THE UNIVERSITY OF CHICAGO

ENGINEERING HALOGENASES FOR SELECTIVE FUNCTIONALIZATION OF BIOACTIVE MOLECULES

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

JAMES T. PAYNE

CHICAGO, ILLINOIS

DECEMBER 2015

To my wife Dana

to offer you a fraction of my gratitude

for your enduring love and support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

LIST OF FIGURES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

APPENDIX I

APPENDIX II

APPENDIX III

LIST OF SCHEMES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

ACKNOWLEDGEMENTS

First, I would like to thank those with whom I have worked on the research described herein over the past half-decade: my research advisor, Professor Jared Lewis, whose support and instruction have indelibly shaped the scientist I am today; my coworkers on these projects, Dr. Landon Durak, Dr. Catherine Poor, Mary Andorfer, Kyle Kunze, and Dr. Duo-Sheng Wang, to all of whom I am indebted for their invaluable contributions that made this work possible; the members of the Lewis group, both past and present, all of whom have been friendly and talented individuals; and all my colleagues at the University of Chicago, the aforementioned included but many others as well, whose conversations, jokes, and commiserations have been essential to my success.

In the lattermost category fall Dr. Landon Durak, Ken Ellis-Guardiola, Mary Andorfer, Dr. Nathan La Porte, Michael Boles, Jeff Montgomery, Anthony Martinez, Noumaan Shamsi, and many others that I exclude only for the sake of brevity. Graduate school has given me the opportunity to experience a wide range of interesting science, and as a bonus an experiment here and there has given me a positive result, but even more impressive is the diversity of interesting and exceptionally talented people to whom I have been exposed. Leaving the friends I have made will be the hardest part of graduating, but I know we will remain friends for many years to come.

The community at the University of Chicago has always been welcoming and intellectually stimulating. I would like to thank the students, professors, instrumental staff, and departmental staff for always being generous with their time to help and instruct me. I would especially like to thank the members of my thesis committee, Professors Chuan He and Bryan Dickinson, for aiding me in this final step of my graduate career. I would also like to thank all the members of the CBC

and CBI, especially Professor Joe Piccirilli, for the learning opportunities they have offered me over the years.

I am also in debt to the friends and family outside of the University who have offered me an extraordinary amount of support over the years. They have accommodated my irregular schedule and allowed me to vent my frustrations with an astonishing degree of patience. I extend my gratitude to my parents, Laurie Braun and John Payne, and my brothers, David and Paul, for the nearly three decades of love they have shown me. I would also like to thank Uly for his resolute calm when I most needed it. I also thank my friends, especially James Mitchell, who listened to my many complaints and still supported me despite lacking any familial obligation to do so. I would also like to thank my friends Mike and Dexter, both of whom showed me nothing but kindness for the regrettably short time they were given. You are both missed.

Lastly, I thank my wife, Dana. Our relationship has seen profound change over the course of my graduate career, as we've gone from dating, to engaged, to married, but her patience, understanding, and love have never wavered. I don't know how I'll ever repay the debt I owe to her, but thankfully I have the rest of my life to try.

ABSTRACT

This dissertation describes the engineering of halogenating enzymes (halogenases) to facilitate the site-selective functionalization of bioactive molecules. While most small-molecule methods of halogenation have selectivities dictated by the electronics of the substrate, there exist enzymes that are able to functionalize (an albeit limited set of) substrates at significantly electronically disfavored positions. Prior to the work described within this dissertation, the extremely limited protein engineering efforts performed on these enzymes had been unable to increase their activities on substrates larger than their native substrates. The work described herein entails the first such engineering of one halogenase in particular, RebH, to allow the functionalization of a range of large, bioactive substrates with selectivities not observed with small-molecule strategies.

Chapter I provides a background to halogenases and biocatalysis. Halogenation is a vital process in the production of a significant percentage of pharmaceuticals and agrochemicals, both to facilitate chemical synthesis and to tailor the biological properties of the products. Despite this, there is still a lack of methodology to install halogen substituents at electronically disfavored positions on many substrates of interest. A number of halogenases, some with selectivities for electronically disfavored positions, have been reported and characterized, and in this chapter a summary of the existing halogenase literature will be provided. A unique advantage of using enzymes as catalysts rather than small molecules is the ability to genetically introduce mutations to alter the properties of your catalyst; by doing so iteratively, one can recapitulate natural selection toward the goals of the researchers on laboratory timescales in a process known as directed evolution. While this process had not been applied to halogenases prior to the work described in this dissertation, it is a well-developed technique with years of use and success. A brief

introduction to directed evolution will also be provided in this chapter as a prelude to the description of its application to halogenases.

Chapter II describes the enhancement of the expression of soluble RebH in *E. coli* and the exploration of the scope of wild-type RebH. A number of halogenases were explored in *E. coli* and RebH was selected for efforts to improve its solubility. Once this was accomplished, it allowed wild-type RebH to be tested on preparative scales on a range of unnatural substrates. Activity on several unnatural substrates was observed with high selectivity for electronically disfavored positions. The work described in this chapter was necessary to allow the directed evolution described in the following chapter.

Chapter III describes the directed evolution of RebH. The improvement of the thermostability of RebH by directed evolution will be briefly summarized, which provided the first step of the expansion of the substrate scope of RebH toward large, bioactive molecules. The directed evolution of RebH to expand its substrate scope, as well as preparative-scale bioconversions of a range of important bioactive substrates and kinetic analysis of the evolution lineage, will then be described. The engineered RebH variants developed in this process have proven to be broadly useful for a diverse range of substrates, and the screening of a library of compounds will be described to illustrate this point. Lastly, halogenation provides for a useful synthetic handle for further functionalization, and the development of a facile procedure for subsequent cross-coupling of crude products from bioconversions will be briefly discussed.

Chapter IV describes the use of RebH variants to perform enantioselective halogenations *via* desymmetrization on a range of symmetric substrates. Symmetric, prochiral dianiline substrates were synthesized, halogenated using the engineered RebH variants, the development of which was described in Chapter III, and the monochlorinated products found to be

enantioenriched. It was then demonstrated that protein engineering can be used to alter the yield and enantioselectivity of these bioconversions. The work described in this last chapter marks the first reported application of this class of enzymes in enantioselective catalysis.

PREFACE

The compounds described in each chapter of this dissertation are numbered independently; a given compound may therefore have a different number in different chapters. All experimental details, references, and notes for a given chapter are included at the end of said chapter; a given reference may accordingly have a different number in different chapters as well. All NMR spectra are included in the appendices: the NMR spectra for Chapter Two are included in Appendix I; the NMR spectra for Chapter Three are include in Appendix II; and the NMR spectra for Chapter Four are included in Appendix III.

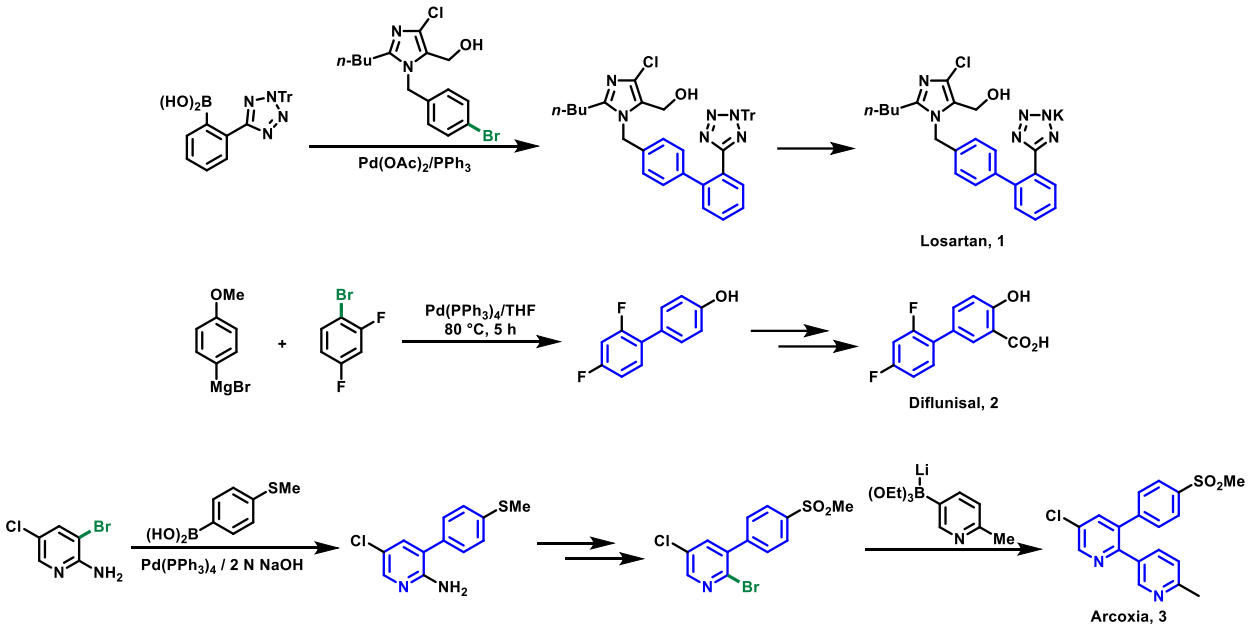
CHAPTER I

AN INTRODUCTION TO HALOGENASES AND DIRECTED EVOLUTION

1.1 THE IMPORTANCE OF SITE-SELECTIVE HALOGENATION

Halogenation plays a vital role in the development of bioactive molecules; for example, an estimated one-quarter of all pharmaceuticals and agrochemicals are halogenated.[1] Although already impressive that such a significant fraction of the compounds we depend on for our health and for agriculture directly possess a halogen substituent in the final product, halogenation is also involved in the production of an even greater number of compounds. Halogenated intermediates are ubiquitous throughout synthetic organic chemistry, with one very important example being the use of aryl and alkyl halides in cross-coupling chemistry.[2] Especially prevalent is the use of aryl halides to form biaryl compounds, a motif that is widely represented in pharmaceuticals[3] displaying a wide range of biological activities.[4] Three of many examples of marketed drugs possessing a biaryl motif that can be prepared through cross-coupling from an aryl halide are the antihypertensive losartan (**1**), the NSAID diflunisal (**2**), and the COX-2 inhibitor etoricoxib (Arcoxia, **3**) (Scheme 1.1).[5] While these examples demonstrate the utility of aryl halides on the industrial scale, their use by medicinal chemists in drug development is also ubiquitous. An analysis of the publication output in 2008 of three large pharmaceutical companies, Pfizer, GlaxoSmithKline, and AstraZeneca, showed that 30% of compounds produced contain an aryl chloride, nearly 40% contain a biaryl motif, and Suzuki and Sonogoshira couplings utilizing halides are two of the most common reactions performed.[6]

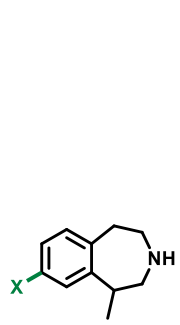
Scheme 1.1. Pharmaceutical syntheses using aryl halides.[a]



[a] Resultant biaryl motifs are shown in blue, functionalized aryl halide C-X bonds are shown in green.

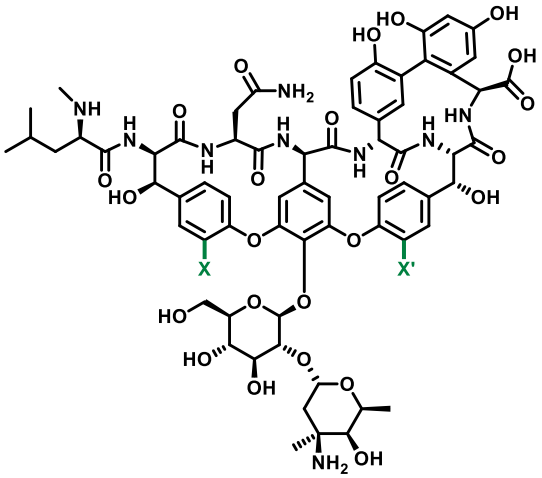
As first mentioned, halogenation is not only significant for the role it plays in the syntheses of many important compounds we use, but halogen substituents are present in a large fraction of the final products, as well. This is, in part, due to the profound change that halogenation can impart on the biological activity of a compound (Figure 1.1).

Figure 1.1. Examples of the impact of halogenation on bioactivity.



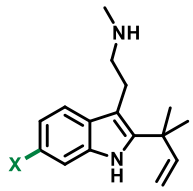
lorcaserin, 4

substituent	$\frac{EC_{50}(5HT_{2C})^{[a]}}{EC_{50}(5HT_{2B})}$
X = H	13
Cl	79



vancomycin, 5

substituent	AC/AC(H,H)[b]
X = H, X' = H	1
Cl, X' = H	2.4
Cl, X' = Cl	6.4



desformylflustrabromine, 6

substituent	IC_{50}[c]
X = H	negligible
Br	174 μM

[a] Activites shown for lorcaserin are the ratios of the half-maximal effective concentrations for the $5HT_{2C}$ versus the $5HT_{2B}$ receptor subtypes.[9] [b] Activities shown for vancomycin are the ratios of the affinity coefficients for an Ac-D-ala-D-ala peptide (the motif targeted in bacterial cells) of the indicated compound to that of the X = H, X' = H compound.[7] [c] Activities shown for desformylflustrabromine are half-maximal inhibitory concentrations for biofilm formation by *E. coli*.[8]

In many of these examples, both the presence and the location of the halogen are necessary for the observed change in biological activity. For example, the anti-obesity drug lorcaserin (**4**) has a significantly increased ratio of activity on the $5HT_{2C}$ versus the $5HT_{2B}$ receptor when chlorinated at the 8-position compared with the unchlorinated analogue, but when the chlorine substituent is at the 7-position, this is diminished by four-fold.[9] As the $5HT_{2C}$ receptor is involved in regulating appetite, while the $5HT_{2B}$ receptor regulates heart rate, it is vital that the final drug

have as high as possible of a ratio of activity between these two serotonin receptor subtypes in order to minimize undesirable cardiac side effects. Similarly, the 6- and 9-chlorinated analogues had approximately 100-fold decreased activity on the $5HT_{2C}$ receptor. Therefore the ability to produce solely the 8-chlorinated product, rather than the 6-, 7-, or 9-chlorinated analogues, is essential. Differing specificity for receptor subtypes indicates that the presence of the chlorine affects substrate binding and thus the pharmacodynamics of lorcaserin. The ability of halogenation to affect pharmacokinetics is also well documented. In a study of over 220,000 compounds conducted by Pfizer, it was found that arene chlorination is able to predictably alter the rate of liver microsomal clearance of compounds.[10] Based on the site of chlorination on a given molecule, the metabolic clearance could be either increased or decreased relative to the unchlorinated compound. Thus one can envision that a medicinal chemist may wish to site-selectively chlorinate a drug candidate in order to decrease its metabolic clearance to create a more effective analogue, but lacking such a site-selective method, would instead create regioisomers with increased metabolic clearance.

Most small-molecule methods of halogenation are guided solely by the inherent electronic properties of the substrate. Methods using electrophilic aromatic substitution are frequently used, which in addition to lacking the ability to selectively halogenate relatively electronically deactivated positions also requires harsh reaction conditions, using chlorinated solvents, oxidized halogen sources, transition metal catalysts, and, in some cases, high temperatures.[11] A number of "green" methods of varying utility have been developed to overcome the harsh reaction conditions, but the lack of regioselectivity persists with these methods. There have been some attempts to develop site-specific small-molecule halogenation catalysts, but these methods, such as rhodium-catalyzed bromination of arenes, depend on the installation of directing groups for their site-

selectivity.[12] This use of directing groups adds steps and waste to a synthetic scheme, can only access sites *ortho*- to the directing group, and now require the site-specific installation of a directing group, which is frequently just as untenable of a problem. For example, to selectively synthesize 7-chloroindole using a recently reported procedure, one must first install a silyl directing group on the 1-position nitrogen using a ruthenium catalyst, then borylate the 7-position using an iridium catalyst, then exchange the iridium using copper(II) chloride.[13] This method requires three separate transition metals to access a single position of indole, and is not suitable to functionalize indole natural products that already contain substituents off the 1-position, as many natural products do. Methods such as these that rely on the electronic properties of the substrate or on the installation of directing groups are relying on *substrate* control. A method that instead used *catalyst* control to guide halogenation to specific sites of the substrate would prove to be a great boon to synthetic chemists.

1.2 AN INTRODUCTION TO HALOGENASES

In enzymes, nature has developed an exquisite means to achieve the catalyst control that would be so desirable for synthetic applications. By genetically encoding combinations of a limited set of amino acids, enzymes with a wide range of functionalities have been created, often with a well-defined 3D structure that allows for precise orientation of substrates and control over their reactivity. Given that these enzymes are genetically encoded, tailoring the catalyst to a new set of conditions is greatly simplified compared with the operations required to redesign a small-molecule catalyst; rather than develop an entirely new synthetic scheme for each small-molecule catalyst analogue one would wish to screen, mutations simply need to be introduced to produce perturbations in the enzyme catalyst. This is part of the process of natural selection, which has

played a major role in the development of the cornucopia of nature's enzymes; the recapitulation of this process in the laboratory setting will be discussed in greater detail in section 1.3.

Not all enzymes provide the sort of site-selectivity discussed in section 1.1; different classes of enzymes can effect similar transformations using distinct mechanisms, and only some of these mechanisms may allow for the particular control a chemist desires. When examining halogenases, enzymes that are able to catalyze the conversion of a C-H bond to a C-X bond, four main classes are apparent: S-adenosylmethionine (SAM) dependent halogenases; iron or vanadium dependent haloperoxidases; α-ketoglutarate (αKG) dependent halogenases (ADHs); and flavin dependent halogenases (FDHs). The latter three classes all proceed through oxidation of the halide to generate the reactive species, which then halogenates the substrate through an electrophilic mechanism, whereas the SAM dependent halogenases instead use unoxidized halide directly in a nucleophilic mechanism. These halogenases are also notable for their capability to perform fluorination to form C-F bonds, a property that is not observed in any of the other three classes. Because of the difference in mechanism and the unique value of fluorination, the SAM dependent halogenases will be discussed first.

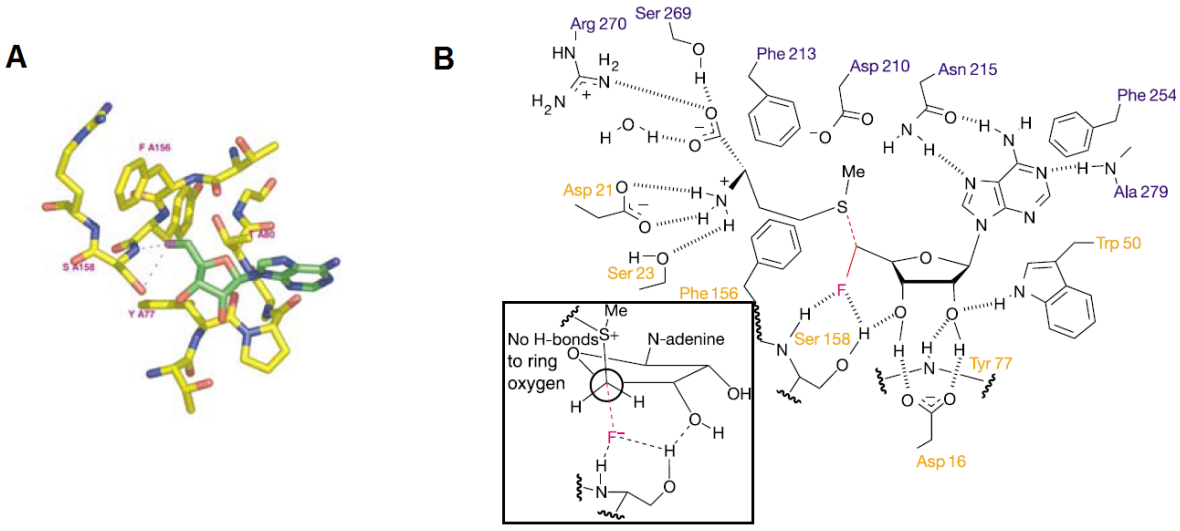
1.2.1 SAM DEPENDENT HALOGENASES

An estimated 20-30% of all drugs and agrochemicals contain at least one fluorine molecule.[14] The C-F bond is widely used as isostere for the C-OH and C=O bonds and the introduction of fluorine in place of oxygen thus does not tend to disrupt the pharmacodynamics of a drug.[14] The C-F bond, in contrast with the C-OH bond however, is strong enough to avoid metabolic cleavage, and fluorines are thus commonly introduced in order to prolong the half-life of a drug candidate.[15] Despite the abundance of organofluorine compounds in drugs, relatively few organofluorine natural products have been discovered. While approximately 2000 each of natural

organochlorine and organobromine compounds had been discovered as of 2003, only approximately 100 organofluorine compounds had been discovered by the same date.[16] This is most likely due to the fact that the fluoride ion in aqueous media is highly solvated, and a significant entropic barrier must be overcome to desolvate this ion to generate the naked, nucleophilic species.[17] Perhaps as a result of the difficult of this task, relatively few fluorinases are known with little variation between them, as is discussed below.

Despite the significant barrier of fluoride desolvation in water, *Streptomyces cattleya* has been observed to form carbon-fluorine bonds using inorganic fluoride as its source.[18] When grown in the presence of fluoride, this organism secretes fluoroacetate and 4-fluorothreonine, and this activity was pinpointed to a single enzyme that generates 5'-fluoro-5'-deoxyadenosine (5'-FDA) and L-methionine from S-adenosyl-L-methionine (SAM) and fluoride. This 5'-fluoro-5'-deoxyadenosine synthase (5'-FDAS) was successfully crystallized, and its structure was determined, which shed light on how it overcomes the aforementioned entropic barrier of desolvation of the fluoride ion.[19] An active site serine forms two hydrogen bonds with fluoride, one via the backbone amide and a second via the side chain hydroxyl (bifurcated with the adenosyl 3'-hydroxyl), thus taking the place of the solvating water molecules (Figure 1.2). The pocket in which fluorine binds has a small radius of 1.4-1.6 Å, which helps explain the preference for fluoride over water or other halides. 5'-FDAS has been found to accept chloride as well to form 5'-chloro-5'-deoxyadenosine (5'-ClDA) when the assay is performed in the presence of an L-amino acid oxidase to remove methionine and pull the reaction equilibrium toward the product.[20] A crystal structure of this enzyme with bound 5'-ClDA revealed that this is accomplished despite the small fluorine binding pocket by displacing the chlorine by 1.3 Å relative to the fluorine.

Figure 1.2. Representations of the 5'-FDAS active site.[a]

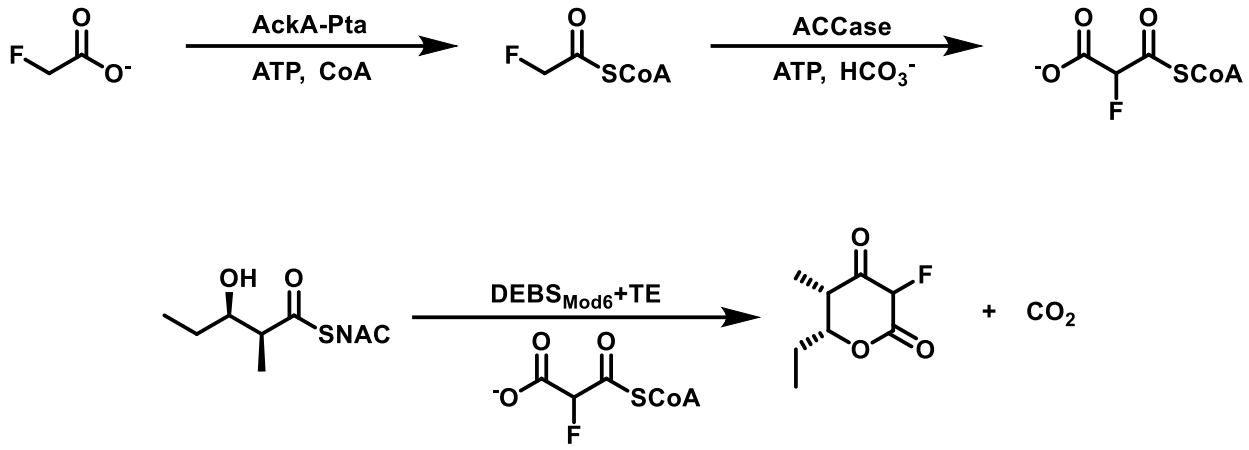


[a] Both images are from reference 18. A) Pymol structure of 5'-FDAS active site (PDB: 1RQR) showing interactions of S158 side chain hydroxyl (red) and backbone amide (blue) with 5'-FDA fluorine (mauve). B) Illustrative representation of same active site with bound methionine also shown.

While the tremendous importance of fluorination would lead one to suspect that this fluorinase would be the immediate target of engineering to expand its substrate scope to allow late-stage functionalization of drug candidates, 5'-FDAS has several significant limitations that hinder such a prospect. Attempts have been made to explore the substrate scope of 5'-FDAS, and 2'-deoxy-5'-FDA (2'd-FDA), as well as the chlorinated analogue, 2'-deoxy-5'ClDA (2'd-ClDA) were both accepted (these reactions were probed in the reverse direction, owing to the ease of synthesis of 5'-FDA analogues relative to analogues of SAM).[21] But even with the relatively minor perturbation of removing the adenosyl 2'-hydroxyl, a 90% decrease in the rate of 5'-FDAS was observed with 2'd-FDA versus 5'-FDA. Another group found via structural analysis that the 2' position is the most likely position at which FDAS is likely to accept modifications and

accordingly tested 2'-alkynyl substrates to allow for post-fluorination functionalization.[22] It is probable that the very precise organization of the substrate required, e.g. the bifurcated hydrogen bond formed between the fluoride, active site serine side chain hydroxyl, and the 5'-FDA adenosyl 3'-hydroxyl group, will make the task of expanding the substrate scope very challenging. Furthermore, the activity of 5'-FDAS on its native substrate possesses a k_{cat} of only 0.07 min^{-1}, and further decreases in this activity, such as that seen with 2'd-FDA, will only further complicate matters. In order to generate fluorinated analogues of complex, bioactive molecules, one group instead recently reported the modification of a polyketide biosynthetic pathway, 6-deoxyerythronolide B synthase (DEBS), such that fluorinated substrates can be accepted (Scheme 1.2).[23] A fluorinated analogue of an advanced intermediate in this pathway, 2-fluoro-2-desmethyltriketide lactone, was observed, indicating the feasibility of this approach to ultimately producing fluorinated analogues of complex polyketide natural products (and potentially from fed fluoride, once the biosynthetic pathway for fluoroacetyl-CoA can also be incorporated). While this is still significantly less desirable than direct, late-stage enzymatic functionalization of complex natural products since new pathways will have to be constructed for each analogue desired and the structural diversity of these analogues will likely be limited, this may still prove to provide novel, synthetically intractable fluorinated drug analogues.

Scheme 1.2. Pathway for integration of fluoroacetic acid into desmethyltriketide lactone.[a]

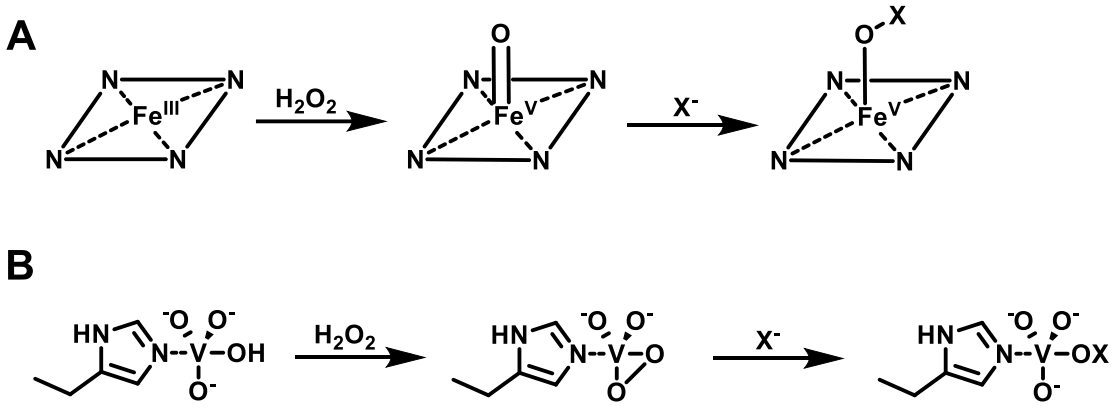


[a] Adapted from reference 23. AckA-Pta = acetate kinase-phosphotransacetylase pair; ACCase = acetyl-CoA carboxylase; SNAC = Nacetylcysteamine thioester; $DEBS_{Mod6}$+TE = sixth module of 6-deoxyerythronolide B synthase including terminal thioesterase.

1.2.2 HALOPEROXIDASES

The oldest discovered class of halogenases are the haloperoxidases, which utilize either iron- or vanadium-based cofactors to generate the active halogenating species. Haloperoxidases, like ADHs and FDHs, use an oxidative mechanism involving hydrogen peroxide to generate HOX (X = Cl, Br, I), an electrophilic species, in contrast to the mechanism described above for SAM dependent halogenases, in which the halide is used directed for nucleophilic attack. The reaction with H_2O_2 generates an iron(IV)-oxo heme intermediate (there is speculation that non-heme iron haloperoxidases exist, but the vast majority have iron in a heme center) or a vanadium peroxide species (without change in the oxidation state of the vanadium atom itself), which then reacts with halide to ultimately generate hypohalous acid (Scheme 1.3).[24] Haloperoxidases can be capable of chlorination, bromination, and iodination, but the electronegativity of fluorine prevents oxidation by hydrogen peroxide.[25]

Scheme 1.3. Partial mechanisms of haloperoxidases.[a]



[a] Shown specifically is the generation of hypohalite (-OX) reactive intermediates in the active sites of A) heme- and B) vanadium-dependent haloperoxidases; adapted from reference 24.

For the purpose of engineering a halogenase capable of selective halogenation of relatively electronically deactivated positions of complex, bioactive molecules, the vast majority of reported haloperoxidases lack the potential for such site-selectivity. This is because the hypohalous acid that is generated by the haloperoxidase is typically not held in the active site, but instead diffuses freely out of the enzyme to react with the substrate(s) in solution. As such, the site-selectivity of haloperoxidases is determined simply by the inherent electronics of the substrate, as is seen with small molecule halogenation reagents. Haloperoxidases thus do not provide improved selectivity over these reagents. Recent reports of the discovery of a stereoselective[26] and a regioselective[27] vanadium chloroperoxidase hold promise for future efforts to engineer variants with altered selectivity.

1.2.3 α-KETOGLUTARATE DEPENDENT HALOGENASES (ADHs)

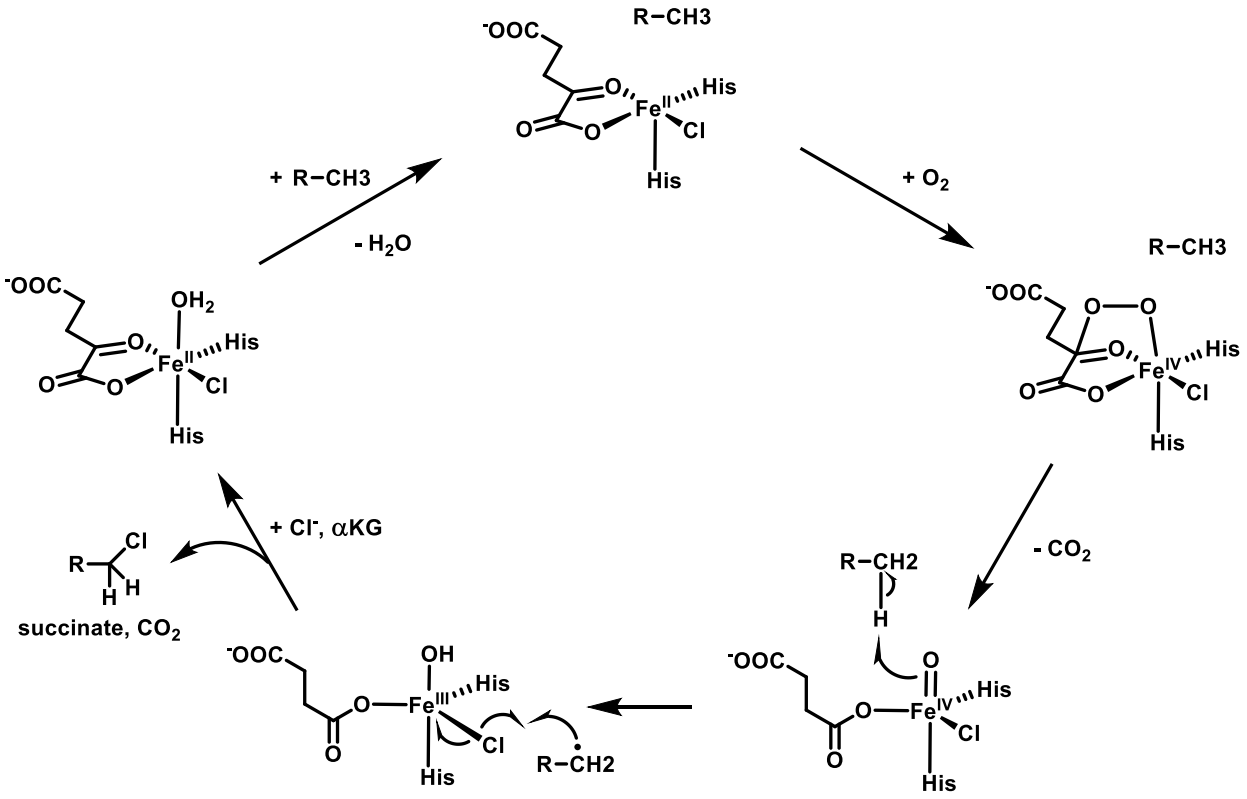
ADHs are the most recently discovered class of halogenases, with the first member, CmaB, reported in 2005.[28] As the heme-dependent haloperoxidases are analogous to the well-studied

cytochrome P450s, which are heme-dependent monooxygenases, ADHs are analogous to another common class of microbial oxygenases, which utilize non-heme iron and decarboxylate an α-ketoglutarate cosubstrate. Studies of biosynthetic gene clusters for halogenated natural products revealed that numerous natural products lacked homologs to the already reported flavin-dependent halogenases (discussed in section 1.2.4), and that the sites of halogenation on many of these products were aliphatic, rather than aromatic carbons. Given the lack of site-selectivity of the haloperoxidases and the fact that the FDHs had only been known to halogenated relatively reactive aromatic carbons, the discovery of the ADHs filled in the unexplained gap left by site-selective halogenation of unreactive aliphatic carbon centers.

ADHs, like FDHs (although in a different mechanism which will be discussed below, in section 1.2.4), are O_2 dependent, using O_2 and release of CO_2 from the αKG cofactor to oxidize Fe(II) to a reactive Fe(IV) oxo species.[24] Similar to P450s, the Fe(IV) oxo species then chlorinates the substrate through a radical rebound mechanism, ultimately releasing water and succinate as byproducts (Scheme 1.4).[29] Since both non-heme iron α-ketoglutarate dependent hydroxylases and halogenases proceed through such a radical rebound mechanism with an Fe(III)-OH species, it is initially puzzling why no halogenation is observed with the hydroxylases and no hydroxylation with the halogenases. The lower reduction potential of the chlorine radical ($Cl\bullet + e^- \rightarrow Cl^-$, 1.36 V) relative to the hydroxyl radical ($HO\bullet + e^- \rightarrow HO^-$, 2.02 V) may explain most of the preference of the halogenase for halogenation,[30] and an explanation for the absence of halogenation by hydroxylases is suggested by the crystal structure of one such ADH, SyrB2. SyrB2 is from the biosynthetic gene cluster for syringomycin E found in *Pseudomonas syringae*, and this ADH chlorinates the terminal methyl of the side chain of a threonine residue.[31] While the non-heme iron α-ketoglutarate dependent hydroxylases have the side chain of an aspartic acid or glutamic acid

residue as one of the ligands on iron, such a residue is conspicuously absent from the SyrB2 crystal structure, instead replaced with an alanine that leaves suitable space for a chloride. Further corroborating the hypothesis that chloride acts to complete the ligand set on iron in ADHs, it has been observed that apo-SyrB2 binds Fe^{2+} three orders of magnitude more tightly when in the presence of chloride.[24]

Scheme 1.4. Catalytic mechanism of ADHs.[a]



[a] Adapted from references 24 and 29. In αKG dependent hydroxylases the chloride is replaced with an aspartic or glutamic acid residue and radical rebound occurs on the iron-oxygen bond.

Given the site-selectivity of ADHs, combined with their ability to halogenate aliphatic positions of substrates, one might think that ADHs would be the logical first choice for an engineering effort directed toward the site-selective halogenation of complex substrates of interest

at relatively electronically disfavored positions. Significant limitations to employing ADHs in such a manner exist. First and foremost, until recently, all reported ADHs were only known to possess activity on substrates tethered to a carrier protein via a phosphopantatheine arm. Such a restriction requires that the carrier protein is also capable of loading the substrate of interest, which limits potential scope to substrates with an acyl group that can be conjugated to the phosphopntatheine arm, and may necessitate the engineering not only of the ADH of interest, but also its conjugate carrier protein. In the case of the aforementioned ADH SyrB2, the substrate scope of its conjugate carrier protein, SyrB1, was explored. The relatively small change of the substrate from L-threonine to L-serine, which differs by only a single methyl group, resulted in a 60-fold decrease in the catalytic efficiency for this process,[32] indicating that the scope of the carrier protein would be a significant limitation to expanding the scope of the ADH in this case. The simultaneous engineering of multiple proteins may increase the difficulty of the endeavor exponentially. Furthermore, the requirement of using a loaded carrier protein in the reaction assay adds practical difficult to the screening step of the directed evolution cycle, which is already the major bottleneck of many engineering efforts (see section 1.3 below).

Because of the limitations posed by the necessity of the carrier protein, the recent report of the first example of an ADH that is able to act on a substrate not bound to any carrier protein is tremendously encouraging for protein engineering efforts directed at ADHs.[33] This ADH, WelO5, is from the pathway for the biosynthesis of welwitindolinone in the algae *Hapalosiphon welwitschii*. WelO5 is capable of the halogenation of an unactivated methylene on the substrate 12-*epi*-Fischerindole U to form 12-*epi*-Fischerindole G, as well as the analogous halogenation of the related substrate 12-*epi*-Hapalindole C to form 12-*epi*-Hapalindole E. Welwitindolinones have been a popular target for total syntheses, occupying the efforts of at least 60 researchers,[34] and the

potential to use engineered WelO5 to biosynthetically create synthetically intractable derivatives of these compounds is an exciting prospect. The discovery of WelO5, and potential related ADHs that function on free-standing substrates, will likely prove to be pivotal for the engineering of halogenases capable of site-selective halogenation of alkylic positions on bioactive substrates.

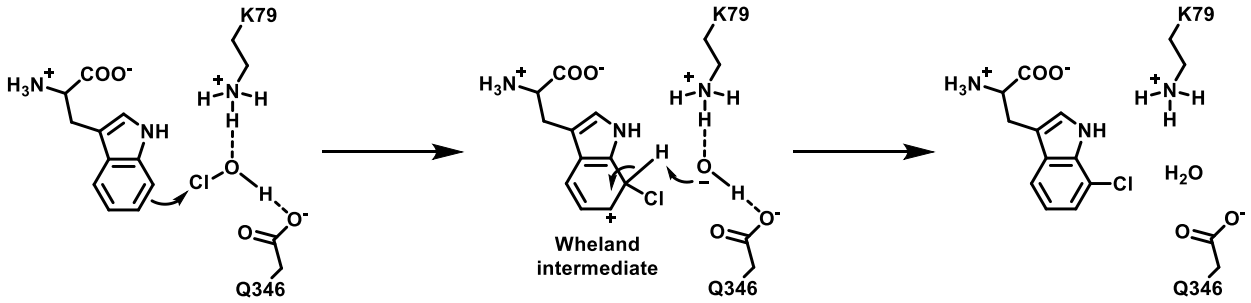
1.2.4 FLAVIN DEPENDENT HALOGENASES (FDHs)

The last class of halogenases to be discussed are the flavin dependent halogenases (FDHs), the engineering of which will be the main focus of this thesis. Like the ADHs, FDHs are O_2 dependent, but while ADHs use an iron cofactor, FDHs instead use an FAD cofactor to form an FADH-OOH peroxo species.[24] This reaction requires the reduced form of FAD, $FADH_2$, which in turn is regenerated by a separate NADH dependent flavin reductase. The FADH-OOH peroxo species then reacts with halide (both Cl^- and Br^- have been reported to be accepted by several FDHs) to generate an equivalent of hypohalous acid (HOX), water, and the regenerated FAD cofactor. The hypohalous acid then travels to the active site of the FDH, a journey of approximately 10 Å in some cases, where it then interacts with a conserved active site lysine (Scheme 1.5).[35]

The precise role of the hypohalous acid is, at this point, debated, as the hypohalous acid may merely is coordinate to the lysine and is still the active halogenating agent (Scheme 1.5),[35] or the hypohalous acid might react with the lysine to form a haloamine species which is responsible for halogenation.[36] It has been argued that a haloamine would not have the necessary redox potential to perform the halogenations reported by FDHs and thus that hypohalous acid must be the active species; however, the protonation state of the theoretical haloamine in the enzyme active site has not been determined, and if protonated would be significantly more active than the calculations performed would indicate, even ignoring further mitigating effects from protein tertiary structure.[35] In support of the theory that a haloamine is formed from the hypohalous acid,

Walsh and coworkers were able to isolate and characterize a long-lived, catalytically competent haloamine intermediate of RebH, an L-tryptophan 7-halogenase from the rebeccamycin biosynthetic pathway in *Lechevalieria aerocolinogenes*.[36] Further discussion of FDHs need not concern itself with the validity of either of these theories – the fact that the halogenating agent is in some way held by the active site lysine, whether that is through coordination with or reaction with said lysine, is the crucial factor of significance with this class of enzymes.

Scheme 1.5. One proposed mechanism[a] of the FDH PrnA.[35]



[a] The hypohalous acid is depicted as the active halogenating agent in this scheme.

This restriction of the halogenating agent by the active site lysine allows for the site-specific halogenation of substrates at electronically disfavored positions. While the position may be electronically disfavored relative to other positions on the molecule, it must still be sufficiently activated to undergo halogenation by a halenium intermediate, and it is still unclear what the maximum difference in reactivities of different sites can be overridden. Although the hypohalous acid/haloamine ultimately halogenates the substrate through electrophilic aromatic substitution, the same mechanism employed by the non-site-specific small molecule halogenation agents described in section 1.1, the lysine is able to coordinate the halogenating agent proximal to only a single position of the substrate, and thus only that position is allowed to access the halogenating agent. This is analogous to the site-selectivity seen with ADHs, in which the substrate is held in a

specific conformation proximal to the Fe(IV) oxo species, except that the halogenating agent travels away from the point of generation to a separate substrate binding pocket. This partial separation of some of the catalytic machinery, namely the flavin binding site, from the substrate binding pocket may facilitate protein engineering efforts that aim to affect substrate binding without disrupting said catalytic machinery.

To date, most characterized FDHs that are able to functionalize a freestanding substrate, i.e. one not bound to a peptide or acyl carrier protein, utilize L-tryptophan as their native substrate.[37] These include the aforementioned RebH and PrnA, L-tryptophan 7-halogenases, as well as other halogenases capable of selective 5- or 6-halogenation of L-tryptophan (see Table 2.1). In addition to these tryptophan halogenases, some phenolic halogenases have been characterized that also function on freestanding substrates. For example, ChlA has been found to dichlorinate the substrate (2,4,6-trihydroxyphenyl)-1-hexan-1-one in the biosynthesis of DIF-1.[38] Furthermore, the halogenase TiaM has been found to halogenate the orsellinic acid moiety of a large substrate in the course of tiacumicin B biosynthesis,[39] while the halogenase Rdc2 was found to also halogenate an orsellinic acid moiety present in the smaller resorcylic acid lactone monocillin I in the biosynthesis of radicicol.[40] As is seen with ADHs, there exist numerous FDHs that are only known to functionalize substrates that are tethered to peptides or acyl carrier proteins.[41] For the same reasons discussed for ADHs, engineering of FDHs that require the substrate is bound to a peptide or acyl carrier protein is likely to be significantly more difficult than engineering FDHs capable of functionalizing freestanding substrates. Of the FDHs characterized to functionalize freestanding substrates, the L-tryptophan halogenases are the best studied.

Rebeccamycin halogenase (RebH) nicely illustrates the power of substrate binding to control FDH selectivity. As mentioned above, RebH is able to selectively functionalize the 7-position of the amino acid L-tryptophan.[42] One would expect the 2-position of L-tryptophan to be more reactive than the benzene positions, but despite this, the RebH catalyzed chlorination of tryptophan produces only 7-chloro-L-tryptophan with no trace of the significantly electronically favored 2-chloro-L-tryptophan. This is due to the aforementioned ability of FDHs to hold the substrate such that only a single C-H bond of the substrate is proximal to the halogenating agent with such specificity that not even the nearby, and electronically comparable, other benzene positions are halogenated. Furthermore, there also exist FDHs capable of selective 6-halogenation[43] and 5-halogenation[44] of L-tryptophan. Taken together, these halogenases are a powerful demonstration of the ability of FDHs to effect exquisite selective halogenation at a range of electronically disfavored positions.

While the selective halogenation of L-tryptophan is impressive, medicinal chemists are often interested in functionalizing compounds significantly larger and more complex than L-tryptophan. Although many large, complex halogenated natural products exist, the halogenation event usually occurs relatively early in the biosynthetic pathway on a relatively small intermediate (e.g. L-tryptophan), which then undergoes numerous further functionalizations to furnish the final product. Researchers that have intended to make halogenated derivatives of complex natural products have taken advantage of biosynthetic pathways that utilize similar early intermediates, thus incorporating FDHs that functionalize these intermediates to provide a halogenated derivative that is carried through the biosynthetic pathway to yield a halogenated derivative of the normally non-halogenated natural product. For example, chlorinated and brominated analogues of a number of strictosidine derivatives, including dihydroakuammicine, ajmalicine, catharanthine, and

tabersonine, have been generated via incorporation of the aforementioned L-tryptophan 7-halogenase RebH or the L-tryptophan 5-halogenase PyrH into their biosynthetic pathways.[45] Another L-tryptophan 7-halogenase, PrnA,[46] has been incorporated into the biosynthetic pathway for pacidamycin, furnishing chloropacidamycin, which was then further chemically functionalized via cross-coupling methodology to yield arylated derivatives of pacidamycin.[47] This combined chemoenzymatic functionalization of complex natural products will be discussed further in Section 3.2.4. Additionally, the site of halogenation in halogenated natural products has been altered by substituting a halogenase with different regioselectivity than the native halogenase into the biosynthetic pathway. This has been accomplished with rebeccamycin, which natively uses RebH to chlorinate the 7-position of L-tryptophan, which is then further functionalized to yield the indolocarbazole rebeccamycin.[48] Two halogenases with different regioselectivities, PyrH, an L-tryptophan 5-halogenase, and thal, an L-tryptophan 6-halogenase, were substituted for RebH, resulting in production of analogues of rebeccamycin with indole 5- and 6-chlorination, respectively.

Although powerful, these metabolic engineering strategies for producing halogenated derivatives of complex natural products have significant drawbacks. First of all, the construction of novel biosynthetic pathways is still a laborious procedure, and producing a new pathway for each desired compound is prohibitively time-consuming. Second, this is only an option for biosynthetic pathways of compounds that have been well characterized and contain an intermediate (e.g. tryptophan) that can be halogenated by FDH. Third, the steps of the biosynthetic pathway downstream must tolerate the addition of the halogen atom to their substrates. And lastly, this method is incapable of producing halogenated derivatives of non-natural molecules of interest, such as pharmaceuticals. We have already seen that FDHs are attractive candidates for site-

selective halogenation of large, bioactive molecules, as they are able to selectively functionalize substrates at electronically disfavored positions without requiring that the substrate be bound to any acyl carrier protein. The main limitation of reported FDHs is that they are only known to functionalize a relatively limited set of small molecules.[49] As such, the discovery of an FDH with the ability to functionalize a range of large molecules of interest, while still preserving regioselective halogenation at electronically disfavored positions, would be of great interest. Such a discovery has not yet occurred, leading us to speculate on the possibility of developing such a halogenase from existing FDHs.

1.3 THE USE OF DIRECTED EVOLUTION TO ENGINEER ENZYMES

As discussed in sections 1.2.1-1.2.4, the different classes of halogenases each offer distinct advantages and disadvantages relative to one another. For example, while SAM dependent halogenases are capable of fluorination, they have limited substrate scopes, with only a single substrate utilized in nature by these enzymes. While ADHs, unlike most haloperoxidases, are capable of site-selective halogenation at electronically deactivated positions, they have primarily only been reported to functionalize substrates bound to acyl carrier proteins, limiting their use on non-native substrates. It is worth noting again that the recent report of WelO5, which acts on a freestanding substrate, may open new avenues toward engineering ADHs. FDHs, on the other hand, are capable of site-selective halogenation at electronically deactivated positions on a range of small molecule substrates, but prior to the work discussed in Section 3, these substrates were largely limited to small amino acids.

While each of these enzyme classes presents its own limitations, distinct advantages are apparent over small molecule halogenation agents, as discussed in section 1.1. A major advantage is the ease with which enzymes can be modified. Enzymes are genetically encoded, meaning that

the synthesis of the enzyme can be carried out by the cellular machinery of an organism used as an expression host. Because of this, extremely complex enzymes with extraordinarily finely tuned structures (including substrate binding sites) can be produced with relative ease, provided one can competently handle the expression host. Furthermore, the ability of researchers to modify the genes encoding enzymes has rapidly increased in the past several decades, resulting in simple tools for the introduction of point mutations or random mutagenesis of a gene in the most industrially prevalent expression hosts, e.g. *Escherichia coli* or *Saccharomyces cerevisiae*. Once one has introduced the desired mutation(s) into an enzyme, its production in an expression host is typically identical to the original, i.e. parent, enzyme. While it may seem unnecessary to say the preceding sentences explicitly, the contrast between the production of enzyme catalysts versus small molecule catalysts is profound. Small molecule catalysts are synthesized by the physical manipulations of a chemist. The production of analogues of such a catalyst will require a new synthetic scheme for each catalyst desired, with different considerations potentially required for each such scheme. It may be that a researcher does not know exactly what permutation is necessary to his or her catalyst to effect a desired change in activity, but would like to access a large library of catalysts with slight variations to the parent catalyst. By modifying the gene encoding for an enzyme, one can readily produce as many such variants as one pleases; libraries of a few thousand enzyme variants are regularly screened in each step of a directed evolution project. Synthesizing even a thousand small molecule catalysts in hopes of finding one with the desired modification in activity is a far more daunting task.

Perhaps the key advantage of enzyme catalysts over small molecule catalysts is not the ease of production of analogues, but the extremely fine *tunability* of these analogues. Enzymes often range from tens to hundreds of ångströms in diameter, while the substrate may only occupy

a tiny fraction of that volume. Thus one can introduce permutations into the enzyme structure that are quite distal to the substrate binding site, but influence the overall structure of the enzyme in such a way that a fraction of an ångström shift in a key residue is effected, resulting in a slight change in binding of the substrate. This may allow for the substrate to be bound in such a way that a different position is proximal to the functionalizing species, or that a substrate with what would normally be considered a very minor permutation is preferred or excluded to a remarkable degree. For example, a theophylline responsive riboswitch has been reported that binds caffeine with 1000-fold lower affinity than theophylline, despite the fact that these substrates differ by only a single methyl group.[50]

Such selectivity is essential for a catalyst that is required to discriminate both between multiple similar positions on a single molecule and between multiple similar molecules in an environment, both of which situations arise frequently in biological contexts. Using the aforementioned L-tryptophan 7-halogenase RebH as an example, RebH must discriminate between the electronically similar 4-, 5-, 6- and 7-positions on the indole ring of L-tryptophan (all of which are significantly less electron rich than the 2-position). And since RebH is genetically encoded, its activity is tunable through the introduction of mutations to its genetic sequence. While 7-chloro-L-tryptophan is desired by the producing organism for rebeccamycin biosynthesis, a chemist might instead be interested in producing 5-chloro-L-tryptophan, or a 7-chlorinated larger indole, or even a larger indole that is chlorinated at the 7-position. Such a project can be tackled iteratively, with multiple rounds of mutation and screening to progressively enhance the desired activity; to evolve RebH to chlorinate a larger indole at the 5-position, for example, one might over several rounds increase the activity for 7-chlorination on this large indole until the desired

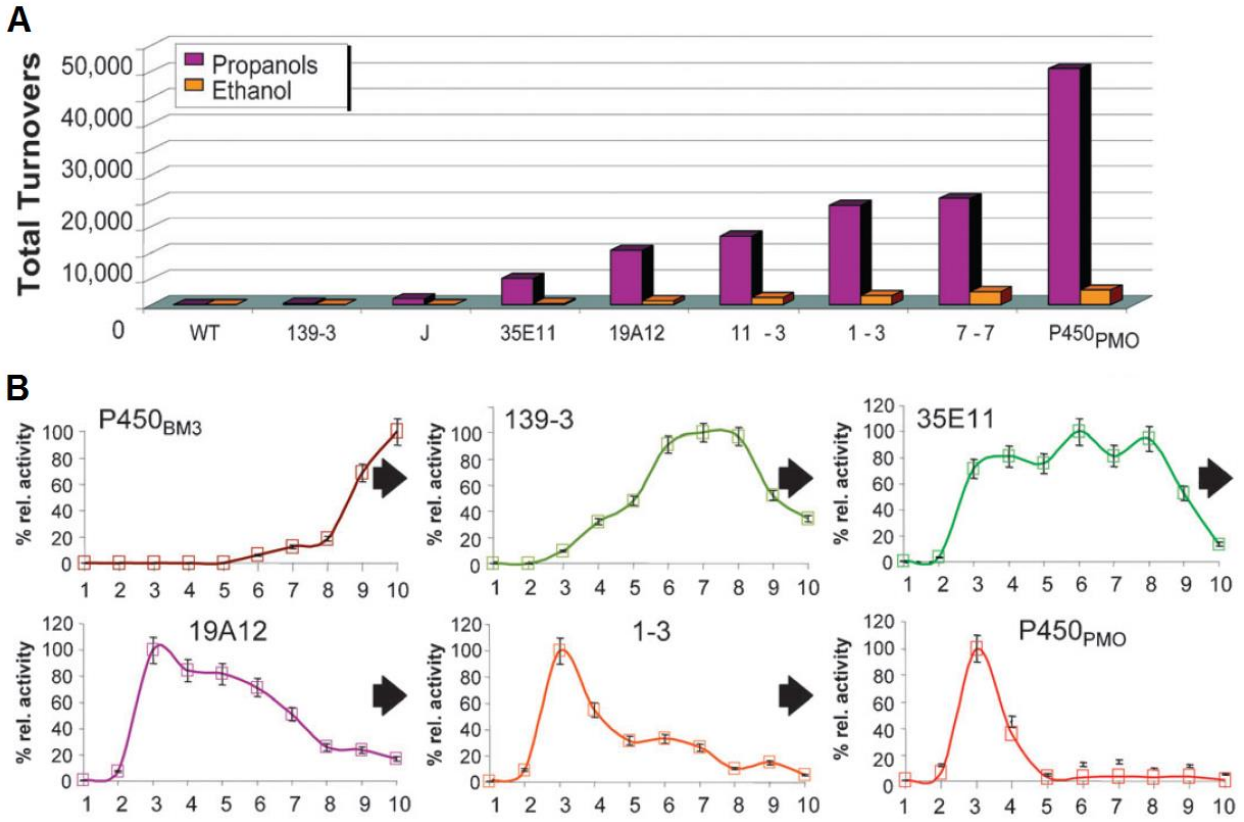
level of activity is seen, then over further rounds change the selectivity from the 7- to the 5-position.

Splitting the development of a novel biocatalyst into multiple iterative steps is immensely powerful, as it allows one not to immediately set out to look for the *best* catalyst, but simply a *better* catalyst. While the sequence space of an enzyme is prohibitively large to attempt to find a single best variant (20^n, where n is the number of amino acid residues in the enzyme - for RebH, n>500, so there are a potential $>20^{500}$ RebH variants), there likely exist many, many *better* variants within that sequence space, plenty of which likely differ by only a single amino acid substitution from the parent sequence.[51] By finding one of these better variants and then using that variant as the parent for the next round to find one that is even better yet than the first, enough rounds can be combined to effect the overall desired change in activity. If one is using random mutagenesis, in which changes are introduced throughout the entirety of the genetic sequence of the enzyme with no intended bias in the location of mutations, the crystal structure, mechanism, or any other information about the enzyme need not be known in order to simply look for gradual increases in the desired activity. However, any information, such as the mechanism or crystal structure, can inform structure-guided mutagenesis, which can potentially reduce the size of the sequence space under exploration and enhance the percentage of productive mutants developed.

As described above, using iterative rounds of mutation and screening, a project such as the evolution of a large indole 5-selective halogenase from a L-tryptophan 7-selective halogenase can be split into separate projects in which the activity and selectivity are separately altered. It may be necessary to even further break down such a project, though; perhaps the target large indole differs too significantly from L-tryptophan, such that it is not likely that 1-2 mutations will show any measurable activity on the target. In such a case, a technique known as substrate walking can be

employed to increase the activity on substrates with increasing resemblance to the target. One can envision that a target molecule differs from the native substrate by the addition of two ethyl substituents, but no activity is seen with the parent enzyme on this target substrate, making incremental increases in activity (from an initial activity of zero) difficult to find. However, sufficient activity is seen on an intermediate substrate with only one of the two ethyl groups present, and screening can first be performed on this intermediate substrate until a high-activity variant is developed that has sufficiently elevated activity on the target substrate (with both ethyl groups present) to now allow for direct screening on the target. Two drawbacks to substrate walking are immediately apparent: first, the intermediate substrates must be available, either commercially or through synthesis; and second, the technique relies on the assumption that increases in activity on the intermediate substrate will result in increases in activity on the target substrate (or the next intermediate, if multiple intermediate substrates are employed). Despite this, availability of substrates for screening is often not a limiting factor and the latter assumption often holds true; gratifyingly, substrate walking has been successfully employed for the development of numerous useful engineered enzyme variants. Examples can be found for enzyme classes as diverse as cytochrome P450 monooxygenases,[52] transaminases,[53] tRNA synthetases,[54] and monoamine oxidases.[55] One especially powerful example from the first class of enzymes, the cytochrome P450 monooxygenases, is the directed evolution of $P450_{BM3}$ by Arnold and coworkers to hydroxylate propane and ethane. $P450_{BM3}$ natively hydroxylates long-chain (C_{12}-C_{20}) fatty acids, but through a screening effort initially on octane and then on propane, variants were developed with total turnover numbers on propane (>40,000) exceeding those of wild-type $P450_{BM3}$ on its native substrates (Figure 1.3). Furthermore, the final variants also show significant activity on ethane to produce ethanol.

Figure 1.3. Evolutionary lineage of $P450_{BM3}$ ethane hydroxylases.



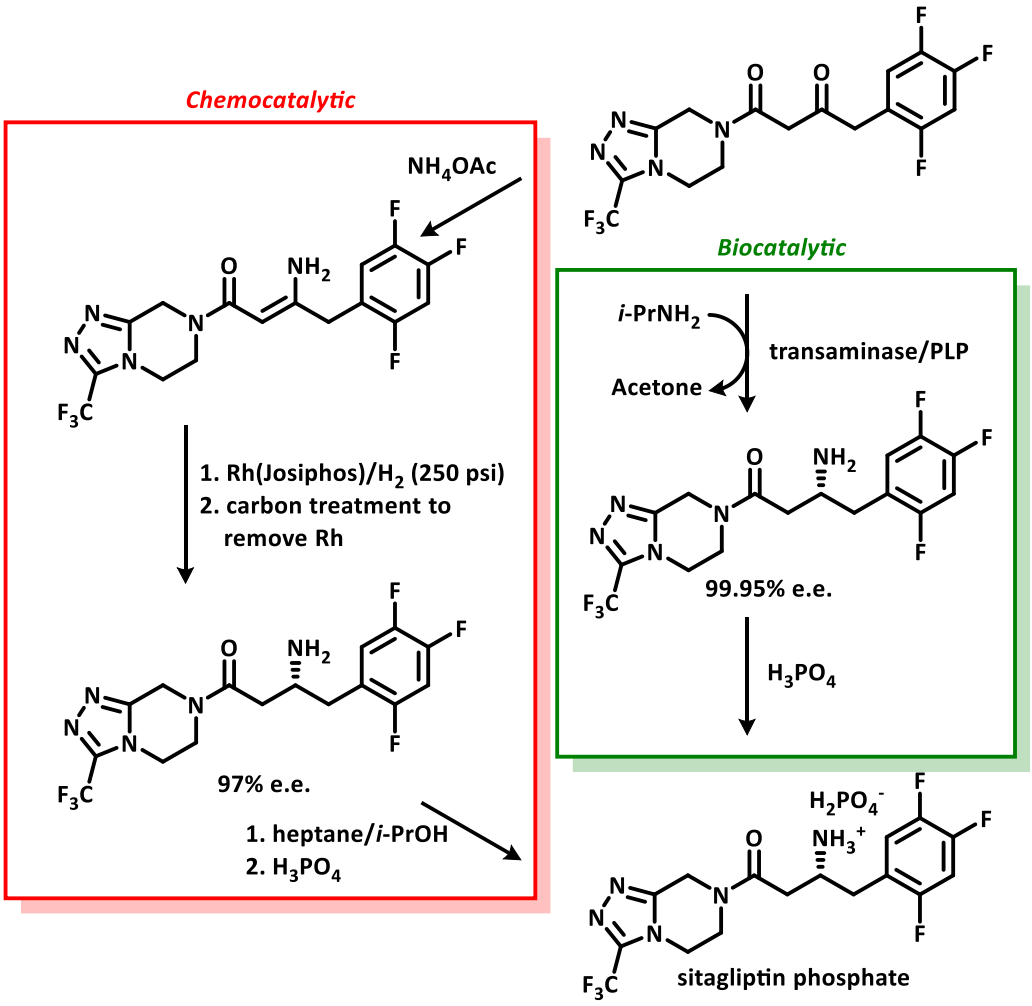
[a] Image is from reference 30. A) Change in total turnovers on propane and ethane over evolutionary lineage.[52] B) Change in substrate specificity over evolutionary lineage.[56]

Another example of substrate walking is the engineering of a transaminase for the production of sitagliptin.[53] This effort was not only a powerful demonstration of substrate walking but also stands as one of the most illustrative examples to date of the potential benefits to be derived from directed evolution. Merck had formerly employed a chemocatalytic route for the production of sitagliptin from a prositagliptin ketone using asymmetric hydrogenation catalyzed by Rh[Josiphos]. Asymmetric hydrogenation is one of the most common methods to install chirality in industrial chemistry,[57] and Merck was able to obtain an impressive 97% e.e. with this route. However, because rhodium was used as the catalyst, a subsequent carbon treatment is required to

remove residual rhodium, and while impressive, 97% e.e. is not sufficiently high for dosing to patients, so an additional recrystallization step is also necessary to further increase the e.e. Both of these added steps resulted in decreases in yield and increases in cost, so Merck collaborated with Codexis to develop an alternative biocatalytic route that could eliminate these steps.

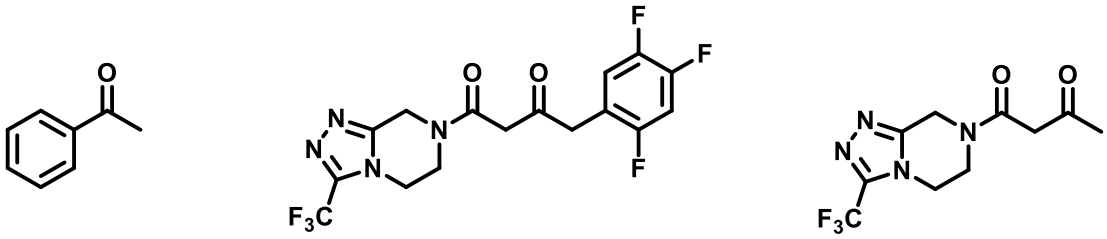
Beginning with a transaminase with no detectable activity on the target prositagliptin ketone (**7b**), structure-guided mutagenesis was used to first increase activity on a truncated version of the target substrate (**7c**, Figure 1.4). Site-saturation mutagenesis was used to uncover a single S223P mutation that afforded an 11-fold increase in activity. A combinatorial library targeting several sites that were expected to interact with the portion of the substrate that had been removed to form the truncated substrate then furnished a variant with an additional four mutations that showed the first detectable activity on the non-truncated prositagliptin ketone; it was found, though, that when these same mutations were introduced into the parent transaminase in the absence of the S223P mutation, no detectable activity was observed. Thus, substrate walking enabled a route to an enzyme variant active on the desired substrate when direct screening on this substrate would have failed to uncover the same variant.

Scheme 1.6. Alternative biocatalytic route for the production of sitagliptin.[a]



[a] Adapted from reference 53.

Figure 1.4. Substrate walking for the production of sitagliptin.



7a 7b 7c

[a] Adapted from reference 53. Native substrates of wild-type transaminases are typically limited to small, acetophenone sized substrates (7a); the target substrate (7b) was not initially accepted, so the truncated substrate (7c) was initially used.

The low activity that was observed at this point was still far from the level required for a useful enzyme for industrial scale production of sitagliptin; the enzyme variant at this stage was still affording only an approximate 0.1 turnovers per day. Not only would the rate need to be significantly increased, but it would need to see large increases in its organic solvent tolerance (>25% DMSO), loading of prositagliptin substrate and isopropyl amine cosubstrate (100 g/L and 1 M, respectively), as well as its ability to work at higher temperatures and pHs (>40 °C and pH 8.5). These changes were introduced gradually over the subsequent 10 rounds of evolution, as further mutations were introduced through site-directed mutagenesis (aided by computer design), homology models, and random mutagenesis. After the introduction of 27 mutations in a total 11 rounds of evolution (including the first on the truncated substrate), a transaminase variant was developed capable of working in the above specified conditions, furnishing product at >99.95% e.e. Remarkably, of these 27 mutations, only 10 are predicted to interact directly with the substrate; while some of the remainder are located at the predicted dimer-dimer interface (improvements in dimer stability in the increasingly harsh reaction conditions was thought to be a major contributor

to increases in turnovers), others are distal from both sites. Mutations such as these, located on residues that were not or would not have been the targets of site-directed mutagenesis that often have unknown but significant impacts on activity, offer strong testimony to the value of random mutagenesis when trying to optimize an enzyme to the highest degree possible. Ultimately, the use of the engineered transaminase variant as a biocatalytic alternative to the formerly employed chemocatalytic route for the production of sitagliptin resulted in increases in yield and productivity, while waste was decreased (including the elimination of rhodium waste through the omission of the asymmetric hydrogenation, meaning that carbon treatment is no longer necessary), as was the overall cost of production. This engineered transaminase is now used to produce sitagliptin, a multi-billion dollar drug, on industrial scale.

1.4 CONCLUSIONS

The development of the biocatalytic route for the production of sitagliptin is certainly a stunning success story for directed evolution, but it is now merely one of many.[58] Using directed evolution, any enzyme property for which one can screen (e.g. turnovers, organic solvent tolerance, optimum pH, etc.) can be optimized, and with substrate walking, one has a means by which to find variants with activity on substrates on which the parent had no detectable activity. As long as the enzyme possesses the necessary machinery for a chemical reaction, that enzyme can theoretically be engineered to perform that reaction on any substrate. Many groups are working on methods to expand the range of chemical reactions that can be performed by enzymes beyond those developed by nature, for example through the design of *de novo* enzymes[59] and the construction of artificial metalloenzymes.[60] But many of the chemical reactions for which nature has already designed catalysts still have great potential to be utilized to meet human needs. Halogenation is but one example of such a chemical reaction, but it stands as one of especially

pressing need. Because of the ubiquity of halogenation in bioactive molecules and the dearth of suitable small-molecule reactions for selective halogenation, site-selective halogenases are poised to offer great benefit to organic chemists. But first, halogenases must be engineered to work on the sorts of substrates and in the sorts of conditions in which organic chemists are interested.

1.5 REFERENCES

[1] Herrera-Rodriguez, L. N.; Khan, F.; Robins, K. T.; Meyer, H.-P. Perspectives on biotechnological halogenation Part I: Halogenated products and enzymatic halogenation. *Chem. Today* **29**, 31 (2011).

[2] Metal-Catalyzed Cross-Coupling Reactions (Eds.: Diederich, F.; Stang, P. J.), Wiley-VCH, New York, 1998.

[3] Horton, D. A.; Bourne, G. T.; Smythe, M. L. The Combinatorial Synthesis of Bicyclic Privileged Structures or Privileged Substructures. *Chem. Rev.* **103**, 893-930 (2003).

[4] Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **39**, 2887-2893 (1996).

[5] King, A. O.; Yasuda, N. Palladium-Catalyzed Cross-Coupling Reactions in the Synthesis of Pharmaceuticals. *Topics Organomet. Chem.* **6**, 205-245 (2004).

[6] Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *J. Med. Chem.* **54**, 3451-3479 (2011).

[7] 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.* **107**, 6652-6658 (1985).

[8] Bunders, C. A.; Minvielle, M. J.; Worthington, R. J.; Ortiz, M.; Cavanagh, J.; Melander, C. Intercepting Bacterial Indole Signaling with Flustramine Derivatives. *J. Am. Chem. Soc.* **133**, 20160 (2011).

[9] Smith, B. M.; Smith, J. M.; Tsai, J. H.; Schultz, J. A.; Gilson, C. A.; Estrada, S. A.; Chen, R. R.; Park, D. M.; Prieto, E. B.; Gallardo, C. S.; Sengupta, D.; Dosa, P. I.; Covel, J. A.; Ren, A.; Webb, R. R.; Beeley, N. R. A.; Martin, M.; Morgan, M.; Espitia, S.; Saldana, H. R.; Bjenning, C.; Whelan, K. T.; Grottick, A. J.; Menzaghi, F.; Thomsen, W. J. Discovery and Structure-Activity Relationship of (1*R*)-8-Chloro-2,3,4,5-tetrahydro-1-methyl-1*H*-3-benzazepine (Lorcaserin), a Selective Serotonin 5-HT_{2C} Receptor Agonist for the Treatment of Obesity. *J. Med. Chem.* **51**, 305-313 (2008).

[10] Sun, H.; Keefer, C. E.; Scott, D. O. Systematic and Pairwise Analysis of the Effects of Aromatic Halogenation and Trifluoromethyl Substitution on Human Liver Microsomal Clearance. *Drug Metab. Lett.* **5**, 232-242 (2011).

[11] Smith, K.; El-Hiti, G. A. Regioselective Control of Electrophilic Aromatic Substitution Reactions. *Curr. Org. Synth.* **1**, 253-274 (2004).

[12] Schröder, N.; Wencel-Delord, J.; Glorius, F. High-Yielding, Versatile, and Practical [Rh(III)Cp*]-Catalyzed *Ortho* Bromination and Iodination of Arenes. *J. Am. Chem. Soc.* **134**, 8298-8301 (2012).

[13] Robbins, D. W.; Boebel, T. A.; Hartwig, J. F. Iridium-Catalyzed, Silyl-Directed Borylation of Nitrogen-Containing Heterocycles. *J. Am. Chem. Soc.* **132**, 4068-4069 (2010).
[14] Müller, K.; Faeh, C.; Diederich, F. Fluorine in Pharmaceuticals: Looking Beyond Intuition. *Science* **317**, 1881-1886 (2007).
[15] Park, B. K.; Kitteringham, N. R.; O'Neill, P. M. Metabolism of Fluorine-Containing Drugs. *Annu. Rev. Pharmacol. Toxicol.* **41**, 443-470 (2001).
[16] Gribble, G. W. The diversity of naturally produced organohalogens. *Chemosphere* **52**, 289-297 (2003).
[17] Yang, Z.-Z.; Li, X. Ion Solvation in Water from Molecular Dynamics Simulation with the ABEEM/MM Force Field. *J. Phys. Chem. A* **109**, 3517-3520 (2005).
[18] Sanada, M.; Miyano, T.; Iwadere, S.; Williamson, J. M.; Arison, B. H.; Smith, J. L.; Douglas, A. W.; Liesch, J. M.; Inamine, E. Biosynthesis of fluorothreonine and fluoroacetic acid by the thienamycin producer, *Streptomyces cattleya*. *J. Antibiot.* **39**, 259-265 (1986).
[19] Dong, C.; Huang, F.; Deng, H.; Schaffrath, C.; Spencer, J. B.; O'Hagan, D.; Naismith, J. H. Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* **427**, 561-565 (2004).
[20] Deng, H.; Cobb, S. L.; McEwan, A. R.; McGlinchey, R. P.; Naismith, J. H.; O'Hagan, D.; Robinson, D. A.; Spencer, J. B. The Fluorinase from *Streptomyces cattleya* Is Also a Chlorinase. *Angew. Chem. Int. Ed.* **45**, 759-762 (2006).
[21] Cobb, S. L.; Deng, H.; McEwan, A. R.; Naismith, J. H.; O'Hagan, D.; Robinson, D. A. Substrate specificity in enzymatic fluorination. The fluorinase from *Streptomyces cattleya* accepts 2'-deoxyadenosine substrates. *Org. Biomol. Chem.* **4**, 1458-1460 (2006).
[22] Thompson, S.; Zhang, Q.; Onega, M.; McMahon, S.; Fleming, I.; Ashworth, S.; Naismith, J. H.; Passchier, J.; O'Hagan, D. A Localized Tolerance in the Substrate Specificity of the Fluorinase Enzyme enables "Last-Step" ^{18}F Fluorination of a RGD Peptide under Ambient Aqueous Conditions. *Angew. Chem. Int. Ed.* **53**, 8913-8918 (2014).
[23] Walker, M. C.; Thuronyi, B. W.; Charkoudian, L. K.; Lowry, B.; Khosla, C.; Chang, M. C. Y. Expanding the Fluorine Chemistry of Living Systems Using Engineered Polyketide Synthase Pathways. *Science* **341**, 1089-1094 (2013).
[24] Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; Garneau-Tsodikova, S.; Walsh, C. T. Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. *Chem. Rev.* **106**, 3364-3378 (2006).
[25] Handbook of Metalloproteins. Part 7. Vanadium Haloperoxidases (Eds.: Wever, R.; Hemrika, W.), Wiley-VCH, New York, 2006.
[26] Bernhardt, P.; Okino, T.; Winter, J. M.; Miyanaga, A.; Moore, B. S. A Stereoselective Vanadium-Dependent Chloroperoxidase in Bacterial Antibiotic Biosynthesis. *J. Am. Chem. Soc.* **133**, 4268-4270 (2011).
[27] Diethelm, S.; Teufel, R.; Kaysser, L.; Moore, B. S. A Multitasking Vanadium-Dependent Chloroperoxidase as an Inspiration for the Chemical Synthesis of the Merochlorins. *Angew. Chem. Int. Ed.* **53**, 11023-11026 (2014).
[28] Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191-1194 (2005).
[29] Wong, S. D.; Srnec, M.; Matthews, M. L.; Liu, L. V.; Kwak, Y.; Park, K.; Bell III, C. B.; Alp, E. E.; Zhao, J.; Yoda, Y.; Kitao, S.; Seto, M.; Krebs, C.; Bollinger Jr., J. M.; Solomon, E. I. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499**, 320-323 (2013).

[30] Lewis, J. C.; Coelho, P. S.; Arnold, F. H. Enzymatic functionalization of carbon-hydrogen bonds. *Chem. Soc. Rev.* **40**, 2003-2021 (2011).
[31] Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**, 368-371 (2006).
[32] Vaillancourt, F. H.; Yin, J.; Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme Fe^{II} α-ketoglutarate and O_2-dependent halogenase. *Proc. Natl. Acad. Sci.* **102**, 10111-10116 (2005).
[33] Hillwig, M. L.; Liu, X. A new family or iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Bio.* **10**, 921-923 (2014).
[34] Wood, J. L. Total synthesis: Welwitindolinone is well worth it. *Nat. Chem.* **4**, 341-343 (2012).
[35] Flecks, S.; Patallo, E. P.; Zhu, X.; Ernyei, A. J.; Seifert, G.; Schneider, A.; Dong, C.; Naismith, J. H.; van Pée, K.-H. New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angew. Chem. Int. Ed.* **47**, 9533-9536 (2008).
[36] 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* **46**, 1284-1292 (2007).
[37] van Pée, K.-H. Halogenating Enzymes for Selective Halogenation Reactions. *Curr. Org. Chem.* **16**, 2583-2597 (2012).
[38] Neumann, C. S.; Walsh, C. T.; Kay, R. R. A flavin-dependent halogenase catalyzes the chlorination step in the biosynthesis of *Dictyostelium* differentiation-inducing factor 1. *Proc. Natl. Aca. Sci.* **107**, 5798-5803 (2010).
[39] Xiao, Y.; Li, S.; Niu, S.; Ma, L.; Zhang, G.; Zhang, H.; Zhang, G.; Ju, J.; Zhang, C. Characterization of Tiacumicin B Biosynthetic Gene Cluster Affording Diversified Tiacumicin Analogues and Revealing a Tailoring Dihalogenase. *J. Am. Chem. Soc.* **133**, 1092-1105 (2011).
[40] Zeng, J.; Zhan, J. A Novel Fungal Flavin-Dependent Halogenase for Natural Product Biosynthesis. *ChemBioChem* **11**, 2119-2123 (2010).
[41] Wagner, C.; König, G. M. Handbook of Marine Natural Products: Mechanisms of Halogenation of Marine Secondary Metabolites (Eds.: Fattorusso, E.; Gerwick, W. H.; Taglialatela-Scafati, O.), Wiley-Springer Netherlands, 2012.
[42] 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.* **102**, 3960-3965 (2005).
[43] Seibold, C.; Schnerr, H.; Rumpf, J.; Kunzendorf, A.; Hatscher, C.; Wage, T.; Ernyei, A. J.; Dong, C.; Naismith, J. H.; van Pée, K.-H. A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocat. Biotrans.* **24**, 401-408 (2006).
[44] Zehner, S.; Kotzsch, A.; Bister, B.; Süssmuth, R. D.; Méndez, C.; Salas, J. A.; van Pée, K.-H. A Regioselective Tryptophan 5-Halogenase Is Involved in Pyrroindomycin Biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* **12**, 445-452 (2005).
[45] Runguphan, W.; Qu, X.; O'Connor, S. E. Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461-464 (2010).
[46] Keller, S.; Wage, T.; Hohaus, K.; Hölzer, M.; Eichhorn, E.; van Pée, K.-H. Purification and Partial Characterization of Tryptophan 7-Halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew. Chem. Int. Ed.* **39**, 2300-2302 (2000).
[47] Roy, A. D.; Grüschow, S.; Cairns, N.; Goss, R. J. M. Gene Expression Enabling Synthetic Diversification of Natural Products: Chemogenetic Generation of Pacidamycin Analogs. *J. Am. Chem. Soc.* **132**, 12243-12245 (2010).

[48] Sánchez, C.; Zhu, L.; Braña, A. F.; Salas, A. P.; Rohr, J.; Méndez, C.; Salas, J. A. Combinatorial biosynthesis of antitumor indolocarbazole compounds. *Proc. Natl. Acad. Sci.* **102**, 461-466 (2005).
[49] Hölzer, M.; Burd, W.; Reißig, H.-U.; van Pée, K.-H. Substrate Specificity and Regioselectivity of Tryptophan 7-Halogenase from *Pseudomonas fluorescens* BL915. *Adv. Synth. Catal.* **343**, 591-595 (2001).
[50] Suess, B.; Fink, B.; Berens, C.; Stentz, R.; Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.* **32**, 1610-1614 (2004).
[51] Romero, P. A.; Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Bio.* **10**, 866-876 (2009).
[52] Fasan, R.; Chen, M. M.; Crook, N. C.; Arnold, F. H. Engineered Alkane-Hydroxylating Cytochrome $P450_{BM3}$ Exhibiting Nativelike Catalytic Properties. *Angew. Chem. Int. Ed.* **46**, 8414-8418 (2007).
[53] 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* **329**, 305-309 (2010).
[54] Xie, J.; Liu, W.; Schultz, P. G. A Genetically Encoded Bidentate, Metal-Binding Amino Acids. *Angew. Chem. Int. Ed.* **46**, 9239-9242 (2007).
[55] Ghislieri, D.; Green, A. P.; Pontini, M.; Willies, S. C.; Rowles, I.; Frank, A.; Grogan, G.; Turner, N. J. Engineering an Enantioselective Amine Oxidase for the Synthesis of Pharmaceutical Building Blocks and Alkaloid Natural Products. *J. Am. Chem. Soc.* **135**, 10863-10869 (2013).
[56] Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H. Evolutionary History of a Specialized P450 Propane Monooxygenase. *J. Mol. Biol.* **383**, 1069-1080 (2008).
[57] Blaser, H. U.; Spindler, F.; Studer, M. Enantioselective catalysis in fine chemicals production. *Appl. Catal. A: Gen.* **221**, 119-143 (2001).
[58] Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the third wave of biocatalysis. *Nature* **485**, 185-194 (2012).
[59] Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas III, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **319**, 1387-1391 (2008).
[60] Lewis, J. C. Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *ACS Catal.* **3**, 2954-2975 (2013).

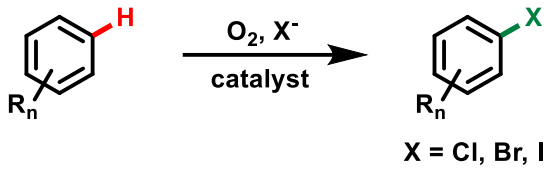
CHAPTER II

ENHANCING THE EXPRESSION OF WILD-TYPE REBH AND AN EXPLORATION OF ITS SUBSTRATE SCOPE

2.1 INTRODUCTION

As discussed in section 1.1, halogenated organic compounds are essential building blocks for chemical synthesis, for example in cross-coupling reactions that form biaryl compounds,[1] and are also widely represented in bioactive compounds such as pharmaceuticals and agrochemicals.[2] The presence of the halogen atom can often have beneficial effects on the bioactivity of such compounds,[3] and the site of halogenation can lead to drastically different biological activities (see Figure 1.1).[4] Despite their ubiquity and utility in modern society, however, the installation of halogen atoms onto a specific site of a compound frequently necessitates the use of activated or prefunctionalized starting materials and wasteful multistep functional group conversion sequences,[5] the design and execution of such sequences potentially differing significantly for each halogenated isomer one wishes to prepare. Arene halogenation via electrophilic aromatic substitution (EAS) often suffers from poor regioselectivity and results in undesired polyhalogenated side products while requiring harsh reaction conditions and harmful stoichiometric reagents, e.g. chlorine gas for chlorinations.[6] While oxidative methods to catalytically generate halogenating agents from halide salts have been developed to address the latter limitation, few utilize oxygen as the terminal oxidant, and none have solved the issue of regioselectivity (Scheme 2.1).[7]

Scheme 2.1. General scheme for oxidative halogenation.



The development of a site-selective oxidative halogenation method would be of great use to organic chemists in the syntheses of many high-value compounds. As established in the previous paragraph, two pitfalls of contemporary halogenation methods we wished to overcome were the lack of site-selectivity and the use of harsh reaction conditions. In both of these areas, enzymes are known to excel, as they often offer unparalleled site-selectivity and operate in the mild conditions associated with most biological milieu. The background given for halogenases in sections 1.2.1-1.2.4 demonstrates that, especially with regard to site-selectivity, the FAD-dependent halogenases (FDHs),[8] first identified in the mid 1990s,[9] are particularly attractive for engineering efforts. Detailed mechanistic investigations of FDHs have been spearheaded by the van Pée[10] and Walsh[11] groups, focusing primarily on tryptophan FDHs, and have served to elucidate the possible mechanisms of these enzymes. A separate NADH-dependent flavin reductase supplies $FADH_2$ that reacts in a flavin-binding site of the FDH with O_2 to form a flavin peroxide, FADH-OOH, which reacts with halide to generate hypohalous acid, HOX. This hypohalous acid then travels through a tunnel within the enzyme to either coordinate with an active-site lysine residue[10] or react with said residue to form a haloamine species.[11] In either case, the HOX or haloamine acts as a source of halenium ion and is proximal to only a single C-H bond of the substrate, which is held by the enzyme to affect the site-selectivity of the halogenation. These enzymes are thus able to override the electronic preference of L-tryptophan for halogenation at the 2-position and instead direct

halogenation to the relatively electronically disfavored 5-,[17] 6-,[18] or 7-position;[15,16] FDH homologs have been identified that are each capable of site-selective halogenation of only one of these sites.

FDHs thus have the potential to overcome the lack of site-selectivity and the harsh, environmentally detrimental conditions associated with current halogenation methods, but have thus far remained largely unexplored for preparative scale syntheses. Reported FDHs have typically had their activities determined through analytical scale reactions (micrograms of substrate) of their native substrates, often L-tryptophan.[16] The substrate scope of the L-tryptophan halogenase PrnA expressed in *Pseudomonas fluorescens* was explored by van Pée and coworkers. This enzyme halogenated a variety of substituted indoles, but for all non-native substrates, halogenation was invariably observed at the most electronically activated 2-position.[28] The same group was able to alter the selectivity on the native substrate, L-tryptophan, to furnish some 5-chloro or 5-bromo-L-tryptophan, but roughly 75% of the product was still 7-halogenated.[12] In order to effect the halogenation of non-native, bioactive substrates at electronically disfavored positions, some groups have utilized PrnA, PyrH (an L-tryptophan 5-halogenase), Thal (an L-tryptophan 6-halogenase), and especially RebH (an L-tryptophan 7-halogenase) in metabolic engineering efforts to halogenate early intermediates that are further functionalized biosynthetically;[13] some of these efforts are summarized in section 1.2.4. In the course of one such effort, O'Connor and coworkers demonstrated that wild-type RebH is capable of site-selective halogenation of tryptamine at the 7-position, in contrast to PrnA, for which only 2-halogenation of tryptamine was observed.[29] These authors also introduced point mutations in order to shift the preference of RebH for tryptamine versus L-tryptophan.

Inspired by these examples, we set out to obtain high expression of an active, soluble halogenase in *Escherichia coli*, an industrially tractable, easily genetically manipulated organism

that is a workhorse of molecular biology. Such high expression would then facilitate preparative scale halogenations and allow facile exploration of the scope and selectivity of our halogenase of interest. In addition to preparative scale reactions, high expression titers would also enable high-throughput screening of small scale cultures of halogenase variants, a prerequisite for the directed evolution. Thus, we hoped that by developing a robust expression system for a halogenase with a suitably broad substrate scope, we would be well poised for subsequent halogenase engineering efforts, which will be described in Chapters 3 and 4. Large portions of this work were performed with Mary Andorfer, and a paper has been published describing these results.[14]

2.2 RESULTS AND DISCUSSION

2.2.1 THE SELECTION OF A SUITABLE HALOGENASE

While several FDHs had been reported as of the beginning of this project, none to date had been explicitly reported to express in high titers in *E. coli*, with many lacking reports of any expression in *E. coli*. For example, the aforementioned study exploring the substrate scope of PrnA was conducted by expressing PrnA in its native host, *Pseudomonas fluorescens*, in order to obtain high enough quantities of isolated enzyme to perform multiple preparative scale reactions. Due to our interest in the development of engineered halogenase variants capable of site-selective halogenation of bioactive indoles (see Chapter 3), we decided to focus on L-tryptophan halogenases. A number of these halogenases, along with their reported selectivities and kinetic parameters, are shown in Table 2.1:

Table 2.1. Reported L-tryptophan FDH properties.

Enzyme	Site-Selectivity	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} μM^{-1})	Reference
RebH	7-position	1.4	2.0	0.7	15
PrnA	7-position	0.093	160	0.0006	16
PyrH	5-position	0.5	150	0.003	17
thal	6-position	2.8	110	0.025	18
KtzQ	7-position	0.19	≤2	0.1	19
KtzR	6-position	0.08	808	0.0001	19

Initially, three of the most well-studied of these tryptophan halogenases – PrnA, RebH, and PyrH – were selected to explore the level of expression of soluble enzyme each affords in *E. coli*. All of the tryptophan halogenases described in Table 2.1 were later studied in our lab and all were found to give comparable levels of expression to that described below for PrnA, RebH, and PyrH. These three halogenases were cloned into either pET22 or pET28a plasmid and used to transform BL21(DE3) *E. coli*, but subsequent SDS-PAGE analysis of expression cultures showed almost solely insoluble protein (see Figure 2.1 for RebH expression; PrnA and PyrH showed comparable expression levels). Given that PrnA, RebH, and PyrH all possess a GC percentage of 65%, while the overall GC content of a common laboratory strain of *E. coli* is 50.79%,[20] we hypothesized that using a halogenase with a GC content closer to 50% might fold more favorably in the intracellular milieu.

A BLAST search revealed two putative halogenases with high homology for the reported halogenase RebH with GC contents closer to 50%. One of these (Mmar) is expressed by *Mycobacterium marinum* and possesses 59% GC content and 64% sequence identity to RebH. The second (Pcurd) is expressed by *Paenibacillus curdlanolyticus* and possesses 51% GC content and 56% sequence identity to RebH. Both of these genes were cloned from the genomic DNA of their respective organisms of origin into pET22 and used to transform BL21(DE3) *E. coli*. SDS-PAGE

analysis of these cultures revealed no improvement in the ratio of soluble to insoluble protein relative to the reported halogenases was observed (expression levels were all comparable to that seen for RebH in Figure 2.1). Given the equally poor solubility seen with all the halogenases examined thus far, RebH was selected for studies to attempt to improve its solubility. This choice was made because RebH has the most favorable reported kinetic parameters of the halogenases reported to date (see table 2.1) and its purification from *E. coli* had been previously reported,[15] although without specific mention of the isolated yield of protein.

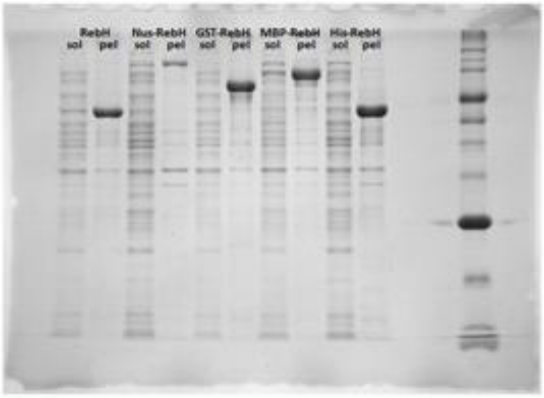
2.2.2 ENHANCING THE EXPRESSION OF SOLUBLE REBH

Based on literature reports of the beneficial effect upon solubility obtained by fusing a very well-folded protein (a fusion solubility partner) to a protein of interest,[21] we decided to investigate if doing so could have such an effect on the solubility of RebH. The Bottomley group at Monash University generously provided us with plasmids to allow for the facile incorporation of a gene of interest into four different constructs in the pLIC vector (a modified pET21 plasmid): one which attaches the N-utilizing substrance A (NusA) fusion solubility partner; one which attaches the maltose binding protein (MBP) fusion solubility partner; one which attaches the glutathione S-transferase (GST) fusion solubility partner; and one which is an identical construct but attaches only a hexahistidine (His) tag as a control.[22]

Using sequence and ligation independent cloning,[23] RebH was cloned into each of these four constructs. Each of the four constructs, Nus-RebH, GST-RebH, MBP-RebH, and His-RebH, were expressed alongside pET28a/RebH (5 mL expression cultures, induced with 0.1 mM IPTG at OD_{600} = 0.6, expressed for 20 hours at 37 °C). The results of this test showed that there were no extraordinary increases in the amount of soluble RebH obtained with the use of any fusion partner (Fig. 2.3). Shortening the expression time from 20 hours to 3 hours resulted in a significant increase

in the quantity of soluble enzyme obtained with the Nus-RebH and MBP-RebH fusions. These fusions were expressed and purified on the 750 mL scale, but the purified enzyme showed essentially no activity. While the solubility partners and RebH are separated by a TEV cleavage site and thus the solubility partners could be cleaved to liberate potentially active RebH, the additional protease cleavage step was not desired for the throughput of our planned directed evolution strategy.

Figure 2.1. Impact of fusion solubility partners on RebH.[a]

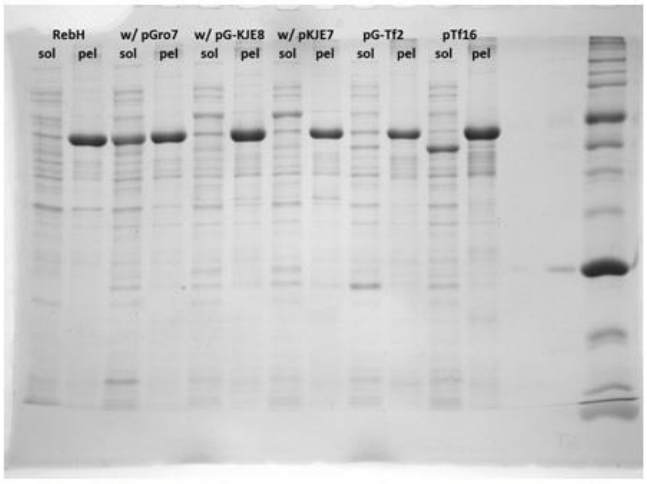


[a] Leftmost two lanes show the soluble and insoluble fraction of the untagged protein.

We then decided to test another method commonly employed to improve the solubility of expressed proteins that would not require any covalent modification of our protein of interest – coexpression with chaperone proteins.[24] Five commercially available chaperone plasmids were tested, each encoding for one to five chaperone proteins (in the form of different combinations of the dnaK-dnaJ-grpE, groEL-groES, and tig chaperones). Each was cotransformed into the *E. coli* BL21 (DE3) pET28/RebH system and all five chaperone coexpression systems were expressed alongside pET28a/RebH (5 mL expression cultures, induced with 0.1 mM IPTG at OD_{600} = 0.6, expressed for 20 hours at 37 °C). The results of this test showed that the pGro7 chaperone plasmid,

encoding the groEL-groES chaperone proteins, led to a significant increase in the quantity of soluble RebH obtained (Figure 2.2).

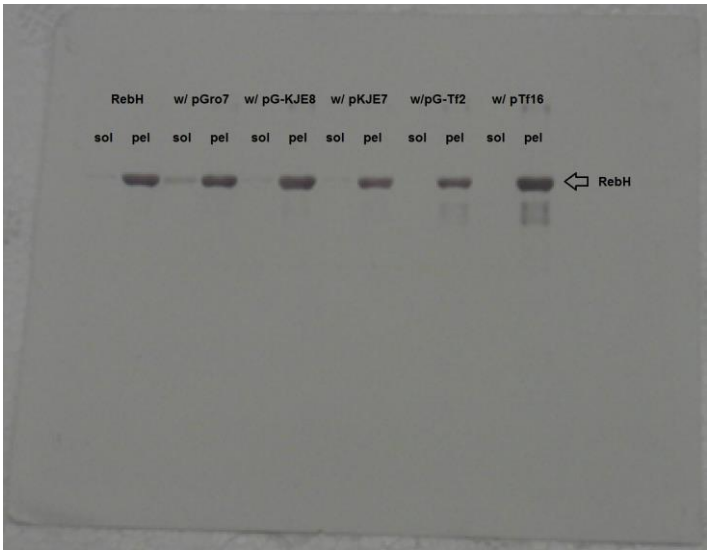
Figure 2.2. Impact of chaperone plasmid coexpression on RebH.[a]



[a] Leftmost two lanes show the soluble and insoluble fraction of RebH without chaperone coexpression.

Given the similar molecular weights of RebH (62 kD) and GroEL (58 kD), we expected that the soluble protein band was partially composed of GroEL. Western blotting was thus used to directly visualize the increase in the amount of soluble RebH furnished from chaperone coexpression (using a Mouse Anti-His antibody, which was then detected with a Goat Anti-Mouse Antibody conjugated to horseradish peroxidase). As seen in Figure 2.3, a significant amount of the RebH/groEL band in this unoptimized system was composed of groEL, but nonetheless, an approximately fourfold increase in the amount of soluble RebH was obtained upon coexpression with pGro7.

Figure 2.3. Western blot of RebH coexpression with chaperones.



In order to quickly optimize expression in this system, a functional halogenation assay (as described in the "General Procedure for 10 mg Bioconversions" subsection of Section 2.4.2, except on a 75 µL scale in 1.5 mL microcentrifuge tubes) was employed for further optimization of the expression medium, IPTG concentration, L-arabinose concentration, OD_{600} at induction, time of addition of L-arabinose, expression time, expression temperature, expression volume, and other factors. After optimization, large-scale cultures were used to determine isolated yields of soluble RebH obtained with the best expression conditions. A yield of 111 mg RebH per liter of culture with pGro7 coexpression was obtained, while only 15 mg RebH per liter of culture were obtained under the reported conditions, meaning that an over sevenfold improvement was obtained with the groEL/groES chaperones. To verify the beneficial effect of the chaperones themselves, as opposed to simply the varied expression conditions, a culture of RebH without the chaperone was grown using the optimized conditions and found to yield only 6 mg RebH per liter of culture (Table 2.2).

Table 2.2. Yields after purification under reported[15] and optimized conditions.[a]

Conditions	RebH	RebF
Reported Conditions	15 mg/L	3 mg/L
New Conditions	6 mg/L	5 mg/L
With pGro7 (RebH) or MBP (RebF)	111 mg/L	33 mg/L

[a] Purification performed using Ni-NTA affinity chromatography as described in Section 2.4.1.

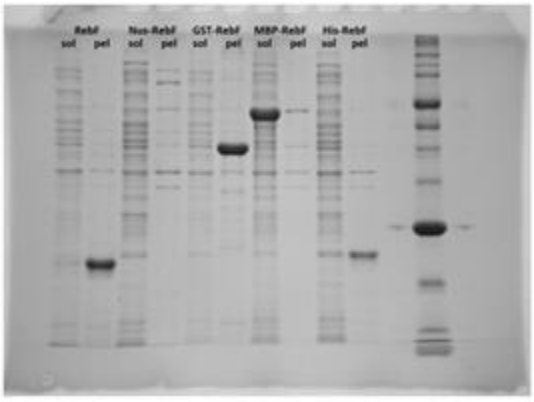
2.2.3 ENHANCING THE EXPRESSION OF SOLUBLE REBF AND DEVELOPING A COFACTOR REGENERATION SYSTEM

Given that RebH is a flavin-dependent halogenase, it is necessary to provide a supply of reduced FAD to the enzyme for the purpose of reducing molecular oxygen to generate FAD-peroxide. This in turn then reacts with halide to produce hypohalous acid, which then is itself either the active halogenating species, or which then further reacts with the active-site lysine to generate a haloamine that is the active halogenating species. While there have been reports that DTT can be used to reduce FAD in place of a flavin reductase,[25] we observed lower total turnovers with DTT than we did with a full cofactor regeneration system. Other authors have reported the use of exogenously added endogenous *E. coli* flavin reductases in FDH catalyzed bioconversions, such as SsuE,[16] but we decided to first investigate the use of RebF, the flavin reductase partner to RebH from the rebeccamycin biosynthetic gene cluster.

Upon expressing RebF in pET28a in BL21(DE3) *E. coli*, a similar situation was observed to that seen with RebH; a significant amount of expressed protein was obtained, but it was almost entirely insoluble. Following the same strategy described for RebH, the aforementioned fusion solubility partners – NusA, GST, and MBP – were cloned onto RebF to examine their effect on RebF's solubility. Unlike with RebH, the effect of these solubility partners was immediately

apparent upon expressing each construct, with the MBP-RebF fusion showing very high expression of almost fully soluble protein (Figure 2.4). After optimization of expression medium, temperature, time, inducer concentration, and OD_{600} at induction, large scale cultures were grown and the yield of isolated MBP-RebF was compared to that obtained for RebF. Approximately 33 mg/L of soluble MBP-RebF were obtained, while only 3 mg/L of RebF were obtained following the reported procedures (Table 2.2). Again, the optimized conditions were used in the absence of the MBP fusion partner to confirm that the impact on yield was not simply due to the change in conditions, and only 5 mg/L of RebF were obtained.

