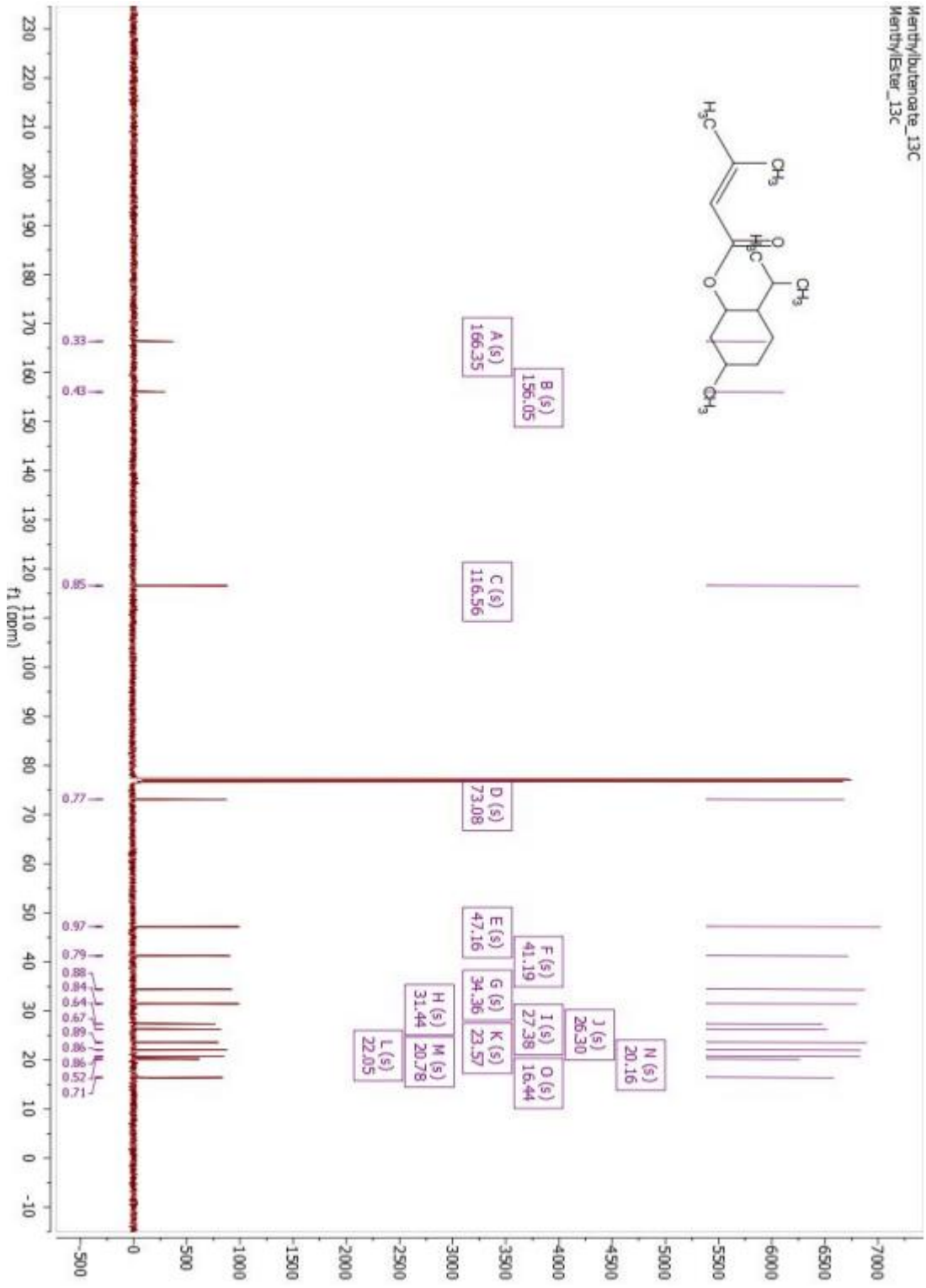


Figure AIII. 28: 13CNMR of substrate 8



APPENDIX IV

NMR Spectra for Compounds from Chapter 5

Figure AIV. 1: 1HNMR of (rac)- (2-methylcyclopropyl)methyl (2Z,4E)-5-(1-hydroxy-2,6,6-trimethyl-4-oxocyclohex-2-en-1-yl)-3-methylpenta-2,4-dienoate