Figure 2.4. Impact of fusion solubility partners on RebF.[a]



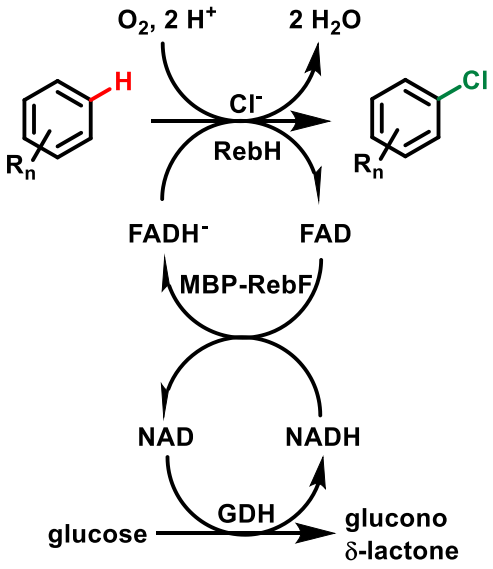
[a] Leftmost two lanes show the soluble and insoluble fraction of the untagged protein.

Gratifyingly, the MBP-RebF fusion is still fully active without any requisite cleavage of the solubility partner. Curiously, an *increase* in activity relative to RebF without the MBP partner was observed ($k_{cat,RebF}$ = 32 min^{-1}; $k_{cat,MBP\text{-}RebF}$ = 80 min^{-1} – while this observed $k_{cat,RebF}$ is lower than that previously reported,[15] multiple trials with both identically prepared enzymes were performed in our hands and all confirmed the same relative rates). While retention of some activity with the attachment of a fusion tag has been reported,[26] we have not been able to find precedent for an

increase in activity. Dynamic light scattering experiments have not shown increased aggregation for RebF without the MBP tag (data not shown), which we thought may have been leading to a still soluble, but inactive, form.[27]

With the ability to obtain large amounts of RebH (>100 mg/L) and RebF (>30 mg/L) now in hand (Table 2.2), we next investigated the use of a glucose dehydrogenase to regenerate the NADH that is consumed by RebF in the course of regenerating $FADH_2$. Thankfully, multiple glucose dehydrogenases are commercially available from Codexis as lyophilized powders and we found that these integrated well into our overall cofactor regeneration system (Scheme 2.2). The overall system was optimized to determine the optimum stoichiometry of NAD, FAD, halide, glucose, RebH, MBP-RebF, and GDH. We found that RebH and MBP-RebF show no loss of activity after months of storage frozen in 25 mM HEPES, 10% glycerol, pH = 7.4, so all the components are easily stored and handled. RebH can also be easily lyophilized and reconstituted without any significant loss in activity.

Scheme 2.2. Cofactor regeneration system.[a]



2.2.4 EXPLORING THE SUBSTRATE SCOPE OF WILD-TYPE REBH

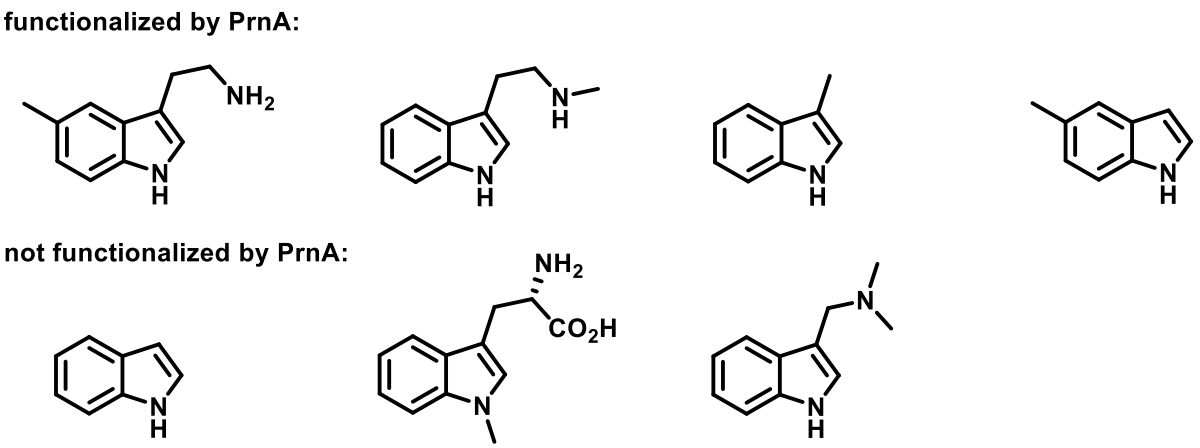
The established cofactor regeneration system described in section 2.2.3 allowed us to begin to explore the activity of RebH. First, the reported native substrate was found to be halogenated well by RebH in the presence of sodium chloride and the other reaction components, giving full conversion to a monochlorinated product at <1 mol% enzyme loading. Higher enzyme loadings and longer reaction times show that a substantial percentage of the monochlorinated product is converted to a dichlorinated product; the regioselectivity of the second site of halogenation was not verified, but we postulate that the product is 6,7-dichlorotryptophan, as this product has been reported for the similar L-tryptophan halogenase, KtzQ.[19] In addition, running the reaction with sodium bromide in place of sodium chloride gives slightly lower but still substantial conversion to the monobrominated product, with dibromination also seen at higher enzyme loadings and reaction times.

To explore whether or not wild-type RebH possesses activity on unnatural substrates – an initial, even if low, activity on other substrates of interest would provide a valuable starting point for evolution to expand the substrate scope of RebH – it was first necessary to test the cosolvent tolerance of RebH. The native substrate, L-tryptophan, is soluble in the reaction buffer (25 mM HEPES, pH = 7.4), but many of the substrates we intended to test are not soluble in aqueous buffer and instead require methanol, DMSO, or other organic solvents. Reactions using L-tryptophan were run with the addition of 5% of methanol, ethanol, isopropanol, acetone, and DMSO, and in every case no significant loss in activity was observed.

The tolerance of RebH for the above mentioned organic solvents allowed us to then test a range of unnatural substrates, with varying degrees of similarity to L-tryptophan, to see if any initial activity could be observed. Prior to this work, the primary exploration of substrate scope of an

FDH was conducted by the van Pée group on the closely related L-tryptophan 7-halogenase PrnA.[28] In this study, the authors tested a range of substituted indoles and found that PrnA demonstrated activity on some, but lacked activity on others. As mentioned in Section 2.1, in all cases except the native substrate, L-tryptophan, PrnA was found to give the electronically favored 2-substituted product. These substrates were an immediate choice to test with RebH to identify potential differences in the scope of the two related halogenases. The substrates shown in Figure 2.5 were tested and RebH was found to have some to high activity on them all.

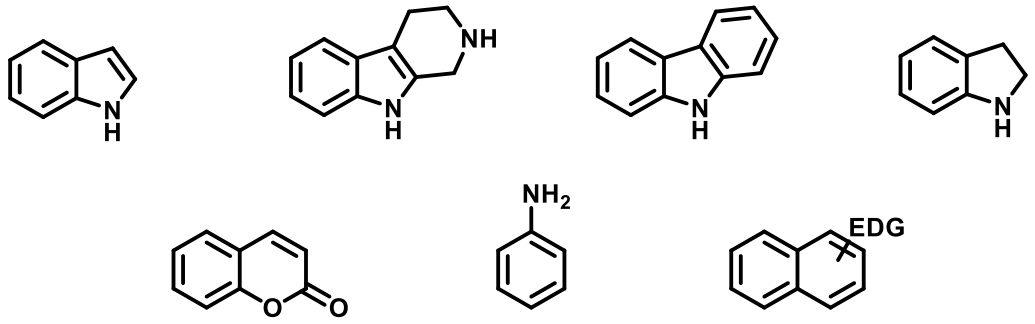
Figure 2.5. Substrates previously screened with PrnA.[28]



We were then interested in exploring an even further range of substrates than those reported for PrnA, especially to see if RebH possessed activity on non-indole substrates. Approximately 60 substrates were initially screened, and from these, an impressive diversity of scaffolds was accepted by wild-type RebH (Figure 2.6). This included a further expansion of indoles screened beyond those reported with PrnA (as shown in Figure 2.5), including tryptamine, tryptophol, D-tryptophan (the opposite enantiomer of the native substrate, L-tryptophan), and the tricycle tryptoline (Figure 2.7). In addition, several other classes of aromatic compounds were accepted, including coumarins, naphthalenes, anilines, and others (Figure 2.6). The halogenated coumarins

were found to have a shifted wavelength of maximum emission relative to the unhalogenated coumarins, which offered the possibility of using the halogenation of coumarins as a fluorometric screen for enhanced activity, thermostability, or organic solvent tolerance. Unfortunately, the simple coumarins explored at this point did not offer a significant enough change in signal to allow for reliable detection of improved conversion in crude lysate (data not shown).

Figure 2.6. Range of scaffolds functionalized by RebH.



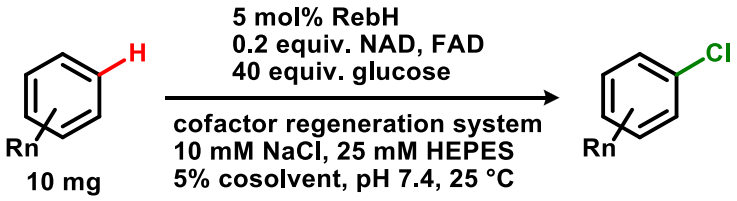
2.2.5 SELECTIVITIES OF WILD-TYPE REBH PREPARATIVE SCALE REACTIONS

As wild-type RebH was capable of halogenating a wide range of small substrates, we then wanted to determine the selectivity of RebH toward these substrates. While PrnA was reported to halogenate a limited range of substrates, as described above, it was found to halogenate all accepted substrates invariably at the 2-position. This selectivity is in accordance with that expected for chemical halogenation, as this position is electronically favored compared to the relatively deactivated benzene (4- through 7-) positions. Given that we intended to ultimately develop engineered halogenase variants with improved regioselectivities compared to small-molecule halogenation methods, we would intend to begin directed evolution with a substrate on which wild-type RebH has novel regioselectivity. As such, it was important for us to determine if wild-type RebH was affording novel regioselectivity on the non-native substrates on which it had displayed activity, and thus if these non-native substrates would offer a starting point for further evolution.

We were encouraged by a report that wild-type RebHdisplays such regioselectivity in halogenating tryptamine,[29] as it does so selectively at the 7-position, and we then resolved to determine the site of halogenation on other non-native substrates.

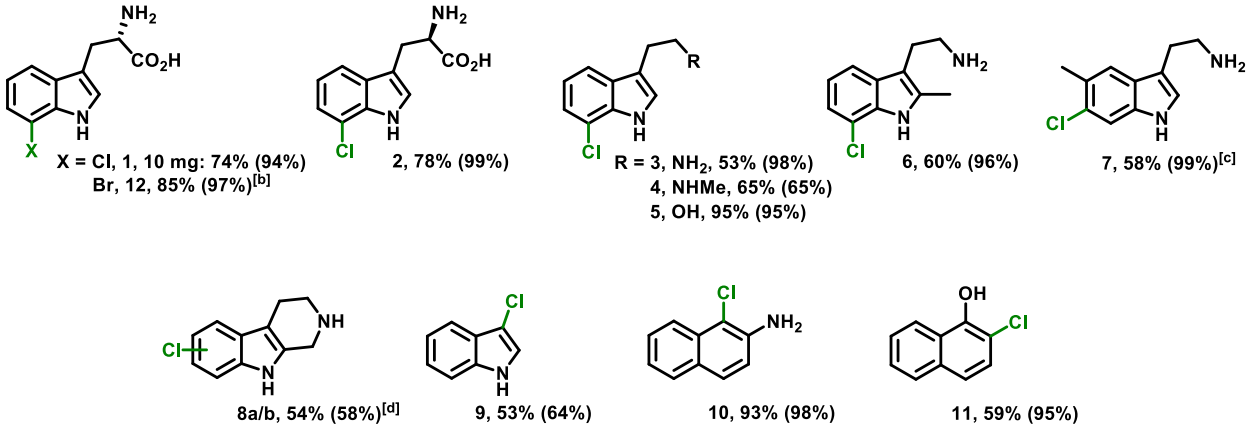
In order to determine the site of halogenation, the bioconversions using RebH must be carried out at such a scale as to afford an isolable quantity of product sufficient for NMR characterization. Such characterization can be readily performed on ~10 mg of product, but this represents a scale over 1000 times larger than the analytical scale reactions previously described. The increased expression of RebH described in Section 2.2.2, as well as the development of a robust cofactor regeneration system described in Section 2.2.3, enabled such a scale-up in a practical time frame. Parallel 750 mL cultures of RebH and MBP-RebFwere grown in shake-flasks and the proteins were purified and used to perform 10 mg scale bioconversions of a subset of substrates (Scheme 2.3). The bioconversions were worked up and purified using standard chemical techniques (as described in Section 2.4.2) and the sites of halogenation determined using 1D and 2D NMR. The products, with the determined site of halogenation highlighted in green, are shown in Table 2.3.

Scheme 2.3. General scheme for RebH-catalyzed arene halogenation.[a]



[a] Cofactor regeneration system is described in detail in Table 2.3.

Table 2.3. Yields for preparative RebH-catalyzed aromatic halogenation reactions.[a]



[a] Yields of isolated products and HPLC conversions (in brackets) are provided. The cofactor regeneration system consisted of 0.2 equiv FAD, 0.2 equiv NAD, 0.5 mol% RebF, 50 U/mL glucose dehydrogenase, and 40 equiv glucose. [b] 100 mm NaBr was used in place of 10 mm NaCl. [c] 6,7-dichloro-5-methyltryptamine was also isolated in 23% yield. [d] 10 mol% RebH loading was used. A nearly 1:1 mixture of 5- and 6-halogenation was observed.

Gratifyingly, wild-type RebH furnished novel regioselectivities on a number of the substrates tested. The native substrate, L-tryptophan, resulted in the reported regioselectivity at the 7-position (**1**), and surprisingly the opposite enantiomer, D-tryptophan, afforded the same regioselectivity (**2**). A range of tryptamine analogues, including tryptamine (**3**), N-Ω-methyltryptamine (**4**), tryptophol (**5**), and 2-methyltryptamine (**6**), all were selectively halogenated at the 7-position as well, with no trace of 2-halogenation detected for any of these substrates, despite PrnA affording only 2-halogenation on non-native indoles. Interestingly, 6-halogenation of 5-methyltryptamine was observed in 58% yield (**7**), indicating that additional sites of the indole benzene ring can be halogenated despite the presence of the more electron-rich pyrrole ring. At longer reaction times, dihalogenation of these substrates occurred as has been reported for the

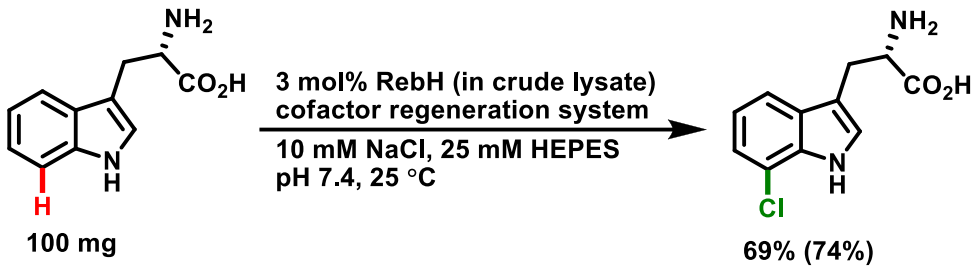
halogenases KtzQ and KtzR,[19] so reaction monitoring was required to ensure isolation of high yields of mono-halogenated products.

Halogenation of a number of bicyclic and tricyclic carbocycles and heterocycles lacking pendent functionality was also examined on the preparative scale. The tricyclic 2,3-disubstituted indole tryptoline was converted to a nearly 1:1 mixture of 6- and 7-chlorotryptoline (**8a/b**; reported as 5- and 6-chlorotryptoline, respectively, as they were originally numbered following indole nomenclature rather than carbazole nomenclature), providing a promising starting point for evolving halogenases with activity on more complex heterocyclic compounds.[30] Comparing this result with those for 2-methyltryptamine (**6**, 7-halogenation) and 5-methyltryptamine (**7**, 6-halogenation) also demonstrates that different sites of the benzene ring of nonnatural indole substrates can be halogenated by RebH. Again contrasting with reports for PrnA,[28] RebH provided high conversion of indole to halogenated products (**9**). This substrate was particularly susceptible to dihalogenation, and the reported 53% yield of 3-chloroindole is the maximum obtained before this process became significant. Substituted naphthalenes were also viable substrates, and monochlorinated compounds **10** and **11** were isolated in high yields. While these latter substrates were halogenated at their most activated sites, they nonetheless illustrate the ability of RebH to accept substrates significantly different from L-tryptophan.

To further demonstrate the synthetic utility of RebH, we confirmed that the bromination capabilities of this enzyme[15] could be translated to the preparative scale, and addition of NaBr to the reaction medium provided 7-bromotryptophan (**12**) in 85% isolated yield. The reaction scale was also increased to enable chlorination of 100 mg of tryptophan using crude cell lysate, rather than purified enzyme, as a catalyst (Scheme 2.4). While a significant decrease in conversion was observed relative to the 10 mg reaction conducted using purified enzyme, a 69% product yield was

still obtained. A strong dependence of yield on the reaction surface area/volume ratio was also observed and suggests that improvements may be achieved by controlling dissolved oxygen concentration. Oxygen is required for RebH-catalyzed halide oxidation but can also oxidize $FADH_2$ in solution, so control of O_2 concentration will be required to balance these processes.

Scheme 2.4. Halogenation of 100 mg L-tryptophan using crude lysate.[a]



[a] Cofactor regeneration is the same as described in Table 2.3.

Finally, the differing catalytic efficiency of RebH on representative substrates was examined by conducting reactions with enzyme loadings below those required for maximal substrate conversion (see Section 2.4.4 for representative kinetic data). As expected, decreased total turnover numbers (TTNs) were observed for non-native substrates (Table 2.4). The increasing K_m values for tryptamine (**3**) and tryptoline (**8**) are consistent with their increasing structural variation from tryptophan. The large structural differences between these substrates and 2-aminonaphthalene (**10**) make direct comparisons difficult, but the relative efficacy of this substrate presumably results from its strong electronic activation. These data clearly showed that significant improvements to catalyst efficiency or stability are desirable, both of which are goals we would later pursue.

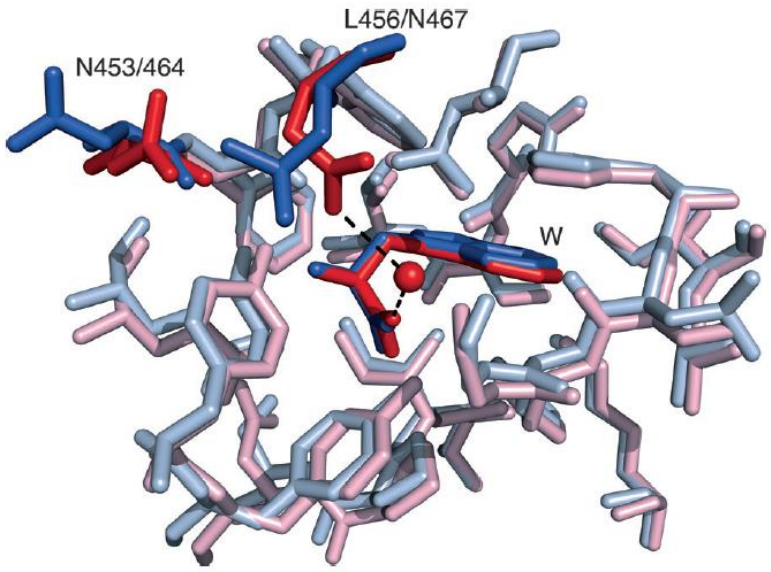
Table 2.4. Catalytic parameters for halogenation of select substrates.[a]

Substrate	RebH (mol%)	Conv. (%)	TTN	k_{cat} (min^{-1})[b]	K_m (μM)	k_{cat}/K_m (min^{-1} μM^{-1})
L-tryptophan[c]	0.5	75	165	1.1	7.3	0.15
tryptamine	1	15	19	0.023	9.0	0.0026
tryptoline	5	30	3.6	0.027	216	0.00013
2-aminonaphthalene	1	32	26	0.59	14	0.042

[a] 75 μL reactions were conducted as described in Table 2.3 using the indicated RebH loading and 0.5 mM phenol as an internal standard, quenched with an equal volume of methanol, and analyzed by HPLC. [b] Initial rate data were used to construct Hanes–Woolf plots to determine k_{cat} and K_m. [c] Reported values for the k_{cat} and K_m of RebH on L-tryptophan are 1.4 min^{-1} and 2.0 μM,[15] respectively.

Several notable differences between the reactivity of RebH and that reported for PrnA deserve comment given the significant homology of both the complete sequences (55%) and the active sites (Figure 2.7) of these two enzymes. First, PrnA was reported to have no activity on indole, gramine, or 1-methyl-L-tryptophan,[28] whereas RebH halogenated each of these substrates. The simplest explanation for this difference is that low conversion using PrnA precluded identification of the halogenated products. RebH has higher catalytic efficiency than PrnA (reported k_{cat} values on L-tryptophan are 1.4 min^{-1} versus 0.093 min^{-1} – see Table 2.1), and our improved reaction conditions would have further improved this advantage. We suspect that this low activity may have made efforts to distinguish between different regioisomers difficult and that PrnA may display similar regioselectivity to RebH on some non-native substrates; analytical-scale tests in our hands with PrnA are consistent with this conclusion (data not shown).

Figure 2.7. Comparison of structures of RebH and PrnA.[a]



[a] Selected residues and a water molecule located within 5 Å of tryptophan in PrnA (purple/lavender) and RebH (red/pink) shown. PDB files 2A1 and 2AQJ used for RebH[31] and PrnA,[32] respectively. Bold colors denote the tryptophan substrates (W), residues with different identities (L456/N467) or conformations (N453/464) in PrnA or RebH, respectively, and a water molecule.

Second, PrnA was reported to catalyze 2-halogenation of both 5- and N-Ω-methyltryptamine,[28] while RebH catalyzes 6- and 7-halogenation of these substrates, respectively. Crystal structures for both of these enzymes bound to FAD, chloride, and L-tryptophan are available.[31,32] Analysis of residues (24 total) within 5 Å of the substrate tryptophan shows only one pair of residues (N467 and L456) that differ in identity and only one additional pair of residues (N464 and N453) displaying a notable conformational difference (Figure 2.7). In RebH, the side chain of N467 forms a water-mediated hydrogen bond to the substrate L-tryptophan. Similar hydrogen bonding could help position other substrates for halogenation at electronically disfavored

sites during catalysis. The structurally analogous residue in PrnA is L456, which cannot form a hydrogen bond, and this difference may explain the apparent differences in selectivity between PrnA and RebH.

2.3 CONCLUSIONS

The expression of an active FDH in high yield is an essential precursor to all future work involving preparative scale or high-throughput applications. In order to achieve this, RebH was selected from a set of halogenases and its expression increased significantly through coexpression with the chaperone proteins groEL and groES and optimization of expression conditions. This, along with a similar effort for the partner flavin reductase, RebF, and the establishment of a robust cofactor regeneration system, allowed the exploration of the substrate scope of wild-type RebH. Halogenation, including both chlorination and bromination, of a range of medicinally relevant indoles and naphthalenes was possible on up to a 100 mg scale with high yields, and it was determined that many of these non-native substrates are halogenated selectively at relatively electronically disfavored positions. These results contrasted somewhat with those reported for PrnA, a structurally homologous halogenase, which provided a narrower substrate scope and only enabled halogenation of non-native substrates at their most electronically activated positions. Overall, this work demonstrated the unique potential of RebH as a catalyst for regioselective oxidative halogenation and set the stage for the future engineering of RebH to improve its thermostability, expand its substrate scope, and alter its regioselectivity.

2.4 EXPERIMENTAL

2.4.1 GENERAL EXPERIMENTAL PROCEDURES

Materials:

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Deuterated solvents were obtained from Cambridge Isotope labs. Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Oligonucleotides were purchased from Integrated DNA Technologies (San Diego, CA), and the sequences of the primers used in this study are reported below.

Plasmids pET-28a/RebF and pET-28a/RebH in BL-21 (DE3) *E. coli* were provided by the Walsh group of Harvard Medical School, Boston, MA.[15] The pLIC-MBP plasmid was provided by the Bottomley group of Monash University, Clayton, Australia.[22] The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan). DH5α and BL21 (DE3) *E. coli* cells were purchased from Invitrogen (Carlsbad, CA). SacII restriction enzyme, T4 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. All genes were confirmed by sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser using method Ec3. Ni-nitrilotriacetic

acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were following when using both products (for Ni-NTA resin, 5 mL resin were used, with buffers delivered by a peristaltic pump at a rate of 1 mL/min, in a 4 °C cold cabinet). Amicon® 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. The glucose dehydrogenase CDX-901, FAD, and NAD were purchased from Codexis (Redwood City, CA). NADH was purchased from Chem-Impex International (Wood Dale, IL). DOWEX™ 50WX8 strong cation exchange resin was purchased from Sigma-Aldrich. Biotage reverse phase columns (SNAP-KP-C18-HS) and amine-modified silica columns (SNAP-KP-NH) were purchased from Biotage.

General Procedures:

Standard molecular cloning procedures were followed.[34] Reactions were monitored using HPLC (Agilent 1200 UHPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H2O/0.1% TFA, solvent B = CH3CN; 0-10 min, B = 15%; 10-20 min, B = 15-100%; 20-24 min, B = 100%). Reverse phase preparative chromatography was carried out using a Biotage Isolera One. 1H and 13C NMR spectra were recorded at 500 MHz and 126 MHz, respectively, on a Bruker DMX-500 or DRX-500 spectrometer, and chemical shifts are reported relative to residual solvent peaks.[33] High-resolution mass spectra were obtained at the University of Chicago Mass Spectrometry Service Center on an Agilent Technologies 6224 TOF LC/MS. Low resolution mass spectra were obtained at the University of Chicago Mass Spectrometry Service Center on a Varian Saturn 2000, CP-3800 GC-MS/MS.

Sonication Conditions:

5 mL cultures: sonication was performed on a Qsonica S-4000 Sonicator with a microtip using the following procedure: 6x30s with 30s rests, 30% duty cycle delivering 15 W, while cooling on a circulating ice-water bath.

750 mL cultures: sonication was performed on a Qsonica S-4000 Sonicator with a 0.5" horn using the following procedure: 5x1m with 1m rests, 20% duty cycle delivering 40-50 W, while cooling on a circulating ice-water bath.

2.4.2 SPECIFIC EXPERIMENTAL PROECURES

Cloning of MBP-RebF Fusion Plasmid: pET-28a/RebF was isolated from the cell stock obtained from the Walsh group using a Qiagen Miniprep Kit. Chemically competent DH5α *E. coli* stocks were prepared using the calcium chloride method described in Sambrook and Russell, p. 1.116.[34] RebF was amplified from the isolated pET-28/RebF described using the following oligonucleotides:

Table 2.5. Primers used to clone RebF into the fusion plasmids.

#	Primer Name	Sequence
1	RebFFusNdeF	5' – CCA GGG AGC AGC CTC GCA TAT GAC GAT CGA GTT CGA CAG ACC CG – 3'
2	RebFFusHindR	5' – GCA AAG CAC CGG CCT CGT TAC ATA TGT CCC TCC GGT GTC CAC ACG GCG – 3'

The PCR conditions were as follows: 1x Phusion HF buffer, 0.2 mM dNTPs each, 0.2 μM forward primer, 0.2 μM reverse primer, and 0.02 U/μl Phusion polymerase. The final PCR reaction volumes were 50 μL each, and were carried out in an Eppendorf Mastercycler pro S thermal cycler. The PCR procedure was as follows: 95 °C for 1 minute; 94 °C for 3 minutes; 30 cycles of 98 °C for 30 seconds, 50 °C for 1 minute, and 72 °C for 3 minutes; and ending with 72 °C for 5 minutes. Formation of pLIC-MBP/RebF was performed as described in Cabrita *et al.*,[22] with the ligation product transformed into the chemically competent DH5α *E. coli* described above and plated onto LB/agarose plates containing 0.1 mg/mL ampicillin. A glycerol stock was prepared from 3.0 mL of overnight culture in LB with 0.1 mg/mL ampicillin, which was spun down (3,000g, 10 min) and resuspended in 350 μL fresh LB with 0.1 mg/mL ampicillin, then diluted with 500 μL 50% v/v water/glycerol. Glycerol stocks were all stored at -80 °C. The pLIC-MBP/RebF plasmid was isolated from this DH5α *E. coli* using a Qiagen Miniprep Kit. Electrocompetent BL21(DE3) *E. coli* stocks were prepared using the method described in Sambrook and Russell, p. 1.120.[34] Isolated pLIC-MBP/RebF plasmid was transformed by electroporation into electrocompetent BL21(DE3) *E. coli* and plated onto LB/agarose plates containing 0.1 mg/mL ampicillin. A glycerol stock was prepared as described above and stored at -80 °C.

Cloning of RebH with pGro7 Chaperone Plasmid: Electrocompetent DH5α *E. coli* stocks were prepared as described above. The pGro7 plasmid was transformed by electroporation into electrocompetent DH5α *E. coli* and then plated on LB/agarose plates containing 0.02 mg/mL chloramphenicol. A glycerol stock was prepared from 3.0 mL of overnight culture in LB with 0.02 mg/mL chloramphenicol, which was spun down (3,000g, 10 min) and resuspended in 350 μL fresh LB with 0.02 mg/mL chloramphenicol, then diluted with 500 μL 50% v/v water/glycerol. pGro7

plasmid was then isolated from this cell stock using a Qiagen Miniprep Kit. Electrocompetent pET-28a/RebH BL21(DE3) *E. coli* stocks were prepared as described above. The isolated pGro7 plasmid was transformed by electroporation into electrocompetent pET-28a/RebH BL21(DE3) *E. coli* and plated on LB/agarose plates containing 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin. A glycerol stock was prepared from 3.0 mL of overnight culture in LB with 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin, which was spun down (3,000g, 10 min) and resuspended in 350 μL fresh LB with 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin, then diluted with 500 μL 50% v/v water/glycerol.

Small-Scale Expression of MBP-RebF and RebH: For an initial comparison of improvements to solubility, RebH with and without pGro7 and RebF with and without the MBP fusion tag were expressed in BL21(DE3) *E. coli* overnight at 37 °C in 5 mL LB medium containing the requisite antibiotic(s) (100 μg/mL ampicillin for the fusion tag, 50 μg/mL kanamycin and 20 μg/mL chloramphenicol for the chaperone coexpression, and 50 μg/mL kanamycin for the controls). 5 mL fresh media were inoculated with 50 μL of overnight culture, and growth was continued at 37 °C until the cultures reached an OD_{600} of 0.6. At this point, 100 μM isopropyl β-D-thiogalactopyranoside (IPTG) and 2 mg/mL L-arabinose (for pGro7 chaperone coexpression) were added to induce expression, and incubation was continued for 20 hours at 37 °C. The cells were harvested by centrifugation at 3,000 g for 10 minutes at 4 °C and resuspended in 1 mL of 25 mM HEPES pH 7.4. The cells were lysed by sonication and pelleted by centrifugation at 16,100 g for 10 minutes at 4 °C. The supernatant was decanted, the pellets were each resuspended in 1 mL of 25 mM HEPES pH 7.4, and the soluble (supernatant) and insoluble (resuspended pellet) proteins for each sample were examined by SDS-PAGE (Figs. 2.4 and 2.6).

Large-Scale Expression and Purification of MBP-RebF: Primary cultures of pLIC-MBP/RebF BL21(DE3) *E. coli* were grown (7.5 mL LB containing 0.1 mg/mL ampicillin in 15 mL Falcon culture tubes) from the glycerol stock described above at 37 °C and 250 rpm overnight. 7.5 mL of primary culture were used to inoculate 750 mL of LB containing 0.1 mg/mL ampicillin in a 2.8 L, non-beveled glass flask. The cultures were then grown at 37 °C and 250 rpm to OD_{600} = 0.6, at which point isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 100 µM. The cultures were then incubated at 37 °C and 250 rpm for an additional three hours, at which time the cells were pelleted by centrifugation (4,000g, 30 min, 4 °C) and stored at -20 °C until further use.

The pellets were each resuspended in 30 mL of HEPES (25 mM, pH 7.4), transferred to 50 mL conical tubes, and lysed by sonication. The lysates were then clarified by centrifugation (24,000g, 30 min, 4 °C). The clarified lysate was then purified by Ni-NTA affinity chromatography. The elution fractions were then concentrated to 3 mL using an Amicon® 15 mL 30 kD cutoff centrifugal filter and exchanged five times into HEPES/glycerol buffer (25 mM, pH 7.5, 10% v/v glycerol). Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -20 °C until use.

Large-Scale Expression and Purification of RebH: Primary cultures of pET-28a/RebH & pGro7 BL21(DE3) *E. coli* were grown (7.5 mL LB containing 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin in 15 mL Falcon culture tubes) from the glycerol stock described above at 37 °C and 250 rpm overnight. 3.75 mL of primary culture were used to inoculate 750 mL of TB containing 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin in a 2.8 L, non-beveled glass flask. The cultures were then grown at 37 °C and 250 rpm to OD_{600} = 0.6, at which point isopropyl

β-D-1-thiogalactopyranoside (IPTG) and L(+)-arabinose were added to final concentrations of 100 μM and 2 mg/mL, respectively. The temperature was then reduced to 30 °C and the shake rate left at 250 rpm for an additional 20 hours, at which time the cells were pelleted by centrifugation (4,000g, 30 min, 4 °C) and stored at -20 °C until further use.

The pellets were each resuspended in 30 mL of HEPES (25 mM, pH 7.4), transferred to 50 mL conical tubes, and lysed by sonication. The lysates were then clarified by centrifugation (24,000g, 30 min, 4 °C). The clarified lysate was then purified by Ni-NTA affinity chromatography. The elution fractions were then concentrated to 3 mL using an Amicon® 15 mL 30 kD cutoff centrifugal filter and exchanged five times into HEPES/glycerol buffer (25 mM, pH 7.5, 10% v/v glycerol). Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -20 °C until use.

NADH oxidation by MBP-RebF: The activities of RebF and the MBP-RebF fusion were measured by monitoring the decrease in absorbance at 340 nM, which is caused by oxidation of NADH to NAD. For kinetic characterization, 1 μM RebF or MBP-RebF was incubated with 0.01-1.0 mM NADH, 50 μM FAD, and 10 mM NaCl in HEPES (25 mM, pH 7.5) in a 96-well microtiter plate. Reactions were monitored at 340 nm using a Tecan infinite M200pro plate reader and the resulting data were fit to the equation $y = (B_{max} * x) / (K_d + x)$ using SigmaPlot to obtain values for k_{cat}.

Determination of Kinetic Parameters for RebH: Rates were determined by monitoring the conversion of 75-215 μM substrate in the presence of NAD (100 μM final concentration), FAD (100 μM final concentration), NaCl (10 mM final concentration), MBP-RebF (2.5 μM final

concentration), glucose dehydrogenase (50 U/mL final concentration CDX-901), glucose (20 mM final concentration), and phenol as an internal standard (0.5 mM final concentration) at a final volume of 75 µL in a microtiter plate. RebH was added at a final concentration of either 1 µM for L-tryptophan and 2-aminonaphthalene or 25 µM for tryptamine and tryptoline. The reactions were left shaking at 600 rpm at room temperature, then quenched at 5-120 minutes (all time points were collected in triplicate) by addition of 75 µL of MeOH. The precipitated protein was then removed by centrifugation and the reactions were filtered and analyzed by HPLC using the method described in the General Procedures. Product formation was quantitated by calculating the ratio of product to internal standard and fitting that value to a calibration curve prepared from known concentrations of each material. The kinetic parameters (K_m and k_{cat}) for each substrate were determined using the Hanes-Woolf plots (see Sec. 2.4.4) constructed from the substrate concentrations and the observed initial rates.

Determination of Total Turnover Numbers (TTNs) for RebH: Reactions were assembled in triplicate containing substrate (1 equiv., 0.5 mM final concentration), NAD (0.2 equiv., 100 µM final concentration), FAD (0.2 equiv., 100 µM final concentration), NaCl (20 equiv., 10 mM final concentration), MBP-RebF (0.005 equiv., 2.5 µM final concentration), glucose dehydrogenase (50 U/mL final concentration CDX-901), glucose (40 equiv., 20 mM final concentration), and phenol as an internal standard (1 equiv., 0.5 mM final concentration) at a final volume of 75 µL in a microtiter plate. RebH was added at a final concentration of either 2.5 µM for L-tryptophan, 5 µM for tryptamine and 2-aminonaphthalene, or 25 µM for tryptoline. The reactions were left shaking at 600 rpm at room temperature, then quenched after 12 hours by addition of 75 µL of MeOH. The precipitated protein was then removed by centrifugation and the reactions were filtered and

analyzed by HPLC using the method described in the General Procedures. Product formation was quantitated by calculating the ratio of product to internal standard and fitting that value to a calibration curve prepared from known concentrations of each material.

General Procedure for 10 mg Bioconversions: A solution of 10 mg substrate (1 equiv., 0.5 mM final concentration) in 5 mL HEPES buffer (25 mM, pH 7.4) or isopropanol was added to a crystallization dish (125 x 65 or 100 x 50 mm). Solutions in HEPES buffer (25 mM, pH 7.4) of NAD (0.2 equiv., 100 µM final concentration), FAD (0.2 equiv., 100 µM final concentration), NaCl (20 equiv., 10 mM final concentration), and a glucose dehydrogenase (50 U/mL final concentration CDX-901) were added. Solutions in HEPES/glycerol buffer (25 mM, pH 7.5, 10% glycerol v/v) of RebH (0.05 equiv., 25 µM final concentration) and MBP-RebF (0.005 equiv., 2.5 µM final concentration) were added. The bioconversion was diluted with HEPES buffer (25 mM, pH 7.4) to the appropriate reaction volume and an aqueous solution of glucose (40 equiv., 20 mM final concentration) was added to initiate the cofactor regeneration cycle. The dish was covered with perforated aluminum foil and stirred with a magnetic stir bar at 60 rpm.

Reactions were monitored by HPLC as described in the General Procedures and were quenched with aqueous HCl (5.0 M, until pH < 2) upon completion. NaCl was added to saturation. Precipitated protein was filtered out through a pad of Celite and was washed with water and/or CH_2Cl_2 until all product had been rinsed from the Celite pad, as indicated by HPLC. The filtrate was either extracted into CH_2Cl_2 or submitted to strong cation exchange chromatography. The crude material thus obtained was purified by either normal or reverse phase chromatography.

Procedure for 100 mg L-Tryptophan Bioconversion: A solution of 100 mg L-tryptophan (1 equiv., 0.5 mM final concentration) in 10 mL HEPES buffer (25 mM, pH 7.4) was added to a crystallization dish (190 x 100 mm). Solutions in HEPES buffer (25 mM, pH 7.4) of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and CDX-901 (50 U/mL final concentration) were added. Sixteen 750 mL cultures of RebH coexpressing pGro7 were grown, expressed, spun down, resuspended, lysed, and clarified as described above, then pooled to furnish approximately 500 mL of clarified lysate in HEPES buffer (25 mM, pH 7.4). The rate of conversion of L-tryptophan was monitored for a volume of this lysate and compared to the rate of conversion of L-tryptophan using a known concentration of purified RebH. From this comparison, an effective RebH concentration in the lysate was calculated. Lysate was added such that the effective final concentration of RebH in the reaction was 15 μM (0.03 equiv.). A solution in HEPES/glycerol buffer of MBP-RebF (0.005 equiv., 2.5 μM final concentration) was added. The bioconversion was diluted with HEPES buffer (25 mM, pH 7.4) to the appropriate reaction volume and an aqueous solution of glucose (40 equiv., 20 mM final concentration) was added to initiate the cofactor regeneration cycle. The reaction was split evenly into two identical crystallization dishes (190 x 100 mm). The dishes were covered with perforated aluminum foil and stirred with magnetic stir bars at 60 rpm.

The reaction was monitored by HPLC as described in Section 2.4.1. Product concentration was not observed to increase between 20 and 22 hours, and the reaction was quenched with aqueous HCl (5.0 M, until pH < 2) and submitted to centrifugation (3,000g, 10 min) for initial protein removal. The supernatant was collected, and the pellet was washed through three rounds of re-suspension in acidic water (1 M HCl used to adjust pH < 2) and centrifugation (3,000g, 10

min). The supernatants were combined and passed through a Celite pad for further protein removal. The filtrate was submitted to strong cation exchange was followed as described below. Elution fractions from ion exchange were pooled, dry loaded onto Celite, and purified by reverse phase chromatography (Biotage SNAP-KP-C18-HS, gradient from pure H_2O to 15% CH_3CN/H_2O).

Procedure for Strong Cation Exchange: After protein precipitation and filtration through Celite, aqueous filtrate was submitted to strong cation exchange to remove salts if the product could not be extracted into organic solvent. DOWEX™ 50WX8 resin was slurry-packed with methanol in a 250 mL chromatography column. The resin was washed with ~300 mL of methanol and ~300 mL of deionized water. The resin was acidified with HCl (1 M) until the pH of flow through was less than 2. The resin was washed with deionized water until the pH was between 6-7. Acidic filtrate (pH < 2) was added to resin. The resin was washed with ~500 mL of deionized water. The product was eluted with NH_4OH (1 M) until product was no longer eluting from the column, as indicated by HPLC. Product-containing fractions were concentrated to dryness using a rotary evaporator under high vacuum.

2.4.3 DETAILED ISOLATION AND CHARACTERIZATION

7-chloro-L-tryptophan (1): The bioconversion was conducted according to the general procedure in a 100 x 50 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, submitted to strong cation exchange

chromatography, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford **1** in 74% yield (8.6 mg, 0.036 mmol). ^{1}H NMR (500 MHz, Deuterium Oxide with one drop TFA) δ 7.23 (d, *J* = 8.0 Hz, 1H), 7.01 (s, 1H), 6.92 (d, *J* = 7.6 Hz, 1H), 6.76 (t, *J* = 7.8 Hz, 1H), 4.03 (dd, *J* = 7.3, 5.4 Hz, 1H), 3.18 – 3.04 (m, 2H). ^{13}C NMR (126 MHz, Deuterium Oxide) δ 177.45, 133.29, 128.61, 125.37, 121.24, 120.13, 117.34, 116.50, 109.97, 55.51, 27.84. HRMS (ESI-TOF) calcd for $C_{11}H_{12}N_2O_2Cl$ $[M + H]^+$: 239.0529 and 241.0502, found: 239.0534 and 241.0501.

7-chloro-D-tryptophan (2): The bioconversion was conducted according to the general procedure in a 100 x 50 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, submitted to strong cation exchange chromatography, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford **2** in 78% yield (9.1 mg, 0.038 mmol). ^{1}H NMR (500 MHz, Deuterium Oxide with one drop TFA) δ 7.22 (d, *J* = 8.0 Hz, 1H), 6.99 (s, 1H), 6.93 (d, *J* = 7.6 Hz, 1H), 6.77 (t, *J* = 7.8 Hz, 1H), 4.04 (dd, *J* = 7.1, 5.4 Hz, 1H), 3.15-3.03 (m, 2H). ^{13}C NMR (126 MHz, Deuterium Oxide) δ 133.40, 128.53, 125.59, 121.40, 120.28, 117.33, 116.61, 109.37, 55.22, 27.04. HRMS (ESI-TOF) calcd for $C_{11}H_{12}N_2O_2Cl$, $[M + H]^+$: 239.0587 and 241.0561, found: 239.0529 and 241.0502. The carboxylate carbon was not detected because of low sample concentration. In the NOESY

spectrum of this product, resonances between the singlet at 6.99 ppm and other aryl protons are present. When examining the cross-peaks of these resonances, it was found they could not be phased properly and thus were not reliable. COSY and HMQC spectra were collected to ensure that these resonances were not true NOEs, and that the proposed structure was thus consistent with the acquired spectra. This is observed for all tryptophan products as well as 7-chlorotryptamine.

7-chloro-tryptamine (3): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 20% CH_3CN/water 0.1%TFA) to afford **3** in 53% yield (10.2 mg of **3**·TFA, 0.033 mmol). ^{1}H NMR (500 MHz, Deuterium Oxide) δ 7.48 (d, *J* = 8.0 Hz, 1H), 7.23 (s, 1H), 7.17 (d, *J* = 7.6 Hz, 1H), 7.01 (t, *J* = 7.8 Hz, 1H), 3.19 (t, *J* = 7.0 Hz, 2H), 3.02 (t, *J* = 7.0 Hz, 2H). ^{13}C NMR (126 MHz, Deuterium Oxide) δ 133.48, 128.15, 125.00, 121.46, 120.27, 117.05, 116.73, 110.36, 39.63, 22.64. HRMS (ESI-TOF) calcd for $C_{10}H_{12}N_2Cl$ $[M + H]^+$: 195.0689 and 197.0661, found: 195.0627 and 197.0664.

7-chloro-*N*-Ω-methyltryptamine (4): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite.

The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 25% CH_3CN/water 0.1%TFA) to afford **4** in 65% yield (12.0 mg of **4**·TFA, 0.037 mmol). ^{1}H NMR (500 MHz, 75:25 Deuterium Oxide:Acetonitrile-*d*$_3$) δ 7.91 (d, *J* = 7.9 Hz, 1H), 7.65 (s, 1H), 7.57 (d, *J* = 7.5 Hz, 1H), 7.43 (t, *J* = 7.8 Hz, 1H), 3.60 (t, *J* = 7.5 Hz, 2H), 3.46 (t, *J* = 7.4 Hz, 2H), 2.99 (s, 3H). ^{13}C NMR (126 MHz, 75:25 Deuterium Oxide:Acetonitrile-*d*$_3$) δ 132.88, 127.94, 124.19, 120.75, 119.63, 116.67, 115.96, 109.76, 48.51, 32.09, 21.01. HRMS (ESI-TOF) calcd for $C_{10}H_{13}N_2Cl$ $[M + H]^+$: 209.0846 and 211.0818, found: 209.0808 and 211.0782.

7-chlorotryptophol (5): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. Purification by flash column chromatography (SiO_2, 30% ethyl acetate/hexanes) afforded **5** in 95% yield (11.5 mg, 0.059 mmol). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.28 (s, 1H), 7.53 (d, *J* = 7.9 Hz, 1H), 7.21 (d, *J* = 7.6 Hz, 1H), 7.15 (s, 1H), 7.07 (t, *J* = 7.8 Hz, 1H), 5.3 (s, 1H), 3.91 (t, *J* = 6.4 Hz, 2H), 3.03 (t, *J* = 6.4 Hz, 2H). ^{13}C NMR (126 MHz, Chloroform-*d*) δ 133.66, 128.91, 123.03, 121.54, 120.26, 117.51, 116.72, 113.64, 62.61, 28.79. HRMS (ESI-TOF) calcd for $C_{10}H_{11}NOCl$ $[M + H]^+$: 196.0529 and 198.0502, found: 196.0489 and 198.0465.

7-chloro-2-methyltryptamine (6): The bioconversion was conducted according to the general procedure in a 100 x 50 mm crystallization dish. Substrate was added in a solution of isopropanol

(5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2 and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-NH). The crude material was purified by normal phase chromatography (gradient from 100% CH_2Cl_2 to 5% MeOH/ CH_2Cl_2) to afford **6** in 60% yield (7.2 mg, 0.034 mmol). ^{1}H NMR (500 MHz, Acetonitrile-d_3) δ 9.25 (s, 1H), 7.42 (d, J = 7.8 Hz, 1H), 7.06 (d, J = 8.3 Hz, 1H), 6.96 (t, J = 7.7 Hz, 1H), 2.81 – 2.72 (m, 4H), 2.39 (s, 3H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 133.89, 131.93, 130.39, 119.29, 119.08, 116.47, 114.88, 109.65, 42.71, 28.28, 11.26. HRMS (ESI-TOF) calcd for $C_{11}H_{14}N_2Cl$ $[M + H]^+$: 209.0846 and 211.0818, found: 209.0788 and 211.0762.

6-chloro-5-methyltryptamine (7): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After the reaction was quenched with acid as described above, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 25% CH_3CN/water 0.1%TFA) to afford **7** in 58% yield (10.7 mg of **7**·TFA, 0.033 mmol). ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.00 (s, 1H), 7.87 (s, 3H), 7.49 (s, 1H), 7.41 (s, 1H), 7.23 (s, 1H), 3.06 (s, 2H), 2.98 – 2.90 (m, 2H), 2.39 (s, 3H). ^{13}C NMR (126 MHz, MeOD–d_4) δ 137.36, 129.48, 127.41, 127.17, 125.24, 120.23, 112.63, 110.06, 41.14, 24.39, 20.52. HRMS (ESI-TOF) calcd for $C_{11}H_{13}N_2Cl$ $[M + H]^+$: 209.0846 and 211.0818, found: 208.9847 and 210.9836.

6-chloro-tryptoline (8-a) and 7-chloro-tryptoline (8-b): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated products was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 30% CH_3CN/water 0.1%TFA) to afford a mixture of **8-a** and **8-b** in 54% yield (10.1 mg of **8-a**·TFA and **8-b**·TFA, 0.031 mmol). ^{1}H NMR (500 MHz, Methanol-d_4) δ 7.46 (d, J = 1.9 Hz, 0.56H), 7.43 (d, J = 8.4 Hz, 0.39H), 7.35 (d, J = 1.8 Hz, 0.37H), 7.31 (d, J = 8.6 Hz, 0.56H), 7.10 (dd, J = 8.6, 2.0 Hz, 0.57H), 7.03 (dd, J = 8.4, 1.8 Hz, 0.39H), 4.43 (m, 2H), 3.57 (t, J = 6.1 Hz, 2H), 3.05 (m, 2H). ^{13}C NMR (126 MHz, Methanol-d_4) δ 138.55, 136.56, 129.17, 128.60, 128.28, 127.52, 126.26, 126.21, 123.44, 121.04, 120.02, 118.47, 113.51, 112.18, 107.29, 106.86, 43.59, 42.11, 42.08, 19.32, 19.30. HRMS (ESI-TOF) calcd for $C_{11}H_{12}N_2Cl$ $[M + H]^+$: 207.0689 and 209.0662, found: 207.0633 and 209.0610.

3-chloroindole[35] (9): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. Purification by flash column chromatography (SiO_2, 5% ethyl acetate/hexanes) afforded **9** in 53% yield (6.9 mg, 0.046 mmol). ^{1}H NMR (500 MHz, Methylene Chloride-d_2) δ 8.25 (s, 1H), 7.61 (d, J = 7.9 Hz, 1H), 7.41 (d, J = 8.2 Hz, 1H), 7.25 (t, J = 8.2 Hz, 1H), 7.22 (d, J = 2.6 Hz, 1H), 7.19 (t, J = 8.0 Hz, 1H). ^{13}C NMR

(126 MHz, Methylene Chloride-d_2) δ 135.38, 125.61, 123.42, 121.43, 120.76, 118.28, 111.92, 106.40. GCMS (EI) calcd for C_8H_6NCl [M]: 151.0 and 153.0, found: 151.0 and 153.0.

1-chloro-2-aminonaphthalene (10): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. Purification by flash column chromatography (SiO_2, 1% ethyl acetate/hexanes) afforded **10** in 93% yield (11.5 mg, 0.065 mmol). ^{1}H NMR (500 MHz, Methylene Chloride-d_2) δ 8.03 (d, J = 8.5 Hz, 1H), 7.72 (d, J = 8.1 Hz, 1H), 7.61 (d, J = 8.7 Hz, 1H), 7.51 (t, J = 8.3 Hz, 1H), 7.30 (t, J = 8.1 Hz, 1H), 7.04 (d, J = 8.7 Hz, 1H), 4.37 (s, 3H). ^{13}C NMR (126 MHz, Methylene Chloride-d_2) δ 141.18, 132.13, 128.76, 128.43, 128.11, 127.78, 123.13, 122.38, 118.15, 111.38. HRMS (ESI-TOF) calcd for $C_{10}H_9NCl$ $[M + H]^+$: 178.0424 and 180.0396, found: 178.0376 and 180.0350.

2-chloronaphthol[36] (11): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. Purification by flash column chromatography (SiO_2, hexanes) afforded **11** in 59% yield (7.3 mg, 0.041 mmol). ^{1}H NMR (500 MHz, Methylene Chloride-d_2) δ 8.24 – 8.16 (m, 1H), 7.80-7.82 (m, 1H), 7.57 – 7.48 (m, 2H), 7.44 – 7.35 (m, 2H), 6.10 (s, 2H). ^{13}C NMR (126 MHz, Methylene Chloride-d_2) δ 133.63, 127.97,

127.05, 126.48, 126.28, 124.76, 122.21, 121.30. GCMS (EI) calcd for $C_{10}H_7OCl$ [M]: 178.0 and 180.0, found: 178.0 and 180.0.

7-bromo-L-tryptophan (12): The bioconversion was conducted according to the general procedure in a 100 x 50 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, submitted to strong cation exchange chromatography, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford **12** in 85% yield (11.8 mg, 0.042 mmol). ^{1}H NMR (500 MHz, Deuterium Oxide) δ 7.54 (d, *J* = 7.9 Hz, 1H), 7.28 (d, *J* = 7.6 Hz, 1H), 7.19 (s, 1H), 6.93 (t, *J* = 7.8 Hz, 1H), 3.80 (dd, *J* = 7.8, 4.8 Hz, 1H), 3.27-3.23 (m, 1H), 3.11-3.07 (m, 1H). ^{13}C NMR (126 MHz, Deuterium Oxide) δ 173.23, 136.44, 129.78, 127.50, 126.23, 122.48, 119.33, 109.55, 106.14, 54.81, 27.45. HRMS (ESI-TOF) calcd for $C_{11}H_{12}N_2O_2Br$ $[M + H]^+$: 283.0082 and 285.0063, found: 283.0017 and 285.0015.

2.4.4 HANES-WOOLF PLOTS

Figure 2.8. Hanes-Woolf plot of RebH conversion of L-tryptophan.

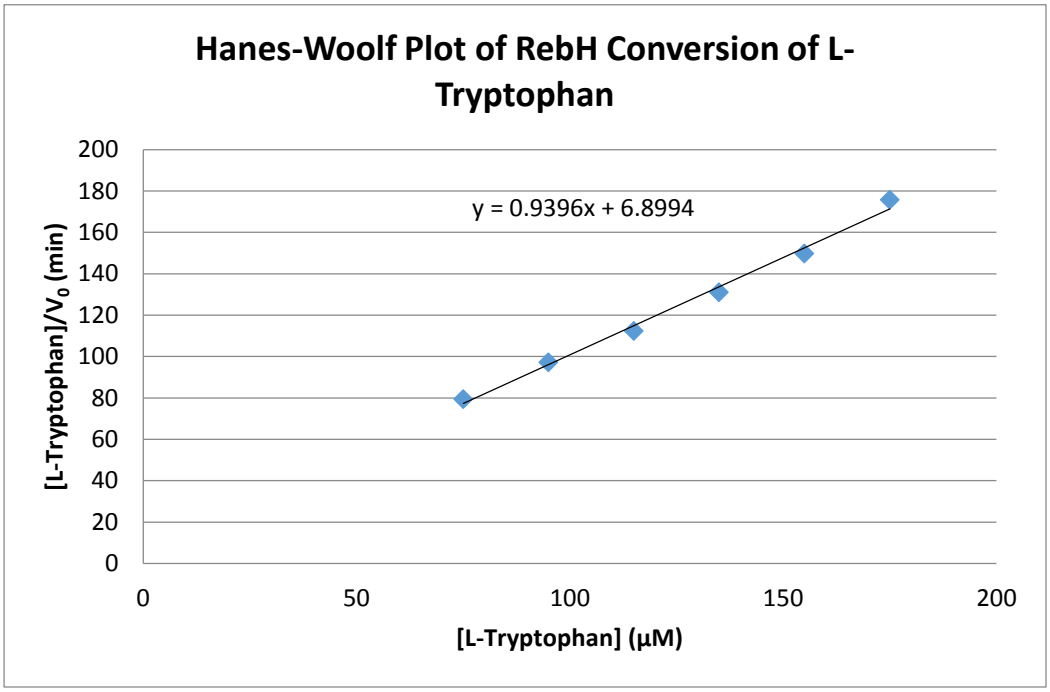


Figure 2.9. Hanes-Woolf plot of RebH conversion of tryptamine.

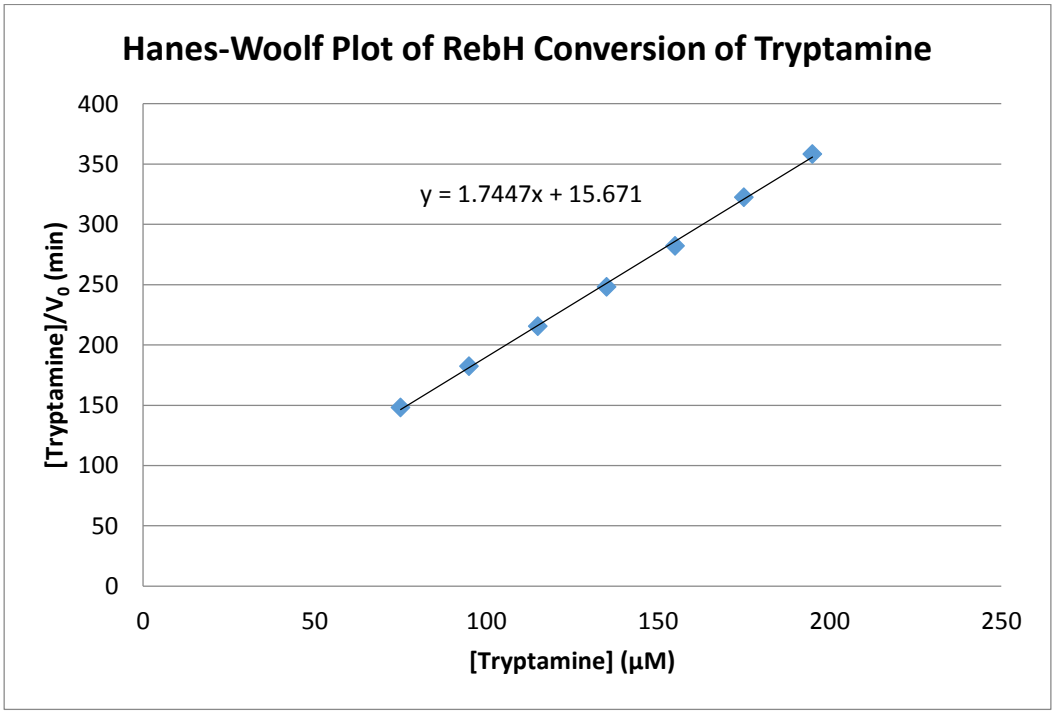


Figure 2.10. Hanes-Woolf plot of RebH conversion of tryptoline.

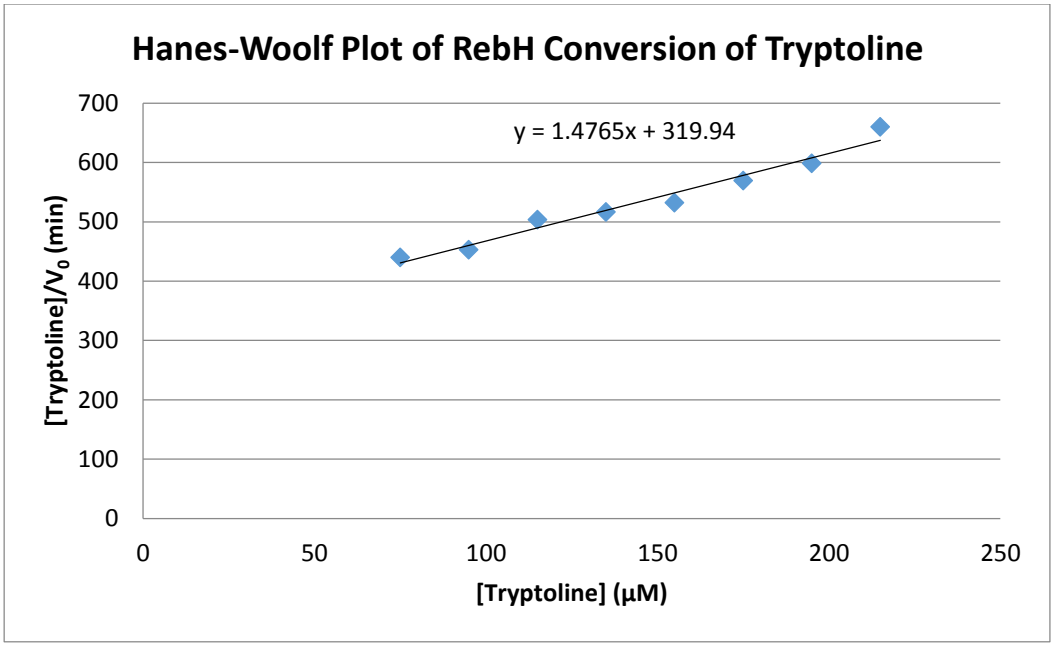
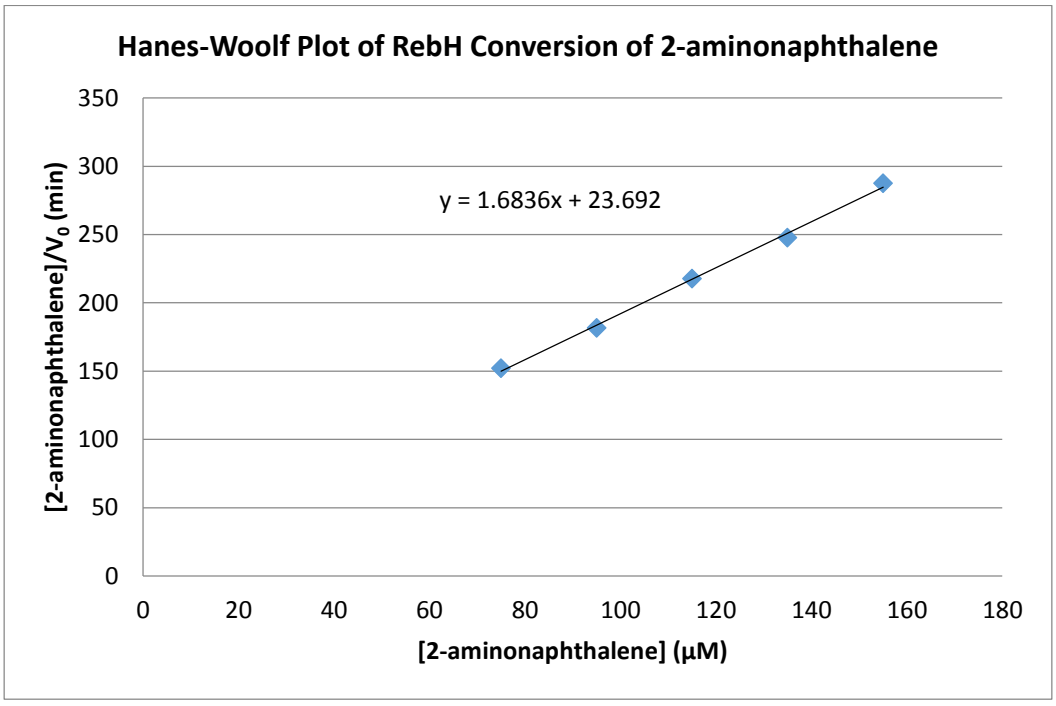


Figure 2.11. Hanes-Woolf plot of RebH conversion of 2-aminonaphthalene.



2.4.5 REPRESENTATIVE REACTION TIME COURSES

Figure 2.12. Time course of 10 mg RebH-catalyzed bioconversion of D-tryptophan.

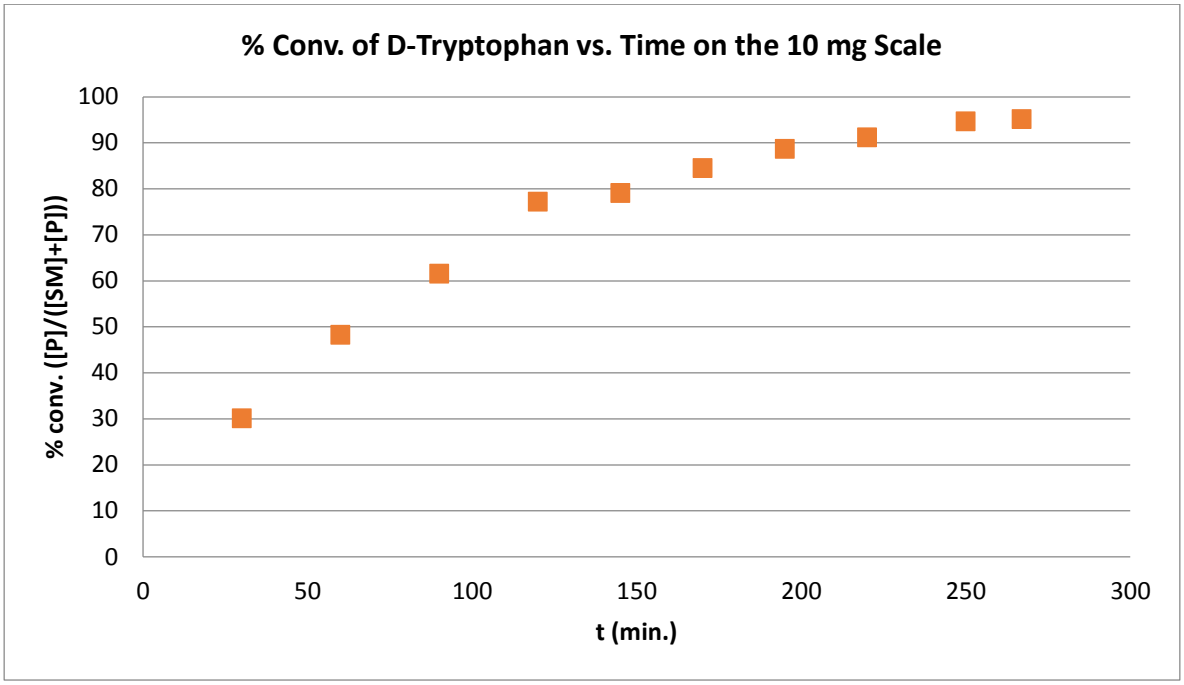
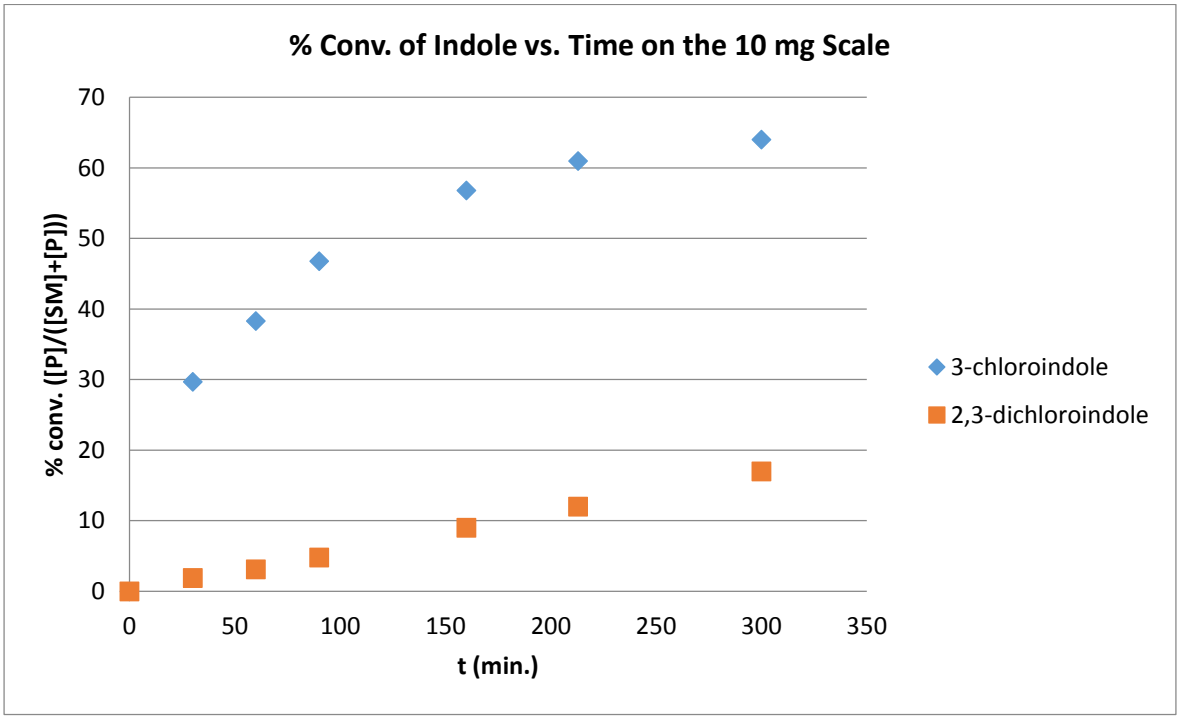


Figure 2.13. Time course of 10 mg RebH-catalyzed bioconversion of indole.



2.5 ACKNOWLEDGEMENTS

We would like to thank Mary Andorfer for her work and discussions that made Section 2.2.5 of this thesis possible. We would also like to thank Sarah Iqbal for her help in optimizing lysis conditions. We would also like to thank the Bottomley group for the donation of the fusion protein plasmids, the Walsh group for the donation of pET28/RebH and pET28/RebF, and the van Pée group for the donation of the PrnA plasmid. Lastly, we would like to thank the Rawal and Yamamoto groups for the donation of several of the screened substrates.

2.6 REFERENCES

[1] Metal-Catalyzed Cross-Coupling Reactions (Eds.: Diederich, F.; Stang, P. J.), Wiley-VCH, New York, 1998.

[2] Herrera-Rodriguez, L. N.; Khan, F.; Robins, K. T.; Meyer, H.-P. Perspectives on biotechnological halogenation Part I: Halogenated products and enzymatic halogenation. *Chem. Today* **29**, 31 (2011).

[3] a) 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.* **107**, 6652-6658 (1985); b) Bunders, C. A.; Minvielle, M. J.; Worthington, R. J.; Ortiz, M.; Cavanagh, J.; Melander, C. Intercepting Bacterial Indole Signaling with Flustramine Derivatives. *J. Am. Chem. Soc.* **133**, 20160 (2011).

[4] a) Smith, B. M.; Smith, J. M.; Tsai, J. H.; Schultz, J. A.; Gilson, C. A.; Estrada, S. A.; Chen, R. R.; Park, D. M.; Prieto, E. B.; Gallardo, C. S.; Sengupta, D.; Dosa, P. I.; Covel, J. A.; Ren, A.; Webb, R. R.; Beeley, N. R. A.; Martin, M.; Morgan, M.; Espitia, S.; Saldana, H. R.; Bjenning, C.; Whelan, K. T.; Grottick, A. J.; Menzaghi, F.; Thomsen, W. J. Discovery and Structure-Activity Relationship of (1*R*)-8-Chloro-2,3,4,5-tetrahydro-1-methyl-1*H*-3-benzazepine (Lorcaserin), a Selective Serotonin 5-HT_{2C} Receptor Agonist for the Treatment of Obesity. *J. Med. Chem.* **51**, 305-313 (2008); b) Sun, H.; Keefer, C. E.; Scott, D. O. Systematic and Pairwise Analysis of the Effects of Aromatic Halogenation and Trifluoromethyl Substitution on Human Liver Microsomal Clearance. *Drug Metab. Lett.* **5**, 232-242 (2011).

[5] "Formation of Carbon-Halogen Bonds (Cl, Br, I)": Sasson, Y. in *Patai's Chemistry of Functional Groups*, Wiley-VCH, New York, **2009**.

[6] Smith, K.; El-Hiti, G. A. Regioselective control of electrophilic aromatic substitution reactions. *Curr. Org. Synth*. **1**, 253 – 274 (2004).

[7] Podgorsek, Z.; Zupan, M.; Iskra, J. Oxidative halogenation with "green" oxidants: oxygen and hydrogen peroxide. *Angew. Chem. Int. Ed.* **48**, 8424 – 8450 (2009).

[8] Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; Garneau-Tsodikova, S.; Walsh, C. T. Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. *Chem. Rev.* **106**, 3364-3378 (2006).

[9] Dairi, T.; Nakano, T.; Aisaka, K.; Katsumata, R.; Hasegawa, M. Cloning and Nucleotide Sequence of the Gene Responsible for Chlorination of Tetracycline. *Biosci. Biotech. Biochem.* **59**, 1099-1106 (1995).
[10] Flecks, S.; Patallo, E. P.; Zhu, X.; Ernyei, A. J.; Seifert, G.; Schneider, A.; Dong, C.; Naismith, J. H.; van Pée, K.-H. New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angew. Chem. Int. Ed.* **47**, 9533-9536 (2008).
[11] 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* **46**, 1284-1292 (2007).
[12] Lang, A.; Polnick, S.; Nicke, T.; William, P.; Patallo, E. P.; Naismith, J. H.; van Pée, K.-H. Changing the Regioselectivity of the Tryptophan 7-Halogenase PrnA by Site-Directed Mutagenesis. *Angew. Chem. Int. Ed.* **50**, 2951-2953 (2011).
[13] a) Runguphan, W.; Qu, X.; O'Connor, S. E. Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461-464 (2010); b) Roy, A. D.; Grüschow, S.; Cairns, N.; Goss, R. J. M. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J. Am. Chem. Soc.* **132**, 12243–12245 (2010); c) Sánchez, C.; Zhu, L.; Braña, A. F.; Salas, A. P.; Rohr, J.; Méndez, C.; Salas, J. A. Combinatorial biosynthesis of antitumor indolocarbazole compounds. *Proc. Natl. Acad. Sci.* **102**, 461-466 (2005).
[14] Payne, J. T.; Andorfer, M. C.; Lewis, J. C. Regioselective Arene Halogenation using the FAD-Dependent Halogenase RebH. *Angew. Chem. Int. Ed.* **52**, 5271-5274 (2013). Note: we previously referred to the 6- and 7-halogenation of tryptoline as 5- and 6-halogenation, by analogy to the numbering used for L-tryptophan, but here we use the established numbering for carbazole derivatives.
[15] 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.* **102**, 3960-3965 (2005).
[16] Keller, S.; Wage, T.; Hohaus, K.; Hölzer, M.; Eichhorn, E.; van Pée, K.-H. Purification and Partial Characterization of Tryptophan 7-Halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew. Chem. Int. Ed.* **39**, 2300-2302 (2000).
[17] Zehner, S.; Kotzsch, A.; Bister, B.; Süssmuth, R. D.; Méndez, C.; Salas, J. A.; van Pée, K.-H. A Regioselective Tryptophan 5-Halogenase Is Involved in Pyrroindomycin Biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* **12**, 445-452 (2005).
[18] Seibold, C.; Schnerr, H.; Rumpf, J.; Kunzendorf, A.; Hatscher, C.; Wage, T.; Ernyei, A. J.; Dong, C.; Naismith, J. H.; van Pée, K.-H. A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocat. Biotrans.* **24**, 401-408 (2006).
[19] Heemstra, Jr., J. R.; Walsh, C. T. Tandem Action of the O_2- and $FADH_2$-Dependent Halogenases KtzQ and KtzR Produce 6,7-Dichlorotryptophan for Kutzneride Assembly. *J. Am. Chem. Soc.* **130**, 14024-14025 (2008).
[20] Riley, M.; Abe, T.; Arnaud, M. B.; Berlyn, M. K.; Blattner, F. R.; Chaudhuri, R. R.; Glasner, J. D.; Horiuchi, T.; Keseler, I. M.; Kosuge, T.; Mori, H.; Perna, N. T.; Plunkett III, G.; Rudd, K. E.; Serres, M. H.; Thomas, G. H.; Thomson, N. R.; Wishart, D.; Wanner, B. L. Escherichia coli K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Res.* **34**, 1-9 (2006).
[21] Esposito, D.; Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17**, 353-358 (2006).
[22] Cabrita, L. D.; Dai, W.; Bottomley, S. P. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* **6**, 12 (2006).

[23] Li, M. Z.; Elledge, S. J. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat. Meth.* **4**, 251-256 (2007).
[24] Sørensen, H. P.; Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* **4**, 1 (2005).
[25] Glenn, W. S.; Nims, E.; O'Connor, S. E. Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *J. Am. Chem. Soc.* **133**, 19346-19349 (2011).
[26] Kapust, R. B.; Waugh, D. S. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668-1674 (1999).
[27] Nominé, Y.; Ristriani, T.; Laurent, C.; Lefévre, J. F.; Weiss, E.; Travé, G. Formation of soluble inclusion bodies by HPV E6 oncoprotein fused to maltose-binding protein. *Protein Expr. Purif.* **23**, 22-32 (2001).
[28] Hölzer, M.; Burd, W.; Reißig, H.-U.; van Pée, K.-H. Substrate Specificity and Regioselectivity of Tryptophan 7-Halogenase from *Pseudomonas fluorescens* BL915. *Adv. Synth. Catal.* **343**, 591-595 (2001).
[29] Glenn, W. S.; Nims, E.; O'Connor, S. E. Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *J. Am. Chem. Soc.* **133**, 19346-19349 (2011).
[30] Lewis, J.C.; Arnold, F. H. Catalysts on Demand: Selective Oxidations by Laboratory-Evolved Cytochrome P450 BM3. *Chimia* **6**, 309-312 (2009).
[31] Bitto, E.; Bingman, C. A.; Singh, S.; Phillips, Jr., G. N. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70**, 289–293 (2008).
[32] Dong, C.J.; Flecks, S.; Unversucht, S.; Haupt, C.; van Pée, K.-H.; Naismith, J. H. Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science* **309**, 2216 – 2219 (2005).
[33] Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem*. **62**, 7512-7515 (1997).
[34] Sambrook, J.; Frisch, E. F.; Maniatis, T. (**1989**) *Molecular cloning: a laboratory manual*. 2, Cold Spring Harbor
Laboratory Press, New York.
[35] Wagger, J.; Groselj, U.; Svete, J.; Stanovnik, B. Synthesis of Racemic, N-Benzylated Neoechinulin A and Isoneoechinulin A1. *Synlett* **8**, 1197-1200 (2010).
[36] Ding, Z.; Xue, S.; Wulff, W.D. A Succinct Synthesis of the Vaulted Biaryl Ligand Vanol via a Dienone–Phenol Rearrangement. *Chemistry – An Asian Journal* **6**, 2130-2146 (2011).

CHAPTER III

THE ENGINEERING OF REBH VARIANTS WITH EXPANDED SUBSTRATE SCOPE

3.1 INTRODUCTION

Halogenated aromatic compounds are essential building blocks in chemical synthesis,[1] and as such, halogenation is important in the production of intermediates in chemical synthesis. Halogen substituents are also found in end-product pharmaceuticals[2] and agrochemicals;[3] an estimated one quarter of all such compounds are halogenated.[4] In these cases, the halogen does not serve as a handle for further functionalization, but instead the presence of the halogen substituent itself significantly impacts molecular pharmacology, which is highlighted by the effects of chlorination or bromination on a diverse range of drugs spanning antibiotics,[5] anticancer agents,[6] and psychoactive compounds[7] (see Figure 1.1). It has also been demonstrated that the presence of a halogen substituent can greatly alter drug metabolism. In one study of the microsomal clearance of over 220,000 compounds, chlorination was found to predictably increase or decrease the metabolic half-life of a drug, depending on the location of the substitution.[8]

Despite the importance of halogen substituents, conventional arene halogenation methods, such as those proceeding via electrophilic aromatic substitution (EAS), often suffer from poor regioselectivity and require harsh reaction conditions.[9] It is therefore notable that flavin-dependent halogenase (FDH) catalyzed arene halogenation, proposed to proceed via EAS,[10] provides high regioselectivity using air and halide salts as the terminal oxidant and halogen source, respectively. As discussed in Chapter 2, the tryptophan 7-halogenase RebH,[11] a FDH from the rebeccamycin

biosynthetic pathway, can halogenate a range of indoles and naphthalenes on a preparative scale.[12] In contrast to previous reports for another FDH, PrnA,[13] RebH was found to halogenate many of these substrates at sites other than those most electronically activated, providing regioselectivity that would not be obtained from conventional EAS. The unnatural substrates halogenated by RebH in both our[12] and others'[14] work were, however, comparable in size to the native substrate, tryptophan (**1**, Figure 3.2), and increasing structural differences led to lower turnover numbers and less favorable kinetic parameters.[12] Given a recent report in which cross-linked RebH aggregates were used for gram scale halogenation,[15] the narrow substrate scope and low activity of RebH on unnatural substrates stood as key limitations to its general utility for preparative halogenation.

Several groups have used RebH and other FDHs in metabolic engineering efforts that involve typtophan halogenation and conversion of halogenated tryptophan into halogenated natural product derivatives.[16] A point mutant of RebH has been reported to preferentially halogenate tryptamine over tryptophan, and direct halogenation of this smaller alkaloid precursor has aided the aforementioned metabolic engineering efforts.[17] On the other hand, no examples in which FDHs are used for late stage halogenation of complex, biologically active compounds on a preparative scale have been reported, presumably due to substrate scope limitations. We sought to take advantage of the fact that wild-type RebH will halogenate several non-native but small substrates regioselectively and engineer variants of RebH to functionalize significantly larger, bioactive substrates while preserving the ability for site-selective halogenation. Large portions of this work were performed with Dr. Catherine Poor and Dr. Landon Durak, and together two papers have been written describing these results.[18]

3.2 RESULTS AND DISCUSSION

3.2.1 A BACKGROUND TO THE EVOLUTION OF THERMOSTABLE REBH VARIANTS

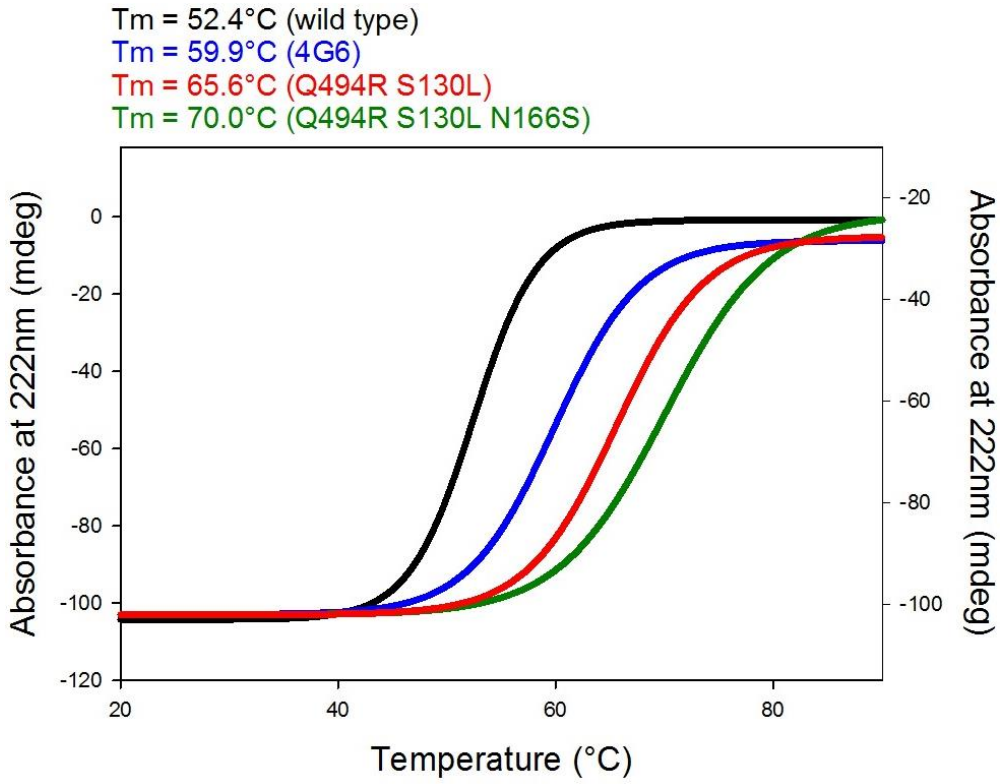
As we intended to evolve RebH to improve its substrate scope and other properties, we expected to mutate numerous amino acid residues in RebH. Broadly speaking, the stability of an enzyme correlates with its evolvability.[19] Most amino acid mutations will tend to be destabilizing to the protein structure, and mutations made for the purpose of expanding substrate scope are no exception. Thus, as we introduce progressively more scope-expanding mutations into RebH, we can expect its stability to decrease until it may reach a point where it rapidly degrades in the necessary reaction conditions, even if they are remarkably mild. One solution to this quandary is to introduce stabilizing mutations prior to evolving the desired trait to compensate preemptively for the expected destabilization that will occur.

While the preparative scale reactions described in Section 2.2.5 were being performed, the directed evolution of RebH to improve its thermostability was performed by a postdoctoral research in our lab, Dr. Catherine Poor.[20] Random mutagenesis was performed on the entirety of the RebH gene, and just over 1000 colonies were picked in each of three rounds of evolution. After the colonies were picked, arrayed into culture plates, expressed, and lysed to furnish the crude extract of the RebH variants, this lysate was subjected to a heat pretreatment of increasing temperature in each round of evolution. The lysates were then screened for residual enzyme activity, and those that were least disrupted by the heat pretreatment and thus gave the highest residual activity were picked, characterized, and used as the parent in the following round of mutagenesis. In this way, the thermostability of RebH was increased dramatically in just three rounds of evolution; by introducing eight amino acid mutations, the T_m (the midpoint of the

thermal unfolding transition curve as measured by circular dichroism) of the most stabilized variant was increased by nearly 18 °C relative to wild-type RebH (Figure 3.1). The thermostabilization of RebH was significant both as the first example of the directed evolution of RebH – and, for that matter, of any FDH - and importantly furnished stabilized mutants that could serve as starting points for future evolutionary lineages.

Scheme 3.1. Evolutionary lineage of thermostabilized RebH variants.

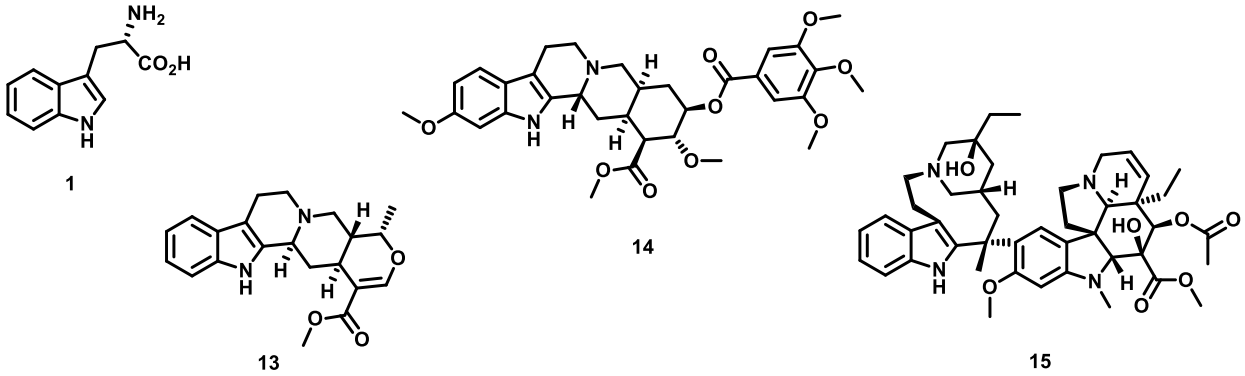
Figure 3.1. Thermal unfolding transition curves of thermostabilized RebH variants.[20]



3.2.2 THE EVOLUTION OF A REBH VARIANT TO HALOGENATE A BACTERIAL BIOFILM INHIBITOR

As described in sections 2.2.4, 2.2.5, and 3.1, the substrates that RebH has been reported to functionalize, while relatively diverse for a wild-type enzyme, are limited largely to substrates comparable in size or smaller than the native substrate, L-tryptophan (**1**). Many pharmaceuticals, however, are significantly larger than L-tryptophan. For example, while L-tryptophan has a molecular weight of 204 daltons (Da), the top five best-selling pharmaceuticals in 2012 were the proton pump inhibitor esomeprazole (Nexium), the antipsychotic aripiprazole (Abilify), the statin rosuvastatin (Crestor), the corticosteroid fluticasone propionate (Advair Diskus), and the serotonin-norepinephrine reuptake inhibitor duloxetine (Cymbalta), which have molecular weights of 345, 448, 482, 501, and 297 Da, respectively.[21] If we look specifically at drugs that bear structural similarity to L-tryptophan, namely drugs that contain an indole moiety, we see that many of the most important indole drugs are also much larger than L-tryptophan. Three specific examples of such drugs are the antihypertensive ajmalicine (**13**), the antipsychotic reserpine (**14**), and the anticancer drug vinblastine (**15**).[22] These three compounds have molecular weights of 352, 609, and 811, respectively, all significantly higher than L-tryptophan, and all with significant added complexity extending off of the 2- and 3-positions of the indole moiety (Figure 3.2).

Figure 3.2. Structures of bioactive indoles.



For the reasons discussed in Section 3.1 as well as at greater length in Section 1.1, the halogenation of bioactive indoles (such as those shown in Figure 3.2) potentially adds significant value to these compounds. The ability to site-selectively modify these compounds at late-stage is especially valuable when compared with halogenating an early synthetic intermediate and then performing difficult and laborious synthetic manipulations to arrive at a halogenated derivative of the target of interest. However, the fact that wild-type RebH is only known to work on substrates comparable in size to or smaller than L-tryptophan precludes such late-stage functionalization of large substrates. To overcome this limitation, we sought to engineer RebH to accept larger, bioactive indoles as substrates, particularly focusing on expanding the active site of RebH to allow the added complexity in the 2- and 3-positions that is seen in many bioactive indoles.

As an initial target we wished to functionalize with RebH variants with expanded substrate scope, we selected the debrominated analogue (**3**, Figure 3.3) of deformylflustrabromine (dFBr).[5] Deformylflustrabromine, a tryptamine metabolite first isolated from *Flustra foliacea*, is brominated at the indole 6-position, and it has been shown that the presence of this bromine substituent is necessary to evoke the biological activity of dFBr, which is its ability to inhibit the growth of bacterial biofilms. Bacterial biofilms play a significant role in the pathology of many

human diseases, especially intransigent infections associated with the implantation of medical devices.[23] Without the presence of this bromine substituent at the indole 6-position, no inhibition of the growth of bacterial biofilms is observed. However, to the best of our knowledge, no one has prepared and explored the biological activities of the 4-, 5-, or 7-brominated analogues, or any of the chlorinated analogues, likely due in part to the difficulty of preparing these site-selectively halogenated derivatives. We therefore saw the development of engineered halogenases capable of site-selective halogenation of the debrominated analogue of dFBr to be of use in the production and biological evaluation of these derivatives, while this compound offered a prime target for expansion of the RebH active site to accommodate the bulky inverse prenyl substituent at its 2-position.

Figure 3.3. Substrate walking progression to expand substrate scope of RebH.

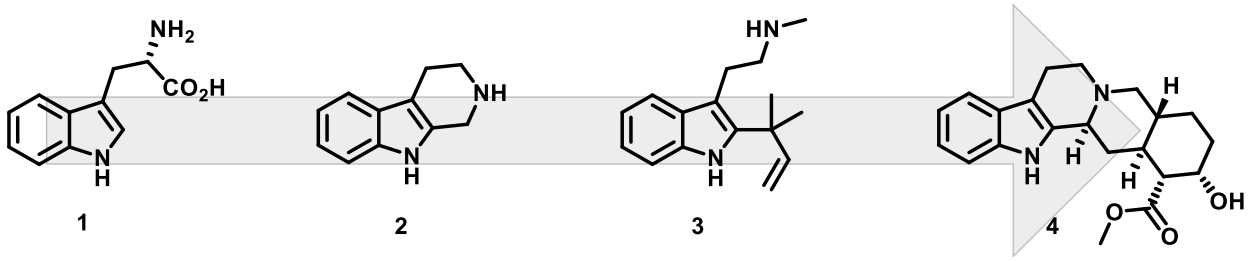


Table 3.1. Evolutionary lineage and conversions of directed evolution of RebH.

Generation	Enzyme	Mutation[a]	Method (substrate)[b]	%Conv. **1** (%Loading)[c]	%Conv. **2** (%Loading)	%Conv. **3** (%Loading)	%Conv. **4** (%Loading)
0	RebH	-	-	53 (0.2)	3 (0.5)	<1 (5)	1 (5)
1	**1-PVM**	S2P/M71V/K145M	epPCR[d] (**1**)	43 (0.2)	5 (0.5)	<1 (5)	5 (5)
2	**2-T**	N467T	epPCR (**1**)	53 (0.2)	7 (0.5)	<1 (5)	9 (5)
3	**3-SS**	G112S/N470S	epPCR (**2**)	22 (0.2)	64 (0.5)	6 (5)	9 (5)
3	**3-S**	S112G	point mut.(**3**)[e]	43/5[f] (0.2)	50 (0.5)	29 (5)	39 (5)
4	**4-V**	A442V	epPCR (**3**)	39/3[f] (0.2)	43 (0.5)	48 (5)	38 (5)

[a] Mutations relative to the parent on the previous line. [b] Method used to introduce mutations and (substrate used in the screening effort noted). [c] Conversions of substrates determined by UPLC and (mol% loadings of enzymes used). [d] Error-prone PCR. [e] Point mutation introduced via SOE PCR.[24] [f] Second number refers to conversion to dihalogenated product.

We found that RebH variant 1-PVM from our thermostability lineage provided increased conversion of numerous substrates, including tryptamine and tryptoline (**2**, Figure 3.3). Initially, we intended to further improve 1-PVM activity on **2**, as we hoped that variants with improved activity on this tricyclic molecule would also have activity on larger tryptophan-based natural products, including **3**. Both RebH and 1-PVM, however, provided insufficient conversion of **2** for reliable UPLC analysis under the conditions used for high throughput screening (crude lysate in 96 well plates). Because 1-PVM is both more stable and has nearly twofold higher activity on **2** than RebH (Table 3.1), it was used as the parent for our first round of evolution. We hoped that improving 1-PVM activity on **1**, which was similar to that of RebH and could be detected by UPLC, would provide variants with sufficient activity on **2** to enable further evolution using this substrate.

A library of 1-PVM variants, each containing an average of 1-2 amino acid mutations, was generated by error-prone PCR, and 1080 colonies were picked and expressed in 96 well plates. Halogenation of **1** was assayed using UPLC, and any variants that provided increased conversion

relative to 1-PVM included in each plate were submitted to a secondary screen for increased conversion of **2**. All hits were verified by gene sequencing and large scale reactions conducted using purified enzyme. From this library, one variant possessing a single N467T mutation (and therefore called 2-T) gave a modest 1.2-fold increase in conversion of **1**, and, when submitted to the secondary screen, showed a 1.3-fold increase in conversion of **2**. Although also modest, this improvement elevated the conversion of **2** to nearly 10% under the conditions used for screening. We found that at this level of conversion, we could reliably detect 1.5-fold increases in conversion of **2**, and thus decided to screen directly on **2** for our next round of mutagenesis.

The second round library was constructed as described above but using 2-T as a parent, and 1080 variants were again picked and expressed. Reactions were conducted using **2** as the substrate, and any variants that showed 1.5-fold or higher activity in lysate were grown and purified to confirm their activity. Several mutations giving activity above this threshold were found, and one of these, N470S, led to nearly 7-fold higher conversion of **2**. This mutation and a second mutation identified in this round of screening, G112S, were combined to give variant 3-SS, which provided 9-fold higher conversion of **2** relative to 2-T.

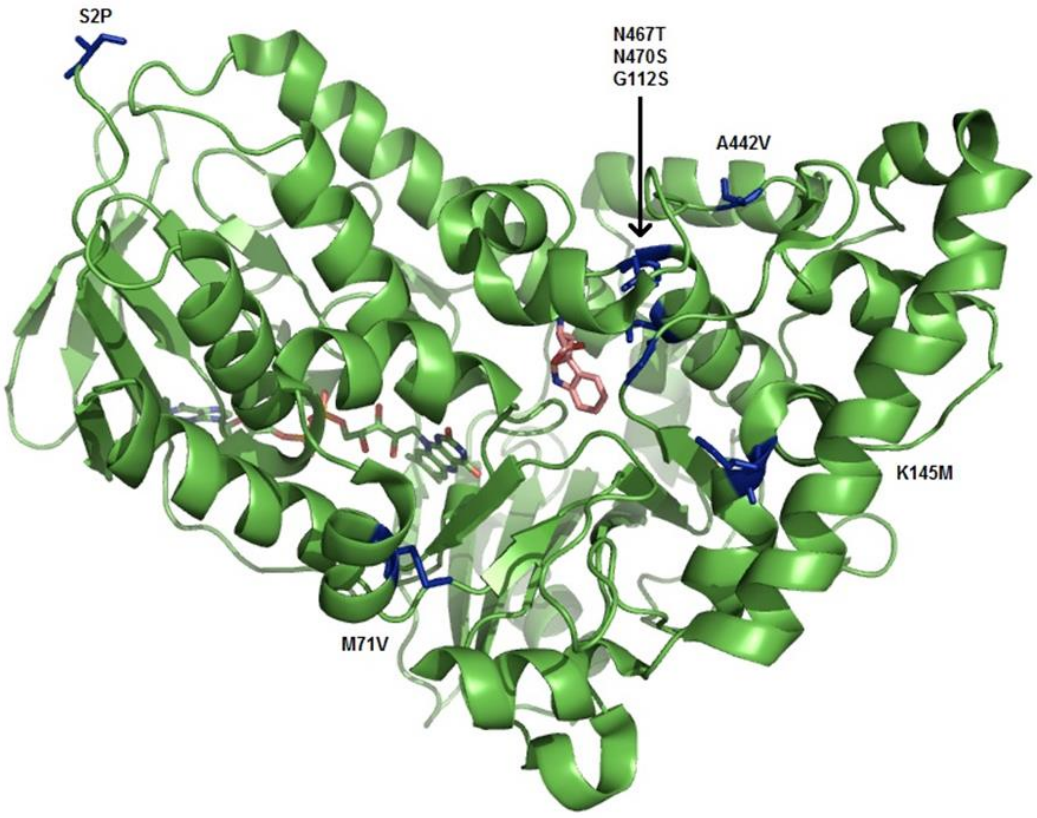
Many of the mutations found in 3-SS are in the enzyme active site (N467T, N470S, and G112S) and could impact the binding of **2** and thus the halogenation selectivity on this substrate.[21] Steady-state kinetic analysis of these variants was conducted to probe the effects of these mutations on kinetic parameters (Table 3.2). We found that the catalytic efficiency scaled accordingly with the increased conversion observed for successive variants. While this effect appears to have been solely due to increased k_{cat} for variants 1-PVM and 2-T, the final variant, 3-SS, possesses greatly increased k_{cat} and greatly decreased K_m, resulting in a catalytic efficiency nearly 70-fold greater than WT.

Table 3.2. Kinetic data for chlorination of tryptoline (2).