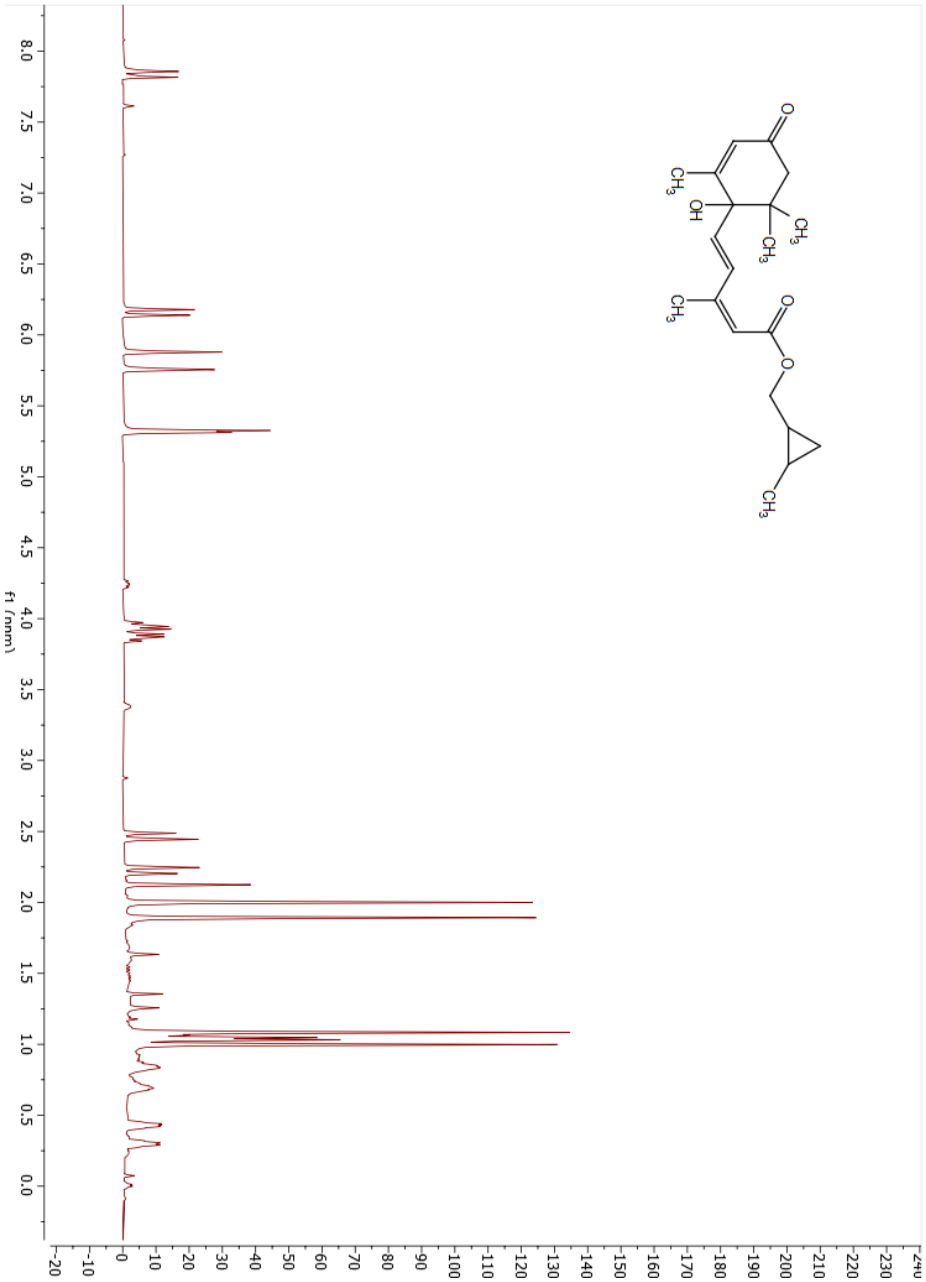


Figure AIV. 2: 13CNMR of (rac)- (2-methylcyclopropyl)methyl (2Z,4E)-5-(1-hydroxy-2,6,6-trimethyl-4-oxocyclohex-2-en-1-yl)-3-methylpenta-2,4-dienoate