Variant	Mol %	k_{cat} [min^{-1}]	K_m [μM]	k_{cat}/K_m [min^{-1} μM^{-1}]
RebH	5	0.013	144	9.05×10^{-5}
1-PVM	5	0.040	215	1.86×10^{-4}
2-T	3	0.056	236	2.37×10^{-4}
3-SS	0.5	0.537	8.6	6.24×10^{-3}

At this point, in order to continue the evolution of RebH further to functionalize **3**, the aforementioned debrominated analogue of deformylflustrabromine, this compound was synthesized in five linear steps following reported literature procedures.[25] While wild-type RebH has no quantifiable activity on this compound, 3-SS did provide low conversion. Notably, we found that variant 3-S, which lacks the G112S mutation but still possesses N470S, gave 5-fold higher conversion of **3** than did 3-SS. As the conversion of **3** afforded by 3-S was sufficient to observe low activity in crude lysate in 96 well plates, we screened for improved activity on **3** for our third round of evolution just as was done on **2** for our second round. After verifying the hits with purified enzyme, one variant was found with nearly 1.7-fold increased activity on **3**. This variant, 4-V, possesses a single A442V mutation that is quite distant from the active site of the enzyme (Figure 3.4). It is well established that mutations distant from enzyme active sites can have profound effects on their activity,[26] and such mutations are especially difficult to find with structure-guided mutagenesis.

Figure 3.4. RebH structure and mutations to expand substrate scope.[a]



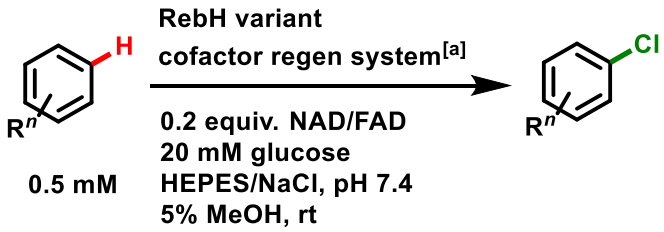
[a] A previously reported crystal structure of wild-type RebH was used (PDB entry 2OA1),[35] with residues that are mutated in variants 3-SS and 4-V shown in blue (bound L-tryptophan as substrate is shown in pink).

3.2.3 AN EXPLORATION OF THE SUBSTRATE SCOPE AND SELECTIVITIES OF ENGINEERED REBH VARIANTS

As described in section 2.2.5, we previously had explored the selectivities of wild-type RebH on small, non-native substrates by performing preparative scale bioconversions and characterizing the products. In the course of this analysis, we performed a 10 mg, preparative scale bioconversion of tryptoline (**2**) using 10 mol% wild-type RebH, furnishing 58% conversion to a

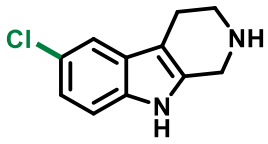
roughly equal mixture of the 6- and 7-chlorinated products. We performed a second preparative scale bioconversion of **2** on the same 10 mg scale with 3-SS, dropping the enzyme loading 10-fold to 1 mol%, and saw a significant increase in conversion to 93% (Scheme 3.2). Upon isolating and characterizing the product, it was found that 3-SS, in contrast to wild-type RebH, gave almost exclusively the 6-chlorinated product (Table 3.3). This shows that improved activity need not come at the expense of selectivity (or vice versa) and that both were simultaneously increased in this case.

Scheme 3.2. General scheme for RebH-variant-catalyzed arene halogenation.

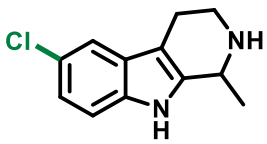


[a] The cofactor regeneration system consisted of 0.5 mol% MBP-RebF and 50 U ml glucose dehydrogenase. NAD = nicotinamide adenine dinucleotide, FAD = flavin adenine dinucleotide, MBP = maltose binding protein.

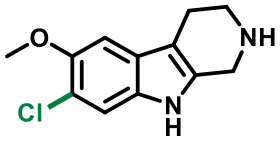
Table 3.3. Substrate scope of 3-SS.[a]



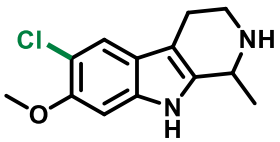
2
1% 3-SS, 78% (93)



5
1% 3-SS, 83% (89%)



6
2% 3-SS, 80% (98%)

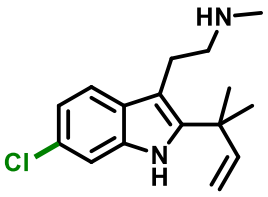


7
5% 3-SS, 79% (76%)

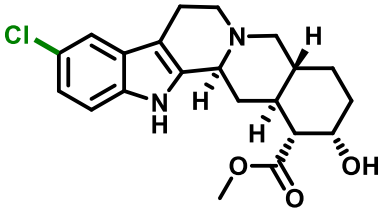
[a] Reactions were performed on preparative scale (10 mg) and selectivities observed with 3-SS, mol% enzyme loading, isolated yield, and (HPLC conversion) are shown.

While the activity of 3-SS on tryptoline was very substantially increased relative to that of wild-type RebH, and 4-V displayed significant activity on **3** whereas wild-type RebH gave no quantifiable activity, we also wished to explore if these engineered variants displayed increased breadth of scope. We began by examining the scope of 3-SS and found that it is particularly effective for halogenating tricyclic indoles similar to **2** (Table 3.3). Eleagnine (**5**), an alkaloid from *Chrysophyllum albidum* with potent analgesic effects,[27] has the structure of **2** with an additional methyl group at the 1-position. We found this added steric bulk distal from the indole moiety interesting, as we aimed to further expand the substrate scope of RebH to encompass compounds significantly larger in the area of this methyl group. Although RebH shows very low conversion of **5**, 3-SS has high activity on this compound - over 60-fold higher than that observed with wild-type RebH (Table 3.5). With only 1 mol% enzyme loading of 3-SS, 89% conversion of **5** was seen on the 10 mg scale. We found that 3-SS also gave good conversion of two tryptoline derivatives with substituents on the benzene ring of the indole moiety: pinoline (**6**), or 6-methoxytryptoline, a metabolite with monoamine oxidase A inhibition activity,[28] which was chlorinated solely at the 7-position; and tetrahydroharmine (**7**), or 7-methoxyeleagnine, an alkaloid from *Banisteriopsis caapi* with antiviral and antifungal activity,[29] which was chlorinated solely at the 6-position. These two substrates show that these RebH variants are able to tolerate substituents at multiple positions on the indole moiety, which occur frequently in biologically active natural products.

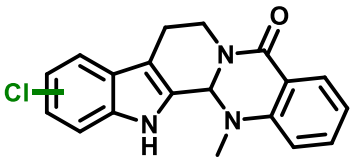
Table 3.4. Substrate scope of 4-V.[c]



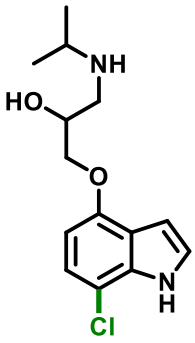
3
5% 4-V, 68% (72%)



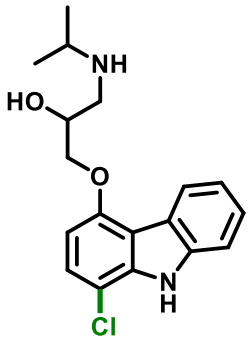
4
5% 4-V, 54% (56%)[a]



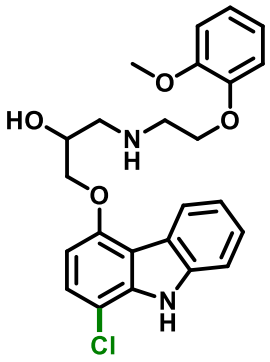
8
5% 4-V, (36%)[b]



9
1% 4-V, 93% (97%)



10
1% 4-V, 79% (91%)



11
2% 4-V, 69% (98%)

[a] In addition to the major product of 10-chlorination shown, 11-chlorination was also observed to afford a minor product. [b] Multiple closely eluting products were observed and were not individually isolated, so only total conversion is shown. [c] Reactions were performed on preparative scale (10 mg) and selectivities observed with 4-V, mol% enzyme loading, isolated yield, and (HPLC conversion) are shown.

We then used 4-V for preparative chlorination of **3**, and found that this enzyme provided exclusively 6-chlorination, providing a facile synthesis of the 6-chloro analogue of dFBr. The high activity of 4-V on **3**, despite the bulk of the inverse prenyl group, encouraged us to explore compounds that differed significantly from those halogenated by wild-type RebH and 3-SS. We first looked at two pentacyclic compounds with well studied biological activities: yohimbine (**4**), an alkaloid from *Catharansus roseus* that is also an α_2 adrenergic receptor antagonist;[30] and

evodiamine (**8**), an alkaloid from the *Tetradium* family of plants.[31] Both of these compounds are significantly larger than any others we had assayed to this point – **4** has over twice the molecular weight of **2** – but to our surprise, both showed significant conversion with 4-V (Table 3.4). While 4-V produces two distinct monochlorinated derivatives of **8**, as well as one dichlorinated derivative, only monochlorinated products were observed with **4**, and at a sufficient level of conversion for a preparative scale halogenation. After performing the preparative scale reaction, we observed that **4** is chlorinated to 56% conversion in a roughly 3:1 ratio at the 10- and 11-positions, respectively.

Given that we had demonstrated that RebH variants could accept greatly expanded bulk on the pyrrole moiety of these unnatural substrates, as well as substitution at the 5- and 6-positions on the indole moiety, we decided to explore the impact of large substitutions at the 4-position of the indole ring. Pindolol (**9**), a nonselective beta blocker,[32] possesses a sizeable substituent with alcohol and amine functionalities via an ether linkage at the indole 4-position. We found that 4-V was able to fully convert this compound to the 7-chlorinated product, even at lower (<1%) enzyme loadings. A similar compound, carazolol (**10**),[33] possesses a carbazole rather than an indole moiety, and we found that this compound was selectively halogenated at the analogous position. This was the first time we had observed selective conversion of a carbazole in high yield by any RebH variant, and encouraged by this result we decided to test carvedilol (**11**), another carbazole with an even bulkier substituent at the 5-position (analogous to the indole 4-position) that is a nonselective beta blocker/alpha-1 blocker.[34] Carvedilol has a molecular weight of over 400 g/mole, twice the size of the native tryptophan and even larger than **4**. Despite this added steric bulk, we found that 4-V was able to again give full conversion of **11** at only 2% enzyme loading, with exquisite selectivity to only a single chlorinated product. The confined nature of the RebH active site in

crystal structures with bound tryptophan substrate[35] makes rationalizing the binding of the larger substrates explored in this work difficult, but it may be that some of these, particularly 11, are not fully contained within the active site.

Table 3.5. Ratio of improvement to activity versus wild-type RebH.

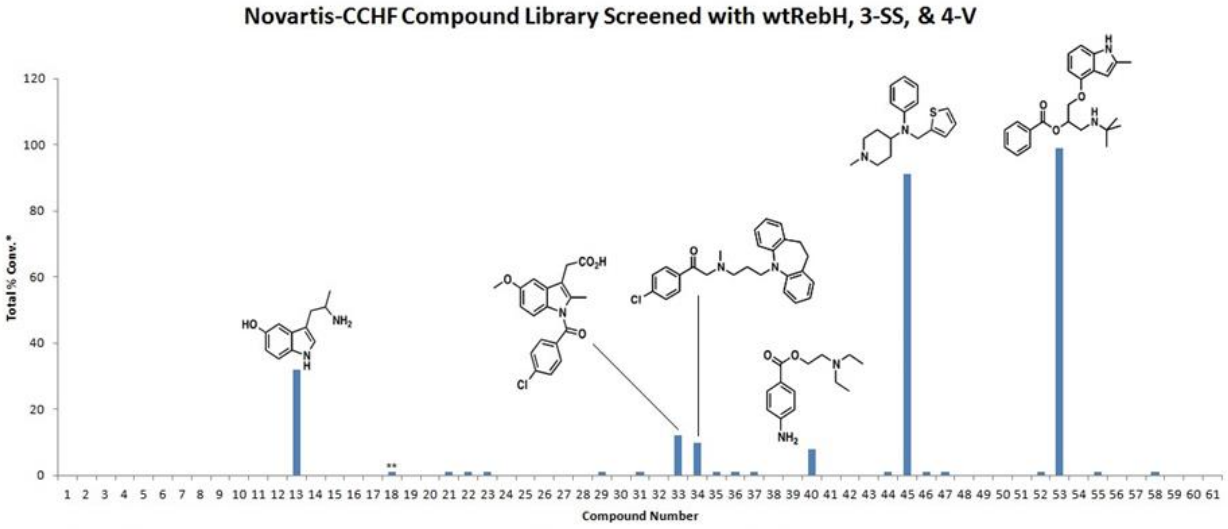
Substrate	Variant	Mol %	Activity Ratio[a]
Tryptoline (**2**)	3-SS	0.5	65.5
Eleagnine (**5**)	3-SS	0.5	67.1
Pinoline (**6**)	3-SS	0.5	2.0
Tetrahydroharmine (**7**)	3-SS	5	17.6
Desbromo-dFBr (**3**)	4-V	5	N/A[b]
Yohimbine (**4**)	4-V	5	38.0
Evodiamine (**8**)	4-V	5	16.5
Pindolol (**9**)	4-V	0.2	1.3
Carazolol (**10**)	4-V	0.2	4.9
Carvedilol (**11**)	4-V	0.5	8.2

[a] Activity ratio is ratio of conversion seen with variant tested vs WT. Reaction conditions were those shown in Scheme 3.2. [b] WT showed no quantifiable activity, and thus a ratio cannot be determined.

Encouraging by the surprisingly broad substrate scopes of 3-SS and especially 4-V, we next applied these variants to an even wider diversity of substrates. As our research group participates in the Center for C-H Functionalization (CCHF), an NSF-funded Center for Chemical Innovation broadly focused upon developing C-H functionalization technologies and promoting their application in organic synthesis, we were provided a library of 61 compounds from Novartis. This library of compounds represents a diverse breadth of bioactive chemical structures that all pose a challenge for selective C-H functionalization, and thus the members of the CCHF seek to find new techniques for the selective functionalization of these often very high-value targets. We screened all 61 compounds with wild-type RebH, 3-SS, and 4-V and found that a remarkable 34% of the compounds in the library showed at least trace halogenation, while 6 of these 61 (~10%)

showed some to nearly full conversion (Figure 3.5). Given that we had no input in selecting the compounds that went into this library, and thus did not bias the selection toward scaffolds that RebH variants are likely to accept, and screened only three RebH variants, substantial conversion of 10% of these compounds represents a remarkable substrate scope.

Figure 3.5. Novartis library screening with best hits shown.



[a] If multiple monochlorinated and/or dichlorinated products were observed, the sum of all is shown. Conversions are based on UV absorbances at 254 nm, unadjusted for response factors.

[b] Compounds that gave only trace, non-quantifiable conversion is shown as 1% conversion for visibility.

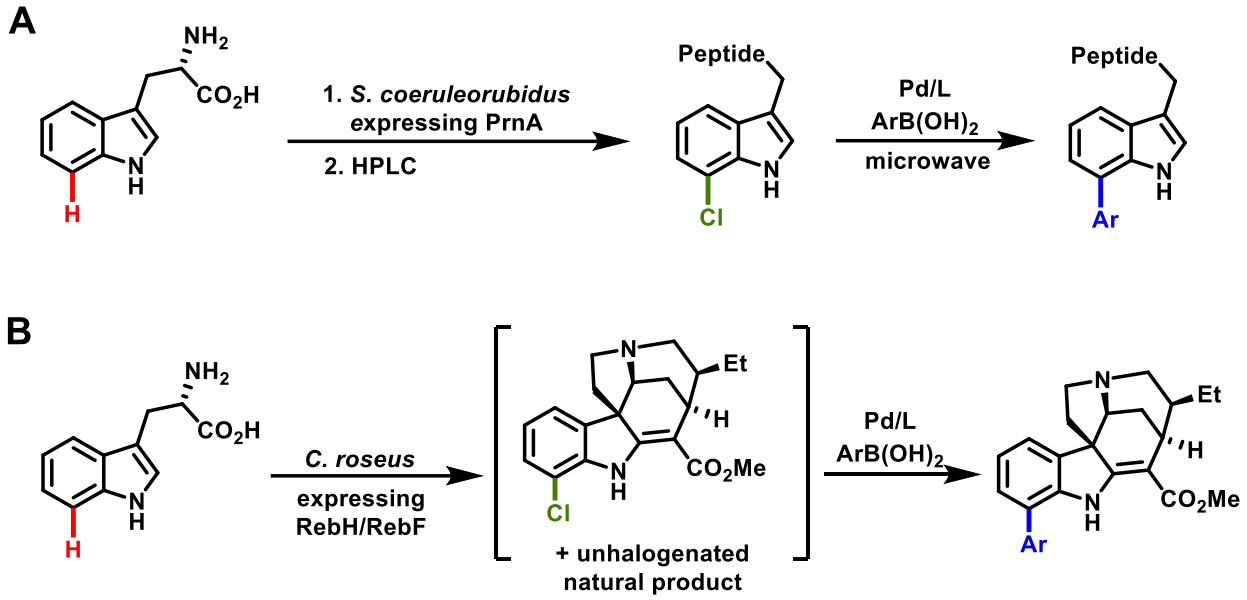
3.2.4 THE USE OF ENGINEERED REBH VARIANTS IN CHEMOENZYMATIC REACTIONS

Until this point, the engineered halogenases discussed were used for the synthesis of halogenated end-products; however, the presence of the halogen substituent also offers a handle for further functionalizations to make awide range of potential products. Combining an enzymatic

functionalization with a subsequent small-molecule transformation is perhaps best exemplified by the work that has been done with cytochromes P450. Cytochromes P450 have been extensively engineered to perform preparative oxygenations, especially hydroxylations, on a wide variety of substrates.[36] Just as is seen with halogenation, the oxygenation event itself can impart interesting bioactivities to these substrates,[37] but the installed oxygen atom can also serve as a synthetic handle for subsequent transformations,[38] such as deoxyfluorination.[39]

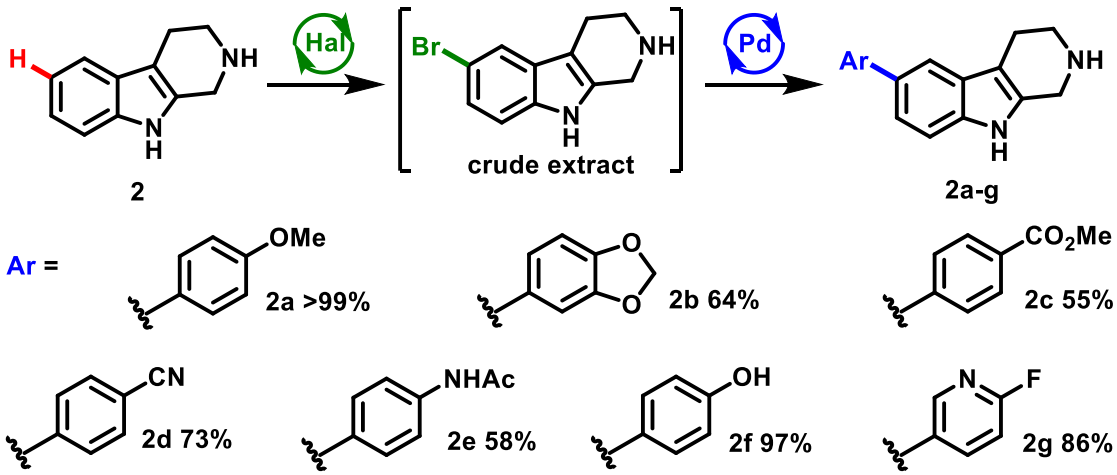
Similarly, there is some precedent for the combination of halogenase-catalyzed selective C-H halogenation with small-molecule palladium-catalysis to furnish the chemoenzymatic arylation of compounds. For example, Goss demonstrated that PrnA (another FAD-dependent tryptophan 7-halogenase; see Section 2.2.1) could be expressed in *S. coeruleorubidus* to chlorinate tryptophan, which the organism subsequently incorporated into cyclic polypeptides (Scheme 3.3A).[16b] These polypeptides were then purified by HPLC and subjected to microwave-mediated Suzuki-Miyaura conditions to furnish arylated polypeptide derivatives. More recently, O'Connor reported that RebH could be expressed in *C. roseus* to prepare halo-indole alkaloids that could be arylated via Suzuki cross-coupling (Scheme 3.3B).[40] Unlike the P450 examples noted above, however, both of these reports were limited to the native substrate (L-tryptophan, although another report from O'Connor[17] utilized a point-mutant of RebH to functionalize the smaller substrate tryptamine) and selectivity (7-position) of the halogenases used, and compatibility with only a limited range of cross-coupling reactions was demonstrated.

Scheme 3.3. Previously reported halogenase chemoenzymatic syntheses.



Given that our group had engineered RebH variants capable of site-selective, late-stage halogenation of large, bioactive compounds, as described in Sections 3.2.2 and 3.2.3, we sought to combine the halogenations catalyzed by these engineered variants with subsequent small-molecule transformations to afford a range of chemoenzymatic functionalizations on complex molecules. A student in our group, Dr. Landon Durak, found that a combination of $Pd(OAc)_2$ and water-soluble SPhos was able to provide quantitative conversion of 6-bromotryptoline from crude organic extracts of bioconversions to yield the 6-aryltryptoline product. We then performed a series of 10 mg bioconversions of tryptoline (**2**) using the engineered variant 3-SS to allow Dr. Durak to test a series of boronic acids as partners for the subsequent Suzuki reaction and found that a range of aryltryptolines were able to be isolated in high yield (Table 3.6). A variety of functional groups including ethers (**2a/b**), esters (**2c**), nitriles (**2d**), amides (**2e**), hydroxyl (**2f**), and a substituted pyridine (**2g**) were tolerated by this protocol, indicating that aryl boronic acids can be coupled to substrates bearing an unprotected indole (N-H) core.

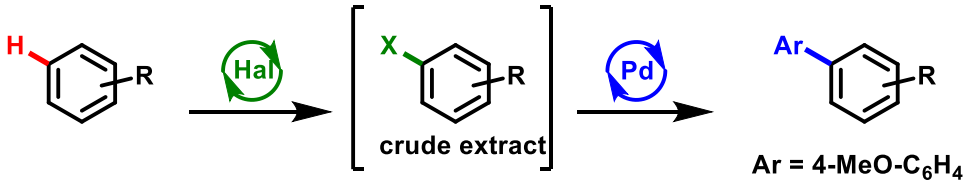
Table 3.6. Boronic acid scope of chemoenzymatic arylation.[a]

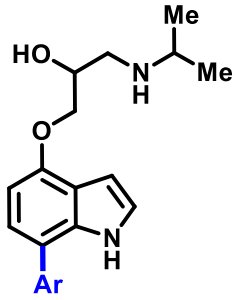


[a] Hal: 1 mol% 3-SS, 0.05 mol% MBP-RebF, 9 U/mL GDH, 35 U/mL catalase, 20 equiv. NaBr, 40 equiv. glucose, 0.2 equiv. NAD & FAD, 3.5% v/v *i*-PrOH/phosphate buffer (25 mM, pH 8.0), 22 °C, 16h. Pd: 5 mol% $Pd(OAc)_2$ & SPhos-SO_3Na, 1.5 equiv. $ArB(OH)_2$, 50% v/v *i*-PrOH/phosphate buffer (170 mM, pH 8.5), 90 °C, 1h.

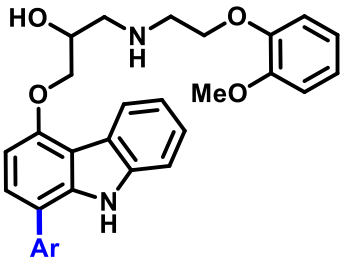
Arene scope was then evaluated by functionalizing different bioactive arenes (Table 3.7) that RebH variants 3-SS and 4-V had been found to functionalize.[18a] It was also discovered in the course of analyzing the library of compounds from Novartis, as described in Section 3.2.3 (see Figure 3.5), that RebH variant 4-V also accepted the antihistamine thenalidine (**12**), which does not bear an indole moiety but undergoes halogenation at the *para* position of its aniline core. The crude extracts from preparative scale halogenation reactions of each of the aforementioned arenes were submitted to Pd-catalyzed Suzuki cross-coupling conditions to furnish the corresponding arylated analogues. Notably, both the halogenation and cross-coupling steps tolerated the installation of both bromine and chlorine, further highlighting the flexibility of this methodology.

Table 3.7. Arene scope of chemoenzymatic arylation.[a]

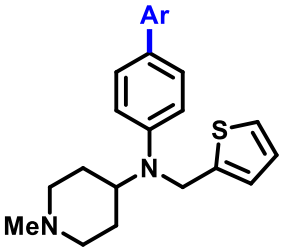




9a
99% (X=Br, 1% 4-V)



11a
93% (X=Cl, 2% 4-V)[b]

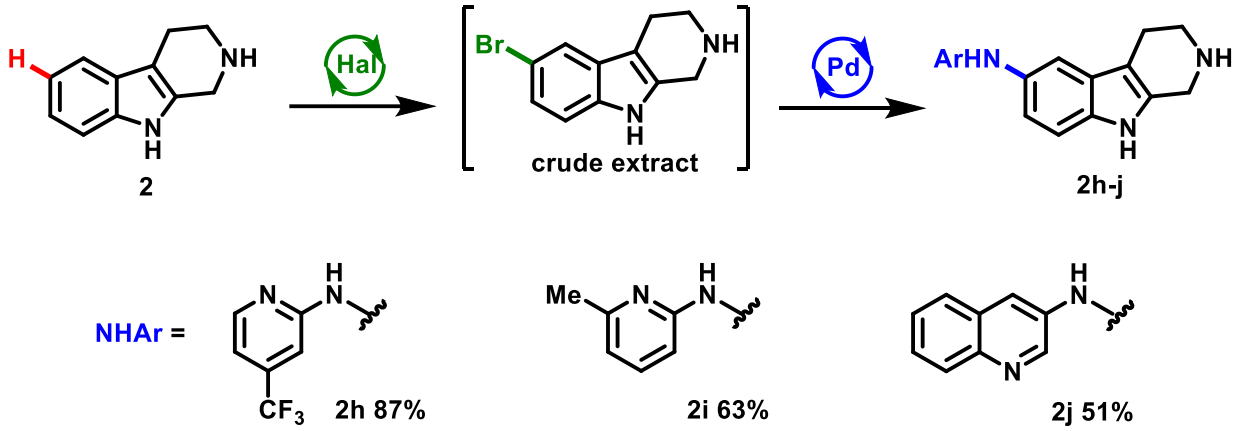


12a
74% (X=Br, 5% 4-V)

[a] Hal: as in Table 3.6 with the halide source and RebH variant/loading indicated. Pd: as in Table 3.6 using 1.5 equiv. 4-MeO-C_6H_4-$B(OH)_2$. [b]5 mol% $Pd(OAc)_2$ & SPhos, 1.5 equiv. 4-MeO-C_6H_4-$B(OH)_2$, 2 equiv. K_3PO_4, 20% v/v H2O/dioxane, 100 °C, 12h.

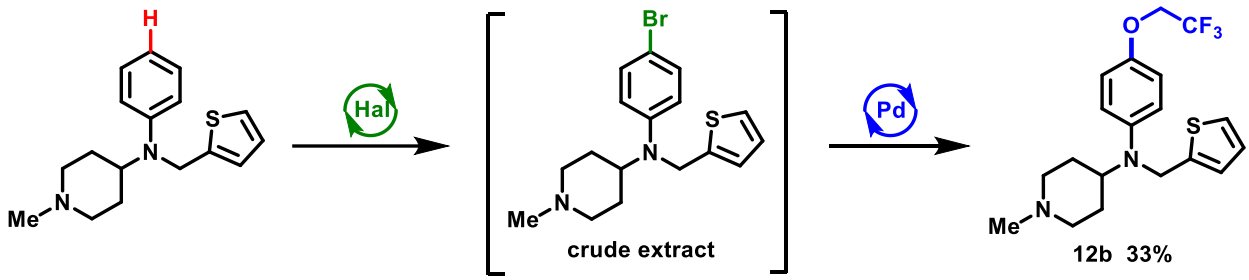
To expand the reaction scope beyond C-C bond formation, crude extracts from enzymatic halogenation reactions were also subjected to Buchwald-Hartwig amination[41] and alkoxylation[42] conditions. Tryptoline underwent smooth two-step C-H amination with aminopyridine derivatives to afford secondary amines (Table 3.8, **1h-j**). Attempts to perform a C-H alkoxylation on tryptoline, however, were unsuccessful in our hands. To our knowledge, there are no reports of Pd-catalyzed alkoxylation reactions on unprotected indoles, and our results corroborate its difficulty. Fortunately, thenalidine underwent two-step C-H alkoxylation providing trifluoroethoxythenalidine (**12b**) in useful, albeit low, yield (Scheme 3.4) illustrating that substrates compatible with alkoxylation can be used in our chemoenzymatic procedure.

Table 3.8. Chemoenzymatic amination of tryptoline.[a]



[a] Hal: as in Table 3.6. Pd: 3 mol% Pd(OAc)2, 3 mol% BrettPhos, 3 equiv. $ArNH_2$, 6 equiv. NaOt-Bu, dioxane, 100 °C, 14h.

Scheme 3.4. Chemoenzymatic alkoxylation of thanalidine.[a]



[a] Hal: as in Table 3.6 using 5 mol% 4-V. Pd: 0.5 mol% $[(allyl)PdCl]_2$, 1.5 mol% RockPhos, 2 equiv. CF_3CH_2OH & Cs_2CO_3, PhMe, 90 °C, 14h.

3.3 CONCLUSIONS

In summary, beginning from a thermostablized variant of RebH, 1-PVM, RebH variants with significantly expanded substrate scope were identified by via three rounds of random mutagenesis and screening against progressively larger substrates. These variants catalyzed the

selective halogenation of a number of unnatural substrates displaying a range of sizes, substitutions, and biological activities. Given the established impact that halogenation has on the biological activities of compounds and the dearth of methods to perform selective halogenations such as those we have described, we believe that these variants could prove useful for producing high-value derivatives of biologically active compounds. While the mutations discovered offer great insight for future targeted libraries,[43] the discovery of crucial mutations distant from the active site, such as A442V in variant 4-V, highlight the value of random mutagenesis in discovering important mutations that are difficult to predict.[44] We hope to gain insight into the role of this mutation and the others identified during our evolution efforts by using protein crystallography with bound substrates. Furthermore, the ability to couple enzymatic halogenations with subsequent small-molecule reactions greatly expands the breadth of novel compounds with potentially improved bioactivities that can be accessed. Lastly, another student in our group, Mary Andorfer, has engineered RebH variants with altered regioselectivity on tryptamine;[45] we hope to combine mutations that alter regioselectivity with those that expand substrate scope to ideally access multiple sites on a range of complex, bioactive structures.

3.4 EXPERIMENTAL

3.4.1 GENERAL EXPERIMENTAL PROCEDURES

Materials:

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Debromodeformylflustrabromine (starting material for halogenation to produce **3)** was synthesized according to previous reports.[25] Deuterated solvents were obtained from Cambridge Isotope labs. Silicycle silica gel plates (250 mm, 60 F254) were used for

analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Oligonucleotides were purchased from Integrated DNA Technologies (San Diego, CA). Plasmids pET-28a/RebF and pET-28a/RebH in BL-21 DE3 *E. coli* were provided by the Walsh group of Harvard Medical School, Boston, MA.[11] The pLIC-MBP plasmid was provided by the Bottomley group of Monash University, Clayton, Australia.[46] The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan). BL21(DE3) *E. coli* cells were purchased from Invitrogen (Carlsbad, CA). T7 DNA ligase, 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). Library colonies were picked using an automated colony picker (Norgren Systems). Qiagen Miniprep Kits were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. All genes were confirmed by sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser using method Ec2. Ninitrilotriacetic acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were following when using both products (for Ni-NTA resin, 5 mL resin were used, with buffers delivered by a peristaltic pump at a rate of 1 mL/min, in a 4 °C cold cabinet). Amicon® 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. Biotage reverse phase columns (SNAP KP-C18-HS) were purchased from Biotage. HPLC analyses were performed using HPLC grade acetonitrile (Fisher Scientific), 18 MΩ water from a Milli-Q purification system (model No. QGARD00D2), and trifluoroacetic acid (Oakwood Chemicals). Water and isopropanol used in

Suzuki-Miyaura reactions were deoxygenated by sparging with N_2 for 30 minutes. Anhydrous dioxane was purchased from Acros in an AcroSeal bottle and used as received. Toluene was obtained from an Innovative Technologies solvent purification system (solvent deoxygenated by sparging with N_2 and dried over alumina). $[(allyl)PdCl]_2$ was prepared according to the literature.[47] Thenalidine was obtained from Novartis through the CCHF. Glucose dehydrogenase (GDH, product No. GDH-105), and NAD (product No. NAD-004626) were purchased from Codexis (Redwood City, CA). FAD (product No. 00151) was purchased from Chem-Impex International (Wood Dale, IL). Catalase (product No. C1345-1G) was purchased from Aldrich. Biotage reversed-phase columns were purchased from Biotage (FSUL-0401-0012). AeraSeal film was purchased from Research Products International (product No. 202504).

General Procedures:

Standard molecular cloning procedures were followed.[48] Reactions were monitored using UPLC (Agilent 1200 UPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; C18 4.6 x 50 mm column, 3.5 μM particle size; and C18 2.1 x 50 mm column, 1.8 μM particle size; solvent A = H2O/0.1% TFA, solvent B = CH3CN). Gel filtration was performed using a HiLoad 16/600 Superdex 200 column (GE Healthcare Life Sciences). Reverse phase preparative chromatography was carried out using a Biotage Isolera One. 1H spectra were recorded at 500 MHz on a Bruker DMX-500 or DRX-500 spectrometer at room temperature, and chemical shifts are reported relative to residual solvent peaks with coupling constants reported in Hz.[49] In some cases the carbon atoms in the trifluoroacetate ion cannot be observed in a reasonable number of scans.[50] Mass spectra were obtained from the University of Chicago mass spectrometry facility using an Agilent Technologies 6224 TOF LC/MS.

3.4.2 SPECIFIC EXPERIMENTAL PROCEDURES

Library construction, expression, and screening: The procedures used for library construction, expression, and screening were adapted from those previously reported for the evolution of RebH for increased thermostability.[20] All genes encoding RebH were cloned into pET-28a between the NdeI and HindIII digestion sites. Mutant libraries were constructed by error-prone PCR, using Taq polymerase with 150 μM $MnCl_2$. PCR was performed in a volume of 50 μL with conditions of 95 °C 30 s, (95 °C 30 s, 55 °C 30 s, 72 °C 90 s) for 20 cycles, 72 °C 10 min. Beneficial mutations were recombined via overlap extension[24] with PCR conditions of 98 °C 30 s, (98 °C 10 s, 72 °C 50 s) for 35 cycles, 72 °C 10 min. Plasmids were transformed by electroporation into *E. coli* containing the chaperone pGro7. Library colonies were picked using an automated colony picker (Norgren Systems) and arrayed in 1-ml 96-well plates containing 300 μL LB with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol. Cells were grown overnight at 37 °C, 250 rpm, and 50-100 μL of overnight culture was used to inoculate 1 mL TB (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol) in 2-mL 96-well plates. Following growth at 37 °C, 250 rpm, to an OD_{600} = 0.9-1.0, enzyme expression was induced with IPTG and arabinose to final concentrations of 10 μM and 0.2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cultures were harvested by centrifugation and stored at -80 °C until use.

Cell pellets were thawed and suspended in 100 μL HEPES buffer (25 mM, pH 7.4) containing 0.75 mg/mL lysozyme. After incubation at 37 °C, 250 rpm, cells were flash frozen in liquid nitrogen and thawed in a 37 °C water bath. Ten microliters of DNaseI at 1 mg/mL were added and the cells incubated at 37 °C, 250 rpm, for 15 min. After centrifugation, 50 μL of supernatant were transferred to a microtiter plate for screening.

Similar to what was previously described for the evolution of RebH for increased thermostability,[20] a combined stock solution containing MBP-RebF (2.5 μM final concentration) and glucose dehydrogenase (9 U/mL final concentration) was added to 50 μL lysate. A second combined stock solution containing all small molecule components, including NaCl (10 mM final concentration), FAD (100 μM final concentration), NAD (100 μM final concentration), substrate (from a 10 mM stock solution in MeOH, 0.5 mM final concentration), and glucose (20 mM final concentration), all in 25 μM HEPES buffer (pH 7.4), was added to the reaction mixtures to initiate. Reactions were mixed, the plates sealed, and left overnight on the benchtop (increased activity on L-tryptophan and tryptoline) or shaken at 600 rpm (increased activity on debromodeformylflustrabromine). Reactions were quenched with an equal volume of methanol and centrifuged, and the supernatant was filtered and analyzed for chlorination of substrate via UPLC (Agilent 1200 UPLC with an Agilent Eclipse Plus C18 2.1 x 50 mm column, 1.8 μM particle size; solvent A = H_2O/0.1 % TFA, solvent B = CH_3CN; 0-0.5 min, B = 16%; 0.5-1.5 min, B = 16-80%).

Enzyme purification: The MBP-RebF and RebH variants used for analytical and 10 mg bioconversions was grown, expressed, lysed and purified according to a previous report.[12] An overnight starter culture was used to inoculate 50 mL TB (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol). Following growth at 37 °C, 250 rpm, until OD_{600} = 0.6-0.8, enzyme expression was induced with IPTG and arabinose to final concentrations of 100 μM and 2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cultures were harvested by centrifugation and stored at -80 °C until use. Cell pellets were thawed, suspended in 30 mL 25 mM HEPES (pH 7.4) and lysed by sonication while kept on ice (Qsonica S-4000 with

a 0.5" horn; 5 x 1 min with 1 min rests, 20 % duty cycle delivering 40-50 W). After clarification by centrifugation, MBP-RebF and RebH variants were purified by Ni-NTA affinity chromatography and exchanged into a buffer of 25 mM HEPES (pH 7.4) and 10 % glycerol. Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -20 °C until use.

General Procedure for Analytical Bioconversions[20]: Substrate (37.5 nmol) was added to a 1.5 mL Eppendorf tube as a 10 mM solution in MeOH (evodiamine, **8**, was added as a solution in DMSO). Solutions of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and glucose dehydrogenase (9 U/mL final concentration GDH) were added to the reaction. This was diluted such that the final reaction volume was 75 μL with HEPES buffer, and RebH (0.005-0.05 equiv., 2.5-25 μM final concentration) and MBP-RebF (0.005 equiv., 2.5 μM final concentration) were added as solutions of HEPES/glycerol buffer (25 mM HEPES, pH 7.4, 10% glycerol v/v). The reaction was initiated with a solution of 1 M glucose (40 equiv., 20 mM final concentration), the tube was closed, and incubated at 25 °C at 600 rpm. Reactions were quenched by addition of a reaction volume of MeOH after 12 hours. These reactions were analyzed by UPLC (Agilent 1200 UPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN). The following method was used for all substrates: 0-10 min, B = 15%; 10-20 min, B = 15-100%; 20-24 min, B = 100%. Alternate buffer salts, concentrations, and pHs were tested as well, and we observed that up to 2.5-fold increased conversions could be obtained on relative to the conditions described above by using 300 mM HEPES, pH 7.4 for certain substrates

(L-tryptophan, tryptoline, pindolol, and carazolol all showed increased conversions with this buffer substitution).

General Procedure for 10 mg Bioconversions (without subsequent cross-coupling)[12]: Substrate (10.0 mg) was added to a crystallization dish (100 x 50 mm) as a solution in MeOH. Solutions of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and a glucose dehydrogenase (9 U/mL final concentration GDH) were added to the reaction. This was diluted to the appropriate volume with HEPES buffer, and RebH (0.01-0.05 equiv., 5-25 μM final concentration) and MBP-RebF (0.005 equiv., 2.5 μM final concentration) were added as solutions of HEPES/glycerol buffer (25 mM HEPES, pH 7.4, 10% glycerol v/v). The reaction was initiated with a solution of 1 M glucose (40 equiv., 20 mM final concentration), sealed with an AeraSeal film, and left on the benchtop at room temperature without shaking. These reactions were analyzed by UPLC (Agilent 1200 UPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN). The following method was used for all substrates: 0-10 min, B = 15%; 10-20 min, B = 15-100%; 20-24 min, B = 100%. Each reaction was allowed to continue for 3 days, at which time no additional conversion was seen by UPLC and enzyme had visibly precipitated out. The bioconversions were quenched with HCl (5 M, until pH<2) and saturated with NaCl. Precipitated protein was filtered out through a pad of Celite and the filtrate brought to pH>12 through addition of NaOH (5M). The filtrate was extracted into CH_2Cl_2. The crude material was purified by reverse phase chromatography (Biotage SNAP-KP-C18-HS, gradient from pure H_2O to 40% CH_3CN/H_2O).

Determination of Kinetic Parameters for RebH Variants: Kinetic parameters were determined as was previously described.[12] Rates were determined by monitoring the conversion of 75-215 μM tryptoline in the presence of NAD (100 μM final concentration), FAD (100 μM final concentration), NaCl (10 mM final concentration), MBP-RebF (2.5 μM final concentration), glucose dehydrogenase (9 U/mL final concentration), glucose (20 mM final concentration), and phenol as an internal standard (10 mM in MeOH, 0.5 mM final concentration) at a final volume of 75 μL in a microtiter plate. RebH variants were added at a final concentration of either 25 μM for wild-type RebH and 1-PVM, 15 μM for 2-T, or 2.5 μM for 3-SS. The reactions were left shaking at 600 rpm at room temperature, then quenched at 15-60 minutes (all time points were collected in triplicate) by addition of 75 μL of MeOH. The precipitated protein was then removed by centrifugation and the reactions were filtered and analyzed by UPLC using the method described in Section 3.4.1. Product formation was quantitated by calculating the ratio of product to internal standard and fitting that value to a calibration curve prepared from known concentrations of each material. The kinetic parameters (K_m and k_{cat}) for each substrate were determined using the Hanes-Woolf plots (see Section 3.4.5) constructed from the substrate concentrations and the observed initial rates.

General Procedure for 10 mg Scale Bioconversions (with subsequent cross-coupling): Substrate (10.0 mg), NaBr/NaCl (20 equiv., 10 mM final concentration), glucose (40 equiv., 20 mM final concentration), glucose dehydrogenase (9 U/mL final concentration GDH), and catalase (35 U/mL final concentration) were added as solids to a crystallization dish (100 x 50 mm) containing a magnetic stir bar. This was diluted to an appropriate volume with phosphate buffer (25 mM K_2HPO_4, pH 8.0) and isopropanol (3.5% v/v) was added. 10 mM aqueous solutions of

FAD and NAD were prepared and added to the reaction mixture (0.20 equiv., 0.10 mM final concentration each). Stocks of RebH and RebF, stored in a HEPES/glycerol buffer (25 mM HEPES, pH 7.4, 10% glycerol v/v) following purification, were thawed in an ice water bath. RebH (0.01-0.05 equiv., 5-25 µM final concentration) and MBP-RebF (0.0005 equiv. 0.25 µM final concentration) were added as solutions. The reaction vessel was sealed with an AeraSeal film and gently stirred at room temperature for 16 hours. The reactions were monitored by UPLC (Agilent 1200 UPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 µm particle size; solvent A = H_2O/0.1% TFA, solvent B = MeCN) until maximum conversion was observed. The following method was used for all substrates: 0-10 min, B = 15%; 10-20 min, B = 15-100%; 20-24 min, B = 100%. The bioconversion were quenched with HCl (5 M, until pH<2) and saturated with solid NaCl. Precipitated protein was filtered out through a pad of Celite and the filtrate was brought to pH>12 through addition of NaOH (5M). The filtrate was extracted into CH_2Cl_2, and the solvent removed by rotoary evaporator.

General Procedure for Suzuki-Miyaura Coupling on Crude Extract of Bioconversion: The crude extracts from an enzymatic halogenation were transferred to a 50 mL round bottomed flask and Ar-$B(OH)_2$ (1.5 equiv.), $Pd(OAc)_2$ (0.05 equiv.), sodium 2′-dicyclohexylphosphino-2,6-dimethoxy-1,1′-biphenyl-3-sulfonate hydrate (water soluble Sphos, 0.05 equiv.), and a magnetic stir bar were added. The flask was equipped with a reflux condenser and capped with a rubber septum. The apparatus was purged by three cycles of evacuation and N_2 refill. A deoxygenated 1:1 mixture (20 mL) of isopropanol and phosphate buffer (170 mM K_2HPO_4, pH 8.5) was added via syringe and the mixture was allowed to stir under reflux in an oil bath. After 1 hour, the reaction vessel was allowed to cool to room temperature, and the isopropanol was removed by

rotary evaporation. The aqueous solution was extracted with CH_2Cl_2 (3x, 10 mL each), and the combined organic extracts were dried over Na_2SO_4 and concentrated by rotary evaporation. The crude mixture was purified by reversed-phase chromatography (Biotage SNAP-KP-C18-HS, gradient from pure H_2O to 60% MeCN/H_2O) and isolated as the trifluoroacetate (TFA) salt.

General Procedure for Buchwald-Hartwig Amination on Crude Extract of Bioconversion: The crude extracts from an enzymatic halogenation were transferred to a 20 mL round scintillation vial and $ArNH_2$ (3 equiv.), $Pd(OAc)_2$ (0.03 equiv.), BrettPhos (0.03 equiv.), NaO*t*-Bu (6 equiv.) and a magnetic stir bar were added. The vial was transferred to an inert atmosphere dry box, dioxane (1 mL) was added, and the vial was sealed with a teflon lined cap. The vial was removed from the dry box and the mixture was allowed to stir in a 100 °C oil bath. After 14 hours, the reaction vessel was allowed to cool to room temperature, and the contents filtered over silica, eluting with 150 mL 4:1 CH_2Cl_2/MeOH. The filtrate was collected and concentrated by rotary evaporation. The crude mixture was purified by reversed-phase chromatography (Biotage SNAP-KP-C18-HS, gradient from pure H_2O to 60% MeCN/H_2O) and isolated as the trifluoroacetate (TFA) salt.

3.4.3 DETAILED SYNTHESES AND CHARACTERIZATION

6-chlorotryptoline (1): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of tryptoline. RebH variant 3-SS was added to a final concentration of 5 μM (0.01 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the general procedure to afford **1** in 78% yield (14.5 mg of **1**·TFA, 0.045 mmol). ^{1}HNMR spectrum was consistent with previous

reports of this compound.[12] ^{1}H NMR (500 MHz, MeOD) δ 7.47 (d, *J* = 1.7 Hz, 1H), 7.32 (d, *J* = 8.6 Hz, 1H), 7.11 (dd, *J* = 8.6, 1.8 Hz, 1H), 4.44 (s, 2H), 3.58 (t, *J* = 6.1 Hz, 2H), 3.05 (t, *J* = 6.0 Hz, 2H). Only trace (<5%) 7-chlorotryptoline could be observed (consistent with previous reports of this compound[12]), but the low abundance (in an already small amount of material) and the fact that one of the 7-chlorotryptoline peaks is partially obscured by an impurity makes precise quantitation difficult. HRMS (ESI-TOF) calc'd for $C_{11}H_{11}N_2Cl$ [M + H]+: 207.0689 and 209.0660, found: 207.0645 and 209.0626.

6-chloroeleagnine (5): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of eleagnine. RebH variant 3-SS was added to a final concentration of 5 μM (0.01 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the general procedure to afford **5** in 83% yield (14.9 mg of **5**·TFA, 0.045 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.48 (d, *J* = 1.8 Hz, 1H), 7.33 (d, *J* = 8.6 Hz, 1H), 7.12 (dd, *J* = 8.6, 2.0 Hz, 1H), 4.78 (q, *J* = 6.6 Hz, 1H), 3.73 (m, 1H), 3.47 (m, 1H), 3.14 – 2.96 (m, 2H), 1.73 (d, *J* = 6.8 Hz, 3H). ^{13}C NMR (126 MHz, MeOD) δ 161.88, 161.61, 138.47, 128.35, 123.89, 120.44, 116.58, 111.55, 106.67, 42.05, 21.37, 18.83. HRMS (ESI-TOF) calc'd for $C_{12}H_{13}N_2Cl$ [M + H]+: 221.0846 and 223.0816, found: 221.0799 and 223.0795. Note: NOESY for this compound was performed using the free base of **5**, not the TFA salt.

7-chloropinoline (6): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of pinoline. RebH variant 3-SS was added to a final

concentration of 10 μM (0.02 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the general procedure to afford **6** in 80% yield (13.9 mg of **6**·TFA, 0.040 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.37 (s, 1H), 7.09 (s, 1H), 4.41 (s, 2H), 3.88 (s, 3H), 3.58 (t, *J* = 6.0 Hz, 2H), 3.05 (t, *J* = 5.9 Hz, 2H). ^{13}C NMR (126 MHz, MeOD) δ 151.75, 136.08, 120.36, 118.68, 105.09, 94.89, 55.43, 49.24, 41.11, 29.26, 22.80, 18.07, 16.32. HRMS (ESI-TOF) calc'd for $C_{12}H_{13}N_2OCl$ [M + H]+: 237.0795 and 239.0765, found: 237.0762 and 239.0747.

6-chlorotetrahydroharmine (7): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of tetrahydroharmine. RebH variant 3-SS was added to a final concentration of 25 μM (0.05 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the general procedure to afford **7** in 79% yield (13.3 mg of **7**·TFA, 0.037 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.46 (s, 1H), 7.01 (s, 1H), 4.74 (q, *J* = 6.6 Hz, 1H), 3.88 (s, 3H), 3.70 (m, 1H), 3.44 (m, 1H), 3.12 – 2.91 (m, 2H), 1.71 (d, *J* = 6.8 Hz, 3H). ^{13}C NMR (126 MHz, MeOD) δ 153.16, 137.49, 121.77, 120.09, 106.50, 96.30, 56.84, 50.64, 42.51, 30.67, 24.21, 19.47, 17.73. HRMS (ESI-TOF) calc'd for $C_{13}H_{15}N_2OCl$ [M + H]+: 251.0951 and 253.0922, found: 251.0897 and 253.0895.

6-chlorodebromodeformyllflustrabromine (3): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of debromodeformylflustrabromine. RebH variant 4-V was added to a final concentration of 25 μM (0.05 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the

general procedure to afford **3** in 68% yield (11.0 mg of **3**·TFA, 0.028 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.45 (d, *J* = 8.4 Hz, 1H), 7.34 (s, 1H), 6.99 (d, *J* = 8.4 Hz, 1H), 6.19 (m, 1H), 5.17 (m, 2H), 3.15 (m, 4H), 2.74 (s, 3H), 1.54 (d, *J* = 1.5 Hz, 6H). ^{13}C NMR (126 MHz, MeOD) δ 145.86, 127.34, 126.67, 119.15, 117.77, 110.96, 110.48, 104.12, 49.26, 38.75, 32.27, 29.25, 26.90, 22.80, 21.54. HRMS (ESI-TOF) calc'd for $C_{16}H_{21}N_2Cl$ [M + H]+: 277.1472 and 279.1442, found: 277.1432 and 279.1422.

10-chloroyohimbine and 11-chloroyohimbine (referred to together as 4): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of yohimbine. RebH variant 4-V was added to a final concentration of 25 μM (0.05 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the general procedure to afford **4** in 54% yield (7.65 mg of **4**·TFA, 0.015 mmol). ^{1}H NMR (500 MHz, MeOD) δ 11.11 (s, 0.5H), 11.07 (s, 0.2H), 7.51 (d, *J* = 1.5 Hz, 1H), 7.48 (d, *J* = 8.5 Hz, 0.4H), 7.38 (s, 0.4H), 7.35 (d, *J* = 8.6 Hz, 1H), 7.15 (dd, *J* = 8.7, 1.8 Hz, 1H), 7.08 (dd, *J* = 8.4, 1.6 Hz, 0.4H), 4.77 (d, *J* = 11.9 Hz, 1H), 4.34 (d, *J* = 2.1 Hz, 1H), 3.87 – 3.82 (m, 1H), 3.81 (s, 3H), 3.66 – 3.55 (m, 2H), 3.21 (t, *J* = 12.1 Hz, 2H), 3.10 (dd, *J* = 16.5, 5.3 Hz, 1H), 2.93 (d, *J* = 14.5 Hz, 1H), 2.47 (dd, *J* = 11.6, 2.4 Hz, 1H), 2.38 – 2.28 (m, 1H), 1.99 (dd, *J* = 13.8, 3.0 Hz, 1H), 1.74 (t, *J* = 13.5 Hz, 2H), 1.63 (m, 1H), 1.57 – 1.43 (m, 2H). ^{13}C NMR (126 MHz, MeOD) δ 172.92, 122.36, 118.90, 117.29, 112.44, 66.74, 61.52, 57.97, 52.67, 51.40, 50.91, 38.18, 34.24, 31.66, 31.49, 30.53, 24.47, 23.93, 21.99, 18.64. The previously described peak assignments represent a roughly 3:1 mixture of 10-chloroyohimbine and 11-chloroyohimbine, respectively – note the peaks in the aryl region of ^{1}H NMR (7.51 (d, *J* = 1.5 Hz, 1H), 7.48 (d, *J* = 8.5 Hz, 0.4H), 7.38 (s, 0.4H), 7.35 (d, *J* = 8.6 Hz, 1H), 7.15 (dd, *J* = 8.7, 1.8 Hz, 1H), 7.08 (dd, *J* = 8.4, 1.6 Hz,

0.4H)). All peaks observed in the ^{13}C NMR observed with a signal:noise of >2:1 are reported – higher quality ^{13}C NMR data were not able to be obtained in a reasonable number of scans, given the small amount of isolated material available and the high percentage of tertiary and quarternary carbons present in these two compounds. HRMS (ESI-TOF) calc'd for $C_{21}H_{25}N_2O_3Cl$ [M + H]+: 389.1632 and 391.1602, found: 389.1601 and 391.1596.

7-chloropindolol (9): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of pindolol. RebH variant 4-V was added to a final concentration of 5 μM (0.01 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the general procedure to afford **9** in 93% yield (14.9 mg of **9**·TFA, 0.037 mmol). ^1H NMR (500 MHz, MeOD) δ 7.21 (d, *J* = 3.0 Hz, 1H), 7.02 (d, *J* = 8.2 Hz, 1H), 6.62 (d, *J* = 3.1 Hz, 1H), 6.51 (d, *J* = 8.3 Hz, 1H), 4.36 – 4.25 (m, 1H), 4.15 (m, 2H), 3.47 (m, 1H), 3.37 – 3.17 (m, 2H, partially obscured by methanol residual peak), 1.37 (dd, *J* = 6.3, 5.1 Hz, 7H). ^{13}C NMR (126 MHz, MeOD) δ 150.79, 123.96, 120.58, 120.18, 109.60, 100.88, 99.39, 69.83, 65.59, 50.68, 47.22, 17.93, 17.40. HRMS (ESI-TOF) calc'd for $C_{14}H_{19}N_2O_2Cl$ [M + H]+: 283.1213 and 285.1184, found: 283.1181 and 285.1159.

8-chlorocarazolol (10): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of carazolol. RebH variant 4-V was added to a final concentration of 5 μM (0.01 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the general procedure to afford **10** in 79% yield (11.8 mg of **10**·TFA, 0.027 mmol). ^1H NMR (500 MHz, MeOD) δ 8.27 (d, *J* = 7.9 Hz,

1H), 7.53 (d, *J* = 8.2 Hz, 1H), 7.44 – 7.37 (m, 1H), 7.31 (d, *J* = 8.5 Hz, 1H), 7.20 (t, *J* = 7.5 Hz, 1H), 6.71 (d, *J* = 8.5 Hz, 1H), 4.44 (m, 1H), 4.30 (m, 2H), 3.49 (m, 1H), 3.44 – 3.27 (m, 2H, partially obscured by methanol residual peak), 1.38 (dd, *J* = 6.4, 5.0 Hz, 6H). ^{13}C NMR (126 MHz, MeOD) δ 154.88, 140.86, 139.40, 126.58, 126.25, 123.88, 123.57, 120.57, 111.96, 109.94, 102.64, 71.14, 67.01, 52.04, 48.61, 19.34, 18.78. HRMS (ESI-TOF) calc'd for $C_{18}H_{21}N_2O_2Cl$ [M + H]+: 333.1370 and 335.1340, found: 333.1337 and 335.1315.

8-chlorocarvedilol (11): The 10 mg bioconversion was conducted according to the general procedure, using 10.0 mg (1 equiv.) of carvedilol. RebH variant 4-V was added to a final concentration of 10 μM (0.02 equiv.). The reaction was monitored by UPLC. After reaction completion, the reaction mixture was purified according to the general procedure to afford **11** in 69% yield (9.4 mg of **11**·TFA, 0.017 mmol). ^{1}H NMR (500 MHz, MeOD) δ 8.24 (d, *J* = 7.8 Hz, 1H), 7.50 (d, *J* = 8.1 Hz, 1H), 7.35 (t, *J* = 7.6 Hz, 1H), 7.30 (d, *J* = 8.4 Hz, 1H), 7.07 (t, *J* = 7.5 Hz, 1H), 6.99 (m, 3H), 6.93 – 6.85 (m, 1H), 6.72 (d, *J* = 8.5 Hz, 1H), 4.53 - 4.51 (m, 1H), 4.41 – 4.22 (m, 4H), 3.77 (s, 3H), 3.68 – 3.41 (m, 4H). ^{13}C NMR (126 MHz, MeOD) δ 153.47, 149.63, 146.98, 139.42, 137.99, 125.12, 124.81, 122.78, 122.54, 122.14, 120.91, 119.20, 115.08, 113.62, 111.86, 110.48, 108.55, 101.21, 69.82, 65.32, 64.74, 54.95, 50.21, 46.91, 29.26, 22.80. HRMS (ESI-TOF) calc'd for $C_{24}H_{25}N_2O_4Cl$ [M + H]+: 441.1537 and 443.1524, found: 441.1537 and 443.1552.

6-(4-methoxyphenyl)-2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indole (2a): 2a was prepared from tryptoline (10 mg, 0.058 mmol) and 4-methoxyphenylboronic acid following the general procedure for enzymatic bromination with 3-SS and Suzuki-Miyaura Coupling outlined above in >99% yield

(23.17 mg of **2a**·TFA, 0.059 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.64 (s, 1H), 7.55 (d, *J* = 8.5 Hz, 2H), 7.39 (s, 2H), 6.98 (d, *J* = 8.5 Hz, 2H), 4.46 (s, 2H), 3.83 (s, 3H), 3.61 (t, *J* = 5.9 Hz, 2H), 3.13 (t, *J* = 5.6 Hz, 2H). ^{13}C NMR (126 MHz, MeOD) δ 160.08, 137.43, 136.34, 129.31, 129.04, 128.09, 127.06, 122.81, 121.35, 116.72, 115.16, 112.49, 55.76, 43.77, 42.30, 19.50. HRMS calculated for $C_{18}H_{19}N_2O$ $[M + H]^+$: 279.1498, found: 279.1515.

6-(benzo[*d*][1,3]dioxol-5-yl)-2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indole (2b): 2b was prepared from tryptoline (10 mg, 0.058 mmol) and 3,4-(methylenedioxy)phenylboronic acid following the general procedure for enzymatic bromination with 3-SS and Suzuki-Miyaura Coupling outlined above in 64% yield (15.1 mg **2b**·TFA, 0.037 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.62 (s, 1H), 7.41 – 7.31 (m, 2H), 7.09 (m, 2H), 6.87 (d, *J* = 8.0 Hz, 1H), 5.96 (s, 2H), 4.45 (s, 2H), 3.60 (t, *J* = 6.0 Hz, 2H), 3.12 (t, *J* = 5.7 Hz, 2H). ^{13}C NMR (126 MHz, MeOD) δ 149.52, 147.84, 138.30, 137.55, 134.26, 128.04, 127.16, 122.96, 121.45, 117.00, 112.50, 109.35, 108.60, 107.39, 102.31, 43.77, 42.30 19.49. HRMS calculated for $C_{18}H_{17}N_2O_2$ $[M + H]^+$: 293.1290, found: 293.1291.

methyl 4-(2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indol-6-yl)benzoate (2c): 2c was prepared from tryptoline (10 mg, 0.058 mmol) and 4-methoxycarbonylphenylboronic acid following the general procedure for enzymatic bromination with 3-SS and Suzuki-Miyaura Coupling outlined above in 55% yield (13.5 mg **2c**·TFA, 0.032 mmol). ^{1}H NMR (500 MHz, MeOD) δ 8.07 (d, *J* = 8.1 Hz, 2H), 7.91 – 7.64 (m, 3H), 7.48 (m, *J* = 27.9, 8.5 Hz, 2H), 4.47 (s, 2H), 3.92 (s, 3H), 3.62 (t, *J* = 5.7 Hz, 2H), 3.15 (br, 2H). ^{13}C NMR (126 MHz, MeOD) δ 181.67, 168.72, 148.56, 138.32, 132.89, 131.02, 129.08, 128.01, 127.98, 127.60, 123.00, 117.80, 112.87, 107.73, 52.53, 43.76, 42.29, 19.48. HRMS calculated for $C_{19}H_{19}N_2O_2$ $[M + H]^+$: 307.1447, found: 307.1446.

4-(2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indol-6-yl)benzonitrile (2d): 2d was prepared from tryptoline (10 mg, 0.058 mmol) and 4-cyanophenylboronic acid following the general procedure for enzymatic bromination with 3-SS and Suzuki-Miyaura Coupling outlined above in 73% yield (16.5 mg **2d**·TFA, 0.043 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.83 (m, 3H), 7.76 (m, 2H), 7.49 (m, 3H), 4.47 (s, 2H), 3.62 (t, *J* = 6.0 Hz, 2H), 3.15 (t, *J* = 6.0 Hz, 2H). ^{13}C NMR (126 MHz, MeOD) δ 148.49, 138.48, 133.64, 132.09, 128.77, 128.24, 127.86, 122.85, 120.08, 117.95, 113.03, 110.65, 107.81, 43.71, 42.23, 19.46. HRMS calculated for $C_{18}H_{16}N_3$ $[M + H]^+$: 274.1344, found: 274.1345.

***N*-(4-(2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indol-6-yl)phenyl)acetamide (2e): 2e** was prepared from tryptoline (10 mg, 0.058 mmol) and (4-acetamido)phenylboronic acid following the general procedure for enzymatic bromination with 3-SS and Suzuki-Miyaura Coupling outlined above in 58% yield (14.1 mg **2e**·TFA, 0.034 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.70 (s, 1H), 7.60 (s, 4H), 7.45 – 7.39 (m, 2H), 4.47 (s, 2H), 3.62 (t, *J* = 6.0 Hz, 2H), 3.14 (t, *J* = 5.8 Hz, 2H), 2.15 (s, 3H). ^{13}C NMR (126 MHz, MeOD) δ 170.03, 152.37, 138.37, 133.82, 131.77, 128.54, 128.27, 122.84, 121.64, 121.56, 116.98, 113.95, 112.59, 43.79, 42.32, 23.79, 19.50. HRMS calculated for $C_{19}H_{20}N_3O$ $[M + H]^+$: 306.1606, found: 306.1614.

4-(2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indol-6-yl)phenol (2f): 2f was prepared from tryptoline (10 mg, 0.058 mmol) and 4-hyroxyphenylboronic acid following the general procedure for enzymatic bromination with 3-SS and Suzuki-Miyaura Coupling outlined above in 97% yield in roughly 90% purity (21.2 mg **2f**·TFA, 0.056 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.61 (s, 4H), 7.45 (d, *J* = 8.4 Hz, 8H), 7.37 (s, 6H), 6.84 (d, *J* = 8.4 Hz, 8H), 4.45 (s, 8H), 3.68 – 3.48 (m, 10H),

3.12 (t, J = 5.7 Hz, 7H). ^{13}C NMR (126 MHz, MeOD) δ 137.32, 135.29, 134.47, 129.09, 128.07, 126.95, 126.09, 122.81, 121.63, 113.96, 112.43, 107.28, 43.79, 42.32, 19.51. HRMS calculated for $C_{17}H_{17}N_2O$ $[M + H]^+$: 265.1341, found: 265.1356.

6-(6-fluoropyridin-3-yl)-2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indole (2g): 2g was prepared from tryptoline (10 mg, 0.058 mmol) and 6-fluoro-3-pyridinylboronic acid following the general procedure for enzymatic bromination with 3-SS and Suzuki-Miyaura Coupling outlined above in 86% yield (19.0 mg **2g**·TFA, 0.050 mmol). ^{1}H NMR (500 MHz, MeOD) δ 8.44 (d, J = 2.2 Hz, 1H), 8.20 (m, 1H), 7.75 (s, 1H), 7.61 – 7.25 (m, 2H), 7.13 (dd, J = 8.4, 2.4 Hz, 1H), 4.48 (s, 2H), 3.62 (t, J = 6.0 Hz, 2H), 3.15 (t, J = 5.8 Hz, 2H). ^{13}C NMR (126 MHz, MeOD) δ 146.12 (d, $J_{C\text{-}F}$ = 13.8 Hz), 141.72 (d, $J_{C\text{-}F}$ = 8.0 Hz), 138.17, 137.86 (d, J = 4.3 Hz), 129.38, 128.26, 127.80, 122.69, 117.66, 113.10, 110.56, 110.27, 107.64, 43.72, 42.24, 19.46. ^{19}F NMR (470 MHz, MeOD) δ -75.51 (s), -76.97 (s) (TFA). HRMS calculated for $C_{16}H_{15}FN_3$ $[M + H]^+$: 268.1250, found: 268.1265.

***N*-(4-(trifluoromethyl)pyridin-2-yl)-2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indol-6-amine (2h): 2h** was prepared from tryptoline (10 mg, 0.058 mmol) and 2-amino-4-trifluromethylpyridine following the general procedure for enzymatic bromination with 3-SS and Buchwald-Hartwig amination outlined above in 87% yield (22.4 mg **2h**·TFA, 0.050 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.96 (d, J = 6.6 Hz, 1H), 7.62 – 7.51 (m, 2H), 7.39 (s, 1H), 7.17 (dd, J = 8.5, 1.7 Hz, 1H), 7.13 (d, J = 6.7 Hz, 1H), 4.49 (s, 2H), 3.60 (t, J = 5.8 Hz, 2H), 3.11 (br 2H). ^{13}C NMR (126 MHz, MeOD) δ 160.75, 139.54, 137.93, 129.09, 128.69, 123.38, 120.71, 120.53, 118.89, 116.54,

114.47, 112.28, 109.35, 107.86, 43.60, 42.12, 19.36. ^{19}F NMR (470 MHz, MeOD) δ -68.18, -77.67 (TFA). HRMS calculated for $C_{17}H_{16}F_3N_4$ $[M + H]^+$: 333.1327, found: 333.1327.

***N*-(6-methylpyridin-2-yl)-2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indol-6-amine (2i):** **2i** was prepared from tryptoline (10 mg, 0.058 mmol) and 2-amino-6-methylpyridine following the general procedure for enzymatic bromination with 3-SS and Buchwald-Hartwig amination outlined above in 63% yield (14.3 mg **2i**·TFA, 0.036 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.81 (dd, *J* = 17.4, 9.8 Hz, 2H), 7.51 (s, 1H), 7.14 (dd, *J* = 8.5, 1.6 Hz, 1H), 6.84 (d, *J* = 8.9 Hz, 1H), 6.80 (d, *J* = 7.0 Hz, 1H), 4.49 (s, 2H), 3.60 (t, *J* = 6.0 Hz, 2H), 3.09 (t, *J* = 5.7 Hz, 2H). ^{13}C NMR (126 MHz, MeOD) δ 149.46, 145.92, 137.41, 128.97, 128.70, 128.50, 121.07, 116.37, 114.25, 113.94, 113.23, 110.36, 107.64, 43.62, 42.13, 19.36, 19.07. HRMS calculated for $C_{17}H_{19}N_4$ $[M + H]^+$: 279.1609, found: 279.1621.

***N*-(quinolin-3-yl)-2,3,4,9-tetrahydro-1*H*-pyrido[3,4-*b*]indol-6-amine (2j):** **2j** was prepared from tryptoline (10 mg 0.058 mmol) and 3-aminoquinoline following the general procedure for enzymatic bromination with 3-SS and Buchwald-Hartwig amination outlined above in 51% yield (12.7 mg **2j**·TFA, 0.030 mmol). ^{1}H NMR (500 MHz, MeOD) δ 8.70 (s, *N*-H, 1H), 8.04 (s, 1H), 7.94 (d, *J* = 7.7 Hz, 1H), 7.82 (d, *J* = 8.3 Hz, 1H), 7.63 (m, 2H), 7.42 (m, *J* = 8.5 Hz, 2H), 7.12 (d, *J* = 8.6 Hz, 1H), 4.44 (s, 2H), 3.57 (br, 3H), 3.05 (br, 3H). ^{13}C NMR (126 MHz, MeOD) δ 143.91, 137.94, 136.13, 133.41, 131.69, 130.48, 130.03, 128.51, 128.10, 128.02, 123.35, 122.66, 120.98, 119.69, 113.69, 113.30, 107.25, 43.73, 42.26, 19.45. HRMS calculated for $C_{20}H_{19}N_4$ $[M + H]^+$: 315.1609, found: 315.1619.

1-(2-(2-methoxyphenoxy)ethylamino)-3-(1-(4-methoxyphenyl)-9*H*-carbazol-4-yloxy)propan-2-ol (11a): 11a was prepared from carvedilol (10 mg, 0.025 mmol) and 4-methoxyphenylboronic acid following the general procedure for enzymatic chlorination with 4-V. The crude extracts from the bioconversion were transferred to a 20 mL round scintillation vial and $Pd(OAc)_2$ (0.28 mg, 0.0012 mmol, 0.05 equiv.), Sphos (0.50 mg, 0.0012 mmol, 0.05 equiv.), 4-MeO-C_6H_4-$B(OH)_2$ (5.61 mg, 0.037 mmol, 1.5 equiv), K_3PO_4 (10.4 mg, 0.043 mmol, 2 equiv.), and a magnetic stir bar were added. The vial was transferred to an inert atmosphere dry box, dioxane (0.8 mL) and water (0.2 mL) were added, and the vial was sealed with a teflon lined cap. The vial was removed from the dry box and the mixture was allowed to stir in a 100 °C oil bath. After 12 hours, the reaction vessel was allowed to cool to room temperature, and the contents filtered over silica, eluting with 150 mL 4:1 CH_2Cl_2/MeOH. The filtrate was collected and concentrated by rotary evaporation. The crude mixture was purified by reversed-phase chromatography (Biotage SNAP-KP-C18-HS, gradient from pure H_2O to 60% MeCN/H_2O) and isolated in 93% yield (14.3 mg **11a**·TFA, 0.023 mmol). ^{1}H NMR (500 MHz, MeOD) δ 8.28 (d, *J* = 7.8 Hz, 1H), 7.56 (dd, *J* = 6.6, 4.8 Hz, 2H), 7.46 (d, *J* = 8.1 Hz, 1H), 7.32 – 7.25 (m, 2H), 7.08 (d, *J* = 8.6 Hz, 2H), 7.05 – 6.95 (m, 4H), 6.93 – 6.88 (m, 1H), 6.81 (d, *J* = 8.2 Hz, 1H), 4.56 (dt, *J* = 9.5, 5.8 Hz, 1H), 4.42 (dd, *J* = 9.9, 4.6 Hz, 1H), 4.31 (dt, *J* = 13.5, 6.7 Hz, 3H), 3.87 (s, 3H), 3.78 (s, 3H), 3.67 (dd, *J* = 12.7, 2.9 Hz, 1H), 3.60 – 3.55 (m, 2H), 3.50 (dd, *J* = 12.6, 9.9 Hz, 1H). ^{13}C NMR (126 MHz, MeOD) δ 160.36, 155.19, 151.14, 148.43, 141.07, 140.21, 132.82, 130.53, 127.21, 125.84, 124.25, 123.69, 123.56, 122.35, 120.47, 120.04, 116.69, 115.46, 113.97, 113.38, 111.75, 102.35, 71.05, 66.83, 66.23, 56.42, 55.84, 51.76, 48.38. HRMS calculated for $C_{31}H_{32}N_2O_5$ $[M + H]^+$: 513.2389, found: 513.2400.

1-(isopropylamino)-3-(7-(4-methoxyphenyl)-1*H*-indol-4-yloxy)propan-2-ol (9a): 9a was prepared from pindolol (10 mg, 0.040) and 4-methoxyphenylboronic acid following the general procedure for enzymatic bromination with 4-V and Suzuki-Miyaura Coupling outlined above in >99% yield (19.4 mg **9a**·TFA, 0.041 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.50 (d, *J* = 8.7 Hz, 2H), 7.23 – 7.12 (m, 1H), 7.10 – 6.95 (m, 3), 6.68 – 6.56 (m, 2H), 4.38 – 4.29 (m, 1H), 4.25 (dd, *J* = 9.9, 4.9 Hz, 1H), 4.15 (dd, *J* = 9.9, 6.0 Hz, 1H), 3.48 (dt, *J* = 13.1, 6.5 Hz, 1H), 3.37 (dd, *J* = 12.6, 2.9 Hz, 1H), 3.23 (dd, *J* = 12.6, 9.6 Hz, 1H), 1.38 (dd, *J* = 6.5, 4.5 Hz, 6H). ^{13}C NMR (126 MHz, MeOD) δ 160.22, 152.36, 133.10, 130.22, 124.99, 124.83, 122.76, 121.42, 120.62, 115.36, 102.08, 99.94, 71.07, 67.08, 55.79, 52.10, 19.33, 18.81. HRMS calculated for $C_{21}H_{27}N_2O_3$ $[M + H]^+$: 355.2021, found: 355.2036.

***N*-(4'-methoxybiphenyl-4-yl)-1-methyl-*N*-(thiophen-2-ylmethyl)piperidin-4-amine (12a): 12a** was prepared from thenalidine (10 mg, 0.035 mmol) and 4-methoxyphenylboronic acid following the general procedure for enzymatic bromination with 4-V and Suzuki-Miyaura Coupling outlined above in 74% yield (13.1 mg **12a**·TFA, 0.026 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.44 (dd, *J* = 15.2, 8.5 Hz, 4H), 7.23 (d, *J* = 4.8 Hz, 1H), 7.06 – 6.88 (m, 5H), 4.65 (s, 2H), 4.06 (t, *J* = 11.5 Hz, 1H), 3.80 (s, 3H), 3.58 (d, *J* = 11.2 Hz, 2H), 3.18 (t, *J* = 11.9 Hz, 2H), 2.87 (s, 3H), 2.27 – 2.09 (m, 2H), 2.00 (d, *J* = 12.0 Hz, 2H). ^{13}C NMR (126 MHz, MeOD) δ 128.36, 128.25 , 128.11, 128.04, 127.76, 125.51, 125.15, 118.01, 115.21, 115.16, 55.74, 55.49, 47.04, 43.82, 31.11, 28.46 (Some aryl carbons cannot be distinguished due to overlap). HRMS calculated for $C_{24}H_{29}N_2OS$ $[M + H]^+$: 393.2000, found: 393.2000.

1-methyl-*N*-(thiophen-2-ylmethyl)-*N*-(4-(2,2,2-trifluoroethoxy)phenyl)piperidin-4-amine (12b): 12b was prepared from thenalidine (10 mg, 0.035 mmol) following the general procedure for enzymatic bromination with 4-V. The crude extracts from the bioconversion were transferred to a 20 mL round scintillation vial and $[(allyl)PdCl]_2$ (0.06 mg, $1.74x10^{-4}$ mmol, 0.005 equiv.), RockPhos (0.24 mg, $5.22x10^{-4}$ mmol, 0.015 equiv.), Cs_2CO_3 (20.6 mg, 0.070 mmol, 2 equiv.), and a magnetic stir bar were added. The vial was transferred to an inert atmosphere dry box, toluene (1 mL) and trifluoroethanol (5.3 µL, 0.070 mmol, 2 equiv.) were added, and the vial was sealed with a teflon lined cap. The vial was removed from the dry box and the mixture was allowed to stir in a 90 °C oil bath. After 14 hours, the reaction vessel was allowed to cool to room temperature, and the contents filtered over silica, eluting with 150 mL 4:1 CH_2Cl_2/MeOH. The filtrate was collected and concentrated by rotary evaporation. The crude mixture was purified by reversed-phase chromatography (Biotage SNAP-KP-C18-HS, gradient from pure H_2O to 60% MeCN/H_2O) and isolated in 33% yield (5.6 mg **12b**·TFA, 0.011 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.24 (dd, *J* = 5.1, 1.1 Hz, 1H), 7.06 (m, 2H), 6.94 (m, 3H), 6.91 – 6.87 (m, 1H), 4.63 (s, 2H), 4.45 (dd, *J* = 17.0, 8.5 Hz, 2H), 3.80 (t, *J* = 11.7 Hz, 1H), 3.56 (d, *J* = 12.2 Hz, 2H),3.11 (t, *J* = 11.7 Hz, 2H), 2.85 (s, 3H), 2.21 (d, *J* = 13.6 Hz, 2H), 1.89 (m, *J* = 23.9, 11.2 Hz, 2H). ^{13}C NMR (126 MHz, MeOD) δ 161.32, 127.59, 126.52, 125.84, 124.10, 123.30, 120.17, 117.73, 116.83, 67.38, 67.10, 55.26, 49.85, 43.72, 28.59. ^{19}F NMR (470 MHz, MeOD) δ -75.89, -77.13 (TFA). HRMS calculated for $C_{19}H_{24}F_3N_2OS$ $[M + H]^+$: 385.1561, found: 385.1574.

3.4.4 DEMONSTRATIONS OF NOVEL SELECTIVITY

Enzyme-Free Chlorinations: To demonstrate that the selectivity and/or reactivity we observed using our RebH variants 3-SS and 4-V (Tables 3.3 & 3.4), we attempted to perform chlorinations in the absence of our enzymes to compare the activities and selectivities observed.

Three substrates were tested: **2**, tryptoline; **6**, pinoline; and **11**, carvedilol. Three chlorination conditions were initially tested, the first using N-chlorosuccinimide in MeCN[51] and the latter two using sodium hypochlorite in aqueous buffer (one in 25 mM HEPES, pH 7.4, and the second in 25 mM HEPES, pH 5.0).[52] In all three reaction conditions on all three substrates, 1 equiv. of the chlorinating reagent was added and the reactions proceeded at room temperature, at 600 rpm, for 12 hours. After addition of an equal volume of MeOH and analysis by LCMS, no conversion of **2** or **6** was observed, but low conversion of **11** was seen to at least two monochlorinated products and at least two dichlorinated products.

To try to see conversion of **2** or **6**, the same reaction chlorinating reagents described above were again employed, but this time under harsher reaction conditions. 10 equivs. of the chlorinating reagent were added to each reaction, and the reactions were heated to 50 °C, at 600 rpm, for 2 days. After addition of an equal volume of MeOH and analysis by LCMS, higher consumption of **11** was seen with the same product distribution described above with additional appearance of trichlorinated product as well (see chromatograms shown below). However, there were still no mono-, di-, or trichlorinated products observed in any of the reactions of **2** or **6**.

Figure 3.6. Analysis of NaOCl chlorination of carvedilol (11).[a]

[a] UV (280 nm) chromatogram and extracted ion chromatograms (for starting material, monochlorination, dichlorination, and trichlorination) shown for the chlorination of 11, carvedilol, using 10 equiv. NaOCl in 25 mM HEPES, pH 7.4 at 50 °C for 2 days. Majority of starting material still remains, but at least two resolved monochlorination products, as well as di- and trichlorination products are visible.

We therefore conclude from this experiment that with at least substrate **11** we observe novel selectivity, as we have isolated and characterized only a single monochlorinated product whereas at least two were observed in chlorinations in the absence of enzyme. Furthermore, our trials on substrates **2** and **6** were not able to produce cleanly chlorinated derivatives of these substrates, even at high loadings of chlorinating agents at elevated temperatures for long reaction times, whereas we saw immediate conversion to a single monochlorinated derivative of these substrates at room temperature.

Preserved Regioselectivity on Small Substrates:

As an additional test that we have not reduced the selectivity imparted by our enzyme variants 3-SS and 4-V relative to that established for wild-type RebH (we speculated that by expanding the active site to accommodate larger substrates, we might have reduced substrate binding such that we would only halogenate the most electronically activated site(s)), we retested halogenations of two small substrates with which we reported novel regioselectivity with wild-type RebH. Chlorination reactions of both tryptamine and tryptophol (both of which we previously reported[12] and both of which are unprotected at the 2-position, which is by far the most electronically activated position on these substrates) were set up with wild-type RebH, 3-SS, and 4-V. After the reactions were quenched and analyzed by UPLC, the products were seen to perfectly coelute to the best of our ability to resolve any potential regioisomers by HPLC. We therefore conclude that 3-SS and 4-V are providing the same regioselectivity we previously reported on these substrates with wild-type RebH, 7-chlorination, and therefore we are not simply halogenating the most electronically activated site with these enzyme variants.

3.4.5 HANES-WOOLF PLOTS

All measurements were conducted in triplicate. Error bars on saturation curves represent standard deviations of the triplicate measurements of rate. Hanes-Woolf plots were then constructed from the averages of these three rates.

Figure 3.7. Saturation curve of wild-type RebH conversion of tryptoline.

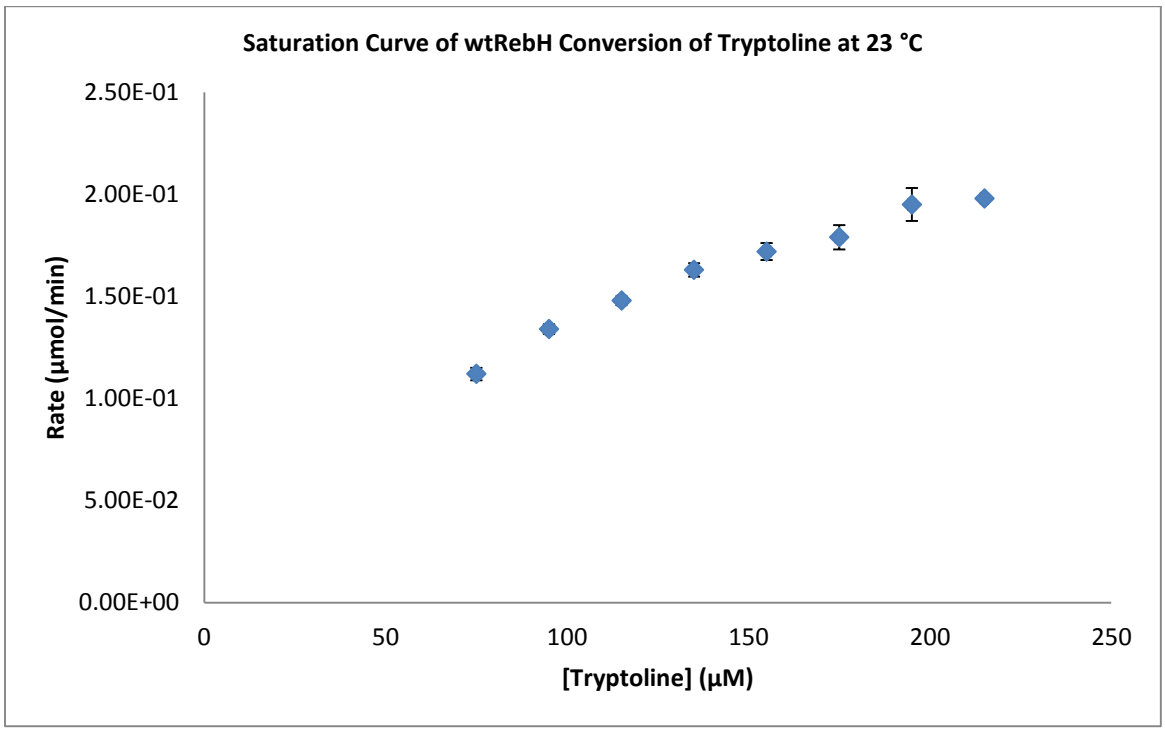


Figure 3.8. Hanes-Woolf plot of wild-type RebH conversion of tryptoline.

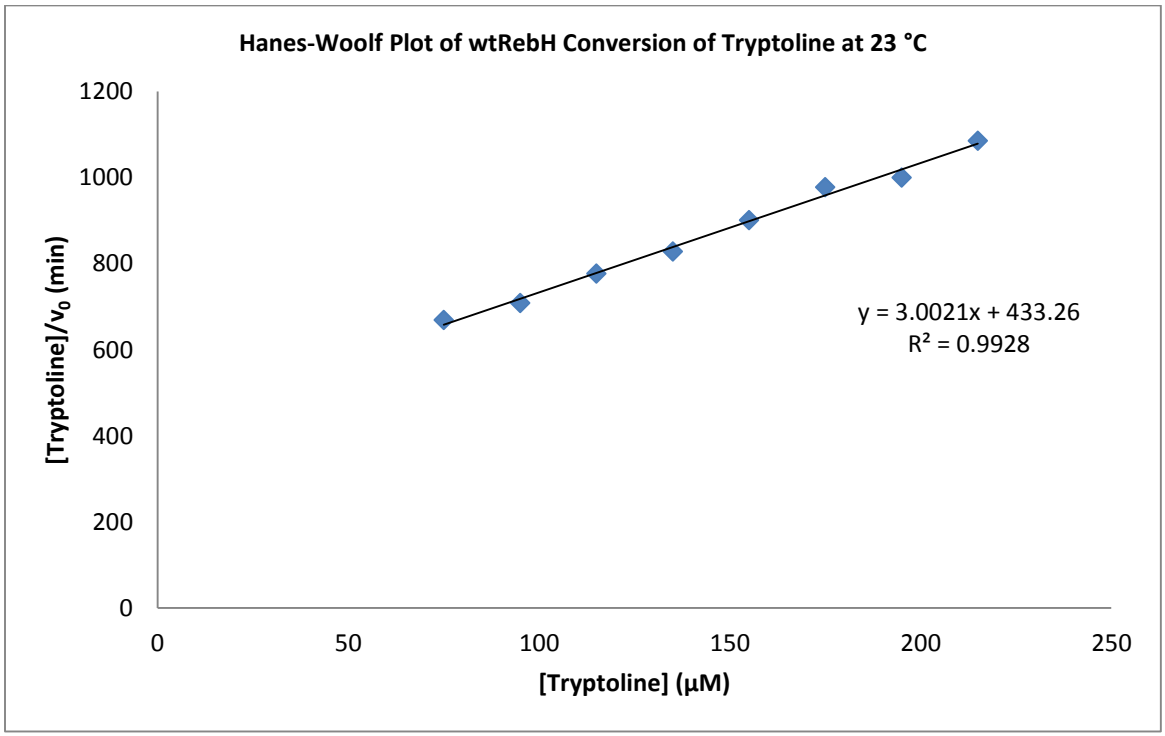


Figure 3.9. Saturation curve of 1-PVM conversion of tryptoline.

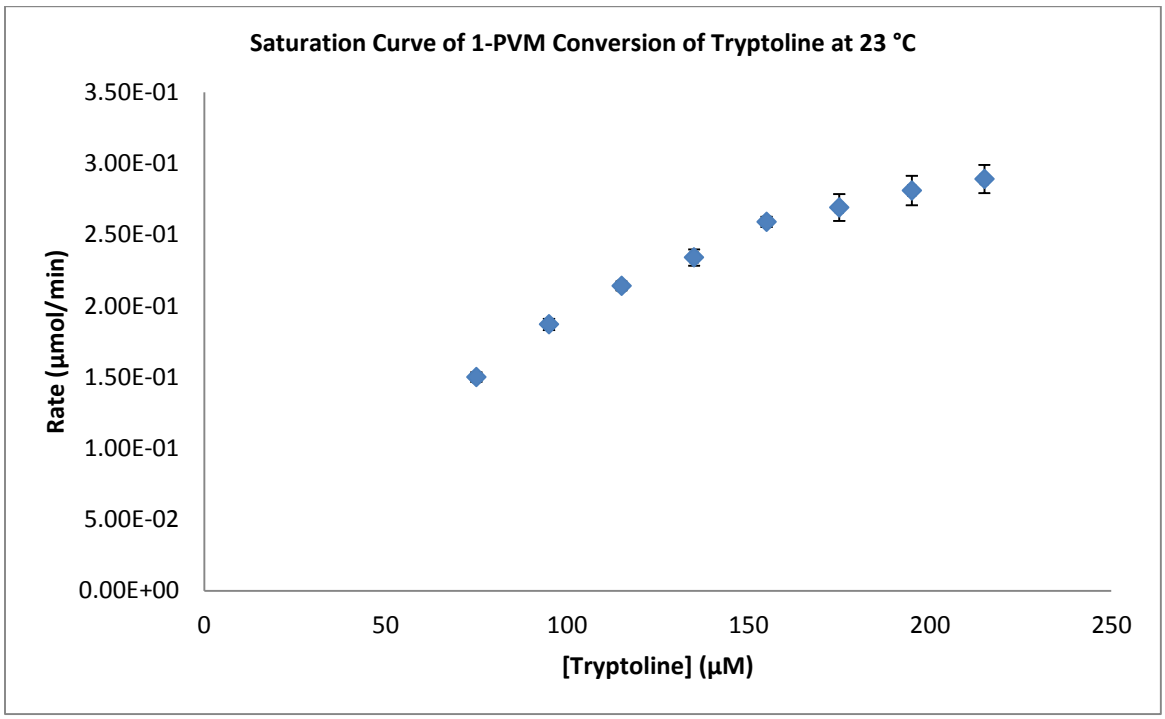


Figure 3.10. Hanes-Woolf plot of 1-PVM conversion of tryptoline.

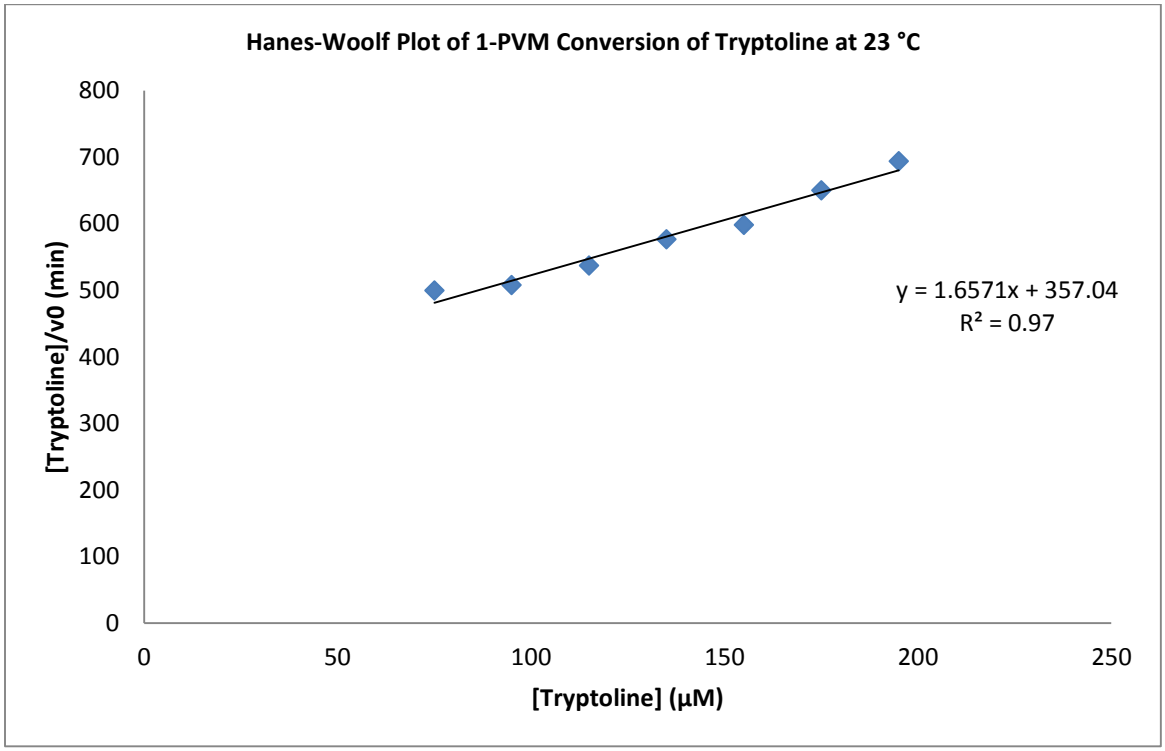


Figure 3.11. Saturation curve of 2-T conversion of tryptoline.

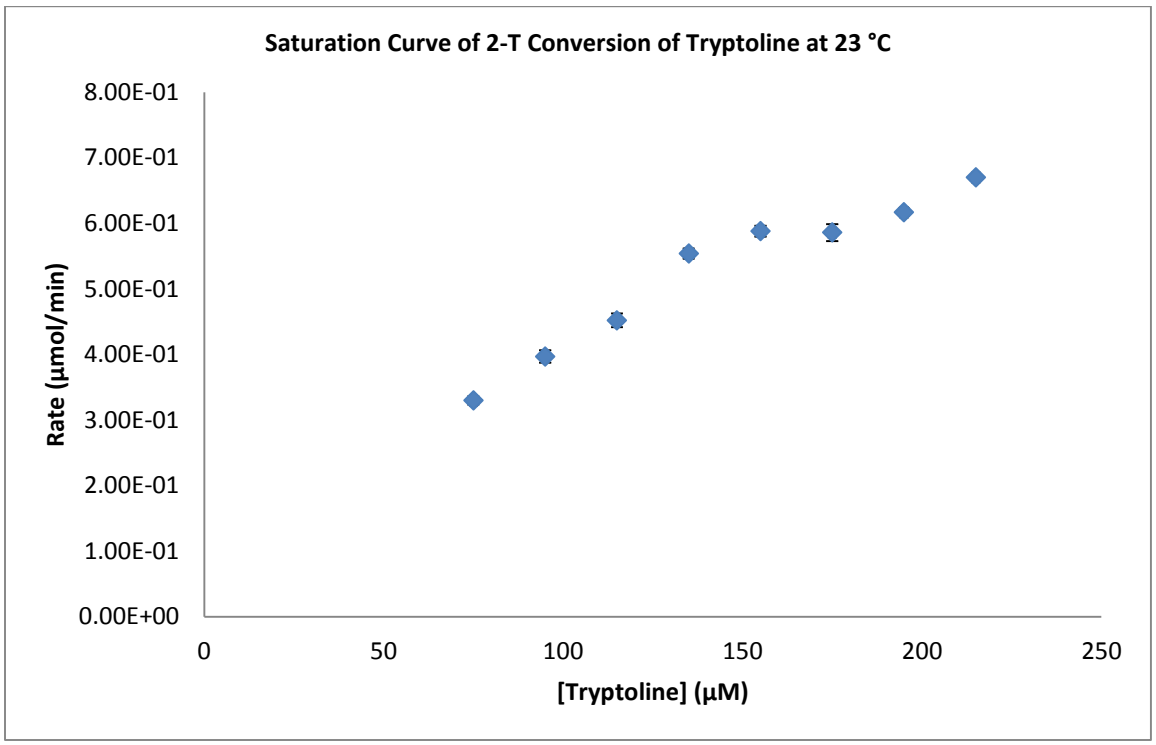


Figure 3.12. Hanes-Woolf plot of 2-T conversion of tryptoline.

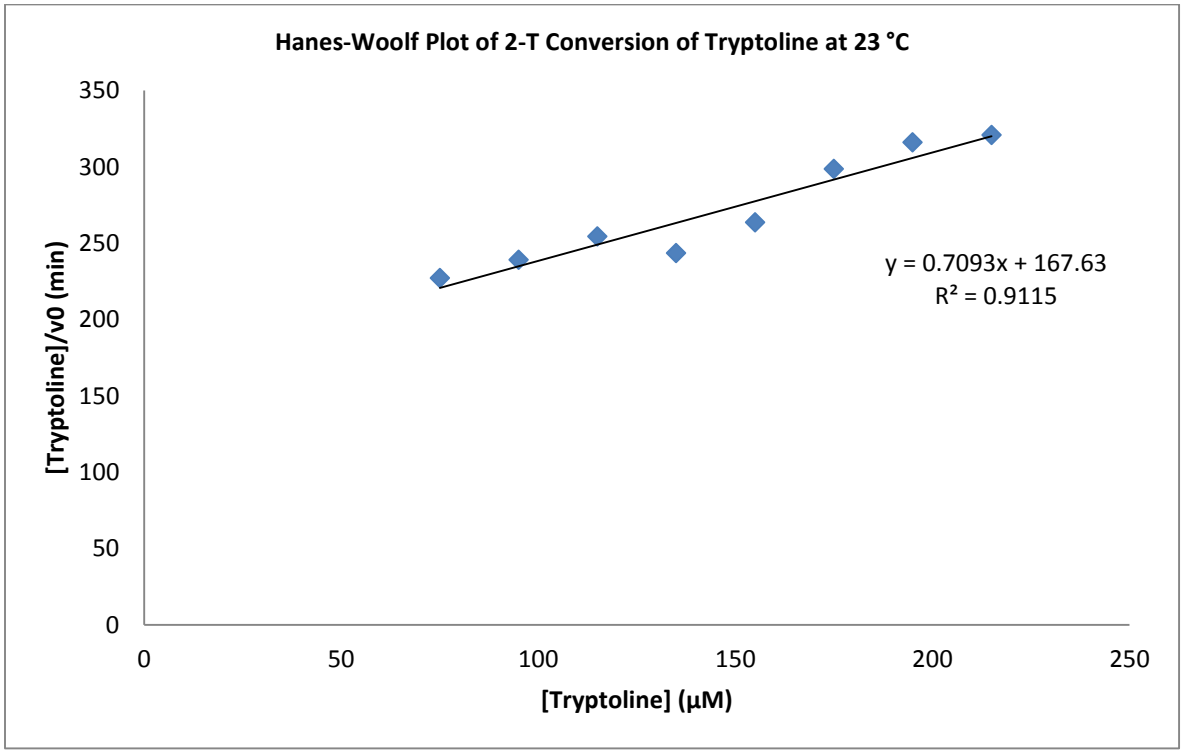


Figure 3.13. Saturation curve of 3-SS conversion of tryptoline.