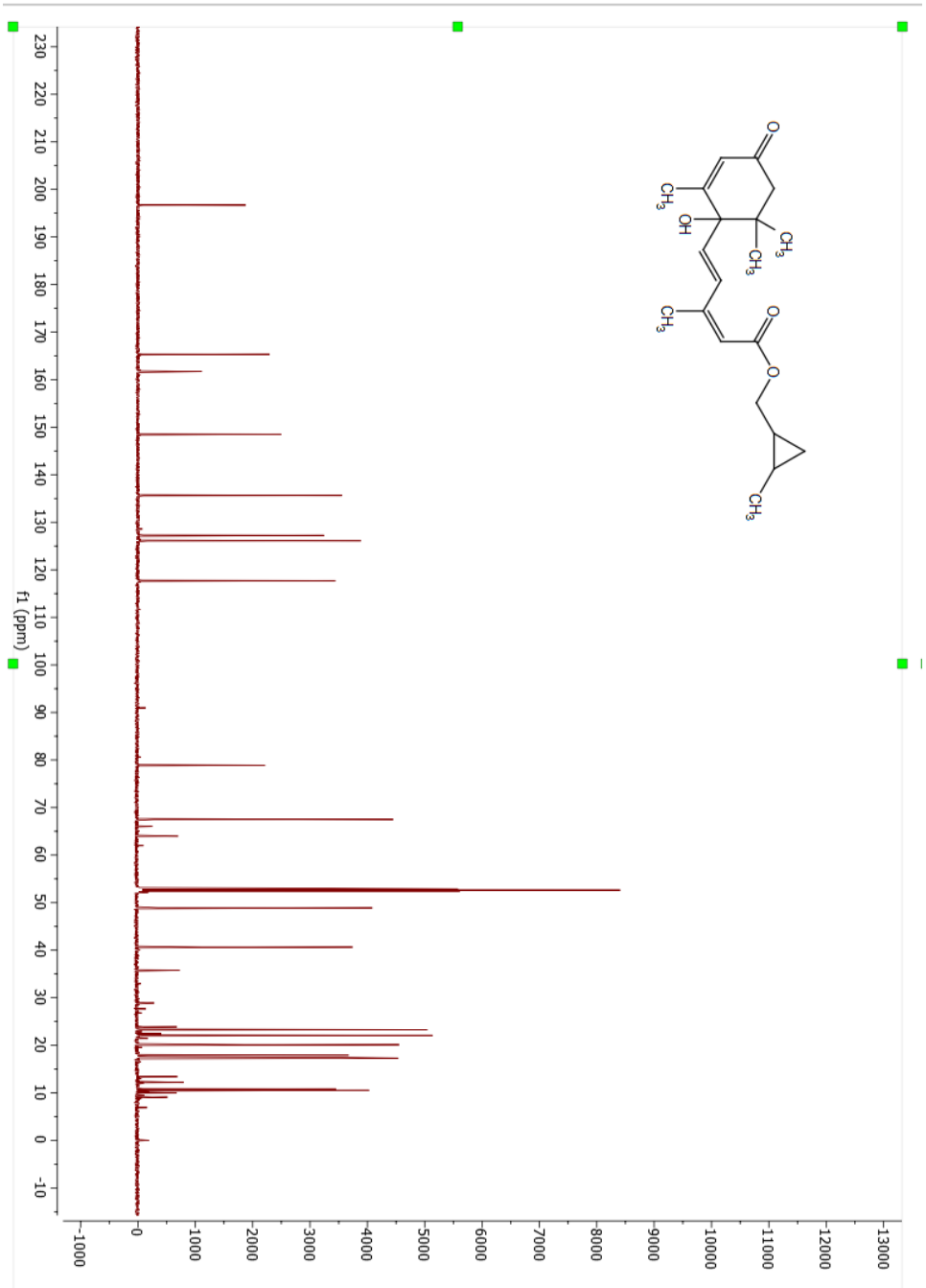


Figure AIV. 3: 1HNMR 4-oxo-2-phenylchroman-7-yl 1-(2,2-difluorobenzo[d][1,3]dioxol-5-yl)cyclopropane-1-carboxylate

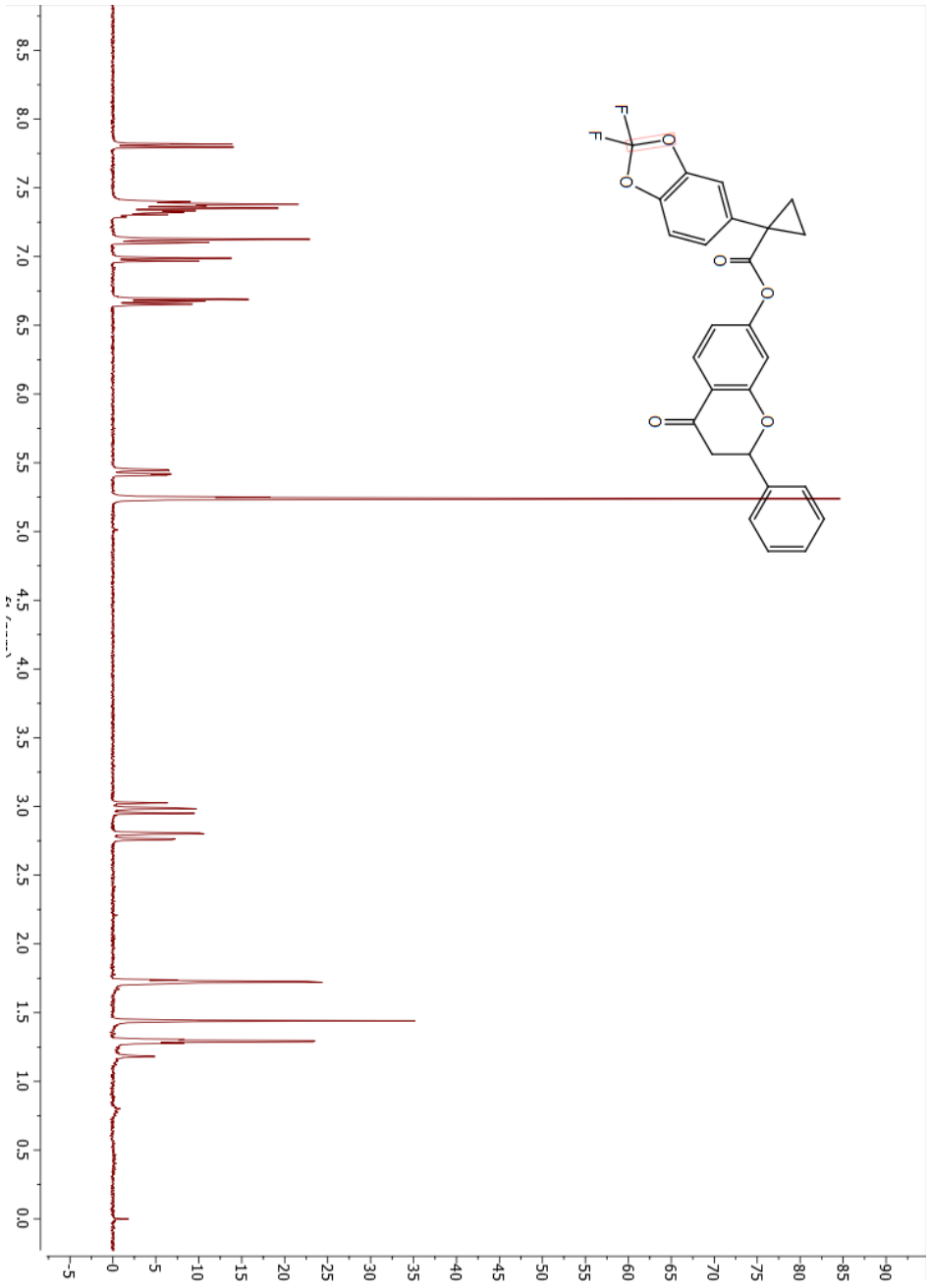


Figure AIV. 4: 13CNMR 4-oxo-2-phenylchroman-7-yl 1-(2,2-difluorobenzo[d][1,3]dioxol-5-yl)cyclopropane-1-carboxylate