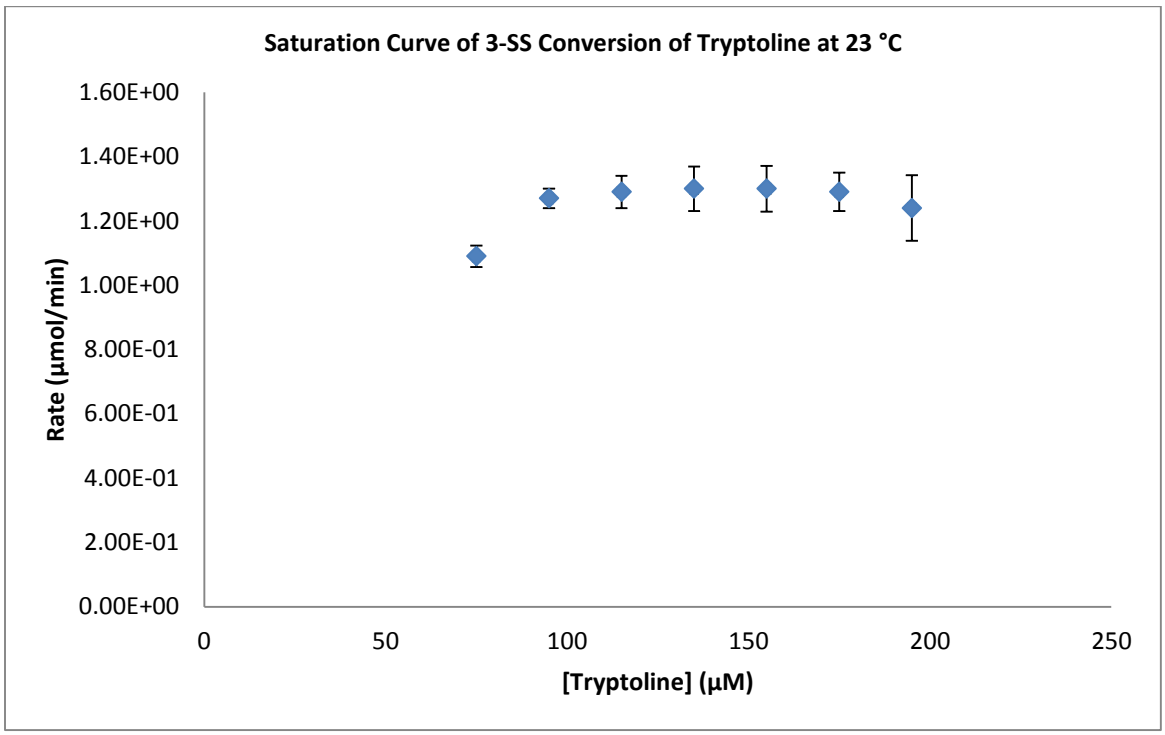
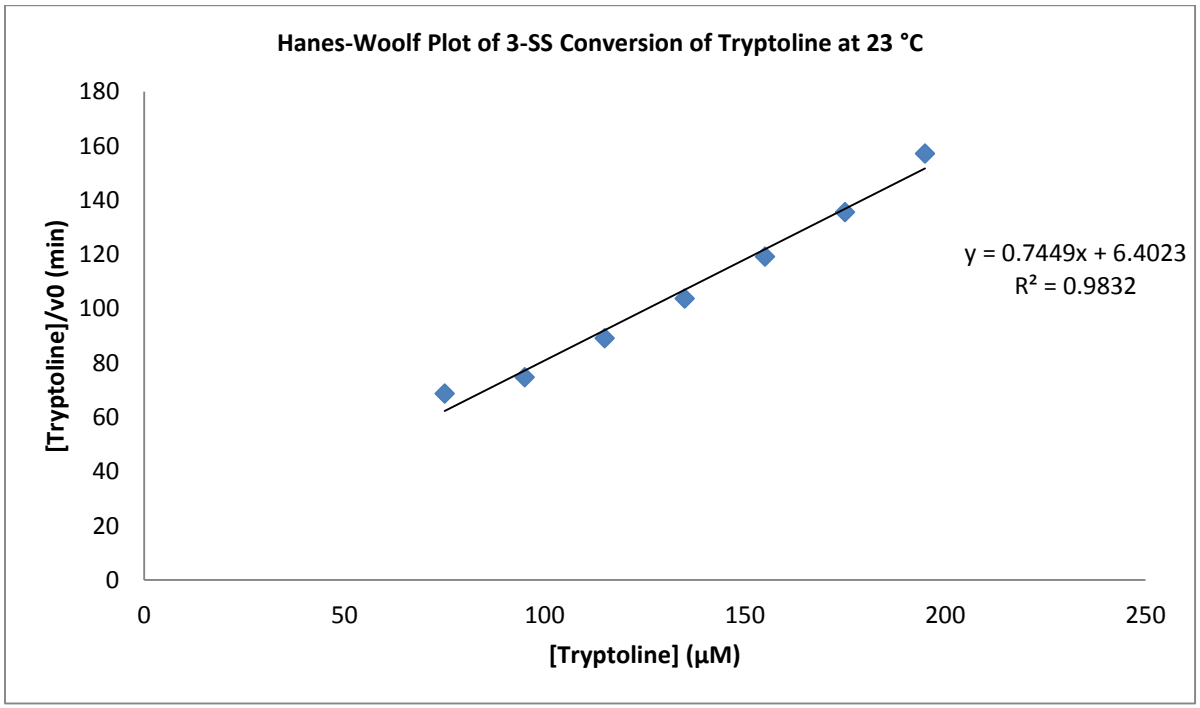


Figure 3.14. Hanes-Woolf plot of 3-SS conversion of tryptoline.



3.5 ACKNOWLEDGEMENTS

We would like to thank Dr. Catherine Poor for her work and discussions that made Section 3.2.2 of this thesis possible. We would also like to thank Dr. Landon Durak for his work and discussions that made Section 3.2.4 of this thesis possible. In addition, large portions of the text and figures in Section 3.2.4 were adapted with permission from the manuscript written by Dr. Landon Durak. In addition, we would like to thank The Novartis Institutes for Biomedical Research for the donation of thenalidine to the Center for C-H Functionalization (CCHF). Lastly, we would also like to thank Kyle Kunze for aiding in the synthesis of one of the batches of the dFBr starting material.

3.6 REFERENCES

[1] Metal-Catalyzed Cross-Coupling Reactions (Eds.: Diederich, F.; Stang, P. J.), Wiley-VCH, New York, 1998.

[2] a) Hernandes, M. Z., Cavalcanti, S. M. T., Moreira, D. R. M., de Azevedo Junior, W. F., & Leite. A. C. L. Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design. *Curr. Drug Targets* **11**, 303–314 (2010); b) Naumann, K. Influence of chlorine substituents on biological activity of chemicals: a review. *Pest Manag. Sci.* **56**, 3–21 (2000).

[3] Jeschke, P. The unique role of halogen substituents in the design of modern agrochemicals. *Pest Manag. Sci.* **66**, 10–27 (2010).

[4] Herrera-Rodriguez, L. N., Khan, F., Robins, K. T., Meyer, H.-P. Perspectives on biotechnological halogenation. Part I: Halogenated products and enzymatic halogenation. *Chem. Today* **29**, 31 (2011).

[5] Bunders, C. A., Minvielle, M. J., Worthington, R. J., Ortiz, M., Cavanagh, J., Melander, C. Intercepting Bacterial Indole Signaling with Flustramine Derivatives. *J. Am. Chem. Soc.* **133**, 20160 (2011).

[6] Williams, P. G., Buchanan, G. O., Feling, R. H., Kauffman, C. A., Jensen, P. R., Fenical, W. New Cytotoxic Salinosporamides from the Marine Actinomycete Salinispora tropica. *J. Org. Chem.* **70**, 6196–6203 (2005).

[7] Smith, B. M.; Smith, J. M.; Tsai, J. H.; Schultz, J. A.; Gilson, C. A.; Estrada, S. A.; Chen, R. R.; Park, D. M.; Prieto, E. B.; Gallardo, C. S.; Sengupta, D.; Dosa, P. I.; Covel, J. A.; Ren, A.; Webb, R. R.; Beeley, N. R. A.; Martin, M.; Morgan, M.; Espitia, S.; Saldana, H. R.; Bjenning, C.; Whelan, K. T.; Grottick, A. J.; Menzaghi, F.; Thomsen, W. J. Discovery and Structure-Activity Relationship of (1*R*)-8-Chloro-2,3,4,5-tetrahydro-1-methyl-1*H*-3-benzazepine (Lorcaserin), a Selective Serotonin 5-HT_{2C} Receptor Agonist for the Treatment of Obesity. *J. Med. Chem.* **51**, 305-313 (2008).

[8] Sun, H.; Keefer, C. E.; Scott, D. O. Systematic and Pairwise Analysis of the Effects of Aromatic Halogenation and Trifluoromethyl Substitution on Human Liver Microsomal Clearance. *Drug Metab. Lett.* **5**, 232-242 (2011).
[9] Smith, K.; El-Hiti, G. A. Regioselective control of electrophilic aromatic substitution reactions. *Curr. Org. Synth*. **1**, 253 – 274 (2004).
[10] Dong, C.J.; Flecks, S.; Unversucht, S.; Haupt, C.; van Pée, K.-H.; Naismith, J. H. Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science* **309**, 2216 – 2219 (2005).
[11] 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.* **102**, 3960-3965 (2005).
[12] Payne, J. T.; Andorfer, M. C.; Lewis, J. C. Regioselective Arene Halogenation using the FAD-Dependent Halogenase RebH. *Angew. Chem. Int. Ed.* **52**, 5271-5274 (2013). Note: we previously referred to the 6- and 7-halogenation of tryptoline as 5- and 6-halogenation, by analogy to the numbering used for L-tryptophan, but here we use the established numbering for carbazole derivatives.
[13] Hölzer, M.; Burd, W.; Reißig, H.-U.; van Pée, K.-H. Substrate Specificity and Regioselectivity of Tryptophan 7-Halogenase from *Pseudomonas fluorescens* BL915. *Adv. Synth. Catal.* **343**, 591-595 (2001).
[14] Frese, M.; Guzowska, P. H.; Voß, H.; Sewald, N. Regioselective Enzymatic Halogenation of Substituted Tryptophan Derivatives using the FAD-Dependent Halogenase RebH. *ChemCatChem* **6**, 1270–1276 (2014).
[15] Frese, M.; Sewald, N. Enzymatic Halogenation of Tryptophan on a Gram Scale. *Angew. Chem. Int. Ed.* **54**, 298-301 (2015).
[16] a) Runguphan, W.; Qu, X.; O'Connor, S. E. Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461-464 (2010); b) Roy, A. D.; Grüschow, S.; Cairns, N.; Goss, R. J. M. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J. Am. Chem. Soc.* **132**, 12243–12245 (2010); c) Sánchez, C.; Zhu, L.; Braña, A. F.; Salas, A. P.; Rohr, J.; Méndez, C.; Salas, J. A. Combinatorial biosynthesis of antitumor indolocarbazole compounds. *Proc. Natl. Acad. Sci.* **102**, 461-466 (2005).
[17] Glenn, W. S.; Nims, E.; O'Connor, S. E. Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *J. Am. Chem. Soc.* **133**, 19346-19349 (2011).
[18] a) Payne, J. T.; Poor, C. B.; Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angew. Chem. Int. Ed.* **54**, 4226-4230 (2015); b) Durak, L. J.; Payne, J. T.; Lewis, J. C. Late-Stage Diversification of Biologically Active Molecules via Chemoenzymatic C-H Functionalization. Manuscript in preparation.
[19] Bloom, J. D., Labthavikul, S. T., Otey, C. R., & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci*. **103**, 5869-5874 (2006).
[20] Poor, C. B.; Andorfer, M. C.; Lewis, J. C. Improving the Stability and Catalyst Lifetime of the Halogenase RebH By Directed Evolution. *ChemBioChem* **15**, 1286-1289 (2014).
[21] McGrath, N. A., Brichacek, M., & Njardarson, J. T. A Graphical Journey of Innovative Chemical Architectures That Have Improved Our Lives. *J. Chem. Educ.* **87**, 1348-1349 (2010).
[22] Zhang, M.-Z., Chen, Q., & Yang, G.-F. A review on recent developments of indole-containing antiviral agents. *Eur. J. Med. Chem.* **89**, 421-441 (2015).
[23] Hall-Stoodley, L., Costerton, J. W., & Stoodley, P. Bacterial Biofilms: From the Natural Environment to Infectious Diseases. *Nat. Rev. Microbiol.* **2**, 95-108 (2004).

[24] Heckman, K. L.; Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2**, 924–932 (2007).
[25] Adla, S. K., Golz, G., Jones, P. G., & Lindel, T. Study on the NBS-Induced Rearrangement of 2-tert-Prenyltryptamines. *Synthesis* **13**, 2161-2170 (2010).
[26] a) Shimotohno, A., Oue, S., Yano, T., Kuramitsu, S., & Kagamiyama, H. Demonstration of the Importance and Usefulness of Manipulating Non-Active-Site Residues in Protein Design. *J. Biochem*. **129**, 943–948 (2001); b) Spiller, B., Gershenson, A., Arnold, F. H., & Stevens, R. C. A structural view of evolutionary divergence. *Proc. Natl. Acad. Sci. USA* **96**, 12305–12310 (1999); c) Romero, P. A., & Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009); d) Fasan, R., Chen, M. M., Crook, N. C., & Arnold, F. H. Engineered Alkane-Hydroxylating Cytochrome $P450_{BM3}$ Exhibiting Nativelike Catalytic Properties. *Angew. Chem. Int. Ed.* **46**, 8414–8418 (2007).
[27] Idowu, T. O., Iwalewa, E. O., Aderogba, M. A., Akinpelu, B. A., & Ogundaini, A. O. Antinociceptive, Anti-inflammatory and Antioxidant activities of Eleagnine: An alkaloid Isolated from *Chrysophyllum albidum* Seed Cotyledons. *J. Biol. Sci.* **6**, 1029–1034 (2006).
[28] Pähkla, R., Harro, J., & Rägo, L. Behavioural effects of pinoline in the rat forced swimming, open field and elevated plus-maze tests. *Pharmacol. Res.* **34**, 73– 78 (1996).
[29] Song, H., Liu, Y., Liu, Y., Wang, L., & Wang, Q. Synthesis and Antiviral and Fungicidal Activity Evaluation of β-Carboline, Dihydro-β-carboline, Tetrahydro-β-carboline Alkaloids, and Their Derivatives. *J. Agric. Food Chem.* **62**, 1010–1018 (2014).
[30] Tam, S. W., Worcel, M., & Wyllie, M. Yohimbine: a clinical review. *Pharmacol. Ther*. **91**, 215–243 (2001).
[31] Kobayashi, Y., Nakano, Y., Kizaki, M., Hoshikuma, K., Yokoo, Y., & Kamiya, T. Capsaicin-Like Anti-Obese Activities of Evodiamine from Fruits of *Evodia rutaecarpa*, a Vanilloid Receptor Agonist. *Planta Med*. **67**, 628–633 (2001).
[32] Aellig, W. H. Pindolol - a β-adrenoceptor blocking drug with partial agonist activity: clinical pharmacological considerations. *Br. J. Clin. Pharmacol*. **13**, 187S–192S (1982).
[33] Innis, R. B., Correa, F. M. A., & Snyder, S. H. Carazolol, an extremely potent β-adrenergic blocker - binding to β-receptors in brain membranes. *Life Sci.* **24**, 2255–2264 (1979).
[34] Stafylas, P. C. & Sarafidis, P. A. Carvedilol in hypertension treatment. *Vasc. Health Risk Manag*. **4**, 23–30 (2008).
[35] Bitto, E.; Bingman, C. A.; Singh, S.; Phillips, Jr., G. N. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70**, 289–293 (2008).
[36] Fasan, R. Tuning P450 enzymes as oxidation catalysts. *ACS Catal.* **2**, 647–666 (2012).
[37] Sawayama, A. M.; Chen, M. M. Y.; Kulanthaivel, P.; Kuo, M.-S.; Hemmerle, H.; Arnold, F. H. A panel of cytochrome P450 BM3 variants to produce drug metabolites and diversify lead compounds. *Chemistry* **15**, 11723–11729 (2009).
[38] Lewis, J. C.; Bastian, S.; Bennett, C. S.; Fu, Y.; Mitsuda, Y.; Chen, M. M.; Greenberg, W. A.; Wong, C.-H.; Arnold, F. H. Combinatorial alanine substitution enables rapid optimization of cytochrome $P450_{BM3}$ for selective hydroxylation of large substrates. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 16550–16555 (2009).
[39] Rentmeister, A.; Arnold, F. H.; Fasan, R. Chemo-enzymatic fluorination of unactivated organic compounds. *Nat. Chem. Biol.* **5**, 26–28 (2009).
[40] Runguphan, W.; O'Connor, S. E. Diversification of Monoterpene Indole Alkaloid Analogs through Cross-Coupling. *Org. Lett.* **15**, 2850–2853 (2013).

[41] Maiti, D.; Fors, B. P.; Henderson, J. L.; Nakamura, Y.; Buchwald, S. L. Palladium-catalyzed coupling of functionalized primary and secondary amines with aryl and heteroaryl halides: two ligands suffice in most cases. *Chem. Sci.* **2**, 57 (2011).
[42] Wu, X.; Fors, B. P.; Buchwald, S. L. A Single Phosphine Ligand Allows Palladium-Catalyzed Intermolecular C-O Bond Formation with Secondary and Primary Alcohols. *Angew. Chem. Int. Ed.* **50**, 9943–9947 (2011).
[43] a) Reetz, M. T.; Bocoa, M.; Carballeira, J. D.; Zha, D.; Vogel, A. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chem. Int. Ed.* **44**, 4192–4196 (2005); b) Lewis, J. C.; Mantovani, S. M.; Fu, Y.; Snow, C. D.; Komor, R. S.; Wong, C.-H.; Arnold, F. H. Combinatorial Alanine Substitution Enables Rapid Optimization of Cytochrome $P450_{BM3}$ for Selective Hydroxylation of Large Substrates. *ChemBioChem* **11**, 2502–2505 (2010).
[44] a) Shimotohno, A.; Oue, S.; Yano, T.; Kuramitsu, S.; Kagamiyama, H. Demonstration of the Importance and Usefulness of Manipulating Non-Active-Site Residues in Protein Design. *J. Biochem*. **129**, 943–948 (2001); b) Spiller, B.; Gershenson, A.; Arnold, F. H.; Stevens, R. C. A structural view of evolutionary divergence. *Proc. Natl. Acad. Sci. USA* **96**, 12305–12310 (1999); c) Romero, P. A.; Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009); d) Fasan, R.; Chen, M. M.; Crook, N. C.; Arnold, F. H. Engineered Alkane-Hydroxylating Cytochrome $P450_{BM3}$ Exhibiting Nativelike Catalytic Properties. *Angew. Chem. Int. Ed.* **46**, 8414-8418 (2007).
[45] Andorfer, M. C.; Park, H. J.; Vergara-Coll, J.; Lewis, J. C. Directed Evolution of RebH for Catalyst-Controlled Halogenation of Aromatic C-H Bonds. Manuscript in preparation.
[46] Cabrita, L. D.; Dai, W.; Bottomley, S. P. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* **6**, 12 (2006).
[47] Tatsuno, Y.; Yoshida, T.; Otsuka, S. *Inorganic Syntheses*; Angelici, R. J., Ed.; Inorganic Syntheses; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1990; Vol. 28.
[48] Sambrook, J.; Frisch, E. F.; Maniatis, T. (**1989**) *Molecular cloning: a laboratory manual*. 2, Cold Spring Harbor
Laboratory Press, New York.
[49] Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem*. **62**, 7512-7515 (1997).
[50] Durak, L. J.; Lewis, J. C. Transmetalation of Alkyl Ligands from $Cp^*(PMe_3)IrR^1R^2$ to $(cod)PtR^3X$. *Organometallics* **32** (11), 3153–3156 (2013).
[51] Zyuzin, I. N. Chlorination of Geminal Bis(alkoxy-*NNO*-azoxy) Compounds with Sodium Hypochlorite. *Rus. J. Org. Chem.* **50**, 1252-1257 (2014).
[52] Nickson, T. E.; Roche-Dolson, C. A. A Convenient Procedure for the Chlorination of Deactivated Anilines. *Synthesis* **1985(6/7)**, 669-670 (1985).

CHAPTER IV

REMOTE DESYMMETRIZATIONS CATALYZED BY ENGINEERED HALOGENASES

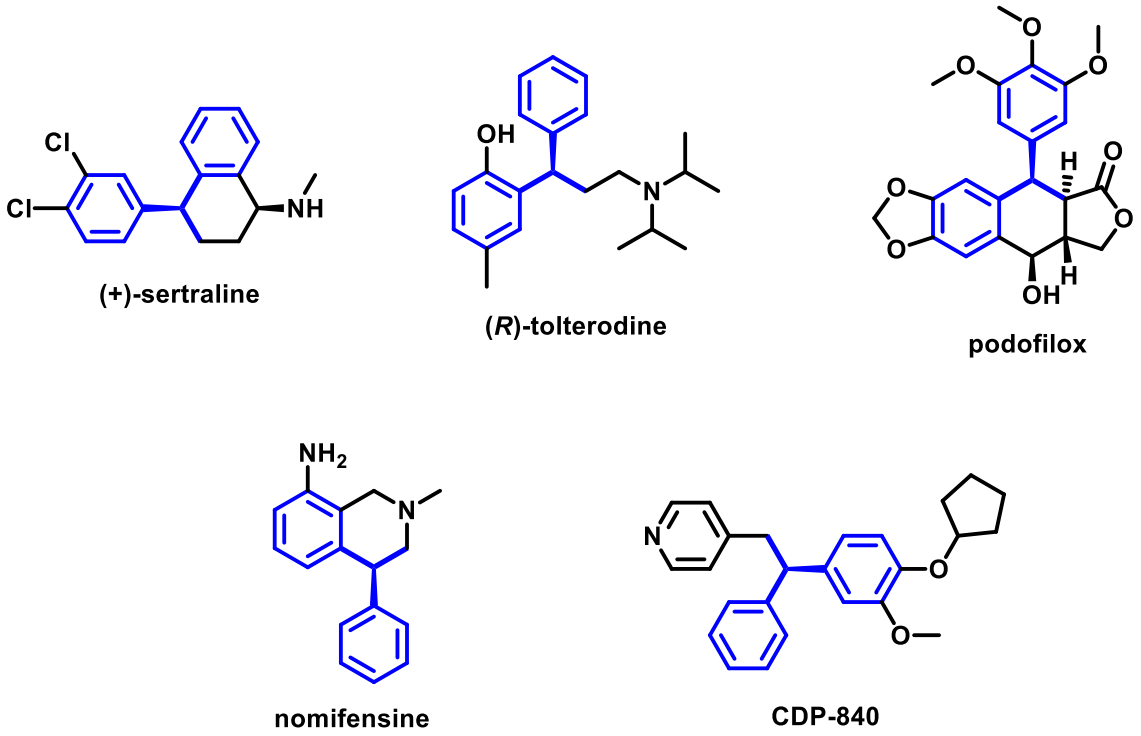
4.1 INTRODUCTION

The FDHs described so far catalyze regioselective halogenations of druglike molecules, but the active sites of these enzymes suggest that they might also catalyze enantioselective halogenations. To date, there have been no reports of the use of FDHs, either in nature or in the laboratory, for the asymmetric catalysis. It was recently reported that the vanadium-dependent chloroperoxidase NapH1 catalyzes a highly stereoselective chlorination-cyclization reaction in the course of the biosynthesis of the antibiotic napyradiomycin.[1] This report is especially exciting, given that many haloperoxidases are known to allow the generated hypohalous acid to diffuse out of the enzyme active site, thus preempting the possibility of using the secondary sphere coordination of the enzyme active site to enable regio- or stereoselective catalysis.[2] Given that the halogenating intermediate in FDH-catalyzed halogenations in held in the enzyme active site,[3] thus giving rise to the regioselectivity of FDHs as has been described in Sections 2 and 3 for RebH, FDHs also offer the potential to be exploited to perform stereoselective halogenations.

Stereochemistry is a major consideration in the synthesis of pharmaceuticals, and the market share of enantipure drugs has increased steadily in recent years.[4] One important class of chiral pharmaceuticals are chiral 1,1-diarylmethanes, or those bearing a diaryl tertiary chiral center.[5] There exist numerous examples of bioactive compounds in this class bearing specifically a diphenyl tertiary chiral center, such as (+)-sertraline (Zoloft), (R)-tolterodine (Detrol), podofilox

(Condylox), CDP-840, and nomifensine (Figure 4.1).[6] The ability to enzymatically install chirality in 1,1-diphenylmethane compounds would be useful to medicinal chemists, and the ability to do so in the course of performing a late-stage halogenation would be especially beneficial. Numerous chiral 1,1-diphenylmethane pharmaceuticals do, in fact, possess arene halogenation as a motif in their structures, such as the antihistamines cetirizine and clemastine and the antihypertensive fenoldopam.[5]

Figure 4.1. Bioactive compounds possessing a diphenyl tertiary chiral center.[a]

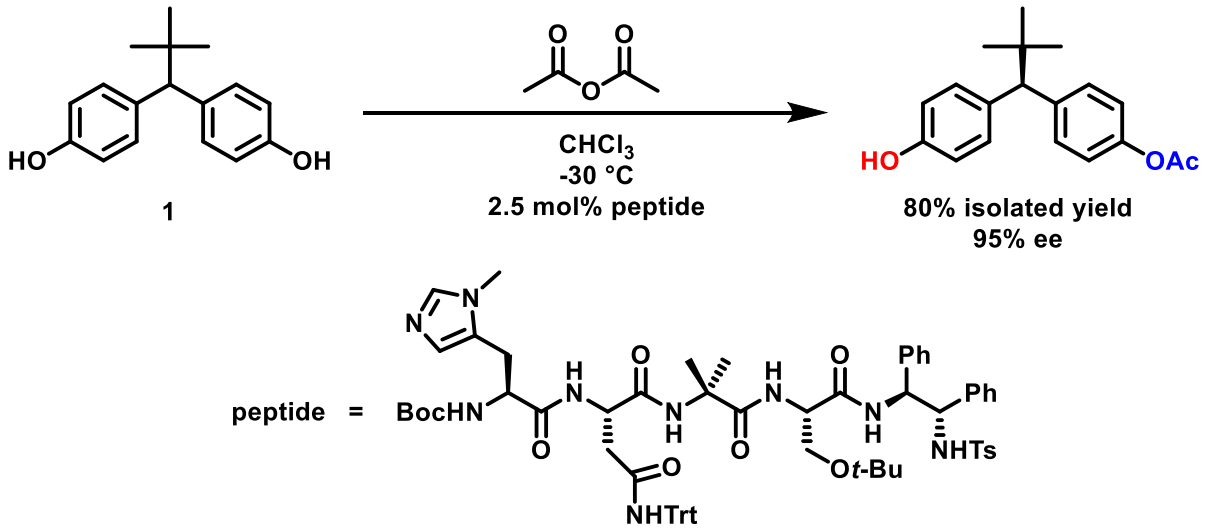


[a] The diphenyl tertiary chiral center motif is highlighted in blue.

By starting with a symmetric 1,1-diarylmethane with a third substituent at the 1-position, one can asymmetrically functionalize one of the two arenes while the third substituent is in a single conformation (either up or down with respect to the plane of the arenes). This results in the remote desymmetrization of the compound to furnish a functionalized, chiral product. Such

desymmetrizations have been performed by Scott Miller's group on 1,1-diphenyl compounds bearing symmetrical 4,4'-dihydroxy substituents (Scheme 4.1).[7] By screening libraries of naturally occurring lipases, they were first able to identify an enzyme capable of the desymmetrization of the related acetyl-protected compound, but found that high enantiomeric excess (e.e.) (99%) was only attainable after significant overhydrolysis of the starting material to the diphenolmethane, resulting in a secondary kinetic resolution but limiting the overall yield. They then screened libraries of synthesized peptide catalysts to develop a catalyst capable of starting with the diphenolmethane and acetylating to produce the desired product without relying on a secondary kinetic resolution to afford the desired enantiopurity. After six iterative rounds of peptide development, their final peptide catalyst was able to afford 95% ee and 80% yield.

Scheme 4.1. Summary of the peptide-catalyzed remote desymmetrization of 1.[a]



[a] Adapted from reference 7.

As mentioned above, the chiral active site of FDHs that imparts their regioselectivity could also be conceivably used to impart stereoselectivity to this desymmetrization reaction. Due to our work on engineering RebH variants with expanded substrate scope, we had several variants at our

disposal capable of functionalizing different ranges of substrates that vary significantly in size and shape.[8] In addition, other evolution efforts in our group furnished additional potentially useful variants,[9] and we had also acquired several other natural FDHs as well. We were thus encouraged that we could use our halogenase variants to expand the scope of desymmetrizations available to medicinal chemists.

4.2 RESULTS AND DISCUSSION

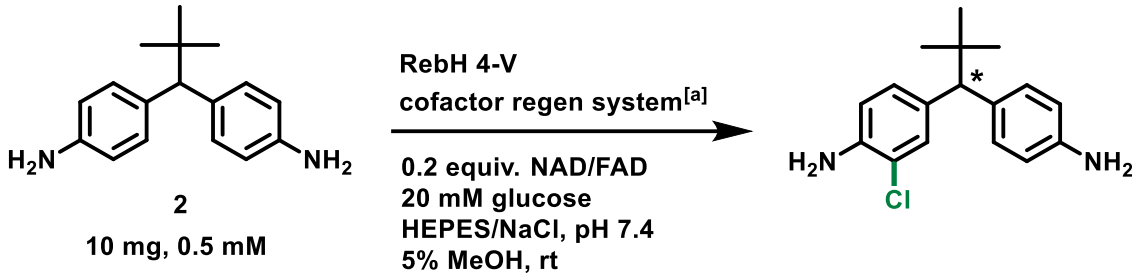
4.2.1 THE DESYMMETRIZATION OF SUBSTRATE 2 WITH REBH 4-V

The Miller group focused on the desymmetrization of the diphenolmethane **1** as shown in Scheme 4.2. We knew from previous studies that RebH variants catalyze the halogenation of anilines of comparable size, such as thenalidine, as described in Section 3.2.4. We therefore decided to begin our studies on desymmetrization with **2**, the dianiline analogue of **1**, which was synthesized as previously reported.[10] Because of the large size of the *tert*-butyl group, we expected that this substrate would be best tolerated by 4-V, the RebH variant we had previously engineered to functionalize yohimbine, carvedilol, and other large, bioactive molecules.[8]

A preparative scale (10 mg) bioconversion of **2** was performed using 5 mol% 4-V (Scheme 4.2), the product of which was isolated, purified, and determined to be monochlorinated *ortho* to one of the amine groups. A small amount of the starting material was also chlorinated with N-chlorosuccinimide (NCS) in order to afford the racemic form of the same product. Using the racemate, a method was developed on the UPLC to resolve the two enantiomers, and the purified product of the bioconversion was then compared to the racemic standard. To our delight, almost exclusive a single enantiomer was produced, providing an enantiomeric ratio (e.r.) for this *tert*-butyl substrate with 4-V of 99:1 (Figure 4.2). This result is especially surprising when one notes

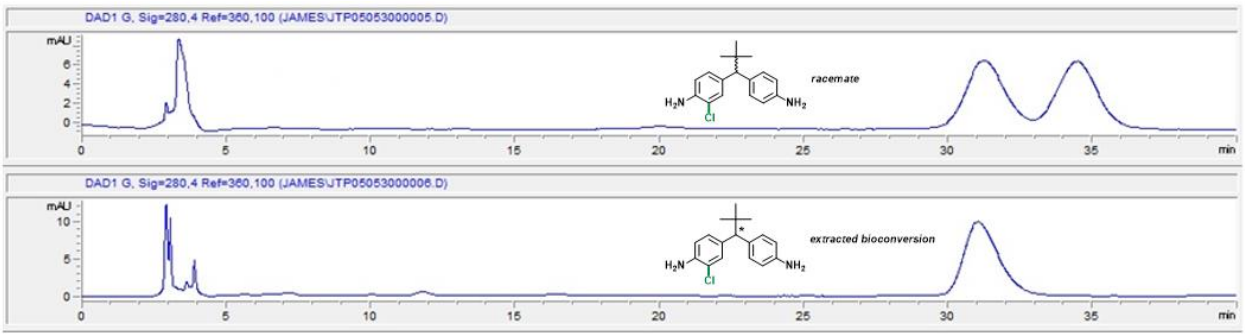
that 4-V was engineered for an entirely different purpose, namely to have an expanded substrate scope; the high enantioselectivity observed in this desymmetrization reaction is entirely fortuitous. We are currently attempting to crystallize the monochlorinated product of this bioconversion in order to determine its absolute chirality.

Scheme 4.2. Reaction conditions for the 4-V-catalyzed halogenation/desymmetrization of 2.



[a] The cofactor regeneration system consisted of 0.5 mol% MBP-RebF and 50 U ml glucose dehydrogenase. NAD = nicotinamide adenine dinucleotide, FAD = flavin adenine dinucleotide, MBP = maltose binding protein.

Figure 4.2. Representative HPLC traces showing enantioselective desymmetrization.[a]



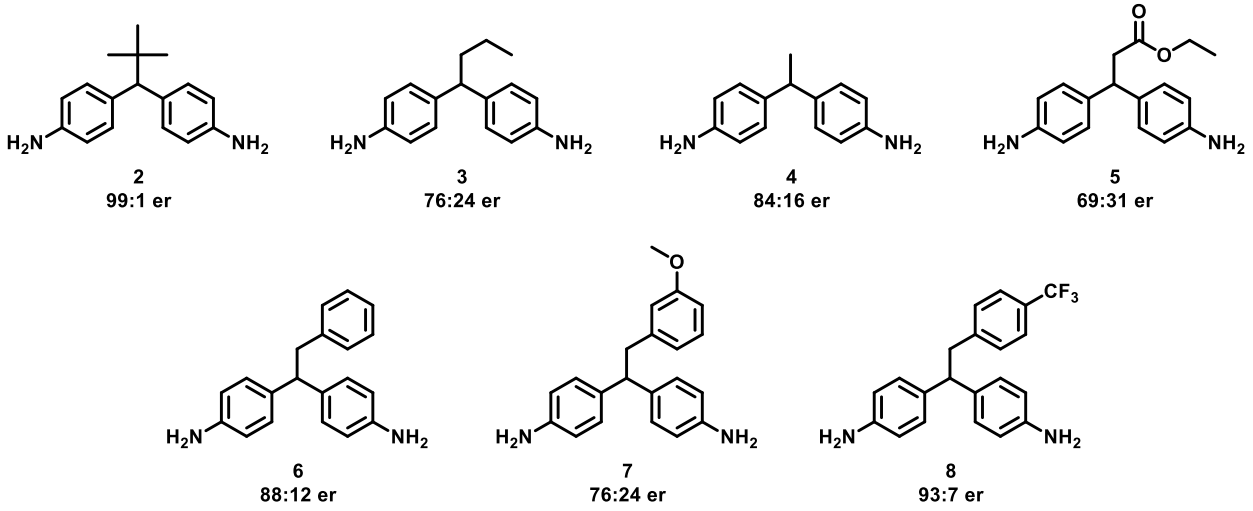
[a] Chromatograms were collected at 280 nm. Top chromatogram shows racemic standard of chlorinated **2** while the bottom chromatogram shows the product of the 4-V-catalyzed bioconversion of **2** (as shown in Scheme 4.2).

4.2.2 SYNTHESIS AND DESYMMETRIZATION OF DIARYLMETHANES WITH 4-V

Having observed unexpectedly high enantioselectivity on the *tert*-butyl substrate with 4-V, we then wished to explore the degree to which we could vary the *tert*-butyl group and still see enantioselectivity. While the large size of the *tert*-butyl group likely hinders substrate access to the enzyme active site, the same large size also likely facilitates resolution of the two potential conformations in which the substrate could bind in the active site, which would then give rise to the two possible enantiomeric products. We therefore expected to see lower e.r. values for analogous substrates with groups smaller than *tert*-butyl.

With the help of Kyle Kunze and Dr. Duo-Sheng Wang, six additional substrates were synthesized as described in Section 4.4.2. These six substrates possessed, in place of the *tert*-butyl group of **2**, an *n*-propyl group, a methyl group, an ethyl ester group, a benzyl group, a *p*-CF_3-benzyl group, or a *m*-OMe-benzyl group (Table 4.1). As was done for **2**, chlorinations with NCS were performed to generate the racemic, monochlorinated products for each of these substrates, from which supercritical fluid chromatography (SFC) methods were developed to resolve the two enantiomers. It was then found that the e.r. could be determined from the crude organic extract of a bioconversion conducted using 1 mg of substrate, so bioconversions on that scale were performed with 5 mol% 4-V for each substrate, and the crude organic extract for each was isolated. These crude extracts were then resolved by SFC to determine the e.r. obtained with 4-V for each substrate (Table 4.1).

Table 4.1. Desymmetrization substrate scope of 4-V.[a]



[a] The e.r. observed for the monochlorinated product (using bioconversion conditions described in Scheme 4.2) is shown beneath each substrate.

With these substrates, a range of e.r. values were observed with 4-V. For the *n*-propyl substrate (**3**), an e.r. of 76:24 was obtained, while for the methyl substrate (**4**), a higher e.r. of 84:16 was seen. As expected, both of these values are lower than that seen for the *tert*-butyl substrate (**2**), but to our surprise, a higher e.r. was seen with the methyl substrate than for the *n*-propyl substrate. We are currently attempting to obtain crystal structures of 4-V with these substrates bound to determine the origin of enantioselectivity in order to elucidate how these unexpected differences may have arisen. For the ethyl ester substrate (**5**), an e.r. of 69:31 was obtained; while we have previously seen that RebH variants can tolerate a wide range of functionalities on their substrates, this was the first desymmetrization we had performed with a non-alkyl group in place of the *tert*-butyl group. The three benzylic substrates, **6**, **7**, and **8**, gave e.r. values of 88:12, 93:7, and 76:24, respectively.

The high enantioselectivities observed thus far could result from differing degrees of contribution of two processes; either from enantiotopic group selection or from a secondary kinetic resolution of the monochlorinated product. We observed formation of a dichlorinated product in these bioconversions and determined that the second chlorination event occurs on the opposite aniline ring than the first chlorination event (data not shown). If dichlorination occurs predominantly on one enantiomer of the monochlorinated product preferentially, the opposite enantiomer would be concurrently enriched. To see if this is occurring, we monitored the change in e.r. over time, as the quantity of dichlorinated product increases, for the *n*-Pr substrate (**3**) (Table 4.2). These results show that despite a substantial increase in the quantity of dichlorinated product over time, the e.r. increases from only 72:28 to 78:22, indicating that a kinetic resolution cannot be the major source of enantioinduction and that the high enantioselectivities observed are the result of enantiotopic group selection by 4-V.

Table 4.2. Change in e.r. over time in the 4-V-catalyzed bioconversion of 3.

Time (hr)	Observed e.r.	Ratio of Mono-/Di-halogenation[a]
2	72:28	13.1:1
4	74:26	8.2:1
6	75:25	6.4:1
8	75:25	3.4:1
24	78:22	1.4:1
48	78:22	1.4:1

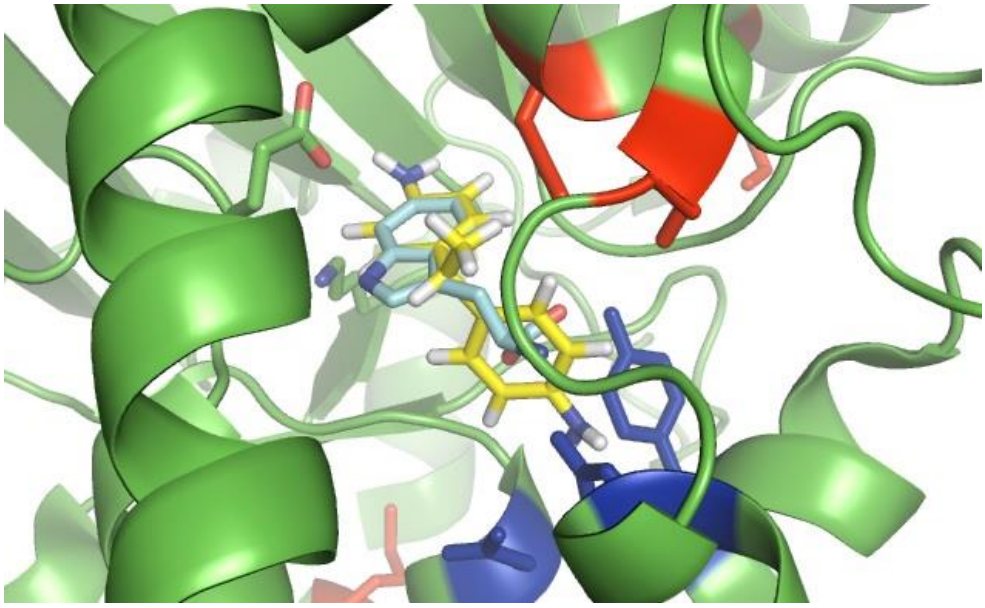
[a] Ratio of mono/di indicates the observed ratio of peak areas (measured by UV absorbance at 254 nm) of the monochlorinated product to the dichlorinated product.

4.2.3 ENGINEERING REBH VARIANTS FOR IMPROVED DESYMMETRIZATION

As stated above, the e.r. values observed with 4-V on these various substrates are all the more remarkable when one takes into account how 4-V was engineered to accommodate entirely different substrates; the fact that it halogenates the dianilinemethane substrates shown in Table 4.1, and furthermore that it does so enantioselectively, is entirely fortuitous. Given that fact, it is very likely that 4-V is far from fully optimized for this purpose. While we had numerous other halogenase variants on hand to explore with these substrates, we also hoped that rational design might be able further optimize activity on these substrates.

Using a reported crystal structure for wild-type RebH,[11] the *t*-Bu substrate **2** was modeled in the active site in a configuration that overlaps the observed position of chlorination with the 7-position of L-tryptophan, the site of chlorination on the native substrate (Figure 4.3). This configuration places the *tert*-butyl group in the vicinity of a loop containing residues N467 and N470, which are mutated from the wild-type sequence in variant 4-V, which is consistent with its elevated activity on this substrate relative to wild-type RebH (*vide infra*). In addition, this configuration places the aniline not undergoing halogenation in the vicinity of three moderate to large residues: N54, Y455, and E461. Previous reports have implicated carboxylic acid containing amino acid residues to be involved in forming a salt bridge with the amino group in the enzyme substrate,[12] while another group installed an amine containing lysine residue to form a salt bridge to a carboxylic acid in their substrate.[13] We hoped to engineer a similar interaction into 4-V by mutating these three large residues in the vicinity of the aniline amino group to eight aspartic or glutamic acid so as to install a nearby carboxylic acid.

Figure 4.3. Model of *t*-Bu substrate 2 in RebH active site.[a]

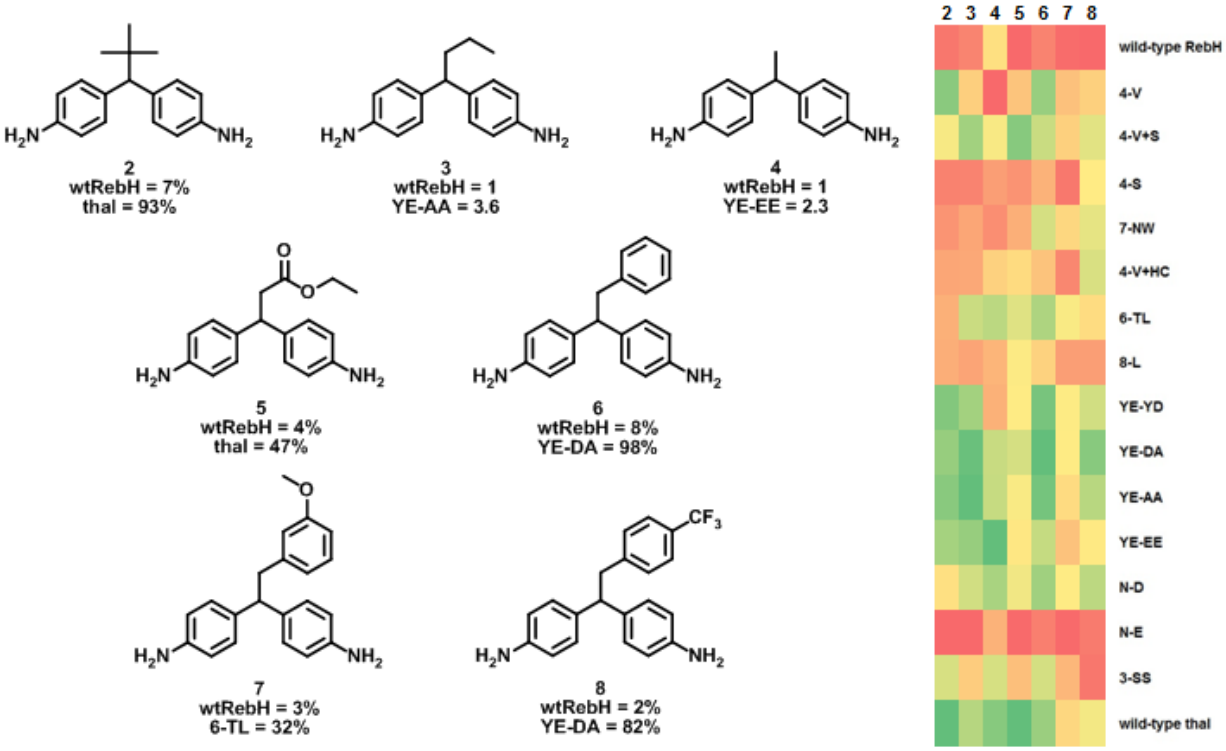


[a] Model created in Pymol using a previously reported crystal structure of wild-type RebH (PDB entry 2OA1).[11] Bound L-tryptophan is shown in light blue with substrate **2** overlaid in yellow. Residues N467 and N470 are shown in red, residues E357 and K79 are explicitly shown in green, and residues N54, Y455, and E461 are shown in blue.

Each residue was mutated to glutamic acid, aspartic acid, alanine, or left unmutated, in combinations such that either only one, two, or all three residues were affected. The racemic chlorinated products of each substrate were used to prepare calibration curves, and then bioconversions were set up using each of the engineered salt bridge variants on each desymmetrization substrate. In addition, numerous other halogenase variants were also screened: other variants engineered to have expanded substrate scope, variants engineered to have altered regioselectivity,[9b] 4-V engineered to contain some of the mutations that resulted in the altered regioselectivity of the aforementioned described variants, as well as wild-type RebH and other naturally occurring FDHs. The mutations present in each of these variants are described in Table

4.4. The results from all of these bioconversions were then fitted to the calibration curves to determine the percent conversion observed with each halogenase variant on each desymmetrization substrate.

Table 4.3. Halogenase variant conversions of desymmetrization substrates.[a]



[a] The percent conversion observed with wild-type RebH and the halogenase variant that showed the highest conversion on that particular substrate is given underneath each substrate (for substrates **3** and **4** conversions relative to wild-type RebH are given rather than percent conversions). For the heat map shown at the right, red represents a lower conversion, while green represents a higher conversion.

From these tests, we found that the halogenase variants gave a wide range of conversions, with the results for each substrate differing greatly. In Table 4.3, the variant that afforded the

highest conversion on each substrate is shown, while each conversion is represented on the heat map shown at right; quantitative charts showing the conversions of each variant on each substrate can be found in Section 4.4.4 (Figures 4.5-4.11). From the heat map, some general trends can be observed. Wild-type RebH shows generally low activity on each substrate, especially when compared with wild-type thal. The engineered variant 4-V shows improved activity relative to wild-type RebH on every substrate except **4**, while YE-DA shows improved activity on all of the substrates, even when compared with 4-V. When substrates **7** and **8** are compared, it appears that *meta*-substitution on the benzyl group is much more disfavorable than *para*-substitution is; a direct comparison of the effects of substitution at these two positions on the benzyl group is in progress.

For the *tert*-butyl substrate **2**, the variant that provided the highest observed conversion remarkably was thal, a tryptophan 6-halogenase from the biosynthetic pathway for thienodolin in *Streptomyces albogriseolus*,[14] which gave over 90% conversion (Figure 4.5). 4-V provided nearly fifteen-fold higher conversion of this substrate than did wild-type RebH, and two of the salt bridge variants gave slightly higher conversions than 4-V. For the *n*-propyl (**3**) and methyl (**4**) substrates, issues with the calibration curves prevent actual conversions from being obtained, but relative conversions between the variants can still be determined and are thus reported relative to the observed conversion with wild-type RebH. With the methyl substrate, 4-V was found to actually give decreased conversion relative to wild-type RebH, which may be consistent with the observed preference 4-V has for larger substrates relative to wild-type RebH, while **4** is the smallest substrate tested in this set. However, the salt bridge variant of 4-V in which Y455 is mutated to glutamic acid, while N54 and E461 are left unchanged, gave over twofold higher conversion than did wild-type RebH and nearly six-fold higher conversion than did 4-V (Figure 4.7). Thal again gave high conversion, as did several other salt bridge variants and the altered regioselectivity variant 6-TL.

The *n*-propyl substrate also showed improvements with the salt bridge variants, with some furnishing nearly fourfold higher conversion than wild-type RebH did (Figure 4.6).

With the ethyl ester substrate (**5**), less than 4% conversion was seen with wild-type RebH; however, two other variants stood out from amongst the rest, with 4-V+S (where the N470S mutation from the altered regioselectivity variants was engineered into 4-V) giving over tenfold higher conversion, and Thal again performing best, giving almost thirteen times the conversion than did wild-type RebH at 47% conversion (Figure 4.8). The salt bridge variant in which Y455 was mutated to aspartic acid, E461 was mutated to alanine, and N54 was left unchanged gave the best results for the benzyl substrate (**6**) at 98% conversion (Figure 4.9), again nearly thirteen times higher than observed with wild-type RebH. This same variant also performed best with the *p*-CF_3-benzyl substrate (**8**), furnishing 82% conversion, an almost forty-fold increase over wild-type RebH (Figure 4.11). However, for the *m*-OMe-benzyl substrate (**7**), the altered regioselectivity variant 6-TL gave the highest observed conversion at 32%, over tenfold higher than wild-type RebH (Figure 4.10).

While different halogenase variants clearly have a significant effect on the conversion observed for each of these substrates, this increase in conversion may come at the expense of enantioselectivity; the enzyme active site may simply be more readily accepting these substrates independent of their confirmation, resulting in higher rates of product of both enantiomers of product. To determine if the high enantioselectivities observed with 4-V are preserved with the alternative halogenase variants that showed higher conversion, each of these variants were used to perform 0.5 mg scale bioconversions. These bioconversions were worked up, the products were examined by SFC as described for the bioconversions with 4-V, and the e.r. values were determined for each product (Table 4.4).

Table 4.4. Enantiomeric ratios afforded by halogenase variants.[a]

	Halogenase Variant										
Substrate	4-V[8]	YE-DA[b]	YE-AA[c]	YE-YD[d]	YE-EE[e]	N-D[f]	thal[14]	6-TL[g]	7-NW[h]	4-V+S[i]	3-SS[8]
***t*-Bu (2)**	**99:1 (12.1)[j]**	99:1 (11.4)	95:5 (12.2)	93:7 (12.5)	96:4 (10.7)	87:13 (5.6)	-	-	-	-	90:10 (8.1)
***n*-Pr (3)**	76:24 (2.0)	66:34 (3.6)	70:30 (3.6)	-	**78:22 (3.2)**	-	64:36 (3.0)	71:29 (2.8)	-	76:24 (3.1)	-
Me (4)	84:16 (0.4)	85:15 (1.5)	82:18 (1.5)	-	90:10 (2.3)	76:24 (1.7)	-	85:15 (1.6)	-	-	**95:5 (1.4)**
Ester (5)	69:31 (3.3)	-	51:49 (4.6)	68:32 (4.3)	-	47:53 (5.1)	-	**78:22 (6.1)**	-	-	65:35 (3.2)
Benzyl (6)	88:12 (9.8)	88:12 (12.9)	**95:5 (11.8)**	94:6 (11.7)	-	87:13 (9.3)	-	53:47 (8.5)	-	-	83:17 (6.3)
***m*-OMe-Bz (7)**	76:24 (7.0)	75:25 (10.1)	-	**87:13 (10.2)**	-	87:13 (10.2)	73:27 (8.5)	66:34 (10.9)	-	-	-
***p*-CF_3-Bz (8)**	93:7 (10.8)	**95:5 (37.4)**	92:8 (27.8)	86:14 (22.6)	-	92:8 (26.8)	-	-	89:11 (18.4)	-	-

[a] Detailed reaction conditions are those described in Section 4.4.2 using 0.5 mg of substrate. The highest observed enantiomeric ratio for each substrate is shown in bold. [b] YE-DA = 4-V with Y455D and E461A. [c] YE-AA = 4-V with Y455A and E461A. [d] YE-YD = 4-V with E461D. [e] YE-EE = 4-V with Y455E. [f] N-D = 4-V with N54D. [g] 6-TL = wild-type RebH with N470S, S448P, Q494R, L380F, R509Q, Y455W, S110P, F111L, S130L, N166S, I52T, and F465L.9[b] [h] 7-NW = 6-TL with T52N and L465W. [i] 4-V+S = 4-V with N470S. [j] Term in parentheses refers to the ratio of the conversion observed to that observed with wild-type RebH.

Despite significant increases in conversion, many of these variants did not show significantly reduced enantioselectivities, with some variants even providing higher conversions *and* enantioselectivies. For example, variant YE-DA provided a nearly fourfold increase in conversion on **8** relative to 4-V (and nearly fortyfold higher than wild-type RebH), but still showed an increase in enantioselectivity to afford an e.r. of 95:5. Even more remarkable is the activity of variant 3-SS on the methyl substrate **4**; 3-SS gave nearly fourfold higher conversion than 4-V did

on this substrate and also showed a significant increase in enantioselectivity to afford an e.r. of 95:5. The enzyme active site of 3-SS is somehow able to discriminate between the different possible positions of this very small methyl group to such a degree that 95% of the product is a single enantiomer. These results clearly illustrate the potential to further engineer these halogenases to provide even higher yields without compromising, and even further increasing, the remarkably high enantioselectivities they afford. Further illustrated is the importance of screening a panel of halogenase variants, as different variants gave the best observed conversion on different substrates, while separate variants often gave the best observed enantioselectivies.

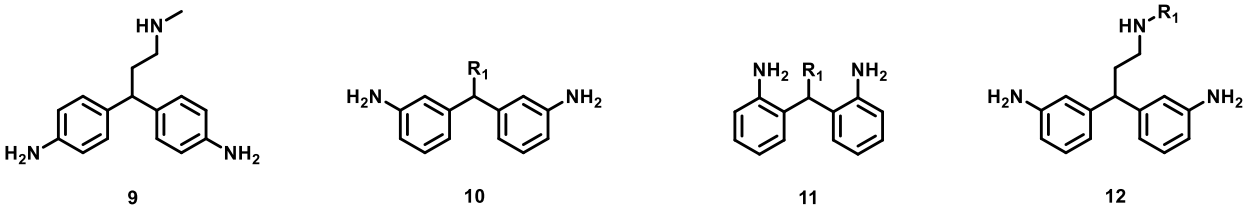
4.3 CONCLUSIONS

Using 4-V, a variant of RebH that was engineered to functionalize large, bioactive indoles and carbazoles, we were able to perform a remote desymmetrization of a dianilinemethane substrate stereoselectively, such that the two enantiomers were produced in a ratio of 99:1. A variety of related substrates were synthesized and also accepted by 4-V, furnishing a range of observed ers. Variants of 4-V were thus engineered with the intention of installing a salt bridge between the amino group of the substrates and installed carboxylic acid containing residues, and with every substrate tested, at least one of these engineered salt bridge variants gave enhanced conversion relative to 4-V. Numerous other halogenase variants also gave an interesting range of conversions of these substrates, with up to 98% conversion obtained at 5 mol% enzyme loading. The enantioselectivities of these other halogenase variants are currently being explored, as are additional substrates; preliminary results already show that alternative halogenase variants have the potential to not only provide increased conversion without compromising enantioselectivity, but can provide increased enantioselectivity *in addition to* increased conversion. Furthermore, we hope to develop conditions to crystallize these halogenase variants with these desymmetrization

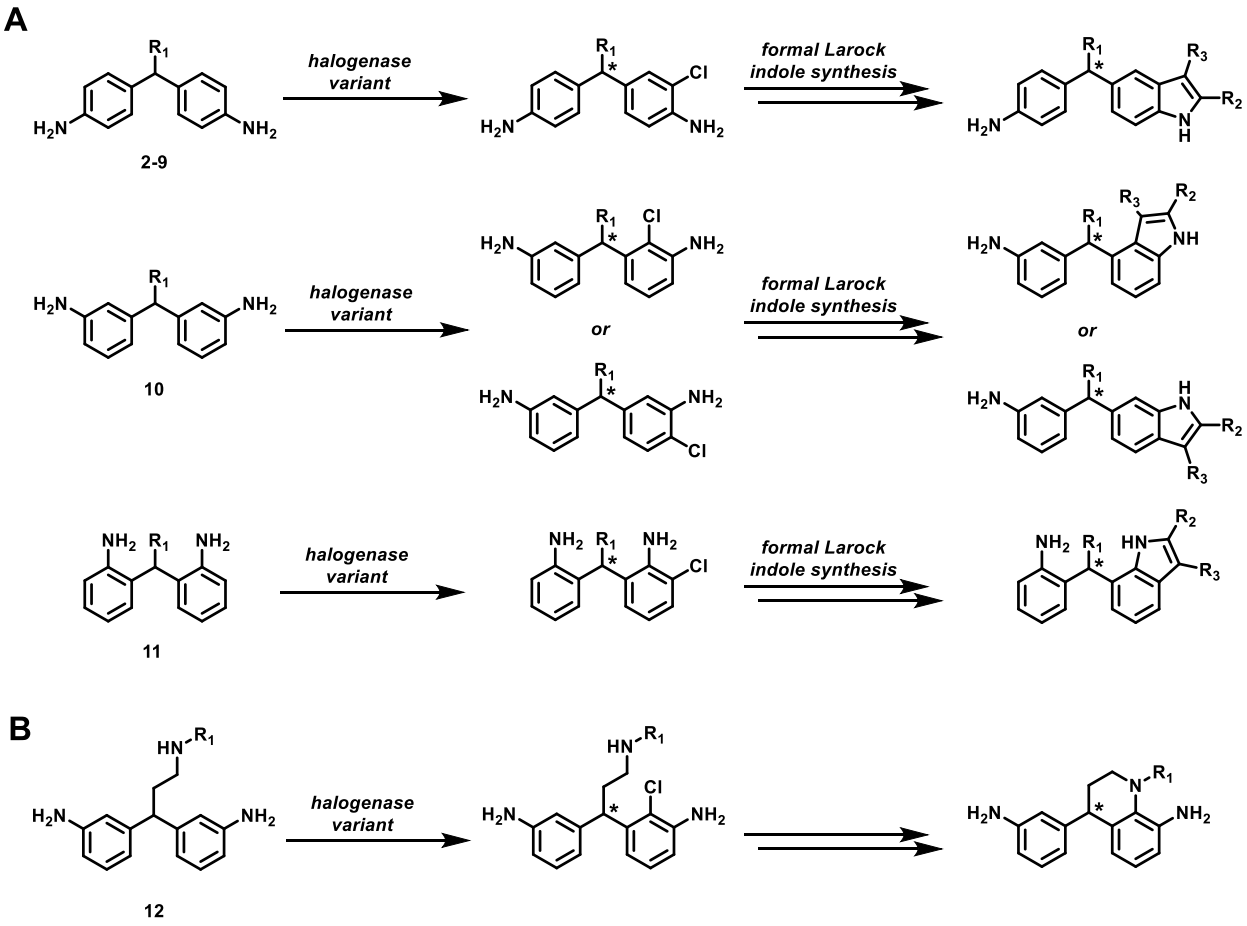
substrates bound to elucidate the origins of the enantioselectivity and inform future engineering efforts. These results already indicate that halogenases can play a useful role in the asymmetric syntheses of a range of useful medicinal scaffolds.

In addition, we are investigating further compounds as potential substrates for asymmetric halogenation (Figure 4.4). Some of these substrates will include thus far unexplored functionalities on the non-aniline group (e.g. **9**), different positions of the aniline amino groups (**10** and **11**), or a combination of these two properties (e.g. **12**). Using the principle of chemoenzymatic functionalization described in Section 3.2.4, the products of these asymmetric halogenations can be further employed as intermediates in the synthesis of more complex, pharmaceutically important scaffolds (Scheme 4.3). The halogenated products can be alkynylated via a Sonogashira coupling and then cyclized to form an indole to accomplish a formal Larock indole synthesis (Scheme 4.3A).[15] If substrates with the aniline amino groups at different positions (e.g. Figure 4.4, **10** and **11**) are successfully asymmetrically halogenated, these products will then lead to differently substituted indoles. In this way, a route to the preparation of chiral 4-, 5-, 6-, or 7-substituted indoles may potentially be accessed. Furthermore, if substrates analogous to **12** (Figure 4.4) are also asymmetrically halogenated, subsequent functionalization will allow access to chiral, substituted tetrahydroquinolines (Scheme 4.3B). The tetrahydroquinoline scaffold is well represented in bioactive products and has been an important target of synthetic efforts in recent years.[16] We thus anticipate that future research efforts will further expand the potential applications of these halogenases in the preparation of a range of useful, bioactive compounds that are otherwise difficult to prepare.

Figure 4.4. Potential future substrates for asymmetric halogenation.



Scheme 4.3. Potential chemoenzymatic functionalizations of halogenated products.



4.4 EXPERIMENTAL

4.4.1 GENERAL EXPERIMENTAL PROCEDURES

Materials:

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Compounds **2**,[10] **3**,[17] **4**,[18] and **6**[19] were synthesized according to previous reports. Deuterated solvents were obtained from Cambridge Isotope labs. Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Oligonucleotides were purchased from Integrated DNA Technologies (San Diego, CA). BL21(DE3) *E. coli* cells were purchased from Invitrogen (Carlsbad, CA). T7 DNA ligase, 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. All genes were confirmed by sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser using method Ec2. Ninitrilotriacetic acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were following when using both products (for Ni-NTA resin, 5 mL resin were used, with buffers delivered by a peristaltic pump at a rate of 1 mL/min, in a 4 °C cold cabinet). Amicon® 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. Biotage reverse phase columns (SNAP KP-C18-HS) were purchased from Biotage. HPLC analyses were performed using HPLC grade acetonitrile (Fisher

Scientific), 18 MΩ water from a Milli-Q purification system (model No. QGARD00D2), and trifluoroacetic acid (Oakwood Chemicals). SFC analyses were performed using bone-dry grade CO_2 (Cylinder Gas Operations, University of Chicago, Chicago, IL), HPLC grade methanol (Fisher Scientific), and isobutanol (Sigma-Aldrich). Glucose dehydrogenase (GDH, product No. GDH-105), and NAD (product No. NAD-004626) were purchased from Codexis (Redwood City, CA). FAD (product No. 00151) was purchased from Chem-Impex International (Wood Dale, IL). AeraSeal film was purchased from Research Products International (product No. 202504).

General Procedures:

Standard molecular cloning procedures were followed,[20] and the same PCR conditions were used as previously reported.8 Point mutations were inserted via sequence-overlap extension PCR.[21] Reactions were monitored using UPLC (Agilent 1200 UPLC with a 1290 DAD detector (G4212A) with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; C18 4.6 x 50 mm column, 3.5 μM particle size; and C18 2.1 x 50 mm column, 1.8 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN). Reverse phase preparative chromatography was carried out using a Biotage Isolera One. 1H spectra were recorded at 500 MHz on a Bruker DMX-500 or DRX-500 spectrometer at room temperature, and chemical shifts are reported relative to residual solvent peaks with coupling constants reported in Hz.[22] Mass spectra were obtained from the University of Chicago mass spectrometry facility using an Agilent Technologies 6224 TOF LC/MS.

4.4.2 SPECIFIC EXPERIMENTAL PROCEDURES

Enzyme purification: The MBP-RebF and RebH variants used for analytical and 10 mg bioconversions was grown, expressed, lysed and purified according to a previous report.[8] all halogenase variants, an overnight starter culture was used to inoculate either 50 mL or 750 mL TB (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol for the pGro7 plasmid) in a 250 mL Erlenmeyer flask or a 2.8 L Fernbach flask, respectively. Following growth at 37 °C, 250 rpm, until OD_{600} = 0.6-0.8, enzyme expression was induced with IPTG and arabinose to final concentrations of 100 μM and 2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cultures were harvested by centrifugation and stored at -80 °C until use. Cell pellets were thawed, suspended in 30 mL 25 mM HEPES (pH 7.4) and lysed by sonication while kept on ice (Qsonica S-4000 with a 0.5" horn; 5 x 1 min with 1 min rests, 20 % duty cycle delivering 40-50 W). After clarification by centrifugation, MBP-RebF and RebH variants were purified by Ni-NTA affinity chromatography and exchanged into a buffer of 25 mM HEPES (pH 7.4) and 10 % glycerol. Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -20 °C until use.

Chlorination to Produce Racemic Standards: To the starting material (**2-8**, 0.05-0.1 mmol) in 10 mL acetonitrile in a 25 mL round-bottomed flask equipped with a reflux condenser was added 1 equivalent of *N*-chlorosuccinimide. The reaction was then stirred at 80 °C for 48 hours. The solvent was then removed and the crude material was purified by reverse-phase chromatography using a Biotage.

General Procedure for Analytical Bioconversions[8]: Substrate (37.5 nmol) was added to a 1.5 mL Eppendorf tube as a 10 mM solution in MeOH. Solutions of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and glucose dehydrogenase (9 U/mL final concentration GDH) were added to the reaction. This was diluted such that the final reaction volume was 75 μL with HEPES buffer, and halogenase variant (0.005-0.05 equiv., 2.5-25 μM final concentration) and MBP-RebF (0.005 equiv., 2.5 μM final concentration) were added as solutions of HEPES/glycerol buffer (25 mM HEPES, pH 7.4, 10% glycerol v/v). The reaction was initiated with a solution of 1 M glucose (40 equiv., 20 mM final concentration), the tube was closed, and incubated at 25 °C at 600 rpm. Reactions were quenched by addition of a reaction volume of MeOH after 16 hours. These reactions were analyzed by UPLC (Agilent 1200 UPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN). The following method was used for all substrates: 0-10 min, B = 15%; 10-20 min, B = 15-100%; 20-24 min, B = 100%.

General Procedure for 10 mg Bioconversions (without subsequent cross-coupling)[8]: Substrate (0.1 mg-10.0 mg) was added to a 20 mL scintillation vial (for 0.1-1 mg bioconversions) or a crystallization dish (100 x 50 mm) (for 10 mg bioconversions) as a solution in MeOH. Solutions of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and a glucose dehydrogenase (9 U/mL final concentration GDH) were added to the reaction. This was diluted to the appropriate volume with HEPES buffer, and halogenase variant (0.01-0.05 equiv., 5-25 μM final concentration) and MBP-RebF (0.005 equiv., 2.5 μM final concentration) were added as solutions of HEPES/glycerol buffer

(25 mM HEPES, pH 7.4, 10% glycerol v/v). The reaction was initiated with a solution of 1 M glucose (40 equiv., 20 mM final concentration), sealed with an AeraSeal film, and left on the benchtop at room temperature without shaking. These reactions were analyzed by UPLC (Agilent 1200 UPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 µM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN). The following method was used for all substrates: 0-10 min, B = 15%; 10-20 min, B = 15-100%; 20-24 min, B = 100%. Reactions were monitored by UPLC and in most cases stopped after 16 hours. The bioconversions were quenched with HCl (5 M, until pH<2) and saturated with NaCl. For 0.1-1 mg bioconversions, precipitated protein was spun down by centrifugation (15,000 g for 10 min.), and the supernatant was decanted and brought to pH>12 through addition of NaOH (5M). For 10 mg bioconversions, precipitated protein was filtered out through a pad of Celite, and the filtrate was brought to pH>12 through addition of NaOH (5M). The filtrate was extracted into CH_2Cl_2. The crude material was either analyzed directedly or purified by reverse phase chromatography (Biotage SNAP-KP-C18-HS, gradient from pure H_2O to 40% CH_3CN/H_2O).

SFC Analysis: To determine the e.r. values of the products, analyses were carried out using an Agilent 1200 UPLC with a 1260 Infinity SFC Control Module (G4301A) supplying supercritical CO_2 to channel A with a 1260 DAD VL+ detector (G1315C). The crude product of a typically 0.5 mg bioconversion was resuspended in 200 µL of MeOH, of which 15 µL were injected into a 5 µL loop. A Daicel Chiral Technologies (West Chester, PA, USA) IC-3 column was used with an isocratic method of 65% CO_2, 35% MeOH/25 mM isobutanol, 3.0 mL/min, for 10 minutes for all substrates except the *t*-Bu substrate, **2**, which was analyzed using an Agilent 1200 UPLC with a 1290 DAD detector (G4212A) with a Phenomenex Lux 3u Cellulose-1 250 x 4.6 mm column with

an isocratic method of 70% H_2O/0.1% triethylamine, 30% ACN/0.1% triethylamine, 1.0 mL/min, for 40 minutes. E.r. values were determined by taking the ratios of the areas of the UV peaks (at 254 nm) that match the retention times observed for identical analyses of the racemic products.

4.4.3 DETAILED SYNTHESES AND CHARACTERIZATION

Note: characterization of these compounds is currently ongoing, with some compounds being re-synthesized in order to obtain higher quality data.

***t*-Bu dianilinemethane (2):** This compound was synthesized as described in a previous report.[10] ^{1}H NMR (500 MHz, MeOD) δ 7.19 (d, *J* = 8.4 Hz, 4H), 6.66 (d, *J* = 8.5 Hz, 4H), 3.51 (s, 1H), 0.98 (s, 9H). HRMS (ESI-TOF) calc'd for $C_{17}H_{23}N_2$ [M + H]+: 255.1861, found: 255.1182.

***n*-Pr dianilinemethane (3):** This compound was synthesized as described in a previous report.[17] ^{1}H NMR (500 MHz, MeOD) δ 6.98 (d, *J* = 8.3 Hz, 4H), 6.66 (d, *J* = 8.4 Hz, 4H), 3.68 (t, *J* = 7.8 Hz, 1H), 1.91 (dd, *J* = 15.4, 7.8 Hz, 2H), 1.25 (dd, *J* = 15.1, 7.5 Hz, 2H), 0.92 (t, *J* = 7.4 Hz, 3H). HRMS (ESI-TOF) calc'd for $C_{16}H_{21}N_2$ [M + H]+: 241.1705, found: 241.1023.

Methyl dianilinemethane (4): This compound was synthesized as described in a previous report.[18] ^{1}H NMR (500 MHz, MeOD) δ 6.96 (d, *J* = 8.4 Hz, 5H), 6.67 (d, *J* = 8.3 Hz, 5H), 3.91 (q, *J* = 7.2 Hz, 1H), 1.51 (d, *J* = 7.2 Hz, 3H). HRMS (ESI-TOF) calc'd for $C_{14}H_{17}N_2$ [M + H]+: 213.1392, found: 213.0716.

Ethyl ester dianilinemethane (5): To a stirred solution of 0.5 mmol (129.1 mg) of bis(4-nitrophenyl)methane in 4 mL THF at 0 °C were added 1 equiv. (0.5 mmol, 56.1 mg) KO*t*Bu and 1.1 equiv. (0.55 mmol, 145.4 mg) 18-C-6. This solution was stirred for 10 minutes before it was cooled to -78 °C prior to the addition of 1 equiv. (0.5 mmol, 83.5 mg) of ethyl 2-bromoacetate in 1 mL THF, after which the reaction was allowed to warm to room temperature. After purification, this furnished 141 mg (82% yield) of ethyl 3,3-bis(4-nitrophenyl)propanoate. The entirety of this was then stirred at room temperature in 4 mL MeOH, to which 14 mg of Pd/C (10%) were added and allowed to react under a H_2 atmosphere in an autoclave. After purification, this afforded 116 mg (quantitative yield) of **5**. ^{1}H NMR (500 MHz, MeOD) δ 7.01 – 6.94 (d, 4H), 6.70 – 6.63 (d, 4H), 4.26 (t, *J* = 8.2 Hz, 1H), 4.00 (q, *J* = 7.1 Hz, 2H), 2.93 (d, *J* = 8.2 Hz, 2H), 1.11 (t, *J* = 7.1 Hz, 3H). HRMS (ESI-TOF) calc'd for $C_{11}H_{14}NO_2$ [M – C_6H_7N (elimination to lose aniline to form α,β-unsaturated system) + H]+: 192.l025, found: 192.0402.

Benzyl dianilinemethane (6): This compound was synthesized as described in a previous report.[19] ^{1}H NMR (500 MHz, MeOD) δ 7.15 – 7.00 (m, 5H), 6.97 (d, *J* = 8.3 Hz, 4H), 6.63 (d, *J* = 8.4 Hz, 4H), 4.02 (t, *J* = 7.9 Hz, 1H), 3.23 (d, *J* = 7.9 Hz, 2H). HRMS (ESI-TOF) calc'd for $C_{20}H_{21}N_2$ [M + H]+: 289.1705, found: 289.0049.

***m*-Methoxybenzyl dianilinemethane (7):** To a stirred solution of 0.5 mmol (129.1 mg) of bis(4-nitrophenyl)methane in 4 mL THF at 0 °C were added 1 equiv. (0.5 mmol, 56.1 mg) KO*t*Bu and 1.1 equiv. (0.55 mmol, 145.4 mg) 18-C-6. This solution was stirred for 10 minutes before it was cooled to -78 °C prior to the addition of 1 equiv. (0.5 mmol, 100.5 mg) of *m*-MeO-benzyl bromide

in 1 mL THF, after which the reaction was allowed to warm to room temperature. After purification, this furnished 154 mg (81% yield) of 4 4,4'-(2-(3-methoxyphenyl)ethane-1,1-diyl)bis(nitrobenzene). The entirety of this was then stirred at room temperature in 2 mL MeOH, to which 20 mg of Pd/C (10%) were added and allowed to react under a H_2 atmosphere in a balloon. This afforded 128 mg (99% yield) of **7**. ^{1}H NMR (500 MHz, MeOD) δ 7.04 (t, *J* = 7.9 Hz, 1H), 6.97 (d, *J* = 8.2 Hz, 4H), 6.63 (m, 6H), 6.54 (s, 1H), 3.99 (t, *J* = 7.8 Hz, 1H), 3.63 (s, 3H), 3.20 (d, *J* = 7.9 Hz, 2H). HRMS (ESI-TOF) calc'd for $C_{21}H_{23}N_2O$ [M + H]+: 319.1811, found: 319.0985.

***p*-Trifluoromethylbenzyl dianilinemethane (8):** To a stirred solution of 0.5 mmol (129.1 mg) of bis(4-nitrophenyl)methane in 4 mL THF at 0 °C were added 1 equiv. (0.5 mmol, 56.1 mg) KO*t*Bu and 1.1 equiv. (0.55 mmol, 145.4 mg) 18-C-6. This solution was stirred for 10 minutes before it was cooled to -78 °C prior to the addition of 1 equiv. (0.5 mmol, 119.5 mg) of *p*-CF_3-benzyl bromide in 1 mL THF, after which the reaction was allowed to warm to room temperature. After purification, this furnished 190 mg (91% yield) of 4,4'-(2-(4-(trifluoromethyl)phenyl)ethane-1,1-diyl)bis(nitrobenzene). The entirety of this was then stirred at room temperature in 2 mL MeOH, to which 20 mg of Pd/C (10%) were added and allowed to react under a H_2 atmosphere in a balloon. This afforded 137 mg (85% yield) of **8**. ^{1}H NMR (500 MHz, MeOD) δ 7.40 (d, *J* = 8.0 Hz, 2H), 7.17 (d, *J* = 7.9 Hz, 2H), 6.97 (d, *J* = 8.3 Hz, 4H), 6.64 (d, *J* = 8.3 Hz, 4H), 4.03 (t, *J* = 7.9 Hz, 1H), 3.29 (d, *J* = 8.0 Hz, 2H). HRMS (ESI-TOF) calc'd for $C_{21}H_{20}F_3N_2$ [M + H]+: 357.1579, found: 357.0754.

Racemic monochlorinated *t*-Bu dianilinemethane (2a): This compound was prepared from 0.2 mmol (51 mg) of **2** as described in Section 4.4.2. After reaction completion, a significant amount of dichlorinated material was observed, which was removed by normal phase flash chromatography (eluting with 20:1 DCM:MeOH) before the reaction mixture was purified as described in Section 4.4.2 to afford **2a** in 14% yield (11.1 mg of **2a**•TFA, 0.028 mmol). ^{1}H NMR (500 MHz, MeOD) δ 7.59 (d, *J* = 8.5 Hz, 2H), 7.28 (m, 3H), 7.18 (dd, *J* = 8.3, 2.0 Hz, 1H), 6.80 (d, *J* = 8.3 Hz, 1H), 3.72 (s, 1H), 1.01 (s, 9H). HRMS (ESI-TOF) calc'd for $C_{17}H_{22}N_2Cl$ [M + H]+: 289.1472 and 291.1442, found: 289.0621 and 291.0592.

Racemic monochlorinated *n*-Pr dianilinemethane (3a): This compound was prepared from 0.091 mmol (22 mg) of **3** as described in Section 4.4.2. After reaction completion, the reaction mixture was purified as described in Section 4.4.2 to afford **3a** (this compound was prepared in a collaboration and a yield was not reported). ^{1}H NMR (500 MHz, MeOD) δ 7.37 (d, *J* = 8.4 Hz, 2H), 7.22 (d, *J* = 8.4 Hz, 2H), 7.08 (d, *J* = 1.8 Hz, 1H), 6.97 (dd, *J* = 8.3, 1.9 Hz, 1H), 6.78 (d, *J* = 8.2 Hz, 1H), 3.85 (t, *J* = 7.9 Hz, 1H), 1.98 (td, *J* = 9.0, 2.7 Hz, 2H), 1.31 – 1.20 (m, 2H), 0.95 (t, *J* = 7.4 Hz, 3H). HRMS (ESI-TOF) calc'd for $C_{16}H_{20}N_2Cl$ [M + H]+: 275.1315 and 277.1286, found: 275.0503 and 277.0475.

Racemic monochlorinated methyl dianilinemethane (4a): This compound was prepared from 0.11 mmol (23.4 mg) of **4** as described in Section 4.4.2. After reaction completion, the reaction mixture was purified as described in Section 4.4.2 to afford **4a** (this compound was prepared in a collaboration and a yield was not reported). ^{1}H NMR (500 MHz, MeOD) δ 7.36 (d, *J* = 8.4 Hz,

2H), 7.25 (d, *J* = 8.4 Hz, 2H), 7.06 (d, *J* = 1.8 Hz, 1H), 6.95 (dd, *J* = 8.3, 1.9 Hz, 1H), 6.79 (d, *J* = 8.2 Hz, 1H), 4.09 (dd, *J* = 14.4, 7.2 Hz, 1H), 1.59 (d, *J* = 7.2 Hz, 3H). HRMS (ESI-TOF) calc'd for $C_{14}H_{16}N_2Cl$ [M + H]+: 247.1002 and 249.0973, found: 247.0222 and 249.0189.

Racemic monochlorinated ethyl ester dianilinemethane (5a): This compound was prepared from 0.035 mmol (10 mg) of **5** as described in Section 4.4.2. After reaction completion, the reaction mixture was purified as described in Section 4.4.2 to afford **5a** (this compound was prepared in a collaboration and a yield was not reported). ^{1}H NMR (500 MHz, MeOD) δ 7.06 (d, *J* = 2.0 Hz, 1H), 7.00 (d, *J* = 8.4 Hz, 2H), 6.93 (d, *J* = 8.2 Hz, 1H), 6.76 (d, *J* = 8.3 Hz, 1H), 6.70 (d, *J* = 8.3 Hz, 2H), 4.26 (t, *J* = 8.1 Hz, 1H), 4.02 (q, *J* = 7.2 Hz, 2H), 2.94 (d, *J* = 8.2 Hz, 2H), 1.13 (t, *J* = 7.1 Hz, 3H). HRMS (ESI-TOF) calc'd for $C_{11}H_{13}NO_2Cl$ [M – C_6H_7N (elimination to lose aniline to form α,β-unsaturated system) + H]+: 226.0635 and 228.0605, found: 225.9876 and 227.9844.

Racemic monochlorinated benzyl dianilinemethane (6a): This compound was prepared from 0.087 mmol (25 mg) of **6** as described in Section 4.4.2. After reaction completion, the reaction mixture was purified as described in Section 4.4.2 to afford **6a** (this compound was prepared in a collaboration and a yield was not reported). ^{1}H NMR (500 MHz, MeOD) δ 7.22 (d, *J* = 8.3 Hz, 2H), 7.15 (t, *J* = 7.3 Hz, 2H), 7.12 – 7.03 (m, 4H), 7.03 – 6.97 (m, 2H), 6.95 (dd, *J* = 8.3, 1.9 Hz, 1H), 6.74 (d, *J* = 8.3 Hz, 1H), 4.15 (t, *J* = 7.9 Hz, 1H), 3.28 (d, *J* = 8.0 Hz, 2H). HRMS (ESI-TOF) calc'd for $C_{20}H_{20}N_2Cl$ [M + H]+: 323.1315 and 325.1286, found: 323.0427 and 325.0394.

Racemic monochlorinated *m*-methoxybenzyl dianilinemethane (7a): This compound was prepared from 0.072 mmol (23 mg) of **7** as described in Section 4.4.2. After reaction completion, the reaction mixture was purified as described in Section 4.4.2 to afford **7a** (this compound was prepared in a collaboration and a yield was not reported). ^{1}H NMR (500 MHz, MeOD) δ 7.30 (d, *J* = 8.4 Hz, 2H), 7.14 – 7.02 (m, 4H), 6.97 (dd, *J* = 8.3, 1.9 Hz, 1H), 6.76 (d, *J* = 8.2 Hz, 1H), 6.67 (dd, *J* = 13.3, 4.9 Hz, 2H), 6.60 (s, 1H), 4.17 (t, *J* = 8.0 Hz, 1H), 3.68 (s, 3H), 3.27 (dd, *J* = 8.0, 2.8 Hz, 2H). HRMS (ESI-TOF) calc'd for $C_{21}H_{22}N_2OCl$ [M + H]+: 353.1421 and 325.1391, found: 353.0489 and 355.0458.

Racemic monochlorinated *p*-trifluoromethylbenzyl dianilinemethane (8a): This compound was prepared from 0.007 mmol (25 mg) of **8** as described in Section 4.4.2. After reaction completion, the reaction mixture was purified as described in Section 4.4.2 to afford **8a** (this compound was prepared in a collaboration and a yield was not reported). ^{1}H NMR (500 MHz, MeOD) δ 7.57 – 7.36 (m, 3H), 7.30 - 7.24 (m, 3H), 7.05 (d, *J* = 1.7 Hz, 1H), 6.98 (d, *J* = 8.4 Hz, 1H), 6.92 (d, *J* = 8.3 Hz, 1H), 6.74 (t, *J* = 8.0 Hz, 1H), 6.66 (d, *J* = 8.3 Hz, 1H), 4.04 (dd, *J* = 12.1, 6.8 Hz, 1H), 3.24 (d, *J* = 7.3 Hz, 2H). HRMS (ESI-TOF) calc'd for $C_{21}H_{18}F_3N_2Cl$ [M + H]+: 391.1189 and 393.1159, found: 391.1401 and 393.1960.

4.4.4 HALOGENASE VARIANT CONVERSIONS

Figure 4.5. Halogenase variant conversions of *t*-Bu substrate 2.

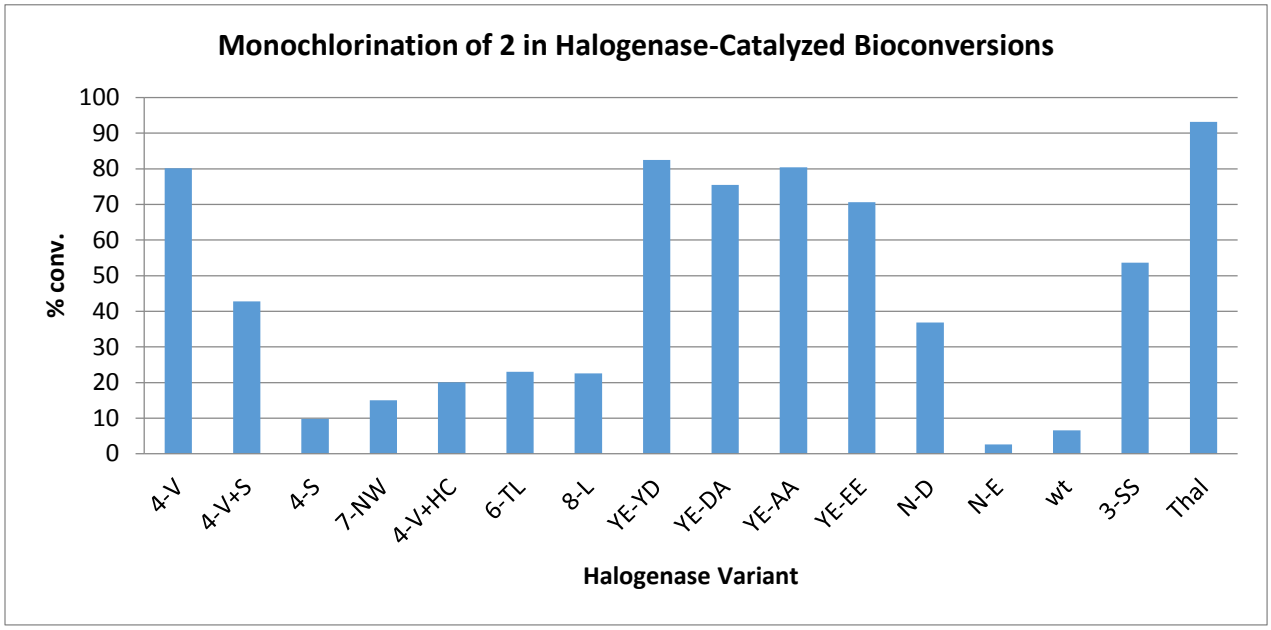


Figure 4.6. Halogenase variant conversions of *n*-Pr substrate 3.

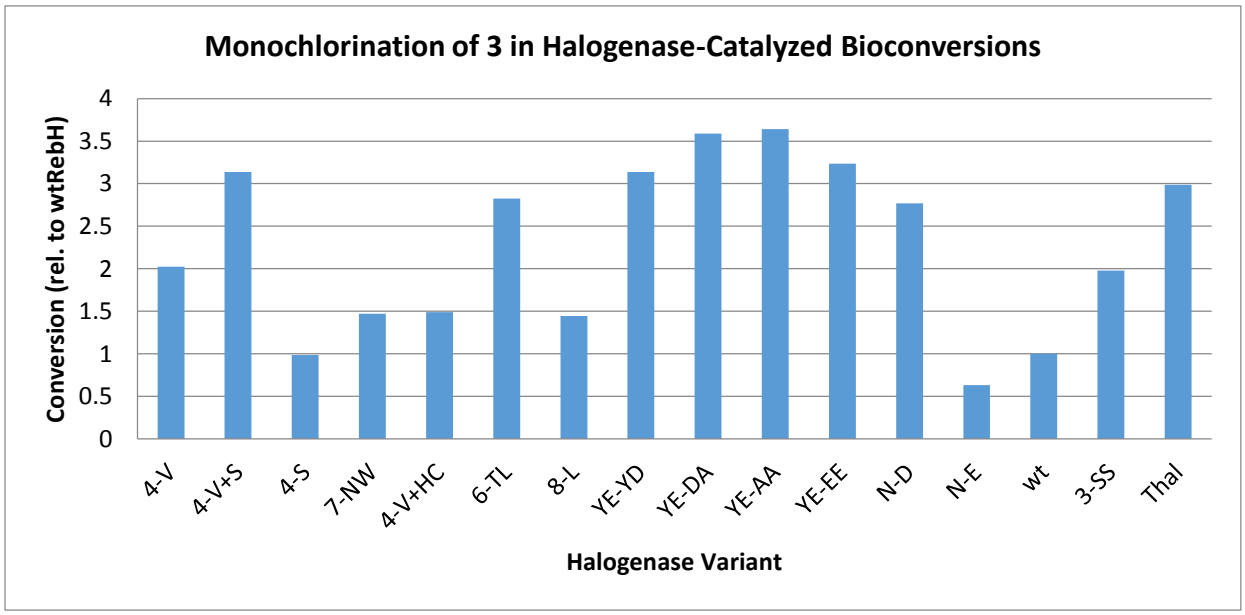


Figure 4.7. Halogenase variant conversions of methyl substrate 4.

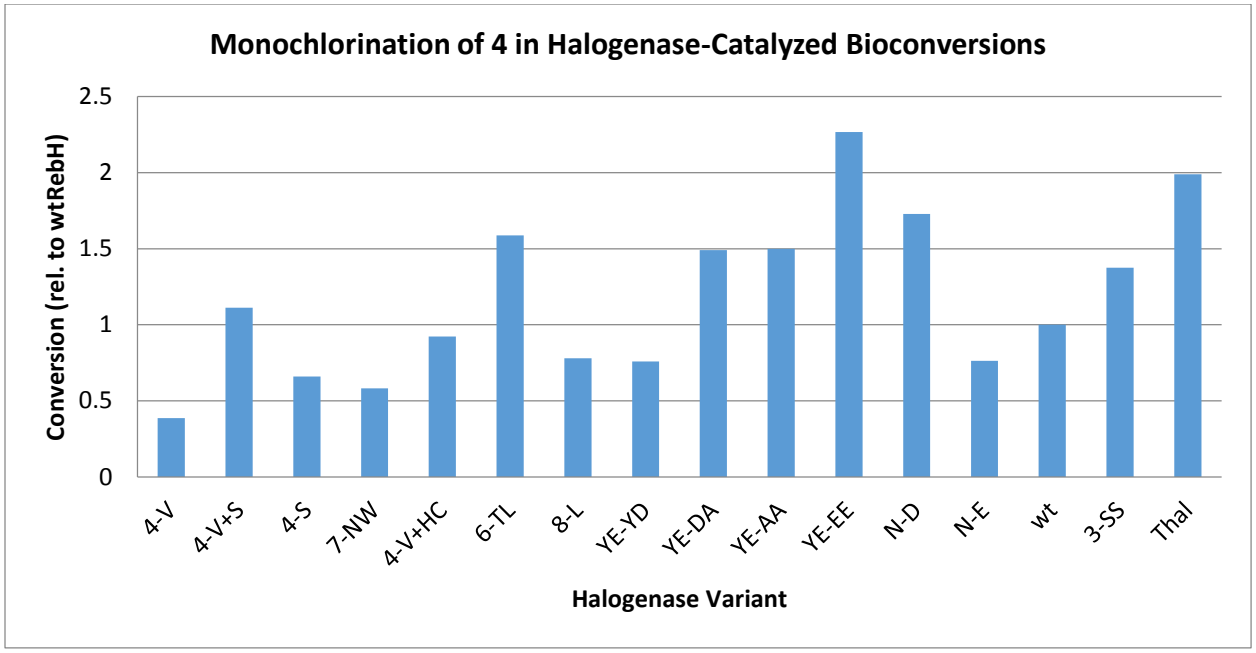


Figure 4.8. Halogenase variant conversions of ethyl ester substrate 5.

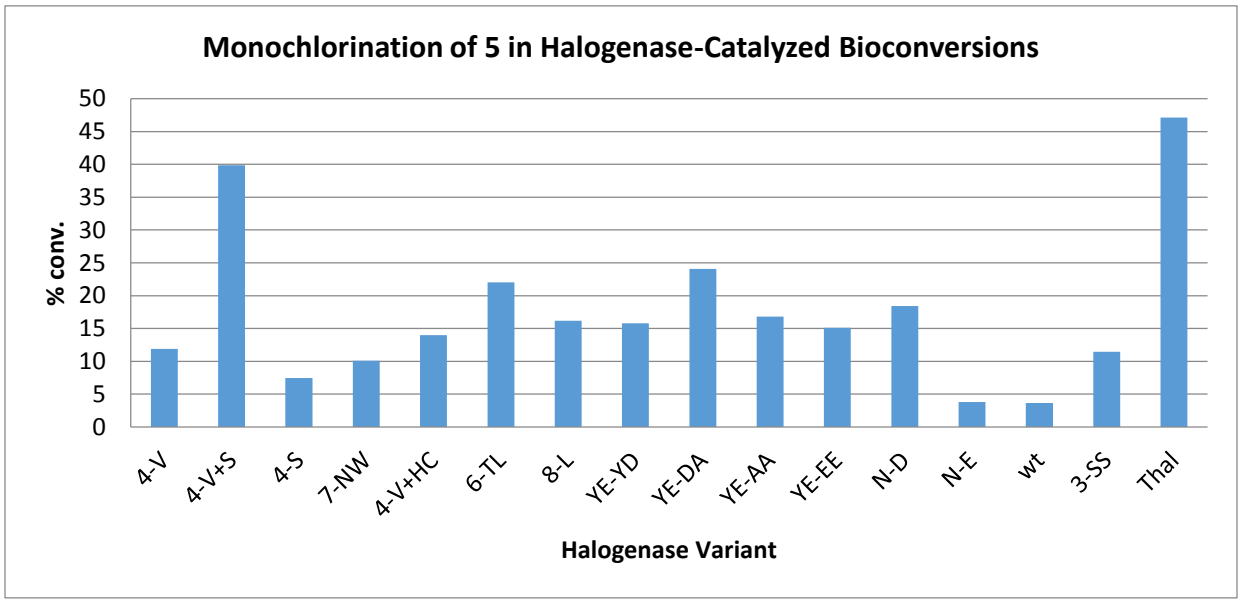


Figure 4.9. Halogenase variant conversions of benzyl substrate 6.

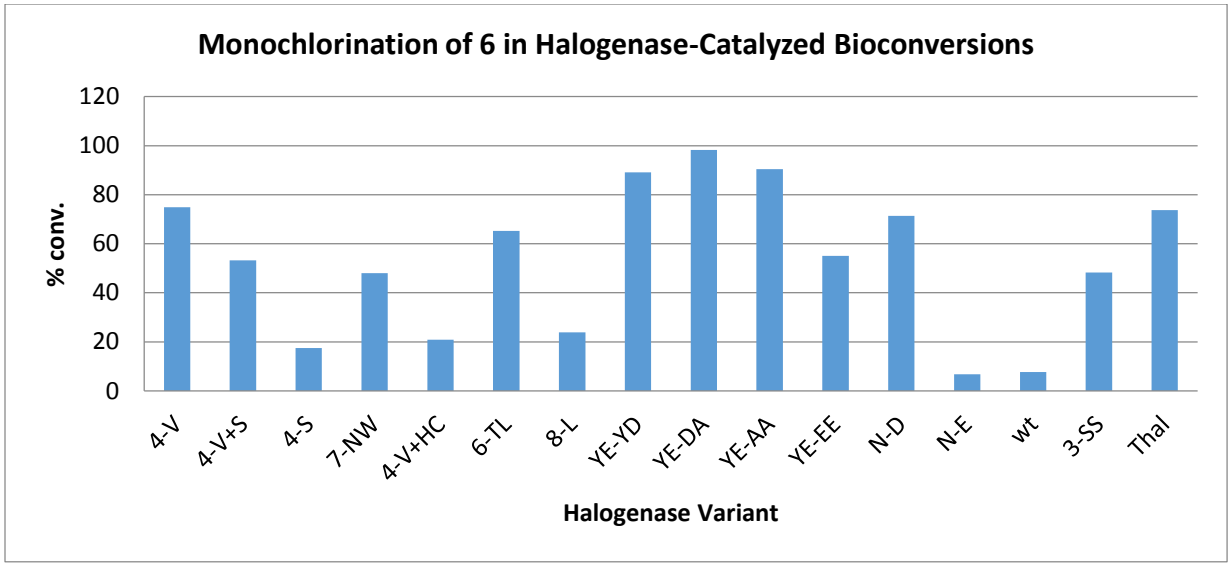


Figure 4.10. Halogenase variant conversions of *m*-MeO-benzyl substrate 7.

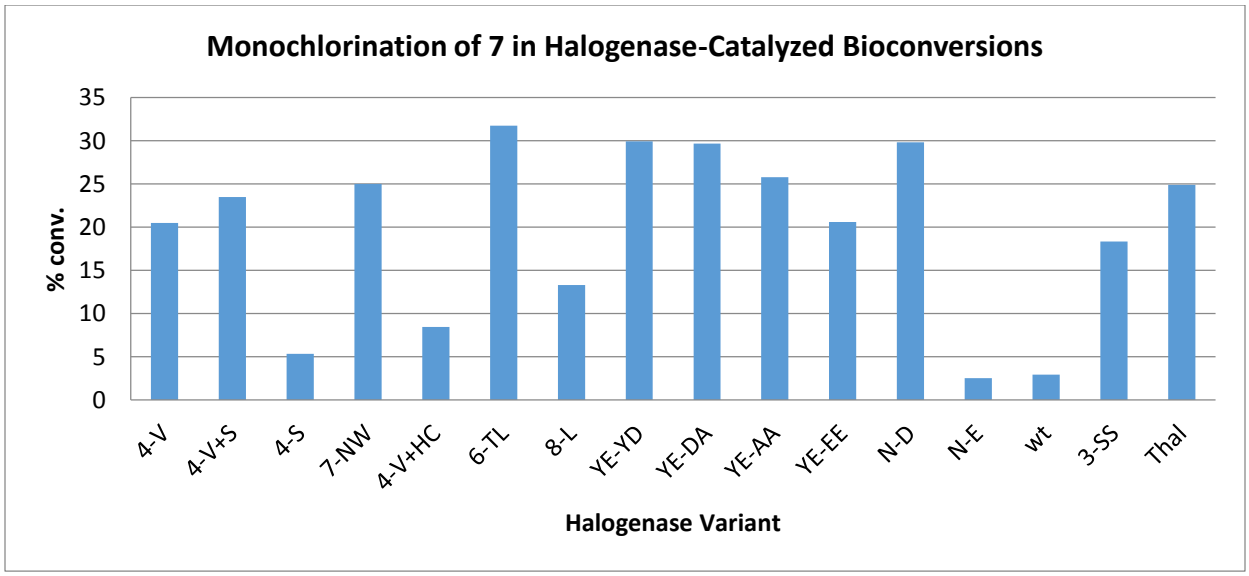
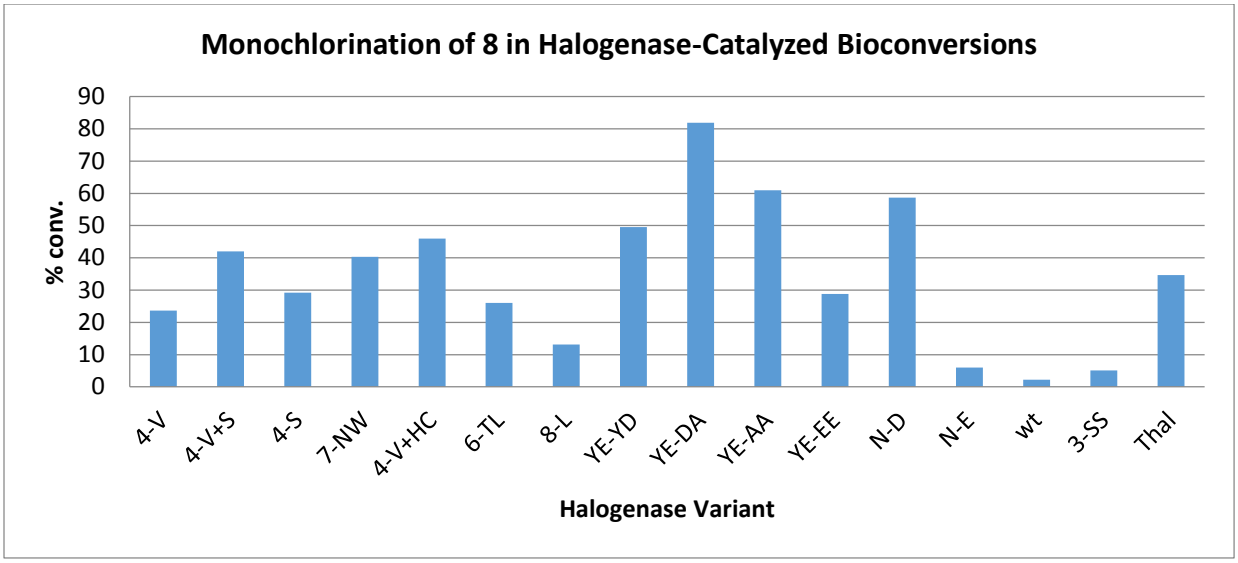


Figure 4.11. Halogenase variant conversions of *p*-CF_3-benzyl substrate 8.



4.5 ACKNOWLEDGEMENTS

We would like to thank Mary Andorfer for the cloning of RebH variants 4-V+S and 4-V+HC, as well as for the cloning of thal into pET28. We would also like to thank Kyle Kunze and Dr. Duo-Sheng Wang for their work on synthesizing a number of the desymmetrization substrates and their racemic chlorinated products, as described in Section 4.2.2.

4.6 REFERENCES

[1] Bernhardt, P.; Okino, T.; Winter, J. M.; Miyanaga, A.; Moore, B. S. A stereoselective vanadium-dependent chloroperoxidase in bacterial antibiotic biosynthesis. *J. Am. Chem. Soc.* **133**, 4268-4270 (2011).
[2] Neumann, C. S.; Fujimori, D. G.; Walsh, C. T. Halogenation strategies in natural product biosynthesis. *Chem. Biol.* **15**, 99-109 (2008).
[3] Blasiak, L. C.; Drennan, C. L. Structural perspective on enzymatic halogenation. *Accounts Chem. Res.* **42**, 147-155 (2009).
[4] Caner, H.; Groner, E.; Levy, L.; Agranat, I. Trends in the development of chiral drugs. *Drug Disc. Tod.* **9**, 105-110 (2004).
[5] Ameen, D.; Snape, T. J. Chiral 1,1-diaryl compounds as important pharmacophores. *MedChemComm* **4**, 893-907 (2013).
[6] Xu, B.; Li, M.-L.; Zuo, X.-D.; Zhu, S.-F.; Zhou, Q.-L. Catalytic Asymmetric Arylation of α-Aryl-α-diazoacetates with Aniline Derivatives. *J. Am. Chem. Soc.* **137**, 8700-8703 (2015).