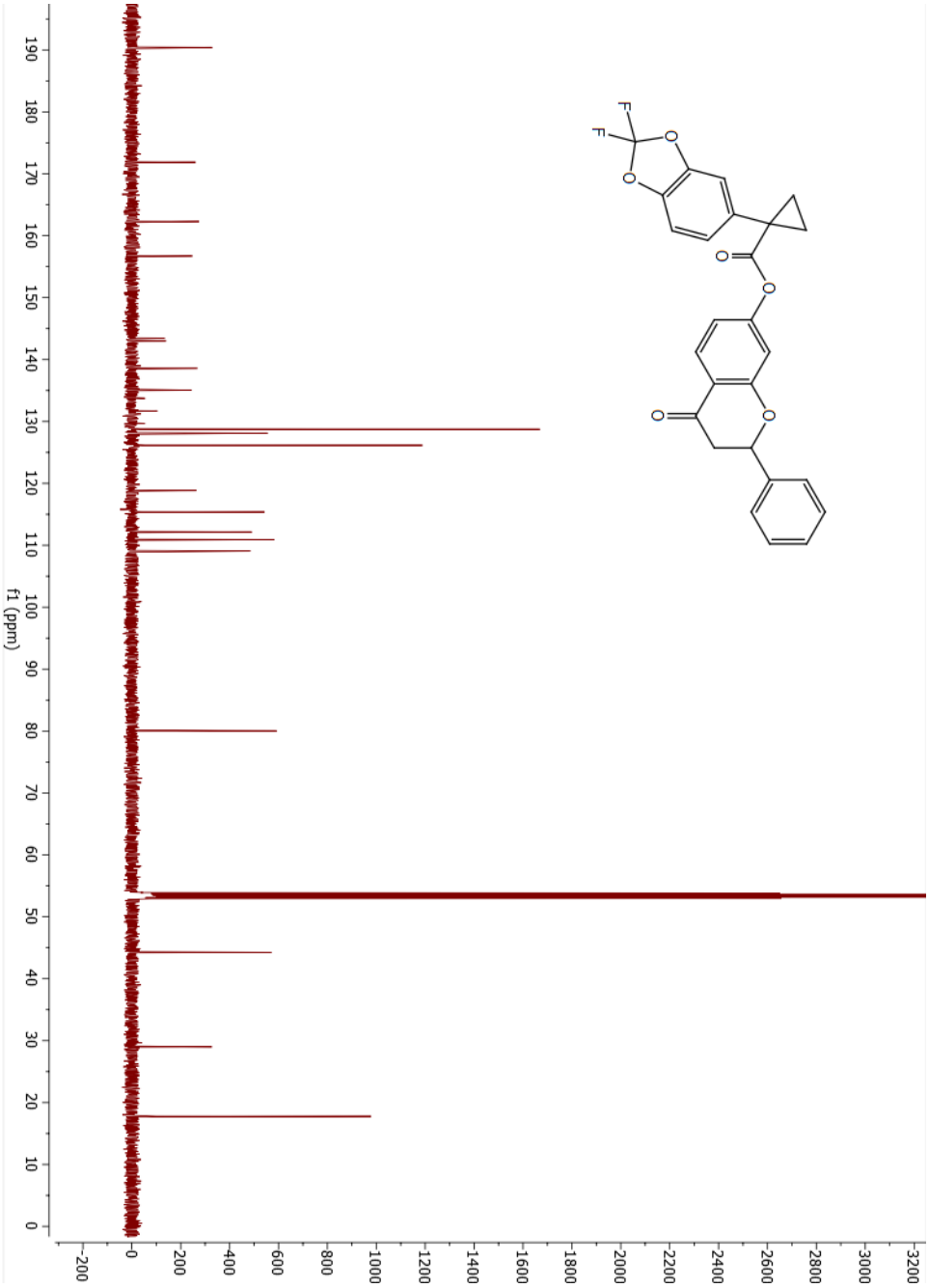


Figure AIV. 5: 1HNMR 4-oxo-2-phenylchroman-7-yl 1-(4-fluorophenyl)cyclopropane-1-carboxylate

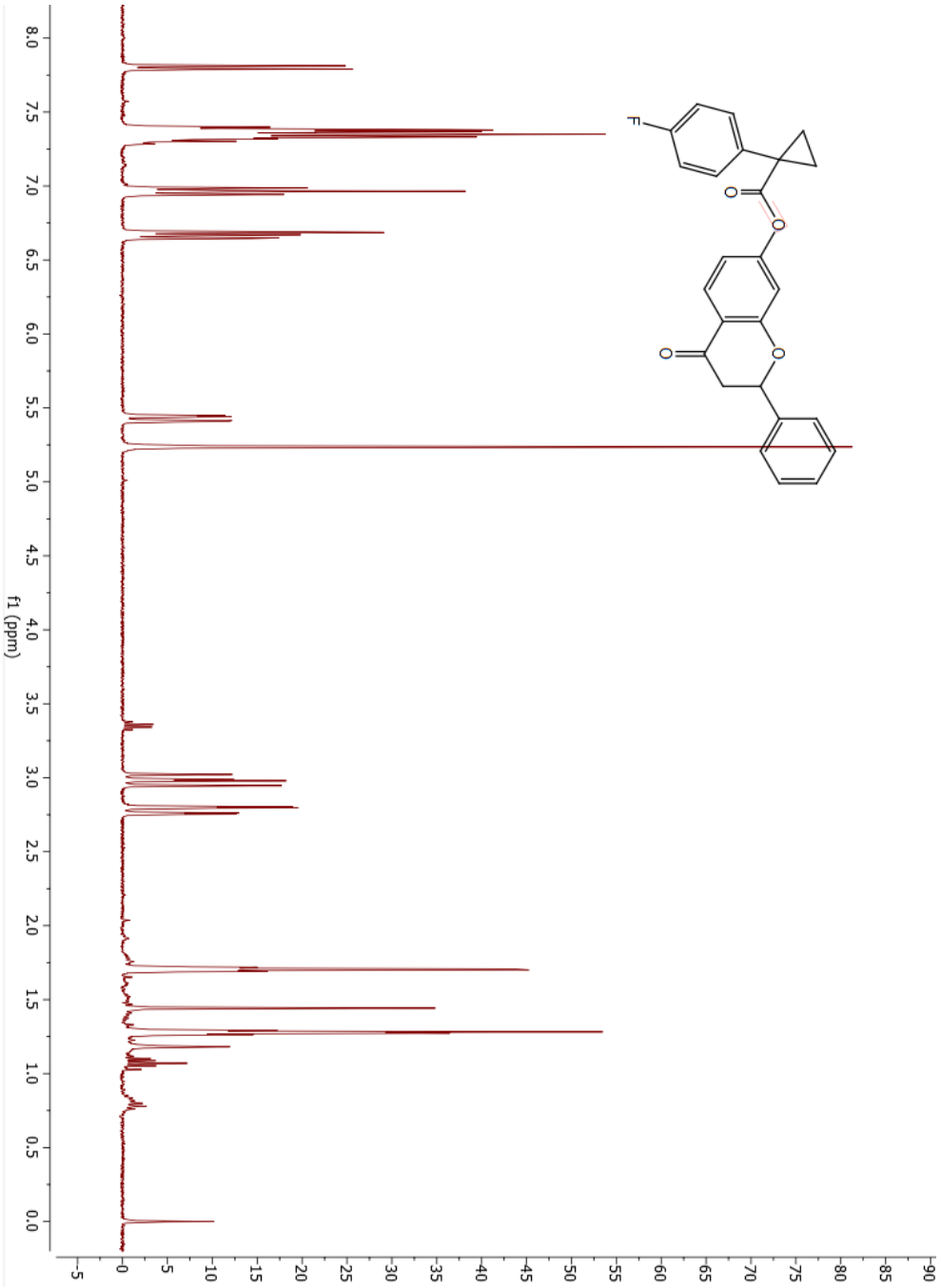


Figure AIV. 6: 13CNMR 4-oxo-2-phenylchroman-7-yl 1-(4-fluorophenyl)cyclopropane-1-carboxylate