[7] Lewis, C. A.; Gustafson, J. L.; Chiu, A.; Balsells, J.; Pollard, D.; Murry, J.; Reamer, R. A.; Hansen, K. B.; Miller, S. J. A Case of Remote Asymmetric Induction in the Peptide-Catalyzed Desymmetrization of a Bis(phenol). *J. Am. Chem. Soc.* **130**, 16358-16365 (2008).
[8] Payne, J. T.; Poor, C. B.; Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angew. Chem. Int. Ed.* **54**, 4226-4230 (2015).
[9] a) Poor, C. B.; Andorfer, M. C.; Lewis, J. C. Improving the Stability and Catalyst Lifetime of the Halogenase RebH By Directed Evolution. *ChemBioChem* **15**, 1286-1289 (2014); b) Andorfer, M. C.; Park, H. J.; Vergara-Coll, J.; Lewis, J. C. Directed Evolution of RebH for Catalyst-Controlled Halogenation of Aromatic C-H Bonds. Manuscript in preparation.
[10] Kimura, T.; Hosokawa-Muto, J.; Kamatari, Y. O.; Kuwata, K. Synthesis of GN8 derivatives and evaluation of their antiprion activity in TSE-infected cells. *Bioorg. Med. Chem. Let.* **21**, 1502-1507 (2011).
[11] Bitto, E.; Bingman, C. A.; Singh, S.; Phillips, Jr., G. N. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70**, 289–293 (2008).
[12] Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M. The Structural Basis for Substrate Anchoring, Active Site Selectivity, and Product Formation by P450 PikC from Streptomyces venezuelae. *J. Biol. Chem.* **281**, 26289-26297 (2006).
[13] 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*. **6**, 3454-3460 (2015).
[14] Seibold, C.; Schnerr, H.; Rumpf, J.; Kunzendorf, A.; Hatscher, C.; Wage, T.; Ernyei, A. J.; Dong, C.; Naismith, J. H.; van Pée, K.-H. A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocatal. Biotrans.* **24**, 401-408 (2006).
[15] Li, J.J. (**2011**) "Larock Indole Synthesis". *Name Reactions in Heterocyclic Chemistry II*, John Wiley & Sons, ISBN 978-0-470-08508-0, pp. 143–166.
[16] a) Sridharan, V.; Suryavanshi, P. A.; Menéndez, J. C. Advances in the Chemistry of Tetrahydroquinolines. *Chem. Rev.* **111**, 7157-7259 (2011); b) Li, J.-J.; Mei, T.-S.; Yu, J.-Q. Synthesis of Indolines and Tetrahydroisoquinolines from Arylethylamines by Pd^{II}-Catalyzed C-H Activation Reactions. *Angew. Chem. Int. Ed.* **47**, 6452-6455 (2008).
[17] Ghatge, N. D.; Khune, G. D. Synthesis of diamines. *Ind. Chem. J.* **13**, 22-32 (1978).
[18] Hamatani, T.; Massuda, T. Improved preparation of 1,1-bis(4-aminophenyl)ethane. *Jpn. Kokai Tokkyo Koho*, Patent 2001253856. 18 September 2001.
[19] Kuwata, K.; Kimura, T.; Junji, M. Preparation of bis[4-[2-(pyrrolidin-1-yl)acetamido]phenyl] methane derivatives as prion protein structure transformation inhibitors for prevention and/or treatment of prion diseases. Patent 2010131717. 18 November 2010.
[20] Sambrook, J.; Frisch, E. F.; Maniatis, T. (**1989**) *Molecular cloning: a laboratory manual*. 2, Cold Spring Harbor Laboratory Press, New York.
[21] Heckman, K. L.; Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2**, 924–932 (2007).
[22] Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem*. **62**, 7512-7515 (1997).

APPENDIX I

NMR SPECTRA FOR COMPOUNDS FROM CHAPTER II

Figure AI.1. ^{1}H NMR spectrum of 1.

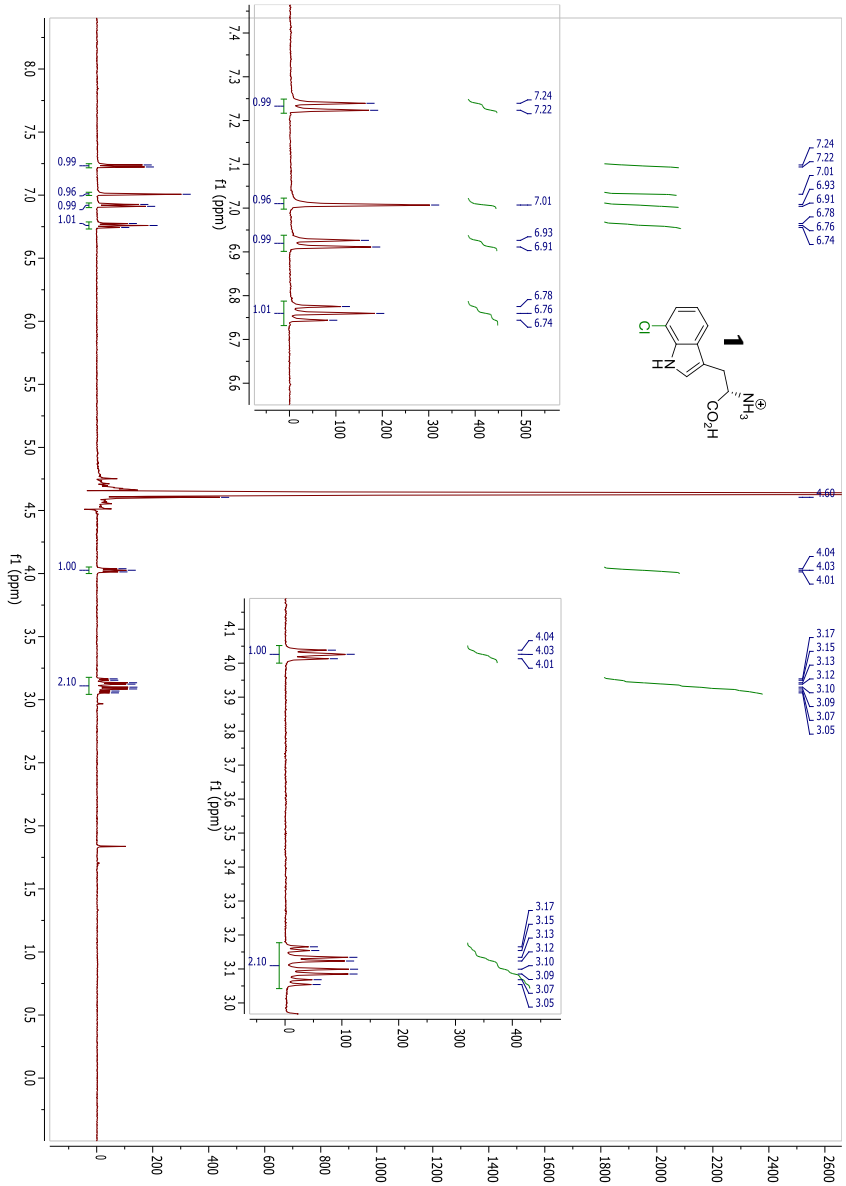


Figure AI.2. ^{13}C NMR spectrum of 1.

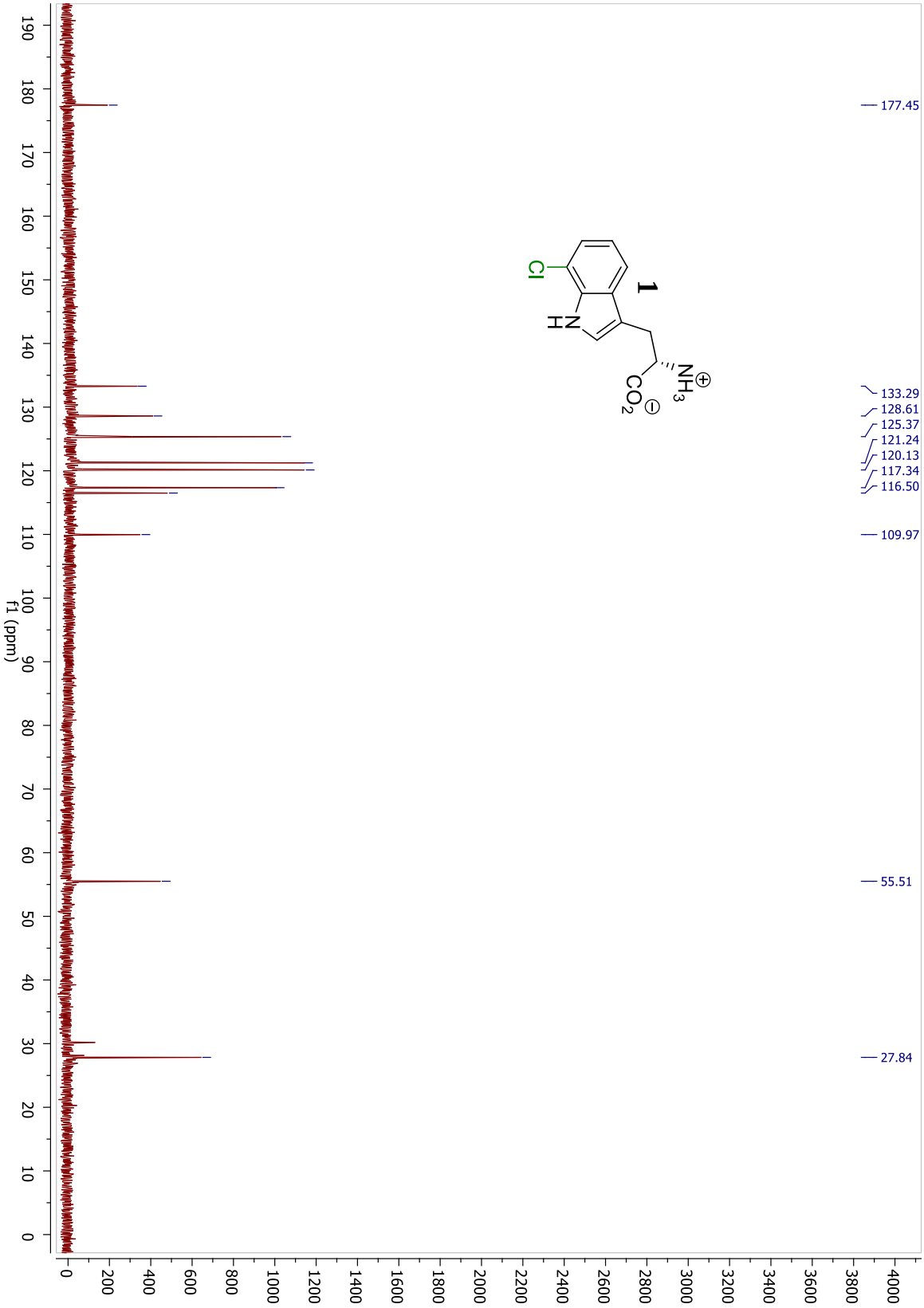
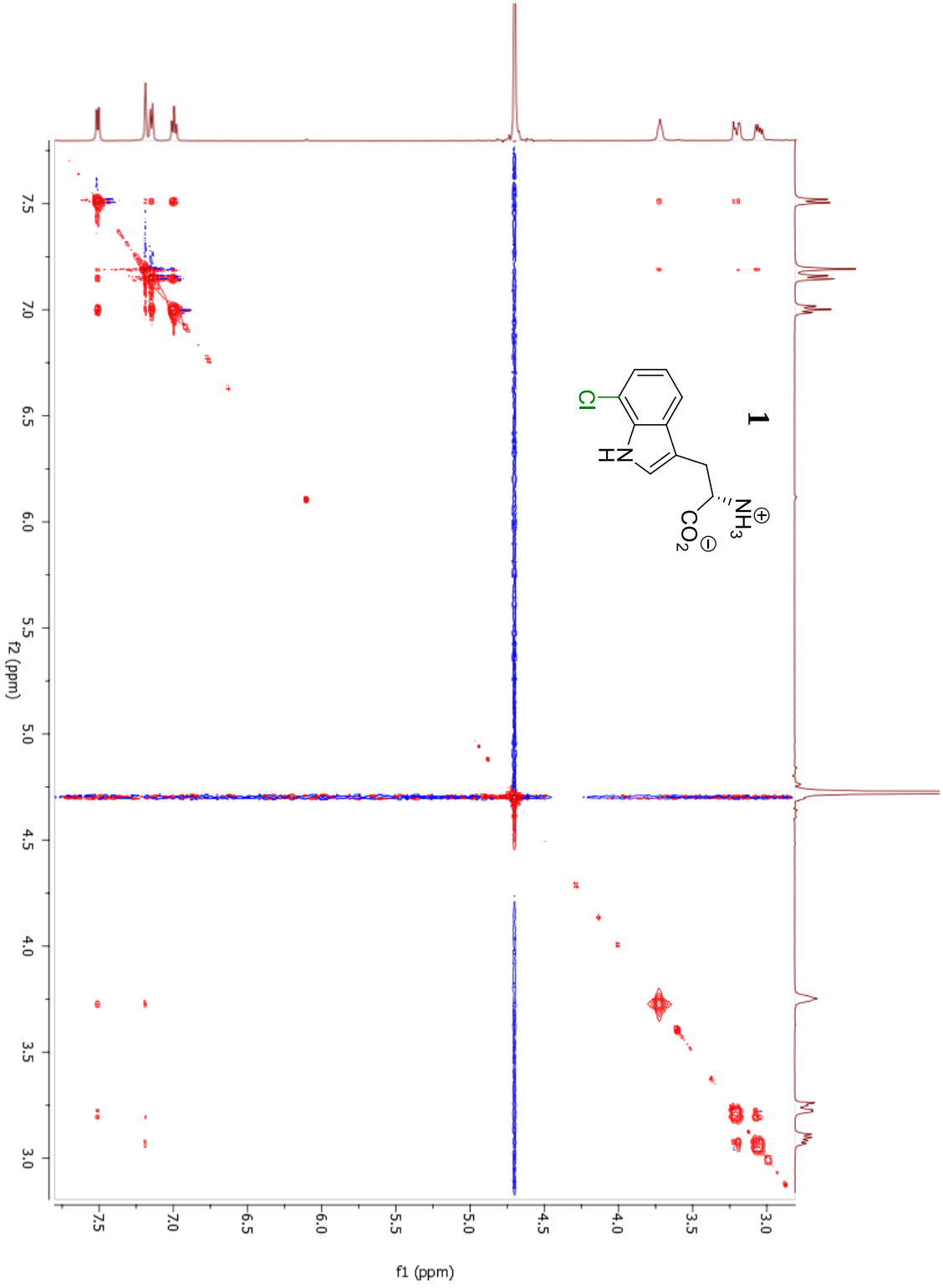


Figure AI.3. NOESY spectrum of 1.[a]



[a] Resonances between peaks at 7.19 and 3.72 ppm, and between peaks at 7.51 and 3.72 ppm demonstrate chlorination at the 7 position.

Figure AI.4. ^{1}H NMR spectrum of 2.

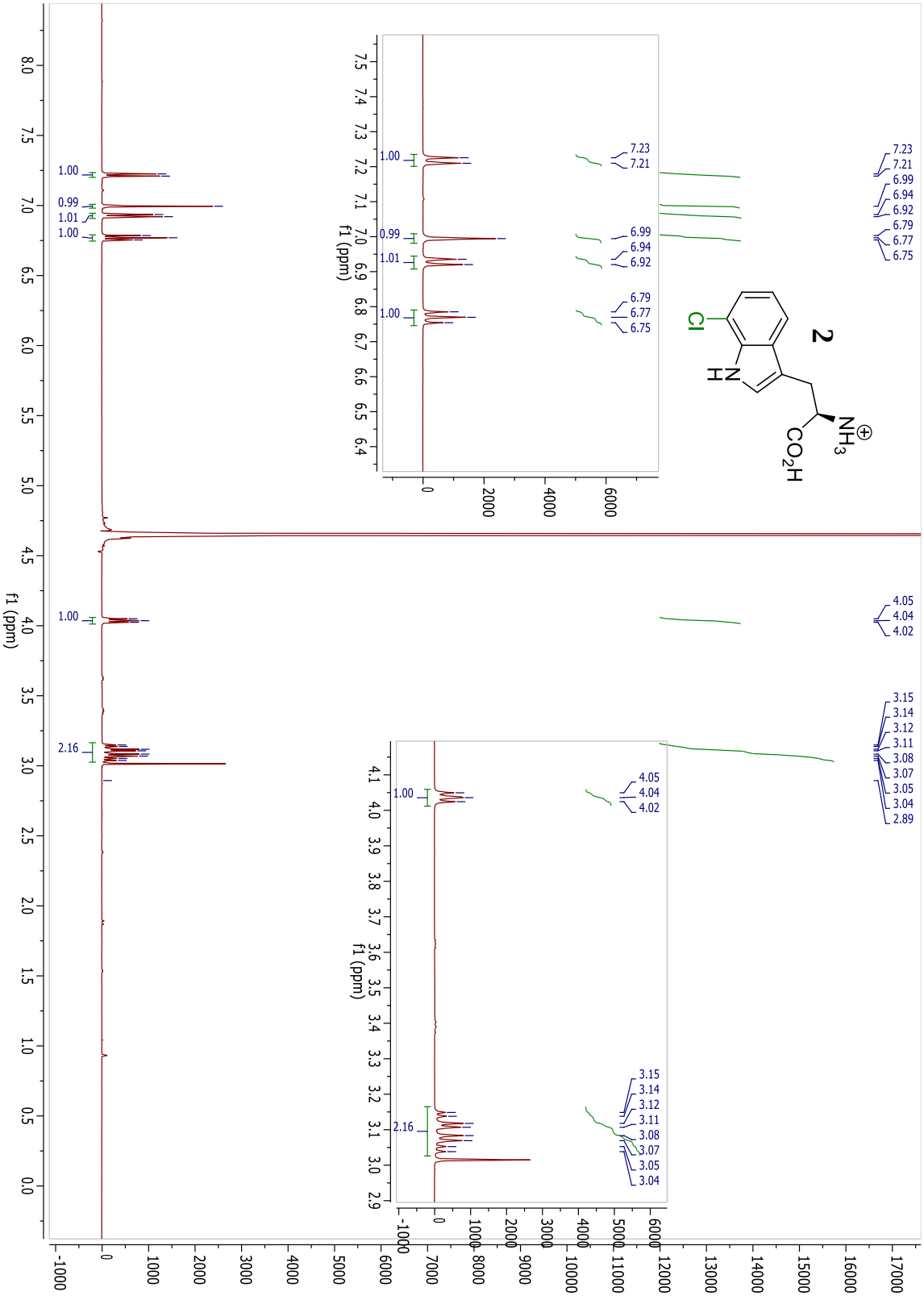


Figure AI.5. ^{13}C NMR spectrum of 2.

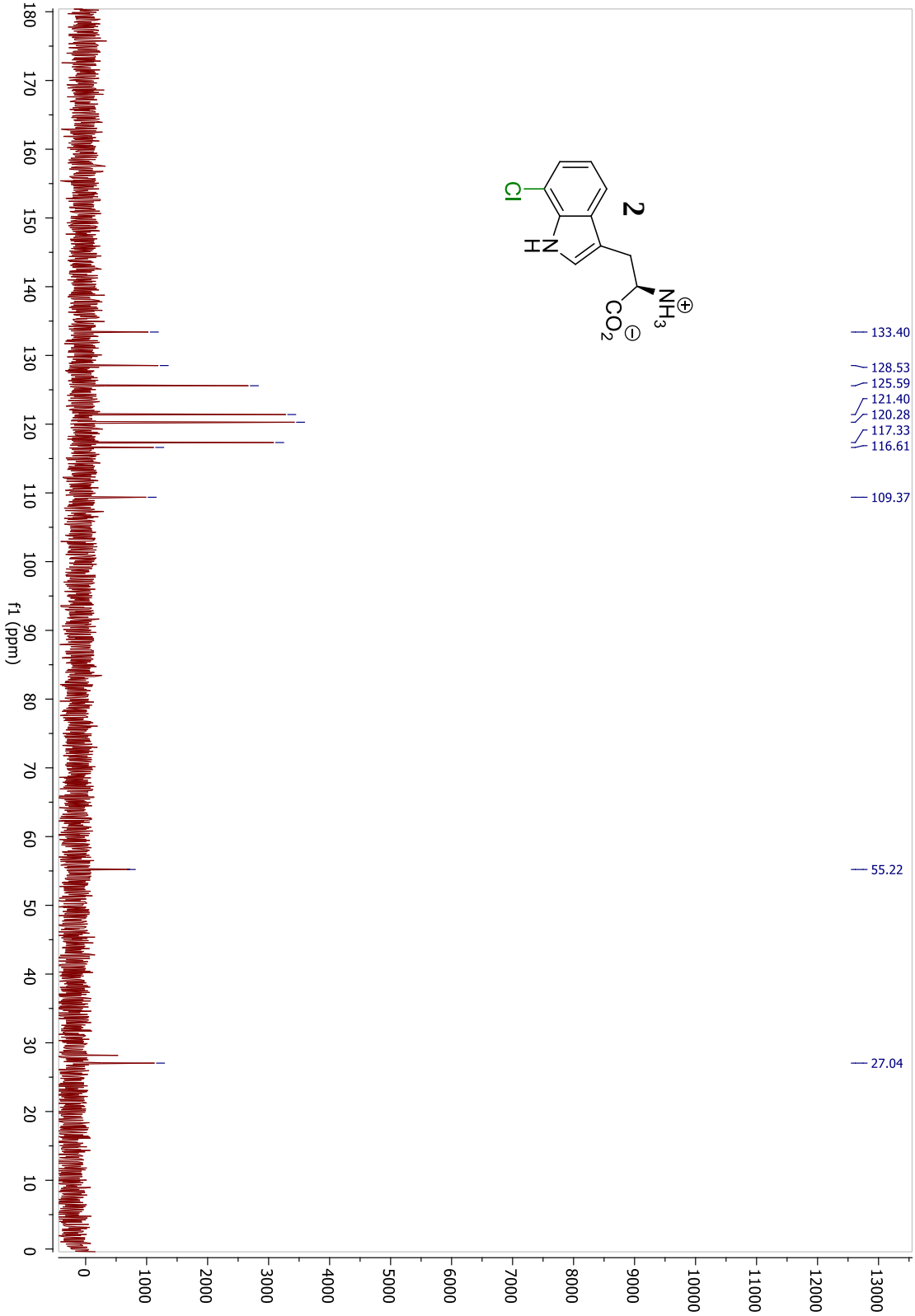
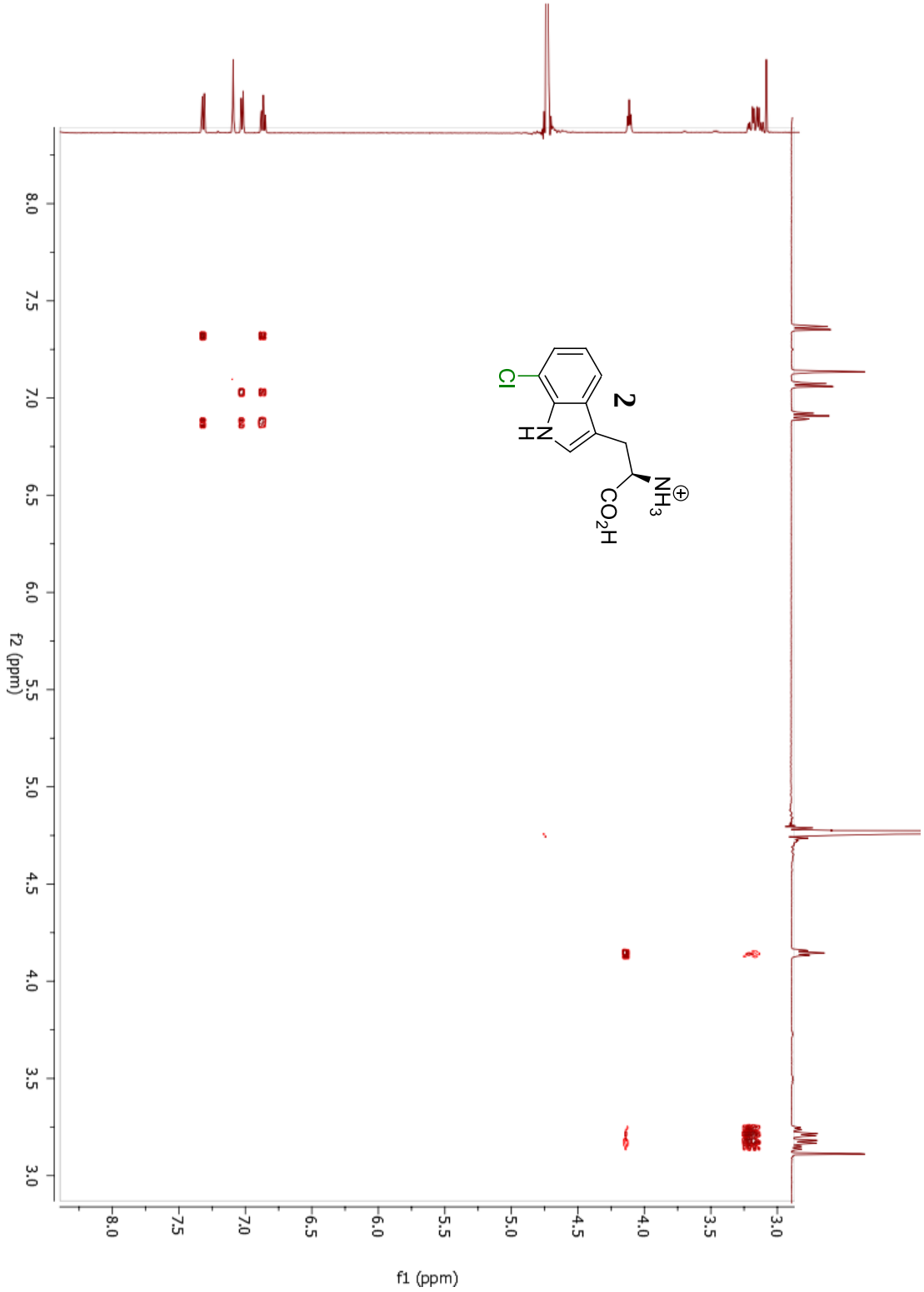
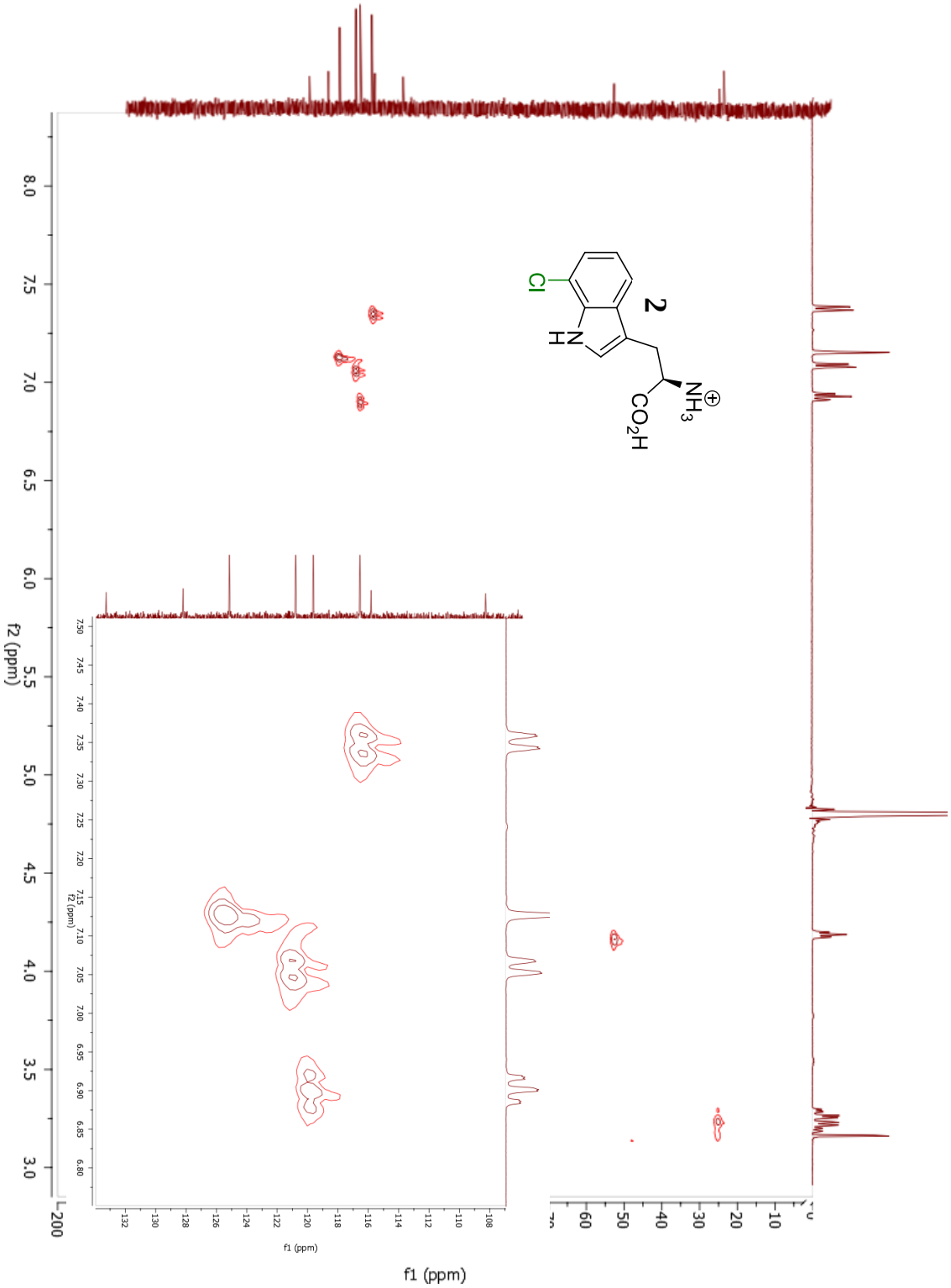


Figure AI.6. COSY spectrum of 2.[a]



[a] Resonances between peaks at 7.22 and 6.77 ppm, and between peaks at 6.93 and 6.7 ppm support chlorination at the 7 position.

Figure AI.7. HMQC spectrum of 2.[a]



[a] Resonances support chlorination at the 7 position.

Figure A1.8. ^{1}H NMR spectrum of 3.

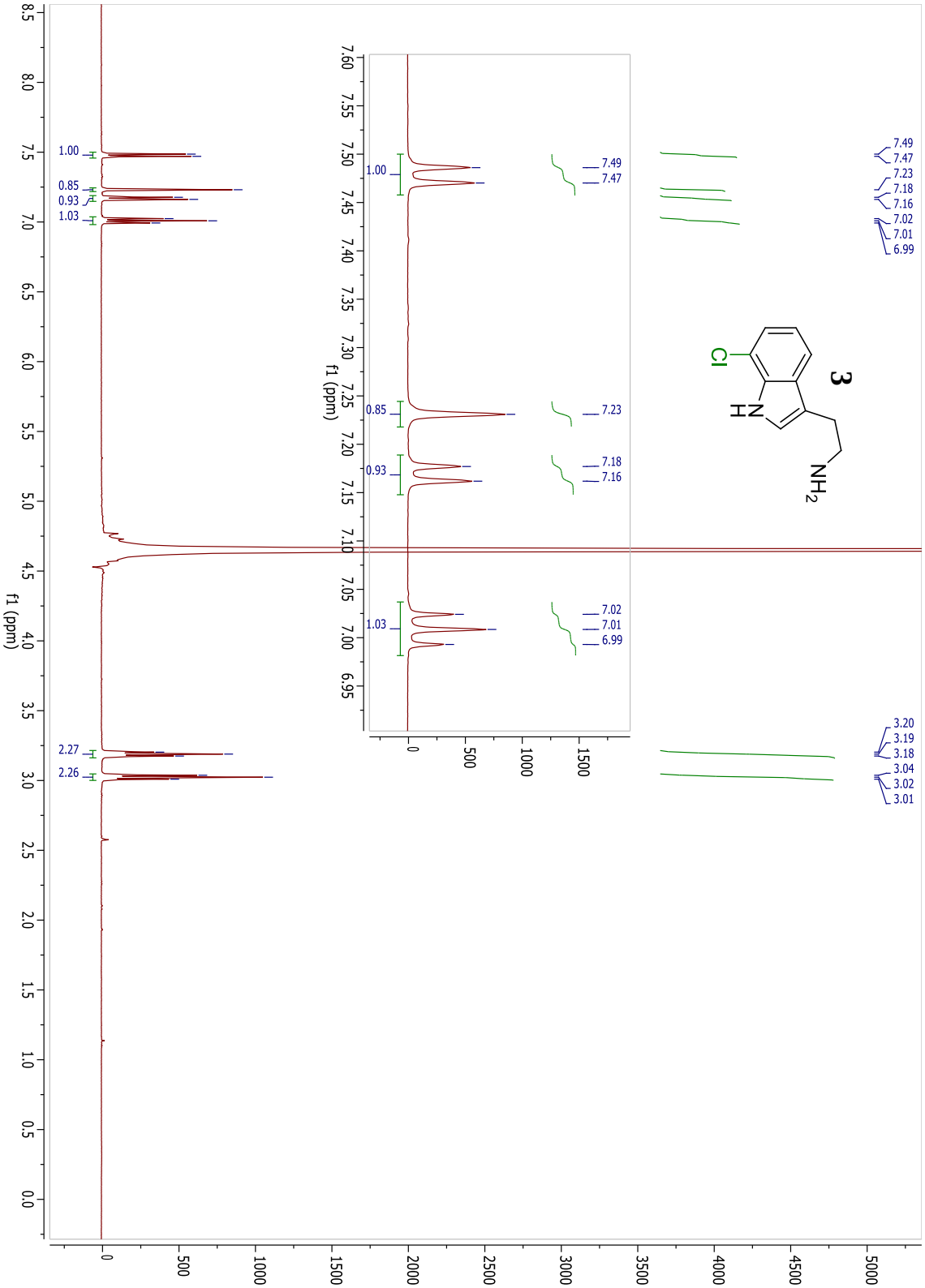


Figure AI.9. ^{13}C NMR spectrum of 3.

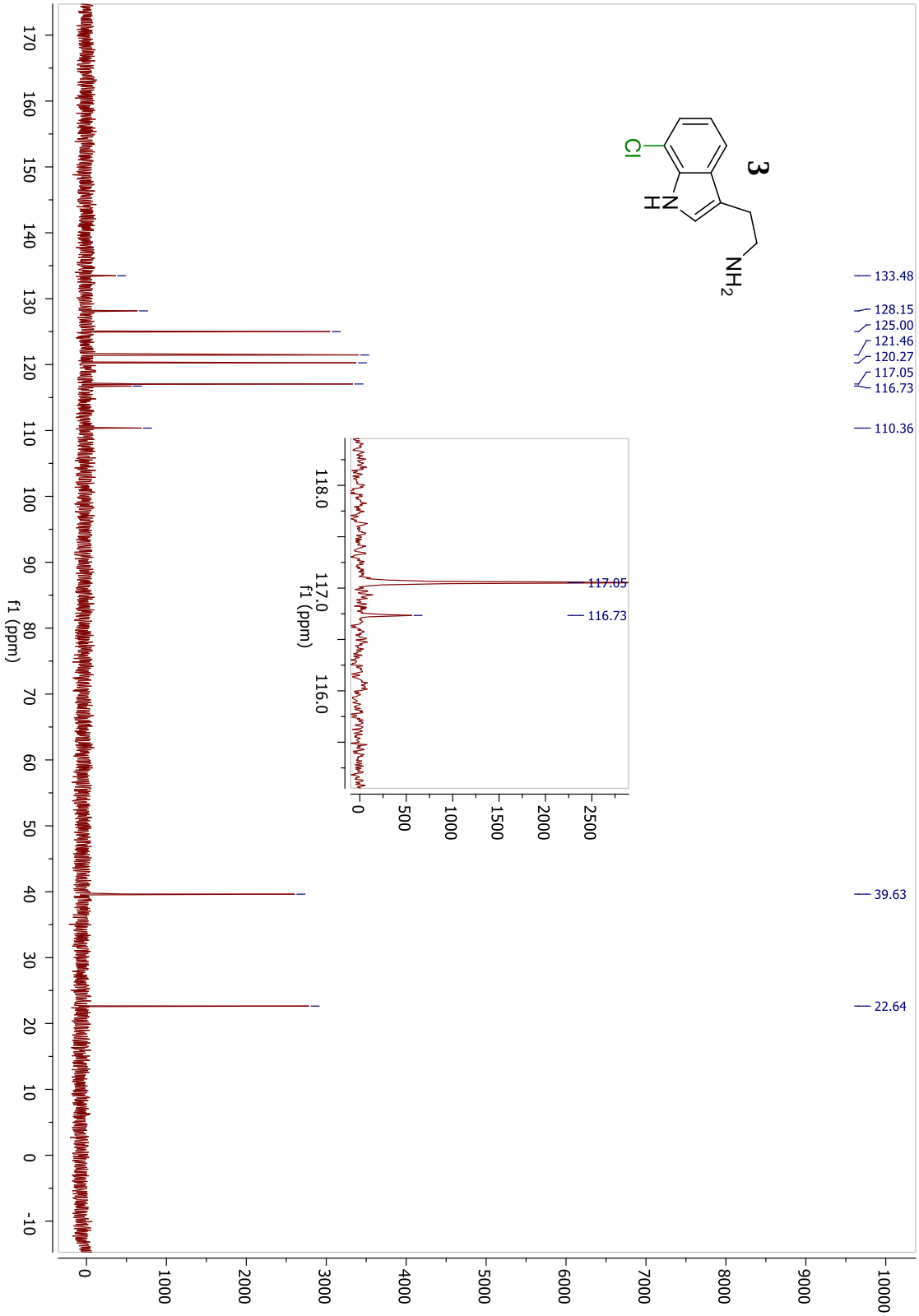
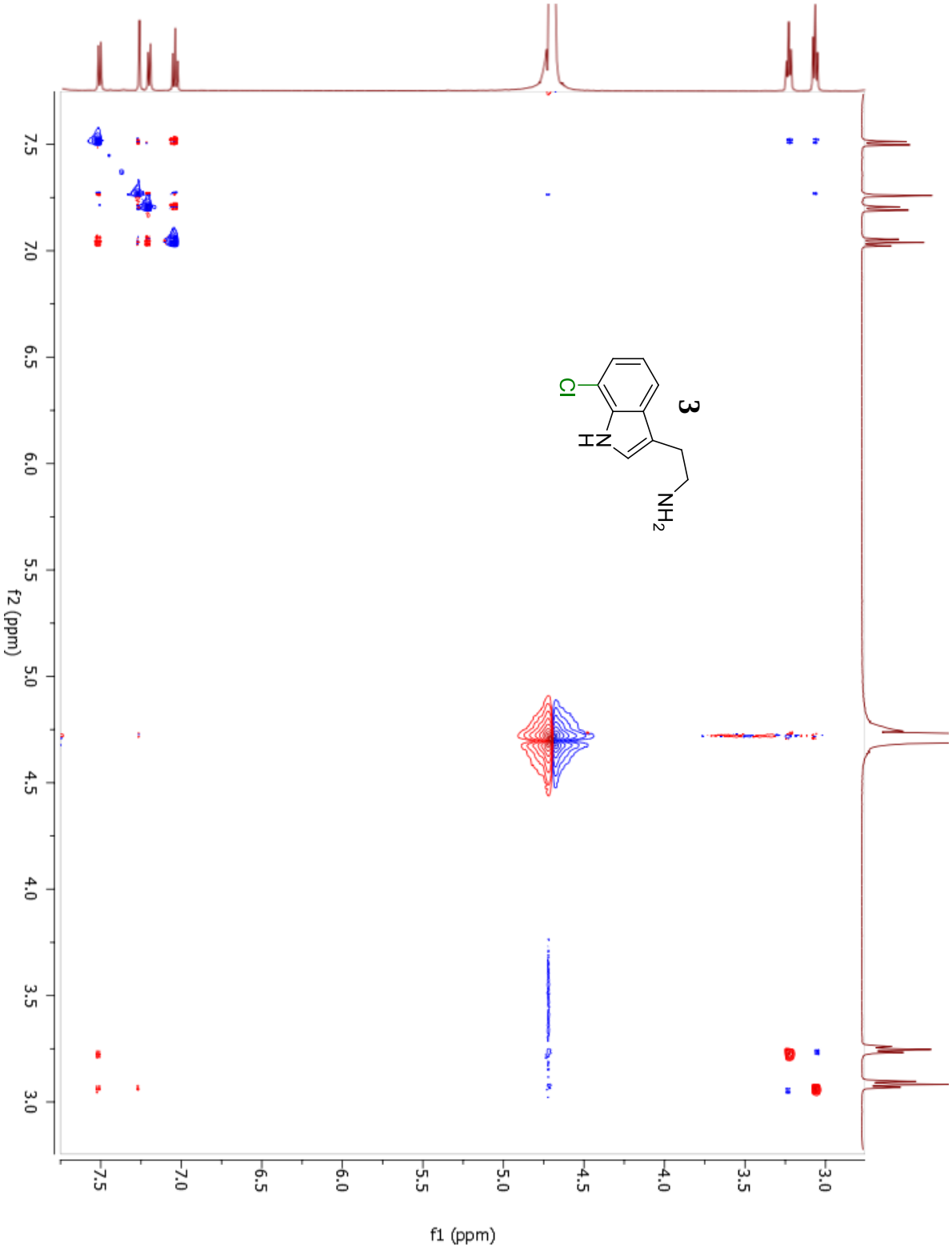


Figure AI.10. NOESY spectrum of 3.[a]



[a] Resonances between peaks at 7.48 and 3.02 ppm, and between peaks at 7.23 and 3.02 ppm demonstrate chlorination at the 7 position.

Figure AI.11. ^{1}H NMR spectrum of 4.

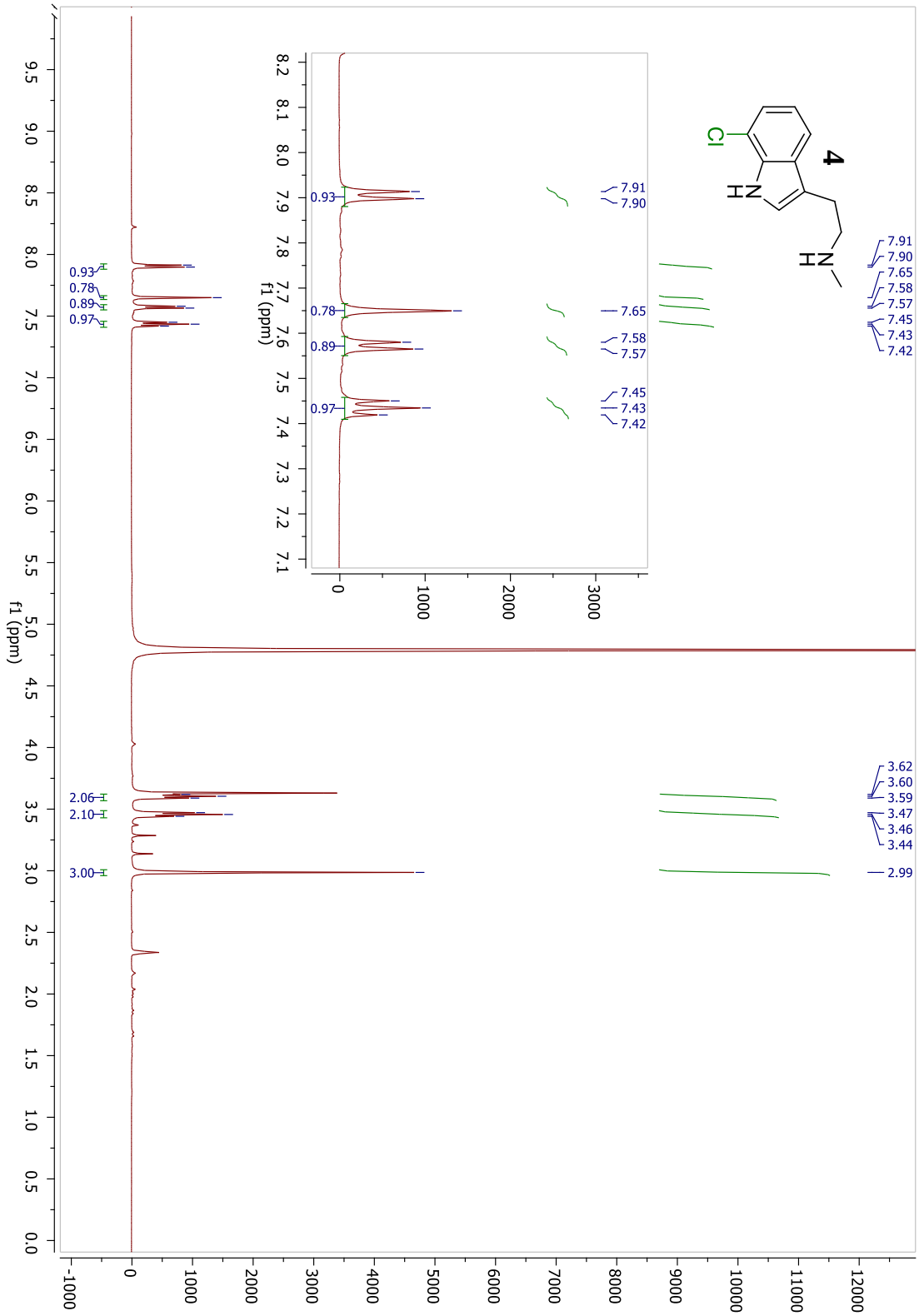


Figure AI.12. ^{13}C NMR spectrum of 4.

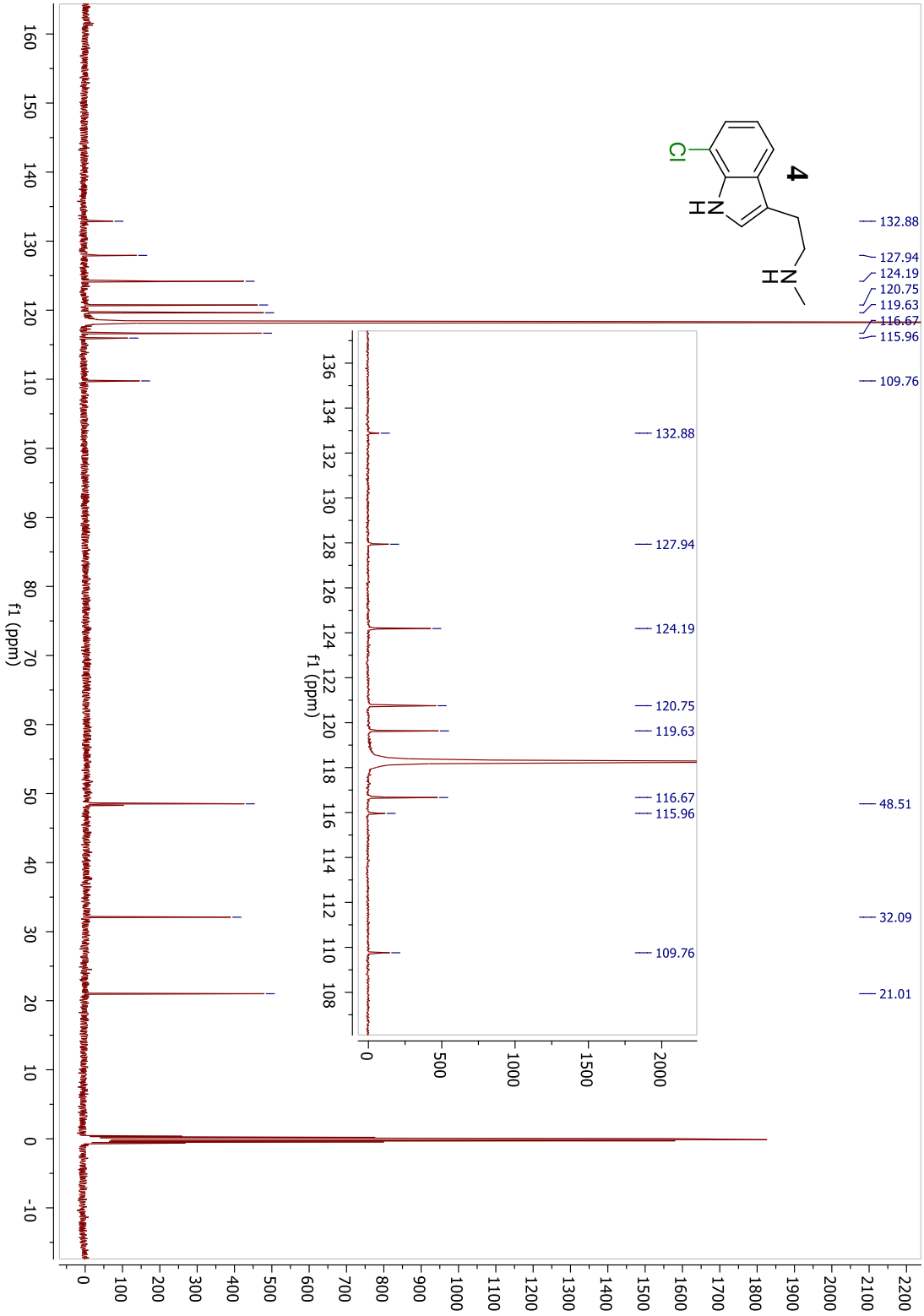
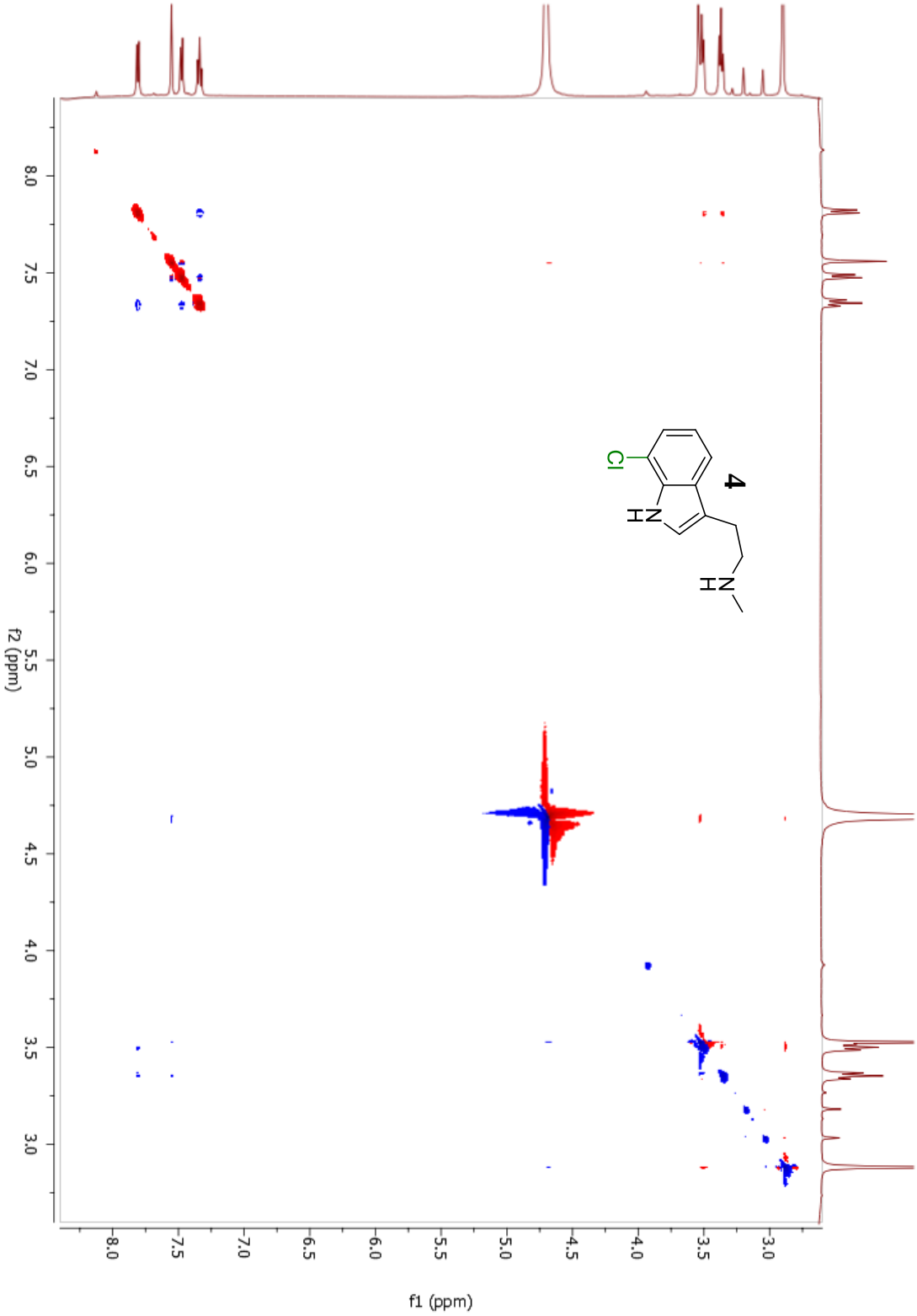


Figure AI.13. NOESY spectrum of 4.[a]



[a] Resonances between peaks at 7.90 and 3.46 ppm, and between peaks at 7.65 and 3.46 ppm demonstrate chlorination at the 7 position.

Figure AI.14. ^{1}H NMR spectrum of 5.

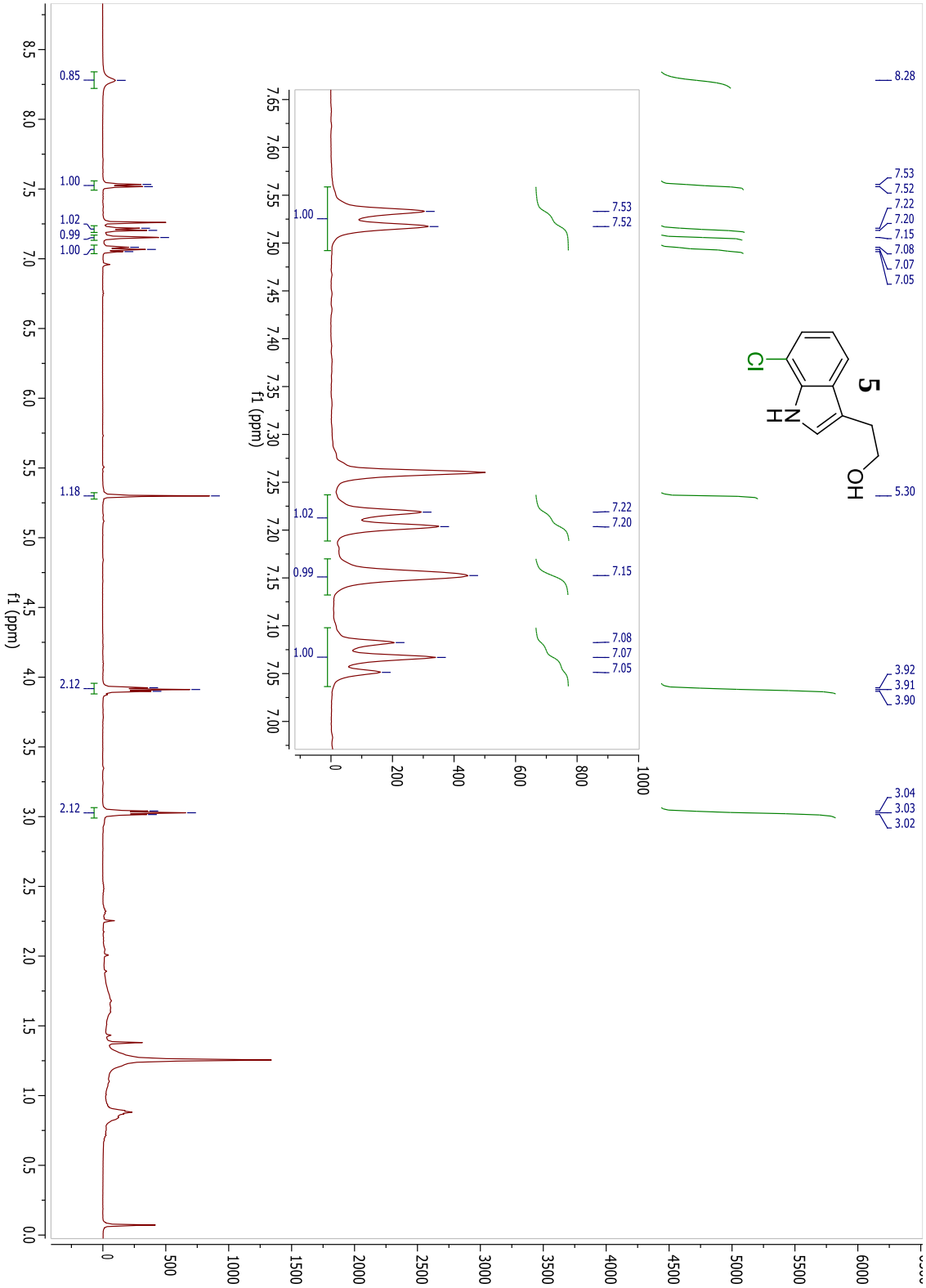


Figure AI.15. ^{13}C NMR spectrum of 5.

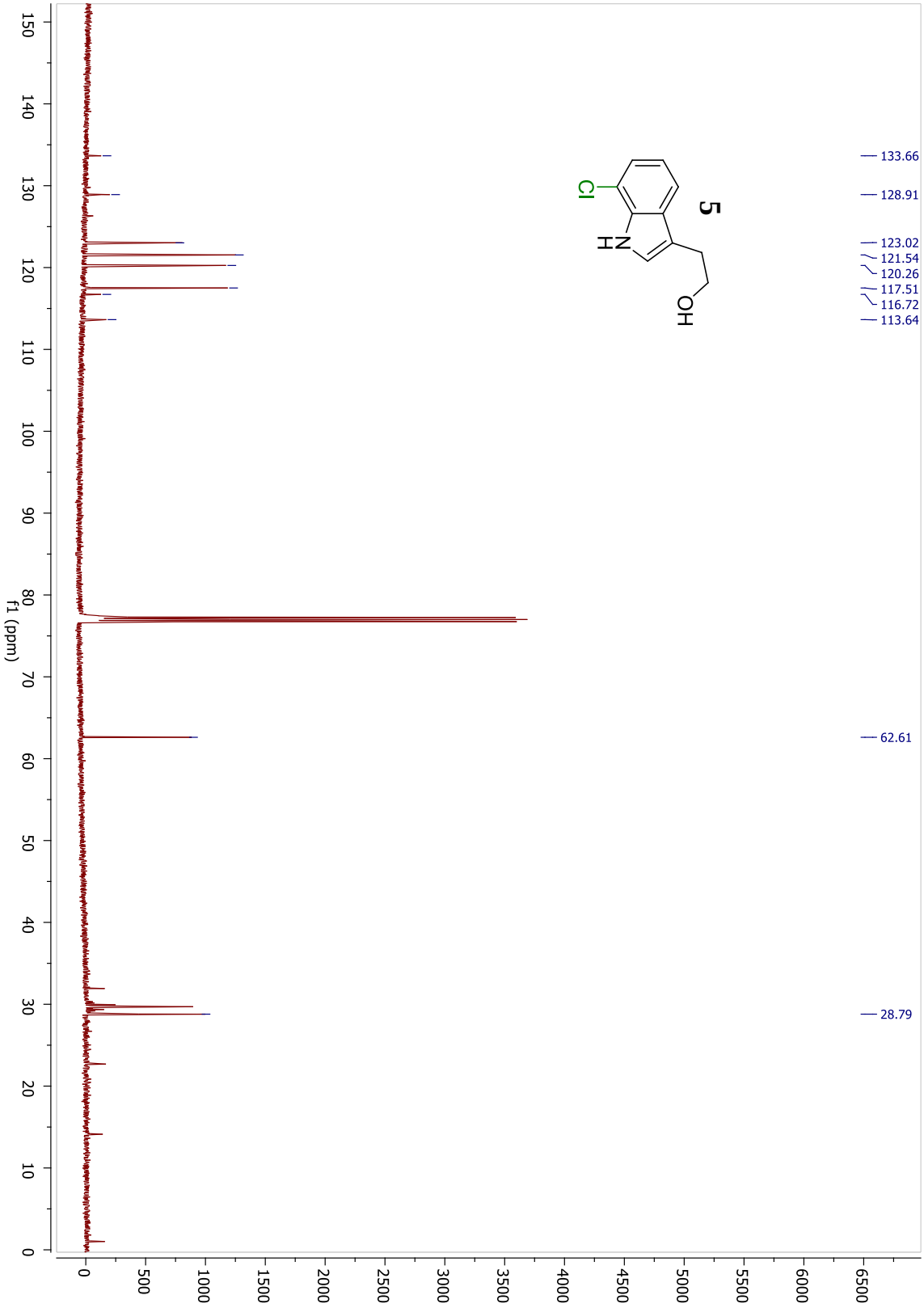
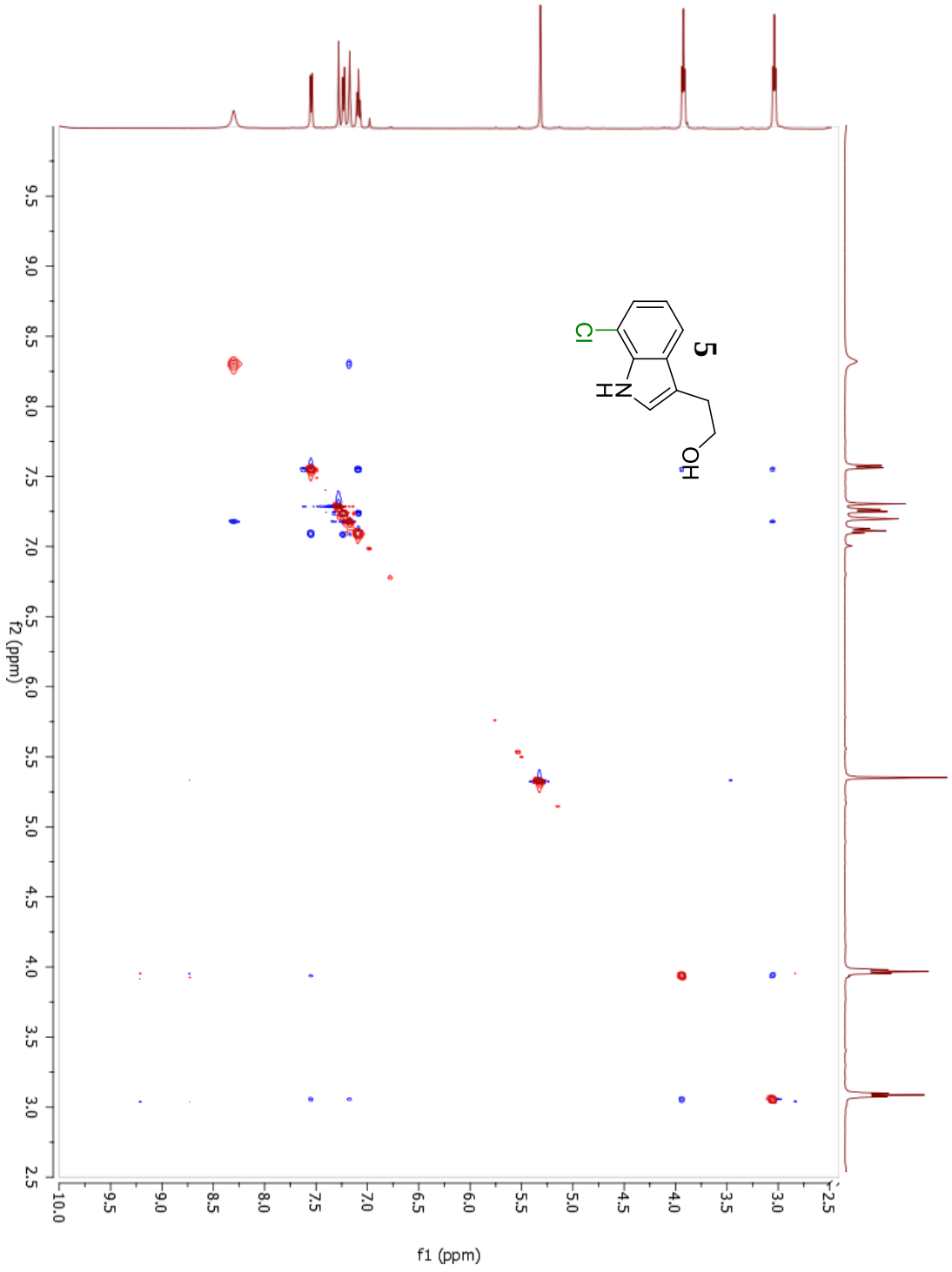


Figure AI.16. NOESY spectrum of 5.[a]



[a] Resonances between peaks at 7.53 and 3.03 ppm, and between peaks at 7.15 and 3.03 ppm demonstrate chlorination at the 7 position.

Figure AI.17. ^{1}H NMR spectrum of 6.

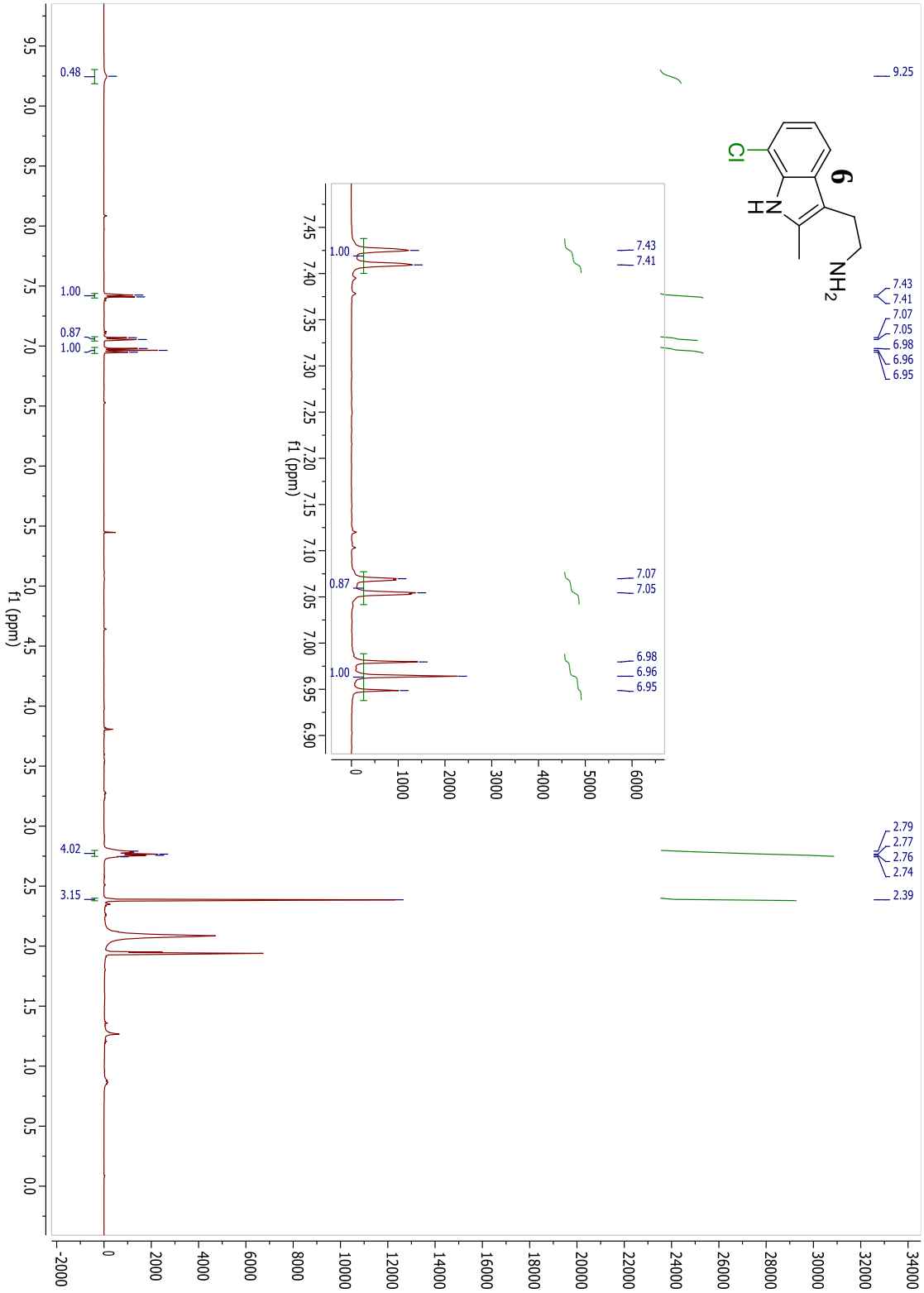


Figure AI.18. ^{13}C NMR spectrum of 6.

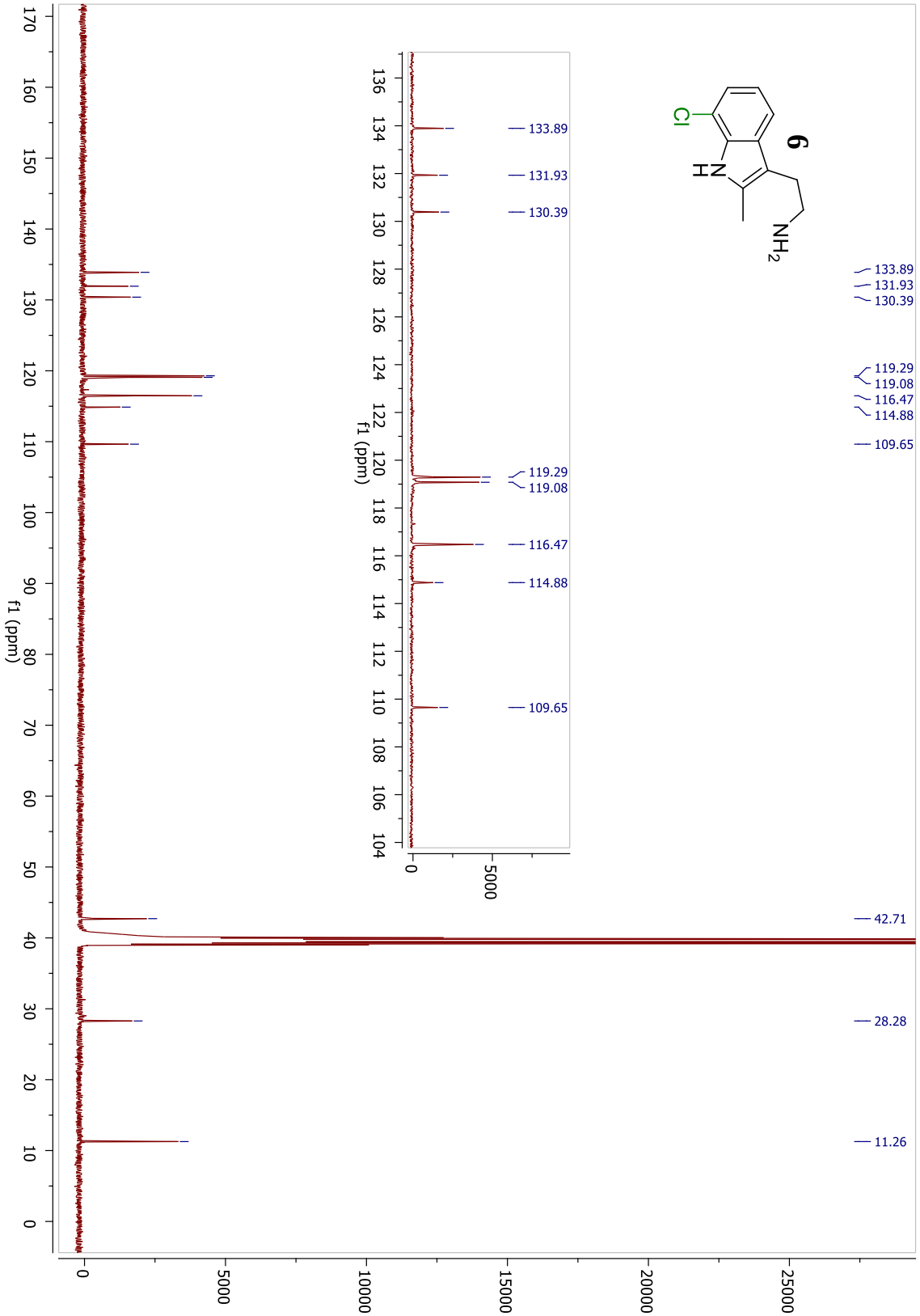
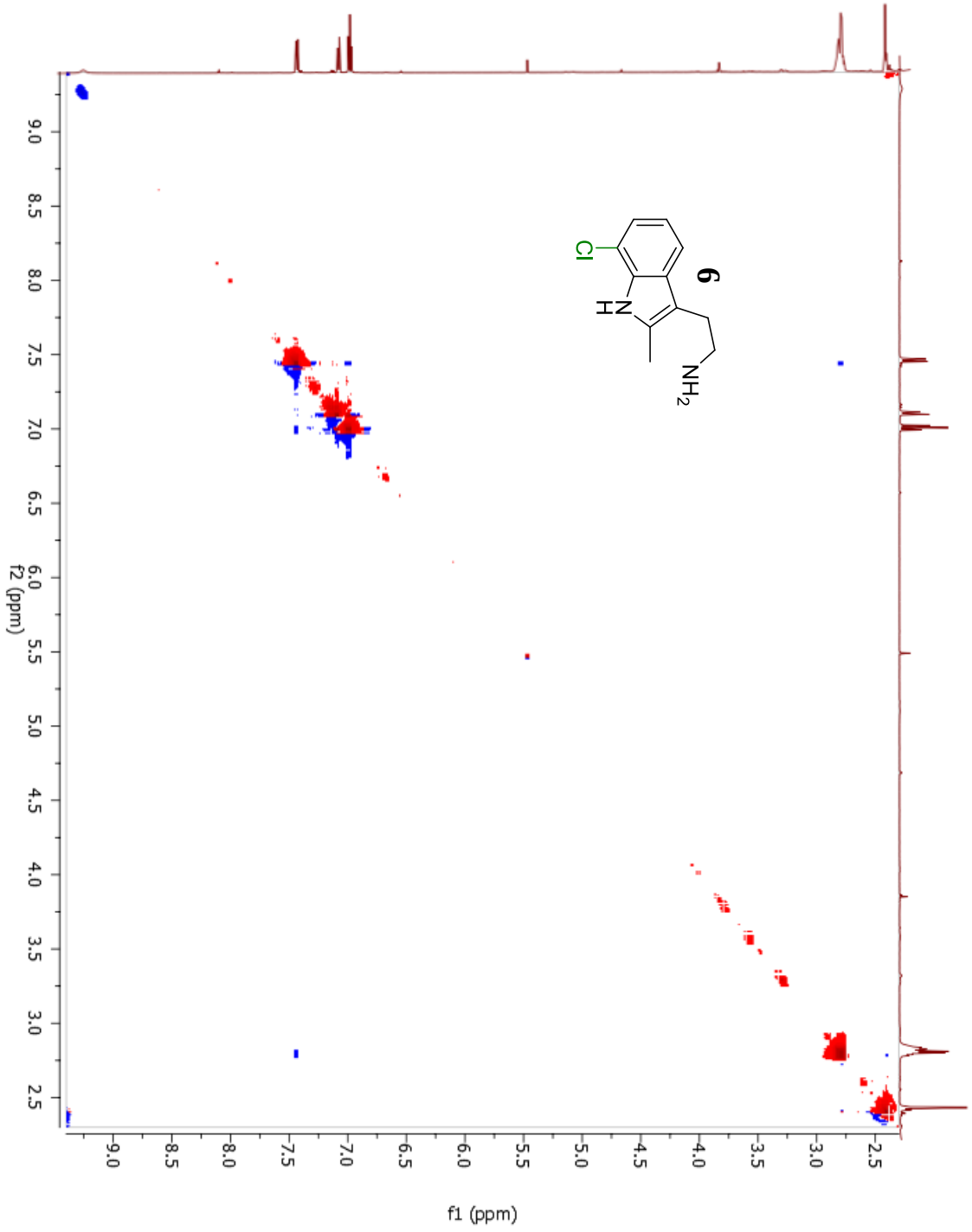
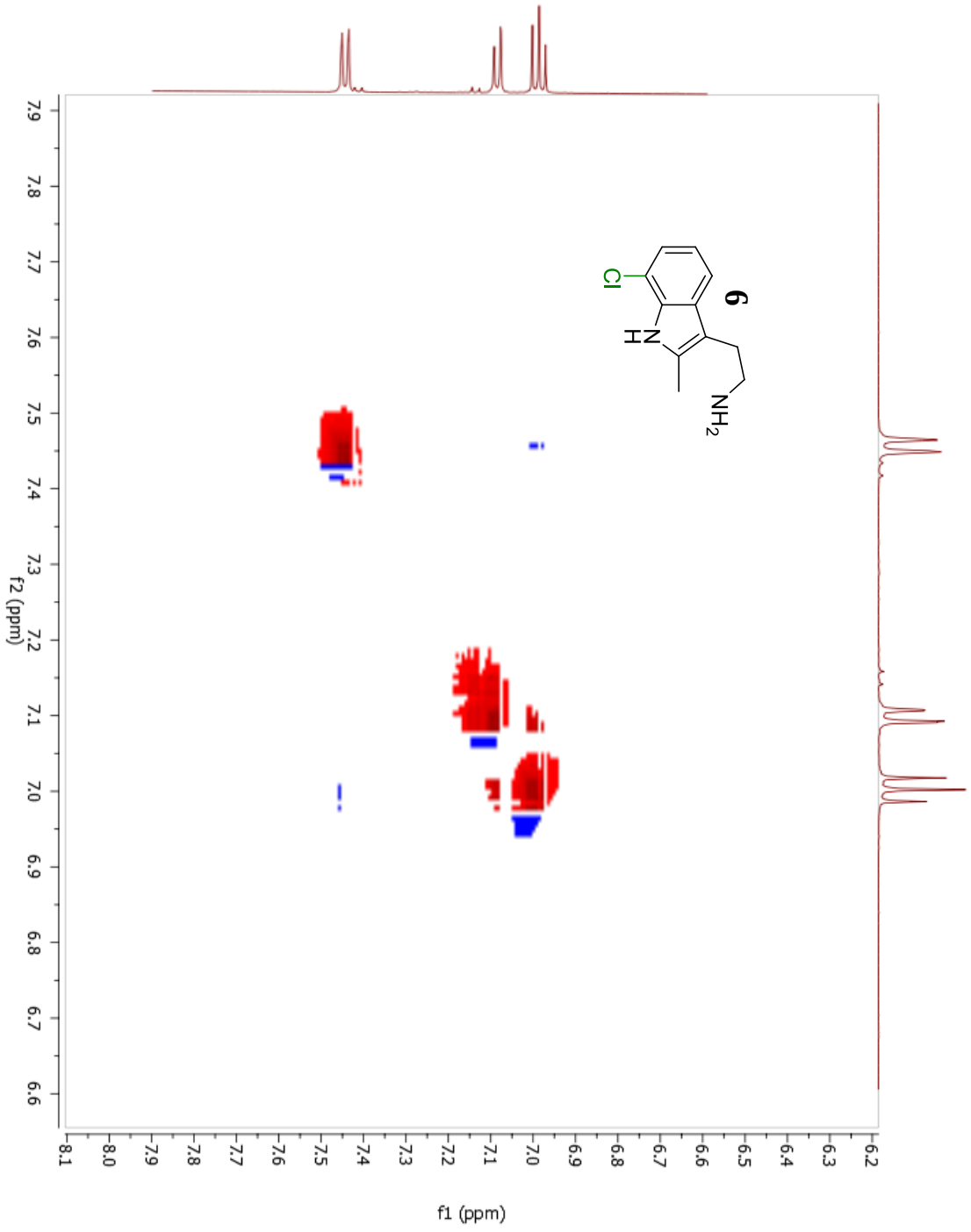


Figure AI.19. NOESY spectrum of 6.[a]



[a] The resonance between peaks at 7.45 and 2.79 ppm demonstrate chlorination at the 7 position.

Figure AI.20. Closeup of aryl region of NOESY spectrum of 6.[a]



[a] Resonances between peaks at 7.45 and 7.00 ppm, and between peaks at 7.09 and 7.00 ppm support chlorination at the 7 position.

Figure AI.21. ^{1}H NMR spectrum of 7.

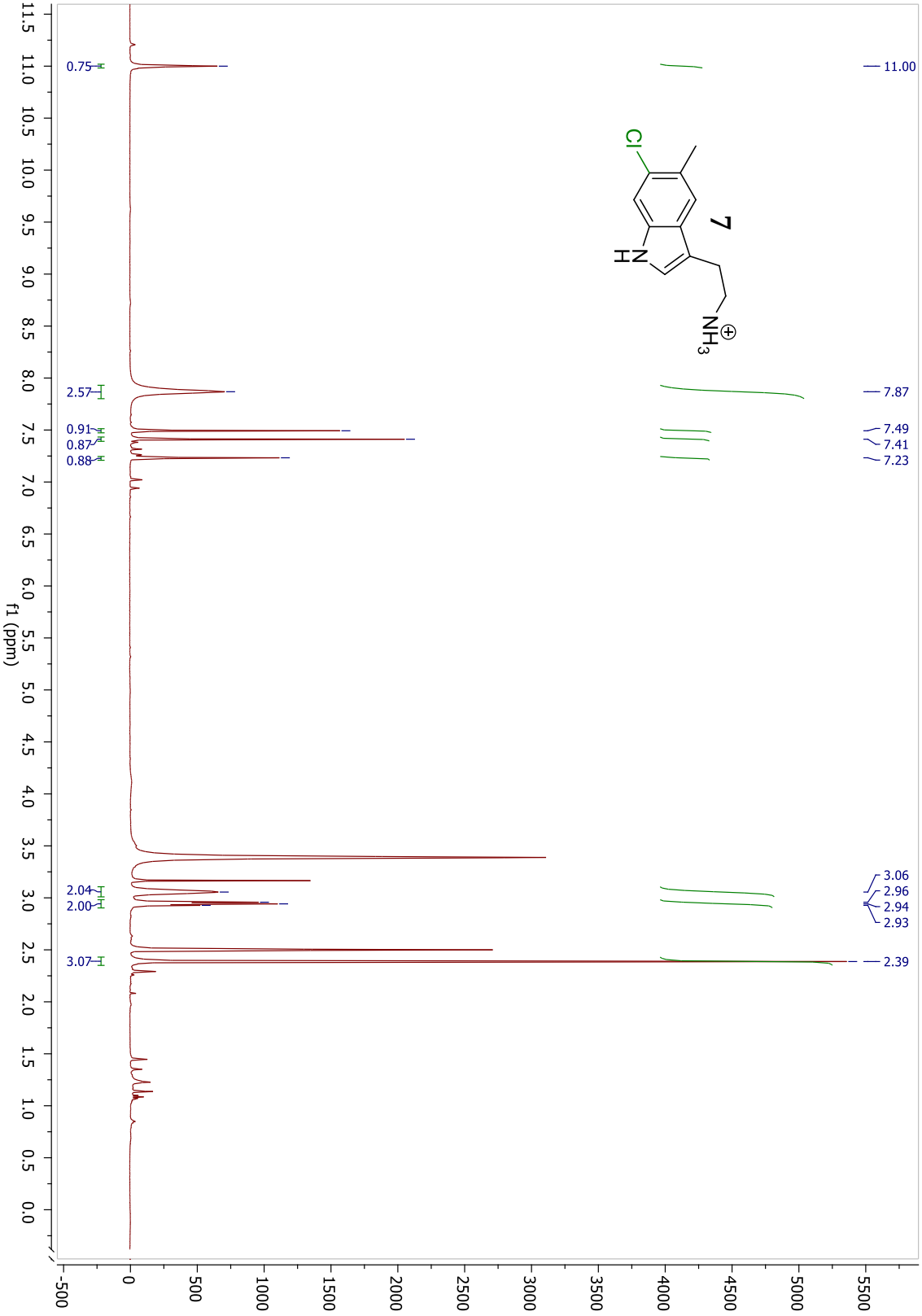


Figure AI.22. ^{13}C NMR spectrum of 7.

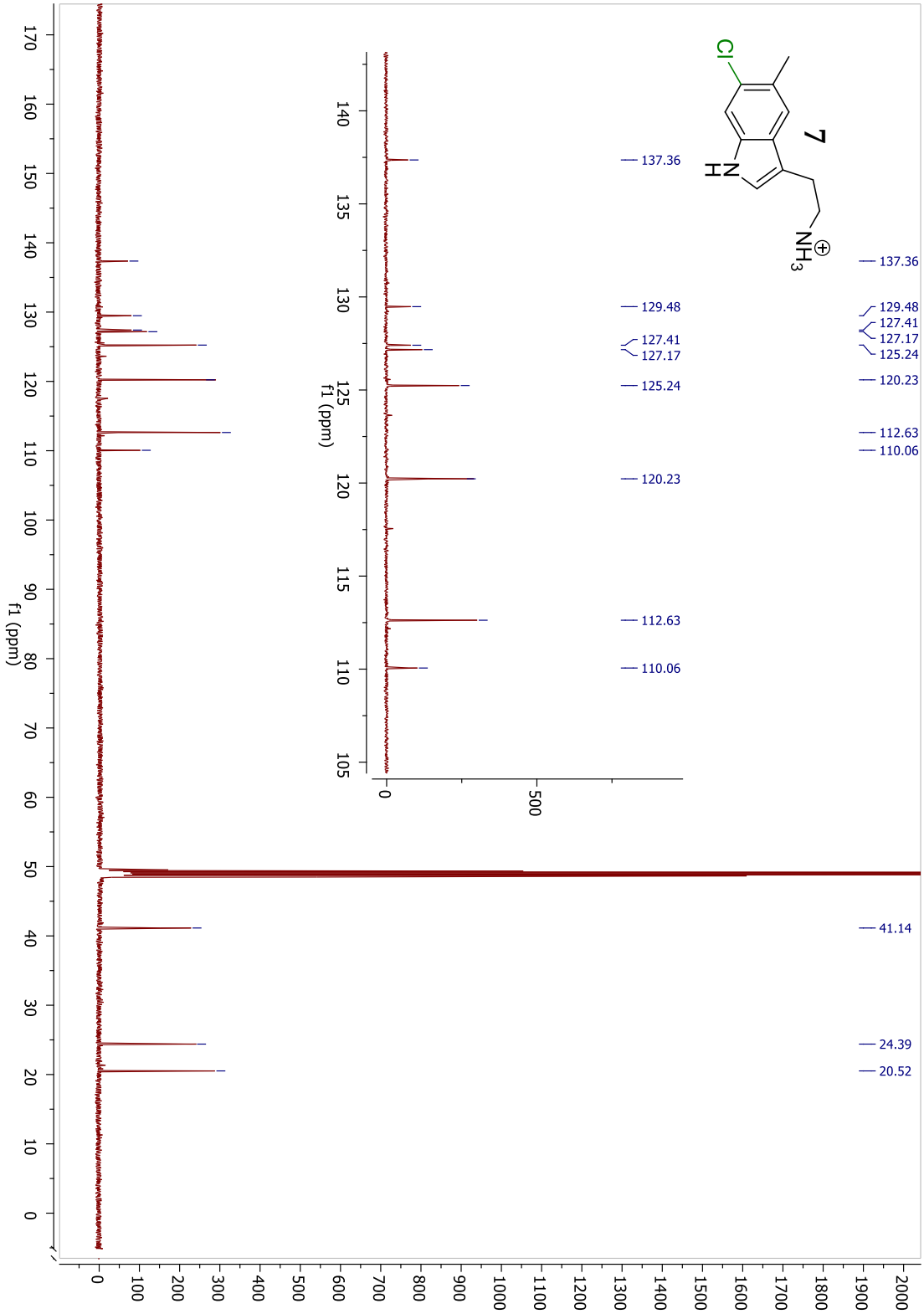
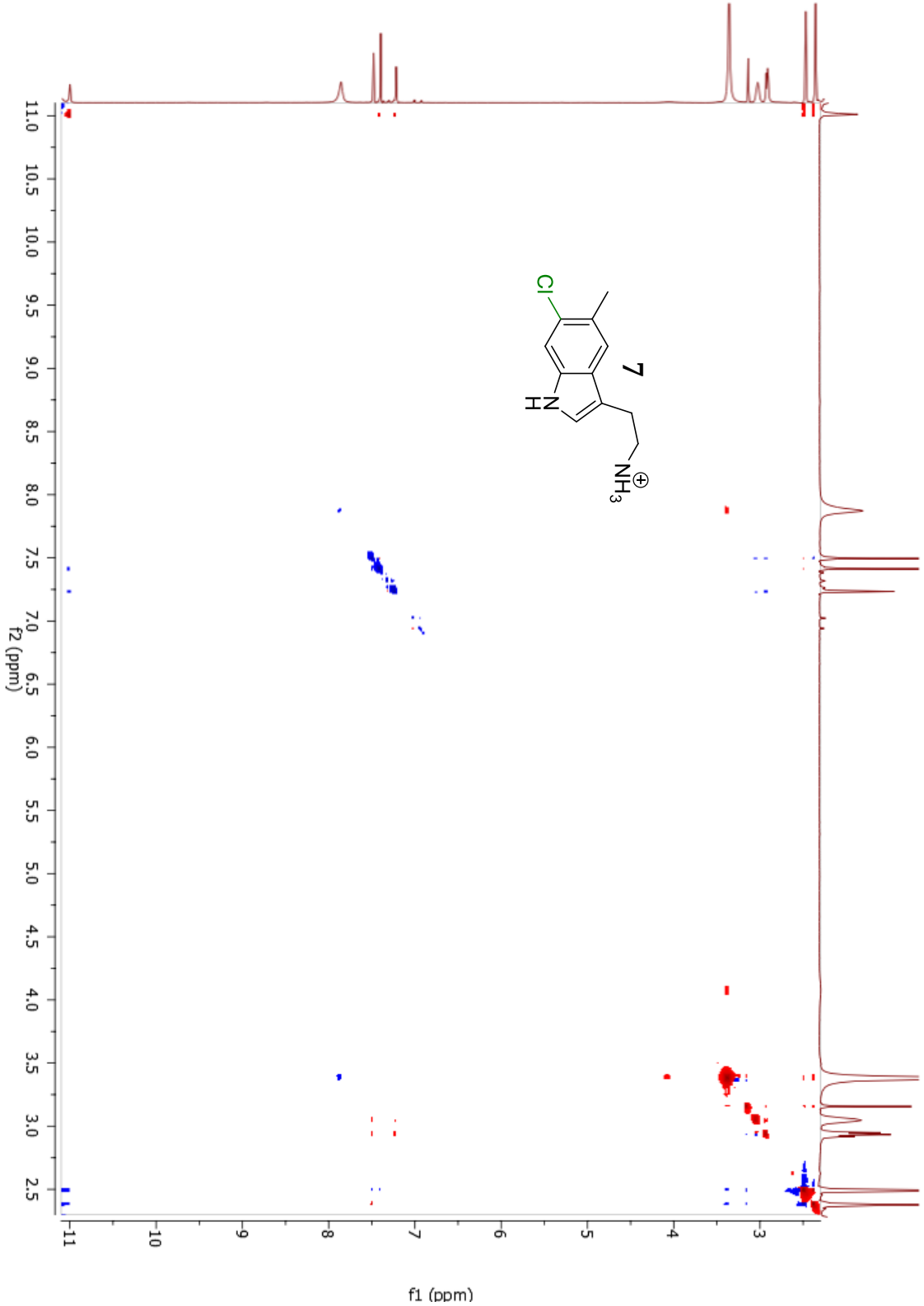
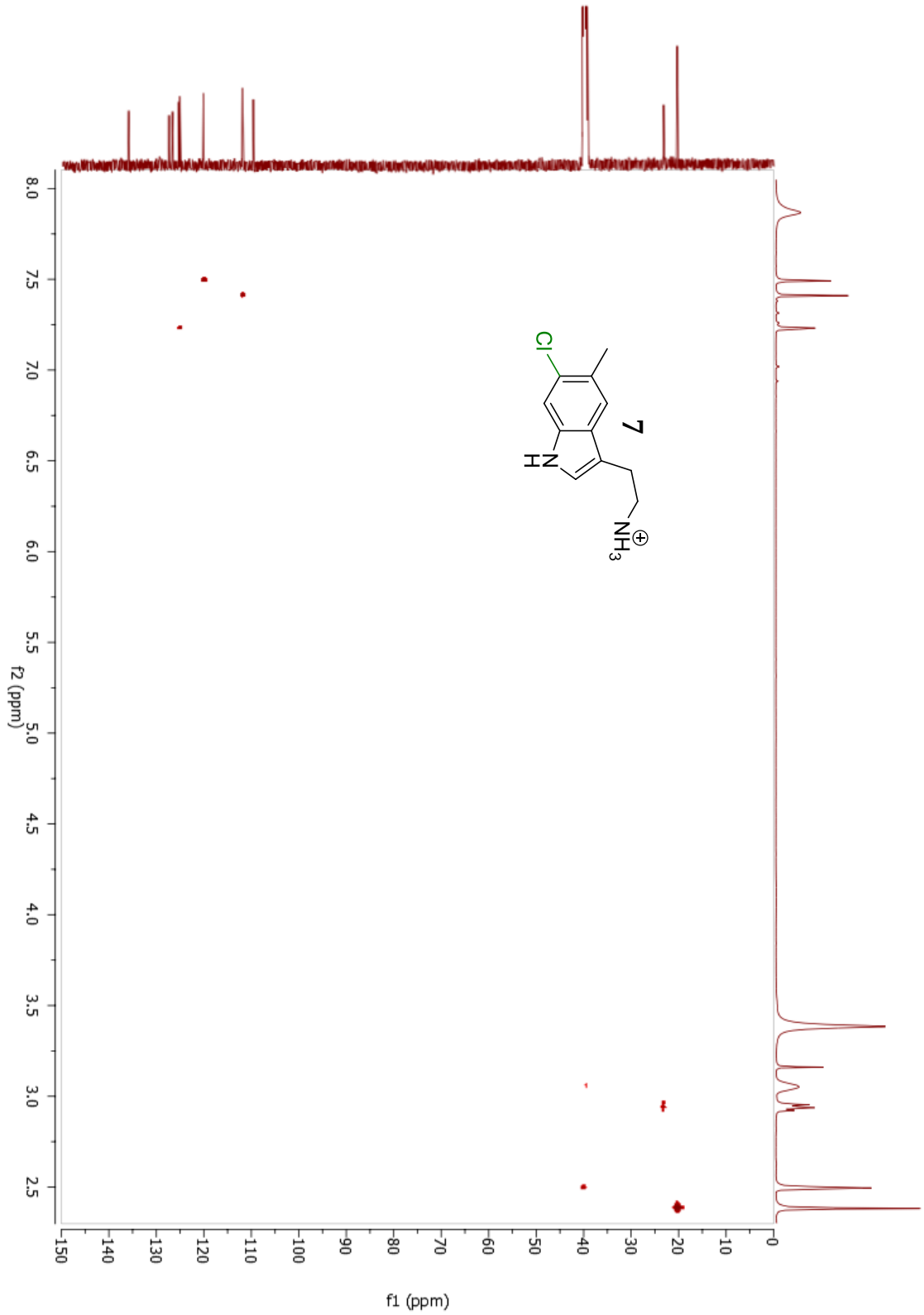


Figure AI.23. NOESY spectrum of 7.[a]



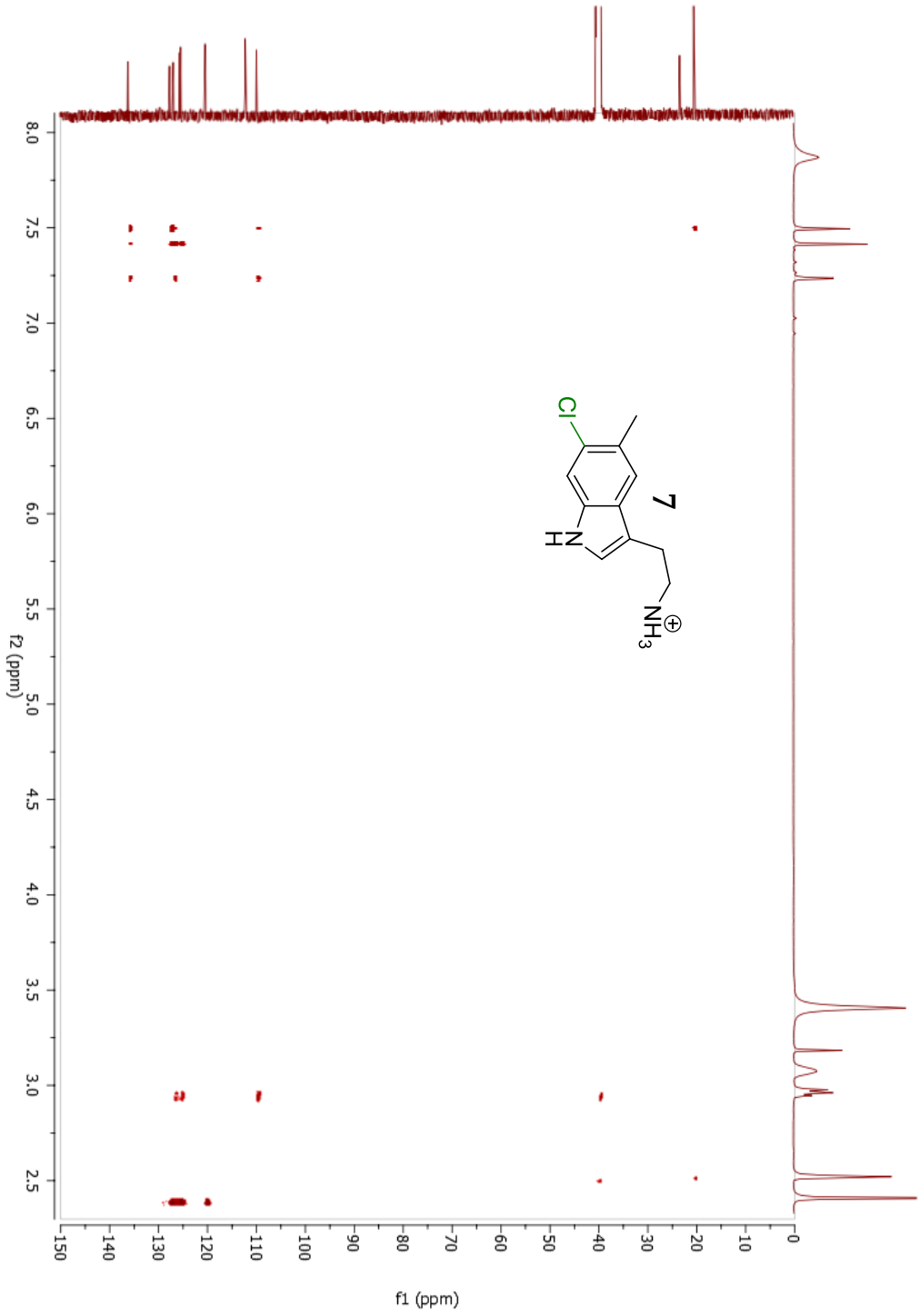
[a] Resonances between peaks at 7.50 and 2.95 ppm, between peaks at 7.24 and 2.95 ppm, and between peaks at 11.01 and 7.42 ppm support chlorination at the 6 position.