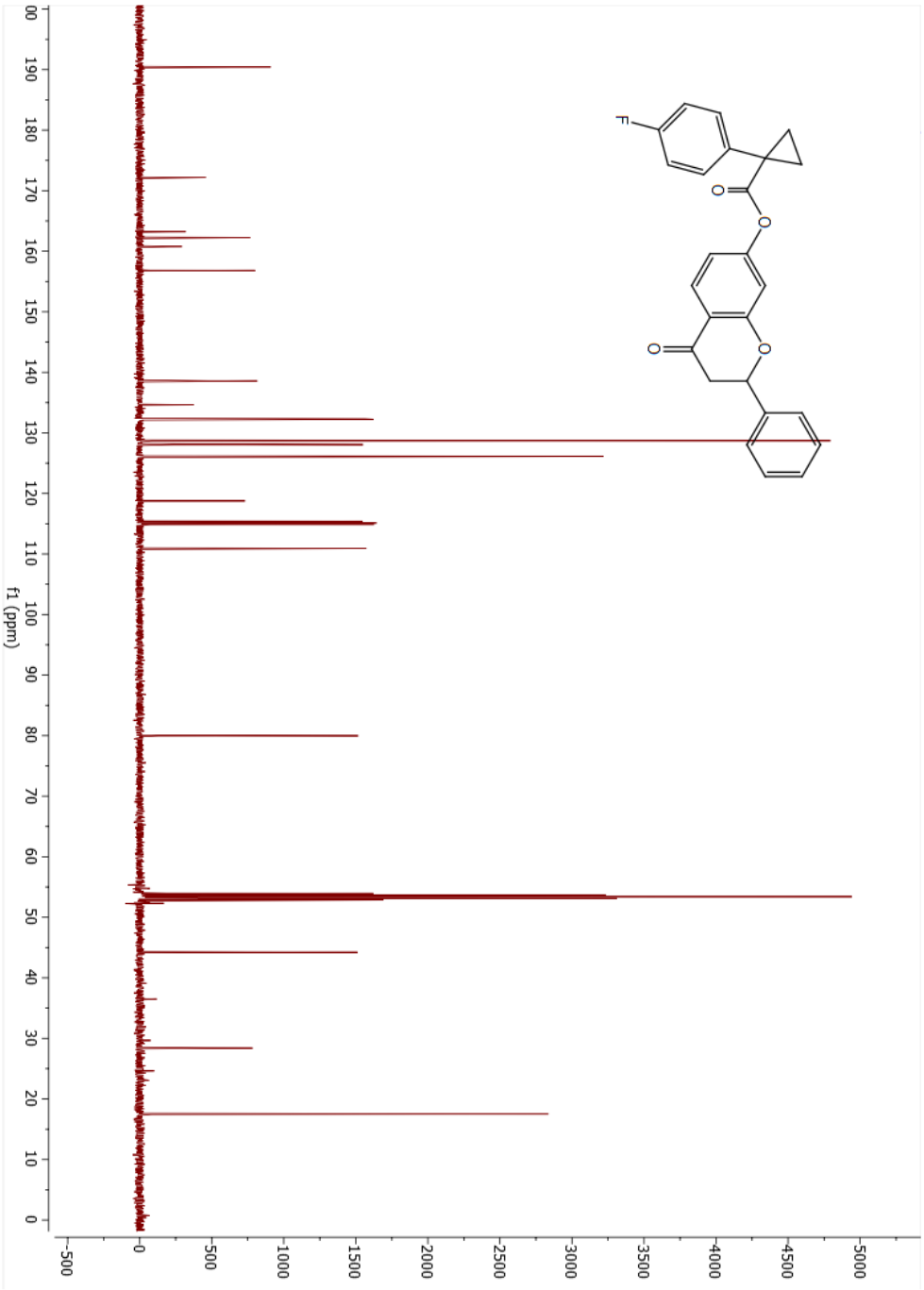
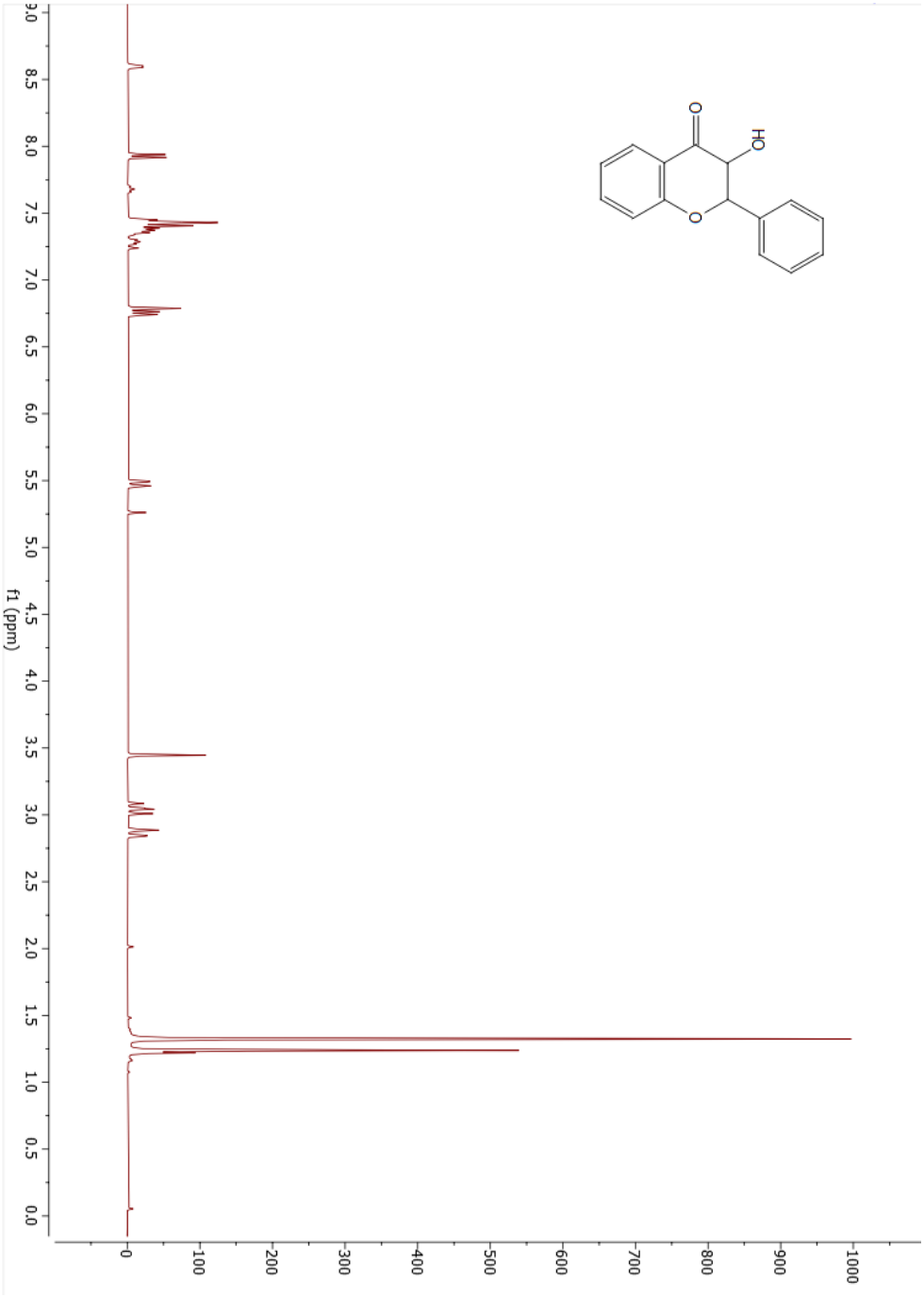


Figure AIV. 7: 1HNMR of BM3 variant F8 hydroxylated product 3-hydroxy-2-phenylchroman-4-one



APPENDIX V

NMR Spectra for Compounds from Chapter 6

Figure AV. 1: 1HNMR of 3-(quinolin-5-yl)aniline

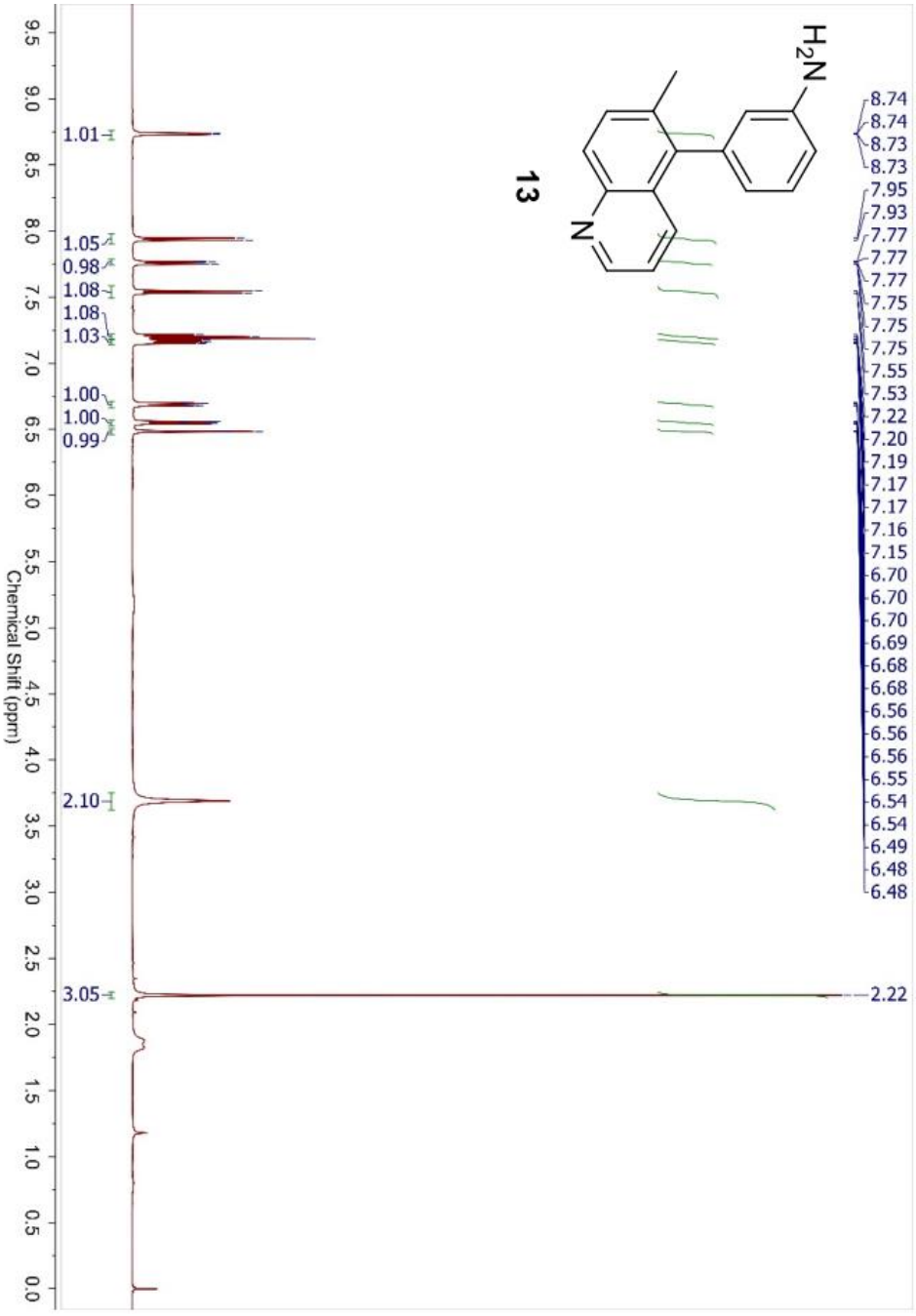


Figure AV. 2: 13CNMR of 3-(quinolin-5-yl)aniline