Figure AI.24. HMQC spectrum of 7.[a]



[a] The resonance between peaks at 2.40 and 20.54 ppm identifies the 5-methyl carbon, which is necessary to analyze the HMBC spectrum.

Figure AI.25. HMBC spectrum of 7.[a]



[a] The resonance between peaks at 7.50 and 20.54 ppm confirms the identity of the proton on C4. The HMBC of the starting material shows a clear resonance between the 5-methyl carbon and the proton at C6. This resonance does not appear in this product, thus supporting chlorination at the 6 position.

Figure AI.26. ^{1}H NMR spectrum of 8a/b.

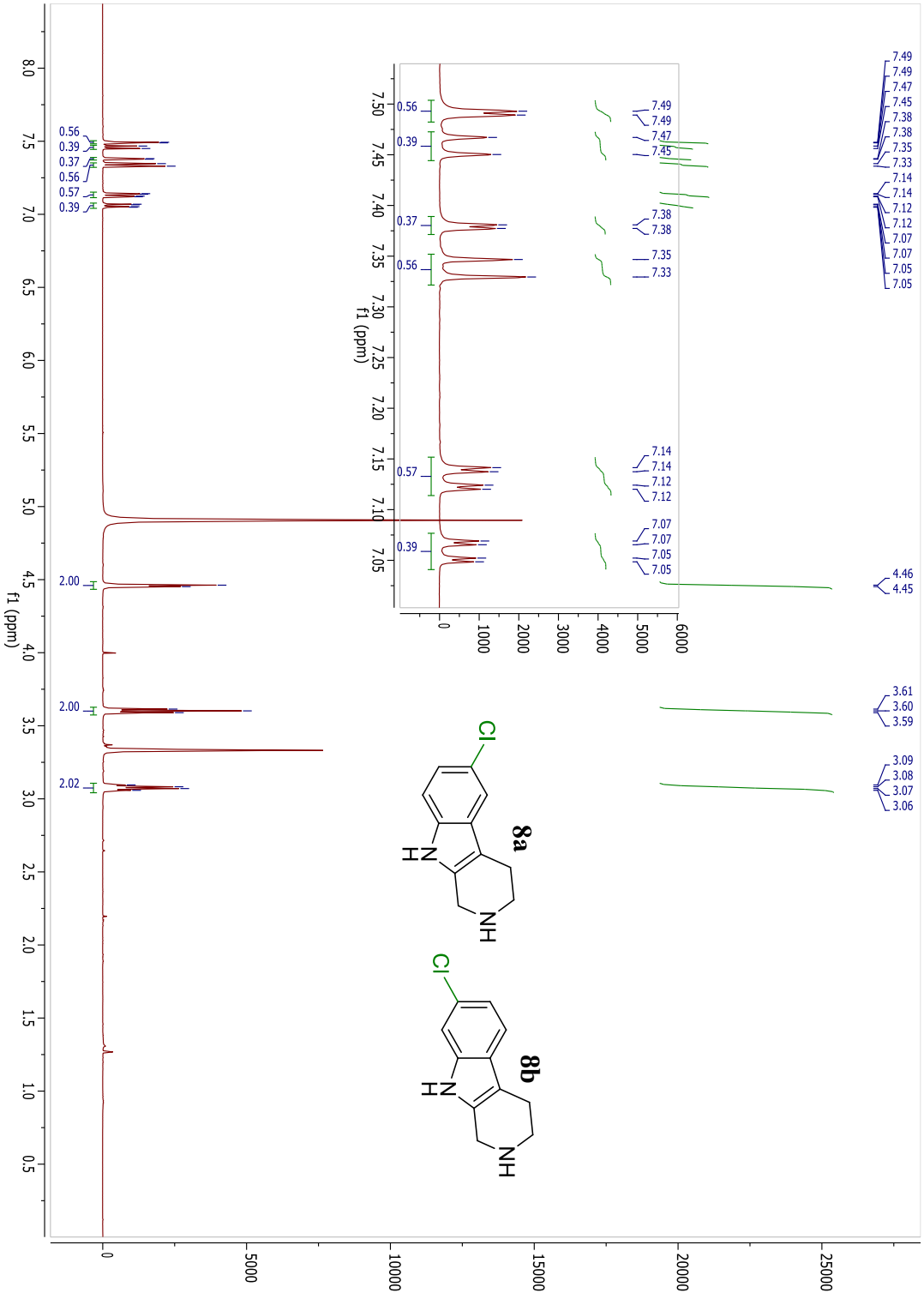


Figure AI.27. ^{13}C NMR spectrum of 8a/b.

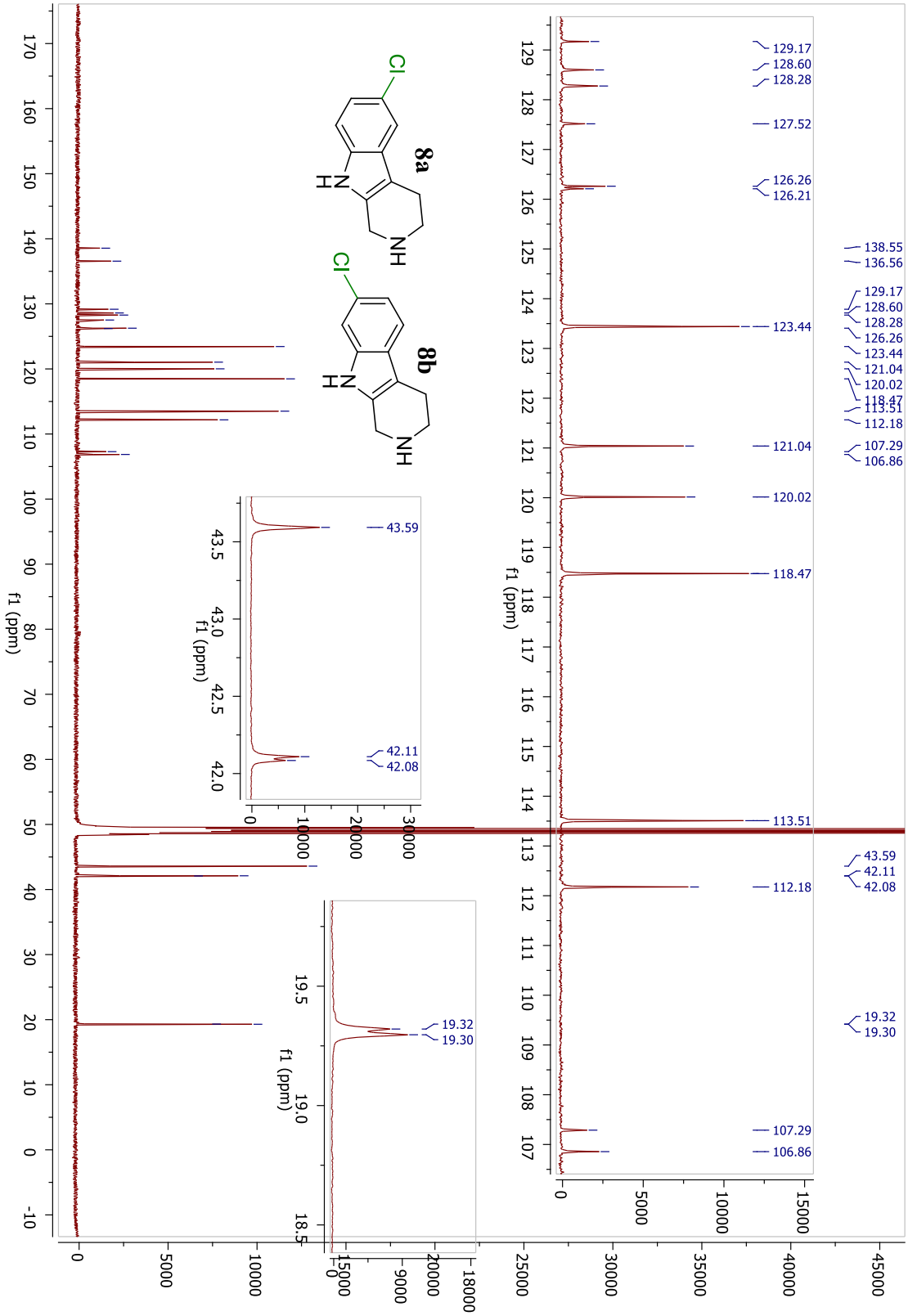


Figure AI.28. ^{1}H NMR spectrum of 9.

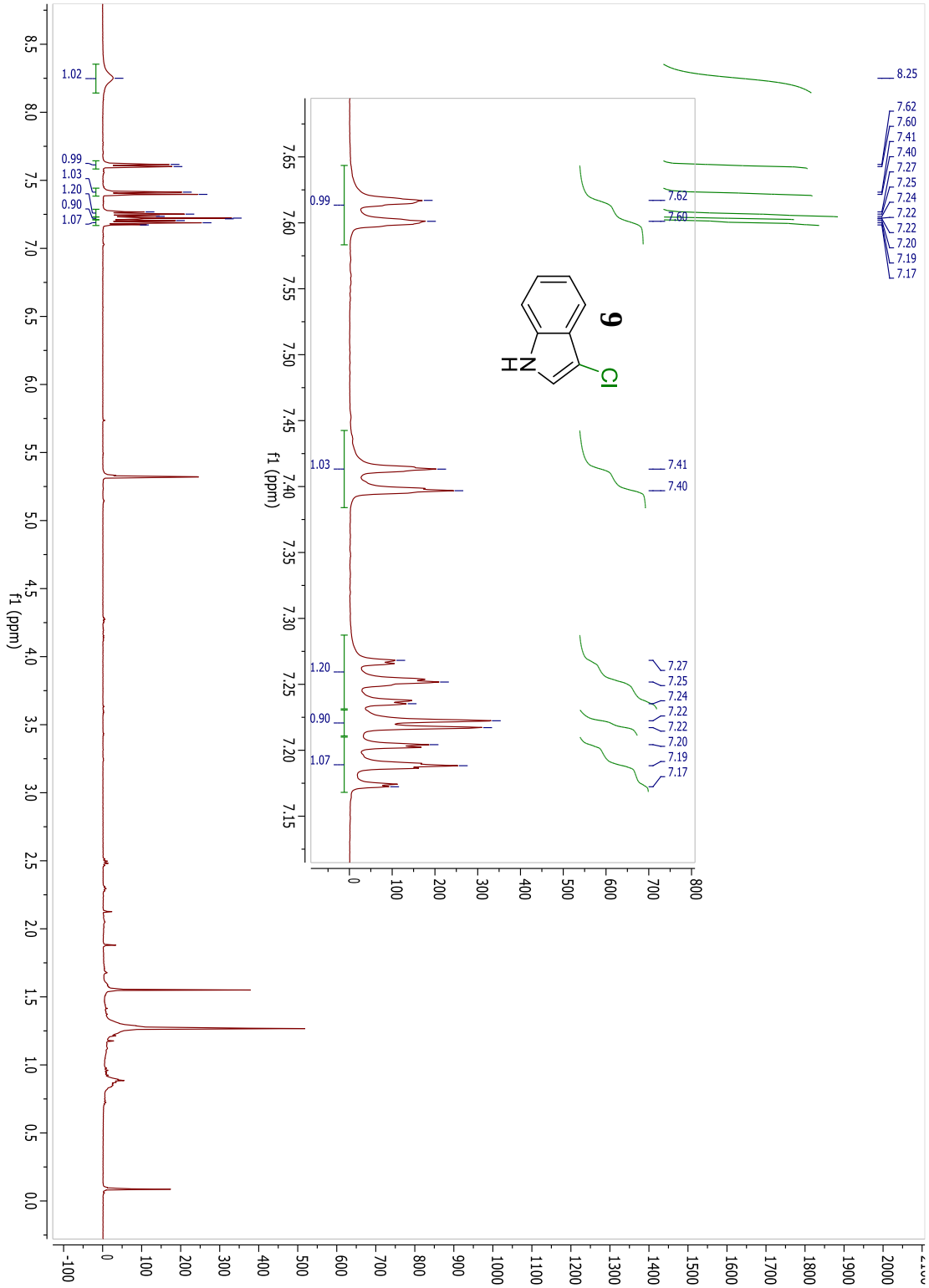


Figure AI.29. ^{13}C NMR spectrum of 9.

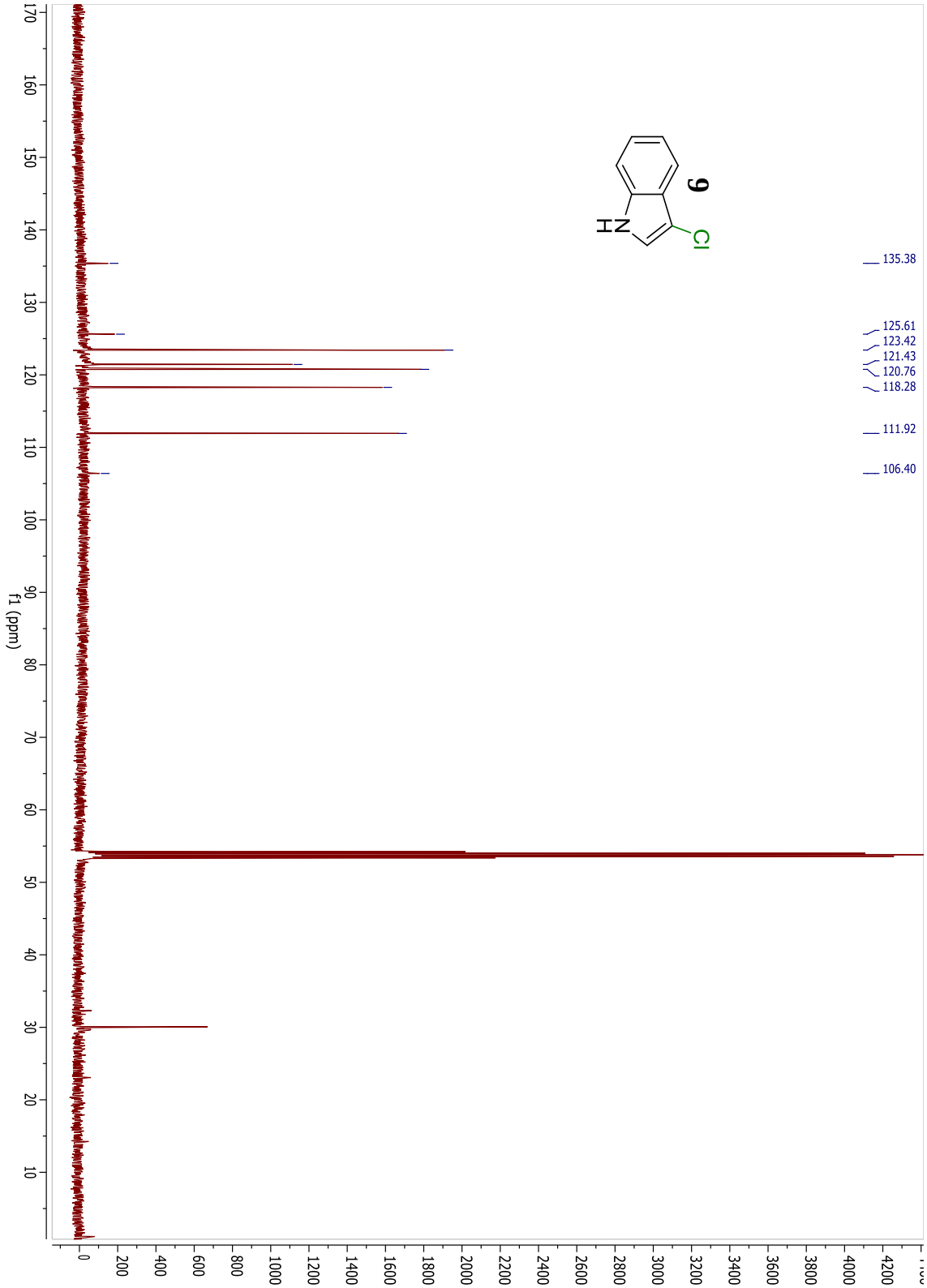


Figure AI.30. ^{1}H NMR spectrum of 10.

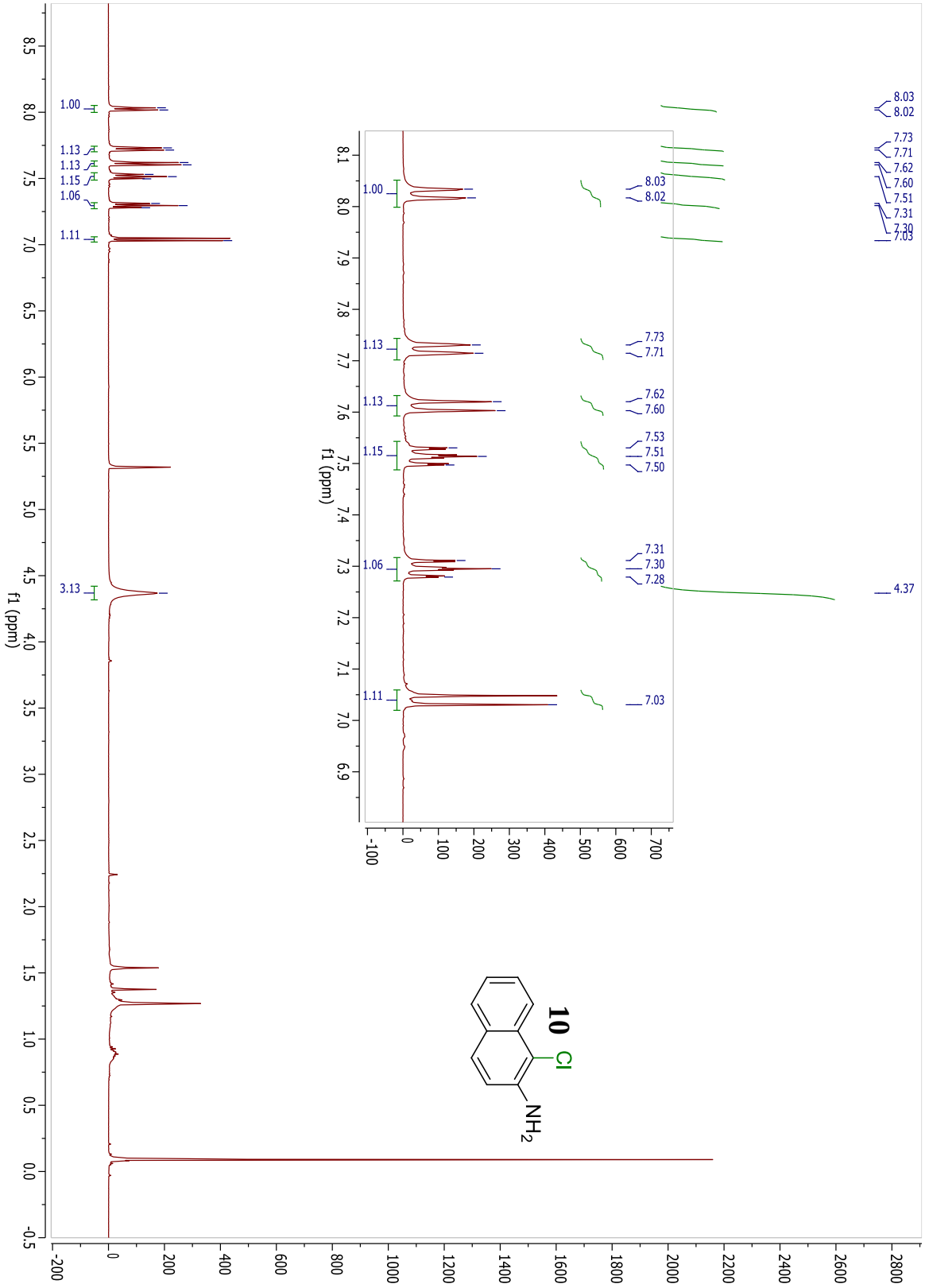


Figure AI.31. ^{13}C NMR spectrum of 10.

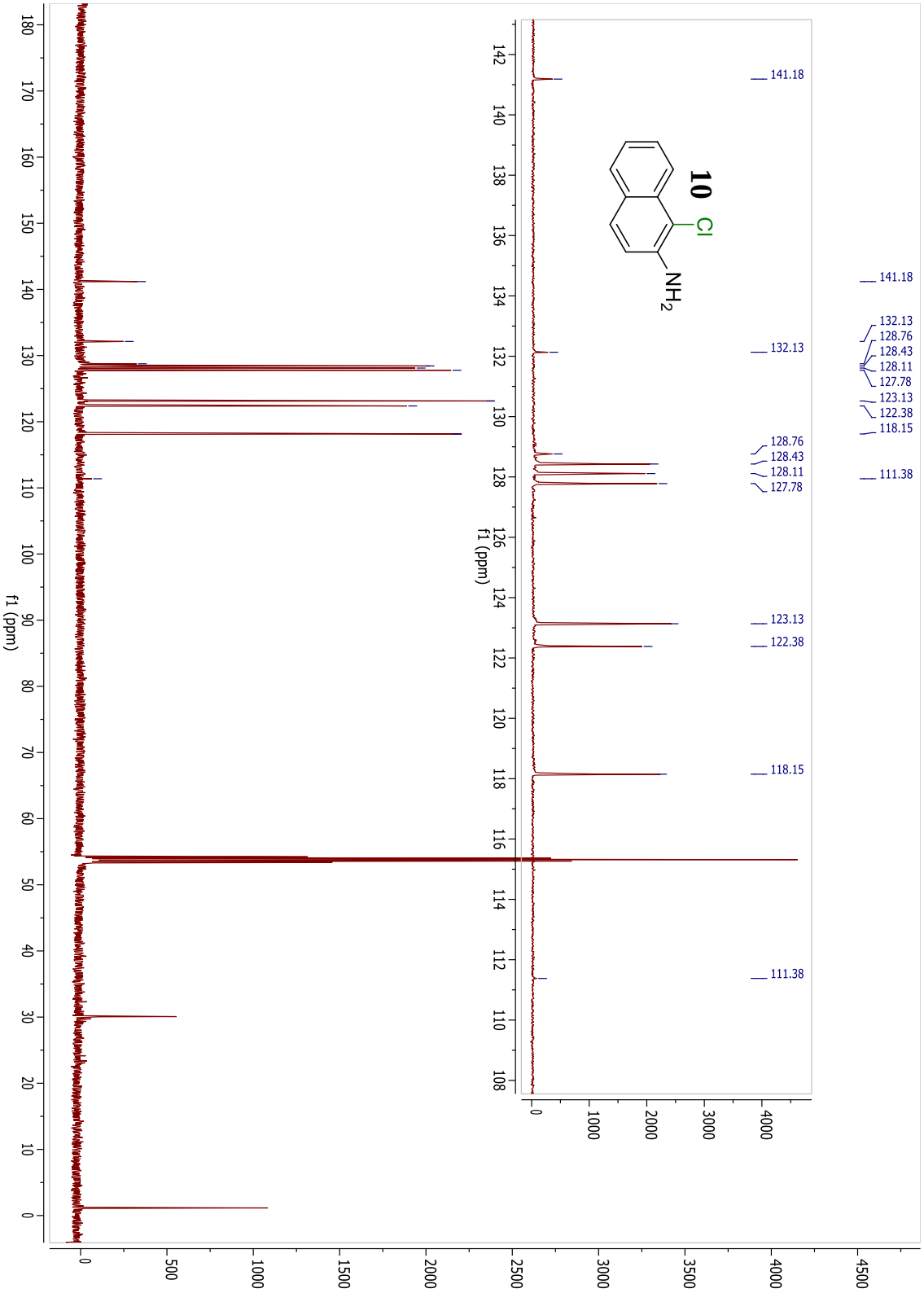


Figure AI.32. NOESY spectrum of 10.

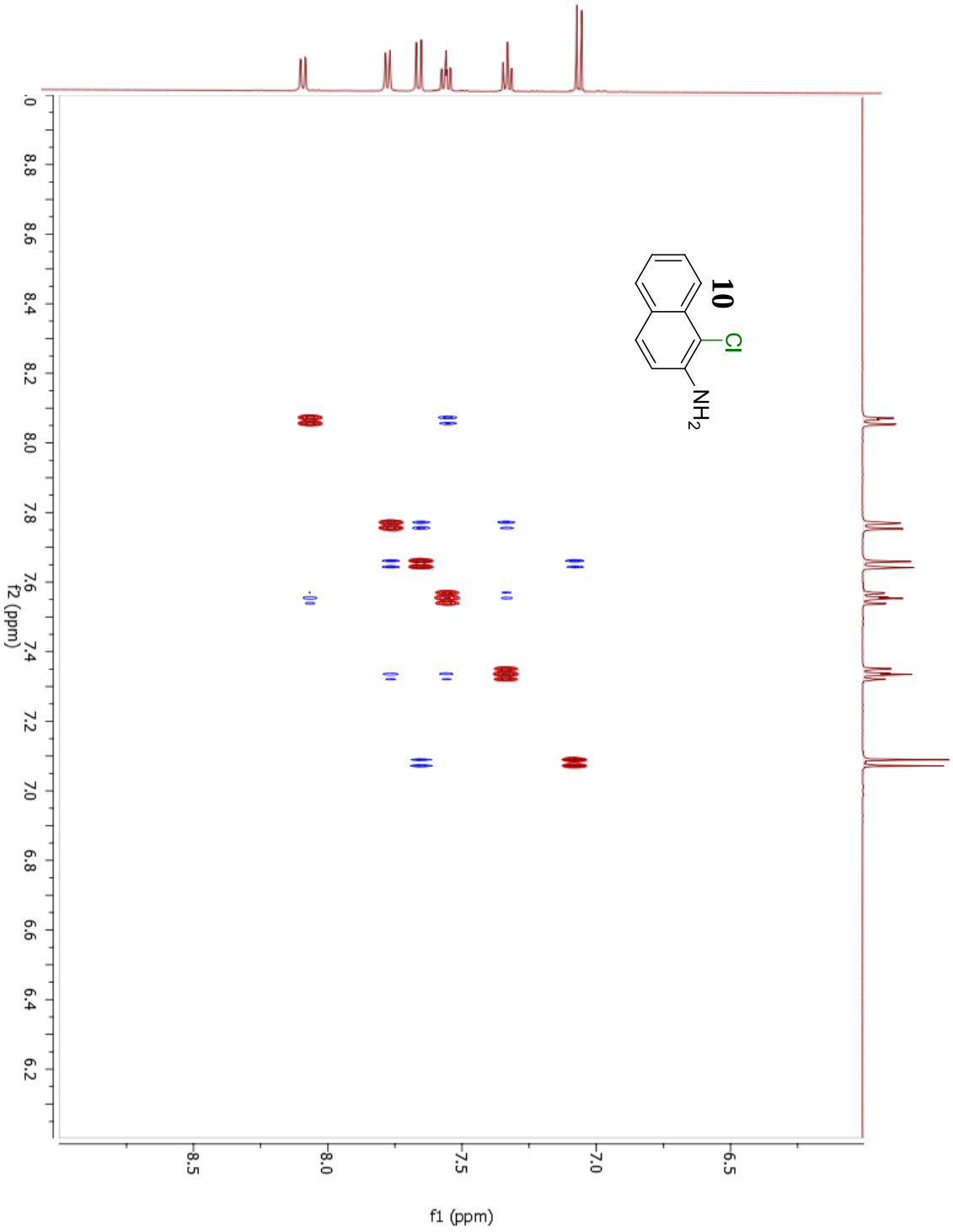


Figure AI.33. ^{1}H NMR spectrum of 11.

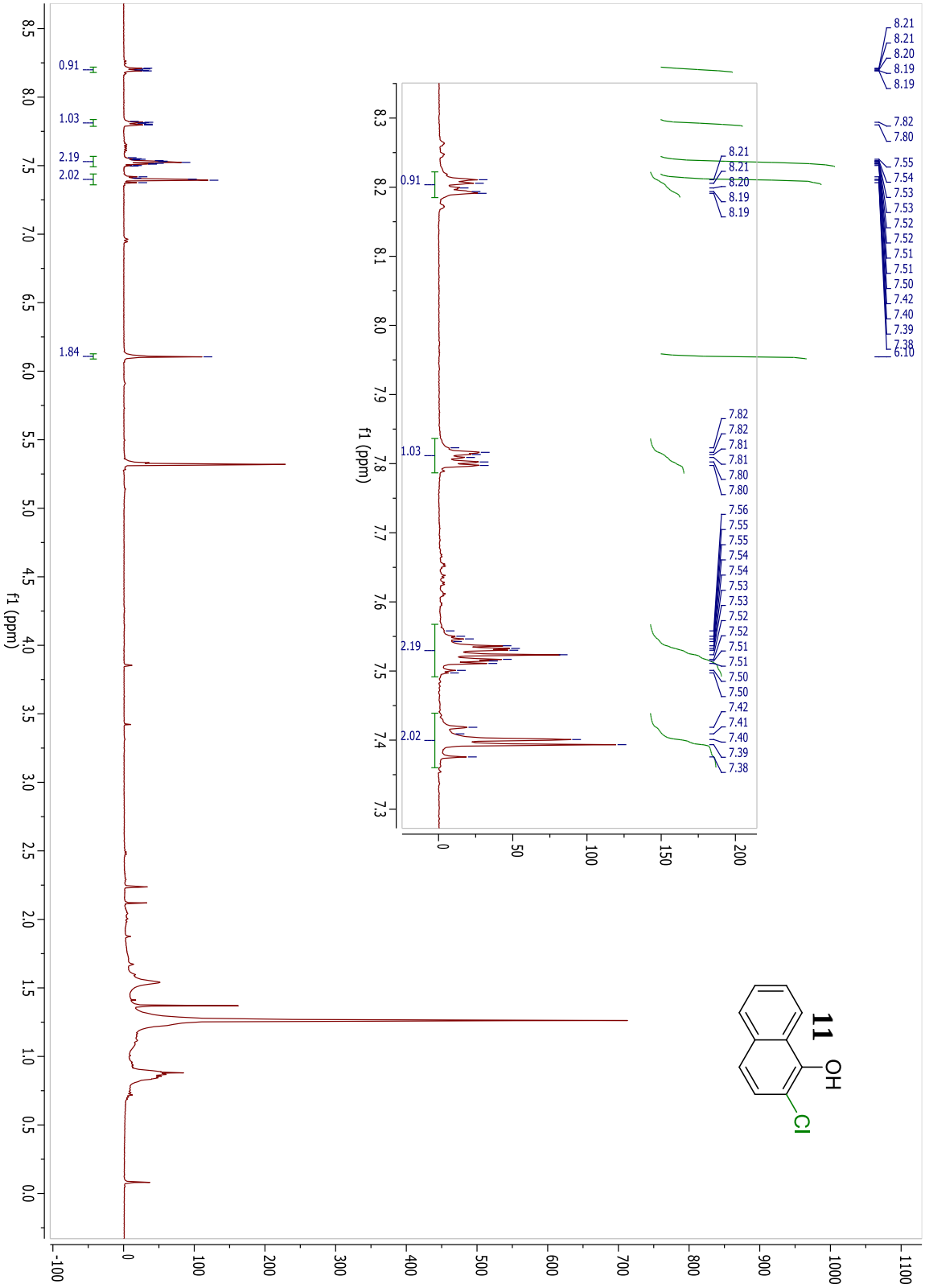


Figure AI.34. ^{13}C NMR spectrum of 11.

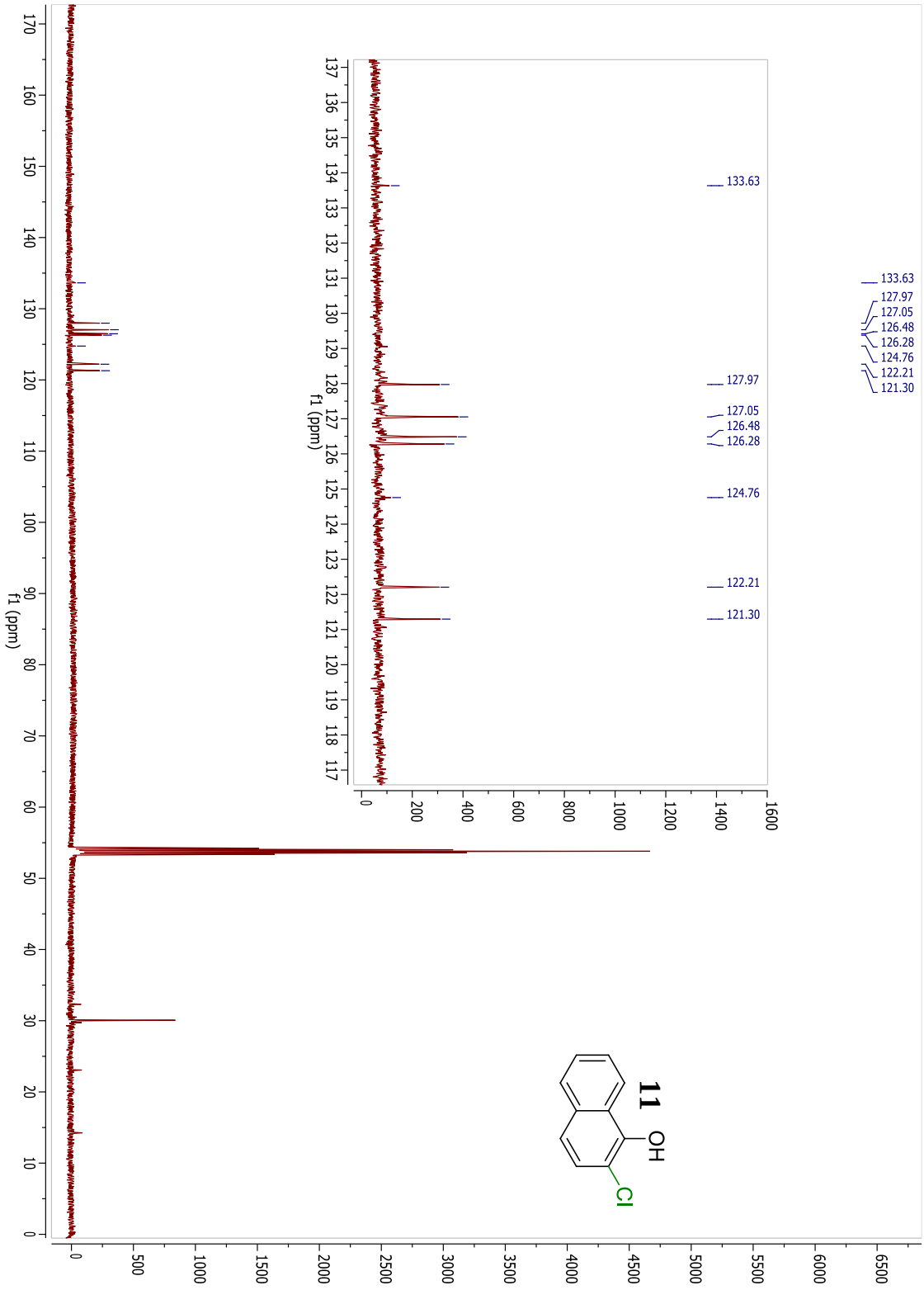


Figure AI.35. ^{1}H NMR spectrum of 12.

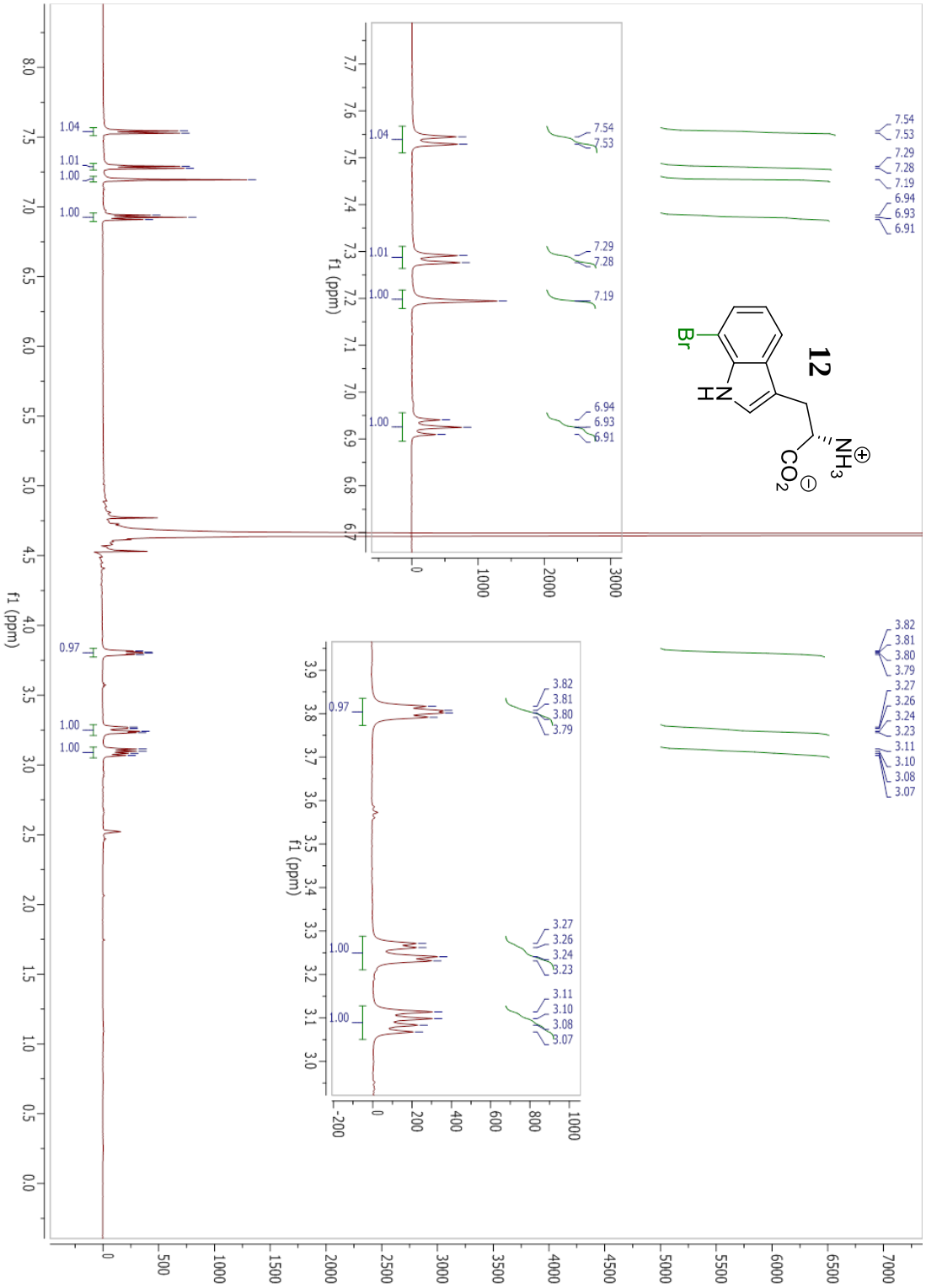


Figure AI.36. ^{13}C NMR spectrum of 12.

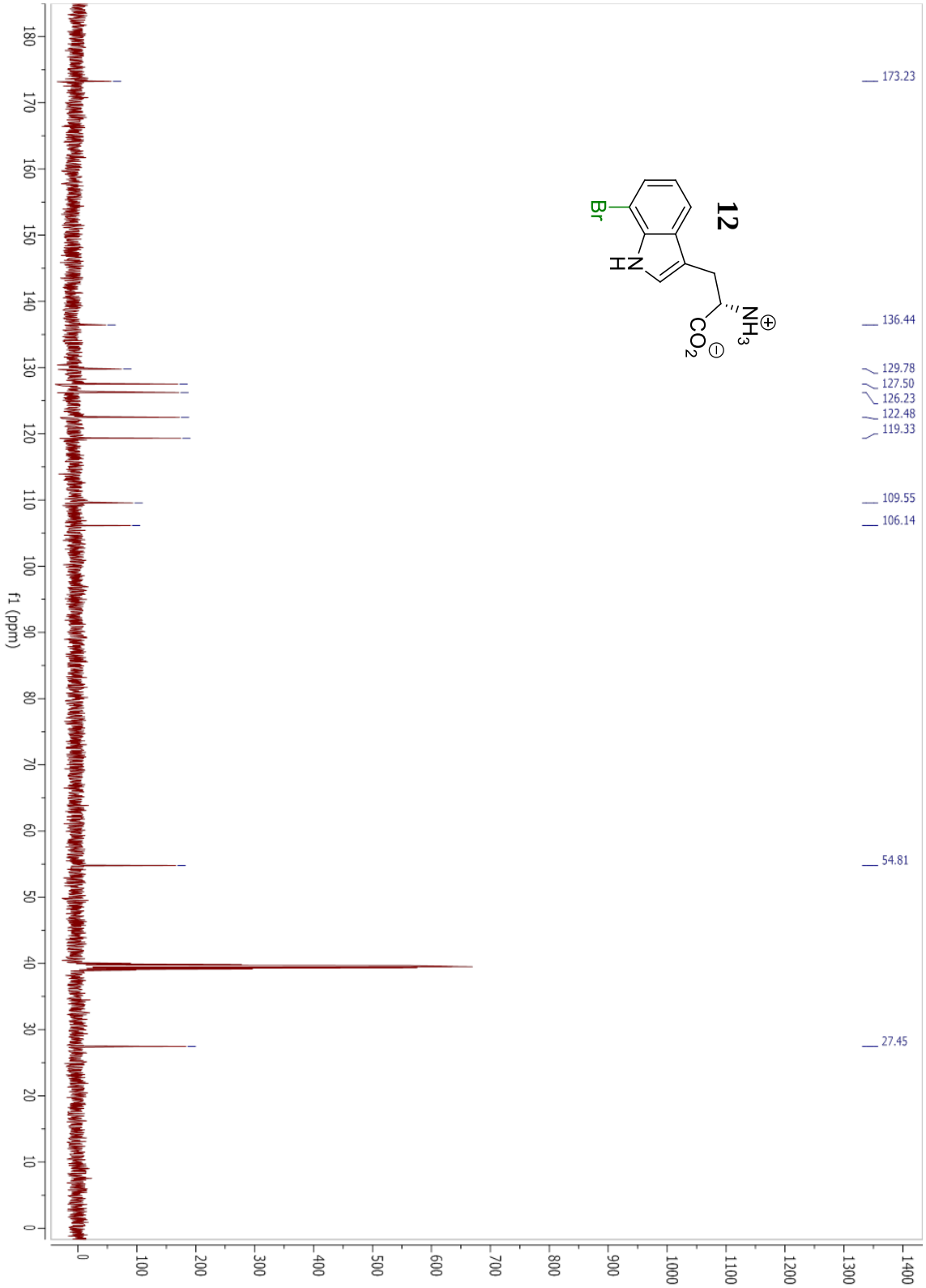
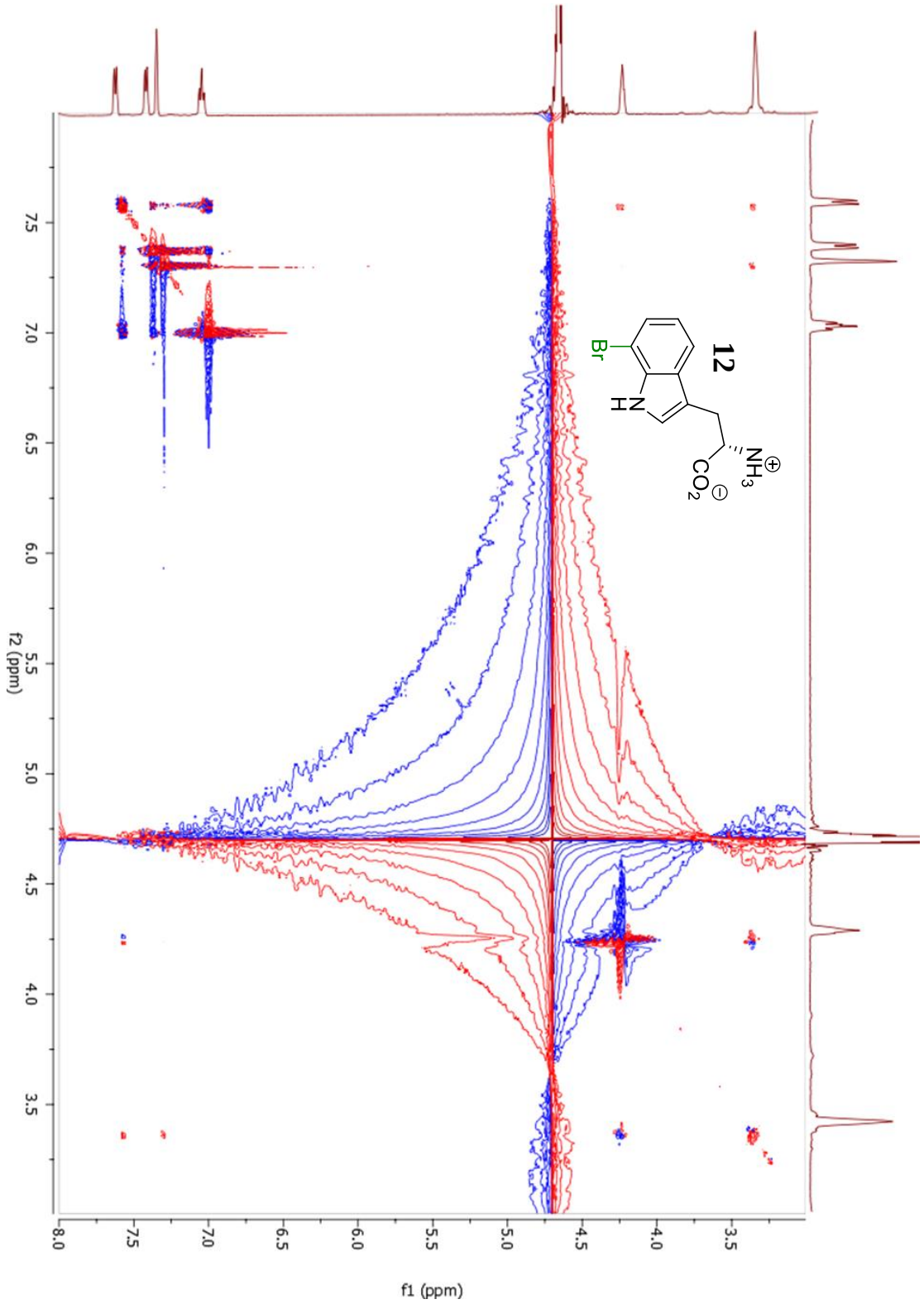


Figure AI.37. NOESY spectrum of 12.[a]



[a] Resonances between peaks at 7.19 and 3.42 ppm, and between peaks at 7.54 and 3.42 ppm demonstrate chlorination at the 7 position.

APPENDIX II

NMR SPECTRA FOR COMPOUNDS FROM CHAPTER III

Figure AII.1. ^{1}H NMR spectrum for 5.

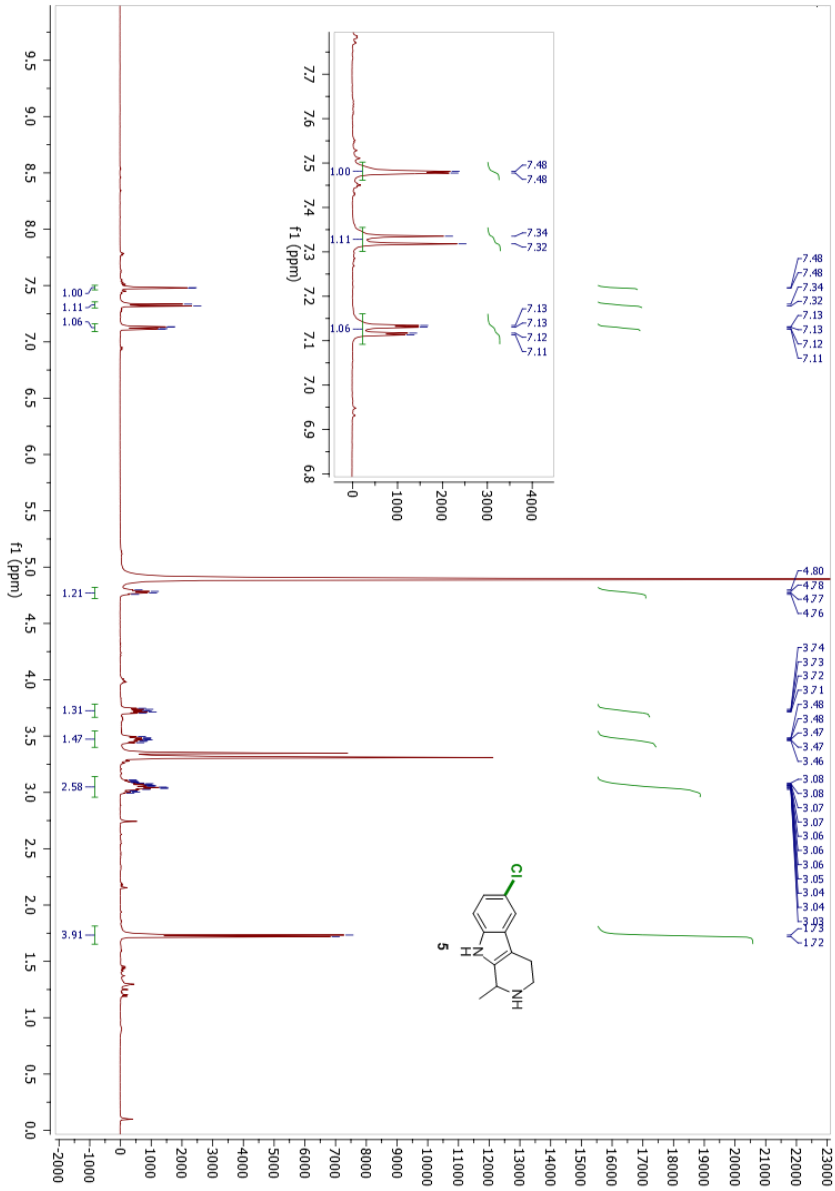


Figure AII.2. ^{13}C NMR spectrum for 5.

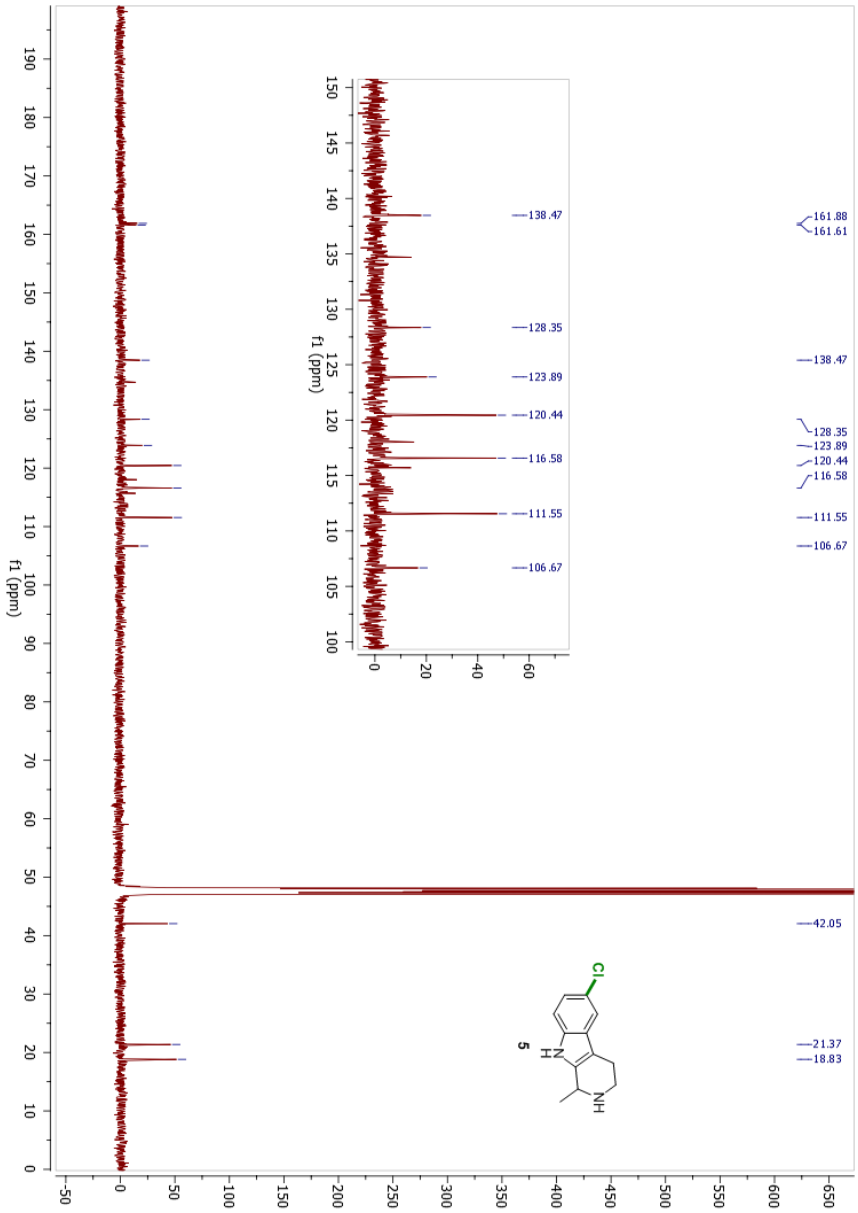
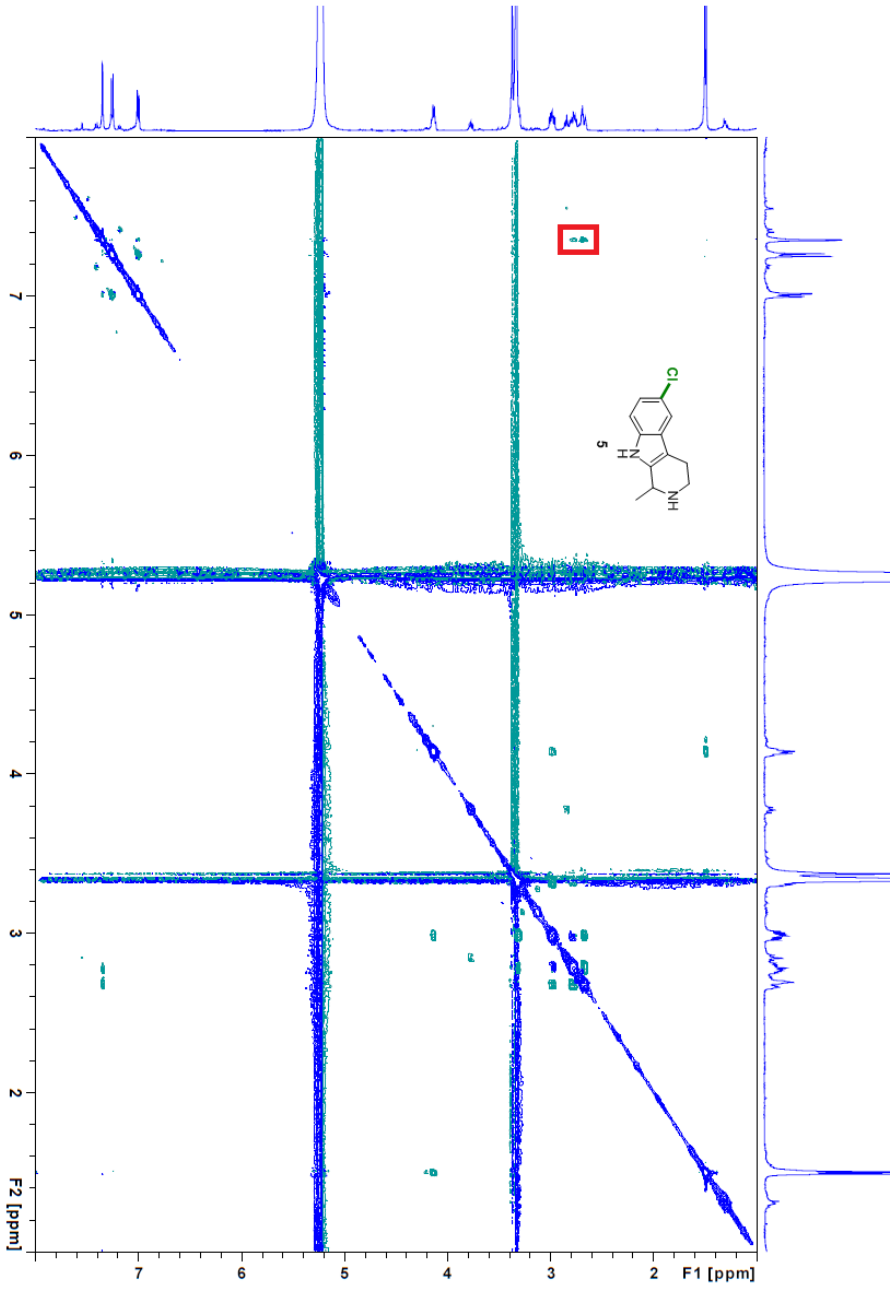


Figure AII.3. NOESY spectrum for 5.[a]



[a] Cross-peak between peaks at ~7.4 ppm (carbazole 5-position) and ~2.9 ppm (carbazole 4-position) demonstrates chlorination at the carbazole 6-position.

Figure AII.4. ^{1}H NMR spectrum for 6.

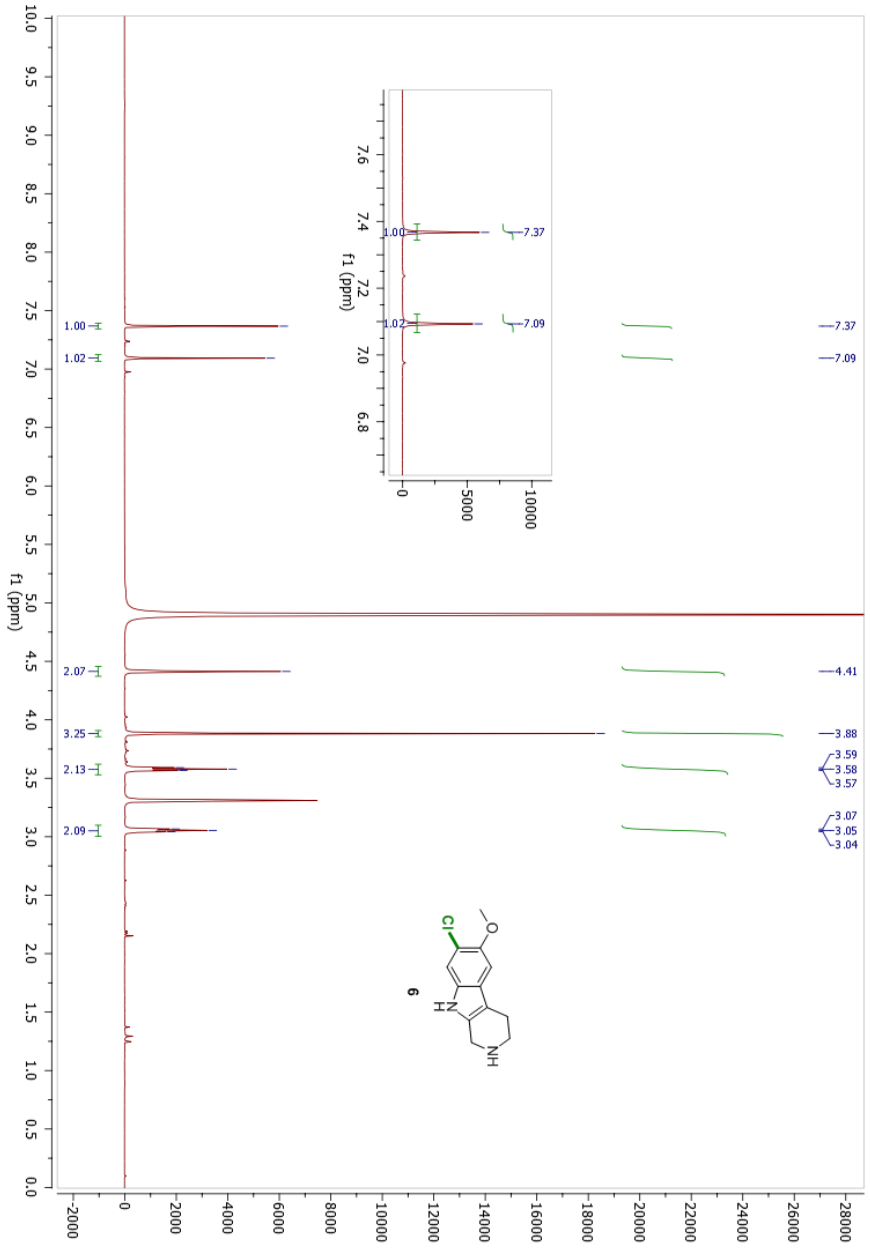


Figure AII.5. ^{13}C NMR spectrum for 6.

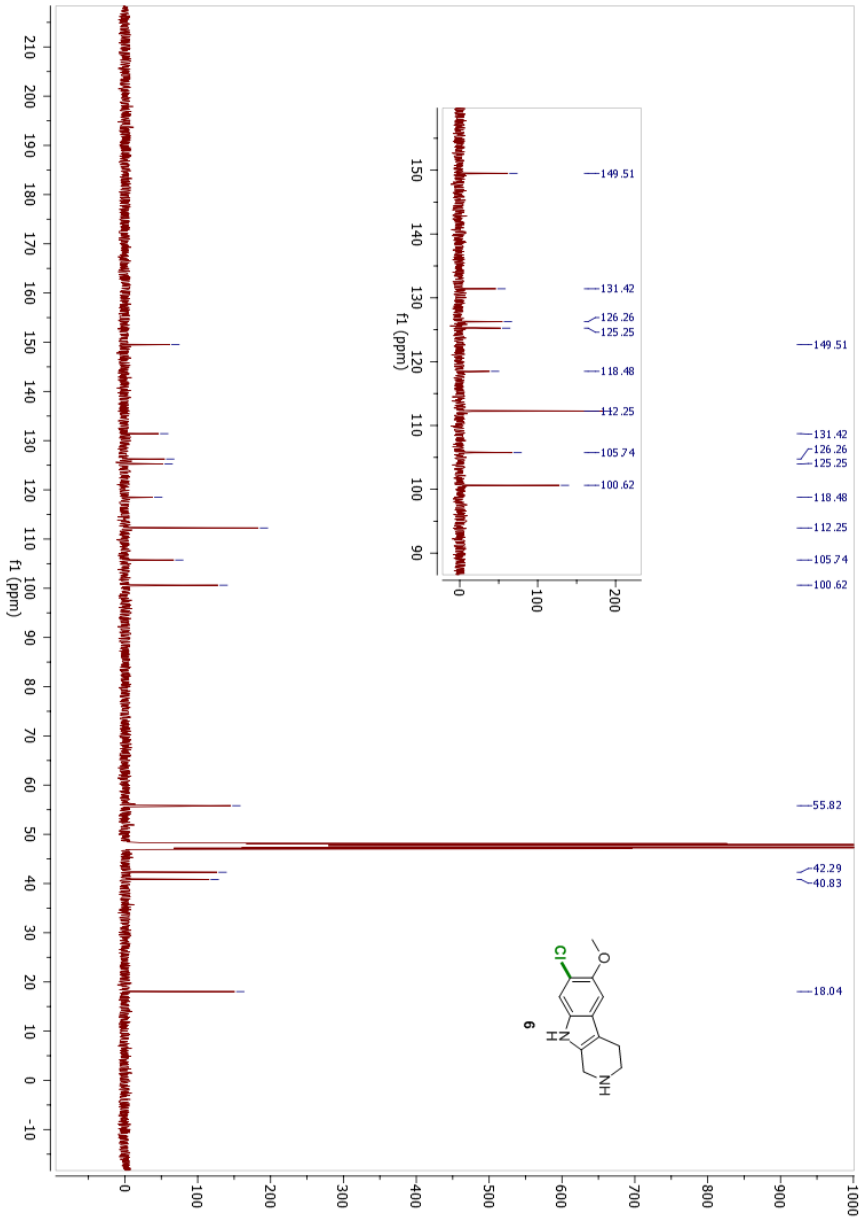
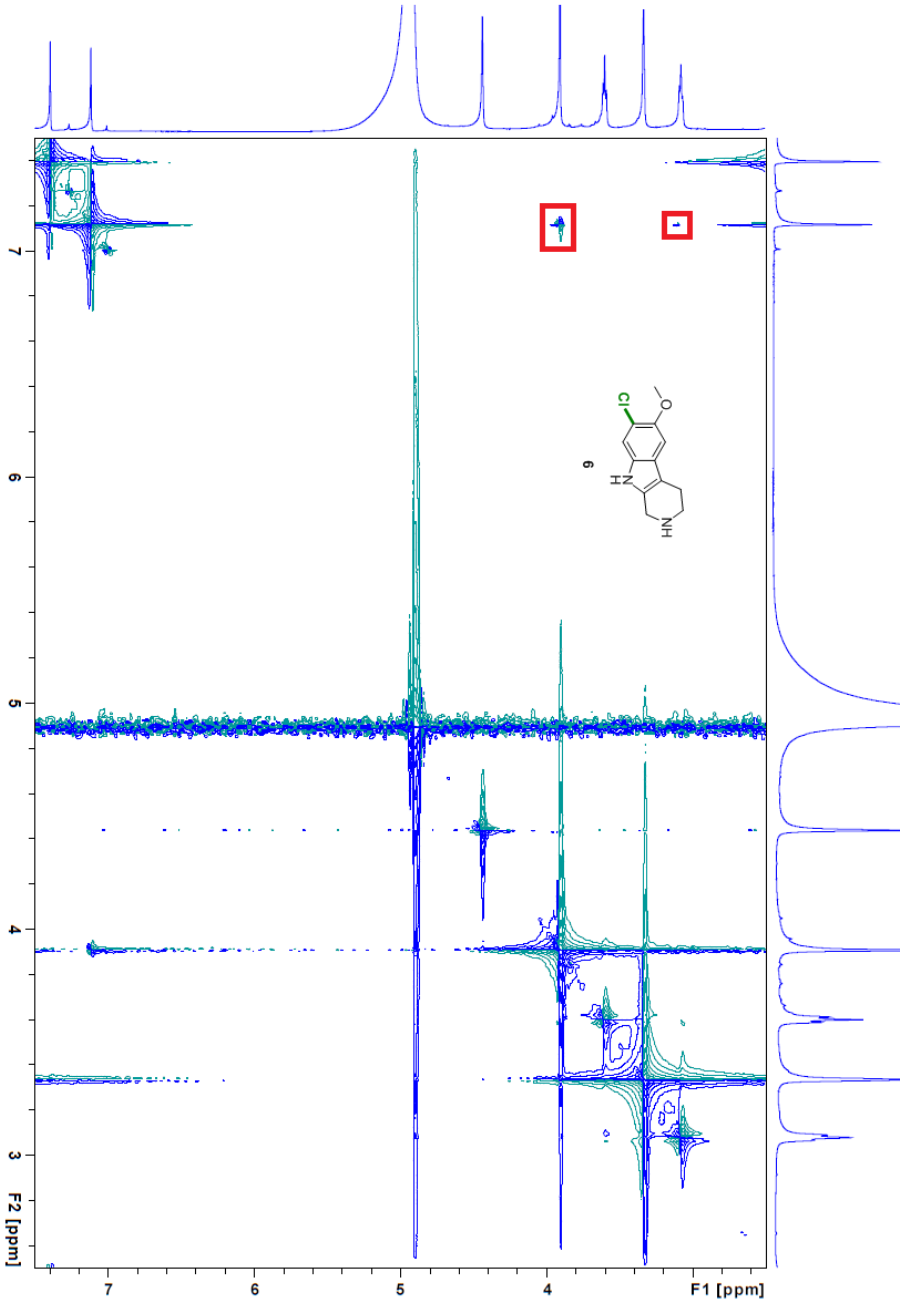


Figure AII.6. NOESY spectrum for 6.[a]



[a] Cross-peak between peaks at ~7.1 ppm (carbazole 5-position) and ~3.1 ppm (carbazole 4-position), along with absence of cross-peak between peaks at ~3.9 ppm (carbazole 6-methoxy protons) and ~7.4 ppm (which we therefore conclude to be the carbazole 8-position), demonstrates chlorination at the carbazole 7-position.

Figure AII.7. ^{1}H NMR spectrum for 7.

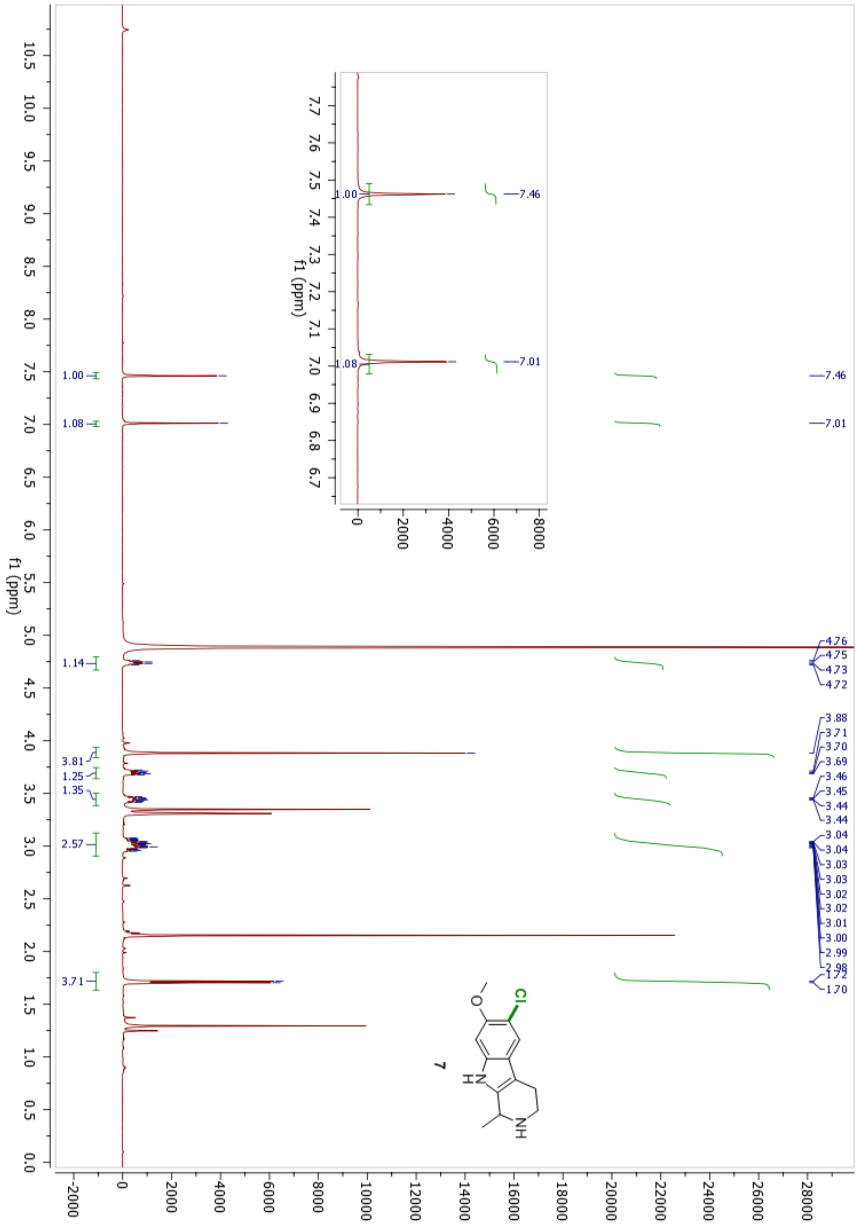


Figure AII.8. ^{13}C NMR spectrum for 7.

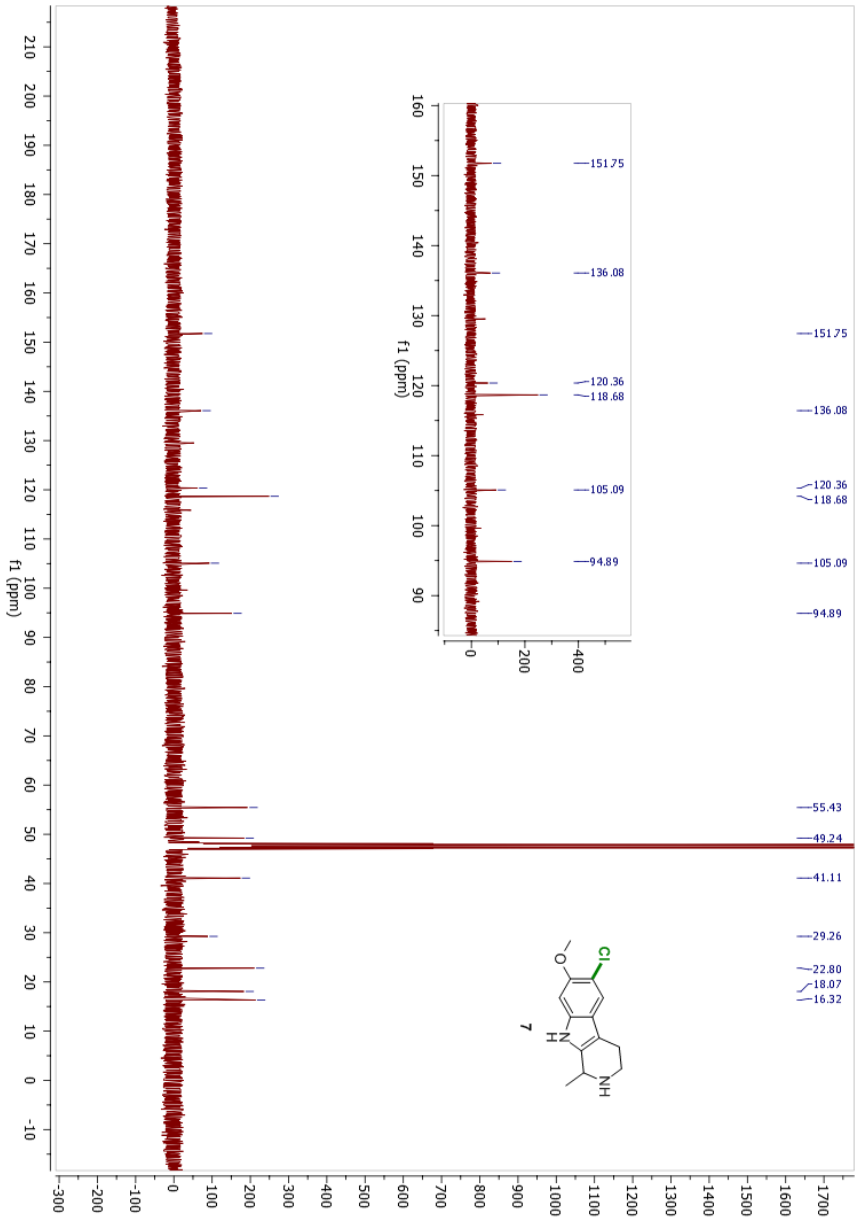
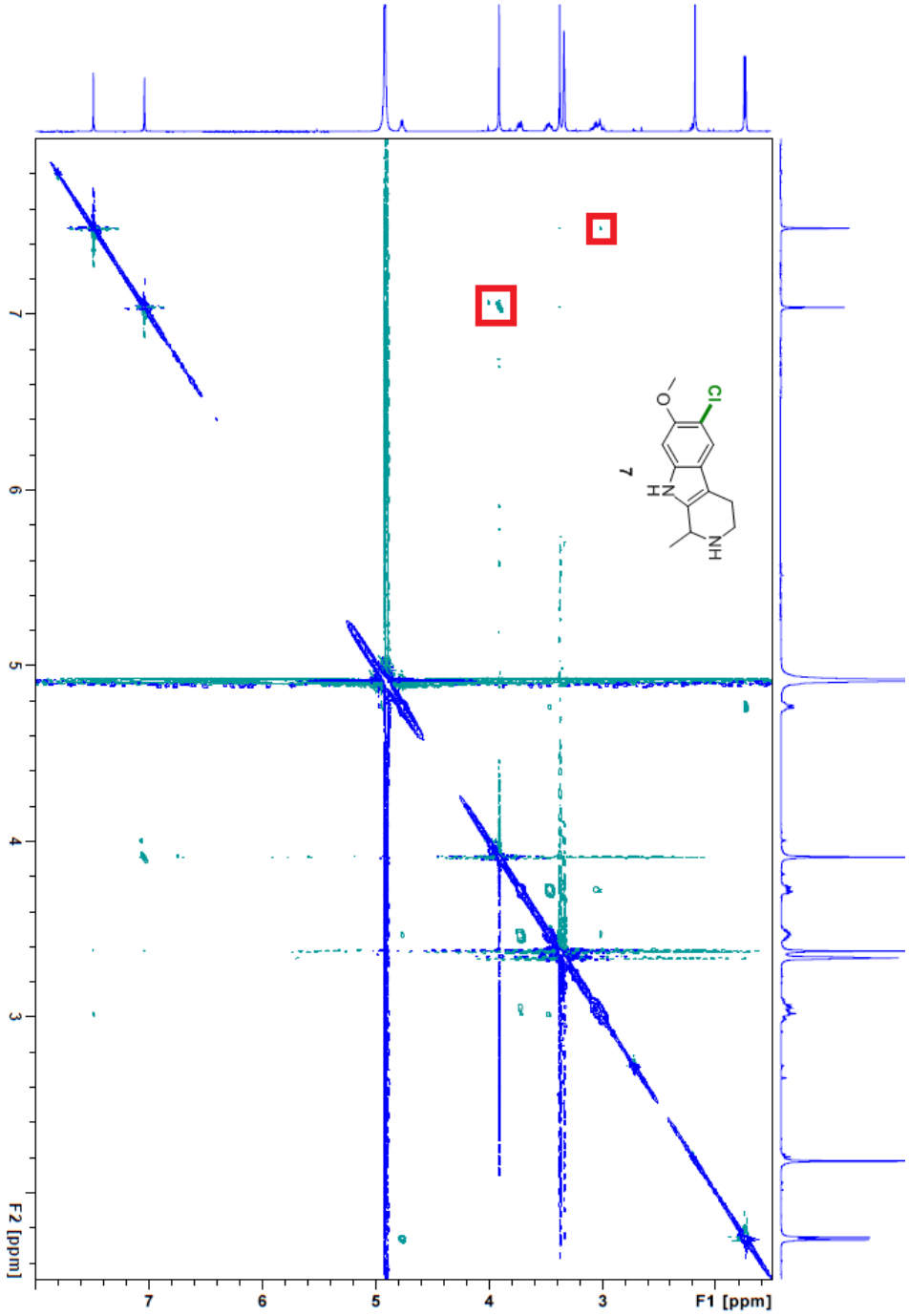


Figure AII.9. NOESY spectrum for 7.[a]



[a] Cross-peak between peaks at ~7.5 ppm (carbazole 5-position) and ~3.0 ppm (carbazole 4-position), along with cross-peak between peaks at ~3.9 ppm (carbazole 6-methoxy protons) and ~7.1 ppm (carbazole 8-position), along with concurrent absence of cross-peak between peaks at ~3.9 ppm (carbazole 6-methoxy protons) and ~7.5 ppm (further corroborating that this is the carbazole 5-position), demonstrates chlorination at the carbazole 7-position.

Figure AII.10. ^{1}H NMR spectrum for 3.

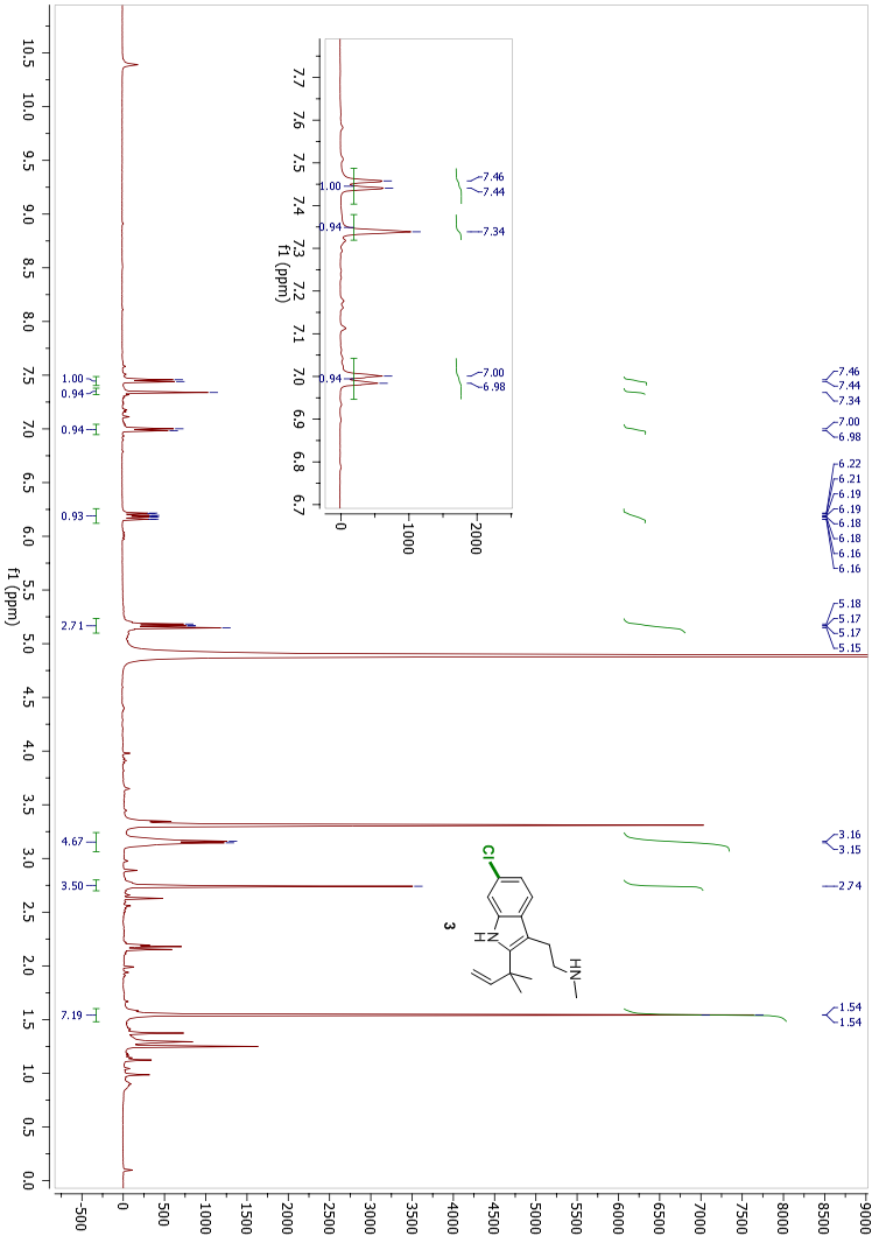


Figure AII.11. ^{13}C NMR spectrum for 3.

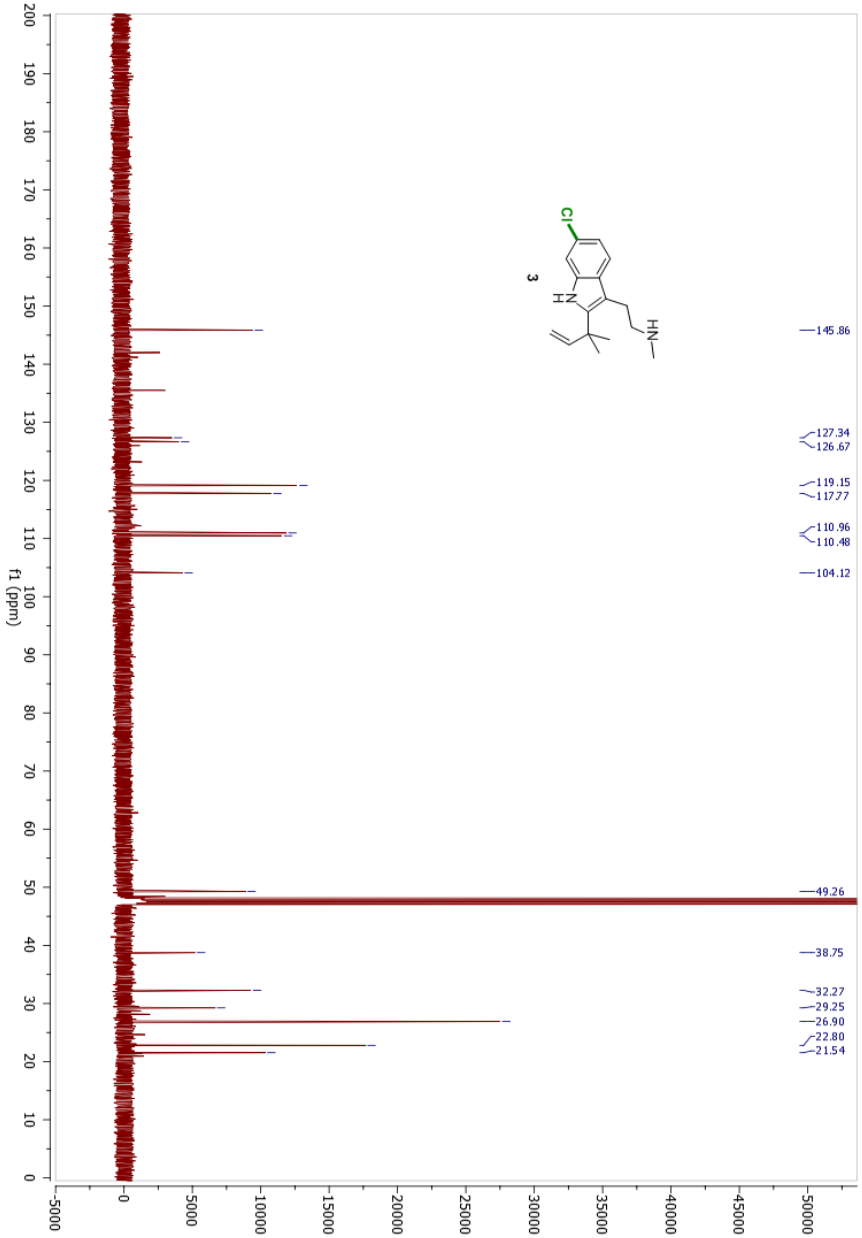
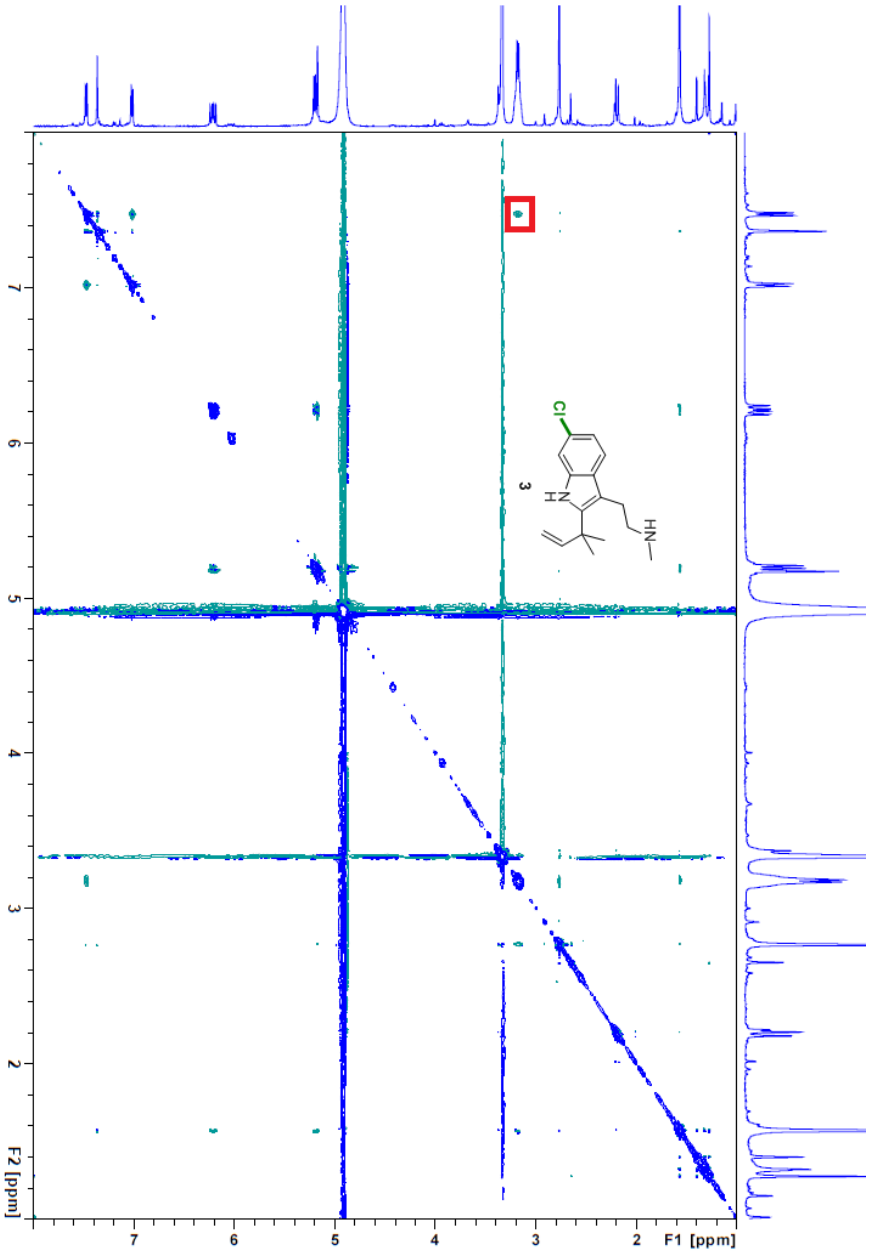


Figure AII.12. NOESY spectrum for 3.[a]



[a] Cross-peak between peaks at ~7.5 ppm (tryptamine 4-position) and ~3.2 ppm (tryptamine β-position) demonstrates chlorination at the tryptamine 6-position.

Figure AII.13. ^{1}H NMR spectrum for 4.

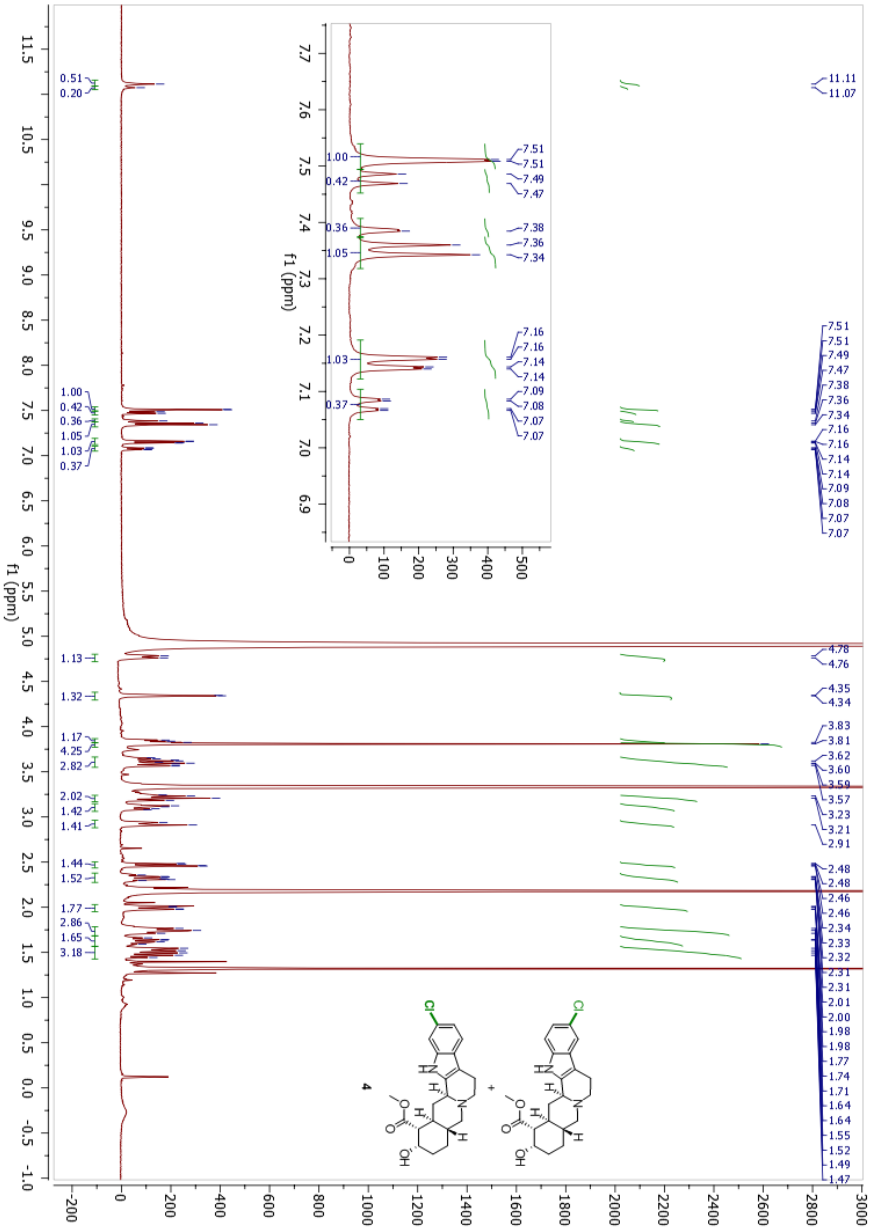


Figure AII.14. ^{13}C NMR spectrum for 4.

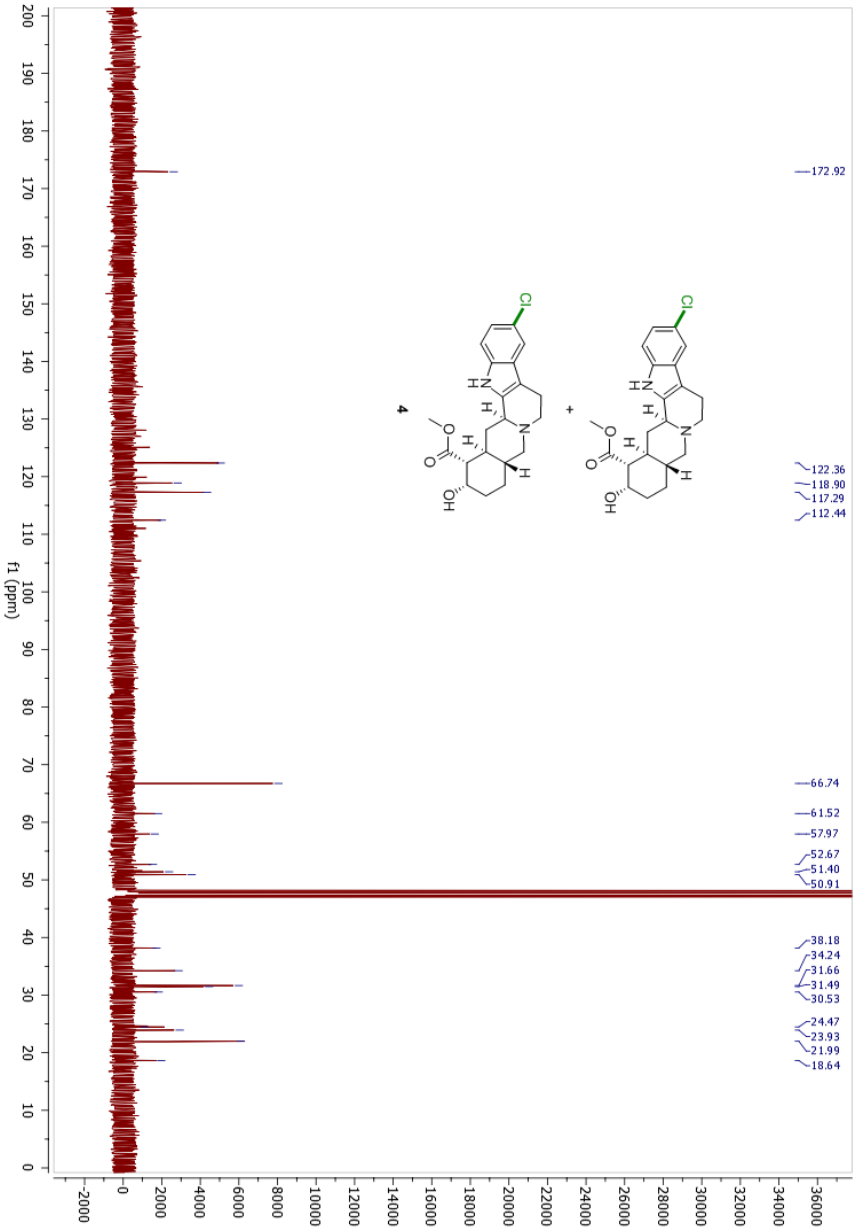
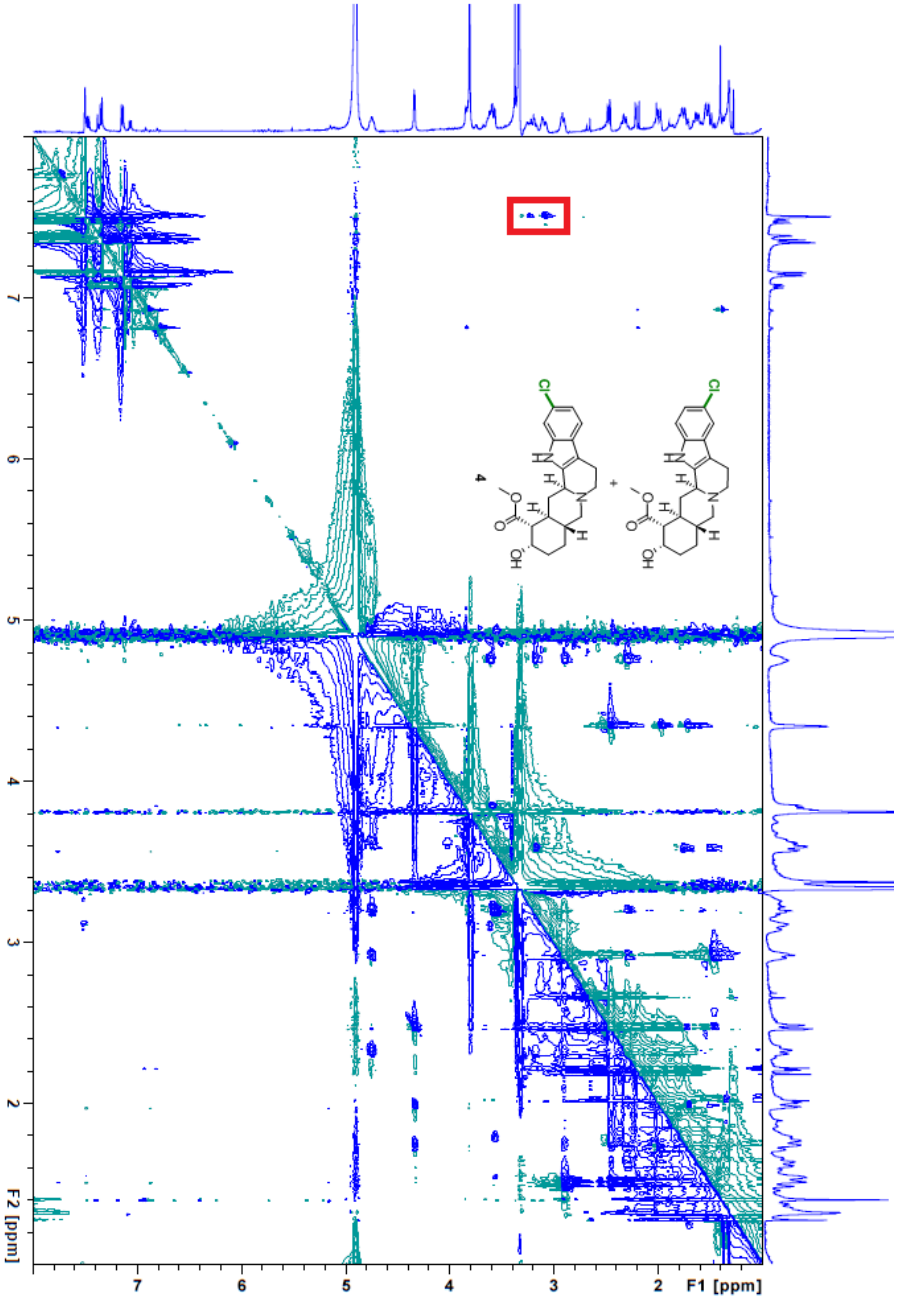


Figure AII.15. NOESY spectrum for 4.[a]



[a] Cross-peaks between peaks at ~7.5 ppm (yohimban 9-position) and ~3.1 ppm (yohimban 6-position) demonstrates chlorination of major product occurs at the yohimban 10-position and chlorination of minor product occurs at the yohimban 11-position.

Figure AII.16. ^{1}H NMR spectrum for 9.

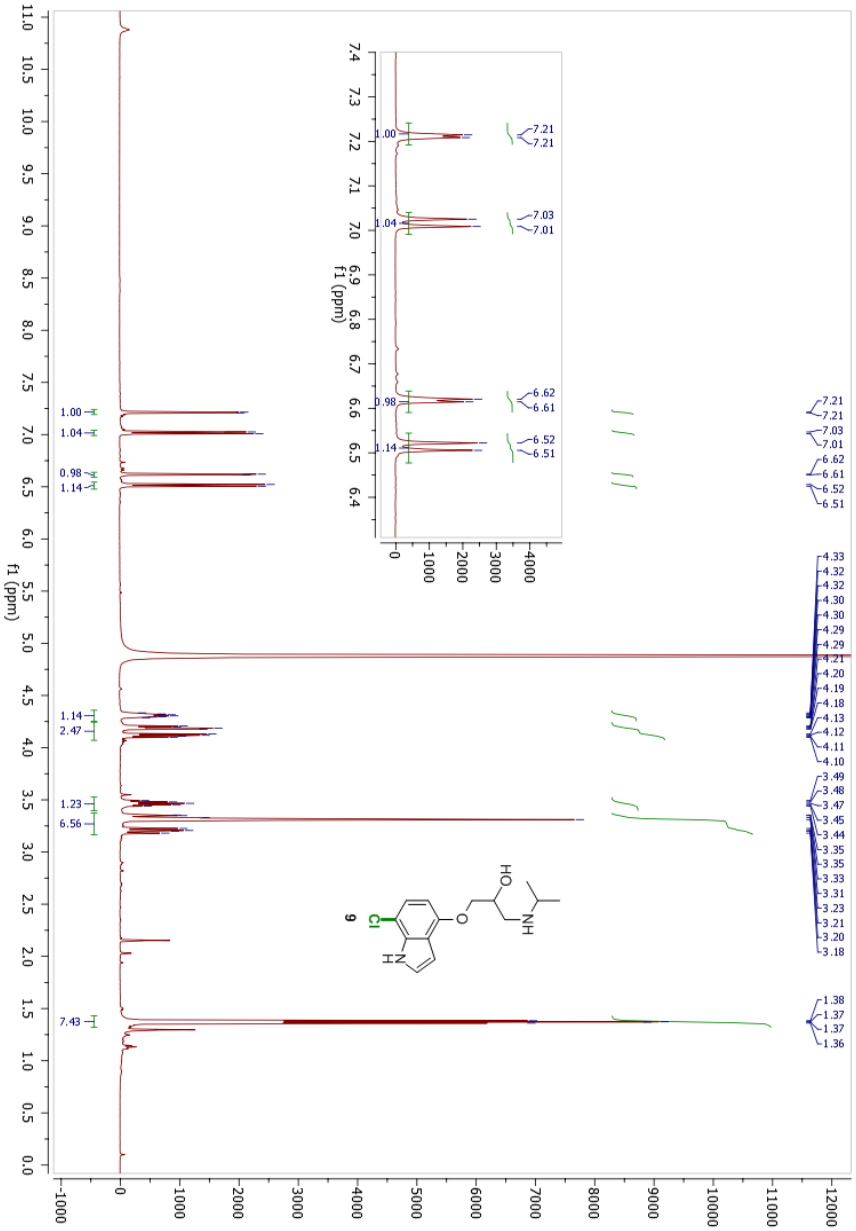


Figure AII.17. ^{13}C NMR spectrum for 9.

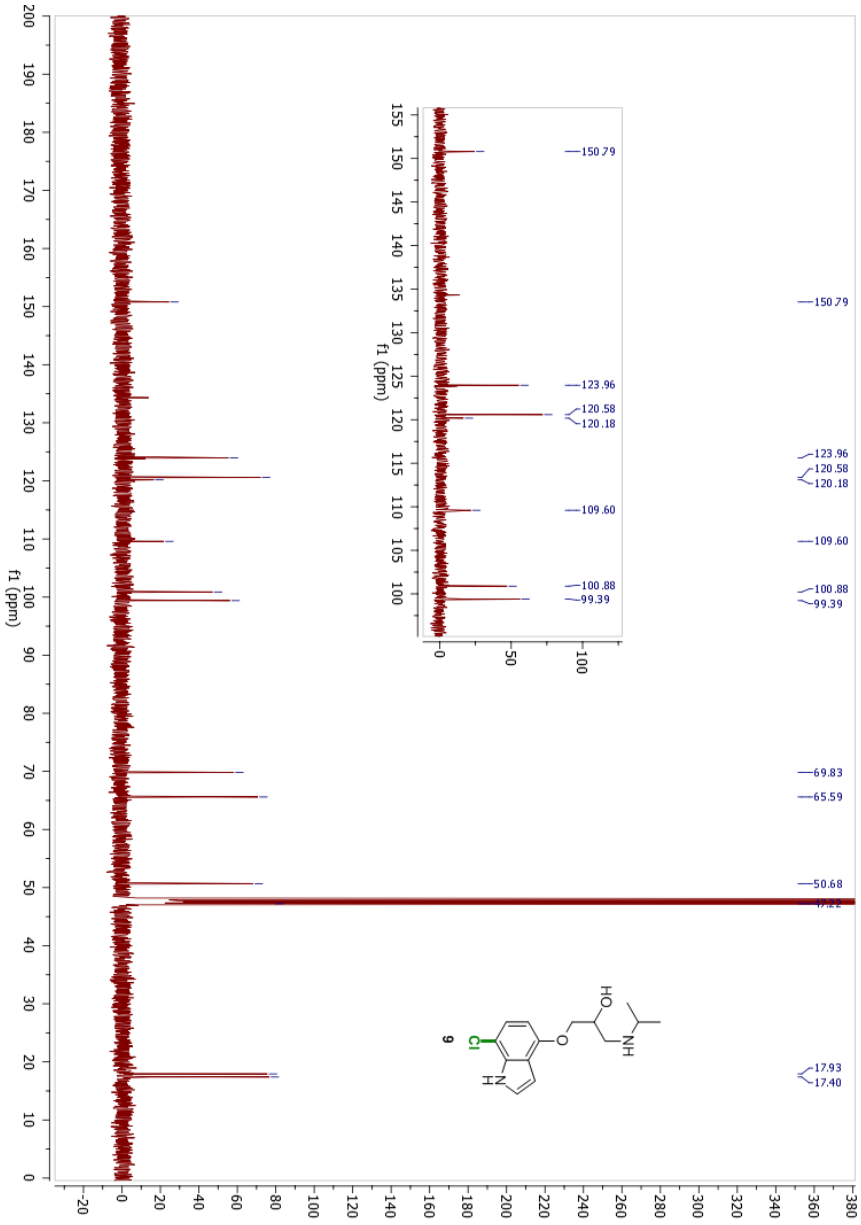
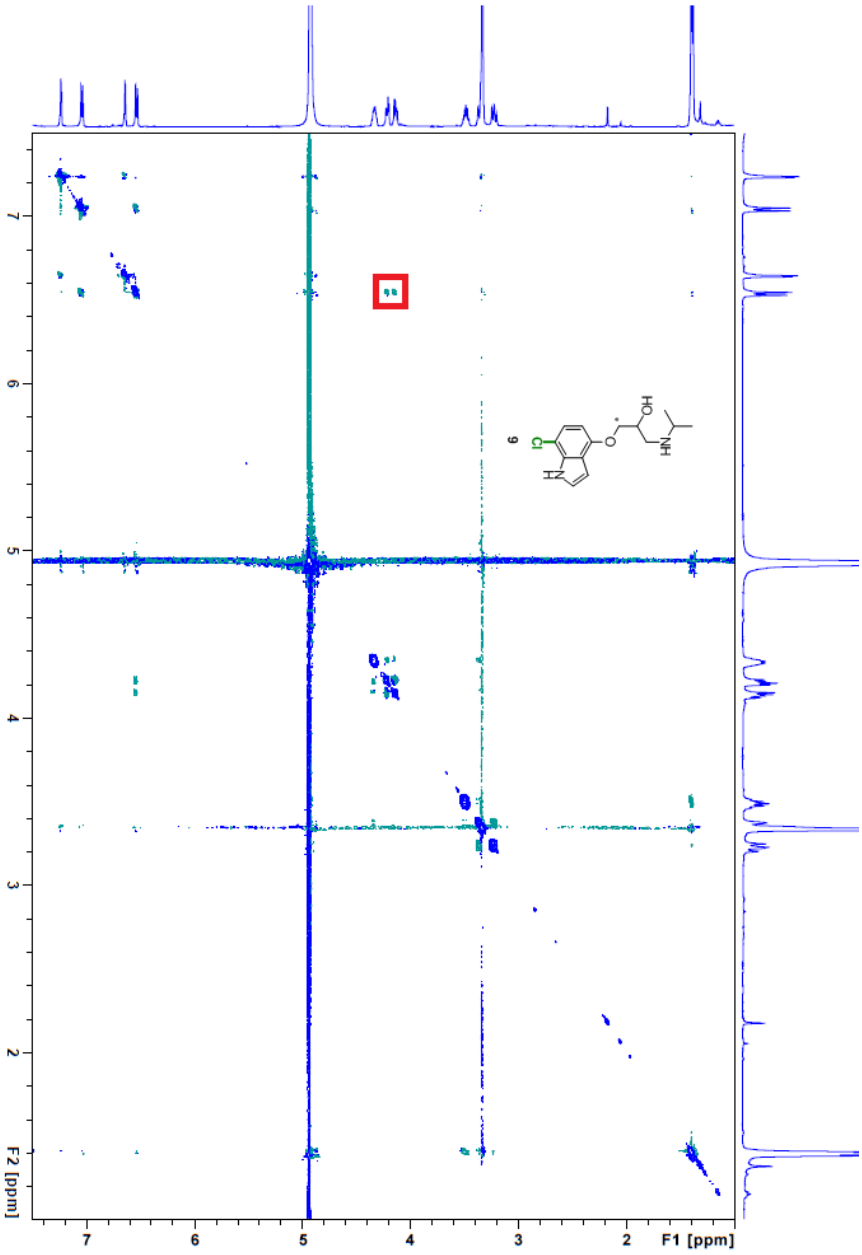


Figure AII.18. NOESY spectrum for 9.[a]



[a] Cross-peak between peaks at ~6.5 ppm (indole 5-position) and ~4.2 ppm (position marked by *) demonstrates chlorination occurs at the indole 7-position.

Figure AII.19. ^{1}H NMR spectrum for 10.

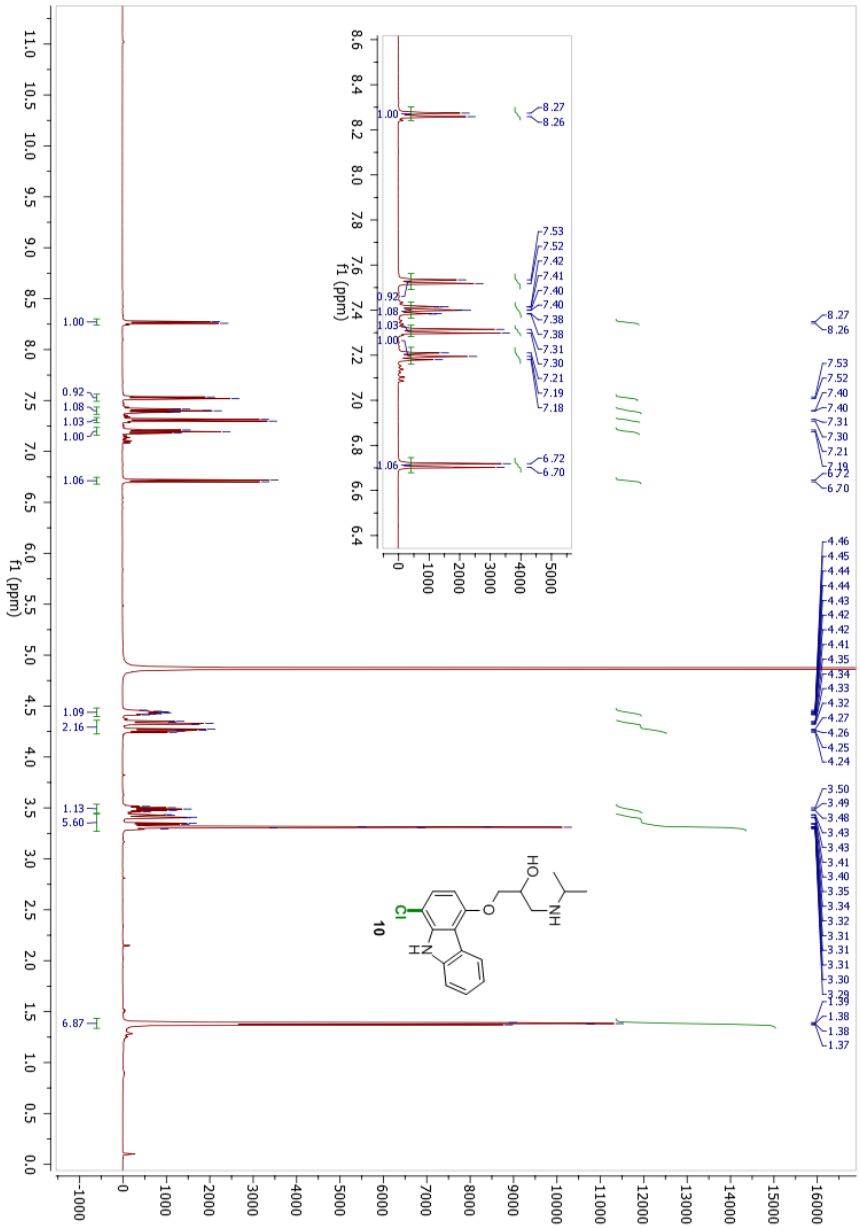


Figure AII.20. ^{13}C NMR spectrum for 10.

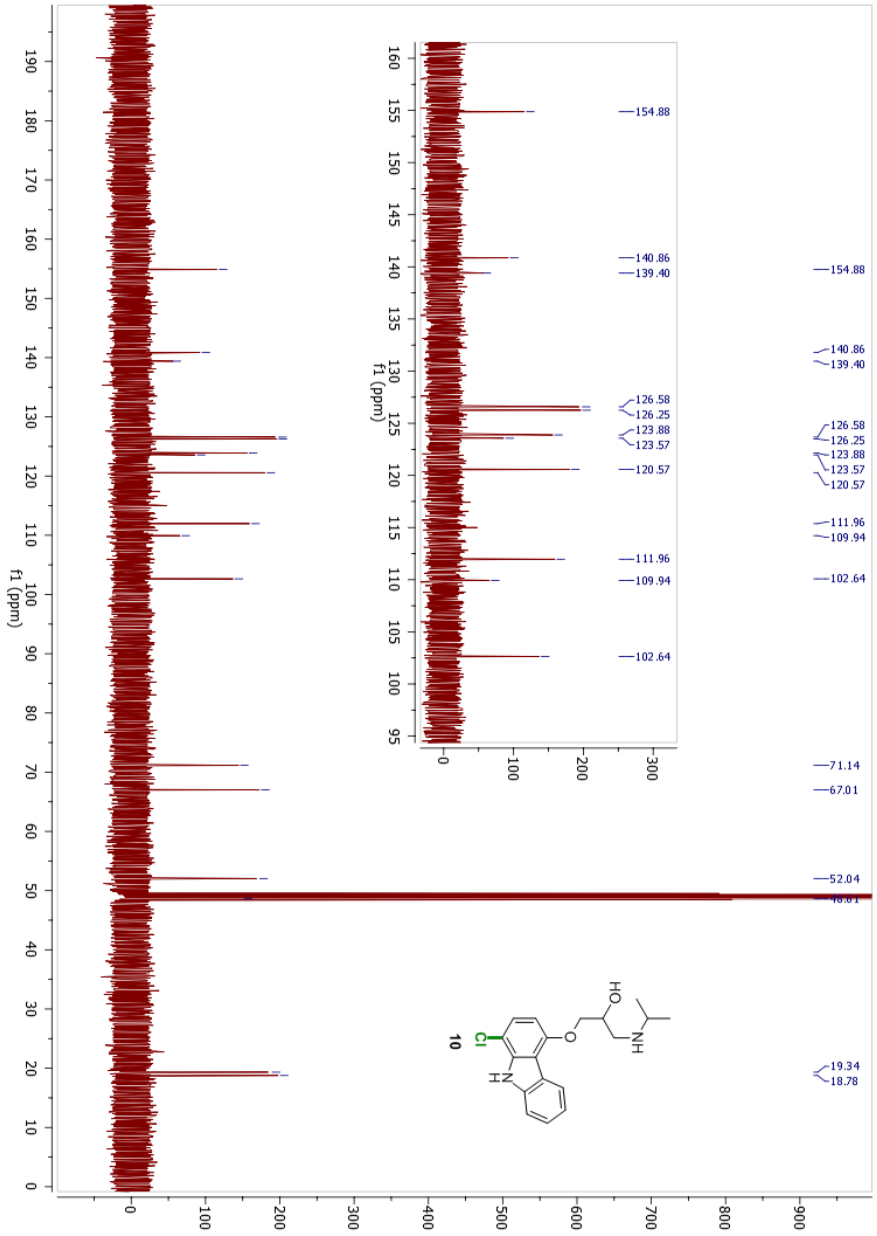
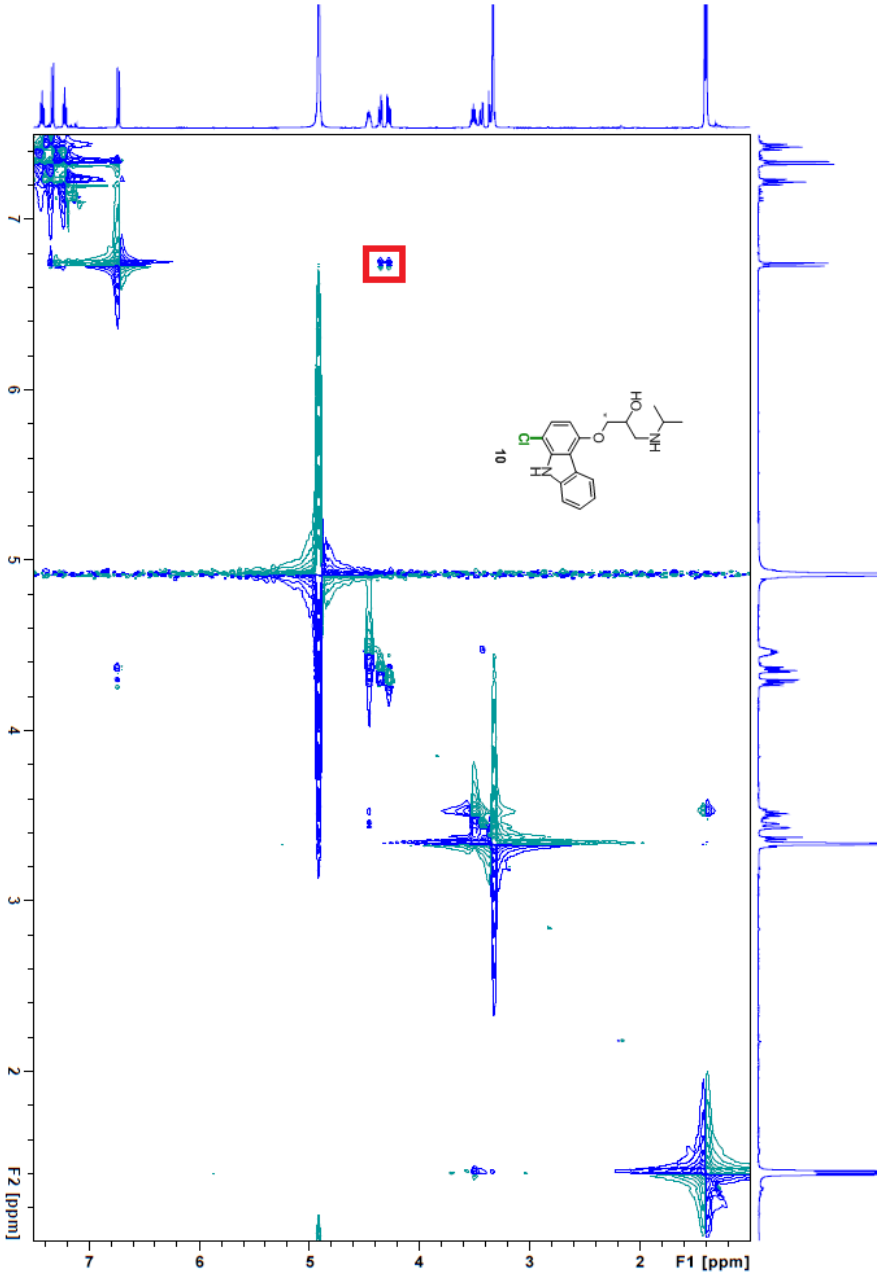


Figure AII.21. NOESY spectrum for 10.[a]



[a] Cross-peak between peaks at ~6.7 ppm (carbazole 6-position) and ~4.4 ppm (position marked by *) demonstrates chlorination occurs at the carbazole 8-position.

Figure AII.22. ^{1}H NMR spectrum of 11.

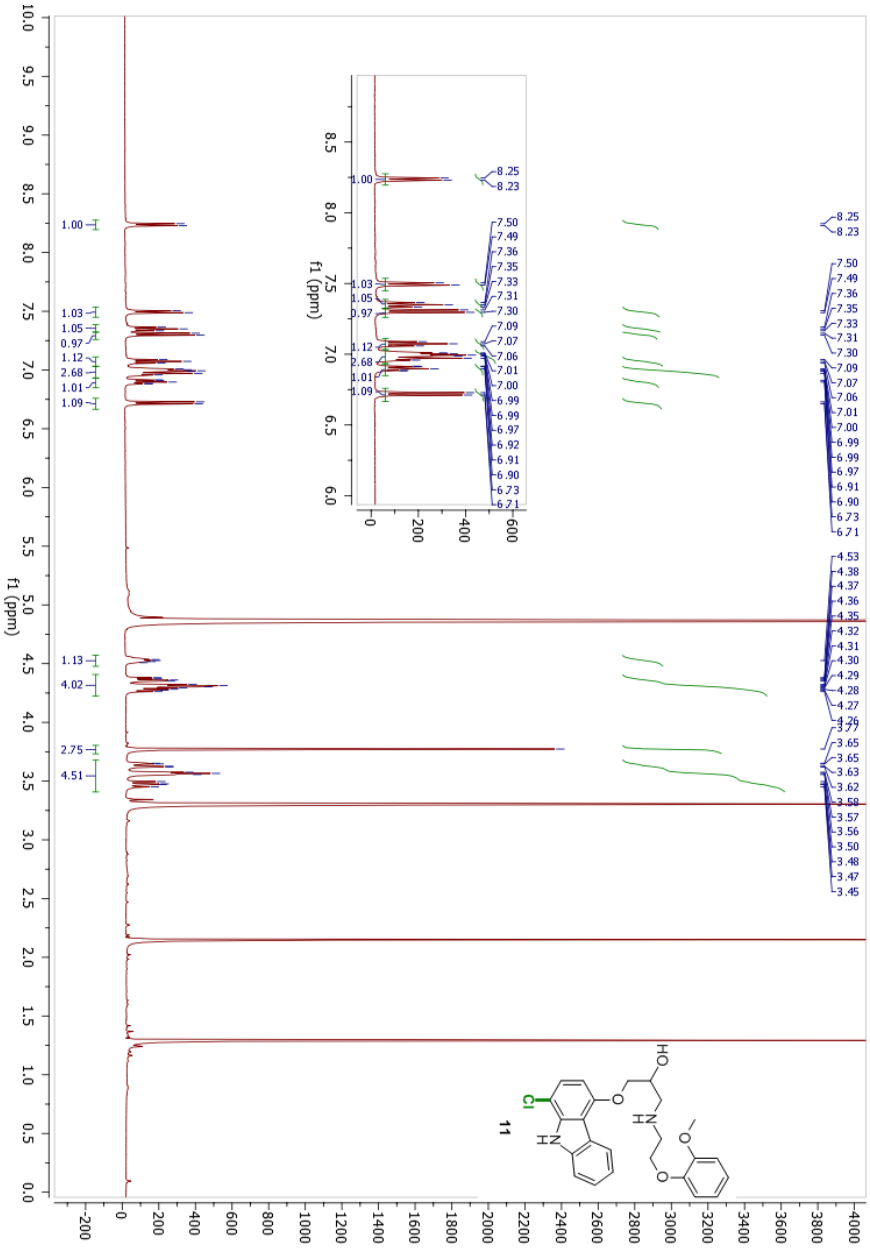


Figure AII.23. ^{13}C NMR spectrum for 11.

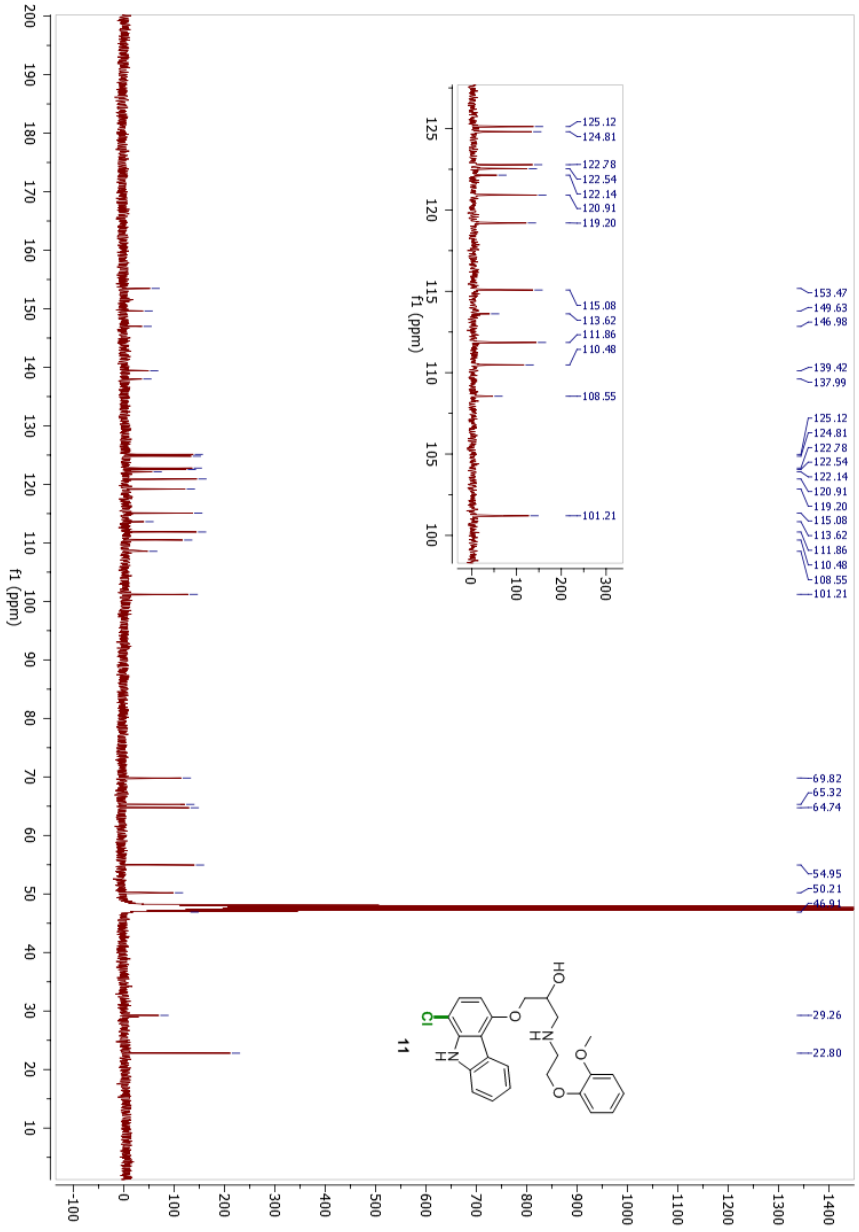
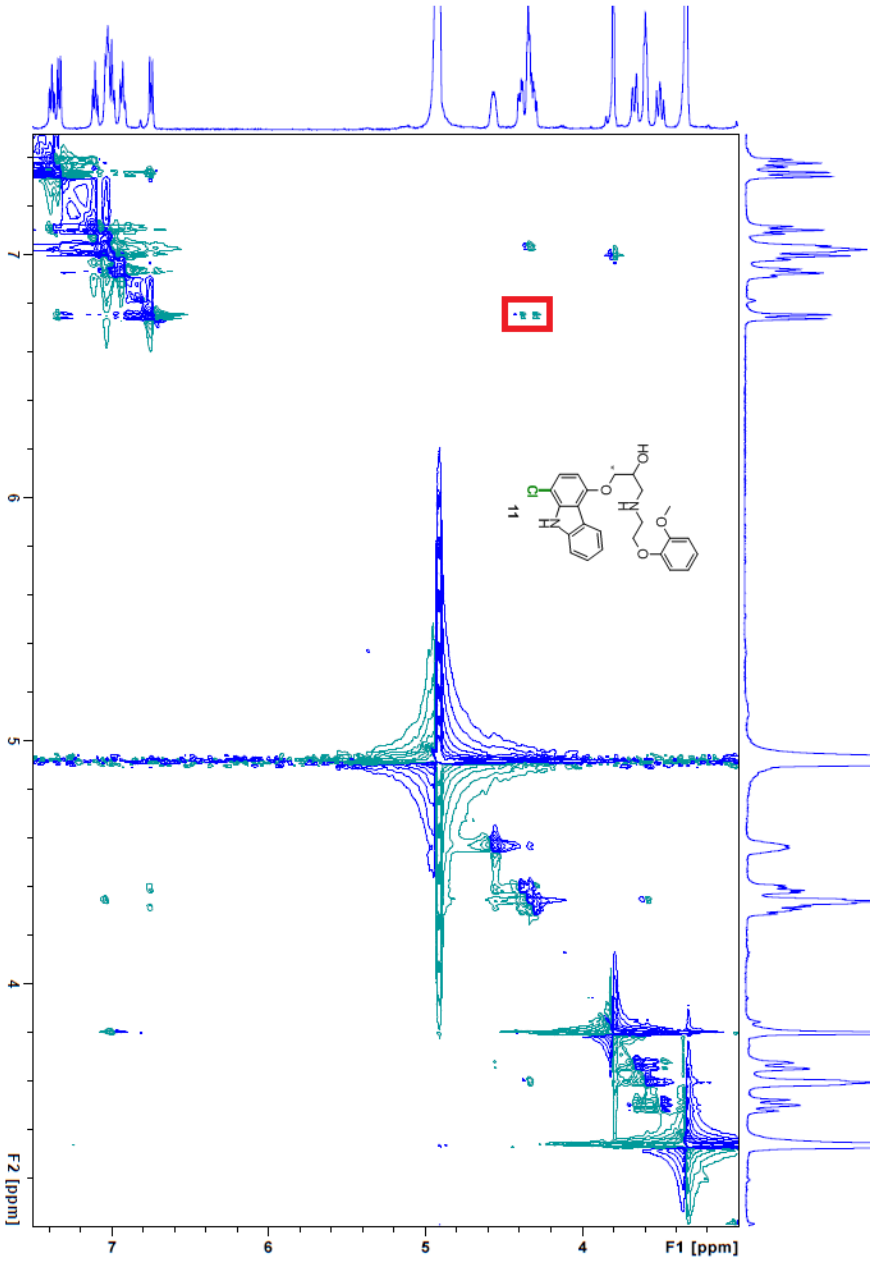
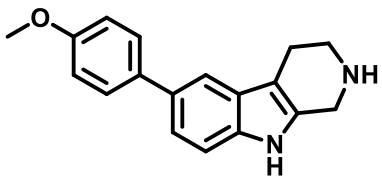


Figure AII.24. NOESY spectrum for 11.[a]



[a] Cross-peak between peaks at ~6.8 ppm (carbazole 6-position) and ~4.4 ppm (position marked by *) demonstrates chlorination occurs at the carbazole 8-position.

Figure AII.25. ^{1}H NMR spectrum for 2a.



2a

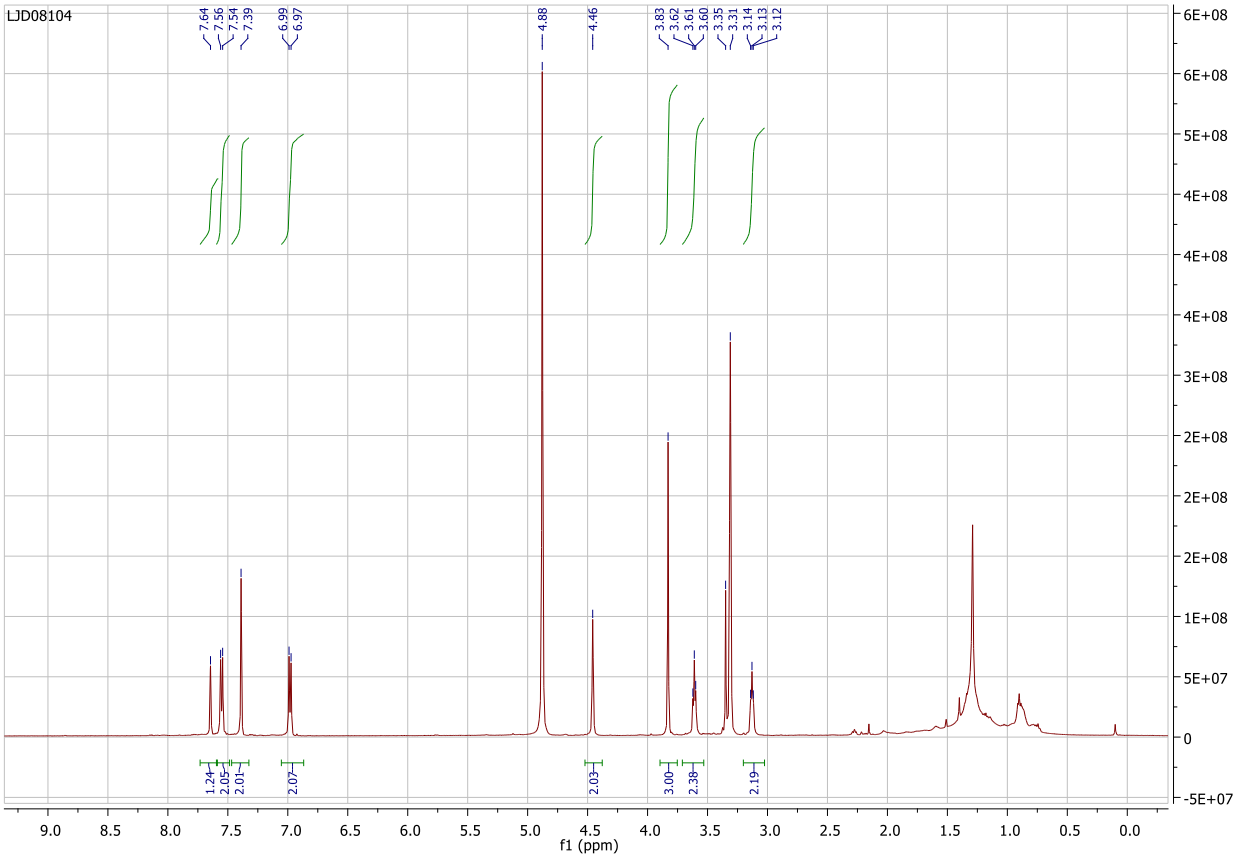
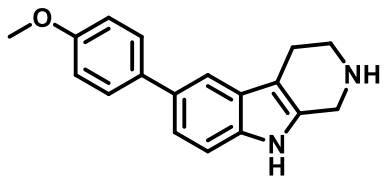


Figure AII.26. ^{13}C NMR spectrum for 2a.



2a

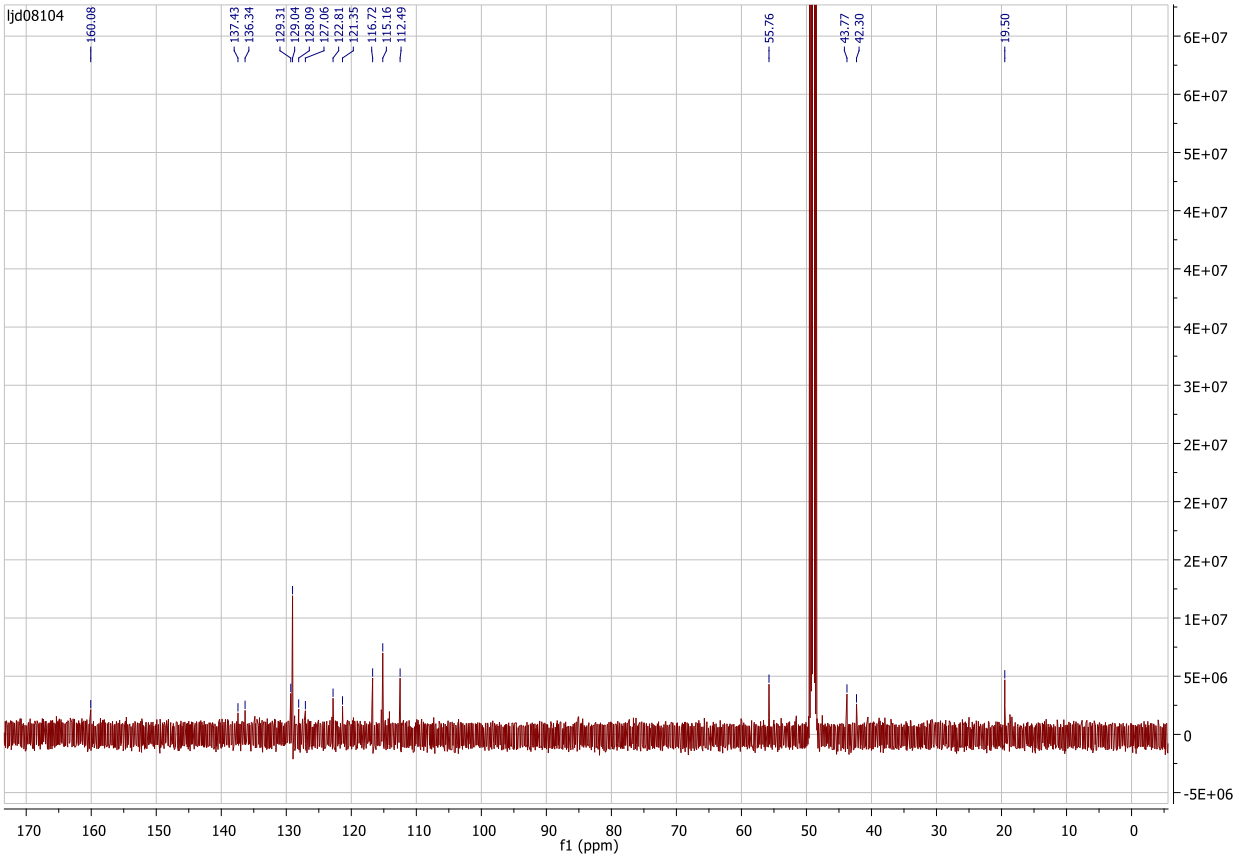
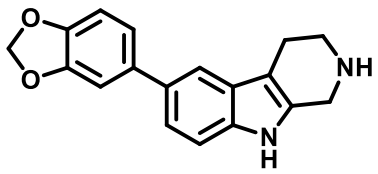


Figure AII.27. ^{1}H NMR spectrum for 2b.



2b

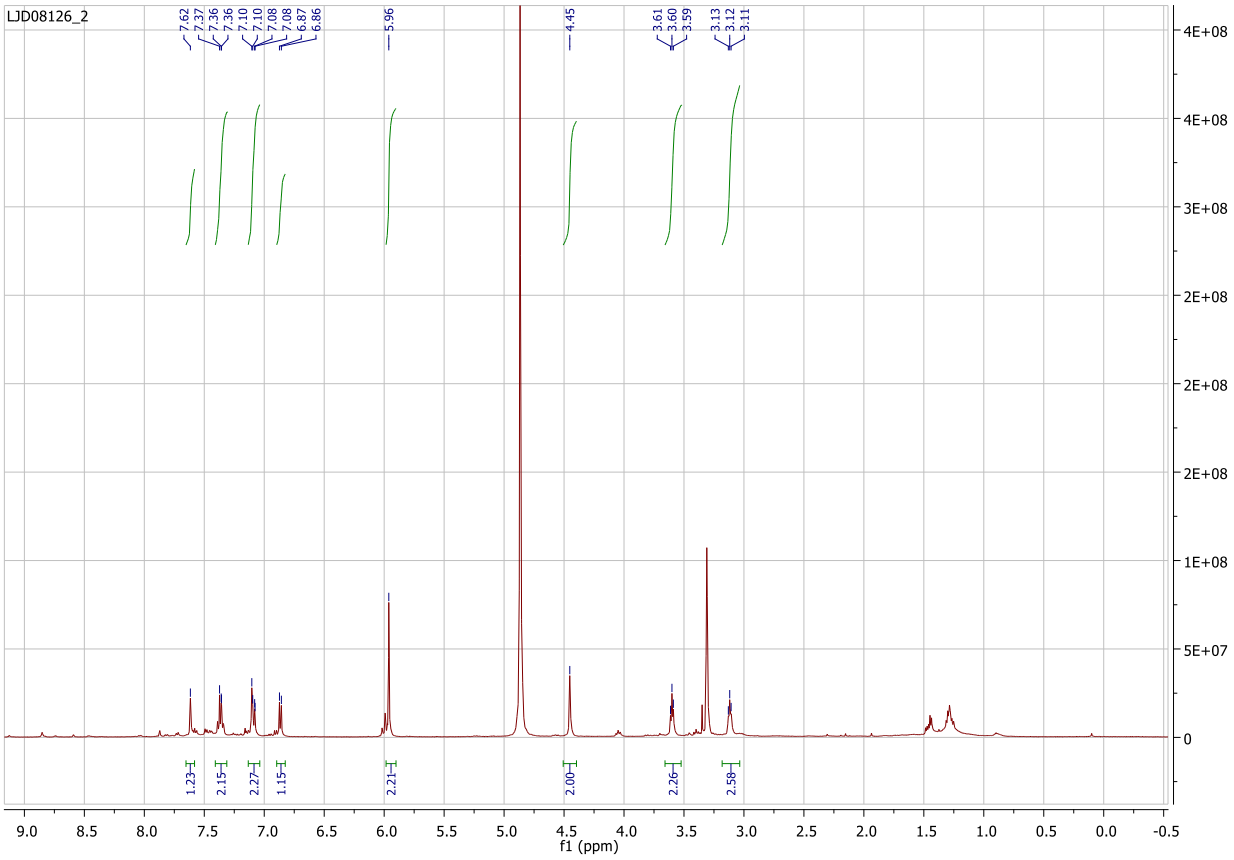
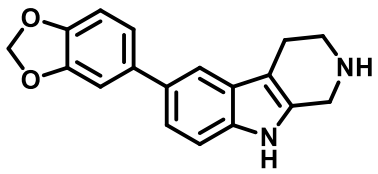


Figure AII.28. ^{13}C NMR spectrum for 2b.



2b

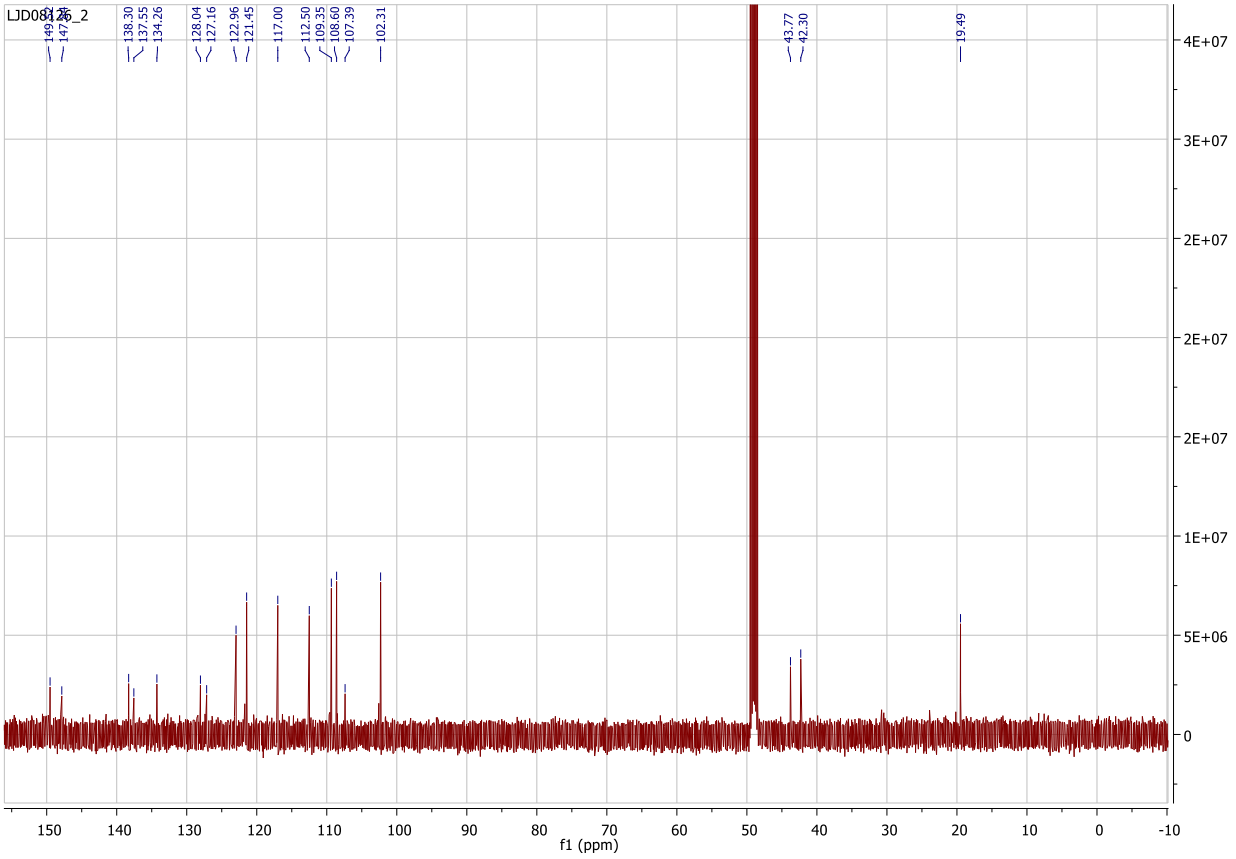
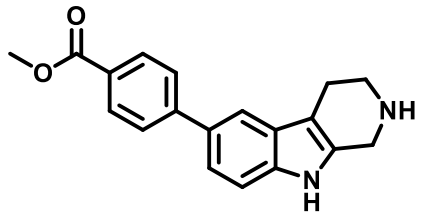


Figure AII.29. ^{1}H NMR spectrum for 2c.



2c

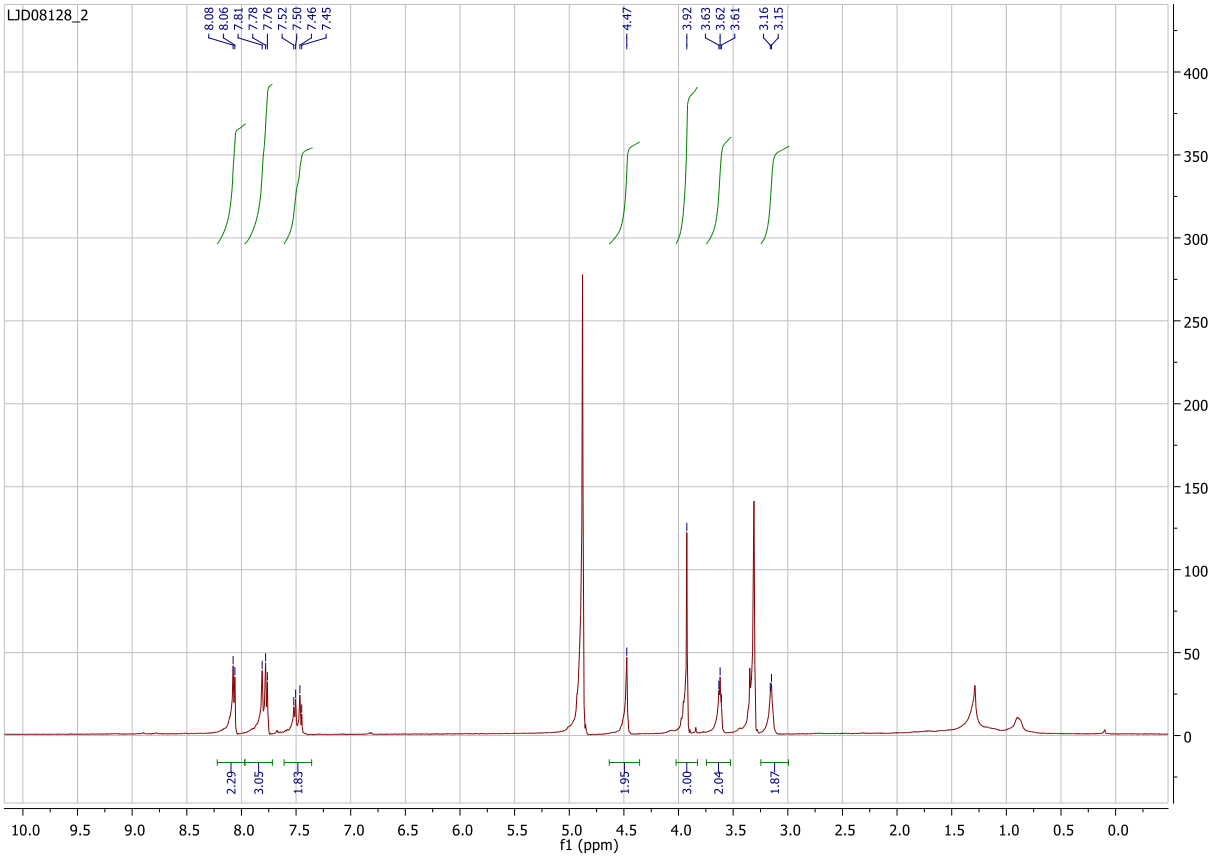
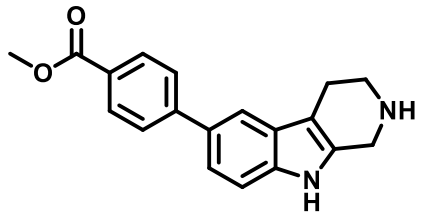


Figure AII.30. ^{13}C NMR spectrum for 2c.



2c

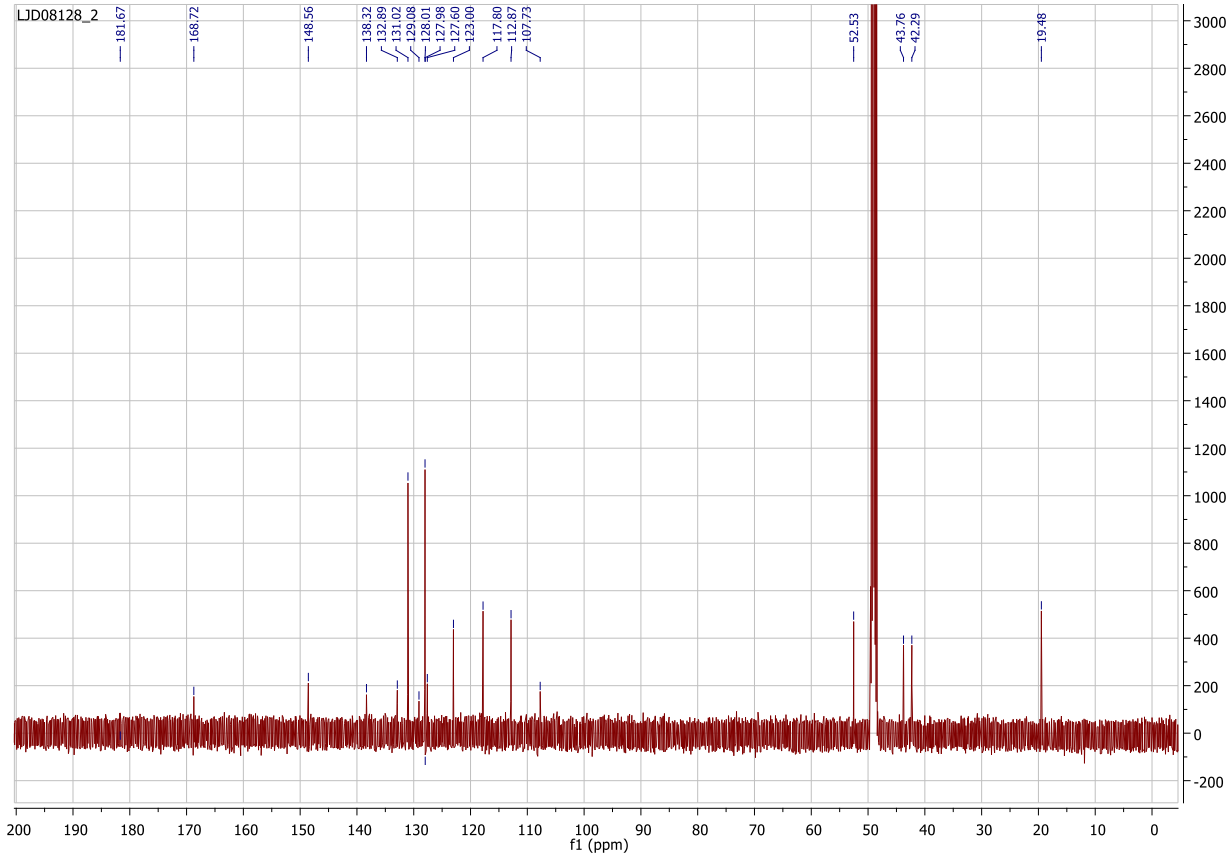
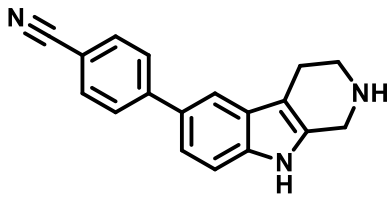


Figure AII.31. ^{1}H NMR spectrum for 2d.



2d

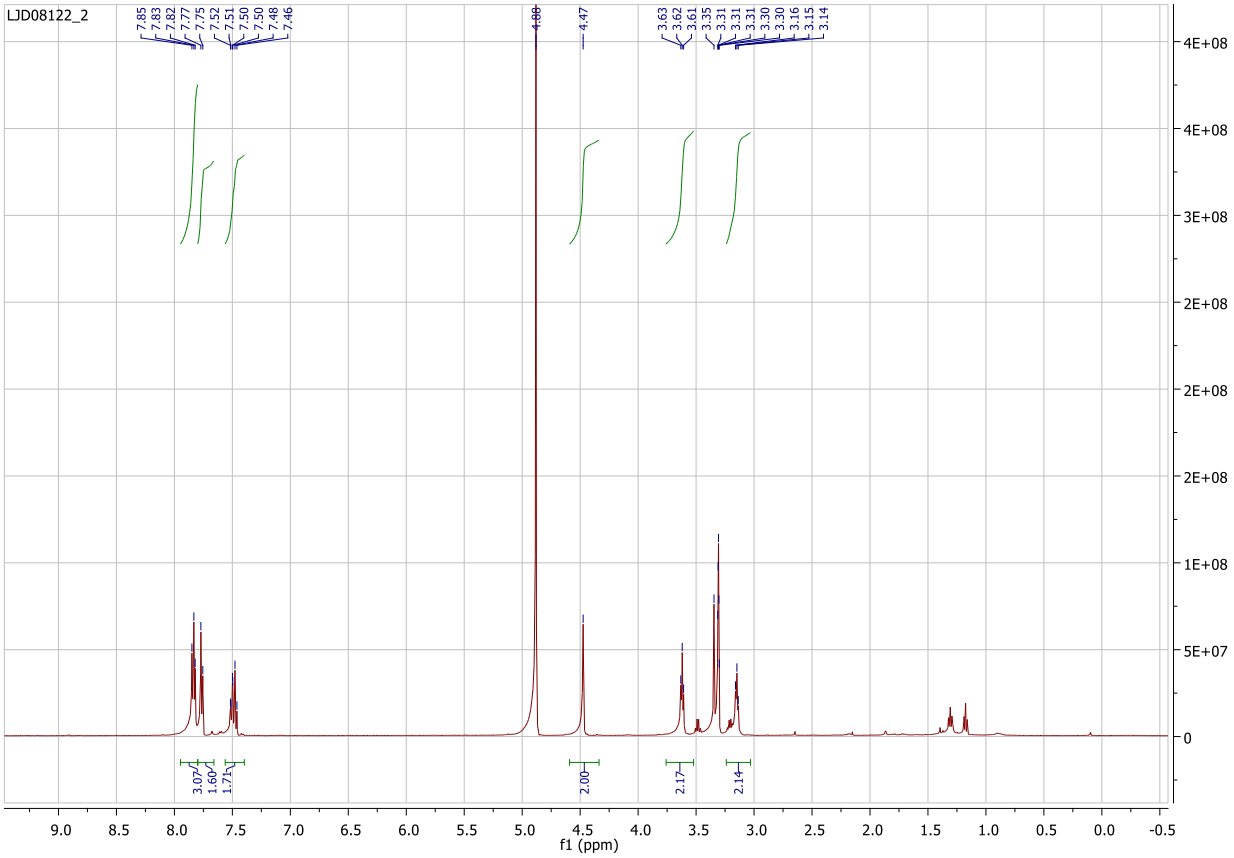
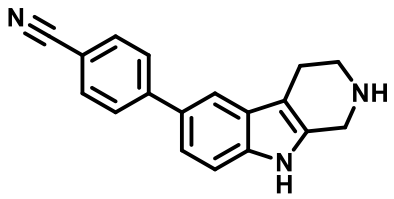


Figure AII.32. ^{13}C NMR spectrum for 2d.



2d

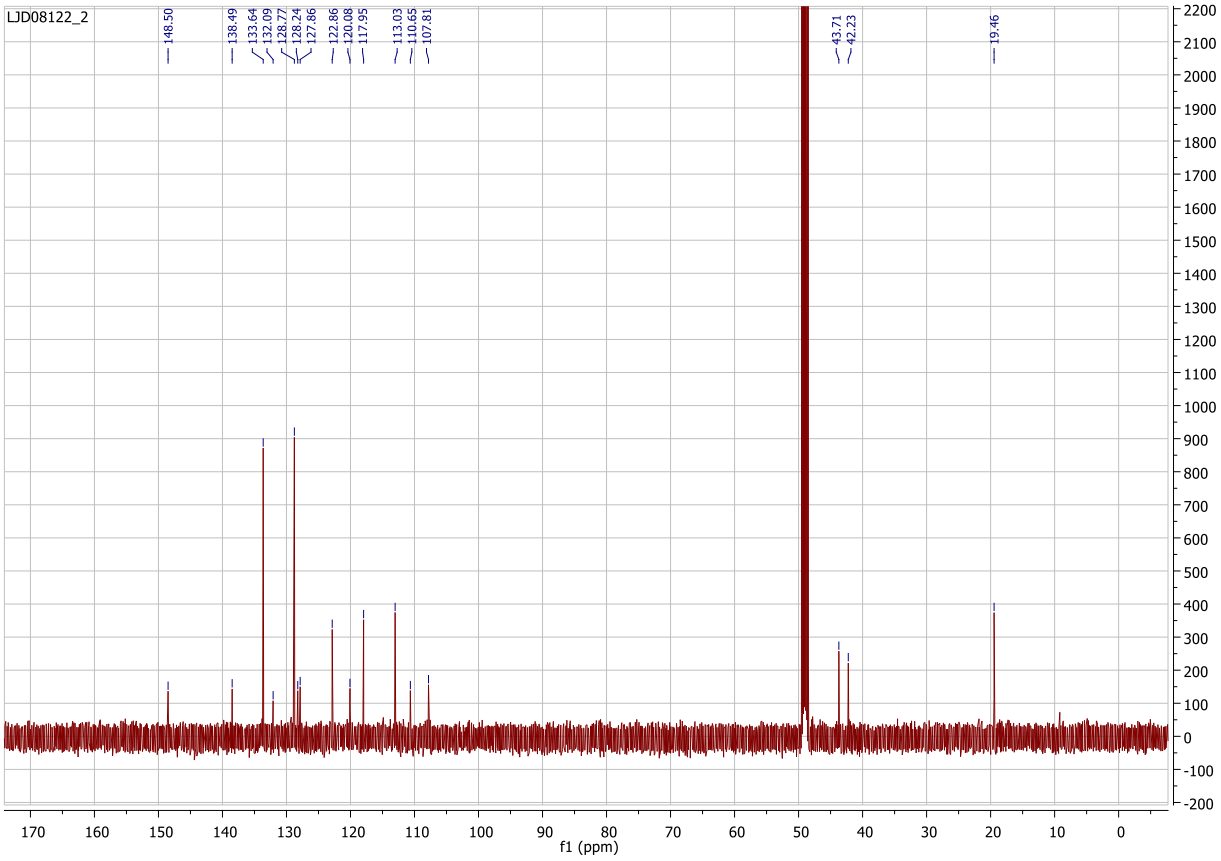
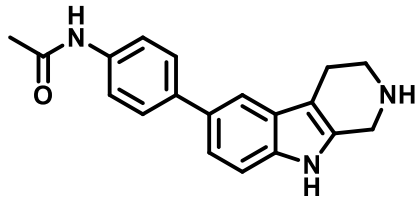


Figure AII.33. ^{1}H NMR spectrum for 2e.



2e

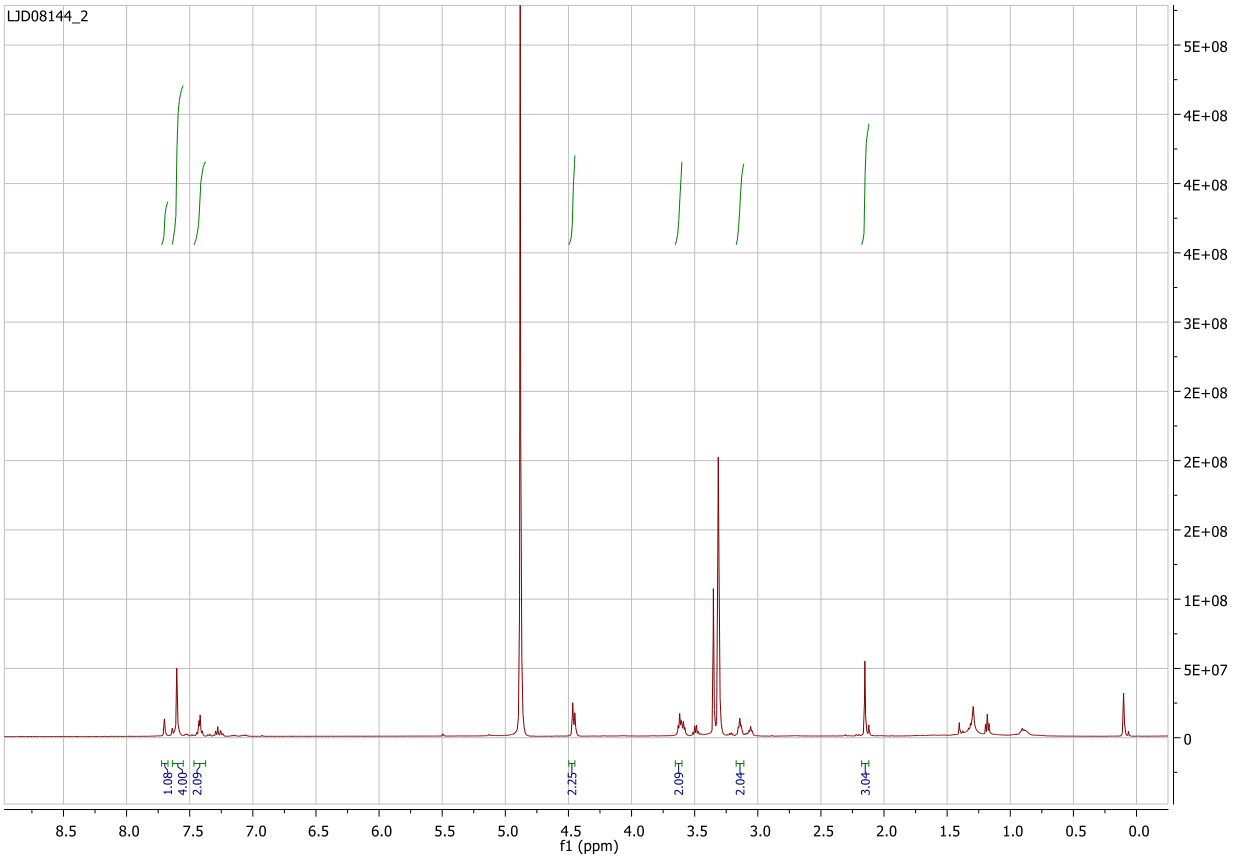
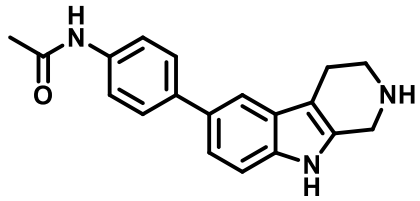


Figure AII.34. ^{13}C NMR spectrum for 2e.



2e

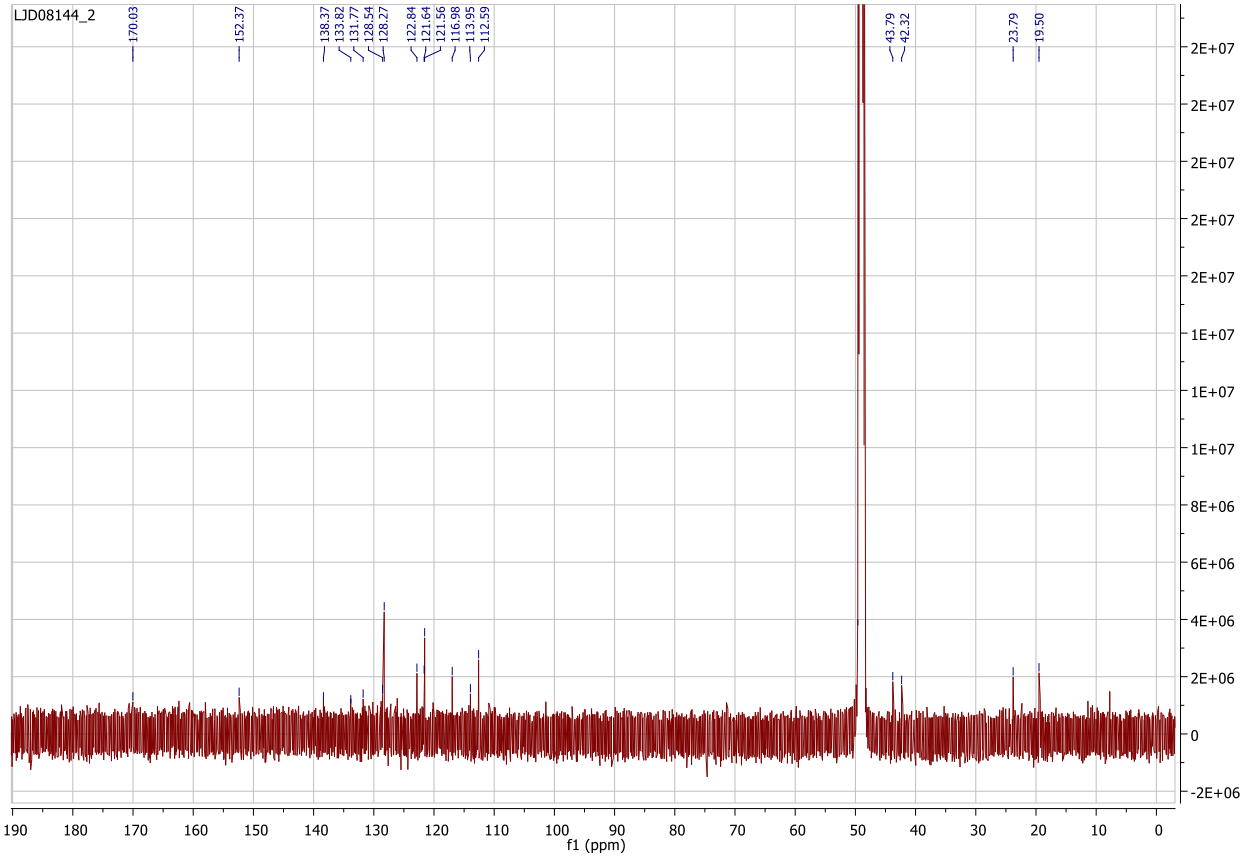
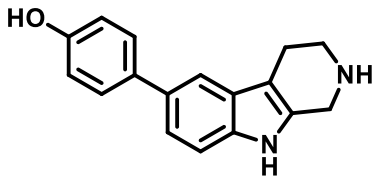


Figure AII.35. ^{1}H NMR spectrum for 2f.



2f

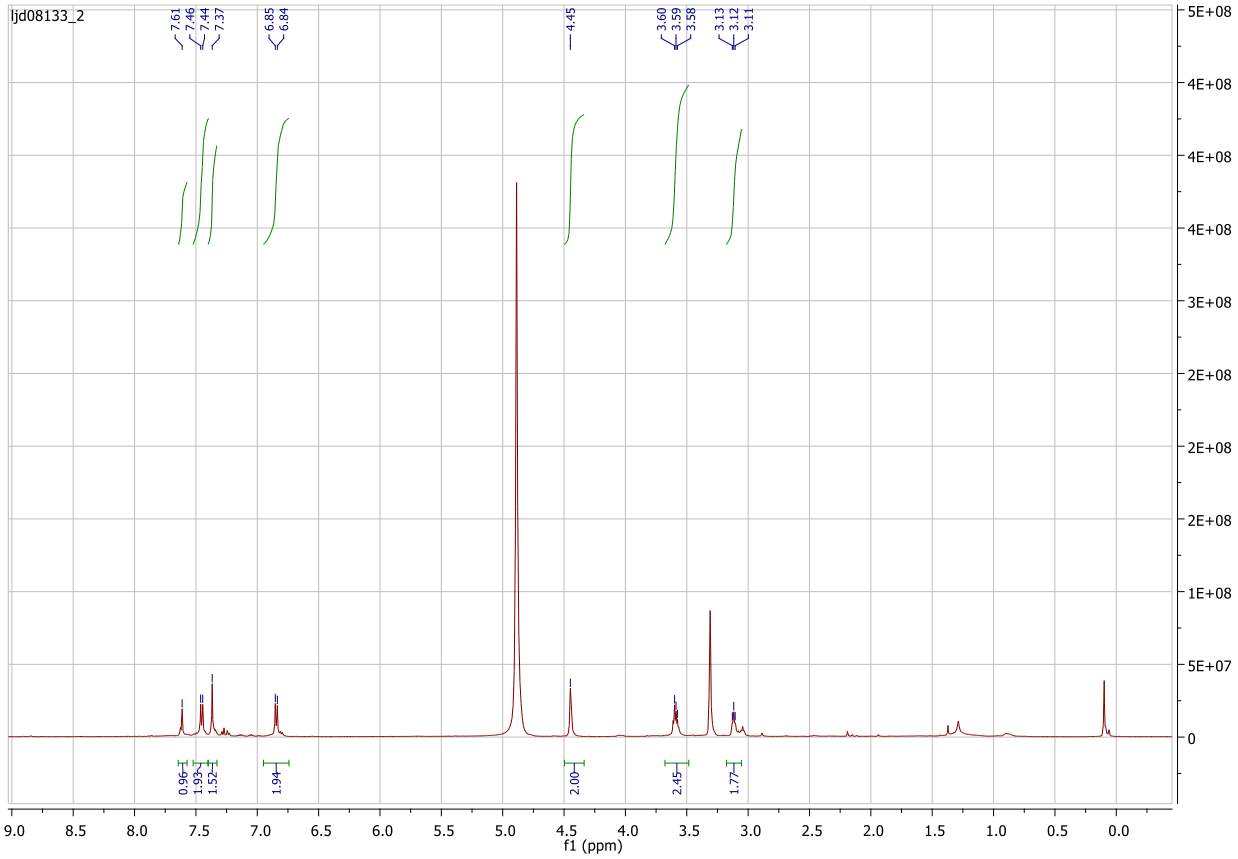
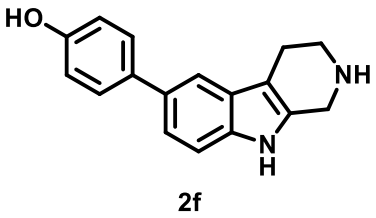


Figure AII.36. ^{13}C NMR spectrum for 2f.



2f

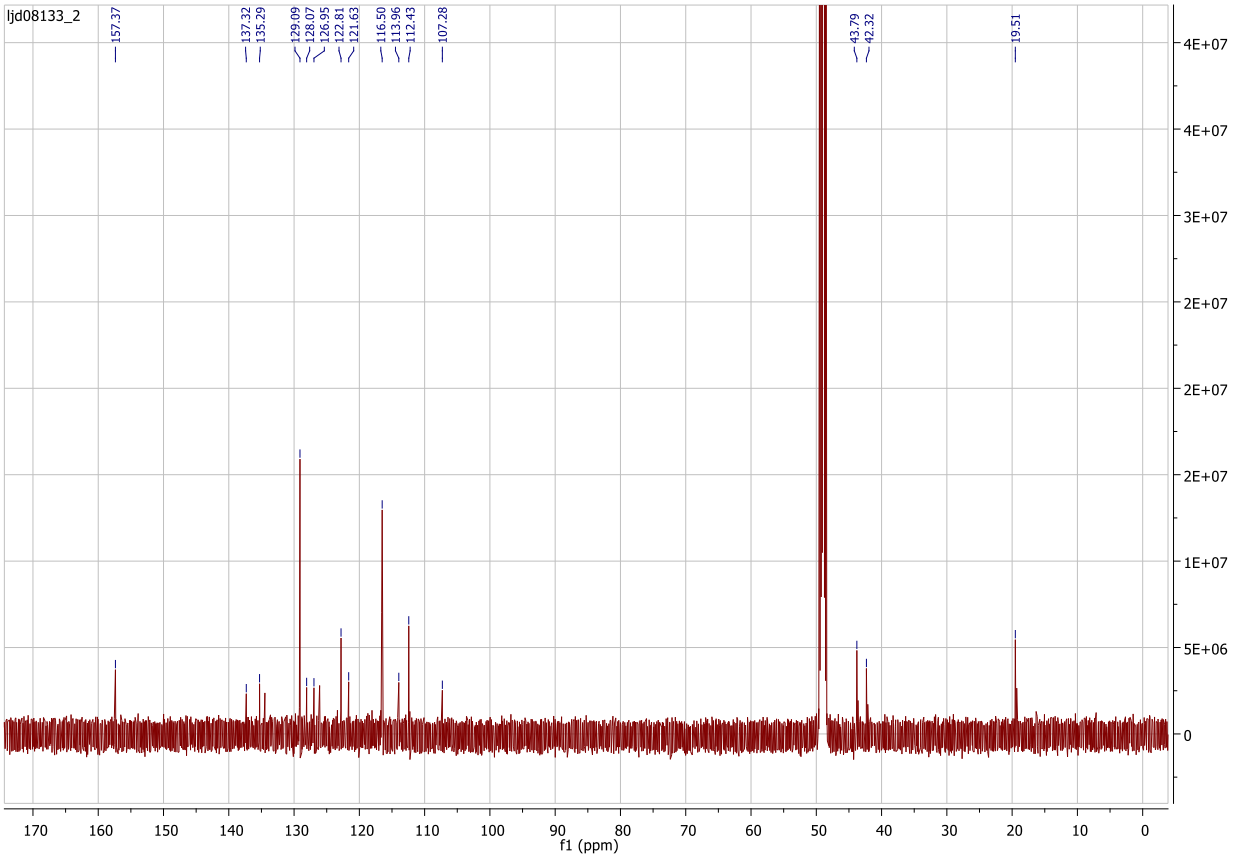
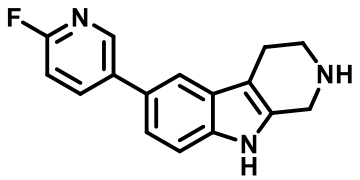


Figure AII.37. ^{1}H NMR spectrum for 2g.



2g

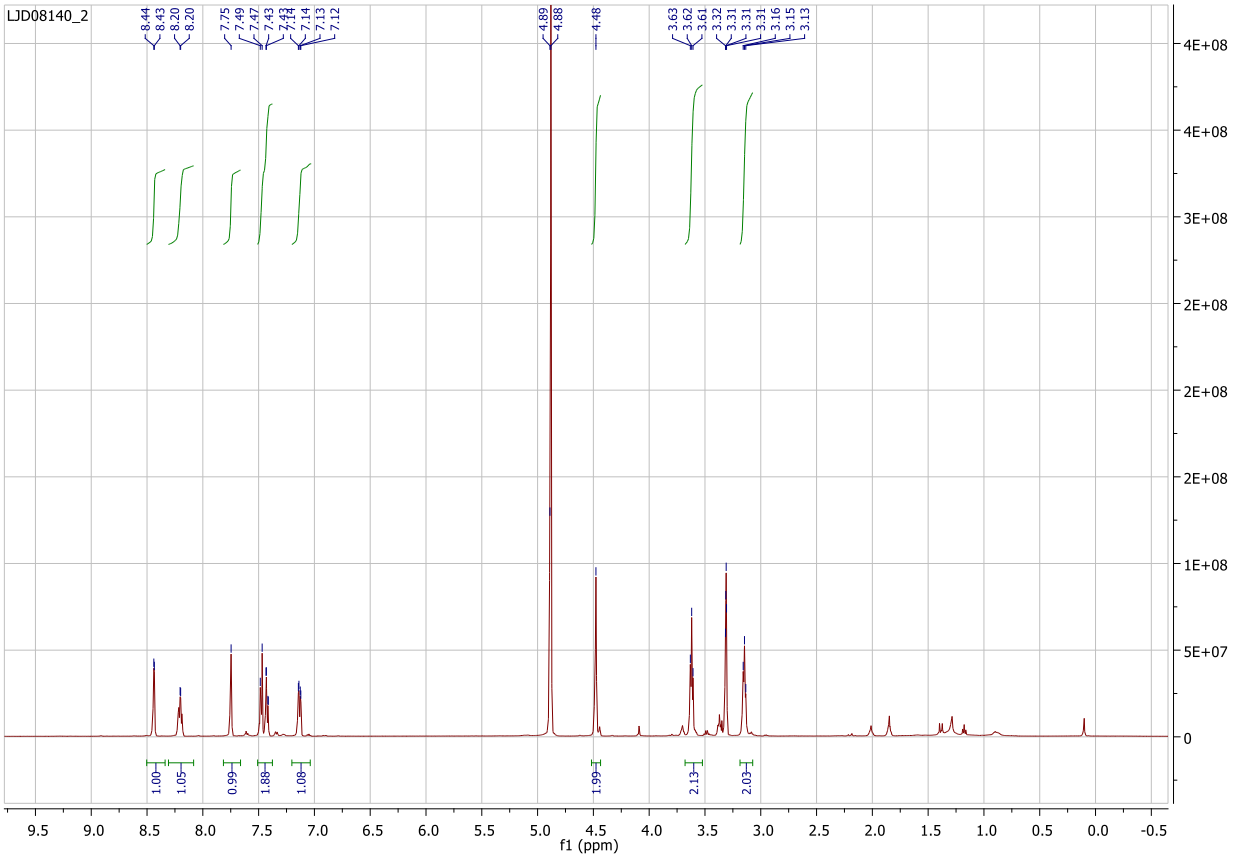
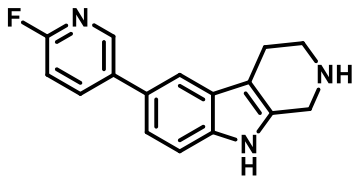


Figure AII.38. ^{13}C NMR spectrum for 2g.



2g

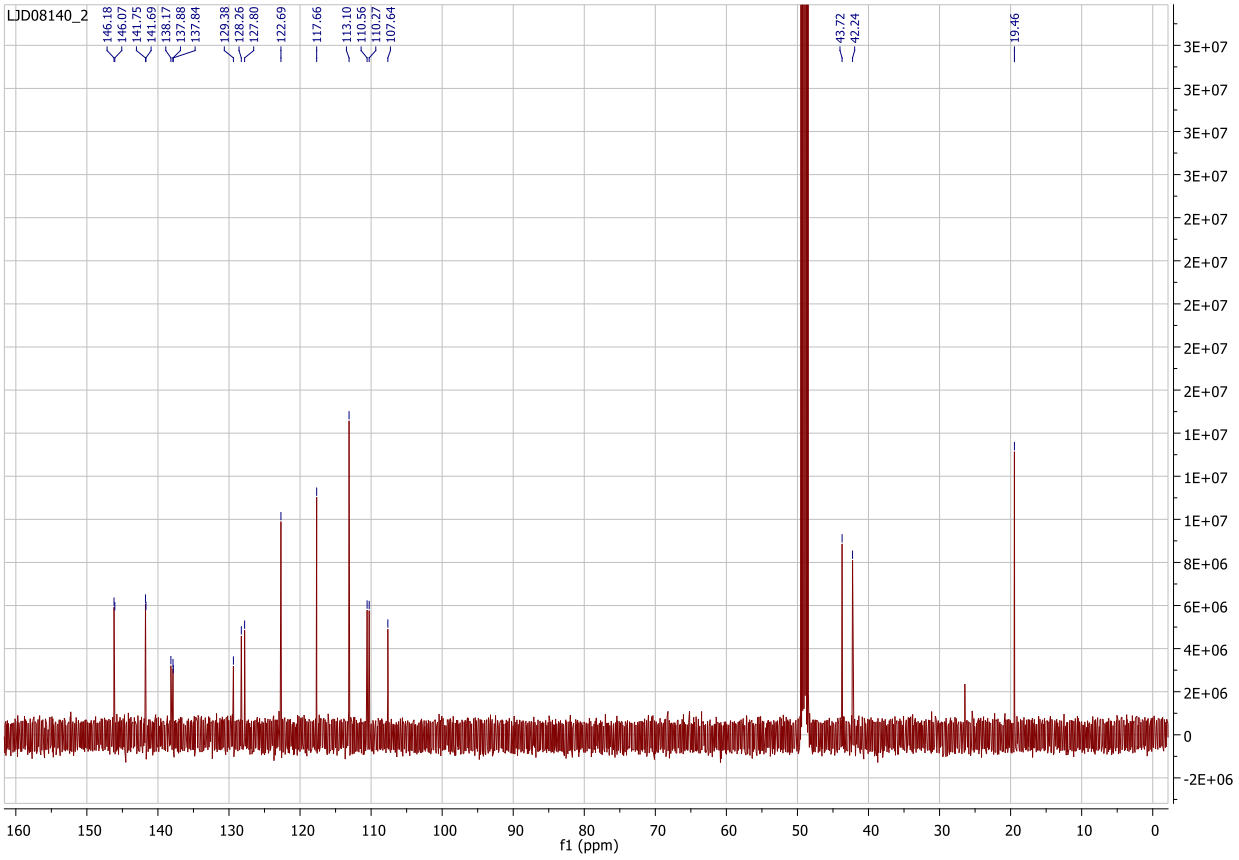
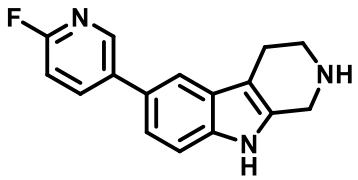


Figure AII.39. ^{19}F NMR spectrum for 2g.



2g

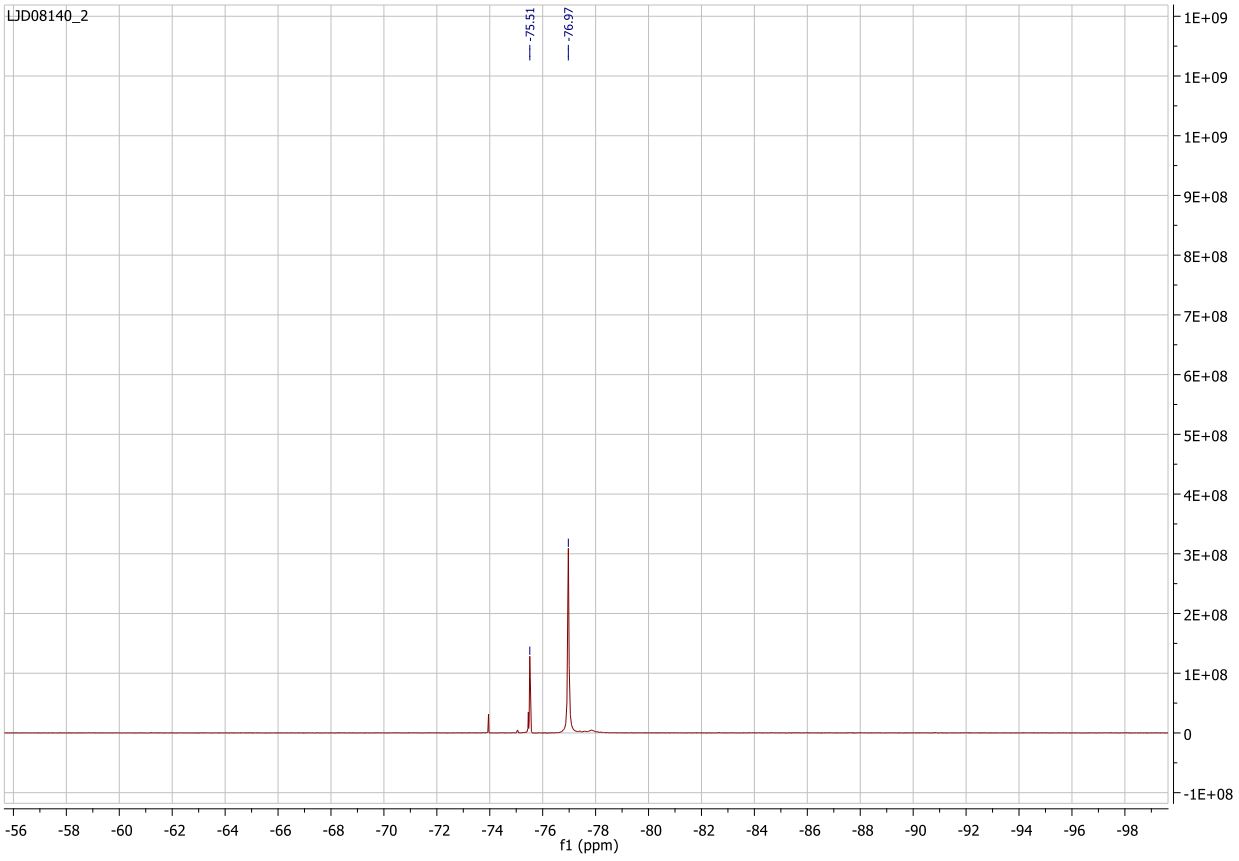
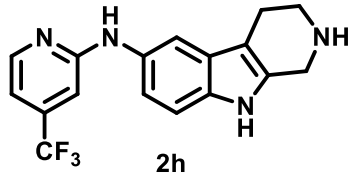


Figure AII.40. ^{1}H NMR spectrum for 2h.



2h

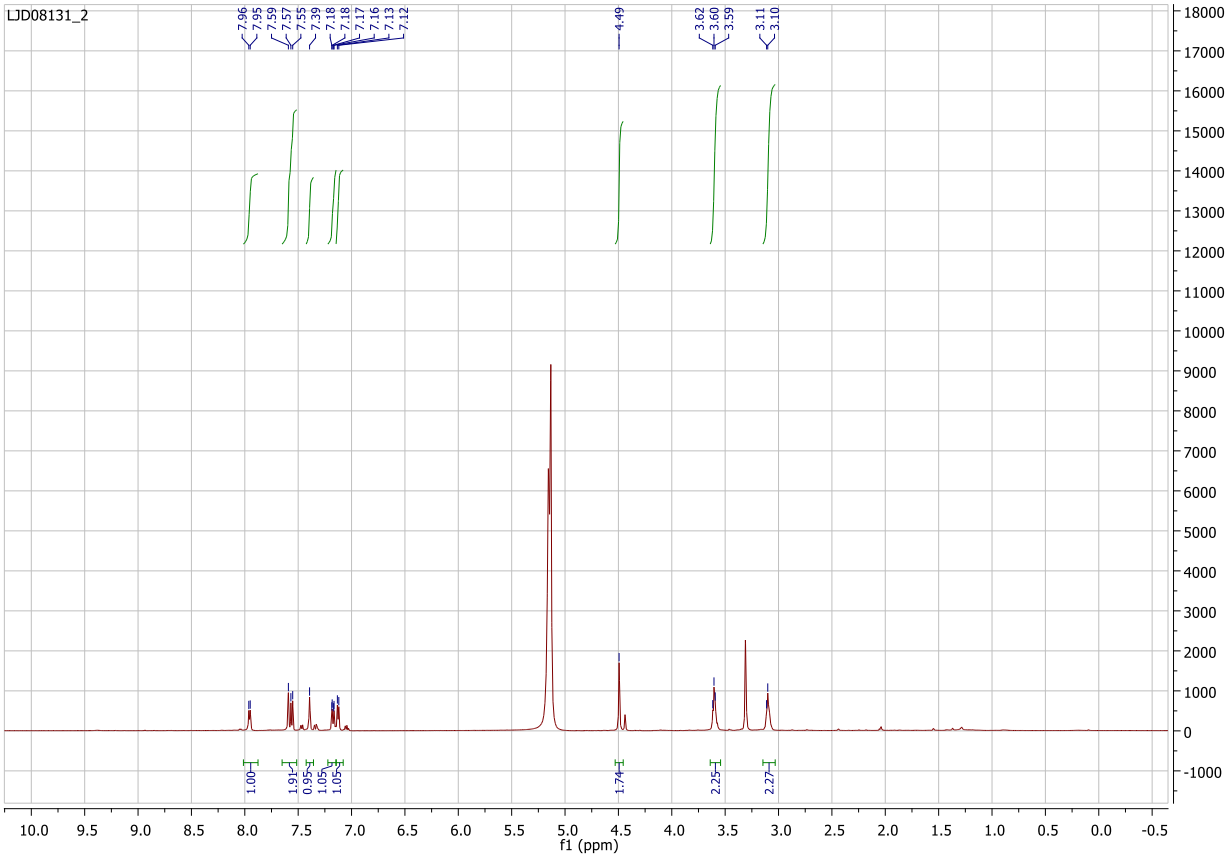
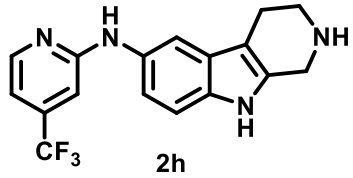


Figure AII.41. ^{13}C NMR spectrum for 2h.



2h

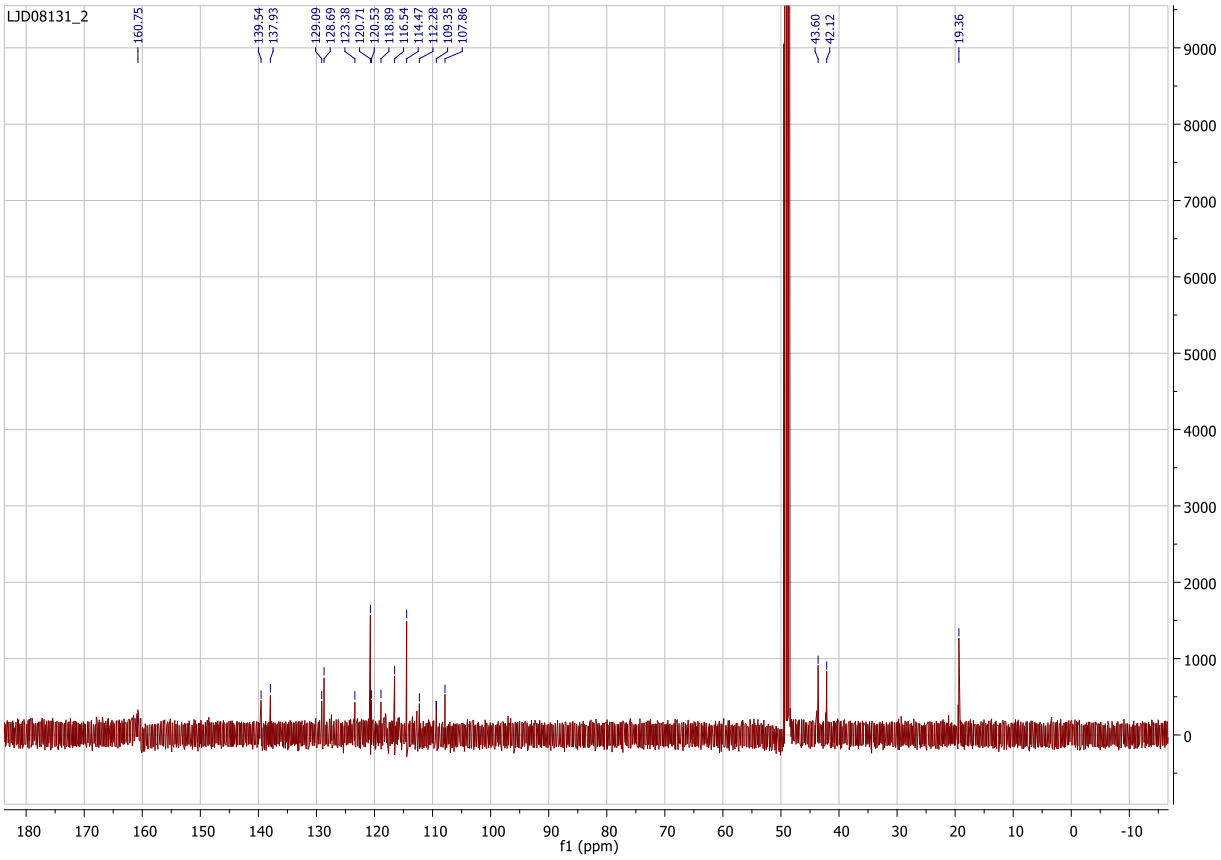
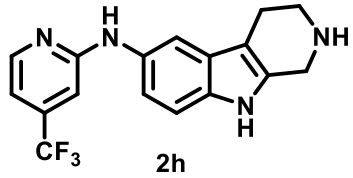


Figure AII.42. ^{19}F NMR spectrum for 2h.



2h

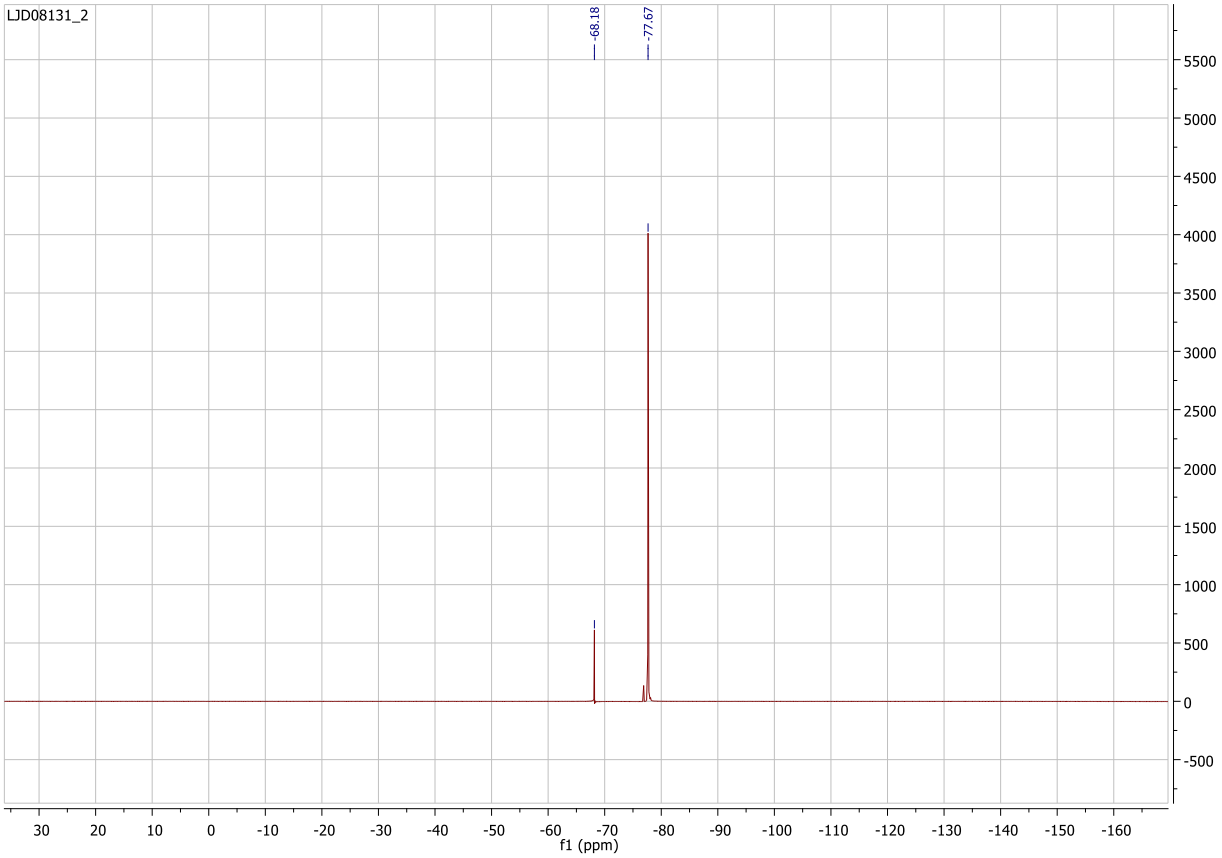
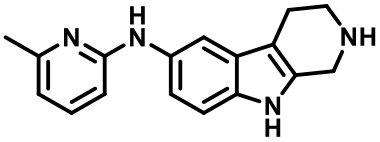


Figure AII.43. ^{1}H NMR spectrum for 2i.



2i

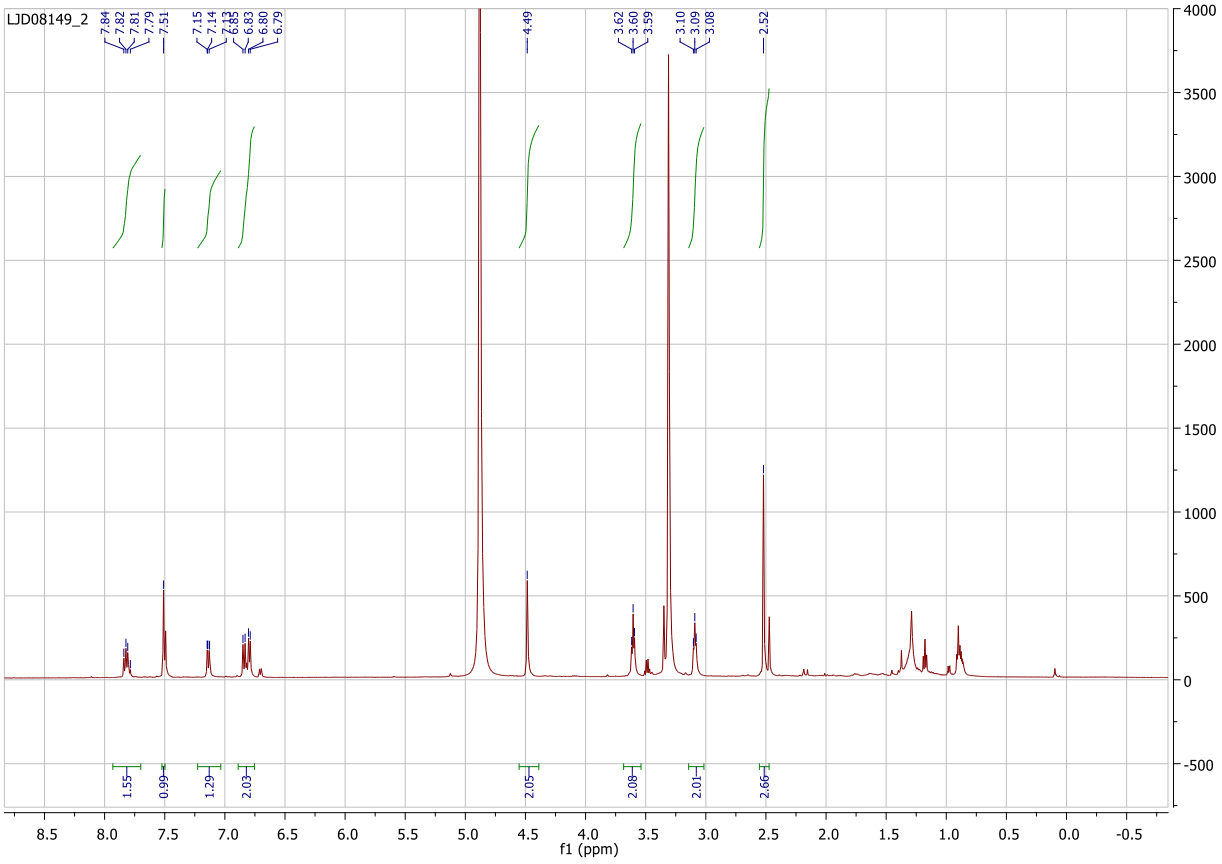
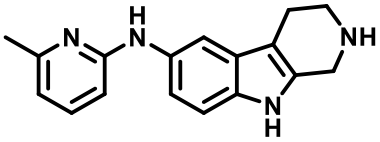


Figure AII.44. ^{13}C NMR spectrum for 2i.



2i

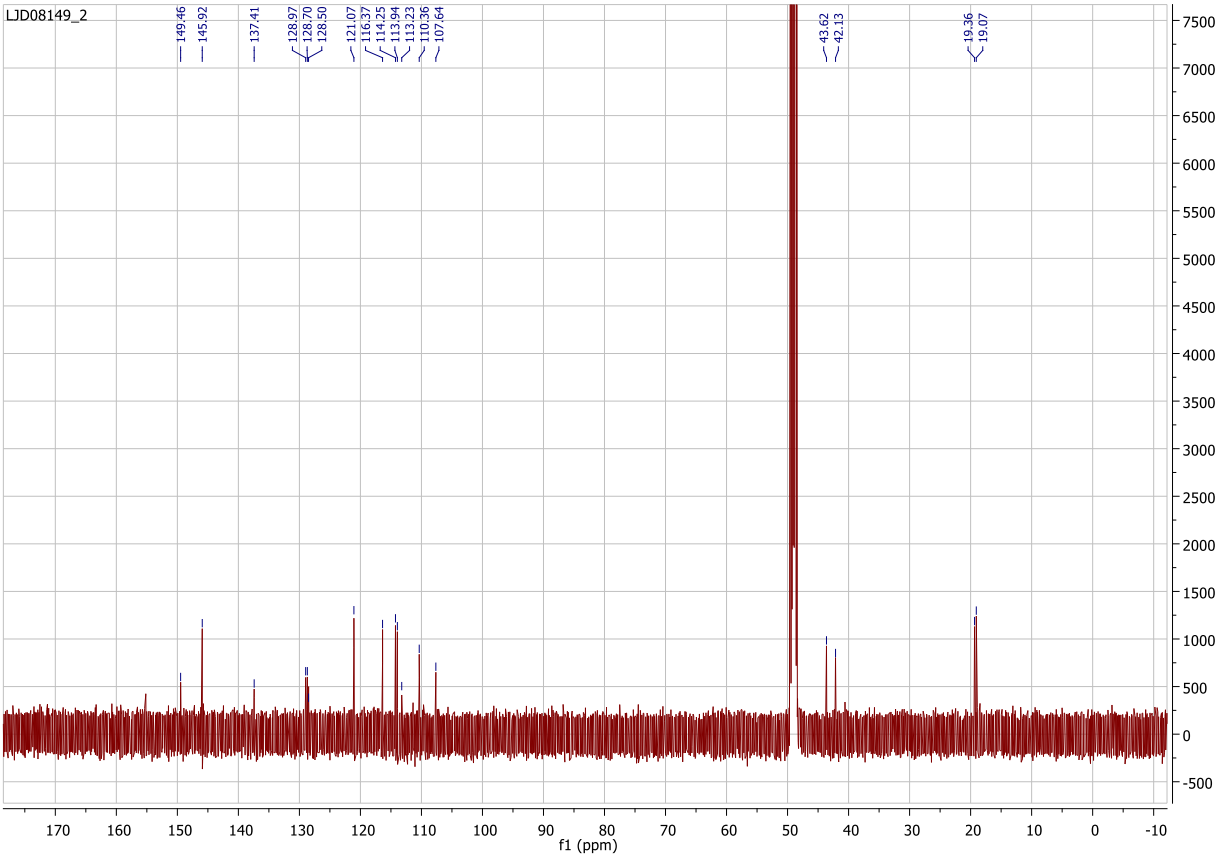
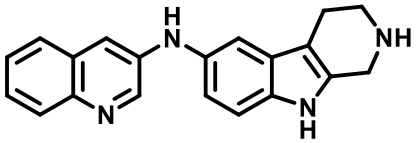


Figure AII.45. ^{1}H NMR spectrum for 2j.



2j

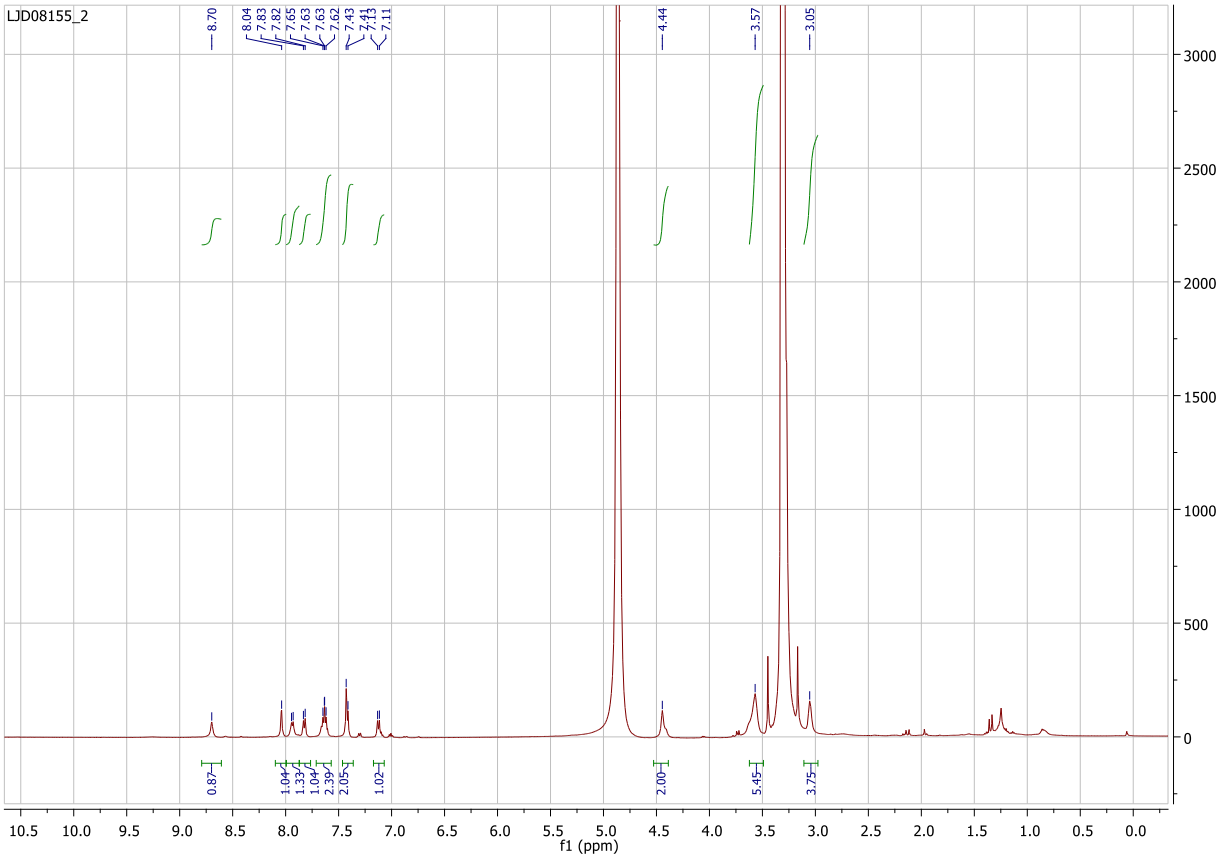
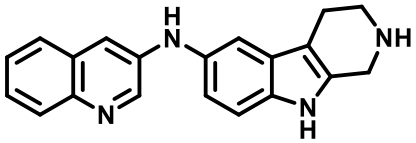


Figure AII.46. ^{13}C NMR spectrum for 2j.



2j

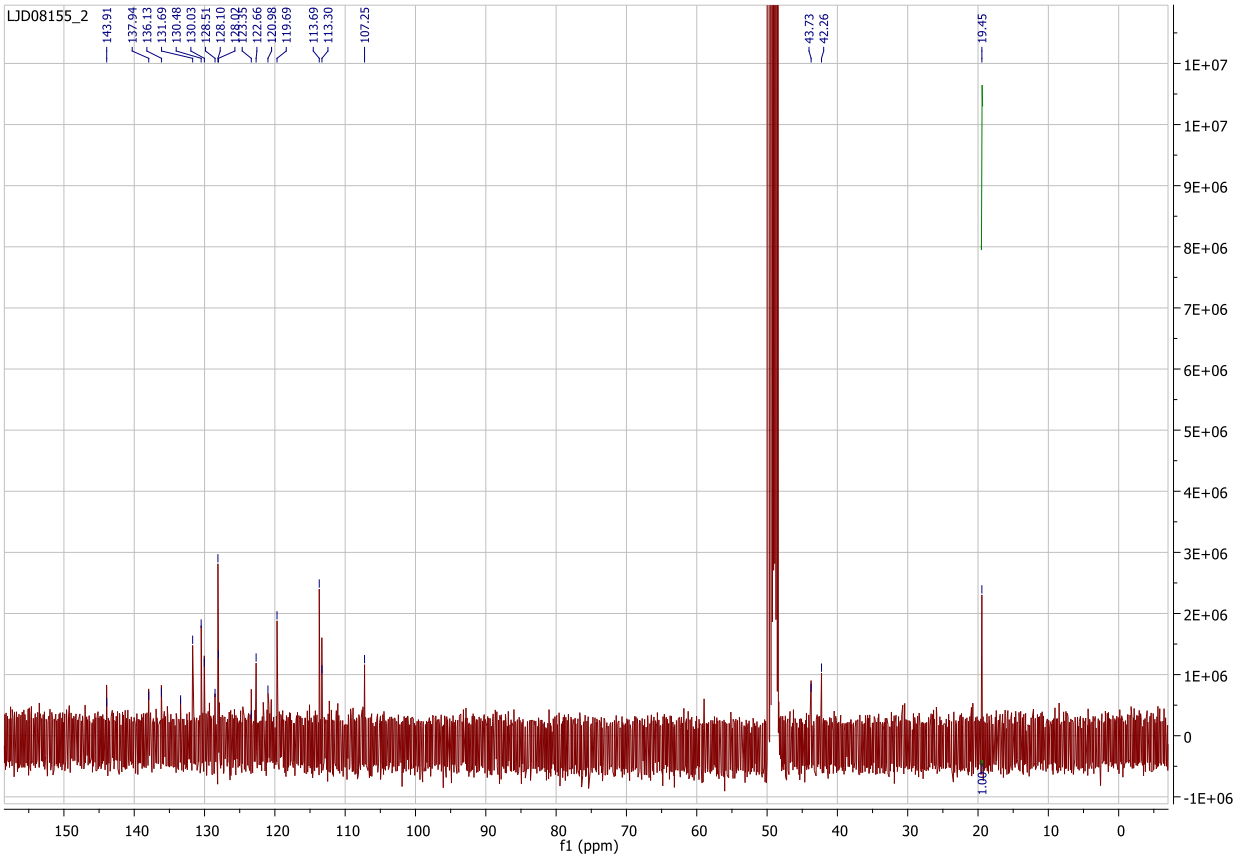
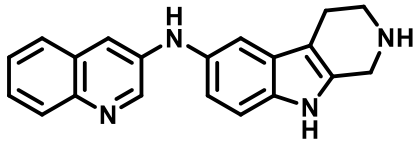


Figure AII.47. Closeup of ^{13}C NMR spectrum for 2j.



2j

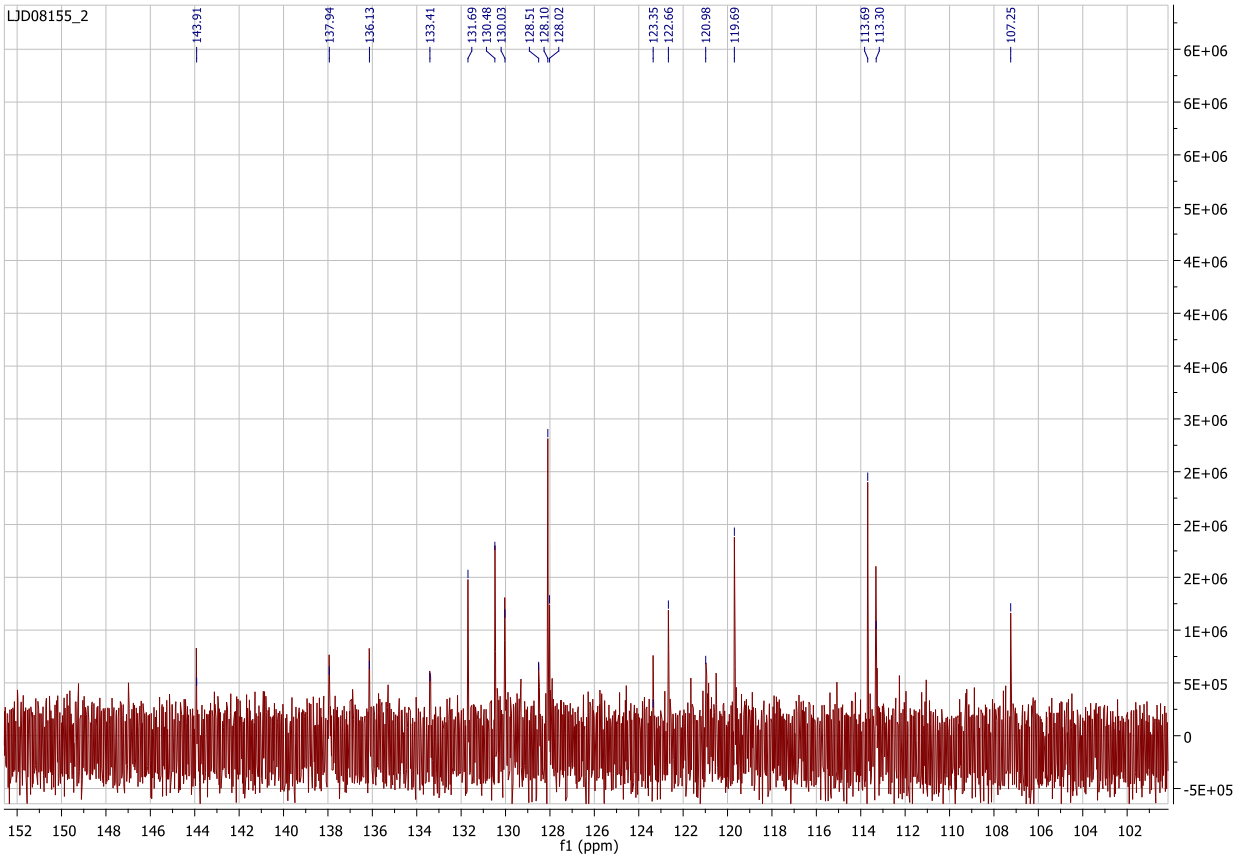
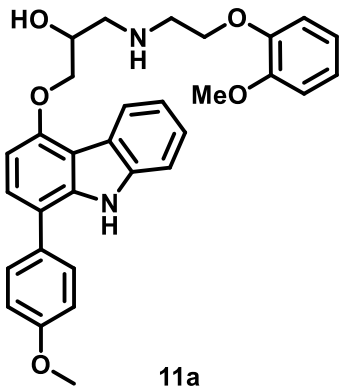


Figure AII.48. ^{1}H NMR spectrum for 11a.



11a

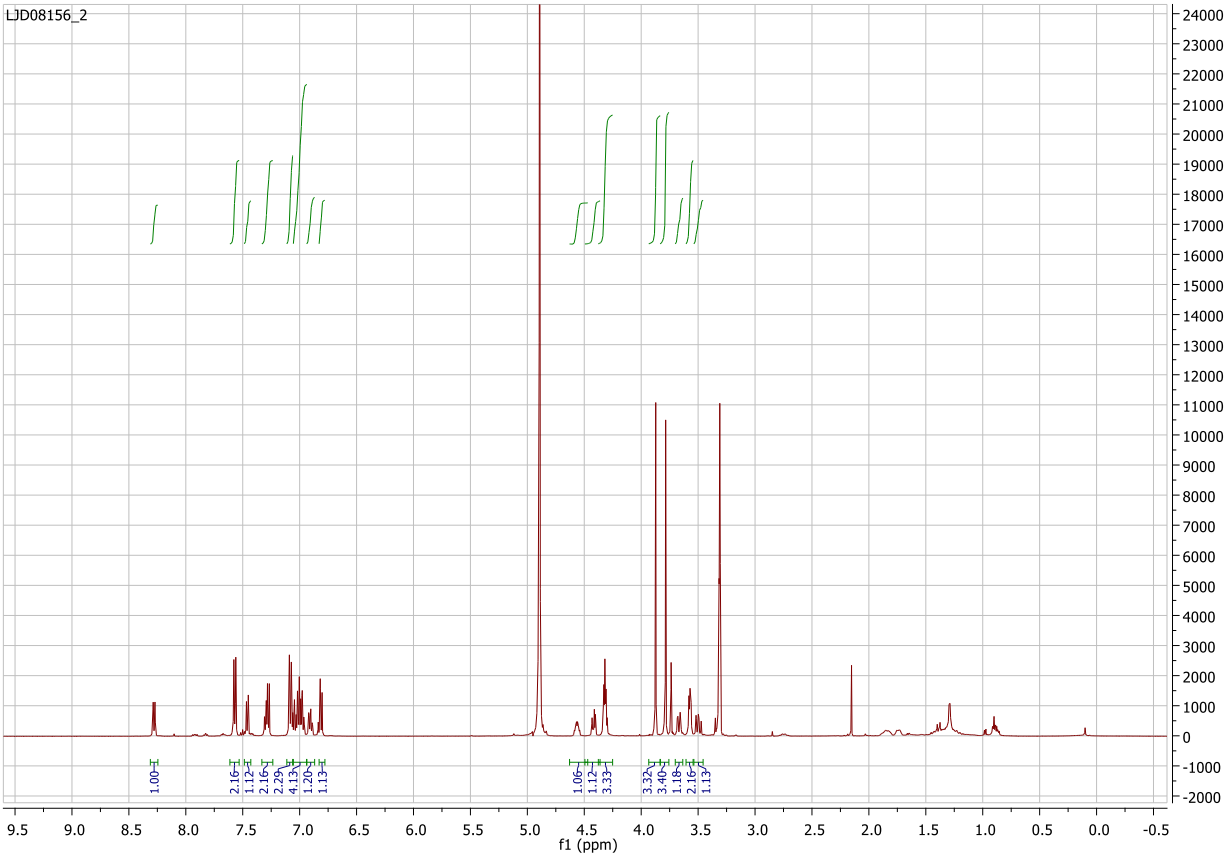
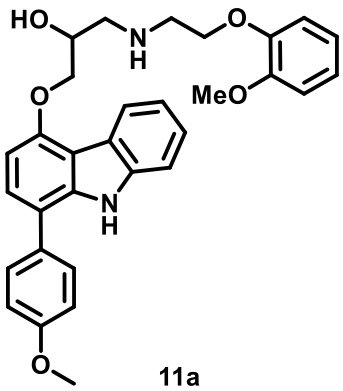


Figure AII.49. ^{13}C NMR spectrum for 11a.



11a

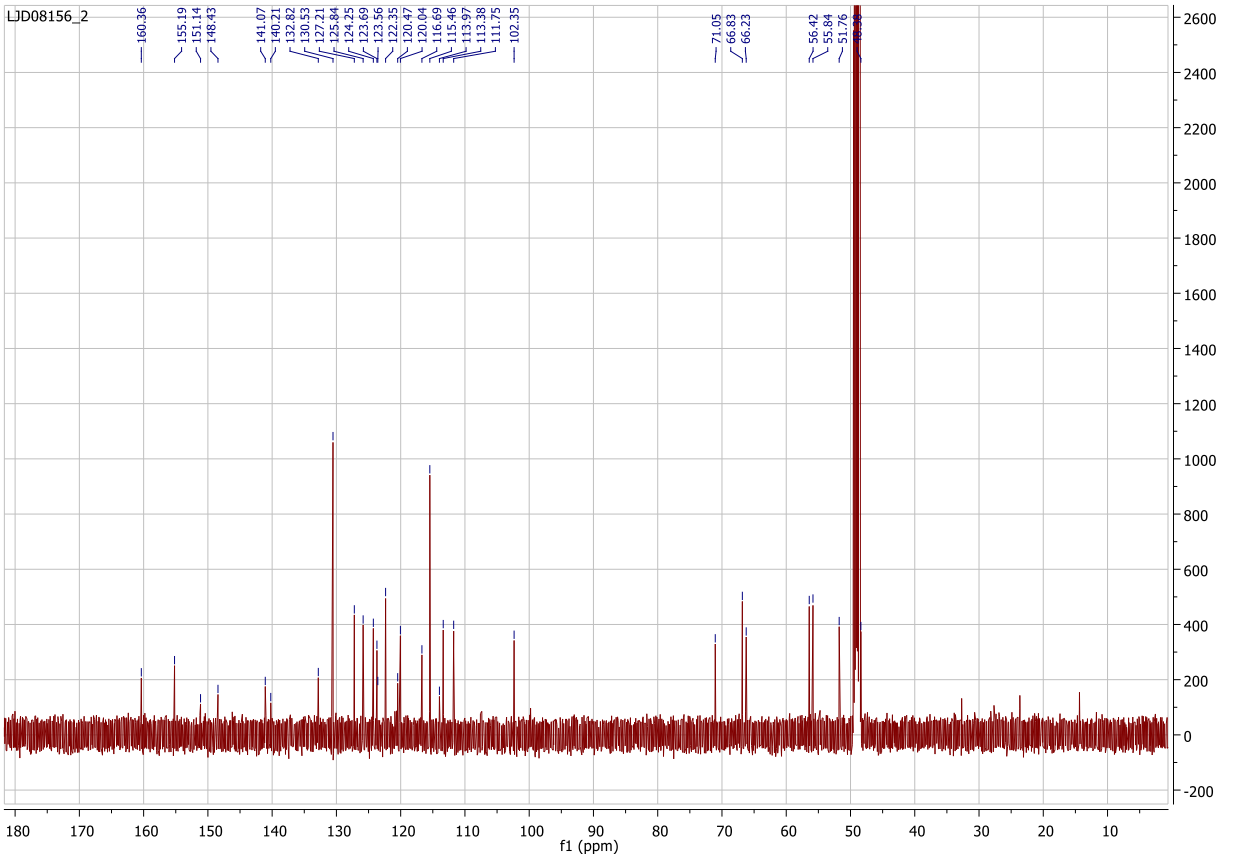
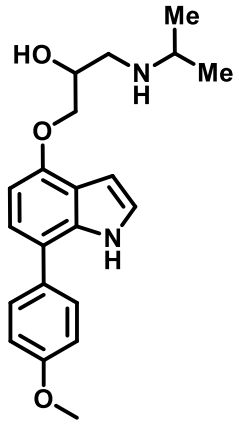


Figure AII.50. ^{1}H NMR spectrum for 9a.



9a

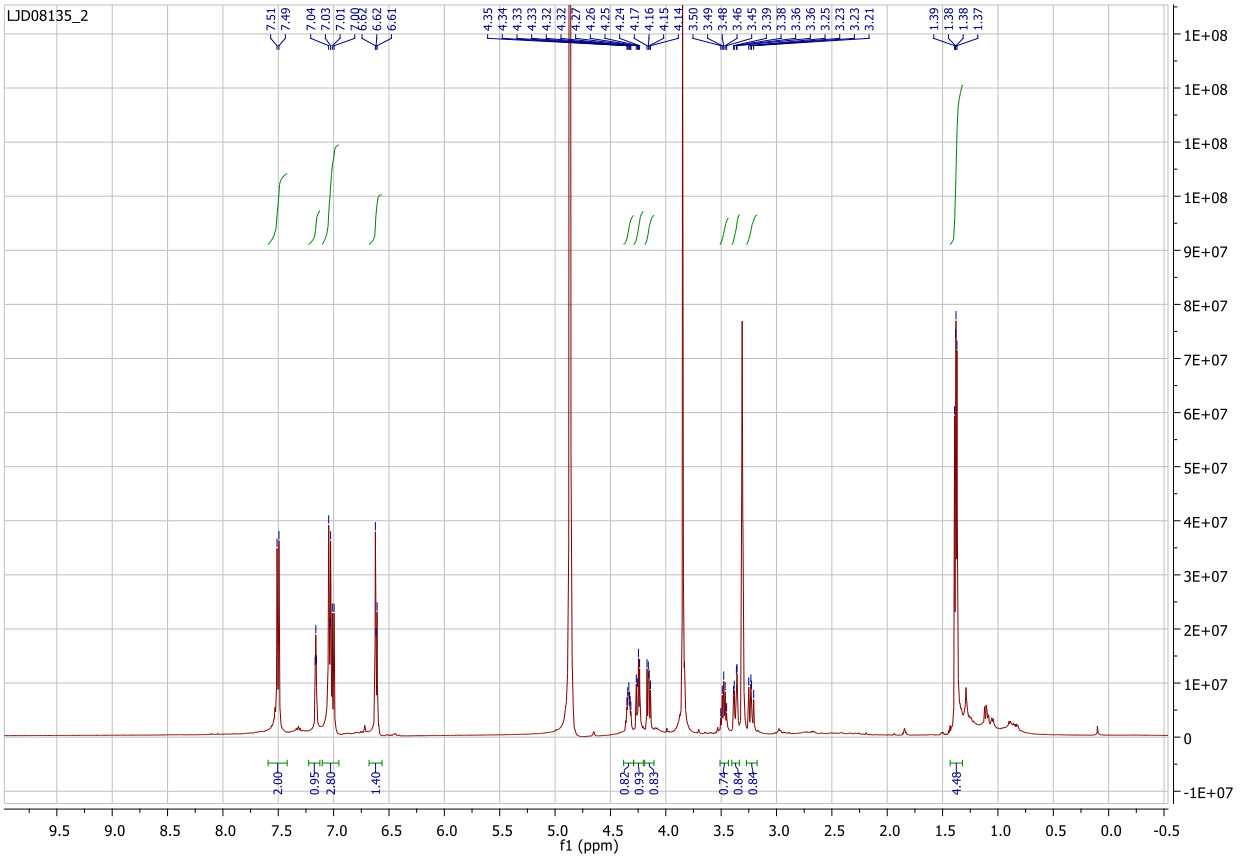
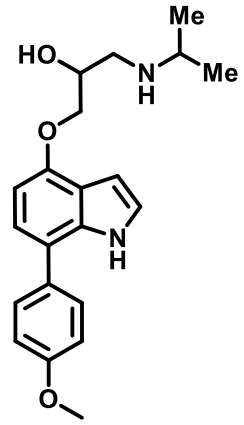


Figure AII.51. ^{13}C NMR spectrum for 9a.



9a

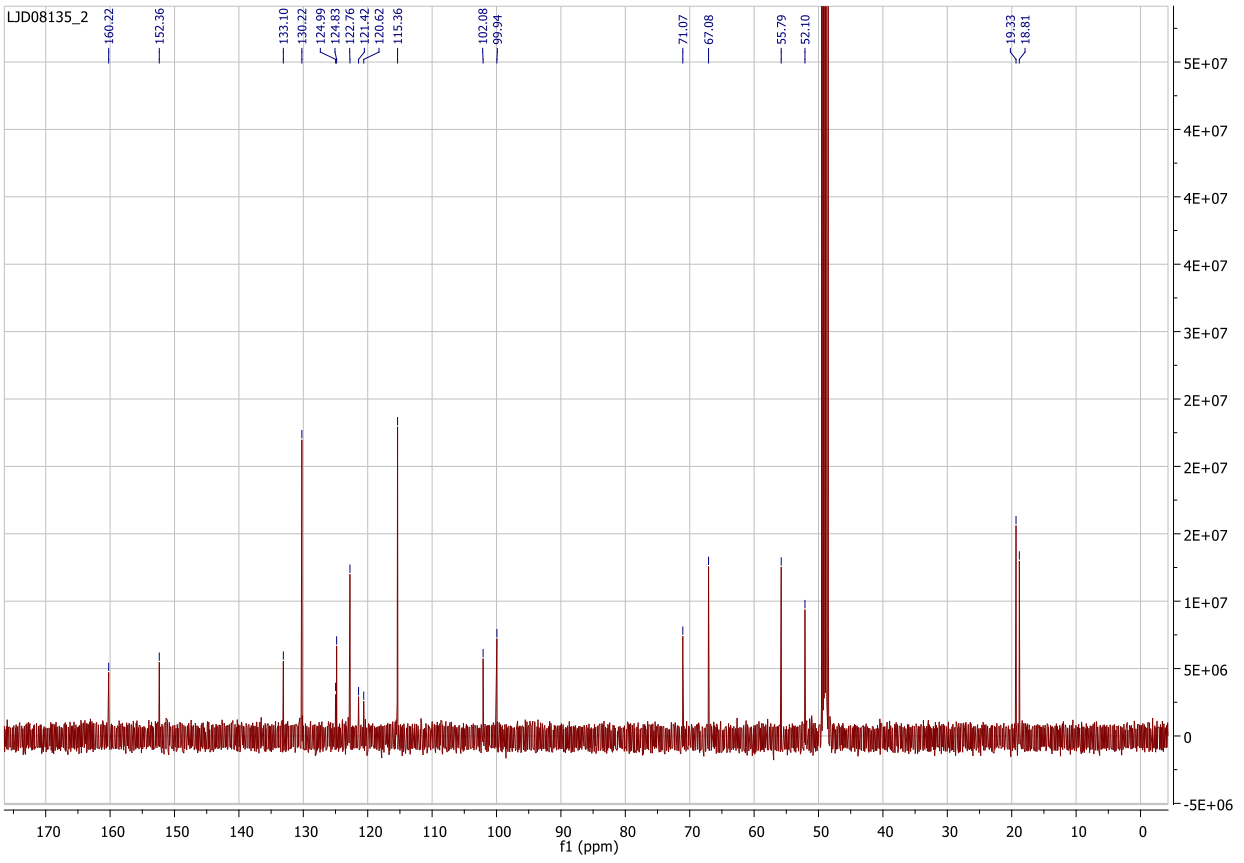
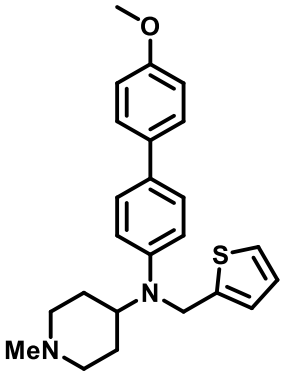


Figure AII.52. ^{1}H NMR spectrum for 12a.



12a

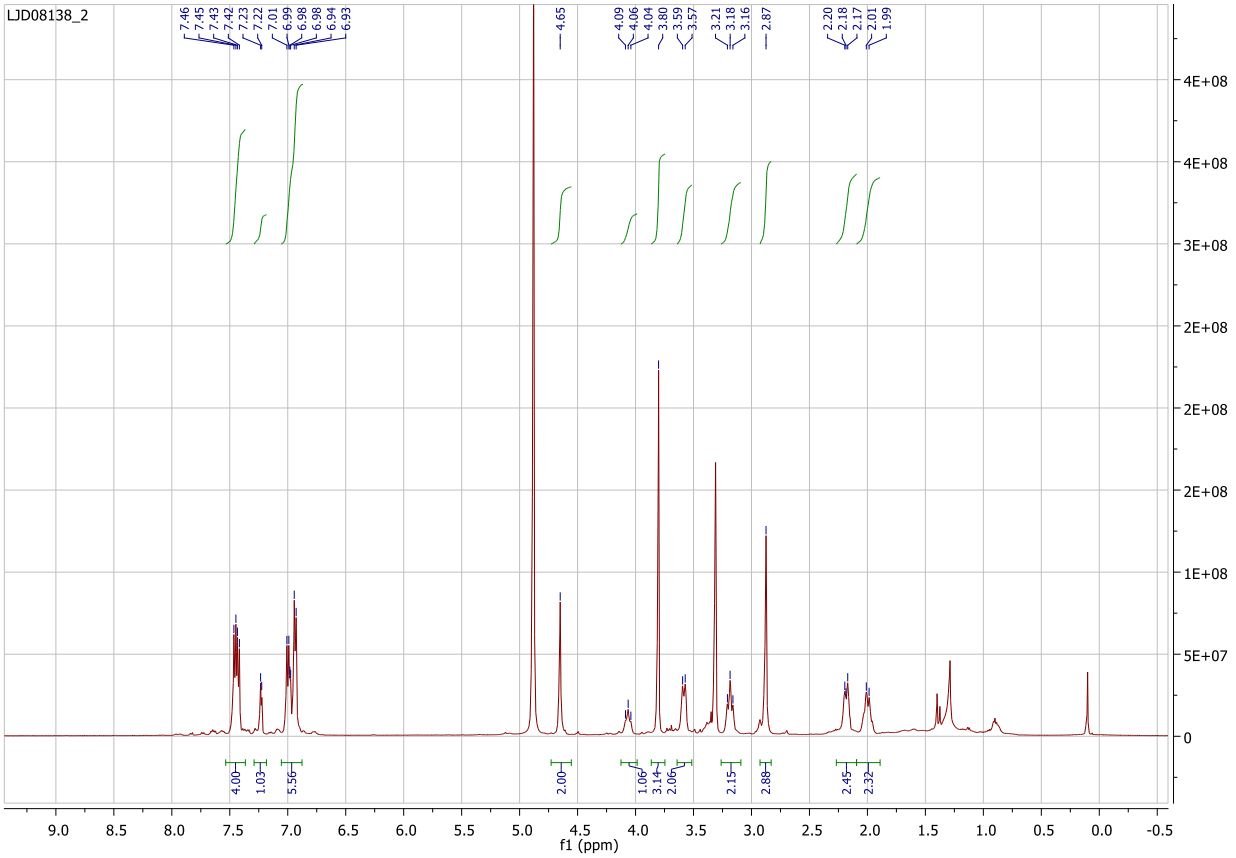
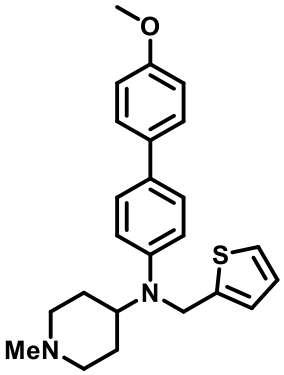


Figure AII.53. ^{13}C NMR spectrum for 12a.



12a

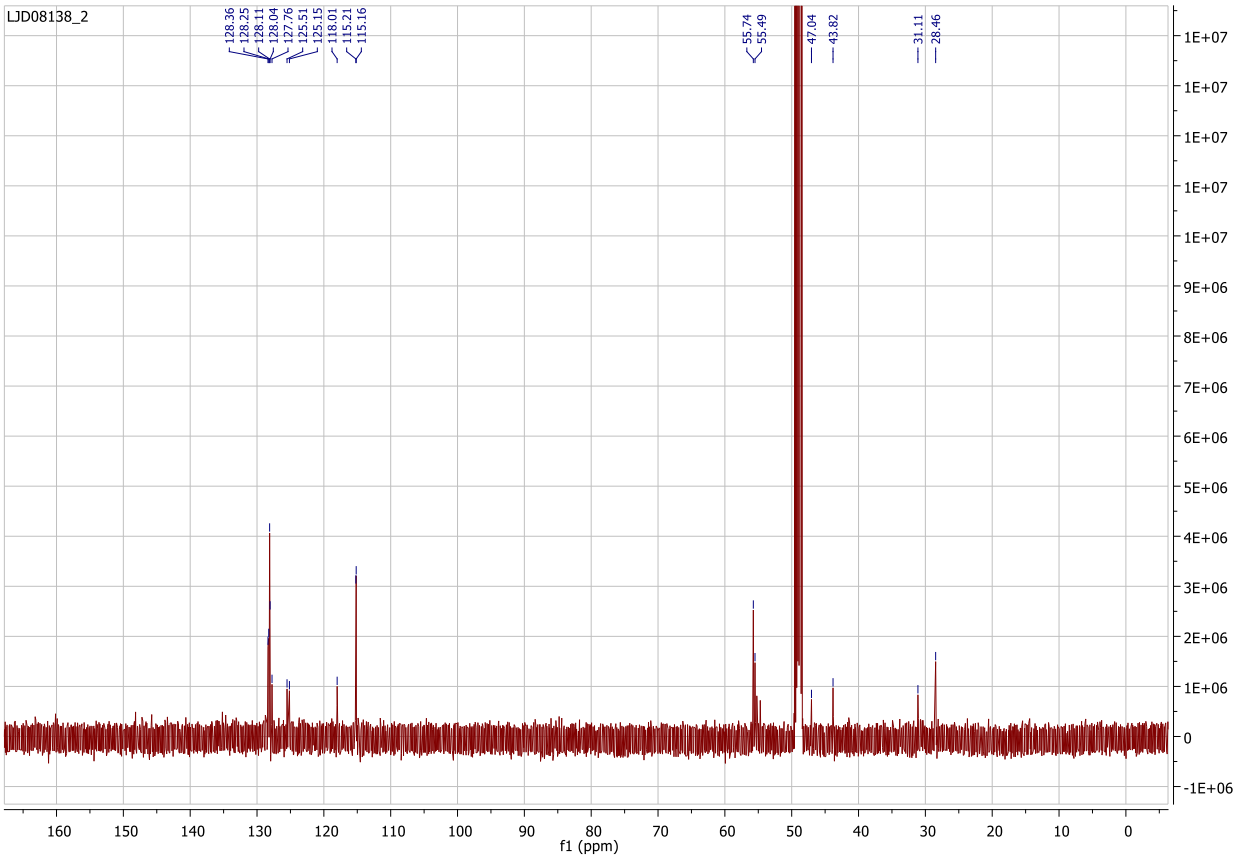
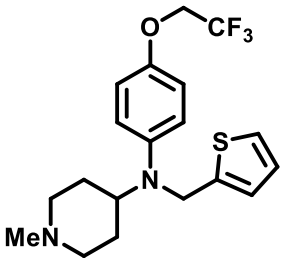


Figure AII.54. ^{1}H NMR spectrum for 12b.



12b

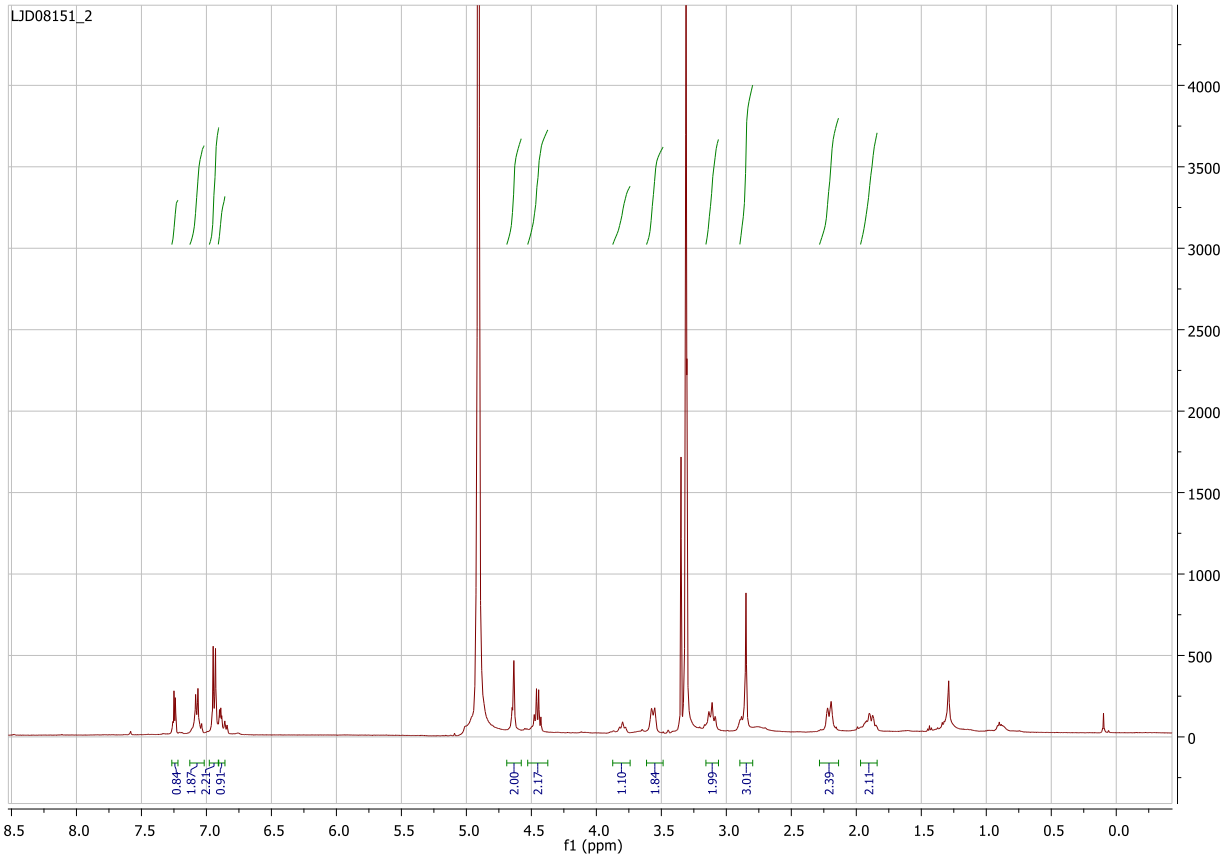
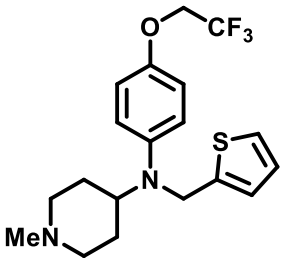


Figure AII.55. ^{13}C NMR spectrum for 12b.



12b

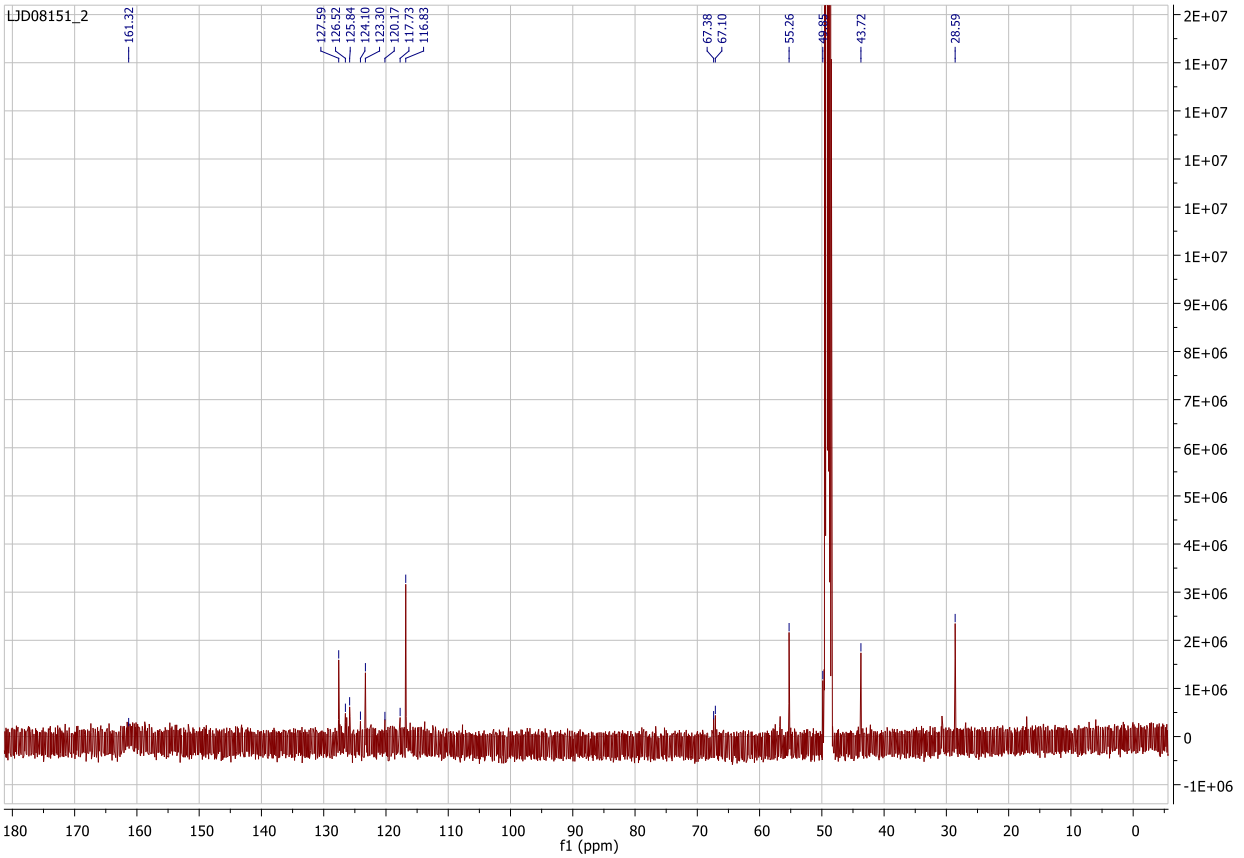
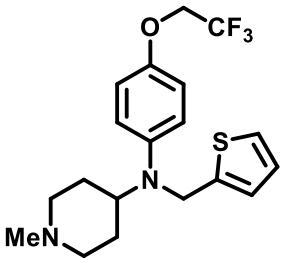
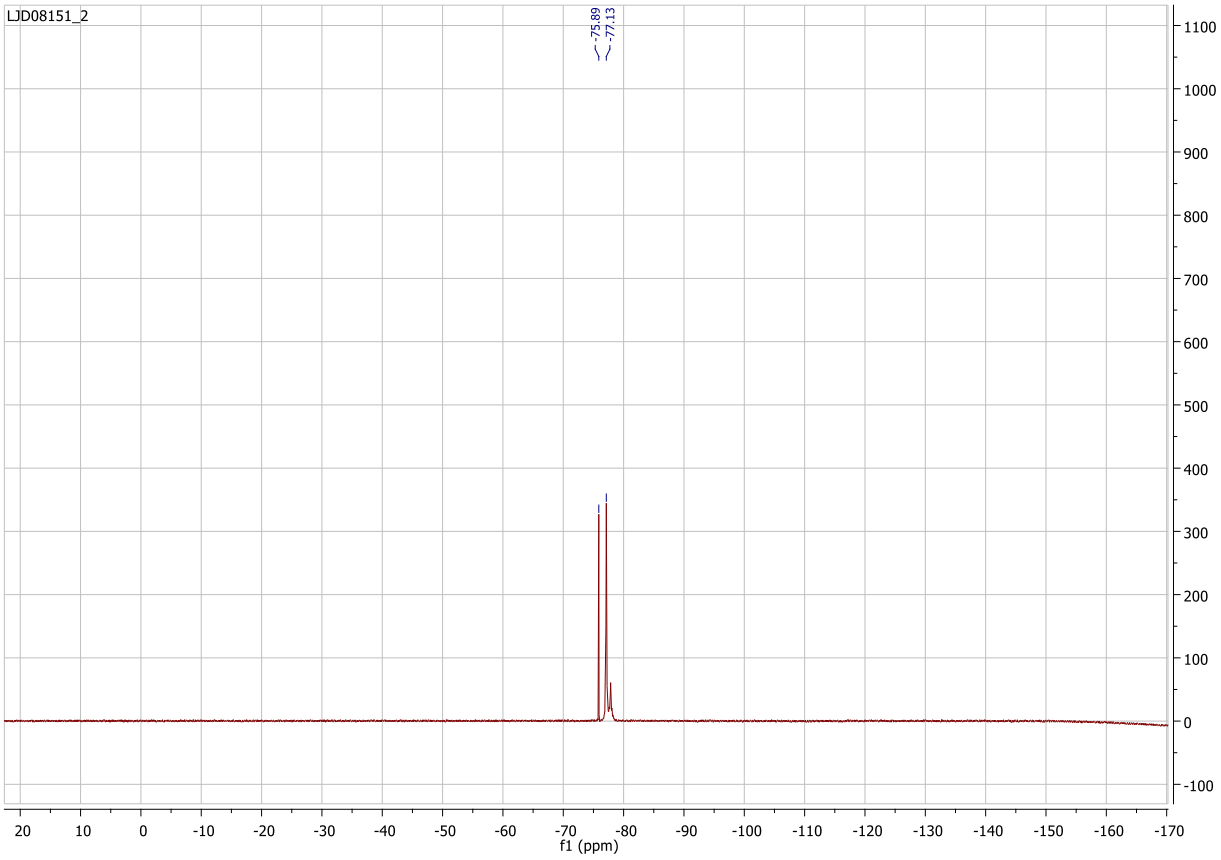


Figure AII.56. ^{19}F NMR spectrum for 12b.



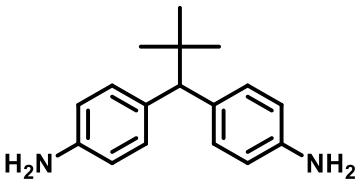
12b



APPENDIX III

NMR SPECTRA FOR COMPOUNDS FROM CHAPTER IV

Figure AIII.1. ^{1}H NMR spectrum of 2.



2

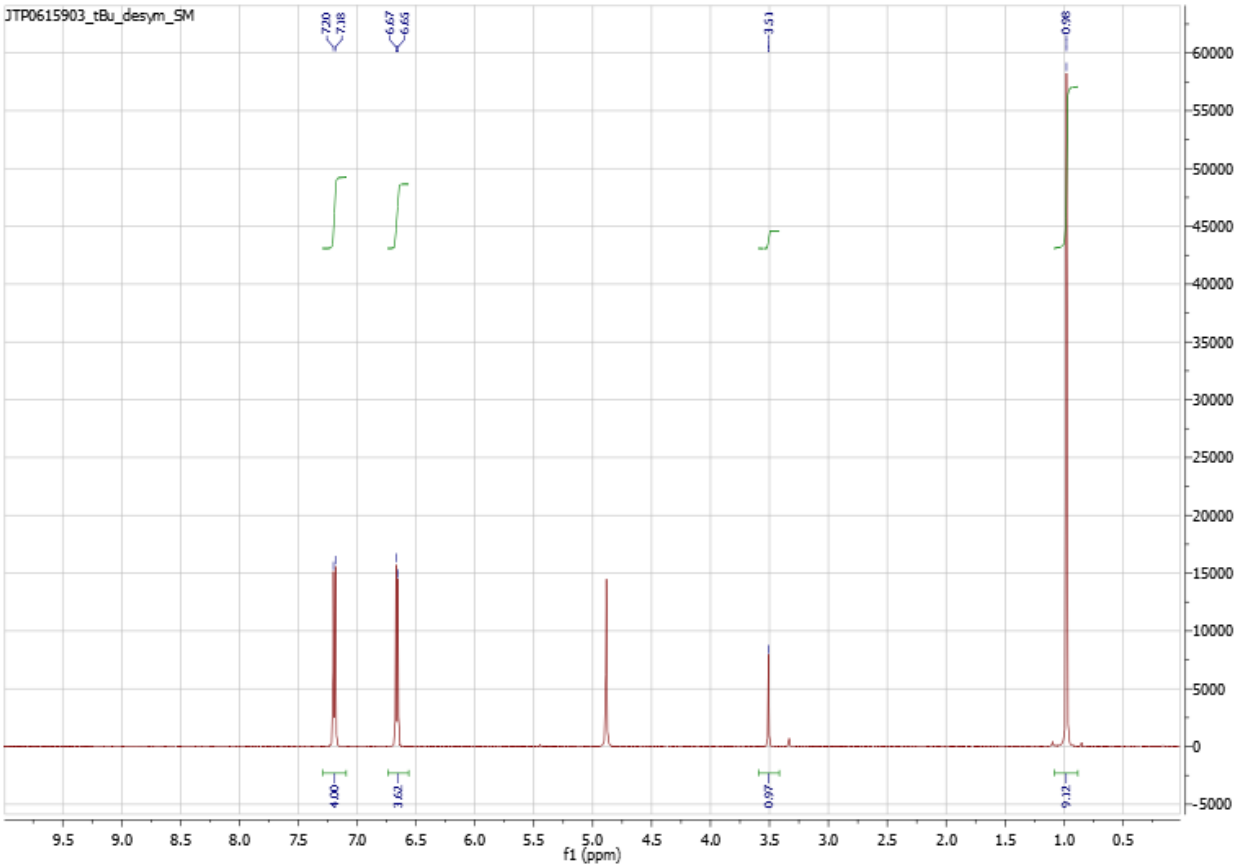
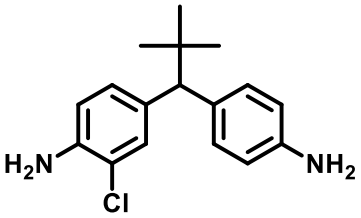


Figure AIII.2. ^{1}H NMR spectrum of 2a.



2a

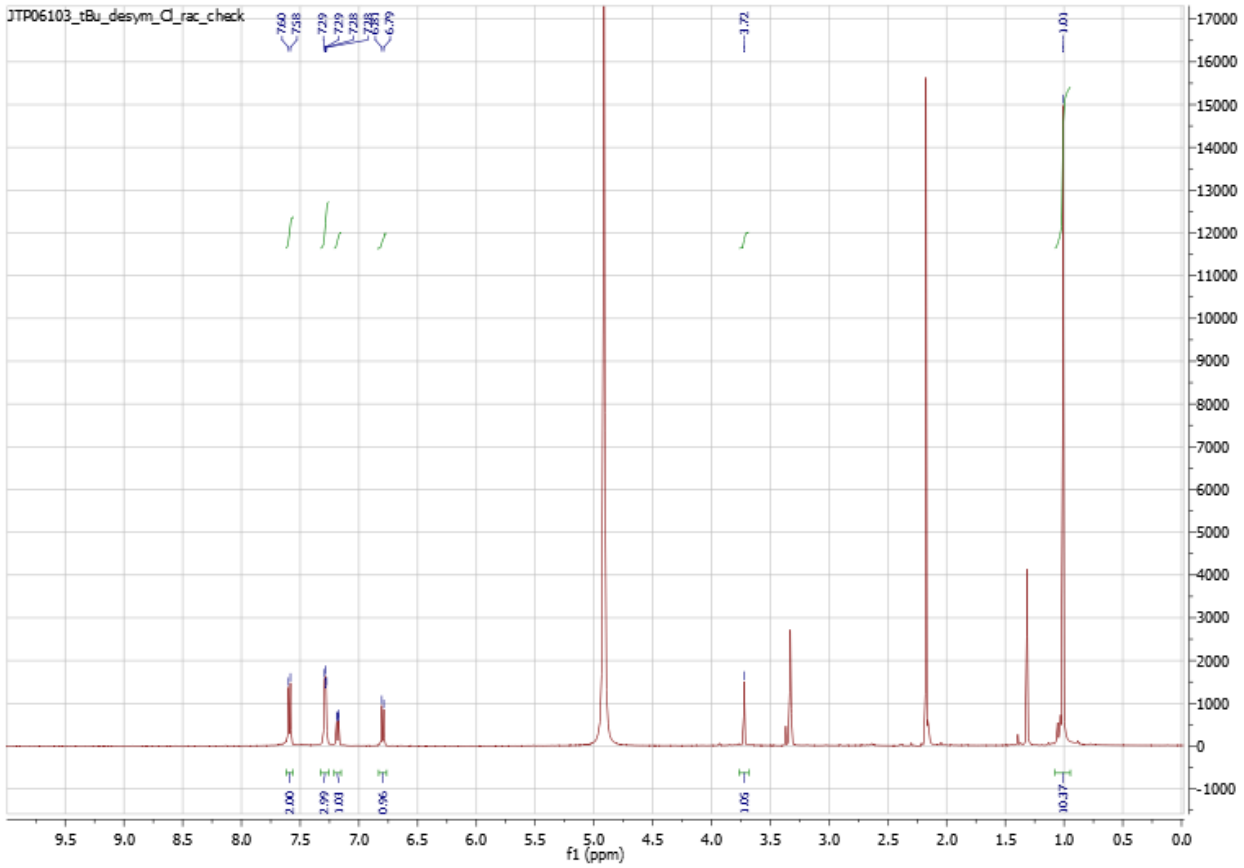
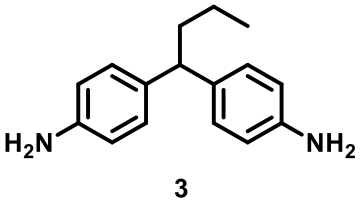


Figure AIII.3. ^{1}H NMR spectrum of 3.



3

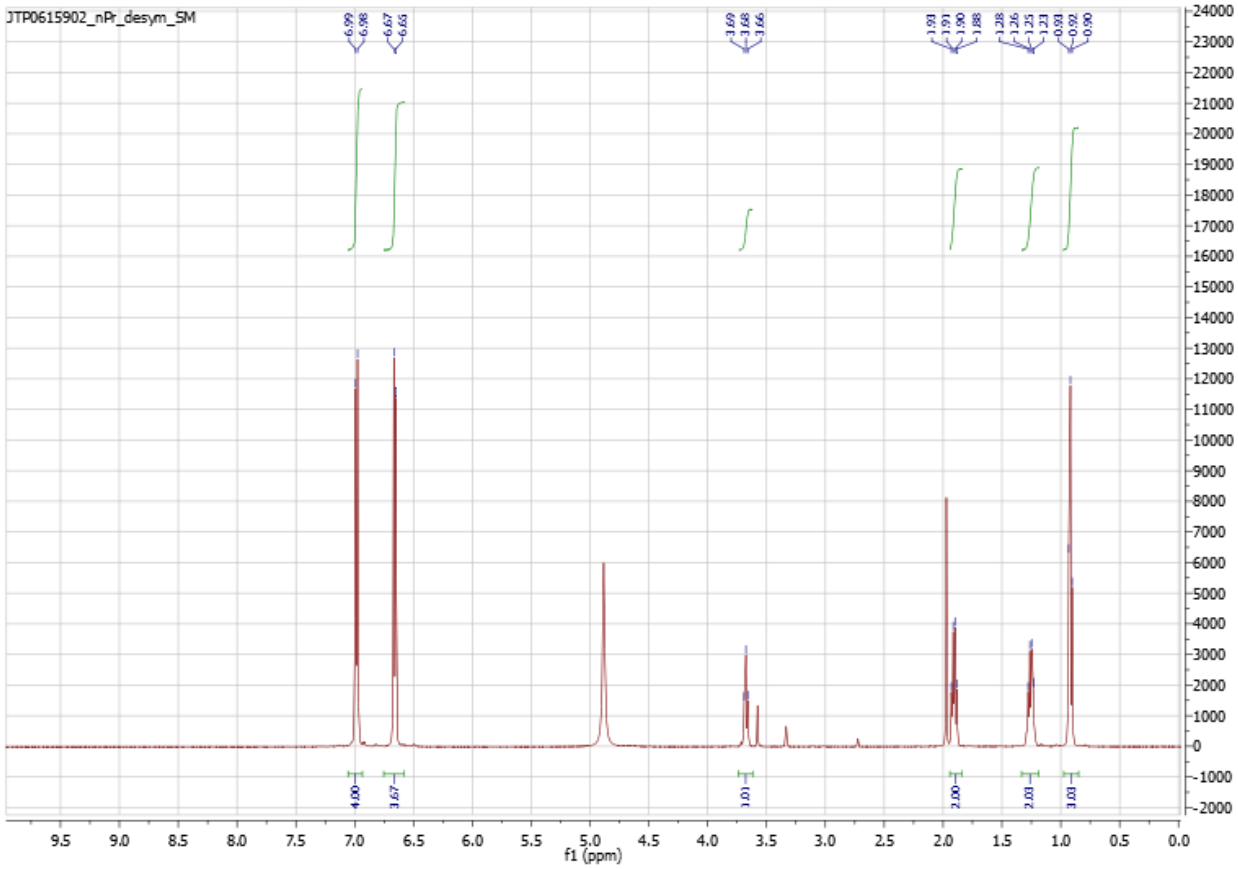
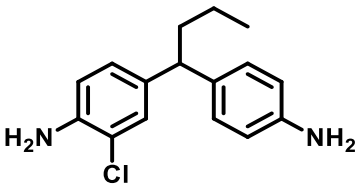


Figure AIII.4. ^{1}H NMR spectrum of 3a.



3a

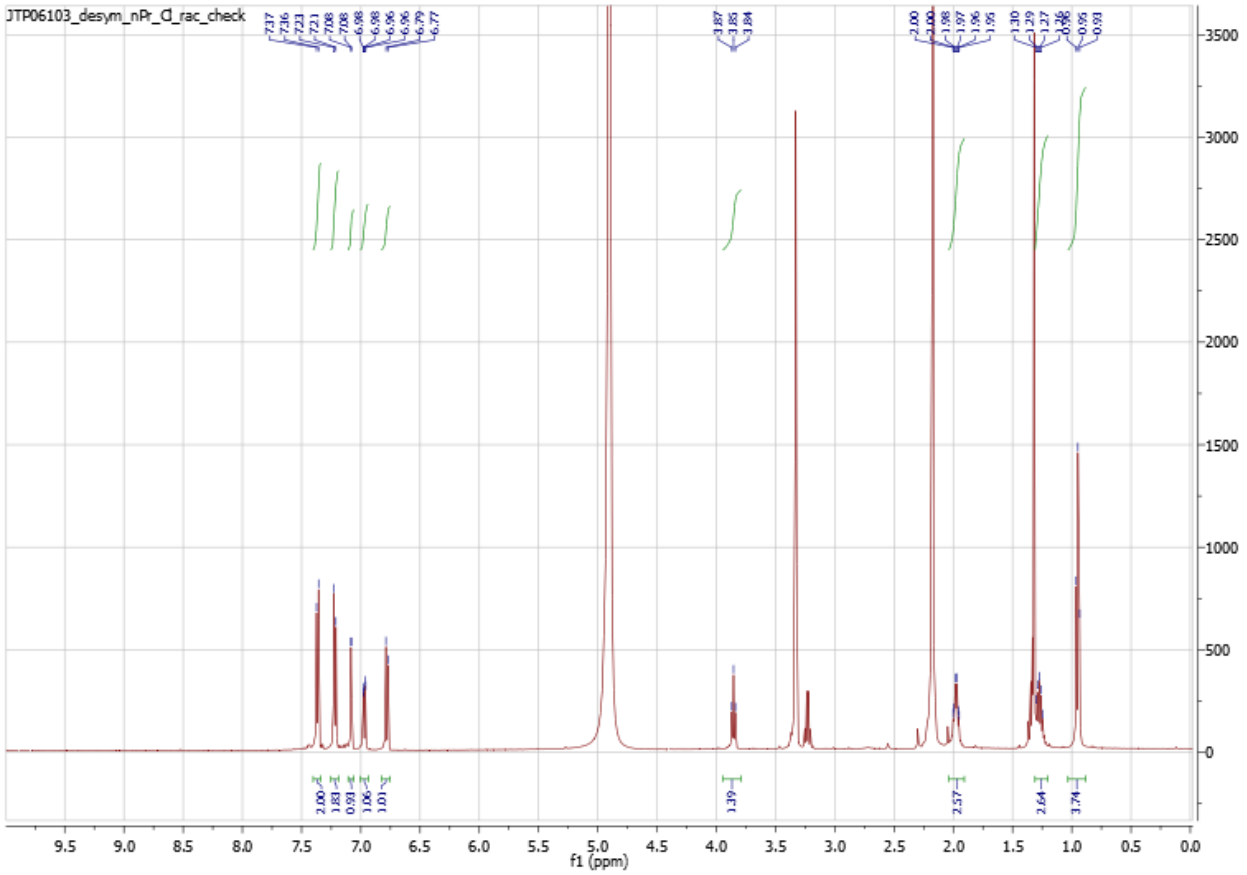
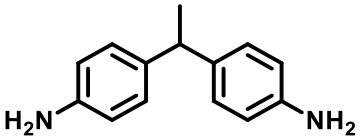


Figure AIII.5. ^{1}H NMR spectrum of 4.



4

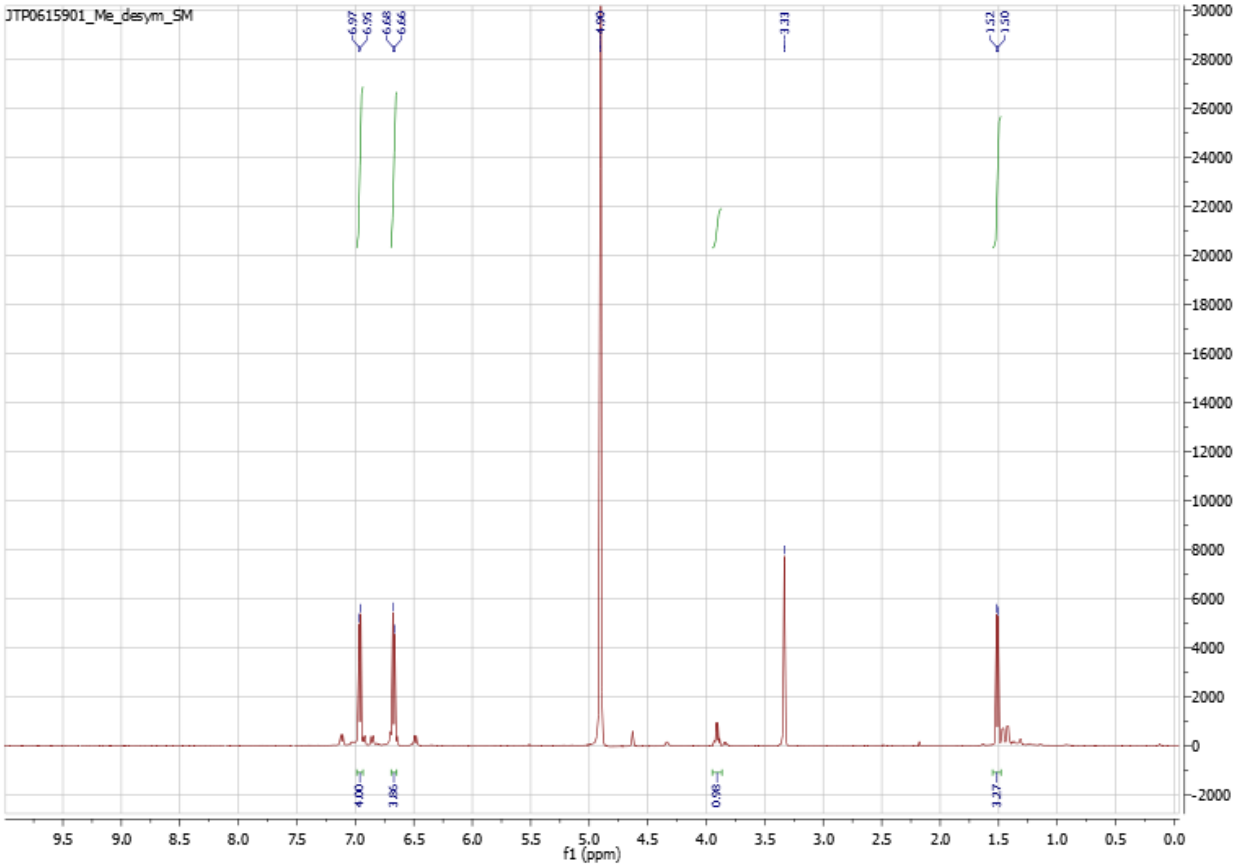
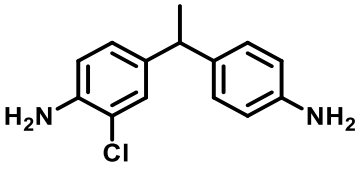


Figure AIII.6. ^{1}H NMR spectrum of 4a.



4a

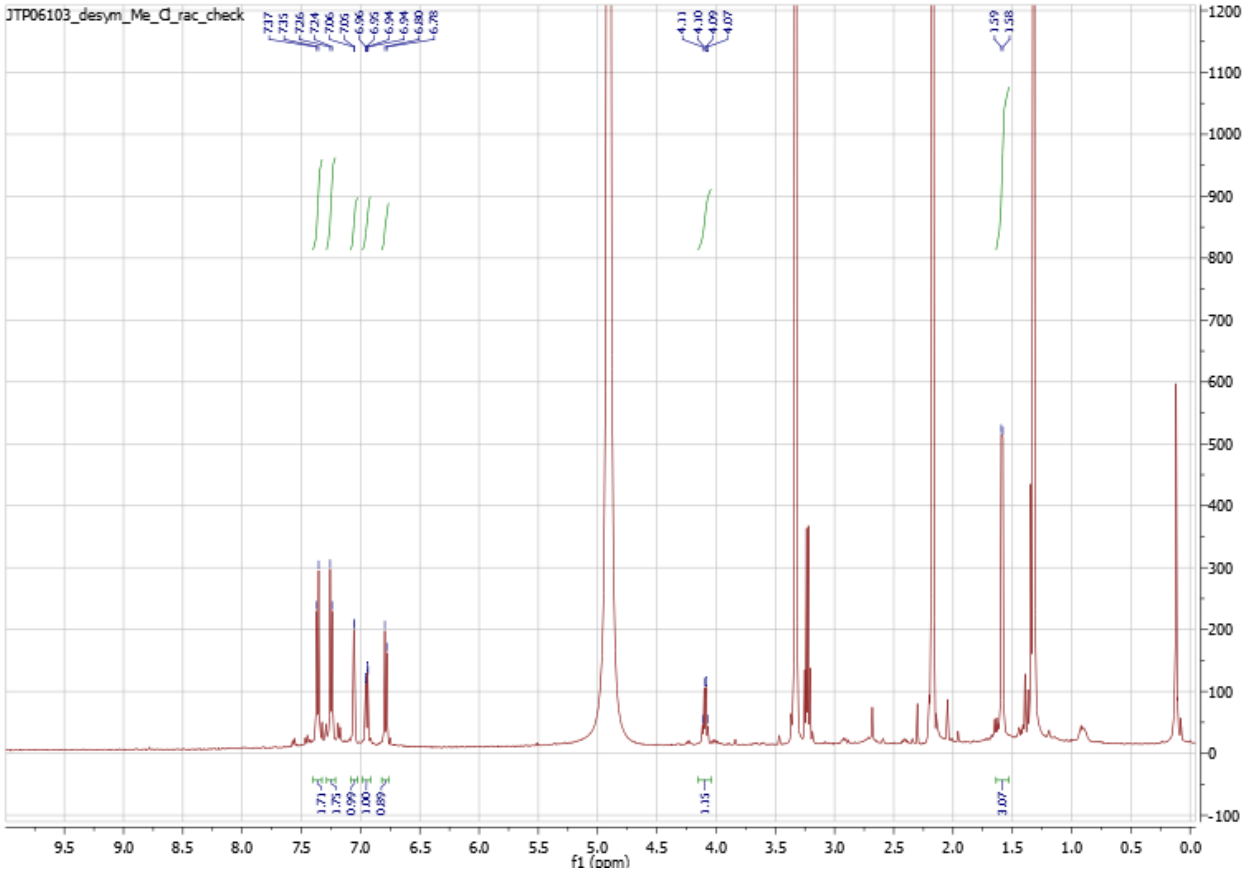
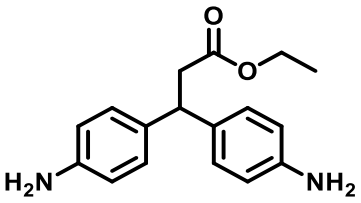


Figure AIII.7. ^{1}H NMR spectrum of 5.



5

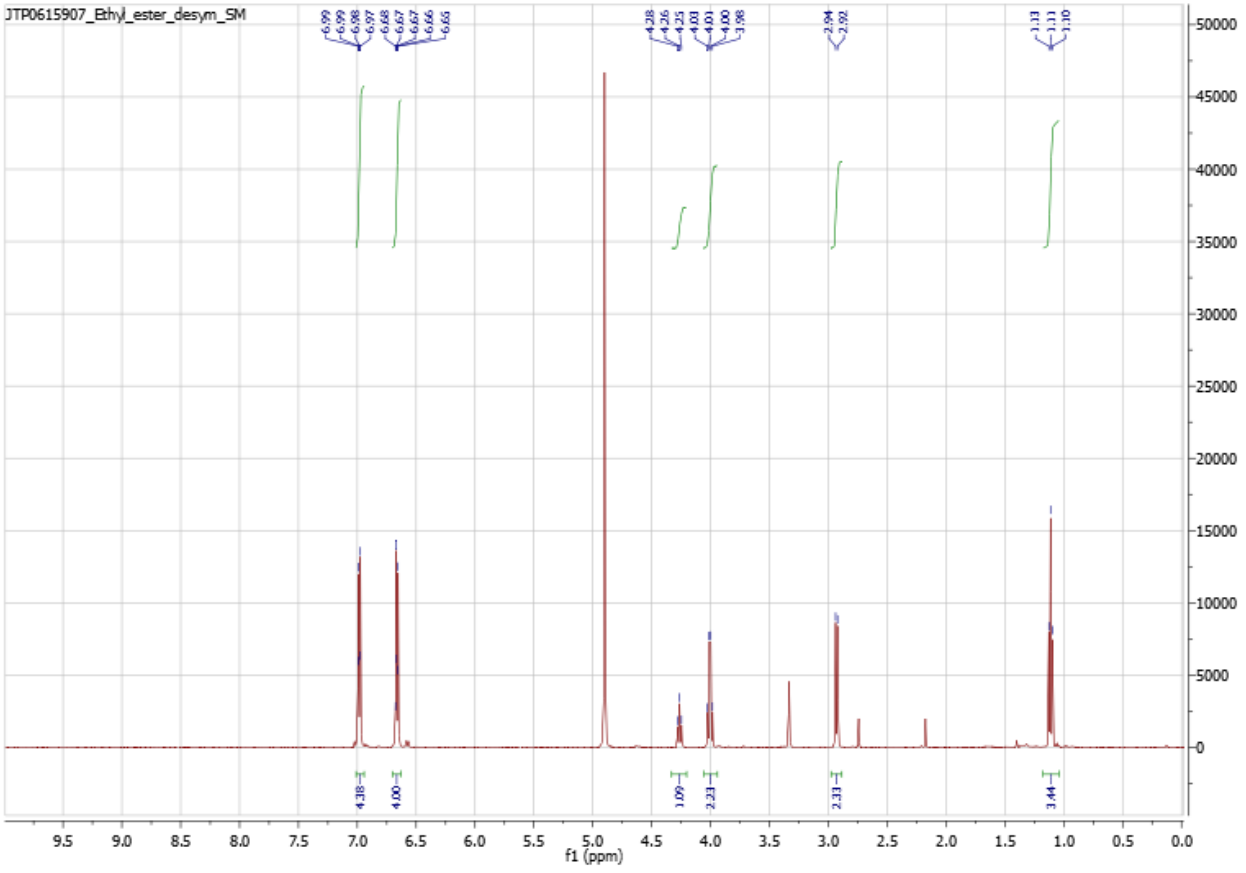
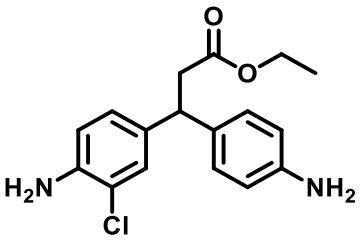


Figure AIII.8. ^{1}H NMR spectrum of 5a.



5a

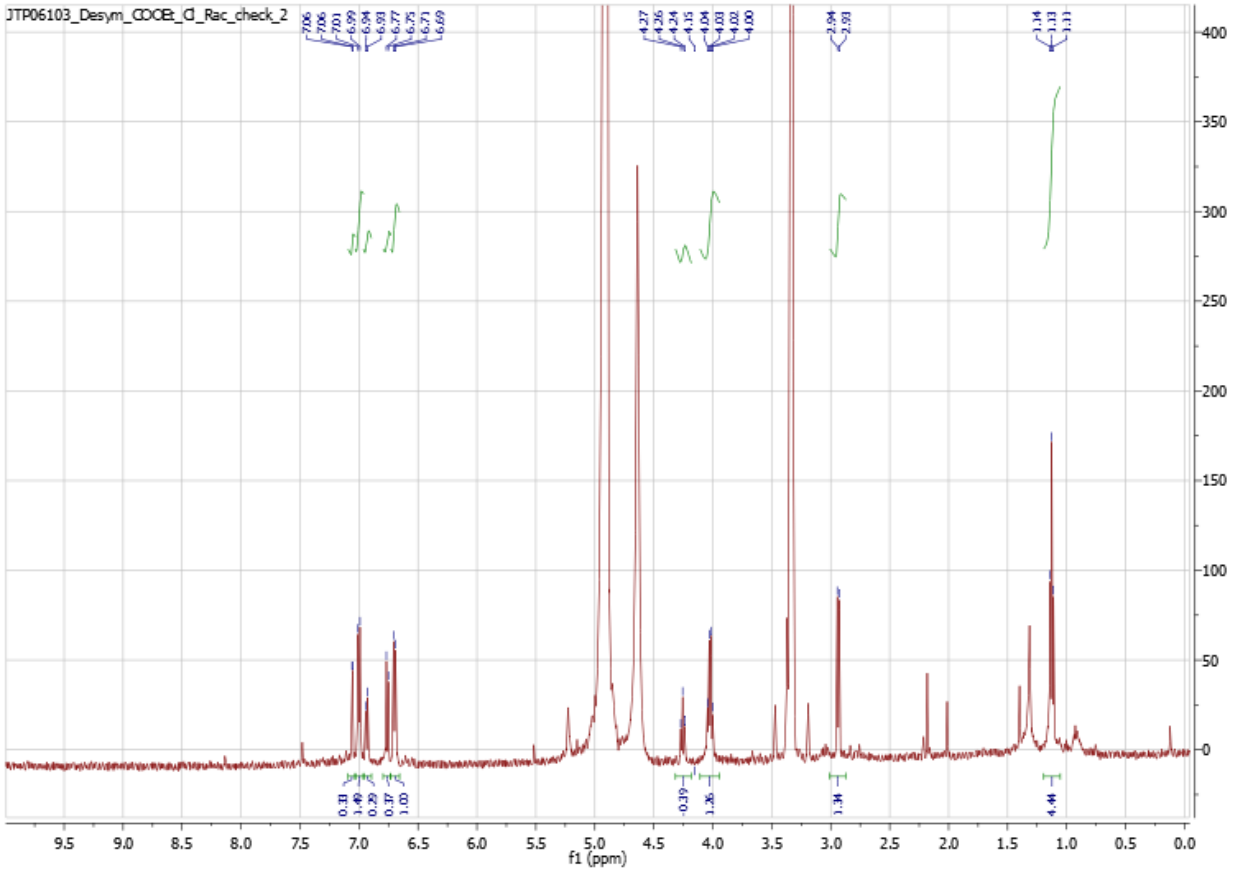
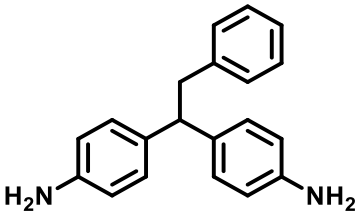


Figure AIII.9. ^{1}H NMR spectrum of 6.



6

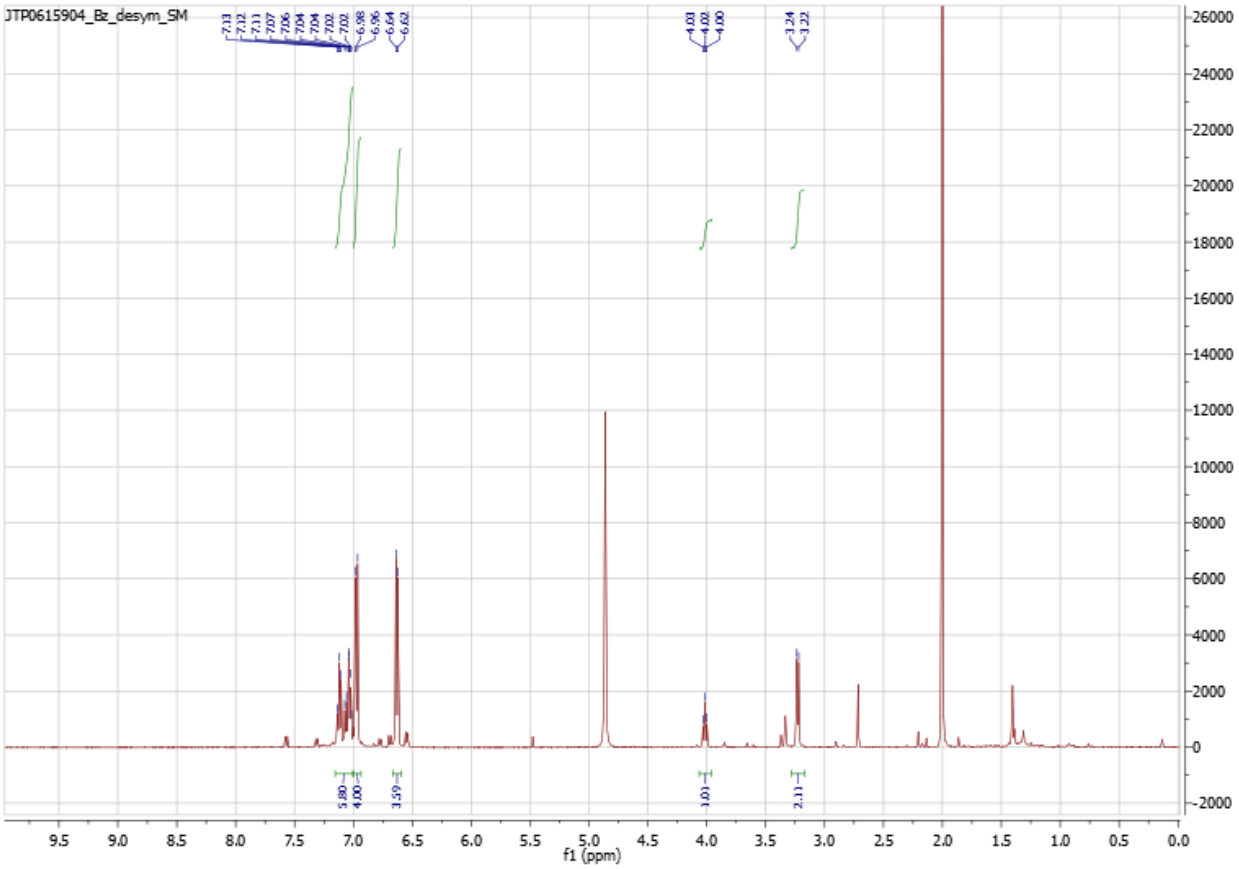
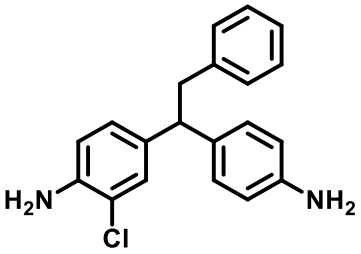


Figure AIII.10. ^{1}H NMR spectrum of 6a.



6a

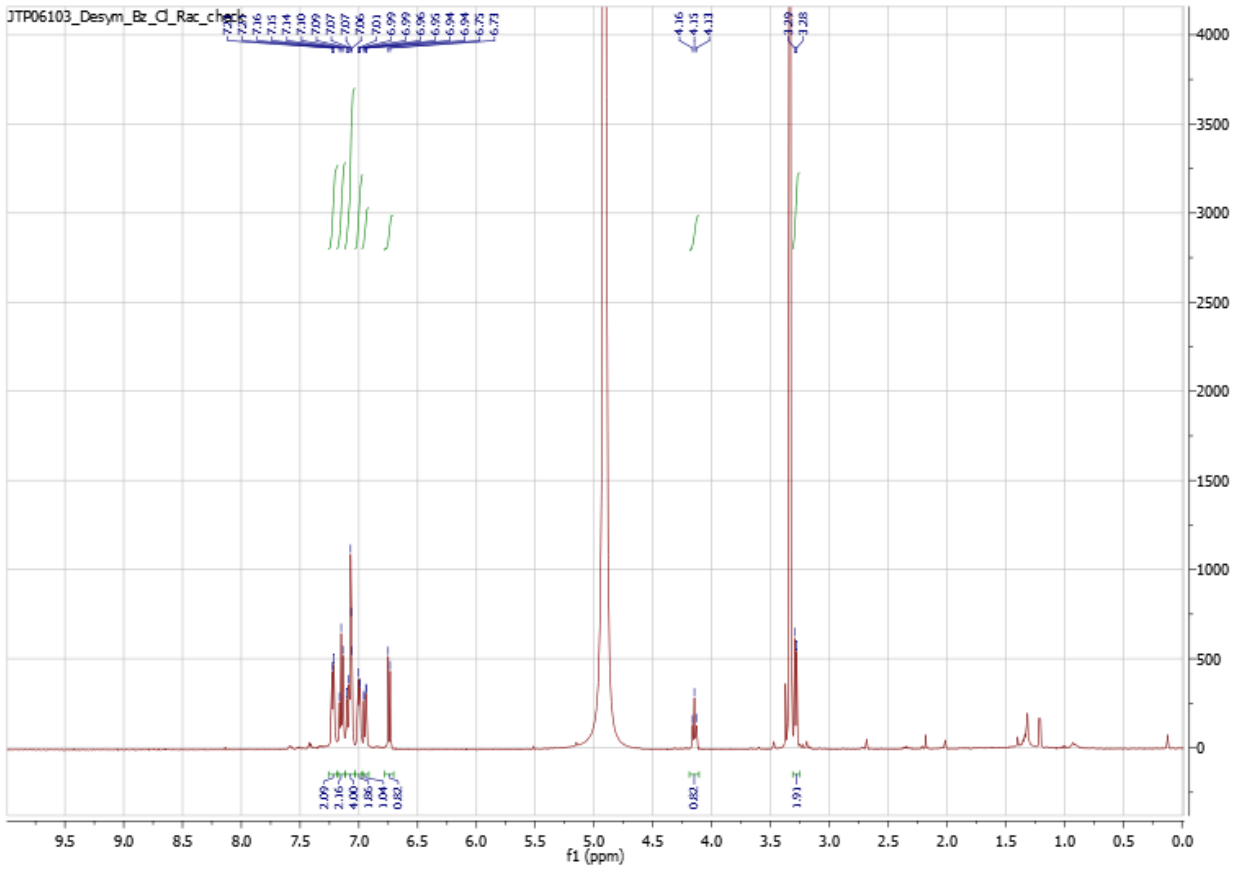
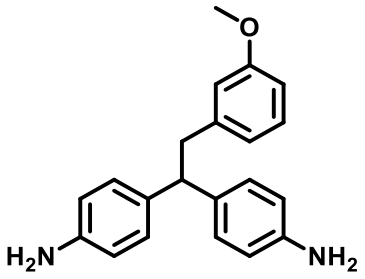


Figure AIII.11. ^{1}H NMR spectrum of 7.



7

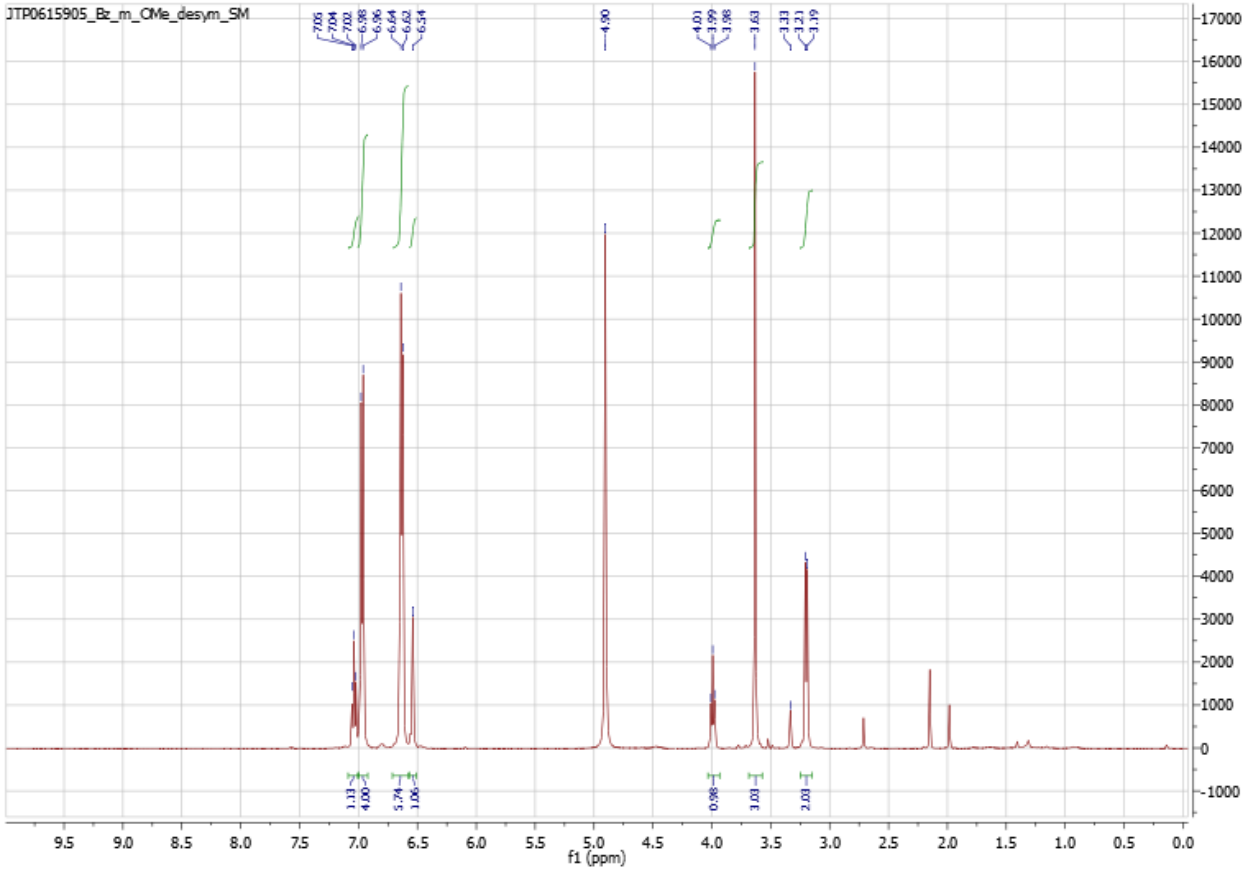
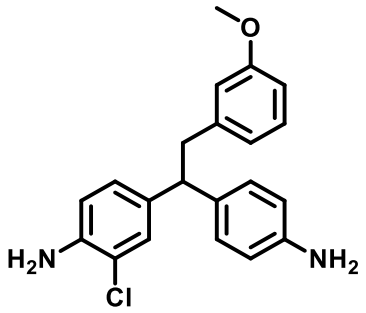


Figure AIII.12. ^{1}H NMR spectrum of 7a.



7a

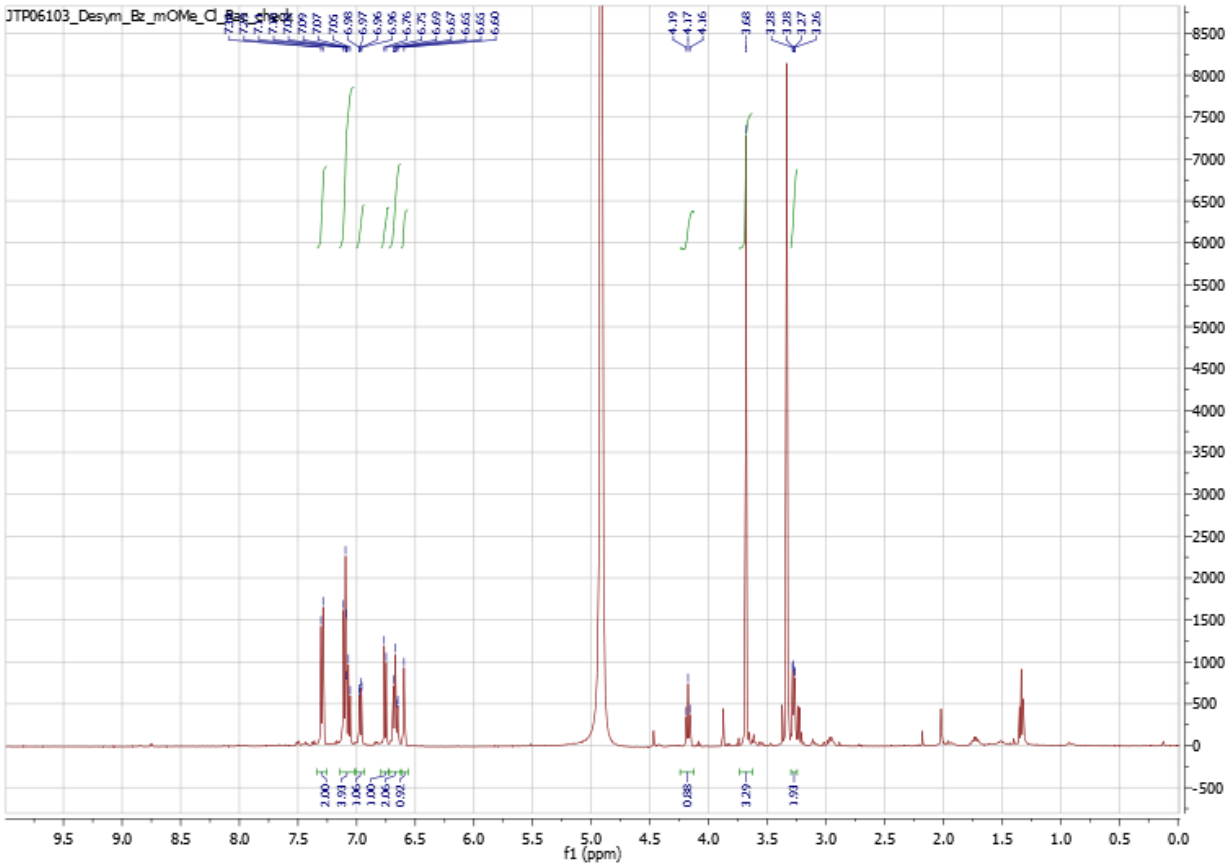
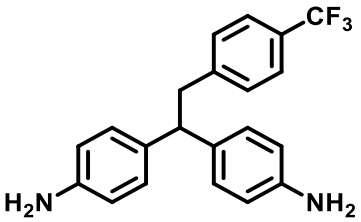


Figure AIII.13. ^{1}H NMR spectrum of 8.



8

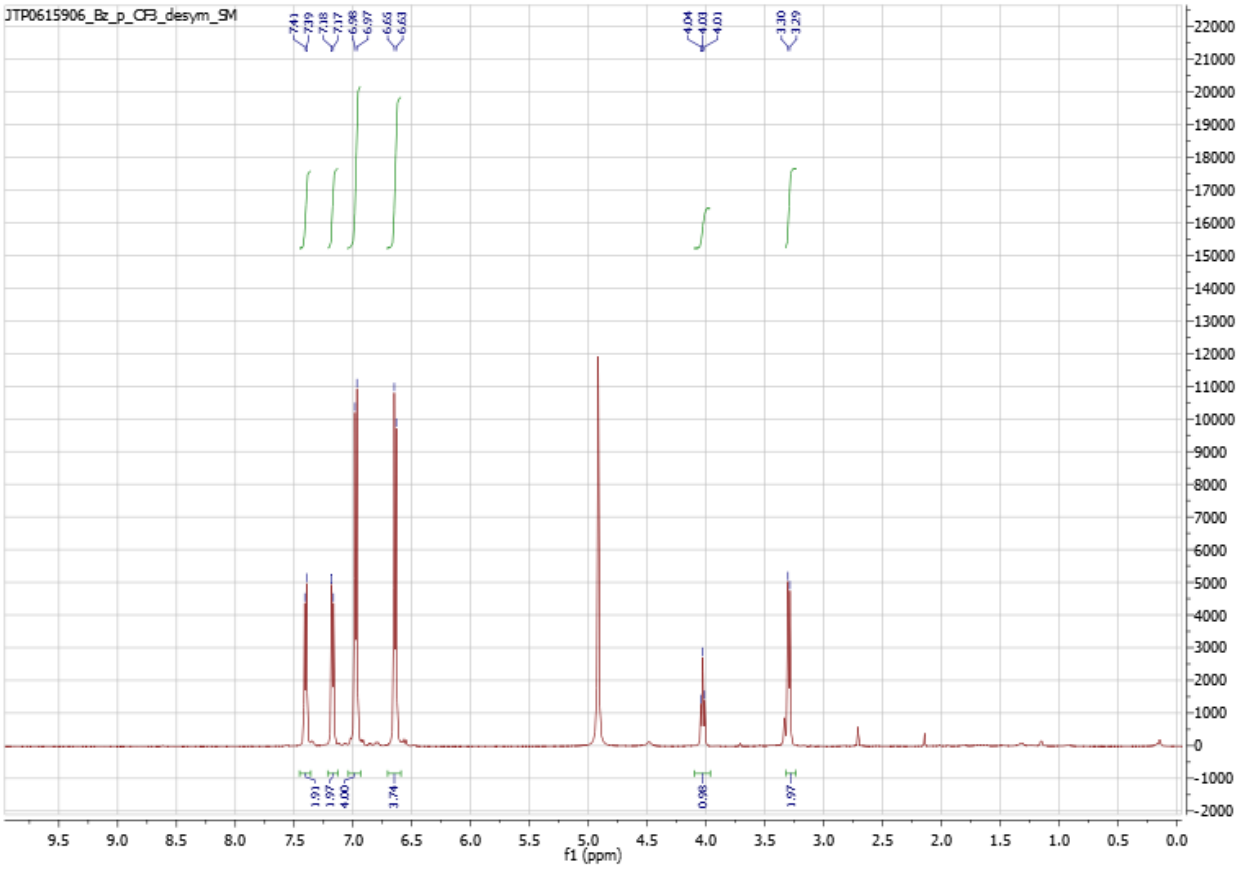
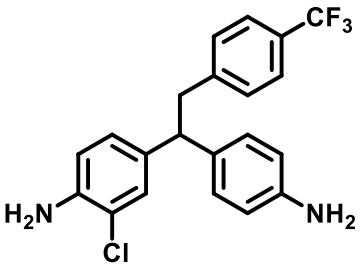


Figure AIII.14. ^{1}H NMR spectrum of 8a.



8a